

PŘIHLÁŠKA VYNÁLEZU

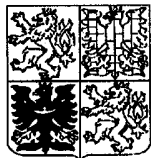
zveřejněná podle § 31 zákona č. 527/1990 Sb.

(21) Číslo dokumentu:

3979-98

(19)

ČESKÁ
REPUBLIKA



ÚŘAD
PRŮMYSLOVÉHO
VLASTNICTVÍ

(22) Přihlášeno: **02. 06. 97**

(32) Datum podání prioritní přihlášky: **06.06.96, 12.08.96**

(31) Číslo prioritní přihlášky: **96/659453, 96/689587**

(33) Země priority: **US, US**

(40) Datum zveřejnění přihlášky vynálezu: **14. 07. 99**
(Věstník č. 7/99)

(86) PCT číslo: **PCT/US97/09472**

(87) PCT číslo zveřejnění: **WO 97/46704**

(13) Druh dokumentu: **A3**

(51) Int. Cl.⁶:

C 12 G 1/68

(71) Přihlášovatel:

**LYNX THERAPEUTICS, INC., Hayward, CA,
US;**

(72) Původce:

**Albrecht Glenn, Redwood City, CA, US;
Brenner Sydney, Cambridge, GB;
Lloyd David H., Daly City, CA, US;
Dubridge Robert B., Belmont, CA, US;
Pallas Michael C., San Bruno, CA, US;**

(74) Zástupce:

**Zelený Pavel JUDr., Hálkova 2, Praha 2,
12000;**

(54) Název přihlášky vynálezu:

**Způsob sekvenční analýzy ligací
kódovaného adaptoru**

(57) Anotace:

Způsob sekvenční analýzy založený na ligaci jednoho či více souborů kódovaných adaptorů na konec cílového polynukleotidu. Kódované adaptory, jejichž vyčnívající vlákna vytvářejí dokonale spárované duplexy s komplementárními vyčnívajícími vlákny cílového polynukleotidu, se připojují a identita nukleotidů ve vyčnívajících vláknech se určuje oligonukleotidovou značkou nesenou kódovaným adaptorem. Toto stanovení či "dekódování" se provádí specifickou hybridizací značeného komplementu značky s jeho odpovídající značkou na navázaném adaptoru.

CZ 3979-98 A3

Způsob sekvenční analýzy ligací kódovaného adaptoru

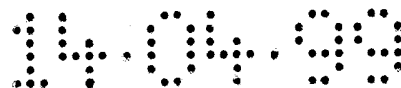
Oblast techniky

Tento vynález se týká obecně způsobů stanovení sekvence nukleotidů v polynukleotidu a konkrétněji způsobu identifikace terminálních nukleotidů polynukleotidu specifickou ligací kódovaných adaptorů.

Dosavadní stav techniky

Způsoby volby pro sekvenční analýzu DNA se téměř ve všech oblastech vědeckého a průmyslového použití zakládají na přístupu terminace dideoxy-řetězce, který navrhnul jako první Sanger, například Sanger a kol., Proc. Natl. Acad. Sci., 74, 5463-5467 (1977). Tento způsob se různě zlepšoval a užívá se v různých formách pro veškeré komerční přístroje pro sekvenční analýzu DNA, například Hunkapiller a kol., Science, 254, 59-67 (1991).

Způsob terminace řetězce vyžaduje vytvoření jednoho nebo více souborů značených fragmentů DNA, při kterém mají všechny společný původ a každý je zakončen známou bází. Soubor nebo soubory fragmentů se potom musí rozdělit podle velikosti pro obdržení informace o sekvenci. Rozdělení podle velikosti se obvykle uskutečňuje gelovou elektroforézou s vysokým rozlišením, která musí být schopna od sebe odlišit velice velké fragmenty, které se svou velikostí neliší o více než o jeden nukleotid. Přes mnohá významná zlepšení, jako jsou separace s kapilárním uspořádáním a použití negelových elektroforetických separačních médií, tento způsob sám o sobě nezajišťuje snadnou minituarizaci či paralelní provádění ve velkém měřítku.



Jako alternativa ke způsobům sekvenování DNA podle Sangera bylo zkoumáno několik způsobů zvaných "báze po bázi" nebo způsoby "jednotlivé báze", například Cheeseman, US patent č. 5 302 509 Tsien a kol., mezinárodní patentová přihláška WO 91/06678, Rosenthal a kol., mezinárodní patentová přihláška WO 93/21340, Canard a kol., Gene, 148, 1-6 (1994) a Metzker a kol., Nucleic Acid Research, 22, 4259-4267 (1994). Tyto způsoby lze charakterizovat stanovením jednotlivé nukleotidu na cyklus chemických či biochemických operací bez požadavku separačního kroku. Proto způsoby "báze po bázi", pokud se mohou uskutečnit podle předpokladu, nabízejí možnost provádění mnoha tisíců reakcí sekvenování paralelně, například na cílových polynukleotidech připojených k mikročásticím nebo na uspořádáních na pevné fázi, například mezinárodní patentová přihláška PCT/US95/12678 (WO 96/12039).

Naneštěstí způsoby sekvenční analýzy "báze po bázi" nenalezly široké uplatnění vzhledem k četným problémům, jako jsou neúčinné chemické postupy, které znemožňují stanovení více než několika nukleotidů v úplné operaci sekvenční analýzy. Navíc při způsobech "báze po bázi", které vyžadují enzymatické manipulace, vznikají problémy s přístrojovým vybavením pro automatizované zpracování. Jestliže se provádí řada enzymatických kroků v reakčních komorách s vysokými poměry povrchu k objemu a při malých rozměrech kanálů, mohou enzymy přilnout k povrchu a tak způsobit problémy při promývání a následných krocích. Akumulace bílkovin též ovlivňuje molekulární reporterové systémy, zejména ty, které užívají fluorescenční značky a tím se stává interpretace měření založených na těchto systémech obtížnou a nevýhodnou. Tyto a podobné obtíže významně zpomalily použití způsobů sekvenč-

ní analýzy "báze po bázi" pro účely paralelního sekvenování.

Důležitý pokrok způsobů sekvenční analýzy "báze po bázi" by se mohl dosáhnout zejména v automatizovaných systémech, pokud by byl k dispozici alternativní způsob stanovení terminálních nukleotidů v polynukleotidech, který by minimalizoval nebo eliminoval opakující se cykly zpracování s použitím více enzymů.

Podstata vynálezu

Předmětem tohoto vynálezu je poskytnutí způsobu sekvenční analýzy DNA, která nemá nevýhody současných způsobů "báze po bázi".

Dalším cílem tohoto vynálezu je poskytnout způsob sekvenční analýzy DNA, který lze použít pro paralelní či současné aplikace na tisíce fragmentů DNA přítomných v jedné reakční nádobě.

Dalším cílem tohoto vynálezu je poskytnout způsob sekvenční analýzy DNA umožňující identifikaci terminální části cílového polynukleotidu s minimálním počtem enzymatických kroků.

Dalším cílem tohoto vynálezu je poskytnout soubor kódovaných adaptorů pro identifikaci sekvence většího počtu terminálních nukleotidů v jednom nebo více cílových polynukleotidech.

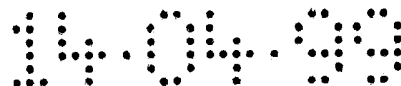
Tento vynález plní tyto záměry poskytnutím způsobu sekvenční analýzy nukleových kyselin na základě ligace jednoho nebo více souborů kódovaných adaptorů na konec cílo-



vého polynukleotidu (nebo na konce více cílových polynukleotidů, pokud jsou použity při způsobu paralelního sekvenování). Každý kódovaný adaptor zahrnuje vyčnívající vlákno a oligonukleotidovou značku vybranou z minimálně zkřížené hybridizujícího souboru oligonukleotidů. Nastává ligace kódovaných adaptorů, jejichž vyčnívající vlákna vytvářejí dokonale spárované duplexy s komplementárními vyčnívajícími vlákny cílového polynukleotidu. Po ligaci se stanoví identita a uspořádání nukleotidů ve vyčnívajících vláknech neboli dojde k dekódování specifickou hybridizací komplementu značené značky s jeho odpovídající značkou na navázaném adaptéru.

Jestliže například kódovaný adaptor s vyčnívajícím vláknem čtyř nukleotidů 5'-AGGT vytváří dokonale spárovaný duplex s komplementárním vyčnívajícím vláknem cílového polynukleotidu a naváže se, lze stanovit čtyři komplementární nukleotidy, 3'-TCCA, na polynukleotidu jedinečnou oligonukleotidovou značkou vybranou ze souboru 256 značek pro každou možnou nukleotidovou sekvenci vyčnívajících vláken. Komplementy značek se aplikují na navázané adaptory za podmínek umožňujících specifickou hybridizaci pouze těch komplementů značek, které vytvářejí dokonale spřažené duplexy (nebo triplexy) s oligonukleotidovými značkami navázaných adaptorů. Komplementy značek se mohou aplikovat individuálně nebo jako jedna či více směsí pro určení identity oligonukleotidových značek a tím i sekvencí vyčnívajících vláken.

Podle dále uvedeného podrobnějšího vysvětlení lze užít kódované adaptory v sekvenční analýze buď (i) pro identifikaci jednoho či více nukleotidů jako kroku způsobu, který opakuje cykly ligace, identifikace a štěpení, jak popisuje Brenner, US patent 5 599 675 a publikace PCT č. WO



95/27080 nebo (ii) jako způsob identifikace "stand alone", při kterém se aplikují soubory kódovaných adaptorů na cílové polynukleotidy takovým způsobem, že každý soubor je schopen identifikovat nukleotidovou sekvenci jiné části cílového polynukleotidu, což v dalším ztělesnění značí, že se sekvenční analýza provádí s jednotlivou ligací pro každý soubor s následnou identifikací.

Důležitou funkcí kódovaných adaptorů je použití oligonukleotidových značek, které jsou členy minimálně zkříženě hybridizujícího souboru oligonukleotidů, jak se například popisuje v mezinárodních patentových přihláškách PCT/US95/12791 (WO 96/12041) a PCT/US96/09513 (WO 96/41011). Sekvence oligonukleotidů v takovém souboru se liší od sekvencí každého dalšího členu téhož souboru alespoň o dva nukleotidy. Proto žádný člen tohoto souboru nemůže vytvářet duplex (či triplex) s komplementem některého jiného členu s méně než dvěma případy chybného spárování. Je vhodné, aby se každý člen minimálně zkříženě hybridizujícího souboru lišil od každého jiného členu o co největší počet nukleotidů konzistentní s rozměrem souboru požadovaným pro dané použití. Jestliže se například užívají delší oligonukleotidové značky, jako jsou 12- až 20-mery pro aplikaci značek na kódované adaptory, potom je vhodné, aby byl rozdíl mezi členy minimálně zkříženě hybridizujícího souboru významně vyšší než dvě. Vhodnější je, aby se každý člen tohoto souboru lišil od každého jiného členu souboru alespoň o čtyři nukleotidy. Ještě vhodnější je, jestliže se každý člen takového souboru liší od každého jiného členu souboru alespoň o šest nukleotidů. Komplementy oligonukleotidových značek podle tohoto vynálezu se zde uvádějí jako "komplementy značek".

Oligonukleotidové značky mohou být jednovláknové a mohou být zvoleny pro specifickou hybridizaci na jednovláknové komplementy značek vytvářením duplexů. Oligonukleotidové značky mohou též být dvojevláknové a mohou se volit pro specifickou hybridizaci na jednovláknové komplementy značek vytvářením triplexů. Je vhodné, aby oligonukleotidové značky kódovaných adaptorů byly dvojevláknové a jejich komplementy značek jednovláknové, takže specifická hybridizace značky s jejími komplementy nastává vytvářením triplexní struktury.

Je vhodné, aby způsob podle tohoto vynálezu zahrnoval dva kroky: (a) ligaci kódovaného adaptoru na konec polynukleotidu s podmínkou, že adaptor má oligonukleotidovou značku zvolenou z minimálně hybridizujícího souboru oligonukleotidů a vyčnívající vlákno komplementární k vyčnívajícímu vláknu polynukleotidu a (b) identifikaci jednoho či více nukleotidů ve vyčnívajícím vláknu polynukleotidu specifickou hybridizací komplementu značky s oligonukleotidovou značkou kódovaného adaptoru.

Přehled obrázků na výkresech

Obrázky 1A až 1E znázorňují schematicky použití kódovaných adaptorů pro stanovení sekvencí terminálních nukleotidů většího počtu značených polynukleotidů.

Obrázek 2 znázorňuje jev selfligace identických polynukleotidů zakotvených na pevnou fázi.

Obrázek 3A znázorňuje kroky preferovaného způsobu vynálezu, při kterém se na cílový polynukleotid navazuje dvojevláknový adaptor mající blokovaný uhlík 3'.



Obrázek 3B znázorňuje použití preferovaného ztělesnění způsoby sekvenční analýzy DNA postupnými cykly ligace a štěpení.

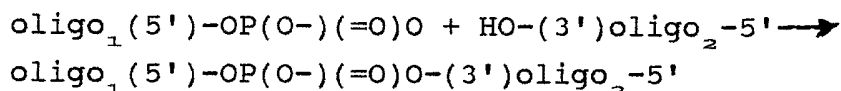
Obrázek 4 znázorňuje údaje ze stanovení terminálních nukleotidů testovaného polynukleotidu s použitím způsobu podle tohoto vynálezu.

Obrázek 5 je schematickým znázorněním průtokové komory a detekčního zařízení pro pozorování dvojrozměrného uspořádání mikročástic s nanesenými molekulami cDNA pro sekvenční analýzu.

Dále se definují použité pojmy.

Pojem "kódovaný adaptor" se zde používá jako synonymum s pojmem "kódovaná sonda" z údajů o prioritní přihlášce US patentu ser. č. 08/689 587.

Pojem "ligace", jak se zde užívá, znamená vytvoření kovalentní vazby mezi konci jednoho nebo více (obvykle dvou) oligonukleotidů. Tento pojem se obvykle vztahuje k vytvoření fosfodiesterové vazby, které je výsledkem následující reakce obvykle katalyzované ligázou:



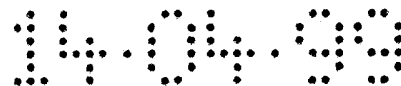
kde oligo_1 a oligo_2 jsou buď dva různé oligonukleotidy nebo různé konce téhož oligonukleotidu. Tento pojem zahrnuje neenzymatické vytvoření fosfodiesterových vazeb stejně tak jako vytvoření jiných než fosfodiesterových kovalentních



vazeb mezi konci oligonukleotidů, jako jsou fosforothioatové vazby, disulfidové vazby a podobně. Reakce ligace je obvykle řízená templátem v tom smyslu, že konce oligo₁ a oligo₂ se přivedou do juxtapozice specifickou hybridizací s vláknem templátu. Zvláštním případem ligace řízené templátem je ligace dvojvláknových oligonukleotidů s komplementárními vyčnívajícími vlákny.

"Komplement" nebo "komplement značky", jak se užívá zde v odkazu na oligonukleotidové značky, znamená oligonukleotid, s kterým se oligonukleotidová značka specificky hybridizuje s vytvořením dokonale spárovaného duplexu či triplexu. Ve ztělesněních, ve kterých vede specifická hybridizace k triplexu, se může oligonukleotidová značka zvolit jako dvojvláknová nebo jednovláknová. Proto je při vytvoření triplexů pojem "komplement" míněn tak, že zahrnuje buď dvojvláknový komplement jednovláknové oligonukleotidové značky nebo jednovláknový komplement dvojvláknové oligonukleotidové značky.

Pojem "oligonukleotid", jak se zde užívá, zahrnuje lineární oligomery přirozených či modifikovaných monomerů nebo spojení, včetně deoxyribonukleosidů, ribonukleosidů, jejich anomerních forem, peptidových nukleových kyselin (PNAs) a podobně, schopné specifické vazby na cílový polynukleotid způsobem pravidelného uspořádání interakcí mezi monomery, jako je Watsonův-Crickův typ párování bází, stacking bází, Hoogsteenovy či reverzní Hoogsteenovy typy párování bází a podobně. Monomery se obvykle spojují fosfodiesterovými vazbami nebo jejich analogy s vytvořením oligonukleotidů o rozměrech několika málo monomerních jednotek, například 3 až 4, až několika desítek monomerních jednotek, například 40 až 60. Při znázornění oligonukleotidu sekvencí



písmen, jako je "ATGCCTG", se rozumí, že tyto nukleotidy jsou v pořadí 5' → 3' zleva doprava a že "A" značí deoxyadenosin, "C" značí deoxycytidin, "G" značí deoxyguanosin a "T" značí thymidin, pokud se neuvádí jinak. Oligonukleotidy podle tohoto vynálezu obvykle zahrnují čtyři přirozené nukleotidy, avšak mohou též zahrnovat jiné než přirozené analogy nukleotidů. Pro toho, kdo má zkušenost v oboru, je zřejmé, kdy lze použít nukleotidy přirozeného nebo jiného původu, například v případě požadavků na enzymatické zpracování se obvykle požadují oligonukleotidy obsahující přirozené nukleotidy.

Pojem "dokonale spárovaný" při odkazu na duplex znamená, že poly- nebo oligonukleotidová vlákna vytvářející duplex poskytují takovou dvojvláknovou strukturu s jinými vlákny, že se každý nukleotid v každém vláknu podrobuje Watsonovu-Crickovu párování bází s nukleotidem v jiném vlákně. Tento termín též zahrnuje párování nukleosidových analogů, jako je deoxyinosin, nukleosidy s 2-aminopurinovými bázemi a podobně, které se mohou použít. Při odkazu na triplex tento termín znamená, že se triplex skládá z dokonale spárovaného duplexu a třetího vlákna, ve kterém se každý nukleotid podrobuje Hoogsteenově nebo reverzní Hoogsteenově asociaci se základním párem dokonale spárovaného duplexu. Opačně "chybné spárování" v duplexu mezi značkou a oligonukleotidem znamená, že se některý pár či trojice nukleotidů v duplexu či triplexu nepodrobil Watsonovu-Crickovu a/nebo Hoogsteenovu a/nebo reverznímu Hoogsteenovu navázání.

Pojem "nukleosid", jak se zde užívá, zahrnuje přirozené nukleosidy včetně forem 2'-deoxy a 2'-hydroxy, jak například popisují Kornberg a Baker, DNA Replication, 2. vydání (Freeman, San Francisco, 1992). Pojem "analogy" v odkazu



na nukleosidy zahrnuje syntetické nukleosidy mající modifikované zbytky bází a/nebo modifikované zbytky cukrů, jak popisuje například Scheit, *Nucleotide Analogs* (John Wiley, New York, 1980), Uhlman a Peyman, *Chemical Reviews*, 90, 543-584 (1990) a podobně, s jedinou podmínkou, že jsou schopny specifické hybridizace. Tyto analogy zahrnují syntetické nukleosidy určené pro zvyšování vazebných schopností, snižování komplexity, zvyšování specifičnosti a podobně.

Pojem "stanovení sekvence" nebo "stanovení sekvence nukleotidů", jak se zde užívá v odkazu na polynukleotidy, zahrnuje stanovení částečné či úplné informace o sekvenci polynukleotidů. To znamená, že tento termín zajišťuje sekvenční srovnávání, fingerprinting a podobné hladiny informací o cílovém polynukleotidu, stejně tak jako výslovnou identifikaci a uspořádání nukleosidů, obvykle každého nukleosidu v cílovém polynukleotidu. Tento termín též zahrnuje stanovení identifikace, uspořádání a umístění jednoho, dvou nebo tří ze čtyř typů nukleotidů v rámci cílového polynukleotidu. V některých ztělesněních se může například stanovení sekvence provést identifikací uspořádání a umístění jednoho typu nukleotidu, například cytosinů, v cílovém polynukleotidu "CATCGA...", takže jeho sekvenci lze znázornit binárním kódem, například "100101..." pro "C-(nikoliv C)-(nikoliv C)-C-(nikoliv C)-C..." a podobně.

Termín "komplexita", jak se zde užívá v odkazu na skupinu polynukleotidů, znamená počet různých druhů molekuly přítomných v této skupině.

Dále se uvádí podrobný popis tohoto vynálezu.

Tento vynález zahrnuje ligaci adaptorů specificky



hybridizovaných s koncem nebo konci jednoho nebo více cílových polynukleotidů. Sekvenční informace o oblasti, kde nastává specifická hybridizace, se obdrží "dekódováním" oligonukleotidových značek takto navázaných kódovaných adaptorů. V jednom aspektu tohoto vynálezu jsou vícečetné soubory kódovaných adaptorů navázány na cílový polynukleotid ve střídavě uspořádaných místech štěpení, takže kódované adaptory zajišťují sekvenční informaci pro každou z většího množství částí cílového polynukleotidu. Takové části mohou být rozpojené, překrývající se nebo styčné, avšak přednostně jsou tyto části styčné a společně umožňují identifikaci sekvence nukleotidů rovné součtu délek jednotlivých částí. V tomto aspektu existuje pouze jednotlivá ligace kódovaných adaptorů následovaná identifikací "dekódováním" značek navázaných adaptorů. V dalším aspektu tohoto vynálezu se používají kódované adaptory jako identifikační krok ve způsobu zahrnujícím opakované cykly ligace, identifikace a štěpení, jak se popisuje podrobněji níže.

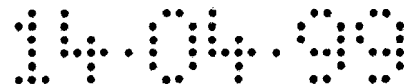
V tomto druhém ztělesnění využívá tento vynález nukleáz, jejichž rozpoznávací místa jsou oddělená od jejich míst štěpení. Takovými preferovanými nukleázami jsou restriční endonukleázy typu II. Tyto nukleázy se využívají pro vytváření vyčnívajících vláken cílových polynukleotidů, na která jsou navázány kódované adaptory. Množství sekvenčních informací obdržených v daném ztělesnění vynálezu závisí částečně na tom, kolik takových nukleáz se použije, a na délce vytvořeného vyčnívajícího vlákna při štěpení.

Důležitým aspektem tohoto vynálezu je schopnost paralelní sekvenční analýzy mnoha cílových nukleotidů. V tomto aspektu zahrnuje způsob podle tohoto vynálezu následující kroky: (a) takové připojení první oligonukleotidové značky



z repertoáru značek ke každému polynukleotidu v populaci polynukleotidů, že každá první oligonukleotidová značka z repertoáru se volí z prvního minimálně zkříženě hybridizujícího souboru, (b) takové odebrání vzorku populace polynukleotidů, že v podstatě všechny rozdílné polynukleotidy v populaci mají připojené rozdílné první oligonukleotidové značky, (c) navázání jednoho či více kódovaných adaptorů ke konci každého z polynukleotidů v populaci, kde každý kódovaný adaptor má druhou oligonukleotidovou značku zvolenou z druhého minimálně zkříženě hybridizujícího souboru a vyčnívající vlákno komplementární k vyčnívajícímu vláknu polynukleotidu populace, (d) třídění polynukleotidů populace specifickou hybridizací prvních oligonukleotidových značek s příslušnými komplementy, kde tyto příslušné komponenty jsou připojeny jako jednotné populace v podstatě identických oligonukleotidů v prostorově diskrétních oblastech na jednom nebo více pevných nosičích a (e) identifikace jednoho nebo více nukleotidů v těchto vyčnívajících vláknech polynukleotidů specifickou hybridizací komplementu značky s každou druhou oligonukleotidovou značkou jednoho či více kódovaných adaptorů. V tomto ztělesnění může být alespoň jeden kódovaný adaptor připojen na konec polynukleotidů buď před tříděním nebo po třídění polynukleotidů na pevných nosičích prvními nukleotidovými značkami. V preferovaném ztělesnění zahrnují kódované adaptory restriční místo endonukleázy typu II, což umožňuje odštěpení kódovaných adaptorů z polynukleotidů a zkrácení polynukleotidů po identifikaci sekvence.

V souladu s tímto preferovaným ztělesněním tento způsob dále zahrnuje opakované cykly ligace, identifikace a štěpení, takže v každém cyklu identifikuje alespoň jeden nukleotid. Přednostně se identifikují 2 až 6 nukleotidů a v každém cyklu se stanoví jejich pořadí.



Sekvenční analýza bez cyklů ligace a štěpení

Sekvenční analýza bez cyklů ligace a štěpení v souladu s tímto vynálezem se ilustruje ztělesněním podle obrázků 1A až 1E. V tomto ztělesnění se připraví k cílových nukleotidů, jak se popisuje níže a jak též popisuje Brenner, mezinárodní patentová přihláška PCT/US95/12791 (WO 96/12041) a PCT/US96/09513 (WO 96/41011). Vzorek se tedy odebírá z populace polynukleotidů konjugovaných s oligonukleotidovými značkami označenými malými písmeny "t". Tyto značky se někdy nazývají polynukleotidové značky pro třídění nebo "první" oligonukleotidové značky. Konjugáty značka-polynukleotid vzorku se amplifikují například polymerázovou řetězovou reakcí (PCR) nebo klonováním s obdržáním 1 až k populací konjugátů označených (14)-(18) na obr. 1A. Přednostně se konce konjugátů proti značkám "t" připraví pro ligaci jednoho či více adaptorů, kde každý obsahuje rozpoznávací místo pro nukleázu, jejíž místo štěpení je oddělené od místa rozpoznávání. Ve znázorněném ztělesnění se používají tyto tři adaptory, které se zde udávají jako "adaptory štěpení". Počet těchto použitých adaptorů závisí na několika faktorech včetně množství požadovaných sekvenčních informací, dostupnosti nukleáz typu II s vhodnými charakteristikami dosahu a štěpení a podobně. Přednostně se používá jeden až tři adaptory štěpení a adaptory jsou vybrány tak, aby se přizpůsobily různým nukleázám typu II schopným vytváření vyčnívajících vláken o nejméně čtyřech nukleotidech po štěpení.

Jestliže se metoda podle vynálezu používá pro charakterizaci sekvenování populace cDNA, potom mohou být konjugá-



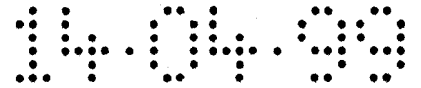
ty značka-polynukleotid před ligací na adaptory štěpení štěpeny restriční endonukleázou s vysokým výskytem rozpoznávacích míst, jako je Taq I, Alu I, HinP1 I, Dpn II, Nla III a podobně. Pro enzymy, jako je Alu I, které odstupují z tuppých konců, se může střídavě uspořádaný konec vytvořit T4 DNA polymerázou, jak například popisuje Brenner, mezinárodní patentová přihláška PCT/US95/12791 citovaná výše a Kuijper a kol., Gene, 112, 147-155 (1992). Jestliže se cílové polynukleotidy připraví štěpením s Taq I, potom jsou k dispozici pro ligaci následující konce:

cgannnn ... -3'
tnnnn ... -5'

Exemplární soubor tří adaptorů štěpení se tedy může vytvořit následujícím způsobem:

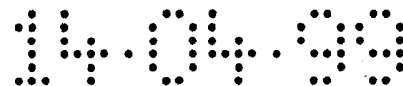
- (1) NN ... NGAAGA cgannnnnnnnnnnnnnnnnnnnnn ... -3'
NN ... NCTTCTGCp tnnnnnnnnnnnnnnnnnnnnnn ... -5'
=====
- (2) NN ... NGCAGCA cgannnnnnnnnnnnnnnnnnnnnn ... -3'
NN ... NCGTTCGTGCp tnnnnnnnnnnnnnnnnnnnnnn ... -5'
=====
- (3) NN ... NGGGA cgannnnnnnnnnnnnnnnnnnnnn ... -3'
NN ... NCCCTGCp tnnnnnnnnnnnnnnnnnnnnnn ... -5'
=====

kde se adaptory štěpení (1), (2) a (3) označují velkými písmeny s příslušnými rozpoznávacími místy nukleáz Bbs I, Bbv I a Bsm FI podtrženými a 5' fosfátem označeným jako "p". Dvakrát podtržené části cílového polynukleotidu označují polohy vyčnívajících vláken po ligaci a štěpení. Ve všech případech je cílový polynukleotid ponechán s vyčnívajícím vlák-



nem 5' ze čtyř nukleotidů. Mnohá ztělesnění se mohou zřejmě vytvořit s použitím různých počtů a typů nukleáz. Jak diskutuje Brenner, US patent 5 599 675 a WO 95/27080, je vhodné, aby před štěpením byla vnitřní místa Bbs I, Bbv I a Bsm FI blokována, například methylací, aby se zabránilo nežádoucímu štěpení na vnitřních místech cílového polynukleotidu.

Při návratu zpět ke znázorněnému ztělesnění lze pozorovat, že se adaptory štěpení A_1 , A_2 a A_3 navazují (20) v koncentračním poměru 1:1:1 k cílovým polynukleotidům s poskytnutím konjugátů znázorněných na obr. 1B tak, že v každé populaci konjugátů značka-polynukleotid je zhruba stejný počet konjugátů majících připojené A_1 , A_2 a A_3 . Po ligaci (20) se polynukleotidy postupně štěpí každou z nukleáz adaptoru štěpení a navazují na soubor kódovaných adaptorů. Nejprve se cílové polynukleotidy štěpí (22) nukleázou adaptoru štěpení A_1 a poté se první soubor kódovaných adaptorů navazuje na výsledná vyčnívající vlákna. Štěpení vede zhruba k jedné třetině polynukleotidů každého typu, to jest t_1 , t_2 , ..., t_k schopných ligace. Preferuje se použití kódovaných adaptorů jako jedné nebo více směsí adaptorů, které společně obsahují každou možnou sekvenci vyčnívajícího vlákna. Reakční podmínky se volí takovým způsobem, že se navazují pouze kódované adaptory, jejichž vyčnívající vlákna vytvářejí dokonale spárované duplexy s vlákny cílového polynukleotidu s vytvořením kódovaných konjugátů (28), (30) a (32) (obr. 1C). Velká písmena "T" s indexy dole označují, že kódované adaptory pro značení nesou jedinečné oligonukleotidové značky. Tyto oligonukleotidové značky nesené kódovanými adaptory se někdy nazývají značky pro zajištění označení kódovaných adaptorů nebo jako "druhé" oligonukleotidové značky. Jak se popisuje podrobněji níže, jednovláknové oligonukleotidové značky užívané pro třídění přednostně obsahu-



jí pouze tři ze čtyř nukleotidů, takže lze použít stripping-
govou reakci T4 DNA polymerázy, například Kuijper a kol.
(citováno výše), pro přípravu cílových polynukleotidů pro
nanesení na pevný nosič. Na druhé straně mohou oligonukleo-
tidové značky užívané pro zajištění označení obsahovat
všechny čtyři nukleotidy.

Jak se uvádí výše, zahrnují kódované adaptory vyční-
vající vlákno (24) a oligonukleotidovou značku (26). Proto,
jestliže štěpení "A₁" t₁-polynukleotidových konjugátů vede
k následujícím koncům:

5'- ... nnnnnnnnn
3'- ... nnnnnnnnnacct

potom by oligonukleotidová značka T₂₄ mohla mít následující
strukturu (číslo identifikace sekvence: 1):

tggattctagagagagagagagagagagagagag -3'
aagatctctctctctctctctctctctctctc

kde může být dvojitá část jedním ze souboru 48 (=12 po-
loh nukleotidu x 4 typy nukleotidu) dvojitých
20-merních oligonukleotidových značek, který vytváří dokona-
le spárovaný triplex s jedinečným komplementem značky a vy-
tváří triplex nejméně s šesti nesprávnými spárováními se
všemi ostatními komplementy značek. Kódované adaptory v tom-
to příkladu se mohou navazovat na cílové polynukleotidy
v jedné směsi či ve více směsích o celkovém počtu 768 (3
x 256) členů. Kódovaný adaptor může případně též zahrnovat
oblast vymezující vzdálenost, jak ukazuje příklad popsany
výše, kde sekvence 4 nukleotidů "ttct" slouží jako vymezova-
cí oblast mezi vyčnívajícím vláknem a oligonukleotidovou



značkou.

Po ligaci prvního souboru kódovaných adaptorů (28), (30) a (32) se polynukleotidové konjugáty štěpí (34) nukleázou adaptoru štěpení A_2 a poté se aplikuje druhý soubor kódovaných adaptorů pro vytvoření konjugátů (36), (38) a (40) (obr. 1D). Konečně se štěpí konjugáty značka-polynukleotid (42) nukleázou adaptoru štěpení A_3 a poté se aplikuje třetí soubor kódovaných adaptorů pro vytvoření konjugátů (44), (46) a (48) (obr. 1E). Po dokončení sledu štěpení a ligací kódovaných adaptorů se směs nanese (50) na jeden nebo více pevných nosičů prostřednictvím oligonukleotidových značek t_1 až t_k , jak se popisuje podrobněji níže, a jak uvádí Brenner, například PCT/US95/12791 nebo PCT/US96/09513. Jestliže se analyzuje jednotlivý cílový polynukleotid, potom zřejmě není třeba užívat více oligonukleotidových značek t_1, t_2, \dots, t_k . V takovém ztělesnění lze použít biotin nebo podobný zbytek pro zakotvení konjugátu adaptoru kódovaného polynukleotidem a není třeba žádné třídění. Také uspořádání kroků štěpení, ligace a nanesení na pevný nosič závisí na daném prováděném ztělesnění. Například se mohou konjugáty značka-polynukleotid nejprve nanést na pevný nosič s následnou ligací adaptoru štěpení, štěpením a ligací kódovaných adaptorů nebo se mohou napřed navázat adaptory štěpení s následným nanesením štěpením a ligací kódovaných adaptorů atd.

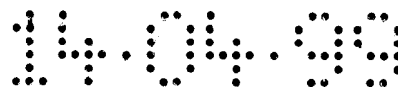
Po navázání kódovaných adaptorů na konce cílového polynukleotidu v souladu s tímto vynálezem se obdrží sekvenční informace postupnou aplikací komplementů označené značky pro immobilizované cílové polynukleotidy, buď jednotlivě nebo ve směsích za podmínek umožňujících vytvoření dokonale spárovaných duplexů a/nebo triplexů mezi nukleotidovými značkami



kódovaných adaptorů a jejich příslušnými komplementy značky. Počty a komplexita směsí závisí na několika faktorech včetně typu systému užitého pro označení, délky částí, jejichž sekvence mají být identifikovány, na tom, zda se použijí analogy snižující komplexitu a podobně. Pro ztělesnění znázorněné v obrázcích 1A až 1E se preferuje použití jednotlivého fluorescenčního barviva pro označení každého z 48 (=3 x 16) komplementů značek. Komplementy značek se aplikují individuálně pro identifikaci nukleotidů každé čtyřnukleotidové části cílového polynukleotidu (to jest 4 komplementy značek na každých 12 poloh pro celkové množství 48). Části o různých délkách budou zřejmě vyžadovat různé počty komplementů značek, například v souladu s tímto ztělesněním bude pětinnukleotidová část vyžadovat 20 komplementů značek, dvojnukleotidová část bude vyžadovat 8 komplementů značek, atd. Komplementy značek se aplikují za podmínek dostatečně přísných proto, aby se vytvořily pouze dokonale spárované duplexy, měří se signály z fluorescenčních označení na specificky hybridizovaných komplementech značek a komplementy značek se vymyjí z kódovaných značek před aplikací další směsí. 16 komplementů značek odpovídá jeden po druhém následujícím sekvencím 4-merních částí cílové sekvence:

| | | | |
|------|------|------|------|
| ANNN | NANN | NNAN | NNNA |
| CNNN | NCNN | NNCN | NNNC |
| GNNN | NGNN | NNGN | NNNG |
| TNNN | NTNN | NNTN | NNNT |

kde "N" je kterýkoliv z nukleotidů A, C, G nebo T. Proto se činí čtyři oddělené dotazy pro každou nukleotidovou polohu, jeden pro každý typ nukleotidu. Toto ztělesnění zahrnuje značný stupeň nadbytečnosti (používá se celkem 16 komplementů značek pro identifikaci 4 nukleotidů) pro získání zvýšené



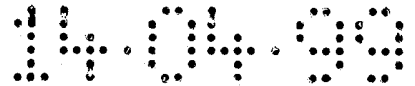
spolehlivosti stanovení nukleotidu.

Alternativně lze aplikovat 12 směsí 4 komplementů značek po sobě s použitím 4 spektrálně rozlišitelných fluorescenčních barviv takových, že mezi barvivy a typy nukleotidu existuje vzájemná příslušnost jednoho barviva vůči jednomu nukleotidu. Například směs 4 komplementů značek může identifikovat nukleotid "x" v sekvenci vyčnívajícího vlákna "nnxn", takže první fluorescenční značka se pozoruje, jestliže x=A, druhá fluorescenční značka se pozoruje, jestliže x=C a třetí fluorescenční značka se pozoruje, jestliže x=G, atd.

Další sekvenční informace lze obdržet s použitím výše popsaného ztělesnění ve způsobu analogickém způsobu "multi-stepping", který popisuje Brenner, mezinárodní patentová přihláška PCT/US95/03678 (WO 95/27080). V tomto ztělesnění čtvrtý adaptor, který se zde uvádí jako "stepping adaptor" se naváže na konce cílových polynukleotidů spolu s adaptory štěpení A_1 , A_2 a A_3 v koncentračním poměru 3:1:1:1. Tím se zhruba polovina přístupných konců naváže na "stepping adaptor". Tento stepping adaptor zahrnuje rozpoznávací místo pro nukleázu typu II umístěné takovým způsobem, že jeho dosah (definovaný níže) umožní štěpení cílových polynukleotidů na konci stanovované sekvence prostřednictvím adaptorů A_1 , A_2 a A_3 . Následuje příklad stepping adaptoru, který se může použít s výše popsaným souborem adaptorů štěpení:

```
NN ... NCTGGAGA      cgannnnnnnnnnnnnnnnnnnnnn ... -3'  
NN ... NGACCTCTGCp   tnnnnnnnnnnnnnnnnnnnnnn ... -5'  
==
```

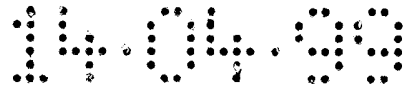
kde, podobně jako výše, rozpoznávací místo nukleázy, v tomto



případě BpM I, je podtrženo jednou a nukleotidy v místě štěpení jsou podtrženy dvakrát. Cílové polynukleotidy štěpené nukleázou stepping adaptoru mohou být navázány na další soubor adaptorů štěpení A₄, A₅ a A₆, které mohou obsahovat rozpoznávací místa nukleázy, která jsou stejná nebo rozdílná od těch, která jsou obsažena v adaptorech štěpení A₁, A₂ a A₃. To, zda se požaduje zvětšený soubor kódovaných adaptorů, závisí na tom, zda lze v signálním měřicím zařízení tolerovat reakce štěpení a ligace. Jestliže je, podobně jako výše, požadavek na minimalizaci enzymatických reakcí ve spojení s měřením signálu, potom je třeba použít přídatné soubory kódovaných adaptorů. To znamená, že tam, kde se vyžaduje více než 768 oligonukleotidových značek a komplementů značek se šesti reakcemi štěpení produkujícími vyčnívající vlákna, každé o čtyřech nukleotidech, požaduje se 1536 oligonukleotidových značek a komplementů značek (24 směsí, každá po 64 komplementech značek). Například adaptory štěpení A₄, A₅ a A₆ se stejnými nukleázovými rozpoznávacími místy jako A₁, A₂ a A₃, které by se mohly být použít pro stepping adaptor ukázaný výše jsou následující:

- (4) NN ... NGAAGACNN nnnnnnnnnnnnnnnnnnnnn ... -3'
- NN ... NCTTCTGp nnnnnnnnnnnnnnnnnnnnn ... -5'
- =====
- (5) NN ... NGCAGCACNN nnnnnnnnnnnnnnnnnnnnn ... -3'
- NN ... NCGTCGTGp nnnnnnnnnnnnnnnnnnnnn ... -5'
- =====
- (6) NN ... NGGGACNN nnnnnnnnnnnnnnnnnnnnn ... -3'
- NN ... NCCCTGp nnnnnnnnnnnnnnnnnnnnn ... -5'
- =====

kde jsou místa štěpení označena dvojitým podtržením. Adaptory štěpení A₄, A₅ a A₆ se přednostně aplikují jako směsi, takže je reprezentováno každé dvojnukleotidové vyčnívající vlákno.

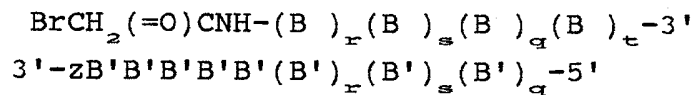


Když se kódované adaptory navázaly, připraví se cílové polynukleotidy pro nanesení na pevné nosiče, přednostně mikročástice, jak uvádí Brenner, mezinárodní patentová přihláška PCT/US95/12791 (WO 96/12041). Stručně řečeno jsou oligonukleotidové značky pro třídění učiněny jednovláknovými s použitím "stripping" reakce s T4 DNA polymerázou, například Kuijper a kol. (citováno výše). Jednovláknové oligonukleotidové značky se specificky hybridizují a navazují na své komplementy značek na mikročásticích. Mikročástice s navázanými značkami se poté analyzují v přístroji, jak popisuje Brenner (citováno výše), což umožňuje sekvenční dodávku, specifickou hybridizaci a odstranění označených komplementů značek.

Ve ztělesněních, ve kterých se kódované adaptory navazují na cílový polynukleotid (nebo populaci cílových polynukleotidů) pouze jednou, existuje několik neenzymatických způsobů ligace řízených templáty, které lze použít v souladu s tímto vynálezem. Takové způsoby ligace zahrnují, avšak bez omezení na tyto případy, ty, které uveřejnili Shabarova, *Biochimie*, 70, 1323-1334 (1988), Dolinnaya a kol., *Nucleic Acids Research*, 16, 3721-3738 (1988), Letsinger a kol., US patent 5 476 930, Gryaznov a kol., *Nucleic Acids Research*, 22, 2366-2369 (1994), Kang a kol., *Nucleic Acids Research*, 23, 2344-2345 (1995), Gryaznov a kol., *Nucleic Acids Research*, 21, 1403-1408 (1993), Gryaznov, US patent 5 571 677 a podobně. Neenzymatická ligace se přednostně provádí způsobem Letsingera a kol. (citováno výše). V tomto způsobu reaguje kódovaný adaptor mající 3'-bromacetylovaný konec s polynukleotidem mající komplementární vyčnívající vlákno a thiofosforylovou skupinu na svém konci 5'. Příklad kódovaného adaptoru používajícího tento chemický postup má násle-



dující strukturu:



kde B a B' jsou nukleotidy a jejich komponenty a z, r, s, q a t se popisují níže. Symboly Br, C, H a N mají svůj běžný chemický význam. Jak se vysvětluje v citacích výše v reakci řízené templáty, reaguje 3'-bromacetylovaný oligonukleotid spontánně s oligonukleotidem majícím 5'-thiofosforylovou skupinu ve vodném prostředí s vytvořením thiofosforylacylaminového spojení. Thiofosforylová skupina se snadno připojuje k 5'-hydroxylové skupině cílového polynukleotidu zpracováním T4 kinázou za přítomnosti adenosin-5'-O-(1-thiotrifosfatu), to jest g-S-ATP, jak popisuje Kang a kol. (citováno výše).

Sekvenční analýza s cykly ligace a štěpení

Kódované adaptory lze použít při způsobu sekvenční analýzy DNA založeném na adaptorech zahrnujícím opakované cykly ligace, identifikace a štěpení, jako je způsob, který popisuje Brenner, US patent 5 599 675 a publikace PCT č. WO 95/27080. Stručně řečeno obsahuje tento způsob následující kroky: (a) navázání kódovaného adaptoru na konec polynukleotidu, kde kódovaný adaptor má mít nukleázové rozpoznávací místo nukleázy, jejíž místo štěpení je oddělené od tohoto rozpoznávacího místa, (b) identifikace jednoho či více nukleotidů na konci polynukleotidů podle identity kódovaného adaptoru, který je k němu navázán, (c) štěpení polynukleotidů s nukleázou rozpoznávající nukleázové rozpoznávací místo nukleázy kódovaného adaptoru, takže se polynukleotid zkrátí o 1 nebo více nukleotidů a (d) opakování uvedených kroků (a)



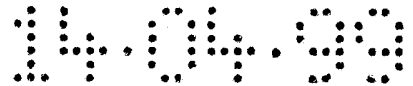
až (c) až po stanovení nukleotidové sekvence polynukleotidu. Při identifikačním kroku se následné soubory komplementů značek hybridizují s příslušnými značkami přenášenými kódovanými adaptory navázanými na konci cílových polynukleotidů, jak se popisuje výše. Typ a sekvence nukleotidů ve vyčnívajících vláknech polynukleotidů se identifikují označením přenášeným specificky hybridizovaným komplementem značky a souborem, ze kterého daný komplement značky pochází, jak se popisuje výše.

Oligonukleotidové značky a komplementy značek

Oligonukleotidové značky se používají pro různé účely v preferovaných ztělesněních tohoto vynálezu. Oligonukleotidové značky se používají tak, jak popisuje Brenner, mezinárodní patentová přihláška PCT/US95/12791 a PCT/US96/09513 (WO 96/12041 a WO 96/41011) pro třídění velkých počtů polynukleotidů, například několika tisíc až několika set tisíc ze směsi do jednotných populací identických polynukleotidů pro analýzu a dále se používají pro dodání označení kódovaným adaptorům, jejichž počet je v rozmezí několika málo desítek až několika tisíc. Pro první použití se typicky požadují velké počty či repertoáry značek, a proto je syntéza jednotlivých oligonukleotidových značek problematická. V těchto ztělesněních se preferuje syntéza značek. Na druhé straně tam, kde se nevyžadují mimořádně velké repertoáry značek - jako při dodání označení kódovaným adaptorům - lze odděleně syntetizovat oligonukleotidové značky minimálně zkříženě hybridizujícího souboru a mohou se též syntetizovat kombinatoricky.

Jak popisuje Brenner (citováno výše), nukleotidové sekvence oligonukleotidů minimálně zkříženě hybridizujícího

souboru se výhodně vyhodnotí jednoduchými počítačovými programy, jako jsou příklady programů, jejichž zdrojové kódy jsou v dodatcích I a II. Podobné počítačové programy se snadno napíší pro výčet oligonukleotidů minimálně zkříženě hybridizujících souborů pro kterékoliv ztělesnění tohoto vý-
nálezu. Tabulka I níže poskytuje vodítko, pokud jde o veli-
kost souborů minimálně zkříženě hybridizujících oligonukleo-
tidů pro uvedené délky a hodnoty rozdílů mezi nukleotidy.
Výše popsané počítačové programy byly použity pro generování
hodnot.



Tabulka I

Minimálně zkříženě hybridizující soubory slov obsahujících čtyři nukleotidy

| Délka slova oligonukleo- tidů | Nukleotido- vý rozdíl mezi oligo- nukleotidy minimálně zkříženě hybridizu- jícího souboru | Maximální velikost minimálně zkříženě hybridizu- jícího souboru | Velikost repertoáru se třemi slovy | Velikost repertoáru se čtyřmi slovy |
|-------------------------------------|---|---|---|--|
|-------------------------------------|---|---|---|--|

| | | | | |
|----|---|-----|--------------------|-------------------|
| 4 | 3 | 11 | 1331 | 14 641 |
| 6 | 4 | 25 | 15 625 | $3,9 \times 10^5$ |
| 6 | 5 | 4 | 64 | 256 |
| 8 | 4 | 225 | $1,14 \times 10^7$ | |
| 8 | 5 | 56 | $1,75 \times 10^5$ | |
| 8 | 6 | 17 | 4913 | |
| 12 | 8 | 62 | | |

Soubory obsahující několik sto až několik tisíc nebo dokonce několik desítek tisíc oligonukleotidů se mohou syntetizovat přímo řadou paralelních syntetických způsobů, jak například popisuje Frank a kol., US patent 4 689 405, Frank a kol., Nucleic Acids Research, 11, 4365-4377 (1983), Matson a kol., Anal. Biochem., 224, 110-116 (1995), Fodor a kol., mezinárodní patentová přihláška PCT/US93/04145 (WO 93/22684), Pease a kol., Proc. Natl. Acad. Sci., 91, 5022-5026 (1994), Southern a kol., J. Biotechnology, 35,



217-227 (1994), Brennan, mezinárodní patentová přihláška PCT/US94/05896 (WO 94/27719), Lashkari a kol., Proc. Natl. Acad. Sci., 92, 7912-7915 (1995) a podobně.

Komplementy značek, ať syntetizované kombinatoricky či individuálně, se volí tak, aby měly podobné duplexní či triplexní stability jedna vůči druhé, takže dokonale spárované hybridy mají podobné nebo v podstatě identické teploty tání. To umožňuje snadnější odlišení nesprávně spárovaných komplementů značek od dokonale spárovaných komplementů značek při aplikaci na kódované adaptory, například promytím za přísných podmínek. Pro kombinatoricky syntetizované komplementy značek lze vytvářet minimálně zkříženě hybridizující soubory z podjednotek, které přinášejí zhruba ekvivalentní příspěvky k duplexní stabilitě jako každá jiná podjednotka ze souboru. Vodítka pro provádění těchto selekcí poskytují uveřejněné způsoby selekce optimálních trimerů PCR a pro výpočty duplexních stabilit, například Rychlik a kol., Nucleic Acids Research, 17, 8543-8551 (1989) a 18, 6409-6412 (1990), Breslauer a kol., Proc. Natl. Acad. Sci., 83, 3746-3750 (1986), Wetmur, Crit. Rev. Biochem. Mol. Biol., 26, 227-259 (1991) a podobně. Když se požadují menší počty oligonukleotidových značek, jako pro dodávku označení kódovaným adaptorům, lze použít počítačové programy dodatků I a II pro vytvoření a vypsání seznamu sekvencí minimálně zkříženě hybridizujících souborů oligonukleotidů, které se používají přímo (to jest bez vzájemného zapojení do "vět"). Takové seznamy se mohou podrobit dalšímu skreeningu ohledně přídatných kritérií, jako je obsah GC, distribuce nesprávných spárování, teoretická teplota tání a podobně pro vytvoření přídatných minimálně zkříženě hybridizujících souborů.

Pro kratší značky, například okolo 30 nukleotidů nebo



méně, se pro výpočet duplexní stability preferuje algoritmus popsany Rychlikem a Wetmurem a pro delší značky, například okolo 30 až 50 nukleotidů nebo více, algoritmus uveřejněný Suggsem a kol., str. 683 až 693 v Brown, redakce, ICN-UCLA Symp. Dev. Biol., 23 (Academic Press, New York, 1981) lze s výhodou použít. Je zřejmé, že pro toho, kdo má zkušenost v oboru vytváření souborů minimálně zkříženě hybridizujících podjednotek v rámci tohoto vynálezu, je k dispozici řada způsobů. Například pro minimalizaci účinků různých energií stackingu bází terminálních nukleotidů při tvorbě podjednotek je možno zajistit podjednotky, které mají stejné terminální nukleotidy. Tímto způsobem bude při vzájemném spojení podjednotek součet energií stackingu bází všech připojených terminálních nukleotidů stejný, čímž se snižuje či eliminuje variabilita teplot tání značky.

Ve značkách o více podjednotkách lze též přidat "slovo" terminálních nukleotidů ukázané níže kurzívou ke každému konci značky, takže se vždy vytvoří dokonalé spárování mezi ním a podobným terminálním "slovem", na kterémkoliv jiném komplementu značky. Taková zvětšená značka by byla ve formě:

| | | | | | | |
|------------|-------------------------|-------------------------|-----|----------------------------------|--------------------------------|------------|
| <i>W</i> | <i>W</i> ₁ | <i>W</i> ₂ | ... | <i>W</i> _{<i>k</i>-1} | <i>W</i> _{<i>k</i>} | <i>W</i> |
| <i>W</i> ' | <i>W</i> ' ₁ | <i>W</i> ' ₂ | ... | <i>W</i> ' _{<i>k</i>-1} | <i>W</i> ' _{<i>k</i>} | <i>W</i> ' |

kde *W* označená ' znamenají komplementy. Při koncích značek vytvářejících vždy dokonale spárované duplexy budou všechna nesprávně spárovaná slova vnitřními nesprávnými spárováními, čímž se snižuje stabilita duplexů značka-komplement, které by jinak měly nesprávně spárovaná slova na svých koncích. Je dobře známo, že duplexy s vnitřními chybnými spárováními



jsou významně méně stabilní než duplexy se stejným chybným spárováním na konci.

Pro oligonukleotidové značky užitě pro třídění jsou preferovaným ztělesněním minimálně zkříženě hybridizujících souborů soubory, jejichž podjednotky tvoří až tři ze čtyř přirozených nukleotidů. Jak bude diskutováno podrobněji níže, nepřítomnost jednoho typu nukleotidu v oligonukleotidových značkách umožňuje nanesení cílových polynukleotidů na pevné nosiče použitím exonukleázové aktivity 5'--3' DNA polymerázy. Následuje příklad minimálně zkříženě hybridizujícího souboru podjednotek, kde každá obsahuje čtyři nukleotidy zvolené ze skupiny zahrnující A, G a T:

Tabulka II

| | | | | |
|-----------|-------|-------|-------|-------|
| Slovo: | W_1 | W_2 | W_3 | W_4 |
| Sekvence: | GATT | TGAT | TAGA | TTTG |
| Slovo: | W_5 | W_6 | W_7 | W_8 |
| Sekvence: | GTAA | AGTA | ATGT | AAAG |

V tomto souboru by každý člen vytvářel duplex mající tři chybně spárované báze s komplementem každého jiného členu.

Pro oligonukleotidové značky použité pro zajištění označení kódovaných adaptorů se používají všechny čtyři nukleotidy.

Oligonukleotidové značky podle tohoto vynálezu a jejich komplementy se výhodně syntetizují na automatickém syntetizátoru DNA, například Applied Biosystems, Inc. (Foster City, California) model 392 nebo 394 DNA/RNA Synthesizer



s použitím standardních chemických postupů, jako je fosforamiditový postup uveřejněný například v Beaucage a Iyer, *Tetrahedron*, 48, 2223-2311 (1992), Molko a kol., US patent 4 980 460, Koster a kol., US patent 4 725 677, Caruthers a kol., US patenty 4 415 732, 4 458 066 a 4 973 679 a podobně. Alternativní chemické postupy vedoucí například k nepřírozeným skupinám hlavního řetězce, jako u peptidových nukleových kyselin (PNA), N3'→P5' fosforamidatů a podobně lze též použít. V některých ztělesněních mohou značky obsahovat přirozené nukleotidy, které umožňují zpracování a manipulaci pomocí enzymů, zatímco odpovídající komplementy mohou obsahovat nepřírozené nukleotidové analogy, jako jsou peptidové nukleové kyseliny či podobné sloučeniny, které napomáhají vytváření stabilnějších duplexů během třídění. V případě značek užitých pro zajištění označení kódovaných adaptorů mohou být oligonukleotidové značky i komplementy značek vytvářeny z nepřírozených nukleotidů či analogů s podmínkou, že ligace nastává buď chemicky nebo enzymaticky.

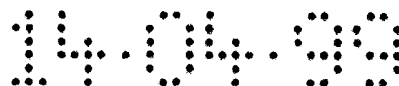
Dvojvláknové formy značek se mohou připravit odděleně syntézou komplementárních vláken s následným smícháním za podmínek, které umožňují tvorbu duplexů. Alternativně lze dvojvláknové značky vytvářet nejprve syntézou jednovláknového repertoáru spřaženého do známé sekvence oligonukleotidů, která slouží jako vazebné místo primerů. Druhé vlákno se poté syntetizuje kombinací jednovláknového repertoáru s primerem a prodloužením pomocí polymerázy. Tento další způsob popisuje Oliphant a kol., *Gene*, 44, 177-183 (1986). Tyto duplexní značky se potom mohou vkládat do klonujících vektorů spolu s cílovými polynukleotidy pro třídění a manipulaci cílového polynukleotidu v souladu s tímto vynálezem.

Když se používají komplementy mající až čtyři nukleo-

14.04.99

tidy, které mají zvýšené vazebné vlastnosti, jako PNA nebo fosforamidaty oligonukleotidu N3'--P5', lze třídění provádět vytvářením smyček D mezi značkami obsahujícími přirozené nukleotidy a jejich PNA nebo fosforamidatovými komplementy jako alternativa k "strippingové" reakci s použitím 3'→5' exonukleázové aktivity DNA polymerázy pro zajištění jednovláknové značky.

Oligonukleotidové značky pro třídění mohou být v rozmezí délek od 12 do 60 nukleotidů nebo párů bází. Přednostně jsou oligonukleotidové značky v rozmezí délek 18 až 40 nukleotidů či párů bází. Ještě lepší je rozmezí délek oligonukleotidů od 25 do 40 nukleotidů či párů bází. Ve smyslu preferovaných a ještě preferovanějších počtů podjednotek mohou být tato rozmezí vyjádřena následujícím způsobem:



Tabulka III

Počty podjednotek značek v preferovaných ztělesněních

| Monomery v podjednotce | Nukleotidy v oligonukleotidové značce | | |
|---------------------------|---------------------------------------|---------------|---------------|
| | (12-60) | (18-40) | (25-40) |
| 3 | 4-20 podjedn. | 6-13 podjedn. | 8-13 podjedn. |
| 4 | 3-15 podjedn. | 4-10 podjedn. | 6-10 podjedn. |
| 5 | 2-12 podjedn. | 3-8 podjedn. | 5-8 podjedn. |
| 6 | 2-10 podjedn. | 3-6 podjedn. | 4-6 podjedn. |

Nejvhodnější oligonukleotidové značky pro třídění jsou jednovláknové a specifická hybridizace nastává Watsonovým-Crickovým párováním s komplementy značek.

Dává se přednost repertoárům jednovláknových oligonukleotidových značek pro třídění, které obsahují alespoň 100 členů, ještě vhodnější jsou repertoáry těchto značek obsahující nejméně 1000 členů a nejvhodnější jsou repertoáry těchto značek obsahující nejméně 10000 členů.

Preferují se repertoáry komplementů značek pro zajištění označení obsahující alespoň 16 členů, vhodnější repertoáry těchto značek obsahují alespoň 64 členů. Ještě vhodnější jsou tyto repertoáry komplementů značek obsahující 16 až 1024 členů, například počet pro identifikaci nukleotidů ve vyčnívajících vláknech v délce od 2 do 5 nukleotidů. Nejvhodnější jsou repertoáry komplementů značek obsahující od 64 do 256 členů. Repertoáry žádaných rozměrů se volí přímým vytvářením souborů slov či podjednotek o žádaném rozměru, například pomocí počítačových programů dodatků I a II



nebo se repertoáry vytvářejí generací souboru slov, která se potom používají v kombinatorické syntéze pro obdržení repertoáru žádaného rozměru. Přednostně je délka jednovláknových komplementů značek pro zajištění označení mezi 8 až 20. Vhodnější délka je mezi 9 až 15.

Triplexní značky

Ve ztělesněních, ve kterých nastává specifická hybridizace vytváření triplexů, sledují sekvence značek stejné principy jako při značkách s vytvářením duplexů, avšak existují další omezení na volbu sekvencí podjednotek. Obecně je připojení třetího vlákna Hoogsteenovým způsobem vazby nejstabilnější podél homopyrimidinových-homopurinových řetězců dvojitěvláknového cílového nukleotidu. Obvykle vytvářejí základní triplety motivy T-A*T nebo C-G*C (kde "-" značí Watsonovo-Crickovo párování a "*" značí Hoogsteenův způsob vazby), avšak jsou možné i jiné motivy. Například Hoogsteenovo párování bází umožňuje paralelní a antiparalelní orientace mezi třetím vláknem (Hoogsteenovo vlákno) a vláknem duplexu bohatým na puriny, ke kterému se třetí vlákno váže v závislosti na podmínkách a složení těchto vláken. V literatuře existují rozsáhlá vodítka ohledně volby vhodných sekvencí orientace podmínek typu nukleosidu (například, zda se použijí ribozové či deoxyribozové nukleotidy), modifikací bází (například methylovaný cytosin a podobně) pro maximalizaci či jinou regulaci triplexní stability podle požadavku jednotlivých ztělesnění například Roberts a kol., Proc. Natl. Acad. Sci., 88, 9397-9401 (1991), Roberts a kol., Science, 258, 1463-1466 (1992), Roberts a kol., Proc. Natl. Acad. Sci., 93, 4320-4325 (1996), Distefano a kol., Proc. Natl. Acad. Sci., 90, 1179-1183 (1993), Mergny a kol., Biochemistry, 30, 9791-9798 (1991), Cheng a kol., J. Am.



Chem. Soc., 114, 4465-4474 (1992), Beal a Dervan, Nucleic Acids Research, 20, 2773-2776 (1992), Beal a Dervan, J. Am. Chem. Soc., 114, 4976-4982 (1992), Giovannangeli a kol., Proc. Natl. Acad. Sci., 89, 8631-8635 (1992), Moser a Dervan, Science, 238, 645-650 (1987), McShan a kol., J. Biol. Chem., 267, 5712-5721 (1992), Yoon a kol., Proc. Natl. Acad. Sci., 89, 3840-3844 (1992), Blume a kol., Nucleic Acids Research, 20, 1777-1784 (1992), Thuong a Helene, Angew. Chem. Int. Ed. Engl., 32, 666-690 (1993), Escude a kol., Proc. Natl. Acad. Sci., 93, 4365-4369 (1996) a podobně. Podmínky pro anelování jednovláknových či duplexních značek na jejich jednovláknové či duplexní komplementy jsou dobře známy, například Ji a kol., Anal. Chem., 65, 1323-1328 (1993), Cantor a kol., US patent č. 5 482 836 a podobně. Použití triplexních značek při třídění má výhodu, že se nepožaduje "strippingová" reakce s polymerázou pro expozici značky pro anelování na její komplement.

Přednostně jsou oligonukleotidovými značkami podle tohoto vynálezu používajícími triplexní hybridizaci dvojevláknové DNA a odpovídající komplementy značek jsou jednovláknové. Ještě vhodněji se používá 5-methylcytosin místo cytosinu v komplementech značek pro rozšíření rozmezí stability pH triplexu vytvářeného mezi značkou a jejím komplementem. Preferované podmínky pro vytváření triplexů se úplně popisují ve výše citovaných odkazech. Stručně řečeno nastává hybridizace v koncentrovaném roztoku soli například v 1,0 M roztoku chloridu sodného, 1,0 M roztoku octanu draselného nebo podobně při pH pod 5,5 (nebo 6,5, pokud se použije 5-methylcytosin). Teplota hybridizace závisí na délce a složení značky, avšak pro 18- až 20-merní značku či delší je vhodná hybridizace při teplotě místnosti. Promytí se může provádět s méně koncentrovanými roztoky soli, například

s 10 mM roztokem octanu sodného, 100 mM roztokem chloridu hořečnatého, při pH 5,8, při teplotě místnosti. Značky se mohou eluovat ze svých komplementů značek inkubací v podobném roztoku soli při pH 9,0.

Minimálně zkříženě hybridizující soubory oligonukleotidových značek vytvářející triplexy se mohou generovat počítačovým programem dodatku II nebo podobnými programy. Příklad souboru dvojitých 8-merních slov je popsán níže velkými písmeny s odpovídajícími komplementy malými písmeny. Každé takové slovo se liší od každého jiného slova souborem tří párů bází.

Tabulka IV

Příklady minimálně zkřížené hybridizujících souborů dvojvláknových 8-merních značek

| | | | |
|-------------|-------------|-------------|-------------|
| 5'-AAGGAGAG | 5'-AAAGGGA | 5'-AGAGAAGA | 5'-AGGGGGGG |
| 3'-TTCTCTC | 3'-TTTCCCCT | 3'-TCTCTTCT | 3'-TCCCCCCC |
| 3'-ttcctctc | 3'-tttcccct | 3'-tctcttct | 3'-tccccccc |
| 5'-AAAAAAA | 5'-AAGAGAGA | 5'-AGGAAAAG | 5'-GAAAGGAG |
| 3'-TTTTTTTT | 3'-TTCTCTCT | 3'-TCCTTTTC | 3'-CTTTCCTC |
| 3'-tttttttt | 3'-ttctctct | 3'-tccttttc | 3'-ctttcctc |
| 5'-AAAAAGG | 5'-AGAAGAGG | 5'-AGGAAGGA | 5'-GAAGAAGG |
| 3'-TTTTTCCC | 3'-TCTTCTCC | 3'-TCCTTCCT | 3'-CTTCTTCC |
| 3'-tttttccc | 3'-tcttctcc | 3'-tccttcct | 3'-cttcttcc |
| 5'-AAAGGAAG | 5'-AGAAGGAA | 5'-AGGGGAAA | 5'-GAAGAGAA |
| 3'-TTTCCTTC | 3'-TCTTCCTT | 3'-TCCCCTTT | 3'-CTTCTCTT |
| 3'-tttccttc | 3'-tcttcttt | 3'-tccccttt | 3'-cttctctt |



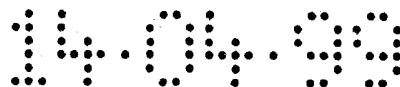
Tabulka V

Velikost repertoáru různých dvojevláknových značek vytvářejících triplexy s jejich komplementy značek

| Délka slova oligonukleotidů | Nukleotidový rozdíl mezi oligonukleotidy minimálně zkříženě hybridizujícího souboru | Maximální velikost minimálně zkříženě hybridizujícího souboru | Velikost repertoáru se čtyřmi slovy | Velikost repertoáru s pěti slovy |
|-----------------------------|---|---|-------------------------------------|----------------------------------|
| 4 | 2 | 8 | 4096 | $3,2 \times 10^4$ |
| 6 | 3 | 8 | 4096 | $3,2 \times 10^4$ |
| 8 | 3 | 16 | $6,5 \times 10^4$ | $1,05 \times 10^6$ |
| 10 | 5 | 8 | 4096 | |
| 15 | 5 | 92 | | |
| 20 | 6 | 768 | | |
| 20 | 7 | 484 | | |
| 20 | 8 | 189 | | |
| 20 | 9 | 30 | | |

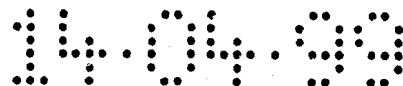
Syntéza a struktura adaptorů

Kódované adaptory a adaptory štěpení se vhodným způsobem syntetizují na automatizovaných syntetizátorech DNA s použitím standardních chemických způsobů, jako jsou fosforamiditové chemické postupy zveřejněné například v následujících odkazech: Beaucage a Iyer, *Tetrahedron*, **48**, 2223-2311 (1992), Molko a kol., US patent č. 4 980 460,



Koster a kol., US patent č. 4 725 677, Caruthers a kol., US patenty č. 4 415 732, 4 458 066 a 4 973 679 a podobně. Alternativní chemické způsoby vedoucí například k jiným než přirozeným skupinám základního řetězce, jako jsou fosforothioatové, fosforoamidatové a podobně, se mohou též použít s podmínkou, že výsledné oligonukleotidy jsou kompatibilní s činidly ligace a/nebo štěpení v daném ztělesnění. Obvykle se po syntéze komplementárních vláken tato vlákna kombinují s vytvořením dvojláknového adaptoru. Vyčnívající vlákno z kódovaného adaptoru se může syntetizovat jako taková směs, že ve vyčnívající části je zastoupena každá možná sekvence. Tyto směsi se snadno připravují dobře známými způsoby, například jak popisuje Telenius a kol., *m Genomics*, 13, 718-725 (1992), Welsh a kol., *Nucleic Acids Research*, 19, 5275-5279 (1991), Grothues a kol., *Nucleic Acids Research*, 21, 1321-1322 (1993), Hartley, evropská patentová přihláška 90304496.4 (publikace EP č. 395398) a podobně. Obecně tyto techniky jednoduše vyžadují aplikaci směsí aktivovaných monomerů na rostoucí oligonukleotidy během kroků spojování, pokud se požaduje uvedení většího počtu nukleotidů. Jak se diskutuje výše, v některých ztělesněních se může požadovat snížení komplexity adaptorů. Toho lze dosáhnout s použitím analogů snižující komplexitu, jako je deoxyinosin, 2-aminopurin či podobně, jak se popisuje v Kong Thoo Lin a kol., *Nucleic Acids Research*, 20, 5149-5152 nebo US patent č. 5 002 867, Nichols a kol., *Nature*, 369, 492-493 (1994) a podobně.

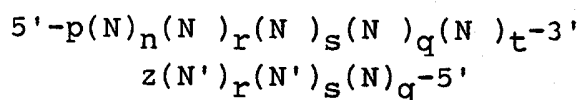
V některých ztělesněních se může požadovat syntéza kódovaných adaptorů nebo adaptorů štěpení jako jednoduchého polynukleotidu obsahujícího self-komplementární oblasti. Po syntéze se self-komplementární oblasti anelují s vytvořením adaptoru s vyčnívajícím vláknem na jednom konci a jednovlák-



novou smyčkou na druhém konci. V těchto ztělesněních může oblast smyčky přednostně obsahovat 3 až 10 nukleotidů nebo jiných srovnatelných vazebných zbytků, například alkyl-etherových skupin, jak uvádí US patent č. 4 914 210. Jsou dostupné mnohé způsoby pro připojení reaktivních skupin k bázím nebo internukleosidovým spojením pro označení, jak se diskutuje v odkazech citovaných níže.

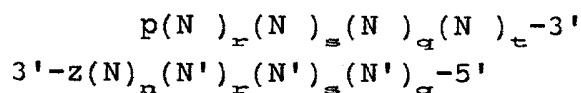
Při použití konvenčních ligáz v tomto vynálezu, jak se popisuje podrobněji níže, může být v některých ztělesněních konec adaptoru 5' fosforylován. 5' monofosfát se může připojit ke druhému oligonukleotidu buď chemicky nebo enzymaticky pomocí kinázy, například Sambrook a kol., *Molecular Cloning: A Laboratory Manual*, 2. vydání (Cold Spring Harbor Laboratory, New York, 1989). Chemickou fosforylací popisují Horn a Urdea, *Tetrahedron Lett.*, 27, 4705 (1986) a činidla pro provedení uveřejněných způsobů jsou komerčně dostupná, jako například 5' Phosphate-ONTM od Clontech Laboratories (Palo Alto, Kalifornie).

Kódované adaptory podle tohoto vynálezu mohou mít několik ztělesnění v závislosti například na tom, zda se používají jednovláknové nebo dvojitěvláknové značky, zda se používá větší počet značek, zda se používá vyčnívající vlákno 5' nebo 3', zda se používá blokovácí skupina 3' a podobně. Vzorce pro několik ztělesnění kódovaných adaptorů se udávají níže. Preferované struktury kódovaných adaptorů s použitím jednovláknové značky jsou následující:





nebo



kde N je nukleotid a N' je jeho komplement, p je fosfátová skupina, z je 3' hydroxylová nebo 3' blokující skupina, n je celé číslo mezi 2 a 6 včetně, r je celé číslo větší než nebo rovné nule, s je celé číslo, které je buď mezi 4 až 6, pokud má adaptor nukleázové rozpoznávací místo nebo 0, pokud neexistuje nukleázové rozpoznávací místo, q je celé číslo větší než nebo rovné 0 a t je celé číslo mezi 8 a 20 včetně. Ještě vhodnější je n 4 nebo 5 a t mezi 9 a 15 včetně. Pokud kódovaný adaptor obsahuje nukleázové rozpoznávací místo, volí se oblast "r" nukleotidových párů takovým způsobem, že se předem určený počet nukleotidů odštěpuje od cílového polynukleotidu, když se používá nukleázové rozpoznávací místo. Rozměr "r" v konkrétním ztělesnění závisí na dosahu nukleázy (ve smyslu definovaném v US patentu 5 599 675 a WO 95/27080) a na počtu nukleotidů, které mají být odštěpeny od cílového polynukleotidu. Přednostně je r mezi 0 a 20, ještě lépe je r mezi 0 a 12. Oblast "q" nukleotidových párů je segment vymezující vzdálenost mezi nukleázovým rozpoznávacím místem a oblastí značky kódované sondy. Oblast "q" nukleotidu může zahrnovat další nukleázová rozpoznávací místa označující zbytky generující signál či podobně. Jednovláknový oligonukleotid o "t" nukleotidech je "t-merní" oligonukleotidová značka zvolená z minimálně zkříženě hybridizujícího souboru.

Skupina blokující 3'"z" může mít různé formy a může zahrnovat téměř jakoukoliv chemickou složku, která zabraňuje ligaci a která neinterferuje s ostatními kroky tohoto způsobu.

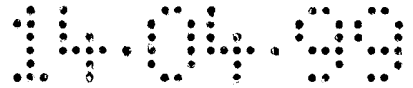


bu, například s odstraněním vlákna s blokovaným 3', s ligací či podobně. Skupiny blokující 3' zahrnují například, avšak nikoliv s omezením pouze na tyto případy, atom vodíku (to jest 3' deoxy), fosfátovou skupinu, fosforothioatovou skupinu, acetylovou skupinu a podobně. Přednostně je skupina blokující 3' fosfátovou skupinou vzhledem k výhodě při přidávání této skupiny během syntézy vlákna s blokovaným 3' a výhodnosti odstraňování skupiny fosfatázou pro dosažení schopnosti ligace vlákna pomocí ligázy. Oligonukleotid mající fosfát 3' se může syntetizovat s použitím způsobu popsaného v kapitole 12 v práci Eckstein, redakce, *Oligonucleotides and Analogues: A Practical Approach* (IRL Press, Oxford, 1991).

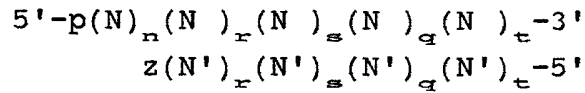
Další skupiny blokující 3' jsou dostupné chemickými způsoby vyvinutými pro reverzibilní nukleotidy ukončující řetězec ve schématech sekvenování báze po bázi, jak se například popisuje v následujících odkazech: Cheeseman, US patent 5 302 509, Tsien a kol., mezinárodní patentová přihláška WO 91/06678, Canard a kol., *Gene*, 148, 1-6 (1994) a Metzker a kol., *Nucleic Acids Research*, 22, 4259-4267 (1994). Zhruba řečeno umožňují tyto chemické způsoby chemické či enzymatické odstraňování specifických blokovacích skupin (obvykle majících připojené označení) pro vytvoření volné hydroxylové skupiny na konci 3' základního vlákna.

Jestliže z je skupina blokující 3', je to fosfátová skupina a dvojvláknová část adaptoru obsahuje nukleázové rozpoznávací místo nukleázy, jejíž rozpoznávací místo je oddělené od jejího místa štěpení.

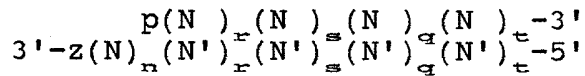
Když se použijí dvojvláknové oligonukleotidové značky, které specificky hybridizují s jednovláknovými komple-



menty značek s vytvořením triplexních struktur, preferují se kódované značky podle tohoto vynálezu ve formě:

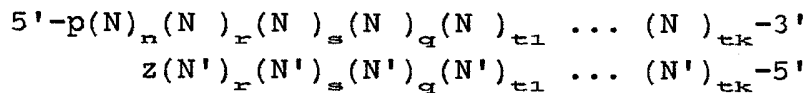


nebo

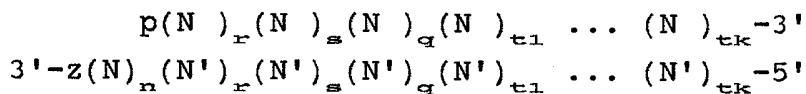


kde N, N', p, q, r, s, z a n se definují výše. V tomto ztělesnění se preferuje t jako celé číslo v rozmezí 12 až 24.

Je zřejmé, že existují přídatné struktury obsahující elementy výše popsaných základních návrhů, které jsou známé tomu, kdo má zkušenost v oboru. Například kódované adaptory podle tohoto vynálezu zahrnují ztělesnění s větším počtem značek, jako je:



nebo



kde kódovaný adaptor zahrnuje k dvojvláknových značek. Preferuje se $t_1 = t_2 = \dots t_k$ a k 1, 2 nebo 3.

Označování komplementů značek

Komplementy značky podle tohoto vynálezu se mohou



označovat různými způsoby pro dekódování oligonukleotidové značky včetně přímého či nepřímého připojení radioaktivních zbytků, fluorescenčních zbytků, kolorimetrických zbytků, chemiluminiscenčních zbytků a podobně. Mnohé souborné referáty o způsobech značení DNA a tvorbě adaptorů DNA poskytují vodítko použitelné pro tvorbu adaptoru podle tohoto vynálezu. Tyto referáty zahrnují práce Matthews a kol., Anal. Biochem., 169, 1-25 (1988), Haugland, Handbook of Fluorescent Probes and Research Chemicals (Molecular Probes, Inc., Eugene, 1992), Keller a Manak, DNA Probes, 2. vydání (Stockton Press, New York, 1993) a Eckstein, redakce, Oligonucleotides a Analogues: A Practical Approach (IRL Press, Oxford, 1991), Wetmur, Critical Reviews in Biochemistry and Molecular Biology, 26, 227-259 (1991) a podobně. Mnohé konkrétní způsoby použitelné pro tento vynález lze nalézt v následujících odkazech: Fung a kol., US patent 4 757 141, Hobbs, Jr., a kol., US patent 5 151 507, Cruickshank, US patent 5 091 519 (syntéza funkcionalizovaných oligonukleotidů pro připojení reporterových skupin), Jablonski a kol., Nucleic Acids Research, 14, 6115-6128 (1986) (konjugáty enzym-oligonukleotid), Ju a kol., Nature Medicine, 2, 246-249 (1996) a Urdea a kol., US patent 5 124 246 (rozvětvené DNA). Místa připojení označujících zbytků nejsou kritická s podmínkou, že tato označení neinterferují s kroky ligace a/nebo štěpení.

Preferuje se použití jednoho nebo více fluorescenčních barviv pro označení komplementů značky, jak například uveřejňuje Menchen a kol., US patent 5 188 934, Bergot a kol., mezinárodní patentová přihláška PCT/US90/05565 (WO 91/05060). Pojem "zbytek vytvářející fluorescenční signál" tak, jak se zde používá, značí signální prostředky přenášející informaci prostřednictvím vlastností absorpce a/nebo



emise fluorescence jedné či více molekul. Tyto fluorescenční vlastnosti zahrnují intenzitu fluorescence, dobu trvání fluorescence, emisní spektrální charakteristiky, přenos energie a podobně.

Ligace adaptorů a zabránění self-ligaci

V souladu s preferovaným ztělesněním tohoto vynálezu se navazují adaptory štěpení na konce cílových polynukleotidů pro přípravu těchto konců pro případnou ligaci kódovaných adaptorů. Ligace se přednostně provádí enzymaticky s použitím ligázy při standardním způsobu. Mnohé ligázy jsou známé a jsou vhodné pro použití v tomto vynálezu, například Lehman, *Science*, 186, 790-797 (1974), Engler a kol., *DNA Ligases*, str. 3 až 30 v Boyer, redakce, *The Enzymes*, sv. 15B (Academic Press, New York, 1982) a podobně. Preferované ligázy zahrnují T4 DNA ligázu, T7 DNA ligázu, DNA ligázu *E. coli*, Taq ligázu, Pfu ligázu a Tth ligázu. Způsoby jejich použití jsou dobře známy, například Sambrook a kol. (citováno výše), Barany, *PCR Methods and Applications*, 1, 5-16 (1991), Marsh a kol., *Strategies*, 5, 73-76 (1992) a podobně. Obecně požadují ligázy přítomnost 5' fosfátové skupiny pro ligaci na 3' hydroxylovou skupinu přiléhajícího vlákna. To se výhodně zajistí alespoň pro jedno vlákno cílového polynukleotidu volbou nukleázy ponechávající 5' fosfát například jako Fok I.

Zvláštní problém může vzniknout při zacházení s konci polynukleotidů nebo s adaptory, které jsou schopné self-ligace, jak znázorňuje obr. 2, kde jsou čtyřnukleotidová vyčnívající vlákna zakotvených polynukleotidů navzájem komplementární (114). Tento problém je zvláště závažný ve ztělesněních, ve kterých se analyzované polynukleotidy



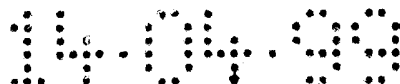
(112) prezentují adaptorům jako jednotlivé populace identických polynukleotidů připojených na pevný nosič (110). V těchto situacích se mohou volné konce zakotvených polynukleotidů zkroutit a vytvořit spolu dokonale spárované duplexy (116). Jsou-li vlákna 5' konců fosforylovaná, polynukleotidy se v přítomnosti ligázy snadno spojují. Analogický problém též existuje pro dvojlátkové adaptory. Zejména pokud jsou jejich vlákna 5' fosforylována, může se vlákno 5' adaptoru navazovat na volnou hydroxylovou skupinu 3' nebo na jiný adaptor, jakmile jsou nukleotidové sekvence jejich vyčnívajících vláken komplementární. Dojde-li k self-ligaci, nejsou pro analýzu či zpracování k dispozici ani vyčnívajcí vlákna adaptoru ani vyčnívajcí vlákna cílových polynukleotidů. To opět vede ke ztrátě či vymizení signálu vyvolávaných jako odpověď na správné ligace adaptorů s cílovými polynukleotidy. Je-li pravděpodobnost palindromového 4-meru vyskytujícího se v náhodné sekvenci stejná jako pravděpodobnost opakovaného páru nukleotidů (6,25 %), lze u způsobů založených na adaptoru pro sekvenování de novo očekávat vysoký stupeň selhání po několika cyklech následkem self-ligace. Jestliže k tomu dojde, stává se další analýza polynukleotidu nemožnou.

Výše popisované problémy se mohou řešit provedením vynálezu v následujících krocích, které se znázornují na obr. 3A pro příklad preferovaného ztělesnění: (a) ligace (120) kódovaného adaptoru na konec polynukleotidu (122), kde tento konec polynukleotidu má defosforylovanou 5' hydroxylovou skupinu a konec kódovaného adaptoru (124), který se má navazovat, má první vlákno (126) a druhé vlákno (128) a druhé vlákno kódovaného adaptoru má blokující skupinu 3' (130), (b) odstranění blokující skupiny 3' druhého vlákna po ligaci, například promytím (132) nebo enzymatickým či che-



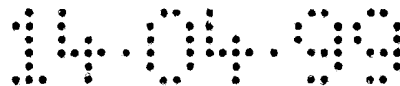
mickým odstraněním skupiny in situ, to jest zpracováním fosfatázou, pokud je blokující skupinou fosfát, (c) fosforylace (134) hydroxylové skupiny 5' polynukleotidu, (d) ligace (136) druhého vlákna (142) majícího neblokovaný zbytek 3' pro regeneraci kódovaného adaptoru (138) a (e) identifikace (144) jednoho či více nukleotidů na konci polynukleotidu identitou kódovaného adaptoru, který je k němu navázán, například fluorescenčním označením (140) komplementu značky. Kódované adaptory a cílové polynukleotidy se mohou kombinovat pro ligaci buď jednotlivě nebo jako směsí. Například stejný jednotlivý druh adaptoru majícího definovanou sekvenci se může kombinovat s jednotlivým druhem polynukleotidu majícího společnou (a možná neznámou) nukleotidovou sekvenci, nebo jednotlivý druh adaptoru majícího definovanou sekvenci se může kombinovat se směsí polynukleotidů, jako je větší množství uniformních populací identických polynukleotidů připojených k různým pevným nosičům ve stejné reakční nádobě, jak popisuje například Brenner a kol. v mezinárodní patentové přihlášce PCT/US96/09513 (WO 41011) nebo směs kódovaných adaptorů, zejména směsí mající různé nukleotidové sekvence ve svých vyčnívajících vláknech se mohou kombinovat s jednotlivým druhem polynukleotidu, nebo směs kódovaných adaptorů se může kombinovat se směsí polynukleotidů. Když se používá pojem "adaptor" nebo "kódovaný adaptor" v jednotném čísle, zahrnuje tento pojem směsí adaptorů majících různé sekvence vyčnívajících vláken stejně jako jednotlivý druh adaptoru mající stejnou sekvenci vyčnívajícího vlákna způsobem analogickým použití pojmu "sonda".

Kromě odstranění táním lze 3' deoxy odstranit z druhého vlákna polymerázovou "výměnnou" reakcí popsanou Kuijperem a kol., *Gene*, 112, 147-155 (1992), Aslanidisem a kol., *Nucleic Acid Research*, 18, 6069-6074 (1990) a podob-



ně. Stručně řečeno lze použít 5'--3' exonukleázovou aktivitu T4 DNA polymerázy a podobných enzymů pro výměnu nukleotidů v základním vlákne s jejich trifosfátovými protějšky v roz-
toku například Kuijper a kol. (citováno výše). Touto reakcí může být tedy 3' dideoxynukleotid vyměněn za 2'-deoxy-3'-
-hydroxynukleotid z reakční směsi, která zajišťuje možnost připojení druhého vlákna na cílový polynukleotid po zpracování polynukleotidovou kinázou.

Preferované ztělesnění používající cykly ligace a štěpení zahrnuje následující kroky: (a) ligaci (220) kódovaného adaptoru na konec polynukleotidu (222), kde tento konec polynukleotidu má defosforylovanou hydroxylovou skupinu 5', konec dvojvláknového adaptoru, který se má navazovat (224), má první vlákno (226) a druhé vlákno (228), druhé vlákno dvojvláknového adaptoru má skupinu blokující 3' (230) a dvojvláknový adaptor má nukleázové rozpoznávací místo (250) nukleázy, jejíž rozpoznávací místo je oddělené od jejího místa štěpení, (b) odstranění skupiny blokující 3' po ligaci, například vymytím z druhého vlákna (232), (c) fosforylaci (234) hydroxylové skupiny 5' polynukleotidu, (d) ligaci (236) druhého vlákna (242) majícího neblokovaný zbytek 3' pro regeneraci dvojvláknového adaptoru (238) a nukleázové rozpoznávací místo (250), (e) identifikaci (244) jednoho či více nukleotidů na konci polynukleotidu identitou navázaného adaptoru, (f) štěpení (252) polynukleotidu nukleázou, která rozpoznává rozpoznávací místo, takže je polynukleotid zkrácen o jeden nebo více nukleotidů a rozpoznávací místo je umístěné ve znázorněném adaptoru (224) takovým způsobem, že štěpení (254) odstraňuje dva nukleotidy z polynukleotidu (222), (g) defosforylaci (256) konce 5' polynukleotidu a (h) opakování (258) kroků (a) až (g).



Obvykle se před ligací konce polynukleotidů, které se mají analyzovat, připraví zpracováním jednou či více restrikčními endonukleázami s obdržení předem určených štěpení, obvykle s vyčnívajícimi vlákny 3' nebo 5', to jest "lepkavé" konce. Taková zpracování obvykle zanechávají fosforylovaná vlákna 5'. Přednostně se tyto fosforylované konce 5' defosforylují zpracováním fosfatázou, jako je telecí střevní alkalická fosfatáza či podobným enzymem s použitím standardních způsobů, jak popisuje Sambrook a kol., *Molecular Cloning*, druhé vydání (Cold Spring Harbor Laboratory, New York, 1989). Při odstranění fosfátů 5' se stanou cílové polynukleotidy neschopnými ligace v přítomnosti ligázy. Krok defosforylace přednostně zanechává volný hydroxyl 5'.

Preferované nukleázy

"Nukleáza" jako pojem užívaný v souladu s tímto vynálezem znamená kterýkoliv enzym, kombinaci enzymů či jiných chemických činidel nebo kombinace chemických činidel a enzymů, které při aplikaci na navázaný komplex, jak se podrobněji diskutuje níže, odštěpují navázaný komplex s poskytnutím zvětšeného adaptoru a zkráceného cílového polynukleotidu. Nukleáza podle tohoto vynálezu nemusí být jednotlivým proteinem nebo se skládat výhradně z kombinace proteinů. Klíčovým rysem nukleázy nebo kombinace činidel používaných jako nukleáza je, že její (jejich) místo štěpení je oddělené od jejího (jejich) rozpoznávacích míst. Vzdálenost mezi rozpoznávacím místem nukleázy a jejím místem štěpení se zde bude uvádět jako "dosah". Podle dohody se "dosah" definuje dvěma celými čísly, která poskytují počet nukleotidů mezi rozpoznávacím místem a hydrolyzovanými fosfodiesterovými vazbami každého vlákna. Například vlastnosti rozpoznávání a štěpení Fok I se obvykle znázorňují jako "GGATG(9/13)", neboť roz-



poznává a odpojuje dvojláknovou DNA následujícím způsobem
(číslo identifikace sekvence: 2):

```
5'- ... NNGGATGNNNNNNNNNN          NNNNNNNNNN ...
3'- ... NNCCTACNNNNNNNNNNNNNNNNN          NNNNNN ...
```

kde nukleotidy vyznačené polotučným písmem jsou místa rozpoznávání Fok I a N jsou libovolné nukleotidy a jejich komplementy.

Je důležité, aby nukleáza štěpila cílový polynukleotid teprve, když vytvoří komplex se svým rozpoznávacím místem. Nukleáza přednostně zanechává vyčnívající vlákno na cílovém poli nukleotidu po štěpení.

Přednostně jsou nukleázy používané v tomto vynálezu přirozené proteinové endonukleázy (i), jejichž rozpoznávací místo je oddělené od místa štěpení a (ii) jejichž štěpení vede k vyčnívajícímu vláknu na cílovém polynukleotidu. Nejvhodněji se používají v tomto vynálezu restriční endonukleázy třídy II, jak popisuje například Szybalski a kol., *Gene*, 100, 13-26 (1991), Roberts a kol., *Nucleic Acids Research*, 21, 3125-3137 (1993) a Livak a Brenner, US patent 5 093 245. Příklady nukleáz třídy II pro použití v tomto vynálezu zahrnují Alw XI, Bsm AI, Bbv I, Bsm FI, Sts I, Hga I, Bsc AI, Bbv II, Bce fI, Bce 851, Bcc I, Bcg I, Bsa I, Bsg I, Bsp MI, Bst 71 I, Ear I, Eco 571, Esp 31, Fau I, Fok I, Gsu I, Hph I, Mbo II, Mme I, Rle AI, Sap I, Sfa NI, Taq II, Tth 111II, Bco 51, Bpu AI, Fin I, Bsr DI a jejich isoschizomery. Preferované nukleázy zahrnují Bbv I, Fok I, Hga I, Ear I a Sfa NI. Bbv I je nejvhodnější nukleázou.

Přednostně se před nukleázovými kroky štěpení, obvyk-



le na počátku sekvenční operace, zpracuje cílový nukleotid pro blokování míst rozpoznávání a/nebo míst štěpení použité nukleázy. To zabraňuje nežádoucímu štěpení cílového polynukleotidu následkem náhodného výskytu nukleázových rozpoznávacích míst uvnitř cílového polynukleotidu. Blokování lze dosáhnout různými způsoby včetně methylace a zpracování sekvenčně specifických aptamerů, DNA vazebných proteinů nebo oligonukleotidů vytvářejících triplexy. Jestliže se použijí přirozené proteinové endonukleázy, lze rozpoznávací místa výhodně blokovat methylací cílového polynukleotidu kognátovou methylázou použité nukleázy. To znamená, že pro většinu, ne-li pro všechny bakteriální restriční endonukleázy typu B, existuje tak zvaná "kognátová" methyláza, která methyluje jejich rozpoznávací místo. Řadu takových methyláz popisuje Roberts a kol. (citováno výše) a Nelson a kol., *Nucleic Acids Research*, 21, 3139-3154 (1993) a jsou komerčně dostupné z různých zdrojů, zejména od New England Biolabs (Beverly, MA). Alternativně, použije-li se při přípravě cílových polynukleotidů pro sekvenování krok PCR, lze užít 5-methylcytosin během amplifikace takovým způsobem, že se přírodní cytosin nahradí methylovanými cytosiny v amplikonu. Tento způsob má další výhodu eliminace nutnosti zpracovat cílový polynukleotid navázaný na pevný nosič jiným enzymem.

Je zřejmé, že jedna z běžných praxí v oboru by mohla spojovat vlastnosti výše popsaných ztělesnění pro vytvoření dalších ztělesnění v souladu s tímto vynálezem, která však nejsou výslovně popsána výše.

Poskytuje se řada souprav pro provedení různých ztělesnění tohoto vynálezu. Obecně soupravy podle tohoto vynálezu zahrnují kódované adaptory, adaptory štěpení a komplementy označených značek. Soupravy dále zahrnují nukleázová

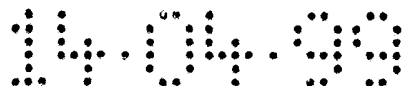


činidla, ligační činidla a návody pro provádění daného ztělesnění tohoto vynálezu. Ve ztělesněních používajících přirozené proteinové endonukleázy a ligázy se mohou užívat ligázové pufrы a nukleázové pufrы. V některých případech mohou být tyto pufrы identické. Takové soupravy mohou též zahrnovat metylázu a její reakční pufr. Přednostně soupravy též zahrnují jeden nebo více pevných nosičů, například mikročástic nesoucích komplementy značky pro třídění a zakotvování cílových polynukleotidů.

Připojování značek na polynukleotidy pro třídění na pevných nosičích

Důležitým aspektem tohoto vynálezu je třídění a připojování populací polynukleotidů, například z knihovny cDNA na mikročástice nebo na oddělené oblasti na pevném nosiči takovým způsobem, že každá mikročástice či oblast má v podstatě pouze jeden typ připojeného polynukleotidu. Tohoto záměru se dosáhne zajištěním toho, že v podstatě všechny různé polynukleotidy mají připojené rozdílné značky. Tato podmínka je opět splněna použitím vzorku plného souboru konjugátů značka-polynukleotid pro analýzu. Je přijatelné, aby identické polynukleotidy měly rozdílné značky, neboť to pouze vede k tomu, že stejný polynukleotid je zpracován či analyzován dvakrát v různých místech. Takové odebírání vzorků lze provádět buď otevřeně - například odebráním malého objemu z větší směsi - po připojení značek polynukleotidům (to lze provést inherentně jako sekundární účinek způsobů použitých pro zpracování polynukleotidů a značek) nebo se může vzorek odebrat jak otevřeně tak v rámci inherentní části kroků zpracování.

Přednostně se při tvorbě knihovny cDNA, kde v podsta-



tě veškeré rozdílné cDNA mají rozdílné značky, používá repertoár značek, jejichž komplexita či počet rozdílných značek značně převyšuje celkový počet mRNA odebraných ze vzorku buňky či tkáně. Přednostně je komplexita repertoáru značek nejméně 10x vyšší než populace polynukleotidů a ještě lépe je komplexita repertoáru značek 100x vyšší než populace polynukleotidů. Níže se uveřejňuje způsob tvorby knihovny cDNA s použitím směsi primerů obsahující plný repertoár příkladů značek o devíti slovech. Taková směs primerů obsahujících značky má komplexitu 8^9 neboli okolo $1,34 \times 10^8$. Jak ukazují Winslow a kol., *Nucleic Acids Research*, 19, 3251-3253 (1991), mRNA pro tvorbu knihovny se může extrahovat z pouhých 10 až 100 savčích buněk. Jelikož jednotlivá savčí buňka obsahuje okolo 5×10^5 kopií molekul mRNA o zhruba $3,4 \times 10^4$ různých typech, lze standardními způsoby izolovat mRNA ze zhruba 100 buněk neboli (teoreticky) okolo 5×10^7 molekul mRNA. Srovnání tohoto čísla s komplexitou směsi primerů ukazuje, že bez jakýchkoliv dalších kroků a dokonce předpokladu, že se mRNA převádějí na cDNA s dokonalou účinností (1% účinnost nebo menší je přesnější), způsob tvorby knihovny cDNA vede k populaci obsahující ne více než 37 % celkového počtu různých značek. To jest, bez jakéhokoliv otevřeného kroku odběru vzorků vytváří způsob inherentně vzorek obsahující 37 % nebo méně repertoáru značek. Pravděpodobnost obdržení dvojice za těchto podmínek je okolo 5 %, což je v preferovaném rozmezí. S mRNA z 10 buněk se redukuje frakce repertoáru značek na pouhých 3,7 % dokonce i za předpokladu, že všechny kroky zpracování nastávají se 100% účinností. Ve skutečnosti jsou účinnosti kroků zpracování pro tvorbu knihoven cDNA velmi nízké (pravidlo palce), takže dobrá knihovna by měla obsahovat okolo 10^8 klonů cDNA z mRNA extrahované z 10^6 savčích buněk.



Při použití větších počtů mRNA ve výše popsaném způsobu nebo pro větší množství polynukleotidů obecně, kde počet takových molekul překračuje komplexitu repertoáru značek, směs konjugátů značka-polynukleotid potenciálně obsahuje každé možné párování značek a typů mRNA či polynukleotidu. V těchto případech lze provést otevřený odběr vzorků odstraněním objemu vzorku po sériovém zředění výchozí směsi konjugátů značka-polynukleotid. Míra požadovaného zředění závisí na množství výchozího materiálu a účinnostech kroků zpracování, kteréžto údaje lze snadno zjistit.

Jestliže se mRNA extrahuje z 10^6 buněk (což by odpovídalo zhruba $0,5 \mu\text{g}$ poly(A)⁺RNA) a jsou-li přítomny primery zhruba v desetinásobném přebytku koncentrace - jak je tomu v obvyklém způsobu, například Sambrook a kol., Molecular Cloning, 2. vydání, str. 8.61 ($10 \mu\text{l}$ $1,8 \text{ kb}$ mRNA při 1 mg/ml se rovná zhruba $1,68 \times 10^{-11}$ molů a $10 \mu\text{l}$ 18-merního primeru při 1 mg/ml se rovná zhruba $1,68 \times 10^{-9}$ molů) potom by se celkový počet konjugátů značka-polynukleotid v knihovně cDNA jednoduše rovnal nebo byl menší než výchozí počet mRNA nebo okolo 5×10^{11} vektorů obsahujících konjugáty značka-polynukleotid (to opět předpokládá, že každý krok při tvorbě cDNA - syntéza prvního vlákna a syntéza druhého vlákna a ligace do vektoru - nastává s dokonalou účinností), což je velice konzervativní odhad. Skutečný počet je významně nižší.

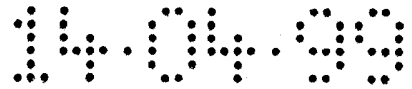
Jestliže je vzorek n konjugátů značka-polynukleotid náhodně odebrán z reakční směsi - což lze uskutečnit odebráním objemu vzorku, popisuje se pravděpodobnost odebrání konjugátů se stejnou značkou Poissonovým rozdělením $P(r) = e^{-\lambda} (\lambda^r) / r!$, kde r je počet konjugátů majících stejnou značku a $\lambda = np$, kde p je pravděpodobnost volby



dané značky. Jestliže $n = 10^6$ a $p = 1/(1,34 \times 10^8)$, potom $\lambda = 0,00746$ a $P(2) = 2,76 \times 10^{-5}$. Vzorek jednoho milionu molekul tedy poskytuje očekávaný počet dvojic dostatečně v preferovaném rozmezí. Takový vzorek se snadno obdrží následujícím způsobem. Předpokládejme, že 5×10^{11} mRNA se dokonale převede na 5×10^{11} vektorů s konjugáty značka-cDNA jako vložkami a že v reakčním roztoku o objemu $100 \mu\text{l}$ je 5×10^{11} vektorů. Čtyři desetinásobná zředění se mohou provést přenosem $10 \mu\text{l}$ z původního roztoku do nádoby obsahující $90 \mu\text{l}$ příslušného pufru, jako je TE. Tento způsob se může opakovat pro tři další zředění pro obdržení $100 \mu\text{l}$ roztoku obsahujícího 5×10^5 molekul vektorů na ml. Alikvót $2 \mu\text{l}$ z tohoto roztoku poskytuje 10^6 vektorů obsahujících konjugáty značka-cDNA jako vložky. Tento vzorek se poté amplifikuje přímou transformací kompetentní hostitelské buňky s následným pěstováním.

Jak se ovšem uvádí výše, žádný krok ve výše popsaném způsobu neprobíhá s dokonalou účinností. Zejména pokud se používají vektory pro amplifikaci vzorku konjugátů značka-polynukleotid, je krok transformace hostitele velice neúčinný. Obvykle se vychytá hostitelem a replikuje ne více než 1 % vektorů. Proto by pro tento způsob amplifikace by byla požadována dokonce nižší zředění k obdržení vzorku o 10^6 konjugátů.

Repertoár oligonukleotidových značek se může konjugovat s populací polynukleotidů různými způsoby včetně přímé enzymové ligace, amplifikace, například prostřednictvím PCR s použitím primerů obsahujících sekvence značek a podobně. Počáteční krok ligace poskytuje značně velkou populaci konjugátů značka-polynukleotid, takže jednotlivá značka je obecně připojena ke mnoha různým polynukleotidům. Avšak, jak



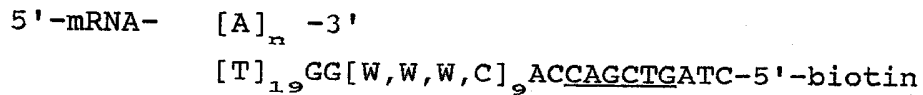
se uvádí výše, odebrání dostatečně malého vzorku konjugátů pravděpodobnost obdržení "dvojic", to jest stejná značka na různých polynukleotidech se může učinit zanedbatelnou. Obecně čím větší vzorek, tím vyšší pravděpodobnost obdržení dvojice. Proto nastává otázka volby mezi velkým vzorkem konjugátů značka-polynukleotid, který například zajišťuje adekvátní pokrytí cílového polynukleotidu v mnoha možnostech sekvenční operace či adekvátní reprezentací rychle se měnícího půlu mRNA a volbou malého vzorku zajišťujícího minimální počet přítomných dvojic. V řadě ztělesnění přítomnost dvojic pouze zvyšuje přídavný zdroj šumu nebo v případě sekvenování menší komplikaci při snímání a zpracování signálu neboť mikročástice poskytující více fluorescenčních signálů se mohou prostě ignorovat.

Termín "v podstatě všechny", jak se zde užívá v odkazu na značky připojované k molekulám, zejména polynukleotidům, má odrážet statistickou povahu způsobu odebrání vzorku pro obdržení populace konjugátů značka-molekula, které v podstatě neobsahují dvojice. Smysl v podstatě všechny ve smyslu skutečného procentického podílu konjugátů značka-molekula závisí na tom, jak se značky používají. Přednostně pro sekvenční analýzu nukleových kyselin v podstatě všechny znamená, že alespoň 80 % polynukleotidů má připojené jedinečné značky. Lepší je, pokud to znamená, že alespoň 90 % polynukleotidů má připojené jedinečné značky. Ještě lepší je, jestliže to znamená, že alespoň 95 % polynukleotidů má připojené jedinečné značky a nejlepší je, pokud to znamená, že alespoň 99 % polynukleotidů má připojené jedinečné značky.

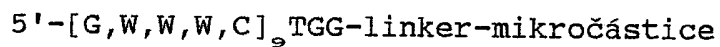
Je vhodné, jestliže populace polynukleotidů obsahuje informační RNA (mRNA), aby byly nukleotidové značky spojeny



reverzní transkripcí mRNA se souborem primerů přednostně obsahujícím komplementy sekvencí značek. Příklad souboru takových primerů by mohl mít následující sekvenci:



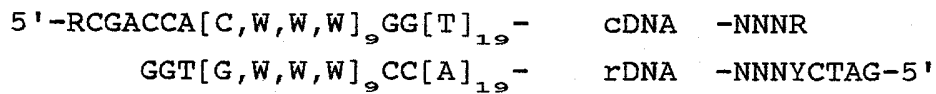
kde "[W,W,W,C]" představuje sekvenci oligonukleotidové značky o devíti podjednotkách, každá ze čtyř nukleotidů a "[W,W,W,C]" představuje výše popsané sekvence podjednotek, to jest "W" představuje T nebo A. Podtržené sekvence určují zvolenou restrikci endonukleázového místa, které lze použít pro uvolnění polynukleotidu z připojení na pevný nosič biotinem, pokud se používá. Pro výše popsaný primer by mohl být komplement připojený na mikročástici ve formě



Po reverzní transkripci se mRNA odstraní například zpracováním RNázou H a druhé vlákno cDNA se syntetizuje například s použitím primeru v následující formě (číslo identifikace sekvence: 3):

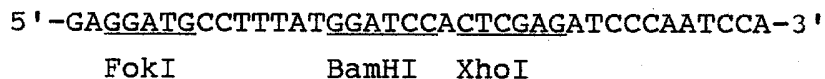


kde N je jeden z A, T, G nebo C, R je nukleotid obsahující purin a Y je nukleotid obsahující pyrimidin. Tento daný primer vytváří restrikční místo Bst YI ve výsledné dvojvláknové DNA, které spolu s místem Sal I umožňuje klonování do vektoru například s místy Bam HI a Xho I. Po zpracování Bst YI a Sal I by byl tento příklad konjugátu ve formě:



S konjugáty polynukleotid-značka lze potom zacházet pomocí standardních způsobů molekulární biologie. Například výše popsaný konjugát, který je ve skutečnosti směsí, se může nanést na komerčně dostupné klonovací vektory, například Stratagene Cloning System (La Jolla, CA), přenést na hostitele, jako jsou komerčně dostupné hostitelské bakterie, který se potom pěstuje pro zvýšení množství konjugátů. Klonovací vektory se poté mohou izolovat s použitím standardních způsobů, například Sambrook a kol., Molecular Cloning, 2. vydání (Cold Spring Harbor Laboratory, New York, 1989). Alternativně lze použít příslušné adaptory a primery takovým způsobem, že populaci konjugátů lze zvětšit pomocí PCR.

Přednostně, pokud se použije způsob sekvenční analýzy založený na ligáze, klonují se fragmenty zpracované Bst YI a Sal I do vektoru zpracovaného Bam HI-/Xho I majícího následující "single-copy" restrikční místa:



To přidává místo Fok I, což umožňuje iniciaci sekvenovacího procesu, který se podrobněji diskutuje níže.

Značky mohou být konjugovány s cDNA existujících knihoven standardními klonovacími způsoby. cDNA se vyříznou ze svého existujícího vektoru, izolují a naváží na vektor obsahující repertoár značek. Přednostně je vektor obsahující značku linearizován štěpením dvěma restrikčními enzymy, takže vyříznuté cDNA se mohou navázat v předem určené orien-



taci. Koncentrace linearizovaného vektoru obsahujícího značku je v podstatném přebytku nad linearizací vložek cDNA, takže ligace zajišťuje inherentní odebrání vzorků značek.

Obecný způsob expozice jednovláknové značky po amplifikaci zahrnuje zpracování konjugátu obsahujícího cílový polynukleotid s exonukleázovou aktivitou 5'--3' T4 DNA polymerázy nebo podobného enzymu. Při použití v přítomnosti jednotlivého deoxynukleosidtrifosfatu bude taková polymeráza štěpit nukleotidy od 3' ustupujících konců přítomných na netemplátovém vlákně dvojláknového fragmentu až do dosažení jednotlivého deoxynukleosidtrifosfatu na templátovém vlákně. Když se dosáhne tohoto nukleotidu, reakce 5'--3' ve skutečnosti zanikne, neboť aktivita polymerázy přidává nukleotidy rychleji než je odstraňuje aktivita působící excisi. Následkem toho se jednovláknové značky vytvořené třemi nukleotidy snadno připraví pro nanesení na pevné nosiče.

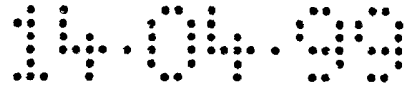
Tento způsob se může též použít pro preferenční metylaci vnitřních míst Fok I cílového polynukleotidu s ponecháním jednotlivého místa Fok I na konci polynukleotidu nemethylovaného. Nejprve se místo Fok I učiní jednovláknovým použitím polymerázy s deoxycyditrifosfatem. Dvojláknová část fragmentu se poté methyloje, načež se jednovláknový konec obsadí DNA polymerázou v přítomnosti všech čtyř nukleosidtrifosfátů, čímž se regeneruje místo Fok I. Tento způsob lze zřejmě zobecnit pro jiné endonukleázy než Fok I.

Po přípravě oligonukleotidových značek pro specifickou hybridizaci, například tak, že se učiní jednovláknovými, jak se popisuje výše, se smísí polynukleotidy s mikročásticemi obsahujícími komplementární sekvence značek za podmíněk, které dávají přednost vytváření dokonale spárovaných



duplexů mezi značkami a jejich komplementy. V literatuře existují rozsáhlá vodítka pro vytvoření těchto podmínek. Příklady odkazů poskytující tato vodítka jsou Wetmur, *Critical Reviews in Biochemistry and Molecular Biology*, 26, 227-259 (1991), Sambrook a kol., *Molecular Cloning: A Laboratory Manual*, 2. vydání (Cold Spring Harbor Laboratory, New York, 1989) a podobně. Přednostně jsou hybridizační podmínky dostatečně přísné tak, aby se stabilní duplexy vytvořily pouze z dokonale spárovaných sekvencí. Za těchto podmínek se mohou polynukleotidy specificky hybridizované svými značkami navazovat na komplementární sekvence připojené k mikročásticím. Konečně se mikročástice promyjí pro odstranění polynukleotidů s nenavázanými a/nebo chybně spárovanými značkami.

Když se jako nosiče pro syntézu používají mikročástice CPG, je hustota komplementů značek na povrchu mikročástic obvykle vyšší, než je třeba pro některé sekvenovací operace. To znamená, že ve způsobech sekvenování, které vyžadují následné zpracování připojených polynukleotidů řadou enzymů, se mohou hustě umístěné polynukleotidy snažit inhibovat přístup relativně velkých enzymů k polynukleotidům. V těchto případech se polynukleotidy přednostně mísí s mikročásticemi takovým způsobem, že jsou komplementy značek přítomny ve významném přebytku, například od 10:1 do 100:1 či více oproti polynukleotidům. Tím se zajišťuje, že hustota polynukleotidů na povrchu mikročástic není tak vysoká, aby inhibovala přístup enzymů. Přednostně je průměrná interpolynukleotidová vzdálenost na povrchu mikročástice řádu 30 až 100 nm. Vodítka při volbě poměrů pro standardní nosiče CPG a Bollootiniho kuličky (typ pevného skleněného nosiče) lze nalézt v Maskos a Southern, *Nucleic Acid Research*, 20, 1679-1684 (1992). Přednostně se pro sekvenční aplikace nanáší na standardní kuličky CPG o průměru 20 až 50 μm okolo 10^5 polynukleotidů



a na glycidalmethakrylatové (GMA) kuličky, které lze obdržet od Bangs Laboratories (Carmel, IN), o průměru 5 až 10 μm se nanáší několik desítek tisíc polynukleotidů, například 4×10^4 až 6×10^4 .

V preferovaném ztělesnění se komplementy značek pro třídění syntetizují na mikročasticích kombinatoricky, takže se na konci syntézy obdrží komplexní směs mikročastic, ze které se odebírá vzorek pro obsazení cílových polynukleotidů. Rozměr vzorku mikročastic bude záviset na několika faktorech včetně rozměru repertoáru komplementů značek, povahy přístroje použitého pro pozorování obsazených mikročastic, například jeho kapacity, tolerance na vícečetné kopie mikročastic se stejným komplementem značky (to jest "dvojice kuliček") a podobně. Následující tabulka poskytuje vodítko ohledně rozměru vzorku mikročastic, průměru mikročastic a přibližných fyzikálních rozměrů uspořádání mikročastic různých rozměrů.



| | | | | |
|--|-----------------|------------------|--------------------|------------------|
| Průměr | | | | |
| mikročástice | 5 μm | 10 μm | 20 μm | 40 μm |
| Maximální počet nukleotidů na- nesených při 1 na 10^{-15}m^2 | | 3×10^5 | $1,26 \times 10^6$ | 5×10^6 |
| Přibližná plocha monovrstvy 10^6 mikročástic | 0,45x0,45 cm | 1x1 cm | 2x2 cm | 4x4 cm |

Pravděpodobnost, že vzorek mikročástic obsahuje daný komple-
ment značky, nebo je přítomen ve více kopiích, lze popsat
Poissonovou distribucí, jak ukazuje následující tabulka.

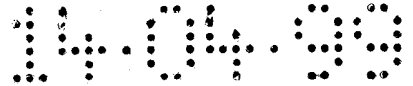


Tabulka VI

| Počet mikro- částic ve vzor- ku (jako zlomek rozměru reper- toáru), m | Zlomek reper- toáru komple- mentů značek přítomných ve vzorku, $1-e^{-m}$ | Zlomek mikročás- tic ve vzor- ku s připo- jeným jedi- nečným komp- lementem značky $m(e^{-m})/2$ | Zlomek mikro- částic ve vzorku nesou- cích stejný komplement jako jedna ji- ná mikročástice ve vzorku ("dvo- jice kuliček") $m^2(e^{-m})/2$ |
|---|---|--|--|
| 1,000 | 0,63 | 0,37 | 0,18 |
| 0,693 | 0,50 | 0,35 | 0,12 |
| 0,405 | 0,33 | 0,27 | 0,05 |
| 0,285 | 0,25 | 0,21 | 0,03 |
| 0,223 | 0,20 | 0,18 | 0,02 |
| 0,105 | 0,10 | 0,09 | 0,005 |
| 0,010 | 0,01 | 0,01 | |

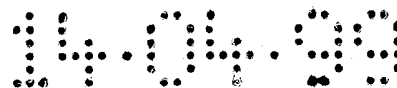
Vysoce specifické třídění a "panning"

Kinetika třídění závisí na rychlosti hybridizace oligonukleotidových značek s jejich komplementy značek, která opět závisí na komplexitě značek v hybridizační reakci. Proto existuje kompromis mezi rychlostí třídění a komplexitou značek takový, že vzrůst rychlosti třídění lze dosáhnout na účet snížení komplexity značek účastnících se hybridizační reakce. Jak se vykládá níže, účinek tohoto vztahu lze zlepšit pomocí "panningu".



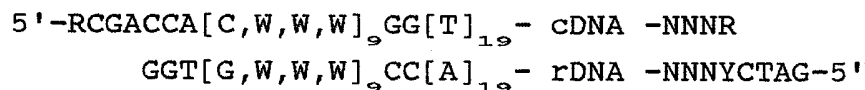
Specifičnost hybridizací se může zvýšit odebráním dostatečně malého vzorku tak, aby bylo vysoké procento jedinečných značek ve vzorku a aby se nejbližší sousední nebo v podstatě všechny značky ve vzorku lišily alespoň o dvě slova. Tato poslední podmínka se může splnit odebráním vzorku obsahujícího počet konjugátů značka-polynukleotid okolo 0,1 % nebo méně velikosti použitého repertoáru. Jestliže jsou například značky vytvořeny osmi slovy zvolenými z tabulky II, vznikne repertoár 8^8 nebo $1,67 \times 10^7$ značek a komplementů značek. V knihovně značka-konjugáty cDNA, jak se popisuje výše, znamená 0,1% vzorek přítomnost okolo 16 700 různých značek. Jestliže se nanese přímo na repertoárový ekvivalent mikročastic nebo v tomto případě na vzorek $1,67 \times 10^7$ mikročastic, potom se obsadí pouze malá podmnožina mikročastic. Hustota obsazených mikročastic se může zvýšit, například pro účinnější sekvenování provedením kroku "panningu" při kterém se použijí konjugáty značka-cDNA pro separaci obsazených částic od neobsazených. Proto v příkladu popsáném výše, i když "0,1 procentní" vzorek obsahuje pouze 16 700 cDNA, kroky odebrání vzorku a panningu se mohou opakovat, až se nahromadí tolik mikročastic, kolik je třeba. Alternativně se mohou obsazené mikročastice oddělit od neobsazených mikročastic přístrojem pro fluorescenčně aktivované třídění buněk (FACS) s použitím konvenčních způsobů, například se mohou konjugáty značka-cDNA fluorescenčně označit způsobem popsáným níže zajištěním fluorescenčně označeného pravého primeru. Po nanesení a třídění FACS může být označení odštěpeno před ligací kódovaných adaptorů, například pomocí Dpn I nebo podobného enzymu, který rozpoznává methylovaná místa.

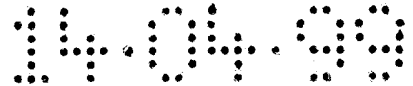
Krok panningu lze provést zajištěním vzorku konjugátů



značka-cDNA, z nichž každý obsahuje záchyťový zbytek na konci opačném či distálním vůči oligonukleotidové značce. Přednostně je záchyťový zbytek zbytkem takového typu, který se může z konjugátů značka-cDNA uvolnit, takže konjugáty značka-cDNA se mohou sekvenovat způsobem jednobázového sekvenování. Tyto zbytky mohou obsahovat biotin, digoxigenin nebo podobné ligandy, oblast vazby triplexu či podobně. Přednostně obsahuje tento zbytek biotinovou složku. Biotin se může připojit ke konjugátům značka-cDNA řadou standardních způsobů. Jestliže jsou na konjugáty značka-cDNA připojeny vhodné adaptory obsahující místa vazby primeru PCR, může se biotin připojit použitím biotinylového primeru při amplifikaci po odebrání vzorku. Alternativně, pokud jsou konjugáty značka-cDNA vložky klonovacích vektorů, může se biotin připojit po excizi konjugátů značka-cDNA digescí s příslušným restričním enzymem s následnou izolací a nanesením na vyčnívající vlákno distální vůči značkám pomocí DNA polymerázy za přítomnosti biotinylovaného uridintrifosfatu.

Po zachycení konjugátu značka-cDNA může být tento konjugát uvolněn z biotinového zbytku řadou způsobů, jako je chemická vazba, která se štěpí redukcí, například Herman a kol., *Anal. Biochem.*, 156, 48-55 (1986) nebo se štěpí fotochemicky, například Olejnik a kol., *Nucleic Acids Research*, 24, 361-366 (1996) nebo se štěpí enzymaticky zavedením restričního místa v primeru PCR. Poslední ztělesnění se může uvést na příkladu úvahou výše popsané knihovny konjugátů značka-polynukleotid:





Následující adaptory se mohou navazovat na konce těchto fragmentů pro umožnění amplifikace pomocí PCR:

5'-XXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXYGAT

Pravý adaptor

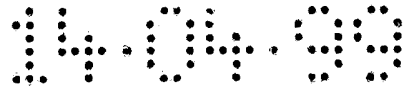
GATCZZACTAGTZZZZZZZZZZZZ-3'
ZZTGATCAZZZZZZZZZZZZ

Levý adaptor

ZZTGATCAZZZZZZZZZZZZ-5'-biotin

Levý primer

kde "ACTAGT" je rozpoznávací místo Spe I (zanechávající střídavě uspořádané štěpení připravené pro jednobázové sekvencování) a X a Z jsou nukleotidy zvolené tak, že teploty anelování a disociace příslušných primerů jsou zhruba stejné. Po ligaci adaptorů a amplifikaci pomocí PCR s použitím biotinylovaného primeru se konjugáty učiní jednovláknovými exonukleázovou aktivitou T4 DNA polymerázy a kombinují se se vzorkem mikročástic, například s ekvivalentem repertoáru, s připojenými komplementy značek. Po anelování za přísných podmínek (pro minimalizaci chybného spárování značek) se konjugáty preferovaně navazují na své komplementy značek a obsazené mikročástice se oddělují od neobsazených mikročástic záchytem na avidinizovaných magnetických kuličkách nebo podobným záchyťovým způsobem.



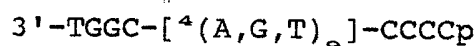
Při návratu úvah zpět ke vzorku tento způsob vede k akumulaci zhruba 10 500 (16 700 x 0,63) obsazených mikročástic s různými značkami, které se mohou uvolnit z magnetických kuliček štěpením pomocí Spe I. Opakováním tohoto způsobu 40x až 50x s novými vzorky mikročástic a konjugátů značka-cDNA lze akumulovat 4 až 5×10^5 cDNA půlováním uvolněných mikročástic. Půlované mikročástice mohou být poté současně sekvenovány jednobázovým sekvenovacím způsobem.

Požadavek, kolikrát je třeba opakovat kroky odběru vzorků a panningu, nebo obecně určení, kolik cDNA se má analyzovat, závisí na daném záměru. Jestliže je záměrem monitorovat změny ve výskytu relativně běžných sekvencí, činících například 5 nebo více procent populace, potom mohou relativně malé vzorky, to jest malá frakce celkového rozměru populace, umožnit statisticky významné odhady relativních výskytů. Na druhé straně, jestliže je třeba monitorovat výskyty vzácných sekvencí, činících například do 0,1 % nebo méně populace, potom se požadují větší vzorky. Obecně existuje přímý vztah mezi velikostí vzorku a spolehlivostí odhadů relativních výskytů založené na tomto vzorku. V literatuře existují rozsáhlé návody ke stanovení správných rozměrů vzorku pro spolehlivé statistické odhady, například Koller a kol., *Nucleic Acids Research*, 23, 185-191 (1994), Good, *Biometrika*, 40, 16-264 (1953), Bunge a kol., *J. Am. Stat. Assoc.*, 88, 364-373 (1993) a podobně. Přednostně se pro monitorování změn genové exprese na základě analýzy řady knihoven cDNA obsahujících 10^5 až 10^8 nezávislých klonů $3,0$ až $3,5 \times 10^4$ různých sekvencí pro analýzu každé knihovny akumuluje vzorek o počtu sekvencí nejméně 10^4 . Vhodnější je nahromadění vzorku o počtu sekvencí nejméně 10^5 pro analýzu každé knihovny a nejvhodnější je nahromadění vzorku o počtu sekvencí 5×10^5 pro analýzu každé knihovny. Alternativně

počet odebraných sekvencí je přednostně dostatečný pro stanovení relativního výskytu přítomné sekvence při frekvenci v rozmezí 0,1 až 5 % s 95% mezí spolehlivosti nepřevyšující 0,1 % velikosti populace.

Tvorba knihovny značek

Příklad knihovny značek se vytvoří následujícím způsobem pro získání chemicky syntetizovaných značek o 9 slovech pro nukleotidy A, G a T definované vzorcem:



ve kterém "[⁴(A,G,T)₉]" značí směs značek, kde každá značka obsahuje devět 4-merních slov z A, G a T a "p" značí 5' fosfát. Tato směs se navazuje na následující vazebné oblasti pravého a levého primeru (číslo identifikace sekvence: 4 a 5):

5' - AGTGGCTGGGCATCGGACCG
TCACCGACCCGTAGCCp

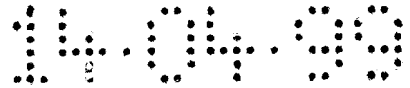
5' - GGGGCCCAGTCAGCGTCGAT
GGGTCAGTCGCAGCTA

levý

pravý

Vazebné oblasti pravého a levého primeru se navazují na výše popsanou směs značek a poté se jednovláknová část navázané struktury obsadí DNA polymerázou, a smísí s pravým a levým primerem ukázaným níže a amplifikuje pro poskytnutí knihovny značek.

Levý primer



5' - AGTGGCTGGGCATCGGACCG

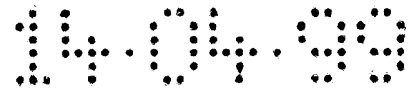
5' - AGTGGCTGGGCATCGGACCG- [^{4'}(A,G,T)₉]-GGGGCCCAGTCAGCGTCGAT
TCACCGACCCGTAGCCTGGC- [^{4'}(A,G,T)₉]-CCCCGGGTCAGTCGCAGCTA
CCCCGGGTCAGTCAGCAGCTA-5'

Pravý primer

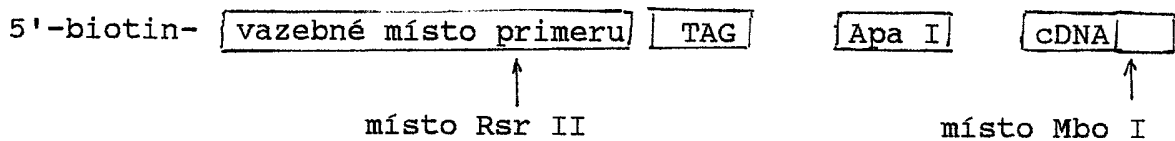
Podtržená část vazebné oblasti levého primeru značí rozpoznávací místo Rsr II. Nejlevější podtržená oblast vazebné oblasti pravého primeru značí rozpoznávací místa pro Bsp 120I, Apa I a Eco O 109I a místo štěpení pro Hga I. Nejpravější podtržená část vazebné oblasti pravého primeru značí rozpoznávací místo pro Hga I. Na základě volby lze levé či pravé primery syntetizovat s připojeným biotinem (s užitím konvenčních činidel dostupných například od Clontech Laboratories, Palo Alto, CA) pro umožnění purifikace po amplifikaci a/nebo štěpení.

Tvorba plasmidové knihovny konjugátů značka-polynukleotid pro sekvenování "podpisu" cDNA s užitím kódovaných adaptorů

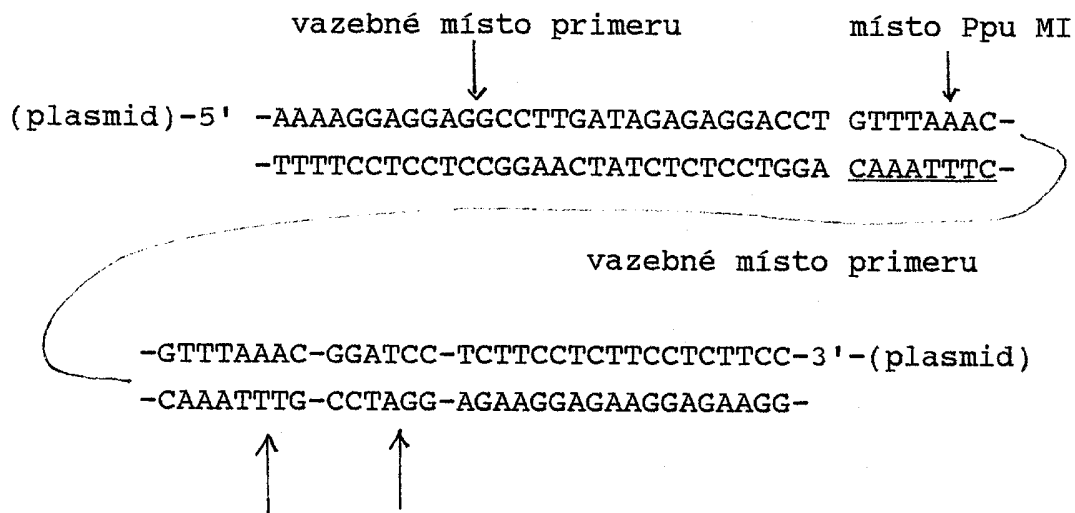
cDNA se připravuje ze vzorku mRNA konvenčními způsoby s použitím pGGCCCT₁₅ (A nebo G nebo C) jako primeru pro syntézu prvního vlákna zakotveného na rozhraní oblasti poly A ribonukleových kyselin mRNA a N₉ (A nebo T)GATC jako primeru pro syntézu druhého vlákna. To znamená, že oba jsou primery degenerované takovým způsobem, že primer druhého vlákna je přítomen ve dvou formách a primer prvního vlákna je přítomen ve třech formách. Sekvence GATC v primeru druhého vlákna odpovídá rozpoznávacímu místu Mbo I. Je možno použít též jiná čtyřbázová rozpoznávací místa, jako jsou rozpoznávací místa pro Bam H1, Sph I, Eco RI a podobně. Přítomnost

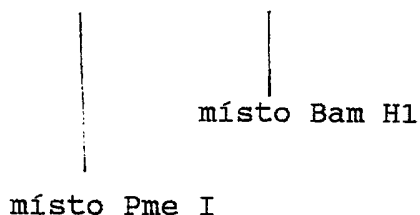


A a T přiléhajících k rozpoznávacímu místu primeru druhého vlákna zajišťuje, že lze použít reakce strippingu a výměny v dalším kroku pro tvorbu pětibázového převisu 5' "GGCCC". Primer prvního vlákna se aneluje se vzorkem mRNA a prodlužuje reverzní transkriptázou a poté se vlákno RNA degraduje aktivitou RNázy H reverzní transkriptázy s ponecháním jednovláknové cDNA. Primer druhého vlákna se aneluje a prodlužuje DNA polymerázou pomocí konvenčních způsobů. Po syntéze druhého vlákna se výsledné cDNA methylojí CpG methylázou (New Englan Biolabs, Beverly, MA) podle návodů výrobce. Vlákna 3' cDNA se potom odříznou výše popsanou strippingovou a výměnnou reakcí s použitím T4 DNA polymerázy za přítomnosti dATP a dTTP a poté se cDNA připojují na knihovnu značek popsanou výše se štěpením pomocí Hga I s poskytnutím:



Odděleně se vytváří klonovací vektor, například s výchozím použitím komerčně dostupného plasmidu, jako je Bluescript phagemid (Stratagene, La Jolla, CA) (identifikační číslo sekvence: 6)





Plasmid se štěpí Ppu MI a Pme I (s obdržením Rsr II kompatibilního konce a konce "flush", takže vložka je orientovaná) a methyluje se DAM methylázou. Struktura obsahující značku se štěpí pomocí Rsr II a poté se navazuje na otevřený plasmid a následně se konjugát štěpí Mbo I a Bam HI pro umožnění ligace a uzavření plasmidu. Plasmidy se potom amplifikují a izolují pro použití podle tohoto vynálezu.

Příklady provedení vynálezu

Příklad 1

Analýza cílového polynukleotidu amplifikovaného z pGEM7Z:
Identifikace nukleotidů cykly ligace a štěpení

V tomto příkladu se segment plasmidu pGEM7Z (Promega, Medison, WI) amplifikuje a připojuje na skleněné kuličky prostřednictvím dvojvláknového DNA linkeru, jehož jedno vlákno se syntetizuje přímo na kuličkách a proto je na ně kovalentně vázáno. Po přípravě konců cílového polynukleotidu pro ligaci na kódované adaptory v každém cyklu ligace a štěpení se směs kódovaných adaptorů (celkem 1024 různých adaptorů) aplikuje na cílové nukleotidy takovým způsobem, že se navazují pouze ty adaptory, jejichž vyčnívající vlákna vytvářejí dokonale spárované duplexy s cílovými polynukleotidy. Každý z 16 fluorescenčně označených komplementů značky se poté nanese na konjugáty polynukleotid-adaptor za podmínek umožňujících hybridizaci pouze správných komplementů



značek. Přítomnost či nepřítomnost fluorescenčního signálu po promytí ukazuje na přítomnost či nepřítomnost daného nukleotidu v daném místě. Protokol sekvenování podle tohoto příkladu je nyní použitelný pro větší množství cílových polynukleotidů tříděných na jednom nebo více pevných nosičích, jak popisuje Brenner, mezinárodní patentová přihláška PCT/US95/12791 a PCT/US96/09513 (WO 96/12041 a WO 96/41011).

47-merní oligonukleotid se syntetizuje přímo na Ballotiniho kuličkách (0,040 až 0,075 mm, Jencons Scientific, Bridgeville, PA) s použitím standardního návodu automatického DNA syntetizátoru. Komplementární vlákno 47-meru se syntetizuje odděleně a čistí pomocí vysokovýkonné kapalinové chromatografie. Při hybridizaci má výsledný duplex restriktivní Bdt XI na konci distálním vzhledem ke kuličce. Komplementární vlákno se hybridizuje s připojeným 47-merem v následující směsi: 25 μ l komplementárního vlákna při 200 pmol/ μ l, 20 mg Ballotiniho kuliček s 47-merem, 6 μ l restriktivního pufru New England Biolabs #3 (ze zásobního roztoku o 10-násobné koncentraci) a 25 μ l destilované vody. Tato směs se zahřívá na 93 °C a poté se pomalu ochladí na 55 °C a následně se přidá 40 jednotek Bst XI (10 jednotek/ μ l) s obdržením reakčního objemu 60 μ l. Tato směs se inkubuje při 55 °C po dobu 2 h a poté se kuličky promyjí 3x TE (pH 8,0).

Segment pGEM7Z, který se má připojit ke kuličkám, se připraví následujícím způsobem: Dva PCR primery se připraví standardními způsoby (číslo identifikace sekvence: 7 a číslo identifikace sekvence:8):

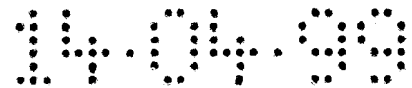
Primer 1: 5'-CTAAACCATTTGGTATGGGCCAGTGAATTGTAATA



Primer 2: 5'-CGCGCAGCCCGCATCGTTTATGCTACAGACTGTC-
AGTGCAGCTCTCCGATCCAAA

Reakční směs PCR obsahuje následující složky: 1 μ l pGEM7Z při 1 ng/ μ l, 10 μ l primeru 1 při 10 pmol/ μ l, 10 μ l primeru 2 při 10 pmol/ μ l, 10 μ l deoxyribonukleotidtrifosfátů při koncentraci 2,5 mM, 10 μ l 10násobného pufru PCR (Perkin-Elmer), 0,5 μ l Taq DNA polymerázy při 5 jednotkách/ μ l a 58 μ l destilované vody, což činí konečný objem 100 μ l. Reakční směs se podrobí 25 cyklům při 93 °C po dobu 30 s, při 60 °C po dobu 15 s a 72 °C po dobu 60 s s obdržením produktu o 172 párech bází, který se postupně zpracuje Bbv I (100 μ l reakční směsi PCR, 12 μ l 10-násobného pufru #1 New England Biolabs, 8 μ l Bbv I při 1 jednotce/ μ l se inkubuje při 37 °C po dobu 6 h) a Bst XI (k reakční směsi Bbv I se přidá 5 μ l 1 M roztoku chloridu sodného, 67 μ l destilované vody a 8 μ l Bst XI při 10 jednotkách/ μ l a výsledná směs se inkubuje při 55 °C po dobu 2 h).

Po průchodu výše popsané reakční směsi vířivým sloupcem Centricon 30 (Amicon, Inc.) podle návodu výrobce se fragment s restrikcí Bbv I/Bst XI navazuje na dvojlátkový linker připojený na Ballotiniho kuličky v následující směsi: 17 μ l fragmentu s restrikcí Bbv I/Bst XI (10 μ g), 10 μ l kuliček (20 mg), 6 ml 10-násobného ligačního pufru (New England Biolabs, uváděný níže jako NEB), 5 μ l T4 DNA ligázy při koncentraci 2000 jednotek/ μ l a 22 μ l destilované vody, tato směs se inkubuje při 25 °C po dobu 4 h a poté se kuličky promyjí třikrát TE (pH 8,0) s poskytnutím následujícího cílového polynukleotidu (číslo identifikace sekvence: 9) pro sekvenování majícího 5' fosfát:



[KULIČKY]--
 . . . AGCTACCCGATC
 . . . TCGATGGGCTAGATTTp-5'

5' fosfát se odstraní zpracováním směsi kuliček alkalickou fosfatázou, například z telecího střeva, kterou lze obdržet od New England Biolabs (Beverly, MA) s použitím návodu výrobce.

Horní vlákna z následujících 16 souborů 64 kódovaných adaptorů (číslo identifikace sekvence: 10 až číslo identifikace sekvence: 25) se syntetizují, každé odděleně, na automatizovaném syntetizátoru DNA (model 392 Applied Biosystems, Foster City) s použitím standardních způsobů. Spodní vlákno, které je stejné pro všechny adaptory, se syntetizuje odděleně a poté se hybridizuje s příslušnými vrchními vlákny:

| Číslo identifikace sekvence | Kódovaný adaptor |
|-----------------------------|--|
| 10 | 5'-pANNNTACAGCTGCATCCcttggcgctgagg pATGCACGCGTAGGG-5' |
| 11 | 5'-pNANNTACAGCTGCATCCctgggcctgtaag pATGCACGCGTAGGG-5' |
| 12 | 5'-pCNNNTACAGCTGCATCCcttgacgggtctc pATGCACGCGTAGGG-5' |
| 13 | 5'-pNCNNTACAGCTGCATCCctgcccgcacagt pATGCACGCGTAGGG-5' |

14.04.99

- 14 5'-pGNNNTACAGCTGCATCCCTtcgcctcggac
pATGCACGCGTAGGG-5'
- 15 5'-pNGNNTACAGCTGCATCCCTgatccgctagc
pATGCACGCGTAGGG-5'
- 16 5'-pTNNNTACAGCTGCATCCCTtccgaaccgac
pATGCACGCGTAGGG-5'
- 17 5'-pNTNNTACAGCTGCATCCCTgaggggatag
pATGCACGCGTAGGG-5'
- 18 5'-pNNANTACAGCTGCATCCCTtcccgtacac
pATGCACGCGTAGGG-5'
- 19 5'-pNNNATACAGCTGCATCCCTgactccccgag
pATGCACGCGTAGGG-5'
- 20 5'-pNNCNTACAGCTGCATCCCTgtggtgcgcg
pATGCACGCGTAGGG-5'
- 21 5'-pNNNCTACAGCTGCATCCCTctacagcagcg
pATGCACGCGTAGGG-5'
- 22 5'-pNNGNTACAGCTGCATCCCTgtcgcgctcgtt
pATGCACGCGTAGGG-5'
- 23 5'-pNNGTACAGCTGCATCCCTcggagcaacct
pATGCACGCGTAGGG-5'
- 24 5'-pNNTNTACAGCTGCATCCCTggtgaccgtag
pATGCACGCGTAGGG-5'
- 25 5'-pNNNTTACAGCTGCATCCCTccccctgctgga
pATGCACGCGTAGGG-5'



kde N a p se definují výše a nukleotidy označené malými písmeny jsou 12-merní oligonukleotidové značky. Každá značka se liší od každé jiné o 6 nukleotidů. Stejná molární množství každého adaptoru se kombinují v pufru NEB č. 2 (New England Biosciences, Beverly, MA) s vytvořením směsi při koncentraci 1000 pmol/ μ l.

Každý z 16 komplementů značek se syntetizuje odděle- ně jako aminoderivatizovaný oligonukleotid a každý se ozna- čuje fluoresceinovou molekulou (například FAM, NHS-esterem fluoresceinu od Molecular Probes, Eugene, OR), která se při- pojuje na konce 5' komplementu značky polyethylenglykolovým linkerem (Clontech Laboratories, Palo Alto, CA). Sekvence komplementů značek jsou jednoduše 12-merní komplementy zna- ček ukázaných výše.

Ligace adaptorů na cílový polynukleotid se provádí ve směsi obsahující 5 μ l kuliček (20 mg), 3 μ l 10-násobného li- gázového pufru, 5 μ l směsi adaptoru (25 nM), 2,5 μ l NEB T4 DNA ligázy (2000 jednotek/ μ l) a 14,5 μ l destilované vody. Tato směs se inkubuje při 16 °C po dobu 30 min a poté se ku- ličky promyjí 3x TE (pH 8,0).

Po centrifugaci a odstranění TE se 3' fosfáty naváza- ných adaptorů odstraní zpracováním směsi polynukleotid-kuličky telecí střevní alkalickou fosfatázou (CIP) (New England Biolabs, Beverly, MA) s použitím návodu výrobce. Po odstranění 3' fosfátů se může CIP inaktivovat proteolytickou digescí, například s použitím PronázyTM (od Boeringer Mannheim, Indianapolis, IN) nebo ekvivalentní pro- teázy podle návodu výrobce. Směs polynukleotid-kuličky se poté promyje a zpracuje směsí T4 polynukleotid kinázy a T4 DNA ligázy (New England Biolabs, Beverly, MA) po přidání 5'



fosfátu do mezery mezi cílový polynukleotid a adaptor k ukončení ligace adaptorů na cílový polynukleotid. Směs kuličky-polynukleotid se poté promyje TE.

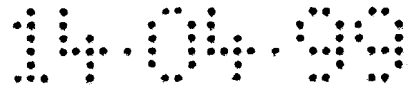
Každý označený komplement značky se odděleně aplikuje na směs polynukleotid-kuličky za podmínek umožňujících vytvoření dokonale spárovaných duplexů pouze mezi oligonukleotidovými značkami a jejich příslušnými komplementy a poté se směs promyje za přísných podmínek a měří se přítomnost či nepřítomnost fluorescenčního signálu. Komplementy se aplikují v roztoku obsahujícím komplement značky o koncentraci 25 nM, chlorid sodný o koncentraci 50 mM, hořčík o koncentraci 3 mM, Tris-HCl o koncentraci 10 mM (pH 8,5) při teplotě 20 °C, inkubují se po dobu 10 min a poté se promyjí stejným roztokem (bez komplementu značky) po dobu 10 min při 55 °C.

Po identifikaci čtyř nukleotidů, jak se popisuje výše, se odštěpí kódované adaptory z polynukleotidů působením Bbv I s použitím návodu výrobce. Po počáteční ligaci a identifikaci se cyklus ligace, identifikace a štěpení opakuje třikrát pro obdržení sekvence 16 terminálních nukleotidů cílového polynukleotidu. Obr. 4 ilustruje relativní fluorescenci od každého ze čtyř komplementů značek aplikovaného pro identifikaci nukleotidů v polohách 5 až 16 (z nejdistanějšího konce vzhledem ke kuličce k nejproximálnějšímu konci směrem ke kuličce).

Příklad 2

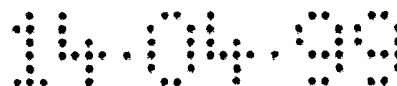
Tvorba a třídění knihovny cDNA pro sekvenování "podpisu" kódovanými adaptory

V tomto příkladu se vytváří knihovna cDNA, ve které



se na každou cDNA připojuje oligonukleotidová značka obsahující 8 čtyřnukleotidových "slov". Jak se popisuje výše, je repertoár oligonukleotidových značek tohoto rozměru dostatečně velký (okolo 10^8), takže pokud se cDNA syntetizují z populace zhruba 10^6 mRNA, potom je vysoká pravděpodobnost, že každá cDNA bude mít pro třídění jedinečnou značku. Po extrakci mRNA se provede syntéza prvního vlákna za přítomnosti 5-Me-dCTP (pro blokování určitých restričních míst cDNA) a biotinylované směsi primerů obsahující oligonukleotidové značky. Po konvenční syntéze druhého vlákna se konjugáty značka-cDNA štěpí Dpn II (který není ovlivňován 5-Me-deoxycytosiny), biotinylované podíly se oddělí z reakční směsi s použitím magnetických kuliček potažených streptavidinem a konjugáty tag-cDNA se získají jejich odštěpením od magnetických kuliček přes místo Bsm BI přenášené biotinylovaným primerem. Fragment Bsm BI-Dpn II obsahující konjugát značka-cDNA se poté vloží do plasmidu a amplifikuje. Po izolaci plasmidu se konjugáty značka-cDNA amplifikují z plasmidu pomocí PCR za přítomnosti 5-Me-dCTP s použitím biotinylovaných a fluorescenčně značených primerů obsahujících předem definovaná restriční endonukleázová místa. Po afinitní purifikaci magnetických kuliček potažených streptavidinem se konjugáty značka-cDNA odštěpí od kuliček, zpracují T4 DNA polymerázou za přítomnosti dGTP, aby byly značky učiněny jednovláknovými a poté se kombinují s repertoárem kuliček GMA s připojenými komplementy značek. Po přísné hybridizaci a ligaci se kuličky GMA třídí pomocí FACS pro obdržení obohacené populace kuliček GMA obsazených cDNA. Obohacená populace obsazených kuliček GMA se imobilizuje v dvojrozměrném uspořádání v průtokové komoře, kde nastává sekvence báze po bázi s použitím kódovaných adaptorů.

Zhruba 5 μ g poly(A⁺) mRNA se extrahuje z kvasinek



DBY746 s použitím konvenčních návodů. Syntéza prvního a druhého vlákna cDNA se provede kombinací 100 až 150 molů následujícího primeru (číslo identifikace sekvence: 26):

5'-biotin-ACTAATCGTCTCACTATTTAATTAA [W,W,W,G]₈CC(T)₁₈V-3'

=====

s poly(A+)mRNA s použitím soupravy pro syntézu cDNA Strata-gene (La Jolla, CA) cDNA Synthesis Kit podle návodu výrobce. To vede k cDNA, jejichž deoxycytosiny prvního vlákna jsou methylované v poloze uhlíku 5. Ve vzorci znázorněném výše je "V" G, C nebo A, "[W,W,W,G]" je čtyřnukleotidové slovo zvolené z tabulky II, jak se popisuje výše, jednou podtržená část je rozpoznávací místo Bsm BI a dvakrát podtržená část je rozpoznávací místo Pac I. Po frakcionaci rozměrů (souprava GIBCO-BRL cDNA Size Fractionation Kit) s použitím konvenčních návodů se cDNA zpracuje Dpn II (New England Bioscience, Beverly, MA) s použitím návodu výrobce a provede se afinitní čištění s magnetickými kuličkami potaženými streptavidinem (kuličky M-280, Dynal A.S., Oslo, Norway). DNA zachycená kuličkami se zpracuje Bsm BI pro uvolnění konjugátu značka-cDNA pro klonování do modifikovaného vektoru pBCSK⁻ (Stratagene, La Jolla, CA) s použitím standardních návodů. Vektor pBCSK⁻ se modifikuje přidáním místa Bbs I vložením následujícího fragmentu (číslo identifikace sekvence: 27) do vektoru zpracovaného Kpn I/Eco RV.

CGAAGACCC

3'-CATGGCTTCTGGGGATA-5'

konjugát značka-cDNA zpracovaný Bsm BI/Dpn II se vkládá do pBCSK⁻, který byl předběžně zpracován Bbs I a Bam HI. Po ligaci se vektor přenesse na hostitele doporučeného výrobcem pro amplifikaci.



Po izolaci vektoru pBCSK⁻ ze standardního plasmidového minipreparátu se konjugáty značka-cDNA amplifikují PCR v přítomnosti 5-Me-dCTP s použitím 20-merních primerů komplementárních vektorových sekvencí připojených na konce vložené části značka-cDNA. Primer umístěný "směrem vzhůru", to jest přiléhající ke značce, se biotinyluje a primer umístěný "směrem dolů", to jest přilehlý k cDNA, se označí fluoresceinem. Po amplifikaci se provede afinitní čištění PCR produktu a poté štěpení pomocí Pac I pro uvolnění fluorescenčně označených konjugátů značka-cDNA. Značky konjugátů jsou učiněny jednovláknovými jejich zpracováním T4 DNA polymerázou za přítomnosti dGTP. Po ukončení reakce se konjugát značka-cDNA čistí extrakcí směsí fenol-chloroform a spojí s kuličkami GMA 5,5 mm nesoucími komplementy značky, kde každá značka má 5'fosfát. Hybridizace se provede za přísných podmínek za přítomnosti termostabilní ligázy, takže se navazují pouze značky vytvářející dokonale spárované duplexy se svými komplementy. Kuličky GMA se promyjí a obsazené kuličky se koncentrují tříděním FACS s použitím fluorescenčně označených cDNA pro identifikaci obsazených kuliček GMA. Konjugáty značka-cDNA připojené ke kuličkám GMA se zpracují Dpn II pro odstranění fluorescenčního označení a dále alkalickou fosfatázou pro přípravu cDNA pro sekvenování.

Následující adaptor štěpení (číslo identifikace sekvence: 28) se navazuje na cDNA zpracované Dpn II fosfatázou:

5'-pGATCAGCTGCTGCAAATTT
pTCGACGACGTTTAAA

poté se odstraní 3'fosfát alkalickou fosfatázou, vlákno 5'



cdNA se zpracuje T4 DNA kinázou a naváže se zářez mezi adaptorem štěpení a cdNA. Po štěpení pomocí Bbv I se kódované adaptory z příkladu 1 navazují na konce cdNA, jak se popisuje výše.

Průtoková komora (500) schematicky znázorněná na obr. 5 se připraví leptáním dutiny mající vstup kapaliny (502) a výstup (504) ve skleněné desce (506) použitím standardních mikroobráběcích způsobů, například Ekstrom a kol., mezinárodní patentová přihláška PCT/SE91/00327 (WO 91/16966), Brown, US patent 4 911 782, Harrison a kol., Anal. Chem., 64, 1926-1932 (1992) a podobně. Rozměr průtokové komory (500) je takový, že obsazené mikročástice (508), například kuličky GMA, mohou být uspořádány v dutině (510) do těsně uložené monovrstvy o 100 až 200 tisících kuliček. Dutina (510) se připraví v uzavřené komoře se vstupem a výstupem anodickým navázáním krycího sklíčka (512) na naleptanou skleněnou desku (506), například Pomerantz, US patent 3 397 279. Činidla se odměřují do průtokové komory pomocí čerpadel s injekční stříkačkou (514 až 520) ventilovým blokem (522) ovládaným mikroprocesorem, který se běžně užívá v automatizovaných syntetizátorech DNA a peptidů, například Bridgham a kol., US patent 4 668 479, Hood a kol., US patent 4 252 769, Barstow a kol., US patent 5 203 368, Hunkapiller, US patent 4 703 913 a podobně.

Tři cykly navázání, identifikace a štěpení se provedou v průtokové komoře (500) s obdržáním sekvencí 12 nukleotidů na koncích každé z okolo 100000 cdNA. Nukleotidy cdNA se identifikují hybridizací komplementů značek s kódovanými adaptory jak se popisuje v příkladu 1. Specificky hybridizované komplementy značek se detekují excitací jejich fluorescenčních označení osvětlovacím paprskem (524) ze světelného



zdroje (526), kterým může být laser, rtuťová lampa či podobně. Osvětlovací svazek (524) prochází filtrem (528) a excituje fluorescenční označení na komplementech značek specificky hybridizovaných s kódovanými adaptory v průtokové komoře (500). Výsledná fluorescence (530) se sbírá konfokálním mikroskopem (532), prochází filtrem (534) a směřuje na kameru CCD (536), která vytváří elektronické zobrazení uspořádání kuliček pro zpracování a analýzu pracovní stanicí (538). Přednostně jsou cDNA po každém kroku navázání a štěpení zpracovány PronázouTM nebo podobným enzymem. Kódované adaptory a T4 DNA ligáza (Promega, Madison, WI) při zhruba 0,75 jednotek na μl procházejí průtokovou komorou při průtoku okolo 1 až 2 $\mu\text{l}/\text{min}$ po dobu 20 až 30 min při 16 °C a poté se 3' fosfáty odstraní z adaptorů a cDNA se připraví pro navázání druhého vlákna průchodem směsí alkalické fosfatázy (New England Bioscience, Beverly, MA) při 0,02 jednotkách/ μl a T4 DNA kinázy (New England Bioscience, Beverly, MA) při 7 jednotkách na μl průtokovou kamerou při 37 °C při průtoku 1 až 2 $\mu\text{l}/\text{min}$ po dobu 15 až 20 min. Ligace se uskuteční T4 DNA ligázou (0,75 jednotek/ml, Promega) průtokovou komorou po dobu 20 až 30 min. Komplementy značek při koncentraci 25 nm procházejí průtokovou komorou při rychlosti průtoku 1 až 2 $\mu\text{l}/\text{min}$ po dobu 10 min při 20 °C a poté se fluorescenční označení přenášená komplementy značek osvětlí a sbírá se fluorescence. Komplementy značek se oddělí od kódovaných adaptorů průchodem hybridizačního pufru proudovou komorou při rychlosti průtoku 1 až 2 $\mu\text{l}/\text{min}$ při 55 °C po dobu 10 min. Kódované adaptory se odštěpí od cDNA průchodem Bbv I (New England Biosciences, Beverly, MA) při koncentraci 1 jednotka/ μl a průtoku 1 až 2 $\mu\text{l}/\text{min}$ po dobu 20 min při 37 °C.



Dodatek I

Příklad počítačového programu pro tvorbu minimálně
zkříženě hybridizujících souborů (jednovláknová
značka/jednovláknový komplement)

Program tagN

c

c

c Program tagN vytváří minimálně zkříženě hybridizující
c soubory podjednotek s danou i) délkou podjednotky N a
c ii) počáteční sekvencí podjednotky. tagN předpokládá, že
c se ve značkách používají pouze 3 ze čtyř přirozených
c nukleotidů.

c

c

znak*1 podj1 (20)

celé číslo*2 msoub (10000,20), nbáze(20)

c

c

napište(*,*) "VLOŽTE DÉLKU PODJEDNOTKY"

čtěte (*,100)npodj

100 formátujte(i2)

c

c

napište(*,*) "VLOŽTE SEKVENCI PODJEDNOTKY"

čtěte(*,110) (podj1(k), k=1,npodj)

110 formátujte(20a1)

c

c

ndiff=10

c

c



```
c      Položte a=1 c=2 g=3 a t=4
c
c
provedte 800 kk=1,npodj
jestliže(podj1(kk).eq."a"), potom
    msoub(1,kk)=1
    endif
        jestliže(podj1(kk).eq."c"), potom
            msoub(1,kk)=2
            endif
                jestliže(podj1(kk).eq."g"), potom
                    msoub(1,kk)=3
                    endif
jestliže(podj1(kk).eq."t"), potom
    msoub(1,kk)=4
    endif
800 pokračujte
c
c
c      Vytvořte soubor podjednotek lišících se od
      podj1 nejméně o ndiff nukleotidů.
c
c
      jj=1

provedte 1000 k1=1,3
      provedte 1000 k2=1,3
        provedte 1000 k3=1,3
          provedte 1000 k4=1,3
            provedte 1000 k5=1,3
              provedte 1000 k6=1,3
                provedte 1000 k7=1,3
                  provedte 1000 k8=1,3
```

14.04.99

provedte 1000 k9=1,3
provedte 1000 k10=1,3
provedte 1000 k11=1,3
provedte 1000 k12=1,3
provedte 1000 k13=1,3
provedte 1000 k14=1,3
provedte 1000 k15=1,3
provedte 1000 k16=1,3
provedte 1000 k17=1,3
provedte 1000 k18=1,3
provedte 1000 k19=1,3
provedte 1000 k20=1,3

c

c

nbáze(1)=k1
nbáze(2)=k2
nbáze(3)=k3
nbáze(4)=k4
nbáze(5)=k5
nbáze(6)=k6
nbáze(7)=k7
nbáze(8)=k8
nbáze(9)=k9
nbáze(10)=k10
nbáze(11)=k11
nbáze(12)=k12
nbáze(13)=k13
nbáze(14)=k14
nbáze(15)=k15
nbáze(16)=k16
nbáze(17)=k17
nbáze(18)=k18
nbáze(19)=k19

```

                                nbáze(20)=k20
c
c
                                proved'te 1250 nn=1, jj
c
                                n=0
                                proved'te 1200 j=1,npodj
                                jestliže(msoub(nn,j).eq.1 .a. nbáze(j).ne.1 .nebo
1                                msoub(nn,j).eq.2 .a. nbáze(j).ne.2 .nebo
2                                msoub(nn,j).eq.3 .a. nbáze(j).ne.3 .nebo
3                                msoub(nn,j).eq.4 .a. nbáze(j).ne.4) .potom
                                n=n+1
                                endif
1200                                pokračujte
c
c
                                jestliže(n.lt.ndiff) potom
                                přejděte na 1000
                                endif
1250                                pokračujte
c
c
                                jj=jj+1
                                napište(*,130) (nbáze(i),i=1,npodj),jj
                                proved'te 1100 i=1,npodj
                                msoub(jj,i)=nbáze(i)
1100                                pokračujte
c
c
1000                                pokračujte
c
c
                                napište(*,*)
```



130 formátujte(10x,20(1x,i1),5x,i5)

napište(*,*)

napište(*,120) jj

120 formátujte(1x,"počet slov=",i5)

c

c

konec

c

c

c



Dodatek II

Příklad počítačového programu pro tvorbu minimálně
zkříženě hybridizujících souborů (dvojvláknová
značka/jednovláknový komplement značky)

Program 3tagN

```
c
c
c      Program 3tagN vytváří minimálně zkříženě
c      hybridizující soubory triplexních slov dané i)
c      délky N--podjednotky, ii) počáteční sekvenci
c      podjednotky a iii) identity nukleotidů
c      vytvářejících podjednotky, to jest zda podjednotky
c      obsahují všechny čtyři nukleotidy,
c      nebo nějaký podsoubor nukleotidů.
```

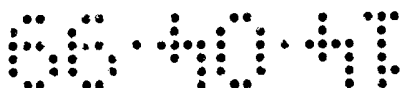
```
c
c      znak*1 podj1(20)
c      celé číslo*2 msoub(10000,20) nbáze(20)
```

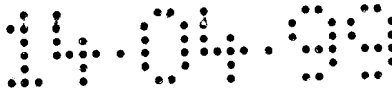
```
c
c      npodj=20
c      ndiff=6
```

```
c
c      napište(*,*) "VLOŽTE SEKVENCI PODJEDNOTKY: pouze a a g"
c      čtete (*,110)(podj1(k),k=1,npodj)
110 formátujte(20a1)
```

```
c
c
```

Vytvořte soubor slov lišících se od podj1 nejméně
o tři ndiff nukleotidů.





c
c
c
c
c

Přeložte a & g na čísla s a=1 & g=2

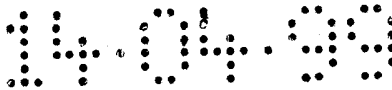
```
provedte 800 kk=1,npodj
jestliže(podj1(kk).eq."a"), potom
    msoub(1,kk)=1
endif
    jestliže(podj1(kk).eq."g"), potom
        msoub(1,kk)=2
    endif
```

800 pokračujte

c
c
c
c
c

jj=1

```
provedte 1000 k1=1,2
    provedte 1000 k2=1,2
        provedte 1000 k3=1,2
            provedte 1000 k4=1,2
                provedte 1000 k5=1,2
                    provedte 1000 k6=1,2
                        provedte 1000 k7=1,2
                            provedte 1000 k8=1,2
                                provedte 1000 k9=1,2
                                    provedte 1000 k10=1,2
provedte 1000 k11=1,2
    provedte 1000 k12=1,2
        provedte 1000 k13=1,2
            provedte 1000 k14=1,2
                provedte 1000 k15=1,2
```

proved'te 1000 k16=1,2

proved'te 1000 k17=1,2

proved'te 1000 k18=1,2

proved'te 1000 k19=1,2

proved'te 1000 k20=1,2

c
c

nbáze(1)=k1

nbáze(2)=k2

nbáze(3)=k3

nbáze(4)=k4

nbáze(5)=k5

nbáze(6)=k6

nbáze(7)=k7

nbáze(8)=k8

nbáze(9)=k9

nbáze(10)=k10

nbáze(11)=k11

nbáze(12)=k12

nbáze(13)=k13

nbáze(14)=k14

nbáze(15)=k15

nbáze(16)=k16

nbáze(17)=k17

nbáze(18)=k18

nbáze(19)=k19

nbáze(20)=k20

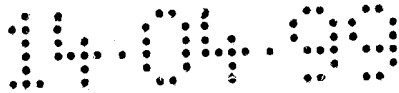
c
c

proved'te 1250 nn=1, jj

c

n=0

proved'te 1200 j=1,npodj



```

        jestliže(msoub(nn,j).eq.1 .a. nbáze(j).ne.1 .nebo
1          msoub(nn,j).eq.2 .a. nbáze(j).ne.2) .potom
          n=n+1
          endif
1200      pokračujte
c
c
        jestliže(n.lt.ndiff) potom
          přejděte na 1000
          endif
1250      pokračujte
c
c
          jj=jj+1
          napište(*,130) (nbáze(i),i=1,npodj),jj
          proveďte 1100 i=1,npodj
          msoub(jj,i)=nbáze(i)
1100      pokračujte
c
c
1000      pokračujte
c
c
          napište(*,*)
130      formátujte(5x,20(1x,i1),5x,i5)
          napište(*,*)
          napište(*,120) jj
120      formátujte(1x,"počet slov=",i5)
c
c
        konec
```



SEZNAM SEKVENCÍ

- (1) OBECNÉ INFORMACE:
 - (i) PŘIHLAŠOVATEL: Lynx Therapeutics, Inc.
 - (ii) NÁZEV VYNÁLEZU: Způsob sekvenční analýzy ligací kódovaného adaptoru
 - (iii) POČET SEKVENCÍ: 28
 - (iv) ADRESA PRO KORESPONDENCI:
 - (A) SPOLEČNÍCI: Dehlinger & Associates
 - (B) ULICE: 350 Cambridge Avenue, Suite 250
 - (C) MĚSTO: Palo Alto
 - (D) STÁT: CA
 - (E) ZEMĚ: USA
 - (F) POŠTOVNÍ KÓD: 94306
 - (v) FORMA PRO ČTENÍ NA POČÍTAČI:
 - (A) TYP MÉDIA: Floppy disk
 - (B) POČÍTAČ: IBM PC kompatibilní
 - (C) OPERAČNÍ SYSTÉM: PC-DOS/MS-DOS
 - (D) PROGRAMOVÉ VYBAVENÍ: PatentIn Release #1.0, verze #1.25
 - (vi) ÚDAJE O SOUČASNÉ PŘIHLÁŠCE:
 - (A) ČÍSLO PŘIHLÁŠKY:
 - (B) DATUM PODÁNÍ:
 - (vii) ÚDAJE O PRIORITNÍ PŘIHLÁŠCE:
 - (A) ČÍSLO PŘIHLÁŠKY: US 08/689 587
 - (B) DATUM PODÁNÍ: 12. srpna 1996



(vii) ÚDAJE O PRIORITYNÍ PŘIHLÁŠCE:

- (A) ČÍSLO PŘIHLÁŠKY: US 08/659 453
- (B) DATUM PODÁNÍ: 6. června 1996

(viii) ÚDAJE O ZÁSTUPCI:

- (A) JMÉNO: Vincent M. Powers
- (B) ČÍSLO REGISTRACE: 36 246
- (C) REFERENČNÍ ČÍSLO: 5525-0029.41/808-1wo

(ix) TELEKOMUNIKAČNÍ INFORMACE:

- (A) TELEFON: (415) 324-0880
- (B) FAX: (415) 324-0960

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 28 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 1:

TGGATTCTAG AGAGAGAGAG AGAGAGAG

28

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 2:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 16 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 2:



NNGGATGNNN NNNNNN

16

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 3:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 11 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: jedno
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 3

NRRGATCYNN N

11

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 4:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 20 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 4:

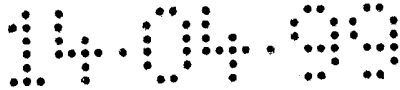
AGTGGCTGGG CATCGACCG

20

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 5:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 20 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě



(D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 5:

GGGGCCCAGT CAGCGTCGAT 20

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 6:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 70 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 6:

AAAAGGAGGA GGCCTTGATA GAGAGGACCT GTTTAAACGT TTAAACGGAT 50
CCTCTTCCTC TTCCTCTTCC 70

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 7:

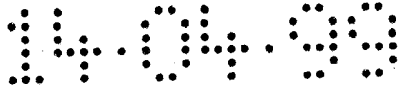
(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 34 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: jedno
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 7:

CTAAACCATT GGTATGGGCC AGTGAATTGT AATA 34

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 8:



(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 55 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: jedno
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 8:

CGCGCAGCCC GCATCGTTTA TGCTACAGAC TGTCAGTGCA 40
GCTCTCCGAT CCAAA 55

(2) INFORMACE PRO SEKVENCÍ IDENTIFIKAČNÍ ČÍSLO: 9:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 16 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 9:

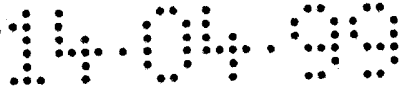
TCGATGGGCT AGATTT 16

(2) INFORMACE PRO SEKVENCÍ IDENTIFIKAČNÍ ČÍSLO: 10:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 10:



ANNNTACAGC TGCATCCCTT GCGCTGAGG

30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 11:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 11:

NANNTACAGC TGCATCCCTG GGCCTGTAAG

30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 12:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 12:

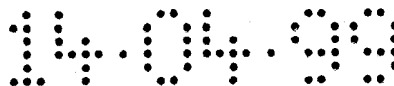
CNNNTACAGC TGCATCCCTT GACGGGTCTC

30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 13:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární



(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 13:

NCNNTACAGC TGCATCCCTG CCCGCACAGT 30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 14:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 14:

GNNNTACAGC TGCATCCCTT CGCCTCGGAC 30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 15:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 15:

NGNNTACAGC TGCATCCCTG ATCCGCTAGC 30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 16:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů



- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 16:

TNNNTACAGC TGCATCCCTT CCGAACCCGC 30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 17:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 17:

NTNNTACAGC TGCATCCCTG AGGGGGATAG 30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 18:

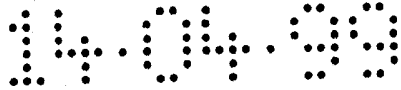
(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 18:

NNANTACAGC TGCATCCCTT CCCGCTACAC 30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 19:



(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 19:

NNNATACAGC TGCATCCCTG ACTCCCCGAG 30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 20:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 20:

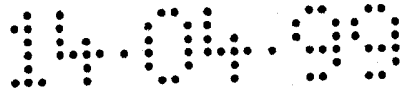
NNCNTACAGC TGCATCCCTG TGTTGCGCGG 30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 21:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 21:



NNNCTACAGC TGCATCCCTC TACAGCAGCG

30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 22:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 22:

NNGNTACAGC TGCATCCCTG TCGCGTCGTT

30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 23:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 23:

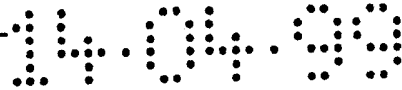
NNNGTACAGC TGCATCCCTC GGAGCAACCT

30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 24:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární



(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 24:

NNTNTACAGC TGCATCCCTG GTGACCGTAG 30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 25:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 30 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 25:

NNNTTACAGC TGCATCCCTC CCCTGTCGGA 30

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 26:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 78 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: jedno
- (D) TOPOLOGIE: lineární

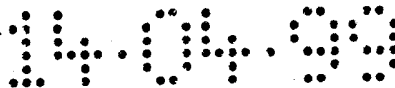
(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 26:

ACTAATCGTC TCACTATTTA ATTAANNNNN NNNNNNNNNN 40

NNNNNNNNNN NNNNNNGGT TTTTTTTTTT TTTTTTPTV 78

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 27:

(i) SEKVENČNÍ CHARAKTERISTIKY:



- (A) DÉLKA: 17 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 27:

ATAGGGGTCT TCGGTAC 17

(2) INFORMACE PRO SEKVENCI IDENTIFIKAČNÍ ČÍSLO: 28:

(i) SEKVENČNÍ CHARAKTERISTIKY:

- (A) DÉLKA: 19 nukleotidů
- (B) TYP: nukleová kyselina
- (C) POČET VLÁKEN: dvě
- (D) TOPOLOGIE: lineární

(xi) POPIS SEKVENCE: IDENTIFIKAČNÍ ČÍSLO SEKVENCE: 28:

GATCAGCTGC TGCAAATTT 19

P A T E N T O V É N Á R O K Y

1. Způsob stanovení nukleotidové sekvence na konci polynukleotidu, v y z n a č u j í c í s e t í m, že obsahuje následující kroky:

připojení jednoho či více kódovaných adaptorů na konec polynukleotidu, kde každý kódovaný adaptor má oligonukleotidovou značku zvolenou z minimálně zkříženě hybridizujícího souboru oligonukleotidů a vyčnívající vlákno komplementární části k vlákna polynukleotidu a

identifikaci jednoho či více nukleotidů v každé této části vlákna polynukleotidu specifickou hybridizací komplementu značky s každou oligonukleotidovou značkou jednoho či více kódovaných připojených adaptorů.

2. Způsob podle nároku 1, v y z n a č u j í c í s e t í m, že krok připojení zahrnuje připojení většího počtu různých kódovaných adaptorů k danému konci tohoto polynukleotidu, takže vyčnívající vlákna řady různých kódovaných adaptorů jsou komplementární s řadou různých částí tohoto vlákna daného polynukleotidu, takže existuje vztah jeden-na-jeden mezi různými kódovanými adaptory a různými částmi tohoto vlákna.

3. Způsob podle nároku 2, v y z n a č u j í c í s e t í m, že tyto různé části tohoto vlákna daného nukleotidu jsou navzájem přilehlé.

4. Způsob podle kteréhokoliv z nároků 1 až 3, v y z n a č u j í c í s e t í m, že toto vyčnívající vlákno těchto kódovaných adaptorů obsahuje od 2 do 6 nukleo-



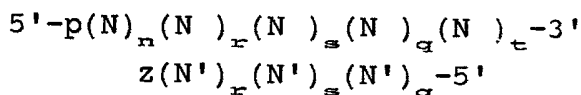
tidů a krok identifikace zahrnuje specifickou hybridizaci těchto komplementů značek s těmito oligonukleotidovými značkami, takže identita každého nukleotidu v těchto částech tohoto polynukleotidu se určuje postupně.

5. Způsob podle kteréhokoliv z nároků 1 až 4, v y z n a č u j í c í s e t í m, že tento způsob identifikace dále zahrnuje zajištění několika souborů komplementů značek ekvivalentních počtu nukleotidů, které se mají identifikovat v těchto částech tohoto polynukleotidu.

6. Způsob podle nároku 5, v y z n a č u j í c í s e t í m, že tento krok identifikace dále zahrnuje zajištění těchto komplementů značek v každém z těchto souborů, které jsou schopny ukázat přítomnost předem určeného nukleotidu signálem vytvořeným zbytkem generujícím fluorescenční signál a přitom pro každý typ nukleotidů existuje jiný zbytek generující fluorescenční signál.

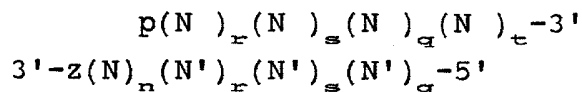
7. Způsob podle kteréhokoliv z nároků 1 až 6, v y z n a č u j í c í s e t í m, že tyto oligonukleotidové značky těchto kódovaných adaptorů jsou jednovláknové a tyto komplementy značek pro oligonukleotidové značky jsou jednovláknové, takže dochází ke specifické hybridizaci mezi oligonukleotidovou značkou a jejím příslušným komplementem značky Watsonovým-Crickovým párováním bází.

8. Způsob podle nároku 7, v y z n a č u j í c í s e t í m, že kódované adaptory jsou ve formě:





nebo

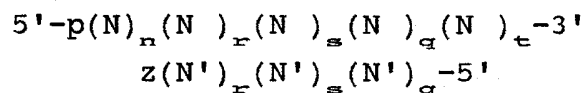


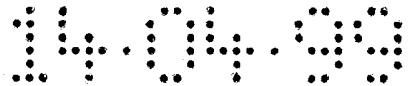
kde N je nukleotid a N' je jeho komplement, p je fosfátová skupina, z je 3' hydroxylová skupina nebo 3' blokující skupina, n je celé číslo od 2 do 6 včetně, r je celé číslo od 0 do 18 včetně, s je celé číslo buď od 4 do 6 včetně, pokud má kódovaný adaptor nukleázové rozpoznávací místo, nebo je rovno 0, pokud neexistuje nukleázové rozpoznávací číslo, q je celé číslo větší než nebo rovné 0 a t je celé číslo větší než nebo rovné 8.

9. Způsob podle nároku 8, v y z n a č u j í c í s e t í m, že r je mezi 0 a 12 včetně, t je celé číslo mezi 8 a 20 včetně a z je fosfátová skupina.

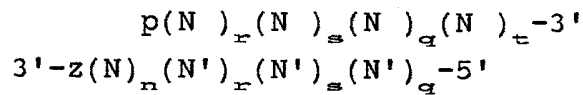
10. Způsob podle kteréhokoliv z nároků 1 až 6, v y z n a č u j í c í s e t í m, že tyto nukleotidové značky těchto kódovaných adaptorů jsou dvojitě vláknové a tyto komplementy značek s oligonukleotidovými značkami jsou jednovláknové, takže nastává specifická hybridizace mezi oligonukleotidovou značkou a jejím příslušným komplementem značky tvorbou Hoogsteenova nebo reverzního Hoogsteenova triplexu.

11. Způsob podle nároku 10, v y z n a č u j í c í s e t í m, že kódované adaptory jsou ve formě:





nebo



kde N je nukleotid a N' je jeho komplement, p je fosfátová skupina, z je 3' hydroxylová skupina nebo 3' blokující skupina, n je celé číslo od 2 do 6 včetně, r je celé číslo od 0 do 18 včetně, s je celé číslo buď od 4 do 6 včetně, pokud má kódovaný adaptor nukleázové rozpoznávací místo, nebo je rovno 0, pokud neexistuje nukleázové rozpoznávací místo, q je celé číslo větší než nebo rovné 0 a t je celé číslo větší než nebo rovné 8.

12. Způsob podle nároku 11, v y z n a č u j í c í s e t í m, že r je od 0 do 12 včetně, t je celé číslo od 12 do 24 včetně a z je fosfátová skupina.

13. Způsob podle kteréhokoliv z nároků 1 až 12, v y z n a č u j í c í s e t í m, že každý člen tohoto minimálně zkříženě hybridizujícího souboru se liší od každého jiného členu alespoň o 6 nukleotidů.

14. Způsob stanovení nukleotidových sekvencí většího počtu polynukleotidů, v y z n a č u j í c í s e t í m, že zahrnuje kroky:

(a) připojení první oligonukleotidové značky z repertoáru značek na každý polynukleotid v populaci polynukleotidů takovým způsobem, že se každá první oligonukleotidová značka z tohoto repertoáru volí z prvního minimálně zkříženě hybridizujícího souboru,



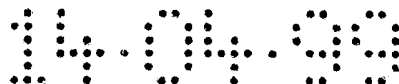
(b) odběr vzorku populace polynukleotidů s vytvořením vzorku polynukleotidů, takovým způsobem, že v podstatě všechny rozdílné polynukleotidy ve vzorku mají připojené rozdílné první oligonukleotidové značky,

(c) třídění polynukleotidů vzorku specifickou hybridizací prvních oligonukleotidových značek s jejich odpovídajícími komplementy, kde příslušné komplementy jsou připojené jako jednotné populace v podstatě identických oligonukleotidů v prostorově diskrétních oblastech na jednom nebo více pevných nosičích,

(d) připojení jednoho či více kódovaných adaptorů ke konci polynukleotidů ve vzorku, kde každý kódovaný adaptor má druhou oligonukleotidovou značku zvolenou z druhého minimálně zkříženě hybridizujícího souboru a vyčnívající vlákno komplementární k vyčnívajícímu vláknu polynukleotidu populace a

(e) identifikace většího počtu nukleotidů v těchto vyčnívajících vláknech polynukleotidů specifickou hybridizací komplementu značky s každou značkou druhého oligonukleotidu jednoho či více kódovaných adaptorů.

15. Způsob podle nároku 14, v y z n a č u j í c í s e t í m, že dále zahrnuje kroky (f) odštěpení těchto kódovaných adaptorů od těchto polynukleotidů nukleázou, která má nukleázové rozpoznávací místo oddělené od svého místa štěpení, takže se vytváří nové vyčnívající vlákno na tomto konci každého z těchto nukleotidů a (g) opakování kroků (d) až (f).



16. Způsob identifikace populace molekul mRNA,
v y z n a č u j í c í s e t í m, že obsahuje tyto kroky:

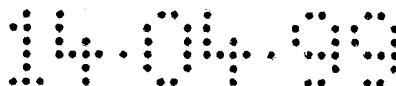
(a) vytvoření populace molekul cDNA z populace molekul mRNA takovým způsobem, že každá molekula cDNA má připojenou první oligonukleotidovou značku, která se volí z prvního minimálně zkříženě hybridizujícího souboru,

(b) odebrání vzorku populace molekul z cDNA pro vytvoření vzorku molekul cDNA takovým způsobem, že v podstatě všechny rozdílné molekuly cDNA mají připojené rozdílné první oligonukleotidové značky,

(c) třídění molekul cDNA specifickou hybridizací prvních oligonukleotidových značek s jejich příslušnými komplementy, které jsou připojené jako jednotné populace v podstatě identických komplementů v prostorově diskrétních oblastech na jednom či více pevných nosičích,

(d) připojení jednoho či více kódovaných adaptorů ke konci molekul cDNA v populaci takovým způsobem, že každý kódovaný adaptor má druhou oligonukleotidovou značku zvolenou z druhého minimálně zkříženě hybridizujícího souboru a vyčnívající vlákno komplementární k vyčnívajícímu vláknu molekuly cDNA vzorku a

(e) stanovení identity a pořadí většího počtu nukleotidů v každém z těchto vyčnívajících vláken molekul cDNA specifickou hybridizací komplementů značky s každou druhou oligonukleotidovou značkou jednoho či více kódovaných adaptorů, kde se populace molekul mRNA identifikuje frekvenční distribucí částí sekvencí molekul cDNA.



17. Způsob podle nároku 16, v y z n a č u j í c í s e t í m, že dále zahrnuje tyto kroky (f) odštěpení těchto kódovaných adaptorů od těchto polynukleotidů nukleázou mající nukleázové rozpoznávací místo oddělené od svého místa štěpení, takže se vytváří nové vyčnívající vlákno na tomto konci každé z těchto molekul cDNA a (g) opakování kroků (d) až (f).

18. Způsob stanovení nukleotidové sekvence na konci polynukleotidu, v y z n a č u j í c í s e t í m, že obsahuje následující kroky:

(a) navázání kódovaného adaptoru ke konci polynukleotidu, kde tento kódovaný adaptor má oligonukleotidovou značku zvolenou z minimálně zkřížené hybridizujícího souboru oligonukleotidů a vyčnívající vlákno komplementární k části vlákna polynukleotidu,

(b) identifikace jednoho či více nukleotidů v této části vlákna polynukleotidu specifickou hybridizací komplementu značky s oligonukleotidovou značkou navázaného kódovaného adaptoru,

(c) odštěpení kódovaného adaptoru od konce polynukleotidu nukleázou mající rozpoznávací nukleázové místo oddělené od svého místa štěpení, takže se na konci polynukleotidu vytváří nové vyčnívající vlákno a

(d) opakování kroků (a) až (c).

19. Způsob podle nároku 18, v y z n a č u j í c í s e t í m, že toto vyčnívající vlákno tohoto kódovaného adaptoru obsahuje od 2 do 6 nukleotidů a krok identifikace



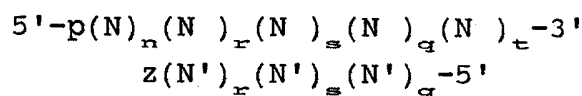
zahrnuje specifickou hybridizaci těchto postupných komple-
mentů značek s touto oligonukleotidovou značkou takovým způ-
sobem, že identita každého nukleotidu v této části tohoto
polynukleotidu se stanoví postupně.

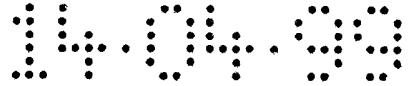
20. Způsob podle nároku 18 nebo 19,
v y z n a č u j í c í s e t í m, že tento krok identifi-
kace dále zahrnuje poskytnutí několika souborů komplementů
značek ekvivalentních počtu nukleotidů, které se mají iden-
tifikovat v dané části tohoto polynukleotidu.

21. Způsob podle nároku 20, v y z n a č u j í c í
s e t í m, že tento krok identifikace dále zahrnuje po-
skytnutí těchto komplementů značek v každém z těchto souborů,
které jsou schopny ukázat přítomnost předem určeného
nukleotidu signálem generovaným zbytkem generujícím fluores-
cenční signál s podmínkou, že pro každý typ nukleotidu existuje
jiný zbytek generující fluorescenční signál.

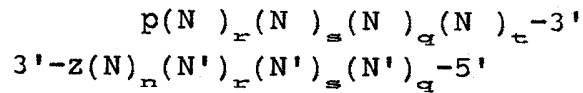
22. Způsob podle kteréhokoliv z nároků 18 až 21,
v y z n a č u j í c í s e t í m, že tyto oligonukleoti-
dové značky těchto kódovaných adaptorů jsou jednovláknové
a tyto komplementy značek pro tyto oligonukleotidové značky
jsou jednovláknové, takže dochází ke specifické hybridizaci
mezi oligonukleotidovou značkou a jejím příslušným komple-
mentem značky Watson-Crickovým párováním bází.

23. Látkové složení, v y z n a č u j í c í s e
t í m, že obsahuje dvojitý oligonukleotidový adaptor
mající formu:





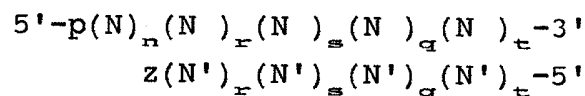
nebo



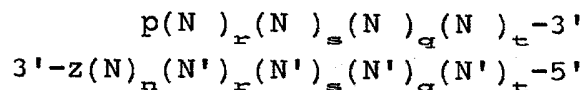
kde N je nukleotid a N' je jeho komplement, p je fosfátová skupina, z je 3' hydroxylová skupina nebo 3' blokující skupina, n je celé číslo od 2 do 6 včetně, r je celé číslo od 0 do 18 včetně, s je celé číslo buď od 4 do 6 včetně, pokud má kódovaný adaptor nukleázové rozpoznávací místo, nebo je rovno 0, pokud neexistuje nukleázové rozpoznávací místo, q je celé číslo větší než nebo rovné 0 a t je celé číslo větší než nebo rovné 8.

24. Látkové složení podle nároku 23, ve kterém r je od 0 do 12 včetně, t je celé číslo od 8 do 20 včetně, z je fosfátová skupina a daný jednovláknový zbytek $(N)_t$ je členem minimálně zkřížené hybridizujícího souboru.

25. Látkové složení obsahující dvojvláknový oligonukleotidový adaptor ve formě:

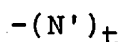
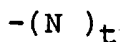


nebo



kde N je nukleotid a N' je jeho komplement, p je fosfátová skupina, z je 3' hydroxylová skupina nebo 3' blokující skupina, n je celé číslo od 2 do 6 včetně, r je celé číslo od 0 do 18 včetně, s je celé číslo buď od 4 do 6 včetně, pokud má kódovaný adaptor nukleázové rozpoznávací místo, nebo je rovno 0, pokud neexistuje nukleázové rozpoznávací místo, q je celé číslo větší než nebo rovné 0 a t je celé číslo větší než nebo rovné 8.

26. Látkové složení podle nároku 25,
v y z n a č u j í c í s e t í m, že r je od 0 do 12
včetně, t je celé číslo od 12 do 24 včetně, z je fosfátová
skupina a daný dvojvláknový zbytek

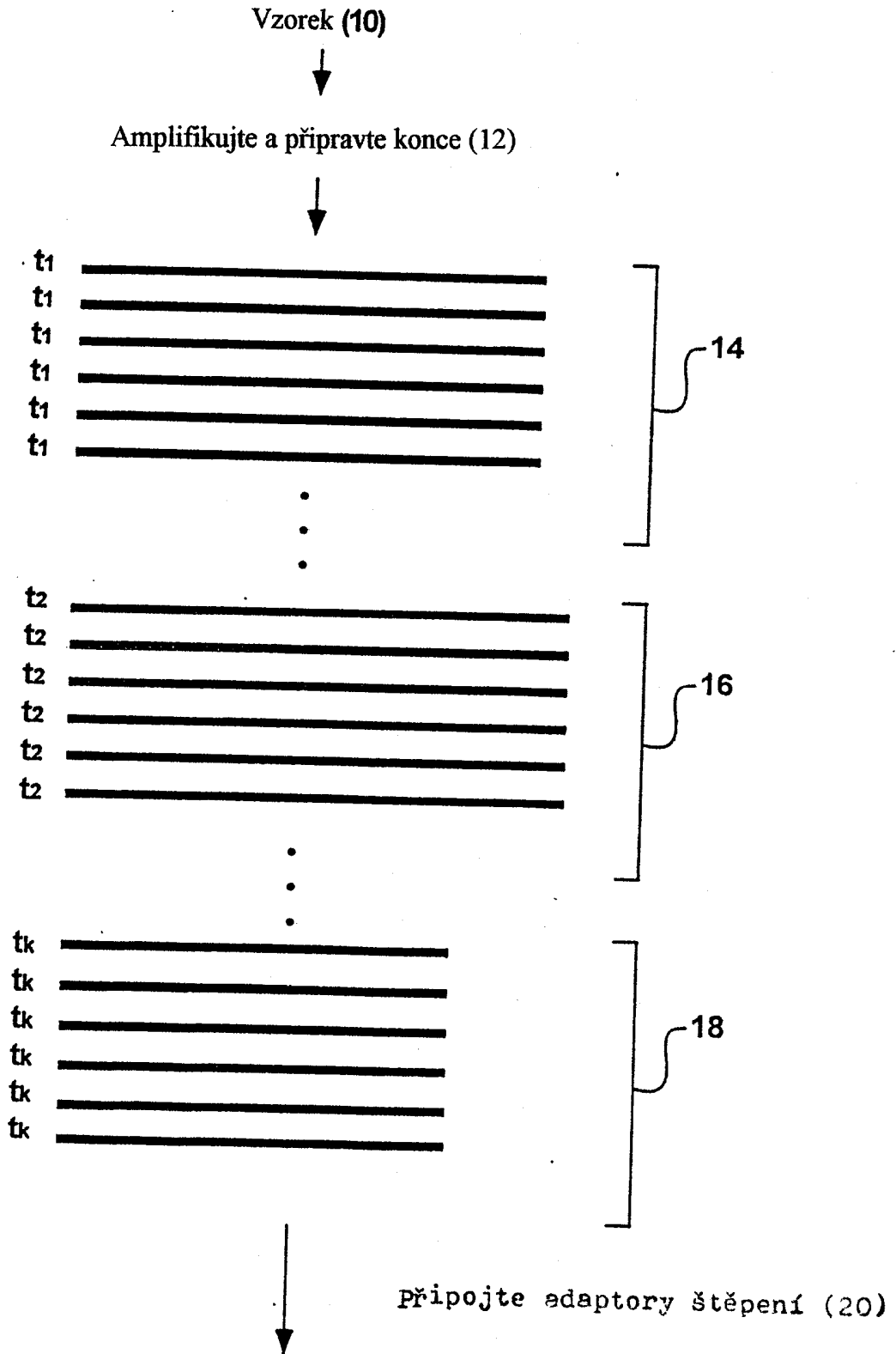


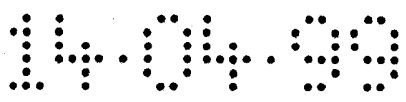
je členem minimálně zkříženě hybridizujícího souboru.

27. Látkové složení podle kteréhokoliv z nároků 23 až
26, v y z n a č u j í c í s e t í m, že n se rovná 4
a členové tohoto minimálně zkříženě hybridizujícího souboru
se navzájem od sebe liší alespoň o šest nukleotidů.

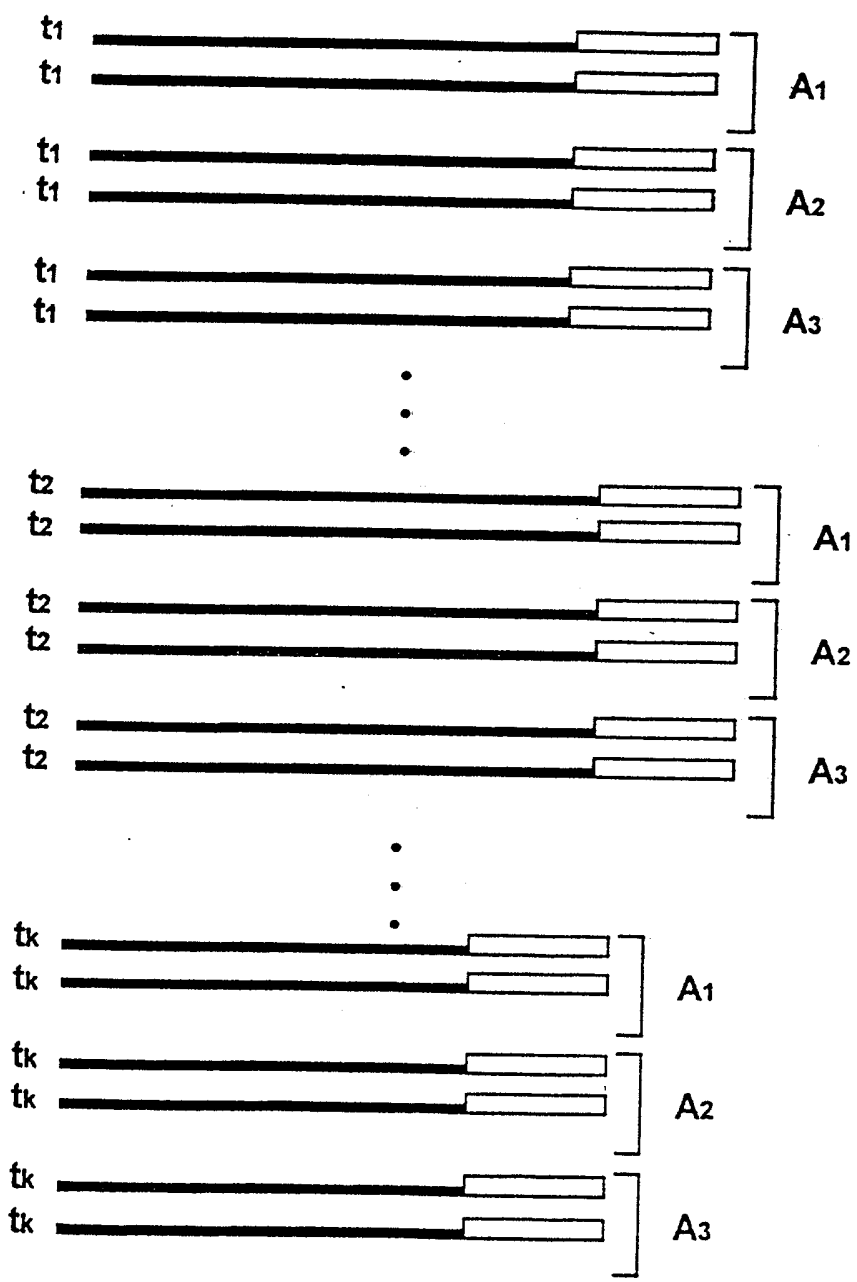
14.04.99

Obr. 1A

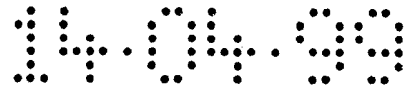




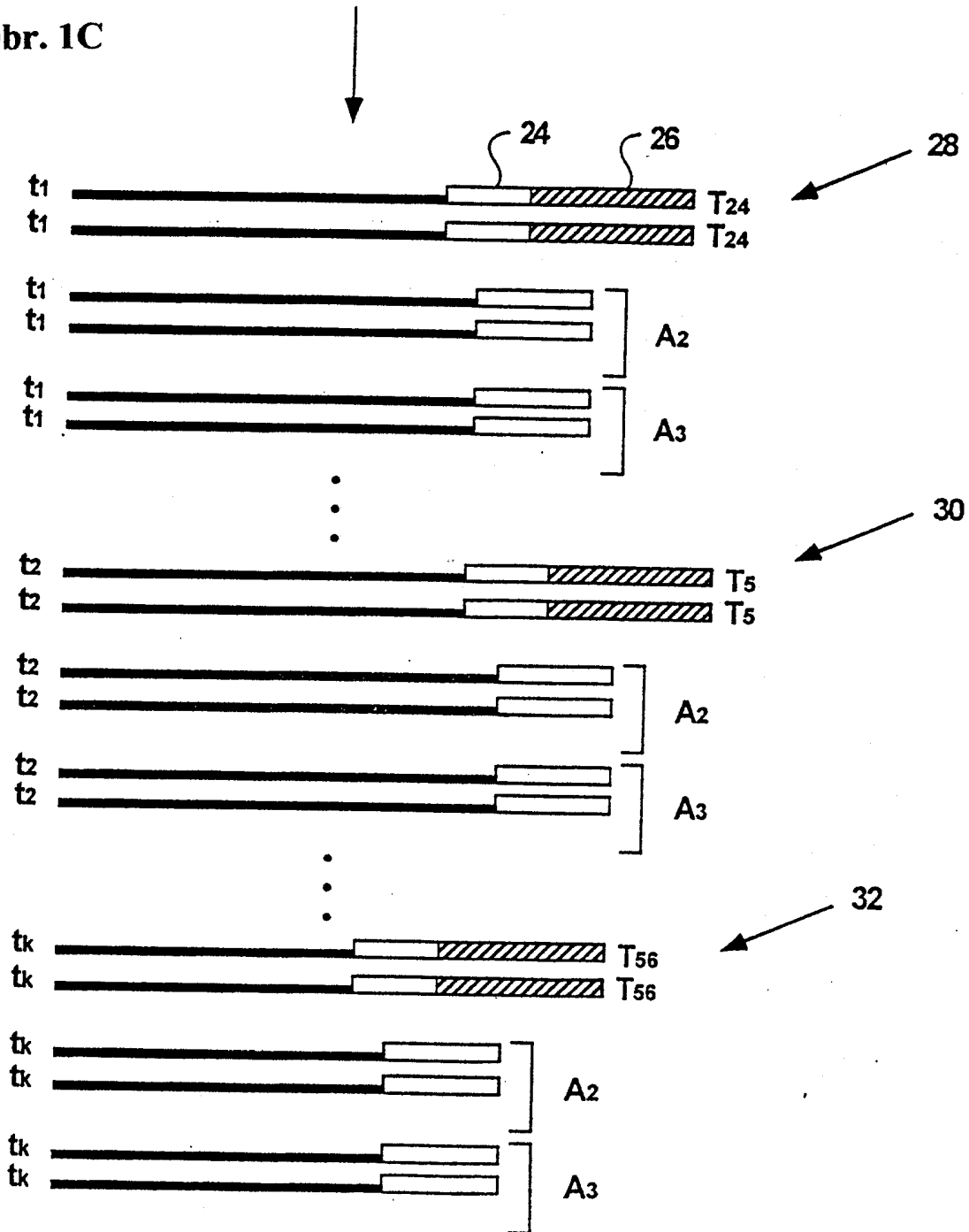
Obr. 1B



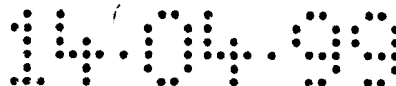
Odštěpte endonukleázou A1 a
připojte první soubor kódovaných sond (22)



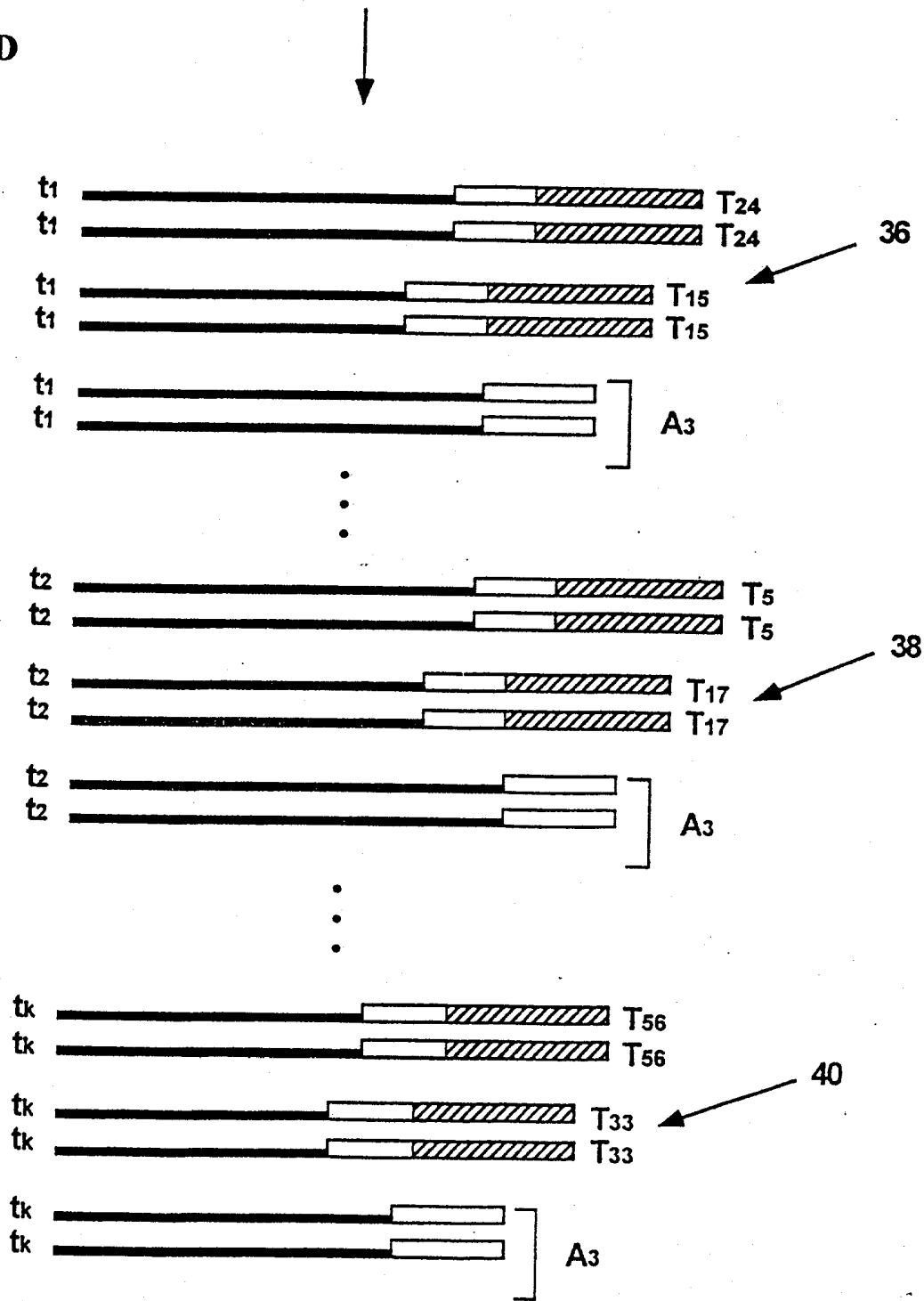
Obr. 1C



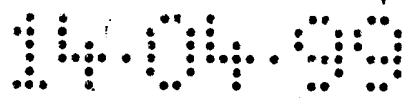
Odštěpte endonukleázou A2 a připojte druhý soubor kódovaných sond (34)



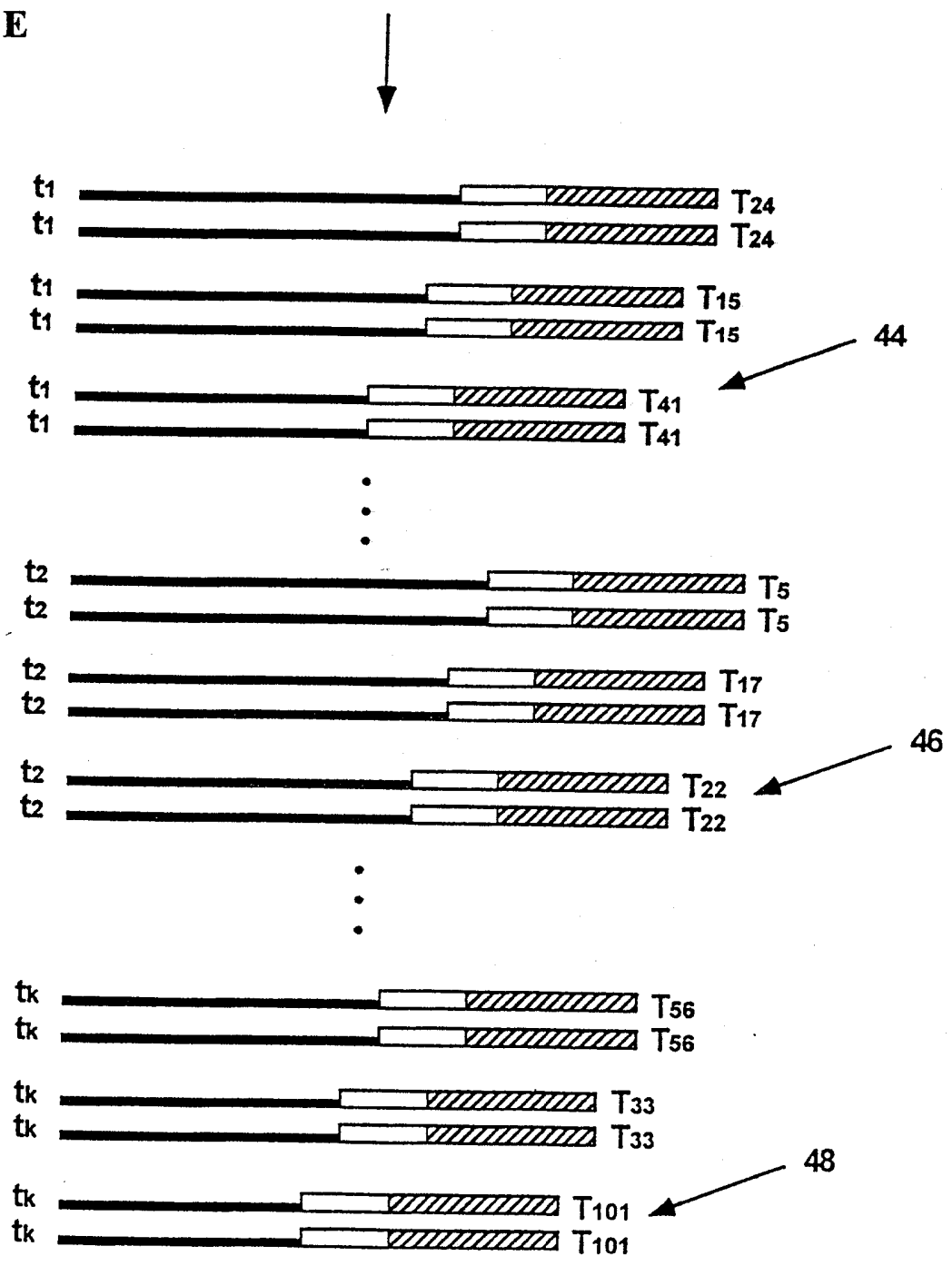
Obr. 1D



Odštěpte endonukleázou A₃ a
připojte třetí soubor kódovaných
sond (42)



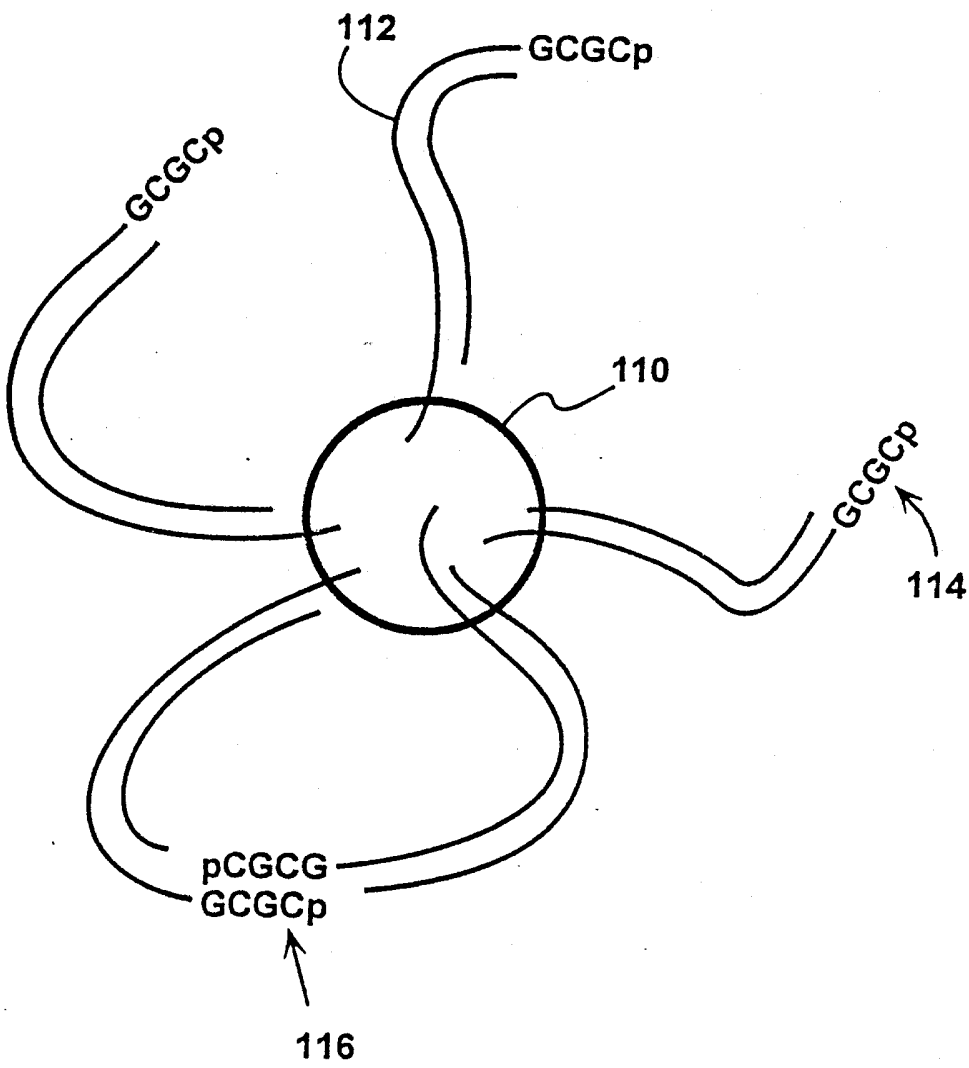
Obr. 1E



Naneste na pevný nosič
a hybridizujte komplemen-
ty značek (50)

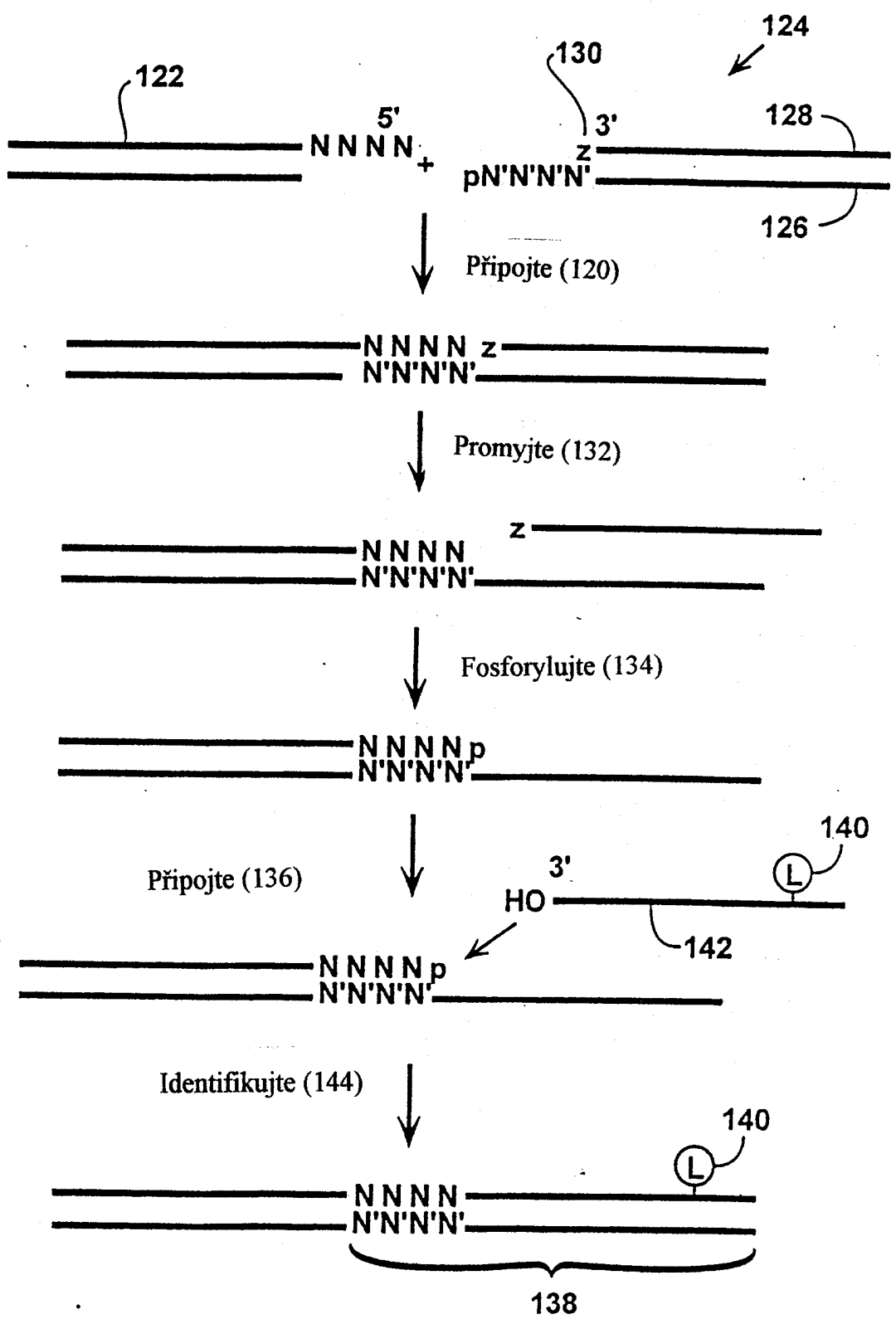
14-04-99

Obr. 2



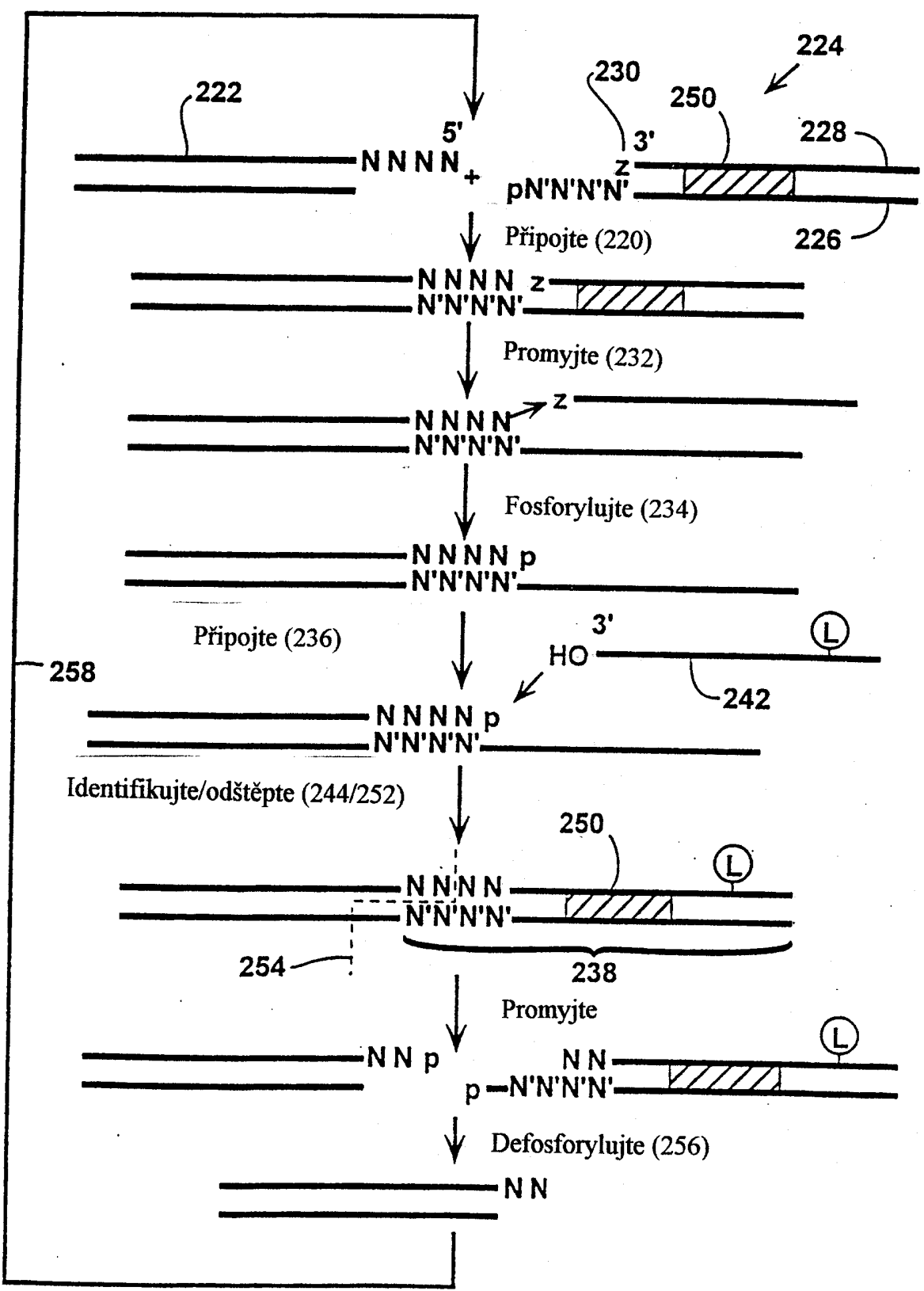
140400

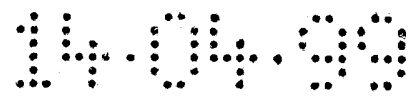
Obr. 3A



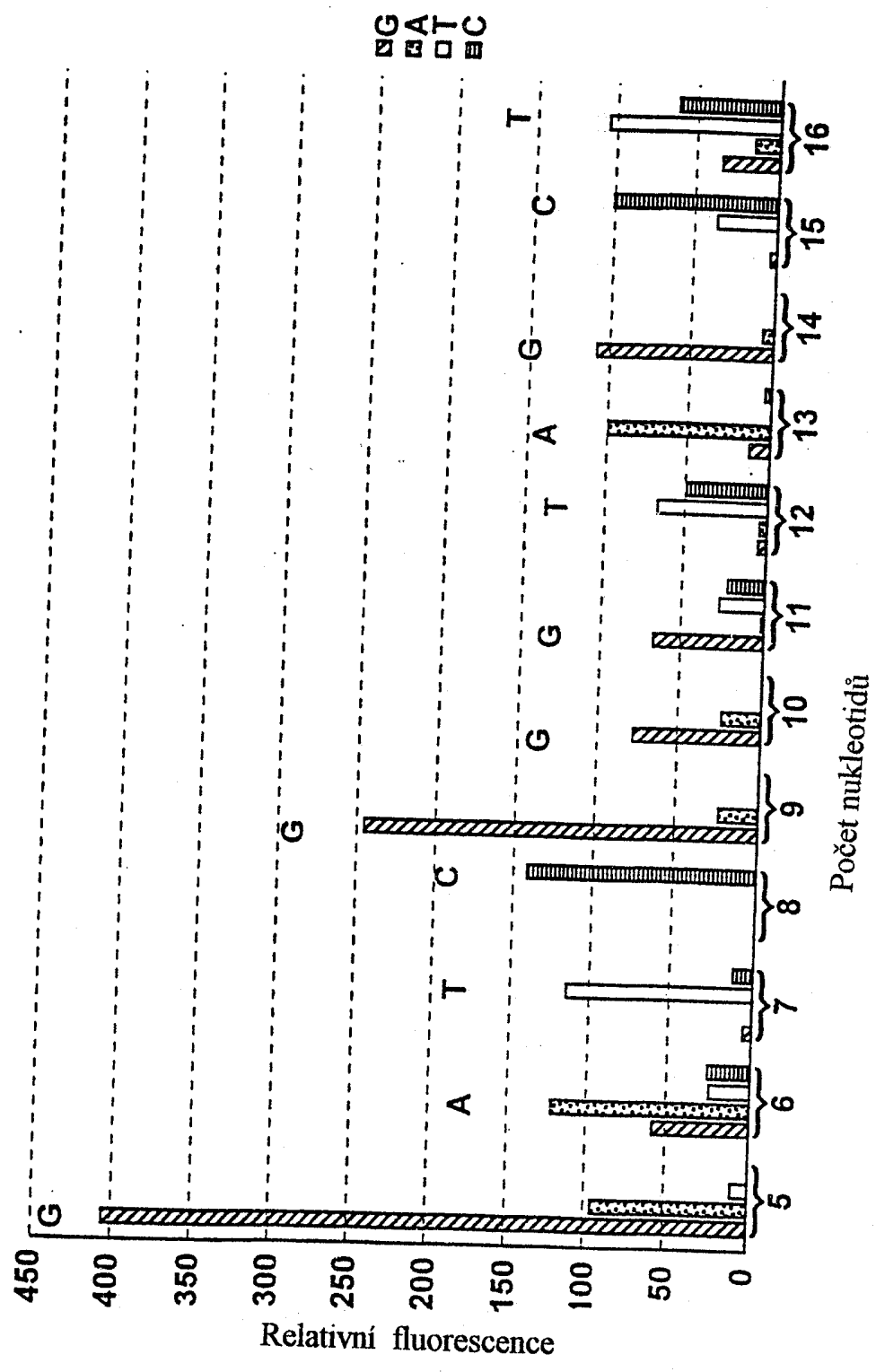


Obr. 3B





Obr. 4



Obr. 5

