

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4124377号
(P4124377)

(45) 発行日 平成20年7月23日 (2008. 7. 23)

(24) 登録日 平成20年5月16日 (2008. 5. 16)

(51) Int. Cl.

F I

C 1 2 N 15/09 (2006. 01)

C 1 2 N 15/00 Z N A A

C 1 2 Q 1/68 (2006. 01)

C 1 2 Q 1/68 A

請求項の数 28 (全 55 頁)

(21) 出願番号 特願平10-500755
 (86) (22) 出願日 平成9年6月2日 (1997. 6. 2)
 (65) 公表番号 特表2000-515006 (P2000-515006A)
 (43) 公表日 平成12年11月14日 (2000. 11. 14)
 (86) 国際出願番号 PCT/US1997/009472
 (87) 国際公開番号 W01997/046704
 (87) 国際公開日 平成9年12月11日 (1997. 12. 11)
 審査請求日 平成16年4月1日 (2004. 4. 1)
 (31) 優先権主張番号 08/659, 453
 (32) 優先日 平成8年6月6日 (1996. 6. 6)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 08/689, 587
 (32) 優先日 平成8年8月12日 (1996. 8. 12)
 (33) 優先権主張国 米国 (US)

(73) 特許権者

ソレクサ・インコーポレイテッド
 アメリカ合衆国 カリフォルニア 945
 45, ヘイワード, インダストリアル ブ
 ールバード 25861

(74) 代理人

弁理士 山田 行一

(74) 代理人

弁理士 野田 雅一

(74) 代理人

弁理士 池田 成人

(72) 発明者

アルプレッチ, グレン
 アメリカ合衆国 カリフォルニア 940
 61, レッドウッド シティ, オリバー
 ストリート 1113

最終頁に続く

(54) 【発明の名称】 コードアダプターの連結による配列決定

(57) 【特許請求の範囲】

【請求項 1】

ポリヌクレオチド末端のヌクレオチド配列を決定する方法であって、該方法は以下の工程

:

(a) 複数の異なるコードアダプターを該ポリヌクレオチドに適用する工程であって、こ
 こで各コードアダプターは、(i) 最小にクロスハイブリダイズするオリゴヌクレオチド
 のセットから選択されたオリゴヌクレオチドタグ、および(ii) 該ポリヌクレオチドの鎖
 の一部分に相補的な突出鎖、を含む2本鎖デオキシリボ核酸であり、

ここで各オリゴヌクレオチドタグは、該最小にクロスハイブリダイズするセットの他のど
 のオリゴヌクレオチドとも少なくとも2つのヌクレオチド異なり、そして該各オリゴヌ
 レオチドタグは、公知の様式で該突出鎖に対応する、工程、

(b) 該ポリヌクレオチドの該末端に、コードアダプターを連結する工程であって、該コ
 ードアダプターの突出末端は該末端と完全に適合した二重鎖を形成する、工程、

および

(c) 該ポリヌクレオチドの該末端において複数のヌクレオチドの各々について、標識さ
 れたタグ相補物を、該末端に連結された該各コードアダプターの該オリゴヌクレオチドタ
 グに特異的にハイブリダイズさせる工程であって、該ハイブリダイズされたタグ相補物が
 、該ヌクレオチドの同一性に対して既知の様式で対応し、これにより、該ポリヌクレオチ
 ドの該末端における該複数のヌクレオチドの各々を同定する工程、

を包含する、方法。

【請求項 2】

請求項 1 に記載の方法であって、前記連結する工程が、前記複数の異なるコードアダプターの前記突出鎖が前記ポリヌクレオチドの前記鎖の複数の異なる部分に相補的であり、そして該異なるコードアダプターと該鎖の該異なる部分との間に 1 対 1 の対応が存在するように、該複数の異なるコードアダプターを該ポリヌクレオチドの前記末端に連結することを包含する、方法。

【請求項 3】

前記ポリヌクレオチドの前記鎖の前記異なる部分が連続している、請求項 2 に記載の方法。

【請求項 4】

請求項 1 から 3 のいずれか 1 つに記載の方法であって、前記コードアダプターの前記突出鎖が、2 から 6 のヌクレオチドを含み、そして前記同定する工程が、前記ポリヌクレオチドの前記部分における各ヌクレオチドの同一性が連続的に決定されるように、前記タグ相補物を前記オリゴヌクレオチドタグに特異的にハイブリダイズさせることを包含する、方法。

【請求項 5】

請求項 1 から 4 のいずれか 1 つに記載の方法であって、前記同定する工程が、前記ポリヌクレオチドの前記部分において同定されるヌクレオチドの数に等しいタグ相補物の多数のセットを提供することをさらに包含する、方法。

【請求項 6】

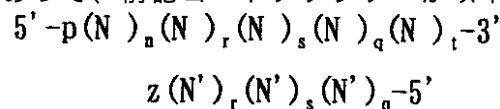
請求項 5 に記載の方法であって、前記同定する工程が、蛍光シグナル発生部分により発生されるシグナルによって、予め決定されたヌクレオチドの存在を示し得る前記セットのそれぞれにおいて、前記タグ相補物を提供することをさらに包含し、ヌクレオチドの各種類について異なる蛍光シグナル発生部分が存在する、方法。

【請求項 7】

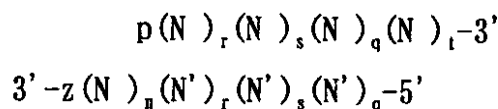
請求項 1 から 6 のいずれか 1 つに記載の方法であって、前記コードアダプターの前記オリゴヌクレオチドタグが 1 本鎖であり、そして該オリゴヌクレオチドタグに対する前記タグ相補物が 1 本鎖であり、その結果、オリゴヌクレオチドタグとそのそれぞれのタグ相補物との間の特異的なハイブリダイゼーションがワトソン-クリック塩基対形成により生じる、方法。

【請求項 8】

請求項 7 に記載の方法であって、前記コードアダプターが以下の形態を有し：



または



ここで、N はヌクレオチドであり、N' はその相補物であり、p はリン酸基であり、z は 3' 水酸基または 3' ブロッキング基であり、n は 2 から 6 の間の整数であり、r は 0 から 18 の間の整数であり、s は、コードアダプターがヌクレアーゼ認識部位を有する場合はいつでも 4 から 6 の間のいずれかである整数であるか、またはヌクレアーゼ認識部位がない場合はいつでも 0 であり、q は 0 を超える整数であるかもしくは 0 に等しく、そして t は 8 を超える整数であるかもしくは 8 に等しい、方法。

【請求項 9】

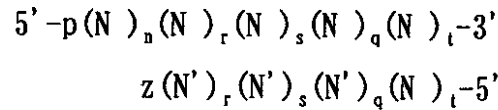
請求項 8 に記載の方法であって、ここで r が 0 から 12 の間であり、t が 8 から 20 の間の整数であり、そして z がリン酸基である、方法。

【請求項 10】

請求項 1 から 6 のいずれか 1 つに記載の方法であって、前記コードアダプターの前記オリゴヌクレオチドタグが 2 本鎖であり、そして該オリゴヌクレオチドタグに対する前記タグ相補物が 1 本鎖であり、その結果、オリゴヌクレオチドタグとそのそれぞれのタグ相補物との間の特異的なハイブリダイゼーションがフーグスティーンまたは逆フーグスティーン 3 重鎖の形成により生じる、方法。

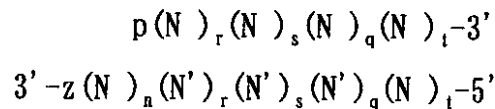
【請求項 1 1】

請求項 1 0 に記載の方法であって、前記コードアダプターは以下の形態を有し：



10

または



ここで、N はヌクレオチドであり、N' はその相補物であり、p はリン酸基であり、z は 3' 水酸基または 3' ブロッキング基であり、n は 2 から 6 の間の整数であり、r は 0 から 18 の間の整数であり、s は、該コードアダプターがヌクレアーゼ認識部位を有する場合はいつでも 4 から 6 の間のいずれかである整数であるか、またはヌクレアーゼ認識部位がない場合はいつでも 0 であり、q は 0 を超える整数であるかもしくは 0 に等しく、そして t は 8 を超える整数であるかもしくは 8 に等しい、方法。

20

【請求項 1 2】

請求項 1 1 に記載の方法であって、r が 0 から 12 の間であり、t が 12 から 24 の間の整数であり、そして z がリン酸基である、方法。

【請求項 1 3】

請求項 1 から 1 2 のいずれか 1 つに記載の方法であって、前記最小にクロスハイブリダイズするセットのメンバーは、他のどのメンバーとも少なくとも 6 ヌクレオチド異なる、方法。

【請求項 1 4】

複数のポリヌクレオチドのヌクレオチド配列を決定する方法であって、該方法は以下の工程：

30

(a) タグレパートリー由来の第 1 のオリゴヌクレオチドタグを、ポリヌクレオチド集団中の各ポリヌクレオチドに付着させる工程であって、ここで、該レパートリーの各第 1 のオリゴヌクレオチドタグは、第 1 の最小にクロスハイブリダイズするオリゴヌクレオチドのセットから選択され、そして該第 1 の最小にクロスハイブリダイズするセットの各オリゴヌクレオチドは、該第 1 のセットの他のどのオリゴヌクレオチドとも少なくとも 2 つのヌクレオチド異なる、工程；

(b) 該ポリヌクレオチドの集団をサンプリングして、サンプル中の実質に全ての異なるポリヌクレオチドが、付着された異なる第 1 のオリゴヌクレオチドタグを有するように、ポリヌクレオチドのサンプルを形成する工程；

40

(c) 該第 1 のオリゴヌクレオチドタグとそれらのそれぞれの相補物とを特異的にハイブリダイズさせることにより、該サンプルの該ポリヌクレオチドを選別する工程であって、該それぞれの相補物は、1 つ以上の固相支持体の上の空間的に別々の領域に実質的に同一のオリゴヌクレオチドの均一な集団として付着される、工程

(d) 複数の異なるコードアダプターを該ポリヌクレオチドに適用する工程であって、ここで各コードアダプターは、(i) オリゴヌクレオチドの第 2 の最小にクロスハイブリダイズするセットから選択されたオリゴヌクレオチドタグ、および(ii) 該サンプル中のポリヌクレオチドの該末端部分に相補的な突出鎖を含む、2 本鎖デオキシリボ核酸であり、該第 2 の最小にクロスハイブリダイズするセットの各オリゴヌクレオチドタグは、該第 2 のセットの他のどのオリゴヌクレオチドとも少なくとも 2 ヌクレオチド異なり、そして該

50

各オリゴヌクレオチドタグは、公知の様式で該突出鎖に対応する、工程；および
 (e) 該ポリヌクレオチドの該末端に、コードアダプターを連結する工程であって、該コードアダプターの突出末端は該末端と完全に適合した二重鎖を形成する、工程、
 (f) 該ポリヌクレオチドの該末端において複数のヌクレオチドの各々について、標識されたタグ相補物を、該末端に連結された該各コードアダプターの該オリゴヌクレオチドタグに特異的にハイブリダイズさせる工程であって、該ハイブリダイズされたタグ相補物が、該ヌクレオチドの同一性に対して既知の様式で対応し、これにより、該ポリヌクレオチドの該末端における該複数のヌクレオチドの各々を同定する工程、
 を包含する、方法

【請求項 15】

請求項 14 に記載の方法であって、(f) 新たな突出鎖が前記ポリヌクレオチドのそれぞれの前記末端に形成されるように、その切断部位から離れたヌクレアーゼ認識部位を有するヌクレアーゼを用いて前記コードアダプターを該ポリヌクレオチドから切断する工程、および (g) 工程 (d) から (f) を反復する工程、をさらに包含する、方法。

【請求項 16】

mRNA 分子の集団を同定する方法であって、該方法は以下の工程：

(a) 各 cDNA 分子が、付着された第 1 のオリゴヌクレオチドタグを有するように、該 mRNA 分子の集団から cDNA 分子の集団を形成する工程であって、該第 1 のオリゴヌクレオチドタグが、第 1 の最小にクロスハイブリダイズするオリゴヌクレオチドのセットから選択され、ここで該第 1 の最小にクロスハイブリダイズするセットの各オリゴヌクレオチドは、該第 1 のセットの他のどのオリゴヌクレオチドとも少なくとも 2 つのヌクレオチド異なる、工程；

(b) 該 cDNA 分子の集団をサンプリングして、実質的に全ての異なる cDNA 分子が、付着された異なる第 1 のオリゴヌクレオチドタグを有するように、cDNA 分子のサンプルを形成する工程；

(c) 該第 1 のオリゴヌクレオチドタグをそれらのそれぞれの相補物と特異的にハイブリダイズさせることにより該 cDNA 分子を選別する工程であって、該それぞれの相補物は、1 つ以上の固相支持体上の空間的に別々の領域において実質的に同一の相補物の均一な集団として付着される、工程；

(d) 複数の異なるコードアダプターを該ポリヌクレオチドに適用する工程であって、ここで各コードアダプターは、(i) 第 2 の最小にクロスハイブリダイズするオリゴヌクレオチドのセットから選択されたオリゴヌクレオチドタグ、および (ii) 該集団中の cDNA 分子の該末端の一部分に相補的な突出鎖を含む、2 本鎖デオキシリボ核酸であり、ここで該第 2 の最小にクロスハイブリダイズするセットの各オリゴヌクレオチドタグは、該第 2 のセットの他のどのオリゴヌクレオチドとも少なくとも 2 つのヌクレオチド異なり、そして該各オリゴヌクレオチドタグは、公知の様式で該突出鎖に対応する、工程；

(e) 該ポリヌクレオチドの該末端に、コードアダプターを連結する工程であって、該コードアダプターの突出末端は該末端と完全に適合した二重鎖を形成する、工程；

(f) 該ポリヌクレオチドの該末端において複数のヌクレオチドの各々について、標識されたタグ相補物を、該末端に連結された該各コードアダプターの該オリゴヌクレオチドタグに特異的にハイブリダイズさせる工程であって、該ハイブリダイズされたタグ相補物が、該ヌクレオチドの同一性に対して既知の様式で対応し、これにより、該ポリヌクレオチドの該末端における該複数のヌクレオチドの各々を同定する工程、ならびに

(g) 該 cDNA 分子の配列の一部分の頻度分布により、該 mRNA 分子の集団を同定する工程、
 を包含する、方法。

【請求項 17】

請求項 16 に記載の方法であって、(f) 新たな突出鎖が前記 cDNA 分子のそれぞれの前記末端に形成されるように、その切断部位から離れたヌクレアーゼ認識部位を有するヌクレアーゼを用いて前記コードアダプターを前記ポリヌクレオチドから切断する工程、および (g) 工程 (d) から (f) を反復する工程、

をさらに包含する、方法。

【請求項 18】

ポリヌクレオチド末端のヌクレオチド配列を決定する方法であって、該方法は以下の工程：

(a) 複数の異なるコードアダプターを該ポリヌクレオチドに適用する工程であって、ここで各コードアダプターは、(i) 最小にクロスハイブリダイズするオリゴヌクレオチドのセットから選択されたオリゴヌクレオチドタグ、および(ii) 該ポリヌクレオチドの該末端の一部分に相補的な突出鎖を含む、2本鎖デオキシリボ核酸であり、そして、ここで各オリゴヌクレオチドタグは、該最小にクロスハイブリダイズするセットの他のどのオリゴヌクレオチドとも少なくとも2つのヌクレオチド異なり、そして該各オリゴヌクレオチドタグは、公知の様式で該突出鎖に対応する、工程；

(b) 該ポリヌクレオチドの該末端に、コードアダプターを連結する工程であって、該コードアダプターの突出末端は該末端と完全に適合した二重鎖を形成する、工程；

(c) 該ポリヌクレオチドの該末端において複数のヌクレオチドの各々について、標識されたタグ相補物を、該末端に連結された該各コードアダプターの該オリゴヌクレオチドタグに特異的にハイブリダイズさせる工程であって、該ハイブリダイズされたタグ相補物が、該ヌクレオチドの同一性に対して既知の様式で対応し、これにより、該ポリヌクレオチドの該末端における該複数のヌクレオチドの各々を同定する工程

(d) 新たな突出鎖が該ポリヌクレオチドの末端に形成されるように、その切断部位から離れたヌクレアーゼ認識部位を有するヌクレアーゼを用いて、該コードアダプターを該ポリヌクレオチド末端から切断する工程；および

(e) 工程(a)から(d)を反復する工程、
を包含する、方法。

【請求項 19】

請求項 18 に記載の方法であって、前記コードアダプターの前記突出鎖が 2 から 6 のヌクレオチドを含み、そして前記同定する工程が、前記ポリヌクレオチドの前記部分における各ヌクレオチドの同一性が連続的に決定されるように、連続的に前記タグ相補物を前記オリゴヌクレオチドタグに特異的にハイブリダイズさせる工程を包含する、方法。

【請求項 20】

請求項 19 に記載の方法であって、前記同定する工程が、前記ポリヌクレオチドの前記部分において同定されたヌクレオチドの数に等しいタグ相補物の多数のセットを提供することをさらに包含する、方法。

【請求項 21】

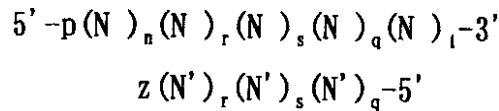
請求項 20 に記載の方法であって、前記同定する工程が、蛍光シグナル発生部分により発生されるシグナルにより予め決定されたヌクレオチドの存在を示し得る前記セットのそれぞれにおいて、前記タグ相補物を提供することをさらに包含し、ヌクレオチドの各種類について異なる蛍光シグナル発生部分が存在する、方法。

【請求項 22】

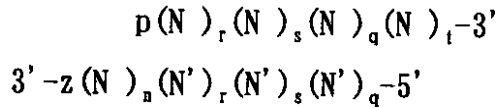
請求項 18 から 21 のいずれか 1 つに記載の方法であって、ここで前記コードアダプターの前記オリゴヌクレオチドタグが 1 本鎖であり、そして前記オリゴヌクレオチドタグに対する前記タグ相補物が 1 本鎖であり、その結果、オリゴヌクレオチドタグとそのそれぞれのタグ相補物との間の特異的なハイブリダイゼーションがワトソン-クリック塩基対形成により生じる、方法。

【請求項 23】

以下の形態を有する複数の 2 本鎖オリゴヌクレオチドアダプターを含む物質の組成物：



または



ここで、Nはヌクレオチドであり、N'はその相補物であり、pはリン酸基であり、zは3'水酸基または3'ブロッキング基であり、nは2から6の間の整数であり、rは0から18の間の整数であり、sは、コードアダプターがヌクレアーゼ認識部位を有する場合はいつでも4から6の間のいずれかである整数であるか、またはヌクレアーゼ認識部位がない場合はいつでも0であり、qは0を超える整数であるかもしくは0に等しく、そしてtは8を超える整数であるかもしくは8に等しく、そして(N)_tは、1本鎖であり、そして最小にクロスハイブリダイズするオリゴヌクレオチドのセットの各オリゴヌクレオチドが、該セットの他のどのオリゴヌクレオチドとも少なくとも2塩基対異なるような該セットから選択される、組成物。

10

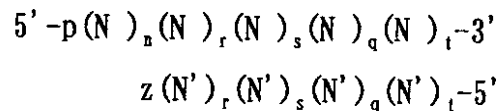
【請求項24】

請求項23に記載の組成物であって、ここでrが0から12の間であり、tが8から20の間の整数であり、そしてzがリン酸基である、組成物。

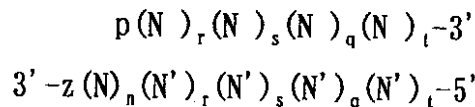
20

【請求項25】

以下の形態を有する複数の2本鎖オリゴヌクレオチドアダプターを含む物質の組成物：



または



30

ここで、Nはヌクレオチドであり、N'はその相補物であり、pはリン酸基であり、zは3'水酸基または3'ブロッキング基であり、nは2から6の間の整数であり、rは0から18の間の整数であり、sは、コードアダプターがヌクレアーゼ認識部位を有する場合はいつでも4から6の間のいずれかである整数であるか、またはヌクレアーゼ認識部位がない場合はいつでも0であり、qは0を超える整数であるかもしくは0に等しく、そしてtは8を超える整数であるかもしくは8に等しく、その結果、

(N)_t

(N')_t

は、最小にクロスハイブリダイズするオリゴヌクレオチドのセットの各オリゴヌクレオチドが、該セットの他のどのオリゴヌクレオチドとも少なくとも2つの塩基対異なるように、該セットから選択される2本鎖部分を形成する、組成物。

40

【請求項26】

請求項25に記載の組成物であって、ここでrが0から12の間であり、tが12から24の間の整数であり、そしてzがリン酸基である、組成物。

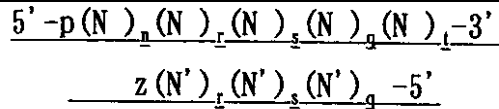
【請求項27】

請求項23から26のいずれか1つに記載の組成物であって、ここで前記最小にクロスハイブリダイズするセットのメンバーは、他のどのメンバーとも少なくとも6つのヌクレオチド異なる、組成物。

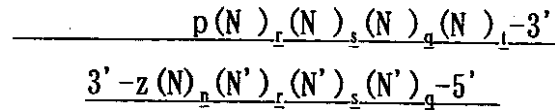
【請求項28】

50

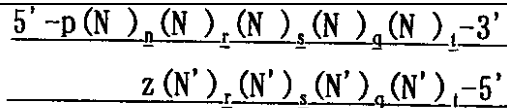
以下の形態を有する複数の 2 本鎖オリゴヌクレオチドアダプターを含む物質の組成物：



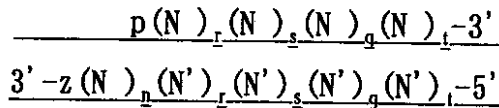
または



ここで、各 (N)_i は、ユニークな一本鎖オリゴヌクレオチドタグであり、各セットのオリゴヌクレオチドが少なくとも 2 ヌクレオチド分該セットの一つおきのオリゴヌクレオチドごとに異なるように最小限にクロスハイブリダイズするセットのオリゴヌクレオチドから選択され、t は 8 と 20 との間の両端を含む整数であるか、または、



または



ここで、

各 $(N)_t$ は、ユニークな二本鎖オリゴヌクレオチドタグであり、該セットの $(N')_t$

各オリゴヌクレオチドが少なくとも 2 塩基対分該セットの一つおきのオリゴヌクレオチド
ごとに異なるように最小限にクロスハイブリダイズするセットのオリゴヌクレオチドから
選択され、 t は 1 2 と 2 0 との間の両端を含む整数であり、

N はヌクレオチドであり、N' はその相補物であり、

p はリン酸基であり、

z は3' 水酸基または3' ブロッキング基であり、

n は 2 から 6 の間の整数であり、

r は 0 から 18 の間の整数であり、

s は、4 から 6 の間の両端を含む整数であるか、または

該アダプターが、二本鎖部分において、ヌクレアーゼのヌクレアーゼ認識部位を含み、該ヌクレアーゼの認識部位が、該ヌクレアーゼの開裂部位とは別個であり、

q は、 0 またはそれより大きな整数である、

組成物。

【発明の詳細な説明】

発明の分野

本発明は、一般的に、ポリヌクレオチドのヌクレオチド配列を決定するための方法、より詳細には、コードアダプターの特異的連結によりポリヌクレオチドの末端ヌクレオチドを同定する方法に関する。

発明の背景

ほとんど全ての科学的および商業的適用のための選り抜きのDNA配列決定法は、Sanger（例えば、Sangerら、Proc.Natl.Acad.Sci., 74:5463-5467(1977)）により開拓されたジデオキシ鎖終結アプローチに基づく。この方法はいくつかの方法により改善され、そして種々の形態で、全ての市販されているDNA配列決定器機において使用される（例えば、Hunkapillerら、Science, 254:59-67(1991)）。

鎖終結法は、各々が共通の起源を有し、そして各々が既知の塩基で終結する、標識DNAフラグメントの1セット以上の作製を必要とする。次いで、フラグメントのセット（単数ま

たは複数)は、配列情報を得るためにサイズにより分離されなければならない。サイズ分離は、通常、高分離度ゲル電気泳動により達成され、これはたった1つのヌクレオチドでサイズが異なる非常に大きなフラグメントを区別する能力を有さなければならない。多くの著しい改善(例えば、細孔配列(capillary array)および非ゲル電気泳動分離媒体の使用を用いる分離)にもかかわらず、この技術は、小型化または大規模に並行する実行に容易に向いていない。

DNA配列決定へのSangerに基づくアプローチに対する代替として、いくつかのいわゆる「1塩基ずつ」または「単一塩基」配列決定アプローチが調べられた(例えば、Cheeseman、米国特許第5,302,509号; Tsienら、国際出願WO 91/06678; Rosenthalら、国際出願WO 93/21340; Canardら、Gene, 148:1-6(1994); およびMetzkerら、Nucleic Acids Research, 22:4259-4267(1994))。これらのアプローチは、化学的または生化学的操作のサイクルあたりの単一ヌクレオチドの決定より特徴づけられ、そして分離工程を必要としない。従って、それらが考えられるように実行され得る場合、「1塩基ずつ」のアプローチは、多くの配列決定反応を、例えば、微粒子または固相配列に付着させた標的ポリヌクレオチド上で並行して行う可能性を約束する(例えば、国際特許出願PCT/US95/12678(WO96/12039))。

不運なことに、「1塩基ずつ」の配列決定スキームは、非常に多くの問題(例えば、完全な配列決定操作における数ヌクレオチドより多くのいずれかの決定を妨げる非効率的な化学的性質)のために広範な適用を有さなかった。さらに、酵素操作を必要とする1塩基ずつのアプローチにおいて、さらなる問題は、自動化プロセッシングに使用される器具使用で生じる。一連の酵素的工程が高い表面对容積比および狭いチャンネル面積を有する反応チャンバーにおいて行われる場合、酵素は表面成分に固着し得、これは洗浄および連続プロセッシング工程を非常に難しくする。タンパク質の蓄積はまた、分子レポーターシステム、特に蛍光標識を使用するものに影響を与え、そしてこのようなシステムに基づく測定の解釈を難しくかつ不便にする。これらのおよび類似の困難は、並行する配列決定の努力への「1塩基ずつ」の配列決定スキームの適用を著しく示した。

複数の酵素を使用する反復プロセッシングサイクルを最小化するか、または削除する代替のアプローチが、ポリヌクレオチドの末端ヌクレオチドの決定に利用可能な場合、特に自動化システムにおいて、1塩基ずつの配列決定技術における重要な進歩が遂げられ得る。

発明の要旨

従って、本発明者らの発明の目的は、現在の1塩基ずつのアプローチの欠点をこうむらないDNA配列決定スキームを提供することである。

本発明者らの発明の別の目的は、共通の反応容器に存在する何千のDNAフラグメントに対して並行してまたは同時に適用しやすいDNA配列決定の方法を提供することである。

本発明者らの発明のさらなる目的は、最小の酵素的工程で標的ポリヌクレオチドの末端部分の同定を可能にするDNA配列決定の方法を提供することである。

本発明者らの発明のなお別の目的は、1つ以上の標的ポリヌクレオチドの複数の末端ヌクレオチドの配列を同定するためのコードアダプターの1セットを提供することである。

本発明者らの発明は、コードアダプターの1つ以上のセットの標的ポリヌクレオチド末端への(または並行した配列決定操作において使用される場合、複数の標的ポリヌクレオチド末端への)連結に基づく核酸配列分析の方法を提供することによりこれらのおよび他の目的を提供する。各コードアダプターは、突出鎖および最小にクロスハイブリダイズするオリゴヌクレオチドのセットから選択されるオリゴヌクレオチドタグを含む。コードアダプター(その突出鎖は、標的ポリヌクレオチドの相補的突出鎖と完全にマッチした2重鎖を形成する)が連結される。連結後、突出鎖におけるヌクレオチドの同一性および順序が決定されるか、または連結アダプター上のその対応するタグに相補的な標識タグを特異的にハイブリダイズさせることにより「解読(decode)」される。

例えば、4つのヌクレオチドの突出鎖(例えば、5'-AGGT)を有するコードアダプターが、標的ポリヌクレオチドの相補的な突出鎖と完全にマッチした2重鎖を形成し、そして連結される場合、ポリヌクレオチド上の4つの相補的ヌクレオチド3'-TCCAは、突出鎖の全て

10

20

30

40

50

の可能な4つのヌクレオチド配列について1つである、256のこのようなタグのセットから選択される唯一のオリゴヌクレオチドタグにより同定され得る。タグ相補物は、連結されたアダプターのオリゴヌクレオチドタグと共に完全にマッチする2重鎖（または3重鎖）を形成するそれらのタグ相補物のみの特異的なハイブリダイゼーションを可能にする条件下で、連結アダプターに適用される。タグ相補物は、オリゴヌクレオチドタグ、従って突出鎖の配列の同一性を決定するために、個々にまたは1つ以上の混合物として適用され得る。

以下により十分に説明されるように、コードアダプターは、(i) Brenner米国特許第5,599,675号およびPCT公報第WO 95/27080号に記載されるように、連結、同定、および切断の反復サイクルを含むプロセスの工程として1つ以上のヌクレオチドを同定するために、または(ii)「スタンド(stand)単独」同定方法（ここで、コードアダプターのセットが標的ポリヌクレオチドに適用され、その結果各セットが標的ポリヌクレオチドの異なる部分のヌクレオチド配列を同定し得る）としてのいずれかで、配列分析において使用され得る；すなわち、後者の実施態様において、配列分析は、各セットの単一の連結、それに続く同定を用いて行われる。

コードアダプターの重要な特性は、例えば、国際特許出願PCT/US95/12791(WO 96/12041)およびPCT/US96/09513(WO 96/41011)に記載されるような、最小にクロスハイブリダイズするオリゴヌクレオチドのセットのメンバーであるオリゴヌクレオチドタグの使用である。このようなセットのオリゴヌクレオチドの配列は、少なくとも2つのヌクレオチドで同じセットの全ての他のメンバーの配列と異なる。従って、このようなセットの各メンバーは、2未満のミスマッチを有する任意の他のメンバーの相補物と2重鎖（または3重鎖）を形成し得ない。好ましくは、最小にクロスハイブリダイズするセットの各メンバーは、特定の適用に必要とされるセットのサイズに一致したできるだけ多くのヌクレオチドで全ての他のメンバーと異なる。例えば、標識をコードアダプターに送達するために12~20マーのようなより長いオリゴヌクレオチドタグが使用される場合、最小にクロスハイブリダイズするセットのメンバー間の差異は、好ましくは、2より著しく大きい。好ましくは、このようなセットの各メンバーは、少なくとも4つのヌクレオチドで全ての他のメンバーと異なる。より好ましくは、このようなセットの各メンバーは少なくとも6つのヌクレオチドで全ての他のメンバーと異なる。本発明のオリゴヌクレオチドタグの相補物は、「タグ相補物」と本明細書中で呼ばれる。

オリゴヌクレオチドタグは1本鎖であり得、そして2重鎖形成による1本鎖タグ相補物への特異的ハイブリダイゼーションのために設計され得る。オリゴヌクレオチドタグはまた2本鎖であり得、そして3重鎖形成による1本鎖タグ相補物への特異的ハイブリダイゼーションのために設計され得る。好ましくは、コードアダプターのオリゴヌクレオチドタグは2本鎖であり、そしてそれらのタグ相補物は1本鎖であり、その結果その相補物とのタグの特異的ハイブリダイゼーションは、3重鎖構造の形成により生じる。

好ましくは、本発明の方法は以下の工程：(a)コードアダプターをポリヌクレオチド末端に連結する工程であって、アダプターはオリゴヌクレオチドの最小にクロスハイブリダイズするセットから選択されたオリゴヌクレオチドタグおよびポリヌクレオチドの突出鎖に相補的な突出鎖を有する工程；ならびに(b)タグ相補物をコードアダプターのオリゴヌクレオチドタグに特異的にハイブリダイズさせることにより、ポリヌクレオチドの突出鎖における1つ以上のヌクレオチドを同定する工程、を包含する。

【図面の簡単な説明】

図1A~1Eは、複数のタグ化ポリヌクレオチドの末端ヌクレオチド配列を決定するためのコードアダプターの使用を模式的に示す。

図2は、固相支持体に固定される同一ポリヌクレオチドの自己連結の現象を示す。

図3Aは、ブロックされた3'炭素有する2本鎖アダプターが標的ポリヌクレオチドに連結される、本発明の好ましい方法における工程を示す。

図3Bは、連結および切断の段階を追ったサイクルによるDNA配列決定方法における好ましい実施態様の使用を示す。

10

20

30

40

50

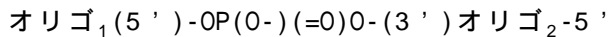
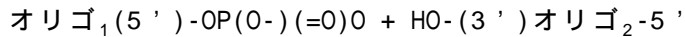
図4は、本発明の方法を用いる試験ポリヌクレオチドの末端ヌクレオチドの決定によるデータを示す。

図5は、配列決定のためにcDNA分子で充填された微粒子の平面配列を観察するための、フローチャンバーおよび検出装置の模式図である。

定義

本明細書中で使用される用語「コードアダプター」は、優先権書類である米国特許出願第08/689,587号の用語「コードプローブ」と同義に使用される。

本明細書中で使用される用語「連結」は、1つ以上(通常2つ)のオリゴヌクレオチドの末端間での共有結合の形成を意味する。通常、この用語は、以下の反応により生じるホスホジエステル結合の形成をいい、これは通常リガーゼにより触媒される：



ここで、オリゴ₁およびオリゴ₂は、2つの異なるオリゴヌクレオチドまたは同じオリゴヌクレオチドの異なる末端のいずれかである。この用語は、ホスホジエステル結合の非酵素的形成、ならびにオリゴヌクレオチド末端間での非ホスホジエステル共有結合(例えば、ホスホロチオエート結合、ジスルフィド結合など)の形成を含む。連結反応は、通常テンプレートにより駆動され、オリゴ₁およびオリゴ₂の末端は、テンプレート鎖への特異的ハイブリダイゼーションにより並置して導かれる。テンプレート駆動連結の特殊なケースは、相補的突出鎖を有する2つの2本鎖オリゴヌクレオチドの連結である。

オリゴヌクレオチドタグに関して本明細書中で用いられる「相補物」または「タグ相補物」は、オリゴヌクレオチドタグが特異的にハイブリダイズして完全にマッチした2重鎖または3重鎖を形成するオリゴヌクレオチドをいう。特異的ハイブリダイゼーションが3重鎖を生じる実施態様では、オリゴヌクレオチドタグは2本鎖または1本鎖のいずれかであるように選択され得る。従って、3重鎖が形成される場合、用語「相補物」は、1本鎖オリゴヌクレオチドタグの2本鎖相補物または2本鎖オリゴヌクレオチドタグの一本鎖相補物のいずれかを包含することを意味する。

本明細書で用いられる用語「オリゴヌクレオチド」は、モノマー対モノマーの相互作用の通常のパターン(例えば、ワトソン-クリック(Watson-Crick)型の塩基対形成、塩基の積み重ね、フーグスティーン(Hoogsteen)または逆フーグスティーン型の塩基対形成など)により標的ポリヌクレオチドに特異的に結合し得る、天然のまたは改変されたモノマーまたは連結物(デオキシリボヌクレオシド、リボヌクレオシド、それらのアナログ型、ペプチド核酸(PNA)などを含む)の直鎖状オリゴマーを含む。通常、モノマーは、ホスホジエステル結合またはそのアナログにより連結され、少数のモノマーユニット(例えば3~4)から数10のモノマーユニット(例えば、40~60)のサイズの範囲のオリゴヌクレオチドを形成する。オリゴヌクレオチドが一連の文字(例えば、「ATGCCTG」)により示される場合はいつも、他に記載されない限り、ヌクレオチドは左から右に5' 3'の順であり、「A」はデオキシアデノシンを示し、「C」はデオキシシチジンを示し、「G」はデオキシグアノシンを示し、および「T」はチミジンを示すことが理解される。通常、本発明のオリゴヌクレオチドは、4つの天然のヌクレオチドを含む；しかし、それらはまた、非天然のヌクレオチドアナログを含み得る。天然または非天然のヌクレオチドを有するオリゴヌクレオチドが用いられ得る場合、例えば、酵素によるプロセッシングが必要とされる場合、通常、天然のヌクレオチドからなるオリゴヌクレオチドが必要であることが当業者には明らかである。

二重鎖に関して「完全にマッチした」は、二重鎖を作製するポリヌクレオチド鎖またはオリゴヌクレオチド鎖が、各鎖中の全てのヌクレオチドが他鎖中のヌクレオチドとワトソン-クリック塩基対形成を受けるように、互いに2本鎖構造を形成することを意味する。この用語はまた、用いられ得るヌクレオシドアナログ(例えば、デオキシイノシン、2-アミノプリン塩基を有するヌクレオシドなど)の対形成を含む。三重鎖に関しては、この用語は、三重鎖が完全にマッチした二重鎖および第3鎖からなり、ここで、全てのヌクレオチドが、完全にマッチした二重鎖の塩基対とのフーグスティーンまたは逆フーグスティーン

10

20

30

40

50

会合を受けることを意味する。逆に、二重鎖におけるタグとオリゴヌクレオチドとの間の「ミスマッチ」は、二重鎖または三重鎖におけるヌクレオチドの対またはトリプレット(triplet)がワトソン-クリックおよび/またはフーグスティーンおよび/または逆フーグスティーン結合を受け得ないことを意味する。

本明細書中で用いられる「ヌクレオシド」は、天然のヌクレオシドを含み、これは2'-デオキシ形態および2'-ヒドロキシル形態(例えば、KornbergおよびBaker、DNA Replication、第2版(Freeman, San Francisco, 1992)に記載される)を含む。ヌクレオシドに関して「アナログ」は、改変された塩基部分および/または改変された糖部分を有する合成ヌクレオシド(例えば、Scheit, Nucleotide Analogs(John Wiley, New York, 1980); Uhlmann およびPeyman, Chemical Reviews, 90:543-584(1990)などに記載される)などを含む。ただし、それらは特異的にハイブリダイズし得る。このようなアナログは、結合特性を増強する、複雑度を減少させる、特異性を増大させるなどのために設計された合成ヌクレオシドを含む。

本明細書中で使用される、ポリヌクレオチドに関する「配列決定」または「ヌクレオチド配列を決定すること」は、ポリヌクレオチドの部分的ならびに全配列情報の決定を含む。すなわち、この用語は、標的ポリヌクレオチドについての配列比較、フィンガープリンティング、および同様のレベルの情報、ならびに標的ポリヌクレオチドにおけるヌクレオシド、通常各ヌクレオシドの明確な同定および順序付けを含む。この用語はまた、標的ポリヌクレオチド内の4つの型のヌクレオチドのうちの1つ、2つ、または3つの同一性、順序、および位置の決定を含む。例えば、いくつかの実施態様において、配列決定は、標的ポリヌクレオチド「CATCGC」内の単一の型のヌクレオチド(例えば、シトシン)の順序および位置を同定することによりもたらされ得、その結果、その配列は、二進符号(例えば、「C-(Cでない)-(Cでない)-C-(Cでない)-C...」について「100101...」)などとして示される。

本明細書中で使用される、ポリヌクレオチドの集団に関する用語「複雑度」は、この集団に存在する分子の異なる種の数意味する。

発明の詳細な説明

本発明は、1つ以上の標的ポリヌクレオチドの末端(単数または複数)に特異的にハイブリダイズするコードアダプターの連結を含む。特異的ハイブリダイゼーションが起こる領域についての配列情報は、このように連結されたコードアダプターのオリゴヌクレオチドタグを「解読」することにより得られる。本発明の1つの局面において、コードアダプターの複数セットは、付着(staggered)切断点で標的ポリヌクレオチドに連結され、その結果コードアダプターは、標的ポリヌクレオチドの複数の部分のそれぞれから配列情報を提供する。このような部分は、ばらばらであり得るか、重複し得るか、または連続的であり得る;しかし、好ましくは、この部分は連続的であり、そして個々の部分の長さの和に等しいヌクレオチド配列の同定を共に可能にする。この局面において、コードアダプターの単一の連結、それに続く連結されたアダプターのタグを「解読すること」による同定のみが存在する。本発明の別の局面において、コードアダプターは、以下により十分に記載される、連結、同定、および切断の反復サイクルを含むプロセスにおいて同定工程として使用される。

後者の実施態様において、本発明は、その認識部位がそれらの切断部位から離れているヌクレアーゼを使用する。好ましくは、このようなヌクレアーゼは、II型制限エンドヌクレアーゼである。このヌクレアーゼは、コードアダプターが連結される標的ポリヌクレオチド上に突出鎖を作製するために使用される。本発明の所定の実施態様で得られる配列情報量は、このようなヌクレアーゼがどれくらい使用されるか、および切断の際にどれくらいの長さの突出鎖が生じるかに部分的に依存する。

本発明の重要な局面は、多くの標的ポリヌクレオチドを並行して配列決定する能力である。この局面において、本発明の方法は、以下の工程を包含する:(a)レパートリー由来の各第1のオリゴヌクレオチドタグが第1の最小にクロスハイブリダイズするセットから選択されるように、タグのレパートリー由来の第1のオリゴヌクレオチドタグを、ポリヌ

クレオチドの集団中の各ポリヌクレオチドに付着させる工程；(b) 集団中の実質に全ての異なるポリヌクレオチドが、付着された異なる第1のオリゴヌクレオチドタグを有するように、ポリヌクレオチドの集団をサンプリングする工程；(c) 1つ以上のコードアダプターを集団中の各ポリヌクレオチドの末端に連結する工程であって、各コードアダプターは、第2の最小にクロスハイブリダイズするセットから選択される第2のオリゴヌクレオチドタグおよび集団のポリヌクレオチドの突出鎖に相補的な突出鎖を有する、工程；(d) 第1のオリゴヌクレオチドタグとそれらのそれぞれの相補物とを特異的にハイブリダイズさせることにより集団のポリヌクレオチドを選別する工程であって、それぞれの相補物は、1つ以上の固相支持体上の空間的に別々の領域において実質的に同一のオリゴヌクレオチドの均一の集団として付着される、工程；および(e) タグ相補物を1つ以上のコードアダプターの各第2のオリゴヌクレオチドタグに特異的にハイブリダイズさせることにより、ポリヌクレオチドの上記の突出鎖における1つ以上のヌクレオチドを同定する工程。この実施態様において、1つ以上のコードアダプターが、ポリヌクレオチドが第1のオリゴヌクレオチドタグにより固相支持体上に選別される前または後のいずれかに、ポリヌクレオチド末端に連結され得る。好ましい実施態様において、コードアダプターは、コードアダプターがポリヌクレオチドから切断され、そしてポリヌクレオチドが配列同定後に短縮されるのを可能にする、II型制限エンドヌクレアーゼ部位を含む。

好ましい実施態様によれば、この方法は、連結、同定、および切断の反復サイクルをさらに含み、その結果1つ以上のヌクレオチドが各サイクルで同定される。好ましくは、2～6のヌクレオチドが同定され、そしてそれらの順序が各サイクルで決定される。

連結および切断のサイクルを使用しない配列分析

本発明に従う連結および切断のサイクルを使用しない配列分析は、図1A～1Eの実施態様により示される。この実施態様において、k個の標的ポリヌクレオチドは、以下、およびBrenner、国際特許出願PCT/US95/12791(WO 96/12041)およびPCT/US96/09513(WO 96/41011)に記載のように調製される。すなわち、サンプルは、小さな「t」により示されるオリゴヌクレオチドタグに結合したポリヌクレオチドの集団から採取される。これらのタグは、選別のためのオリゴヌクレオチドタグまたは「第1の」オリゴヌクレオチドタグと呼ばれる。サンプルのタグ-ポリヌクレオチド結合体は、図1Aにおける(14)～(18)により示される、結合体の1～k個の集団を得るために、例えば、ポリメラーゼ連鎖反応(PCR)またはクローニングにより増幅される。好ましくは、(小さな「t」)タグの反対の結合体の末端は、1つ以上のアダプターを連結するために調製され、これらのそれぞれは、その切断部位がその認識部位から離れている、ヌクレアーゼの認識部位を含む。示された実施態様において、「切断アダプター」と本明細書中で呼ばれる3つのこのようなアダプターが使用される。使用されるこのようなアダプターの数は、所望される配列情報の量、適切な勢力範囲および切断特徴を有するII型ヌクレアーゼの利用可能性などを含むいくつかの要因に依存する。好ましくは、1～3つの切断アダプターが使用され、そしてこのアダプターは、切断後に少なくとも4つのヌクレオチドの突出末端を作製し得る、異なるII型ヌクレアーゼを適応させるための設計である。

本発明の方法が、cDNA集団の配列決定をするサインに適用される場合、切断アダプターの連結前に、タグ-ポリヌクレオチド結合体は、高頻度の認識部位を有する制限エンドヌクレアーゼ(例えば、TaqI、AluI、HinP1I、DpnII、NlaIIIなど)で切断され得る。平滑末端を残すAluIのような酵素について、付着末端は、例えば、上記で引用されたBrenner、国際特許出願PCT/US95/12791およびKuijperら、Gene, 112:147-155(1992)に記載のようにT4 DNAポリメラーゼを用いて産生され得る。標的ポリヌクレオチドがTaqIでの切断により調製される場合、以下の末端は連結に利用可能である：

cgannnn...-3'

tnnnn...-5'

従って、3つの切断アダプターの例示的なセットが、以下のように構築され得る：

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

ここで、切断アダプター(1)、(2)、および(3)は大文字で示され、ヌクレアーゼ BbsI、BbvI、およびBsmFIのそれぞれの認識部位は下線で示され、そして5'リン酸は「p」として示される。標的ポリヌクレオチドの二重下線部分は、連結および切断後の突出鎖の位置を示す。全ての場合において、標的ポリヌクレオチドは、4つのヌクレオチドの5'突出鎖が残される。明らかに、多くの異なる実施態様は、ヌクレアーゼの異なる数および種類を用いて構築され得る。Brenner、米国特許第5,599,675号およびWO 95/27080に記載されるように、好ましくは、切断前に、内部BbsI、BbvI、およびBsmFI部位は、標的ポリヌクレオチドの内部部位での所望でない切断を妨げるために、例えば、メチル化によりブロックされる。

示された実施態様に戻って、切断アダプター A_1 、 A_2 、および A_3 は、図1Bに示される結合体を得るために、 k 個の標的ポリヌクレオチドに対して1:1:1の濃度比で連結され(20)、その結果、タグ-ポリヌクレオチド結合体の各集団内に、付着された A_1 、 A_2 、および A_3 を有するほぼ等しい数の結合体が存在する。連結(20)後に、標的ポリヌクレオチドは、切断アダプターの各ヌクレアーゼで連続的に切断され、そして1セットのコードアダプターに連結される。最初に、標的ポリヌクレオチドは、切断アダプター A_1 のヌクレアーゼで切断され(22)、この後、第1のセットのコードアダプターが、得られた突出鎖に連結される。この切断は、連結に利用可能な各型(すなわち、 t_1 、 t_2 、... t_k)の標的ポリヌクレオチドの約1/3をもたらす。好ましくは、コードアダプターは、突出鎖の全ての可能な配列をまとめて含む、1つ以上のアダプター混合物として適用される。その突出鎖が標的ポリヌクレオチドの突出鎖と完全にマッチした2重鎖を形成するコードアダプターのみが連結して、コード結合体を形成するように、反応条件が選択される(28)、(30)、および(32)(図1C)。小文字付きの大文字「T」は、唯一のオリゴヌクレオチドタグが標識のためにコードアダプターにより保持されることを示す。コードアダプターにより保持されるオリゴヌクレオチドタグは、時々、標識をコードアダプターに送達するためのタグまたは「第2の」オリゴヌクレオチドタグと呼ばれる。以下により十分に記載されるように、選別のために使用される1本鎖オリゴヌクレオチドタグは、好ましくは、4つのヌクレオチドのうちの3つのみからなり、その結果、T4 DNAポリメラーゼ「ストリッピング(stripping)」反応(例えば、Kuijperら(上記))が、固相支持体へのロードのための標的ポリヌクレオチドを調製するために使用され得る。他方では、標識を送達するために使用されるオリゴヌクレオチドタグは、全てのヌクレオチドからなり得る。

上記のように、コードアダプターは、突出鎖(24)およびオリゴヌクレオチドタグ(26)を含む。従って、 t_1 -ポリヌクレオチド結合体の「 A_1 」切断が以下の末端：

5'-...nnnnnnnnnn

3'-...nnnnnnnnnnacct

を生じる場合、オリゴヌクレオチドタグ $T_{2,4}$ は、以下の構造(配列番号1)を有し得る：
 tggattctagagagagagagagagagag-3'

aagatctctctctctctctctctctc

ここで、2本鎖部分は、唯一のタグ相補物と完全にマッチした3重鎖を形成し、そして全ての他のタグ相補物と少なくとも6つのミスマッチで3重鎖を形成する、48個(=12個の

10

20

30

40

50

ヌクレオチド位置 × 4 種類のヌクレオチド) の 2 本鎖 20 マーオリゴヌクレオチドタグの 1 セットのうちの 1 つであり得る。本実施例におけるコードアダプターは、計 768 (3 × 256) メンバーの 1 つ以上の混合物中の標的ポリヌクレオチドに連結され得る。必要に応じて、コードアダプターはまた、4 つのヌクレオチド配列「ttct」が突出鎖とオリゴヌクレオチドタグとの間のスパーサーとして作用する、上記の実施例において示されるようなスパーサー領域を含み得る。

第 1 のセットのコードアダプターの連結 (28)、(30)、および (32) 後、タグ-ポリヌクレオチド結合体は、切断アダプター A₂ のヌクレアーゼで切断され (34)、この後、第 2 のセットのコードアダプターが結合体を形成するために適用される (36)、(38)、および (40) (図 1D)。最後に、タグ-ポリヌクレオチド結合体は、切断アダプター A₃ のヌクレアーゼで切断され (42)、この後、第 3 のセットのコードアダプターが結合体を形成するために適用される (44)、(46)、および (48) (図 1E)。コードアダプターの連続する切断および連結の完了後、混合物は、以下により十分に記載され、そして Brenner (例えば、PCT/US95/12791 または PCT/US96/09513) により教示されるように、オリゴヌクレオチドタグ t₁ ~ t_k を介して 1 つ以上の固相支持体にロードされる (50)。単一の標的ポリヌクレオチドが分析される場合、明らかに複数のオリゴヌクレオチドタグ、t₁、t₂、... t_k は必要ない。このような実施態様において、選別が必要とされない場合、ピオチンまたは同様の部分が、ポリヌクレオチド-コードアダプター結合体を固定するために使用され得る。また、切断、連結、および固相支持体へのローディング工程の順序は、実行される特定の実施態様に依存する。例えば、タグ-ポリヌクレオチド結合体は、最初に、固相支持体にロードされ、続いて切断アダプターの連結、その切断、およびコードアダプターの連結が行われ得る；または切断アダプターが、最初に連結され、続いてローディング、切断、およびコードアダプターの連結が行われ得るなどである。

コードアダプターが本発明に従って標的ポリヌクレオチドの末端に連結された後、配列情報は、コードアダプターのオリゴヌクレオチドタグとそれらの個々のタグ相補物との間に完全にマッチした 2 重鎖および / または 3 重鎖の形成を可能にする条件下で、個々にまたは混合物としてのいずれかで固定化標的ポリヌクレオチドに標識タグ相補物を連続的に適用することにより得られる。混合物の数および複雑度は、使用される標識システムの型、部分 (その配列が同定される) の長さ、複雑度を低減するアナログが使用されるか否かなどを含む、いくつかの要因に依存する。好ましくは、図 1a ~ 1e に示される実施態様について、単一の蛍光色素が、48 (= 3 × 16) のタグ相補物のそれぞれを標識するために使用される。タグ相補物は、標的ポリヌクレオチドの各 4 つのヌクレオチド部分のヌクレオチドを同定するために個々に適用される (すなわち、計 48 について各 12 位置に対する 4 つのタグ相補物)。明らかに、異なる長さの部分が、タグ相補物の異なる数を必要とし、例えば、本実施例によれば、5 ヌクレオチド部分は 20 のタグ相補物を必要とし、2 ヌクレオチド部分は 8 タグ相補物を必要とするなどである。タグ相補物は、完全にマッチした 2 重鎖のみが形成されるような十分にストリンジェントな条件下で適用され、特異的にハイブリダイズしたタグ相補物上の蛍光標識からのシグナルが測定され、そしてタグ相補物は、次の混合物が適用され得るようにコードタグから洗浄される。16 個のタグ相補物が、標的配列の 4 マー部分の以下の配列と 1 対 1 の対応を有する：

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

ここで「N」はヌクレオチド A、C、G、または T のいずれか 1 つである。従って、各ヌクレオチド位置に対して 4 つの別々の疑問符 (interrogation)、各種類のヌクレオチドの 1 つが作製される。この実施態様は、ヌクレオチド決定の増大した信頼性の代わりに、有意な程度の重複性 (計 16 個のタグ相補物が 4 つのヌクレオチドを同定するために使用さ

10

20

30

40

50

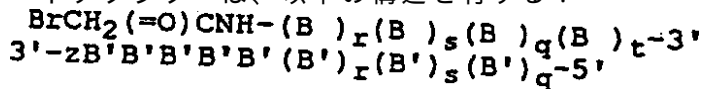
さらなる配列情報は、Brenner、国際特許出願PCT/US95/03678(WO 95/27080)に開示される「マルチステッピング(multi-stepping)」プロセスに類似するプロセスにおいて上記の実施態様を用いて得られ得る。この実施態様において、本明細書中で「ステッピングアダプター」といわれる第4のアダプターは、例えば3:1:1:1の濃度比で、切断アダプターA₁、A₂、およびA₃と共に標的ポリヌクレオチドの末端に連結される。従って、利用可能な末端の約半分は、ステッピングアダプターに連結される。ステッピングアダプターは、その区域(reach)(以下に定義)が、切断アダプターA₁、A₂、およびA₃を介して決定される配列の末端で標的ポリヌクレオチドの切断を可能にするように配置されるII型ヌクレアーゼの認識部位を含む。上記の切断アダプターのセットと共に使用され得るステッピングアダプターの例は、以下の通りである：

ここで、上記のように、ヌクレアーゼ（この場合、BpMI）の認識部位は一本下線が引かれ、そして切断部位のヌクレオチドは二重下線が引かれる。ステップングアダプターのヌクレアーゼで切断された標的ポリヌクレオチドは、切断アダプターA₁、A₂、およびA₃に含まれるものと同じかまたは異なるヌクレアーゼ認識部位を含み得る、さらなるセットの切断アダプターA₄、A₅、およびA₆に連結され得る。コードアダプターの拡大したセットが必要とされるか否かは、切断および連結反応がシグナル測定装置に許容され得るか否かに依存する。上記のように、シグナル測定に関して酵素反応を最小化することが所望される場合、コードアダプターのさらなるセットが使用されなければならない。すなわち、上記の768のオリゴヌクレオチドタグおよびタグ相補物が、4つのヌクレオチドそれぞれの突出鎖を産生する6つの切断反応に必要とされる場合、1536のオリゴヌクレオチドタグおよびタグ相補物（64のタグ相補物それぞれの24の混合物）が必要とされる。A₁、A₂、およびA₃と同じヌクレアーゼ認識部位を有し、そして上記に示されるステップングアダプターと共に使用され得る、例示的な切断アダプターA₄、A₅、およびA₆は、以下の通りである：

一旦コードアダプターが連結されると、標的ポリヌクレオチドは、Brenner、国際特許出願PCT/US95/12791(WO 96/12041)に開示されるように、固相支持体、好ましくは微粒子へのローディングのために調製される。簡単には、選別のためのオリゴヌクレオチドタグは、T4 DNAポリメラーゼでの「ストリッピング」反応（例えば、Kuijperら（上記））を用いて1本鎖にされる。1本鎖オリゴヌクレオチドタグは、特異的にハイブリダイズされ、そして微粒子上のそれらのタグ相補物に連結される。次いで、ロードされた微粒子は、連続送達、特異的ハイブリダイゼーション、およびコードアダプターに対する標識化タグ相

補物の除去を可能にする、Brenner（上記）に記載のような装置において分析される。

コードアダプターが一度だけ標的ポリヌクレオチド（または標的ポリヌクレオチドの集団）に連結される実施態様において、本発明に従って使用され得る連結のいくつかの非酵素的テンプレート駆動方法が存在する。このような連結方法は、Shabarova, Biochimie 70: 1323-1334(1988) ; Dolinnayaら、Nucleic Acids Research, 16:3721-3738(1988) ; Letsingerら、米国特許第5,476,930号 ; Gryaznovら、Nucleic Acids Research, 22:2366-2369(1994) ; Kangら、Nucleic Acids Research, 23:2344-2345(1995) ; Gryaznovら、Nucleic Acids Research, 21:1403-1408(1993) ; Gryaznov、米国特許第5,571,677号に開示されるもの、および同様の参考文献を含むが、それらに限定されない。好ましくは、非酵素的連結は、Letsingerら（上記）の方法により行われる。この方法において、3' プロモアセチル化末端を有するコードアダプターは、相補的突出鎖およびその5' 末端にチオホスホリル基を有するポリヌクレオチドと反応される。このような化学的性質を使用する例示的なコードアダプターは、以下の構造を有する：



ここで、BおよびB' はヌクレオチドおよびそれらの相補物であり、そしてz、r、s、q、およびtは以下に記載されるとおりである。Br、C、H、およびNはそれらの通常の化学的意味を有する。上記の参考文献において説明されるように、テンプレート駆動反応において、3' プロモアセチル化オリゴヌクレオチドは、水性条件下で5' チオホスホリル基を有するオリゴヌクレオチドと自発的に反応して、チオホスホリルアセチルアミノ連結を形成する。チオホスホリル基は、Kangら（上記）に記載されるように、アデノシン-5'-O-(1-チオトリホスフェート)（すなわち、g-S-ATP）の存在下でT4キナーゼでの処理により標的ポリヌクレオチドの5' 水酸基に容易に付着される。

連結および切断のサイクルを用いる配列分析

コードアダプターは、連結、同定、および切断のサイクルの反復を含むDNA配列決定のアダプターに基づく方法（例えば、Brenner、米国特許第5,559,675号およびPCT公報第WO 95/27080号に記載される方法）において使用され得る。簡単には、このような方法は、以下の工程を含む：（a）コードアダプターをポリヌクレオチド末端に連結する工程であって、コードアダプターは、その切断部位がその認識部位とは離れている、ヌクレアーゼのヌクレアーゼ認識部位を有する、工程；（b）ポリヌクレオチド末端の1つ以上のヌクレオチドを、それに連結されるコードアダプターの正体により同定する工程；（c）コードアダプターのヌクレアーゼ認識部位を認識するヌクレアーゼでポリヌクレオチドを切断し、その結果ポリヌクレオチドは1つ以上のヌクレオチドにより短くされる、工程；および（d）ポリヌクレオチドの前記のヌクレオチド配列が決定されるまで、前記の工程（a）から（c）を反復する工程。同定工程において、タグ相補物の連続セットは、上記のように、標的ポリヌクレオチド末端に連結されたコードアダプターにより保持されるそれぞれのタグに特異的にハイブリダイズされる。ポリヌクレオチドの突出鎖におけるヌクレオチドの型および配列は、上記のように、特異的にハイブリダイズされたタグ相補物およびタグ相補物が生じたセットにより保持される標識により同定される。

オリゴヌクレオチドタグおよびタグ相補物

オリゴヌクレオチドタグは、本発明の好ましい実施態様において2つの異なる目的のために使用される：オリゴヌクレオチドタグは、混合物から分析のための同一ポリヌクレオチドの均一集団へ多数のポリヌクレオチド（例えば、数千から数10万）を選別するために、Brenner、国際特許出願PCT/US95/12791およびPCT/US96/09513(WO 96/12041およびWO 96/41011)に記載されるように使用され、そしてそれらは、標識を数十から数千の範囲からなるコードアダプターに送達するために使用される。前者の使用のために、多くの数すなわちレパートリーのタグが代表的に必要とされ、従って、個々のオリゴヌクレオチドタグの合成は問題である。これらの実施態様において、タグの組合せ合成が好ましい。他方では、極端に多くのレパートリーのタグが、例えば、標識をコードアダプターに送達するため

10

20

30

40

50

に必要とされない場合、オリゴヌクレオチドタグの最小にクロスハイブリダイズするセットは、組合せ的に合成されるのと同様に別々に合成され得る。

Brenner (上記) に記載のように、オリゴヌクレオチドの最小にクロスハイブリダイズするセットのヌクレオチド配列は、単純なコンピュタープログラム (例えば、そのソースコード (source code) が Appendix I および II に列挙されるプログラムによって例示される) によって簡便に数値化される。同様のコンピュタープログラムは、本発明の任意の実施態様についてオリゴヌクレオチドの最小にクロスハイブリダイズするセットを列挙するために容易に書かれる。以下の表 I は、示された長さおよび数のヌクレオチド差異の数についての最小にクロスハイブリダイズするオリゴヌクレオチドのセットのサイズに関するガイダンスを提供する。上記のコンピュタープログラムを使用して、数値を生成した。

表 I

4つのヌクレオチドからなるワードの最小にクロスハイブリダイズするセット

オリゴヌクレオチド ワードの長さ	最小にクロスハイブリ ダイズするセットの オリゴヌクレオチド間 のヌクレオチド差	最小にクロスハイブリ ダイズするセット の最大の大きさ	3ワードを有する レパートリーの 大きさ	4ワードを有する レパートリーの 大きさ
4	3	11	1331	14,641
6	4	25	15,625	3.9×10^5
6	5	4	64	256
8	4	225	1.14×10^7	
8	5	56	1.75×10^5	
8	6	17	4913	
12	8	62		

数百から数千のオリゴヌクレオチドまたは数万のオリゴヌクレオチドでさえも含むセットは、例えば、Frankら、米国特許第4,689,405号; Frankら、Nucleic Acids Research, 11: 4365-4377(1983); Matsonら、Anal. Biochem., 224:110-116(1995); Fodorら、国際出願PCT/US93/04145(WO 93/22684); Peaseら、Proc. Natl. Acad. Sci., 91:5022-5026(1994); Southernら、J. Biotechnology, 35:217-227(1994); Brennan、国際出願PCT/US94/05896(WO 94/27719); Lashkariら、Proc. Natl. Acad. Sci., 92:7912-7915(1995)などに開示されるような、種々の並行合成アプローチにより直接合成され得る。

好ましくは、混合物中のタグ相補物は、組合せ的に合成されたかまたは個々に合成されたかにかかわらず、完全にマッチしたハイブリッドが類似したまたは実質的に同一の融解温度を有するような互いに類似した二重鎖または三重鎖安定性を有するように選択される。これは、ミスマッチタグ相補物が、コードアダプターに適用される場合 (例えば、ストリンジェントな条件下で洗浄することにより)、完全にマッチしたタグ相補物とより容易に区別されることを可能にする。組合せ的に合成されたタグ相補物について、最小にクロスハイブリダイズするセットは、このセット中の全ての他のサブユニットとほぼ同等の二重鎖安定性に寄与するサブユニットから構築され得る。このような選択を行うためのガイダンスは、最適PCRプライマーを選択し、そして二重鎖安定性を計算するための公開された技術 (例えば、Rychlikら、Nucleic Acids Research, 17:8543-8551(1989)および18:6409-6412(1990); Breslauerら、Proc. Natl. Acad. Sci., 83:3746-3750(1986); Wetmur, Crit. Rev. Biochem. Mol. Biol., 26:227-259(1991)など) により提供される。より少数のオリゴヌクレオチドタグが、例えば、標識をコードアダプターに送達するために必要とされる場合、Appendix I および II のコンピュタープログラムを使用して、直接使用され得る (すなわち、「センテンス (Sentence)」への連鎖なく) 最小にクロスハイブリダイズするセットのオリゴヌクレオチドの配列を作成し、そして列挙し得る。このようなリストは

、さらなる最小にクロスハイブリダイズするセットを形成するために、さらなる基準（例えば、GC含有量、ミスマッチの分布、理論的融解温度など）についてさらにスクリーニングされ得る。

より短いタグ（例えば、約30ヌクレオチド以下）について、RychlikおよびWetmurにより記載されるアルゴリズムが二重鎖安定性を計算するために好ましく、そしてより長いタグ（例えば、約30～35ヌクレオチド以上）について、編集者Brown、ICN-UCLA Symp.Dev.Biol.,第23巻（Academic Press, New York, 1981）においてSuggsら、683-693頁により開示されるアルゴリズムが簡便に使用され得る。明らかに、本発明の範囲内で最小にクロスハイブリダイズするサブユニットのセットを設計するために当業者に利用可能な多くのアプローチがある。例えば、サブユニットが組み立てられた場合、末端ヌクレオチドの異なる塩基スタッキングエネルギーの影響を最小にするために、同じ末端ヌクレオチドを有するサブユニットが提供され得る。このように、サブユニットが連結される場合、全ての隣接末端ヌクレオチドの塩基スタッキングエネルギーの合計は同じであり、それによりタグ融解温度の変化を低減するか、または除去する。

複数サブユニットタグにおいて、以下のイタリック体で示される末端ヌクレオチドの「ワード」はまた、完全なマッチがそれと任意の他のタグ相補物上の同様の末端「ワード」との間でいつも形成されるように、タグの各末端に付加され得る。このように増大したされたタグは以下の形態を有する：

<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> '

ここでプライム記号（'）が付いた*W*'は相補物を示す。タグ末端は完全にマッチした二重鎖を常に形成し、全てのミスマッチワードは内部ミスマッチであり、それにより、それらの末端でミスマッチワードを他に有するタグ相補物二重鎖の安定性を低減する。内部ミスマッチを有する二重鎖が、末端で同じミスマッチを有する二重鎖より著しく安定でないことは周知である。

オリゴヌクレオチドタグが選別のために用いられる場合、最小にクロスハイブリダイズするセットの好ましい実施態様は、そのサブユニットが4つの天然ヌクレオチドのうち3つから構成される実施態様である。以下により十分に議論されるように、オリゴヌクレオチドタグの1つの型のヌクレオチドの非存在は、標的ポリヌクレオチドがDNAポリメラーゼの5' 3'エキソヌクレアーゼ活性の使用により固相支持体にロードされることを可能にする。以下は、A、G、およびTからなる群より選択される4つのヌクレオチドを含む各サブユニットの例示的な最小にクロスハイブリダイズするセットである：

表II

ワード：	<i>w</i> ₁	<i>w</i> ₂	<i>w</i> ₃	<i>w</i> ₄
配列：	GATT	TGAT	TAGA	TTTG
ワード：	<i>w</i> ₅	<i>w</i> ₆	<i>w</i> ₇	<i>w</i> ₈
配列：	GTAA	AGTA	ATGT	AAAG

このセットにおいて、各メンバーは、全ての他のメンバーの相補物と3つのミスマッチ塩基を有する二重鎖を形成する。

オリゴヌクレオチドタグが標識をコードアダプターに送達するために使用される場合、4つ全てのヌクレオチドが使用される。

本発明のオリゴヌクレオチドタグおよびそれらの相補物は、例えば、以下の参考文献：BeaucageおよびIyer, Tetrahedron, 48:2223-2311(1992)；Molkoら、米国特許第4,980,460号；Kosterら、米国特許第4,725,677号；Caruthersら、米国特許第4,415,732号；同第4,458,066号；および同第4,973,679号などに開示される標準的な化学（例えばホスホールア

10

20

30

40

50

ミダイト化学)を用いて、自動化DNA合成機(例えば、Applied Biosystems, Inc.(Foster City, California)モデル392または394DNA/RNA Synthesizer)で簡便に合成される。例えば、非天然骨格基(例えば、ペプチド核酸(PNA)、N3' P5' ホスホールアミドなど)をもたらず代替の化学もまた使用され得る。いくつかの実施態様において、タグは、酵素によるプロセッシングまたは操作を可能にする天然に生じるヌクレオチドを含み得るが、一方対応するタグ相補物は、選別の間により安定な二重鎖の形成を促進する、非天然ヌクレオチドアナログ(例えば、ペプチド核酸または同様の化合物)を含み得る。標識をコードアダプターに送達するために使用されるタグの場合、オリゴヌクレオチドタグおよびタグ相補物の両方は、連結が化学的または酵素的のいずれかで起こり得るのであれば、非天然ヌクレオチドまたはアナログから構築され得る。

10

タグの2本鎖形態は、相補鎖を別々に合成し、続いて二重鎖形成を可能にする条件下で混合することにより作られ得る。あるいは、2本鎖タグは、プライマー結合部位として働く既知のオリゴヌクレオチド配列に連結される1本鎖レパートリーを最初に合成することにより形成され得る。次いで、第2鎖は、1本鎖レパートリーとプライマーとを組合せ、そしてポリメラーゼで伸長することにより合成される。この後者のアプローチは、Oliphantら、Gene, 44:177-183(1986)に記載される。次いで、このような二重鎖タグは、本発明に従う標的ポリヌクレオチドの選別および操作のために、標的ポリヌクレオチドと共にクローニングベクターに挿入され得る。

増強した結合特徴を有するヌクレオチド(例えば、PNAまたはオリゴヌクレオチドN3' P5' ホスホールアミドなど)から構成されるタグ相補物が使用される場合、選別は、タグを1本鎖にするためにDNAポリメラーゼの3' 5' エキソヌクレアーゼ活性を使用する「ストリッピング」反応に対する代替として、天然のヌクレオチドを含むタグとそれらのPNAまたはホスホールアミダイト相補物との間のDループの形成により実行され得る。

20

選別のためのオリゴヌクレオチドタグは、12~60のヌクレオチドまたは塩基対の長さの範囲であり得る。好ましくは、オリゴヌクレオチドタグは、18~40のヌクレオチドまたは塩基対の長さの範囲であり得る。より好ましくは、オリゴヌクレオチドタグは、25~40のヌクレオチドまたは塩基対の長さの範囲であり得る。好ましいおよびより好ましい数のサブユニットに関して、これらの範囲は、以下のように表され得る：

表III

好ましい実施態様におけるタグ中のサブユニットの数

30

サブユニット

中のモノマー オリゴヌクレオチドタグ中のヌクレオチド

	(12~60)	(18~40)	(25~40)
3	4~20サブユニット	6~13サブユニット	8~13サブユニット
4	3~15サブユニット	4~10サブユニット	6~10サブユニット
5	2~12サブユニット	3~8サブユニット	5~8サブユニット
6	2~10サブユニット	3~6サブユニット	4~6サブユニット

40

最も好ましくは、選別のためのオリゴヌクレオチドタグは1本鎖であり、そして特異的ハイブリダイゼーションは、タグ相補物とのワトソン-クリック型塩基対形成を生じる。

好ましくは、選別のための1本鎖オリゴヌクレオチドタグのレパートリーは少なくとも1000のメンバーを含み；より好ましくは、このようなタグのレパートリーは少なくとも10000のメンバーを含み；そして最も好ましくは、このようなタグのレパートリーは少なくとも10,000のメンバーを含む。

好ましくは、標識を送達するためのタグ相補物のレパートリーは少なくとも16のメンバーを含み；より好ましくは、このようなタグのレパートリーは、少なくとも64のメンバーを

50

含む。さらにより好ましくは、タグ相補物のこのようなレパートリーは、16～1024のメンバー（例えば、長さが2～5ヌクレオチドの突出鎖のヌクレオチドを同定するためのメンバー）を含む。最も好ましくは、タグ相補物のこのようなレパートリーは、64～256のメンバーを含む。所望のサイズのレパートリーは、例えば、Appendix IおよびIIのコンピュータプログラムの助けにより、所望のサイズのワードのセットもしくはサブユニットを直接作製することにより選択されるか、またはレパートリーは、1セットのワードを作製し、次いでこれを組合せ的合成スキームにおいて使用して所望のサイズのレパートリーを得ることにより形成される。好ましくは、標識を送達するための1本鎖タグ相補物の長さは、8～20である。より好ましくは、長さは9～15である。

三重鎖タグ

特異的ハイブリダイゼーションが三重鎖形成を介して生じる複数の実施態様において、タグ配列のコーディングは、二重鎖形成タグと同じ原理に従う；しかし、サブユニット配列の選択に対してさらなる制約がある。一般的に、フーグスティーン型の結合を介した第3鎖会合は、2本鎖標的におけるホモピリミジン-ホモプリン進路に沿って最も安定である。通常、塩基トリプレットは、T-A^{*}TまたはC-G^{*}Cモチーフ（ここで「^{*}」はワトソン-クリック型塩基対形成を示し、そして「^{*}」はフーグスティーン型結合を示す）において形成する；しかし、他のモチーフもまた可能である。例えば、フーグスティーン型塩基対形成は、鎖の状態および組成に依存して、第3鎖（フーグスティーン鎖）と第3鎖が結合する二重鎖のプリンリッチ鎖との間の平行および逆平行方向を可能にする。特定の実施態様において所望のように三重鎖安定性を最大にするか、さもなければ調節するために、適切な配列、方向、条件、ヌクレオチド型（例えば、リボースまたはデオキシリボースヌクレオシドが使用されるか否か）、塩基修飾（例えば、メチル化シトシンなど）を選択するための文献中に広範なガイダンスがある（例えば、Robertsら、Proc.Natl.Acad.Sci., 88:9397-9401(1991)；Robertsら、Science, 258:1463-1466(1992)；Robertsら、Proc.Natl.Acad.Sci., 93:4320-4325(1996)；Distefanoら、Proc.Natl.Acad.Sci., 90:1179-1183(1993)；Mergnyら、Biochemistry, 30:9791-9798(1991)；Chengら、J.Am.Chem.Soc., 114:4465-4474(1992)；BealおよびDervan、Nucleic Acids Research, 20:2773-2776(1992)；BealおよびDervan、J.Am.Chem.Soc., 114:4976-4982(1992)；Giovannangelisら、Proc.Natl.Acad.Sci., 89:8631-8635(1992)；MoserおよびDervan、Science, 238:645-650(1987)；McShanら、J.Biol.Chem., 267:5712-5721(1992)；Yoonら、Proc.Natl.Acad.Sci., 89:3840-3844(1992)；Blumeら、Nucleic Acids Research, 20:1777-1784(1992)；ThuongおよびHelene、Angew.Chem.Int.Ed.Engl. 32:666-690(1993)；Escudeら、Proc.Natl.Acad.Sci., 93:4365-4369(1996)など）。1本鎖または二重鎖タグをそれらの1本鎖または二重鎖相補物にアニーリングさせるための条件は周知である（例えば、Jiら、Anal.Chem. 65:1323-1328(1993)；Cantorら、米国特許第5,482,836号など）。選別における三重鎖タグの使用は、その相補物へのアニーリングのためにタグを暴露するための、ポリメラーゼを用いた「ストリッピング」反応を必要としない利点を有する。

好ましくは、三重鎖ハイブリダイゼーションを用いる本発明のオリゴヌクレオチドタグは2本鎖DNAであり、そして対応するタグ相補物は1本鎖DNAである。より好ましくは、5-メチルシトシンは、タグとその相補物との間で形成される三重鎖のpH安定性の範囲を広げるためにタグ相補物中のシトシンの代わりに使用される。三重鎖を形成するための好ましい条件は、上記の参考文献に充分に開示される。簡単には、ハイブリダイゼーションは、濃縮塩溶液（例えば、1.0M NaCl、1.0M酢酸カリウムなど）中でpH5.5未満で（または、5-メチルシトシンが使用される場合6.5）起こる。ハイブリダイゼーション温度は、タグの長さおよび組成に依存する；しかし、18～20マータグまたはそれより長いタグについては、室温でのハイブリダイゼーションが適切である。洗浄は、あまり濃縮されていない塩溶液（例えば、10mM酢酸ナトリウム、100mM MgCl₂、pH5.8）で室温で行われ得る。タグは、pH9.0の同様の塩溶液中のインキュベーションによりそれらのタグ相補物から溶出され得る。

三重鎖を形成するオリゴヌクレオチドタグの最小にクロスハイブリダイズするセットは、

AppendixIIのコンピュータプログラムまたは同様のプログラムにより作製され得る。2本鎖8マーワードの例示的なセットは、小文字の対応する相補物と共に大文字で以下に列挙される。このような各ワードは、3つの塩基対によりこのセット中の他の各ワードと異なる。

表IV

例示的な最小にクロスハイブリダイズする、2本鎖8マーワードのセット

5'-AAGGAGAG 3'-TTCCTCTC 3'-ttcctctc	5'-AAAGGGGA 3'-TTTCCCCT 3'-tttcccct	5'-AGAGAAGA 3'-TCTCTTCT 3'-tctcttct	5'-AGGGGGGG 3'-TCCCCCCC 3'-tccccccc
5'-AAAAAAGA 3'-TTTTTTTT 3'-tttttttt	5'-AAGAGAGA 3'-TTCTCTCT 3'-ttctctct	5'-AGGAAAAG 3'-TCCTTTTC 3'-tccttttc	5'-GAAAGGAG 3'-CTTCCTCT 3'-cttctctc
5'-AAAAAGGG 3'-TTTTTCCC 3'-tttttccc	5'-AGAAGAGG 3'-TCTTCTCC 3'-tcttctcc	5'-AGGAAGGA 3'-TCCTTCCT 3'-tccttcct	5'-GAAGAAGG 3'-CTTCTTCC 3'-cttcttcc
5'-AAAGGAAG 3'-TTTCCTTC 3'-tttccttc	5'-AGAAGGAA 3'-TCTTCCTT 3'-tcttcctt	5'-AGGGGAAA 3'-TCCCCTTT 3'-tccccttt	5'-GAAGAGAA 3'-CTTCTCTT 3'-cttctctt

10

表V

三重鎖のタグ相補物と三重鎖を形成する、 種々の2本鎖タグのレパートリーサイズ

20

オリゴヌクレオチド のワードの長さ	最小にクロスハイブリダイズ するセットのオリゴヌクレオ チド間のヌクレオチド差	最小にクロスハイブ リダイズするセット の最大の大きさ	4ワードを有する レパートリーの大きさ	5ワードを有する レパートリーの大きさ
4	2	8	4096	3.2×10^4
6	3	8	4096	3.2×10^4
8	3	16	6.5×10^4	1.05×10^6
10	5	8	4096	
15	5	92		
20	6	768		
20	7	484		
20	8	189		
20	9	30		

30

アダプターの合成および構造

40

コードアダプターおよび切断アダプターは、例えば、以下の参考文献：BeaucageおよびIyer, Tetrahedron, 48:2223-2311(1992) ; Molkoら、米国特許第4,980,460号 ; Kosterら、米国特許第4,725,677号 ; Caruthersら、米国特許第4,415,732号 ; 同第4,458,066号 ; および同第4,973,679号などにおいて開示されるような標準的な化学（例えば、ホスホールアミダイト化学）を用いて自動化DNA合成機で簡便に合成される。例えば、非天然骨格基（例えば、ホスホロチオエート、ホスホールアミダイトなど）をもたらす代替の化学もまた、得られたオリゴヌクレオチドが特定の実施態様において使用される連結および/または切断試薬に適合するのであれば使用され得る。代表的に、相補鎖の合成後、この鎖は、2本鎖アダプターを形成するために組み合わされる。コードアダプターの突出鎖は、全ての可能な配列が突出部分部分において示されるように混合物として合成され得る。このよう

50

な混合物は、周知の技術（例えば、Teleniusら、Genomics, 13:718-725(1992) ; Welshら、Nucleic Acids Research, 19:5275-5279(1991) ; Grothuesら、Nucleic Acids Research、21:1321-1322(1993) ; Herley, 欧州特許出願90304496.4(EP公報第395398)などに開示される）を用いて容易に合成される。一般的に、これらの技術は、単純に、複数のヌクレオチドを導入することが所望されるカップリング工程間で活性化モノマーの混合物の増大するオリゴヌクレオチドへの適用を要求する。上記のように、いくつかの実施態様において、アダプターの複雑度を低減することが所望され得る。これは、例えば、Kong Thoo Linら、Nucleic Acids Research, 20:5149-5152または米国特許第5,002,867号 ; Nicholsら、Nature, 369:492-493(1994)などにおいて教示されるように、複雑度を低減したアナログ（例えば、デオキシイノシン、2-アミノプリンなど）を用いて達成され得る。

10

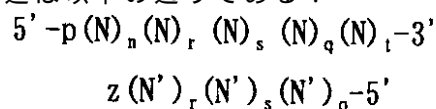
いくつかの実施態様において、自己相補性領域を含む単一のポリヌクレオチドとしてコードアダプターまたは切断アダプターを合成することが所望され得る。合成後、自己相補性領域はアニールされて、1つの末端の突出鎖およびもう1つの末端の1本鎖ループとアダプターを形成させる。好ましくは、このような実施態様において、ループ領域は、約3～10のヌクレオチドまたは米国特許第4,914,210号に開示されるような他の匹敵する連結部分（例えば、アルキルエーテル基）を含み得る。多くの技術は、以下に引用される参考文献に議論されるように、標識のために反応基の塩基またはヌクレオシド連結への付着に利用可能である。

従来のリガーゼが以下により十分に記載されるように本発明において使用される場合、アダプターの5'末端は、いくつかの実施態様においてリン酸化され得る。ーリン酸塩は、化学的またはキナーゼを用いて酵素的のいずれかで第2のオリゴヌクレオチドに付着され得る（例えば、Sambrookら、Molecular Cloning: A Laboratory Manual, 第2版（Cold Spring Harbor Laboratory, New York, 1989））。化学的リン酸化は、HornおよびUrdea、Tetrahedron Lett., 27:4705(1986)により記載され、そして開示されたプロトコルを行うための試薬は、市販されている（例えば、Clontech Laboratories(Palo Alto, California)からの5' Phosphate-ONTM）。

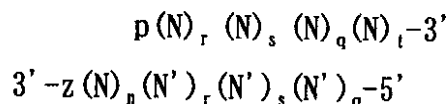
20

本発明のコードアダプターは、例えば、1本鎖または2本鎖タグが使用されるか否か、複数のタグが使用されるか否か、5'突出鎖または3'突出鎖が使用されるか否か、3'プロッキング基が使用されるか否かなどに依存して、いくつかの実施態様を有し得る。コードアダプターのいくつかの実施態様のための式は以下に示される。1本鎖タグを用いたコードアダプターに好ましい構造は以下の通りである：

30



または



ここで、Nはヌクレオチドであり、そしてN'はその相補物であり、pはリン酸基であり、zは3'水酸基または3'プロッキング基であり、nは、両端を含む2～6の整数であり、rは0以上の整数であり、sは、コードアダプターがヌクレアーゼ認識部位を有する時はいつでも4～6の整数であるか、またはヌクレアーゼ認識部位がない場合はいつでも0であり、qは0以上の整数であり、そしてtは、両端を含む8～20の整数である。より好ましくは、nは4または5であり、そしてtは両端を含む9～15である。コードアダプターがヌクレアーゼ認識部位を含む時はいつでも、「r」ヌクレオチド対の領域は、この部位を認識するヌクレアーゼが適用された時はいつでも、予め決定された数のヌクレオチドが標的ポリヌクレオチドから切断されるように選択される。特定の実施態様における「r」のサイズは、ヌクレアーゼの区域（この用語は、米国特許第5,599,675号およびWO 95/27080において定義される）ならびに標的ポリヌクレオチドから切断されようと努められる

40

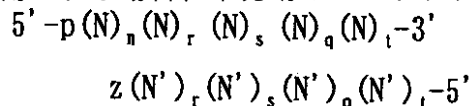
50

ヌクレオチド数に依存する。好ましくは、 r は 0 ~ 20 であり ; より好ましくは、 r は 0 ~ 12 である。「 q 」ヌクレオチド対の領域は、ヌクレアーゼ認識部位とコードプローブのタグ領域との間のスペーサーセグメントである。「 q 」ヌクレオチドの領域は、さらなるヌクレアーゼ認識部位、標識またはシグナル発生部分などを含み得る。1 本鎖オリゴヌクレオチドの「 t 」ヌクレオチドは、最小にクロスハイブリダイズするセットから選択される「 t マー」のオリゴヌクレオチドタグである。

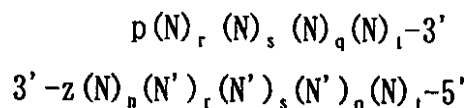
3' ブロッキング基「 z 」は種々の形態を有し得、そして連結を排除し、そしてこの方法の他の工程 (例えば、3' ブロック鎖の除去、連結など) を妨げない、ほとんど任意の化学的存在物を含み得る。例示的な 3' ブロッキング基は、水素 (すなわち、3' デオキシ)、リン酸、ホスホロチオエート、アセチルなどを含むが、これらに限定されない。好ましくは、3' ブロッキング基は、3' ブロック鎖の合成間にこの基を付加における簡便さ、およびこの鎖をリガーゼを用いて連結させ得るためのホスファターゼを用いたこの基の除去における簡便さのためリン酸である。3' リン酸を有するオリゴヌクレオチドは、編集者 Eckstein、Oligonucleotide and Analogues: A Practical Approach (IPL Press, Oxford, 1991) の第 12 章において記載されるプロトコルを用いて合成され得る。

さらなる 3' ブロッキング基は、例えば、以下の参考文献に開示される、1 塩基ずつの配列決定スキームにおける可逆的鎖終結ヌクレオチドのために開発された化学から利用可能である: Cheeseman、米国特許第 5,302,509 号; Tsien ら、国際出願 WO 91/06678 号; Canard ら、Gene, 148:1-6(1994); および Metzker ら、Nucleic Acids Research, 22:4259-4267(1994)。およそ、これらの化学は、プライミング鎖の 3' 末端の生産的遊離水酸基への特異的ブロッキング基 (通常、付随標識を有する) の化学的または酵素的除去を可能にする。好ましくは、 z が 3' ブロッキング基の場合、これはリン酸基であり、そしてアダプターの 2 本鎖部分は、その認識部位がその切断部位から離れているヌクレアーゼのヌクレアーゼ認識部位を含む。

三重鎖構造を形成するために 1 本鎖タグ相補物と特異的にハイブリダイズする、2 本鎖オリゴヌクレオチドタグが使用される場合、本発明のコードタグは以下の形態を有する:

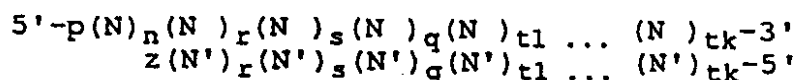


または

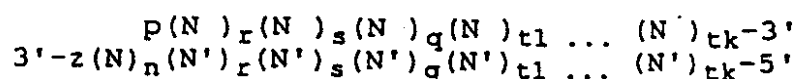


ここで、 N 、 N' 、 p 、 q 、 r 、 s 、 z 、および n は上記のように定義される。好ましくは、この実施態様において、 t は 12 ~ 24 の範囲の整数である。

明らかに、当業者には明らかである上記に示された基本的設計の要素を含むさらなる構造がある。例えば、本発明のコードアダプターは、以下のような複数のタグを有する実施態様を含む。



または



ここで、コードアダプターは k 個の 2 本鎖タグを含む。好ましくは、 $t_1=t_2=t_k$ であり、そして k は 1、2、または 3 のいずれかである。

タグ相補物の標識

本発明のタグ相補物は、放射性部分、蛍光部分、比色定量部分、化学発光部分などの直接

的または間接的付着を含む、オリゴヌクレオチドタグを解読するための種々の方法で標識され得る。DNAを標識し、そしてDNAアダプターを構築するための方法論の多くの包括的な概説は、本発明のアダプターの構築に適用可能なガイダンスを提供する。このような概説は、Matthewsら、Anal Biochem., 第169巻、1-25頁 (1988) ; Haugland, Handbook of Fluorescent Probes and Research Chemicals(Molecular Probes, Inc., Eugene, 1992) ; KellerおよびManak, DNA Probes, 第2版 (Stockton Press, New York, 1993) ; ならびに編集者Eckstein、Oligonucleotides and Analogues:A Prctical Approach(IRL Press, Oxford, 1991) ; Wetmur, Critical Reviews in Biochemistry and Molecular Biology, 26:227-259(1991)などを含む。本発明に適用可能な多くのより特定した方法論は、以下の参考文献のサンプルにおいて開示される：Fungら、米国特許第4,757,141号 ; Hobbs, Jrら、米国特許第5,151,507号 ; Cruickshank, 米国特許第5,091,519号 ; (レポーター基の付着のための官能化オリゴヌクレオチドの合成) ; Jablonskiら、Nucleic Acids Research, 14: 6115-6128(1986) (酵素-オリゴヌクレオチド結合体) ; Juら、Nature Medicine, 2:246-249(1996) ; およびUrdeaら、米国特許第5,124,246号 (分枝DNA) 。標識部分の付着部位は、このような標識が連結および/または切断工程を妨げなければ重要でない。

好ましくは、1つ以上の蛍光色素が、例えば、Menchenら、米国特許第5,188,934号 ; Bergotら、PCT出願PCT/US90/05565(WO 91/05060)により開示されるように、タグ相補物の標識として使用される。本明細書中で使用する用語「蛍光シグナル発生部分」は、1つ以上の分子の蛍光吸収および/または発光特性による情報を伝達するシグナリング手段を意味する。このような蛍光特性は、蛍光強度、蛍光寿命、発光スペクトル特徴、エネルギー転移などを含む。

アダプターの連結および自己連結の防止

本発明の好ましい実施態様によれば、切断アダプターは、コードアダプターの最後の連結のためのこのような末端を調製するために標的ポリヌクレオチドの末端に連結される。好ましくは、連結は、標準的なプロトコルにおいてリガーゼを用いて酵素的に行われる。多くのリガーゼは公知であり、そして本発明での使用に適する(例えば、Lehman, Science, 186:790-797(1974) ; 編集者Boyer, The Enzyme, 第15B巻 (Academic Press, New York, 1982) 中のEnglerら、DNA Ligases, 3-30頁など)。好ましいリガーゼは、T4 DNAリガーゼ、T7 DNAリガーゼ、E. coli DNAリガーゼ、Taqリガーゼ、Pfuリガーゼ、およびTthリガーゼを含む。それらの使用のためのプロトコルは周知である：例えば、Sambrookら、(上記で引用) ; Barany, PCR Methods and Applications, 1 : 5-16(1991) ; Marshら、Strategies, 5:73-76(1992) ; など。一般的に、リガーゼは、5'リン酸基が隣接鎖の3'水酸基への連結のために存在することを必要とする。これは、5'リン酸基を残すヌクレアーゼ(例えば、FokI)を選択することにより、標的ポリヌクレオチドの少なくとも1つの鎖に簡便に提供される。

固定されたポリヌクレオチドの4ヌクレオチドの突出鎖が互いに相補的である(114)、図2に示されるように、特別な問題は、自己連結し得るポリヌクレオチド末端またはアダプターのいずれかの処理において生じ得る。この問題は、分析されるポリヌクレオチド(112)が、固相支持体(110)に付着された同一のポリヌクレオチドの均一集団としてアダプターに提示される実施態様において特に重篤である。これらの状況において、固定ポリヌクレオチドの遊離末端は、互いに完全にマッチした2重鎖(116)を形成するためにねじれ得る。この末端の5'鎖がリン酸化されている場合、ポリヌクレオチドは、リガーゼの存在下で容易に連結される。類似の問題はまた、2本鎖アダプターについて存在する。すなわち、それらの5'鎖がリン酸化される時はいつでも、1つのアダプターの5'鎖は、それらの突出鎖のヌクレオチド配列が相補的である時はいつでも、もう1つのアダプター遊離3'水酸基に連結され得る。自己連結が起こる場合、アダプターの突出鎖も標的ポリヌクレオチドの突出鎖のどちらも、分析またはプロセッシングに利用可能でない。次に、これは、アダプターの標的ポリヌクレオチドへの正しい連結に応答して発生されるシグナルの喪失または消失を導く。ランダム配列でパリンδροーム4マーの生じる確率は、反復対のヌクレオチドの確率(6.25%)と同じであるので、デノボ配列決定のためのアダプター

に基づく方法は、自己連結のために数サイクル後に高い期待値の失敗を有する。これが生じる場合、ポリヌクレオチドのさらなる分析は不可能になる。

上記の問題は、好ましい実施態様のために図3Aに示されるように、以下の工程で本発明を実施することにより手段を講じ得る：(a) ポリヌクレオチド(122)の末端にコードアダプターを連結する工程(120)であって、ここでポリヌクレオチド末端は脱リン酸化5'水酸基を有し、そして連結されるコードアダプター(124)の末端は第1鎖(126)および第2鎖(128)を有し、そしてコードアダプターの第2鎖は3'ブロッキング基(130)を有する、工程；(b) 例えば、洗浄により(132)、またはインサイチュでこの基を酵素的もしくは化学的に除去することにより(例えば、ブロッキング基がリン酸である場合、ホスファターゼでの処理により)、連結後に第2鎖の3'ブロッキング基を除去する工程；(c) ポリヌクレオチドの5'水酸基をリン酸化する工程(134)；(d) 非ブロック3'部分を有する第2鎖(142)を連結して、コードアダプター(138)を再生する工程(136)；および(e) それに連結されたコードアダプターの正体により(例えば、蛍光標識(140)タグ相補物を介して)、ポリヌクレオチド末端の1つ以上のヌクレオチドを同定する工程(144)。コードアダプターおよび標的ポリヌクレオチドは、単一でまたは混合物としてのいずれかで連結のために組み合わせられ得る。例えば、定義配列を有する単一の種類のアダプターは、共通の(そしておそらく未知の)ヌクレオチド配列を有する単一の種類のポリヌクレオチドと組み合わせられ得るか；または定義配列を有する単一の種類のアダプターは、ポリヌクレオチドの混合物(例えば、同じ反応容器中の異なる固相支持体に付着された同一のポリヌクレオチドの複数の均一集団)と組み合わせられ得るか(例えば、Brennerら、国際出願PCT/US96/09513(WO 96/41011)に記載される)；または、コードアダプターの混合物、特にそれらの突出鎖に異なるヌクレオチド配列を有する混合物は、単一のポリヌクレオチドと混合され得る；またはコードアダプターの混合物は、ポリヌクレオチドの混合物と組み合わせられ得る。用語「アダプター」または「コードアダプター」が単数形で使用される場合、これは、用語「プローブ」の語法と類似した様式で、異なる配列の突出鎖を有するアダプターの混合物ならびに同じ配列の突出鎖を有する単一の種類のアダプターを含むことを意味する。

融解による除去に加えて、3'デオキシは、Kuijperら、Gene, 112:147-155(199.2)；Aslanidisら、Nucleic Acids Research, 18:6069-6074(1990)；および同様の参考文献により開示される、ポリメラーゼ「交換」反応により第2鎖から除去され得る。簡単には、T4 DNAポリメラーゼおよび同様の酵素の5'→3'エキソヌクレアーゼ活性を使用して、溶液中にそれらの三リン酸対応物でプライミング鎖のヌクレオチドを交換し得る(例えば、Kuijperら(上記))。従って、このような反応により、3'ジデオキシヌクレオチドは、反応混合物から2'-デオキシ-3'ヒドロキシヌクレオチドで交換され得、これはポリヌクレオチドキナーゼでの処理後に、第2鎖を標的ポリヌクレオチドに連結可能にする。

連結および切断のサイクルを使用する好ましい実施態様は、以下の工程を含む：(a) ポリヌクレオチド(222)の末端にコードアダプターを連結する工程(220)であって、ポリヌクレオチド末端は脱リン酸化5'水酸基を有し、そして連結される2本鎖アダプター(224)の末端は第1鎖(226)および第2鎖(228)を有し、2本鎖アダプターの第2鎖は3'ブロッキング基(230)を有し、そして2本鎖アダプターは、その認識部位がその切断部位から離れているヌクレアーゼのヌクレアーゼ認識部位(250)を有する、工程；(b) 例えば、第2鎖を洗浄することにより(232)、連結後に3'ブロッキング基を除去する工程；(c) ポリヌクレオチドの5'水酸基をリン酸化する工程(234)；(d) 2本鎖アダプター(238)およびヌクレアーゼ認識部位(250)を再生するために、非ブロック3'部分を有する第2鎖(242)を連結する工程(236)；(e) それに連結されたアダプターの同定により、ポリヌクレオチド末端の1つ以上のヌクレオチドを同定する工程(244)；(f) 認識部位を認識するヌクレアーゼでポリヌクレオチドを切断し(252)、その結果ポリヌクレオチドは1つ以上のヌクレオチドまで短くされ、認識部位は、切断(254)がポリヌクレオチド(222)から2つのヌクレオチドを除去するように示されたアダプター(224)において配置される、工程；(g) ポリヌクレオチドの5'末端を脱リン酸化する工

10

20

30

40

50

程 (256) ; および (h) 工程 (a) から (g) を反復する工程 (258) 。

代表的には、連結前に、分析されるポリヌクレオチドの末端は、通常3'または5' 突出鎖 (すなわち、「粘着」末端) を有する、予め決定された切断を生成する1つ以上の制限エンドヌクレアーゼでそれらを消化することにより調製される。このような消化は、通常、5' 鎖をリン酸化させておく。好ましくは、これらの5' リン酸化末端は、標準的なプロトコル (例えば、Sambrookら、Molecular Cloning, 第2版 (Cold Spring Harbor Laboratory, New York, 1989) を用いて、ホスファターゼ (例えば、子ウシ腸アルカリホスファターゼまたは同様の酵素) での処理により脱リン酸化される。5' リン酸の除去により、標的ポリヌクレオチドは、リガーゼの存在下では連結され得なくなる。好ましくは、脱リン酸化工程は、遊離5' 水酸基を残す。

10

好ましいヌクレアーゼ

本発明に従って使用される用語としての「ヌクレアーゼ」は、以下により十分に議論される連結された複合体に適用された場合に、連結された複合体を切断して、増大したアダプターおよび短くされた標的ポリヌクレオチドを生成する、任意の酵素、酵素の組み合わせ、もしくは他の化学試薬、または化学試薬および酵素の組合せを意味する。本発明のヌクレアーゼは、単一のタンパク質であったり、単にタンパク質の組合せから構成される必要はない。ヌクレアーゼまたはヌクレアーゼとして使用される試薬の組み合わせの重要な特性は、その (それらの) 切断部位がその (それらの) 認識部位から離れていることである。ヌクレアーゼの認識部位とその切断部位との間の距離は、本明細書中でその「区域」と呼ばれる。取り決めにより、「区域」は、各鎖の認識部位と加水分解ホスホジエステル結合との間のヌクレオチドの数を与える2つの整数により定義される。例えば、FokIの認識および切断特性は、代表的に、これが以下のように (配列番号2) 2本鎖DNAを認識および切断するので、「GGATG (9/13)」として示される:

20

5'-... NNGGATGNNNNNNNNNN NNNNNNNNNN ...

3'-... NNCCTACNNNNNNNNNNNNNNNN NNNNNN ...

ここで、ボールドのヌクレオチドはFokIの認識部位であり、そしてNは任意のヌクレオチドおよびそれらの相補物である。

ヌクレアーゼのみが、その認識部位との複合体を形成した後に標的ポリヌクレオチドを切断することは重要であり; そして好ましくは、ヌクレアーゼは、切断後に標的ポリヌクレオチド上に突出鎖を残す。

30

好ましくは、本発明で使用されるヌクレアーゼは、(i) その認識部位がその切断部位から離れており、そして(ii) 切断が標的ポリヌクレオチド上に突出鎖をもたらず、天然のタンパク質エンドヌクレアーゼである。最も好ましくは、II型制限エンドヌクレアーゼは、例えば、Szybalskiら、Gene, 100:13-26 (1991) ; Robertsら、Nucleic Acids Research, 21:3125-3137 (1993) ; およびLivakおよびBrenner、米国特許第5,093,245号に記載のように本発明におけるヌクレアーゼとして使用される。本発明での使用のための例示的なII型ヌクレアーゼは、AlwXI、BsmAI、BbvI、BsmFI、StsI、HgaI、BscAI、BbvII、Bcefl、Bce85I、BcclI、BcglI、Bsal、BsgI、BspMI、Bst71I、EarI、Eco57I、Esp3I、FauI、FokI、GsuI、HphI、MboII、MmeI、RleAI、SapI、SfaNI、TaqII、Tth111II、Bco5I、BpuAI、FnlI、BsrDI、およびそのアイソシゾマーを含む。好ましいヌクレアーゼは、BbvI、FokI、HgaI、EarI、およびSfaNIを含む。BbvIは、最も好ましいヌクレアーゼである。

40

好ましくは、ヌクレアーゼ切断工程前に、通常配列決定操作の開始で、標的ポリヌクレオチドは、使用されるヌクレアーゼの認識部位および/または切断部位をブロックするように処理される。これは、標的ポリヌクレオチドにおける内部位置でのヌクレアーゼ認識部位の偶然の出現による標的ポリヌクレオチドの所望でない切断を妨げる。ブロッキングは、メチル化および配列特異的アプタマー (aptamer) 、DNA結合タンパク質、または3重鎖を形成するオリゴヌクレオチドによる処置を含む、種々の方法において達成され得る。天然タンパク質エンドヌクレアーゼが使用される時はいつでも、認識部位は、使用されるヌクレアーゼのコグネイトメチラーゼで標的ポリヌクレオチドをメチル化することにより、都合良くブロックされ得る。すなわち、全てではないにしても、大部分のII型細菌性制限

50

エンドヌクレアーゼについて、その認識部位をメチル化する、いわゆる「コグネイト」メチラーゼが存在する。多くのこのようなメチラーゼは、Robertsら（上記）およびNelsonら、Nucleic Acids Research, 21:3139-3154（1993）に開示され、そして種々の供給源（特に、New England Biolabs（Beverly, MA）から市販されている。あるいは、PCR工程が配列決定のための標的ポリヌクレオチドの調製において使用される場合、5-メチルシトシン三リン酸は、天然のシトシンがアンプリコンにおいてメチル化シトシンで置換されるために増幅の間に使用され得る。この後者のアプローチは、固相支持体に結合した標的ポリヌクレオチドを別の酵素で処置する必要性を排除する追加の利点を有する。

明らかに、当業者は、本発明に従ってなおさらなる実施態様を設計するために上記で示されるが、上記で明白に示されなかった実施態様の特性を組み合わせ得る。

本発明の異なる実施態様を行うために種々のキットが提供される。一般的に、本発明のキットは、コードアダプター、切断アダプター、および標識タグ相補物を含む。キットは、ヌクレアーゼ試薬、連結試薬、および本発明の特定の実施態様を行うための使用説明書をさらに含む。天然のタンパク質エンドヌクレアーゼおよびリガーゼを使用する実施態様においては、リガーゼ緩衝液およびヌクレアーゼ緩衝液が含まれ得る。いくつかの場合において、これらの緩衝液は同一であり得る。このようなキットはまた、メチラーゼおよびその反応緩衝液を含み得る。好ましくは、キットはまた、1つ以上の固相支持体（例えば、標的ポリヌクレオチドを選別および固定するためにタグ相補物を保持する微粒子）を含む。

固相支持体上での選別のための、タグのポリヌクレオチドへの付着

本発明の重要な局面は、各微粒子または領域が実質的に付着された1種類のみのポリヌクレオチドを有するような、（例えば、cDNAライブラリーから）固相支持体上の微粒子または別の領域へのポリヌクレオチド集団の選別および付着である。この目的は、実質的に全ての異なるポリヌクレオチドが付着された異なるタグを有することを保証することにより達成される。この状態は、順に、分析のためにタグ-ポリヌクレオチド結合体の全集合のサンプルを採取することにより達成される。（2つの異なる位置で2回操作されるか、または分析される同じポリヌクレオチドを単に生じるように、同一のポリヌクレオチドが異なるタグを有することが許容である。）このようなサンプリングは、タグがポリヌクレオチドに付着された後、例えば、より多くの混合物から小容量を採取することにより、明白に行われ得るか（これは、ポリヌクレオチドおよびタグをプロセスするために使用される技術の二次的効果として本質的に行われ得る）、またはサンプリングは、明白におよびプロセッシング工程の本質的部分としての両方で行われ得る。

好ましくは、実質的に全ての異なるcDNAが異なるタグを有するcDNAライブラリーの構築において、その複雑度または異なるタグの数が細胞または組織サンプルから抽出されるmRNAの総数を大きく超える、タグレパートリーが使用される。好ましくは、タグレパートリーの複雑度は、ポリヌクレオチド集団の少なくとも10倍であり；より好ましくは、タグレパートリーの複雑度は、ポリヌクレオチド集団の少なくとも100倍である。以下に、例示的な9-ワードタグの全レパートリーを含むプライマー混合物を用いたcDNAライブラリーの構築についてのプロトコルが開示される。タグ含有プライマーのこのような混合物は、 8^9 または約 1.34×10^8 の複雑度を有する。Winslowら、Nucleic Acids Research, 19:3251-3253（1991）により示されるように、ライブラリー構築のためのmRNAは、10～100ほどの少ない哺乳動物細胞から抽出され得る。単一の哺乳動物細胞は約 5×10^5 コピーの約 3.4×10^4 の異なる種類のmRNA分子を含むので、標準的技術により、約100の細胞由来のmRNA、すなわち（理論的に）約 5×10^7 mRNA分子を単離し得る。この数とプライマー混合物の複雑度との比較は、任意のさらなる工程を行うことなく、そしてmRNAが完全な効率（1%以下の効率がより正確である）でcDNAに変換されることを保証することさえなく、cDNAライブラリー構築プロトコルが、異なるタグの総数のたった37%を含む集団を生じることを示す。すなわち、いずれの明白なサンプリング工程を少しも行うことなく、このプロトコルは、タグレパートリーの37%以下を含むサンプルを本質的に生じる。これらの条件下で二重に得られる確率は約5%であり、これは好ましい範囲内である。10の細胞からのmRNAを用い

てサンプリングされたタグレパートリーの画分はたった3.7%に低減され、これは全てのプロセッシング工程が100%の効率で起こることを保証さえする。実際に、cDNAライブラリーを構築するためのプロセッシング工程の効率は非常に低く、「経験則」では、良いライブラリーは、 10^6 の哺乳動物細胞から抽出されたmRNA由来の約 10^8 のcDNAクローンを含むはずである。

上記のプロトコルにおけるより多量のmRNAの使用、または一般的により多量のポリヌクレオチドの使用（ここで、このような分子の数はタグレパートリー、タグ-ポリヌクレオチド結合体の混合物の複雑度を越える）は、タグおよびmRNAまたはポリヌクレオチドの型の全ての可能な塩基対形成を潜在的に含む。このような場合、明白なサンプリングは、タグ-ポリヌクレオチド結合体の出発混合物の連続希釈後にサンプル容量を取り出すことにより実行され得る。必要とされる希釈量は、出発物質の量およびプロセッシング工程の効率に依存し、これは容易に評価される。

mRNAが 10^6 の細胞から抽出され（約 $0.5 \mu\text{g}$ のポリ(A)⁺RNAに相当する）、代表的なプロトコル（例えば、Sambrookら、Molecular Cloning, 第2版, 8.61頁）において要求されるように、プライマーが約10~100倍の濃度過剰で存在した場合（ 1 mg/mL の $10 \mu\text{L}$ 1.8kbmRNAは、約 1.68×10^{-11} モルに等しい、そして 1 mg/mL の $10 \mu\text{L}$ 18マープライマーは、約 1.68×10^{-9} モルに等しい）、cDNAライブラリー中のタグ-ポリヌクレオチド結合体の総数は、出発数のmRNAまたはタグ-ポリヌクレオチド結合体を含む約 5×10^{11} のベクターに単に等しいか、またはそれ未満であり（再度、これは、cDNA構築の各工程--第1鎖合成、第2鎖合成、およびベクターへの連結--が完全な効率で起こることを保証する）、これは非常に控えめな評価である。実際の数著しく少ない。

n のタグ-ポリヌクレオチド結合体のサンプルが反応混合物からランダムに取り出される（これは、サンプル容量を採取することにより実施され得る）場合、同じタグを有する結合体を取り出す確率は、ポアソン分布 $P(r) = e^{-r} \cdot r^r / r!$ （ここで、 r は同じタグを有する結合体の数である）および $= np$ （ここで、 p は選択される所定のタグの確率である）により記載される。 $n = 10^6$ および $p = 1 / (1.34 \times 10^8)$ であれば、 $= 0.00746$ および $P(2) = 2.76 \times 10^{-5}$ である。従って、10万の分子のサンプルは、好ましい範囲内で良好に予想された数の二重(double)を生じる。このようなサンプルは、以下のように容易に得られる： 5×10^{11} のmRNAが、挿入物としてタグ-cDNA結合体を有する 5×10^{11} のベクターに変換され、そして 5×10^{11} のベクターが $100 \mu\text{L}$ の容量を有する反応溶液にあることを保証する。4回の10倍連続希釈は、最初の溶液からの $10 \mu\text{L}$ を $90 \mu\text{L}$ の適切な緩衝液（例えばTE）を含む容器に移すことにより行われ得る。このプロセスは、 $1 \mu\text{L}$ あたり 5×10^5 のベクター分子を含む $100 \mu\text{L}$ 溶液を得るために3回のさらなる希釈で反復され得る。この溶液からの $2 \mu\text{L}$ アリコートは、挿入物としてタグ-cDNA結合体を含む 10^6 のベクターを生じる。次いで、このサンプルは、コンピテント宿主細胞の簡単な形質転換により増幅され、続いて培養される。

もちろん、上記のように、上記のプロセス中の工程は完全な効率で進行しない。特に、ベクターがタグ-ポリヌクレオチド結合体のサンプルを増幅するために使用される場合、宿主を形質転換する工程は非常に非効率的である。通常、たった1%のベクターが宿主により取り込まれ、そして複製される。従って、このような増幅方法のために、いっそうより少ない希釈が、 10^6 の結合体のサンプルを得るために必要とされる。

オリゴヌクレオチドタグのレパートリーは、多くの方法（直接酵素的連結、タグ配列を含むプライマーを用いた増幅（例えば、PCRを介した）などを含む）でポリヌクレオチド集団に結合され得る。最初の連結工程は、タグ-ポリヌクレオチド結合体の非常に多くの集団を生成し、その結果、単一のタグが一般的に多くの異なるポリヌクレオチドに付着される。しかし、上記のように、結合体の十分に小さなサンプルを採取することにより、「二重」（すなわち、2つの異なるポリヌクレオチド上の同じタグ）を得る確率は、無視され得る。一般的に、サンプルが大きいほど、二重を得る確率は大きくなる。従って、設計のトレードオフは、タグ-ポリヌクレオチド結合体の大きなサンプルを選択すること（これは、例えば、ショットガン配列決定操作または迅速に変化するmRNAプールの適切な提示に

おける標的ポリヌクレオチドの適切な適用範囲を確実にする)と小さなサンプルを選択すること(これは、最小数の二重が存在することを確実にする)との間に存在する。大部分の実施態様において、二重の存在は、複数の蛍光シグナルを生じる微粒子が簡単に無視され得るように、ノイズのさらなる原因または配列決定の場合におけるスキヤニングおよびシグナルプロセッシング工程での重要でない複雑化を単に付加する。

本明細書中で使用する、タグの、分子(特に、ポリヌクレオチド)への付着に関する用語「実質的に全て」は、二重を本質的に含まないタグ-分子結合体の集団を得るために使用されるサンプリング手順の統計学的性質を反映することを意味する。タグ-分子結合体の実際の割合に関する実質的に全ての意味は、タグが使用された方法に依存する。好ましくは、核酸配列決定について、実質的に全ては、少なくとも80%のポリヌクレオチドが、付着された特有のタグを有することを意味する。より好ましくは、少なくとも90%のポリヌクレオチドが、付着された特有のタグを有することを意味する。さらにより好ましくは、少なくとも95%のポリヌクレオチドが、付着された特有のタグを有することを意味する。そして、最も好ましくは、少なくとも99%のポリヌクレオチドが、付着された特有のタグを有することを意味する。

好ましくは、ポリヌクレオチド集団がメッセンジャーRNA(mRNA)からなる場合、オリゴヌクレオチドタグは、好ましくはタグ配列の相補物を含む1セットのプライマーでmRNAを逆転写することにより付着され得る。このようなプライマーの例示的なセットは、以下の配列を有し得る:

5'-mRNA-[A]_n-3'

[T]₁₉GG[W、W、W、C]₉ACCAGCTGATC-5'-ビオチン

ここで、「[W、W、W、C]₉」は、各4のヌクレオチドの9のサブユニットのオリゴヌクレオチドタグの配列を示し、そして「[W、W、W、C]」は、上記に列挙されたサブユニット配列を示す(すなわち、「W」はTまたはAを示す)。下線の配列は、それが使用される場合、ビオチンを介した固相支持体への付着からポリヌクレオチドを放出するために使用され得る、任意の制限エンドヌクレアーゼ部位を同定する。上記のプライマーについて、微粒子に付着される相補物は以下の形態を有する:

5'-[G、W、W、W]₉TGG-リンカー-微粒子

逆転写後、mRNAは、例えば、RNaseH消化により除去され、そしてcDNAの第2鎖は、例えば、以下の形態(配列番号3)のプライマーを用いて合成される。

5'-NRRGATCYNNN-3'

ここで、NはA、T、G、またはCのいずれか1つであり; Rはプリン含有ヌクレオチド、そしてYはピリミジン含有ヌクレオチドである。この特定のプライマーは得られた2本鎖DNAにBstY1制限部位を作成し、これは、SalI部位と共に、例えばBamHIおよびXhoI部位を有するベクターへのクローニングを容易にする。BstY1およびSalI消化後に、例示的な結合体は以下の形態を有する:

5'-RCGACCA[C、W、W、W]₉GG[T]₁₉-cDNA-NNNR

GGT[G、W、W、W]₉CC[A]₁₉-rDNA-NNNYCTAG-5'

次いで、ポリヌクレオチド-タグ結合体は、標準的な分子生物学技術を用いて操作され得る。例えば、上記の結合体(これは、実際には混合物である)は、市販のクローニングベクター(例えば、Stratagene Cloning System(La Jolla, CA)に挿入され; 宿主(例えば、市販の宿主細菌)にトランスフェクトされ得; 次いで、これは結合体の数を増大させるために培養される。次いで、クローニングベクターは、標準的な技術(例えば、Sambrookら、Molecular Cloning, 第2版(Cold Spring Harbor Laboratory, New York, 1989))を用いて単離され得る。あるいは、適切なアダプターおよびプライマーが使用され得、その結果、結合体集団がPCRにより増大され得る。

好ましくは、リガーゼに基づく配列決定法が使用される場合、BstY1およびSalI消化フラグメントは、以下の単一コピー制限部位を有するBamHI/XhoI消化ベクターにクローニング

10

20

30

40

50

される：

5'-GAGGATGCCTTTATGGATCCACTCGAGATCCCAATCCA-3'

FokI BamHI XhoI

これは、以下により十分に議論される配列決定プロセスの開始を可能にするFokI部位を付加する。

タグは、標準的なクローニング法により、存在するライブラリーのcDNAに結合され得る。cDNAは、それらの存在するベクターから切り出され、単離され、次いでタグレパートリーを含むベクターに連結される。好ましくは、タグ含有ベクターは、2つの制限酵素で切断することにより線状化され、その結果、切り出されたcDNAは予め決定された方向で連結され得る。線状化タグ含有ベクターの濃度は、cDNA挿入物の濃度を超えて実質的に過剰であり、その結果、連結はタグの本質的なサンプリングを提供する。

10

増幅後に1本鎖タグを曝露する一般的な方法は、標的ポリヌクレオチド含有結合体を、T4 DNAポリメラーゼまたは同様の酵素の5' 3' エキソヌクレアーゼ活性で消化することを含む。単一のデオキシヌクレオシド三リン酸の存在下で使用される場合、このようなポリメラーゼは、単一のデオキシヌクレオシド三リン酸の相補物がテンプレート鎖に到達されるまで、2本鎖フラグメントの非テンプレート鎖に存在する3' 凹化末端からヌクレオチドを切断する。このようなヌクレオチドが達成された場合、ポリメラーゼ伸長活性は、切除活性がヌクレオチドを除去するより高い速度でヌクレオチドを付加するので、5' 3' 消化は効果的に停止する。結果として、3つのヌクレオチドで構築された1本鎖タグは、固相支持体へのローディングのために容易に調製される。

20

この技術はまた、標的ポリヌクレオチドの内部FokI部位を優先的にメチル化するが、一方ポリヌクレオチドの末端の単一のFokI部位をメチル化しないままにするために使用され得る。最初に、末端FokI部位は、ポリメラーゼとデオキシシチジン三リン酸とを用いて1本鎖にされる。次いで、このフラグメントの2本鎖部分はメチル化され、この後、1本鎖末端は、4つ全てのヌクレオシド三リン酸の存在下でDNAポリメラーゼを用いてフィルインされ、それによりFokI部位を再生する。明らかに、この手順は、FokI以外のエンドヌクレアーゼに対して一般化され得る。

オリゴヌクレオチドタグが、例えば、上記のようにそれらを1本鎖にすることにより、特定のハイブリダイゼーションのために調製された後、ポリヌクレオチドは、タグとそれらの相補物との間で完全にマッチした2重鎖の形成に好ましい条件下で、タグの相補的配列を含む微粒子と混合される。これらの条件を作成するために文献中に広範なガイダンスがある。このようなガイダンスを提供する例示的な参考文献は、Wetmur, *Critical Reviews in Biochemistry and Molecular Biology*, 26:227-259 (1991); Sambrookら, *Molecular Cloning: A Laboratory Manual*, 第2版 (Cold Spring Harbor Laboratory, New York, 1989) などを含む。好ましくは、ハイブリダイゼーション条件は十分にストリンジェントであり、その結果、その結果、完全にマッチした配列のみが安定な2重鎖を形成する。このような条件下で、それらのタグにより特異的にハイブリダイズされるポリヌクレオチドは、微粒子に付着された相補的配列に連結され得る。最終的に、微粒子は洗浄され、非連結および/またはミスマッチタグを有するポリヌクレオチドを除去される。

30

40

合成支持体として従来より使用されるCPG微粒子が使用される場合、微粒子表面上のタグ相補物の密度は、いくつかの配列決定操作に必要なものより、代表的に大きい。すなわち、種々の酵素での付着されたポリヌクレオチドの連続処理を必要とする配列決定アプローチにおいて、密な間隔のポリヌクレオチドは、ポリヌクレオチドに対して相対的に巨大な酵素の接近を阻害する傾向にあり得る。このような場合、ポリヌクレオチドは、好ましくは微粒子と混合され、その結果、タグ相補物がポリヌクレオチドより著しく過剰に（例えば、10:1~100:1またはそれより多く）存在する。これは、微粒子表面上のポリヌクレオチドの密度が酵素接近を阻害するほど高くないことを確実にする。好ましくは、微粒子表面上の平均ポリヌクレオチド間の間隔は、およそ30~100nmである。標準的なCPG支持体およびBallotiniビーズ（固体ガラス支持体型）のための比の選択におけるガイダンスは、M

50

askosおよびSouthern, Nucleic Acids Research, 20:1679-1684 (1992)に見出される。
好ましくは、配列決定適用のために、20～50 μm 範囲の直径の標準的なCPGビーズは、約 10^5 のポリヌクレオチドでロードされ、そして5～10 μm 範囲の直径のBangs Laboratories (Carmel, IN) から入手可能なグリシダルメタクリレート (GMA) ビーズは、数万のポリヌクレオチド (例えば、 $4 \times 10^4 \sim 6 \times 10^4$) でロードされる。

好ましい実施態様において、選別のためのタグ相補物は組合せ的に微粒子上で合成され；従って、合成の終わりに微粒子の複合体混合物を得、これからローディングタグ化ポリヌクレオチドのためにサンプルが採取される。微粒子サンプルのサイズは、いくつかの因子 (タグ相補物レパートリーのサイズ、ロードされた微粒子を観察するために使用される装置の性質 (例えば、その能力)、同じタグ相補物を有する複数コピーの微粒子 (すなわち、「ビーズ二重」) に対する寛容などを含む) に依存する。以下の表は、微粒子サンプルサイズ、微粒子直径、および種々の直径の微粒子のパックされた配置のおおよその物理的寸法についてのガイダンスを提供する。

微粒子直径 5 μm 10 μm 20 μm 40 μm

1/ 10^5 平方オングストロー

ム1でロードされた最大 3×10^5 1.26×10^6 5×10^6
ポリヌクレオチド数

10^6 の微粒子 45×45cm 1×1cm 2×2cm 4×4cm

の単層のおよその面積

微粒子サンプルが所定のタグ相補物を含むか、または複数コピーに存在する確率は、以下の表に示されるようにポアソン分布により記載される。

表VI

サンプル中の微粒子数 (レパートリーサイズの 画分として)	サンプル中に存在する タグ相補体の レパートリー画分	付着された特有のタグ 相補体を有するサンプル 中の微粒子の画分	サンプル中の1つの他の微粒子 と同じタグ相補体を保持 するサンプル中の微粒子の画分 (「二重ビーズ」)
m	$1-e^{-m}$	$m(e^{-m})/2$	$m^2(e^{-m})/2$
1.000	0.63	0.37	0.18
.693	0.50	0.35	0.12
.405	0.33	0.27	0.05
.285	0.25	0.21	0.03
.223	0.20	0.18	0.02
.105	0.10	0.09	0.005
.010	0.01	0.01	

高い特異性の選別およびパンニング (panning)

選別の反応速度論は、オリゴヌクレオチドタグのそれらのタグ相補物へのハイブリダイゼーション速度に依存し、これは順にハイブリダイゼーション反応におけるタグの複雑度に依存する。従って、トレードオフは選別速度とタグ複雑度との間に存在し、その結果、選別速度の増大が、ハイブリダイゼーション反応に関与するタグの複雑度の低減を犠牲にして達成され得る。以下に説明されるように、このトレードオフの影響は、「パンニング」により改良され得る。

ハイブリダイゼーションの特異性は十分に小さなサンプルを採取することにより増大され得、その結果、サンプル中の高い割合のタグが特有であり、そしてサンプル中の実質的に全てのタグの最も近い隣接物 (neighbor) が少なくとも2つのワードで異なる。この後者

の条件は、使用されるレポトリサイズの約0.1%以下である多数のタグ-ポリヌクレオチド結合体を含むサンプルを採取することにより満たされ得る。例えば、タグが表IIから選択される8つのワードで構築される場合、 8^8 すなわち 1.67×10^7 のタグおよびタグ相補物のレポトリが生成される。上記のタグ-cDNA結合体のライブラリーにおいて、0.1パーセントサンプルは、約16,700の異なるタグが存在することを意味する。これが微粒子のレポトリ等価物またはこの実施例における 1.67×10^7 の微粒子サンプルに直接ロードされるのであれば、サンプリングされた微粒子のまばらなサブセットのみがロードされる。ロードされた微粒子の密度は、サンプリングされたタグ-cDNA結合体を使用してロードされなかった微粒子からロードされた微粒子を分離する「パンニング」工程に着手することにより（例えば、より効率的な配列決定のために）増大され得る。従って、上記の例において、「0.1パーセント」のサンプルが16,700のcDNAしか含まなくても、サンプリングおよびパンニング工程は、所望するほど多くのロードされた微粒子が蓄積されるまで反復され得る。あるいは、ロードされた微粒子は、従来のプロトコルを用いて蛍光活性化細胞選別（FACS）装置によりロードされていない微粒子から分離され得る（例えば、タグ-cDNA結合体は、蛍光標識された右プライマーを提供することにより以下に記載の技術において蛍光標識され得る。ローディングおよびFACS選別後、標識は、例えば、メチル化部位を認識するDpnIまたは同様の酵素により、コードアダプターに連結される前に切断され得る。

10

パンニング工程は、タグ-cDNA結合体（これらはそれぞれ、オリゴヌクレオチドタグに反対のまたは遠位の末端に捕獲部分を含む）のサンプルを提供することにより実行され得る。好ましくは、この捕獲部分は、タグ-cDNA結合体から放出され得る型であり、その結果、タグ-cDNA結合体は単一塩基配列決定法で配列決定され得る。このような部分は、ビオチン、ジゴキシゲニン、または同様のリガンド、3重鎖結合領域などを含み得る。好ましくは、このような捕獲部分は、ビオチン成分を含む。ビオチンは、多くの標準的技術によりタグ-cDNA結合体に付着され得る。PCRプライマー結合部位を含む適切なアダプターがタグ-cDNA結合体に付着される場合、ビオチンは、サンプリング後の増幅においてビオチン化プライマーを用いることにより付着され得る。あるいは、タグ-cDNA結合体がクローニングベクターの挿入物である場合、ビオチンは、適切な制限酵素での消化によりタグ-cDNA結合体を切り出し、続いて単離し、そしてビオチン化ウリジン三リン酸の存在下でDNAポリメラーゼを用いてタグに対して遠位の突出鎖をフィルインした後に付着され得る。

20

30

タグ-cDNA結合体が捕獲された後、多くの方法において、例えば、還元により切断されるか（例えば、Hermanら、Anal.Biochem., 156:48-55（1986））、または光化学的に切断されるか（例えば、Olejnikら、Nucleic Acids Research, 24:361-366（1996））、またはPCRプライマーに制限部位を導入することにより酵素的に切断される化学的連結により、ビオチン部分から放出され得る。後者の実施態様は、上記のタグ-ポリヌクレオチド結合体のライブラリーを考慮することにより例示され得る：

5'-RCGACCA[C, W, W]₉GG[T]₁₉-cDNA-NNNR

GGT[G, W, W]₉CC[A]₁₉-rDNA-NNNYCTAG-5'

以下のアダプターは、PCRによる増幅を可能にするために、これらのフラグメントの末端に連結され得る：

40

5'-XXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXYGAT

右アダプター

GATCZZACTAGTZZZZZZZZZZ-3'
ZZTGATCAZZZZZZZZZZ

10

左アダプター

ZZTGATCAZZZZZZZZZZ-5'- ビオチン

左プライマー

ここで「ACTAGT」はSpeI認識部位であり（これは、単一塩基配列決定に容易な互い違い切断を残す）、そしてX'およびZ'は選択されたヌクレオチドであり、その結果、それぞれのプライマーのアニールングおよび解離温度がほぼ同じである。アダプターの連結およびビオチン化プライマーを用いたPCRによる増幅後に、結合体のタグは、T4 DNAポリメラーゼのエキソヌクレアーゼ活性により1本鎖にされ、そして結合体は、付着されたタグ相補物を有する微粒子サンプル（例えば、レパートリー等価物）と組み合わせられる。（タグの誤った付着を最小にするために）ストリンジентな条件下でのアニールング後、結合体は、好ましくは、それらのタグ相補物に連結され、そしてロードされた微粒子は、アビジン化磁気ビーズまたは同様の捕獲技術により、ロードされていない微粒子から分離される。

20

例に帰ると、このプロセスは、異なるタグ（これは、SpeIでの切断により磁気ビーズから放出され得る）を有する約10,500（=16,700×0.63）のロードされた微粒子の蓄積をもたらす。新たな微粒子およびタグ-cDNA結合体のサンプルを用いてこのプロセスを40～50回反復することにより、 $4 \sim 5 \times 10^5$ のcDNAが、放出された微粒子をプールすることにより蓄積され得る。次いで、プールされた微粒子は、単一塩基配列決定技術により同時に配列決定され得る。

30

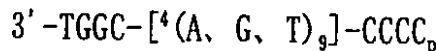
サンプリングおよびパンニング工程を反復する回数の決定、またはより一般的には、分析するcDNA量の決定は、目的による。目的が、例えば集団の5%以上を構成する、相対的に共通な配列の発生量における変化をモニターすることであるならば、相対的に小さなサンプル（すなわち、総集団サイズの小さな画分）は、相対的な発生量の統計学的に有意な評価を可能にし得る。他方では、例えば、集団の0.1%以下を構成する希な配列の発生量をモニターしようとするなら、大きなサンプルが必要とされる。一般的に、サンプルサイズとサンプルに基づく相対的発生量評価の信頼性との間に直接的な関連がある。信頼できる統計学的評価をするための適切なサンプルサイズの決定についての文献において広範なガイダンスがある（例えば、Kollerら、Nucleic Acids Research, 23:185-191（1994）；Good, Biometrika, 40:16-264（1953）；Bungeら、J.Am.Stat.Assoc., 88:364-373（1993）など）。好ましくは、 $3.0 \sim 3.5 \times 10^4$ の異なる配列の $10^5 \sim 10^8$ の独立したクローンを含む一連のcDNAライブラリーの分析に基づく遺伝子発現における変化をモニターするために、少なくとも 10^4 の配列サンプルが各ライブラリーの分析のために蓄積される。より好ましくは、少なくとも 10^5 の配列サンプルは、各ライブラリーの分析のために蓄積される；そして最も好ましくは、少なくとも 5×10^5 の配列サンプルが各ライブラリーの分析のために蓄積される。あるいは、サンプリングされた配列の数は、好ましくは、集団サイズの0.1%を超えずに、95%の信頼限界で、0.1%～5%の範囲内の頻度で存在する配列の相対的発生量を評価するために十分である。

40

50

タグライブラリーの構築

例示的なタグライブラリーは、以下の式により定義されるヌクレオチド A、G、および T の化学的に合成された 9-ワードタグを形成するために以下のように構築される。



ここで「 $[{}^4(A, G, T)_9]$ 」は、各タグが A、G、T の 9 つの 4-マーワードからなる、タグ混合物を示し；「p」は 5'リン酸を示す。この混合物は、以下の右および左プライマー結合領域（配列番号 4 および 5）に連結される：



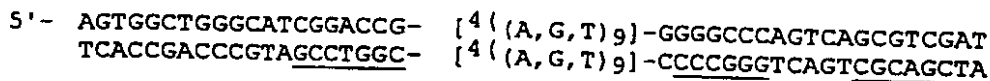
10

左

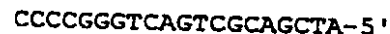
右

右および左プライマー結合領域は上記のタグ混合物に連結され、この後、連結された構造の 1 本鎖部分は DNA ポリメラーゼでフィルインされ、次いで以下に示される右および左プライマーと混合され、そして増幅されて、タグライブラリーを生じる。

左プライマー



20



右プライマー

左プライマー結合領域の下線部分は、RsrII 認識部位を示す。右プライマー結合領域の最も左の下線領域は、Bsp120I、ApaI、および EcoO 109I の認識部位および HgaI の切断部位を示す。右プライマー結合領域の最も右の下線領域は、HgaI の認識部位を示す。必要に応じて、右または左プライマーは、増幅および / または切断後の精製を容易にするために、（例えば、Clontech Laboratories, Palo Alto, CA から入手可能な従来の試薬を用いて）付着されたビオチンと共に合成され得る。

30

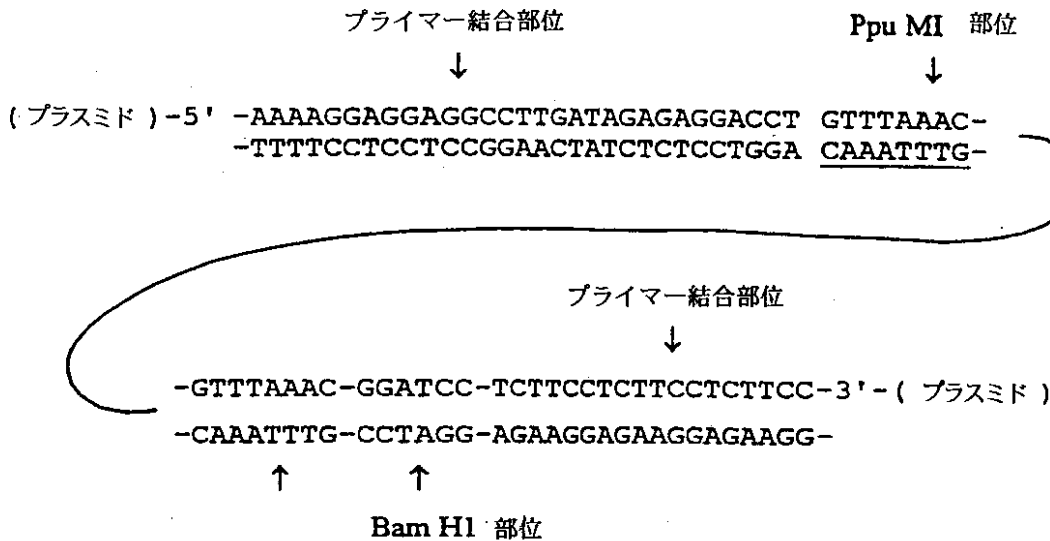
コードアダプターを用いた cDNA 「サイン (signature)」配列決定のためのタグ-ポリヌクレオチド結合体のプラスミドライブラリーの構築

cDNA は、mRNA のポリ A 領域の境界に固定された第 1 鎖合成のプライマーとして pGGCCCT₁₅（A または G または C）および第 2 鎖合成のプライマーとして N₈（A または T）GATC を用いた従来のプロトコルにより mRNA サンプルから生成される。すなわち、両方は、第 2 鎖プライマーが 2 つの形態で存在し、そして第 1 鎖プライマーが 3 つの形態で存在するような縮重プライマーである。第 2 鎖プライマーの GATC 配列は、MboI の認識部位に対応し；他の 4 つの塩基認識部位もまた、BamHI、SphI、EcoRI などのように使用され得る。第 2 鎖プライマーの認識部位に隣接した A および T の存在は、ストリッピングおよび交換反応が、「GGCC C」の 5 塩基 5' 突出を作成するために次の工程で使用され得る。第 1 鎖プライマーは mRNA サンプルにアニールされ、そして逆転写酵素で伸長され、この後 RNA 鎖は、逆転写酵素の RNaseH 活性により分解され、1 本鎖 cDNA が残る。第 2 鎖プライマーはアニールされ、そして従来のプロトコルを用いて DNA ポリメラーゼで伸長される。第 2 鎖の合成後、得られた cDNA は、製造者のプロトコルを用いて CpG メチラーゼ（New England Biolabs, Beverly, MA）でメチル化される。次いで、cDNA の 3' 鎖は、dATP および dTTP の存在下で T4 DNA ポリメラーゼを用いた上記のストリッピングおよび交換反応で切り戻され、この後、cDNA は HgaI で予め切断された上記のタグライブラリーに連結されて、以下の構築物を生じる：

40



別々に、以下のクローニングベクターは、例えば、市販のプラスミド（例えば、Bluescript ファージミド（Stratagene, La Jolla, CA））から開始して構築される（配列番号6）。



Pme I 部位

プラスミドは、PpuMIおよびPmeIで切断され（挿入物が指向されるようにRsrII適合可能末端および平滑末端を生じる）、次いでDAMメチラーゼでメチル化される。タグ含有構築物はRsrIIで切断され、次いで開裂したプラスミドに連結され、この後結合体はMboIおよびBamHIで切断され、プラスミドの連結およびクローニングを可能にする。次いで、プラスミドは、本発明に従う使用のために増幅され、そして単離される。

実施例 1

pGEM7Zから増幅された標的ポリヌクレオチドの配列決定：連結および切断のサイクルによるヌクレオチドの同定

本実施例において、プラスミドpGEM7Z（Promega, Madison, WI）のセグメントを増幅し、そして2本鎖DNAリンカー（このうちの一方の鎖をビーズ上で直接合成する（従って、ビーズに共有結合する））を介してガラスビーズに付着させる。標的ポリヌクレオチドの末端をコードアダプターへの連結のために調製した後、連結および切断の各サイクルにおいて、コードアダプターの混合物（全部で1024の異なるアダプター）を標的ポリヌクレオチドに適用し、その結果、その突出鎖が標的ポリヌクレオチドと完全にマッチした2重鎖を形成するアダプターのみが連結される。次いで、16の蛍光標識されたタグ相補物のそれぞれを、正しいタグ相補物のみのハイブリダイゼーションを可能にする条件下でポリヌクレオチド-アダプター結合体に適用する。洗浄後の蛍光シグナルの存在または非存在は、特定の位置の特定のヌクレオチドの存在または非存在を示す。本実施例の配列決定プロトコルは、Brenner、国際特許出願PCT/US95/12791およびPCT/US96/09513（WO 96/12041およびWO 96/41011）に記載されるように、1つ以上の固相支持体に選別される複数の標的ポリヌクレオチドに適用可能である。

47マーオリゴヌクレオチドを、標準的自動化DNA合成機プロトコルを用いてBallotiniビーズ（0.040~0.075mm、Jencons Scientific, Bridgeville, PA）上に直接合成する。その47マーに対する相補鎖を別々に合成し、そしてHPLCにより精製する。ハイブリダイズされると、得られた2重鎖はビーズから遠位の末端にBstXI制限部位を有する。相補鎖を、以下の混合物中で付着された47マーにハイブリダイズさせる：25 μLの200pmol/μL相補鎖；20mgの47マーを有するBallotiniビーズ；6 μLのNew England Biolabs#3制限緩衝液（10 × ストック溶液から）；および25 μLの蒸留水。混合物を93 °Cに加熱し、次いで55 °Cにゆ

っくりと冷却し、この後40ユニットのBstXI (10ユニット/ μ L)を、60 μ Lの反応容量にするように添加する。混合物を55℃で2時間インキュベートし、この後ビーズをTE (pH8.0)中で3回洗浄する。

ビーズに付着したpGEM7Zのセグメントを以下のように調製する：2つのPCRプライマーを、標準的なプロトコルを用いて調製した(配列番号7および配列番号8)：

プライマー 1: 5'-CTAAACCATTTGGTATGGGCCAGTGAATTGTAATA

プライマー 2: 5'-CGCGCAGCCCCGCATCGTTTATGCTACAGACTGTC-

AGTGCAGCTCTCCGATCCAAA

10

PCR反応混合物は以下からなる：1 μ Lの1ng/ μ L pGEM7Z；10 μ Lの10pmol/ μ Lプライマー1；10 μ Lの10pmol/ μ Lプライマー2；10 μ Lの2.5mMデオキシリボヌクレオチド三リン酸；10 μ Lの10 \times PCR緩衝液(Perkin-Elmer)；0.5 μ Lの5ユニット/ μ L Taq DNAポリメラーゼ；および100 μ Lの最終容量を得るための58 μ L蒸留水。反応混合物を93℃30秒；60℃15秒；および72℃60秒の25サイクルに供して、172塩基対産物を得、これをBbvI (100 μ LのPCR反応混合物、12 μ Lの10 \times #1 New England Biolabs緩衝液、8 μ Lの1ユニット/ μ L BbvIが37℃で6時間インキュベートされる)およびBstXI (BbvI反応混合物に添加する：5 μ Lの1M NaCl、67 μ Lの蒸留水、および8 μ Lの10ユニット/ μ L BstXI、得られた混合物を55℃で2時間インキュベートする)で連続的に消化する。

上記の反応混合物を製造者のプロトコルに従ってCentricon 30 (Amicon, Inc.) スピンカラムに通過させた後、BbvI/BstXI制限処理フラグメントを、以下の混合物：17 μ LのBbvI/BstXI制限処理フラグメント(10 μ g)、10 μ Lのビーズ(20mg)、6mlの10 \times 連結緩衝液(New England Biolabs、NEBとして以下で呼ばれる)、2000単位/ μ Lの5 μ L T4 DNAリガーゼ、および22 μ Lの蒸留水(この混合物を、25℃で4時間インキュベートする)中でBallotiniビーズに付着された2本鎖リンカーに連結し、この後ビーズをTE (pH8.0)で3回洗浄し、5'リン酸を有する配列決定のための以下の標的ポリヌクレオチド(配列番号9)を残す：

. . . AGCTACCCGATC
 [ビーズ]-- . . . TCGATGGGCTAGATTTp-5'

20

5'リン酸を、製造者のプロトコルを用いてNew England Biolabs (Beverly, MA) から入手可能なアルカリホスファターゼ(例えば、子ウシ腸由来)でビーズ混合物を処理することにより除去する。

以下の64コードアダプターの16セット(配列番号10～配列番号25)の上部鎖のそれぞれを、標準的方法を用いて自動化DNA合成機(モデル392 Applied Biosystems, Foster City)で別々に合成する。全てのアダプターについて同じである下部鎖を別々に合成し、次いでそれぞれの上部鎖にハイブリダイズさせる：

30

配列番号	コードアダプター	
10	5'-pANNNTACAGCTGCATCCCTtggcgctgagg pATGCACGCGTAGGG-5'	
11	5'-pNANNTACAGCTGCATCCCTtgggcctgtaag pATGCACGCGTAGGG-5'	
12	5'-pCNNNTACAGCTGCATCCCTtgacgggtctc pATGCACGCGTAGGG-5'	10
13	5'-pNCNNTACAGCTGCATCCCTgcccgcacagt pATGCACGCGTAGGG-5'	
14	5'-pGNNNTACAGCTGCATCCCTtcgcctcggac pATGCACGCGTAGGG-5'	
15	5'-pNGNNTACAGCTGCATCCCTgatccgctagc pATGCACGCGTAGGG-5'	
16	5'-pTNNNTACAGCTGCATCCCTtccgaacccgc pATGCACGCGTAGGG-5'	20
17	5'-pNTNNTACAGCTGCATCCCTgaggggtag pATGCACGCGTAGGG-5'	
18	5'-pNNANTACAGCTGCATCCCTtcccgtacac pATGCACGCGTAGGG-5'	
19	5'-pNNNATACAGCTGCATCCCTgactccccgag pATGCACGCGTAGGG-5'	
20	5'-pNNCNTACAGCTGCATCCCTgtgttgcgcg pATGCACGCGTAGGG-5'	
21	5'-pNNNCTACAGCTGCATCCCTctacagcagcg pATGCACGCGTAGGG-5'	30
22	5'-pNNGNTACAGCTGCATCCCTgtcgcgtcgtt pATGCACGCGTAGGG-5'	
23	5'-pNNNGTACAGCTGCATCCCTcggagcaacct pATGCACGCGTAGGG-5'	
24	5'-pNNTNTACAGCTGCATCCCTggtagccgtag pATGCACGCGTAGGG-5'	
25	5'-pNNNTTACAGCTGCATCCCTccccctgtcgga pATGCACGCGTAGGG-5'	40

ここで、Nおよびpは上記で定義された通りであり、そして小文字で示されたヌクレオチドは12マーのオリゴヌクレオチドタグである。各タグは、6ヌクレオチドで互いに異なる。等モル量の各アダプターを、1000pmol/μlの濃度の混合物を形成するために、NEB緩衝液2番(New England Biosciences, Beverly, MA)中で混合する。

各16タグ相補物を、アミノ誘導体オリゴヌクレオチドとして別々に合成し、そしてポリエチレングリコールリンカー(Clontech Laboratories, Palo Alto, CA)を通してタグ相補物の5'末端に付着される、フルオレセイン分子(例えば、Molecular Probes, Eugene, OR)から入手可能なFAM、フルオレセインのNHSエステル)でそれぞれ標識する。タグ相補物の配列は、単に、上記に列挙されたタグの12マー相補物である。

標的ポリヌクレオチドへのアダプターの連結を、5 μ l ビーズ (20mg)、3 μ l NEB 10 \times リガーゼ緩衝液、5 μ l アダプター混合物 (25nM)、2.5 μ l NEB T4 DNAリガーゼ (2000ユニット/ μ l) および14.5 μ l の蒸留水からなる混合物中で行う。混合物を、16 $^{\circ}$ で30分間インキュベートし、この後、ビーズをTE (pH8.0) で3回洗浄する。

TEの遠心分離および除去後、連結アダプターの3'リン酸を、製造者のプロトコルを用いて子ウシ腸アルカリホスファターゼ (CIP) (New England Biolabs, Beverly, MA) でポリヌクレオチド-ビーズ混合物を処理することにより除去する。3'リン酸の除去後、CIPを、製造者のプロトコルを用いて、例えば、PronaseTM (Boeringer Mannheim, Indianapolis, INから入手可能) または等価なプロテアーゼを用いたタンパク質分解消化により不活化し得る。次いで、ポリヌクレオチド-ビーズ混合物を洗浄し、そして標的ポリヌクレオチドへのアダプターの連結を完全にするために、T4ポリヌクレオチドキナーゼおよびT4 DNAリガーゼ (New England Biolabs, Beverly, MA) の混合物で処理して、標的ポリヌクレオチドとアダプターとの間のギャップに5'リン酸を付加する。次いで、ビーズ-ポリヌクレオチド混合物をTEで洗浄する。

別々に、各標識タグ相補物を、オリゴヌクレオチドタグとそれらのそれぞれの相補物との間でのみ完全にマッチした2重鎖の形成を可能にする条件下で、ポリヌクレオチド-ビーズ混合物に適用し、この後混合物をストリンジェントな条件下で洗浄し、そして蛍光シグナルの存在または非存在を測定する。タグ相補物を、25nMタグ相補物、50mM NaCl、3mM Mg、10mM Tris-HCl (pH8.5) からなる溶液中で20 $^{\circ}$ で適用し、10分間インキュベートし、次いで同じ溶液 (タグ相補物なしで) で55 $^{\circ}$ で10分間洗浄する。

4つのヌクレオチドを上記のように同定した後、コードアダプターを、製造者のプロトコルを用いてBbvIでポリヌクレオチドから切断する。最初の連結および同定後、連結、同定、および切断のサイクルを3回反復して、標的ポリヌクレオチドの16末端ヌクレオチドの配列を得る。図4は、5~16位 (ビーズから最も遠位~ビーズから最も近位) のヌクレオチドを同定するために適用された4つのタグ相補物のそれぞれからの相対的蛍光を示す。

実施例2

コードアダプターを用いたサイン配列決定のためのcDNAライブラリーの構築および選別

本実施例においてcDNAライブラリーを構築し、ここで8つの4-ヌクレオチド「ワード」からなるオリゴヌクレオチドタグを各cDNAに連結する。上記のように、このサイズのオリゴヌクレオチドタグのレパートリーは十分に大きく (約10⁸)、その結果、cDNAが約10⁶のmRNAの集団から合成される場合、各cDNAは選別のための特有のタグを有する確率が高い。mRNA抽出後、第1鎖合成を、(特定のcDNA制限部位をブロックするために) 5-Me-dCTPおよびオリゴヌクレオチドタグ含有ビオチン化プライマー混合物の存在下で行う。従来の第2鎖合成後、タグ-cDNA結合体をDpnII (5-Me-デオキシシトシンにより影響されない) で切断し、ビオチン化部分をストレプトアビジンコート磁気ビーズを用いて反応混合物から分離し、そしてタグ-cDNA結合体を、それらをビオチン化プライマーにより運ばれるBsmBI部位を介して磁気ビーズから切断することにより回収する。次いで、タグ-cDNA結合体を含むBsmBI-DpnIIフラグメントをプラスミドに挿入し、そして増幅する。プラスミドの単離後、タグ-cDNA結合体を、予め定義されたエンドヌクレアーゼ部位を含むビオチン化および蛍光標識化プライマーを用いて、5-Me-dCTPの存在下でPCRによりプラスミドから増幅する。ストレプトアビジンコート磁気ビーズでのアフィニティ精製後、タグ-cDNA結合体をビーズから切断し、タグを1本鎖にするためにdGTPの存在下でT4 DNAポリメラーゼで処理し、次いで付着されたタグ相補物を有するGMAビーズのレパートリーと混合する。ストリンジェントなハイブリダイゼーションおよび連結後、GMAビーズを、cDNAでロードされたGMAビーズの富化集団を生成するためにFACSを介して選別する。ロードされたGMAビーズの富化集団を、1塩基ずつの配列がコードアダプターを用いて起こるフローチャンバー中で平面配列に固定する。

約5 μ gのポリ (A⁺) mRNAを、従来のプロトコルを用いてDBY746酵母細胞から抽出する。第1および第2鎖cDNA合成を、製造者のプロトコルに従ってStratagene (La Jolla, CA) cDNA合成キットを用いて、100~150pmolの以下のプライマー (配列番号26) :

5'-ビオチン-ACTAATCGTCTCACTATTTAATTAA[W、W、W、G]₈CC(T)₁₈V-3'

とポリ(A⁺) mRNAとを組み合わせることにより行う。これは、第1鎖デオキシシトシンが5炭素位置でメチル化されているcDNAをもたらす。上記の式において、「V」はG、C、またはAであり、「W、W、W、G」は上記の表IIから選択される4-ヌクレオチドワードであり、一重下線部分はBsmBI認識部位であり、二重下線部分はPacI認識部位である。従来のプロトコルを用いたサイズ画分(GIBCO-BRL cDNAサイズ画分キット)後に、cDNAを製造者のプロトコルを用いてDpnII(New England Bioscience, Beverly, MA)で消化し、ストレプトアビジンコート磁気ビーズ(M-280ビーズ、DynaI A.S., Oslo, Norway)でアフィニティ精製する。ビーズにより捕獲されたDNAをBsmBIで消化して、標準的なプロトコルを用いて改変pBCSK⁺ベクター(Stratagene, La Jolla, CA)へのクローニングのためにタグ-cDNA結合体を放出する。以下のフラグメント(配列番号27)をKpnI/EcoRV消化ベクターに挿入することによってBbsI部位を付加することにより、pBCSK⁺ベクターを改変する。

CGAAGACCC

3'-CATGGCTTCTGGGGATA-5'

BsmBI/DpnII消化タグ-cDNA結合体を、予めBbsIおよびBamHIで消化したpBCSK⁺に挿入する。連結後に、増幅のために、製造者の推奨する宿主にベクターをトランスフェクトする。標準的プラスミドミニプレップから上記のpBCSK⁺ベクターを単離した後に、タグ-cDNA結合体を、タグ-cDNA挿入物に隣接するベクター配列に相補的な20-マープライマーを用いて5-Me-dCTPの存在下でPCRにより増幅する。「上流」プライマー(すなわちタグに隣接)をビオチン化し、そして「下流」プライマー(すなわちcDNAに隣接)をフルオレセインで標識する。増幅後、PCR産物をアフィニティ精製し、次いでPacIで切断して、蛍光標識されたタグ-cDNA結合体を放出させる。結合体のタグを、それらをdTTPの存在下でT4 DNAポリメラーゼを用いて処理することにより1本鎖にする。反応物をクエンチングした後、タグ-cDNA結合体をフェノール-クロロホルム抽出により精製し、そしてタグ相補物(各タグ相補物は5'リン酸を有する)を保持する5.5mm GMAビーズと組み合わせる。それらの相補物と完全にマッチした2重鎖を形成するタグのみが連結されるように、ハイブリダイゼーションを耐熱性リガーゼの存在下でストリンジェントな条件下で行う。GMAビーズを洗浄し、そしてロードされたビーズを蛍光標識されたcDNAを用いてFACS選別により濃縮し、ロードされたGMAビーズを同定する。GMAビーズに付着されたタグ-cDNA結合体をDpnIIで消化して蛍光標識を除去し、そして配列決定のためのcDNAを調製するためにアルカリホスファターゼで処理する。

以下の切断アダプター(配列番号28)を、DpnII消化およびホスファターゼ処理cDNAに連結する：

5'-pGATCAGCTGCTGCAAATTT

pTCGACGACGTTTAAA

この後、3'リン酸をアルカリホスファターゼにより除去し、cDNAの5'鎖をT4 DNAキナーゼで処理し、そして切断アダプターとcDNAとの間のニックを連結する。BbvIによる切断後、実施例1のコードアダプターを上記のcDNAの末端に連結する。

図5に模式的に示されたフローチャンバー(500)を、標準的マイクロ機械加工技術(例えば、Ekstromら、国際特許出願PCT/SE91/00327(WO 91/16966)；Brown, 米国特許第4,911,782号；Harrisonら、Anal.Chem.64:1926-1932(1992)など)を用いて、ガラス板(506)中の液体入口(502)および出口(504)を有する空洞をエッチングすることにより調製する。フローチャンバー(500)の寸法は、ロードされた微粒子(508)(例えば、GMAビーズ)が、100万~200万ビーズの密にパックされた平面単層で空洞内(510)で配置され得るようなものである。空洞(510)は、エッチングされたガラス板(506)へのガラスカバースリップ(512)のアノード結合により入口および出口を有して密封チャンバー内に作られる(例えば、Pomerantz, 米国特許第3,397,279号)。試薬は、シリンジポンプ(514

～520) から、一般的に自動化DNAおよびペプチド合成機で使用されるようにマイクロプロセッサにより制御される弁ブロック(522)を通してフローチャンバーに供給される(例えば、Bridghamら、米国特許第4,668,479号; Hoodら、米国特許第4,252,769号; Barstowら、米国特許第5,203,368号; Hunkapiller, 米国特許第4,703,913号など)。

連結、同定、および切断の3サイクルを、約100,000cDNAのそれぞれの末端に12ヌクレオチドの配列を生じさせるために、フローチャンバー(500)において行う。cDNAのヌクレオチドを、実施例1に記載のようにタグ相補物をコードアダプターにハイブリダイズさせることにより同定する。特異的にハイブリダイズしたタグ相補物を、それらの蛍光標識を光源(526)(レーザー、水銀アークランプなどであり得る)からの照明ビーム(524)で励起することにより検出する。照明ビーム(524)はフィルター(528)を通して通過し、そしてフローチャンバー(500)内のコードアダプターに特異的にハイブリダイズしたタグ相補物上の蛍光標識を励起する。生じた蛍光(530)を共焦点顕微鏡(532)により集め、フィルター(534)を通過させ、そしてCCDカメラ(536)に指向させる。これは、ワークステーション(538)によるプロセッシングおよび分析のためにビーズ配列の電気画像を作成する。好ましくは、各連結および切断工程後、cDNAをPronaseTM、または同様の酵素で処理する。コードアダプターおよび約0.75ユニット/ μ LのT4 DNAリガーゼ(Promega, Madison, WI)を、16 で約20～30分間、約1～2 μ L/分の流速でフローチャンバーに通過させる。この後、3'リン酸をアダプターから除去し、そして0.02ユニット/ μ Lのアルカリホスファターゼ(New England Bioscience, Beverly, MA)および7ユニット/ μ LのT4 DNAキナーゼ(New England Bioscience, Beverly, MA)の混合物を15～20分間、1～2 μ L/分の流速で37 でフローチャンバーを通過させることにより第2鎖連結のためのcDNAを調製する。連結を、20～30分間フローチャンバーに通したT4 DNAリガーゼ(0.75ユニット/mL, Promega)により達成する。25nM濃度のタグ相補物を、20 で10分間、1～2 μ L/分の流速でフローチャンバーに通過させ、この後、タグ相補物によりもたらされる蛍光標識を照明し、そして蛍光を集める。タグ相補物を、ハイブリダイゼーション緩衝液を55 で10分間、1～2 μ L/分の流速で37 でフローチャンバーに通過させることにより、コードアダプターから融解する。コードアダプターを、1ユニット/ μ LのBbvI(New England Bioscience, Beverly, MA)を37 で20分間、1～2 μ L/分の流速で通過させることによりcDNAから切断する。

補遺 I

最小にクロスハイブリダイズするセット(1本鎖タグ/1本鎖タグ相補物)を作成するための例示的なコンピュータープログラム


```

Program tagN
c
c
c
c      Program tagN generates minimally cross-hybridizing
c      sets of subunits given i) N--subunit length, and ii)
c      an initial subunit sequence. tagN assumes that only
c      3 of the four natural nucleotides are used in the tags.
c
character*1 subl(20)
integer*2 mset(10000,20), nbase(20)
c
c
c      write(*,*)'ENTER SUBUNIT LENGTH'
c      read(*,100)nsub
c      format(i2)
c
c
c      write(*,*)'ENTER SUBUNIT SEQUENCE'
c      read(*,110)(subl(k),k=1,nsub)
c      format(20a1)
c
c      ndiff=10
c
c
c      Let a=1 c=2 g=3 & t=4
c
c
c      do 800 kk=1,nsub
c      if(subl(kk).eq.'a') then
c          mset(1,kk)=1
c      endif
c          if(subl(kk).eq.'c') then
c              mset(1,kk)=2
c          endif
c              if(subl(kk).eq.'g') then
c                  mset(1,kk)=3
c              endif
c                  if(subl(kk).eq.'t') then
c                      mset(1,kk)=4
c                  endif
c
c      continue
c
c
c      Generate set of subunits differing from
c      subl by at least ndiff nucleotides.
c
c      jj=1

```

```

do 1000 k1=1,3
  do 1000 k2=1,3
    do 1000 k3=1,3
      do 1000 k4=1,3
        do 1000 k5=1,3
          do 1000 k6=1,3
            do 1000 k7=1,3
              do 1000 k8=1,3
                do 1000 k9=1,3
                  do 1000 k10=1,3
do 1000 k11=1,3
  do 1000 k12=1,3
    do 1000 k13=1,3
do 1000 k14=1,3
  do 1000 k15=1,3
    do 1000 k16=1,3
      do 1000 k17=1,3
        do 1000 k18=1,3
          do 1000 k19=1,3
            do 1000 k20=1,3
c
c
nbase(1)=k1
nbase(2)=k2
nbase(3)=k3
nbase(4)=k4
nbase(5)=k5
nbase(6)=k6
nbase(7)=k7
nbase(8)=k8
nbase(9)=k9
nbase(10)=k10
nbase(11)=k11
nbase(12)=k12
nbase(13)=k13
nbase(14)=k14
nbase(15)=k15
nbase(16)=k16
nbase(17)=k17
nbase(18)=k18
nbase(19)=k19
nbase(20)=k20
c
c
do 1250 nn=1,jj
  n=0
  do 1200 j=1,nsup
    if(mset(nn,j).eq.1 .and. nbase(j).ne.1 .or.
1      mset(nn,j).eq.2 .and. nbase(j).ne.2 .or.
2      mset(nn,j).eq.3 .and. nbase(j).ne.3 .or.
3      mset(nn,j).eq.4 .and. nbase(j).ne.4) then
      n=n+1
    endif
    continue
1200
c
c
if(n.lt.ndiff) then
  goto 1000
endif
1250
c
c
  
```

10

20

30

40

```

        jj=jj+1
        write(*,130) (nbase(i),i=1,nsub),jj
        do 1100 i=1,nsub
            mset(jj,i)=nbase(i)
            continue
1100
c
c
1000      continue
c
c
        write(*,*)
130      format(10x,20(1x,i1),5x,i5)
        write(*,*)
        write(*,120) jj
120      format(1x,'Number of words=',i5)
c
c
        end
c
c
c      *****
c      *****

```

10

補遺II

最小にクロスハイブリダイズするセット（2本鎖タグ/1本鎖タグ相補物）を作製するための例示的なコンピュータプログラム


```

do 1000 k4=1,2
  do 1000 k5=1,2
    do 1000 k6=1,2
      do 1000 k7=1,2
        do 1000 k8=1,2
          do 1000 k9=1,2
            do 1000 k10=1,2
              do 1000 k11=1,2
                do 1000 k12=1,2
                  do 1000 k13=1,2
                    do 1000 k14=1,2
                      do 1000 k15=1,2
                        do 1000 k16=1,2
                          do 1000 k17=1,2
                            do 1000 k18=1,2
                              do 1000 k19=1,2
                                do 1000 k20=1,2
                                  c
                                  c
                                  nbase(1)=k1
                                  nbase(2)=k2
                                  nbase(3)=k3
                                  nbase(4)=k4
                                  nbase(5)=k5
                                  nbase(6)=k6
                                  nbase(7)=k7
                                  nbase(8)=k8
                                  nbase(9)=k9
                                  nbase(10)=k10
                                  nbase(11)=k11
                                  nbase(12)=k12
                                  nbase(13)=k13
                                  nbase(14)=k14
                                  nbase(15)=k15
                                  nbase(16)=k16
                                  nbase(17)=k17
                                  nbase(18)=k18
                                  nbase(19)=k19
                                  nbase(20)=k20
                                  c
                                  c
                                  do 1250 nn=1,jj
                                    n=0
                                    do 1200 j=1,nsup
                                      if(mset(nn,j).eq.1 .and. nbase(j).ne.1 .or.
1                                      mset(nn,j).eq.2 .and. nbase(j).ne.2) then
                                        n=n+1
                                        endif
1200                                     continue
                                      c
                                      c
                                      if(n.lt.ndiff) then
                                        goto 1000
                                      endif
1250                                     continue
                                      c

```

10

20

30

40

```

c
      jj=jj+1
      write(*,130) (nbase(i),i=1,nsub),jj
      do 1100 i=1,nsub
        mset(jj,i)=nbase(i)
1100      continue
c
c
1000  continue
c
c
      write(*,*)
130   format(5x,20(1x,i1),5x,i5)
      write(*,*)
      write(*,120) jj
120   format(1x,'Number of words=',i5)
c
c
      end

```

配列表

- (1) 一般的情報 :
- (i) 出願人 :
- (A) 名称 : リンクス セラピューティクス , インコーポレイテッド 10
- (ii) 発明の名称 : コードアダプターの連結による大規模平行特性 (signature) 配列決定
- (iii) 配列数 : 28
- (iv) 通信住所 :
- (A) 名称 : デーリンガー アンド アソシエイツ
- (B) 番地 : ケンブリッジ アベニュー 350 , スイート 250
- (C) 市 : パロアルト
- (D) 州 : カリフォルニア
- (E) 国 : アメリカ合衆国
- (F) 郵便番号 : 94306 30
- (v) コンピューター読み出し形態 :
- (A) 媒体型 : フロッピー ディスク
- (B) コンピューター : IBM PC互換用
- (C) OS : PC-DOS/MS-DOS
- (D) ソフトウェア : パテントイン リリース #1.0 , バージョン #1.25
- (vi) 現在の出願データ :
- (A) 出願番号 :
- (B) 出願日 :
- (vii) 先願データ :
- (A) 出願番号 : US 08/689,587 40
- (B) 出願日 : 1996年 8 月12日
- (vii) 先願データ :
- (A) 出願番号 : US 08/659,453
- (B) 出願日 : 1996年 6 月 6 日
- (viii) 代理人/事務所情報 :
- (A) 名称 : パワーズ , ヴィンセント エム .
- (B) 登録番号 : 36,246
- (C) 照会/記録番号 : 5525-0029.41/808-1wo
- (ix) 通信情報
- (A) 電話 : (415) 324-0880 50

(B) テレファックス : (415) 324-0960		
(2) 配列番号 1 の情報 :		
(i) 配列の特徴 :		
(A) 長さ : 28ヌクレオチド		
(B) 型 : 核酸		
(C) 鎖の数 : 二本鎖		
(D) トポロジ : 直鎖状		
(xi) 配列 : 配列番号 1 :		
TGGATTCTAG AGAGAGAGAG AGAGAGAG	28	
(2) 配列番号 2 の情報 :		10
(i) 配列の特徴 :		
(A) 長さ : 16ヌクレオチド		
(B) 型 : 核酸		
(C) 鎖の数 : 二本鎖		
(D) トポロジ : 直鎖状		
(xi) 配列 : 配列番号 2 :		
NNGGATGNNN NNNNNN	16	
(2) 配列番号 3 の情報 :		
(i) 配列の特徴 :		
(A) 長さ : 11ヌクレオチド		20
(B) 型 : 核酸		
(C) 鎖の数 : 一本鎖		
(D) トポロジ : 直鎖状		
(xi) 配列 : 配列番号 3 :		
NRRGATCYNN N	11	
(2) 配列番号 4 の情報 :		
(i) 配列の特徴 :		
(A) 長さ : 20ヌクレオチド		
(B) 型 : 核酸		
(C) 鎖の数 : 二本鎖		30
(D) トポロジ : 直鎖状		
(xi) 配列 : 配列番号 4 :		
AGTGGCTGGG CATCGGACCG	20	
(2) 配列番号 5 の情報 :		
(i) 配列の特徴 :		
(A) 長さ : 20ヌクレオチド		
(B) 型 : 核酸		
(C) 鎖の数 : 二本鎖		
(D) トポロジ : 直鎖状		
(xi) 配列 : 配列番号 5 :		40
GGGGCCCAGT CAGCGTCGAT	20	
(2) 配列番号 6 の情報 :		
(i) 配列の特徴 :		
(A) 長さ : 70ヌクレオチド		
(B) 型 : 核酸		
(C) 鎖の数 : 二本鎖		
(D) トポロジ : 直鎖状		
(xi) 配列 : 配列番号 6 :		
AAAAGGAGGA GGCCTTGATA GAGAGGACCT GTTTAAACGT TTAAACGGAT	50	
CCTCTTCCTC TTCCTCTTCC	70	50

(2) 配列番号7の情報:

(i) 配列の特徴:

(A) 長さ: 34ヌクレオチド

(B) 型: 核酸

(C) 鎖の数: 一本鎖

(D) トポロジー: 直鎖状

(xi) 配列: 配列番号7:

CTAAACCATT GGTATGGGCC AGTGAATTGT AATA 34

(2) 配列番号8の情報:

(i) 配列の特徴:

10

(A) 長さ: 55ヌクレオチド

(B) 型: 核酸

(C) 鎖の数: 一本鎖

(D) トポロジー: 直鎖状

(xi) 配列: 配列番号8:

CGCGCAGCCC GCATCGTTTA TGCTACAGAC TGTCAGTGCA 40

GCTCTCCGAT CAAA 55

(2) 配列番号9の情報:

(i) 配列の特徴:

(A) 長さ: 16ヌクレオチド

20

(B) 型: 核酸

(C) 鎖の数: 二本鎖

(D) トポロジー: 直鎖状

(xi) 配列: 配列番号9:

TCGATGGGCT AGATTT 16

(2) 配列番号10の情報:

(i) 配列の特徴:

(A) 長さ: 30ヌクレオチド

(B) 型: 核酸

(C) 鎖の数: 二本鎖

30

(D) トポロジー: 直鎖状

(xi) 配列: 配列番号10:

ANNNTACAGC TGCATCCCTT GCGCTGAGG 30

(2) 配列番号11の情報:

(i) 配列の特徴:

(A) 長さ: 30ヌクレオチド

(B) 型: 核酸

(C) 鎖の数: 二本鎖

(D) トポロジー: 直鎖状

(xi) 配列: 配列番号11:

40

NANNTACAGC TGCATCCCTG GGCCTGTAAG 30

(2) 配列番号12の情報:

(i) 配列の特徴:

(A) 長さ: 30ヌクレオチド

(B) 型: 核酸

(C) 鎖の数: 二本鎖

(D) トポロジー: 直鎖状

(xi) 配列: 配列番号12:

CNNNTACAGC TGCATCCCTT GACGGGTCTC 30

(2) 配列番号13の情報:

50

(i) 配列の特徴：

(A) 長さ：30ヌクレオチド

(B) 型：核酸

(C) 鎖の数：二本鎖

(D) トポロジ：直鎖状

(xi) 配列：配列番号13：

NCNNTACAGC TGCATCCCTG CCCGCACAGT

30

(2) 配列番号14の情報：

(i) 配列の特徴：

(A) 長さ：30ヌクレオチド

10

(B) 型：核酸

(C) 鎖の数：二本鎖

(D) トポロジ：直鎖状

(xi) 配列：配列番号14：

GNNNTACAGC TGCATCCCTT CGCCTCGGAC

30

(2) 配列番号15の情報：

(i) 配列の特徴：

(A) 長さ：30ヌクレオチド

(B) 型：核酸

(C) 鎖の数：二本鎖

20

(D) トポロジ：直鎖状

(xi) 配列：配列番号15：

NGNNTACAGC TGCATCCCTG ATCCGCTAGC

30

(2) 配列番号16の情報：

(i) 配列の特徴：

(A) 長さ：30ヌクレオチド

(B) 型：核酸

(C) 鎖の数：二本鎖

(D) トポロジ：直鎖状

(xi) 配列：配列番号16：

TNNNTACAGC TGCATCCCTT CCGAACCCGC

30

(2) 配列番号17の情報：

(i) 配列の特徴：

(A) 長さ：30ヌクレオチド

(B) 型：核酸

(C) 鎖の数：二本鎖

(D) トポロジ：直鎖状

(xi) 配列：配列番号17：

NTNNTACAGC TGCATCCCTG AGGGGGATAG

30

(2) 配列番号18の情報：

(i) 配列の特徴：

(A) 長さ：30ヌクレオチド

(B) 型：核酸

(C) 鎖の数：二本鎖

(D) トポロジ：直鎖状

(xi) 配列：配列番号18：

NNANTACAGC TGCATCCCTT CCCGCTACAC

30

(2) 配列番号19の情報：

(i) 配列の特徴：

(A) 長さ：30ヌクレオチド

50

(B) 型：核酸		
(C) 鎖の数：二本鎖		
(D) トポロジー：直鎖状		
(xi) 配列：配列番号19：		
NNNATACAGC TGCATCCCTG ACTCCCCGAG	30	
(2) 配列番号20の情報：		
(i) 配列の特徴：		
(A) 長さ：30ヌクレオチド		
(B) 型：核酸		
(C) 鎖の数：二本鎖		10
(D) トポロジー：直鎖状		
(xi) 配列：配列番号20：		
NNCNTACAGC TGCATCCCTG TGTTCGCGCG	30	
(2) 配列番号21の情報：		
(i) 配列の特徴：		
(A) 長さ：30ヌクレオチド		
(B) 型：核酸		
(C) 鎖の数：二本鎖		
(D) トポロジー：直鎖状		
(xi) 配列：配列番号21：		20
NNNCTACAGC TGCATCCCTC TACAGCAGCG	30	
(2) 配列番号22の情報：		
(i) 配列の特徴：		
(A) 長さ：30ヌクレオチド		
(B) 型：核酸		
(C) 鎖の数：二本鎖		
(D) トポロジー：直鎖状		
(xi) 配列：配列番号22：		
NNGNTACAGC TGCATCCCTG TCGCGTCGTT	30	
(2) 配列番号23の情報：		30
(i) 配列の特徴：		
(A) 長さ：30ヌクレオチド		
(B) 型：核酸		
(C) 鎖の数：二本鎖		
(D) トポロジー：直鎖状		
(xi) 配列：配列番号23：		
NNNGTACAGC TGCATCCCTC GGAGCAACCT	30	
(2) 配列番号24の情報：		
(i) 配列の特徴：		
(A) 長さ：30ヌクレオチド		40
(B) 型：核酸		
(C) 鎖の数：二本鎖		
(D) トポロジー：直鎖状		
(xi) 配列：配列番号24：		
NNTNTACAGC TGCATCCCTG GTGACCGTAG	30	
(2) 配列番号25の情報：		
(i) 配列の特徴：		
(A) 長さ：30ヌクレオチド		
(B) 型：核酸		
(C) 鎖の数：二本鎖		50

(D) トポロジー：直鎖状		
(xi) 配列：配列番号25：		
NNNTTACAGC TGCATCCCTC CCCTGTCGGA	30	
(2) 配列番号26の情報：		
(i) 配列の特徴：		
(A) 長さ：78ヌクレオチド		
(B) 型：核酸		
(C) 鎖の数：一本鎖		
(D) トポロジー：直鎖状		
(xi) 配列：配列番号26：		10
ACTAATCGTC TCACTATTTA ATTAANNNNN NNNNNNNNNN	40	
NNNNNNNNNN NNNNNNNGGT TTTTTTTTTT TTTTTTV	78	
(2) 配列番号27の情報：		
(i) 配列の特徴：		
(A) 長さ：17ヌクレオチド		
(B) 型：核酸		
(C) 鎖の数：二本鎖		
(D) トポロジー：直鎖状		
(xi) 配列：配列番号27：		
ATAGGGGTCT TCGGTAC	17	20
(2) 配列番号28の情報：		
(i) 配列の特徴：		
(A) 長さ：19ヌクレオチド		
(B) 型：核酸		
(C) 鎖の数：二本鎖		
(D) トポロジー：直鎖状		
(xi) 配列：配列番号28：		
GATCAGCTGC TGCAAATTT	19	
		30

【図 1 A】

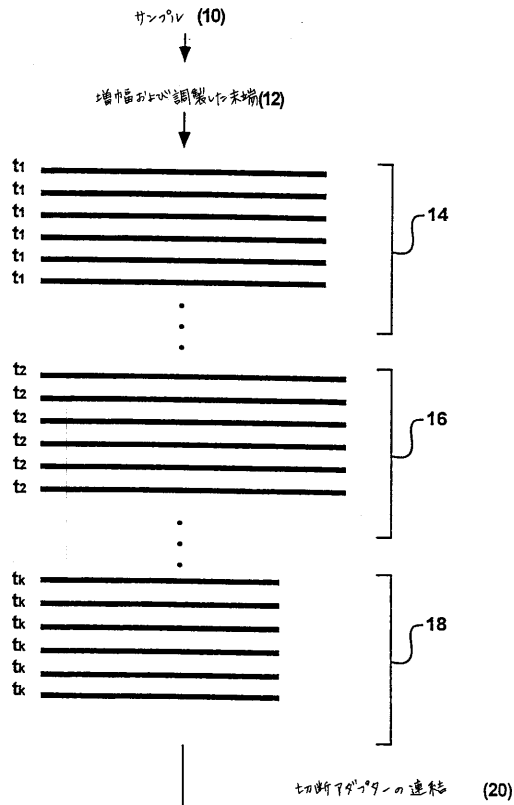


Fig. 1A

【図 1 B】

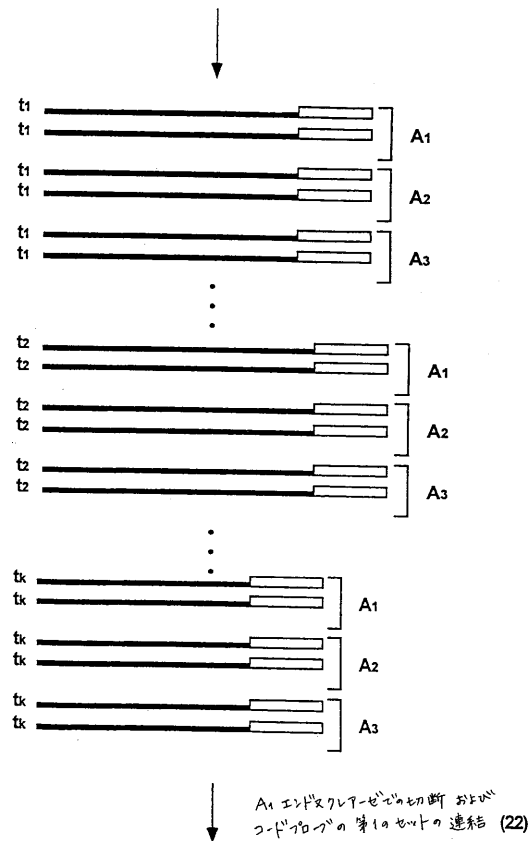


Fig. 1B

【図 1 C】

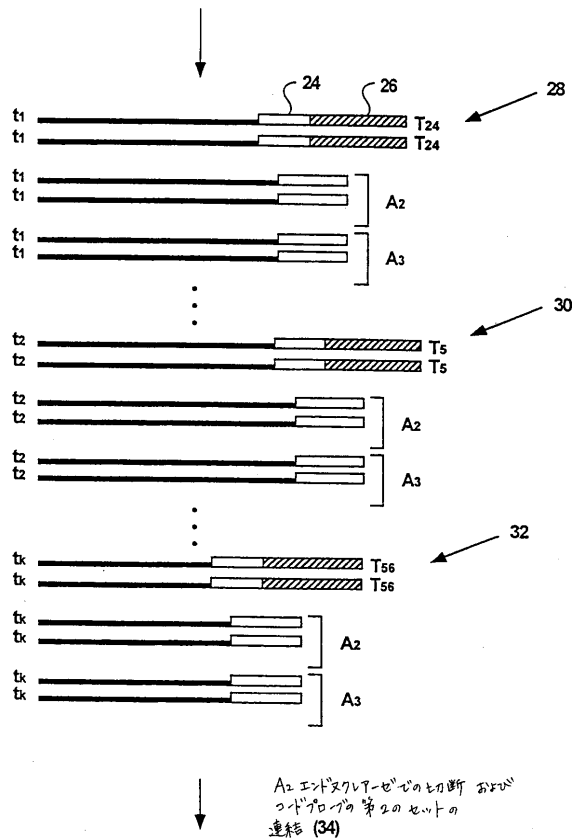


Fig. 1C

【図 1 D】

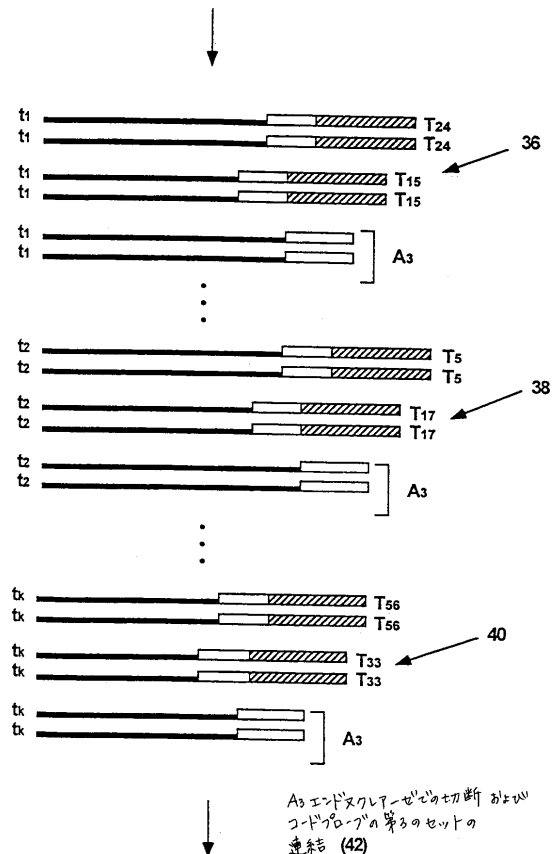


Fig. 1D

【図 1 E】

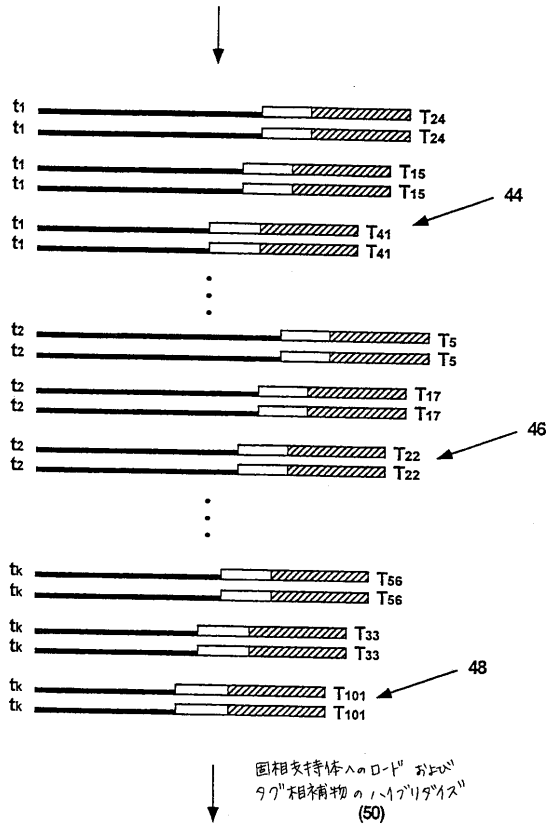


Fig. 1E

【図 2】

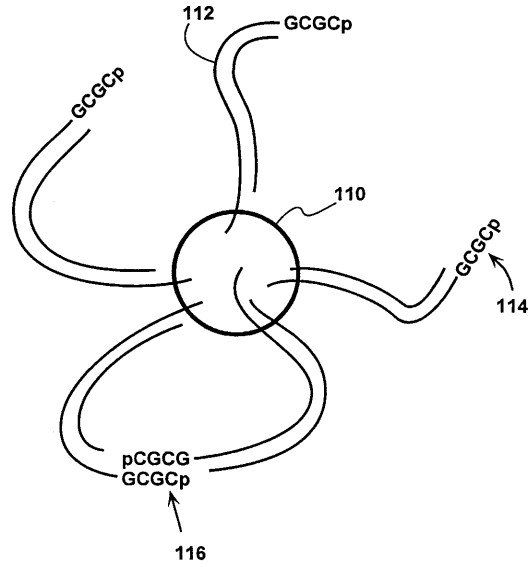


Fig. 2

【図 3 A】

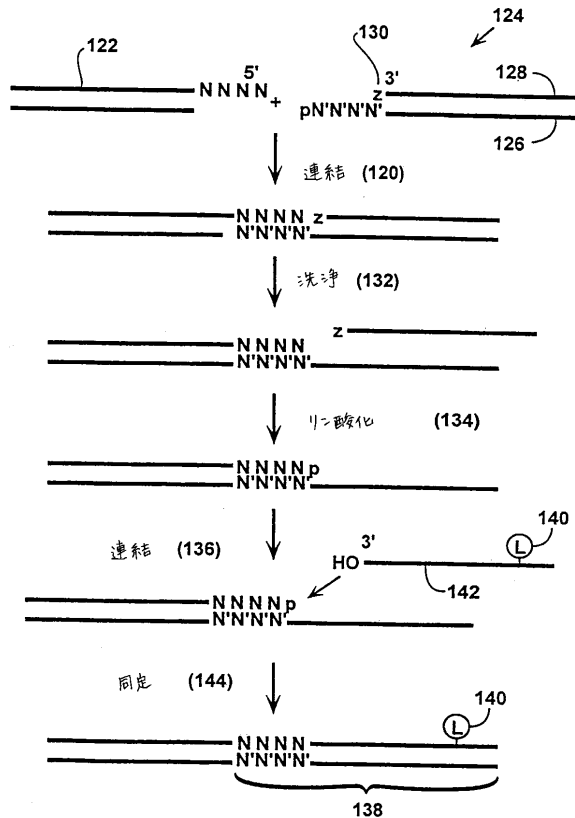


Fig. 3A

【図 3 B】

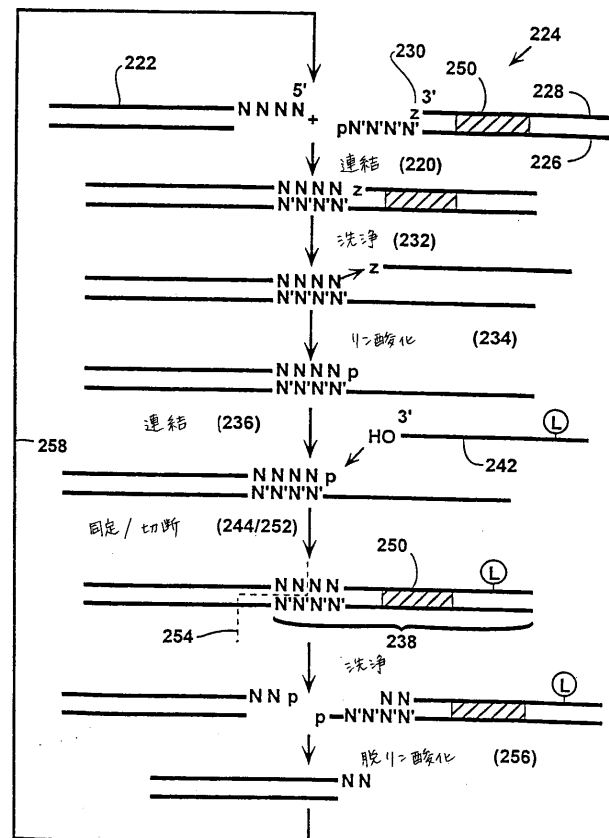


Fig. 3B

【図 4】

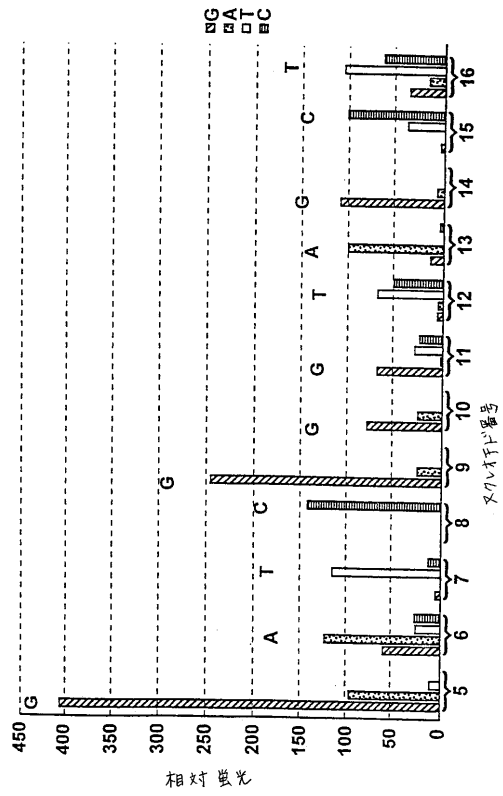
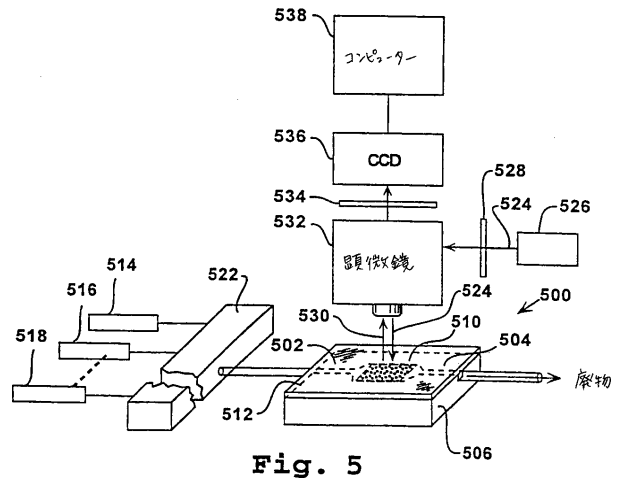


Fig. 4

【図 5】



フロントページの続き

- (72)発明者 ブレナー, シドニー
イギリス国 シービー 2 3 ピー ジェイ ケンブリッジ, セント エドワーズ パッセージ 17
ビー
- (72)発明者 ロイド, デイビッド エイチ.
アメリカ合衆国 カリフォルニア 94014, デイリイ シティ, ナンバー 1, ポイント パ
シフィック ドライブ 850
- (72)発明者 ダブリッジ, ロバート ビー.
アメリカ合衆国 カリフォルニア 94002, ベルモント, ホーリー ロード 825
- (72)発明者 パラス, マイケル シー.
アメリカ合衆国 カリフォルニア 94066, サン ブルノ, オーク アベニュー 408

審査官 斎藤 真由美

- (56)参考文献 国際公開第95/020053 (WO, A1)
国際公開第96/012014 (WO, A1)

(58)調査した分野(Int.Cl., D B 名)

C12N 15/00 - 90
C12N 1/00 - 9/99
C12P 1/00 - 41/00
C12Q 1/00 - 70
G01N 33/50 - 98
PubMed, MEDLINE (STN)
BIOSIS/WPI (DIALOG)