

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第3954092号  
(P3954092)

(45) 発行日 平成19年8月8日(2007.8.8)

(24) 登録日 平成19年5月11日(2007.5.11)

(51) Int.C1.

F 1

C 12 N 15/09 (2006.01)  
C 12 Q 1/68 (2006.01)C 12 N 15/00  
C 12 Q 1/68A  
A

請求項の数 7 (全 29 頁)

(21) 出願番号 特願平7-503084  
 (86) (22) 出願日 平成6年6月24日(1994.6.24)  
 (65) 公表番号 特表平9-501561  
 (43) 公表日 平成9年2月18日(1997.2.18)  
 (86) 国際出願番号 PCT/US1994/007106  
 (87) 国際公開番号 WO1995/000530  
 (87) 国際公開日 平成7年1月5日(1995.1.5)  
 審査請求日 平成11年7月30日(1999.7.30)  
 (31) 優先権主張番号 08/082,937  
 (32) 優先日 平成5年6月25日(1993.6.25)  
 (33) 優先権主張国 米国(US)

(73) 特許権者 399125757  
 アフィメトリックス インコーポレイテッド  
 アメリカ合衆国 カリフォルニア州 95051 サンタクララ セントラル エクスプレス ウェイ 3420  
 (74) 代理人 100078282  
 弁理士 山本 秀策  
 (74) 代理人 100062409  
 弁理士 安村 高明  
 (74) 代理人 100113413  
 弁理士 森下 夏樹

最終頁に続く

(54) 【発明の名称】核酸配列のハイブリダイゼーションと配列決定

## (57) 【特許請求の範囲】

## 【請求項 1】

標的よりも少數の塩基をもつ多数の核酸プローブにより標的核酸を配列決定する方法であつて：

(a) そのプローブをその標的と接触させる段階と；  
 (b) その標的に最も高い結合アフィニティーを有する第1プローブを同定する段階と；  
 (c) 少なくとも2の異なるプローブを含んでなる伸長プローブの第1セットを選ぶ段階であつて、その少なくとも2の異なるプローブは、第1プローブの配列の全てもしくは一部に加え、その第1プローブ配列の1以上の可能性のある伸長を含む、段階と；および  
 (d) その伸長プローブの第1セットの他のものよりも強くその標的に特異的にハイブリダイズするその伸長プローブの第1セットの中の1を同定し、それによりその伸長プローブの第1セット中の1がその標的核酸の中の1つの塩基を同定する段階とを含んで成り、

上記同定の段階(d)が、1塩基ミスマッチ・プローブであつて、伸長プローブの第1セットのA, C, T, U、及びGモノ置換の中の少なくとも2を含んで成るものと同定し；その1塩基ミスマッチ・プローブのハイブリダイゼーション・アフィニティー・データを記録し；そして

そのハイブリダイゼーション・アフィニティー・データがその塩基ミスマッチ・プローブの予想ハイブリダイゼーション・アフィニティー・データを確信させるときその第1プローブの正しい伸長として伸長プローブの第1セットの中の1を選ぶ、

段階をさらに含み、

その予想されるハイブリダイゼーション・データが：

伸長プローブの末端においてミスマッチをもつプローブ／標的複合体についてのより高い結合アフィニティー；及び

その複合体の内部においてミスマッチをもつプローブ／標的複合体についてのより低い結合アフィニティー、

を含んで成る

方法。

【請求項 2】

核酸プローブの実質的に全てが  $n$  ヌクレオチドを含んで成り、そして伸長プローブが第 1 プローブの  $n - 1$  ヌクレオチドを含んで成る、請求項 1 に記載の方法。 10

【請求項 3】

伸長プローブの第 1 セットから反対の方向に伸長する伸長プローブの第 2 セットを選び；そして

その伸長プローブの第 2 セットの他のものよりも強く標的に特異的にハイブリダイズする  
その伸長プローブの第 2 セットの中の 1 を同定し、それにより、その伸長プローブの第 2 セットの中の 1 が標的核酸内の第 2 の塩基を同定する、

段階をさらに含んで成る、請求項 1 に記載の方法。

【請求項 4】

伸長プローブのセットを選ぶ段階を繰り返し、そして 5 回以上伸長プローブを同定する、  
段階をさらに含んで成る、請求項 1 に記載の方法。 20

【請求項 5】

ハイブリダイゼーション・データが伸長プローブの中の 1 についてのハイブリダイゼーション値に対して正規化され；そして同定の段階 (d) が、その伸長プローブの中の 1 つの正規化された値よりも高い正規化されたハイブリダイゼーション値をもたない末端 1 塩基ミスマッチ・プローブをもつ伸長プローブの中の 1 を選ぶことを含んで成る、  
請求項 1 に記載の方法。

【請求項 6】

同定の段階 (d) が、他の伸長プローブよりも高い、標的にに対する結合アフィニティーを示す伸長プローブのセットの中の 1 を選ぶ段階を含んで成る、請求項 1 に記載の方法。 30

【請求項 7】

同定の段階 (d) が、適当にプログラムされたコンピューター内で行われる、請求項 1 に記載の方法。

【発明の詳細な説明】

政府の権利

本明細書中に記載する発明は、エネルギー省と Affymax との間の契約第 DE - FG03 - 92ER 81 275 号（認可第 21012 - 92 - 11）の中で又はその下で；並びに NIH 契約第 1R01HG 00813 - 01 号の中で又はその下で生じた。

発明の背景

本発明は、核酸分析、検出、及び配列決定の分野に関する。より特に、1 の態様においては、本発明は、核酸のアレイ (arrays) の合成、核酸のハイブリダイジング、一本鎖プローブと標的核酸から成る 2 本鎖核酸内のミスマッチの検出、及び DHA 又は RNA あるいは他のポリマーの配列の決定のための改良技術を提供する。 40

多くの分野において核酸の配列を決定することは重要である。なぜなら、例えば核酸は酵素、構造タンパク質、及び生物学的機能の他のエフェクターをコードしているからである。ポリペプチドをコードする核酸のセグメントに加えて、遺伝子発現の制御及び調節に関する多くの核酸配列が存在する。

ヒト：ゲノム計画は、核酸配列決定技術を使用した計画の 1 例である。この計画は、ヒト生物のゲノムの完全な配列を決定することを目的としている。このような配列は、各々の特定の個人の配列とは必ずしも一致しないであろうけれども、その生物一般に関する有意

義な情報と特定の個人からのゲノム・セグメント内に含まれる特定配列を提供するであろう。このヒト・ゲノム計画は、さらに詳細な研究に有用なマッピング情報をも提供するであろう。

配列決定計画、例えばヒト・ゲノム計画の要求におけるよりも明らかな、高迅速で、正確で、そして安価が配列決定技術についての必要性は、どこにもない。ヒト・ゲノムの配列決定を完成するためには、約  $3 \times 10^9$ 、又は30億の、塩基対の決定が必要であろう。

配列決定のために今日典型的に使用される手順は、Sanger et al., Proc. Natl. Acad. Sci. USA (1977) 74: 5463-5467、とMaxam et al., Methods in Enzymology (1980) 65: 499-559中に記載された方法を含む。このサンガー法は、鎖延長停止ジデオキシ・ヌクレオチドによる酵素的伸長を使用する。マキサム・ギルバート法は、塩基特異的解裂反応を示す化学反応を使用する。両方法は、多数の複雑な操作、例えば同種DNA断片の単離、サンプルの骨の折れる、かつ、のろい調製、分離用ゲルの調製、そのゲルへのサンプルの適用、そのゲル上でのサンプルの電気泳動、最終ゲルの調製、及びその手順の結果の分析を必要とする。

他の技術も核酸の配列決定のために提案してきた。全目的をもって引用により本明細書中に取り込むPCT特許公開第92/10588号は、標識された、標的核酸の配列が支持体上の核酸プローブのアレイへのハイブリダイゼーションにより決定されるような1の改良技術を記載している。各プローブは、その支持体上に位置的に区別できる場所に位置する。この標識された標的がその支持体に晒されるとき、それは、相補的ヌクレオチド配列を含む位置において結合する。その結合位置におけるプローブの配列の知識を通じて、ある者は、標的核酸のヌクレオチド配列を決定することができる。この技術は、ひじょうに大きなアレイの核酸プローブが使用されるとき、特に効率的である。このようなアレイは、Pirrung et al.に付与された米国特許第5,143,854号中に記載された技術に従って形成されることができる。また、米国出願逐次番号第07/805,727号をも参照のこと。両方を、全目的をもって引用により本明細書中に取り込む。

これらの核酸プローブが標的よりも短い長さをもつとき、ある者は、より短いプローブからのアフィニティー・データに基づき、より大きな標的の配列を決定するための再構築技術を使用することができる。Drmanac et al.への米国特許第5,202,231号、とSouthernへのPCT特許公開第89/10977号を参照のこと。この困難性を克服するための1つの技術は、ハイブリダイゼーション又はSBHによる配列決定といわれてきた。例えば、12-merの標的DNA 5'-AGCCTAGCTGAAが全部で8ヌクレオチドのプローブのアレイと混合されると仮定する。この標的が正確に相補的なヌクレオチド配列をもつようなプローブだけと結合する場合、65,536のオクタマーの中のたった5つのプローブ (3'-TCGGATCG, CGGATCGA, GGATCGAC, GATCGACT, とATCGACT) がこの標的にハイブリダイズするであろう。これらのハイブリダイジング・プローブからの重複配列の整列は、その元の12-merの標的の相補物を再構築する：

TCGGATCG  
CGGATCGA  
GGATCGAC  
GATCGACT  
ATCGACTT  
TCGGATCGACTT

多くの楽観に会ったにもかかわらず、従来技術は、特定の限定とも会った。例えば、従事者は、他の困難性の中で、部分的にミスマッチの配列のハイブリダイゼーションのために標的核酸にハイブリダイズしたプローブのアレイを分析することにおける実質的な困難性に遭遇した。本発明は、このようなアレイによる配列決定における有意な進歩を提供する。

#### 発明の要約

核酸（オリゴヌクレオチド）を合成し、ハイブリダイズし、分析し、そして配列決定するための改良された技術を本発明により提供する。

本発明の1の態様に従って、標的オリゴヌクレオチドが多数の、より短い長さの固定化さ

10

20

30

40

50

れたプローブに晒される。これらのプローブを、集合的に、“アレイ (array)” という。本法においては、ある者は、標的に高いアフィニティーをもつコア・プローブを最初に固定し、そして次にこのコア・プローブと比較したとき 1 塩基のミスマッチをもつ全てのプローブの結合特性を評価することにより、標的核酸がそのアレイ内のプローブに相補的であるかどうかを同定する。この 1 塩基ミスマッチ・プローブが特徴的な結合又はアフィニティー・パターンを示す場合、このときそのコア・プローブは、その標的核酸の少なくとも一部に正確に相補的である。

本法は、コア・プローブの“左”と“右”伸長といわれることができるプローブの結合アフィニティーを評価することによりそのアレイ内のいずれのプローブよりも大きな標的核酸を配列決定するために拡大されることができる。コアの正しい左と右伸長は、1 塩基ミスマッチ・プローブの特定のハイブリダイゼーション・パターン及び / 又は最も強い結合アフィニティーを示すようなものである。1 塩基ミスマッチ・プローブの結合アフィニティー特定は、その 3 又は 5 末端上にミスマッチをもつプローブ / 標的複合体が内部ミスマッチをもつプローブ / 標的複合体よりも安定である特性パターンに従う。この方法は、次に、核酸標的の配列を提供するために、コア・プローブの追加の左と右伸長を決定するために繰り返される。

いくつかの態様、例えば診断においては、標的は、特定の配列をもつと予想される。標的が予想配列をもつかどうかを決定するために、相補的プローブ並びに 1 塩基ミスマッチ・プローブの全ての全サブセット又はその中のいくつかのサブセットを含むプローブのアレイが合成される。このようなプローブへのこの標的のハイブリダイゼーション・パターンの分析を通じて、その標的が予想配列をもつかどうか決定でき、そしてそうでない場合その標的の配列が場合により決定されることができる。

核酸標的の分析のためのキットも、本発明により提供される。1 の態様に従えば、キットは、核酸プローブのアレイを含む。このプローブは、標的核酸に対する完全な相補物を含むことができる。これらのプローブは、その完全に相補的なプローブの 1 塩基置換物であるプローブをも含む。このキットは、完全相補物の 1 以上の A , C , T , G 、及び / 又は U 置換を含むことができる。このようなキットは、特定の遺伝子配列についての標的の分析を含むさまざまな用途、例えば遺伝子疾患のための分析におけるものをもつであろう。本明細書中における本発明の性質と利点のさらなる理解は、本明細書の残りの部分と添付図面を参照することにより理解されるであろう。

#### 【図面の簡単な説明】

図 1 は、オリゴヌクレオチドの光指定合成 (light-directed synthesis) を説明する。光保護されたヒドロキシル (OH) を担持する表面 (2) が、その光脱保護された領域内で遊離のヒドロキシル (OH) を生成するフォトリトグラフ・マスク (photolithographic mask) (M<sub>1</sub>) を通して照射される。これらのヒドロキシル基は次に、5 - 光保護されたデオキシヌクレオシド・ホスホルアミジット (phosphoramidite) (例えば、T - X) にカップリングされる。新たなマスク (M<sub>2</sub>) を、その表面上に新たなパターンを照射するために使用し、そして第 2 の光保護されたホスホルアミジット (例えば、C - X) を次にカップリングする。照射とカップリングのラウンドを、オリゴヌクレオチド・プローブの所望のセットが得られるまで繰り返す。場合により標識 (\*) をもつ、標的 (R) を、これらのオリゴヌクレオチドに晒す。標的がアレイに結合する位置を、その標的の配列を決定するために使用する；

図 2 は、オリゴヌクレオチドのハイブリダイゼーションと熱解離を説明し、プローブのアレイにハイブリダイズした標的核酸 (5 - GCGTAGGC - フルオレセイン) の蛍光スキャンを示す。この支持体表面は、488nm アルゴン・イオン・レーザー励起を使用して Zeiss Axio scop 20 顕微鏡により走査された。520nm を超える蛍光エミッションを、光子計数モードにおいて操作された冷却光増幅装置 (Hamamatsu 934-02) を使用して検出した。シグナル強度は、この画像の右に示したスケール上に示される。温度は において各パネルの右に示される；

図 3 は、ハイブリダイゼーションの配列特異性を説明する。(A) は、各合成部位における

10

20

20

30

40

40

50

るプローブ組成の係数である。3 - CGCATCCG表面固定プローブ(本明細書中S - 3 - CGCATCCGという。)がストライプ1, 3、と5内で合成され、そしてプローブS - 3 - CGCTTCCGがストライプ2, 4、と6内で合成された。(B)は、標的核酸(10nM 5 - GCGTAGGC - フルオレセイン)とのこの支持体のハイブリダイゼーションを示す蛍光画像である。ハイブリダイゼーションを15分間、15において $6 \times \text{SSPE}$ , 0.1% Triton X - 100中で行った。(C)は、(B)のハイブリダイゼーション溶液に添加された第2核酸(10nM 5 - GCGAAGGC)とのハイブリダイゼーションを示す蛍光画像である。(D)は、(1) (C)からのフルオレセイン化標的の高湿度解離；及び(2) 15分間15における標的核酸(10nM 5 - GCGAAGGC)との支持体のインキュベーション、の後のハイブリダイゼーション結果を示す蛍光画像である。(E)は、(D)のハイブリダイゼーション溶液に添加された第2核酸(10nM 5 - GCGTAGGC)とのハイブリダイゼーションを示す蛍光画像である；

図4は、 $4^4$ テトラヌクレオチドの組合せ合成(combinatorial synthesis)を説明する。ラウンド1においては、合成領域の1/4が、第1MeNPoc-ヌクレオシド(このケースにおいてはT)のカップリングのためにマスク1を通しての照射により活性化される。ラウンド1のサイクル2においては、マスク2は、その合成支持体の異なる1/4セクションを活性化し、そして異なるヌクレオシド(C)がカップリングされる。さらなるアレイのリトグラフィー的亜分割及び化学的カップリングは、完全な256テトラヌクレオチドのセットを生り出す；

図5Aと5Bは、256オクタヌクレオチドのアレイへのハイブリダイゼーションを説明する。図5Aは、15における15分間 $6 \times \text{SSPE}$ , 0.1% Triton X - 100中での標的核酸(10nM 5 - GCGGCAGGC - フルオレセイン)とのアレイのハイブリダイゼーション後の蛍光画像である。図5Bは、S - 3 CG(A + G + C + T) $^4$ CGの合成の間に作られた各々のプローブがどこに位置するかを示すマトリックス解読器である。プローブ配列S - 3 - CGCGCCCCGを含む部位を暗領域として示す。本明細書中に使用する組合せ合成の表記法(notation)は、全目的をもって引用により本明細書中に取り込む米国出願逐次番号第07/624,120号中に十分に記載されている；

図6A～6Cは、k - merプローブを使用したn - mer標的を配列決定するための技術を説明する。図6Aは、支持体上のプローブにハイブリダイズした標的を説明する。図6Bと6Cは、通常の結合アフィニティー対ミスマッチ位置のプロットを説明する；

図7は、ハイブリダイゼーション実験の蛍光画像を説明する；

図8は、1塩基ミスマッチの関数としてハイブリダイゼーション事件をグラフにより説明する；

図9は、ミスマッチの対の関数としての蛍光強度を説明する；

図10は、1塩基ミスマッチ実験の蛍光画像を説明する；

図11A～11Cは、各種1塩基ミスマッチ特性を説明する；

図12A～12Dは、より短いk - 員プローブからのハイブリダイゼーション結果に基づくn - 員(そのヌクレオチジ内のモノマーの数)の標的オリゴヌクレオチドのヌクレオチド配列の決定方法を説明する。特に、図12A～12Dは、4 - 塩基プローブにより10塩基標的を配列決定するための本発明の適用について説明する；

図13は、ヌクレオチド配列の決定のためのコンピューター・システムを説明する；

図14は、標的核酸のミスマッチ分析のための、そして、その決定のためのコンピューター・プログラムを説明する；

図15Aと15Bは、いくつかの可能性の中の選択により標的核酸のヌクレオチド配列を決定するためのコンピューター・プログラム、並びにこのコンピューター・プログラムにおける使用のためのスコアリング・ルーチン(Scoring routine)の例を説明する；

図16は、ヌクレオチド配列の決定における使用のための指定グラフを説明する；

図17Aと17Bは、1塩基ミスマッチ特性を使用した野生型と突然変異の分析を説明する；

図18は、1塩基ミスマッチ・テストの蛍光画像である；そして

図19A～19Dは、核酸配列同定のための技術を説明する。

10

20

30

40

50

## 好ましい態様の説明

## 目 次

- A . 合成
- B . ハイブリダイゼーション
- C . ミスマッチ分析
- D . 適用
- E . 結論

定義

プローブ - 典型的には固体表面上で形成された公知の組成又はモノマー配列の分子であつて、標的分子に晒され又は晒されてもよく、そしてそのプローブがその標的にハイブリダイズされたかどうかを決定するために検査され又は検査されてもよいもの。“コア ( core )”プローブは、標的に強いアフィニティーを示すプローブである。“伸長 ( extension )”プローブは、コア・プローブ配列の全部又は一部に加えそのコア・プローブ配列の 1 以上の可能性のある伸長を含むプローブである。本明細書は、プローブの 3' - 末端における伸長として“左”伸長と、そしてプローブの 5' - 末端における伸長として“右”伸長という。但し、反対の表記法も明らかに採用されることができるであろう。

10

標的 - それについて、その組成又はモノマー配列を研究することが望ましい、典型的には、未知の組成又はモノマー配列をもつ分子。標的は、より大きな分子の一部、例えば、より長い核酸内の数塩基であることができる。

n - 塩基ミスマッチ - n が 1 以上である、コアプローブ内の対応モノマーとは異なる、n モノマーをその中にもつプローブ。

20

A , T , C , G , U - は、それぞれ、以下のヌクレオチド、アデニン、チミン、シトシン、グアニン、ヒドロキシ尿嘧啶についての略号である。

ライブラリー - 1 以上の支持体内にしばしば形成される、所定ヌクレオチド配列の核酸プローブのコレクションであって、標的核酸のハイブリダイゼーション試験において使用されるもの。

## A . 合成

光指定オリゴヌクレオチド合成のための方法を図 1 中に示す。このような戦略は、本発明者の譲受人に譲渡され、そして全目的をもって引用により本明細書中に取り込む米国特許第 5,143,854 号中により詳細に記載されている。

30

図 1 において説明する光指定合成法においては、光不安定性保護基又は基 ( X ) により誘導体化された表面 ( 2 ) を、フォトリトグラフ・マスク ( photolithographic mask ) M<sub>1</sub> を通して照射して、反応性ヒドロキシル ( OH ) 基を露出させる。(光不安定性の保護基により 5' - ヒドロキシルにおいて保護された) ホスホルアミジット活性化ヌクレオシドのシリーズの第 1 の ( T - X ) を次に表面全体に晒す。先行する照射の間に先に晒された部位においてのみカップリングが生じる。

上記カップリング反応が完了した後、支持体を濯ぎ、そしてその表面を再び、新たな又は平行移動されたマスク M<sub>2</sub> を通して照射して、カップリングのための異なる基を露出させる。(光不安定性の保護基により 5' - ヒドロキシルにおいて再び保護された) 新たなホスホルアミジット活性化ヌクレオシド C - X を添加し、そして上記露出部位にカップリングさせる。上記工程を、光脱保護とカップリングのサイクルを通じて繰り返して、支持体上にオリゴヌクレオチド・プローブの所望のセットを作り出す。フォトリトグラフィーが使用されるので、この工程は微細にされることができる。さらに、反応が光により場所的にアドレスされた部位においてのみ生じるので、各部位におけるプローブのヌクレオチド配列は正確に知られ、そして標的分子(標的核酸又は、他の態様においては、タンパク質、例えばレセプタのいずれか)との、各部位におけるオリゴヌクレオチド・プローブの相互作用を評価することができる。

40

5' - O - ( - メチル - 6' - ニトロピペロニルオキシカルボニル ) - N - アシリル - 2' - デオキシヌクレオシド、又は MeNPoc - N - アシリル - デオキシヌクレオシド、MeNPoc - dT , MeNPoc - dC<sup>1-bu</sup> , MeNPoc - dG<sup>PAC</sup> 、及び MeNPoc - dA<sup>PAC</sup> を含む光保護デオキシヌクレオシ

50

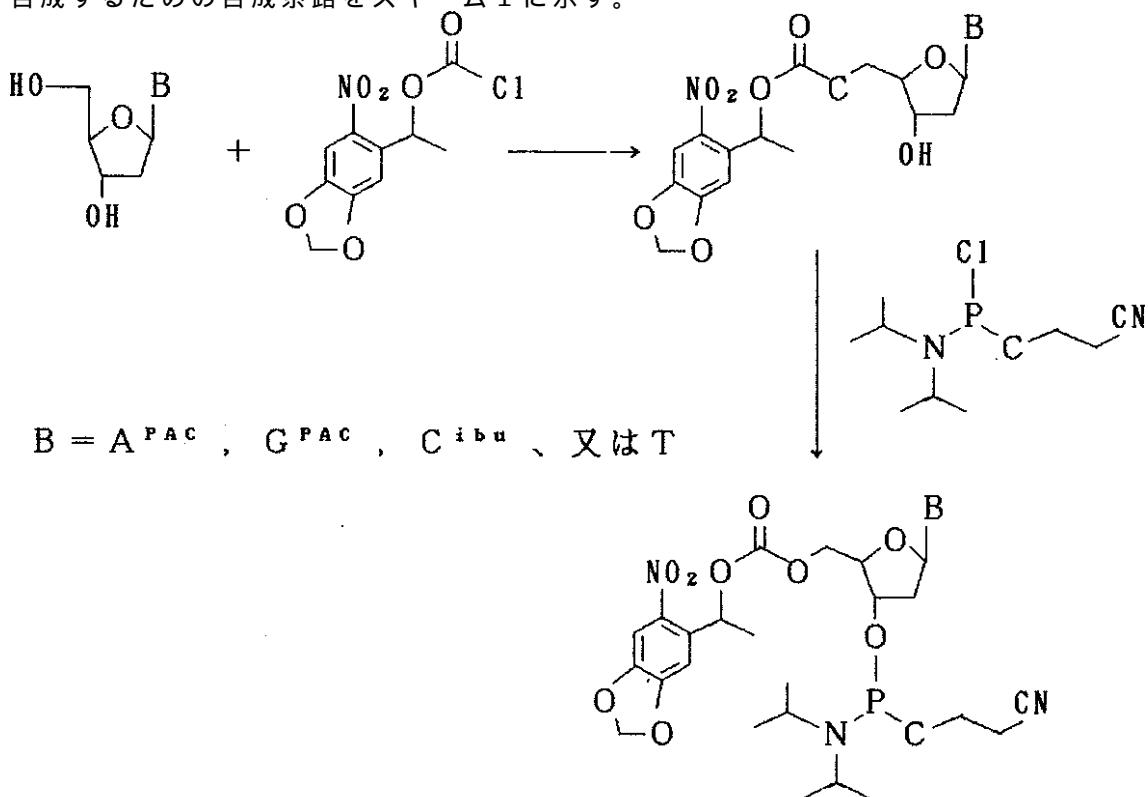
ドを本工程のために開発した。保護基化学は、PCT特許公開第92/10092号及び1990年12月6日に出願された米国出願逐次番号第07/624,120号、及び1992年12月2日に出願された第07/971,181号であって、共に本発明の譲受人に譲渡され、そして全目的をもって引用により本明細書中に取り込まれたものの中にかなり詳細に開示されている。

### 実施例

#### 1. 保護基

塩基は280nm領域内に強い  $\lambda^*$  が遷移をもつて、光除去性保護基の脱保護波長は、不所望のヌクレオシド光化学を回避するために280nmよりも長い波長でなければならない。さらに、4つのデオキシヌクレオシドの光脱保護速度は、光が、全照射合成部位内で、ヒドロキシル（又は他の官能基、例えばスルフヒドリル又はアミノ基）を均しく脱保護する 10 であろうように、近似していなければならない。

これらの基準に適合するために、1セットの、5'-O-（-メチル-6-ニトロピペロニルオキシカルボニル）-N-アシル-2'-デオキシヌクレオシド（MeNPoc-N-アシル-デオキシヌクレオシド）が光指定合成のために開発され、そして保護されたヌクレオシドの光速度論的挙動が測定された。5'-O-（-メチル-6-ニトロピペロニルオキシカルボニル）-N-アシル-2'-デオキシヌクレオシド・ホスホルアミジットを合成するための合成系路をスキームIに示す。



スキーム I

第1段階において、N-アシル-2'-デオキシヌクレオシドを1-(2-ニトロ-4,5-メチレンジオキシフェニル)-エタン-1-クロロホルメートと反応させて、5'-MeNPoc-N-アシル-2'-デオキシヌクレオシドを作る。第2段階において、その3'-ヒドロキシルを、標準的な手順を用いて2-シアノエチル-N,N-ジイソプロピルクロロホスホルアミジットと反応させて、5'-MeNPoc-N-アシル-2'-デオキシヌクレオシド-3'-O-ジイソプロピルクロロホスホルアミジットを作る。これらの試薬は、4においてアルゴン下で乾燥保存されるとき長時間安定であった。

4つのデオキシヌクレオシド、MeNPoc-dT, MeNPoc-dC<sup>ibu</sup>, MeNPoc-dG<sup>PAC</sup>、とMeNPoc-dA<sup>PAC</sup>のそれぞれの0.1mM溶液をジオキサン中で調製した。アリコート（200mL）を、さまざまな時間にわたり狭い経路（2mm）の石英キュベット内で14.5mW/cm<sup>2</sup>の365nm光で照射した。4～5の時間点を、各塩基について採取し、そしてそれらの溶液を、0.1(v/v) 50

%TFAを含む水中60( v / v ) %の移動相により溶出して、280nmにおけるHPLC装置とヌクレオジル5 - C<sub>8</sub>HPLCカラムにより出発材料の損失について分析した( MeNPoc - dTは水中70( v / v ) %メタノールの移動相を必要とした。)。残ったMeNPoc - N - アシル - デオキシヌクレオシドのピーク面積を計算して、MeNPoc - dT, MeNPoc - dC<sup>ib</sup>u, MeNPoc - dG<sup>PAC</sup>、とMeNPoc - dA<sup>PAC</sup>について、それぞれ28秒、31秒、27秒、と18秒の光分解半減期を得た。その後のリトグラフ実験において、4, 5分間の照射時間(  $9 * t_{1/2}^{\text{MeNPoc - dC}}$  )は、99%以上のMeNPoc保護基の除去を導いた。

光指定合成においては、全体の合成収率は、光脱保護の収率、光脱保護のコントラスト、と化学的カップリングの効率に依存する。光速度論的条件は、好ましくは、光脱保護の収率が99%を超えることを確保するように選ばれる。支持体の通常暗い領域内の不所望の光分解は、その合成忠実度( fidelity )に悪影響を及ぼすことができるが、高吸光度( 500D U )によるリトグラフ・マスクを使用することにより、そして光学表面の注意深いインデックス・マッチング( index matching )により最小化されることができる。ガラス支持体へのDMT - N - アシル - デオキシヌクレオシド・ホスホルアミジットの縮合効率は、95% ~ 99%のレンジ内で測定された。MeNPoc - N - アシル - デオキシヌクレオシド・ホスホルアミジットの縮合効率も、90%を上回って測定された。但し、それらの効率は、合成間で変動することができ、そして監視されなければならない。

## 2. カップリング効率の測定

光保護されたヌクレオシドのカップリング効率を調べるために、4つのMeNPoc - アミジットのそれぞれを最初に、( DMT化学物質を介して )支持体にカップリングした。その支持体の領域を照射し、そしてMeNPoc - ホスホルアミジットを保護基なしで添加した。その支持体の新たな領域を次に照射し；蛍光デオキシヌクレオシド・ホスホルアミジット( FAM - ホスホルアミジット Applied Biosystems )をカップリングして；そしてその支持体をシグナルについて走査した。蛍光標識されたホスホルアミジットが、新たに露出されたヒドロキシル基と先に未反応ヒドロキシル基の両方において反応する場合、その時、2つの部位の間の蛍光強度の比は、そのカップリング効率の尺度を提供する。この計測は、表面光分解の収率が不变に近いということを受け入れる。この又は近似の検定を使用した化学的カップリング収率は、変動性であるが、高く、80 ~ 95%の間のレンジにある。

別個の検定において、化学的カップリング効率は、ヘキサエチレングリコール誘導体化支持体上で測定される。まず、グリコール・リンカーを脱トリチル化し、MeNPoc - デオキシヌクレオシド - O - シアノエチルホスホルアミジットをキャッピングなしでその樹脂にカップリングした。次に、DMT - デオキシヌクレオシド - シアノエチルホスホルアミジット( リポーター - アミジット )をその樹脂にカップリングした。このリポーター - アミジットは、上記第1段階からのいずれかの未反応のヒドロキシル基にカップリングする。このトリチル溶出液を集め、そして吸収分光光度計により定量した。溶出液を、そのデリバリー・ライン内に残った残存トリチルを測定するためにMeNPoc - ホスホルアミジット・カップリング直後にそのラインからも集めた。この検定において、カップリング効率は、100%の、リポーター - アミジットのカップリング効率を呈すると測定される。ヘキサエチレングリコール・リンカーへのMeNPoc - デオキシリボヌクレオシド - O - シアノエチルホスホルアミジットのカップリング効率と16のジヌクレオチドの効率が測定され、そしてDMT - デオキシヌクレオシド・ホスホルアミジットとは区別できなかった。

## 3. オリゴヌクレオチド・プローブの場所的に指定された合成

オリゴヌクレオチド・プローブの合成を始めるために、支持体を調製し、そしてMeNPoc - dC<sup>ib</sup>u - 3 - O - ホスホルアミジットを合成リンカーを通じて合成サポートに付着させた。このサポートに付着させた。このサポートの領域を、リトグラフ・マスクの800 × 1280 μmの孔を通しての照射により合成のために活性化した。7つの追加のホスホルアミジット合成サイクルを( 対応のDMT保護されたデオキシヌクレオシドをもって )行い、S - 3 - CGCATCCGを作り出した。室温において4時間濃NH<sub>4</sub>OHにより、ホスフェートと環外アミン保護基の除去の後、支持体を、水ジャケット付熱制御されたハイブリダイゼーション・チャンバー内に載せた。この支持体を、以下に述べるミスマッチ実験において使用し

10

20

30

40

50

た。

#### B. ハイブリダイゼーション

オリゴヌクレオチド・アレイは、ハイブリダイゼーション研究を含む、多種多様の用途において使用されることができる。ハイブリダイゼーション研究においては、そのアレイは、図1中に示すように、着目のレセプタ(R)に晒されることができる。このレセプタは、適当な標識(\*)、例えばフルオレセインにより標識されることができる。このレセプタが結合した支持体上の位置が測定され、そして、その位置におけるオリゴヌクレオチド・プローブの配列の知識を通じて、ある者は次に、そのレセプタがオリゴヌクレオチドである場合、そのレセプタ配列を決定することができる。

ハイブリダイゼーションにより配列決定(SBH)は、各部位におけるプローブの同一性が知られているアレイを形成するために、多くのプローブを表面に付着させることにより最も効率良く行われる。標識された標的DNA又はRNAは次にそのアレイにハイブリダイズされ、そしてそのハイブリダイゼーション・パターンが、そのアレイ内の全相補的プローブの同一性を決定するために調べられる。ミスマッチ・プローブ/標的複合体が重要でないことを教示する従来の技術とは反対に、本発明は、ミスマッチ・プローブ/標的複合体のハイブリダイゼーション・シグナルが、そのアレイ上の完全にマッチしたプローブ/標的複合体の同一性を同定し又は確認するような分析方法を提供する。

オリゴヌクレオチドのアレイは、光指定合成技術を用いたハイブリダイゼーション研究のために効率的に生成される。以下に討議するように、全テトラヌクレオチドのアレイを16サイクルにおいて作り出した。これは完成するのにたった4時間しか必要としなかった。組合せ戦略が使用されるので、アレイ上の異なる化合物の数は合成の間に指数関数的に増加する。一方、化学的カップリング・サイクルの数は、直線的に増加する。例えば、 $4^8$ (65,536)オリゴヌクレオチドの完全セットにこの合成を拡張することは、16の追加のサイクルが必要であるためにその合成にほんの4時間(又はこれ未満)を追加する。さらに、組合せ合成戦略を、いずれかの所望のプローブ組成を作り出すために行われることができる。例えば、ドデカマー( $4^{12}$ )の完全セットは、48以下の光分解とカップリング・サイクルにおいて作り出されるので( $b^n$ 化合物は多くとも  $b \times n$  サイクルを必要とする)。

、(より短いオリゴヌクレオチドのいずれかのサブセットを含む)ドデカマーのいずれかのサブセットは、48以下の化学的カップリング段階において構築することができる。アレイ内の化合物の数は、合成部位の密度と全体のアレイ・サイズによってのみ制限される。本発明は、片面上25ミクロンの正方形部位内で合成されたプローブをもつアレイにより行われている。この解像度において、65,536オクタヌクレオチドの完全セットは、たった $0.64\text{cm}^2$ の寸法のアレイ内に置かれることができる。1,048,576ドデカヌクレオチドのセットは、この個々のプローブ部位サイズにおいてたった $2.56\text{cm}^2$ のアレイを必要とする。ゲノム配列決定プロジェクトの成功は、効率的なDNA配列決定技術に依存する。最近の方法は、複雑な手順に高く頼り、そして実質的な手作業を必要とする。SBHは、最近の実施における手作業の多くを自動化するための潜在能力を提供する。光指定合成は、SBHのためだけでなく多くの他の用途のためにも微細化されたアレイの大規模生産のための効率的な手段を提供する。

オリゴヌクレオチド・アレイは、一次的な配列決定用途のために使用されることができるけれども、多くの診断方法は、標的核酸配列内のほんの数個のヌクレオチドの位置の分析を含む。1塩基の変更は、プローブ・アレイ上の標的のハイブリダイゼーション・パターン内に多数の変化を引き起こすので、本発明に係るオリゴヌクレオチド・アレイと方法は、ある者が、先に明白されたDNA配列の正確さをチェックし、又は標的核酸内での特定の特異的配列内の変化又は突然変異について走査することを可能にする。例えば、後者は、遺伝子、疾患、品質管理、及び法廷分析のために重要である。オクタヌクレオチド・プローブセットにより、標的核酸内の1塩基変化は、8つの完全ハイブリッドの損失により、そして8つの新たな完全ハイブリッドの生成により、検出されることができる。この1塩基変化は、アレイ上の変更されたミスマッチ・プローブ/標的複合体の形成を通じて検出されることもできる。おそらく、さらに驚ろくべきことに、複雑な核酸内のこのようない

10

20

30

40

50

塩基対の変化は、アレイ上の標的の全ハイブリダイゼーション・パターンを劇的に変更する。本発明に従って、この全ハイブリダイゼーション・パターンにおけるこのような変化は、この分析を実際に単純化するために使用される。

光指定オリゴヌクレオチド・アレイの高い情報含有量は、遺伝子診断テストにかなり有益である。100~1,000の異なる突然変異の配列比較は、1時間に1回の形式におけるものの代わって、同時に検定されることができる。アレイは、多種多様な病因生物の迅速な同定についての遺伝子マーカーを含み、そして、RNA/RNA, RNA/DNA、タンパク質/RNA又はタンパク質/DNA、相互作用の配列特異性を研究するために、構築されることもできる。ある者は、アンチセンス、トリプル・ヘリックス、又は他の用途のために非ワトソン・クリック・オリゴヌクレオチド及び新規の合成ヌクレオシド・アナロブを使用することができる。好適に保護されたRNAモノマーが、RNA合成のために使用されることができ、そして多種多様の合成及び非天然核酸アナロブが、実施者の動機付けに依存して、使用されることができる。例えば、引用により本明細書に取り込む、PCT特許公開第91/19813号、第92/05285号、及び第92/14843号を参照のこと。さらに、オリゴヌクレオチド検定は、オリゴヌクレオチド複合体の形成及び安定性を支配する熱力学と速度論の規則を演繹するために使用されることができる。

### 実施例

#### 1. 表面オリゴヌクレオチドへの標的のハイブリダイゼーション

先に討議した支持体結合オクタヌクレオチド・プローブを、15における15分間のインキュベーションによりハイブリダイゼーション・チャンバー内で5 GCGTAGGC - フルオレセインの標的にハイブリダイズさせた。アレイ表面を次に蛍光外(epifluorescence)顕微鏡(488nmアルゴン・イオン励起)に応答信号を送った(interrogated)。この走査の蛍光画像を図2に示す。この蛍光強度パターンは、プローブの合成を指定するために使用される800×1280μmストライプにマッチする。さらに、シグナル強度は、(ガラス支持体の背景を4倍上回って)高く、そのプローブへの標的的結合を立証する。

標的-プローブ複合体の挙動を、そのハイブリダイゼーション溶液の温度を上昇させることにより調べた。各温度において10分間の平衡化の後、支持体をシグナルについて走査した。2本鎖は、研究下配列について予想される温度レンジ内でメルトした( $T_m = [2^\circ (A + T) + 4^\circ (G + C)]$ の規則から得られた $T_m = 28^\circ$ )。アレイ内のプローブは、標的DNAの再ハイブリダイゼーションにより立証されるように、標的-プローブ複合体の温度変性に対して安定であった。

#### 2. 標的ハイブリダイゼーションの配列特異性

標的ハイブリダイゼーションの配列特異性を立証するために、2つの異なるプローブを800×1280μmのストライプ内で合成した。図3Aは、2つのプローブの位置を同定する。プローブS-3 - CGCATCCGをストライプ1, 3と5内で合成した。プローブS-3 - CGCTTCCGをストライプ2, 4と6内で合成した。図3Bは、15における支持体への5 - GCGTAGGC - フルオレセイン標的のハイブリダイゼーションの結果を示している。これらのプローブはたった1の内部塩基が異なるけれども、上記標的は、(ストライプ1, 3と5内の背景を~500カウント上回って)その相補的配列に特異的にハイブリダイズし、位置2, 4と6(~10カウント)内には検出可能なシグナルはほとんど又は全くない。図3Cは、両配列への標的のハイブリダイゼーションの結果を示す。図3中のすべての位置内のシグナルは、図3Bにおけるシグナルの非存在が1塩基ミスマッチの不安定性にのみ依ることを説明している。標的には、等モル濃度において存在するけれども、図3B中のストライプ2, 4と6内のシグナルの比は、領域1, 3と5内のシグナルよりも約1.6倍高い。この2本鎖は、領域2, 4と6を占める2本鎖よりもわずかに高い予想 $T_m$ をもつ。これらの2本鎖は、15分間で45までその温度を上昇させることにより解離され、そしてそのハイブリダイゼーションは、上記と逆の順番で繰り返され(図3Dと3E)、これは、逆方向におけるハイブリダイゼーションの特異性を立証している。

#### 3. プローブ・マトリックスの組合せ合成とそれへの核酸標的のハイブリダイゼーション

光指定合成においては、製品の位置と組成は、照射と化学的カップリング試薬の順番に依

10

20

30

40

50

存する（完全な説明については、Fodor et al., *Science* (1991) 251: 767-773を参照のこと。）図4において説明するように、256テトラヌクレオチドの合成について考える。マスク1は、合成の第1ラウンドにおける4つのヌクレオシドの第1とのカップリングのために支持体表面の1/4を活性化する。サイクル2において、マスク2は第2ヌクレオシドとのカップリングのためにその支持体の別の1/4を活性化する。この工程がモノヌクレオチドの4つの領域を作り上げるために繰り返される。ラウンド2のマスクは、ラウンド1のものに垂直であり、そしてラウンド2の各サイクルは、4つの新たなヌクレオチドを作り出す。この工程は、図4において説明するように16のヌクレオチドを形成するためにラウンド2まで繰り返される。ラウンド3のマスクは、各カップリングのサイクルが16のトリマーを作り出すように上記合成領域をさらに亜分割する。この支持体の亜分割は、テトラヌクレオチドを形成するためにラウンド4まで繰り返される。このプローブ・マトリックスの合成は、 $(A + C + G + T)^4$ の多項表記法において簡単に表現することができる。この多項式の展開は、256のテトラヌクレオチドを作り出す。

プローブ・マトリックスを生り出すために光指定組合せ合成により合成された256プローブのアレイの適用を、図5Aに説明する。この合成のための多項式は： $3 - CG (A + G + C + T)^4$ により与えられる。この合成マップを図5Bに与える。全ての可能性のあるテトラヌクレオチドを、3 - 末端と5 - 末端におけるCGに隣接するように合成した。

におけるこのアレイへの標的5 - GCGGCGGC - フルオレセインのハイブリダイゼーションは、最も強い位置（2,698カウント）としてS - 3 - CGCCGCG相補プローブを正確に作り出した。以下のミスマッチ：S - 3 - CGCAGCCG（554カウント）、S - 3 - CGCCG 20 ACG（317カウント）、S - 3 - CGCCG<sub>T</sub>CG（272カウント）、S - 3 - CGACGCCG（242カウント）、S - 3 - CGT<sub>T</sub>GCCG（203カウント）、S - 3 - CGCC<sub>C</sub>CCG（180カウント）、S - 3 - CGC<sub>T</sub>GCCG（163カウント）、S - 3 - CGCC<sub>A</sub>CCG（125カウント）、とS - 3 - CGC<sub>C</sub>CCG（78カウント）についても有意な強度が観察された。

### C. ミスマッチ分析

先に討議したアレイを、より短い長さkのプローブのアレイを使用して長さnのオリゴヌクレオチドの核酸配列を決定するための本発明に係る方法において使用することができる。図6は単純な例について説明する。この標的は、配列5 - XXYY - 3 {ここで、XとYは相補的核酸、例えばAとT又はCとGである。}をもつ。討議目的のために、図6における説明を、たった2つの塩基とひじょうに短い配列の使用により単純化するが、この技術は、例えば全部で4つのRNA又はDNAの塩基によるより大きな核酸に容易に拡張されることができる。

標的の配列は、一般に最初から(ab initio)知られていない。ある者は、より短いプローブのアレイを用いて本発明を使用して標的の配列を決定することができる。本実施例においては、全ての可能性のあるXとYの4-merのアレイを合成し、そして次に5mer標的の配列を決定するために使用する。

最初に、“コア（core）”プローブを同定する。このコア・プローブは、本発明に係るミスマッチ分析法を使用して標的内の配列に正確に相補的である。このコア・プローブは、以下の基準の中の1又は両方を使用して同定する：

1. コア・プローブは、他のプローブよりも強い標的への結合アフィニティー、典型的には、（先の分析サイクル内でコア・プローブとして同定されていない）アレイ内のいずれかのプローブの最高の結合アフィニティーを示す。

2. コア・プローブ配列に比較したとき、標的とミスマッチであるプローブは、以下により詳細に討議するような特徴的なパターンを示す。ここで、そのプローブの3 - と5 - 末端においてミスマッチするプローブは、内部の並置においてミスマッチするプローブよりも強く結合する。この特定の実施例においては、選定基準#1は、そのプローブが標的にハイブリダイズしたものとして示されている図6A中に示すように、配列3 - YYXYをもつ標的に最高の結合アフィニティーをもつコア4 - merプローブを同定する。（標的の5 - XXYY位に一致する）プローブ3 - YYXYは、それ故、“コア”プローブとして選ばれる。

選定基準 # 2 は、このコア・プローブが標的核酸に正確に相補的であることを確保するための“チェック (check)”として使用される。この第 2 の選定基準は、そのコア・プローブと比較するとき 1 塩基ミスマッチをもつプローブのハイブリダイゼーション・データ（例えば、支持体上のプローブのアレイにハイブリダイズした標識された標的の蛍光強度、但し、他の技術も当業者によく知られている。）を評価する。この特定のケースにおいては、コア・プローブは S - 3 - YYXY として選ばれた。このコア・プローブの 1 塩基ミスマッチ・プローブは：S - 3 - XYXY, S - 3 - YXXX, S - 3 - YYYY、と S - 3 - YYXY である。これらの 1 塩基ミスマッチの結合アフィニティー特性は、“正しい (correct)” コアが選ばれたことを確かめ、又は類似の結合アフィニティーを示すプローブのセットの中からそのコア・プローブを選ぶために使用される。

10

予想される結合アフィニティー対ミスマッチ位置の例示的、仮定的なプロットを図 6Bにおいて提供する。これらの結合アフィニティー値（典型的には、プローブにハイブリダイズされた標識された標的の蛍光強度、但し、アフィニティーに関する多くの他の要因を使用することもできる。）は、このグラフの左手の位置上に 1 の値としてプロットされた、標的への S - 3 - YYXY の結合アフィニティーに対して全て正規化される。たった 2 つの又クレオチドが本例において含意されるので、1 位（そのプローブの 3 - 末端における又クレオチド）においてミスマッチするプローブについてプロットされた値は、S - 3 - XYXY の正規化された結合アフィニティーである。2 位におけるミスマッチについてプロットされた値は、S - 3 - YXXX の正規化されたアフィニティーである。3 位におけるミスマッチについてプロットされた値は、S - 3 - YYYY の正規化されたアフィニティーであり、そして 4 位におけるミスマッチについてプロットされた値は、S - 3 - YYXX の正規化されたアフィニティーである。上記のように、“アフィニティー (affinity)” は、例えば、標的上の蛍光マーカーからの光子カウントの数を含む、多くの方法において測定されることができる。

20

全部で 3 つのミスマッチのアフィニティーは、この説明においてそのコアよりも低い。その上、アフィニティー・プロットは、そのプローブの 3 - 末端におけるミスマッチが、この特定のケースにおいてはプローブの 5 - 末端におけるミスマッチよりもより少ない影響力をもつことを示している。但し、これは、常にそうではない。さらに、このプローブの末端におけるミスマッチは、このプローブの中央におけるミスマッチよりも少ない妨害をもたらす。図 6B 中において示すようにプロットされたとき“スマイル (Smile)” 形状のグラフをもたらすこれらの特徴は、“正しい” コア・プローブの選定後、又は、標的配列の他の部分に関してのコア・プローブであるミスマッチ・プローブを考慮した後の、1 塩基ミスマッチのほとんどのプロットにおいて見られるであろう。この情報は、コア・プローブの最初の選定、又は正しくマッチしたコア・プローブが選ばれていることを確保するためのチェックのいずれかにおいて使用されることができるであろう。もちろん、特定の情報においては、先のセクション B 中に述べたように、例えば、法延の又は遺伝子の研究、等において必要とされるコアの同定が、全てである。

30

配列決定研究においては、この工程を次に、このコア・プローブの左及び / 又は右伸長のために繰り返す。図 6 において説明する例においては、このコア・プローブの右伸長だけが可能である。このコア・プローブの可能性のある 4 - mer 伸長プローブは 3 - YXYY と 3 - YYXY である。再び同一の選定基準を使用する。3 - YXYY と 3 - YYXY の間では、3 - YYXY が最も強い結合アフィニティーをもつであろうことが通常見い出されるであろうし、そしてこのプローブは、正しいプローブ伸長として選ばれる。この選定は、上記コア・プローブと比較するとき 1 塩基ミスマッチをもつプローブの正規化された結合アフィニティーを再びプロットすることにより確かめられることができる。推定プロットを図 6C に説明する。再び、特徴的な“スマイル” パターンが観察され、これは、“正しい” 伸長、すなわち 3 - YYXY が選ばれていることを示している。この情報から、ある者は、標的の配列が 5 - XXYY であると正しく結論付けるであろう。

40

#### 実施例

1. (A + T)<sup>8</sup> アレイと 1 塩基ミスマッチの安定性

50

20ステップの、4 - レプリカ組合せ合成を、Men Poc - dAとMen Poc - dTを使用して行った。リトグラフ・マークを、1セットの256オクタヌクレオチドの各メンバーが $1.28 \times 1.28\text{cm}$ アレイ上の4つの別々の位置内で合成され、1024の異なる合成部位であって、各々がオリゴヌクレオチド・プローブを含み、各々の部位がサイズ $400 \times 400\text{ }\mu\text{m}$ である部位を作り出すように選んだ。dAアミンの合成とフェノキシアセチル脱保護の後、支持体を、サーモスタッフで調節された染色及びフロー・セル内に置き、15において1nM 5 - AAAA AA - フルオロセインとインキュベートし、そして次にZeiss蛍光外顕微鏡内で走査した。得られた蛍光画像を図7に示す。

1塩基ミスマッチの関数としてのハイブリダイゼーション事件の蛍光強度を図8にグラフにより提供する。1塩基においてそのコア・プローブと異なる各オリゴヌクレオチド・プローブについての4つの独立した強度のそれぞれをプロットする。0位ミスマッチ（すなわち、完全な相補物3 - TTTTTTTT）は、~900カウントにおいてアレイ上に最も明るい位置であり；このアレイの背景シグナルは、約220カウントである。ミスマッチ1位（プローブの3 - 末端）は、~760カウントにおいて次に明るい。これに続く位置の“スマイル”又は“U”形曲線は、プローブ/標的複合体の各位置におけるミスマッチの相対的安定性を示している。この“ミスマッチ・ファミリー”は、プローブのアレイとの核酸の相互作用を特徴付け、そして標的配列の同定を提供し又は確かなものにする。3, 4, 5と6位におけるミスマッチは、より脱安定性であり、そして背景からは事実上区別することができない強度を作り出す。1位におけるミスマッチ（オクタヌクレオチドの3 - 末端が支持体につながれている点）は、8位における対応のミスマッチ（その遊離の5 - 末端）よりもより脱安定性されていない。アレイ合成の均一性と標的ハイブリダイゼーションは、上記の4つの2連の合成部位の間の強度の低い変化を反映している。

本発明に係る方法は、2以上のミスマッチをプローブとの標的ハイブリダイゼーションからの情報をも利用することができる。ミスマッチの対の関数としての蛍光強度を図9に表わす。このケースにおいては、強度データは、完全なマッチが強度1をもつように正規化された。例えば、インデックス1, 8におけるデータは、そのプローブ/標的2本鎖の各末端におけるミスマッチに対応する。2項（インデックス1, 1 ~ 8, 8）を、図8に説明した1ミスマッチに対応する。この最高強度は、プローブ/標的複合体の末端におけるミスマッチの1と対に対応する。

## 2. (G + T)<sup>8</sup>アレイと配列再構築

Men Poc - dGとMen Poc - dTのオクタヌクレオチド・アレイを合成した。合成の形式は、先に討議した(A + T)<sup>8</sup>アレイのためのものと同様であり、そして4の反復（全部で1024）におけるGとTの256オクタヌクレオチドをもたらす。最終脱保護と温度制御された(15)ハイブリダイゼーション・チャンバーへの接続後、プローブ・アレイを1nM 5 - AACCCAAACCC - フルオレセイン標的とインキュベートし、そして走査した。得られた画像を図10に与える。4つの明確であるが重複性の、完全に相補的なオクタヌクレオチド・ハイブリダイゼーションが予想される：3 - TTGGGTTT, TGGGTTTG, GGGTTTGG、とGGTTTGGG。本明細書中に示すように、1塩基対のミスマッチをもつプローブ/標的複合体の中程度の安定性は、中程度のシグナルをもつプローブのファミリーを作り出す。図10の多くの強度特性の粗略な検査は、複雑なパターンを現わした。

本発明により提供される再構築された実践的研究（heuristic）は、図10における複雑なデータ・パターンを使用する。このアルゴリズムは、完全にマッチしたプローブ/標的複合体がより高い蛍光強度をもち、そして完全なマッチと関連の1塩基ミスマッチが典型的には図6に示すものと類似の特性を形成するという一般的な規則として推定される。

最高の強度をもつプローブは、その標的への完全なマッチであるはずである。対応のミスマッチ曲線を図11A ~ 11Cに示す。ある者は、そのプローブが標的に正確に相補的であることを確認するために最高強度をもつプローブ（このケースにおいてはS - 3 - TGTTTG）についてのミスマッチ曲線を最初にプロットする。このプローブが標的の断片に相補的であると仮定して、我々は、その標的の3 - 末端上に1塩基の“伸長”を考える。この場合、2つのプローブの選択がある。2つの8 - merプローブS - 3 - GGGTTTGTとS

10

20

20

30

40

50

- 3 - GGGTTTGGの中の 1 は、標的核酸に正確に相補的であろう。これらの 2 つのプローブのそれについての並びにプローブ S - 3 - TGGGTTTGについてのミスマッチ曲線を図11A に強度値により示す。プローブ S - 3 - TGGGTTTGがプローブ S - 3 - TGGGTTG のもの（典型的な“スマイル”プロット）と最も近似したミスマッチ曲線をもつということに注目のこと。それ故、ある者は、正しい伸長プローブが S - 3 - GGGTTTGGであると結論付けるであろう。

図11B は、標的配列の 3 - 末端を評価するための本工程の反復を示す。プローブ S - 3 - GGGTTTGGはコア S - 3 - GGGTTTGGに最も近似したスマイル形のミスマッチ曲線をもち、そしてプローブ S - 3 - GGTTTGGTはそうではないので、ある者は、プローブ S - 3 - GGGTTTGGが正しい伸長プローブであると正しく結論付けるであろう。この工程は、曲線が正しい形をもつまでか、又は標的の“末端”に到達していることを示す。絶対強度が最高強度のものよりもかなり下回るまでのいずれかまで、繰り返されることができる。同様の方法は、5 - 末端に伸長する標的の配列を提供する。図11C は、完全にマッチしたプローブの全てのミスマッチ曲線を示しており；各曲線は、この標的について予測された無矛盾の形状をもつ。

上記技術は、もちろん、図12A ~ 12D の各パネルにおいて説明するように、いずれかの長さの核酸に容易に延長することができる。図12A において示すように、10 - mer 標的が配列決定されるであろうし、そしてその配列は、5 - N<sub>1</sub> N<sub>2</sub> N<sub>3</sub> N<sub>4</sub> N<sub>5</sub> N<sub>6</sub> N<sub>7</sub> N<sub>8</sub> N<sub>9</sub> N<sub>10</sub> - 3 {ここで、N はいずれかのヌクレオチド又は核酸モノマーであり、そして下付き添字はそのプローブ内のヌクレオチド位置を示し、3 はその 3 - 末端モノマーであることを示す。} により示される。当業者は、プローブが支持体に付着された 5 - 末端をもって合成された場合、本発明に係る方法が、適当な修飾をもって適用されることができるることを理解する。

より短いオリゴヌクレオチドのアレイを、本発明の 1 の態様に従ってより大きなヌクレオチドを配列決定するために使用することができる。図12A ~ 12D に示す特定の例においては、4 - mer (長さ 4 モノマーのオリゴヌクレオチド・プローブ) を未知の 10 - mer 標的を配列決定するために使用する。実際には、より長いプローブと標的が典型的には使用されるが、この例示的な例は、本発明の理解を容易にする。4 - mer アレイの 1 のメンバーは図12A に示され、そして配列 S - 3 - P<sub>3</sub> P<sub>4</sub> P<sub>5</sub> P<sub>6</sub> {ここで、各 P (プローブ) ヌクレオチドは、その用途に依存して A, T, C, U, G と他のモノマーからなる群から選ばれ、そしてその添字は、標的に対する位置を示す。} をもつ。討議目的のために、このハイブリダイゼーション・データは 1 のアレイから入手されると推定される。しかしながら、ある者は、本法を実施するために、多数のアレイ、異なる時に於いて合成されたアレイ、又はさらに個々のプローブを使用することができる。上述のように、4 の長さのプローブを討議を容易にするために選ばれ；実際には、より長いプローブが典型的には使用されるであろう。

S - 3 - P<sub>3</sub> P<sub>4</sub> P<sub>5</sub> P<sub>6</sub> は、標的への強い結合アフィニティーと正しいミスマッチ特性をそれが示すために、そのアレイからのコア・プローブとして選ばれる。全 4 - mer のアレイにおいて、配列 S - 3 - P<sub>3</sub> P<sub>4</sub> P<sub>5</sub> P<sub>6</sub> がそのコア配列として選ばれる。なぜなら、フルオレセイン標識された標的 (図12A 中 5 - N<sub>1</sub> N<sub>2</sub> N<sub>3</sub> N<sub>4</sub> N<sub>5</sub> N<sub>6</sub> N<sub>7</sub> N<sub>8</sub> N<sub>9</sub> N<sub>10</sub> - 3 として示す) が支持体に晒されるとき、その標的が図12A に示す矢印により示されるように、そのプローブにハイブリダイズし、そして高い蛍光強度 (すなわち、多数の光子カウント) がその支持体の他の部分と比較するとき、プローブ S - 3 - P<sub>3</sub> P<sub>4</sub> P<sub>5</sub> P<sub>6</sub> を含む支持体の部分内で観察されるからである。通常、最も強い結合アフィニティーを示す配列が第 1 コア配列として選ばれるであろう。

ある者は、好ましくは、最初に選ばれたコア配列が、1 塩基においてコア・プローブと異なるアレイ内のプローブの蛍光強度を調べることにより標的に完全に相補的であるかどうか確認する。図12B は、1 塩基ミスマッチの相対強度対 S - 3 - P<sub>3</sub> P<sub>4</sub> P<sub>5</sub> P<sub>6</sub> コア・プローブについてのミスマッチの位置の典型的なプロットを定量的に説明する。簡単な例として S - 3 - P<sub>3</sub> P<sub>4</sub> P<sub>5</sub> P<sub>6</sub> においてヌクレオチド C が存在しないと仮定する。図12B は

、配列 S - 3 - P<sub>3</sub> P<sub>4</sub> P<sub>5</sub> P<sub>6</sub> 内への C の置換によりそのコア配列プローブと異なるプローブの正規化された蛍光強度を定性的に説明しており、そして、ここで、 C - 含有ミスマッチ・プローブのいずれも標的内の他の配列に正確に相補的ではない。従って、図12Bは、それらが標的とハイブリダイズされるとき、そのコア・プローブに正規化された、以下のプローブ・セット

S - 3 - CP<sub>4</sub> P<sub>5</sub> P<sub>6</sub> ,  
 S - 3 - P<sub>3</sub> CP<sub>5</sub> P<sub>6</sub> ,  
 S - 3 - P<sub>3</sub> P<sub>4</sub> CP<sub>6</sub> 、及び  
 S - 3 - P<sub>3</sub> P<sub>4</sub> P<sub>5</sub> C

の相対的蛍光強度をプロットする。他の態様においては、平均曲線が、各位置における可能性のあるヌクレオチドの全ての置換（ミスマッチ・プローブの“ファミリー”）についてプロットされ、又は最も高い強度が各位置についてプロットされる。従って、図12Bにおけるグラフの X 軸上の 0 位は、置換がないことを表し、そしてコア・プローブ S - 3 - P<sub>3</sub> P<sub>4</sub> P<sub>5</sub> P<sub>6</sub> への標的ハイブリダイゼーションのための蛍光強度を示す。図12B 中の全ての値がこの値に関して正規化されているので、“置換無し”のケースは、1 の正規化強度をもつ。C が 3 , 4 , 5 と 6 位において置換されるとき、その相対強度値は、通常、小さい。なぜなら、これらの配列のいずれも本例においては標的に正確に相補的でないからである。

3 - 又は 5 - 末端においてミスマッチをもつプローブ / 標的複合体の相対蛍光強度が、プローブ / 標的複合体の中央においてミスマッチをもつ複合体よりも典型的には高い。なぜなら、このプローブの末端におけるミスマッチは、プローブ / 標的複合体の中央におけるミスマッチよりも少なく脱安定化される傾向をもつからである。このプローブの 3 - 末端にミスマッチをもつプローブ / 標的複合体は、推定するに、この態様においては支持体表面へのそのプローブの 3 - 末端の近接のために、そのプローブの 5 - 末端におけるミスマッチをもつものよりも少なく影響を及ぼす（そしてそれ故、より高い蛍光強度をもつ）ことができる。それ故、結合アフィニティーに関する正規化係数対ミスマッチ位置をプロットする曲線は、図12B 中に示すような“ゆがんだスマイル（crooked smile）”の形をもつ傾向がある。

この方法論を使用して、ある者は、そのコア・プローブと同一の配列であるが一端において伸長しており、そして場合により他端においてより短いものをもつアレイ上のプローブを調べることによりそのコア配列を延長することができる。これらのプローブは、標的に完全にハイブリダイズする。プローブを決定するための第 2 のコア・配列として評価されることができる。この工程の繰返しにより、ある者は、その標的のヌクレオチド配列の全体を決定することができる。

本法を説明するために、図12C は、コア・プローブ S - 3 - P<sub>3</sub> P<sub>4</sub> P<sub>5</sub> P<sub>6</sub> の 4 つの可能性のある 4 - 員“左伸長”を示す。図示するように、S - 3 - P<sub>3</sub> P<sub>4</sub> P<sub>5</sub> P<sub>6</sub> に相補的な標的の配列に隣接するヌクレオチドは A , T , C 、又は G のいずれかであり、又はその標的上に隣接するヌクレオチドは全く存在しない（すなわち、P<sub>3</sub> は標的の 5 - 末端である。）。それ故、P<sub>3</sub> P<sub>4</sub> P<sub>5</sub> P<sub>6</sub> コア・プローブの可能性のある左伸長は、プローブ S - 3 - AP<sub>3</sub> P<sub>4</sub> P<sub>5</sub> , S - 3 - TP<sub>3</sub> P<sub>4</sub> P<sub>5</sub> , S - 3 - CP<sub>3</sub> P<sub>4</sub> P<sub>5</sub> 、と S - 3 - GP<sub>3</sub> P<sub>4</sub> P<sub>5</sub> である。この説明の目的のために、T が実際に“正しい”ものであると推定される。なぜなら A は、その標的核酸の相補位置内にあるからである。

図12D における上左のプロットは、S - 3 - AP<sub>3</sub> P<sub>4</sub> P<sub>5</sub> プローブのミスマッチ特性について予想されるハイブリダイゼーション・データを説明する。全てのデータが S - 3 - AP<sub>3</sub> P<sub>4</sub> P<sub>5</sub> に対して正規化されている。2 - 5 位の各々における置換の全てについてのデータ点を示すが、各々の位置における 3 つの値についての平均データも使用することができるであろうし、各の位置における 1 の置換を使用することができ、3 つの値の最高のものを利用してもよく、又はいくつかの他の組合せでもよい。S - 3 - AP<sub>3</sub> P<sub>4</sub> P<sub>5</sub> グラフにおいて示されるように、1 の点は残りのものよりもかなり高い結合アフィニティーを示す。これは、2 位における A についての T 置換である。AP<sub>3</sub> P<sub>4</sub> P<sub>5</sub> グラフ内の残りのデ

10

20

30

40

50

ータは、図12B中に示す通常の“スマイル”特徴をもつ。同様のプロットが図12Bの底部分において示されたCとG置換について顕出された。それぞれのケースにおいては、全データ点は、そのグラフ内の推定“コア”プローブに対して正規化されている。

図12Dの上右部分において示されるT伸長グラフは、3 - AP<sub>3</sub>P<sub>4</sub>P<sub>5</sub>グラフその他のような異常曲線をもたないであろう。なぜなら、3 - TP<sub>3</sub>P<sub>4</sub>P<sub>5</sub>プローブの2位におけるモノ置換のいずれも標的に正確に相補的であるであろうからである。従って、2位におけるA, C, とGの置換は全て、標的に対して1塩基ミスマッチをもつプローブについて予想される特徴的な“スマイル”プロットを作り出す。さらに、T置換プローブノ標的複合体の蛍光強度は、C, G, とAプローブノ標的複合体の蛍光強度よりも通常高いであろう。これらのデータは、各種組合せにおいてどの伸長が“正しい”かどうかを、そしてそれによりその標的核酸の配列を決定するために、使用されることができる。

図12A～12D中に示されたデータから、ある者は、コア・プローブに相補的な配列に対してその標的の左伸長物に正確に相補的なプローブが、その標的内の2位においてAモノマーをもつと結論付ける。

この工程は、グラフのいずれも適当な特徴をもたなくなるまで繰り返され、この時に、標的の末端に達している。同様に、右伸長が、その標的の末端（又は着目の配列の末端）に達するまで評価される。

上記技術は、明らかに、ハイブリダイゼーション・データのマニュアル観察を通して行うことができる。しかしながら、好ましい態様においては、データは、1以上の適切にプログラムされたデジタル・コンピューターを使用して分析される。例示のシステムを、図13中に説明する。図中に示すように、このシステムは、CPUの制御下でオペレートされ、そしてメモリー（304）、例えばハード・ディスク、とメモリー（306）、例えばダイナミック・ランダム・アクセス・メモリー（dynamic random access memory）を含む1又は複数のコンピューターを含む。このコンピューターは、支持体（2）の部分にカップリングされた標識された標的ヌクレオチドからの蛍光強度又は他の関連情報を測定するスキヤニング装置（308）を制御するために使用される。

この支持体（2）は、その上に既知の位置において既知配列のプローブ・ヌクレオチドを含む。ユーザーは、インプット装置（313）を介してインプットを提供する。

蛍光強度又は他の関連情報は、メモリー（304/306）内に保存される。CPU（310）は、その蛍光データを処理して、プリント装置（312）又はディスプレイ（314）の1又は両方にアウトプットを提供する。本明細書中に記載する方法に従ってデータを処理し、そして先に示したようなグラフの形態における、又は核酸モノマーの配列における、又は単純な（+）/（-）アウトプットにおけるアウトプット又はこのようなデータの分析の他の結果を得ることができる。好適なコンピューターは、例えば、IBM PC又は互換性SPARCワーカステーション、又は同様の装置を含む。

図14は、n - mersのアレイを評価するための、そして（ミスマッチ分析のための）正確に相補的な配列又は（配列決定又は他の目的のための）より大きなk - merを同定するための、典型的なコンピューター・プログラムのためのフローチャートである。図中に示すように、システムは最初に、例えば、プローブのいくつかの特定のセットの最も高い結合アフィニティーをもつプローブを選ぶことによりステップ（402）においてコア・プローブを同定する。本法は、しばしば、反復工程においてオペレーション可能であろう。ここでアレイ内の最も高いアフィニティーのプローブは、第1反復（iteration）後に選ばれず、そして他の場合においては、例えば、1, 2, 3、又はそれ以上の最高結合プローブを選び、そして各々について左と右の伸長を行い、そして次にこの情報を保存し、そして最終的なアウトプットを提供する前に他のデータとこの情報を比較することが価値のあることである。これらの結果は、正しい配列の確認において援助することができる。

ステップ（404）において、このシステムは、コアn - merの左伸長の全てを同定する。ステップ（406）において、本システムは：

- どの左伸長物がプリセットされたモノマー置換パターンと最も無矛盾である挙動を示すかを決定し、そしてノ又は

10

20

30

40

50

- その最高の結合アフィニティーを示す左伸長を選択する、  
の 1 又は両方により、適切な左伸長選定する。

上記選択基準その他は、いくつかの態様において、AND法において使用されることがある。すなわち、これらの基準の両方が適合し、又はこのシステムは、1が末端モノマーに達したか又はそのシステムが許容可能な状態で行われないかのいずれかを仮定する。他の態様においては、上記基準の中の 1 が、主な選択機構として選ばれることができ、そして他のものが、注意、すなわち、可能性のある誤った選定、又は他の選定をユーザーに提供するために使用されることができる。

その後、本システムは、選定基準がステップ 408においていくつかの最小標準に適合したかどうかを決定する。そうでない場合、本システムは、その配列の末端がステップ (410) に達したと仮定する。この選定基準が適合する場合、先のコアからの正しい伸長物として選ばれた新たな“コア”によりステップ (404) の始めに戻される。

その後、本工程は、右伸長について有効に繰り返される。ステップ (412) において右伸長が同定される。ステップ (414) において、プリセットされたミスマッチ特性のプローブ及び / 又は高アフィニティー右伸長物が同定される。ステップ (416) において、本システムは、この分子の末端に達しているかどうかを決定する。そうでない場合、この工程はステップ (412) に戻される。そうである場合、その分子は配列決定されていると仮定し、そしてこの工程は、プリンター又は他のアウトプット装置への適当なアウトプットをもって終了される。

ミスマッチ分析の使用方法の他の態様を、図 15a 中に例示する。この図は、未知の標的断片の配列を決定するためにデザインされた典型的なコンピューター・プログラムのためのフローチャートを表す。このような方法は通常、適当にプログラムされたデジタル・コンピューター、例えば IBM PC 又は同等なもの、Sun ワークステーション、あるいは他の類似のコンピューター・システムにおいて行われるであろう。ステップ (502) に示すように、ハイブリダイゼーション・データ（例えば、上記のような蛍光強度、但し当業者によく知られた他の技術を通じて得られた他のデータも使用することができる。）をこのシステムに入れる。いくつかの態様においては、このデータは実験システムから直接的にインプットすることができ、一方、その他においては、これらのデータは、別個のシステムにおいて採取されることができる。ステップ (504) において、このシステムは、このデータ・セット内の“最良”データに対応するプローブ配列を選択する。蛍光強度データを使用する態様においては、この最良データは、最高強度レベルを示すようなプローブに典型的には対応するであろう。このデータ・セットの不履行パーセンテージが使用されることが可能、又は場合により、オペレーターは、そのデータ・セットのサイズ、そのシステムの計算能力、及びそのデータ・セットのサイズのような要因に基づいて選ばれたパーセンテージにより、最強強度データの特定パーセンテージを選ぶためにそのシステムに指令する。ステップ (506) においては、このシステムは、ステップ (504) において選ばれたプローブ配列の組合せを通じて誘導することができる特定長さの全断片のセットを同定する。このように同定された断片を“候補”断片という。

ステップ (508) においては、このシステムは、核酸プローブから測定されたハイブリダイゼーション・データを使用して、上記セット内の各候補断片についての完全にマッチしたプローブについて先に記載されたようなミスマッチ分析アフィニティー・プロットを構築する。ステップ (510) においては、各アフィニティー・プロットは、それが、その標的DNAの他の部分とのそのプローブの交差ハイブリダイゼーションについて考慮した後アフィニティー・プロットの予測された形状にどのくらいよく一致するかを反映するよう格付けされる。

この格付けステップ (510) が一旦終了すると、このプログラムは、各候補断片（ステップ 512）についてのすべてのアフィニティー・プロットからの等級を合計し、そしてステップ (514) において次の候補断片を検査するために前進する。ステップ (516) において、全候補断片が検査されたとき、このシステムは、個々の候補断片についての合計されたアフィニティー・プロット等級を比較する。最後に、ステップ (518) において、このシ 50

ステムは、それについて、その等級がその鑄型との最も近い全体的なマッチを示すところの候補断片（単数又は複数）を選択する。この候補断片の配列は、未知の標的断片の配列に対応する。

図15bは、アフィニティー・プロットを格付けするための1の好ましい態様についてのフローチャートである。最初に、予想された結合アフィニティー対ミスマッチ位置（“鑄型”）の所定プロットをシステムにエンターする（ステップ602）。典型的には、この鑄型は上記のような“スマイル”形を示すであろう。ステップ（604）においては、このシステムは、ステップ606～620において特定されるようにその鑄型と各アフィニティー・プロットを比較する。ステップ（606）と（608）においては、このシステムは、その鑄型内の対応ミスマッチ位置の値と各ミスマッチ位置についてのアフィニティー値を比較する。ステップ（610）においては、このアフィニティー値を、最初に、測定された値が鑄型の値の選ばれた間隔内にあるかどうかを決定するために検査される。いくつかの態様においては、この間隔は、不履行レベルにセットされ；その他においては、それは、その実験誤差の予想される大きさのような要因に基づいて、オペレーターにより選ばれる。

所定のミスマッチ位置における値がその間隔内にある場合、ステップ（612）において、このシステムは、完全候補断片とミスマッチを含む塩基配列（“ミスマッチ配列”）を比較する。ステップ（614）において、このプロセスは、そのミスマッチ配列と上記候補断片のいずれかの部分との間の完全なマッチが存在するかどうかを決定する。このようなマッチが見つかった場合、このシステムは、そのプロットのその等級付けからそのミスマッチ位置を除外し（ステップ616）、そしてステップ（620）に続く；その他の場合、その位置は、その後の計算における包含のためにマークされる（618）。ステップ（620）において、この処理は、各ミスマッチ位置が検査されるまで続けられる。鑄型値と実験値のフィットは次に、ステップ（618）における包含のためにマークされた各ミスマッチ位置についてのステップ（622）の計算において決定される。いくつかの態様においては、この計算は、その間隔内又はその間隔の外側のいずれかとして各点を等級付けすることを含むことができ、その他の態様は、それによりそれぞれの含まれた位置がその予想された値からの逸脱を反映するように等級付けされるような方法であって、根 - 平均 - 二乗（Root - Mean - Squares）分析としてよく知られるような技術を含むものを利用する。最後に、このシステムは、ステップ（624）において各アフィニティー・プロットについて全等級を計算する。

#### 実施例

8 - merプローブのアレイを、Pirrung *et al.* に付与された米国特許第5,143,854号と米国出願逐次番号第07/805,727号（両方を全目的をもって引用により本明細書中に取り込む）中に記載された方法に従って構築した。このアレイを16 - mer標的とインキュベートし、そして公知の方法を使用して蛍光強度について走査した（例えば、引用により全目的をもって本明細書中に取り込む、米国出願逐次番号第08/195,889号を参照のこと。）。得られた蛍光強度データを、これまでの方法に従って分析した。アフィニティー・プロットを作り、そして等級付けした後、69 - 71のレンジにある等級をもつ4つの候補断片：5 - AG TTGTAGTGGATGGT, TGTTGTAGTGGATGGT, GGTTGTAGTGGATGGT、とCGTTGTAGTGGATGGTを同定した。この4つの類似の断片は、その断片の5末端における塩基の同一性においてのみ異なり、それは、他の位置の受ける2本鎖分析の利益を受けない。他の候補断片についての等級からのこれらの4つの候補についての等級の分離は実質的であり；次に高い等級の候補が54の等級を受け取った。

さらに他の本発明の態様においては、先に討議したアレイを、以下の方法により未知の標的の断片の配列を同定するために使用することができる。

長さ“n”的プローブからのハイブリダイゼーション・データのセット（ここで、“n”は、プローブ内の塩基の数に等しく、そして未知の標的断片の長さ末端である。）を、適当にプログラムされたコンピューター・システムにエンターする。本明細書中に企図されたハイブリダイゼーション・データ・セットは、先に記載されたような蛍光強度データのセット、並びに当業者に知られた他の技術を通じて得られたデータのセットを含む。この

10

20

30

40

50

システムは、(Harary, F., Graph Theory, Addison-Wesley, Reading, MA (1969) 及び Ahuja, R.K., Magnanti, T.L., Orlin, J.B., Network Flows, Theory, Algorithms, and Applications, Prentice Hall, New York (1993)、(両方を全目的をもって本明細書中に取り込む。) 中に記載されたような) 指定グラフを構築した。ここで、頂点 (vertices) (又は“節点 (nodes)”) は、上記データ・セットの全 ( $n - 1$ ) - mers に対応する塩基配列であり、そしてその端は、そのセットの全  $n$  - mers に対応する塩基配列である。各端は、その端が最初に含まれた ( $n - 1$ ) - mer から終りに含まれた ( $n - 1$ ) - mer まで接続するよう構築される。標的  $k$  - mer が ACTGTTG として知られており、そして  $n$  - mers が 3 - mers である上記のような指定グラフのサンプルは、図16として包含される。図16においては、3 - mer 配列に対応する端だけが含まれ；その標的塩基配列が知られていない典型的な場合においては、このシステムは、その分析において  $n$  - mer 節点に接続する全ての可能性のある端を含むであろう。

さらに、このシステムは、( $n - 1$ ) - mer 節点のいずれとも異なる、指定グラフ内のソース節点、並びにそのソース節点からそのグラフ内の ( $n - 1$ ) - mer 節点のそれまでの端を作り出す。最終的にこのシステムは、先に作り出された節点のいずれとも異なる、グラフ内の沈み節点 (sink nodes)、並びにそのグラフの ( $n - 1$ ) - mer 節点の各々からその沈み節点までの端を作り出す。

本システムは次に上記グラフの端のそれぞれに値を“コスト (cost)”を指定する。出発節点からの端先導 (edges leading) と沈み節点までの端先導は全てゼロのコストと指定される。他の全ての端は、その端に対応する塩基配列が標的  $k$  - mer 内のセグメントの完全マッチである確率の関数としてのコストを指定される。この確率を、“完全マッチの確率 (probability of a perfect match)”という。好ましい態様においては、各  $n$  - mer 端に指定されたコストは、その端に対応する配列が完全にマッチしない確率により除された、その端に対応する配列がその標的断片内のセグメントの完全マッチである確率の自然対数の負に等しい。いくつかの態様においては、これらの端は、別々の、追加の値“キャパシティー (capacity)”を指定することができ；このような態様については、出発節点から他の節の各々までの端先導、並びに沈み節点までの端先導は、全て 1 のキャパシティーを指定され；このグラフの他の全ての端は、無限キャパシティーを指定される。

出発節点から始まり、そしてその指定端に沿って続き、多くの経路がそのグラフを通じてトレースされることができ、沈み節点に終わり、それが、標的断片の配列におそらく対応することができる。このグラフの各々端がコストを指定されているので、これらの経路の各々についての全コストを決定することができる。この方法の最終段階においては、このシステムは、出発節点に始まり、そして沈み節点において終り、グラフを通してのそれぞれの可能性のある経路についてのこの全コストを決定するためにこのグラフにアルゴリズムを適用する。この段階における使用についてもくろまれるアルゴリズムは、当業者によく知られた最小コスト / 最大フロー・プログラミング・アルゴリズム、例えば、引用により本目的をもって本明細書中に取り込まれる Ahuja *et al.*, Network Flows, Theory, Algorithms, and Applications, Prentice Hall, New York (1993) 中に記載されるものを含む。

このようなアルゴリズムを使用して、このシステムは、それについて完全マッチの最高の組合せの確率が存在する合計コストをもつ経路を選び出す。例えば、最小コスト / 最大フロー・アルゴリズムを使用する態様においては、完全マッチの最大確率をもつ経路は、最小コストをもつ経路に対応する。それ故、この経路に関連する塩基の配列は、その標的についての最も類似する配列に対応する。

好ましい態様においては、個々のプローブについての完全マッチの確率は、以下のように蛍光ハイブリダイゼーション・データから決定される。選ばれたアレイの各々のプローブについて、一連の関連プローブについてのハイブリダイゼーション・データがこのシステムにエンターされ；着目のプローブに関連するプローブの強度のパターンを次に、上記のミスマッチ分析を使用して公知の標的配列とのマッチについて分析する。この段階により考慮される関連プローブのセットは、典型的には、利用可能なデータ・セットのサイズと計算キャパシティーのような要因に基づいてオペレーターにより選ばれるような、1 以上

10

20

30

40

50

の塩基のミスマッチをもつようなプローブを含む。このセットに含まれるデータは、典型的には、そのプローブの強度と塩基組成を含み、そして場合により、そのプローブがその標的内のどこかと完全マッチになることを引き起こすであろう塩基置換に関するデータ（“クロストーク（crosstalk）”）を含むであろう。好ましくは、これらのデータは、標的として既知のDNA配列を使用した大シリーズの実験のために集められるであろうし；異なる実験からの投入データを次に神経ネットワーク（Neural Network）、又は（例えば、Rich,E.,Knight,K.Artificial Intelligence, 2nd, McGraw-Hill, Inc., New York (1991)、これを全目的をもって本明細書中に取り込む、中に記載されているような）他によく知られた学習に基づく分類法を通じて分析して、プローブが強度、塩基組成、及び場合によりクロストークの関数として完全マッチであるかどうかを予測するためのアルゴリズムを開発することができる。このような予測の信頼レベルは、このアルゴリズム内で推定されることができ；この信頼レベルは、コスト値を上記指定グラフの端に指定することにおいて使用される完全マッチの確率に対応する。

#### D. 適用

本明細書中に記載する技術は、特に、標的核酸が特定のヌクレオチド配列又は既知配列と異なるいくつかの他の配列をもつかどうかを決定することが望まれる場合に、広いレンジの適用をもつであろう。例えば、本明細書中の本発明の1の適用は、突然変異検出に見い出される。これらの技術は、診断、法延、生体分析論、その他を含む多種多様な分野に適用されることができる。

例えば、“野生型”の核酸が配列 5 - N<sub>1</sub>N<sub>2</sub>N<sub>3</sub>N<sub>4</sub> {ここで、再びNは、モノマー、例えば核酸中のヌクレオチドを表し、そして添字は位置番号を表す。} をもつと仮定する。標的核酸が、それが 5 - N<sub>1</sub>N<sub>2</sub>N<sub>3</sub>N<sub>4</sub> と同一であるかどうか又はそれがこの配列と異なり、そしてそれが突然変異又は突然変異体配列を含むかどうかを決定するように評価されると仮定する。この標的核酸は、先に討議したように、典型的にはより短いプローブのアレイに最初に晒される。その後、1以上の“コア”配列が同定され、その各々が、その標的が突然変異配列又は突然変異を含まない場合、その標的に高い結合アフィニティーをもつと予想されるであろう。この特定の例においては、高い結合アフィニティーを示すと予想されるであろう1のプローブは、3-merアレイが使用されると仮定して、5 - N<sub>1</sub>N<sub>2</sub>N<sub>3</sub> (3 - P<sub>1</sub>P<sub>2</sub>P<sub>3</sub>) に相補的であるであろう。再び、このプローブ及び/又は標的がより長い核酸分子の一部であることができると認められるであろう。

最初のスクリーリング・ツールとして、3 - P<sub>1</sub>P<sub>2</sub>P<sub>3</sub> プローブへの標的の絶対結合アフィニティーがその標的の最初の3つの位置が予想された配列をもつかどうかを決定するために使用されるであろう。5 - N<sub>1</sub>N<sub>2</sub>N<sub>3</sub>への相補性がその標的への強い結合を示さない場合、その標的が野生型をもたないということが適切に結論付けられることができること。

1塩基ミスマッチ特性も、標的が突然変異体又は野生型配列を含むかどうかを決定するために本発明に従って使用されることがある。図17Aと17Bは、野生型（図17A）と突然変異体（図17B）である標的から生じる典型的な例示プロットを説明する。図示するように、野生型標的のための1塩基ミスマッチは、一般的には、典型的な、スマイル形のプロットに従う。これに反し、標的が特定の位置において突然変異をもつとき、特定コア・プローブへの標的の絶対結合アフィニティーが小さいであろうばかりでなく、その1塩基ミスマッチ特定が予想される挙動から逸脱するであろう。

本発明の1の態様に従って、核酸の選択された群をもつ支持体（他には、核酸の“ライブラリー”と本明細書中にいう。）を、特定の核酸が野生型又は他の予想された核酸と同一であるか又は異なるかの決定において使用される。核酸のライブラリーは、通常、プローブのアレイ又は“プローブ・アレイ”として提供されるであろう。このようなプローブ・アレイは、好ましくは、プローブの同一性がその支持体上のその位置によって決定されるような单一支持体上で形成される。場合により、このような支持体は、その標的のヌクレオチドが野生型と同一であるかどうかを決定するだけではなく、その標的に関する配列情報をも提供するであろう。このような支持体は、上述の分野、例えば、法延、診断

10

20

30

40

50

、その他における用途を見い出されるであろう。単に特定の例により、本発明は、鎌状赤血球貧血に関連する診断、多数のP-53突然変異のいずれかの検出、多数の囊胞性線維症突然変異のいずれかのために、高く多形のHLAクラス1又はクラス2遺伝子（特にクラス2DP, DQとDRベータ遺伝子）に関連するいずれかの特定変異体配列のために、並びに遺伝子疾患、遺伝子素因、及び遺伝子評価に関連する多くの他の配列のために、使用されることができる。

支持体がこのような用途において使用されるとき、その支持体上に特定の長さの可能性のある核酸の全てを提供することは必要でない。実際、本発明を使用して可能性のある配列の全ての比較的小さなサブセットだけを提供することが必要であろう。例えば、標的核酸が特に着目の5-塩基配列を含んで成り、そしてある者がその5-塩基配列内の1置換を検出するために使用されることができる支持体を開発することを欲すると想定する。本発明の1の態様に従って、支持体を、5塩基配列の1塩基ミスマッチ・プローブの全て又はほとんどと一緒に、その表面上に形成された予想5塩基配列により形成されるであろう。従って、支持体上に全ての可能性のある5塩基配列を含むことは必要でないであろう。但し、より大きなアレイがしばしば好ましいであろう。典型的には、本発明に従う支持体上の核酸プローブの長さは、約5と100塩基の間、約5と50塩基の間、約8と30塩基の間、又は約8と15塩基の間であろう。

特定の長さの全ての可能性のあるプローブの中での1塩基ミスマッチ・プローブの選択により、支持体上のプローブの数がかなり制限されることができる。例えば、3塩基配列においては、69の可能性のあるDNA塩基配列が存在するが、予想された配列に正確に相補的なたった1のものと9つの可能性のある1塩基ミスマッチ・プローブが存在するであろう。これらのプローブだけを選ぶことにより、スクリーニングに必要な多様性が減少されるであろう。好ましくは、必ずしもそうではないが、このような1塩基ミスマッチ・プローブの全てが1の支持体上で合成される。支持体はその1塩基ミスマッチに加えて着目の他のプローブを含むようにしばしば形成されるけれども、このような支持体は通常、n-塩基の可能性のあるプローブの全ての50%未満、しばしば、n-塩基の可能性のあるプローブの全ての20%未満、n-塩基の可能性のあるプローブの10%未満、そしてしばしばn-塩基の可能性のあるプローブの5%未満をもつであろう。

核酸プローブは、しばしば、特定の遺伝子配列の分析のためのキットにおいて提供されるであろう。1の態様に従えば、このキットは、着目の標的核酸に相補的なプローブを含むであろう。さらに、キットは、標的の1塩基ミスマッチを含むであろう。このキットは通常、このようなプローブの1以上のC, G, T, A及び/又はUの1塩基ミスマッチを含むであろう。このようなキットは、しばしば、本明細書中の教示に従って特定の核酸サンプルの配列の決定において、相補的プローブと1塩基ミスマッチの使用のための適当な指示を提供されるであろう。本発明の1の態様に従って、本キットは、1塩基ミスマッチだけと共に、その標的に対する相補的を提供する。このようなキットは、しばしば、それが特定の遺伝子の特徴を示すかどうかを決定するための遺伝子材料の適当なサンプルの評価において使用されるであろう。このようなキットは、鎌状赤血球貧血の検出、多数のP-53突然変異の中のいずれかの検出、多数の囊胞性線維症突然変異の検出、高く多形のHLAクラス1又はクラス2遺伝子（特にクラス2DP, DQ及びDRベータ遺伝子）に関連した特定の変異体配列の検出、並びに遺伝子疾患、遺伝子素因、及び遺伝子評価に関連した他の多くの配列の検出における、上述のようなサンプルの評価において使用されることがある。

従って、本発明に従って選ばれるプローブをもつ支持体は、多くの突然変異の検出並びに他の働きを行うことができるであろうが、このような働きを行うために限定された数のプローブだけを必要とするであろう。

#### 実施例

1. (G + T)<sup>8</sup>アレイと示差配列決定 (Differential Sequencing)

(G + T)<sup>8</sup>アレイを調製し、そして(5 - AACCCAAA CCCと比べたとき突然変異体配列を表す) 1 nM 5 - AACCCAACCCC - フルオレセインと共にインキュベートし、そしてそ

10

20

20

30

40

50

の配列が“野生型”又は“突然変異体”であるかどうかをテストするために走査した。得られた画像を図18に与える。4つの重複した、正確に相補性のオクタヌクレオチド・プローブ/標的ハイブリダイゼーションが、ある者が、その標的がプローブ；S-3-TTGG GTTG, TGGGTTGG, GGGTTGGG、とGGTTGGGGにより5-AACCCAAACCCでなければならないと仮定する場合に予想される。これらの結果は、1塩基変更の効果が、特に、アレイ上に形成する異なるミスマッチ・プローブ/標的複合体の数と同一性においてきわめて劇的であるということを立証する。図18においてシグナルを作り出す標的核酸が5-AACCCAAACCC C(すなわち、野生型)であるとある者が仮定する場合、相補的プローブS-3-TTGGG TTTのためのミスマッチ特性を図19Aに示す。このミスマッチ特性は、予想された形状をもたず、そしてプローブ/標的複合体は低い蛍光強度をもつ。8位におけるミスマッチに10 対応する強いピークは、この標的内のこの位置内の“正しい”塩基がおそらくAであるということを示している。なぜなら、AとCだけがこの実験において標的内にあるからである。6位のミスマッチも小さなピークを示す。これに反し、コアとしてS-3-TTGGGT TGプローブ配列を用いた同様のプロットは、“スマイル”形と高い蛍光強度を作り出した。図19Bにおいて、次の8-merプローブについての同一の特性を示す。このピークは、左に1位ミストしており、再び、この配列が標的内の8位において野生型から変化していることを確かなものとする。これらは、その元の11-merの標的断片内の同一の位置に対応していた。これらのデータは、野生型に比べたとき、標的の8位において1塩基の変化が存在することを予測させる。

仮定断片5-AACCCAAACCCに対応するミスマッチ・プローブの特性を図19Cに示す。ある者は、上記配列から下位に“動く(moving)”突然変異体の位置を観察する。最終的に、図19Dにおいて、このミスマッチ・プロットは、予想されたスマイル特徴を伴って、5-30 AACCCAAACCCに相補する4つのプローブに対応することが示される。

#### E. 結論

本発明は、ヌクレオチド配列及び核酸の他の分子との相互作用の研究のための改良方法と装置を提供する。これまでの記載は例示的であり、そして制限的なものではない。本発明の多くの変更は、本開示のレビューに基づき当業者に明らかとなるであろう。単に、実施例によって、本明細書中に記載した本発明の特定のものが、他のポリマー、例えばペプチドとタンパク質への適用をもつであろうし、そして他の合成技術を使用することができる。それ故、本発明の範囲は、これまでの説明を参照して決められるべきではないが、その代わりに、それらの十分な均等の範囲と共に添付クレームを参照して決定されなければならない。

## 【図1】

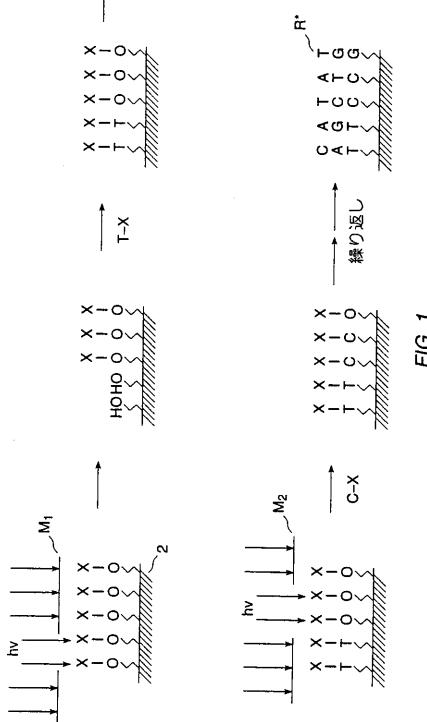


FIG. 1

【 図 2 】

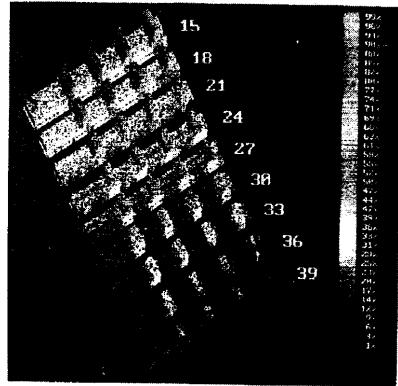


FIG. 2

【 図 3 】

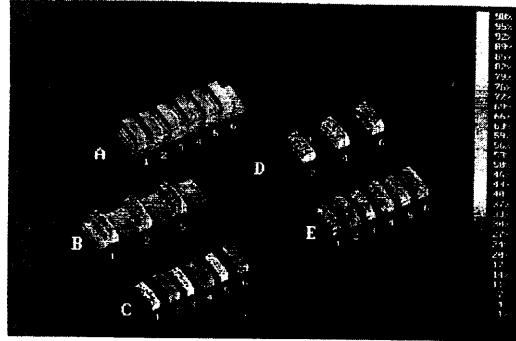


FIG. 3

【 図 4 】

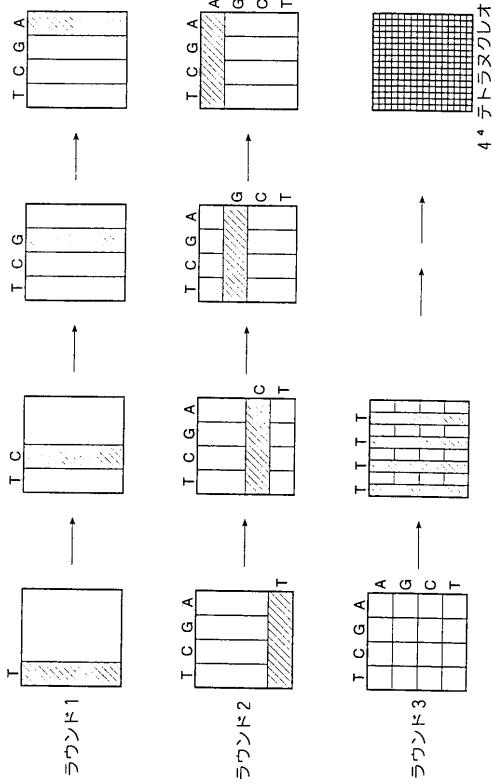


FIG. 4

【図 5 a】

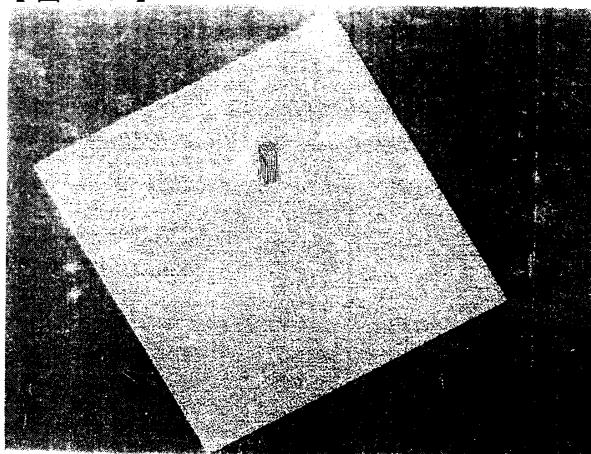


FIG. 5a

【図 5 B】

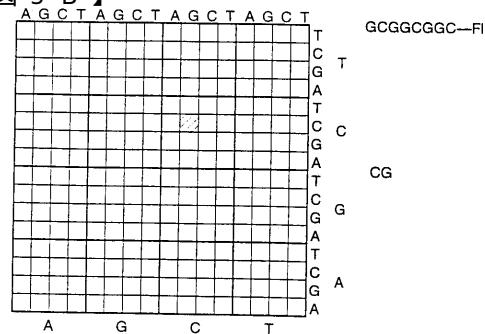


FIG. 5B

【図 6 A】

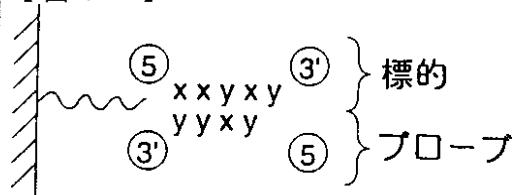


FIG. 6A

【図 6 B】

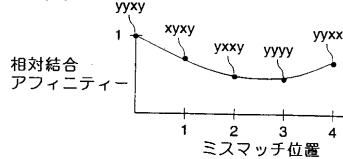


FIG. 6B

【図 6 C】

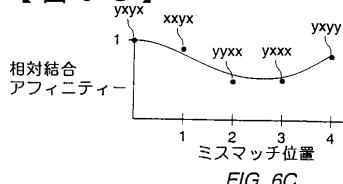


FIG. 6C

【図 7】

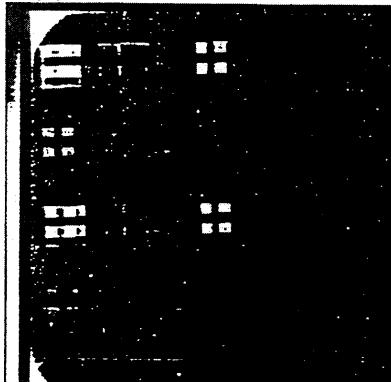


FIG. 7

【図 8】

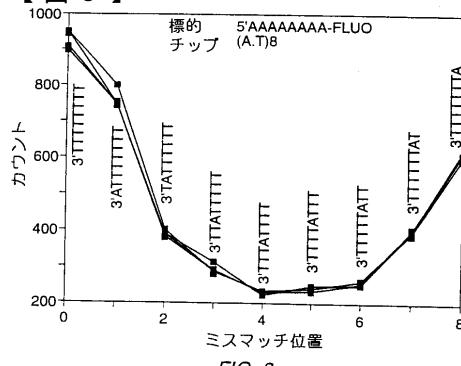


FIG. 8

【図 9】

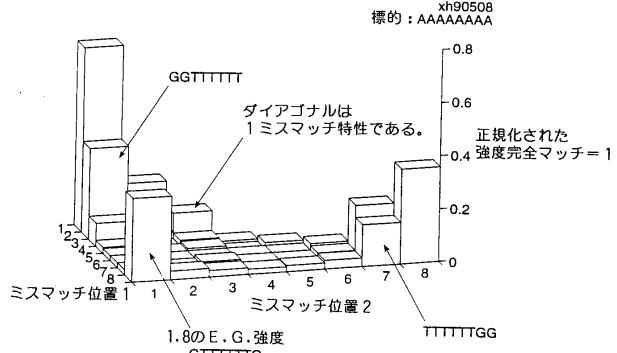


FIG. 9

【図 10】

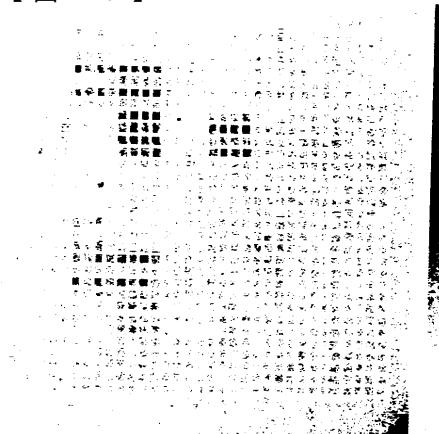


FIG. 10

【図 11 A】

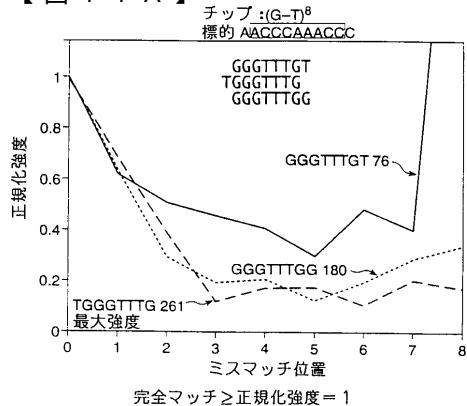


FIG. 11A

【図 11 B】

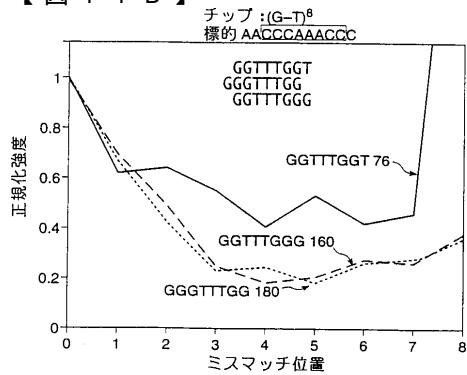


FIG. 11B

【図 11 C】

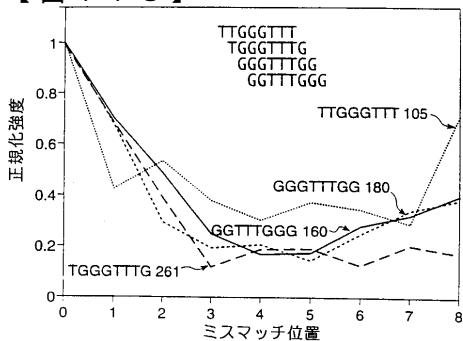


FIG. 11C

【図 12 B】

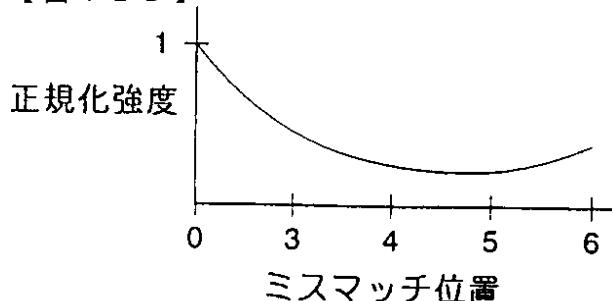


FIG. 12B

【図 12 A】

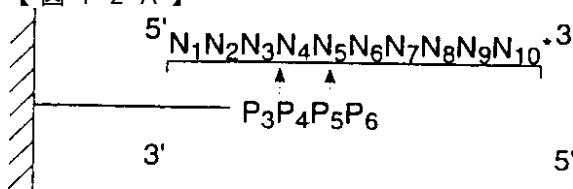


FIG. 12A

【図 12 C】

A P<sub>3</sub> P<sub>4</sub> P<sub>5</sub>  
T P<sub>3</sub> P<sub>4</sub> P<sub>5</sub>  
C P<sub>3</sub> P<sub>4</sub> P<sub>5</sub>  
G P<sub>3</sub> P<sub>4</sub> P<sub>5</sub>

FIG. 12C

【図 1 2 D】

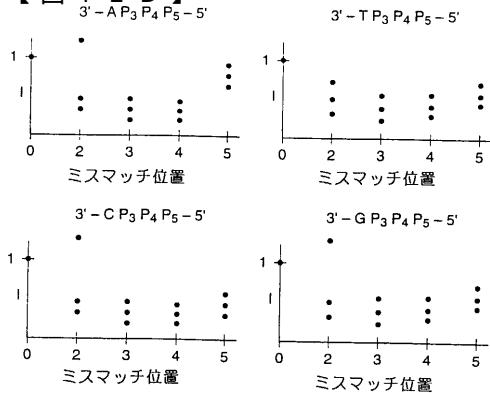


FIG. 12D

【図 1 3】

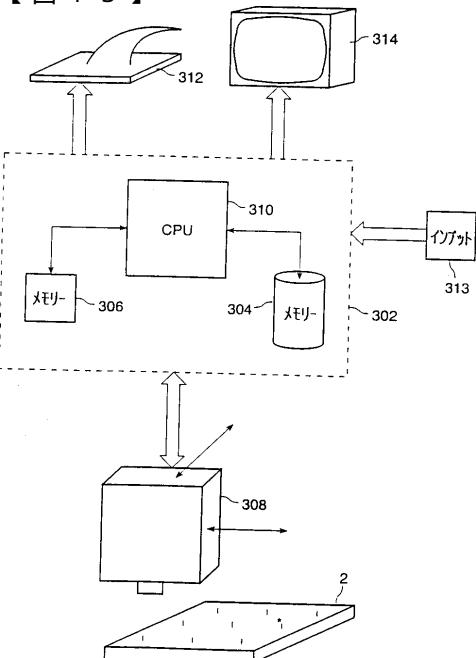


FIG. 13

【図 1 4】

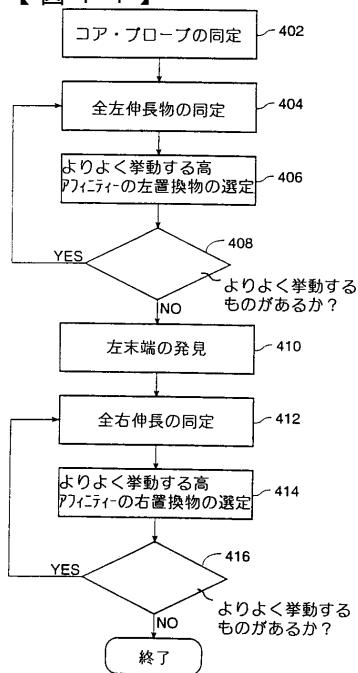


FIG. 14

【図 1 5 A】

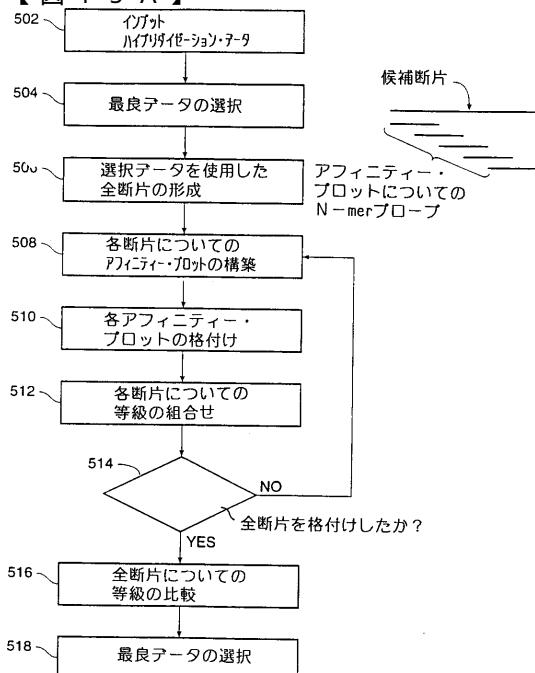
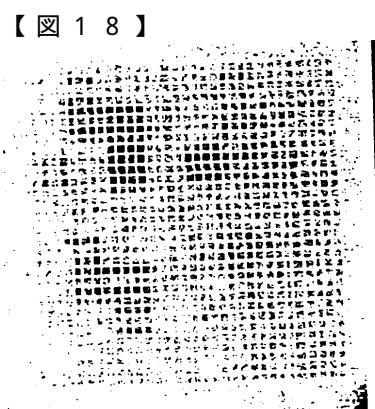
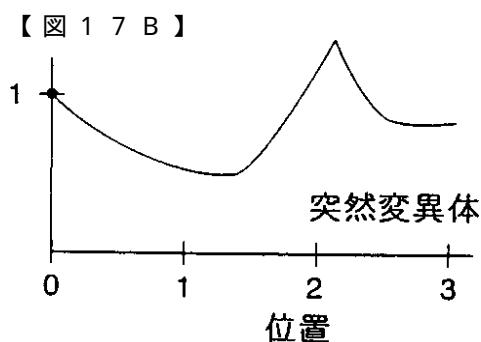
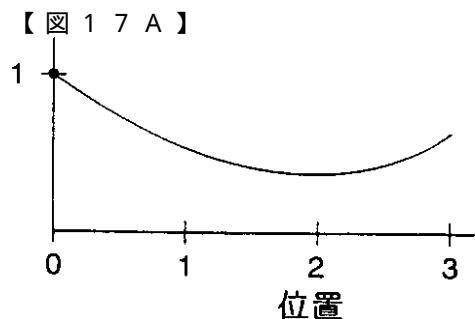
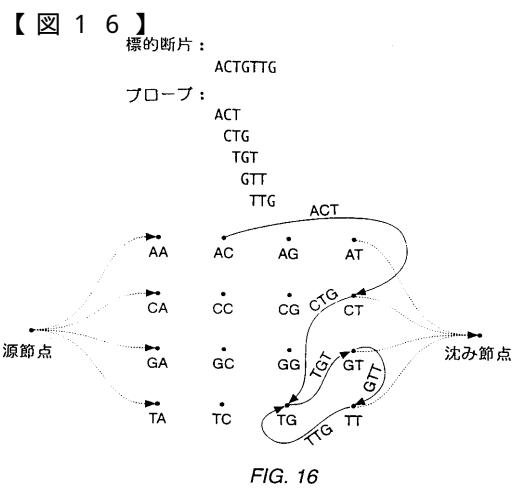
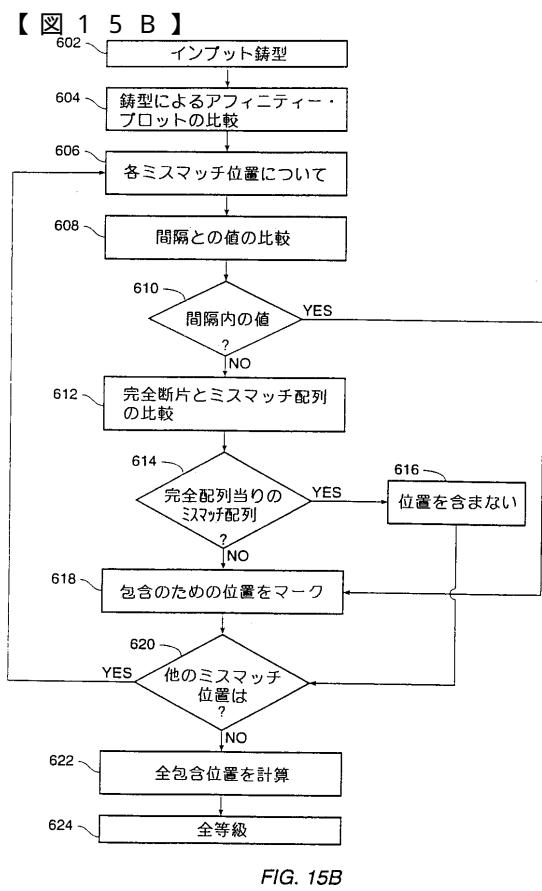
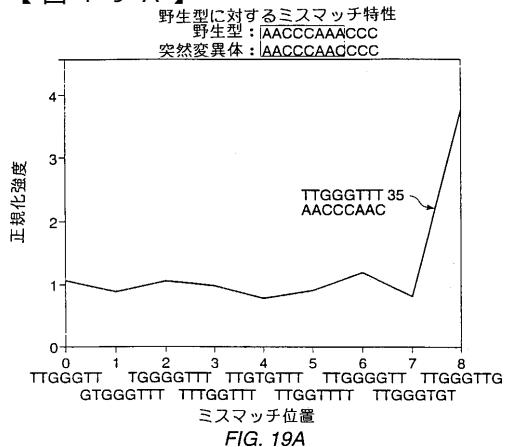


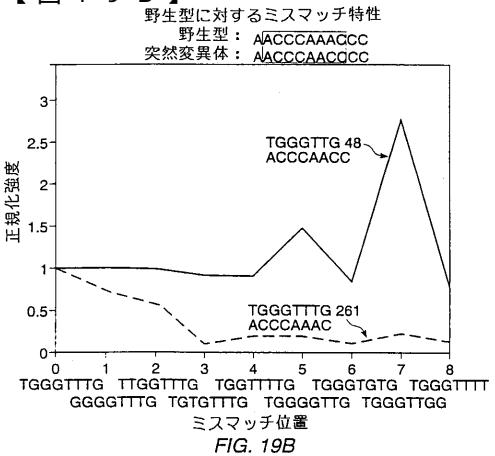
FIG. 15A



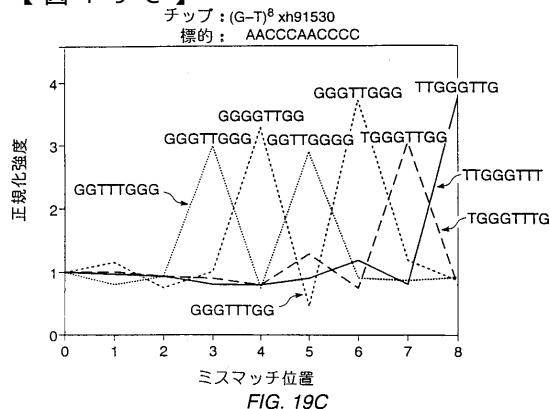
【図 19 A】



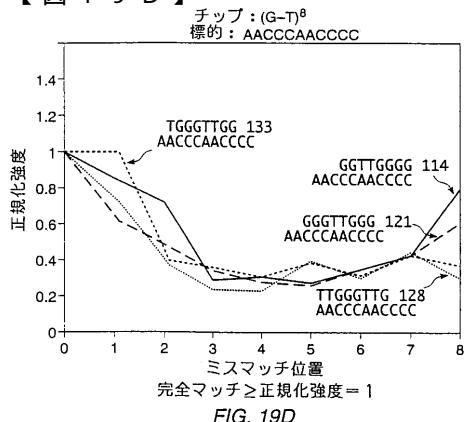
【図 19 B】



【図 19 C】



【図 19 D】



---

フロントページの続き

- (72)発明者 フォドア, スティーブン ピー. エー.  
アメリカ合衆国, カリフォルニア 94303, パロ アルト, ナザン ウェイ 3863
- (72)発明者 リップシュッツ, ロバート ジェイ.  
アメリカ合衆国, カリフォルニア 94301, パロ アルト, パロ アルト アベニュー 970
- (72)発明者 フアン, シャオファ  
アメリカ合衆国, カリフォルニア 94043, マウンテンビュー. ジャクソン ストリート 9  
37
- (72)発明者 ジェボンズ, ルイス キャロル  
アメリカ合衆国, カリフォルニア 94087, シニーベール, ラモーンネ アベニュー 701

審査官 小暮 道明

- (56)参考文献 国際公開第92/010588 (WO, A1)  
特表平04-501362 (JP, A)  
Genomics, Vol.13, No.4(1992)p.1008-1017  
DNA Seq., Vol.1, No.6(1991)p.375-388  
FEBS Lett., Vol.256, No.1-2(1989)p.118-122

(58)調査した分野(Int.Cl., DB名)

C12N 15/00 - 15/90  
C07K 1/00 - 19/00  
PubMed  
BIOSIS/WPI (DIALOG)