

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4588976号
(P4588976)

(45) 発行日 平成22年12月1日 (2010. 12. 1)

(24) 登録日 平成22年9月17日 (2010. 9. 17)

(51) Int. Cl.	F I
C 1 2 N 15/09 (2006. 01)	C 1 2 N 15/00 A
C 1 2 Q 1/68 (2006. 01)	C 1 2 Q 1/68 Z N A A
G O 1 N 33/53 (2006. 01)	G O 1 N 33/53 M
G O 1 N 33/566 (2006. 01)	G O 1 N 33/566

請求項の数 16 (全 45 頁)

(21) 出願番号 特願2002-559836 (P2002-559836)
 (86) (22) 出願日 平成14年1月25日 (2002. 1. 25)
 (65) 公表番号 特表2004-522440 (P2004-522440A)
 (43) 公表日 平成16年7月29日 (2004. 7. 29)
 (86) 国際出願番号 PCT/CA2002/000087
 (87) 国際公開番号 W02002/059354
 (87) 国際公開日 平成14年8月1日 (2002. 8. 1)
 審査請求日 平成17年1月20日 (2005. 1. 20)
 (31) 優先権主張番号 60/263, 710
 (32) 優先日 平成13年1月25日 (2001. 1. 25)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 60/303, 799
 (32) 優先日 平成13年7月10日 (2001. 7. 10)
 (33) 優先権主張国 米国 (US)

(73) 特許権者 502441949
 ルミネックス モレキュラー ダイアグノ
 スティックス, インク.
 カナダ国 オンタリオ州 エム5ジー 1
 ワイ8 トロント スート 900 ユニ
 バーシティ アベニュー 439
 (74) 代理人 100102978
 弁理士 清水 初志
 (74) 代理人 100102118
 弁理士 春名 雅夫
 (74) 代理人 100160923
 弁理士 山口 裕孝
 (74) 代理人 100119507
 弁理士 刑部 俊

前置審査

最終頁に続く

(54) 【発明の名称】 タグおよびタグ相補体としてのポリヌクレオチド、製品、ならびにその使用法

(57) 【特許請求の範囲】

【請求項 1】

各分子が配列番号：1～210から選択されるオリゴヌクレオチドを含む、タグまたはタグ相補体として使用される分子を含む組成物であって、該組成物が少なくとも10種の該分子を含む、組成物。

【請求項 2】

各分子が配列番号：1～7、9、13、15、16、21、23～25、28、31、33、34、37、39、41、42、44、46～48、50～56、58、61～64、67～70、72、74、75、77、79～81、84、87、88、92、94、96、97、100、101、103～105、107、110～113、115、117～119、122～124、126、130、132～134、140、141、174、176、180、182、184、188～190、192、193、196、197、199、および202～207から選択されるオリゴヌクレオチドを含む、請求項1記載の組成物。

【請求項 3】

少なくとも11の分子、または少なくとも12の分子、または少なくとも13の分子、または少なくとも14の分子、または少なくとも15の分子、または少なくとも16の分子、または少なくとも17の分子、または少なくとも18の分子、または少なくとも19の分子、または少なくとも20の分子、または少なくとも21の分子、または少なくとも22の分子、または少なくとも23の分子、または少なくとも24の分子、または少なくとも25の分子、または少なくとも26の分子、または少なくとも27の分子、または少なくとも28の分子、または少なくとも29の分子、または少なくとも30の分子、または少なくとも31の分子、または少なくとも32

10

20

の分子、または少なくとも33の分子、または少なくとも34の分子、または少なくとも35の分子、または少なくとも36の分子、または少なくとも37の分子、または少なくとも38の分子、または少なくとも39の分子、または少なくとも40の分子、または少なくとも41の分子、または少なくとも42の分子、または少なくとも43の分子、または少なくとも44の分子、または少なくとも45の分子、または少なくとも46の分子、または少なくとも47の分子、または少なくとも48の分子、または少なくとも49の分子、または少なくとも50の分子、または少なくとも60の分子、または少なくとも70の分子、または少なくとも80の分子、または少なくとも90の分子、または少なくとも100の分子を含む、請求項1もしくは2記載の組成物。

【請求項4】

10

それぞれの分子は、その相補体へのハイブリダイゼーションによって前記の分子の混合物から区別できるように固相担体に連結される、請求項1から3のいずれか一項記載の組成物。

【請求項5】

それぞれの分子は、固相担体上の定義された位置に連結され、それぞれの前記の分子の定義された位置は、その他の前記の分子について定義された位置と異なる、請求項4記載の組成物。

【請求項6】

それぞれの前記固相担体は微粒子であり、それぞれの前記分子は他の異なる分子とは異なる微粒子に共有結合している、請求項4もしくは5記載の組成物。

20

【請求項7】

それぞれの前記分子はタグ相補体を含む、請求項1から6のいずれか一項記載の組成物。

【請求項8】

1つ以上の固相担体を含み、担体のそれぞれが少なくとも10の空間的に別個の領域を持ち、それぞれの領域は共有結合した同一のタグ相補体の均一な集合であり、タグ相補体は請求項1から3のいずれか一項で定義されたオリゴヌクレオチドの集合から選択される、ポリヌクレオチドの分類および認別用キット。

【請求項9】

組成物の各オリゴヌクレオチドに関するタグ相補体がある、請求項8記載のキット。

【請求項10】

30

1つ以上の固相担体は平面の基質であり、少なくとも10の空間的に別個の領域は少なくとも10の空間的にアドレス可能な領域である、請求項8または9記載のキット。

【請求項11】

1つ以上の固相担体は複数の微粒子である、請求項8もしくは9のいずれか一項記載のキット。

【請求項12】

各微粒子は直径が5 μm から40 μm の範囲である、請求項11記載のキット。

【請求項13】

各微粒子は異なったオリゴヌクレオチドが結合された他の各微粒子から、分光測定により識別可能である、請求項11または12記載のキット。

40

【請求項14】

各核酸分子の遺伝子座における突然変異または多型の存在について、核酸分子に相補的な複数の2本鎖を含む生体試料を解析する方法であって：

(A) タグ配列に相補的な配列を持つ、5'末端を有する増幅した分子を形成するため、第一プライマー対の存在下で2本鎖分子を増幅する段階であって、各プライマーは請求項7で定義されたタグ相補体ファミリーに属するタグ相補体配列に相補的なタグ配列を有する、同一の5'配列を持つ段階、

(B) ポリメラーゼおよび増幅した配列の3'末端に相補的な5'末端を持つ第二プライマー対の存在下で増幅分子を伸長する段階であって、各第二プライマーの3'末端は、最隣接した前記の遺伝子座へと伸長し、それぞれ(i)転写の間、増大するヌクレオチド鎖の3'末端に

50

ポリメラーゼにより組み入れが可能であり、(ii)重合の終結を引き起こし、(iii) それぞれの示差検出が可能である、複数のヌクレオシド三リン酸塩誘導体の存在下で行われる段階、

(C)第二プライマーを(A)のタグ相補体配列を持つタグ相補体に特異的にハイブリダイズさせる段階、

(D)前記の遺伝子座に位置する塩基を同定するため、(B)の第二プライマーに組み込まれたヌクレオチド誘導体を検出する段階

を含み、ここで(A)の各タグの配列は各核酸分子にとって固有であり、段階(A)と(B)は相互の存在下で前記の核酸を用いて行われる方法。

【請求項15】

各核酸分子の遺伝子座における突然変異または多型の存在について、複数の核酸分子を含む生体試料を解析する方法であって：

(a)分子とプライマーをハイブリダイズさせ、プライマーは請求項7で定義されたタグ相補体ファミリーに属するタグ相補体配列に相補的なタグ配列を有する5'配列を持ち、3'末端は最隣接の遺伝子座へと伸長する段階、

(b)それぞれ：(i)増大するヌクレオチド鎖の3'末端に酵素的に組み入れが可能であり、(ii)伸長の終結を引き起こし、(iii)互いに示差検出が可能である、複数のヌクレオシド三リン酸塩誘導体の存在下で、プライマーの3'末端を酵素的に伸長する段階であって、ここで、前記の遺伝子座に存在する可能性のある各ヌクレオチドに相補的な誘導体がある段階、

(c)段階(b)で形成された伸長したプライマーを、(a)のタグ相補体配列を持つタグ相補体に特異的にハイブリダイズさせる段階、

(d)核酸分子の遺伝子座に位置する塩基を同定するため、段階(b)のプライマーに組み込まれたヌクレオチド誘導体を検出する段階

を含み、ここで(a)の各タグは各核酸分子にとって固有であり、段階(a)と(b)は相互の存在下で前記の核酸分子を用いて行われる方法。

【請求項16】

混合物に含まれていると推定されるターゲットの存在を決定する方法であって、以下の段階を含む方法：

(i)第一ラベルでターゲットをラベルする段階、

(ii)ターゲットへの特異的な結合が可能であり、かつ第一タグを含む第一検出部分を提供する段階、

(iii)前記の部分とターゲットの特異的結合を可能にする（又は、引き起こす）条件下で、混合物のサンプルを検出部分に露出する段階、

(iv)請求項7で定義されたタグ相補体ファミリーを提供し、ここでファミリーは、第一タグのファミリーに相補的な配列を持つ第一タグ相補体を含有している段階、

(v)第一タグとそのタグ相補体の特異的ハイブリダイゼーションを可能にする（又は、引き起こす）条件下で、サンプルをタグ相補体ファミリーに露出する段階、

(vi)混合物において前記のターゲットが存在するか否かを決定するため、前記の第一タグ相補体にハイブリダイズした前記の第一検出部分が、前記のラベルされたターゲットに結合されているかどうかを決定する段階。

【発明の詳細な説明】

【技術分野】

【0001】

本発明の属する技術分野

本発明は例えば、分子の分類をするオリゴヌクレオチドタグファミリーの使用に関するものである。タグ相補体への特異ハイブリダイゼーションによって、特定のタグファミリーメンバーは互いに区別される。

【背景技術】

【0002】

本発明の背景

10

20

30

40

50

基本的なプロセスであるオリゴヌクレオチドとアナログの特異ハイブリダイゼーションは、診断アッセイで病気関連のポリヌクレオチドを識別し、ノーベルターゲット・ポリヌクレオチドのクローンをスクリーン、ポリヌクレオチド混合物プロットで特異ポリヌクレオチドを識別、不適切に発現された遺伝子とDNA塩基配列決定を治療学的に妨害するなど幅広い種類の研究、医療用、産業上の利用に使われている。配列特異ハイブリダイゼーションは、高生産性複合核酸アッセイの発展には重要である。ヒトゲノムプロジェクトなどのプロジェクトで収集した膨大な配列情報量をさらに広げるアッセイフォーマットとして、高性能な配列特異ハイブリダイゼーションを成功させるのは非常に難しくなっている。

【0003】

その大部分では、オリゴヌクレオチド使用のハイブリダイゼーションの成功は擬陽性と擬陰性の数を最小化することで決まる。このような問題は一つの試験で多数のハイブリダイゼーションプローブを併用することを非常に難しくしている。そのようななものには複合方式などがあるが、特に遺伝子マイクロアレイにおいての同義遺伝子配列の解析がそうである。例えばいくつかの結合アッセイで、特定のターゲット配列がチップに固定している相補体に重点的に結合すると推定して、複数の核酸分子がチップに結合している。固体担体に結合しているオリゴヌクレオチドタグ使用に関しての研究開発が進められている。このオリゴヌクレオチドタグは特にプローブ配列に結合しているタグ相補体にハイブリダイゼーションするのに使用されるChetverinら（国際公開公報第93/17126号）では核酸を調査、分類するのにセグメント化されたバイナリオリゴヌクレオチドアレイを使用している。これらのアレイは隣接した変数のヌクレオチド配列に結合している定数のヌクレオチド配列をもっており、両方とも共有結合部分をもって固体担体に結合している。またこれらのバイナリアレイは末端配列に従ってストランドを分類するのに使用でき、それにより各ストランドがアレイの特定の位置に結合するので普通のアレイよりも利点がある。この研究の末端配列の設計は、定数の配列と変数の配列で構成されている。Chetverinらの米国特許第6,103,463号と第6,322,971号が2000年8月15日と2001年11月27日に別々に公開されている。

【0004】

このタグ使用で混合物を分類する考えは、バクテリア遺伝学やイースト遺伝学で発された分子タグと相似している（Henselら、Science; 269, 400-403: 1995とSchoemakerら、Nature Genetics; 14, 450-456: 1996）。ここでは各突変異体が異なったDNA配列でタグ付けされている「シグネチャータグ付け」突然変異誘発とよばれる方法で、突然変異遺伝子を約10,000個の複雑な細菌集落の混合物から回収するのに使われる。Baranyら（国際公開公報第9731256号）のタグ付け方法は、「ジップチップ」として知られており、これは核酸分子ファミリーの「ジップコードアドレス」で、これらは互いに違い、グリッド上にセットされているターゲット分子は「ジップコード」と呼ばれる「ジップコードアドレス」に相補するオリゴヌクレオチド配列に結合しており、特にグリッドのアドレス位置にハイブリダイゼーションするのに使われる。アドレスとして使われているこれらのポリヌクレオチド配列ファミリーの選択がアッセイを正確に性能させるのには重要だが、その性能が説明されていない。

【0005】

高度な交又法ハイブリダイゼーション環境の研究には特定なハイブリダイゼーションが要求され、それに使用されるオリゴヌクレオチドファミリーの設計には綿密な選択基準が要求される。このような方法の成功は特異プローブハイブリダイゼーションとその相補体による。核酸分子ファミリーがターゲット配列にクロスハイブリダイゼーションしたり不正確にハイブリダイゼーションすると色々な問題が生じる。不正確なハイブリダイゼーションの為擬陽性が生じたりハイブリッドが形成できない結果、偽陰性が生じたりすることはよくあるが、そのような結果は最小限に押さえないといけない。そのためこの目標を達成するには、核酸ハイブリッドを形成する特定の熱力学的性質を検討すべきである。オリゴヌクレオチドがTm（50%の核酸二本鎖が解離する温度）で知られている相補する配列

で二本鎖を形成する温度は、いくつかの配列相互依存特質によって変動する。それらにはカノニカル対のA-TとG-C（GCが塩基組成物に反映する）の水素結合エネルギーやスタッキング自由エネルギー、そしてこれはあまりないが隣接相互作用などがある。これらのエネルギーは、一般にハイブリダイゼーションアッセイで使われるオリゴヌクレオチド間で大きく変動する。例を挙げれば、標準状態の相補的ターゲットで、一本でGCが40%、もう一本でGCが60%で、24のヌクレオチドから構成されている二本のプローブ配列のハイブリダイゼーションでは、理論上10°Cの融解温度の違いがある(Muellerら、Current Protocols in Mol. Biol.; 15, 5:1993)。全オリゴヌクレオチド配列集団において正確なハイブリダイゼーションを行なうのに不適切な単一温度状態でハイブリダイゼーションが形成されると、ハイブリダイゼーションに問題が生じる。非相補的なプローブのミスマッチハイブリダイゼーションでは、測定可能なミスマッチの安定度で二本鎖を形成することがある(Santaluciaら、Biochemistry; 38: 3468-77, 1999)。特異オリゴヌクレオチド集団の二本鎖ミスマッチは、ミスマッチが二本鎖の安定度の減少を引き起こすハイブリダイゼーション状態で起き、その結果、その特異集団の最小安定の正確な二本鎖よりさらに高いT_mを生じる。例えばハイブリダイゼーションがAT-リッチ完全マッチの二本鎖配列に好ましい状態で行われると、AT-リッチ二本鎖を正確に形成する融解温度よりも高い温度のミスマッチ塩基があるGC-リッチ二本鎖配列をハイブリダイゼーションする可能性がでてくる。そのため複合ハイブリダイゼーション反応に使用できるオリゴヌクレオチド配列ファミリーの設計には、オリゴヌクレオチドの熱力学的性質と、設計されたオリゴヌクレオチド集団内のクロスハイブリダイゼーション挙動を減少あるいは排除する二本鎖形成を考慮に入れるべきである。

10

20

【0006】

複合配列法は1990年7月17日のChurchに交付された米国特許第4,942,124号で説明されている。この方法は少なくともタグ配列で互いに異なる二つのベクトルを必要とする。明細書には一つのベクトルのタグ配列は厳密なハイブリダイゼーション状況でもう一つのベクトルのタグ配列にハイブリダイゼーションしない、と述べられている。これにはあるベクトルでの相補的なタグのプローブはもう一つのベクトルのタグ配列とクロスハイブリダイゼーションしない、という例がある。500~1000 mMのナトリウムリン酸緩衝液で42°Cの温度が模範的な厳密ハイブリダイゼーション状況である。G 残基が欠如している42の20merタグ配列集団全部はChurchの明細書の図3に記載されている。Churchは、配列がいかにか得られたか詳細には記載していないが、まず最初に92本は配列の固有さを保証するのに十分な配列多様性があることから選択されたことは書かれている。

30

【0007】

タグファミリーの発展はそのほかにも試みられている。複合ハイブリダイゼーションアッセイで使用する配列の選択は色々試みられている。アドレス可能なアレイのジップコードとして、またはタグとして使える配列の選択はBrennerと彼の同僚の研究として特許資料に説明されている。米国特許第5,654,413号でオリゴヌクレオチドタグ（および対応するタグ相補体）の集団が説明されている。その説明によると、各オリゴヌクレオチドタグは多数のサブユニットを持っており、各サブユニットは3から6のヌクレオチド長さのオリゴヌクレオチドから構成されている。また各サブユニットは最小限のクロスハイブリダイゼーション集団から選択され、この集団のサブユニットは、集団の他の配列以外にも少なくとも二つのミスマッチを持っている。Brenner 特許明細書の表IIでは前述の基準に従って最小限にクロスハイブリダイゼーションしている模範的な4merサブユニット群の説明がされている。Brennerの研究では、非クロスハイブリダイゼーションオリゴヌクレオチドの作成は、同じ集団の他のサブユニット相補体以外にも少なくとも二つのミスマッチがある二本鎖を形成するサブユニットの使用に依存している。オリゴヌクレオチドタグ作成においてのサブユニットの順序付けは、明確には定義されていない。

40

【0008】

サブユニットを基にしたタグの設計で使われるパラメーターはBaranyら（国際公開公報第9731256号）で検討されている。ポリヌクレオチド配列の設計において、4つの可能性あ

50

るテトラマー集団から長さ24のヌクレオチド(24mer)が、派生され、それぞれ24mer「アドレス」は、最隣接している24mer近所から3テトラマー違うという例がある。さらにBaranyらでは、各テトラマーが他のテトラマーと少なくとも2ヌクレオチド違う場合、各24merは隣のものより少なくとも6ヌクレオチド違うことがさらに検討されている。集団の二配列間にアランメントを形成する時は、組み込みや欠失を考慮せずに決定される。このようにして固有な「ジップコード」配列が作成される。ジップコードはターゲット依存方法(target dependent manner)でラベルに結合すると、固有なジップコードを作り出し、チップ上でそのアドレスにハイブリダイゼーションされる。ジップコードが他のアドレスにクロスハイブリダイゼーションするのを最小限におさえるためには、ハイブリダイゼーション反応が75~80°Cの温度で行わなければならない。ハイブリダイゼーションの高温状態のため、部分的なホモロジーを持っている24merは、完全な相補性で「デッドゾーン」の配列より少なくハイブリダイゼーションする。この厳密なハイブリダイゼーション状態を実行する方法には例えば高温度ハイブリダイゼーションがあり、これはBrennerらによっても実行されている。

10

【0009】

サブユニットを基にした非クロスハイブリダイゼーションタグの設計技術の現状は、厳密な非クロスハイブリダイゼーション挙動を必要とするアッセイにおいて実用的価値がある配列ファミリーを作成するための指導が十分ない。

【0010】

従ってそのようなアレイがただちに多数のアドレス分子を得ることが望ましいが、その半面アドレス分子は分子相補体配列に対して非常に選択的でなければならない。そのようなアレイは、グリッドを作っている分子ファミリーは設計だけのためで、配列が自然に起きるため配列に依存しなくてよいという利点がある。しかし、大きさが十分で各メンバーが他のジップコード分子全体に対し相補体を選択的な分子ファミリー群は(例えば、効果的に低いクロスハイブリダイゼーションおよびクロストークがある)研究者を当惑し続けている。

20

【発明の開示】

【0011】

発明の要約

Benightらの方法(2001年8月16日付け国際公開公報第01/59151号に公開された国際特許出願第PCT/CA 01/00141号に記述されている)を用い、核酸検出アッセイでの使用に最適なハイブリダイゼーション特質を得るのにコンピューターアルゴリズムを使い、100のヌクレオチド配列ファミリーを得ている。100オリゴヌクレオチド配列集団はハイブリダイゼーションアッセイを特徴とし、ファミリーメンバーはクロスハイブリダイゼーション無しで正確に相補配列にハイブリダイゼーションする能力を持っている。これらは表Iの配列番号:1から100までの配列である。この配列集団は110個の補足的な配列を含むように拡大され、非クロスハイブリダイゼーションの特質を持っている最初の100配列の特徴を基に、一緒にまとめることができる。これらの補足された配列は表Iの配列番号:101から210配列として同定されている。これらの配列の採集方法は下記に説明されている。

30

【0012】

表Iからの配列ファミリーの変異体配列ファミリー(タグまたはタグ相補体参照)も本発明の一部である。論議用として、タグ相補体ファミリーの説明をする。

40

【0013】

相補体ファミリーは表Iのようなオリゴヌクレオチドファミリーを基にオリゴヌクレオチド集団から得られる。実施例として、表Iのオリゴヌクレオチドを基にした相補体ファミリーの供給方法を説明する。

【0014】

まず最初に、表Iのオリゴヌクレオチドを基にした配列は次のように表される。

【0015】

(表IA) 一連のオリゴヌクレオチドのワードパターンに対応する数値配列

50

配列識別子

数値パターン

1	1	4	6	6	1	3
2	2	4	5	5	2	3
3	1	8	1	2	3	4
4	1	7	1	9	8	4
5	1	1	9	2	6	9
6	1	2	4	3	9	6
7	9	8	9	8	10	9
8	9	1	2	3	8	10
9	8	8	7	4	3	1
10	1	1	1	1	1	2
11	2	1	3	3	2	2
12	3	1	2	2	3	2
13	4	1	4	4	4	2
14	1	2	3	3	1	1

配列識別子	数値パターン						
15	1	3	2	2	1	4	
16	3	3	3	3	3	4	
17	4	3	1	1	4	4	
18	3	4	1	1	3	3	
19	3	6	6	6	3	5	
20	6	6	1	1	6	5	
21	7	6	7	7	7	5	10
22	8	7	5	5	8	8	
23	2	1	7	7	1	1	
24	2	3	2	3	1	3	
25	2	6	5	6	1	6	
26	4	8	1	1	3	8	
27	5	3	1	1	6	3	
28	5	6	8	8	6	6	
29	8	3	6	5	7	3	
30	1	2	3	1	4	6	
31	1	5	7	5	4	3	
32	2	1	6	7	3	6	
33	2	6	1	3	3	1	
34	2	7	6	8	3	1	
35	3	4	3	1	2	5	20
36	3	5	6	1	2	7	
37	3	6	1	7	2	7	
38	4	6	3	5	1	7	
39	5	4	6	3	8	6	
40	6	8	2	3	7	1	
41	7	1	7	8	6	3	
42	7	3	4	1	6	8	
43	4	7	7	1	2	4	
44	3	6	5	2	6	3	
45	1	4	1	4	6	1	
46	3	3	1	4	8	1	
47	8	3	3	5	3	8	
48	1	3	6	6	3	7	
49	7	3	8	6	4	7	
50	3	1	3	7	8	6	30
51	10	9	5	5	10	10	
52	7	10	10	10	7	9	
53	9	9	7	7	10	9	
54	9	3	10	3	10	3	
55	9	6	3	4	10	6	
56	10	4	10	3	9	4	
57	3	9	3	10	4	9	
58	9	10	5	9	4	8	
59	3	9	4	9	10	7	
60	3	5	9	4	10	8	
61	4	10	5	4	9	3	
62	5	3	3	9	8	10	
63	6	8	6	9	7	10	
64	4	6	10	9	6	4	40
65	4	9	8	10	8	3	
66	7	7	9	10	5	3	
67	8	8	9	3	9	10	
68	8	10	2	9	5	9	

配列識別子

数値パターン

69	9	6	2	2	7	10
70	9	7	5	3	10	6
71	10	3	6	8	9	2
72	10	9	3	2	7	3
73	8	9	10	3	6	2
74	3	2	5	10	8	9
75	8	2	3	10	2	9
76	6	3	9	8	2	10
77	3	7	3	9	9	10
78	9	10	1	1	9	4
79	10	1	9	1	4	1
80	7	1	10	9	8	1
81	9	1	10	1	10	6
82	9	6	9	1	3	10
83	3	10	8	8	9	1
84	3	8	1	9	10	3
85	9	10	1	3	6	9
86	1	9	1	10	3	1
87	1	4	9	6	8	10
88	3	3	9	6	1	10
89	5	3	1	6	9	10
90	6	1	8	10	9	6
91	5	9	9	4	10	3
92	2	10	9	1	9	5
93	10	10	7	2	1	9
94	10	9	9	1	8	2
95	1	8	6	8	9	10
96	1	9	1	3	8	10
97	9	6	9	10	1	2
98	1	10	8	9	9	2
99	1	9	6	7	2	9
100	4	3	9	3	5	1
101	5	11	10	14	12	1
102	7	12	4	13	3	2
103	5	5	4	4	12	9
104	2	13	13	11	13	13
105	10	2	5	4	12	7
106	11	7	4	11	6	4
107	12	12	1	9	11	11
108	12	9	4	14	12	6
109	12	7	13	2	9	11
110	9	11	3	4	1	3
111	10	5	12	11	4	4
112	4	13	7	12	1	5
113	9	13	10	11	11	6
114	10	14	14	10	1	3
115	2	14	1	10	4	5
116	10	12	12	7	11	10
117	9	11	2	12	8	11
118	2	8	5	2	12	14
119	1	8	13	3	7	8
120	9	4	7	5	4	2
121	13	2	12	7	1	12
122	11	10	9	7	5	11

10

20

30

40

配列識別子

数値パターン

123	8	12	2	2	12	7
124	5	2	14	3	4	13
125	1	8	8	1	5	9
126	14	5	11	10	13	3
127	14	1	4	13	2	4
128	4	4	5	11	3	10
129	10	9	2	3	3	11
130	11	4	8	14	3	4
131	5	1	14	8	11	2
132	14	3	11	6	12	5
133	13	4	4	1	10	1
134	6	10	11	6	5	1
135	5	8	12	5	1	7
136	4	5	9	6	9	2
137	13	2	4	4	2	3
138	11	2	2	5	9	3
139	8	1	10	12	2	8
140	12	7	9	11	4	1
141	12	1	4	14	3	13
142	11	2	7	10	4	1
143	3	4	12	11	11	11
144	3	3	4	2	12	11
145	1	5	9	4	2	1
146	6	1	12	2	10	5
147	10	5	1	12	2	14
148	2	11	7	9	4	11
149	7	4	4	5	14	12
150	12	5	2	1	10	12
151	5	9	2	11	6	1
152	12	14	3	6	1	14
153	5	9	11	10	1	4
154	2	5	12	14	10	10
155	4	5	8	4	5	6
156	10	12	4	6	12	5
157	4	2	1	13	6	8
158	9	10	10	14	5	3
159	6	14	10	11	3	3
160	2	9	10	12	5	7
161	13	3	7	10	5	12
162	6	4	1	2	5	13
163	6	1	13	4	14	13
164	2	12	1	14	1	9
165	4	11	13	2	6	10
166	1	10	7	4	5	8
167	7	2	2	10	13	4
168	8	2	11	4	6	14
169	4	8	2	6	2	3
170	7	1	12	11	2	9
171	5	6	10	4	13	4
172	5	10	4	11	9	3
173	3	11	9	3	2	3
174	8	15	6	20	17	19
175	21	10	15	3	7	11
176	11	7	17	20	14	9

10

20

30

40

配列識別子

数値パターン

177	16	6	17	13	21	21
178	10	15	22	6	17	21
179	15	7	17	10	22	22
180	3	20	8	15	20	16
181	17	21	10	16	6	22
182	6	21	14	14	14	16
183	7	17	3	20	10	7
184	16	19	14	17	7	21
185	20	16	7	15	22	10
186	20	10	18	11	22	18
187	18	7	19	15	7	22
188	21	18	7	21	16	3
189	14	13	7	22	17	13
190	19	7	8	12	10	17
191	15	3	21	14	9	7
192	19	6	15	7	14	14
193	4	17	10	15	20	19
194	21	6	18	4	20	16
195	2	19	8	17	6	13
196	12	12	6	17	4	20
197	16	21	12	10	19	16
198	14	14	15	2	7	21
199	8	16	21	6	22	16
200	14	17	22	14	17	20
201	10	21	7	15	21	18
202	16	13	20	18	21	12
203	15	7	4	22	14	13
204	7	19	14	8	15	4
205	4	5	3	20	7	16
206	22	18	6	18	13	20
207	19	6	16	3	13	3
208	18	6	22	7	20	18
209	10	17	11	21	8	13
210	7	10	17	19	10	14

10

20

30

【0016】

ここでは、各数値1から22（数値識別子）は4merを示し、上記リストの配列の数値パターン1から22は、表Iに記載されているオリゴヌクレオチドの四量体オリゴヌクレオチドセグメントパターンと一致する。このオリゴヌクレオチドは非クロスハイブリダイゼーションであることが分かったが、それは実施例の詳細で説明される。各4merはWWWW、WWWX、WWY、WWXW、WWXX、WWXY、WWYW、WWYX、WWYY、WXWW、WXWX、WXWY、WXXW、WXXX、WXXY、WXYW、WXYX、WXYX、WYWW、WYWX、WYWY、WYXW、WYXX、WYXY、WYYW、WYYX、WYYY、XWWW、XWWX、XWWY、XWXW、XWXX、XWXY、XWYW、XWYX、XWYY、XXWW、XXWX、XXWY、XXXW、XXXX、XXXY、XXYW、XXYX、XXYY、XYWW、XYWX、XYWY、XYXW、XYXX、XYXY、XYYW、XYXX、XYYY、YWWW、YWWX、YWWY、YWXW、YWXX、YWXY、YWYW、YWYX、YWYY、YXWW、YXWX、YXWY、YXXW、YXXX、YXXY、YXYW、YXYX、YXYY、YYWW、YYWX、YYWY、YYXW、YYXX、YYXY、YYYW、YYXX、YYYYから構成されている4mer群から選択されている。ここではWとXとYは、ヌクレオチド塩基A、ヌクレオチド塩基G、ヌクレオチド塩基Cなどであり、これらは下記に説明されたルールで作成された塩基割り当てである。

【0017】

この数値パターンでは、1 = WXYX、2 = YWXYなどの例のように4merが数値に割り当てられている。いったん特定の4merが特定の数値に割り当てられると、違う数値位置用には割り当てられない。しかし違う4merを同じ数値に割り当ててことは可能である。例としては、ある位置での数値1は割り与えられたWXYXで、違う位置にある他の数値1は割り与えら

50

れたXXXWである。しかし、その他の数値2から10は、割り与えられたWXYXか XXXWにはなれない。言い換えると1から22は、その他の1から22全部から区別するために表示された81の4merリストから割り当てられた4merである。

【0018】

表Iの特異オリゴヌクレオチドの場合は、1 = WXYX、2 = YWXY、3 = XXXW、4 = YWYX、5 = WYXY、6 = YYWX、7 = YWXX、8 = WYXX、9 = XYYW、10 = XYWX、11 = YYXW、12 = WYYY、13 = XYXW、14 = WYYY、15 = WXYW、16 = WYXW、17 = WXXW、18 = WYYW、19 = XYYX、20 = YXYX、21 = YXXY、22 = XYXYとなる。

【0019】

いったん4merが上記のパターンに基づいた位置に割り当てられると、特異オリゴヌクレオチド集団は塩基A、塩基T/U、塩基G、塩基Cから塩基W、塩基X、塩基Yの適切な割り当てを使って作成することができる。これらの割り当ては次の二つのルールセット中一つから作成されている：

(i) 塩基W、塩基X、塩基Y：

(a) W = A、T/U、G、Cの中の一つ、

X = A、T/U、G、Cの中の一つ、

Y = A、T/U、G、Cの中の一つ、

W、X、Yは他のW、X、Yから区別するため選択されている。

(b) 特定の非選択塩基(i)(a)は、何回でもW、X、Yのうちのどれか一つと置き換えられる、

あるいは

(ii) 塩基W、塩基X、塩基Y

(a) W = Gか C、

X = Aか T/U、

Y = Aか T/U、

X ≠ Y、

(b)(ii)(a)で選択されなかった塩基は、一つ以上の位置で各配列に挿入できる。各挿入位置は、各配列で同じで、又集団の配列の他の配列でも同じである。

【0020】

表Iの特異オリゴヌクレオチドの場合、W = G、X = A、Y = Tとなる。

【0021】

いずれにせよ、これらのルールセットのうちのひとつに従って生成されたオリゴヌクレオチド集団をみると、比較的簡単な方法で特定集団メンバーの修正が可能である。その結果そのような修正を必要とする集団のクロスハイブリダイゼーションの特質をある程度維持しながら違う配列集団を得る。特に配列集団のどの配列の位置でもAとT/UとGとCのうち、三つまで挿入可能である。その代わりとしてまたは追加として、三つまでの塩基を配列集団中のどの配列からでも欠失できる。

【0022】

当業者なら理解できるだろうが、タグファミリー（又はタグ相補体）としての使用を可能にする特質集団のオリゴヌクレオチド集団をみると、集団の全メンバーの順序を逆転させて他の同じ特質をもつファミリーを得ることができる。つまり全メンバーは5'から3'または3'から5'と読みとることができる。

【0023】

本発明の相補体ファミリーは、上記で定義された特定オリゴヌクレオチド集団を基にしている。各ファミリー相補体は集団の異なったオリゴヌクレオチドを基とし、各相補体は少なくともそれが基となっている10の連続するオリゴヌクレオチド（例えば、隣接する）塩基を含んでいる。隣接する塩基配列を選択する時は、集団の各オリゴヌクレオチドの隣接する塩基が、配列内の前記オリゴヌクレオチドの第一塩基の位置と集団の全ヌクレオチドとが同じであるように選択されている集団を優先する。従って、例で言えば20の隣接塩基のヌクレオチド配列が、ヌクレオチド配列が基にされている配列の3塩基から22塩基に

10

20

30

40

50

一致すると、好ましくはヌクレオチド配列全部の20隣接塩基は、ヌクレオチド配列が基にされている配列の3塩基から22塩基に一致する。相補体ファミリーがクロスハイブリダイゼーションを引き起こすインター配列類似を減らすかまたは最小限にしようとする場合、一つ一つの相補体対は特定の相同条件を満たす。特に下記に説明されている一部の例外に関しては、相補体集団内の二つの相補体はいずれも一般的には定義された量の相違がなければならない。

【0024】

ファミリーの相補体対に相違の必要条件があるように、その相違の必要条件を抽象的に理解するため、偽配列が相補体対から生成される。偽配列は塩基のストリングの、対の各相補体からの選択によって相補体対から生成される一本の配列で、ストリングの塩基は両方の相補体で同じ順序で起きる。このような偽配列を作り出す目的は、二つの親配列の配列アイデンティティを偽配列が作り出したものと比較する便利で客観的な方法を作り出すことである。

【0025】

偽配列は共通配列の考えに相似していると考えられる。共通配列は各位置（両配列結合）で最も一般的に選択された塩基を反映する位置にあり、通常は親配列両方からの全部の塩基でなっており、「偽」配列では両方の親配列で同じ順序で起こる塩基だけで構成されている（両配列の交差）。これ以外では偽配列とは、当業者にとっては理解しやすいものである。共通配列はまた通例二つの配列の（又は二本より多い配列でこれは何本の配列が共通配列を作るのに使われるかによる）一般系統発生起源を示すもので、偽配列の定義は先祖の前歴はないがある状態で、クロスハイブリダイゼーションする傾向がある二つの相補的配列間の配列類似点を明確に説明するために作られたものである。

【0026】

偽配列は次の表のような模範配列1および2から生成されることもある。

配列 1: ATGTTTAGTGAAAAGTTAGTATTG
 * •

配列 2: ATGTTTAGTGAATAGTATAGTATTG
 • ♦

偽配列: ATGTTTAGTGAAAAGTTAGTATTG

【0027】

これらの二本の配列から生成された偽配列は、その結果22塩基長となる。これは配列1および2には、同一配列（同一順序）を含んだ22同一塩基がある、ということである。配列を生成されたものから比較した時、偽配列には合計三つの挿入、欠失、ミスマッチがみられる。

ATGT-TAGTGAA-AGT-TAGTATTG

偽配列に関する後者の説明での点線は、それが生成された親配列に対して偽配列の挿入、欠失、ミスマッチの位置を示している。従って配列1のアスタリスク付き「T」と配列2のダイヤモンド付き「A」と二つの点が付いている配列1と配列2のミスマッチ「A-T」は、偽配列を生成するときに欠失したものである。

【0028】

当業者はアスタリスクやダイヤモンドが「挿入／欠失」という言葉を意味することを分かっている。厳密に言えば、変更があるにかかわらず、挿入や欠失は単に見解の一つでしかない。これは、偽配列を得るため配列1の四番目の塩基を欠失、又は配列1を得るのに偽配列の三番目の塩基の後に「T」が挿入可能ということである。

【0029】

親配列の一本から一つの挿入／欠失を除去することによって偽配列を作り出せるかがわかる。それは比較される二本の配列の一本に一塩基がある以外は、偽配列の長さでは二本の親配列が同一ホモロジーを持っていることである。同様に、それぞれの親の一塩基であるミスマッチの塩基対の欠失で偽配列を作り出すことができるかがわかる。それは比較される各配列に一塩基がある以外は、二本の親配列が偽配列の長さでは同一ホモロジーを持っていることである。このような理由で、二本の配列のホモロジーを比較する時、挿入／欠失の効果は塩基対のミスマッチの効果と同等と考えられる。

【0030】

いったん偽配列が生成されると、相補体ファミリー内で生成されていた偽配列の相補体対の適合性は、系統的に評価できる。

10

【0031】

本発明の実施例としては、相補体対から生成された偽配列が次の特質を持っていると、相補体対は相補体ファミリー内の封入に対し適合性を持つ：

(1)偽配列での連続塩基配列は、相補体の連続塩基配列に一致し、それが生成されている第一と第二の相補体は $((3/4 \times L) - 1)$ の塩基長しかない。

(2)偽配列が $(5/6 \times L)$ の長さと同等あるいは長い場合、それが生成されているものから第一と第二の相補体を比較した時、少なくとも3つの挿入／欠失、またはミスマッチを含んでいる。

(3)偽配列が $(11/12 \times L)$ の長さ以下か同等である。

20

【0032】

L_1 は第一の相補体の長さ、 L_2 は第二の相補体の長さであり、 $L = L_1$ であり、または $L_1 \neq L_2$ である場合、 L は L_1 と L_2 の大きいほうとなる。

【0033】

本発明の好適な実施例では、集団の相補体対全部が上記の特質を持っている。この特殊な条件下で、ファミリーの他の相補体と比較したとき、これらの必要条件全部に満たない限られた数の相補体をもつことが有利になることがある。

【0034】

一つのケースでは、最初の相補体では、記載された三条件全部を満たしていない第二相補体が多くても二つ、ファミリーにはある。そのような二つの相補体では、タグ相対と最初の相補体間により大きなクロスハイブリダイゼーションの可能性がある。他のケースでは、第一の相補体には記載された三条件に満たない第二相補体が多くても一つある。

30

【0035】

相補体の特定数や特定部分が記載された三条件を満たしていない相補体ファミリーの設計も本発明では可能である。例えば、互いに比較した時、集団内の一つの相補体対だけが条件を満たしていないところで集団が設計されることがある。このとき二つの対、三つの対、および任意の数の可能な対を全部含んでいることがある。もう一つの方法として、記載された三条件の一つ以上を満たさない他の配列と比較した時、条件に満たない相補体対、たとえば対の10%の相補体対が特定比率を持てば有利になることがある。この数は5%, 15%, 20%, 25%, 30%, 35%, 40%であるかもしれない。

40

【0036】

前述の比較は大部分が適切なコンピューターソフトウェアで般に行われる。偽配列に関してははっきりと理解されるために抽象的に説明されているが、有能なコンピュータープログラマーが対をなす相補体を理論的にさまざまな方法で比較し、同等な結果を得ることはできる。

【0037】

AとGとT/UとCの記号がここでは技術的に意味を持っている。TとUの場合、この発明の内容においてのインターストランド水素結合(Watson-Crick)の結合特質に関しては、これらが互いに同等なことが当業者にはわかるだろう。従って、二つの塩基は互換性があり、T/Uの意味もある。

50

【0038】

天然に発生した塩基のアナログは要求される各場所に挿入できる。アナログはペプチド核酸のような非天然型塩基で、それに対応するヌクレオチド塩基が天然に生じるのと同じ方法で標準のWatson-Crickの対形成を起こす。

【0039】

広い見方では、本発明はタグ又はタグ相補体の使用としての分子から成る組成物で、各分子は表IAの数値パターンを持った配列群を基にしたオリゴヌクレ

オチド集団から選択されたオリゴヌクレオチドから成る。ここでは、

(A)1から22はWWWW、WWWX、WWWY、WWXW、WWXX、WWXY、WWYW、WWYX、WWYY、WXWW、WXWX、WXWY、WXXW、WXXX、WXXY、WXYW、WXYX、WXYX、WXYX、WYWW、WYWX、WYWY、WYXW、WYXX、WYXY、WYYW、WYYX、WYYY、XWWW、XWWX、XWWY、XWXW、XWXX、XWXY、XWYW、XWYX、XWYY、XXWW、XXWX、XXWY、XXXW、XXXX、XXXW、XXYX、XXYY、XYWW、XYWX、XYWY、XYXW、XYXX、XXYX、XYYW、XYXX、XYYY、YWWW、YWWX、YWWY、YWXW、YWXX、YWXY、YWYW、YWYX、YWYY、YXWW、YXWX、YXWY、YXXW、YXXX、YXXY、YXYW、YXYX、YXYX、YYWW、YYWX、YYWY、YYXW、YYXX、YYXY、YYYW、YYYX、YYYYから成る4mer群から選択された4merである。

(B)他の1から22から区別するため1から22が選択されている。

(C)W、X、Yは(i)または(ii)のどちらかが真性の塩基である。

(i)(a)W = A、T/U、G、Cのうちの一つ

X = A、T/U、G、Cのうちの一つ

Y = A、T/U、G、Cのうちの一つ

他のW、X、Yから区別するため各W、X、Yが選択されている。

(b)(i)(a)の選択されていない前記の塩基はW、X、Yのうち一つと何回でも置き換えられる。

(ii)(a)W = GまたはC、

X = AまたはT/U、

Y = AまたはT/U、

X ≠ Y

(b)(ii)(a)で選択されなかった塩基は一つ以上の場所で、各配列に挿入できる。挿入場所は全配列で同じである。

(D)塩基三つまではどの配列のどの場所でも挿入できる。又塩基三つまではどの配列からでも欠失できる。

(E)前記のオリゴヌクレオチド群の全配列は、5'から3'または3'から5'と読むことができる。

前記集団の各オリゴヌクレオチドは、少なくともオリゴヌクレオチドがベースとなる隣接する10塩基の配列を持っており、次のようになる。

(F)(I)GとCの合計を集団の配列全部を合わせたA、T/U、G、C合計で割った商は、約0.1と0.40間で、集団配列の前記商は、合わせた配列の商と0.2より多くは変動しない。

(II)L₁とL₂セット長さの第一と第二配列の対から生成された偽配列では、互いに同一の配列での同一塩基の第一と第二配列各自から選択。

(i)偽配列の連続塩基配列は第一と第二の連続塩基配列と同一であり、そこから生成されている偽配列の連続塩基配列は $((3/4 \times L) - 1)$ 塩基長未満である。

(ii)偽配列の長さが $(5/6 \times L)$ より長い同一であり、偽配列を生成している第一と第二の配列と比較した時、少なくとも三つの挿入／欠失またはミスマッチを含んでいる。かつ、

(iii)偽配列の長さが $(11/12 \times L)$ 以下か同一である。

L = L₁、またはL₁ ≠ L₂であれば、LはL₁とL₂の大きい方となり、

ここにある塩基はどれもアナログによって置き換えられことがある。

【0040】

好適な実施例では、本発明のオリゴヌクレオチド集団は実施例2で試験された配列の数値パターンを基にしている。

【0041】

好ましくは

(G)24mer配列集団では、各 1 = GATT、各2 = TGAT、各3 = AAAG、各4 = TGTA、各5 = GTAT、各6 = TTGA、各7 = TGAA、各8 = GTAA、各9 = ATTG、各10 = ATGA、各11 = TTAG、各12 = GTTA、各13 = ATAG、各14 = GTTT、各15 = GATG、各16 = GTAG、各17 = GAAG、各18 = GTTG、各19 = ATTA、各20 = TATA、各21 = TAAT、各22 = ATATとなる。配列集団では、各 1 = GATT、各2 = TGAT、各3 = AAAG、各4 = TGTA、各5 = GTAT、各6 = TTGA、各7 = TGAA、各8 = GTAA、各9 = ATTG、各10 = ATGA、各11 = TTAG、各12 = GTTA、各13 = ATAG、各14 = GTTT、各15 = GATG、各16 = GTAG、各17 = GAAG、各18 = GTTG、各19 = ATTA、各20 = TATA、各21 = TAAT、各22 = ATATである。定義された条件下では、配列と24mer配列群の異なった配列相補体間での最大のハイブリダイゼーションは、前記配列とその相補体間のハイブリダイゼーション度の30%を超えない。集団の全オリゴヌクレオチドでは、オリゴヌクレオチドと他の集団のオリゴヌクレオチド相補体間での最大のハイブリダイゼーション度はオリゴヌクレオチドとその相補体の50%を超えない。

10

【0042】

従って、選択された集団のオリゴヌクレオチド全部が、最小限で全てクロスハイブリダイゼーションするかどうか定期的に決定するのが可能であることがわかる。前記の定義された条件セットにおいて特に(G)では、配列と別の配列の相補体間の最大のハイブリダイゼーション度が、前記の配列と相補体間のハイブリダイゼーション度の30%を超えることがない。又、各配列と相補体間のハイブリダイゼーション度は、1倍から10倍の間で変動することは事実でもある。さらに好ましいのは1倍から9倍の間であり、より好ましいのは1倍から8倍の間である。最適なオリゴヌクレオチド集団は、配列とその特別の相補体間のハイブリダイゼーション度が1倍から8.25倍の間で変化し、配列と別配列の相補体間のハイブリダイゼーションの最高度は配列とその特別相補体間のハイブリダイゼーション度が10.2%を超えないことが下記の実施例2で立証された。

20

【0043】

好ましくは、配列と24mer配列群の異なった配列相補体間の(G)のハイブリダイゼーションの最高度が25%を超えないことである。さらに好ましくは、配列と24mer配列群の異なった配列の相補体との間の(G)のハイブリダイゼーションの最高度が20%を超えないことである。さらに好ましくは、配列と24mer配列群の異なった配列の相補体との間の(G)のハイブリダイゼーションの最高度が15%を超えないことで、それ以上に好ましくは、配列と24mer配列群の異なった配列の相補体との間の(G)のハイブリダイゼーションの最高度が11%を超えないことである。

30

【0044】

好ましくは、(G)の定義された条件セット下では、配列と他の集団の他の配列の相補体間のハイブリダイゼーションの最高度は配列と前記の24mer配列群の異なった配列の相補体間のハイブリダイゼーションの最高度より15%より大きくなることはない。さらに好ましくは10%、より好ましくは5%より大きくなることである

【0045】

後に説明されている実施例2によると、37°Cの温度のpH 8.0状態で 0.2 M 塩化ナトリウム、0.1 M トリス、および0.08% トリトン X-100の条件下で、微粒子かビーズの場合、24merヌクレオチド配列が共有結合的に固体担体に結合されるとき、配列と24mer配列群の異なった配列の相補体間のハイブリダイゼーションの最高度が10.2%を超えない。

40

【0046】

組成物の他の好ましい面では、(G)の24mer群では、異なった配列の相補体と配列間での最高度のハイブリダイゼーションが、前記の配列とその相補体間のハイブリダイゼーション度の15%を超えない。そして、配列とその相補体間のハイブリダイゼーション度は1から9の間の因子によって変動する。オリゴヌクレオチド集団全部では、オリゴヌクレオチドと他の集団のオリゴヌクレオチドの相補体間のハイブリダイゼーションの最高度はオリゴヌクレオチドとその相補体のハイブリダイゼーション最高度の20%を超えない。

50

【0047】

望ましい点では、各4merはWXXX、WXXY、WXYX、WXYX、WYXX、WYXY、WYYX、WYYY、XWXX、XWXY、XWYX、XWYY、XXWX、XXWY、XXXW、XXYW、XYWX、XYWY、XYXW、XXYW、YWXX、YWXY、YWXY、YWYY、YXWX、YXWY、YXXW、YXYW、YYWX、YYWY、YYXW、YYYWから成る4mer群の中から選択された1から22の数値で示される。

【0048】

他の面では、1の数値で示される各4merは互いに同一である。2の数値で示される各4merは互いに同一である。3の数値で示される各4merは互いに同一である。4の数値で示される各4merは互いに同一である。5の数値で示される各4merは互いに同一である。6の数値で示される各4merは互いに同一である。7の数値で示される各4merは互いに同一である。8の数値で示される各4merは互いに同一である。9の数値で示される各4merは互いに同一である。10の数値で示される各4merは互いに同一である。11の数値で示される各4merは互いに同一である。12の数値で示される各4merは互いに同一である。13の数値で示される各4merは互いに同一である。14の数値で示される各4merは互いに同一である。15の数値で示される各4merは互いに同一である。16の数値で示される各4merは互いに同一である。17の数値で示される各4merは互いに同一である。18の数値で示される各4merは互いに同一である。19の数値で示される各4merは互いに同一である。20の数値で示される各4merは互いに同一である。21の数値で示されている各4merは互いに同一である。22の数値で示される各4merは互いに同一である。

【0049】

他の面では、数値1で示されている4merの少なくともそのうちの一つにWXYX 配列がある。数値2で示されている4merの少なくともそのうちの一つにYWX 配列がある。数値3で示されている4merの少なくともそのうちの一つにXXXW 配列がある。数値4で示されている4merの少なくともそのうちの一つにYWX 配列がある。数値5で示されている4merの少なくともそのうちの一つにWYX 配列がある。数値6で示されている4merの少なくともそのうちの一つにYYWX 配列がある。数値7で示されている4merの少なくともそのうちの一つにYWX 配列がある。数値8で示されている4merの少なくともそのうちの一つにWYXX 配列がある。数値9で示されている4merの少なくともそのうちの一つにXXYW 配列がある。数値10で示されている4merの少なくともそのうちの一つにXYWX 配列がある。数値11で示されている4merの少なくともそのうちの一つにYYXW 配列がある。数値12で示されている4merの少なくともそのうちの一つにWYYY 配列がある。数値13で示されている4merの少なくともそのうちの一つにXXYW 配列がある。数値14で示されている4merの少なくともそのうちの一つにWYY 配列がある。数値15で示されている4merの少なくともそのうちの一つにWXYW 配列がある。数値16で示されている4merの少なくともそのうちの一つにWYXW 配列がある。数値17で示されている4merの少なくともそのうちの一つにWXXW 配列がある。数値18で示されている4merの少なくともそのうちの一つにWYYW 配列がある。数値19で示されている4merの少なくともそのうちの一つにXXYX 配列がある。数値20で示されている4merの少なくともそのうちの一つにYXYX 配列がある。数値21で示されている4merの少なくともそのうちの一つにYXXY 配列がある。数値22で示されている4merの少なくともそのうちの一つにXXYX 配列がある。

【0050】

好ましい面では、組成物の発明は 各1 = WXYX、各2 = YWX、各3 = XXXW、各4 = YWX、各5 = WYX、各6 = YYWX、各7 = YWX、各8 = WYX、各9 = XYYW、各10 = XYWX、各11 = YYXW、各12 = WYYY、各13 = XYXW、各14 = WYYY、各15 = WXYW、各16 = WYXW、各17 = WXXW、各18 = WYYW、各19 = XYYX、各20 = YXYX、各21 = YXXY、各22 = XYYXとなる。

【0051】

広い面においては、本発明は表IAの識別子数値1から173の数値パターンがある配列を基にした配列群から成り立っている組成物である。組成物は(A)の数値1から14で示されている各4merはWXYX、YWX、XXXW、YWX、WYX、YYWX、YWX、WYX、XYYW、XYWX、YYXW、WYYY、XYXW、WYYYから成る4mer群から選択されている。

【0052】

そのような組成物では、数値1で示されている各 4merは互いに同一であり、数値1で示されている各 4merは互いに同一であり、数値2で示されている各 4merは互いに同一であり、数値3で示されている各 4merは互いに同一であり、数値4で示されている各 4merは互いに同一であり、数値5で示されている各 4merは互いに同一であり、数値6で示されている各 4merは互いに同一であり、数値7で示されている各 4merは互いに同一であり、数値8で示されている各 4merは互いに同一であり、数値9で示されている各 4merは互いに同一であり、数値10で示されている各 4merは互いに同一であり、数値11で示されている各 4merは互いに同一であり、数値12で示されている各 4merは互いに同一であり、数値13で示されている各 4merは互いに同一であり、数値14で示されている各 4merは互いに同一であることが好ましい。

10

【0053】

数値1で示されている4merの少なくともそのうちの一つがWXYX 配列を持っており、数値2で示されている4merの少なくともそのうちの一つがYWX 配列を持っており、数値3で示されている4merの少なくともそのうちの一つがXXXW配列を持っており、数値4で示されている4merの少なくともそのうちの一つがYWX 配列を持っており、数値5で示されている4merの少なくともそのうちの一つがYWX 配列を持っており、数値6で示されている4merの少なくともそのうちの一つがYWX 配列を持っており、数値7で示されている4merの少なくともそのうちの一つがYWX 配列を持っており、数値8で示されている4merの少なくともそのうちの一つがYWX 配列を持っており、数値9で示されている4merの少なくともそのうちの一つがYWX 配列を持っており、数値10で示されている4merの少なくともそのうちの一つがYWX 配列を持っており、数値11で示されている4merの少なくともそのうちの一つがYWX 配列を持っており、数値12で示されている4merの少なくともそのうちの一つがYWX 配列を持っており、数値13で示されている4merの少なくともそのうちの一つがYWX 配列を持っており、数値14で示されている4merの少なくともそのうちの一つがYWX 配列を持っていることが又望まれる。

20

【0054】

さらに好ましくは、各 1 = WXYX、各2 = YWX、各3 = XXXW、各4 = YWX、各5 = YWX、各6 = YWX、各7 = YWX、各8 = YWX、各9 = YWX、各10 = YWX、各11 = YWX、各12 = YWX、各13 = YWX、各14 = YWXである。

30

【0055】

他の広い面においては、本発明は表IAの配列識別子1から100の数値パターンがある配列を基にした配列集団から成り立っている組成物である。組成物は(A)の数値1から10で示されている各4merはWXYX、YWX、XXXW、YWX、YWX、YWX、YWX、YWX、YWX、およびYWXから成る4mer集団から選択されている。

【0056】

そのような組成物では、数値1で示されている各4merは互いに同一であり、数値2で示されている各4merは互いに同一であり、数値3で示されている各4merは互いに同一であり、数値4で示されている各4merは互いに同一であり、数値5で示されている各4merは互いに同一であり、数値6で示されている各4merは互いに同一であり、数値7で示されている各4merは互いに同一であり、数値8で示されている各4merは互いに同一であり、数値9で示されている各4merは互いに同一であり、数値10で示されている各4merは互いに同一であることが好ましい。

40

【0057】

また、好ましくは数値1で示されている4merの少なくともそのうちの一つがWXYX配列で、数値2で示されている4merの少なくともそのうちの一つがYWX配列で、数値3で示されている4merの少なくともそのうちの一つがXXXW配列で、数値4で示されている4merの少なくともそのうちの一つがYWX配列で、数値5で示されている4merの少なくともそのうちの一つがYWX配列で、数値6で示されている4merの少なくともそのうちの一つがYWX配列で、数値7で示されている4merの少なくともそのうちの一つがYWX配列で、数値8で示されている4merの少なくともそのうちの一つがYWX配列で、数値9で示されている4merの少なくともそのうちの一つがYWX配列で、数値10で示されている4merの少なくともそのうちの一つがYWX配列で、数値11で示されている4merの少なくともそのうちの一つがYWX配列で、数値12で示されている4merの少なくともそのうちの一つがYWX配列で、数値13で示されている4merの少なくともそのうちの一つがYWX配列で、数値14で示されている4merの少なくともそのうちの一つがYWX配列を持っていることが好ましい。

50

る4merの少なくともそのうちの一つがWYXX配列で、数値9で示されている4merの少なくともそのうちの一つがXYW配列で、数値10で示されている4merの少なくともそのうちの一つがXYWX配列である。

【0058】

より好ましくは、各1 = WXYX、各2 = YWXY、各3 = XXXW、各4 = YWYX、各5 = WYXY、各6 = YYWX、各7 = YWXX、各8 = WYXX、各9 = XYYW、各10 = XYWXである。

【0059】

最も好ましい組成物は(C)(i)(a)において：W = G と Cのうちの一つ；X = A とT/Uのうちの一つ；ならびに(F)の条件を主張するためには Y = A とT/Uのうちの一つとなる。より好ましくは、(C)(i)(a)： W = G；X = A と T/Uのうちの一つ；そして Y = A とT/Uのうちの一つである。さらにより好ましくは、W = G；X = A；そして Y = T/Uである。

10

【0060】

特定のオリゴヌクレオチド配列変異体が最も好ましい配列（表I）の一つに近ければ近いほど、その変異体は最小限にクロスハイブリダイゼーションするオリゴヌクレオチドのメンバーとして、さらに好ましい配列に類似してくるということを当業者は分かっている。

【0061】

配列群（オリゴヌクレオチド）の最小限のクロスハイブリダイゼーションがここで述べられている時は、それは配列群の特定メンバー（オリゴヌクレオチド）のみ群の他の配列（オリゴヌクレオチド）相補体と最小限にハイブリダイゼーションすることを意味している。

20

【0062】

好ましくは、(F)(I)では集団の各配列の商は組み合わせた配列の商と0.1より多く違わない。より好ましくは集団の各配列の商は組み合わせた配列の商と0.05より多く違わない。より好ましくは、集団の各配列の商は組み合わせた配列の商と0.01より多く違わない。

【0063】

又(F)(I)では、組み合わせた集団の配列全部であるAとT/UとG とCの合計で割ったG とCの合計の商は、約0.15と0.35の間であるのが最適である。より好ましくは、0.2と0.3の間である。より好ましくは、0.21と0.29の間である。より好ましくは、0.22と0.28の間である。より好ましくは、0.23と0.27の間である。さらにより好ましくは、0.24と0.26の間である。最も好ましくは、0.25である。

30

【0064】

好ましくは、(D)では二つまでの塩基がどの配列のどの位置にでも挿入できる。または二つまでの塩基がどの配列からも欠失できる。より好ましくは、一つの塩基だけがどの配列のどの位置にでも挿入でき、または一つの塩基がどの配列からも欠失できる。最も好ましくは、どの配列のどの位置からも塩基が挿入されないことである。

【0065】

(D)では、塩基はどの配列からも欠失できないことが好適とされる。最も好ましくは、(D)のどの配列のどの場所からも塩基が挿入あるいは欠失できない。

【0066】

40

好適な組成物では、各集団のオリゴヌクレオチドは、オリゴヌクレオチドがベースになっている少なくとも11の隣接した塩基の配列を持っている。さらに望ましいことは、各集団のオリゴヌクレオチドは、オリゴヌクレオチドがベースになっている少なくとも12の隣接した塩基の配列を持っている。さらに望ましいことは、各集団のオリゴヌクレオチドは、オリゴヌクレオチドがベースになっている少なくとも13の隣接した塩基の配列を持っている。さらに望ましいことは、各集団のオリゴヌクレオチドは、オリゴヌクレオチドがベースになっている少なくとも14の隣接した塩基の配列を持っている。さらに望ましいことは、各集団のオリゴヌクレオチドは、オリゴヌクレオチドがベースになっている少なくとも15の隣接した塩基の配列を持っている。さらに望ましいことは、各集団のオリゴヌクレオチドは、オリゴヌクレオチドがベースになっている少なくとも16の隣接した塩基の配列

50

を持っている。さらに望ましいことは、各集団のオリゴヌクレオチドは、オリゴヌクレオチドがベースになっている少なくとも17の隣接した塩基の配列を持っている。さらに望ましいことは、各集団のオリゴヌクレオチドは、オリゴヌクレオチドがベースになっている少なくとも18の隣接した塩基の配列を持っている。さらに望ましいことは、各集団のオリゴヌクレオチドは、オリゴヌクレオチドがベースになっている少なくとも19の隣接した塩基の配列を持っている。さらに望ましいことは、各集団のオリゴヌクレオチドは、オリゴヌクレオチドがベースになっている少なくとも20の隣接した塩基の配列を持っている。さらに望ましいことは、各集団のオリゴヌクレオチドは、オリゴヌクレオチドがベースになっている少なくとも21の隣接した塩基の配列を持っている。さらに望ましいことは、各集団のオリゴヌクレオチドは、オリゴヌクレオチドがベースになっている少なくとも22の隣接した塩基の配列を持っている。さらに望ましいことは、各集団のオリゴヌクレオチドは、オリゴヌクレオチドがベースになっている少なくとも23の隣接した塩基の配列を持っている。さらに望ましいことは、各集団のオリゴヌクレオチドは、それがベースになっている少なくとも24の隣接塩基の配列を持っている。

【0067】

望ましいことは、各集団のオリゴヌクレオチドは30塩基長までである。さらに望ましいのは、各集団のオリゴヌクレオチドは29塩基長までである。さらに望ましいのは、各集団のオリゴヌクレオチドは28塩基長までである。さらに望ましいのは、各集団のオリゴヌクレオチドは27塩基長までである。さらに望ましいのは、各集団のオリゴヌクレオチドは26塩基長までである。さらに望ましいのは、各集団のオリゴヌクレオチドは25塩基長までである。さらに望ましいのは、各集団のオリゴヌクレオチドは24塩基長までである。

【0068】

一部の好適な態様では、各集団のオリゴヌクレオチドは集団のオリゴヌクレオチド全部の平均5塩基長以内の長さを持っている。さらに好ましくは、各集団のオリゴヌクレオチドは集団のオリゴヌクレオチド全部の平均4塩基長以内の長さを持っている。さらに好ましくは、各集団のオリゴヌクレオチドは集団のオリゴヌクレオチド全部の平均3塩基長以内の長さを持っている。さらに好ましくは、各集団のオリゴヌクレオチドは集団のオリゴヌクレオチド全部の平均2塩基長以内の長さを持っている。さらに好ましくは、各集団のオリゴヌクレオチドは集団のオリゴヌクレオチド全部の平均1塩基長以内の長さを持っている。

【0069】

好ましくは、前記集団の各オリゴヌクレオチドの隣接塩基のストリングはそれを基とした配列内の各ストリングの第一塩基の位置が、集団のヌクレオチド全部と同じであるように選択されている。

【0070】

好適な態様において、組成物は少なくとも10の前述の分子、または少なくとも11の前述の分子、または少なくとも12の前述の分子、または少なくとも13の前述の分子、または少なくとも14の前述の分子、または少なくとも15の前述の分子、または少なくとも16の前述の分子、または少なくとも17の前述の分子、または少なくとも18の前述の分子、または少なくとも19の前述の分子、または少なくとも20の前述の分子、または少なくとも21の前述の分子、または少なくとも22の前述の分子、または少なくとも23の前述の分子、または少なくとも24の前述の分子、または少なくとも25の前述の分子、または少なくとも26の前述の分子、または少なくとも27の前述の分子、または少なくとも28の前述の分子、または少なくとも29の前述の分子、または少なくとも30の前述の分子、または少なくとも31の前述の分子、または少なくとも32の前述の分子、または少なくとも33の前述の分子、または少なくとも34の前述の分子、または少なくとも35の前述の分子、または少なくとも36の前述の分子、または少なくとも37の前述の分子、または少なくとも38の前述の分子、または少なくとも39の前述の分子、または少なくとも40の前述の分子、または少なくとも41の前述の分子、または少なくとも42の前述の分子、または少なくとも43の前述の分子、または少なくとも44の前述の分子、または少なくとも45の前述の分子、または少なくとも46の前述

の分子、または少なくとも47の前述の分子、または少なくとも48の前述の分子、または少なくとも49の前述の分子、または少なくとも50の前述の分子、または少なくとも60の前述の分子、または少なくとも70の前述の分子、または少なくとも80の前述の分子、または少なくとも90の前述の分子、または少なくとも100の前述の分子があり、少なくともオリゴヌクレオチドがベースとなっている配列集団のサイズによる、110の前述の分子、または少なくとも120の前述の分子、または少なくとも130の前述の分子、または少なくとも140の前述の分子、または少なくとも150の前述の分子、または少なくとも160の前述の分子、または少なくとも170の前述の分子、または少なくとも180の前述の分子、または少なくとも190の前述の分子、または少なくとも200の前述の分子を含む。

【0071】

本発明のどのオリゴヌクレオチドファミリーを使用するかによるが、互いに区別できる配列（例えば、最小限にクロスハイブリダイゼーションする）や互いにクロスハイブリダイゼーションする多数の配列を入れるのは適切かもしれないし、適切でないかもしれないが当業者は十分それを理解している。

【0072】

望ましい面としては、本発明は、(II)(i)における偽配列の塩基の任意の連続配列が、それが生成されている各第一配列と第二配列の塩基の連続配列と同じであり、 $(2/3 \times L) - 1$ 塩基長にすぎない組成物である。さらに望ましいのは、偽配列が $(3/4 \times L)$ より長いまたは、同じであれば偽配列が生成されている第一と第二の配列と比べた時、少なくとも三つの挿入／欠失またはミスマッチを含む。さらに望ましいことは、偽配列が長さ $(2/3 \times L)$ より長かったり、同じであればそれが生成されている第一と第二の配列と比べた時、少なくとも三つの挿入／欠失またはミスマッチを含む。

【0073】

他の望ましい面では、(II)(iii)では偽配列長さが $(5/6 \times L)$ 以下か同一である。さらに望ましい面は、偽配列長さが $(3/4 \times L)$ 以下か同一である。

【0074】

他の広い面からいうと、本発明はタグ用かタグ相補体用に使われる分子を含んでいる組成物で、各分子はIA表の実施例2で試験された配列の数値パターンをもっている配列群をベースにしたオリゴヌクレオチド集団から選択されたオリゴヌクレオチドを含む。ここでは、

(A) 1 = WXYX、各 2 = YWXY、各 3 = XXXW、各 4 = YWYX、各 5 = WYXY、各 6 = YYWX、各 7 = YWXX、各 8 = WYXX、各 9 = XYYW、各 10 = XYWX、各 11 = YYXW、各 12 = WYYX、各 13 = XYXW、各 14 = WYYY、各 15 = WXYW、各 16 = WYXW、各 17 = WXXW、各 18 = WYYW、各 19 = XYYX、各 20 = YXYX、各 21 = YXXY、各 22 = XYXY;

(B) 塩基W、塩基X、塩基Y

(i)(a) W = A、T/U、G、Cのうちの一つ

X = A、T/U、G、およびCのうちの一つ

Y = A、T/U、G、およびCのうちの一つ

各W、X、Yは他のW、X、Y全部から区別するため選択されている。

(b)(i)(a)の選択されていない前記塩基はW、X、Yのどれか一つと何回でも置き換えることができる。

(ii)(a) W = Gか C、

X = Aか T/U、

Y = Aか T/U、

X ≠ Y

(b)(ii)(a)で選択されていない塩基は一つ以上の場所で各配列に挿入できる。各挿入場所は全配列で同じである。

(C) 三つまでの塩基はどの配列のどの場所にも挿入できる。または三つまでの塩基はどの配列からも欠失できる。

(D) 前記のオリゴヌクレオチド群の全配列は5'から 3' あるいは 3'から 5'と読むことが

できる。

前記集団の各オリゴヌクレオチドには、オリゴヌクレオチドがベースとなっている少なくとも10の隣接する塩基配列があり、以下ようになる。

(E)GとCの合計を集団の配列A、T/U、G、C全部を合わせた合計で割った商は、約0.1と0.40間で、集団の各配列の前記の商は、合わせた配列の商と0.2より多くは変動しない。

(F)24mer配列集団では、各1 = GATT、各2 = TGAT、各3 = AAAG、各4 = TGTA、各5 = GTAT、各6 = TTGA、各7 = TGAA、各8 = GTAA、各9 = ATTG、各10 = ATGA、各11 = TTAG、各12 = GTTA、各13 = ATAG、各14 = GTTT、各15 = GATG、各16 = GTAG、各17 = GAAG、各18 = GTTG、各19 = ATTA、各20 = TATA、各21 = TAAT、各22 = ATATである。定義された条件下では、配列と24mer配列群の異なった配列の相補体間の最大のハイブリダイゼーション度は、前記配列とその相補体間のハイブリダイゼーション度の30%を超えることはない。集団のオリゴヌクレオチド相補体ではオリゴヌクレオチドと集団の他のオリゴヌクレオチド相補体間のハイブリダイゼーション最高度はオリゴヌクレオチドとその相補体のハイブリダイゼーション度の50%を超えない。ここでのどの塩基もアナログで置き換えられることがある。

10

【0075】

さらに好ましくは、集団の各オリゴヌクレオチドの隣接する塩基は、それが基となっている配列内の各オリゴヌクレオチドの第一塩基の位置が集団の全ヌクレオチドと同じであるように選択されている。

【0076】

20

好ましい面では上記の(E)と(F)の条件に従って、前記集団の各オリゴヌクレオチドはオリゴヌクレオチドがベースとなっている24の隣接する塩基の前記の配列から成っている。

【0077】

さらに好ましくは、(F)の条件に従って、前記集団の各オリゴヌクレオチドはオリゴヌクレオチドがベースとなっている24の隣接する塩基からなる前記配列から成っている。

【0078】

特に好ましい面は、(B)では $W = G$ と C のうちのの一つ; $X = A$ と T/U のうちのの一つ; $Y = A$ と T/U のうちのの一つである。

【0079】

30

さらに好ましいのは、(B)では: $W = G$; $X = A$ と T/U のうちのの一つ; $Y = A$ と T/U のうちのの一つである。

【0080】

他の広い面では、本発明は50のタグあるいはタグ相補体用である最小限クロスハイブリダイゼーションされた分子を含んでいる組成物のことである。ここでは各分子はヌクレオチド塩基配列から成るオリゴヌクレオチドから成っている。定義された条件下では、前記オリゴヌクレオチドと違うオリゴヌクレオチドの相補体間のハイブリダイゼーションの最高度は、前記オリゴヌクレオチドとその相補体間のハイブリダイゼーション度の約10%を超えない。

【0081】

40

このような定義された好ましい条件セットは、ハイブリダイゼーション条件が37℃の温度でpH 8.0状態で 0.2 M 塩化ナトリウムと 0.1 M トリスと0.08% トリトン X-100Aの時に得られるハイブリダイゼーションレベルと同じハイブリダイゼーションレベルを生じる。そして配列は共有結合的に微粒子に結合している。もちろん、これらの条件は好ましくは直接使われる。

【0082】

好ましくは、定義された条件セット下では条件が何であろうと、各オリゴヌクレオチドとその相補体間のハイブリダイゼーション度は1倍から8倍の間で変動する。

【0083】

好ましくは、各オリゴヌクレオチドが同じ長さで、少なくとも20ヌクレオチド塩基長で

50

あることである。さらに好ましいのは、各オリゴヌクレオチドが24ヌクレオチド塩基長であることである。

【0084】

ある態様では、相補体へのハイブリダイゼーションによって前記の分子の混合物から識別可能にするため、組成物の各分子は固相担体に結合されている。そのような分子の各々は固相担体のような定義された場所に結合される。各分子の定義された場所は他の異なった分子の定義された場所と異なる。

【0085】

ある好適な態様では、固相担体は微粒子で、各前記の分子は互いに違った前記の分子に比べて異なった微粒子に共有結合的に結合している。

10

【0086】

本発明はポリヌクレオチドの分類と識別用キットも付いている。そのようなキットには一つ以上の固相担体が付き、担体はそれぞれ一つ以上の空間的に分離した領域を持ち、各領域は実質的には同一の共有結合的に結合しているタグ相補体の一定の集合を持っている。タグ相補体は本発明のオリゴヌクレオチド集団から構成されている。

【0087】

一つ以上の固相担体は、二次元基質でもよく、そこでは一つ以上の空間的に分離した領域が複数の空間的にアドレス可能な領域である。

【0088】

タグ相補体は微粒子に結合することもできる。微粒子は5 μm から40 μm 範囲内の直径であることが望ましい。

20

【0089】

そのようなキットは、分光光度において固有でありそのため従来のラボの技術で識別できる微粒子を含むことが望ましい。もちろんそのようなキットが効果を現すためには、一般には各タイプの微粒子はそれと関連したタグ相補体を一つだけもっており、かつたいてい各タイプの微粒子と関連した（結合した）異なるオリゴヌクレオチドタグ相補体が存在する。

【0090】

本発明には本発明のオリゴヌクレオチドファミリーを使う方法が含まれている。

【0091】

30

このような方法は、核酸の座で突然変異や多型があるかをみるのに生体配列がある生体試料を解析するものである。方法は次のようになる：

(A) タグ配列に相補する配列を持った5'末端の増幅した分子を形成するため、第一プライマーがある核酸分子を増幅する。第一プライマーは5'-配列を持っており、5'-配列は本発明のタグ相補体ファミリーに属しているタグ相補体配列に相補するタグ配列を持っている。

(B) ポリメラーゼと増幅した配列の3'末端に相補する5'末端を持つ第二プライマーが存在する増幅分子を伸長する。第二プライマーの3'末端が最隣接した前記の遺伝子座に伸長し、多数のヌクレオシド三リン酸塩誘導体の存在下ではそれぞれは(i)転写の間、増大するヌクレオシド鎖の3'末端にポリメラーゼの組み入れが可能 (ii)重合の終結を引き起こす (iii) それぞれの微分検出が可能で、前記の増幅した配列の遺伝子座にありうる各ヌクレオシドに相補する前記の誘導体がある。

40

(C) 特に第二プライマーが(A)のタグ相補体配列を持ったタグ相補体にハイブリダイゼーション、そして

(D) 核酸の遺伝子座に位置する塩基を同定するため、(B)の第二プライマーに組み込まれているヌクレオチド誘導体を検出。

【0092】

本発明での他の方法は、複数の核酸分子がある生体試料は、各核酸分子の座で各核酸分子用に突然変異や多型が存在するか解析される。方法は次のようになる：

(A) タグ配列に相補する配列を持った5'末端の増幅した分子を形成するため、第一プライ

50

マーがある核酸分子を増幅する。第一プライマーは5'-配列を持っており、5'-配列は本発明のタグ相補体ファミリーに属しているタグ相補体配列に相補するタグ配列を持っている。

(B)ポリメラーゼと増幅した配列の3'末端に相補する5'末端を持つ第二プライマーが存在する増幅分子を伸長する。第二プライマーの3'末端が最隣接した前記の遺伝子座に伸長し、多数のヌクレオシド三リン酸塩誘導体の存在下ではそれぞれは(i)転写の間、増大するヌクレオシド鎖の3'末端にポリメラーゼの組み入れが可能 (ii)重合の終結を引き起こす(iii) それぞれの微分検出が可能で、前記の増幅した配列の遺伝子座にありうる各ヌクレオシドに相補する前記の誘導体がある。

(C)第二プライマーを(A)のタグ相補体配列を持つタグ相補体に特異的にハイブリダイゼーションする。

10

(D)核酸遺伝子座に位置する塩基を同定するため、(B)の第二プライマーに組み込まれたヌクレオシド誘導体を検出する。

ここでは(A)の各タグは各核酸分子にとって固有である。段階(A)と(B)は 互いの存在下で前記の核分子を使って行われる。

【0093】

他の方法には各核酸分子の座で突然変異や多型の存在を、複数の核酸分子に相補的な二本鎖がある生体試料を使い、各核酸分子用に解析することである：

(A)タグ配列に相補する配列を持った5'末端の増幅した分子を形成するため、第一プライマーがある核酸分子を増幅する。第一プライマーは5'-配列を持っており、5'-配列は本発明のタグ相補体ファミリーに属しているタグ相補体配列に相補するタグ配列を持っている。

20

(B)ポリメラーゼと増幅した配列の3'末端に相補する5'末端を持つ第二プライマーが存在する増幅分子を伸長する。第二プライマーの3'末端が最隣接した前記の遺伝子座に伸長し、多数のヌクレオシド三リン酸塩誘導体の存在下ではそれぞれは(i)転写の間、増大するヌクレオシド鎖の3'末端にポリメラーゼの組み入れが可能 (ii)重合の終結を引き起こす(iii) それぞれの微分検出が可能で、前記の増幅した配列の遺伝子座にありうる各ヌクレオシドに相補する前記の誘導体がある。

(C)第二プライマーを(A)のタグ相補体配列を持つタグ相補体に特異的にハイブリダイゼーションする。

30

(D)核酸遺伝子座に位置する塩基を同定するため、(B)の第二プライマーに組み込まれたヌクレオシド誘導体を検出する。

ここでは(A)の各タグ配列は各核酸分子にとって固有である。段階(A)と(B)は 互いの存在下で前記の核分子を使って行われる。

【0094】

他の面では、本発明は各核酸分子の座で突然変異や染色体の多型の存在を複数の核酸分子の生体試料で解析する方法である。方法は次のようになる：

(a)分子とプライマーをハイブリダイゼーションする。プライマーは5'配列を持ち、5'配列は本発明のタグ相補体ファミリーに属するタグ相補体の配列に相補するタグ配列を持っている。3'末端は最隣接の遺伝子座に伸長する。

40

(b)酵素的に複数のヌクレオシド三リン酸塩誘導体が存在するプライマーの3'末端を伸長する。それぞれは：(i)増大するヌクレオシド鎖の3'末端に酵素的に組み入れが可能 (ii)前記の伸長の終結を引き起こす(iii) 互いに微分検出が可能。そこでは、前記の遺伝子座に存在する可能性のあるヌクレオシドに相補する前記の誘導体がある。

(c)段階(b)で形成された伸長したプライマーを、(a)のタグ相補体配列を持つタグ相補体に特異的にハイブリダイゼーションする。

(d)核酸分子の遺伝子座に位置する塩基を同定するため、段階(b)のプライマーに組み込まれたヌクレオシド誘導体を検出する。

ここでは(a)の各タグは各核酸分子にとって固有であり、段階(a)と(b)は 互いの存在下で、前記の核分子を使って行われる。

50

【0095】

誘導体はジデオキシヌクレオチド三リン酸塩でもよい。

【0096】

各自の相補体は、一つ以上の固相担体にあり、空間的に分離した領域で実質的に同一の相補体の一定した集合として結合されることがある。

【0097】

各タグ相補体には各自の相補体を区別するラベルを入れることができ、段階(d)では、結合したタグとタグ相補体のハイブリダイゼーション複合体別々に異なったラベルの存在を検出することができる。

【0098】

本発明での他の面は、混合物に入っていると推測されるターゲットの存在を決定する方法である。方法は次のようになる：

(i) 第一ラベルでターゲットをラベルする。

(ii) ターゲットへの特異的な結合が可能でありかつ第一タグを含む、最初の検出部分を準備する。

(iii) 分子とターゲットの前記特異的な結合を可能にする（起こす）のに適した条件下で、混合物サンプルを検出部分に露出する。

(iv) 本発明の適切なタグ相補体ファミリーを用意する。このタグ相補体ファミリーには第一タグの配列に相補的な配列を持っている第一タグ相補体を含んでいる。

(v) 第一タグとそのタグ相補体の特異的なハイブリダイゼーションを可能にする（起こす）のに適した条件下で、サンプルをタグ相補体ファミリーに露出する。

(vi) 混合物にある前記のターゲットの存在や欠如を決定するため、最初の前記のタグ相補体にハイブリダイゼーションした前記の最初の検出部分が、前記のラベル付けされたターゲットと結合されているかを決定する。

【0099】

好ましくは、第一のタグ相補体は支持体の特定な位置で固体担体にリンクされている。段階(vi)では、前記の特定の位置での第一ラベルの検出が含まれている。

【0100】

又第一タグ相補体には第二ラベルも含まれており、段階(vi)は部分のハイブリダイゼーションされた複合体と第一タグ相補体に第一ラベルと第二ラベルの存在の検出が述べられている。

【0101】

さらにターゲットは有機分子や抗原、タンパク質、ポリペチド、抗体、核酸からの形成群から選択できる。ターゲットは抗原でもよい、又第一分子は特別その抗原の抗体ともなる。

【0102】

抗原は大抵はポリペチドかタンパク質で、ラベル付け方法は蛍光分子やジゴキシゲニン、ビオチン標識などの共役を含む。

【0103】

核酸はターゲットとして使え、レベル付け方法は蛍光分子、放射性同位元素で識別されたヌクレオチド、ジゴキシゲニン、ビオチン標識などを取り入れることができる。

【0104】

発明の詳細な説明

詳細な態様

本発明では最小限にクロスハイブリダイゼーションしている配列ファミリーを提供する。また本発明は配列ファミリーをオリゴヌクレオチド配列タグとして使って分子の複雑な混合物を分類する方法も含まれている。オリゴヌクレオチド配列タグファミリーは、分類処理の間最小限のクロスハイブリダイゼーションを行なうよう設計されている。このように配列ファミリー内のどの配列も、適切なハイブリダイゼーション状態下ではファミリーから派生した他の配列とクロスハイブリダイゼーションされない。これは当業者によって

10

20

30

40

50

も知られていることである。本発明は特に高度な分析物の並列処理に有効である。

【0105】

オリゴヌクレオチド配列タグファミリー

本発明には24merポリヌクレオチドファミリーも含まれており、これは互いに最小限にクロスハイブリダイゼーションすることが実証されている。従って、このポリヌクレオチドファミリーはタグファミリーととして有効であり、相補体もタグ相補体として有効である。

【0106】

クロスハイブリダイゼーション挙動を示さない配列ファミリーに属するオリゴヌクレオチド配列は、コンピュータープログラム（国際公開公報第01/59151号で説明）で誘導が可能である。プログラムは最小限にクロスハイブリダイゼーションしているポリヌクレオチド配列を、最大数で生成させる方法を使っている。これは次のように要約される。最初に、特定の長さの配列集団が特定のブロック元素数を基に作られる。このように、ポリヌクレオチド配列24ヌクレオチド長(24mer)のファミリーが、各元素が4ヌクレオチドから成る6ブロック元素集団から求められている場合、24merファミリーは6ブロック元素の全位置を考慮にいれながら生成される。この場合、可能な全ポリヌクレオチド配列24ヌクレオチド長を生成する6ブロック元素を組み合わせるには、 6^6 (46,656)の方法がある。

【0107】

ホモロジーの二本の配列間ではこれらの二本の配列間で要求されるホモロジー度を超えないように、制約が配列にはありまたブロックの同一性に規定があり、規定に従い形成されたポリヌクレオチド配列全部は保存されている。配列比較は接続行列を形成するため行われる。接続行列は簡単なグラフで表され、最小限にクロスハイブリダイゼーションしている適切な特質の配列は多数の小集団がある可能性のある簡単なグラフの小集団で見つけることができる。多数の適切な配列がある小集団が見つかり、配列は最小限にクロスハイブリダイゼーションしている配列集団かどうかを決定するのに試しに試験される。この方法は表Iの100非クロスハイブリダイゼーションタグを得るのに使われており、これがこの特許出願の目的でもある。

【0108】

この方法は、ブロックの説明に使われている配列群の選択において理にかなった取り組み方である。たとえば、非標準の塩基かアナログか違ったnヌクレオチドから得ることができる異なった n^4 テトラマーがある。さらに好ましい実施例では、天然ヌクレオチドが使われる時には、 4^4 か256の可能なテトラマーがあり、さらに好ましいのは、たった3塩基が使用済みのA, T, Gの時には、81の可能なテトラマーがある。一番望ましいのは、配列全部にたった一つのGがある時に、32の異なったテトラマーがあることである。

【0109】

ブロック配列は一番最適な天然塩基のサブ集団A, T, Gから構成が可能である。1塩基が欠乏しているブロックから派生された配列は、可能な二次構造形成を減少させたり、自然の核酸によるクロスハイブリダイゼーションの可能性を減少させるような有効な特性を持っている。非クロスハイブリダイゼーションタグ配列ファミリーを構成するのに一番最適なブロック配列集団は、他の集団のブロック配列のように正確な二本鎖形成にほとんど同じ安定度をあたえる。このことは定温で機能するタグ配列を与える。この例では、各配列に1つのGと3つのAかTのような全ブロック配列を一定の塩基組成物で維持することによってできる。好ましくは、ブロック配列の非クロスハイブリダイゼーション集団は定温の配列ブロックから構成されることである。ブロック配列は少なくとも1ミスマッチによって互い異なる。そのような配列選択の指導は、プライマーと、又はプローブ配列の選択方法を用いることができ、この方法は開示されている技法にある。(Robertsonら、Method s Mol Biol;98:121-54 (1998) Rychlikら、Nucleic Acids Research、17:8543-8551 (1989); Breslauerら、Proc Natl Acad Sci.、83:3746-3750 (1986)その他)。追加の配列集団は、当業者に知られている簡単な方法で、非クロスハイブリダイゼーション配列のオリジナル配列ファミリーを推定して設計できる。

【0110】

好ましい100タグファミリーは表Iの配列番号：1から100で表されており、その特徴は、相補する配列で特異な二本鎖構造を形成する配列性能を決定し、クロスハイブリダイゼーションの可能性を評価するのに使われる。100配列は合成されスライドガラスに置いた後、アミン結合によって表面に結合し、相補するタグ配列はCy3のラベルが付けられ、100配列タグファミリーを含んだアレイに個々にハイブリダイゼーションした。二本鎖の組成物は検出されアレイの各位置で定量化した。これは厳しいハイブリダイゼーション状態で有効なクロスハイブリダイゼーションが起こらなかった場合、完全マッチの二本鎖が形成されるという各タグ配列の予想されていた性能であった。ハイブリダイゼーションのサンプルの結果は表1に表されている。100プローブ全部を含んだマイクロアレイが181234プローブに相補するターゲットとハイブリダイゼーションされたとき、図1Aのハイブリダイゼーションのパターンがみられる。対スポットの4セットはターゲットに相補するプローブに対応する。図1Bは類似のアレイが全100ターゲットの配合とハイブリダイゼーションした時のパターンである。この特許の主題である配列ファミリーが非クロスハイブリダイゼーション（タグ）配列ファミリーとして使えるのがこれらの結果でわかる。

10

【0111】

100非クロスハイブリダイゼーション配列ファミリーは、24merオリゴヌクレオチドをさらに構成するために使われるテトラマー配列の追加を組み入れて拡大することができる。一つの例として四つの追加ワードは、上記の10テトラマーを使用した方法から得た配列ファミリーの非クロストーカーとして、新しい配列の発生に含有された。この場合、四つの追加ワードはYYXW (TTAG); WYYX (GTTA); YXXW (ATAG) およびWYYY (GTTT)のワード全部の可能な組み合わせとの可能なホモロジーを避けるため選択され、14の可能なワードを使った6ワードを含んだ配列合計数は、 14^6 か 7,529,536となる。4かそれ以上の類似な塩基(例えば、AAAAか TTTT) または G対のようなハイブリダイゼーション問題の可能性がある反復領域を含んだ配列を除去するため、これらの配列を選び分け、各配列をオリジナルの100非クロスハイブリダイゼーション配列のオリジナルファミリーの配列集団(配列番号:1から 100)と比較し、配列のオリジナルファミリーと15かそれ以上マッチしたホモロジーの最小閾値(挿入と欠失の使用は含まない)を含んだ新しい配列を除去した。これを言い換えれば、少なくとも15/24の最大で単一のホモロジーを得るためオリジナルの100配列の1つ以上をもつ新しい配列が整列できる場合、新しい配列がはずされた。配列対間の「単純ホモロジー」は、ここでは可能なマッチの合計数によって割った二本の整列された配列と比較してマッチ(互いに同一)しているヌクレオチド対数として定義されている。最大数のマッチしているヌクレオチド対を持つのに二本の配列が互いに整列される時、「最大の単純ホモロジー」が得られる。いづれにしても、そのようにして得た新配列集団は「候補配列」として引用されてた。候補配列のひとつは、独断的に選択され配列101として引用されている。全候補配列は配列101に対して点検され、少なくとも15の不連続的なマッチがある配列(例えば、15/24の最大単純ホモロジー(62.5%)は除去された。これは候補配列のより小さな集団を引き起こし、そこから配列102として引用されている他の配列が選択される。候補配列のより小さい集団を、少なくとも15の不連続マッチを含んだ配列を除去している配列102と比較し、そしてこの処理を候補配列が除去されるまで繰り返す。又以前に選択された他の候補体配列と連続マッチする配列が13以上ある場合、候補配列から選択されたどの配列も除去される。

20

30

40

【0112】

表Iの配列番号：1から100と比較したとき、追加のそのように得た73タグ配列集団(配列番号：101から 173)は、表Iのオリジナルの100配列タグの配列と似ていない配列から構成されている。配列の類似はクロスハイブリダイゼーションと直接関連するため(Southernら、Nat. Genet.; 21, 5-9: 1999)、オリジナルの非クロスハイブリダイゼーション配列ファミリーから派生された配列集団(配列番号:1から 173)は、配列番号：1から100の配列と類似したハイブリダイゼーション特質で挙動するのが予想される。

【0113】

50

173の24merオリゴヌクレオチドからなる集団は、次のような配列番号：174から210を持っている配列を含むために拡張した。W=G、X=A、およびY=U/Tである4merのWXYW、XYXW、WXXW、WYYW、XYXX、YXXY、YXXY、YXXYは、可能な24-塩基オリゴヌクレオチドを作成するために配列番号：1から173の作成で使用された14の4merと組み合わせて使われた。集団から除外されたのはGG、AAAA、TTTTの配列パターンを含んだものである。追加の24mer集団を含有するには、配列は少なくともWXYW (GATG)、WYXW (GTAG)、WXXW (GAAG)、およびWYYW (GTTG)のような二つのGを含む一方、6つのGを確実に含む4merのうちの一つを持っていないなければならない。又24merを含有するには、Gの隣接している各対の間に多くても6塩基がなければならない。他の方法では、2つのG間には多くても6つの非Gがあることである。そしてオリゴヌクレオチドの5'末端に最隣接する各G（表I左側）が第1から第7番目の位置（5'-末端位置を第一位置とみなす）のうちの一つを占有することが必要された。候補配列集団は、16/24またはそれ以上の最大の単純ホモロジーか以前の173オリゴヌクレオチド集団（配列番号：1から 173）以外にそれ以上のホモロジーを持った新配列を除去することによって得られた。上記のように、任意の174番目の配列が選択され、比較によって候補配列がそのすぐ後に除去された。この場合、許容された単純ホモロジーの最高度は16/24であった。2つの間に10の連続マッチがある場合、第二配列も除去された（例えば、比較された各配列と同一の10塩基配列がある偽配列の作成は想像上可能であった）。20又はそれ以上の塩基長の偽配列の作成が可能な場合、第二配列も除去された。

10

【0114】

表Iのポリヌクレオチドの特質は、二本の配列間の最大ブロックホモロジーが66 2/3 %を越えないことである。これはコンピュータアルゴリズムが最初に配列を作成したとき、そのような発生を防ぐために設計されていたからである。表Iの特定の配列ファミリーに、表Iのポリヌクレオチドが持っている実証された最小限のクロスハイブリダイゼーション集団の特質を大部分持たせながら配列を修正、付加するのは当業者の技術による。

20

【0115】

表Iには210ポリヌクレオチド配列がある。この210ポリヌクレオチドファミリー全部が最小限クロスハイブリダイゼーションする集団として互いに働きあうので、210のサブ集団である複数のポリヌクレオチドも、ポリヌクレオチドの最小限クロスハイブリダイゼーションする集団として作用することができる。例えば応用では30分子はポリヌクレオチドタグファミリーを使って識別されるので、タグ相補体は表Iの30配列のどの群も使うこともできる。これはあるサブ集団は実際には他のサブ集団よりもっと好適である、ということではない。例では、特定のサブ集団は多様な状況下において許容性があり、そのような状況下でハイブリダイゼーション度が許容度を越える前に、ハイブリダイゼーションが行われる。

30

【0116】

表Iの24塩基長より短いポリヌクレオチドの使用が望ましいかもしれない。非クロスハイブリダイゼーションに重要な2つのファミリー配列間のホモロジー特質を保持するために部分列を選択する限り、表Iの配列につきわずか10塩基しかない部分列を基にした部分列ファミリー（例えば、例示した配列のサブフレーム）を選択することができる。

【0117】

この方法によって行なわれた配列選択はコンピュータプロセスで使われる。従って例では、表IIの第1の24merの隣接する10塩基のストリングとして選択されることができる：GATTTGTATTGATTGAGATTAAAG

40

【0118】

第二の24merからの隣接する塩基のストリングは選択され、最初に選択された配列に対し最大のホモロジーと比較される：

TGATTGTAGTATGTATTGATAAAG

【0119】

系統的な対の比較は、66 2/3 %の最大ホモロジー条件が侵害されていないかを決定するのに行われる。

50

アライメント	マッチ	
GATTTGTATT	1	
ATTGATAAAG		
GATTTGTATT	0	
ATTGATAAAG		
GATTTGTATT	1	
ATTGATAAAG		
GATTTGTATT	1	
ATTGATAAAG		
GATTTGTATT	1	
ATTGATAAAG		
GATTTGTATT	1	
ATTGATAAAG		10
GATTTGTATT	3	
ATTGATAAAG		
GATTTGTATT	1	
ATTGATAAAG		
GATTTGTATT	2	
ATTGATAAAG		
GATTTGTATT	2	
ATTGATAAAG		
GATTTGTATT	5 (*)	
ATTGATAAAG		
GATTTGTATT	3	
ATTGATAAAG		
GATTTGTATT	3	
ATTGATAAAG		20
GATTTGTATT	2	
ATTGATAAAG		
GATTTGTATT	1	
ATTGATAAAG		
GATTTGTATT	1	
ATTGATAAAG		
GATTTGTATT	3	
ATTGATAAAG		
GATTTGTATT	1	
ATTGATAAAG		
GATTTGTATT	0	
ATTGATAAAG		

【0 1 2 0】 30

ここで見られるように2つの選択された部分列の間の最大のコホモロジーは50%（10の合計長さの内5つのマッチ）で、これらの2配列は互い適合性がある。

【0 1 2 1】

10merサブ配列は表Iの第三の24mer配列から選択でき、そのすぐ後適合性を決定するため対を最初の2つの10mer配列のそれぞれと比較した。このように10mer配列ファミリーが発展した。

【0 1 2 2】

10mer配列で示されたアナロジーによって11mer配列、12mer配列、13mer配列、14mer配列、15mer配列、16mer配列、17mer配列、18mer配列、19mer配列、20mer配列、21mer配列、22mer配列、および23mer配列を含有している配列ファミリーを得るのは、本発明の範囲内である。 40

【0 1 2 3】

表Iの24mer配列より長い配列がある配列ファミリーを持つことが望ましいかもしれない。表Iで示されている特定配列ファミリーを得るのは当業者の技術による。一つの可能な方法としては、より長い配列が表Iのオリジナルの非クロスハイブリダイゼーション配列の任意の2つよりは類似せず、タグ配列への余分な塩基の付加がその集団のタグ配列の熱力学特質に大きな変化を引き起こさないように、一つ以上の位置でヌクレオチドや非天然型塩基、あるいはアナログを各配列に挿入することである。例でいえば、GC 内容物は平均20%の分散で、10%~40% 間で維持されなければならない。この塩基挿入方法は40塩基長までの配列ファミリーを得るのに使用できる。 50

【0124】

表Iや表IIあるいはこれら2つの表を合わせた配列のようなタグファミリー（またはタグ相補体ファミリー）として使われる特異配列ファミリーのように、この分野での技術者は変異体ファミリーが同様な働きをすることが直ぐに分かるであろう。

【0125】

再び表Iの配列でいえば、各TはAに転換され、その逆も又同様であり、クロスハイブリダイゼーションの特質には大きな違いは見られない。各Gがa Cに変換されることも事実である。

【0126】

又ファミリーの全配列は従来法として5'-3'方向に作成されるのに使われる、又全配列の作成は反対方向(3'-5')でも使うことができる。

10

【0127】

ここで追加の修正が可能である。Cが配列ファミリーには使われていないという例がある。特定の配列の一つ以上のGの代わりにCの置換は、他のファミリーの配列と共に、少なくとも修正のために選択された特異配列と同じくらいホモロジーの低い配列を産出する。従って、表Iで示されている任意の配列において、一つ以上のGの代わりにCの置換は可能である。同じように、一つ以上のAの代わりにCの置換、または一つ以上のTの代わりにCの置換は可能である。

【0128】

特定のファミリー配列は同じか大体同じ長さであるほうがよい。好ましくは本発明の配列ファミリーの配列全部が、ファミリーの平均の塩基長が5塩基内であるとよい。さらに望ましいのは、配列全部が平均の塩基長が4塩基内であることである。さらに好ましくは、全部あるいはほとんど全部の配列がファミリーの平均塩基長が3塩基内であることである。さらによいのは、全部あるいはほとんど全部の配列がファミリーの平均の塩基長の1塩基内であることである。

20

【0129】

当業者は配列集団をこの特許の主題である配列ファミリーから配列集団を誘導することができ、又不必要なハイブリダイゼーション特質をもつ可能性のある配列を除去することもできる。

【0130】

30

オリゴヌクレオチドファミリーの合成法

本発明のオリゴヌクレオチド配列ができれば標準ホスホラミダイト合成の方法かその他同種の方法で直接合成できるのがよい (Caruthersら、Methods in Enzymology; 154, 287-313: 1987; Lipshutzら、Nature Genet.; 21, 20-24: 1999; Fodorら、Science; 251, 763-773: 1991)。ペプチド核酸か二本鎖安定に有利な修正ヌクレオチドのような非天然型塩基がある代用化学も使われる可能性がある (Haciaら、Nucleic Acids Res.; 27: 4034-4039, 1999; Nguyenら、Nucleic Acids Res.; 27, 1492-1498: 1999; Weilerら、Nucleic Acids Res.; 25, 2792-2799: 1997)。ホスホロチオネートヌクレオチドかホスホロアミドヌクレオチドのような代用ヌクレオチドバックボーンを使つての本発明のオリゴヌクレオチド配列の合成も可能である。段階を追う方法で配列のブロックの追加で合成する方法も使うことができる。(Lyttleら、Biotechniques, 19: 274-280 (1995)。合成は応用では固相担体として使われる基質に直接に行なうことも可能である。又オリゴヌクレオチドは溶液中での使用にまたは第二の担体への結合のために担体から分割されることも可能である。

40

【0131】

固相担体

本発明に使用できる異なった固相担体はいくつかある。それらはスライド、プレート、チップ、メンブラン、ビーズ、微粒子や他の同様なものであるが、それだけとは限らない。固相担体の素材はさまざま、ビニール、ガラス、シリコン、ナイロン、ポリスチレン、シリカゲル、ラテックスや他の同様なものなどである。

【0132】

50

適切な態様としては、タグ相補体配列ファミリーは固体担体への結合を可能にするため誘導体化されるなどである。固体担体の結合に使われる核酸の誘導体化方法の多くは技術的に知られている(Hermanson G., Bioconjugate Techniques; Acad. Press: 1996)。配列タグは共有結合あるいは非共有結合によって固体担体へ結合されることがある(Iannoneら、Cytometry; 39: 131-140, 2000; Matsonら、Anal. Biochem.; 224: 110-106, 1995; Proudnikovら、Anal Biochem; 259: 34-41, 1998; Zammattéoら、Analytical Biochemistry; 280:143-150, 2000)。配列タグは修正核酸を5'末端か 3'末端の位置で組み入れることによって固体担体に結合するためタイミングよく誘導体化される可能性がある。

【0133】

固体担体への結合に役立つ多様な部分（例えば、ビオチン、抗体、その他同種のもの）とそれらを核酸に結合させる方法は技術的に知られている。たとえば、アミン修正核酸塩基(Glen Research社から入手可能)は、二機能性の架橋剤(例えば、bis(sulfosuccinimidyl suberate)はPierce社から入手可能)によって固体担体に結合されることがある(Covalink-NHであるポリスチレン表面を第二級アミノ基で移植したもの、Nunc社から入手可能)。追加のスペーシング部分は捕獲部分と固体担体表面の間の立体障害を減らすのに追加できる。

【0134】

分類のための分析物へのタグの付着

オリゴヌクレオチドタグ配列ファミリーは分析物集合に共役でき、一番望ましいポリヌクレオチド配列には異なる方法があり、直接化学合成、化学共役、連結、増幅、その他同種類のものなどが含まれるがしかしそれだけとは限らない。プライマー配列で合成された配列タグは、PCR増幅のようなターゲットでのプライマーの酵素拡張として使うことができる。

【0135】

プライマー伸長を用いての単一ヌクレオチド多型の検出

病気診断、単一ヌクレオチド多型分析、遺伝子型、発現分析やその他同種類のものではないクロスハイブリダイゼーション配列が適用できる遺伝分析領域が多数ある。プライマー伸長方法(遺伝子ビットとしても知られている(Nikiforovら、Nucleic Acids Res.; 22, 4167-4175: 1994; Headら Nucleic Acids Res.; 25, 5065-5071: 1997))と呼ばれている遺伝分析法は、ゲノムDNA内の特異多型位置に位置するヌクレオチドを識別するには非常に正確な方法である。標準のプライマー伸長反応では、定義された多型位置を含んだゲノムDNA部分は、多型位置の側面にあるプライマーを使ったPCRによって増幅される。多型位置にどのヌクレオチドが存在するかを同定するには、第三プライマーは多型位置がプライマーまで直接に3'のところにあるように合成する。プライマー伸長反応は増幅DNAを持つようにセットする。伸長用のプライマーは最大4までのジデオキシヌクレオチド - 三リン酸エステルで、それぞれ異なった蛍光色素とDNAポリメラーゼ1のDNAポリメラーゼのようなDNAポリメラーゼでラベルされている。ジデオキシヌクレオチド使用では、単一塩基が多型の位置に対応する位置であるプライマー3'末端に追加されるのを確認する。この方法で、特異多型位置にあるヌクレオチドの同一性は、各反応に取り込まれている蛍光色素でラベルされたヌクレオチドの同一性によって決定可能になる。この方法での大きな欠点は処理量が低く、各プライマー伸長反応は違った管でそれぞれ行われる。

【0136】

汎用配列は次のように、プライマー伸長アッセイの処理量を増強するのに使うことができ、複数の多型位置があるゲノムDNA領域はPCRによって増幅される。多数の多型位置があるゲノムDNA領域は、PCR 反応で増幅されるか、一つ以上の多型位置があるいくつかのゲノム領域は多発性PCR 反応で増幅される。プライマー伸長反応は上記で説明されたようにおこなわれる。例外としては、プライマーが使ったものはキメラで、それぞれは固有な汎用タグが5'末端と3'末端の伸長配列にあり、この方法で各遺伝子特異配列は特定の汎用配列に関連する。キメラプライマーは増幅したDNAにハイブリダイゼーションされると、プライマー伸長は上記に説明されたように行なわれ、組み入れられたヌクレオチドの特異蛍

光色素の特質で伸長プライマーの配合プールが起こる。プライマー伸長反応の結果、混合した伸長反応はプライマーで汎用配列の逆相補体であるプローブがあるアレイにハイブリダイゼーションされ、プライマー伸長反応の多くの生成物を別々の点に分離してしまう。蛍光色素は各点にあり、特定の位置に組み入れられたヌクレオチドを同定する。

【0137】

タグ配列ファミリーを用いるキット

非クロスハイブリダイゼーション配列ファミリーはたとえば遺伝分析用キットとして使われることがあり、そのようなキットには、溶液中か固体担体上に非クロスハイブリダイゼーション配列集団が少なくとも1集団ある。できれば、配列が微粒子に結合され、適用に緩衝液と試薬が入っているとよい。試薬には酵素、ヌクレオチド、蛍光ラベル、その他の同様の種類があり、それらには特定の適用が必要であり、正しい使用法の説明書は用意される。

10

【0138】

実施例

実施例1-固形アレイにおける非クロストーク挙動の論証

表Iの非クロストークのオリゴヌクレオチドファミリーと対応する100のオリゴヌクレオチドプローブはインテグレートDNAテクノロジー(IDT、Coralville IA)によって合成され、これらのオリゴヌクレオチドは C_{18} エチレングリコールスパーサーを通して5'末端のオリゴに連結した C_6 アミノ連結基を組み入れた。これらのプローブは次のようにマイクロアレイを準備するのに使われ、プローブは150 mM $NaPO_4$ 、pH 8.5の50 μM の濃度で再懸濁された。プローブはSDDC-IIマイクロアレイ・スポッター(ESI社、Toronto Ontario, Canada)を使用し、スーパーアルデハイド・スライド(Telechem社、Sunnyvale CA)表面におとした。形成されたスポットはおおよそ直径120 μM で中心から中心の間隔が200 μM であり、それぞれのプローブは各マイクロアレイに8回おとされた。スポッティングに続き、アレイはスライド製造社の説明通り原則的に処理され、手短にアレイをPBS/EtOH (3:1)の67 mMの水素化ホウ素ナトリウムで5分間処理し、0.1% SDSで4回洗浄された。この場合アレイを沸騰しなかった。

20

【0139】

100のラベルされたオリゴヌクレオチドのターゲットもまたIDTによって合成された。これらのターゲット配列は100のプローブ配列の逆相補体に一致した。ターゲットはCy3で5'末端にラベルされた。

30

【0140】

それぞれのラベルされたCy3ターゲットオリゴヌクレオチドは、100オリゴヌクレオチドプローブのすべてを含む2つのマイクロアレイに別々にハイブリダイゼーションされ、ハイブリダイゼーションは40 μl の反応で42°Cで2時間行なわれ、10 mMのTrisHCl、50 mMのKCl、0.1%のTween 20、pH 8.3で懸濁し、40 nMのラベルされたターゲットが含まれた。これらは非クロスハイブリダイゼーション配列ファミリー性能を厳格に試験するために設計された低緊縮性ハイブリダイゼーションの状態である。ハイブリダイゼーションは清潔なカバーかけにハイブリダイゼーション溶液を沈着し、気泡を避けるために注意深くマイクロアレイのスライドをカバーかけに置いて行なった。スライドをその後反転させ、培養のため多湿な空気室に移した。ハイブリダイゼーションに続き、カバーかけは外されマイクロアレイは15分間常温でハイブリダイゼーション緩衝溶液にて洗浄され、スライドをその後簡単な遠心分離で乾燥させた。

40

【0141】

ハイブリダイゼーションされたマイクロアレイはスキャンアレイ・ライト(GSI-Lumonic s, Billerica MA)で走査し、それぞれのハイブリダイゼーションしたマイクロアレイの走査に使用したレーザーと光電子増倍管の電圧は、完璧なマッチのスポットから強度の信号を最大にするために最適化された。

【0142】

サンプルのハイブリダイゼーション結果は図1Aと1Bに書かれてある。図1Aは100プロー

50

ブすべてがあるマイクロアレイがプローブ181234に相補するターゲットとハイブリダイゼーションされた時に見られるハイブリダイゼーションパターンを表している。4組の対のスポットはターゲットに相補するプローブと対応し、図1Bは同様のアレイが100すべてのターゲットの混合とハイブリダイゼーションされた時に見られるパターンを表す。

【0143】

図1Aに示された同様の結果が試験されたすべての配列で得られ、タグ(またはタグ相補体)集団である配列番号:1から100が入っているオリゴヌクレオチドを含む分子の使用の可能性も得られた。

【0144】

実施例2-ビーズにおける配列クロストーク挙動

10

表Iの100配列固体群が、最低限クロスハイブリダイゼーションするオリゴヌクレオチドファミリーとしての使用可能性があるか試験をし、選択された100配列は試験時に配列に割り当てられた数字で表Iに各別に示している。

【0145】

試験はLuminex社、米国オースティン、テキサスより入手可能なラミネックス、ラボMAP(商標)のプラットフォームを使用し行なわれ、プローブとして使用された100配列はインテグレートDNAテクノロジー(IDT, Coralville, Iowa, U.S.A.)によってオリゴヌクレオチドとして合成された。C₆アミノリンク群を含むそれぞれのプローブはC₁₂エチレングリコールスパーサーを通してオリゴヌクレオチドの5'末端にカップリングした。C₆アミノリンク分子は、オリゴヌクレオチドを固体担体へ取りつけときに使用されるアミン群を含む6つの炭素スパーサーであり、それぞれが100プローブ配列の逆相補体配列である100オリゴヌクレオチドターゲット(プローブ相補体)もIDTで合成された。各ターゲットはその5'末端でビオチンでラベルされた。すべてのオリゴヌクレオチドは標準の脱塩手順で浄化され、濃度およそ200 μMの滅菌された蒸留水として再構成され、オリゴヌクレオチドの濃度は供給業者により提供された吸光係数を用いて分光測定で決定することができた。

20

【0146】

それぞれのプローブはLuminex¹⁰⁰プロトコルに従ってアミノリンク群でラボMAPシステムのカルボキシ化された蛍光性ミクロスフェアに結合された。それぞれのプローブ配列のミクロスフェアやビーズには、プローブが他のプローブから区別できる固有でスペクト的に識別可能な光吸収の特質がある。ストックビーズペレットは超音波処理で分散しその後渦動された。それぞれのビーズ集団でおよそ500万のミクロスフェア(400 μL)をバリア先端部を使いストック管から除去し、1.5 mLのエッペンドルフ管(USA Scientific)に加え、その後ミクロスフェアは遠心分離され、上澄みは取り除かれ、ビーズは0.2 M MES (2-(N-モルホリノ)エタンスルホン酸)(Sigma社)、pH 4.5の25 μLで再懸濁され、渦動と超音波処理が行なわれた。それぞれのプローブ(25 μL容量において)の1ナノモルをそれらと対応するビーズ集団に加え、1.0 mLの殺菌されたddH₂Oを10 mgのEDC粉末に加えることによって使用直前に準備された容量2.5 μLのEDCクロスリンカー(1-エチル-3-(3-ジメチルアミノプロピル)カルボジイメド、ヒドロクロリド(Pierce)がそれぞれのミクロスフェア集団に加えられた。ビーズ混合はその後定期的に渦動され、暗所にて常温で30分間培養された。さらに新しく準備された次のEDC溶液2.5 μL一定量がその後加えられ、また暗所で30分間培養された。2回目のEDC培養に続き1.0 mLの0.02% Tween-20 (BioShop)がそれぞれのビーズ混合に加えられ、定期的に渦動され、ミクロスフェアは遠心分離され上澄みは取り除かれ、ビーズは1.0 mLの0.1% ナトリウム・ドデシル硫酸エステル(Sigma)にて再懸濁された。ビーズは再び遠心分離され上澄みを取り除かれ、カップリングされたビーズは100 μLの0.1 M MES pH 4.5で再懸濁された。ビーズの濃度はその後ddH₂Oでそれぞれの調合液を100倍に希釈し、ヌーバウアー・ブライトライン・ヘマシトメーターで数え上げて決定され、カップリングされたビーズは個々の集団として光線から保護された2°Cから8°Cの温度で保管された。

30

40

【0147】

各々のビーズ集団での相対的オリゴヌクレオチドプローブの密度は、ビオチンddUTPで

50

末端ラベルされたターミナル・デオキシヌクレオチジル・トランスフェラーゼ(TdT)によって審査され、TdTはラベルされたddNTPで一本鎖DNAの3'末端をラベルするのに使用された。一時的にハイブリダイゼーションに使用される100ビーズ集団プールの180 μL (各ビーズ種類のおよそ4000と同等)がエッペンドルフ管の中にピペットで取られ遠心分離され、上澄みを取り除き、ビーズを1x TdT緩衝液で洗浄した後、5x TdT緩衝液、25mM CoCl_2 と1000 pmolのビオチン-16-ddUTP(すべての試剤はRoche社より購入)からなる標識化反応混合液で培養された。全反応容量は殺菌蒸留水で85.5 μL までもっていき、サンプルは暗所にて37°Cで1時間培養された後、2回目の酵素一定分量が追加され、2度目の1時間の培養が行なわれた。TdT以外のすべての組成物を含む逆の制御としてサンプルは重複して行なわれた。組み込まれていないビオチン-ddUTPを除去する為に、ビーズは200 μL のハイブリダイゼーション緩衝液で3回洗浄され、最終洗浄のあと50 μL のハイブリダイゼーション緩衝液で再懸濁された。ビオチンラベルはSA-PE(ストレプトアビジン・フィコエリトリン接合体)を使用し分光測定で検出した。ビオチンとフィコエリトリンに結合するストレプトアビジンはプローブビーズからスペクトル的に識別可能である。10mg/mLのSA-PEストックはハイブリダイゼーション緩衝液で100倍に希釈され、15 μL の希釈されたSA-PEはそれぞれの反応に直接加えられ、37°Cで15分間培養され、反応はLuminex¹⁰⁰ LabMAPで解析された。収集パラメーターは50 μL のサンプル容量を使いビーズに対し100イベント測量するようセットした。

【0148】

得られた結果は図2に示されている。これから分かる様にビーズの平均蛍光度 (Mean Fluorescent Intensity、MFI)は8.25倍の範囲で277.75から2291.08へと変動する。すべてのオリゴヌクレオチドの標識化反応が終了したと仮定して、この図はターゲット(たとえばラベルされた相補体)がプローブ配列に最大限結合された場合にこの濃度で各種のビーズが得られるシグナル強度を描写している。

【0149】

ターゲットのプローブでのクロスハイブリダイゼーションは次のように評価された。上記で説明されたように、100の異なったビーズ集団に結合された100オリゴヌクレオチドプローブは、基本ビーズ混合物を作成するのに混合され、複合反応が起こるのを可能にした。固定化ミクロスフェアプローブのプールは、その後それぞれのデオキシ化されたターゲットと個々にハイブリダイゼーションされた。このように、各ターゲットは相補的な固定化ビーズ配列で個々に特異ハイブリダイゼーションがあるか試験され、特異ハイブリダイゼーションも反応にある他の99の固定化ビーズのユニバーサル配列にあるか試験された。各ハイブリダイゼーション反応では25 μL のビーズ混合(ハイブリダイゼーション緩衝液のおよそ2500の各ビーズ集団からなる)は、96-穴 Thermowell PCRプレートの各穴に加えられ、37°Cで平衡化され、それぞれのターゲットはハイブリダイゼーション緩衝液で最終濃度0.002 fmol/mLに希釈され、50 μL の最終反応容積を得るため25 μL (50 fmol)を各穴に加えた。ハイブリダイゼーション緩衝液は0.2 M 塩化ナトリウム、0.1 M トリス、0.08% トリトン X-100、pH 8.0からなり、ハイブリダイゼーションは37°Cで30分間行なわれた。それぞれのターゲットは3連で解析され、6つのバックグランドサンプル(例えば、ターゲットなし)がそれぞれのプレートに含まれ、SA-PE接合体を上記で説明したようにレポーターとして使用した。10 mg/mLの SA-PEストックをハイブリダイゼーション緩衝液で100倍に希釈し、希釈された15 μL のSA-PEを非結合のターゲットを取り除かずそれぞれの反応に直接加え、37°Cで15分間培養し、最後にさらに、35 μL のハイブリダイゼーション緩衝液を各穴に加えた。その結果Luminex¹⁰⁰ LabMAPを解析する前に、各穴ごとの最終容量が100 μL になった。収集パラメーターは80 μL のサンプル容量を使い各ビーズに対し100イベント測定するよう設定した。

【0150】

ハイブリダイゼーションのパーセントは正味のMFIが少なくともゼロターゲットバックグラウンドの3倍でイベントが計算され、これはすなわち、計算はサンプルが $(\text{MFI}_{\text{サンプル}} - \text{MFI}_{\text{ゼロターゲットバックグラウンド}}) / \text{MFI}_{\text{ゼロターゲットバックグラウンド}} \geq 3$ である時であ

る。

【0151】

「陽性」のクロストークイベント(例えば、かなりのミスマッチまたはクロス・ハイブリダイゼーション)とは、ミスマッチしたハイブリダイゼーションで生成された正味の中央蛍光強度 (net median fluorescent intensity) ($MFI_{\text{サンプル}} - MFI_{\text{ゼロターゲットバックグラウンド}}$)が、同一の条件下で完全にマッチしたハイブリダイゼーションの10%の任意制限 (arbitrarily set limit) より大きいか同等のイベントと定義された。100のプロープと100のターゲットがあるため $100 \times 100 = 10,000$ の異なった相互作用の可能性がある、そのうちの100が完璧なハイブリダイゼーションにより、残りの9900はミスマッチしたプロープとのターゲットハイブリダイゼーションの結果である。

10

【0152】

これによって得られた結果は図3に示されてある。マッチしたプロープによってはっきりと認識されるターゲット性能がここに表されている。100ターゲットが各100プロープのプールに露出された時に起こりうる9900の非特異ハイブリダイゼーションイベントのうち、6イベントが観察され、これらの6イベントのうち、最も高い非特異イベントが完璧なマッチ対(特異ハイブリダイゼーションイベント)を観察するシグナルの10.2%に等しいシグナルを生成した。

【0153】

それぞれ100のターゲットはミクロスフェアに組み込まれた相補体配列で特異ハイブリダイゼーションがあるか個々に試験され、また他の99ターゲット配列の相補体で非特異ハイブリダイゼーションがあるかも個々に試験された。ターゲット16(プロープ16の相補体、表I)の代表的なハイブリダイゼーションの結果が図4に表されている。これでプロープ16は完璧なマッチのターゲットとだけにハイブリダイゼーションすることが判明したが、その他のどの99のターゲットによってもクロスハイブリダイゼーションは観察されなかった。

20

【0154】

これら全てに限定されないがほとんどの配列がハイブリダイゼーションによって他から区別されると想定して、前述の結果は表Iの210配列かそのサブセットの複合システムへの組み入れの可能性を実証している。すなわち、正確な相補体と他のターゲットの相補体との最低限のハイブリダイゼーションによって、ターゲットのハイブリダイゼーションでそれぞれのターゲットを他のターゲットから区別することが可能である。

30

【0155】

実施例3-ポリヌクレオチドの分類に使われるタグ配列

非クロスハイブリダイゼーション配列タグファミリーまたはサブセットは合成中にオリゴヌクレオチドプロープ配列に装着され、増幅したプロープ配列を生み出すのに使用することができる。非クロスハイブリダイゼーション配列タグでのPCR増幅の可能性と、その後の二次元アレイあるいはビーズアレイの適正な位置にそれぞれ各自の配列をアドレス指定する試験のため次の実験が考案された。24merのタグ配列が5'-3'の特異方法でp53のエクロン特異配列(20mer逆プライマー)に接続された。p53配列を接続することはヌクレオチド遺伝子配列の反対の相補体を意味する。増幅後の一本鎖DNAの次の発生を促進する為に、タグ逆プライマーが5'末端で改質リン酸塩(PO_4)によって合成され、フォワード増幅プライマー(5'-3')である適切なエクソンに対して第二のPCRプライマーも生成された。この時点でフォワードプライマーはCy3-アビジンあるいは同等のものが検出できるよう、5'-修飾ビオチンでラベルされた。

40

【0156】

前述の実践上の実施例は次のようである。ヒトp53腫瘍サプレッサー遺伝子配列のエクソン1では次のようなタグリバースプライマーが生成できた：

222087

222063

5'-PO4-GATTGTAAGATTGATAAAGTGTA-TCCAGGGAAGCGTGTACCGTCGT-3'

標識配列 # 3

エクソン1リバー

エクソン-1リバープライマーの上の数字は、示された塩基のゲノムヌクレオチドの位置を表す。

【0157】

対応するエクソン-1のフォワードプライマー配列は次のようである：

221873

221896

5'-ビオチン-TCATGGCGACTGTCCAGCTTTGTG-3'

10

これらのプライマーの組合せは214 bpプラス24 bpタグ伸長の合計サイズが238bpから成る生成物を増幅する。

【0158】

いったん増幅されると、PCR生成物はQIAクイックPCR浄化キットで浄化され、結果としてDNAを定量化することができる。一本鎖DNAを生成するには、DNAはλ-エキソヌクレアーゼ消化の影響下にあり、その結果、固相アレイに共有結合的に結合しているタグ配列に相補する一本鎖配列(反タグ)の露出を引き起こす。この結果できた生成物は95°Cで5分間過熱され、その後10~50 nMの濃度でアレイに直接加えられた。ハイブリダイゼーションと並列の分類に続き、タグエクソン1配列はCy3-ストレプトアビジンを使用し視覚化され、

20

【0159】

対立遺伝子特異ポリメラーゼ連鎖反応(ASPCR)、対立遺伝子特異プライマー伸長(ASPE)とオリゴヌクレオチドライゲーションアッセイ(OLA)などに限らず、単一ヌクレオチド多型の検出用の数々の方法は、本明細書に説明されたタグ配列と併用して当業者によって行われる。

【0160】

定義

非クロスハイブリダイゼーション：

30

完璧な相補体でないそれぞれ2本の配列間でのハイブリダイゼーションの欠如を表す。

【0161】

クロスハイブリダイゼーション：

一本鎖基質に全部ではないが部分的に相補する一本鎖DNA配列の水素結合。

【0162】

ホモロジー：

DNAの鎖の塩基配列をもとに2本あるいはそれ以上の別々のDNA鎖が互いにどれだけ密に関連しているかを表す。

【0163】

アナログ：

40

ヌクレオチド塩基に似ている化学物質である。塩基は通常DNAには現れないが、多少の構造の違いがあるがDNAに現れる塩基と置き換えることができる。

【0164】

相補体：

DNA配列の反転あるいは「鏡」イメージ。相補DNA配列はすべての「T」に「A」が、すべての「G」に「C」がある。一本鎖DNAの2本の相補鎖、例えばタグ配列とその相補体、は二本鎖分子を作り出す為に合体する。

【0165】

相補DNA(cDNA)：

メッセンジャーRNA鋳型から合成されたDNA。物理地図作成のプロープとして一本鎖型が

50

おもに使用される。

【0166】

オリゴヌクレオチド：

短いヌクレオチドポリマーを示す。ヌクレオチドは天然ヌクレオチド塩基あるいはアナログがある。

【0167】

タグ：

併用した時にクロスハイブリダイゼーションしない最低一つの他のオリゴヌクレオチドで特異的に分析物を分類するのに使用することが出来るオリゴヌクレオチドを示す。

【0168】

類似ホモロジー：

本発明の文中で、配列対は配列間の「ホモロジー」の量をもとにそれぞれを比較する。例として、同一の対の塩基の(絶対)最大数を得るために2つの配列を隣り合って並べた時、2本の配列はそれぞれ50%の「最大ホモロジー」を持つといわれる。同一の塩基数は一つの配列で合計塩基数の50%である。(もし比較される配列が異なった長さならば2つの配列の短い方の塩基合計数である)。最大ホモロジーの判定の例は次の記述のようになる。

【0169】

実施例4-最大ホモロジーの判定

* *

A-A-B-B-C-C

B-D-C-D-D-D (4対の塩基のうち2対が同じ)

* *

A-A-B-B-C-C

B-D-C-D-D-D (3対の塩基のうち2対が同じ)

【0170】

このケースでは、同一の対の塩基の最大数は2で、この最大数を生み出す2つの可能なアライメントがある。可能なペアリングの合計数は6で、 $33 \frac{1}{3} \% (2/6)$ ホモロジーを出す。2つの配列間のホモロジーの最大量はこのように1/3である。

【0171】

実施例5-最大ホモロジーの判定

* * *

A-A-B-B-C-A

A-A-D-D-C-D (6対の塩基のうち3対が同じ)

【0172】

このアライメントでは、同一の対の塩基数は3で、可能な対の塩基の合計数は6、それゆえに2つの配列間のホモロジーは3/6 (50%)である。

*

A-A-B-B-C-A

A-A-D-D-C-D (1対の塩基のうち1つが同じ)

【0173】

このアライメントでは、同一の対の塩基数は1で、それゆえに2つの配列間のホモロジーは1/6 ($16 \frac{2}{3} \%$)である。

【0174】

これら2つの配列間の最大ホモロジーはこのように50%である。

【0175】

ブロック配列：

10

20

30

40

50

ブロックの配列の記号表示を示す。最も一般的な型のブロック配列は特定値、数理変数、あるいはその他の選定が配列のそれぞれの配列ブロックに割り当てられていない表示配列である。

【0176】

インシデンスマトリックス：

ここで使用されているのは離散数学分野で明確に定義された用語である。しかしながら、インシデンスマトリックスは最初に「グラフ」を明確にしなくては定義できず、ここで説明された方法では単純グラフと呼ばれる一般グラフのサブセットが使われた。このサブカテゴリーの部分はさらに次のように定義される。

【0177】

単純グラフ G は対 (V, E) で、 V は単純グラフの頂点集合を表し、 E は単純グラフの非適応のエッジ集合 (un-oriented edges) を表す。エッジは頂点の集合メンバーの2つの成分の配合と定義される。言い換えれば、単純グラフ G では、エッジによって結ばれた頂点の対がいくつかある。本発明者らの出願では、グラフは鋳型配列とDNA配列を表す頂点と配列対の相関的な特質を表すエッジで形成された核酸の配列を基にしている。

【0178】

インシデンスマトリックスは、特定のグラフを説明する数学的オブジェクトである。ここで使用された単純グラフのサブセットでは、単純グラフ $G=(V, E)$ で、予め選択された一定の順序付けの頂点では $V=\{v_1, v_2, \dots, v_n\}$ 、インシデンスマトリックスの要素では $A(G) = [a_{ij}]$ となり、次の規則で定義されている。

(1) エッジの集合のメンバーである頂点の対 $\{v_i, v_j\}$ が $a_{ij}=1$

(2) エッジの集合のメンバーでない頂点の対 $\{v_i, v_j\}$ が $a_{ij}=0$

これはインシデンスマトリックスのまさに明確な定義である。事実、頂点のインデックス $1, 2, \dots, n$ を選択すると、頂点 v_i と v_j がエッジで結ばれている場合は $a_{ij}=1$ の要素で、頂点 v_i と v_j がエッジで結ばれていない場合は $a_{ij}=0$ の要素で、 $(n \times n)$ の正方行列を形成する。

【0179】

ここで使われた用語「クラス性質」の定義するには、用語「完全単純グラフ」と「クリーク」をまず定義しなければならない。ここで説明された方法から得た配列全部が、グラフ G のエッジを定義している配列対の相関的な性質を集团的に共有するには、完全単純グラフが必要になるからである。例えば、最終集合から選ばれたどんな配列対にも規則量より多い「最大単純ホモロジー」をもたない、という許容限度規則を破らないことである。特異ヌクレオチドあるいはヌクレオチド配列に関してすでに判明されている挙動や実験で決定された挙動を基にした補足の「局所」規則が基本的な許容限度規則と併せて配列対に適用されることがある。

【0180】

単純グラフという言葉では、 $G=(V, E)$ は最終グラフではエッジで結ばれていない頂点の対はない(対の配列はない)事を意味する。(なぜならエッジは v_i と v_j によって表される配列が許容限度規則を破らないことを意味するからである。)

【0181】

単純グラフのインシデンスマトリックスは上記の書かれている要素の定義によって作り出されることができるので、完全単純グラフを定義する結果、完全単純グラフの対応するインシデンスマトリックスは1に等しいすべての非対角要素と0に等しいすべての対角線要素があるということになる。というのは、一つの配列だけを使って位置を合わせると、すべての他の配列がエッジで結ばれているので、これは許容限度規則を破ることになる。

【0182】

どの単純グラフでも完全なサブグラフがあることがある。はじめに、グラフのサブグラフの定義は以下になる。単純グラフ $G=(V, E)$ のサブグラフ $G_s=(V_s, E_s)$ は頂点の V の集合である頂点 V_s のサブセットを含む単純グラフであり、 V の集合のなかの V_s の集合の包含は液浸 (数学用語) である。これは2段階で単純グラフ G のサブグラフ $G_s=(V_s, E_s)$ を作り出すことを意味する。はじめに G からいくつかの頂点 V_s を選択する。その後選択した頂点を

10

20

30

40

50

結ぶGからそれらのエッジEsを選択し、選択されていない頂点で結ばれるエッジは選択しない。

【0183】

完全単純グラフであるGのサブグラフが望ましい。アルゴリズム基本の鋳型によって作り出された全配列の単純グラフG、そこから作り出された完全単純グラフの性質を利用して、配列対のペアをなす性質(許容限度規則を破るもの/破らないもの)は全集合の特性に転化され、「クラス特質」と定義される。

【0184】

完全単純グラフである単純グラフGのサブグラフを選択することによって、本明細書に説明された局所規則に関与した試験によるが、許容限度の規則を破る結果集合に配列の対がなく、また上に説明されたように、独立した集合の配列対が選択されることを保証する。この特徴を「望ましいクラス特質」(desired class property)と呼ぶ。

10

【0185】

このように本発明は配列の局部ホモロジーを考慮することによって非クロスハイブリダイゼーション挙動の可能性を削減し、既知の研究法よりさらに正確さを持つと思われる。例えば、本明細書に説明された方法は付加や欠失がある配列アライメントを作り出す為一つの配列を他の配列に対して滑動することも含む (Kaneら、Nucleic Acids Res.; 28, 4552-4557: 2000)。

【0186】

(表I)

20

配列番号 (1)	配列	実施例 2 (2)で指定なし
1	GATTTGTATTGATTGAGATTAAAG	1
2	TGATTGTAGTATGTATTGATAAAG	2
3	GATTGTAAGATTTGATAAAGTGTA	3
4	GATTTGAAGATTATTGGTAATGTA	4
5	GATTGATTATTGTGATTGGAATTG	5
6	GATTTGATTGTAAAAGATTGTTGA	6
7	ATTGGTAAATTGGTAAATGAATTG	7
8	ATTGGATTGATAAAGGTAAATGA	8
9	GTAAGTAATGAATGTAAAAGGATT	
10	GATTGATTGATTGATTGATTGAT	
11	TGATGATTAAAGAAAGTGATTGAT	
12	AAAGGATTTGATTGATAAAGTGAT	
13	TGTAGATTTGTATGTATGTATGAT	10
14	GATTTGATAAAGAAAGGATTGATT	
15	GATTAAAGTGATTGATGATTGTA	11
16	AAAGAAAGAAAGAAAGAAAGTGTA	12
17	TGTAAAAGGATTGATTTGTATGTA	
18	AAAGTGTAGATTGATTAAAGAAAG	
19	AAAGTTGATTGATTGAAAAGGTAT	
20	TTGATTGAGATTGATTTTGAGTAT	
21	TGAATTGATGAATGAATGAAGTAT	15
22	GTAATGAAGTATGTATGTAAAGTAA	
23	TGATGATTTGAATGAAGATTGATT	16
24	TGATAAAGTGATAAAGGATTAAAG	17
25	TGATTTGAGTATTTGAGATTTTGA	18
26	TGTAGTAAGATTGATTAAAGGTAA	
27	GTATAAAGGATTGATTTTGAAAAG	
28	GTATTTGAGTAAGTAATTGATTGA	19
29	GTAAAAAGTTGAGTATTGAAAAAG	
30	GATTTGATAAAGGATTTGTATTGA	
31	GATTGTATTGAAGTATTGTAAAAG	20
32	TGATGATTTTGATGAAAAGTTGA	

30

40

配列番号 (1)	配列	実施例 2 (2) で指定なし	
33	TGATTTGAGATTAAAGAAAGGATT	21	
34	TGATTGAATTGAGTAAAAAGGATT	22	
35	AAAGTGTAAGGATTTGATGTAT		
36	AAAGGTATTTGAGATTTGATTGAA		
37	AAAGTTGAGATTTGAATGATTGAA	23	
38	TGTATTGAAAAGGTATGATTGAA		
39	GTATTGTATTGAAAAGGTAATTGA	24	
40	TTGAGTAATGATAAAGTGAAGATT		
41	TGAAGATTTGAAGTAATTGAAAAG	25	
42	TGAAAAAGGTAGATTTTGAGTAA	26	
43	TGTATGAATGAAGATTTGATTGTA		10
44	AAAGTTGAGTATTGATTGAAAAG	27	
45	GATTTGTAGATTTGTATTGAGATT		
46	AAAGAAAGGATTTGTAGTAAGATT	29	
47	GTAAAAAGAAAGGTATAAAGGTAA	30	
48	GATTAAAGTTGATTGAAAAGTGAA	31	
49	TGAAAAAGGTAATTGATGTATGAA		
50	AAAGGATTAAGTGAAGTAATTGA	33	
51	ATGAATTGGTATGTATATGAATGA	34	
52	TGAAATGAATGAATGATGAAATTG	35	
53	ATTGATTGTGAATGAAATGAATTG	36	
54	ATTGAAAGATGAAAAGATGAAAAG	37	
55	ATTGTTGAAAAGTGTAAATGATTGA	38	
56	ATGATGTAATGAAAAGATTGTGTA	39	
57	AAAGATTGAAAGATGATGTAATTG		
58	ATTGATGAGTATATTGTGTAGTAA	41	20
59	AAAGATTGTGTAATTGATGATGAA		
60	AAAGGTATATTGTGTAATGAGTAA		
61	TGTAATGAGTATTGTAATTGAAAG	43	
62	GTATAAAGAAAGATTGGTAAATGA	44	
63	TTGAGTAATTGAATTGTGAAATGA	45	
64	TGTATTGAATGAATTGTTGATGTA	46	
65	TGTAATTGGTAAATGAGTAAAAAG		
66	TGAATGAAATTGATGAGTATAAAG		
67	GTAAGTAAATTGAAAGATTGATGA	49	
68	GTAAATGATGATATTGGTATATTG	50	
69	ATTGTTGATGATTGATTGAAATGA	51	
70	ATTGTGAAGTATAAAGATGATTGA	52	
71	ATGAAAAGTTGAGTAAATTGTGAT		
72	ATGAATTGAAAGTGATTGAAAAAG	54	30
73	GTAAATTGATGAAAAGTTGATGAT		
74	AAAGTGATGTATATGAGTAAATTG	56	
75	GTAATGATAAAGATGATGATATTG	57	
76	TTGAAAAGATTGGTAATGATATGA		
77	AAAGTGAAAAAGATTGATTGATGA	59	
78	ATTGATGAGATTGATTATTGTGTA		
79	ATGAGATTATTGGATTGTAGATT	60	
80	TGAAGATTATGAATTGGTAAGATT	61	
81	ATTGGATTATGAGATTATGATTGA	62	
82	ATTGTTGAATTGGATTAAAGATGA		
83	AAAGATGAGTAAGTAAATTGGATT		
84	AAAGGTAAGATTATTGATGAAAAG	65	
85	ATTGATGAGATTAAAGTTGAATTG		
86	GATTATTGGATTATGAAAAGGATT		40
87	GATTTGTAATTGTTGAGTAAATGA	67	
88	AAAGAAAGATTGTTGAGATTATGA	68	
89	GTATAAAGGATTTTGAATTGATGA		

配列番号 (1)	配列	実施例 2 (2) で指定なし	
90	TTGAGATTGTAAATGAATTGTTGA		
91	GTATATTGATTGTGTAATGAAAAG		
92	TGATATGAATTGGATTATTGGTAT	70	
93	ATGAATGATGAATGATGATTATTG		
94	ATGAATTGATTGGATTGTAATGAT	71	
95	GATTGTAATTGAGTAAATTGATGA		
96	GATTATTGGATTAAAGGTAAATGA	72	
97	ATTGTTGAATTGATGAGATTTGAT	73	
98	GATTATGAGTAAATTGATTGTGAT		
99	GATTATTGTTGATGAATGATATTG		
100	TGTAAAAGATTGAAAGGTATGATT	75	10
101	GTATTTAGATGAGTTTGTAGATT	76	
102	TGAAGTTATGTAATAGAAAGTGAT		
103	GTATGTATTGTATGTAGTTAATTG	77	
104	TGATATAGATAGTTAGATAGATAG	78	
105	ATGATGATGTATTGTAGTTATGAA	79	
106	TTAGTGAATGTATTAGTTGATGTA		
107	GTTAGTTAGATTATTGTTAGTTAG	80	
108	GTTAATTGTGTAGTTTGTATTGTA		
109	GTTATGAAATAGTGATATTGTTAG		
110	ATTGTTAGAAAGTGATAGATTAAAG	81	
111	ATGAGTATGTTATTAGTGTATGTA	82	
112	TGTAATAGTGAAGTTAGATTGTAT	83	
113	ATTGATAGATGATTAGTTAGTTGA	84	
114	ATGAGTTTGTATTATGAGATTAAAG		20
115	TGATGTTTGATTATGATGTAGTAT	85	
116	ATGAGTTAGTTATGAATTAGATGA		
117	ATTGTTAGTGATGTTAGTAAATTAG	86	
118	TGATGTAAGTATTGATGTTAGTTT	87	
119	GATTGTAAATAGAAAGTGAAGTAA	88	
120	ATTGTGTATGAAGTATTGTATGAT		
121	ATAGTGATGTTATGAAGATTGTTA		
122	TTAGATGAATTGTGAAGTATTTAG	90	
123	GTAAGTTATGATTGATGTTATGAA	91	
124	GTATTGATGTTTAAAGTGTAATAG	92	
125	GATTGTAAAGTAAGATTGTATATTG		
126	GTTTGTATTTAGATGAATAGAAAG	93	
127	GTTTGAATTTGTAATAGTGATTGTA		
128	TGTATGTAGTATTTAGAAAGATGA		30
129	ATGAATTGTGATAAAGAAAGTTAG		
130	TTAGTGTAGTAAGTTTAAAGTGTA	95	
131	GTATGATTGTTTGTAAATTAGTGAT		
132	GTTTAAAGTTAGTTGAGTTAGTAT	96	
133	ATAGTGTATGTAGATTATGAGATT	97	
134	TTGAATGATTAGTTGAGTATGATT	98	
135	GTATGTAAGTTAGTATGATTTGAA		
136	TGTAGTATATTGTTGAATTGTGAT		
137	ATAGTGATTGTATGTATGATAAAG		
138	TTAGTGATTGATGTATATTGAAAG		
139	GTAAGATTATGAGTTATGATGTAA		
140	GTTATGAAATTGTTAGTGTAGATT	99	
141	GTTAGATTGTAGTTTAAAGATAG	100	
142	TTAGTGATTGAAATGATGTAGATT		
143	AAAGTGTAGTTATTAGTTAGTTAG		40
144	AAAGAAAGTGATGATGTTATTAG		
145	GATTGTATATTGTGTATGATGATT		
146	TTGAGATTGTTATGATATGAGTAT		

配列番号 (1)

配列

実施例 2 (2) で指定なし

147	ATGAGTATGATTGTTATGATGTTT	
148	TGATTTAGTGAAATTGTGTATTAG	
149	TGAATGTATGTAGTATGTTTGTTA	
150	GTTAGTATTGATGATTATGAGTTA	
151	GTATATTGTGATTTAGTTGAGATT	
152	GTTAGTTTAAAGTTGAGATTGTTT	
153	GTATATTGTTAGATGAGATTTGTA	
154	TGATGTATGTTAGTTTATGAATGA	
155	TGTAGTATGTAATGTAGTATTTGA	
156	ATGAGTTATGTATTGAGTTAGTAT	
157	TGTATGATGATTATAGTTGAGTAA	10
158	ATTGATGAATGAGTTTGTATAAAG	
159	TTGAGTTTATGATTAGAAAGAAAG	
160	TGATATTGATGAGTTAGTATTGAA	
161	ATAGAAAGTGAAATGAGTATGTTA	
162	TTGATGTAGATTGATGTATATAG	
163	TTGAGATTATAGTGTAGTTTATAG	
164	TGATGTTAGATTGTTTGATTATTG	
165	TGTATTAGATAGTGATTTGAATGA	
166	GATTATGATGAATGTAGTATGTAA	
167	TGAATGATTGATATGAATAGTGTA	
168	GTAATGATTTAGTGTATTGAGTTT	
169	TGTAGTAATGATTGATGATAAAG	
170	TGAAGATTGTTATTAGTGATATTG	
171	GTATTTGAATGATGTAATAGTGTA	20
172	GTATATGATGTATTAGATTGAAAG	
173	AAAGTTAGATTGAAAGTGATAAAG	
174	GTAAGATGTTGATATAGAAGATTA	9
175	TAATATGAGATGAAAGTGAATTAG	
176	TTAGTGAAGAAGTATAGTTTATTG	13
177	GTAGTTGAGAAGATAGTAATTAAT	
178	ATGAGATGATATTTGAGAAGTAAT	
179	GATGTGAAGAAGATGAATATATAT	
180	AAAGTATAGTAAGATGTATAGTAG	14
181	GAAGTAATATGAGTAGTTGAATAT	
182	TTGATAATGTTTGTGTTTGTGTTAG	28
183	TGAAGAAGAAAGTATAATGATGAA	
184	GTAGATTAGTTTGAAGTGAATAAT	32
185	TATAGTAGTGAAGATGATATATGA	
186	TATAATGAGTTGTTAGATATGTTG	30
187	GTTGTGAAATTAGATGTGAAATAT	
188	TAATGTTGTGAATAATGTAGAAAG	40
189	GTTTATAGTGAAATATGAAGATAG	42
190	ATTATGAAGTAAGTTAATGAGAAG	47
191	GATGAAAGTAATGTTTATTGTGAA	
192	ATTATTGAGATGTGAAGTTTGTGTT	48
193	TGTAGAAGATGAGATGTATAATTA	53
194	TAATTTGAGTTGTGTATATAGTAG	
195	TGATATTAGTAAGAAGTTGAATAG	
196	GTTAGTTATTGAGAAGTGTATATA	55
197	GTAGTAATGTTAATGAATTAGTAG	58
198	GTTTGTGTTGATGTGATTGAATAAT	
199	GTAAGTAGTAATTTGAATATGTAG	64
200	GTTTGAAGATATGTTTGAAGTATA	40
201	ATGATAATTGAAGATGTAATGTTG	
202	GTAGATAGTATAGTTGTAATGTTA	66
203	GATGTGAATGTAATATGTTTATAG	69

配列番号 (1)	配列	実施例 2 (2) で指定なし
204	TGAAATTAGTTTGTAAGATGTGTA	74
205	TGTAGTATAAAGTATATGAAGTAG	63
206	ATATGTTGTTGAGTTGATAGTATA	89
207	ATTATTGAGTAGAAAGATAGAAAG	94
208	GTTGTTGAATATTGAATATAGTTG	
209	ATGAGAAGTTAGTAATGTAAATAG	
210	TGAAATGAGAAGATTAATGAGTTT	

(1) 配列番号：1～100を有するオリゴヌクレオチドは実施例1の実験で使用される。

(2) 実施例2の実験で使用されるオリゴヌクレオチドは、実験内で指定された数字によってこの縦列に示されている。 10

【0187】

本明細書において引用した全ての文献は、参照として本明細書に組み入れられる。

【0188】

本明細書に説明された本発明の保護要求範囲は、添付の特許請求の範囲によって定義される。また上記が特許請求の範囲で列挙された要素は、どの特許請求の範囲の要素とも組み合わせることができる。特に、従属請求項の範囲の要素は、基となる請求項の任意の要素、または他の互換性がある任意の要素とも組み合わせることができる。

【0189】

本出願は2001年1月25日と2001年7月10日に出願された、米国特許仮出願第60/263,710号と60/303,799号からの優先権を主張する。これらの両文書は参照として本明細書に組み入れられる。 20

【図面の簡単な説明】

【0190】

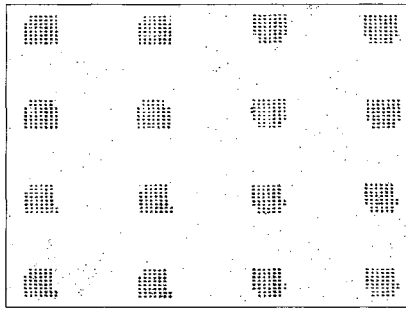
【図1】図1Aおよび1Bは、実施例1に記載されたクロスハイブリダイゼーション実験において得られた結果を例示する。図1Aでは100プローブ(配列番号：1から100)全部が入ったマイクロアレイが、配列番号：3(ターゲット)に相補する配列を持った24merオリゴヌクレオチドとハイブリダイゼーションされた時に見出されたハイブリダイゼーションパターンが示されている。1Bは類似のアレイが100ターゲット全部の混合物とハイブリダイゼーションされたときに観察されるパターンを示している。その例では、オリゴヌクレオチドが配列番号：1から100に相補する配列を持っているなどがある。 30

【図2】完全にマッチした各配列(表Iに示されている)と実施例2で説明されたようにして得た相補体のシグナル(MFI)の強度を示している。

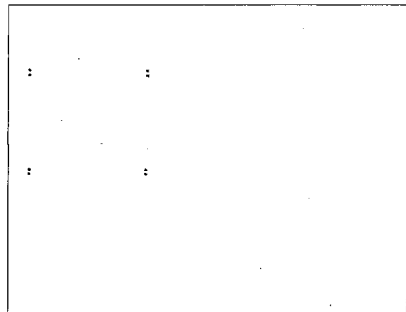
【図3】実施例2で説明された表2の配列を観察したクロスハイブリダイゼーションの三次元表現である。図2での結果は、図面の対角線に沿って再生されている。

【図4】実施例2の100プローブに露出した時、各ターゲット(配列番号：23、ターゲット番号16)について得た結果の例である。各ピーズのMFIはプロットされている。

【図 1】

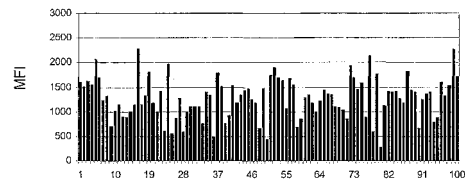


b

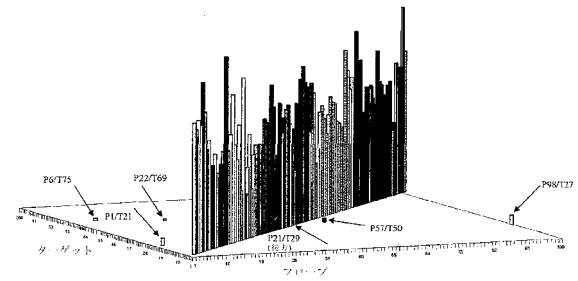


c

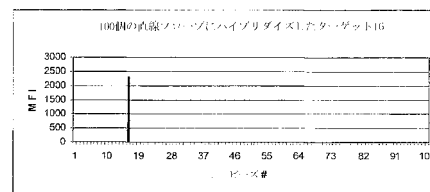
【図 2】



【図 3】



【図 4】



【配列表】

0004588976000001.app

フロントページの続き

- (74)代理人 100142929
弁理士 井上 隆一
- (74)代理人 100148699
弁理士 佐藤 利光
- (74)代理人 100128048
弁理士 新見 浩一
- (74)代理人 100129506
弁理士 小林 智彦
- (74)代理人 100130845
弁理士 渡邊 伸一
- (74)代理人 100114340
弁理士 大関 雅人
- (74)代理人 100114889
弁理士 五十嵐 義弘
- (74)代理人 100121072
弁理士 川本 和弥
- (72)発明者 パンコスカ ペトル
アメリカ合衆国 イリノイ州 エバンストン ヒンマン アベニュー #2シー 901
- (72)発明者 ジャノタ ビット
チェコ共和国 プラハ 7 オベネッカ 27
- (72)発明者 ベナイト アルバート エス.
アメリカ合衆国 イリノイ州 ショウンバーグ バリー ビュー ドライブ 1630
- (72)発明者 ブロック リチャード エス.
アメリカ合衆国 イリノイ州 シカゴ ノース レイク ショア ドライブ 3500
- (72)発明者 リッセリー ピーター ブイ.
アメリカ合衆国 イリノイ州 ティンリー パーク リチャーズ ドライブ 16830
- (72)発明者 ケプラー ダニエル
カナダ国 オンタリオ州 トロント アpartment 1102 ウッド ストリート 33
- (72)発明者 フィールドハウス ダニエル
カナダ国 オンタリオ州 ボルトン チャップリン コート 7

審査官 石丸 聡

- (56)参考文献 国際公開第00/058516 (WO, A1)
欧州特許出願公開第00799897 (EP, A1)
国際公開第96/012014 (WO, A1)
国際公開第97/031256 (WO, A1)
特表2000-510337 (JP, A)
J. Mol. Biol., vol. 292, pages 251-262 (1999)