(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2004-501669
(P2004-501669A)
(43)公表日　平成16年1月22日(2004.1.22)

(51)Int.Cl.[7]	F I		テーマコード（参考）
C12Q 1/68	C12Q 1/68	ZNAZ	4B029
C12M 1/00	C12M 1/00	A	4B063
G06F 17/30	G06F 17/30	170F	5B075

審査請求 未請求　予備審査請求 有　（全 38 頁）

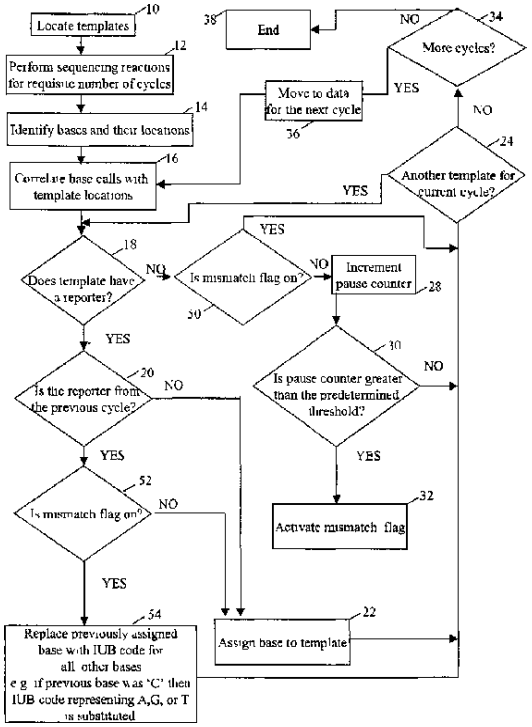
(21)出願番号　特願2002-507300(P2002-507300)
(86)(22)出願日　平成13年7月2日(2001.7.2)
(85)翻訳文提出日　平成15年1月6日(2003.1.6)
(86)国際出願番号　PCT/GB2001/002985
(87)国際公開番号　WO2002/003305
(87)国際公開日　平成14年1月10日(2002.1.10)
(31)優先権主張番号　0016472.3
(32)優先日　平成12年7月5日(2000.7.5)
(33)優先権主張国　イギリス(GB)

(71)出願人　398048914
アマシャム　バイオサイエンス　ユーケイ　リミテッド
イギリス国　エイチ　ピー　７　９　エヌ　エイ　バッキンガムシャー　リトル　チョーフォント　アメルシャム　プレイス　（無番地）
(74)代理人　100062144
弁理士　青山　葆
(74)代理人　100086405
弁理士　河宮　治
(74)代理人　100067035
弁理士　岩崎　光隆

最終頁に続く

(54)【発明の名称】シークエンシング方法および装置

(57)【要約】
塩基付加を用いて未知のヌクレオチド配列を同定する方法において、塩基の配列を鋳型から得て、配列中の塩基を未知塩基として確認し、“未知”インジケーターを配列中に含み、そして未知塩基インジケーターを含有する出力配列を作製する。レポーターの評価およびそれに従って塩基を帰属することにより塩基の配列を鋳型から得る。レポーターが塩基決定の先行するサイクルからであるかどうかを決定し、そして、もしレポーターが塩基決定の先行するサイクルからであれば、この塩基帰属は廃棄される。



【特許請求の範囲】
【請求項１】
塩基付加を用いる未知のヌクレオチド配列を同定する方法であって：
鋳型から塩基の配列を得る；
配列中の塩基を未知塩基として確認しかつ配列中に‘未知’インジケーターを含む；そして
未知塩基のインジケーターを含む出力配列を与える
ステップを含む方法。
【請求項２】
連続した未知塩基の数を計数し、そして連続した未知塩基の数が予め決めた閾値を超えるときに指示を与えることをさらに含む、請求項１に記載の方法。
【請求項３】
予め決めた閾値を超えるときに、先行する塩基を組込みミスしたとしてマークする、請求項２に記載の方法。
【請求項４】
出力配列および参照配列の間に配列アラインメントのステップをさらに含む、請求項１に記載の方法。
【請求項５】
出力配列および参照配列の間に配列アラインメントのステップをさらに含む、請求項２に記載の方法。
【請求項６】
配列をレポーターの評価により決定し、そしてこのレポーターが塩基決定の先行するサイクルからであるかどうかを決定することをさらに含む、請求項１に記載の方法。
【請求項７】
未知のヌクレオチド配列を同定する方法であって：
レポーターの評価により鋳型から塩基の配列を得てそしてそれに従って塩基を帰属する；
このレポーターが塩基決定の先行するサイクルからであるかどうかを決定し；もしこのレポーターが塩基決定の先行するサイクルからであれば、塩基帰属を廃棄し；そして
出力配列を与える
ステップを含む方法。
【請求項８】
塩基付加を用いる未知のヌクレオチド配列を同定するための装置であって：
鋳型から塩基の配列を得るための手段；
配列中の塩基を未知塩基として確認しかつ配列中に‘未知’インジケーターを含むための手段；そして
未知塩基のインジケーターを含む出力配列を与えるための手段
を含む装置。
【請求項９】
塩基付加を用いる未知のヌクレオチド配列を同定するための装置であって：
レポーターの評価により鋳型から塩基の配列を得てそしてそれに従って塩基を帰属するための手段；
このレポーターが塩基決定の先行するサイクルからであるかどうかを決定し、そしてもしこのレポーターが塩基決定の先行するサイクルからであれば、塩基帰属は廃棄されるための手段；そして
出力配列を与えるための手段
を含む装置。
【請求項１０】
塩基伸長を用いる未知のヌクレオチド配列を同定するためのコンピュータプログラム製品であって、コンピュータにロードしたときに、以下の：
鋳型から塩基の配列を得る；

配列中の塩基を未知塩基として確認しかつ配列中に‘未知’インジケーターを含む；そして
未知塩基のインジケーターを含む出力配列を与える
ステップを実施するようにコンピュータを制御するだろうコンピュータプログラム製品。
【請求項１１】
コンピュータプログラムであって、該プログラムをコンピュータ上でランするとき、請求項１～７のいずれか１項の全てのステップを実施するためのプログラムコード手段
を含むコンピュータプログラム。
【発明の詳細な説明】
【０００１】
（技術分野）
本発明は、個別の分子のシークエンシングの間にエラー訂正を可能にする、シークエンシング方法および装置に関する。
【０００２】
（背景技術）
Ｓａｎｇｅｒ，Ｆ．，Ｓ．Ｎｉｃｋｌｅｎ，ａｎｄ Ａ．Ｃｏｕｌｓｏｎ（Ｐｒｏｃ Ｎａｔｌ Ａｃａｄ Ｓｃｉ ＵＳＡ，１９７７．７４（１２）；ｐ．５４６３－７）により本質的に記載されているように、シークエンシングは鎖終結およびゲル分離の方法により日常的に実施される。この方法は、配列中のそれぞれの塩基における終結を表すＤＮＡ断片の混合集団の作製に依存する。それから、これらの断片の電気泳動分離により配列を決定する。
【０００３】
シークエンシングのスループットを増加させる最近の努力は、電気泳動分離ステップを除外する代替法の開発をもたらしている。数多くのこれらの方法は塩基伸長（即ち塩基付加）を利用しそして例えばＷＯ ９３／２１３４０、ＵＳ ５，３０２，５０９およびＵＳ ５，５４７，８３９に記載されている。これらの方法においては、シークエンシング用の試薬を作用させる前に鋳型もしくはプライマーを固体表面上に固定化する。固定化した分子をヌクレオチド類似体の存在下でインキュベートするが、これらは、その位置で水酸基を可逆的に保護する、糖残渣の３’炭素において修飾を有する。このような修飾したヌクレオチドのポリメラーゼによる組込みは、塩基伸長のそれぞれのサイクル間には唯一個のヌクレオチドが付加されることを保証する。それから、付加した塩基は、３’保護基中に組込んでいる標識によって検出される。検出に引き続いて、典型的には光化学的手段によって、保護基を除去して（即ち“切断して”）、次のサイクル間で塩基付加に利用できる遊離水酸基を露出させる。
【０００４】
一般的に、非－分離に基づくアプローチは、与えられた標的からコンセンサス配列を作成すべきそれぞれの標的配列に対する多数の鋳型分子の存在に依存する。かくして、例えば、核酸の離散的スポットを尋問することにより、塩基伸長反応を多重鋳型に応用し得るが、それぞれは、空間的にアドレス可能なアレイに固定化された、多重度の分子を含む。
【０００５】
しかしながら、ターミネーターの組込み／切断の反応、もしくは塩基切除はエラーを起こしやすい。例えば、上述のように、塩基伸長ストラテジーはヌクレオチド類似体を一般的に利用しているが、これらは、レポーター分子、通常蛍光体、の機能を糖部分上の３’位を占めるターミネーターのそれと結合させる。この基とその位置のかさ高い性質は、これらの化合物をポリメラーゼに対して高度に非効果的な基質にさせる。加えて、次の付加を可能にするためのターミネーター基の切断はまた非効率性を招きやすい。それぞれの標的に対して数千の、もしくは好ましくは数百万の、分子の存在下では、５％以下の少なめのエラーでさえも、少数のサイクル内において、それぞれの分子を表す多重度の鎖の間で、同調性の累積的な損失をもたらす。かくして、それぞれのサイクルのシークエンシングごとに、バックグランドノイズが累進的に増加して、それぞれの付加ごとでシグナルの必然

的な劣化になる。これは、入手し得る配列データを持つ塩基数は、特定のシグナルがバックグランドから区別できなくなる前に、限られることを意味する。

【０００６】

単一分子検出の方法における最近の進歩は（例えば、Ｔｒａｂｅｓｉｎｇｅｒ，Ｗ．，ｅｔ ａｌ．，Ａｎａｌ Ｃｈｅｍ．，１９９９．７１（１）；ｐ．２７９－８３およびＷＯ ００／０６７７０に記載されている）、シークエンシングストラテジーを単一の分子へ応用することを可能にする。しかしながら、シークエンシングは、分子のクローン集団に応用すると、他の分子が未修飾のままでいる間に幾らかな分子が反応を受けることになる、確率的プロセスである。かくして、従来のシークエンシング方法においては、組込みミスのようなエラーは、存在する多数の分子がコンセンサスシグナルを得ることを保証するので、普通には重大な意義をもたない。これらの反応を単一分子に応用するときには、結果は効率的に量子化される。

【０００７】

そのような単一分子シークエンシング方法は塩基切除に基づいていて、例えば、Ｈａｗｋｉｎｓ，Ｇ．ａｎｄ Ｌ．Ｈｏｆｆｍａｎ，Ｎａｔｕｒｅ Ｂｉｏｔｅｃｈｎｏｌｏｇｙ，１９９７．ｖｏｌ．１５；ｐ．８０３－８０４およびＵＳ ５，６７４，７４３に記載されている。このストラテジーでは、各塩基が適当なレポーターで標識化されるように単一の鋳型分子を作製する。この鋳型分子をエキソヌクレアーゼで消化しそして切除した塩基をモニターして同定する。これらの方法は、ラムダエキソヌクレアーゼのような高度処理性の酵素を使用するので、長さが数キロ塩基の大きい鋳型を分析する潜在性がある。しかしながら、それぞれの鋳型分子から切除した塩基をリアルタイムで連続的にモニターすることは、並行して分析できる分子の数を制限する。加えて、切除した塩基を生来の光学的もしくは化学的性質に基づいて検出できるように各塩基を適当なレポーターで標識化する場合には、鋳型を作成することが困難となる。

【０００８】

塩基切除に基づく方法（ＢＡＳＳのような）はまた単一分子アプローチに適応されている。

しかしながら、これらの技法はエラーを起こしやすい。特に、修飾したヌクレオチドの組込みは、例えば、修飾したヌクレオチドとのポリメラーゼ作用の減少した効率の結果として、失敗し得る。レポーター分子が蛍光性分子である場合には、蛍光体が無くなったり、損傷されたり、漂白されたり、もしくは切除されなかったりするために、エラーが蛍光の失敗によってまた起こり得る。単一分子レベルでは、これらのような失敗は適切な配列を得る際に失敗をもたらす。

【０００９】

本発明の一つの目的は、エラーの検出を可能にするシークエンシング方法を提供することである。本発明のさらなる目的は、個別の分子の行き先をシークエンシング反応によりモニターすることによって、分析およびエラーの防止、もしくは訂正、を可能にすることである。

【００１０】

（発明の概要）

種々の態様におけるこの発明は上記の個別な請求項で規定されるが、それらについての参照をこれから行わねばならない。有利な特色は付随する請求項に示されている。

【００１１】

簡潔に言えば、ヌクレオチド配列を分析する方法の形態を取るこの発明の一つの好ましい実施態様において、塩基の配列を鋳型から取得し、そして配列中の塩基を未知塩基として確認する。‘未知’インジケーターは未知塩基に相当する位置において配列中に含まれ、そして未知塩基インジケーターを含む出力配列が作製される。この好ましい実施態様において、レポーターの評価およびそれに従って塩基を帰属することにより塩基の配列を鋳型から得る。レポーターが塩基決定の先行するサイクルからであるかどうかを決定し、そして、もしレポーターが塩基決定の先行するサイクルからであれば、この塩基帰属は廃棄さ

れる。
分析すべきヌクレオチド配列はＲＮＡもしくはＤＮＡ配列であってもよい。
【００１２】
（図面の簡単な説明）
ここで、付随する図面を参照して実施例によりこの発明を詳細に説明するが、そこで図１は、核酸分子のような生物分子の配列を決定するための反応の間に得られるデータを解析する方法を図示しかつこの発明の好ましい実施態様を形成する工程図である。
【００１３】
（好ましい実施態様の詳細な説明）
図１は、鋳型から配列情報を得る方法を例示する流れ図を示す。この方法は、（ａ）先行するサイクルから持ち越される塩基を確認しそして（ｂ）塩基の標識失敗もしくは組込みミスから起こり得る休止分子を検出することによりエラーを追及する。このデータ解析法は、以下のように実施する標準的なシークエンシング反応を使用する。第一に、配列データを必要とする核酸分子、鋳型、を顕微鏡のスライドのような固体表面に結合させる。スライドを、例えば蛍光顕微鏡スキャナーにより観察するときにその位置を決定できるように、鋳型を標識化することができる。鋳型の配列における最初の塩基もしくはヌクレオチド、即ちＡ、Ｃ、Ｇもしくは Ｔ、を蛍光的に標識化した塩基もしくは塩基を表す標識を付加する化学反応により検索する。これは、Ａ、Ｃ、Ｇもしくは Ｔのいずれか一つであるか、または四つの異なる区別可能な標識で標識化されたそれら四個全部であり得る。鋳型中の第一の塩基は既知の様式で相補的な塩基に結合するであろう；即ち、ＡはＴと結合し、そしてＣはＧと結合し、かつ逆の場合も同様である。鋳型をポリメラーゼ酵素で伸長するかもしくは標識化したオリゴヌクレオチドをリガーゼで連結することにより、塩基組込みを引き起こすことができる。標識化した塩基の組込みを検出し、そしてその同一性を決定する。それから、その塩基から標識を除去する。それから、このシリーズのステップを鋳型中の次に続く塩基に対して繰り返す。
【００１４】
塩基の付加／組込みを伴う適切な標準的シークエンシング反応としては、ＷＯ　９３／２１３４０、ＵＳ　５，３０２，５０９およびＵＳ　５，５４７，８３９に記載されているような塩基伸長反応ならびにＵＳ　５，７６３，１７５、ＵＳ　５，５９９，６７５、ＵＳ　５，８５６，０９３およびＵＳ　５，７１５，３３０に記載されているような技法が挙げられるが、そこでは連続するラウンドのシークエンシングは、引き続く塩基の組込みに先立って鋳型の塩基切除を伴う。
【００１５】
このシークエンシング反応を実施するときには、エラーが起こり得る。例えば、（ｉ）塩基が間違って組込まれ得る、即ち組込みミスされる、または（ｉｉ）次のサイクルを実施する前に一つのサイクルから標識を除去することに失敗し得る、または（ｉｉｉ）どれか一つのサイクルにおいて塩基の組込みが失敗し得ることである。説明されるこの発明の好ましい実施態様において、シークエンス反応からのデータは、これらのエラーの影響を減少し得るような方式で導入される。
【００１６】
固相上への分子の析出および固定の方法は技術上周知である。核酸を付着させる方法は、例えば、Ｓｃｈｅｎａ（ｅｄ．），ＤＮＡ Ｍｉｃｒｏａｒｒａｙｓ：Ａ ｐｒａｃｔｉｃａｌ ａｐｐｒｏａｃｈ，Ｏｘｆｏｒｄ Ｕｎｉｖｅｒｓｉｔｙ Ｐｒｅｓｓ（１９９９）ＩＳＢＮ：０１９９６３７７６８に総説がある。典型的には、固相はガラスであろうが、非晶性もしくは結晶性シリコンまたはプラスチックのような他の材料を用いることができる。
【００１７】
複数の分子を規則正しいアレイで固相に付着させ得るが、さらに好ましくは、それらをランダム様式で付着させる。分子のランダム付着は、好ましくは、配列情報の光学的分解能に適した密度で分布させると、いかなる数の分子も含み得る。

【００１８】
適切なレポーター部分は種々の既知な報告システムのいずれか一つであってもよい。それは、組込まれたヌクレオシド類似体が容易に検出されるようになる放射性同位元素、例えば、ホスフェートもしくはチオホスフェートまたはＨ－ホスホネート基に組み込まれた^{32}Ｐ、^{33}Ｐ、^{35}Ｓ、もしくは他には^{3}Ｈもしくは^{14}Ｃまたはヨウ素同位元素であってもよい。それは、質量分析法もしくはＮＭＲにより検出可能な同位体であってもよい。それは、シグナル部分、例えば酵素、ハプテン、蛍光体、化学体、化学発光基、ラマン標識もしくは電気化学的標識、もしくは質量分析法による検出に適応するシグナル化合物であってもよい。

【００１９】
それぞれのシークエンシングステップは、個別の鋳型へのレポーター分子の取り付けをもたらすであろうし、かつ組込まれたレポーター部分の検出は塩基の同一性を帰属させることを可能にするであろう。蛍光性のレポーターの場合には、それから、例えば、蛍光顕微鏡法（例えばＰＭＴもしくはＣＣＤを用いて）によりこれらの分子が同定され、そしてレポーターの蛍光性質が、シークエンシング反応で組込まれた塩基への同一性の帰属を可能にするであろう。

【００２０】
一連のラウンドのシークエンシングサイクルからのデータを収集するためには鋳型の位置を特定しなければならない。これは最初のサイクルのシークエンシングで並行して達成し得るが、この場合、最初の塩基中のレポーター分子が鋳型の位置を確認するかまたは鋳型および／もしくはプライマー自体が、シークエンシングサイクリング反応に先立って固相上の位置を検出し得るように、レポーター部分で標識化されていてもよい。それぞれの鋳型分子の位置を知ることにより、シークエンシングのサイクル間に続いて起こる全てのイベントに引き続いてそれぞれの分子の状態をモニターすることが可能になる。付加の引き続く失敗は、例えば、鋳型を含むと知られた位置における蛍光の欠如により、それ自身を顕在化する。刺激の欠如、もしくは化学的損傷のどちらかによるレポーターの失敗はまた、一旦鋳型の位置が決定されていると、決定されることができる。これらの失敗した反応を追跡し、レポーターの失敗によるポテンシャルギャップとして最終配列において処理することができる。もしこれらの分子が引き続くサイクルで関与を再開するならば、これもまた追跡して意味のある配列を得ることができる。単一の塩基ギャップの個別のポイントが確認され得てかつ、複数の同一配列が固体表面上に並べられているならば、シークエンシングアレイ中における鋳型の他のコピーの配列のような参照鎖との比較によりコンセンサス配列を構築することができる。他に、既知の配列であり得る参照鎖との比較により単一の塩基ギャップを確認し得る（例えば、この技法を突然変異の検出に応用する際に）。

【００２１】
かくして、エラー、特に単一分子のシークエンシングに関連するエラーをこのシステムにおいて訂正することが可能であると我々は認識している。訂正の必要なエラーは、レポーターの切断および次のサイクル前の除去の失敗、組込みの失敗、レポーターへの損傷（例えば、蛍光体への損傷）ならびに組込みミスである。

【００２２】
一旦位置を特定すると、位置を特定した分子に対する全てのシークエンシングサイクルの結果は測定可能であろう。二つのセットのヌクレオチド類似体を用いることにより、先行するサイクルから持ち越されているレポーターの確認が可能になる。それ故に、先行するサイクルからのレポーターの再現を確認してモニターすることができる。

【００２３】
鋳型分子の位置を知ることにより、伸長していないように見える鋳型の確認がまた可能になる。上で議論したように、レポーター分子を観察する失敗は組込みの欠如に起因し得るが、レポーター部分への損傷にも起因し得る。しかしながら、損傷した分子の存在は、分解生成物および副反応の生成物を確認して除去することができる修飾ヌクレオチド合成の間での精製プロセスにより効果的に最小化することができるので、蛍光の不在は、それ故

に、修飾したヌクレオチドを組込む失敗の結果である可能性が高い。

【００２４】

もし、いずれかのサイクルのシークエンシングの後で、鋳型分子がいずれのレポーターとも関連しなければ、したがってこのポイントにおいて配列をマークして“休止”と表示する。次のラウンドのシークエンシングにおいて、それから鋳型分子をレポーターと関連させ得る、即ち、“休止”分子は伸長を再開して配列データが得られるようにする。しかしながら、鋳型分子は一つ以上のサイクルに対する関連を欠き続けるであろうし、そして配列はそれぞれのサイクルに対して休止としてマークされるであろう。

【００２５】

シークエンシングの間に作製される位置マーカーは、参照配列と共に作製される配列と比較するかもしくは当業者に既知のアラインメントアルゴリズムの一つを用いてシークエンシング手順間に作製される他の配列と比較するとき、アラインメント中のギャップを解釈するのに有用であろう。

【００２６】

使用した適切なポリメラーゼおよびリガーゼの生来の性質を知ると、組込みミスの位置を予想することが可能である。例えば、ミスマッチした末端塩基を含むプライマー配列は、マッチした配列より10^{2}～10^{6}－倍低い伸長効率を持って、ポリメラーゼに対してより劣った鋳型であることは当業者に既知である（Ｈｕａｎｇ，Ｍ．，Ｎ．Ａｒｎｈｅｉｍ，ａｎｄＭ．Ｇｏｏｄｍａｎ，Ｎｕｃｌｅｉｃ Ａｃｉｄｓ Ｒｅｓ，１９９２．２０（１７）：ｐ．４５６７－７３；Ｔｉｎｄａｌｌ ＫＲ，Ｋ．Ｔ．，Ｂｉｏｃｈｅｍｉｓｔｒｙ，１９８８．２７（１６）：ｐ．６００８－１３；Ｅｓｔｅｂａｎ，Ｊ．，Ｍ．Ｓａｌａｓ，ａｎｄＬ．Ｂｌａｎｃｏ，Ｊ Ｂｉｏｌ Ｃｈｅｍ，１９９３．２６８（４）：ｐ．２７１９－２６を参照）。数回のサイクル間にもしくはシークエンシングプロトコールの最後まで休止のままである分子は、それ故に、末端ミスマッチを含む公算がはるかに高い。そのような休止を受ける鋳型は、それ故に、ミスマッチによる潜在的終結として最後の塩基呼出し位置において標識化される。それから、配列が決定された断片の同定は、参照配列もしくは同一の試料からの他の配列が決定された鋳型へのアラインメントにより達成される。マークした位置で起こるミスマッチは、真の配列を表すことよりむしろ組込みミスの結果である可能性がより高いために、それに従って解釈され得る。

【００２７】

そのために鋳型分子が休止するサイクルの数は、組込まれたレポーターの欠如の連続的検出により計数することができる。偶然に起因する連続的休止の公算に対する閾値を配列データの解析の間にセットすることができる。それ以上ではミスマッチに起因すると分類され得る連続的休止の閾値は、ポリメラーゼ依存性塩基伸長かもしくは配列依存性連結反応のどちらかによる標識化の効率に依存する。例えば、偶然に起因する連続的休止の公算に対する閾値が1×10^{-6}％にセットされるならば、休止がミスマッチとして計数される前に、標識化の異なる効率を考慮して、下記の数の休止を計数する。

【００２８】

【表１】

プライマー伸長もしくは付着末端の連結反応による標識化の効率	１×１０$^{-6}$％の公算カットオフを越える前に発生する休止の数
９９.９％	3
９９.５％	4
９９％	4
９５％	5
９０％	6
８０％	8

【００２９】
比較的大きい確実性をもって、閾値を適当に増加させてもよい。必要な確実性の程度はシークエンシングの応用の許容度に依存するであろう；目的が配列の差を正確に決定するよりむしろ鋳型の断片を単に確認するだけならば、あまり厳格でないカットオフを許容できる。標識組込みのより低い効率の影響はまたシークエンシングの冗長度の程度により相殺することができる。この例における、組込みミスの確率は統計的に処理される。
【００３０】
主として蛍光により、単一分子を映像化して位置を特定することは当業者に既知である（Ｔｒａｂｅｓｉｎｇｅｒ，Ｗ.，ｅｔａｌ.，Ａｎａｌ Ｃｈｅｍ.，１９９９．７１（１）：ｐ．２７９－８３；Ｈａｒｍｓ，Ｇ.，ｅｔａｌ.，Ｂｉｏｐｈｙｓ．Ｊ.，１９９９．１７７：ｐ．２８６４－２８７０；Ｄｅｓｃｈｒｙｖｅｒ，Ｆ.，Ｐｕｒｅ ＆ Ａｐｐｌ．Ｃｈｅｍ，１９９８．７０：２１４７－２１５６；Ｂａｒｔｋｏ，Ａ．ａｎｄ Ｒ．Ｄｉｃｋｓｏｎ，Ｊ Ｐｈｙｓ Ｃｈｅｍ Ｂ，１９９９．１０３：ｐ．１１２３７－１１２４１を参照）。位置および標識のタイプに関する情報を含むデータファイルは、それ故に、容易に作成される。この発明の一つの実施態様において、配列データの解析は、シークエンシング手順の最後でそして全てのシークエンシングデータを得た後で実施される。一つもしくはそれ以上のファイルの中で、このデータを解析して、鋳型の位置を決定しかつこれらの位置に取り付けたいかなるレポーターも同定し得る。それから、そのようなデータを第二の解析に付して、全ての位置を確定した鋳型に対する配列を構築する。
【００３１】
好ましくは、シークエンシング反応のサイクルおよびデータ解析を並行して実施する。この例では、それぞれのサイクルから作製されたデータを解析して、レポーター分子の位置を確定して、それからこれらの位置を鋳型の位置と相関させる。それから、それぞれの位置を確定した鋳型に対する配列をそれぞれの連続したサイクルで構築する。
【００３２】
この発明を具体化するこの好ましい手順をこれから図１を参照して説明する。 図１に図示するシステムにおいて、配列が決定されるべき分子は技術的に記載されているような標準手順により固相上に固定されている。（Ｓｃｈｅｎａ（ｅｄ.），ＤＮＡ Ｍｉｃｒｏａｒｒａｙｓ：Ａ ｐｒａｃｔｉｃａｌ ａｐｐｒｏａｃｈ，Ｏｘｆｏｒｄ Ｕｎｉｖｅｒｓｉｔｙ Ｐｒｅｓｓ（１９９９）ＩＳＢＮ：０１９９６３７７６８に総説がある）。顕微鏡のスライドのような固体表面に結合させた鋳型を標識化し、それでスライドを、例えば蛍光顕微鏡スキャナーにより観察するときに、その位置を決定することができる。ステップ１０において、適切な鋳型の位置が先ず確定される。
【００３３】

今やステップ１２で実施するのは、ポリメラーゼ酵素で鋳型を伸長することにより、もしくは標識化したオリゴヌクレオチドをリガーゼで連結することにより引き起される塩基組込みを伴うシークエンシング反応である。

【００３４】

上述のように、シークエンシングステップは、鋳型の配列中で最初の塩基へのレポーター分子の取り付けをもたらすであろうし、そしてステップ１４では、組込まれるレポーター部分の検出により帰属すべき塩基の同一性を可能にする。次のステップ、ステップ１６は塩基および鋳型の位置を相関させる；最初のサイクルではこれはささいなステップである。それから、鋳型分子がレポーターと関連しているかどうかを決定する。即ち、ステップ１８において、対象とする鋳型がレポーターを持つかどうかをテストする。もしもシークエンシング手順の後で鋳型がレポーターと関連するならば、ステップ２０に手順を移動する。ここで、レポーターが先行するサイクルから来るかどうかを決定するテストを行う。そうでなければ、それからレポーターを確認して、そしてステップ２２で新しい塩基を帰属する。かくして、塩基は正確に同定されて、全てよしとなる。

【００３５】

それから、手順をステップ２４へ移動し、ここではもはや鋳型はないかどうかについてテストする。もしあれば、手順をステップ１８から繰り返す。

もしもステップ２０で塩基に関連するレポーターが先行するサイクルからであると決定されるならば、ステップ２６では何も塩基を帰属しないで、そして手順はステップ２４へおよび、もしあれば次の鋳型へ、直行する。

【００３６】

もしステップ１８で鋳型はレポーターを持たないと見出されるならば、ステップ５０でミスマッチフラグが立っているかどうかについてチェックをする。連続した休止の数が、ステップ３０で行うテストに従って、予め決めた極大値を越えるときには、ミスマッチフラグが活性化される。ミスマッチフラグが立っていなければ、手順をステップ２８へ移動し、休止Ｐを配列中に挿入する。また、起こっている連続休止の数をモニターする、休止カウンターを１個だけ追加する。テストをステップ３０で行って、連続した休止の数が予め決めた閾値または極大値を越えるかどうかを決定する。もし越えていなければ、手順をステップ２４へ移動し、休止を配列中に残す。もし連続した休止の数が予め決めた極大値を越えるならば、先行する塩基をミスマッチしたとして得点し、ステップ３２でミスマッチフラグを活性化して、そして手順をステップ２４へ進める。

【００３７】

休止インジケーターは、未知塩基の表示を与える機能として働く。これは塩基Ａ、Ｃ、ＧおよびＴのいずれか一つであると証明するか、もしくは事実全く塩基でないと証明するかもしれない。未知塩基の可能性を与えることにより、その鋳型についての情報は完全には廃棄されない。むしろ、下記の実施例で説明するように、例えば参照配列と参照して、それをなお使用する。

【００３８】

もしステップ２０でレポーターが先行するサイクルからであると決定されるならば、ステップ５２でミスマッチフラグが立っているかどうかについてチェックをする。もしミスマッチフラグが立っていなければ、手順をステップ２２へ移動し、塩基を帰属する。それから、手順をステップ２４へ移動し、処理すべきもう一つの鋳型があるかどうかを決定する。

【００３９】

もしミスマッチフラグが立っているならば、ステップ５４で、先に帰属した塩基は、ミスマッチした一つを除く全ての他の塩基を表すＩＵＢコードと取り替えられる。これは、もしも先行する塩基が“Ｃ”と表示されたが今やミスマッチしたと知られるならば、この塩基はＡ、ＧもしくはＴのいずれかであることが明らかであるからである。

【００４０】

さらに鋳型がないときには、ステップ２４におけるテストは結果ＮＯとなり、そして手順

をステップ３４へ移動して、そこで完成すべきサイクルがまだあるかどうか、即ち、その分子に対する塩基がまだあるかどうかを決定する。もしあれば、手順を次のサイクルのためのデータ、ステップ３６、へ移動し、その後で処理を再びステップ１６から進めて塩基および鋳型の位置を相関させる。
最終的には、ステップ３４におけるテストは結果ＮＯとなるであろうが、それでステップ３８においてこの手順の終了となる。
【００４１】
次に続く処理は、図１のシステムにより作成したような配列に応用されるもので、例えば、この方法で見出された配列を参照配列と比較することである。この実施例を以下に説明する。
化学反応を伴うステップ１０～１４に引き続く、図１に示すステップをパソコン（ＰＣ）のようなデジタルコンピュータ上に構築する。二つの実施例を、この明細書につけた付録での擬似コードを経由してさらに詳細に示す。最初の擬似コードは、ヌクレオチドが全ての四つの塩基、Ａ、Ｃ、ＧおよびＴ、の混合物により検索されることを仮定し、そして第二の擬似コードは、四つの塩基を別々に順次用いるときに使用する。
【００４２】
本発明は多くの応用を有するが、その幾らかをここに示す。例えば、この方法を用いてＤＮＡおよびＲＮＡのゲノムの配列を決定することができる。さらに、領域もしくは全体のゲノムにおける、領域もしくは全体のゲノムのｍＲＮＡ表現における、または一つもしくはそれ以上の塩基の置換、欠失もしくは挿入から生じる、ゲノムの人工的に作製した表現（例えば、ゲノム領域のＰＣＲ生成物）における、配列変異を同定することができる。
【００４３】
本発明は、ハプロタイピング（個体における染色体対間の配列差を決定する）にもまた定量的なｍＲＮＡ発現分析にも、例えば異なる細胞タイプ（組織）もしくは異なって処理した細胞から誘導した試料間でｍＲＮＡ発現のレベルを比較するのに、応用される。この技法はまた、病原体の検出および同定に使用するために、病原体ゲノムから誘導された配列を同定するのに応用し得る。
【００４４】
ここに示す実施例は、決定された配列におけるエラーを減少させるようにこのシステムにより特定の配列を取り扱う方式に与えられる。
【００４５】
実施例１
シークエンシング反応から次の配列が得られるが：
【化１】

GATCGGCTGACCATGGAC1

式中、１は一個のＴが組込まれていることを示す（そして２＝Ｃ、３＝Ａ、４＝Ｇ）。
サイクルの閾値数のため一層の伸長の失敗は、配列にマークをつけて塩基が休止の閾値数に先立って組込みミスされていることを示すことをもたらす。ここで、１（一つ）は一個のＴが、予め決定された閾値レベル以上の数の休止に先立って、組込まれていることを示し、かくして組込みミスされている可能性がある。それ故に、配列は多分廃棄される。図１に関しては、ＹＥＳがステップ３０において出力されるまで、予め決められた数のステップに対して手順をパス２８と３０に進め、そして先行する塩基をステップ３２でミスマッチしたとしてマークする。１の代わりに他の塩基に対して、２，３もしくは４を用いるが、２はＣを示し、３はＡを示し、そして４はＧを示す。
【００４６】
実施例２
シークエンシング反応から次の配列が得られる。一番目は新しく決定された配列であり、

二番目は参照配列である：
【化２】

GATCGGCTGACCATGGACC1CTGACAGT

GATCGGCTGACCATGGACCTCTGACAGT

サイクルの閾値数より長い休止は、組込みミスとしてＴの代わりに一つの１をマークする。この場合においては、シークエンシングを閾値数の配列の後で再開している。得られた配列を参照配列と比較するとき、配列アラインメントは休止した位置においてＴ．１アラインメントを表示する。それ故に、それは参照配列との現実の塩基差として差引くことができる。この配列アラインメントは図１に図示した処理に追加的な段階を表す。
【００４７】
実施例３
シークエンシングの間に休止に遭遇すると、その位置をＰとしてマークする。もし以下の新しい及び参照配列が得られると：
【化３】

GATCGGCTGACCATGGAPCCTCTGACAGT

GATCGGCTGACCATGGACCTCTGACAGT

Ｐでマークした位置におけるギャップの存在もしくは不存在下における参照配列との配列アラインメントは、それが休止であることを明らかにする。それ故に、全ての配列は隣接してかつ有用である。ここでも配列アラインメントは図１に図示した処理に追加的な段階を表す。
【００４８】
実施例４
シークエンシング反応から次の配列が得られる：
【化４】

GATCGGCTGACCATGGPCCTCTGACAGT

GATCGGCTGACCATGGACCTCTGACAGT

Ｐでマークした位置は失敗したレポーターを持つ塩基の組込みである。マークした位置におけるギャップの存在もしくは不存在下における参照配列との配列アラインメントは、これが配列中のギャップであることを明らかにする。取り出した配列は有用のままである。この例において、Ｐを‘Ｎ’で置換して配列中のギャップを意味することもできる。ここでも配列アラインメントは図１に図示した処理に追加的な段階を表す。
【００４９】
付録
第一擬似コード
シークエンシング反応完了後の配列集合のための擬似コードの例

```
Main()
{
Locate templates();
For (;number of cycles to analyse;);
{
      Correlate reporters with template locations()
      While (there are templates)
      {
            if (template location does not have a reporter)
            {
                  increment pause counter;
                        if (pause counter > threshold) mark
                        preceding base as a mismatch;
            }
                  else if (reporter is from the preceding
                  cycle) discard;
            else identify and assign base to template;
            move to the next template;

      }
            if (more cycles to be analysed) move to data
            for the next sequencing cycle;
      else return;
}
}
```

【００５０】
第二擬似コード
一連の単一塩基シークエンシングのための擬似コード

```
Main()
{
Locate templates();

While (there are cycles)
{
      read data for cycle

For (;four bases;)
{
      Correlate reporters with template locations()
      While (there are templates)
      {

                  if (reporter is from the preceding cycle)
                  discard;
            else identify and assign base to template;
            mark template as extended();
            next template

      }


            increment pause marker to all templates not
            marked extended();
      while (paused templates)
      {
            if (number of pauses have reached the
                  threshold) mark preceding base as
                  misincorporated
      }
            move to next cycle();

}
      Output sequence for analysis();

}
```

【図面の簡単な説明】

【図１】図１は、核酸分子のような生物分子の配列を決定するための反応の間に得られるデータを解析する方法を図示しかつこの発明の好ましい実施態様を形成する工程図である。

【図１】

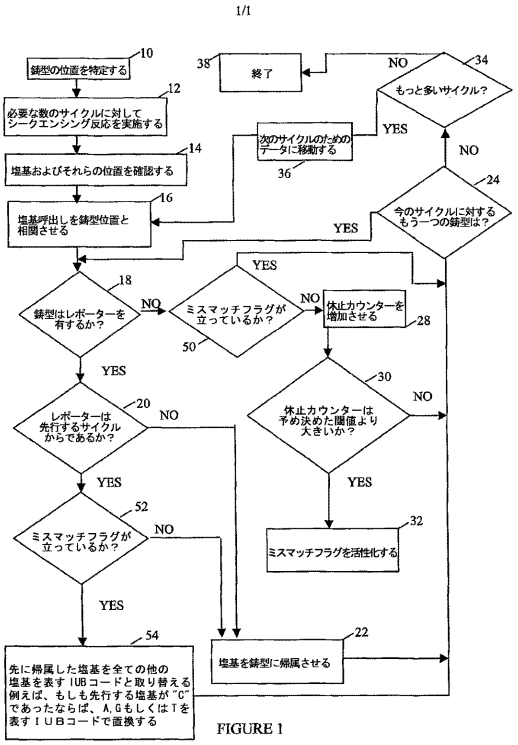

1/1
鋳型の位置を特定する
10
必要な数のサイクルに対して
シークエンシング反応を実施する
12
塩基およびそれらの位置を確認する
14
塩基呼出しを鋳型位置と
相関させる
16
鋳型はレポーターを
有するか？
18
NO
YES
レポーターは
先行するサイクル
からであるか？
20
NO
YES
ミスマッチフラグが
立っているか？
52
NO
YES
先に帰属した塩基を全ての他の
塩基を表すIUBコードと取り替える
例えば、もしも先行する塩基が"C"
であったならば、A,Gもしくは Tを
表すＩＵＢコードで置換する
54
塩基を鋳型に帰属させる
22
ミスマッチフラグが
立っているか？
50
YES
NO
休止カウンターを
増加させる
28
休止カウンターは
予め決めた閾値より
大きいか？
30
NO
YES
ミスマッチフラグを活性化する
32
今のサイクルに対する
もう一つの鋳型は？
24
YES
NO
次のサイクルのための
データに移動する
36
もっと多いサイクル？
34
NO
YES
終了
38

FIGURE 1

【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 January 2002 (10.01.2002)

PCT

(10) International Publication Number
WO 02/03305 A2

(51) International Patent Classification[7]: G06F 19/00, 17/00

(21) International Application Number: PCT/GB01/02985

(22) International Filing Date: 2 July 2001 (02.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0016472.3 5 July 2000 (05.07.2000) GB

(71) Applicant *(for all designated States except US)*: **AMERSHAM PHARMACIA BIOTECH UK LIMITED** [GB/GB]; Amersham Place, Little Chalfont, Buckinghamshire HP7 9NA (GB).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **ODEDRA, Raj** [GB/GB]; Amersham Pharmacia Biotech UK Ltd, Amersham Laboratories, White Lion Road, Amersham, Buckinghamshire HP7 9LL (GB).

(74) Agents: **HAMMER, Catriona, MacLeod** et al.; Nycomed Amersham plc, Amersham Laboratories, White Lion Road, Amersham, Buckinghamshire HP7 9LL (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*
— *entirely in electronic form (except for this front page) and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/03305 A2

(54) Title: SEQUENCING METHOD AND APPARATUS

(57) Abstract: In a method of identifying an unknown nucleotide sequence using base addition, a sequence of bases is obtained from a template, a base in the sequence is identified as an unknown base, an "unknown" indicator is included in the sequence, and an output sequence is generated containing the unknown base indicator. The sequence of bases is obtained from the template by evaluation of a reporter and assigning the bases in accordance therewith. A determination is made as to whether the reporter is from a preceding cycle of base determination, and if the reporter is from a preceding cycle of base determination, the base assignation is discarded.

SEQUENCING METHOD AND APPARATUS

FIELD OF THE INVENTION

The present invention relates to a sequencing method and apparatus that permits error correction during the sequencing of individual molecules.

BACKGROUND OF THE INVENTION

Sequencing is routinely performed by the method of chain termination and gel separation, essentially as described by Sanger, F., S. Nicklen, and A. Coulson (Proc Natl Acad Sci USA, 1977. 74(12); p. 5463-7). The method relies on the generation of a mixed population of DNA fragments representing terminations at each base in the sequence. The sequence is then determined by electrophoretic separation of these fragments.

Recent efforts to increase the throughput of sequencing have resulted in the development of alternative methods that eliminate the electrophoretic separation step. A number of these methods utilise base extension (i.e. base addition) and have been described for example in WO 93/21340, US 5,302,509 and US 5,547, 839. In these methods, the templates or primers are immobilised on a solid surface before exposure to reagents for sequencing. The immobilised molecules are incubated in the presence of nucleotide analogues that have a modification at the 3' carbon of the sugar residue that reversibly blocks the hydroxyl group at that position. The incorporation of such modified nucleotides by a polymerase ensures that only one nucleotide is added during each cycle of base extension. The added base is then detected by virtue of a label that has been incorporated into the 3' blocking group. Following detection, the blocking group is removed (or 'cleaved'), typically, by photochemical means to expose a free hydroxyl group that is available for base addition during the next cycle.

Generally, non-separation-based approaches rely on the presence of large numbers of template molecules for each target sequence to generate a consensus sequence from a given target. Thus, for example, base extension reactions may be applied to multiple templates by interrogating discrete spots of nucleic acid, each comprising a multiplicity of molecules, immobilised in a spatially addressable array.

However, reactions of terminator incorporation/cleavage, or base excision are prone to errors. For example, as described above, base extension strategies have generally utilised nucleotide analogues that combine the functions of a reporter molecule, usually a fluor, with that of a terminator occupying the 3' position on the sugar moiety. The bulky nature of the group and its position renders these compounds highly inefficient substrates for polymerases. In addition, the cleavage of the terminator group to permit subsequent additions is also subject to inefficiencies. In the presence of thousands, or preferably millions, of molecules for each target, even modest errors of less than 5% result in a cumulative loss of synchrony, between the multiplicity of strands representing each molecule, within a small number of cycles. Thus, with each cycle of sequencing the background noise increases progressively with a consequential deterioration of signal with each addition. This means that the number of bases of sequence data that can be obtained is limited before the specific signal becomes indistinguishable from background.

Recent advances in methods of single molecule detection (described, for example, in Trabesinger, W., et al., Anal Chem., 1999. 71(1); p. 279-83 and WO 00/06770) make it possible to apply sequencing strategies to single molecules. However, sequencing, when applied to clonal populations of molecules, is a stochastic process that results in some molecules undergoing reactions while others remain unmodified. Thus, in conventional sequencing methods, errors such as mis-incorporations are not normally of serious significance as the large numbers of molecules present ensure that consensus signal is obtained. When these reactions are applied to single molecules the outcomes are effectively quantized.

One such single molecule sequencing method is based on base excision and described, for example, in Hawkins, G. and L. Hoffman, Nature Biotechnology, 1997. vol.15; p. 803-804 and US 5,674,743. With this strategy, single template molecules are generated such that every base is labelled with an appropriate reporter. The template molecules are digested with exonuclease and the excised bases are monitored and identified. As these methods use highly processive enzymes such as Lambda exonuclease, there is the potential for analysing large templates of several kilobases in length. However, the continuous monitoring of excised bases from each template molecule in real time limits the number of molecules that can be analysed in parallel. In addition, there are difficulties in

generating a template where every base is labelled with an appropriate reporter such that excised bases can be detected on the basis of intrinsic optical or chemical properties.

Methods based on base extension (such as BASS) have also been adapted to a single molecule approach.

However, these techniques are prone to errors. In particular, incorporation of modified nucleotides can fail, for example, as the result of decreased efficiency of polymerase action with modified nucleotides. Where the reporter molecule is a fluorescent molecule, errors can also occur through failure of fluorescence because the fluor is lost, damaged, bleached, or unexcited. At the single molecule level, failures such as these will result in a failure in obtaining adequate sequence.

It is an object of the present invention to provide a sequencing method that enables errors to be detected. It is a further object of the present invention to allow analysis and error prevention, or correction, by monitoring the fate of individual molecules through sequencing reactions.

SUMMARY OF THE INVENTION

The invention in its various aspects is defined in the independent claims below, to which reference should now be made. Advantageous features are set forth in the appendant claims.

Briefly, in a preferred embodiment of the invention which takes the form of a method of analysing a nucleotide sequence, a sequence of bases is obtained from a template, and a base in the sequence is identified as an unknown base. An 'unknown' indicator is included in the sequence at the position corresponding to the unknown base, and an output sequence is generated containing the unknown base indicator. In the preferred embodiment the sequence of bases is obtained from the template by evaluation of a reporter and assigning the bases in accordance therewith. A determination is made as to whether the reporter is from a preceding cycle of base determination, and if the reporter is from a preceding cycle of base determination, the base assignation is discarded.

The nucleotide sequence to be analysed may be an RNA or DNA sequence.

BRIEF DESCRIPTION OF THE DRAWING

The invention will now be described in more detail by way of example with reference to the accompanying drawing in which:

Figure 1 is a flow chart illustrating a method of analysing data obtained during a reaction to determine the sequence of a biological molecule, such as a nucleic acid molecule, and forming a preferred embodiment of the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Figure 1 shows a flow diagram exemplifying a method of obtaining sequence information from a template. The method accounts for errors by (a) identifying bases that are carried over from a preceding cycle and (b) detecting paused molecules that may occur from failure of labelling or misincorporation of bases. The data analysis method makes use of a standard sequencing reaction which is performed as follows. First, a nucleic acid molecule for which sequence data is required, a template, is bound to a solid surface such as a microscope slide. The template can be labelled so that its position can be determined when the slide is viewed through a fluorescent microscope scanner, for example. The first base or nucleotide, i.e. A, C, G, or T, in the sequence of the template is queried by a chemical reaction adding a fluorescently-labelled base or a tag representing that base. This may be any one of A, C, G or T, or all four of them labelled with four different distinguishable labels. The first base in the template will bind to its complementary base in well-known fashion; that is A binds to T, and C binds to G, and vice versa. Base incorporation can be effected by extending the template with a polymerase enzyme or by ligating a labelled oligonucleotide with a ligase. Incorporation of the labelled base is detected and its identity determined. The label from that base is then removed. This series of steps is then repeated for the successive bases in the template.

Suitable standard sequencing reactions involving base addition/incorporation include base extension reactions such as those described in WO 93/21340, US 5,302,509 and US 5,547,839 and techniques such as those described in US 5,763,175, US 5,599,675, US 5,856,093 and US 5,715,330 in which successive rounds of sequencing involve base excision of the template prior to incorporation of the subsequent base.

When this sequencing reaction is performed, errors can occur. For example, (i) a base can be wrongly incorporated, that is misincorporated, or (ii) a label from one cycle can fail to be removed before the next cycle is performed, or (iii) incorporation of a base in any one cycle may fail. In the preferred embodiment of the invention to be described the data from sequence reactions is assimilated in such a way that the effects of these errors can be reduced.

Methods for deposition and fixation of molecules onto solid phases are well known in the art. Methods of attaching nucleic acids, for example, are reviewed in Schena (ed.), DNA Microarrays: A practical approach, Oxford University Press (1999) ISBN: 0199637768. Typically, the solid phase will be glass, although other materials such as amorphous or crystalline silicon or plastics can be used.

A plurality of molecules can be attached to the solid phase in an ordered array but, more preferably, they will be attached in a random manner. A random attachment of molecules may comprise any number of molecules, preferably distributed at a density appropriate for optical resolution of sequence information.

A suitable reporter moiety may be any one of various known reporting systems. It may be a radioisotope by means of which the incorporated nucleoside analogue is rendered easily detectable, for example ^{32}P, ^{33}P, ^{35}S incorporated in a phosphate or thiophosphate or H phosphonate group or alternatively ^{3}H or ^{14}C or an iodine isotope. It may be an isotope detectable by mass spectrometry or NMR. It may be a signal moiety e.g. an enzyme, hapten, fluorophore, chromophore, chemiluminescent group, Raman label, electrochemical label, or signal compound adapted for detection by mass spectrometry.

Each sequencing step will result in the attachment of reporter molecules to individual templates and the detection of the reporter moiety incorporated will permit the identity of the base to be assigned. In the case of fluorescent reporters, these molecules will then be identified by, for example, fluorescence microscopy (e.g. using a PMT or CCD) and the fluorescence property of the reporter will permit the assignment of identity to the base incorporated in the sequencing reaction.

In order to collect data from sequential rounds of sequencing cycles the template must be located. This can be achieved concurrently with the first cycle of sequencing where the reporter molecule in the first base identifies template location or the template and/or primer may itself be labelled with a reporter moiety such that its location on the solid phase may be detected in advance of the sequence cycling reaction. Knowing the location of each template molecule makes it possible to monitor the state of each molecule following all subsequent events during cycles of sequencing. Subsequent failure of addition, for example, manifests itself by lack of fluorescence at a location known to contain a template. Failure of the reporter due either to a lack of stimulus, or chemical damage can also be determined once the location of the template has been determined. These failed reactions can be tracked and treated in the final sequence as potential gaps due to reporter failure. If these molecules resume participation in subsequent cycles this, too, can be tracked and a meaningful sequence obtained. Individual points of single base gaps can be identified and, where multiple identical sequences have been arrayed onto the solid surface, a consensus sequence can be built up through comparisons with reference strands such as sequences of other copies of templates in the sequencing array. Alternatively single base gaps may be identified by comparison with a reference strand which may be the known sequence (e.g. in the application of this technique to mutation detection).

Thus we have appreciated that it is possible in this system to correct errors, particularly errors associated with single molecule sequencing. Errors that need to be corrected are failure of reporter cleavage and elimination before the next cycle, failure of incorporation, damage to reporter (e.g. damage to fluor), and misincorporation.

Once located, all sequencing cycle outcomes for the molecule located will be measurable. Using two sets of nucleotide analogues permits the identification of reporter that has been carried over from the previous cycle. The recurrence of a reporter from the previous cycle can therefore be identified and monitored.

Knowing the location of the template molecule also permits the identification of templates that appear not to have extended. As discussed above, failure to observe a reporter molecule can be due to lack of incorporation, but can also be due to damage to the reporter moiety. However, as the presence of damaged

molecules can be effectively minimised by a purification process during the synthesis of modified nucleotides where breakdown products and products of side reactions can be identified and eliminated, the absence of fluorescence is therefore more likely to be a result of failure to incorporate a modified nucleotide.

If, after any cycle of sequencing, a template molecule is not associated with any reporters, the sequence is marked accordingly at this point to indicate a "pause". In the next round of sequencing, the template molecule may then be associated with a reporter i.e. the "paused" molecule resumes extension allowing sequence data to be obtained. However, the template molecule may continue to lack association with any reporters for more than one cycle, and the sequence will be marked as a pause for each respective cycle.

A positional marker generated during sequencing will be useful for interpreting gaps in alignments when comparing with the sequence generated with reference sequences or with other sequences generated during the sequencing procedure using one of the alignment algorithms known to those skilled in the art.

It is possible to predict positions of mis-incorporation knowing the inherent properties of the pertinent polymerases and ligases used. For example, it is known to those practised in the art that primer sequences that contain a mismatched terminal base are poorer templates for polymerases, with extension efficiencies of between 10^2 to 10^6 -fold lower than matched sequences (see Huang, M., N. Arnheim, and M. Goodman, Nucleic Acids Res, 1992. 20(17): p. 4567-73., Tindall KR, K.T., Biochemistry, 1988. 27(16): p. 6008-13, Esteban, J., M. Salas, and L. Blanco, J Biol Chem, 1993. 268(4): p. 2719-26). Molecules that remain paused for several cycles, or to the end of the sequencing protocol, therefore have a much higher likelihood of containing a terminal mismatch. Templates that undergo such pauses are therefore tagged at the last base call position as potential terminations due to mismatches. Identification of the sequenced fragment is then achieved through alignment to a reference sequence or other sequenced templates from the same sample. Mismatches that occur at marked positions are more likely to be the result of mis-incorporation rather than representing the true sequence and can therefore be interpreted accordingly.

The number of cycles for which a template molecule is paused can be counted by successive detection of a lack of incorporated reporter. A threshold for the likelihood of successive pauses resulting by chance can be set during the analysis of the sequence data. The threshold above which successive pauses can be classed as resulting from a mismatch will be dependent upon the efficiency of labelling either by polymerase dependent base extension, or sequence dependent ligation. For example, if the threshold for the likelihood of successive pauses resulting by chance is set at $1x10^{-6}$% the following numbers of pauses will be counted, taking into account different efficiencies of labelling, before the pause is counted as a mismatch.

Efficiency of labelling by primer extension, or ligation of cohesive termini	Number of pauses encountered before exceeding a likelihood cut off of $1x10^{-6}$%
99.9%	3
99.5%	4
99%	4
95%	5
90%	6
80%	8

For greater certainty, the threshold may be increased appropriately. The degree of certainty required will be dependent on the tolerance of the sequencing application; a less stringent cut off can be tolerated if the aim is simply to identify the template fragments, rather than precisely determine sequence differences. The effect of a lower efficiency of label incorporation can also be offset by the degree of sequencing redundancy. The probability of a misincorporation, in this instance, is dealt with statistically.

Imaging and locating single molecules, principally by fluorescence, is familiar to those practised in the art (see Trabesinger, W., et al., Anal Chem., 1999. 71(1): p. 279-83, Harms, G., et al., Biophys. J., 1999. 177: p. 2864-2870, Deschryver, F., Pure & Appl. Chem, 1998. 70: p. 2147-2156., Bartko, A. and R. Dickson, J Phys Chem B, 1999. 103: p. 11237-11241). Data files that contain information regarding location and type of label are, therefore, readily generated. In one embodiment of this invention, the analysis of sequence data is performed at the end of the sequencing procedure and after all the sequencing data has been acquired. This data, in one or more files, may be analysed to

determine the locations of the templates and identify any attached reporters at these positions. Such data is then subjected to a second analysis to build sequences for all located templates.

Preferably, cycles of sequencing reaction and data analysis are performed concurrently. In this instance, data generated from each cycle is analysed to locate reporter molecules, these locations are then correlated with locations of the templates. The sequences for each located template can then be built on with each successive cycle.

The preferred procedure embodying the invention will now be described with reference to Figure 1.

In the system illustrated in Figure 1, molecules to be sequenced have been fixed onto solid phases by standard procedures as described in the art. (Reviewed in Schena (ed.), DNA Microarrays: A practical approach, Oxford University Press (1999) ISBN: 0199637768). The template, bound to a solid surface such as a microscope slide, is labelled so its position can be determined when the slide is viewed through a fluorescent microscope scanner, for example. At step 10, a relevant template is first located.

Sequencing reactions involving base incorporation which can be effected by extending the template with a polymerase enzyme or by ligating a labelled oligonucleotide with a ligase are now performed, step 12.

As described above, the sequencing step will result in the attachment of a reporter molecule to the first base in the sequence of the template, and the detection of the reporter moiety which is incorporated permits the identity of the base to be assigned, step 14. The next step, step 16, is to correlate the base and template locations; on this first cycle this is a trivial step. A determination is then made as to whether the template molecule is associated with a reporter. That is to say, in step 18 a test is made as to whether the subject template has a reporter or not. If after the sequencing operation the template is associated with a reporter, the procedure moves on to step 20. Here a test is made to determine whether the reporter comes from a previous cycle. If it does not, then it is identified and a new base assigned, step 22. Thus the base has been correctly identified and all is well.

The procedure then moves to step 24 where a test is made as to whether there are any more templates. If so, the procedure repeats from step 18.

If in step 20 it is determined that the reporter associated with the base is from a previous cycle, then no base is assigned, step 26, and the procedure goes straight to step 24 and to the next template, if any.

If in step 18 the template is found not to have a reporter, in step 50 a check is made as to whether the mismatch flag is on. The mismatch flag is activated when the number of consecutive pauses exceeds the predetermined maximum, according to a test made at step 30. If the mismatch flag is not on, the procedure moves to step 28, and a pause P is inserted in the sequence. Also, a pause counter, which monitors the number of consecutive pauses which occur, is incremented by one. A test is the made in step 30 to determine whether the number of consecutive pauses exceeds a predetermined threshold or maximum value. If it does not, the procedure moves to step 24 leaving the pause in the sequence. If the number of consecutive pauses does exceed the predetermined maximum, then the preceding base is scored as mismatched and the mismatch flag is activated, step 32, and the procedure then proceeds to step 24.

The pause indicator serves the function of providing an indication of an unknown base. This may prove to be any one of the bases A, C, G and T, or may in fact prove not to be a base at all. By providing for the possibility of an unknown base the information for that template is not wholly discarded. Rather, it may still be used, for example with reference to a reference sequence, as described in the examples below.

If in step 20 it is determined that the reporter is from a previous cycle, in step 52 a check is made as to whether the mismatch flag is on. If the mismatch flag is not on, then the procedure moves to step 22 and a base is assigned. The procedure than moves to step 24 to determine whether there is another template for processing.

If the mismatch flag is on, at step 54 the previously assigned base is replaced with an IUB code representing all other bases except the one which was mismatched. This is because if the previous base was labelled "C" but is now known to be mismatched, it is clear the base is either A, G or T.

When there are no more templates, the test at step 24 has the result NO, and the procedure moves to step 34, where a determination is made as to whether there are any more cycles to be completed, that is, whether there are any more bases for that molecule. If there are, the procedure moves to the data for the next cycle, step 36, after which the processing proceeds again from step 16, with correlation of the base and template locations.

Eventually the test at step 34 will have the result NO, and that leads to the end of the procedure, step 38.

There may be subsequent processing applied to the sequence as produced by the system of Figure 1, for example to compare the sequence found by the method with a reference sequence. Examples of this are described below.

The steps shown in Figure 1, subsequent to the steps 10 to 14 which involve chemical reactions, are implemented on a digital computer such as a personal computer (PC). Two examples are shown in more detail by way of pseudocode in the Appendix to this specification. The first pseudocode assumes that the nucleotides are queried by a mixture of all four bases A, C, G, and T, and the second pseudocode is for use when the four bases are used separately in sequence.

The present invention has many applications, some of which are given here. For example, the sequence of DNA and RNA genomes can be determined using this method. Further, sequence variations in regions of or entire genomes, mRNA representations of regions of or entire genomes or in artificially generated representations of a genome (eg. PCR products of regions of a genome) which result from substitutions, deletions or insertions of one or more bases can be identified.

The present invention has application in haplotyping (determining sequence differences between chromosome pairs in an individual) and also in quantitative mRNA expression analysis, for example in comparing levels of mRNA expression between samples derived from different cell types (tissues) or differently treated cells. This technique may also be applied to identifying

sequences derived from pathogen genomes for use in pathogen detection and identification.

Examples are now given of the way specific sequences are handled by the system in such a way as to reduce errors in the determined sequence.

Example 1

The following sequence is obtained from a sequencing reaction:

GATCGGCTGACCATGGAC1

wherein 1 indicates a T has been incorporated (and 2=C, 3=A, 4=G).

A failure of further extension for the threshold number of cycles results in marking the sequence to indicate that a base has been misincorporated prior to the threshold number of pauses. Here, a 1 (one) indicates that a T has been incorporated prior to a number of pauses above the predetermined threshold level and thus is likely to have been misincorporated. The sequence may therefore be discarded. Referring to Figure 1, the procedure follows the path 28, 30 for a predetermined number of steps, until a YES is output at step 30, and the preceding base is marked as mismatched in step 32. Instead of a 1, for the other bases 2, 3 or 4 are used, 2 indicating C, 3 indicating A, and 4 indicating G.

Example 2

The following sequences are obtained from a sequencing reaction. The first is a newly determined sequence and the second is a reference sequence:

GATCGGCTGACCATGGACC1CTGACAGT
GATCGGCTGACCATGGACCTCTGACAGT

Pausing for longer than the threshold number of cycles marks a 1 for T as a mis-incorporation. In this case, sequencing has resumed after the threshold number of sequences. When the sequence obtained is compared to the reference sequence, sequence alignment demonstrates a T.1 alignment at the paused position. It can therefore be discounted as a real base difference with the reference sequence. The sequence alignment represents a stage additional to the processing illustrated in Figure 1.

Example 3

When a pause is encountered during sequencing, its position is marked as P. If the following new and reference sequences are obtained:

GATCGGCTGACCATGGAPCCTCTGACAGT
GATCGGCTGACCATGGACCTCTGACAGT

sequence alignment with the reference sequence in the presence or absence of a gap at the position marked with a P reveals that it was a pause. All of the sequence is therefore contiguous and useful. The sequence alignment again represents a stage additional to the processing illustrated in Figure 1.

Example 4

The following sequence is obtained in a sequencing reaction:

GATCGGCTGACCATGGPCCTCTGACAGT
GATCGGCTGACCATGGACCTCTGACAGT

The position marked as P is the incorporation of a base with a failed reporter. Sequence alignment with a reference sequence in the presence or absence of a gap at the marked position reveals that this represents a gap in the sequence. The extracted sequence remains useful. In this instance the P can be substituted with an 'N' to signify a gap in the sequence. The sequence alignment again represents a stage additional to the processing illustrated in Figure 1.

APPENDIX

First Pseudocode

Example of a pseudo code for sequence assembly after completion of the sequencing reactions.

```
Main()
{
Locate templates();
For (;number of cycles to analyse;);
{
      Correlate reporters with template locations()
      While (there are templates)
      {
            if (template location does not have a reporter)
            {
                  increment pause counter;
                        if (pause counter > threshold) mark
                        preceding base as a mismatch;
            }
                  else if (reporter is from the preceding
                  cycle) discard;
            else identify and assign base to template;
            move to the next template;

      }
            if (more cycles to be analysed) move to data
            for the next sequencing cycle;
      else return;
}
}
```

Second Pseudocode

Pseudocode for sequential single base sequencing

```
Main()
{
Locate templates();

While (there are cycles)
{
     read data for cycle

For (;four bases;)
{
     Correlate reporters with template locations()
     While (there are templates)
     {

          if (reporter is from the preceding cycle)
          discard;
     else identify and assign base to template;
     mark template as extended();
     next template

     }


     increment pause marker to all templates not
     marked extended();
     while (paused templates)
     {
          if (number of pauses have reached the
               threshold) mark preceding base as
               misincorporated
     }
          move to next cycle();

}
     Output sequence for analysis();

}
```

CLAIMS

1. A method of identifying an unknown nucleotide sequence using base addition, comprising the steps of:
obtaining a sequence of bases from a template;
identifying a base in the sequence as an unknown base and including an 'unknown' indicator in the sequence; and
providing an output sequence containing the unknown base indicator.

2. A method according to claim 1, further comprising counting the number of consecutive unknown bases, and providing an indication when the number of consecutive unknown bases exceeds a predetermined threshold value.

3. A method according to claim 2, wherein when the threshold is exceeded, the preceding base is marked as misincorporated.

4. A method according to claim 1, further comprising the step of sequence alignment between the output sequence and a reference sequence.

5. A method according to claim 2, further comprising the step of sequence alignment between the output sequence and a reference sequence.

6. A method according to claim 1, wherein the sequence is determined by evaluation of a reporter, and further comprising the step of determining whether the reporter is from a preceding cycle of base determination.

7. A method of identifying an unknown nucleotide sequence, comprising the steps of:
obtaining a sequence of bases from a template by evaluation of a reporter and assigning the bases in accordance therewith;
determining whether the reporter is from a preceding cycle of base determination; if the reporter is from a preceding cycle of base determination, discarding the base assignation; and
providing an output sequence.

8. Apparatus for identifying an unknown nucleotide sequence using base addition, comprising:

means for obtaining a sequence of bases from a template;

means for identifying a base in the sequence as an unknown base and including an 'unknown' indicator in the sequence; and

means for providing an output sequence containing the unknown base indicator.

9. Apparatus for identifying an unknown nucleotide sequence using base addition, comprising:

means for obtaining a sequence of bases from a template by evaluation of a reporter and assigning the bases in accordance therewith;

means for determining whether the reporter is from a preceding cycle of base determination, and if the reporter is from a preceding cycle of base determination, for discarding the base assignation; and

means for providing an output sequence.

10 A computer program product for identifying an unknown nucleotide sequence using base extension which, when loaded into a computer, will control the computer to perform the following steps:

obtain a sequence of bases from a template;

identify a base in the sequence as an unknown base and include an 'unknown' indicator in the sequence; and

provide an output sequence containing the unknown base indicator.

11. A computer program comprising program code means for performing all the steps of any one of claims 1 to 7, when said program is run on a computer.

1/1

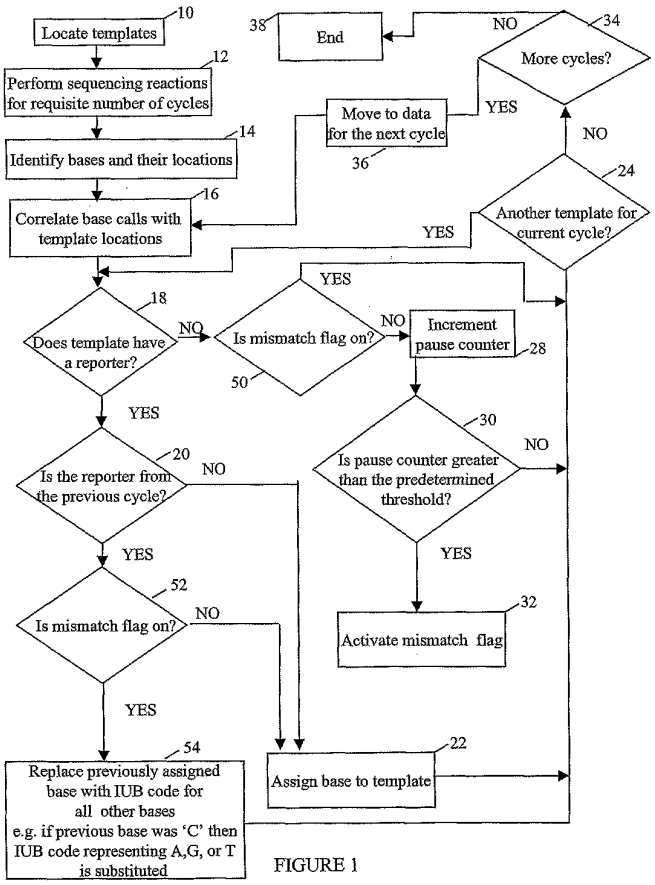

Locate templates
10
Perform sequencing reactions for requisite number of cycles
12
Identify bases and their locations
14
Correlate base calls with template locations
16
Does template have a reporter?
18
NO
YES
Is the reporter from the previous cycle?
20
NO
YES
Is mismatch flag on?
52
NO
YES
Replace previously assigned base with IUB code for all other bases
e.g. if previous base was 'C' then IUB code representing A,G, or T is substituted
54
Assign base to template
22
Is mismatch flag on?
50
YES
NO
Increment pause counter
28
Is pause counter greater than the predetermined threshold?
30
NO
YES
Activate mismatch flag
32
Move to data for the next cycle
36
Another template for current cycle?
24
YES
NO
More cycles?
34
NO
YES
End
38

FIGURE 1

【国際公開パンフレット（コレクトバージョン）】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 January 2002 (10.01.2002)

PCT

(10) International Publication Number
WO 02/003305 A3

(51) International Patent Classification[7]: G06F 19/00, 17/00

(21) International Application Number: PCT/GB01/02985

(22) International Filing Date: 2 July 2001 (02.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0016472.3 5 July 2000 (05.07.2000) GB

(71) Applicant *(for all designated States except US)*: AMERSHAM BIOSCIENCES UK LTD [GB/GB]; Amersham Place, Little Chalfont, Buckinghamshire HP7 9NA (GB).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: ODEDRA, Raj [GB/GB]; Amersham Pharmacia Biotech UK Ltd, The Grove Centre, White Lion Road, Amersham, Buckinghamshire HP7 9LL (GB).

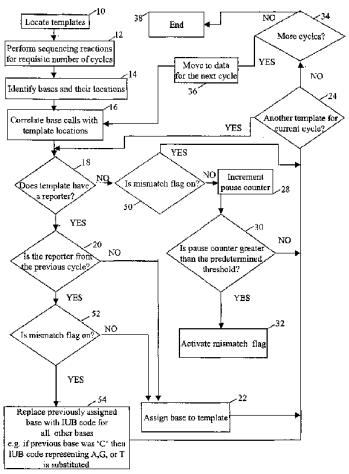
(74) Agents: HAMMER, Catriona, MacLeod et al.; Amersham plc, The Grove Center, White Lion Road, Amersham, Buckinghamshire HP7 9LL (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR SEQUENCE ANALYSIS



(57) Abstract: In a method of identifying an unknown nucleotide sequence using base addition, a sequence of bases is obtained from a template, a base in the sequence is identified as an unknown base, an "unknown" indicator is included in the sequence, and an output sequence is generated containing the unknown base indicator. The sequence of bases is obtained from the template by evaluation of a reporter and assigning the bases in accordance therewith. A determination is made as to whether the reporter is from a preceding cycle of base determination, and if the reporter is from a preceding cycle of base determination, the base assignation is discarded.

WO 02/003305 A3

WO 02/003305 A3

IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
1 August 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 01/02985

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F19/00 G06F17/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 93 05183 A (BAYLOR COLLEGE MEDICINE) 18 March 1993 (1993-03-18) abstract; claim 24	1-11
Y	US 5 552 278 A (BRENNER SYDNEY) 3 September 1996 (1996-09-03) abstract column 13, line 53 - line 67	1-11
A	WO 93 21340 A (MEDICAL RES COUNCIL ;BRENNER SYDNEY (GB); ROSENTHAL ANDRE (GB)) 28 October 1993 (1993-10-28) cited in the application abstract	1-11

[ ] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 April 2002	24/04/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Filloy García, E

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 01/02985

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 9305183	A	18-03-1993	AU	2674092 A	05-04-1993
			WO	9305183 A1	18-03-1993
US 5552278	A	03-09-1996	AT	201238 T	15-06-2001
			AU	685628 B2	22-01-1998
			AU	2379195 A	23-10-1995
			CA	2163662 A1	12-10-1995
			DE	69520917 D1	21-06-2001
			DE	69520917 T2	20-12-2001
			DK	703991 T3	18-06-2001
			EP	0703991 A1	03-04-1996
			ES	2159635 T3	16-10-2001
			JP	8511174 T	26-11-1996
			PT	703991 T	30-10-2001
			WO	9527080 A2	12-10-1995
			US	5599675 A	04-02-1997
			US	5856093 A	05-01-1999
			US	5714330 A	03-02-1998
			US	5831065 A	03-11-1998
WO 9321340	A	28-10-1993	AT	159766 T	15-11-1997
			AU	4020893 A	18-11-1993
			CA	2133956 A1	28-10-1993
			DE	69314951 D1	04-12-1997
			DE	69314951 T2	19-03-1998
			EP	0640146 A1	01-03-1995
			ES	2110604 T3	16-02-1998
			WO	9321340 A1	28-10-1993
			JP	7507681 T	31-08-1995
			US	6087095 A	11-07-2000

Form PCT/ISA/210 (patent family annex) (July 1992)

フロントページの続き

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CR,CU,CZ,DE,DK,DM,DZ,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

(72)発明者 ラジュ・オデドラ
イギリス、エイチピー７・９エルエル、バッキンガムシャー、アマシャム、ホワイト・ライオン・ロード、アマシャム・ラボラトリーズ、アマシャム・ファルマシア・バイオテック・ユーケイ・リミテッド

Ｆターム(参考) 4B029 AA07 BB20 CC03 FA15
4B063 QA13 QQ42 QQ52 QR08 QR20 QR42 QS16 QS31 QS39
5B075 ND20 UU19