PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12Q 1/68, 1/48

A1

(11) International Publication Number: WO 95/00669

(43) International Publication Date: 5 January 1995 (05.01.95)

(21) International Application Number: PCT/US94/07086

(22) International Filing Date: 22 June 1994 (22.06.94)

(30) Priority Data:
9302152-5 22 June 1993 (22.06.93) SE

(71) Applicants *(for all designated States except US)*: PHARMACIA BIOTECH AB [SE/SE]; S-751 82 Uppsala (SE). BAYLOR COLLEGE OF MEDICINE [US/US]; One Baylor Plaza, Houston, TX 77030 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: CASKEY, C., Thomas [US/US]; 6402 Belmont, Houston, TX 77005 (US). SHUMAKER, John [US/US]; 719 Pizer, Houston, TX 77009 (US). METSPALU, Andres [EE/EE]; 4-33 Kaunase Street, Tartu 2400 (EE).

(74) Agents: REYNOLDS, Leo, R. et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).

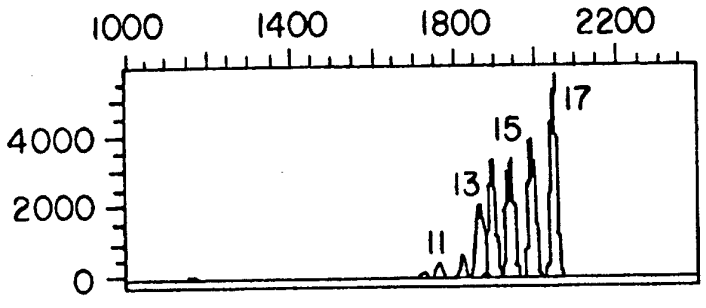
(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PARALLEL PRIMER EXTENSION APPROACH TO NUCLEIC ACID SEQUENCE ANALYSIS

(57) Abstract

A method of analyzing a polynucleotide of interest, comprising providing one or more sets of consecutive oligonucleotide primers differing within each set by one base at the growing end thereof; annealing a single strand of the polynucleotide or a fragment of the polynucleotide to the oligonucleotide primers under hybridization conditions; subjecting the primers to single base extension reactions with a polymerase and terminating nucleotides, the terminating nucleotides being mutually distinguishable; and observing the location and identity of each terminating nucleotide to thereby analyze the sequence or a part of the nucleotide sequence of the polynucleotide of interest, is disclosed. An apparatus comprising a solid support to which is attached at defined locations thereon one or more sets of consecutive oligonucleotide primers differing within each set by one base at the growing end thereof is also described.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

PARALLEL PRIMER EXTENSION APPROACH TO NUCLEIC ACID SEQUENCE ANALYSIS

Background of the Invention

Today, there are two predominant methods for DNA sequence determination: the chemical degradation method (Maxam and Gilbert, Proc. Natl. Acad. Sci. 74:560-564 (1977), and the dideoxy chain termination method (Sanger et al., Proc. Natl. Acad. Sci. 74:5463-5467 (1977)). Most automated sequencers are based on the chain termination method utilizing fluorescent detection of product formation. There are two common variations of these systems: (1) dye-labeled primers to which deoxynucleotides and dideoxynucleotides are added, and (2) primers to which deoxynucleotides and fluorescently labeled dideoxynucleotides are added. In addition, the labeled deoxynucleotides can be used in conjunction with unlabeled dideoxynucleotides. This method is based upon the ability of an enzyme to add specific nucleotides onto the 3′ hydroxyl end of a primer annealed to a template. The base pairing property of nucleic acids determines the specificity of nucleotide addition. The extension products are separated electrophoretically on a polyacrylamide gel and detected by an optical system utilizing laser excitation.

Although both the chemical degradation method and the dideoxy chain termination method are in widespread use, there are many associated disadvantages: for example, the methods require gel-electrophoretic separation. Typically, only 400-800 base pairs can be sequenced from a single clone. As a result, the systems are both time- and labor-intensive. Methods avoiding gel separation have been developed in attempts to increase the sequencing throughput.

Methods have been proposed by Crkvenjakov (Drmanac, et al., Genomics 4:114 (1989); Strezoska et al., (Proc. Natl. Acad.Sci. USA 88:10089 (1991); Drmanac, et al., Science 260: 1649 (1991)) and Bains and Smith (Bains and Smith, J. Theoretical Biol. 135: 303 (1988)). These sequencing by hybridization (SBH) methods potentially can increase the sequence throughput because multiple hybridization reactions are performed simultaneously. This type of system utilizes the information obtained from multiple hybridizations of the polynucleotide of interest, using short oligonucleotides to determine the nucleic acid sequence (Drmanac, United States Patent No. 5,202,231). To reconstruct the sequence requires an extensive computer search algorithm to determine the optimal order of all fragments obtained from the multiple hybridizations.

These methods are problematic in several respects. For example, the hybridization is dependent upon the sequence composition of the duplex of the oligonucleotide and the polynucleotide of interest, so that GC-rich regions are more stable than AT-rich regions. As a result, false positives and false negatives during hybridization detection are frequently present and complicate sequence determination. Furthermore, the sequence of the polynucleotide is not determined directly, but is inferred from the sequence of the known probe, which increases the possibility for error. A great need remains to develop efficient and accurate methods for nucleic acid sequence determination.

Summary of the Invention

The current invention pertains to methods for analyzing, and particularly for sequencing, a polynucleotide of interest, and an apparatus useful in analyzing a polynucleotide of interest. In one embodiment of the current invention, the nucleotide sequence of a

polypeptide of interest is analyzed for the presence of mutations or alterations. In a second embodiment of the current invention, the nucleotide sequence of a polypeptide of interest, for which the nucleotide sequence was not known previously, is determined. The method comprises detecting single base extension events of a set of specific oligonucleotide primers, such that the label and position of each separate extension event defines a base in a polynucleotide of interest.

In one method of the current invention, a solid support is provided. An array of a set or several sets of consecutive oligonucleotide primers of a specified size having known sequences is attached at defined locations to the solid support. The oligonucleotide primers differ within each set by one base pair. The oligonucleotide primers either correspond to at least a part of the nucleotide sequence of one strand of the polynucleotide of interest, if the sequence is known, or represent a set of all possible nucleotide sequences for oligonucleotide primers of the specified size, if the sequence is not known. A polynucleotide of interest, which may be DNA or RNA, or a fragment of the polynucleotide of interest, is annealed to the array of oligonucleotide primers under hybridization conditions, thereby generating "annealed primers". The annealed primers are subjected to single base extension reaction conditions, under which a nucleic acid polymerase and terminating nucleotides, such as dideoxynucleotides (ddNTPs) corresponding to the four known bases (A, G, T and C), are provided to the annealed primers. The terminating nucleotides can also comprise a terminating string of known polynucleotides, such as dinucleotides. As a result of the single base extension reaction, extended primers are generated, in which a terminating nucleotide is added to each of the annealed primers. The terminating nucleotides can be provided to

the annealed primers either simultaneously or sequentially. The terminating nucleotides are mutually distinguishable; i.e., at least one of the nucleotides is labelled to facilitate detection. After addition of the terminating nucleotides, the sequence of the polynucleotide of interest is analyzed by "reading" the oligonucleotide array: the identity and location of each terminating nucleotide within the array on the solid support is observed. The label and position of each terminating nucleotide on the solid support directly defines the sequence of the polynucleotide of interest that is being analyzed.

In a second method of the current invention, the polynucleotide of interest is analyzed for the presence of specific mutations through the use of oligonucleotide primers that are not attached to a solid support. The oligonucleotide primers are tailored to anneal to the polynucleotide of interest at a point immediately preceding the mutation site(s). If more than one mutation site is examined, the oligonucleotide primers are designed to be mutually distinguishable: in a preferred embodiment, the oligonucleotide primers have different mobilities during gel electrophoresis. For example, oligonucleotides of different lengths are used. After the oligonucleotide primers are annealed to the polynucleotide of interest, the annealed primers are subjected to single base extension reaction conditions, resulting in extended primers in which terminating nucleotides are added to each of the annealed primers. As in the first method of the current invention, the terminating nucleotides are mutually distinguishable. After addition of the terminating nucleotides, the sequence of the polynucleotide of interest is analyzed by eluting the extended primers, performing gel electrophoresis, and "reading" the gel: the identity and location of each terminating nucleotides on the gel is observed using standard methods, such as with an automated DNA sequencer.

The label and position of each terminating nucleotide on the gel directly defines the sequence of the polynucleotide of interest that is being analyzed, and indicates whether a mutation is present.

The apparatus of the current invention comprises a solid support having an array of one or more sets of consecutive oligonucleotide primers with known sequences attached to it at defined locations, each oligonucleotide primer differing within each set by one base pair. The set of oligonucleotide primers either corresponds to at least a part of the nucleotide sequence of one strand of the polynucleotide of interest, if the sequence is known, or represents all possible nucleotide sequences for oligonucleotide primers of the specified size, if the sequence is not known.

The current invention provides both direct information, due to the detection of a specific nucleotide addition, and indirect information, due to the known sequence of the annealed primer to which the specific base addition occurred, for the polynucleotide of interest. The ability to determine nucleic acid sequences is a critical element of understanding gene expression and regulation. In addition, as advances in molecular medicine continue, sequence determination will become a more important element in the diagnosis and treatment of disease.

Brief Description of the Drawings

Figure 1 depicts an example of a set of oligonucleotide primers comprising consecutive primers differing by one base pair at the growing end and capable of hybridizing successively along the relevant part(s) or the whole of the polynucleotide of interest.

Figure 2 is a schematic illustration of a single strand template bound to a primer which is in turn attached to a solid support.

Figure 3 illustrates a set of consecutive oligonucleotide primers for a part of the polynucleotide of interest following immediately after the primer illustrated in Figure 2.

Figure 4 illustrates the single base pair additions to all the primers illustrated in Figure 3, as well as the corresponding additions for the corresponding primers related to the complementary strand of the polynucleotide of interest.

Figure 5 is a graphic depiction of the length of extended primers formed utilizing free oligonucleotide primers annealed to a polynucleotide of interest.

Figures 6A, 6B and 6C are graphic depictions of electrophoretograms demonstrating the detection of the presence of a mutation in a polynucleotide of interest.

Figures 7A, 7B and 7C depict the results of a DNA chip-based analysis for a five-base region within the third exon of the HPRT gene.

Detailed Description of the Invention

The current invention pertains to methods for analyzing the nucleotide sequence of a polynucleotide of interest. The method comprises hybridizing all or a fragment of a polynucleotide of interest to oligonucleotide primers, conducting single base extension reactions, and detecting the single base extension events. The method can be used to analyze the sequence of a polypeptide of interest by examining the sequence for the presence of mutations or alterations in the nucleotide sequence, or by determining the sequence of a polypeptide of interest.

As used herein, the term "polynucleotide of interest" refers to the particular polynucleotide for which sequence information is wanted. Representative polynucleotides of interest include oligonucleotides, DNA or DNA fragments, RNA or RNA fragments, as well as genes or portions of

genes. The polynucleotide of interest can be single- or double-stranded. The term "template polynucleotide of interest" is used herein to refer to the strand which is analyzed, if only one strand of a double-stranded polynucleotide is analyzed, or to the strand which is identified as the first strand, if both strands of a double-stranded polynucleotide are analyzed. The term, "complementary polynucleotide of interest" is used herein to refer to the strand which is not analyzed, if only one strand of a double-stranded polynucleotide is analyzed, or to the strand which is identified as the second strand (i.e., the strand that is complementary to the first (template) strand), if both strands of a double-stranded polynucleotide are analyzed. Either one of the two strands can be analyzed. In a preferred embodiment, both strands of a double-stranded polypeptide of interest are analyzed in order to verify sequence information obtained from the template (first) strand by comparison with the complementary (second) strand. Nevertheless, it is not always necessary to analyze both strands. For example, if the polynucleotide of interest is being analyzed for the presence of a single base mutation, and not for the complete base sequence in the mutation region, it is sufficient to analyze a single strand of the polynucleotide of interest.

The methods of the current invention can be used to identify the presence of mutations or alterations in the nucleotide sequence of a polypeptide of interest. To identify mutations or alterations, the sequence of the polynucleotide of interest is compared with the sequence of the native or normal polynucleotide. An "alteration" in the polynucleotide of interest, as used herein, refers to a deviation from the expected sequence (the sequence of the native or normal polynucleotide), including deletions, insertions, point mutations, frame-shifts, expanded

oligonucleotide repeats, or other changes. The portion of the polynucleotide of interest that contains the alteration is known as the "altered" region. The methods can also be used to determine the sequence of a polypeptide of interest having a previously unknown nucleotide sequence.

In one embodiment of the current invention, the polynucleotide of interest is analyzed by annealing the polynucleotide to an array comprising sets of oligonucleotide primers. The oligonucleotide primers in the array have a length N, where N is from about 7 to about 30 nucleotides, inclusive, and is preferably from 20 to 24 nucleotides, inclusive. Each oligonucleotide primer within each set differs by one base pair. The oligonucleotide primers can be prepared by conventional methods (see Sambrook et al., *Molecular Cloning: A Laboratory Manual* (2nd Ed, 1989)). The sets of oligonucleotide primers are arranged into an array, such that the position and nucleotide content of each oligonucleotide primer on the array is known.

The size and nucleotide content of the oligonucleotide primers in the array depend on the polynucleotide of interest and the region of the polynucleotide of interest for which sequence information is desired. To analyze a polynucleotide of interest for the presence of alterations, consecutive primers differing by one base pair at the growing end and capable of hybridizing successively along the relevant part(s) or the whole of the polynucleotide are used. An example of such a primer set is shown in Figure 1. If only one or a few specific positions of the polynucleotide sequence are examined for alterations, the necessary array of oligonucleotide primers covers only the mutation regions, and is therefore small. If the whole or a major part of the polynucleotide of interest is to be analyzed for possible mutations at varying positions, the necessary array is larger. For example, the whole

hypoxanthine-guanine phosphoribosyl-transferase (HPRT) gene can be covered by 900 primers, arranged in a 30 X 30 array; the whole p53 gene requires 700 primers. If both strands of a double-stranded polynucleotide of interest are analyzed for the presence of alterations, the array comprises consecutive oligonucleotide primers for the suspected mutation region of both the template polynucleotide of interest and the complementary polynucleotide of interest. If the polynucleotide of interest has not been sequenced previously, the array includes oligonucleotide primers comprising all possible N-mers.

The array of sets of oligonucleotide primers is immobilized to a solid support at defined locations (i.e., known positions). The immobilized array is referred to as a "DNA chip", which is the apparatus of the current invention. The solid support can be a plate or chip of glass, silicon, or other material. The solid support can also be coated, such as with gold or silver. Coating may facilitate attachment of the oligonucleotide primers to the surface of the solid support. The oligonucleotide primers can be bound to the solid support by a specific binding pair, such as biotin and avidin or biotin and streptavidin. For example, the primers can be provided with biotin handles in connection with their preparation, and then the biotin-labelled primers can be attached to a streptavidin-coated support. Alternatively, the primers can be bound by a linker arm, such as a covalently bonded hydrocarbon chain, such as a C_{10-20} chain. The primers can also be bound directly to the solid support, such as by epoxide/amine coupling chemistry (see Eggers, M.D. et al., Advances in DNA sequencing Technology, SPIE conference proceedings, January 21, 1993). The solid support can be reused, as described in greater detail below.

In another embodiment of the invention, the polynucleotide of interest is analyzed by annealing the polynucleotide to one or more specific oligonucleotide primers that are not attached to a solid support; such oligonucleotide primers are referred to herein as "free oligonucleotide primers". If free oligonucleotide primers are used, the polynucleotide of interest can be attached to a solid support, such as magnetic beads. The free oligonucleotide primers have a length N, as described above, and are prepared by conventional methods (see Sambrook et al., *Molecular Cloning: A Laboratory Manual* (2nd Ed, 1989)). The size and nucleotide content of the free oligonucleotide primers depend on the polynucleotide of interest and the region of the polynucleotide of interest for which sequence information is desired. To analyze a polynucleotide of interest for the presence of alterations, primers capable of hybridizing immediately adjacent to the relevant part(s) of the polynucleotide are used. If more than one position of the polynucleotide sequence is examined for alterations, the free oligonucleotide primers are mutually distinguishable: i.e., the oligonucleotide primers have different mobilities during gel electrophoresis. In a preferred embodiment, oligonucleotides of different lengths are used. For example, an oligonucleotide primer of 10 nucleotides in length is designed to hybridize immediately adjacent to one putative mutation, and an oligonucleotide primer of 12 nucleotides in length is designed to hybridize immediately adjacent to a second putative mutation. Because the oligonucleotide primers are of different lengths, they will migrate to different positions on the gel. Thus, in this manner, the nucleotide content of each oligonucleotide primer can be identified by the position of the oligonucleotide primer on the gel.

The polynucleotide of interest is hybridized to the array of oligonucleotide primers, or to the free nucleotide primers, under high stringency conditions, so that an exact match between the polynucleotide of interest and the oligonucleotide primers is obtained, without any base-pair mismatches (see Sambrook et al., *Molecular Cloning: A Laboratory Manual* (2nd Ed, 1989)). For example, a schematic illustration of a hypothetical polynucleotide of interest annealed to an oligonucleotide primer that is attached to a solid support is shown schematically in Figure 2. In Figure 2, a part of the sequence of the polynucleotide of interest that follows immediately after the portion of the polynucleotide that is bound to the oligonucleotide primer on the array is shown as TGCAACTA. Six corresponding consecutive primers are shown in Figure 3, i.e. primers ending with the pairing bases A, AC, ACG, etc. If the polynucleotide of interest is double-stranded, it can be separated into two single strands either before or after the binding of the polynucleotide of interest to the array oligonucleotide primers. Both the template and the complementary polynucleotide of interest can be analyzed utilizing a single array. Thus, while not shown in Figure 2, appropriate primers corresponding to the complementary polynucleotide of interest are also attached to the solid support in known positions.

When the polynucleotide of interest is hybridized to the array of sets of oligonucleotide primers, or to the free oligonucleotide primers, under hybridization conditions, annealed primers are formed. The term, "annealed primer", as used herein, refers to an oligonucleotide primer (either free or attached to a solid support) to which a polynucleotide of interest is hybridized. The annealed primers are subjected to a single base extension reaction. The "single base extension reaction", as used herein, refers to a reaction in which

the annealed primers are provided with a reaction mixture comprising a DNA polymerase, such as T7 polymerase, and terminating nucleotides under conditions such that single terminating nucleotides are added to each of the annealed primers. The term "terminating nucleotides", as used herein, refers to either single terminating nucleotides, or units of nucleotides, the units preferably being dinucleotides. In a preferred embodiment, the terminating nucleotides are single dideoxynucleotides. The terminating nucleotides can comprise standard nucleotides, and/or nucleotide analogues. The terminating nucleotide added to each annealed primer is thus a base pairing with the template base on the polynucleotide of interest, and is added immediately adjacent to the growing end of the respective primer. An oligonucleotide primer to which a terminating nucleotide has been added through the single base extension reaction is termed an "extended primer". Thus, as schematically shown for both strands of the hypothetical polynucleotide of interest in Figure 4, a single nucleotide is added to each primer in the array; the primer set related to the strand illustrated in Figure 2 is shown to the left in Figure 4, and the other (complementary) strand is shown to the right. The nucleotides added are shown in extra bold type.

The terminating nucleotides preferably comprise dNTPs, and particularly comprise dideoxynucleotides (ddNTPs), but other terminating nucleotides apparent to the skilled person can also be used. If the terminating nucleotides are single nucleotides, then nucleotides corresponding to each of the four bases (A, T, G and C) are utilized in the single base extension reaction. If the terminating nucleotides are dinucleotide units, for example, then nucleotides corresponding to each of the sixteen possible dinucleotides are utilized.

The nucleotides are mutually distinguishable. For example, if the solid support is coated with a free electron metal, such as with gold or silver, surface plasmon resonance (SPR) microscopy allows identification of each nucleotide, by the change of the refractive index at the surface caused by each base extension. Alternatively, at least one of the terminating nucleotides is labelled by standard methods to facilitate detection. Suitable labels include fluorescent dyes, chemiluminescence, and radionuclides. The number of nucleotides that are labelled can be varied. It is sufficient to use three labelled terminating nucleotides, the fourth terminating nucleotide being identified by its "non-label", if single nucleotides are added in the base extension reaction. For example, if one is examining the polynucleotide of interest for the presence of a particular alteration, and not for the complete base sequence in the altered region, three labelled terminating nucleotides are sufficient. Fewer than three labels can also be utilized under appropriate circumstances. An exemplification of the use of two labelled and two unlabelled dNTPs is described below. If a specific alteration is to be investigated, such as a point mutation, only the native or normal nucleotide need be labelled, as a mutation would be indicated by the presence of the "non-label". Alternatively, the expected mutant nucleotide can also be labelled.

After the single base extension reaction has been performed, the identity and location of each terminating nucleotide is observed. If free oligonucleotide primers are used, the extended primers are eluted and separated by gel electrophoresis, and the gel is then analyzed. If oligonucleotide primers attached to an array are used, the array itself is analyzed. The gel or array is analyzed by detecting the labelled, terminating nucleotides bound to the oligonucleotide primers. The labeled, terminating

nucleotides are detected by conventional methods, such as by an optical system. For example, a laser excitation source can be used in conjunction with a filter set to isolate the fluorescence emission of a particular type of terminating nucleotide. Either a photomultiplier tube, a charged-coupled device (CCD), or another suitable fluorescence detection metnod can be used to detect the emitted light from fluorescent terminating nucleotides.

The sequence of the polynucleotide of interest can be analyzed from the label pattern observed on the array or on the gel, since the position of each different primer on the array or on the gel is known, and since the identity of each terminating nucleotide can be determined by its specific label. The label and position of each terminating nucleotide either within the array or on the gel will directly define the sequence of the polynucleotide of interest that is being analyzed. Mutations or alterations in the sequence of the polynucleotide of interest are indicated by alterations in the expected label pattern. For example, assume that the nucleotide sequence shown in Figure 2 contains a mutation: the third base C from the left is replaced by a G in the polynucleotide of interest. The top primer in Figure 3 will still be extended by a C as shown in Figure 4, whereas the next primer will be extended by a C rather than a G. Since this new, unexpected base C can be identified by its specific label and the respective primer location is known, the corresponding base mutation is identified as G.

The following simple example illustrates the ability to obtain complete sequence information and to identify a mutation in a representative polynucleotide of interest. The example utilizes two labelled terminating nucleotides, which give complete sequence information.

Assume a normal polynucleotide of the following base pair composition:

```
+ A C T G C T T A G
- T G A C G A A T C
```

and a corresponding mutant polynucleotide having the following base pair composition, which has a single base mutation in the third base pair:

```
+ A C C G C T T A G
- T G G C G A A T C.
```

Using fluorescent labelling, for example, with a red label ("R") for terminating A and a green label ("G") for terminating C, and no label, i.e., null ("N") for the remaining bases T and G, the following "binary" codes allowing sequence interpretation would be obtained for the normal, mutant and heterozygote sequences, respectively:

```
+ N-G-R-N-G-R-R-N-N      Normal
- R-N-N-G-N-N-N-R-G

+ N-G-G-N-G-R-R-N-N      Mutant (Affected)
- R-N-N-G-N-N-N-R-G

      R
+ N-G-G-N-G-R-R-N-N      Heterozygote (Carrier)
- R-N-N-G-N-N-N-R-G.
```

The presence of such a point mutation will affect the base pairing of the next few oligonucleotide primers to the polynucleotide of interest, and thereby the primer extensions obtained, such that the bases in the vicinity of the mutation (i.e., in the altered region) may not be accurately identified. To optimize identification of bases in the altered region, it is preferred to analyze both strands of such a double-stranded polynucleotide of interest. The few bases that may be difficult to identify on the template polynucleotide of interest, as well as the changed base, will be identified by the base extensions of the primers for the complementary polynucleotide of

interest, as the analysis of the complementary polynucleotide of interest approaches the mutation site from the opposite direction. In the nearest regions on either side of the alteration, the sequence determination is thereby provided by the oligonucleotide primers for one of the two strands.

The sequence of a polynucleotide of interest for which the sequence is previously known can be determined using methods similar to those described above in reference to identification of mutations utilizing an array of oligonucleotide primers. As before, the positions of the terminating nucleotides within the array will directly define the sequence position of each nucleotide in the polynucleotide of interest.

To determine the oligonucleotide sequence, one annealed primer is selected to be the "starting" annealed primer; it is supposed for purposes of analysis that the sequence of the polynucleotide of interest "starts" with this primer. The nucleotide which has been added to the starting annealed primer is detected using standard methods. Then, a second annealed primer which has the same nucleotide sequence as the starting annealed primer, minus the 5′ nucleotide and with the addition of the added nucleotide, is then selected. The terminating nucleotide which has been added to the second annealed primer is detected. These steps are then repeated, using the second annealed primer as the "starting" annealed primer in each repetition, until the sequence of the polynucleotide of interest is determined. For example, if the oligonucleotide primers are 10 nucleotides in length (N = 10), the starting annealed primer is chosen to correspond to the first ten bases of the sequence. The terminating nucleotide of the starting annealed primer is then determined. Next, bases 2-11 (i.e., bases 2-10 of the starting annealed primer plus the terminating nucleotide

extension) are matched to another annealed primer. This primer is the second annealed primer. The terminating nucleotide of the second annealed primer is then determined. These steps are repeated to determine the complete sequence. In this manner, the single base extension reaction automatically links together the set of annealed primers.

After analysis of the polypeptide of interest, the polynucleotide of interest and the terminating nucleotides can be removed from the DNA chip, so that the chip can be reused. In a preferred embodiment, the added terminating nucleotides are capable of being removed from the solid support after analysis of the polynucleotide of interest has been completed. Once the nucleotides are removed, the solid support with the immobilized oligonucleotide primers can be used for a new analysis. The nucleotides can be removed using standard methods, such as enzymatic cleavage or chemical degradation. Enzymatic cleavage, for example, would use a terminating nucleotide which can be removed by an enzyme. The single base extension reaction could result in addition to the oligonucleotide primers of RNA dideoxyTTP or RNA dideoxyCTP by reverse transcriptase or other polymerase. A C/T cleavage enzyme, such as RNase A, can then be used to "strip" off the RNA dideoxynucleotides. Alternatively, sulfur-containing dideoxy-A or dideoxy-G can be used during the single extension reaction; a sulfur-specific esterase, which does not cleave phosphates can then be used to cleave off the dideoxynucleotides. For chemical degradation, a chemically degradable terminating nucleotide can be used. For example, a modified ribonucleotide having its 2′- and 3′-hydroxyl groups esterified, such as by acetyl groups, can be used. After binding of the terminating nucleotide to the annealed primer, the acetyl groups are removed by treatment with a base to expose the 2′- and 3′-hydroxyl groups. The ribose

residue can then be degraded by periodate oxidation, and the residual phosphate group removed from the annealed primer by treatment with a base and alkaline phosphatase.

The method and apparatus of the current invention have uses in detecting mutations, deletions, expanded oligonucleotide repeats, and other genetic abnormalities. For example, the current invention can be used to identify frame shifting mutations caused by insertions or deletions. Furthermore, carrier status of heritable diseases, such as cystic fibrosis, β-thalassemia, α-1, Gaucher's disease, Tay Sach's disease, or Lesch-Nyham syndrome, can be easily determined using the current invention, because both the normal and the altered signals would be detected. Furthermore, mixtures of DNA molecules such as occur in HIV infected patients with drug resistance can be determined. The HIV virus may develop resistance against drugs like AZT by point mutations in the reverse transcriptase (RT) gene. When mutated viruses start to appear in the virus population, both the mutated gene and the normal (wild type) gene can be detected. The greater the proportion is of the mutant, the greater is the signal from the corresponding mutant terminating nucleotide.

The current invention is further exemplified by the following Examples.

EXAMPLE 1 Analyzing the Sequence of a Polynucleotide of Interest Utilizing Free Oligonucleotide Primers

An analysis of the hypoxanthine-guanine phosphoribosyl-transferase (HPRT) gene (the polypeptide of interest) was conducted for three individuals (Patients A, B, and C).

A. Obtaining the Polynucleotide of Interest

The polymerase chain reaction (see Sambrook et al., *Molecular Cloning: A Laboratory Manual* (2nd Ed, 1989), especially chapter 14) was utilized to amplify the polynucleotide of interest. During the reaction, one of the two PCR primers was tagged with a biotin group. Following amplification, the single strand template was captured with streptavidin coated magnetic beads. For a 50 µl PCR reaction, 25 µl of Dynal M-280 paramagnetic beads (Dynal A/S, Oslo, Norway) was used. The supernatant of the beads was removed and replaced with 50 µl of a binding and washing buffer (10 mM Tris-HCl (pH 7.5); 1 mM EDTA; 2 M NaCl). The PCR product was added to the beads and incubated at room temperature for 30 minutes for bead capture of the products. The single stranded polynucleotide of interest was isolated by the addition of 150 µl of 0.15 M NaOH for 5 minutes. The beads were captured, the supernatant was removed, and 150 µl of 0.15 M NaOH was again added for five minutes. Following denaturation, the beads were washed once with 150 µl of 0.15 M NaOH and twice with 1X T7 annealing buffer (40 mM Tris-HCl (pH 7.5); 20 mM $MgCl_2$; 50 mM NaCl). The beads were finally suspended in 70 µl of water. This process both isolates the single-stranded polynucleotide of interest and removes any unincorporated dNTPs remaining after PCR.

B. Analyzing the Sequence of the Polynucleotide of Interest

After single strand isolation, the oligonucleotide primers were annealed to the polynucleotide of interest by heating to 65°C for approximately two minutes and cooling to room temperature over approximately 20 minutes. The 10 µl reaction volume consisted of 7 µl of the polynucleotide

of interest (0.5-1 pmol), 2 μl of 5X T7 annealing buffer, and 1 μl of extension primer (3-9 pmol). The extension reaction was then performed. For the reaction, 1 μl of DTT, 2 μl of T7 polymerase (diluted 1:8) and 1 μl of ddNTPs (final concentration of 0.5 uM) were added. The reaction proceeded at 37°C for two minutes, and then was stopped by the addition of 100 μl of washing buffer (1XSSPE, 0.1% SDS, 30% ethanol). The beads were washed twice with 150 μl of the washing buffer. The extension products were eluted by the addition of 5 μl of formamide andheated to 70°C for two minutes. The beads were captured by the magnet and the supernatant containing the extension products was collected and analyzed on a ABI 373 (Applied Biosystems, Inc.). Oligonucleotide primers of lengths varying from 10 to 17 were used. As shown in Figure 5, extension products were formed efficiently.

1. Deoxynucleotide Labelling - Four Fluorophores

Each ddNTP was labelled by a different fluorophore. ABI Dye Terminator dyes designed for taq polymerase were used: ddG is blue, ddA is green, ddT is yellow, and ddC is red. Four fluorescent ddNTPs were added to each reaction tube. The extension products were purified, gel separated, and analyzed on an ABI 373. Two different bases of exon 3 of the HPRT gene were analyzed: base 16534 (wild type is A) and base 16620 (wild type is C).

The results of the four fluor, single lane, indicated that the presence of mutations could be identified easily. All three patients are wild type for A at base 16534 (data not shown). Electrophoretograms shown in Figures 6A, 6B and 6C indicate that Patient A is wild type (C) at base 16620 (Figure 6A), patient B is a mutated individual (C-->T) at base 16620 (Figure 6B), and patient C is a carrier at base 16620 (both C and T) (Figure 6C).

2. Deoxynucleotide Labelling - Single Fluorophore

Each ddNTP was labelled by the same fluorophore. DuPont NEN fluorescein dyes (NEL 400-404) were used. Each ddNTP appears blue in the ABI 373. Only one fluorescent ddNTP is added to each reaction tube. The extension products were purified, gel separated, and analyzed on an ABI 373. Four lanes on the gel must be used to analyze each base. Two different bases of exon 3 of the HPRT gene were analyzed: base 16534 (wild type is A) and base 16620 (wild type is C).

The results of the single fluor, four lane, demonstrated results that were identical to those obtained using the four fluor, single lane method described in (1), above. This type of assay minimizes the effect of the fluorophore differences during extension product formation and gel separation.

3. Deoxynucleotide Labelling - Biotinlyated Dideoxynucleotides

The ddNTPs are labelled with a biotin group. Four separate reactions are performed, whereby only one of the four ddNTPs is biotinylated. Following the extension reaction, a strepavidin (or avidin) coupled fluorescent group is attached to the biotinylated ddNTPs. Because the biotin group is small, uniform incorporation of the ddNTPs is expected and base-specific differences in extension are minimized. Furthermore, the fluorescent signal can be amplified because the biotin group can bind a steptavidin moiety coupled to multiple fluors.

EXAMPLE 2 Analyzing the Sequence of a Polynucleotide of Interest Utilizing Labelled Deoxynucleotides

An analysis of the hypoxanthine-guanine phosphoribosyl-transferase (HPRT) gene (the polypeptide of

interest) was conducted for three individuals (Patients A, B, and C). The third exon of the HPRT gene was examined.

Microscope glass slides were epoxysilanated at 80°C for eight hours using 25% 3′ glycidoxy propyltriethoxysilane (Aldrich Chemical) in dry xylene (Aldrich Chemical) with a catalytic amount of diisopropylehylamine (Aldrich Chemical), according to Southern (Nucl. Acids Res. 20:1679 (1992), and Genomics 13:1008 (1992)). The DNA chips were made by placing 0.5 μl drops of 5′-amino-linked oligonucleotides (50 μM, 0.1 M NaOH) at 37°C for six hours in a humid environment. The chips were washed in 50°C water for 15 minutes, dried and used. The annealing reaction consisted of adding 2.2 μl of single-stranded DNA (0.1 μM in T7 reaction buffer) to each grid position, heating the chip in a humid environment to 70°C and then cooling slowly to room temperature. A 1 μl drop of 0.1 M DTT, 3 units of Sequenase Version 2.0 (USB), 5 μCi α-^{32}P dNTP (3000 Ci/mmol) (DuPont NEN) and noncompeting unlabeled 18.5 μM ddNTPs (Pharmacia) were added to each grid position for three minutes. The reaction was stopped by washing in 75°C water, and analyzed on a PhosphorImager (Molecular Dynamics).

Figures 7A, 7B and 7C depict the results of a DNA chip-based analysis for a five-base region within the third exon of the HPRT gene. The rows correspond to a particular base under investigation, and the columns correspond to the labeled base. Figure 7A demonstrates the wild type sequence (TCGAG), Figure 7B demonstrates a C-->T mutation, and Figure 7C demonstrates a C-->T mutation.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described specifically herein. Such equivalents are intended to be encompassed in the scope of the following claims.

CLAIMS

What is claimed is:

1. A method of analyzing the sequence of a polynucleotide of interest, comprising the steps of:
   a) annealing a polynucleotide of interest to free oligonucleotide primers having known sequences of N nucleotides in length to generate annealed primers;
   b) subjecting the annealed primers to a single base extension reaction to extend the annealed primers by the addition of a terminating nucleotide;
   c) observing the identity of each terminating nucleotide that has been added to the annealed primers.

2. A method of analyzing the sequence of a polynucleotide of interest, comprising the steps of:
   a) annealing a polynucleotide of interest to oligonucleotide primers having known sequences of N nucleotides in length under hybridization conditions, to generate annealed primers;
   b) subjecting the annealed primers to a single base extension reaction which comprises providing to the annealed primers nucleotides corresponding to each of the four bases, to extend the annealed primers by the addition of a terminating nucleotide;
   c) observing the identity and location of each terminating nucleotide that has been added to the annealed primers.

3. A method of analyzing the sequence of a polynucleotide of interest, comprising the steps of:

a) attaching an array of oligonucleotide primers having known sequences of N nucleotides in length to a solid support at known locations;
b) annealing the polynucleotide of interest to the array of oligonucleotide primers to generate annealed primers;
c) subjecting the annealed primers to a single base extension reaction to extend the annealed primers by the addition of a terminating nucleotide;
d) observing the identity and location of each terminating nucleotide within the array on the solid support.

4. A method of analyzing the sequence of a polynucleotide of interest, comprising the steps of:
a) attaching an array of oligonucleotide primers having known sequences of N nucleotides in length to a solid support at known locations;
b) annealing the polynucleotide of interest to the array of oligonucleotide primers to generate annealed primers;
c) subjecting the annealed primers to a single base extension reaction to extend the annealed primers by the addition of a terminating nucleotide;
d) selecting a starting annealed primer;
e) observing the identity and location of the terminating nucleotide which has been added to the starting annealed primer, to determine the next nucleotide in sequence;
f) selecting a second annealed primer which has the same nucleotide sequence as nucleotides 2 through N of the starting annealed primer nucleotide plus the next nucleotide in sequence as determined in step (e), and

g) repeating steps (e) and (f), using the second annealed primer as the starting annealed primer for each repetition, to determine the sequence of the polynucleotide of interest.

5. A method of analyzing the sequence of a polynucleotide of interest, comprising the steps of:

a) attaching an array of oligonucleotide primers, having known sequences of N nucleotides in length to a solid support at defined locations;

b) annealing the polynucleotide of interest to the array of oligonucleotide primers under hybridization conditions, to generate annealed primers;

c) subjecting the annealed primers to a single base extension reaction which comprises providing to the annealed primers nucleotides corresponding to each of the four bases, to extend the annealed primers by the addition of a terminating nucleotide;

d) observing the identity and location of each terminating nucleotide within the array on the solid support.

6. A method of analyzing the sequence of a polynucleotide of interest, comprising the steps of:

a) attaching an array of oligonucleotide primers, having known sequences of N nucleotides in length to a solid support at defined locations;

b) annealing the polynucleotide of interest to the array of oligonucleotide primers under hybridization conditions, to generate annealed primers;

c) subjecting the annealed primers to a single base extension reaction which comprises providing to

the annealed primers nucleotides corresponding to each of the four bases, to extend the annealed primers by the addition of a terminating nucleotide;

d) selecting a starting annealed primer;

e) observing the identity and location of the terminating nucleotide which has been added to the starting annealed primer, to determine the next nucleotide in sequence;

f) selecting a second annealed primer which has the same nucleotide sequence as nucleotides 2 through N of the starting annealed primer nucleotide plus the next nucleotide in sequence as determined in step (e), and

g) repeating steps (e) and (f), using the second annealed primer as the starting annealed primer for each repetition, to determine the sequence of the polynucleotide of interest.

7. The method of any one of Claims 1 to 6, wherein the single base extension reaction comprises subjecting the annealed primers to a reaction mixture comprising a polymerase and nucleotides corresponding to each of the four bases.

8. The method of any one of Claims 5 to 7, wherein the nucleotides corresponding to each of the four bases are mutually distinguishable.

9. The method of Claim 8, wherein three of the four nucleotides are differently labelled.

10. The method of Claim 9, wherein the three differently labelled nucleotides are fluorescently labelled.

11. The method of any one of Claims 1 to 10, further comprising analyzing the sequence of the complementary polynucleotide of interest.

12. The method of any one of Claims 1 to 11, wherein the terminating nucleotides are dideoxynucleotides.

13. The method of any one of Claims 1 to 12, wherein the length N of the oligonucleotide primers is between 7 and 30 inclusive.

14. The method of any one of Claims 1 to 13, wherein the length N of the oligonucleotide primers is between 20 and 24 inclusive.

15. The method of any one of Claims 1, 2, 13 or 14, wherein the oligonucleotide primers comprise oligonucleotide primers of different lengths.

16. The method of any one of Claims 1 to 15, wherein observing the identity and location of a terminating nucleotide comprises the use of a charge coupled device or a photomultiplier tube.

17. The method of any one of Claims 3 to 14 or 16, wherein the terminating nucleotides are removed from the annealed primers after completed analysis to prepare the solid support for reuse.

18. The method of any one of Claims 1 to 17, wherein the terminating nucleotides are dinucleotides.

19. An apparatus for analyzing the sequence of a polynucleotide of interest, comprising a solid support

having attached thereon at defined locations an array of oligonucleotide primers having known sequences.

20. The apparatus of Claim 19, wherein the oligonucleotide primers are attached to the solid support by a specific binding pair.

21. The apparatus of Claim 20, wherein the specific binding pair is biotin and a molecule selected from the group consisting of: avidin and strepavidin.

REGION OF INTEREST

GATAGCAATC	GCTTACGGTA	ATCCGGCCTG
CTATCGTTAG	CGAATGCCAT	TAGGCCGGAC

SENSE PRIMERS	ANTI-SENSE PRIMERS
5'-GATAGCAATC-3'	3'-GAATGCCATT-5'
ATAGCAATCG	AATGCCATTA
TAGCAATCGC	ATGCCATTAG
AGCAATCGCT	TGCCATTAGG
GCAATCGCTT	GCCATTAGGC
CAATCGCTTA	CCATTAGGCC
AATCGCTTAC	CATTAGGCCG
ATCGCTTACG	ATTAGGCCGG
TCGCTTACGG	TTAGGCCGGA
CGCTTACGGT	TAGGCCGGAC

FIG. 1

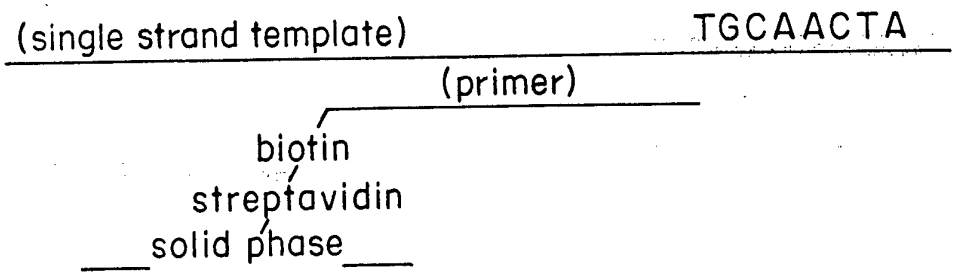


FIG. 2

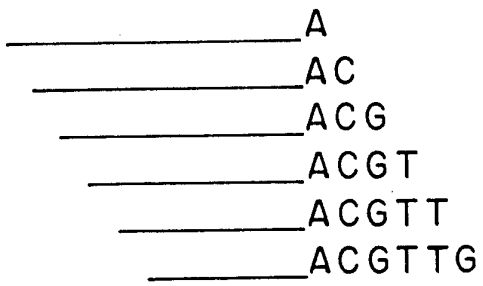


FIG. 3

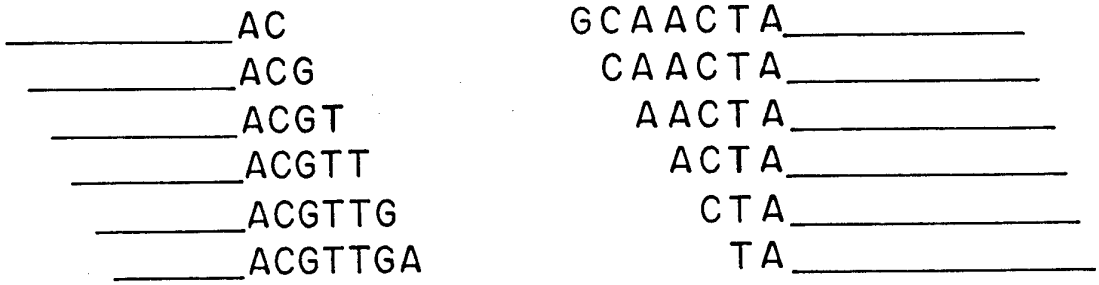


FIG. 4

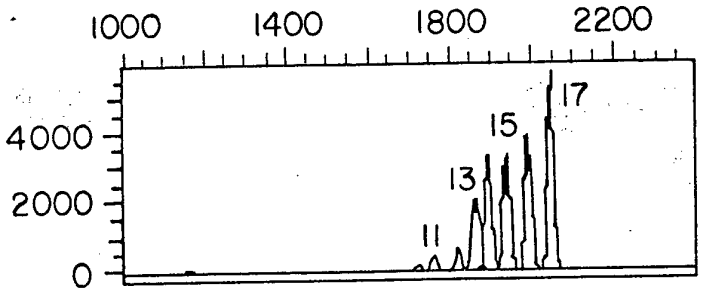


FIG. 5

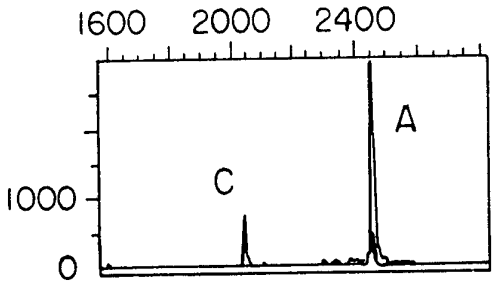


FIG. 6A

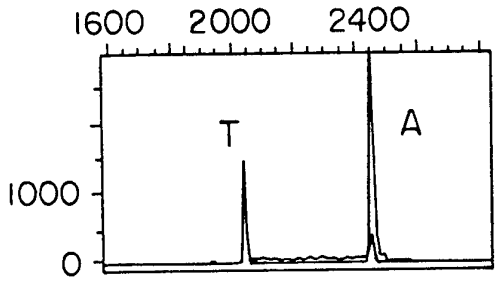


FIG. 6B

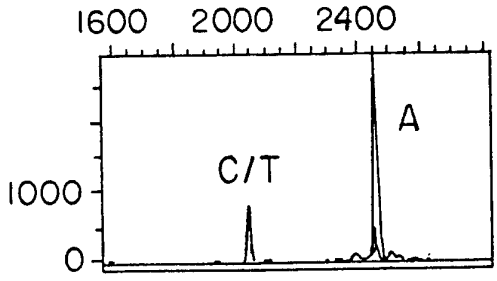


FIG. 6C

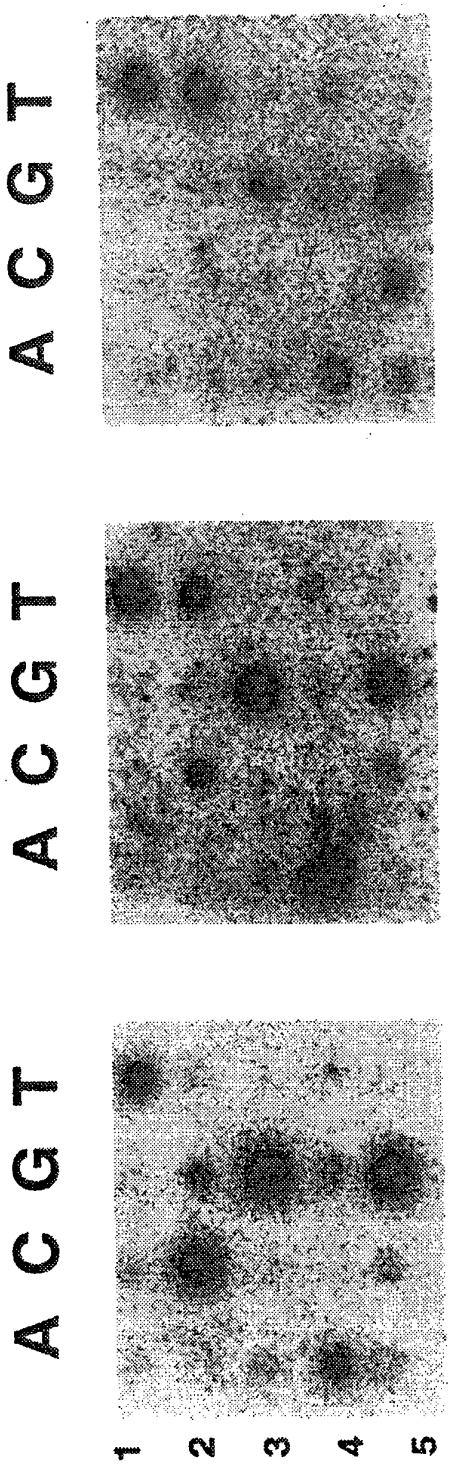


FIG. 7C

FIG. 7B

FIG. 7A

INTERNATIONAL SEARCH REPORT

Ir. tional Application No
PCT/US 94/07086

A. CLASSIFICATION OF SUBJECT MATTER

C 12 Q 1/68,C 12 Q 1/48

According to International Patent Classification (IPC) or to both national classification and IPC 5

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C 12 Q

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)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US, A, 5 202 231 (DRMANAC et al.) 13 April 1993 (13.04.93), abstract; claims. --	1-6
A	CHEMICAL ABSTRACTS, vol. 117, no. 23, issued 1992, December 07 (Columbus, Ohio, USA), R.J.MONNAT et al."Nucleotide sequence analysis of human hypoxanthine phosphoribosyl--transferase (HPRT) gene deletions", page 645, no. 231 604a; & Genomics 1992, 13(3), 777-87. --	1-6
A	CHEMICAL ABSTRACTS, vol. 113,	1-6,

[X] Further documents are listed in the continuation of box C.

[ ] 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
30 September 1994	27 -10- 1994

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	SCHNASS e.h.

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

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 94/07086

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	no. 25, issued 1990, December 17 (Columbus, Ohio, USA), A.EDWARDS et al."Automated DNA sequencing of the human HPRT locus", page 164, no. 225 422r; & Genomics 1990, 6(4), 593-608. --	19
A	CHEMICAL ABSTRACTS, vol. 104, no. 25, issued 1986, June 23 (Columbus, Ohio, USA), S.H.KIM et al." The organiza--tion of the human HPRT gene", page 161, no. 220 010j; & Nucleic Acids Res. 1986, 14(7), 3103-18. ----	1-6

ANHANG

zum internationalen Recherchenbericht über die internationale Patentanmeldung Nr.

ANNEX

to the International Search Report to the International Patent Application No.

PCT/US 94/07086 SAE 93215

ANNEXE

au rapport de recherche international relatif á la demande de brevet international n°

In diesem Anhang sind die Mitglieder der Patentfamilien der im obengenannten internationalen Recherchenbericht angeführten Patentdokumente angegeben. Diese Angaben dienen nur zur Unterrichtung und erfolgen ohne Gewähr.

This Annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The Office is in no way liable for these particulars which are given merely for the purpose of information.

La présente annexe indique les membres de la famille de brevets relatifs aux documents de brevets cités dans le rapport de recherche international visée ci-dessus. Les renseignements fournis sont donnés á titre indicatif et n'engagent pas la responsibilité de l'Office.

Im Recherchenbericht angeführtes Patentdokument Patent document cited in search report Document de brevet cité dans le rapport de recherche	Datum der Veröffentlichung Publication date Date de publication	Mitglied(er) der Patentfamilie Patent family member(s) Membre(s) de la famille de brevets	Datum der Veröffentlichung Publication date Date de publication
US A 5202231	13-04-93	YU A 570/87	31-08-90