(54) Title: METHOD FOR DETECTING SINGLE NUCLEOTIDE POLYMORPHISMS

(57) Abstract: The present invention relates to a method which can be used for detecting mutations, in particular single nucleotide polymorphisms in parallel in different nucleotide sequences. The method comprises the procedural steps of hybridizing d(G$^{Me}$eATC) containing single-stranded sample nucleotide sequences with single 10 stranded reference nucleotide sequences, fixing single-stranded reference nucleotide sequences or single-stranded sample nucleotide sequences before or during the hybridization, or heteroduplexes consisting of reference and sample nucleotide sequences after or during the hybridization, on a surface, incubating with mutS, mutL and mutH for the incision of the complementary region of the is d(G$^{Me}$ATC)-sequence in the reference nucleic acid sequence, if a mismatch is present in the hybridized reference and sample nucleic acid sequences, removing the 3'-ends of the incised reference nucleic acids and incubating with a DNA ploymerase and with labeled esoxynucleotidetriphosphates.

5    Description:

**Method for detecting Single Nucleotide Polymorphisms**

The present invention relates to a method which can be used for detecting

10    mutations, in particular Single Nucleotide Polymorphisms, by enzymatic

processing in parallel in different nucleic acids on a solid surface.

After the completion of the human genome project, the major goal of

pharmacogenomic research will be the continuous discovery of genetic variations

15    within the human population, and the subsequent correlation of these variations

with particular phenotypes. This may allow to develop better diagnostic tools, safer

medications, and will help to study the biological functions of genes. Large-scale

sequencing projects are in progress that aim to draw a map of the most frequently

occuring *Single Nucleotide Polymorphisms* (SNPs), the type of mutation that is

20    expected to contribute most to the genetic variation within the human population.

Several applications, however, require a more targeted approach in which only a

limited number of genes from many individuals is scanned for SNPs that occur in

a medium to low frequency within the human population. If, for instance, the

phenotype of a group of patients suggests that a particular molecular pathway is

25    defective, only those genes that are involved in that pathway might be scanned for

potential SNPs. Several methods for the *de novo* identification of SNPs and other

genetic variations have been developed to circumvent the time-consuming direct

sequencing step (Schafer, A.J. & Hawkins, J.R. DNA variation and the future of

human genetics. *Nature Biotechnol.* 16, 33-39 (1998)), but none of them is

30    compatible with a rapid, parallelized screening on a surface, in particular on a

microarray.

In particular, several protocols use the ability of mismatch repair proteins of the

mutS family to bind to mutant DNA. In these protocols, mismatches are generated

2

by the hybridisation of a DNA that serves as a reference sequence, with the complementary sequence from a different individual; if the two sequences differ, a mismatch is generated, which is discriminated by stronger mutS binding (Jiricny, J., Su, J.J., Wood, S.G. & Modrich, P. Mismatch-containing oligonucleotide

5      duplexes bound by the *E.coli mutS*-encoded protein. *Nucl. Acids Res.* 16, 7843-7853 (1988); Taylor, G.R. & Deeble, J. Enzymatic methods for mutation scanning. *Genetic Analysis: Biomolecular Engineering* 14, 181-186 (1999)).

As DNA microarrays are the method of choice for the parallel analysis of multiple

10     sequences an advanced method for the parallel analysis of DNA heteroduplexes using a fluorescence-labeled mutS protein was developed (PCT/EP01/08127).

On the other hand several methods have been published that make use of cellular mismatch repair systems. In some of the described methods a mismatch repair

15     system consisting of the mutS, mutL and mutH proteins is used. In US 5,376,526 a method for genetic mapping is described wherein nicks are introduced into heterohybrids having mismatches. Finally the gapped heterohybrids are separated from ungapped hybrids for mapping. US 5,571,676 describes a method for mismatch directed sequencing wherein a mutation is identified by sequencing the

20     gapped DNA strands and comparison with the wild type strand. US 5,556,750 and US 5,922, 539 describe methods for eliminating DNA duplex molecules containing base pair mismatches by removing molecules with single stranded regions or by degradation of cleaved heterohybrids. WO 96/41002 describes methods for mismatch dependent DNA sequencing and positional cloning. However, all these

25     methods are not adaptable to a parallel working microarray format, e.g. because of the necessary separation, degradation or sequencing of cleaved or gapped DNA strands, and therefore they are not compatible to high-throughput diagnostic platforms for the parallel analysis of multiple patients with high specificity.

30     Consequently, the invention is based on the object of making available a new method for detecting mutations, in particular Single Nucleotide Polymorphisms (SNPs) in nucleotide sequences, which permits in particular a high sample throughput in a short time and with a high degree of reliability.

It is possible to provide such a method for detecting mutations in nucleotide sequences comprising the procedural steps of

i    hybridizing a single-stranded, d(G$^{Me}$ATC) containing sample nucleotide sequence with a single-stranded reference nucleotide sequence,

ii   fixing single-stranded reference nucleotide sequences or single-stranded sample nucleotide sequences before or during the hybridization, or heteroduplexes consisting of reference and sample nucleotide sequences after or during the hybridization, on a support in a site-resolved manner,

iii  incubating with mutS, mutL and mutH,

iv   incision of the complementary region of the methylated d(G$^{Me}$ATC)-sequence, if a mismatch is present in the hybridized sample nucleotide sequence,

v.   removal of the 3'-end of the incised reference nucleic acids,

vi.  incubating the hybridized sample nucleotide sequences with desoxynucleotidetriphosphates (dNTPs), wherein at least one of said dNTPs contains a detectable label, and with a exonuclease deficient DNA polymerase,

vii. detecting the reference nucleotide sequences comprising a labeled 3'-end.

In a preferred method the 5'-ends of the sample nucleic acid sequence and of the reference nucleic acid sequence are protected. For example, the 5'- ends could be blocked by substituting the 5'-phosphate group by a thiol group during oligonucleotide synthesis.

The following term definitions are introduced for the further description of the invention:

The expression "reference nucleotide sequence" denotes a nucleotide sequence, preferably a DNA sequence, which is used as a comparison sequence;

A "sample nucleotide sequence" is a labeled nucleotide sequence, preferably a DNA sequence, which is to be examined for mutations.

The d(G methyl-A T C) containing sample nucleotide sequences and the reference sequences are obtainable by amplifying a DNA sample with known sequence by PCR for the preparation of the reference nucleic acid sequence and amplifying a DNA sample with unknown sequence for the preparation of the sample nucleotide sequence wherein a primer containing a d($G^{Me}$ATC) sequence is used. Typical primers are polymers of 15 to 35 nucleotides.

The hybridization of the d($G^{Me}$ATC) containing sample nucleotide sequences with single-stranded reference nucleotide sequences can be carried out as described in Frederic N., Ausubel, Short Protocols in Molecular Biology, Section 2.10, Wiley + Sons.

It is possible to provide methods which are particularly suitable as regards increasing sample throughput on the base of an electronically addressable surface in combination with mutS which recognizes mispairings and mutL and mutH, with it being possible to find mutation-specific mispairings reliably and considerably more rapidly than when using conventional passive hybridization techniques.

In particular microelectronic arrays allow a site-specific and individualized hybridisation of nucleic acid strands. Hence they are ideal tools for the generation of heteroduplex nucleic acids by first addressing a biotinylated reference nucleic acid with known sequence to an individual test site, and subsequently addressing a labelled, complementary sample nucleic acid from e.g. a patient sample to the same surface. If the sample DNA contains a mutation when compared to the reference DNA, the heteroduplex will contain a mismatch. This mismatch is site-specific detectable by the present method using the mutS, mutL and mutH recognition and incision system.

In this connection, the fixing of the single-stranded or double-stranded nucleotide sequences, and the hybridization, can be electronically controlled, in particular electronically accelerated.

A particularly preferred embodiment of the claimed method is characterized by a site-resolved, electronically accelerated hybridization, with the hybridization conditions, such as the current strength applied, the voltage applied or the duration of the electronic addressing, being set individually at the respective site.

5   At the same time as, or after, the hybridization, the base mispairing can be incised by adding mutS which recognizes mispairings, mutL and mutH which incises the $d(G^{Me}ATC)$ sequence complementary region of the reference nucleic acid.

The electronic addressing is effected by applying an electric field, preferably
10  between 1.5 V and 2.5 V in association with an addressing duration of between 1 and 3 minutes. Due to the electric charge on the nucleotide sequences to be addressed, their migration is greatly accelerated by an electric field being applied. In this connection, the addressing can be effected in a site-resolved manner; in this case, addressing takes place consecutively to different zones on the chip
15  surface. At the same time, different addressing and hybridization conditions can be set at the individual sites.

When carrying out the detection method according to the invention, nucleotide sequence heteroduplexes consisting of a predetermined nucleotide sequence, i.e.
20  the reference nucleotide sequence, and of the complementary nucleotide sequence from a physiological sample, i.e. the sample nucleotide sequence, are initially produced on a chip surface using electronic addressing. The mispairings which are formed in this connection indicate a mutation in the sample nucleotide sequence and can be incised using mutS which binds to the mispairing site, mutL
25  and mutH which incise the $d(G^{Me}ATC)$ sequence complementary region of the reference nucleic acid. Base mispairing-binding mutS, mutL and mutH are derived e.g. from *E.coli, T. thermophilus* or *T.aquaticus.*

In the method according to the invention, the reference nucleotide sequence, for
30  example, can be employed as a biotinylated oligonucleotide which is either synthesized or prepared by amplification using sequence-specific oligonucleotides, one of which is biotinylated at the 5' end. After that, the reference nucleotide sequence is converted into the single-stranded state by

melting, preferably in a buffer solution having a low salt content, and applied to a predetermined position on a chip by means of electronic addressing. Examples of suitable chips are those marketed by Nanogen (San Diego/USA). The reference nucleotide sequence can be applied, for example, using a Nanogen molecular

5    biology workstation, preferably using the parameters specified by the manufacturer. Unless otherwise indicated, Nanogen's chips and/or their molecular biology workstation is/are used in accordance with the manual which is supplied with them; the method of use is also described in Radtkey et al., Nucl. Acids Res. 28, 2000, e17.

10

The sample nucleotide sequence, which is complementary to the sequence which has already been applied to the chip, can now be loaded onto the chip which has been prepared in this way. For this purpose, dye-labeled and d(G$^{Me}$ATC)-labeled oligonucleotides are synthesized or generated by amplifying the sample

15   nucleotide acid sequence using sequence-specific oligonucleotides one of which is dye-labeled at the 5' end. In this connection, the dye-labeled sample nucleotide sequence constitutes the complementary strand to the biotinylated strand of the reference nucleotide sequence. The sample nucleotide sequence has also to be converted beforehand into the single-stranded state by being melted, for example

20   in a buffer solution having a low salt content, and then applied to the biotinylated reference nucleotide sequence by means of electronic addressing. This results in the formation, by hybridization, of a nucleotide sequence heteroduplex consisting of the reference nucleotide sequence and the sample nucleotide sequence. The heteroduplex can also be prepared on an electronically addressable surface, for

25   example using a Nanogen molecular biology workstation and employing the parameters specified by the manufacturer. Successful hybridization can be monitored optically, and at the same time determined quantitatively, by detecting the dye which is coupled to the heteroduplex.

30   Alternatively, the sample nucleotide sequence can also be biotinylated and electronically addressed, as just described. It is also possible to hybridize in solution, with subsequent electronic addressing and with one of the two nucleotide sequences of the heteroduplex being biotinylated. Apart from derivatizing with

biotin, it is also possible to use other molecular groups, which bind to an electronically addressable surface, for fixing nucleotide sequences. Thus, it is likewise possible, for example, to effect the fixing using introduced thiol groups, hydrazine groups or aldehyde groups.

The support material comprises in a preferred embodiment a permeation layer. E.g. the permeation layer is selected from the group consisting of silicon, silicon dioxide, silicon nitride, controlled porosity glass, metal, metal silicilide, inorganic sol-gels, and hydrogels. Preferred hydrogels are selected from the group consisting of agarose, polyacrylamide, polymethacrylamide, and organic polymer hydrogels.

If the sample nucleotide sequence now exhibits a mutation as compared with the reference nucleotide sequence, there will then be a mispairing in the heteroduplex. Then mutS, which recognizes these mispairings with a high degree of specificity, is used for binding such mispairings and mutL and mutH are added to incise the heteroduplex which is recognized by mutS. The mispairing-recognizing mutS proteins derived from *E.coli* and from *T. thermophilus* are particularly suitable for this purpose. The mispairing-recognizing mutS, mutL and the incising mutH are preferably added in excess. Furthermore it is appropriate to add cofactors such as ATP and $Mg^{2+}$ to the mutS, mut L and mutH comprising reaction mixture, supporting mutH to incise the hemimethylated $d(G^{Me}ATC)$ sequence of the reference nucleic acid and the hybridized sample nucleic acid. (Smith and Modrich, Proc. Natl. Acad. Sci. USA, Vol. 93, p. 4374-4379, 1996)

Preferably the reaction is carried out at low salt conditions, concentration of from 10 to 300 mM salt concentration, in particular from 10 to 150 mM is preferred.

After the incision of the reference nucleic acid sequences, if a mismatch is present in the hybridized reference – sample nucleic acid duplex, the 3`-ends of the incised reference nucleic acids are removed. Removing of said 3`-ends is preferably carried out by increasing the temperature, by washing with low ionic

8

strength or by an enzymatic digestion with a 5'>3' exonuclease, such as lambda exonuclease or T7 exonuclease. If an 5'>3' exonuclease is applied for removing the 3'-ends of the sample nucleic acid sequence 5'-blocked reference and sample nucleic acid sequences are necessary which can not be cleaved by nucleases.

Then the selective digestion of the 3'-end of the reference DNA starting from the new and unblocked 5' end that may be generated as a consequence of a mutH incision could be carried out by adding a 5'>3' exonuclease. By the removal of the 3'-ends of the reference nucleic acid sequence a 5'-overhang is generated which could act as a template for the DNA polymerization step.

The polymerization is carried out by adding a 3'>5' and 5'>3' exonuclease deficient DNA polymerase such as Klenow fragment 3'>5' exo⁻ and by applying desoxynucleotidetriphosphates (dNTPs), wherein at least one of said dNTPs contains a detectable label. The DNA polymerization takes specifically place on the prior mismatched heteroduplexes. The addition of a 3'>5' exonuclease deficient DNA polymerase is sufficient, if 5' protected sample and reference nucleic acids are hybridized. (Polymerization and digestion can be carried out as described in Frederic N., Ausubel, Short Protocols in Molecular Biology, Section 3.5 and 3.11, Wiley + Sons.)

The labeling moiety of the labeled dNTPs is selected from the group consisting of fluorescent moieties, visible dye moieties, radioactive moieties, chemiluminescent moieties. The resulting labeled heteroduplexes are detectable by fluorescence, chemiluminescence, visible light or radioactive radiation scanning.

An advantage of the present method is the reliability concerning the detection of mispairings, based on the mutH endonuclease which incises the reference nucleic acid region complementary to the methylated d(G^{Me}ATC) sequences in amplified DNA with high specificity and activity. Hence the use of the mismatch repair system will allow to detect mismatches with a greater sensitivity than with labeled mutS alone. An additional advantage is that the mutS protein binds to dsDNA

9

containing a mismatch at any position of the DNA sequence and mutS protein binds all kinds of mismatches (A:A, A:G, A:C, G:G, G:T, C:C, C:T, T:T).

Another important advantage of the present method is the possibility of simultaneously detecting mutations in parallel, in an integrated manner in one procedure, in particular on an active electronic array.

The electronic addressing can be effected, for example, on a chip, on which the nucleotide sequences A, B, C..., N are already fixed at sites a, b, c to n, using a mixture containing nucleotide sequences from the group A', B', C', ..., N'. In this case, the nucleotide sequences A/A' to N/N' in each case constitute a reference and sample nucleotide sequence pair. After the electronically accelerated hybridization, the stringency of the hybridization conditions can be increased, for example, by reversing the polarity of the electrical field. This can be effected in a site-resolved manner and consequently be adjusted individually in the case of each site.

On account of the high speed, the reliability of the method and on account of the high degree of parallelization which can be achieved, it is possible, using high sample throughput, to investigate many different samples from patients who are suffering from a hereditary disease. This facilitates the task of achieving a correlation between a clinical syndrome and particular mutations. In addition to this, it is possible to screen more efficiently for mutations which have been acquired during the course of life and which can be correlated with particular diseases.

In addition to the parallel analysis of multiple genes, one major benefit of the active electric microarray technology is that several individuals can be analyzed on the same chip, which is not possible on passive microarrays.

Another embodiment of the present invention are the d(G$^{Me}$ATC) containing nucleic acids for the detection of mutations, e. g. for the detection of SNPs, in particular on an active electronic microarray. A further embodiment are the

d(G^{Me}ATC) containing primers for the preparation of the d(G^{Me}ATC) containing sample nucleic acid sequences by PCR.

Brief Description of the Figures:

Figure 1 describes a preferred method for detecting mutations in nucleotide sequences. Therefore the reference nucleic acid (step (a)) and a sample nucleic acid (step (b)) are prepared by PCR. The PCR of step (b) is carried out with d(G^{Me}ATC) containing forward primers. The single-stranded, d(G^{Me}ATC) containing sample nucleotide sequences and the corresponding single-stranded reference nucleotide sequences are hybridized (step (c)). Then the resulting heteroduplex is incubated with mutS, mutL and mutH (mutSLH) wherein the exonuclease mutH is bound through mutL and the mismatch recognizing mutS to the mismatch containing heteroduplex (step (d)) and the hemimethylated heteroduplex is incised by the mutH exonuclease at the reference nucleic acid region which is complementary to the methylated d(G^{Me}ATC)-sequence of the sample nucleic acid (step (e)).

The 3'-end of the incised reference nucleic acids are removed (step (f)) and the hybridized sample nucleotide sequences are incubated with desoxynucleotidetriphosphates (dNTPs), wherein at least one of said dNTPs contains a detectable label (black circles (step (g)), and with a 3'>5' and 5'>3' exonuclease deficient DNA polymerase (step (g)). The resulting labeled reference nucleotide sequences are detectable by scanning methods.

Figure 2 shows a similar process as described in Figure 1 wherein the 5'-ends of the reference nucleic acid sequence and the sample nucleic acid sequence are protected. After the incision of the mismatch containing hetereoduplex the 3'-end of the reference nucleic acid sequence is removed by applying a 5'>3'-exonuclease (step (f)) followed by the incubation step g) as described in Fig. 1, wherein the addition of only a 3'>5' exonuclease deficient DNA polymerase is sufficient.

11

Patent Claims:

1.   A method for detecting mutations in nucleotide sequences comprising the
     procedural steps of

     i    hybridizing a single-stranded, d(G$^{Me}$ATC) containing sample nucleotide
          sequence with a single-stranded reference nucleotide sequence,

     ii   fixing single-stranded reference nucleotide sequences or single-stranded
          sample nucleotide sequences before or during the hybridization, or
          heteroduplexes consisting of reference and sample nucleotide
          sequences after or during the hybridization, on a support in a site-
          resolved manner,

     iii  incubating with mutS, mutL and mutH,

     iv   incision of the complementary region of the methylated d(G$^{Me}$ATC)-
          sequence, if a mismatch is present in the hybridized sample nucleotide
          sequence,

     v.   removal of the 3'-end of the incised reference nucleic acids,

     vi   incubating the hybridized sample nucleotide sequences with
          desoxynucleotidetriphosphates (dNTPs), wherein at least one of said
          dNTPs contains a detectable label, and with a exonuclease deficient
          DNA polymerase,

     vii  detecting the reference nucleotide sequences comprising labeled 3'-
          ends.

2.   The method of claim 1 wherein the sample nucleotide sequences are
     prepared by PCR with d(G$^{Me}$ATC) containing primers.

3.   The method of any one of claim 1 or 2 wherein the incubation with mutS,
     mutL and mutH is accompanied by the addition of ATP and Mg$^{2+}$.

4.    The method of any one of claims 1 to 3 wherein the 3'-ends of the incised reference nucleic acid sequences are removed by increasing temperature or by washing with low ionic strength.

5.    The method of any one of claims 1 to 4 wherein the 5'-ends of the sample nucleic acid sequence and of the reference nucleic acid sequence are protected.

6.    The method of claim 5 wherein the 3'-ends of the incised reference nucleic acid sequences are removed by addition of a 5'>3' exonuclease.

7.    The method of any one of claims 1 to 5 wherein a 3'>5' exonuclease deficient DNA polymerase is used.

8.    The method of any one of claims 1 to 5 wherein a 3'>5' and 5'>3' exonuclease deficient DNA polymerase is used.

9.    The method of any one of claims 1 to 8 wherein at least one of the added desoxynucleotidetriphosphates (dNTPs) contains a fluorescence label, a chemiluminescence label, a visible dye label, or a radioactive label.

10.   The method of any one of claims 1 to 9 wherein the support employed is an electronically addressable surface.

11.   The method of any one of claims 1 to 10 wherein the support material is a permeation layer over an electrode of an active electronic array.

12.   The method of any one of claims 1 to 11 wherein the support material comprises a material selected from the group consisting of silicon, silicon dioxide, silicon nitride, controlled porosity glass, metal, metal silicilide, inorganic sol-gels, and hydrogels.

13

13. Nucleic acid sequences comprising a d(G$^{Me}$ATC) sequence for the detection of mutations with mutH.

14. Use of nucleic acid sequences comprising a d(G$^{Me}$ATC) sequence for the detection of mutations with mutH.

15. Primer comprising a d(G$^{Me}$ATC) sequence for the preparation of d(G$^{Me}$ATC) containing sample nucleic acids for the detection of mutations with mutH.
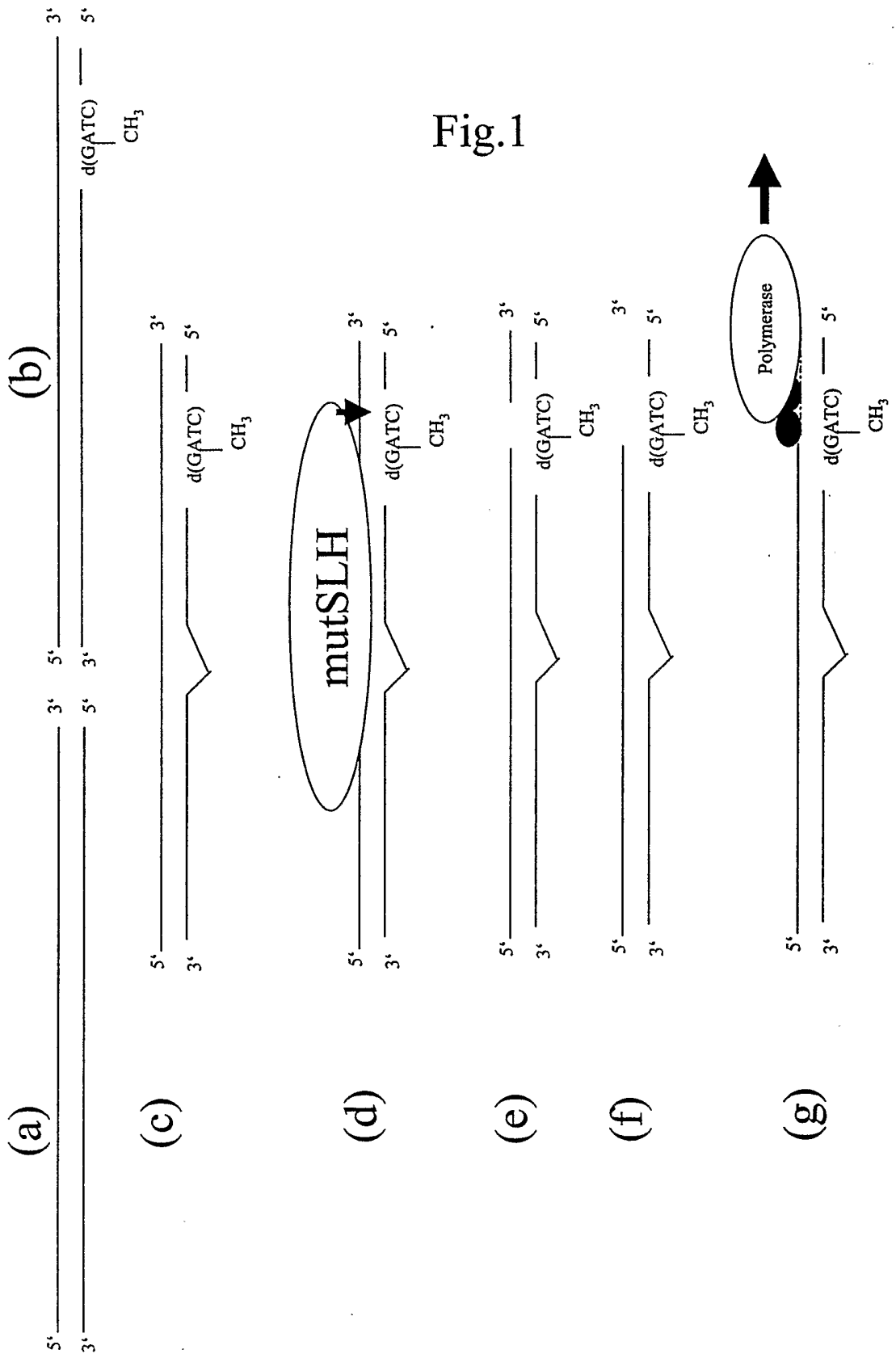
16. Use of primers comprising a d(G$^{Me}$ATC) sequence for the preparation of d(G$^{Me}$ATC) containing sample nucleic acids for the detection of mutations with mutH.
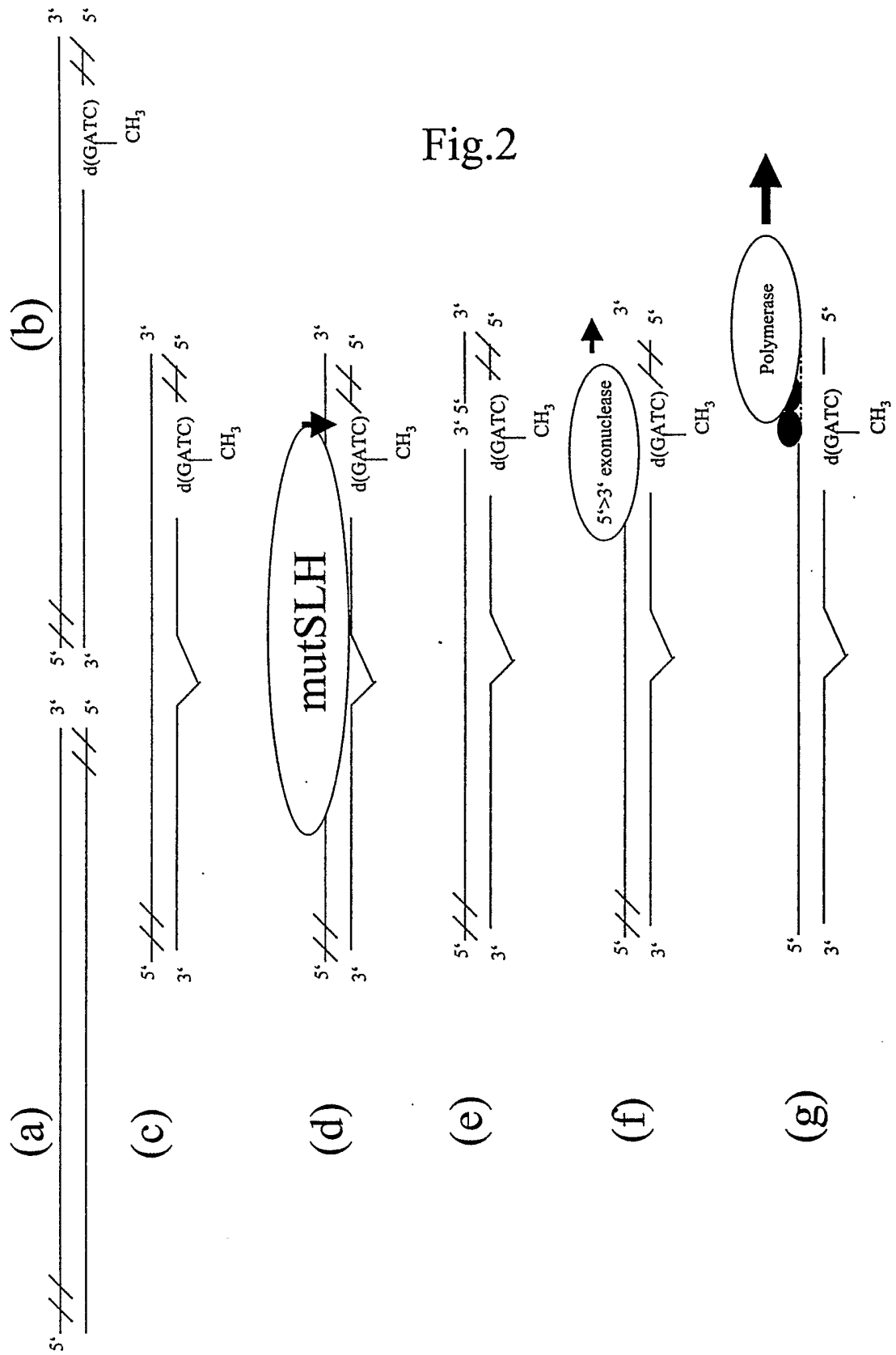
Fig.1

Fig.2