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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 March 2008 (20.03.2008)

PCT

(10) International Publication Number
WO 2008/032058 A2

(51) International Patent Classification:
C12N 15/09 (2006.01)

(21) International Application Number:
PCT/GB2007/003451

(22) International Filing Date:
13 September 2007 (13.09.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0618046.7 13 September 2006 (13.09.2006) GB

(71) Applicant (*for all designated States except US*): **LINGVITAE AS** [NO/NO]; Drammensveien 165, N-0277 Oslo (NO).

(71) Applicant (*for AG only*): **JAPPY, John, William, Graham** [GB/GB]; Gill Jennings & Every LLP, Broadgate House, 7 Eldon Street, London EC2M 7LH (GB).

(72) Inventor; and
(75) Inventor/Applicant (*for US only*): **LEXOW, Preben** [NO/NO]; Lingvitae AS, Drammensveien 165, N-0277 Oslo (NO).

(74) Agent: **GILL JENNINGS & EVERY LLP**; Broadgate House, 7 Eldon Street, London EC2M 7LH (GB).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

WO 2008/032058 A2

(54) Title: METHOD FOR SEQUENCING A POLYNUCLEOTIDE

(57) Abstract: The present invention is a method for sequencing a target polynucleotide, comprising the steps of: (i) ligating the target polynucleotide to a first polynucleotide; (ii) forming a concatemer comprising multiple copies of the product of step (i); (iii) attaching the concatemer to a second polynucleotide such that the second polynucleotide hybridises to portions of the concatemer, but not to regions on the concatemer corresponding to at least a portion of the target polynucleotide; (iv) interrogating one or more bases in those regions not hybridised to the second polynucleotide, to thereby identify the target polynucleotide sequence.

METHOD FOR SEQUENCING A POLYNUCLEOTIDE

Field of the Invention

This invention relates to methods for determining the sequence of polynucleotides.

Background to the Invention

Advances in the study of molecules have been led, in part, by improvement in technologies used to characterise the molecules or their biological reactions. In particular, the study of the nucleic acids DNA and RNA has benefited from developing technologies used for sequence analysis and the study of hybridisation events.

WO-A-00/39333 describes a method for sequencing polynucleotides by converting the sequence of a target polynucleotide into a second polynucleotide having a defined sequence and positional information contained therein. The sequence information of the target is said to be "magnified" in the second polynucleotide, allowing greater ease of distinguishing between the individual bases on the target molecule. This is achieved using "magnifying tags", which are predetermined units of nucleic acid sequence. Each of the bases adenine, cytosine, guanine and thymine on the target molecule is represented by an individual magnifying tag, converting the original target sequence into a magnified sequence. Conventional techniques may then be used to determine the order of the magnifying tags, and thereby determine the specific sequence on the target polynucleotide.

In a preferred sequencing method, each magnifying tag comprises a label, e.g. a fluorescent label, which may then be identified and used to characterise the magnifying tag.

WO-A-04/094664 describes an adaptation of the conversion method disclosed in WO-A-00/39333. In both methods, it is preferred that each magnifying tag comprises two units of distinct sequence which can be used as a binary system, with one unit representing “0" and the other representing “1". Each base on the target is characterised by a combination of the two units, for

example adenine may be represented by "0" + "0", cytosine by "0" +"1", guanine by "1" + "0" and thymine by "1" +"1".

One difficulty with the prior art methods is that the eventual read-out step is often hindered by the need to discriminate between the different
5 magnifying tags or units. It is therefore desirable to identify improvements which permit discrimination to occur.

Summary of the Invention

The present invention provides a method for analysing polynucleotides. The method utilises a concatemer of the target polynucleotide, i.e. repeating
10 the sequence of the target polynucleotide, and then interrogating the various target polynucleotides to reveal the target polynucleotide sequence. The intention is, preferably, to identify one base (nucleotide) of each target polynucleotide on the concatemer with different bases being identified for each target. In this way, all the bases to be identified are more separated
15 than if the bases of the original target polynucleotide were to be sequenced. Increasing the separation allows the eventual read-out technology to discriminate between the units, thereby improving the efficiency of the eventual sequencing/identification step.

According to a first aspect of the present invention, a method for
20 sequencing a target polynucleotide comprises the steps of:

(i) ligating the target polynucleotide to a first polynucleotide;

(ii) forming a concatemer comprising multiple copies of the product of step (i); and

(iii) interrogating one or more bases in multiple copies of the target
25 polynucleotide of the concatemer, to thereby identify the target polynucleotide sequence.

According to a second aspect of the present invention, a method for sequencing a target polynucleotide comprises the steps of:

(i) ligating the target polynucleotide to a first polynucleotide linker
30 of known sequence, to form a circular polynucleotide;

(ii) contacting the circular polynucleotide with a second polynucleotide under conditions which allow a polymerase reaction to

proceed, the second polynucleotide comprising two or more sequences that are complementary to the first polynucleotide except for a region adjacent to the target polynucleotide, such that the polymerised product comprises portions hybridised to the second polynucleotide and non-hybridised portions which correspond to the target polynucleotide and the region adjacent to the target polynucleotide, and

(iii) interrogating a plurality of the non-hybridised portions to identify one or more different bases corresponding to the target polynucleotide, to thereby identify the target polynucleotide sequence.

According to a third aspect of the present invention, a support surface comprises a double-stranded polynucleotide immobilised thereon, wherein one strand is a concatemer of repeating polynucleotide sequences having regions hybridised to the other strand and non-hybridised regions.

Brief Description of the Drawings

The invention is described with reference to the accompanying drawings, wherein:

Figure 1 illustrates the use of a circular polynucleotide to generate the concatemer of the target polynucleotide;

Figure 2 illustrates the hybridisation of a (third) polynucleotide to the sequence adjacent to the non-hybridised target, permitting interrogation with a labelled ddNTP, and

Figure 3 shows the subsequent incorporation of a labelled ddNTP.

Description of the Invention

The term "polynucleotide" is well known in the art and is used to refer to a series of linked nucleic acid molecules, e.g. DNA or RNA. Nucleic acid mimics, e.g. PNA, LNA (locked nucleic acid) and 2'-O-methRNA are also within the scope of the invention.

The reference herein to the bases A, T(U), G and C, relate to the nucleotide bases adenine, thymine (uracil), guanine and cytosine, as will be appreciated in the art. Uracil replaces thymine when the polynucleotide is

RNA, or it can be introduced into DNA using dUTP, again as well understood in the art.

The term "first polynucleotide" is used herein to refer to a polynucleotide of known sequence and length which is used to ligate to the target, preferably to circularise the ligated target. The first polynucleotide acts to provide separation between different copies of the target on the eventual concatemer. The target polynucleotide is linked at either its 5' or 3' end to the first polynucleotide, preferably at both the 5' and 3' ends to form the circular product.

The term "second polynucleotide" is used herein to refer to a polynucleotide intended to hybridise to regions of a concatemer formed with repeated target and first polynucleotide sequences. The second polynucleotide may also be referred to as a "masking" polynucleotide as it acts to prevent interrogation of those regions of the first polynucleotide to which it hybridises. The regions of the first polynucleotide that are not hybridised are said to be "unmasked". The second polynucleotide therefore comprises a repeated sequence complementary to the first polynucleotide, interspersed with a sequence which does not hybridise to either the target or the first polynucleotide. This ensures that the target sequence (or at least a portion of the target sequence) does not hybridise to the second polynucleotide and is therefore available for interrogation in a subsequent step. It is preferable that the second polynucleotide also has a sequence that does not hybridise to a portion of the sequence of the first polynucleotide adjacent to the target. This will be of known sequence and permits the hybridisation of a third polynucleotide sequence adjacent to the target using the second polynucleotide as its complement. This portion of the second polynucleotide will usually be downstream of the target and will typically be from 10 to 40 bases in size, more typically from 15 to 25 bases, and most typically 20 bases. This provides sufficient discrimination for the hybridisation to the third polynucleotide sequence.

The method of the present invention is used to convert a single target polynucleotide sequence into a series of polynucleotides which can each be

interrogated at intervals more spaced apart than that of a single target. This has the benefit of, in effect, separating the bases on the target to permit the ultimate interrogation and read-out steps to be performed with more accuracy and discrimination. The invention relies on the formation of a concatemer of
5 the target polynucleotide which permits subsequent interrogation to be performed on selected bases; the interrogated bases are representative of the bases on the original target polynucleotide.

Having formed the concatemer, one or more of the bases of the target polynucleotide can be interrogated in various ways to reveal their identity.
10 The intention may be to determine the full or partial sequence of the target. Separate individual bases on each target of the concatemer may be targeted and identified. Alternatively, the intention may be to determine a single base on the target, with the multiple targets of the concatemer being used as controls, to ensure that the identified signal is correct. This may be of use in
15 determining single nucleotide polymorphisms (SNPs). Accordingly, the term "sequencing" as used herein is to be given a broad meaning, to include the determination of a single base on the original target, or to determine two or more bases on the original target. In one embodiment, all bases on the original target are identified ultimately, with a single base being identified on
20 each interrogated target on the concatemer. It is preferable if a single base is interrogated (i.e. determined) within a single target on the concatemer, and different bases are interrogated on different copies of the target.

The preferred way of interrogating the concatemer is to hybridise one or more third polynucleotides of defined sequence to the concatemer such that
25 the third polynucleotide hybridises to the region adjacent to the target, permitting interrogation of the base next to it, to occur. Different target sequences on the concatemer can be interrogated at different base positions by modifying the size of the third polynucleotide, as shown in Figures 2 and 3. Universal bases can be added to the third polynucleotides so that the different
30 lengths can be achieved at the same time as retaining the hybrid. This permits control over which base on the target is to be interrogated.

The different sized third polynucleotides can be added sequentially or together. The concentration of each third polynucleotide can be controlled to ensure that each binds to a target. If added together, the different sized third polynucleotides can be labelled so that a distinction can be made for each reaction.

Alternatively, the second polynucleotide (masking polynucleotide) can be designed so that there is a different sequence masking each target region such that this different sequence can be used to hybridise different third polynucleotides depending on the position to be interrogated. It will therefore be possible design different primers depending on which position is to be targeted. These can be added sequentially to carry out interrogation.

Once hybridised, the base(s) can be interrogated by carrying out a polynucleotide extension reaction, using dideoxy nucleotides (ddNTP) that are detectably labelled. The third polynucleotide, acting as a primer, is thereby extended by one base, which can be detected. Further extension is prevented due to the use of the ddNTP which does not permit further extension.

This method allows different bases on different targets to be identified. The eventual read-out step is made more simple by the separation between the labelled bases on the resulting hybrid.

The ddNTPs can be labelled in any convenient way, but preferably are labelled with a fluorophore; a different type for each of the different ddNTPs. The labelling of nculeotides with fluorophores is now widely known in the art, and conventional reagents and procedures can be used.

The target sequence is ligated to the known sequence of the first polynucleotide prior to concatemerisation, so that the concatemer comprises both target polynucleotide sequences and known sequences, so that hybridisation can occur between the concatemer and the second polynucleotide. In this embodiment, the known sequences of the first polynucleotide should be of sufficient length to permit hybridisation with the second polynucleotide to occur. For example, the known sequences should be more than 100 nucleotides, preferably more than 500 nucleotides. This

provides separation between the hybridised sequences and the non-hybridised (target) sequences, which can then be interrogated.

The conditions necessary for carrying out the method of the invention, including temperature, pH, buffer compositions etc., are conventional and will
5 be apparent to those skilled in the art.

Circularisation

In one embodiment, the target and first polynucleotide are circularised to aid the formation of the concatemer. The target may be circularised in any convenient way. In one embodiment, the single-stranded target is hybridized
10 to the 3' end of the first polynucleotide. Both the 5' and the 3' end of the target molecule will hybridize to the first polynucleotide and will be ligated together forming a single-stranded circle. The efficiency of circle ligations is much better with increased complementarity and it is preferred to use at least 6 complementary nucleotides, preferably at least 9 complementary
15 nucleotides for hybridisation to the first polynucleotide. The ligase can be any available ligase, but is preferably T4 DNA ligase, E.coli DNA ligase or Taq DNA ligase.

In an alternative method, a support-bound oligonucleotide can be used to hybridise to the target and to ligate to the first polynucleotide. In one
20 embodiment of this, the hybrid forms a partially double-stranded molecule with an overhang complementary to the first polynucleotide's 3' end. The support oligonucleotide can then be ligated to the first polynucleotide at the 3' end. The 5' end of the target is also complementary to the first polynucleotide and so the target will hybridise to the first polynucleotide bringing the two ends of
25 the target into position for a ligase to join the two ends of the target, forming a circle. The support oligonucleotide acts to help retain the now circularised target at the first polynucleotide, ready for concatemerisation.

The support-bound oligonucleotide will be of a size sufficient to aid hybridisation and circularisation with the target.

Concatemerisation

The target polynucleotide can be concatemerised in any convenient way. In particular a polymerase reaction is used. In one embodiment, the circularised target (first) polynucleotide acts as a template for a polymerase reaction. As the template is a circular molecule, the technique used is commonly known as Rolling Circle Amplification (RCA). Several variants of this method exist, as reviewed in Richardson et al., Genetic engineering, 25, 51-63, the content of which is incorporated herein by reference. Linear RCA utilises one primer, producing one concatemer from each template. Exponential RCA utilises two primers, where one is complementary to the target to be amplified, while the other is complementary to the product generated by the first primer. Hence, the second primer initiates the synthesis of multiple concatemerised copies from one target polynucleotide. Multiply-primed RCA utilises a set of random hexamers as primers. These primers initiate the synthesis of multiple concatemerised copies from one target polynucleotide. Secondary non-specific priming events can occur subsequently on the displaced product strands of the initial RCA step. Several polymerases can be used, including Sequenase, Bst DNA polymerase (large fragment), Klenow exo-DNA polymerase, which are all polymerases operating at 37oC and displaying the strand displacement ability which is preferable for making the concatemers. Also, the heat-stable Vent exo-DNA polymerase may be used. However, the enzyme shown in the literature to be most efficient on acting on circular templates is phi29 polymerase, and this is preferred.

Concatemerisation may also be carried out in ways not dependant on RCA. For example, multiple copies of the target/first polynucleotide can be ligated together using conventional methods, to form a concatemerised product. Other ways will also be evident to the skilled person.

Hybridisation with mask (second) polynucleotide

The hybridisation to the second polynucleotide can be carried out directly as the concatemer is produced. Accordingly, the second polynucleotide can be present during the formation of the concatemer.

In this embodiment, the circular target may be attached (ligated) to the second polynucleotide, so that the polymerase product is formed in proximity to the second polynucleotide, aiding hybridisation.

In an alternative method, the hybridisation can also be separated from the concatemerisation reaction by “blocking” the second polynucleotide using a complementary molecule. The blocking molecule can either be synthesised in a separate reaction and then annealed to the second polynucleotide prior to the concatemerisation reaction. Alternatively, the blocking molecule can be synthesised by a polymerase directly on the second polynucleotide by using a short primer. After the concatemerisation reaction, the blocking molecule can be removed using an exonuclease, and the second polynucleotide is then available for hybridisation to the concatemerised target molecule and its polymerised product.

The second polynucleotide will have at least partial complementarity to the concatemer. This is achieved by knowledge of the first polynucleotide. The intention is to hybridise the sequence corresponding to the first polynucleotide to the second polynucleotide, such that there are non-hybridised portions corresponding to the target sequence which can be interrogated.

Interrogation of the resulting hybrid may be carried out using any convenient read-out technique.

The present invention also relates to support materials which comprise the polynucleotides defined herein. A support surface will comprise a double-stranded polynucleotide immobilised thereon, wherein one strand will be the concatemer molecule and the other strand will be the second polynucleotide, wherein there are hybridised and non-hybridised regions.

Any suitable support material may be used, including conventional glass, ceramic or plastics materials having a suitable surface.

Immobilisation may be carried out using conventional techniques and covalent and non-covalent attachment may be used. Preferably, it is the second polynucleotide that is covalently attached to the support. Suitable linker molecules will be apparent to the skilled person.

The content of all the publications referred to herein are incorporated herein by reference.

CLAIMS

1. A method for the sequencing of a target polynucleotide, comprising the steps of:

5 (i) ligating the target polynucleotide to a first polynucleotide;

(ii) forming a concatemer comprising multiple copies of the product of step (i); and

(iii) interrogating one or more bases in multiple copies of the target polynucleotide of the concatemer, to thereby identify the target polynucleotide

10 sequence.

2. A method according to claim 1, wherein step(III) is carried out by:

(a) attaching the concatemer to a second polynucleotide such that the second polynucleotide hybridises to portions of the concatemer, but not to

15 regions on the concatemer corresponding to at least a portion of the target polynucleotide; and

(b) interrogating one or more bases in those regions not hybridised to the second polynucleotide, to thereby identify the target polynucleotide sequence.

20

3. A method for the sequencing of a target polynucleotide, comprising the steps of:

(i) ligating the target polynucleotide to a first polynucleotide linker of known sequence, to form a circular polynucleotide;

25 (ii) contacting the circular polynucleotide with a second polynucleotide under conditions which allow a polymerase reaction to proceed, the second polynucleotide comprising two or more sequences that are complementary to the first polynucleotide except for a region adjacent to the target polynucleotide, such that the polymerised product comprises

30 portions hybridised to the second polynucleotide and non-hybridised portions which correspond to the target polynucleotide and the region adjacent to the target polynucleotide,

(iii) interrogating a plurality of the non-hybridised portions to identify one or more different bases corresponding to the target polynucleotide, to thereby identify the target polynucleotide sequence.

5 4. A method of according to any preceding claim, the interrogation step is carried out by:

(a) hybridising third polynucleotides to (at least) the regions adjacent to the target polynucleotide, said third polynucleotides being capable of acting as a primer for polynucleotide extension;

10 (b) carrying out a polymerase reaction with a detectably labelled ddNTP such that the base on the target adjacent to that hybridised to the third polynucleotide incorporates a labelled ddNTP; and

(c) identifying the incorporated ddNTP, to thereby identify the complementary base on the target,

15 (d) wherein the third polynucleotides are of known length, such that different bases on the target are interrogated.

5. A method according to any preceding claim wherein consecutive target polynucleotides on the concatemer are interrogated at consecutive base

20 positions, to reveal the sequence of the target.

6. A method according to claim 1, wherein the product of step (i) is a circular polynucleotide.

25 7. A method according to any preceding claim, wherein the second polynucleotide is immobilised on a support surface.

8. A support surface comprising a double-stranded polynucleotide immobilised thereon, wherein one strand is a concatemer of repeating

30 polynucleotide sequences having regions hybridised to the other strand, and non-hybridised regions.

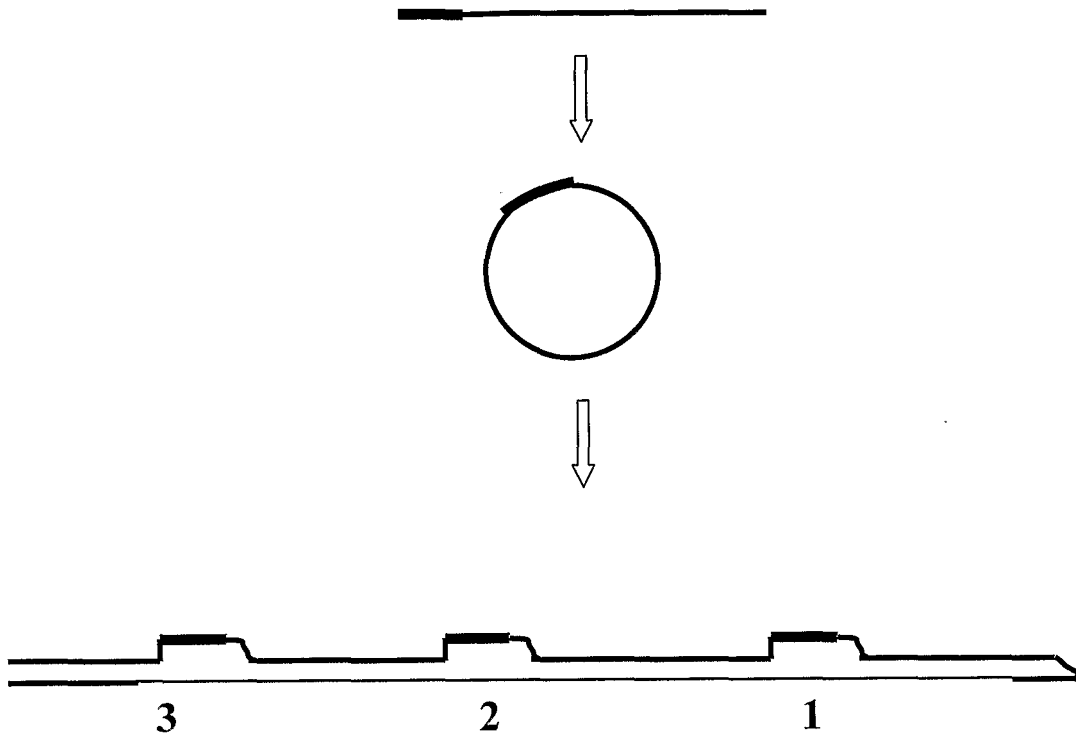

Figure 1

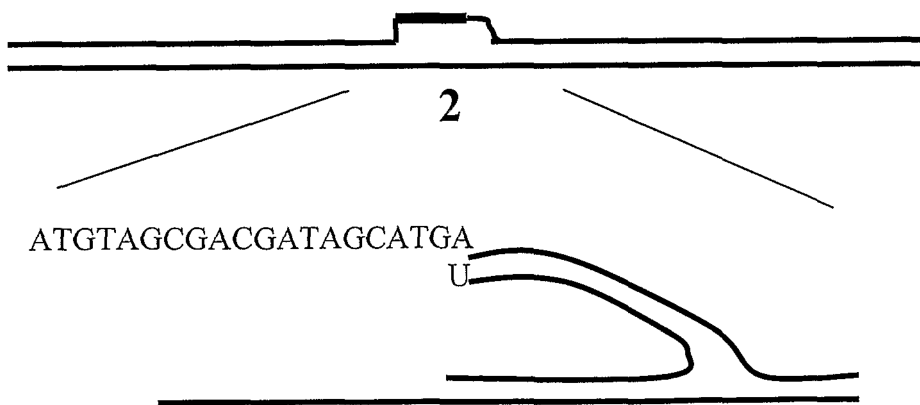

Figure 2

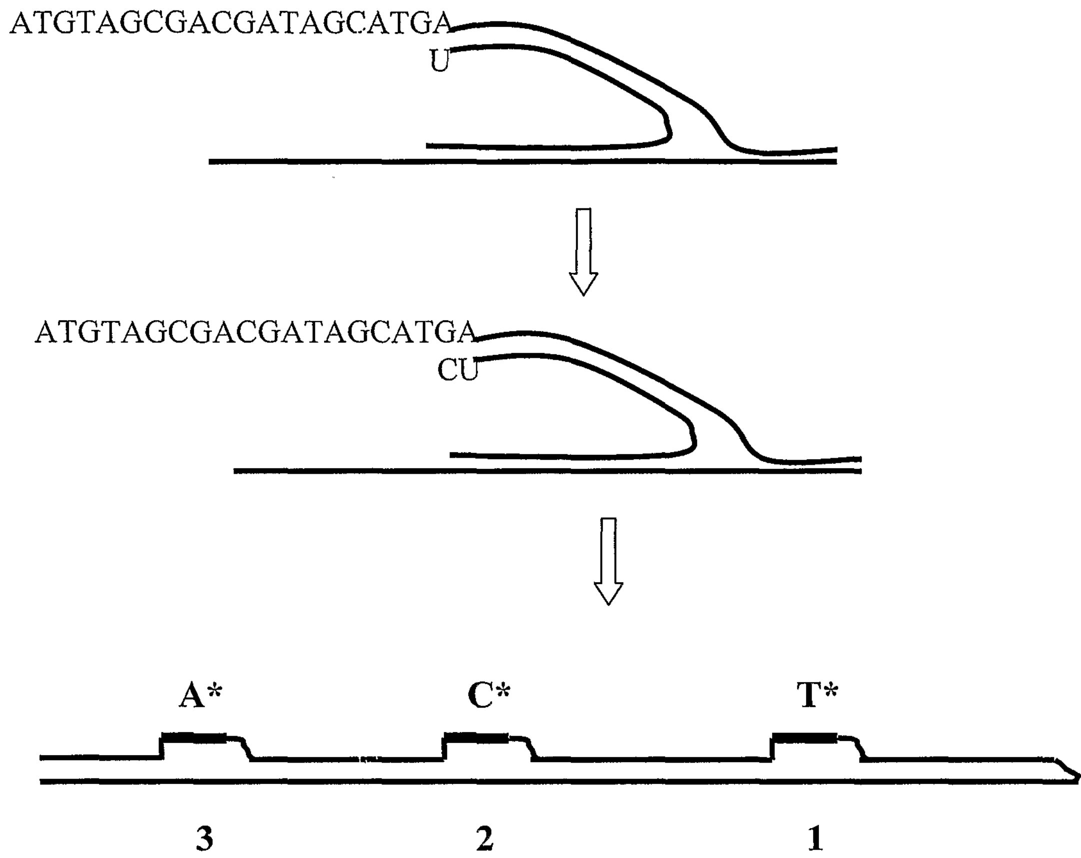

Figure 3