US006087095A

United States Patent [19]

Rosenthal et al.

[11] **Patent Number: 6,087,095**

[45] **Date of Patent: *Jul. 11, 2000**

[54] **DNA SEQUENCING METHOD**

[75] Inventors: **Andre Rosenthal**; **Sydney Brenner**, both of Cambridge, United Kingdom

[73] Assignee: **Medical Research Council**, London, United Kingdom

[*] Notice: This patent is subject to a terminal disclaimer.

[21] Appl. No.: **08/325,224**

[22] PCT Filed: **Apr. 22, 1993**

[86] PCT No.: **PCT/GB93/00848**

§ 371 Date: **Dec. 9, 1994**

§ 102(e) Date: **Dec. 9, 1994**

[87] PCT Pub. No.: **WO93/21340**

PCT Pub. Date: **Oct. 28, 1993**

[30] **Foreign Application Priority Data**

Apr. 22, 1992 [GB] United Kingdom 9208733

[51] **Int. Cl.**[7] **C12Q 1/68**; C12P 19/34; C07N 21/00

[52] **U.S. Cl.** **435/6**; 435/41; 435/172.1; 536/24.33; 536/25.3

[58] **Field of Search** 435/6, 41, 172.1; 935/76, 77, 78; 536/23.1, 24.33, 25.3, 25.32

[56] **References Cited**

FOREIGN PATENT DOCUMENTS

WO 90/13666 11/1990 WIPO .
WO 91/06678 5/1991 WIPO .

OTHER PUBLICATIONS

Drmanac et al., "Sequencing of Megabase Plus DNA by Hybridization: . . . ", Genomics 4:114–128, 1989.

Salmeron et al., "Imaging of Biomolecules with the Scanning Tunneling Microscope: . . . ", J. Vac. Sci. Tech. 8:635, Jan./Feb. 1990.

Maxam et al., "A new method for sequencing DNA", Proc. Natl. Sci. USA 74:560–564, Feb. 1977.

Drmanac et al., "Reliable Hybridization of Oligonucleotides as Short as Six Nucleotides", DNA and Cell Biology 9:527, Nov. 1990.

Sanger et al., "DNA sequencing with chain–terminating inhibitors", Proc. Natl. Acad. Sci. USA 74:5463–5467, Dec. 1977.

Bains et al., "A Novel Method for Nucleic Acid Sequence Determination", J. Theor. Biol. 135:303–307, 1988.

Driscoll et al., "Atomic–scale imaging of DNA using scanning tunnelling microscopy", Nature 346:294, Jul. 1990.

E.D. Hyman, "A New Method of Sequencing DNA", Analytical Biochemistry 174:423–436, 1988.

Khrapko et al., "An oligonucleotide hybridization approach to DNA sequencing", FEBS 256:118–122, Oct. 1989.

P.A. Pevzner, "1–Tuple DNA Sequencing:Computer Analysis", J. Biom. Str. & Dyn. 7:63, 1989.

Jett et al., "Hig–Speed DNA Sequencing: An Approach Based Upon Fluorescence Detection . . . ", J. Biom. Str. & Dyn. 7:301, 1989.

Lindsay et al., Genet. Anal. Tech. Appl., 8:8, 1991.

Nguyen et al., Anal. Chem., 56:348, 1987.

Maskos et al., Cold Spring Harbor Symposium on Genome Mapping and Sequencing, Abstracts, p. 143, 1991.

Allison et al., Scanning Microsc. 4:517, 1990.

Primary Examiner—Nancy Degen
Assistant Examiner—Sean McGarry
Attorney, Agent, or Firm—Fulbright & Jaworski, LLP

[57] **ABSTRACT**

The invention is drawn to a method of DNA sequencing using labeled nucleotides that do not act as chain elongation inhibitors where the label is removed or neutralized for the sequential addition of non-labeled nucleotides.

16 Claims, 1 Drawing Sheet

Columbia Ex. 2005
Illumina, Inc. v. The Trustees
of Columbia University
in the City of New York
IPR2020-00988

DOCKET ALARM

Find authenticated court documents without watermarks at docketalarm.com.

FIG. 1

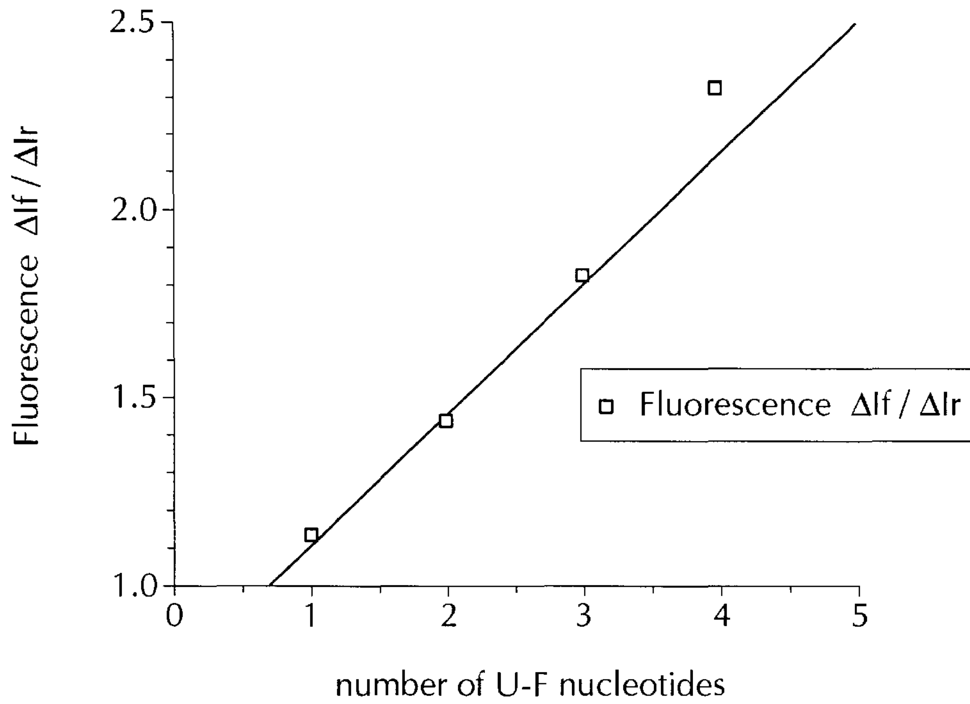
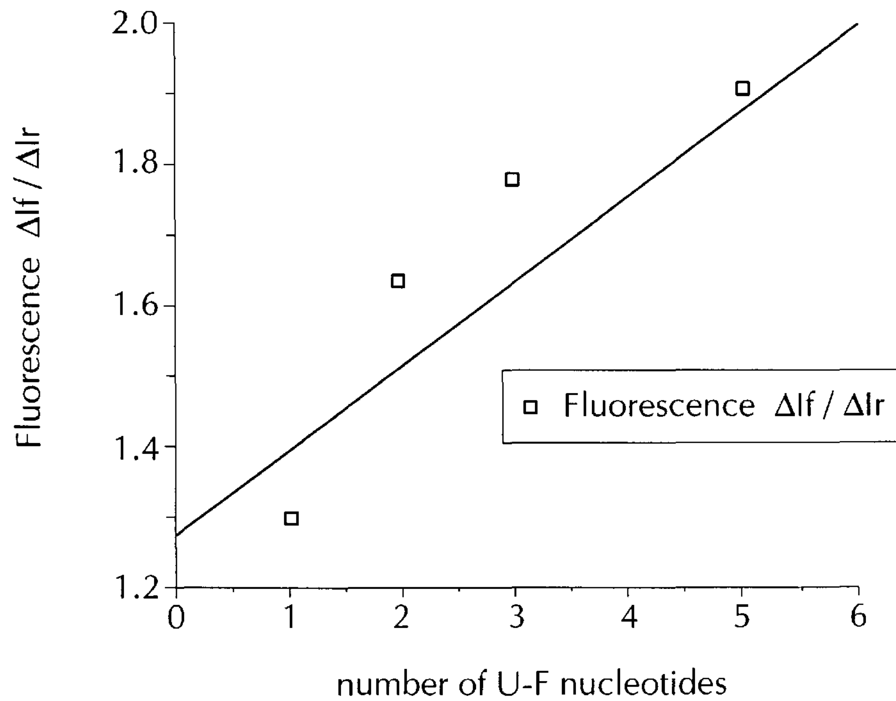


FIG. 2



DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

DNA SEQUENCING METHOD

This application is a 371 of PCT/GB93/00848 filed Apr. 22, 1993.

The present invention relates to a method for sequencing DNA. In particular, the present invention concerns a method for the automated sequencing of large fragments of DNA.

DNA sequence analysis has become one of the most important tools available to the molecular biologist. Current sequencing technology allows sequence data to be obtained from virtually any DNA fragment. This has allowed not only the sequencing of entire genes and other genomic sequences but also the identification of the sequence of RNA transcripts, by the sequencing of cDNA. Currently, emphasis is being placed on genomic sequencing in order to determine the DNA sequence of entire genomes. Ultimately, it is hoped that the sequence of the human genome will be deciphered.

Traditional DNA sequencing techniques share three essential steps in their approaches to sequence determination. Firstly, a multiplicity of DNA fragments are generated from a DNA species which it is intended to sequence. These fragments are incomplete copies of the DNA species to be sequenced. The aim is to produce a ladder of DNA fragments, each a single base longer than the previous one. This can be achieved by selective chemical degradation of multiple copies of the DNA species to be sequenced, as in the Maxam and Gilbert method (A. Maxam and W. Gilbert, PNAS 74, p. 560, 1977). Alternatively, the DNA species can be used as a template for a DNA polymerase to produce a number of incomplete clones, as in the Sanger method (F. Sanger, S. Nicklen and A. Coulson, PNAS 74, p. 5463, 1977). These fragments, which differ in respective length by a single base, are then separated on an apparatus which is capable of resolving single-base differences in size. A thin polyacrylamide gel is invariably used in this process. The third and final step is the determination of the nature of the base at the end of each fragment. When ordered by the size of the fragments which they terminate, these bases represent the sequence of the original DNA species.

Determination of the nature of each base is achievers by previously selecting the terminal base of each fragment. In the Sanger method, for example, dideoxy nucleoside triphosphates (ddNTPs) are used to selectively terminate growing DNA clones at an A, C, G or T residue. This means that four separate reactions need to be performed for each sequencing exercise, each in a separate tube using a different ddNTP. In one tube, therefore, each labelled fragment will terminate with an A residue, while in the next tube with a C residue, and so on. Separation of each croup of fragments side-by-side on a polyacrylamide gel will show the sequence of the template by way of the relative size of the individual fragments.

In the Maxam and Gilbert method, on the other hand, the selectivity is achieved during the chemical degradation process. Chemicals are used which cleave DNA strands at A only, C only, G and A or T and C. Use of limiting concentrations of such chemicals allows partial digestion of the DNA species. As in the Sanger method, four separate reactions must be performed and the products separated side-by-side on a polyacrylamide gel.

The disadvantages of these prior art methods are numerous. They require a number of complex manipulations to be performed, in at least four tubes. They are susceptible to errors due to the formation of secondary structures in DNA, or other phenomena that prevent faithful replication of a DNA template in the Sanger method or which cause base-specificity to be lost by the chemical reactants of the Maxam and Gilbert method. The most serious problems, however, are caused by the requirement for the DNA fragments to be size-separated on a polyacrylamide gel. This process is time-consuming, uses large quantities of expensive chemicals, and severely limits the number of bases which can be sequenced in any single experiment, due to the limited resolution of the gel. Furthermore, reading the gels in order to extract the data is labour-intensive and slow.

A number of improvements have been effected to these sequencing methods in order to improve the efficiency and speed of DNA sequencing. Some of these improvements have related to the sequencing reaction itself. For example, improved polymerase enzymes have been introduced which lead to greater precision in the Sanger method, such as Sequenase® and Taquenase®. Improved reagents have not, however, significantly affected the speed of sequence data generation or significantly simplified the sequencing process.

In the interest of both speed and simplicity, a number of "Automated Sequencers" have been introduced in recent years (reviewed in T. Hunkapiller, R. Kaiser, B. Koop and L. Hood, Science, 254, p. 59, 1991). These machines are not, however, truly automatic sequencers. They are merely automatic gel readers, which require the standard sequencing reactions to be carried out before samples are loaded onto the gel. They do provide a slight increase in speed, however, due to faster reading of the gels and collation of the data generated into computers for subsequent analysis.

Many automated sequencers exploit recent developments which have been made in labelling technology. Traditionally, radioactive labels in the form of ^{32}P or ^{35}S have been used to label each DNA fragment. Recently, however, fluorophores have gained acceptance as labels. These dyes, attached either to the sequencing primer or to nucleotides, are excited to a fluorescent state on the polyacrylamide gel by a laser beam. An automated sequencer, therefore, can detect labelled fragments as they pass under a laser in a reading area. Use of dyes which fluoresce at different wavelengths allows individual labelling of A, G. C and T residues, which permits the products of all four sequencing reactions to be run in a single lane of the gel.

Even incorporating such refinements, however, automated sequencers can still produce no more than about 100 kb of finished sequence per person per year. At this rate, it would take one person 73,000 years to sequence the human genome.

Clearly, if the aim of sequencing the human genome is to be achieved, current sequencing technology is entirely inadequate. In view of this, a few proposals have been made for alternative sequencing strategies which are not merely improvements of the old technology.

One such method, sequencing by hybridisation (SBH), relies on the mathematical demonstration that the sequence of a relatively short (say, 100 kbp) fragment of DNA may be obtained by synthesising all possible N-mer oligonucleotides and determining which oligonucleotides hybridise to the fragment without a single mismatch (R. Drmanac, I. Labat, I. Bruckner and R. Crkvenjakov, Genomics, 4, p. 114, 1989; R. Drmanac, Z. Stvanovic, R. Crkvenjakov, DNA Cell Biology, 9, p. 527, 1990; W. Bains and G. Smith, J. Theor. Biol., 135, pp 303–307, 1988; K. R. Khrapko, et al, FEBS lett., 256, pp. 118–122, 1989; P. A. Pevzner, J. Biomolecular Structure and Dynamics, 7, pp. 63–73, 1989; U. Maskos and E. M. Southern, Cold Spring Harbour Symposium on Genome Mapping and Sequencing, Abstracts, p. 143, 1991). N can be 8, 9 or 10, such sizes being a compromise between the requirement for reasonable hybridisation parameters and manageable library sizes.

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

The technique can be automated by attaching the oligonucleotides in a known pattern on a two-dimensional grid. The fragment to be sequenced is subsequently hybridised to the oligonucleotides on the grid and the oligonucleotides to which the sequence has been hybridised are detected using a computerised detector. Determination of the sequence of the DNA is then a matter of computation. however, errors arise from the difficulty in determining the difference between perfect matches and single base-pair mismatches. Repetitive sequences, which occur quite commonly in the human genome, can also be a problem.

Another proposal involves the fluorescent detection of single molecules (J. Jett et al., J. Biomol. Struct. Dyn., 7, p. 301, 1989: D. Nguyen, et al., Anal. Chem., 56, p. 348, 1987). In this method, a single, large DNA molecule is suspended in a flow stream using light pressure from a pair of laser beams. Individual bases, each of which is labelled with a distinguishing fluorophore, are then cut from the end of the molecule and carried through a fluorescence detector by the flow stream.

Potentially, this method could allow the accurate sequencing of a large number of base pairs—several hundred—per second. However, feasibility of this method is not yet proven.

A third method is sequencing by scanning tunnelling microscopy (STM) (S. Lindsay, et al., Genet. Anal. Tech. Appl., 8, p. 8, 1991: D. Allison et al., Scanning Microsc., 4, p. 517, 1990: R. Driscoll et al., Nature, 346, p. 294, 1990: M. Salmeron et al., J. Vac. Sci. Technol., 8, p. 635, 1990). This technique requires direct three-dimensional imaging of a DNA molecule using STM. Although images of the individual bases can be obtained, interpretation of these images remains very difficult. The procedure is as yet unreliable and the success rate is low.

A fourth method involves the detection of the pyrophosphate group released as a result of the polymerisation reaction which occurs when a nucleotide is added to a DNA primer in a primer extension reaction (E. D. Hyman, Anal. Biochem., 174, p. 423, 1988). This method attempts to detect the addition of single nucleotides to a primer using the luciferase enzyme to produce a signal on the release of pyrophosphate. However, this method suffers a number of drawbacks, not least of which is that dATP is a substrate for luciferase and thus will always give a signal, whether it is incorporated into the chain or not. The added nucleotides are not labelled and no method is disclosed which will allow the use of labelled nucleotides.

In summary, therefore, each of the new approaches to DNA sequencing described above, while solving some of the problems associated with traditional methods, introduces several problems of its own. In general, most of these methods are expensive and not currently feasible.

There is therefore a need for a sequencing method which allows the rapid, unambiguous sequencing of DNA at low cost. The requirements for such a system are that:

1. it should not be based on gel resolution of differently-sized oligomers;
2. it should allow more rapid sequencing than present methods;
3. it should allow several DNA clones to be processed in parallel;
4. the cost of hardware should be reasonable;
5. it should cost less per base of sequence than current technology; and
6. it should be technically feasible at the present time

SUMMARY OF THE INVENTION

The present invention provides such a sequencing system which comprises a method for the sequential addition nucleotides to a primer on a DNA template.

According to a first aspect of the present invention, there is provided a method for determining the sequence of a nucleic acid comprising the steps of:

a) forming a single-stranded template comprising the nucleic acid to be sequenced;

b) hybridising a primer to the template to form a template/primer complex;

c) extending the primer by the addition of a single labelled nucleotide;

d) determining the type of the labelled nucleotide added onto the primer;

e) removing or neutralising the label; and

f) repeating steps (c) to (e) sequentially and recording the order of incorporation of labelled nucleotides.

In the method of the invention, a single-stranded template is generated from a nucleic acid fragment which it is desired to sequence. Preferably, the nucleic acid is DNA. Part of the sequence of this fragment may be known, so that a specific primer may be constructed and hybridised to the template. Alternatively, a linker may be ligated to a fragment of unknown sequence in order to allow for hybridisation of a primer.

The template may be linear or circular. Preferably, the template is bound to a solid-phase support. For example, the template may be bound to a pin, a glass plate or a sequencing chip. The provision of a solid phase template allows for the quick and efficient addition and removal of reagents, particularly if the process of the invention is automated. Additionally, many samples may be processed in parallel in the same vessel yet kept separate.

Preferably, the template is attached to the solid support by means of a binding linker. For example, one of the commercially available universal primers can be ligated to the 5' end of the template or incorporated easily to one of the ends of the templates by the polymerase chain reaction.

The binding linker may be attached to the solid support by means of a biotin/streptavidin coupling system. For example, the surface of the solid support may be derivatised by applying biotin followed by streptavidin. A biotinylated binding linker is then ligated to the template to bind it to the solid support or the biotinylated template generated by PCR is bound to the solid support.

In an alternative embodiment, an unligated binding linker is bound to the solid support by the biotin/streptavidin system. The template is then hybridised to the binding linker. The binding linker may be a separate binding linker, which is not the sequencing primer. Alternatively, the binding linker may also function as the sequencing primer.

Clearly, it is essential in the latter embodiment that the template should possess a region of complementarity with the binding linker bound to the support. Where the template is ligated to a linker, the complementarity may be provided by that linker. Alternatively, the binding linker may be complementary to a unique sequence within the template itself.

Preferably the solid support is derivatised using a mask so as to allow high resolution packaging of the template(s) on the support. An array of template attachment areas can thereby be produced on a glass plate or sequencing chip, allowing parallel processing of a large number of different templates. Where pins are used as the solid support, a single pin is needed for each template. The single pins may be grouped into arrays. It is envisaged that an array of 100×100 pins or attachment areas can be used, to allow the simultaneous processing of 10^4 clones.

The primer is extended by a DNA polymerase in the presence of a single labelled nucleotide, either A, C, G or T.

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

Suitable DNA polymerases are, for example, Sequenase 2.0®, T4 DNA polymerase or the Klenow fragment of DNA polymerase 1 as well as heat-stable polymerases such as Taq polymerase (for example Taquenase®) and Vent polymerase.

In a manually operated procedure using a single template, the labelled nucleotides are used singly and sequentially in order to attempt to add that nucleotide to the primer. The nucleotide will add on to the primer when it is complementary to the next nucleotide in the template. It may take one, two, three or four steps before the appropriate labelled nucleotide is used. However, as soon as it is determined that a labelled nucleotide has been added onto the primer, step (e) can be performed.

In an automated procedure, especially where a large number of templates are being sequenced simultaneously, in step (c) all four labelled nucleotides are used sequentially and it is merely noted which of the labelled nucleotides is added, that is it is determined whether it is the first, second, third or fourth labelled nucleotide which is added.

It has been found that nonspecific end-addition and misincorporation of nucleotides can lead to background problems when the incorporation step has been repeated a number of times. These side reactions are mainly due to the fact that a single nucleotide is present, instead of all four nucleoside triphosphates. In fact, it has been observed that while it is possible to sequence certain templates by the sequential addition of single nucleotides in the absence of the other three, significant problems have been encountered with other templates, particularly those templates containing multiple base repeats, due to non-specific incorporation of a nucleotide which is caused by the polymerase effectively jumping over a non-complementary base.

In order to ensure high accuracy of operation during the primer extension step, it has been found advantageous to carry out step (c) in the presence of chain elongation inhibitors.

Chain elongation inhibitors are nucleotide analogues which either are chain terminators which prevent further addition by the polymerase of nucleotides to the 3' end of the chain by becoming incorporated into the chain themselves, or compete for incorporation without actually becoming incorporated. Preferably, the chain elongation inhibitors are dideoxy nucleotides. Where the chain elongation inhibitors are incorporated into the growing polynucleotide chain, it is essential that they be removed after incorporation of the labelled nucleotide has been detected, in order to allow the sequencing reaction to proceed using different labelled nucleotides. It has been found, as described below, that 3' to 5' exonucleases such as, for example, exonuclease III, are able to remove dideoxynucleotides. This finding allows the use of dideoxynucleotides as chain elongation inhibitors to promote the accuracy of the polymerase in the sequencing method of the invention. Accuracy of the polymerase is essential if 10^4 clones are to be processed simultaneously, since it is high polymerase accuracy which enables the sequencing reaction to be carried out on a single template instead of as four separate reactions.

Alternatively, the chain elongation inhibitors may be deoxynucleoside 5'-[α,β-methylene]triphosphates. These compounds are not incorporated into the chain. Other nucleotide derivatives such as, for example, deoxynucleoside diphosphates or deoxynucleoside monophosphates may be used which are also not incorporated into the chain.

It is furthermore envisaged that blocking groups on the 3' moiety of the deoxyribose group of the labelled nucleotide may be used to prevent nonspecific incorporation. Preferably, therefore, the labelled nucleotide is labelled by attachment of a fluorescent dye group to the 3' moiety of the deoxyribose group, and the label is removed by cleaving the fluorescent dye from the nucleotide to generate a 3' hydroxyl group. The fluorescent dye is preferably linked to the deoxyribose by a linker arm which is easily cleaved by chemical or enzymatic means.

Evidently, when nucleotide analogue chain elongation inhibitors are used, only the analogues which do not correspond to the labelled nucleotide should be added. Such analogues are referred to herein as heterogenous chain elongation inhibitors.

Label is ideally only incorporated into the template/primer complex if the labelled nucleotide added to the reaction is complementary to the nucleotide on the template adjacent the 3' end of the primer. The template is subsequently washed to remove any unincorporated label and the presence of any incorporated label determined. A radioactive label may be determined by counting or any other method known in the art, while fluorescent labels can be induced to fluoresce, for example by laser excitation.

It will be apparent that any label known in the art to be suitable for labelling nucleic acids may be used in the present invention. However, the use of fluorescent labels is currently preferred, due to the sensitivity of detection systems presently available for such labels which do not involve the use of radioactive substances.

Examples of flourescently-labelled nucleotides currently available include fluorescein-12-dUTP, fluorescein-15-dCTP, fluorescein-15-dATP and flurescein-15-dITP. It has proved very difficult to synthesise a suitable fluroescent guanosine compound, so an inosine compound is used in its place. Should a fluorescent guanosine compound become available, its use is envisaged in the present invention.

It has been found advantageous to use a mixture, of unlabelled and labelled nucleotides in the addition step.

When a fluorescent label is used, in order to produce all possible extension products on a template possessing a run of a particular nucleotide, the following ratios were found to be approximately optimal:

Fluorescein—15-dATP/dATP 500:1
Fluorescein—15-dITP/dGTP 500:1
Fluorescein—12-dUTP/dTTP 15:1
Fluorescein—12-dCTP/dCTP 15:1.

Preferably, therefore, the above ratios are used in connection with fluorescently-labelled nucleotides.

By repeating the incorporation and label detection steps until incorporation is detected, the nucleotide on the template adjacent the 3' end of the primer may be identified. Once this has been achieved, the label must be removed before repeating the process to discover the identity of the next nucleotide. Removal of the label may be effected by removal of the labelled nucleotide using a 3'–5' exonuclease and subsequent replacement with an unlabelled nucleotide. Alternatively, the labelling group can be removed from the nucleotide. In a further alternative, where the label is a fluorescent label, it is possible to neutralise the label by bleaching it with laser radiation.

If chain terminators or 3' blocking groups have been used, these should be removed before the next cycle can take place. Preferably, chain terminators are removed with a 3'–5' exonuclease. Preferably, exonuclease III is used. 3' blocking groups may be removed by chemical or enzymatic cleavage of the blocking group from the nucleotide.

Where exonuclease III is used to remove the chain terminators, it is essential to prevent the exonuclease III from chewing back along the growing chain to remove

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE

fastcase®
Smarter legal research.