(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 1 530 578 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
13.03.2013 Bulletin 2013/11

(21) Application number: **03792519.5**

(22) Date of filing: **22.08.2003**

(51) Int Cl.:
C07H 19/06 *(2006.01)* ***C07H 19/10*** *(2006.01)*
C07H 19/16 *(2006.01)* ***C07H 19/20*** *(2006.01)*
C07H 21/00 *(2006.01)* ***C12Q 1/68*** *(2006.01)*
G01N 33/53 *(2006.01)*

(86) International application number:
PCT/GB2003/003686

(87) International publication number:
WO 2004/018497 (04.03.2004 Gazette 2004/10)

(54) MODIFIED NUCLEOTIDES FOR POLYNUCLEOTIDE SEQUENCING

MODIFIZIERTE NUKLEOTIDE FÜR POLYNUKLEOTIDSEQUENCING

NUCLEOTIDES MODIFIES

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LI LU MC NL PT RO SE SI SK TR

(30) Priority: **23.08.2002 US 227131**
23.12.2002 GB 0230037
20.02.2003 GB 0303924

(43) Date of publication of application:
18.05.2005 Bulletin 2005/20

(60) Divisional application:
12180077.5

(73) Proprietor: **Illumina Cambridge Limited**
Little Chesterford
Saffron Walden
Essex CB10 1XL (GB)

(72) Inventors:
- **MILTON, John,**
  c/oSolexa Limited
  Nr. Saffron Walden,
  Essex CB10 1XL (GB)
- **WU, Xiaolin,**
  c/oSolexa Limited
  Nr. Saffron Walden,
  Essex CB10 1XL (GB)
- **SMITH, Mark,**
  c/oSolexa Limited
  Nr. Saffron Walden,
  Essex CB10 1XL (GB)
- **BRENNAN, Joseph,**
  c/oSolexa Limited
  Nr. Saffron Walden,
  Essex CB10 1XL (GB)
- **BARNES, Colin,**
  c/oSolexa Limited
  Nr. Saffron Walden, Essex CB10 1XL (GB)
- **LIU, Xiaohai,**
  c/oSolexa Limited
  Nr. Saffron Walden, Essex CB10 1XL (GB)
- **RUEDIGER, Silke,**
  c/oSolexa Limited
  Nr. Saffron Walden, Essex CB10 1XL (GB)

(74) Representative: **Barnes, Colin Lloyd et al**
Stratagem IPM Limited
Forsters Wing, Anstey Hall
Maris Lane
Trumpington
Cambridge CB2 9LG (GB)

(56) References cited:
WO-A-01/57249
WO-A-02/29003
WO-A-02/29003
WO-A-99/05315
US-A- 5 302 509
US-A- 5 547 839
US-A- 6 087 095
US-A1- 2003 104 437
US-B1- 6 255 475

- **S.G. ZAVGORODNY ET AL.: "S,X-Acetals in nucleoside chemistry. III. Synthesis of 2'- and 3'-azidomethyl derivatives of ribonucleosides" NUCLEOSIDES, NUCLEOTIDES AND NUCLEIC ACIDS, vol. 19, 2000, pages 1977-1991, XP002487670**

Printed by Jouve, 75001 PARIS (FR)

(Cont. next page)

- G.H. VEENEMAN ET AL.: "An efficient approach to the synthesis of thymidine derivatives containing phosphate-isosteric methylene acetal linkages" TETRAHEDRON, vol. 47, 1991, pages 1547-1562, XP002487671
- GUIBE F: "Allylic Protecting Groups and Their Use in a Complex Environment Part II: Allylic Protecting Groups and their Removal through Catalytic Palladium pi-Allyl Methodology" TETRAHEDRON, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 54, no. 13, 1 March 1998 (1998-03-01), pages 2967-3042, XP004109085 ISSN: 0040-4020
- BERGMANN F ET AL: "Allyl as internucleotide protecting group in DNA synthesis to be cleaved off by ammonia" TETRAHEDRON, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 51, no. 25, 19 June 1995 (1995-06-19), pages 6971-6976, XP004819821 ISSN: 0040-4020
- KLOOSTERMAN M ET AL: "THE RELATIVE STABILITY OF ALLYL ETHER, ALLYLOXYCARBONYL ESTER AND PROP-2 ENYLIDENE ACETAL, PROTECTIVE GROUPS TOWARD IRIDIUM, RHODIUM AND PALLADIUM CATALYSTS" TETRAHEDRON LETTERS, ELSEVIER, AMSTERDAM, vol. 26, no. 41, 1 January 1985 (1985-01-01), pages 5045-5048, XP009014324 ISSN: 0040-4039
- KITAMURA M ET AL: "(P(C6H5)3)CpRu+-Catalyzed Deprotection of Allyl Carboxylic Esters" JOURNAL OF ORGANIC CHEMISTRY, AMERICAN CHEMICAL SOCIETY. EASTON, vol. 67, no. 14, 29 May 2002 (2002-05-29), pages 4975-4977, XP002375190 ISSN: 0022-3263
- HAYAKAWA ET AL: "O-Allyl protection of guanine and thymine residues in oligodeoxyribonucleotides" JOURNAL OF ORGANOMETALLIC CHEMISTRY, ELSEVIER-SEQUOIA S.A. LAUSANNE, CH, vol. 58, 1 January 1993 (1993-01-01), pages 5551-5555, XP002094188 ISSN: 0022-328X
- GUIBE F: "Allylic Protecting Groups and Their Use in a Complex Environment Part I: Allylic Protection of Alcohols" TETRAHEDRON, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 53, no. 40, 6 October 1997 (1997-10-06), pages 13509-13556, XP004106236 ISSN: 0040-4020
- T.W. GREEN, P.G.M. WUTS: "Protective groups in organic synthesis" 1999, JOHN WILEY AND SONS, INC , NEW YORK , XP002487672 pages 67-74, 574-576
- P.J. KOCIENSKI: "Protecting groups" 1994, GEORG THIEME VERLAG , STUTTGART , XP002487673 * page 61 - page 68 *
- J. BRUNCKOVA ET AL.: "Intramolecular hydrogen atom abstraction in carbohydrates and nucleosides" TETRAHEDRON LETTERS, vol. 35, 1994, pages 6619-6622, XP002270558
- S. NISHINO ET AL.: "Efficient deanilidation of phosphoranilidates by the use of nitrites and acetic anhydride" HETEROATOM CHEMISTRY, vol. 2, 1991, pages 187-196, XP008027710
- M. KRECMEROVA ET AL.: "Synthesis of 5'-O-phosphonomethyl derivatives of pyrimidine 2'-deoxynucleosides" COLL. CZECH. CHEM. COMMUN., vol. 55, 1990, pages 2521-2536, XP002270559
- P.J.L.M. QUAEDFLIEG ET AL.: "An alternative approach towards the synthesis of (3'-5') methylene acetal linked dinucleosides" TETRAHEDRON LETTERS, vol. 33, 1992, pages 3081-3084, XP002270560
- J.-I. YAMASHITA ET AL.: "Studies on antitumor reagents. VII. Antitumor activities of O-alkoxyalkyl derivatives of 2'-deoxy-5-trifluoromethyluridine" CHEM. PHARM. BULL., vol. 35, 1987, pages 2373-2381, XP002270561

Description

[0001] The invention relates to modified nucleotides. In particular, this invention discloses nucleotides having a removable protecting group, their use in polynucleotide sequencing methods and a method for chemical deprotection of the protecting group.

[0002] 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.

[0003] An example of the technologies that have improved the study of nucleic acids is the development of fabricated arrays of immobilised nucleic acids. These arrays consist typically of a high-density matrix of polynucleotides immobilised onto a solid support material. See, e.g., Fodor et al., Trends Biotech. 12:19-26, 1994, which describes ways of assembling the nucleic acids using a chemically sensitized glass surface protected by a mask, but exposed at defined areas to allow attachment of suitably modified nucleotide phosphoramidites. Fabricated arrays can also be manufactured by the technique of "spotting" known polynucleotides onto a solid support at predetermined positions (e.g., Stimpson et al., Proc. Natl. Acad. Sci. USA 92:6379-6383, 1995).

[0004] Sequencing by synthesis of DNA ideally requires the controlled (i.e. one at a time) incorporation of the correct complementary nucleotide opposite the oligonucleotide being sequenced. This allows for accurate sequencing by adding nucleotides in multiple cycles as each nucleotide residue is sequenced one at a time, thus preventing an uncontrolled series of incorporations occurring. The incorporated nucleotide is read using an appropriate label attached thereto before removal of the label moiety and the subsequent next round of sequencing. In order to ensure only a single incorporation occurs, a structural modification ("blocking group") of the sequencing nucleotides is required to ensure a single nucleotide incorporation but which then prevents any further nucleotide incorporation into the polynucleotide chain. The blocking group must then be removable, under reaction conditions which do not interfere with the integrity of the DNA being sequenced. The sequencing cycle can then continue with the incorporation of the next blocked, labelled nucleotide. In order to be of practical use, the entire process should consist of high yielding, highly specific chemical and enzymatic steps to facilitate multiple cycles of sequencing.

[0005] To be useful in DNA sequencing, nucleotide, and more usually nucleotide triphosphates, generally require a 3'OH-blocking group so as to prevent the polymerase used to incorporate it into a polynucleotide chain from continuing to replicate once the base on the nucleotide is added. There are many limitations on the suitability of a molecule as a blocking group. It must be such that it prevents additional nucleotide molecules from being added to the polynucleotide chain whilst simultaneously being easily removable from the sugar moiety without causing damage to the polynucleotide chain. Furthermore, the modified nucleotide must be tolerated by the polymerase or other appropriate enzyme used to incorporate it into the polynucleotide chain. The ideal blocking group-will therefore exhibit long term stability, be efficiently incorporated by the polymerase enzyme, cause total blocking of secondary or further incorporation and have the ability to be removed under mild conditions that do not cause damage to the polynucleotide structure, preferably under aqueous conditions. These stringent requirements are formidable obstacles to the design and synthesis of the requisite modified nucleotides.

[0006] Reversible blocking groups for this purpose have been described previously but none of them generally meet the above criteria for polynucleotide, e.g. DNA-compatible, chemistry.

[0007] Metzker et al., (Nucleic Acids Research, 22(20): 4259-4267, 1994) discloses the synthesis and use of eight 3'-modified 2-deoxyribonucleoside 5'-triphosphates (3'-modified dNTPs) and testing in two DNA template assays for incorporation activity. The 3'-modified dNTPs included 3'allyl deoxyriboadenosine 5'-triphosphate (3'-allyl dATP). However, the 3'allyl blocked compound was not used to demonstrate a complete cycle of termination, deprotection and reinitiation of DNA synthesis: the only test results presented were those which showed the ability of this compound to terminate DNA synthesis in a single termination assay, out of eight such assays conducted, each conducted with a different DNA polymerase.

[0008] WO02/29003 (The Trustees of Columbia University in the City of New York) describes a sequencing method which, may include the use of an allyl protecting group to cap the 3'-OH group on a growing strand of DNA in a polymerase reaction. The allyl group is introduced according to the procedure of Metzker (infra) and is said to be removed by using methodology reported by Kamal et al (Tet. Let, 40, 371-372, 1999).

[0009] The Kamal deprotection methodology employs sodium iodide and chlorotrimethylsilane so as to generate *in situ* iodotrimethylsilane, in acetonitrile solvent, quenching with sodium thiosulfate. After extraction into ethyl acetate and drying (sodium sulfate), then concentration under reduced pressure and column chromatography (ethyl acetate:hexane; 2:3 as eluant), free alcohols were obtained in 90-98% yield.

[0010] In WO02/29003, the Kamal allyl deprotection is suggested as being directly applicable in DNA sequencing without modification, the Kamal conditions being mild and specific.

[0011] While Metzker reports on the preparation of a 3'allyl-blocked nucleotide or nucleoside and WO02/29003 suggests the use of the allyl functionality as a 3'-OH cap during sequencing, neither of these documents actually teaches

the deprotection of 3'-allylated hydroxyl group in the context of a sequencing protocol. Whilst the use of an allyl group as a hydroxyl protecting group is well known - it is easy to introduce and is stable across the whole pH range and to elevated temperatures - there is to date, no concrete embodiment of the successful cleavage of a 3'-allyl group under DNA compatible conditions, i.e. conditions under which the integrity of the DNA is not wholly or partially destroyed. In other words, it has not been possible hitherto to conduct DNA sequencing using 3'OH allyl-blocked nucleotides.

[0012] The Kamal methodology is inappropriate to conduct in aqueous media since the TMS chloride will hydrolyse preventing the *in situ* generation of TMS iodide. Attempts to carry out the Kamal deprotection (in acetonitrile) in sequencing have proven unsuccessful in our hands.

[0013] The present invention is based on the surprising development of a number of reversible blocking groups and methods of deprotecting them under DNA compatible conditions. Some of these blocking groups are novel per se; others have been disclosed in the prior art but, as noted above, it has not proved possible to utilised these blocking groups in DNA sequencing.

[0014] One feature of the invention derives from the development of a completely new method of allyl deprotection. Our procedure is of broad applicability to the deprotection of virtually all allyl-protected hydroxyl functionality and may be effected in aqueous solution, in contrast to the methodology of Kamal et *al.* (which is effected in acetonitrile) and to the other methods known generally in the prior art which are highly oxygen-and moisture-sensitive. A further feature of the invention derives from the development of a new class of protecting groups. These are based upon acetals and related protecting groups but do not suffer from some of the disadvantages of acetal deprotection known in the prior art.

[0015] The allyl deprotection methodology makes use of a water-soluble transition metal catalyst formed from a transition metal and at least partially water-soluble ligands. In aqueous solution these form at least partially water-soluble transition metal complexes. By aqueous solution herein is meant a liquid comprising at least 20 vol%, preferably at least 50%, for example at least 75 vol%, particularly at least 95 vol% and especially greater than above 98 vol%, ideally 100 vol% of water as the continuous phase.

[0016] As those skilled in the art will appreciate, the allyl group may be used to protect not only the hydroxyl group but also thiol and amine functionalities. Moreover allylic esters may be formed from the reaction between carboxylic acids and allyl halides, for example. Primary or secondary amides may also be protected using methods known in the art. The novel deprotection methodology described herein may be used in the deprotection of all these allylated compounds, e.g. allyl esters and mono- or bisallylated primary amines or allylated amides, or in the deprotection of allylated secondary amines. The method is also suitable in the deprotection of allyl esters and thioethers.

[0017] Protecting groups which comprise the acetal functionality have been used previously as blocking groups. However, removal of such groups and ethers requires strongly acidic deprotections detrimental to DNA molecules. The hydrolysis of an acetal however, results in the formation of an unstable hemiacetal intermediate which hydrolyses under aqueous conditions to the natural hydroxyl group. The inventors have utilised this concept and applied it further such that this feature of the invention resides in utilising blocking groups that included protecting groups to protect intermediate molecules that would normally hydrolyse under aqueous conditions. These protecting groups comprise a second functional group that stabilises the structure of the intermediate but which can be removed at a later stage following incorporation into the polynucleotide. Protecting groups have been used in organic synthesis reactions to temporarily mask the characteristic chemistry of a functional group because it interferes with another reaction.

[0018] Therefore, according to a first disclosure there is provided a modified nucleotide or nucleoside molecule comprising a purine or pyrimidine base and a ribose or deoxyribose sugar moiety having a removable 3'-OH, blocking group covalently attached thereto, such that the 3' carbon atom has attached a group of the structure

-O-Z

wherein Z is any of $-C(R')_2-O-R''$. $-C(R')_2-N(R'')_2$,$-C(R')_2-N(H)R''$, $-C(R')_2-S-R''$ and $-C(R')_2-F$,
wherein each R" is or is part of a removable protecting group;
each R' is independently a hydrogen atom, an alkyl, substituted alkyl, arylalkyl, alkenyl, alkynyl, aryl, heteroaryl, heterocyclic, acyl, cyano, alkoxy, aryloxy, heteroaryloxy or amido group, or a detectable label attached through a linking group; or $(R')_2$ represents an alkylidene group of formula $=C(R''')_2$ wherein each R''' may be the same or different and is selected from the group comprising hydrogen and halogen atoms and alkyl groups; and
wherein said molecule may be reacted to yield an intermediate in which each R" is exchanged for H or, where Z is $-C(R')_2-F$, the F is exchanged for OH, SH or NH_2, preferably OH, which intermediate dissociates under aqueous conditions to afford a molecule with a free 3'OH;
with the proviso that where Z is $-C(R')_2-S-R''$, both R' groups are not H.

[0019] Viewed from another aspect, the disclosure provides a 3'-O-allyl nucleotide or nucleoside which nucleotide or nucleoside comprises a detectable label linked to the base of the nucleoside or nucleotide, preferably by a cleavable linker.

[0020] In a further aspect, the disclosure provides a polynucleotide comprising a 3'-O-allyl nucleotide or nucleoside which nucleotide or nucleoside comprises a detectable label linked to the base of the nucleoside or nucleotide, preferably

by a cleavable linker.

[0021] Viewed from a still further aspect, the disclosure provides a method of converting a compound of formula R-O-allyl, R_2N (allyl) , RNH (allyl) , RN $(allyl)_2$ or R-S-allyl to a corresponding compound in which the allyl group is removed and replaced by hydrogen, said method comprising the steps of reacting a compound of formula R-O-allyl, R_2N (allyl) , RNH(allyl), RN $(allyl)_2$ or R-S-allyl in aqueous solution with a transition metal comprising a transition metal and one or more ligands selected from the group comprising water-soluble phosphine and water-soluble nitrogen-containing phosphine ligands, wherein the or each R is a water-soluble biological molecule.

[0022] In a further aspect the disclosure provides a method of controlling the incorporation of a nucleotide molecule complementary to the nucleotide in a target single-stranded polynucleotide in a synthesis or sequencing reaction comprising incorporating into the growing complementary polynucleotide a molecule according to the invention, the incorporation of said molecule preventing or blocking introduction of subsequent nucleoside or nucleotide molecules into said growing complementary polynucleotide.

[0023] In a further aspect, the disclosure provides a method for determining the sequence of a target single-stranded polynucleotide, comprising monitoring the sequential incorporation of complementary nucleotides, wherein at least one incorporation, and preferably all of the incorporations is of a nucleotide according to the invention as hereinbefore described which preferably comprises a detectable label linked to the base of the nucleoside or nucleotide by a cleavable linker and wherein the identity of the nucleotide incorporated is determined by detecting the label, said blocking group and said label being removed prior to introduction of the next complementary nucleotide.

[0024] From a further aspect, the disclosure provides a method for determining the sequence of a target single-stranded polynucleotide, comprising:

(a) providing a plurality of different nucleotides according to the hereinbefore described invention which nucleotides are preferably linked from the base to a detectable label by a cleavable linker and wherein the detectable label linked to each type of nucleotide can be distinguished upon detection from the detectable label used for other types of nucleotides;
(b) incorporating the nucleotide into the complement of the target single-stranded polynucleotide;
(c) detecting the label of the nucleotide of (b), thereby determining the type of nucleotide incorporated;
(d) removing the label of the nucleotide of (b) and the blocking group; and
(e) optionally repeating steps (b)-(d) one or more times;

thereby determining the sequence of a target single-stranded polynucleotide.

[0025] Additionally, in another aspect, the disclosure provides a kit, comprising:

(a) a plurality of different individual nucleotides of the invention; and
(b) packaging materials therefor.

[0026] The nucleosides or nucleotides according to or used in the methods of the present disclosure comprise a purine or pyrimidine base and a ribose or deoxyribose sugar moiety which has a blocking group covalently attached thereto, preferably at the 3'O position, which renders the molecules useful in techniques requiring blocking of the 3'-OH group to prevent incorporation of additional nucleotides, such as for example in sequencing reactions, polynucleotide synthesis, nucleic acid amplification, nucleic acid hybridisation assays, single nucleotide polymorphism studies, and other such techniques.

[0027] Where the term "blocking group" is used herein, this embraces both the allyl and "Z" blocking groups described herein. However, it will be appreciated that, in the methods as described and claimed herein, where mixtures of nucleotides are used, these very preferably each comprise the same type of blocking, i.e. allyl-blocked or "Z"-blocked. Where "Z"-blocked nucleotides are used, each "Z" group will generally be the same group, except in those cases where the detectable label forms part of the "Z" group, i.e. is not attached to the base.

[0028] Once the blocking group has been removed, it is possible to incorporate another nucleotide to the free 3'-OH group.

[0029] The molecule can be linked via the base to a detectable label by a desirable linker, which label may be a fluorophore, for example. The detectable label may instead, if desirable, be incorporated into the blocking groups of formula "Z". The linker can be acid labile, photolabile or contain a disulfide linkage. Other linkages, in particular phosphine-cleavable azide-containing linkers, may be employed described in greater detail.

[0030] Preferred labels and linkages included those disclosed in WO 03/048387.

[0031] In the methods where nucleotides are incorporated, e.g. where the incorporation of a nucleotide molecule complementary to the nucleotide in a target single stranded polynucleotide is controlled in a synthesis or sequencing reaction of the invention, the incorporation of the molecule may be accomplished via a terminal transferase, a polymerase or a reverse transcriptase.

[0032] Preferably, the molecule is incorporated by a polymerase and particularly from *Thermococcus* sp., such as 9°N. Even more preferably, the polymerase is a mutant 9°N A485L and even more preferably is a double mutant Y409V and A485L.

[0033] In the methods for determining the sequence of a target single-stranded polynucleotide comprising monitoring the sequential incorporation of complementary nucleotides of the invention, it is preferred that the blocking group and the label may be removed in a single chemical treatment step. Thus, in a preferred embodiment of the invention, the blocking group is cleaved simultaneously with the label. This will of course be a feature inherent to those blocking groups of formula Z which incorporate a detectable label.

[0034] Furthermore, preferably the blocked and labelled modified nucleotide constructs of the nucleotide bases A, T, C and G are recognised as substrates by the same polymerase enzyme.

[0035] In the methods described herein, each of the nucleotides can be brought into contact with the target sequentially, with removal of non-incorporated nucleotides prior to addition of the next nucleotide; where detection and removal of the label and the blocking group is carried out either after addition of each nucleotide, or after addition of all four nucleotides.

[0036] In the methods, all of the nucleotides can be brought into contact with the target simultaneously, i.e., a composition comprising all of the different nucleotides is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and the blocking group.

[0037] The methods can comprise a first step and a second step, where in the first step, a first composition comprising two of the four types of modified nucleotides is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and the blocking group, and where in the second step, a second composition comprising the two nucleotides not included in the first composition is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group, and where the first steps and the second step can be optionally repeated one or more times.

[0038] The methods described herein can also comprise a first step and a second step, where in the first step, a composition comprising one of the four nucleotides is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group, and where in the second step, a second composition, comprising the three nucleotides not included in the first composition is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group, and where the first steps and the second step can be optionally repeated one or more times.

[0039] The methods described herein can also comprise a first step and a second step, where in the first step, a first composition comprising three of the four nucleotides is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group and where in the second step, a composition comprising the nucleotide not included in the first composition is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group, and where the first steps and the second step can be optionally repeated one or more times.

[0040] The incorporating step in the methods of the invention can be accomplished via a terminal transferase, a polymerase or a reverse transcriptase as hereinbefore defined. The detectable label and/or the cleavable linker can be of a size sufficient to prevent the incorporation of a second nucleotide or nucleoside into the nucleic acid molecule.

[0041] In certain methods described herein for determining the sequence of a target single-stranded polynucleotide, each of the four nucleotides, one of which will be complementary to the first unpaired base in the target polynucleotide, can be brought into contact with the target sequentially, optionally with removal of non-incorporated nucleotides prior to addition of the next nucleotide. Determination of the success of the incorporation may be carried out either after provision of each nucleotide, or after the addition of all of the nucleotides added. If it is determined after addition of fewer than four nucleotides that one has been incorporated, it is not necessary to provide further nucleotides in order to detect the nucleotides complementary to the incorporated nucleotide.

[0042] Alternatively, all of the nucleotides can be brought into contact with the target simultaneously, i.e., a composition comprising all of the different nucleotide (i.e. A, T, C and G or A, U, C and G) is brought into contact with the target, and non-incorporated nucleotides removed prior to detection and removal of the label(s). The methods involving sequential addition of nucleotides may comprise a first substep and optionally one or more subsequent substeps. In the first substep a composition comprising one, two or three of the four possible nucleotides is provided, i.e. brought into contact with, the target. Thereafter any unincorporated nucleotides may be removed and a detecting step may be conducted to determine whether one of the nucleotides has been incorporated. If one has been incorporated, the cleavage of the linker may be effected. In this way the identity of a nucleotide in the target polynucleotide may be determined. The nascent polynucleotide may then be extended to determine the identity of the next unpaired nucleotide in the target oligonucleotide.

[0043] If the first substep above does not lead to incorporation of a nucleotide, or if this is not known, since the presence of incorporated nucleotides is not sought immediately after the first substep, one or more subsequent substeps may be conducted in which some or all, of those nucleotides not provided in the first substep are provided either, as appropriate, simultaneously or subsequently. Thereafter any unincorporated nucleotides may be removed and a detecting step

conducted to determine whether one of the classes of nucleotide has been incorporated. If one has been incorporated, cleavage of the linker may be effected. In this way the identity of a nucleotide in the target polynucleotide may be determined. The nascent polynucleotide may then be extended to determine the identity of the next unpaired nucleotide in the target oligonucleotide. If necessary, a third and optionally a fourth substep may be effected in a similar manner to the second substep. Obviously, once four substeps have been effected, all four possible nucleotides will have been provided and one will have been incorporated.

[0044] It is desirable to determine whether a type or class of nucleotide has been incorporated after any particular combination comprising one, two or three nucleotides has been provided. In this way the unnecessary cost and time expended in providing the other nucleotide(s) is obviated. This is not a required feature of the invention, however.

[0045] It is also desirable, where the method for sequencing comprises one or more substeps, to remove any unincorporated nucleotides before further nucleotide are provided. Again, this is not a required feature of the invention. Obviously, it is necessary that at least some and preferably as many as practicable of the unincorporated nucleotides are removed prior to the detection of the incorporated nucleotide.

[0046] The kits of the invention include: (a) individual nucleotides according to the hereinbefore described invention, where each nucleotide has a base that is linked to a detectable label via a cleavable linker, or a detectable label linked via an optionally cleavable liner to a blocking group of formula Z, and where the detectable label linked to each nucleotide can be distinguished upon detection from the detectable label used for other three nucleotides; and (b) packaging materials therefor. The kit can further include an enzyme for incorporating the nucleotide into the complementary nucleotide chain and buffers appropriate for the action of the enzyme in addition to appropriate chemicals for removal of the blocking group and the detectable label, which can preferably be removed by the same chemical treatment step.

[0047] The nucleotides/nucleosides are suitable for use in many different DNA-based methodologies, including DNA synthesis and DNA sequencing protocols.

[0048] The invention may be understood with reference to the attached drawings in which:

Fig. 1 shows exemplary nucleotide structures useful in certain aspects of the invention. For each structure, X can be H, phosphate, diphosphate or triphosphate. R_1 and R_2 can be the same or different, and can be selected from H, OH, or any group of claim 1 which can be transformed into an OH, Some suitable functional groups for R_1 and R_2 include the structures shown in Fig. 3 and Fig. 4.

Fig. 2 shows structures of linkers useful in certain aspects of the invention , including (1) disulfide linkers and acid labile linkers, (2) dialkoxybenzyl linkers, (3) Sieber linkers, (4) indole linkers and (5) t-butyl Sieber linkers.

Fig. 3 shows some functional molecules useful in certain aspects of the invention, including some cleavable linkers and some suitable hydroxyl protecting groups. In these structures, R_1 and R_2 may be the same of different, and can be H, OH, or any group of claim 1 which can be transformed into an OH group. R_3 represents one or more substituents independently selected from alkyl, alkoxyl, amino or halogen groups. R_4 and R_5 can be H or alkyl, and R_6 can be alkyl, cycloalkyl, alkenyl, cycloalkenyl or benzyl. X can be H, phosphate, diphosphate or triphosphate.

Fig. 4 is a schematic illustration of some of the Z blocking groups that can be used according to certain aspects of the invention.

Fig. 5 shows two cycles of incorporation of labelled and blocked DGTP, DCTP and dATP respectively (compounds 18, 24 and 32).

Fig. 6 shows six cycles of incorporation of labelled and blocked DTTP (compound 6).

Fig.7 shows the effective blocking by compound 38 (a 3'-0allyl nucleotide

[0049] The present disclosure relates to nucleotide or nucleoside molecules that are modified by the reversible covalent attachment of a 3'-OH blocking groups thereto, and which molecules may be used in reactions where blocked nucleotide or nucleoside molecules are required, such as in sequencing reactions, polynucleotide synthesis and the like.

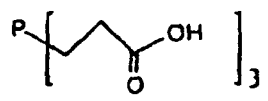
[0050] As used herein the term biological molecule is used to embrace any molecules or class of molecule which performs a biological role. Such molecules include for example, polynucleotides such as DNA and RNA, oligonucleotides and single nucleotides. In addition, peptides and peptide mimetics, such as enzymes and hormones etc., are embraced by the invention. Compounds which comprise a secondary amide linkage, such as peptides, or a secondary amine, where such compounds are allylated on the nitrogen atom of the secondary amine or amide, are examples of compounds of formula R_2N(allyl) in which both R groups belong to the same biological molecule. Particularly preferred compounds however are polynucleotides, (including oligonucleotides) and nucleotides and nucleosides, preferably those which contain one base to which is attached a detectable label linked through a cleavable linker. Such compounds are useful in the determination of sequences of oligonucleotides as described herein.

[0051] Suitable ligands are any phosphine or mixed nitrogen-phosphine ligands known to those skilled in the art, characterised in that the ligands are derivatised so as to render them water-soluble, e.g. by introducing one or more sulfonate, amine, hydroxyl (preferably a plurality of hydroxyl) or carboxylate residues. Where amine residues are present, formation of amine salts may assist the solublisation of the ligand and thus the metal-allyl complex. Examples of appro-

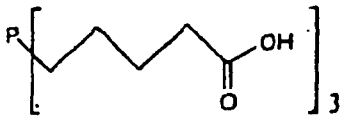
priate ligands are triaryl phosphines, e.g. triphenyl phosphine, derivatised so as to make them water-soluble. Also preferred are trialkyl phosphines, e.g. tri-C_{1-6}-alkyl phosphines such as triethyl phosphines; such trialkyl phosphines are likewise derivatised so as to make them water-soluble. Sulfonate-containing and carboxylate-containing phosphines are particularly preferred; an example of the former 3,3',3"-phosphinidynetris (benzenesulfonic acid) which is commercially available from Aldrich Chemical Company as the trisodium salt; and a preferred example of the latter is tris(2-carboxyethyl) phosphine which is available from Aldrich as the hydrochloride salt.

[0052] The derivatised water-soluble phosphines and nitrogen-containing phosphines described herein may be used as their salts (e.g. as the hydrochloride or sodium salts) or, for example, in the case of the sulfonic and carboxylic acid-containing phosphines described herein, as the free acids. Thus 3,3',3"-phosphinidynetris (benzenesulfonic acid) and tris(2-carboxyethyl)phosphines may be introduced either as the triacids or the trisodium salts. Other appropriate salts will be evident to those skilled in the art. The existence in salt form is not particularly important provided the phosphines are soluble in aqueous solution.

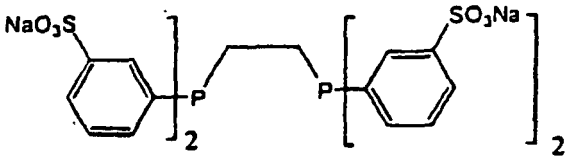
[0053] Other ligands which may be used to include the following:

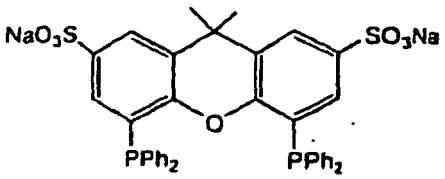


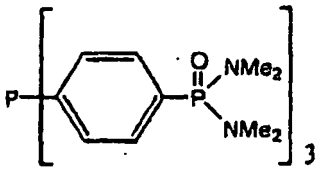
as the hydrochloride salt,neutral compound and the trisodium salt

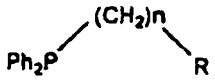


as the hydrochloride salt,neutral compound and the trisodium salt

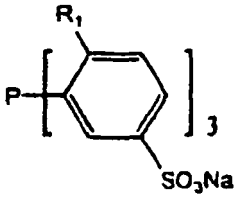




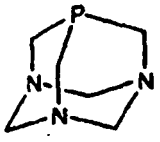


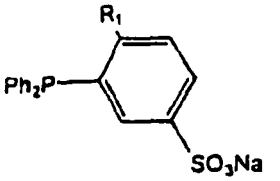


n=1-5
R=PO3Na2,
CO2Na,
SO3Na,
OH

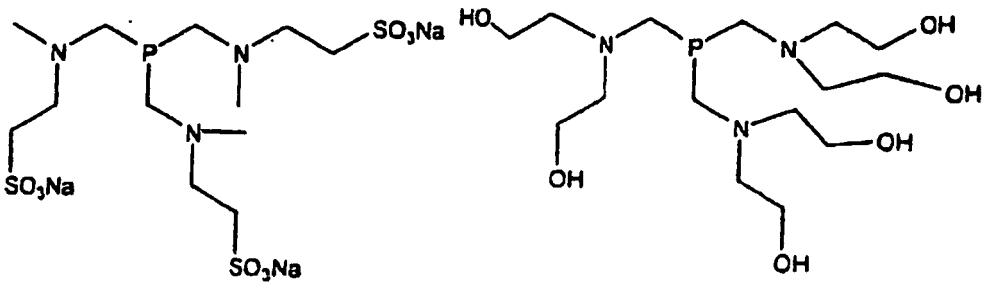


R1=OMe,CO2H,CO2Na





R1=OMe,CO2H,CO2Na



[0054] The skilled person will be aware that the atoms chelated to the transition metal in the water soluble complex may be part of mono- or polydentate ligands. Some such polydentate ligands are shown above. Whilst monodentate ligands are preferred, the invention thus also embraces methods which use water-soluble bi-, tri-, tetra-, penta- and hexadentate water-soluble phosphine and water-soluble nitrogen-containing phosphine ligands

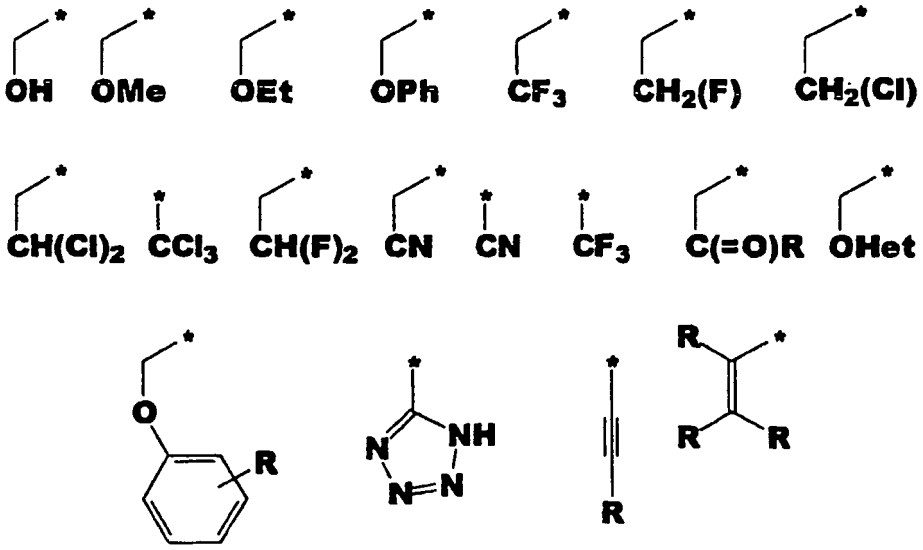
[0055] The skilled person will be able to determine the quantity of ligand best suited to any individual reaction.

[0056] Where the blocking group is any of $C(R')_2$-$N(R")_2$ or -$C(R')_2$-N(H)R", i.e. of formula Z, each R' may be independently H or an alkyl

[0057] The intermediates produced advantageously spontaneously dissociate under aqueous conditions back to the natural 3' hydroxy structure, which permits further incorporation of another nucleotide. Any appropriate protecting group may be used, as discussed herein. Preferably, Z is of formula $C(R')_2$-$N(R")_2$ or -$C(R')_2$-N(H)R"
Particularly preferably, Z is of the formula , -$C(R')_2$-$N(R")_2$. R" R" may be a benzyl group or a substituted benzyl group.

[0058] One example of groups of structure -O-Z wherein Z is -$C(R')_2$-$N(R")_2$ are those in which -$N(R")_2$ is azido (-N_3). One preferred such example is azidomethyl wherein each R' is H. Alternatively, R' in Z groups of formula -$C(R')_2$-N_3 and other Z groups may be any of the other groups discussed herein.

[0059] Examples of typical R' groups include C_{1-6} alkyl, particularly methyl and ethyl, and the following (in which each structure shows the bond which connects the R' moiety to the carbon atom to which it is attached in the Z groups; the asterisks (*) indicate the points of attachment):



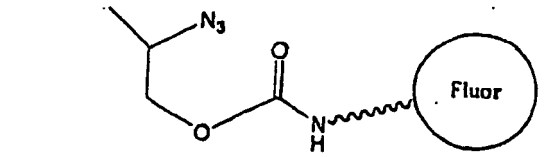
(wherein each R is an optionally substituted C_{1-10} alkyl group, an optionally substituted alkoxy group, a halogen atom or functional group such as hydroxyl, amino, cyano, nitro, carboxyl and the like) and "Het" is a heterocyclic (which may for example be a heteroaryl group). These R' groups shown above are preferred where the other R' group is the same as the first or is hydrogen. Preferred Z groups are of formula $C(R')_2N_3$ in which the R' groups are selected from the structures given above and hydrogen; or in which $(R')_2$ represents an alkylidene group of formula $=C(R'")_2$, e.g. $=C(Me)_2$.

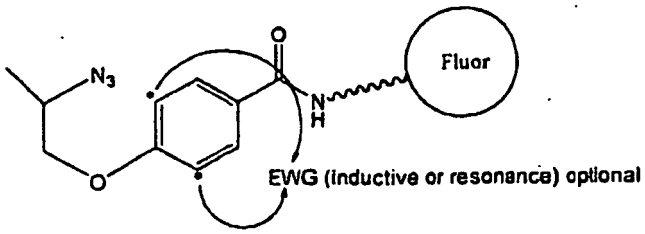
[0060] Where molecules contain Z groups of formula $C(R')_2N_3$, the azido group may be converted to amino by contacting such molecules with the phosphine or nitrogen-containing phosphines ligands described in detail herein. Alternatively, the azido group in Z groups of formula $C(R')_2N_3$ may be converted to amino by contacting such molecules with the thiols, in particular water-soluble thiols such as dithiothreitol (DTT).

[0061] Where an R' group represents a detectable label attached through a linking group, the other R' group or any

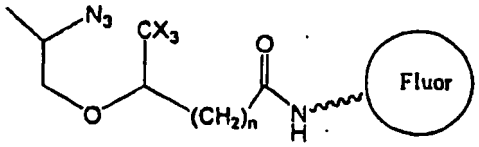
other part of "Z" will generally not contain a detectable label, nor will the base of the nucleoside or nucleotide contain a detectable label. Appropriate linking groups for connecting the detectable label to the 3'blocking group will be known to the skilled person and examples of such groups are described in greater detail hereinafter.

[0062] Exemplary of linkages in R' groups containing detectable labels are those which contain one or more amide bonds. Such linkers may also contain an arylene, e.g. phenylene, group in the chain (i.e. a linking moiety -Ar- where the phenyl ring is part of the linker by way of its 1,4-disposed carbon atoms). The phenyl ring may be substituted at its non-bonded position with one or more substituents such as alkyl, hydroxyl, alkyloxy, halide, nitro, carboxyl or cyano and the like, particularly electron-withdrawing groups, which electron-withdrawing is either by induction or resonance. The linkage in the R' group may also include moieties such a -O-, $-S(O)_q$, wherein $_q$ is 0, 1 or 2 or NH or Nalkyl. Examples of such Z groups are as follows:





X=halogen, electron withdrawing group



(wherein EWG stands for electron-withdrawing group; n is an integer of from 1 to 50, preferably 2-20, e.g. 3 to 10; and fluor indicates a fluorophore). An example of an electron-withdrawing group by resonance is nitro; a group which acts through induction is fluoro. The skilled person will be aware of other appropriate electron-withdrawing groups. In addition, it will be understood that whilst a fluorophore is indicated as being the detectable label present, other detectable groups as discussed in greater detail hereinafter may be included instead.

[0063] Where a detectable label is attached to a nucleotide at the 3'-blocking position, the linker need not be cleavable to have utility in those reactions, such as DNA sequencing, described herein which require the label to be "read" and removed before the next step of the reaction. This is because the label, when attached to the 3'block, will become separated from the nucleotide when the intermediate compounds described herein collapse so as to replace the "Z" group with a hydrogen atom. As noted above, each R" is or is part of a removable protecting group. R" may be a benzyl group or is substituted benzyl group is an alternative embodiment.

[0064] It will be appreciated that where it is possible to incorporate a detectable label onto a group R", the invention embraces this possibility. Thus, where R" is a benzyl group, the phenyl ring may bear a linker group to which is attached a fluorophore or other detectable group. Introduction of such groups does not prevent the ability to remove such R"s and they do not prevent the generation of the desired unstable intermediates during deprotection of blocking groups of formula Z.

[0065] As is known in the art, a "nucleotide" consists of a nitrogenous base, a sugar, and one or more phosphate groups. They are monomeric units of a nucleic acid sequence. In RNA, the sugar is a ribose, and in DNA a deoxyribose, i.e. a sugar lacking a hydroxyl group that is present in ribose. The nitrogenous base is a derivative of purine or pyrimidine. The purines are adenine (A) and guanine (G), and the pyrimidines are cytosine (C) and thymine (T) (or in the context of RNA, uracil (U)). The C-1 atom of deoxyribose is bonded to N-1 of a pyrimidine or N-9 of a purine. A nucleotide is also a phosphate ester or a nucleoside, with esterification occurring on the hydroxyl group attached to C-5 of the sugar.

Nucleotides are usually mono, di- or triphosphates.

[0066] A "nucleoside" is structurally similar to a nucleotide, but is missing the phosphate moieties. An example of a nucleoside analogue would be one in which the label is linked to the base and there is no phosphate group attached to the sugar molecule.

[0067] Although the base is usually referred to as a purine or pyrimidine, the skilled person will appreciate that derivatives and analogues are available which do not alter the capability of the nucleotide or nucleoside to undergo Watson-Crick base pairing. "Derivative" or "analogue" means a compound or molecule whose core structure is the same as, or closely resembles that of, a parent compound, but which has a chemical or physical modification, such as a different or additional side group, or 2' and or 3' blocking groups, which allows the derivative nucleotide or nucleoside to be linked to another molecule. For example, the base can be a deazapurine. The derivatives should be capable of undergoing Watson-Crick pairing. "Derivative" and "analogue" also mean a synthetic nucleotide or nucleoside derivative having modified base moieties and/or modified sugar moieties. Such derivatives and analogs are discussed in, e.g., Scheit, Nucleotide Analogs (John Wiley & Son, 1980) and Uhlman et al., Chemical Reviews 90:543-584, 1990. Nucleotide analogs can also comprise modified phosphodiester linkages, including phosphorothioate, phosphorodithioate, alkyl-phosphonate, phosphoranilidate and phosphoramidate linkages. The analogs should be capable of undergoing Watson-Crick base pairing. "Derivative", "analog" and "modified" as used herein, may be used interchangeably, and are encompassed by the terms "nucleotide" and "nucleoside" defined herein.

[0068] In the context of the present invention, the term "incorporating" means becoming part of a nucleic acid (eg DNA) molecule or oligonucleotide or primer. An oligonucleotide refers to a synthetic or natural molecule comprising a covalently linked sequence of nucleotides which are formed by a phosphodiester or modified phosphodiester bond between the 3' position of the pentose on one nucleotide and the 5' position of the pentose on an adjacent nucleotide.

[0069] The term "alkyl" covers straight chain, branched chain and cycloalkyl groups. Unless the context indicates otherwise, the term "alkyl" refers to groups having 1 to 10 carbon atoms, for example 1 to 8 carbon atoms, and typically from 1 to 6 carbon atoms, for example from 1 to 4 carbon atoms. Examples of alkyl groups include methyl, ethyl, propyl, isopropyl, n-butyl, isobutyl, tert-butyl, n-pentyl, 2-pentyl, 3-pentyl, 2-methyl butyl, 3-methyl butyl, and n-hexyl and its isomers.

[0070] Examples of cycloalkyl groups are those having from 3 to 10 ring atoms, particular examples including those derived from cyclopropane, cyclobutane, cyclopentane, cyclohexane and cycloheptane, bicycloheptane and decalin.

[0071] Where alkyl (including cycloalkyl) groups are substituted, particularly where these form either both of the R' groups of the molecules of the invention, examples of appropriate substituents include halogen substituents or functional groups such as hydroxyl, amino, cyano, nitro, carboxyl and the like. Such groups may also be substituents, where appropriate, of the other R' groups in the molecules of the invention.

[0072] The term amino refers to groups of type NR^*R^{**}, wherein R^* and R^{**} are independently selected from hydrogen, a C_{1-6} alkyl group (also referred to as C_{1-6} alkylamino or di-C_{1-6} alkylamino) .

[0073] The term "halogen" as used herein includes fluorine, chlorine, bromine and iodine.

[0074] The nucleotide molecules of the present invention are suitable for use in many different methods were the detection of nucleotides is required.

[0075] DNA sequencing methods, such as those outlined in U.S. Pat. No. 5,302,509 can be carried out using the nucleotides.

[0076] The present invention can make use of conventional detectable labels. Detection can be carried out by any suitable method, including fluorescence spectroscopy or by other optical means. The preferred label is a fluorophore, which, after absorption of energy, emits radiation at a defined wavelength. Many suitable fluorescent labels are known. For example, Welch et al. (Chem. Eur. J. 5(3):951-960, 1999) discloses dansyl-functionalised fluorescent moieties that can be used in the present invention. Zhu et al. (Cytometry 28:206-211, 1997) describes the use of the fluorescent labels Cy3 and Cy5, which can also be used in the present invention. Labels suitable for use are also disclosed in Prober et al., (Science 238:336-341, 1987); Connell et al. (BioTechniques 5(4):342-38.4, 1987), Ansorge et al. (Nucl. Acids Res. 15(11):4593-4602, 1987) and Smith et al. (Nature 321:674, 1986). Other commercially available fluorescent labels include, but are not limited to, fluorescein, rhodamine (including TMR, texas red and Rox)., alexa, bodipy, acridine, coumarin, pyrene, benzanthracene and the cyanins.

[0077] Multiple labels can also be used in the invention. For example, bi-fluorophore FRET cassettes (Tet. Let. 46: 8867-8871, 2000) are well known in the art and can be utilised in the present invention. Multi-fluor dendrimeric systems (J. Amer. Chem. Soc. 123:8101-8108, 2001) can also be used.

[0078] Although fluorescent labels are preferred, other forms of detectable labels will be apparent as useful to those of ordinary skill. For example, microparticles, including quantum dots (Empodocles et al., Nature 399:126-130, 1999), gold nanoparticles (Reichert et al., Anal. Chem. 72:6025-6029, 2000) and microbeads (Lacoste et al., Proc. Natl. Acad. Sci USA 97(17):9461-9466, 2000) can all be used.

[0079] Multi-component labels can also be used in the invention. A multi-component label is one which is dependent on the interaction with a further compound for detection. The most common multi-component label used in biology is

the biotin-streptavidin system. Biotin is used as the label attached to the nucleotide base. Streptavidin is then added separately to enable detection to occur. Other multi-component systems are available. For example, dinitrophenol has a commercially available fluorescent antibody that can be used for detection.

[0080] The invention has been and will be further described with reference to nucleotides. However, unless indicated otherwise, the reference to nucleotides is also intended to be applicable to nucleosides. The invention will also be further described with reference to DNA, although the description will also be applicable to RNA, PNA, and other nucleic acids, unless otherwise indicated.

[0081] The modified nucleotides of the'invention may use a cleavable linker to attach the label to the nucleotide. The use of a cleavable linker ensures that the label can, if required, be removed after detection, avoiding any interfering signal with any labelled nucleotide incorporated subsequently.

[0082] Generally, the use of cleavable linkers is preferable, particularly in the methods of the invention hereinbefore described except where the detectable label is attached to the nucleotide by forming part of the "Z" group.

[0083] Those skilled in the art will be aware of the utility of dideoxynucleoside triphosphates in so-called Sanger sequencing methods, and related protocols (Sanger-type), which rely upon randomised chain-termination at a particular type of nucleotide. An example of a Sanger-type sequencing protocol is the BASS method described by Metzker (infra). Other Sanger-type sequencing methods will be known to those skilled in the art.

[0084] Sanger and Sanger-type methods generally operate by the conducting of an experiment in which eight types of nucleotides are provided, four of which contain a 3'OH group; and four of which omit the OH group and which are labeled differently from each other. The nucleotides used which omit the 3'OH group - dideoxy nucleotides - are conventially abbreviated to ddNTPs. As is known by the skilled person, since the ddNTPs are labeled differently, by determining the positions of the terminal nucleotides incorporated, and combining this information, the sequence of the target oligonucleotide may be determined.

[0085] The nucleotides of the present invention, it will be recognized, may be of utility in Sanger methods and related protocols since the same effect achieved by using ddNTPs may be achieved by using the novel 3'-OH blocking groups described herein: both prevent incorporation of subsequent nucleotides.

[0086] The use of the nucleotides according to the present invention in Sanger and Sanger-type sequencing methods, wherein the linker connecting the detectable label to the nucleotide may or may not be cleavable, forms a still further aspect of this invention. Viewed from this aspect, the invention provides the use of such nucleotides in a Sanger or a Sanger-type sequencing method.

[0087] Where 3'-OH Z-blocked nucleotides according to the present invention are used, it will be appreciated that the detectable labels attached to the nucleotides need not be connected via cleavable linkers, since in each instance where a labelled nucleotide of the invention is incorporated, no nucleotides need to be subsequently incorporated and thus the label need not be removed from the nucleotide.

[0088] Moreover, it will be appreciated that monitoring of the incorporation of 3'OH blocked nucleotides may be determined by use of radioactive ^{32}P in the phosphate groups attached. These may be present in either the ddNTPs themselves or in the primers used for extension. Where the blocking groups are of formula "Z", this represents a further aspect of the invention.

[0089] Viewed from this aspect, the invention provides the use of a nucleotide having a 3'OH group blocked with a "Z" group in a Sanger or a Sanger-type sequencing method. In this embodiment, a ^{32}P detectable label may be present in either the ddNTPs used in the primer used for extension.

[0090] Cleavable linkers are known in the art, and conventional chemistry can be applied to attach a linker to a nucleotide base and a label. The linker can be cleaved by any suitable method, including exposure to acids, bases, nucleophiles, electrophiles, radicals, metals, reducing or oxidising agents, light, temperature, enzymes etc. The linker as discussed herein may also be cleaved with the same catalyst used to cleave the 3'O-blocking group bond. Suitable linkers can be adapted from standard chemical blocking groups, as disclosed in Greene & Wuts, Protective Groups in Organic Synthesis, John Wiley & Sons. Further suitable cleavable linkers used in solid-phase synthesis are disclosed in Guillier et al. (Chem. Rev. 100:2092-2157, 2000).

[0091] The use of the term "cleavable linker" is not meant to imply that the whole linker is required to be removed from e.g., the nucleotide base. Where the detectable label is attached to the base, the nucleoside cleavage site can be located at a position on the linker that ensures that part of the linker remains attached to the nucleotide base after cleavage.

[0092] Where the detectable label is attached to the base, the linker can be attached at any position on the nucleotide base provided that Watson-Crick base pairing can still be carried out. In the context of purine bases, it is preferred if the linker is attached via the 7-position of the purine or the preferred deazapurine analogue, via an 8-modified purine, via an N-6 modified adenosine or an N-2 modified guanine. For pyrimidines, attachment is preferably via the 5-position on cytosine, thymidine or uracil and the N-4 position on cytosine. Suitable nucleotide structures are shown in Fig. 1. For each structure in Fig. 1 X can be H, phosphate, diphosphate or triphosphate. R_1 and R_2 can be the same or different, and are selected from H, OH, O-allyl, or formula Z as described herein or any other group which can be transformed into an OH, including, but not limited to, a carbonyl, provided that at least one of R_1 and R_2 is O-allyl or formula Z as

described herein. Some suitable functional groups for R_1 and R_2 include the structures shown in Figs. 3 and 4.

[0093] Suitable linkers are shown in Fig. 3 and include, but are not limited to, disulfide linkers (1), acid labile linkers (2, 3, 4 and 5; including dialkoxybenzyl linkers (e.g., 2), Sieber linkers (e.g., 3), indole linkers (e.g., 4), t-butyl Sieber linkers (e.g., 5)), electrophilically-cleavable linkers, nucleophilically cleavable linkers, photocleavable linkers, cleavage under reductive conditions, oxidative conditions, cleavage via use of safety-catch linkers, and cleavage by elimination mechanisms.

A. Electrophilically cleaved linkers.

[0094] Electrophilically cleaved linkers are typically cleaved by protons and include cleavages sensitive to acids. Suitable linkers include the modified benzylic systems such as trityl, p-alkoxybenzyl esters and p-alkoxybenzyl amides. Other suitable linkers include tert-butyloxycarbonyl (Boc) groups and the acetal system.

[0095] The use of thiophilic metals, such as nickel, silver or mercury, in the cleavage of thioacetal or other sulfur-containing protecting groups can also be considered for the preparation of suitable linker molecules.

B. Nucleophilically cleaved linkers.

[0096] Nucleophilic cleavage is also a well recognised method in the preparation of linker molecules. Groups such as esters that are labile in water (i.e., can be cleaved simply at basic pH) and groups that are labile to non-aqueous nucleophiles, can be used. Fluoride ions can be used to cleave silicon-oxygen bonds in groups such as triisopropyl silane (TIPS) or t-butyldimethyl silane (TBDMS).

C. Photocleavable linkers.

[0097] Photocleavable linkers have been used widely in carbohydrate chemistry. It is preferable that the light required to activate cleavage does not affect the other components of the modified nucleotides. For example, if a fluorophore is used as the label, it is preferable if this absorbs light of a different wavelength to that required to cleave the linker molecule. Suitable linkers include those based on 0-nitrobenzyl compounds and nitroveratryl compounds. Linkers based on benzoin chemistry can also be used (Lee et al., J. Org. Chem. 64:3454-3460, 1999).

D. Cleavage under reductive conditions

[0098] There are many linkers known that are susceptible to reductive cleavage. Catalytic hydrogenation using palladium-based catalysts has been used to cleave benzyl and benzyloxycarbonyl groups. Disulfide bond reduction is also known in the art.

E. Cleavage under oxidative conditions

[0099] Oxidation-based approaches are well known in the art. These include oxidation of p-alkoxybenzyl groups and the oxidation of sulfur and selenium linkers. The use of aqueous iodine to cleave disulfides and other sulfur or selenium-based linkers is also within the scope of the invention.

F. Safety-catch linkers

[0100] Safety-catch linkers are those that cleave in two steps. In a preferred system the first step is the generation of a reactive nucleophilic center followed by a second step involving an intra-molecular cyclization that results in cleavage. For example, levulinic ester linkages can be treated with hydrazine or photochemistry to release an active amine, which can then be cyclised to cleave an ester elsewhere in the molecule (Burgess et al., J. Org. Chem. 62:5165-5168, 1997).

G. Cleavage by elimination mechanisms

[0101] Elimination reactions can also be used. For example, the base-catalysed elimination of groups such as Fmoc and cyanoethyl, and palladium-catalysed reductive elimination of allylic systems, can be used.

[0102] As well as the cleavage site, the linker can comprise a spacer unit. The spacer distances e.g., the nucleotide base from the cleavage site or label. The length of the linker is unimportant provided that the label is held a sufficient distance from the nucleotide so as not to interfere with any interaction between the nucleotide and an enzyme.

[0103] In a preferred embodiment the linker may consist of the same functionality as the block. This will make the deprotection and deblocking process more efficient, as only a single treatment will be required to remove both the label

and the block.

[0104] Particularly preferred linkers are phosphine-cleavable azide containing linkers.

[0105] A method for determining the sequence of a target polynucleotide can be carried out by contacting the target polynucleotide separately with the different nucleotides to form the complement to that of the target polynucleotide, and detecting the incorporation of the nucleotides. Such a method makes use of polymerisation, whereby a polymerase enzyme extends the complementary strand by incorporating the correct nucleotide complementary to that on the target. The polymerisation reaction also requires a specific primer to initiate polymerisation.

[0106] For each cycle, the incorporation of the modified nucleotide is carried out by the polymerase enzyme, and the incorporation event is then determined. Many different polymerase enzymes exist, and it will be evident to the person of ordinary skill which is most appropriate to use. Preferred enzymes include DNA polymerase I, the Klenow fragment, DNA polymerase III, T4 or T7 DNA polymerase, Taq polymerase or Vent polymerase. Polymerases engineered to have specific properties can also be used. As noted earlier, the molecule is preferably incorporated by a polymerase and particularly from *Thermococcus sp.,* such as 9°N. Even more preferably, the polymerase is a mutant 9°N A485L and even more preferably is a double mutant Y409V and A485L. An example of one such preferred enzyme is *Thermococcus sp.* 9°N exo -Y409V A485L available from New England Biolabs. Examples of such appropriate polymerases are disclosed in Proc. Natl. Acad. Sci. USA, 1996(93), pp 5281-5285, Nucleic Acids Research, 1999(27), pp 2454-2553 and Acids Research, 2002(30), pp 605-613.

[0107] The sequencing methods are preferably carried out with the target polynucleotide arrayed on a solid support. Multiple target polynucleotides can be immobilised on the solid support through linker molecules, or can be attached to particles, e.g., microspheres, which can also be attached to a solid support material. The polynucleotides can be attached to the solid support by a number of means, including the use of biotin-avidin interactions. Methods for immobilizing polynucleotides on a solid support are well known in the art, and include lithographic techniques and "spotting" individual polynucleotides in defined positions on a solid support. Suitable solid supports are known in the art, and include glass slides and beads, ceramic and silicon surfaces and plastic materials. The support is usually a flat surface although microscopic beads (microspheres) can also be used and can in turn be attached to another solid support by known means. The microspheres can be of any suitable size, typically in the range of from 10 nm to 100 nm in diameter. In a preferred embodiment, the polynucleotides are attached directly onto a planar surface, preferably a planar glass surface. Attachment will preferably be by means of a covalent linkage. Preferably, the arrays that are used are single molecule arrays that comprise polynucleotides in distinct optically resolvable areas, e.g., as disclosed in International Application No. WO00/06770.

[0108] The sequencing method can be carried out on both single polynucleotide molecule and multi-polynucleotide molecule arrays, i.e., arrays of distinct individual polynucleotide molecules and arrays of distinct regions comprising multiple copies of one individual polynucleotide molecule. Single molecule arrays allow each individual polynucleotide to be resolved separately. The use of single molecule arrays is preferred. Sequencing single molecule arrays non-destructively allows a spatially addressable array to be formed.

[0109] The method makes use of the polymerisation reaction to generate the complementary sequence of the target. Conditions compatible with polymerization reactions will be apparent to the skilled person.

[0110] To carry out the polymerase reaction it will usually be necessary to first anneal a primer sequence to the target polynucleotide, the primer sequence being recognised by the polymerase enzyme and acting as an initiation site for the subsequent extension of the complementary strand. The primer sequence may be added as a separate component with respect to the target polynucleotide. Alternatively, the primer and the target polynucleotide may each be part of one single stranded molecule, with the primer portion forming an intramolecular duplex with a part of the target, i.e., a hairpin loop structure. This structure may be immobilised to the solid support at any point on the molecule. Other conditions necessary for carrying out the polymerase reaction, including temperature, pH, buffer compositions etc., will be apparent to those skilled in the art.

[0111] The modified nucleotides of the invention are then brought into contact with the target polynucleotide, to allow polymerisation to occur. The nucleotides may be added sequentially, i.e., separate addition of each nucleotide type (A, T, G or C), or added together. If they are added together, it is preferable for each nucleotide type to be labelled with a different label.

[0112] This polymerisation step is allowed to proceed for a time sufficient to allow incorporation of a nucleotide.

[0113] Nucleotides that are not incorporated are then removed, for example, by subjecting the array to a washing step, and detection of the incorporated labels may then be carried out.

[0114] Detection may be by conventional means, for example if the label is a fluorescent moiety, detection of an incorporated base may be carried out by using a confocal scanning microscope to scan the surface of the array with a laser, to image a fluorophore bound directly to the incorporated base. Alternatively, a sensitive 2-D detector, such as a charge-coupled detector (CCD), can be used to visualise the individual signals generated. However, other techniques such as scanning near-field optical microscopy (SNOM) are available and may be used when imaging dense arrays. For example, using SNOM, individual polynucleotides may be distinguished when separated by a distance of less than

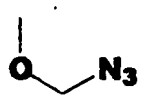
100 nm, e.g., 10 nm to 10 μm. For a description of scanning near-field optical microscopy, see Moyer et al., Laser Focus World 29:10, 1993. Suitable apparatus used for imaging polynucleotide arrays are known and the technical set-up will be apparent to the skilled person.

[0115] After detection, the label may be removed using suitable conditions that cleave the linker and the 3'OH block to allow for incorporation of further modified nucleotides of the invention. Appropriate conditions may be those described herein for allyl group and for "Z" group deprotections. These conditions can serve to deprotect both the linker (if cleavable) and the blocking group. Alternatively, the linker may be deprotected separately from the allyl group by employing methods of cleaving the linker known in the art (which do not sever the O-blocking group bond) followed by deprotection.

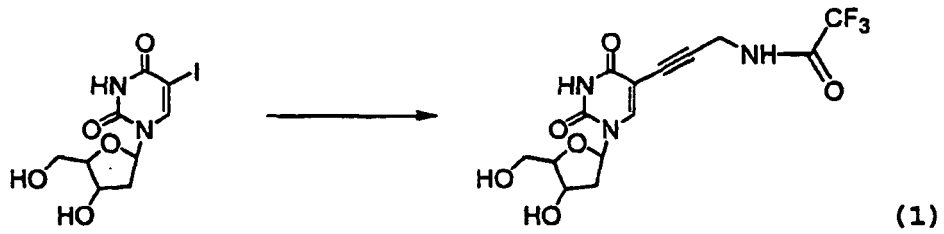
[0116] This invention may be further understood with reference to the following examples which serve to illustrate the invention and not to limit its scope.

3'-OH protected with an azidomethyl group as a protected form of a hemiaminal:

[0117]



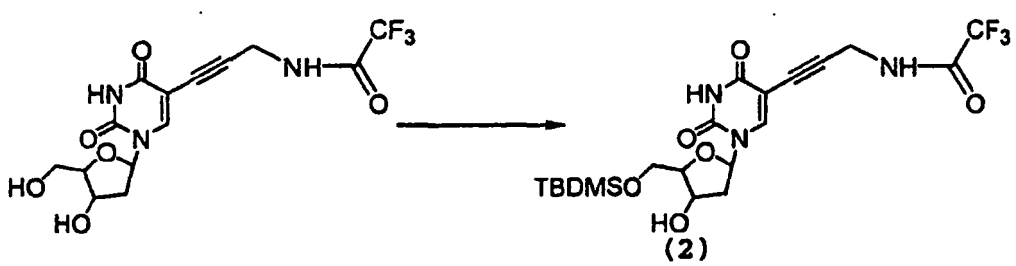
[0118] Nucleotides bearing this blocking group at the 3'position have been synthesised, shown to be successfully incorporated by DNA polymerases, block efficiently and may be subsequently removed under neutral, aqueous conditions using water soluble phosphines or thiols allowing further extension:



(1)

5-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxyuridine (1).

[0119] To a solution of 5-iodo-2'-deoxyuridine (1.05 g, 2.96 mmol) and CuI (114 mg, 0.60 mmol) in dry DMF (21 ml) was added triethylamine (0.9 ml). After stirring for 5 min trifluoro-N-prop-2-ynyl-acetamide (1.35 g, 9.0 mmol) and Pd $(PPh_3)_4$ (330 mg, 0.29 mmol) were added to the mixture and the reaction was stirred at room temperature in the dark for 16 h. Metanol (MeOH) (40 ml) and bicarbonate dowex added to the reaction mixture and stirred for 45 min. The mixture was filtered and the filtrate washed with MeOH and the solvent was removed under vacuum. The crude mixture was purified by chromatography on silica (ethyl acetate (EtOAc) to EtOAc:MeOH 95:5) to give slightly yellow crystals (794 mg, 71 %). ^{1}H NMR (d_6 dimethylsulfoxide (DMSO)) δ 2.13-2.17 (m, 2H, H-2'), 3.57-3.65 (m, 2H, H-5'), 3.81-3.84 (m, 1H, H-4'), 4.23-4.27 (m, 3H, H-3', CH_2N), 5.13 (t, J = 5.0 Hz, 1H, OH), 5.20 (d, J = 4.3 Hz, 1H, OH), 6.13 (t, J = 6.7 Hz, 1H, H-1'), 8.23 (s, 1H, H-6), 10.11 (t, J = 5.6 Hz, 1H, NH), 11.70 (br s, 1H, NH). Mass (-ve electrospray) calcd for $CuH_{14}F_3N_3O_6$ 377.08, found 376.

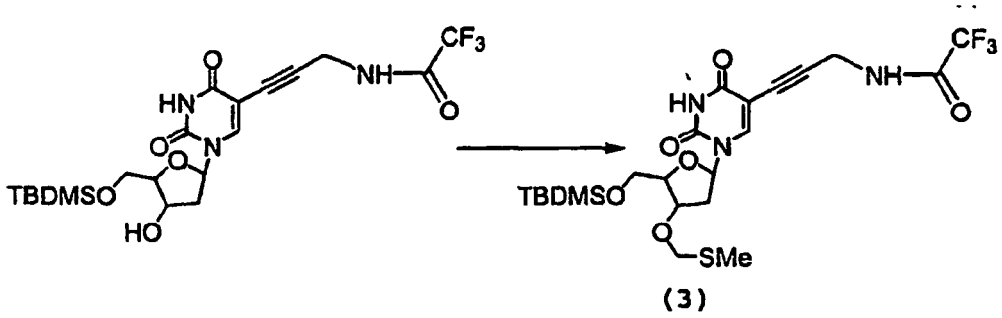


(2)

5'-O-(*tert*-butydimethylsilyl)-5-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxyuridine (2).

[0120] To a solution of (1) (656 mg, 1.74 mmol) in dry DMF (15 ml) was added t-butyldimethylsilylchloride (288 mg, 1.91 mmol) in small portions, followed by imidazole (130 mg, 1.91 mmol). The reaction was followed by TLC and was completed after stirring for 8 h at room temperature. The reaction was quenched with sat. aq. NaCl solution. EtOAc (25 ml) was added to the reaction mixture and the aqueous layer was extracted with EtOAc three times. After drying the combined organics ($MgSO_4$), the solvent was removed under vacuum. Purification by chromatography on silica (EtOAc: petroleum ether 8:2) gave (2) as slightly yellow crystals (676 mg, 83 %). ^{1}H NMR (d_6 DMSO) δ 0.00

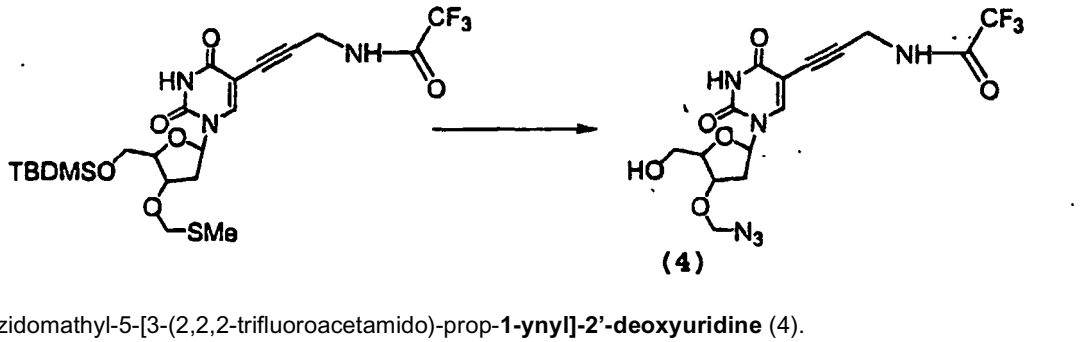
[0121] (s, 6H, CH_3) , 0.79 (s, 9H, tBu) , 1.93 -2.00 (m, 1H, H-2'), 2.06- 2.11 (m, 1H, H-2'), 3.63-3.75 (m, 2H, H-5'), 3.79-3.80 (m, 1H, H-4'), 4.12-4.14 (m, 3H, H-3', CH_2N) , 5.22 (d, *J* = 4.1 Hz, 1H, OH), 6.03 (t, *J* = 6.9 Hz, 1H, H-1'). 7.86 (s, 1H, H-6), 9.95 (t, *J* = 5.4 Hz, 1H, NH), 11.61 (br s, 1H, NH). Mass (-ve electrospray) calcd for $C_{20}H_{28}F_3N_3O_6Si$ 491.17, found 490.



(3)

5'-O-(tert-Butydimethylsilyl)-3'-O-methylthiomethyl-5-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxyuridine (3).

[0122] To a solution of (2) (1.84 g, 3.7 mmol) in dry DMSO (7 ml) was added acetic acid (3.2 ml) and acetic anhydride (10.2 ml). The mixture was stirred for 2 days at room temperature, before it was quenched with sat. aq. $NaHCO_3$ EtOAc (50 ml) was added and the aqueous layer was extracted three times with ethyl acetate. The combined organic layers were washed with sat. aq. $NaHCO_3$ solution and dried ($MgSO_4$) After removing the solvent under reduced pressure, the product (3) was purified by chromatography on silica (EtOAc: petroleum ether 8:2) yielding a clear sticky oil (1.83 g, 89 %) . ^{1}H NMR (d_6 DMSO) : δ 0.00 (s, 6H, CH_3) , 0.79 (s, 9H, tBu), 1.96-2.06 (m, 1H, H-2') , 1.99 (s, 3H, SCH_3), 2.20-2.26 (m, 1H, H-2'), 3.63-3.74 (m, 2H, H-5'), 3.92-3.95 (m, 1H, H-4'), 4.11-4.13 (m, 2H, CH_2), 4.28-4.30 (m, 1H, H-3'), 4.59 (br s, 2H, CH_2), 5.97 (t, *J* = 6.9 Hz, 1H, H-1') , 7.85 (s, 1H, H-6), 9.95 (t, *J* = 5.3 Hz, 1H, NH), 11.64 (s, 1H, NH). Mass (-ve electrospray) calcd for $C_{22}H_{32}F_3N_3O_6SSi$ 551.17, found 550.

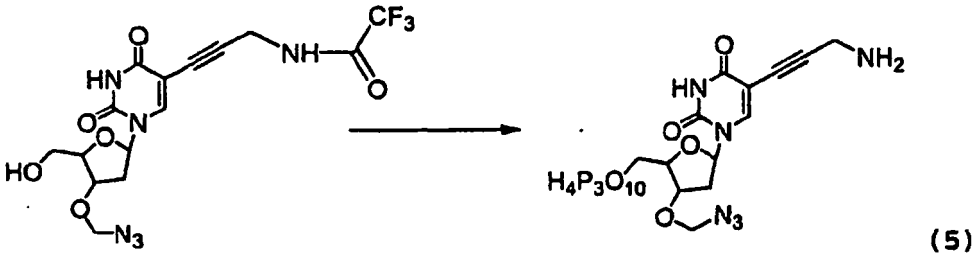


(4)

3'-O-Azidomathyl-5-[3-(2,2,2-trifluoroacetamido)-prop-**1-ynyl]-2'-deoxyuridine** (4).

[0123] To a solution of (3) (348 mg, 0.63 mmol) and cyclohexene (0.32 ml, 3.2 mmol) in dry CH_2Cl_2 (5 ml) at 4 °C, sulfurylchoride (1M in CH_2Cl_2, 0.76 ml, 0.76 mmol) was added drop wise under N_2. After 10 min TLC indicated the full consumption of the nucleoside (3). The solvent was evaporated and the residue was subjected to high vacuum for 20 min. It was then redissolved in dry DMF (3 ml) and treated with NaN_3 (205 mg, 3.15 mmol). The resulting suspension was stirred under room temperature for 2h. The reaction was quenched with CH_2Cl_2 and the organic layers were washed with sat aq. NaCl solution. After removing the solvent, the resulting yellow gum was redissolved in THF (2 ml) and treated with TBAF (1 M in THF, 0.5 ml) at room temperature for 30 min. The solvent was removed and the reaction worked up with CH_2Cl_2 and sat. aq. $NaHCO_3$ solution. The aqueous layer was extracted three times with CH_2Cl_2. Purification by chromatography on silica (EtOAc: petroleum ether 1:1 to EtOAc) gave (4) (100 mg, 37 %) as a pale yellow foam. ^{1}H

NMR (d_6 DMSO) δ 2.15-2.26 (m, 2H, H-2'), 3.47-3.57 (m, 2H, H-5'), 3.88-3.90 (m, 1H, H-4'), 4.14 (d, *J* = 4.7 Hz, 2H, CH_2NH) , 4.24-4.27 (m, 1H, H-3'), 4.75 (s, 2H, CH_2N_3), 5.14 (t, *J* = 5.2 Hz, 1H, OH), 5.96-6.00 (m, 1H, H-1'), 8.10 (s, 1H, H-6), 10.00 (s, 1H, $NHCOCF_3$)), 11.26 (s, 1H, NH).



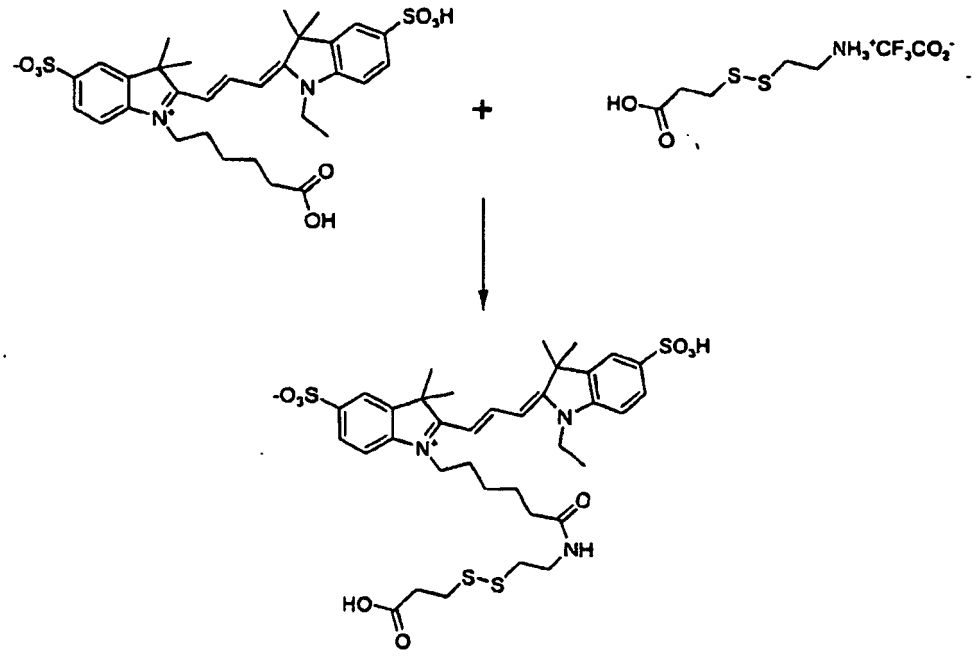
(5)

Preparation of bis (tri-n-butylammonium) pyrophosphate (0.5 M solution in DMF)

[0124] Tetrasodium diphosphate decahydrate (1.5 g, 3.4 mmol) was dissolved in water (34 ml) and the solution was applied to a column of dowex in the H^+ form. The column was eluted with water. The eluent dropped directly into a cooled (ice bath) and stirred solution of tri-n-butylamine (1.6 ml, 6.8 mmol) in EtOH (14 ml). The column was washed until the pH of the eluent increased to 6. The aq. ethanol solution was evaporated to dryness and then co-evaporated twice with ethanol and twice with anhydrous DMF. The residue was dissolved in DMF (6.7 ml). The pale yellow solution was stored over 4Å molecular sieves.

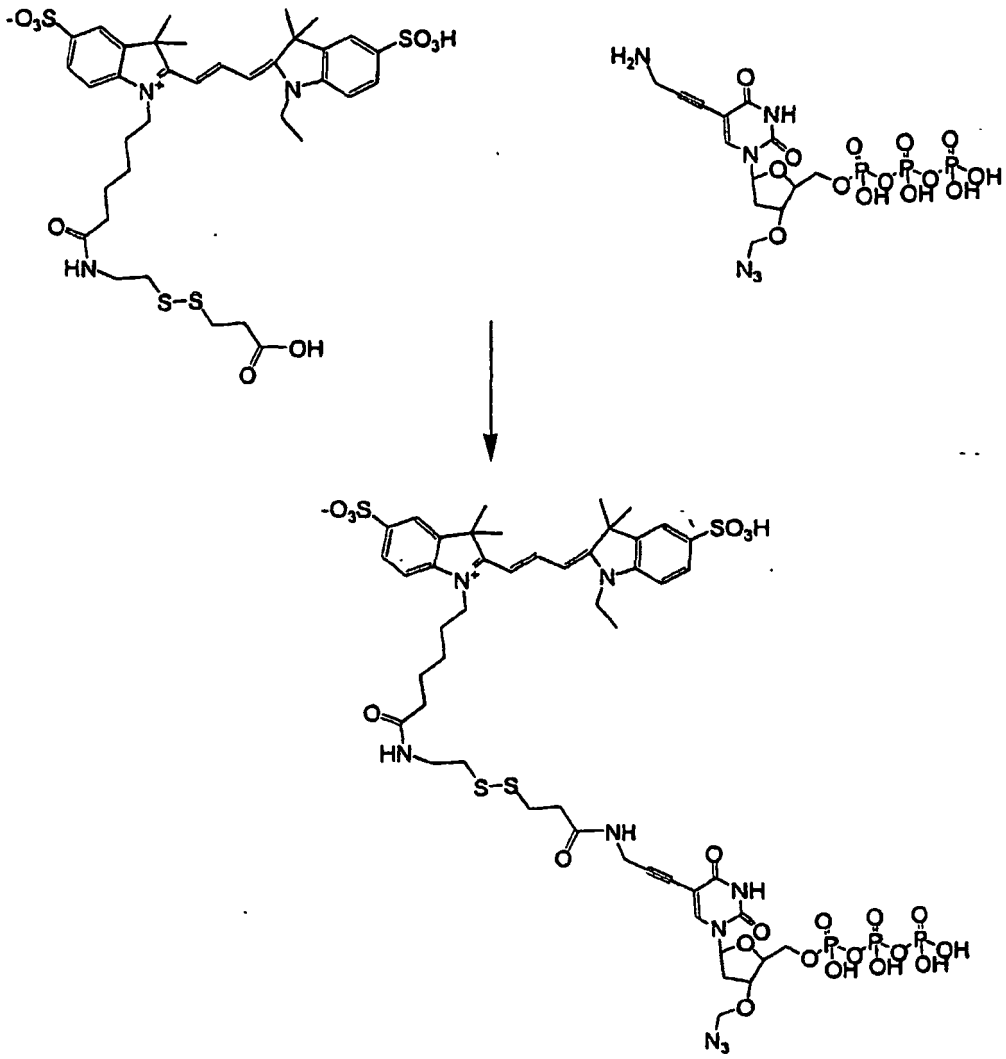
3'-O-Azidomethyl-5-(3-amino-prop-1-ynyl)-2'-deoxyuridine 5'-O-nucleoside triphosphate (5).

[0125] The nucleoside (4) and proton sponge was dried over P_2O_5 under vacuum overnight. A solution of (4) (92 mg, 0.21 mmol) and proton sponge (90 mg, 0.42 mmol) in trimethylphosphate (0.5 ml) was stirred with 4Å molecular sieves for 1 h. Freshly distilled $POCl_3$ (24 μl, 0.26 mmol) was added and the solution was stirred at 4°C for 2h. The mixture was slowly warmed up to room temperature and bis (tri-n-butyl ammonium) pyrophosphate (1.7 ml, 0.85 mmol) and anhydrous tri-n-butyl amine (0.4 ml, 1.7 mmol) was added. After 3 min, the reaction was quenched with 0.1 M TEAB (triethylammonium bicarbonate) buffer (15 ml) and stirred for 3h. The water was removed under reduced pressure and the resulting residue dissolved in concentrated ammonia (p 0.88, 15 ml) and stirred at room temperature for 16 h. The reaction mixture was then evaporated to dryness. The residue was dissolved in water and the solution applied to a DEAE-Sephadex A-25 column. MPLC was performed with a linear gradient of TEAB. The triphosphate was eluted between 0.7 M and 0.8 M buffer. Fractions containing the product were combined and evaporated to dryness. The residue was dissolved in water and further purified by HPLC. HPLC: t_r(5): 18.8 min (Zorbax C18 preparative column, gradient: 5% to 35% B in 30 min, buffer A 0.1M TEAB, buffer B MeCN) The product was isolated as a white foam (76 O.D., 7.6 μmol, 3.8%, ε_{280} = 10000). ^{1}H NMR (D_2O) δ 1.79 (s, CH_2), 2.23-2.30; 2.44-2.50 (2 x m, 2H, H-2'), 3.85 (m, CH_2NH) , 4.10-4.18 (m, 2H, H-5'), 4.27 (br s, H-4'), 4.48-4.50 (m, H-3'), 4.70-4.77 (m, CH_2N_3) , 6.21 (t, *J* = 6.6 Hz, H-1') , 8.32 (s, 1H, H-6). ^{31}P NMR (D_2O) δ -6.6 (m, 1P, P_γ) , -10.3 (d, *J* =18.4 Hz, 1P, P_α), -21.1 (m, 1P, P_β) . Mass (-ve electrospray) calcd for $C_{13}H_{19}N_6O_{14}P_3$ 576.02, found 575.



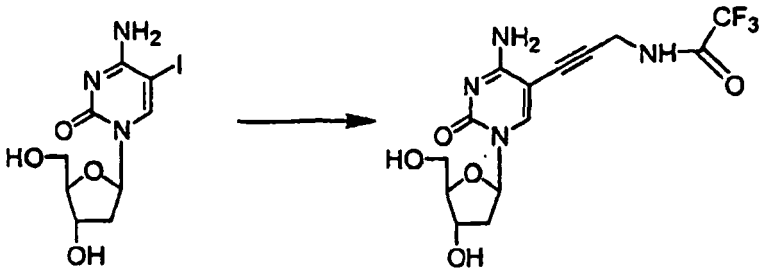
Cy-3disulfide linker.

[0126] The starting disulfide (4.0 mg, 13.1 μmol) was dissolved in DMF (300 μL) and diisopropylethylamine (4 μL) was slowly added. The mixture was stirred at room temperature and a solution of Cy-3 dye (5 mg, 6.53 μmol) in DMF (300 μL) was added over 10 min. After 3.5 h, on complete reaction, the volatiles were evaporated under reduced pressure and the crude residue was HPLC purified on a Zorbax analytical column SB-C18 with a flow rate of 1ml/min in 0.1M triethylammonium bicarbonate buffer (buffer A) and CH_3CN (buffer B) using the following gradient:,5 min 2% B;.31min 55% B; 33 min 95% B;.37 min 95%;.39 min 2% B;.44 min. 2% B. The expected Cy3-disulfide linker was eluted with a t_r: 21.8 min. in 70% yield (based on a UV measurement; ε_{550} 150,000 cm^{-1} M^{-1} in H_2O) as a hygroscopic solid. ^1H NMR (D_2O) δ 1.31-1.20 (m + t, *J* = 7.2 Hz, 5H, CH_2 + CH_3) , 1.56-1.47 (m, 2H, CH_2), 1.67 (s, 12H, 4 CH_3) , 1.79-1.74 (m, 2H, CH_2), 2.11 (t, *J* = 6.9 Hz, 2H, CH_2), 2.37 (t, *J* = 6.9 Hz, 2H, CH_2), 2.60 (t, *J* = 6.3 Hz, 2H, CH_2) , 2.67 (t, *J* = 6.9 Hz, 2H, CH_2), 3.27 (t, *J* = 6.1 Hz, 2H, CH_2), 4.10-4.00 (m, 4H, $2CH_2$), 6.29 (dd, *J* =13.1, 8.1 Hz, 2H, 2 =CH), 7.29 (dd, 2H, *J* = 8.4, 6.1 Hz, 2 =CH), 7.75-7.71 (m, 2H, 2 =CH), 7.78 (s, 2H, =CH), 8.42 (t, *J* =12.8 Hz, 1H, =CH). Mass (-ve electrospray) calcd for $C_{36}H_{47}N_3O_9S_4$ 793.22, found 792 (M-H), 396 [M/2] .



(6)

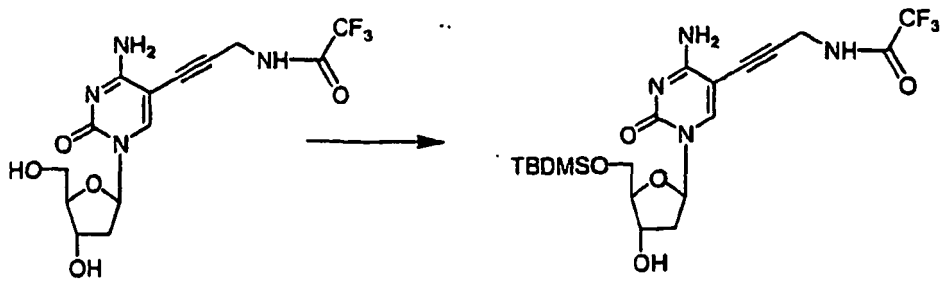
[0127] A mixture of Cy3 disulphide linker (2.5 μmol), disuccinimidyl carbonate (0.96 mg, 3.75 μmol) and DMAP (0.46 mg, 3.75 μmol) were dissolved in dry DMF (0.5 ml) and stirred at room temperature for 10 min. The reaction was monitored by TLC (MeOH:CH_2Cl_2 3:7) until all the dye linker was consumed. Then a solution of (5) (7.5 μmol) and *n*-BU_3N (30 μl, 125 μmol) in DMF (0.2 ml) was added to the reaction mixture and stirred at room temperature for 1 h. TLC (MeOH:CH_2Cl_2 4:6) showed complete consumption of the activated ester and a dark red spot appeared on the baseline. The reaction was quenched with TEAB buffer (0.1M, 10 ml) and loaded on a DEAE Sephadex column (2 x 5 cm). The column was first eluted with 0.1 M TEAB buffer (100 ml) to wash off organic residues and then 1 M TEAB buffer (100 ml). The desired triphosphate analogue (6) was eluted out with 1 M TEAB buffer. The fraction containing the product were combined, evaporated and purified by HPLC. HPLC conditions: t_r(6): 16.1 min (Zorbax C18 preparative column, gradient: 2% to 55% B in 30 min, buffer A 0.1M TEAB, buffer B MeCN). The product was isolated as dark red solid (1.35 μmol, 54%, ε_{550} = 150000). ^{1}H NMR (D_2O) 8 1.17-1.28 (m, 6H 3 × CH_2), 1.41-1.48 (m, 3 H, CH_3), 1.64 (s, 12H, 4 x CH_3), 1.68-1.71 (m, 2H, CH_2), 2.07-2.10 (m, 3H, H-2', CH_2), 2.31-2.35 (m, 1H, H-2'), 2.50-2.54 (m, 2H, CH_2), 2.65 (t, J = 5.9 Hz, 2H, CH_2), 2.76 (t, *J* = 7.0 Hz, 2H, CH_2), 3.26-3.31 (m, 2H, CH_2), 3.88-3.91 (m, 2H CH_2), 3.94-4.06 (m, 3H, CH_2N, H-5'), 4.16 (br s, 1H, H-4'), 4.42-4.43 (m, 1H, H-3'), 4.72-4.78 (m, 2H, CH_2N_3), 6.24 (dd, J = 5.8, 8.2 H_2, H-1') , 6.25 (dd, *J* = 3.5, 8.5 Hz, 2H, H_{Ar}) , 7.24, 7.25 (2d, J = 14.8 Hz, 2 x =CH), 7.69-7.86 (m, 4H, H_{Ar}, H-6), 8.42 (t, J =13.4 Hz, =CH) . ^{31}P NMR (D_2O) δ -4.85 (m, 1P, P_γ), -9.86 (m, 1P, P_α), -20.40 (m, 1P, P_β) . Mass (-ve electrospray) calcd for $C_{49}H_{64}N_9O_{22}P_3S_4$ 1351.23, found 1372 (M-2H+Na), 1270 [M-80] , 1190 [M-160] .



(7)

5-[3-(2,2,2-Trifluoroacetamido)-prop-1-ynyl]-2'-deoxycytidine (7).

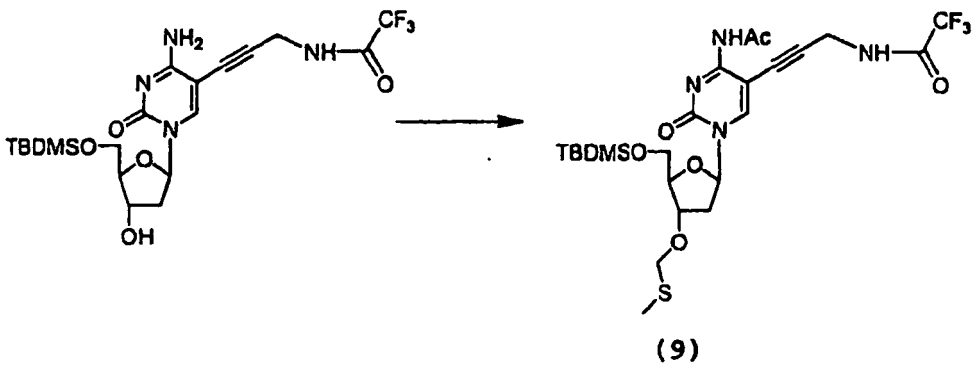
[0128] To a solution of 5-iodo-2'-deoxycytidine (10 g, 28.32 mmol) in DMF (200 ml) in a light protected round bottom flask under Argon atmosphere, was added CuI (1.08 g, 5.67 mmol), triethylamine (7.80 ml, 55.60 mmol), 2,2,2-trifluoro-*N*-prop-2-ynyl-acetamide (12.8 g, 84.76 mmol) and at last $Pd(PPh_3)_4$ (3.27 g, 2.83 mmol). After 18 hours at room temperature, dowex bicarbonate (20 mg) was added and the mixture was stirred for a further 1 h. Filtration and evaporation of the volatiles under reduced pressure gave a residue that was purified by flash chromatography on silica gel (CH_2Cl_2, CH_2Cl_2:EtOAc 1:1, EtOAc:MeOH 9:1).The expected product (7) was obtained as a beige solid in quantitative yield. ^{1}H NMR (D_2O) δ 2.24-2.17 (m, 1H, H-2'), 2.41-2.37 (m, 1H, H-2'), 3.68 (dd, *J* = 12.5, 5.0 Hz, 1H, H-5'), 3.77 (dd, *J* = 12.5, 3.2 Hz, 1H, H-5'), 3.99 (m, 1H, H-4'), 4.27 (s, 2H, CH_2N), 4.34 (m, 1H, H-3'), 6.11 (t, *J* = 6.3 Hz, 1H, H-1'), 8.1 (br s, 1H, NH); MS (ES) : m/z (%) (M-H) 375 (100).



(8)

5'-*O*-(*tert*-Butyldimethylsilyl)-5-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxycytidine (8).

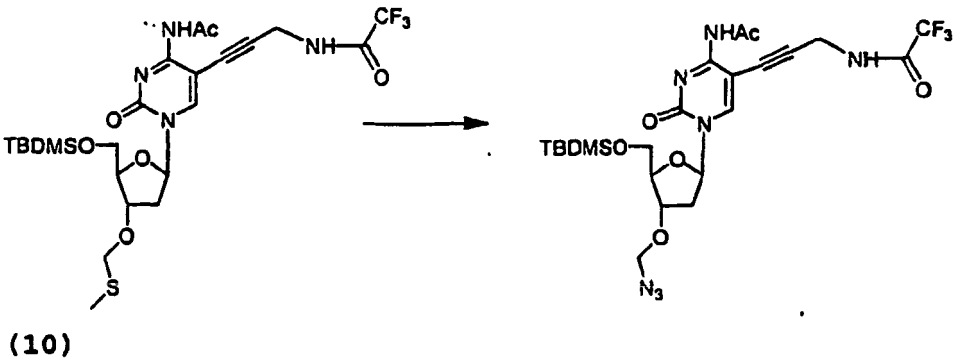
[0129] To a solution of the starting material (7) (1.0 g, 2.66 mmol) and imidazole (200 mg, 2.93 mmol) in DMF (3.0 ml) at 0 °C, was slowly added TBDMSCl (442 mg, 2.93 mmol) in four portions over 1 h. After 2 h, the volatiles were evaporated under reduced pressure and the residue was adsorbed on silica gel and purified by flash chromatography (EtOAc, EtOAc:MeOH 9.5:0.5). The expected product (8) was isolated as a crystalline solid (826 mg, 64%). ^{1}H NMR (d_6 DMSO) δ 0.00 (s, 1H, CH_3); 0.01 (s, 1H, CH_3), 0.79 (s, 9 H, tBu), 1.87-1.80 (m, 1H, H-2'), 2.12 (ddd, *J* = 13.0, 5.8 and 3.0 Hz, 1H, H-2'), 3.65 (dd, *J* = 11.5, 2.9 Hz, 1H, H-5'), 3.74 (dd, *J* = 11.5, 2.5 Hz, 1H, H-5'), 3.81-3.80 (m, 1H, H-4'), 4.10-4.09 (m, 1H, H-3'), 4.17 (d, 2H, *J* = 5.1 Hz, NCH_2), 5.19 (d, 1H, *J* = 4.0 Hz, 3'-OH), 6.04 (t, *J* = 6.6 Hz, 1H, H-1'), 6.83 (br s, 1H, NH*H*), 7.78 (br s, 1H, N*H*H), 7.90 (s, 1H, H-6), 9.86 (t, *J* = 5.1 Hz, 1H, $-H_2CNH$); MS (ES) : m/z (%) $(MH)^+$ 491 (40%).



(9)

4-*N*-Acetyl-5'-*O*-(*tert*-butyldimethylsilyl)-3'-*O*-(methylthiolmethyl)-5-[3-(2,2,2-trifluoroacatamide)-prop-1-ynyl]-2'-deoxycytidine (9).

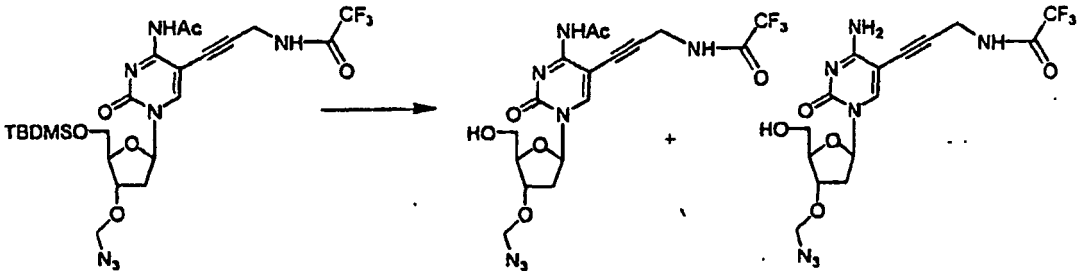
[0130] To a solution of the starting material (8) (825 mg, 1.68 mmol) in DMSO (6.3 ml) and N_2 atmosphere, was slowly added acetic acid (AcOH) (1.3 ml, 23.60 mmol) followed by acetic anhydride (AC_2O) (4.8 ml, 50.50 mmol). The solution was stirred at room temperature for 18 h and quenched at 0 °C by addition of saturated $NaHCO_3$ (20 ml). The product was extracted into EtOAc (3 x 30 ml), organic extracts combined, dried ($MgSO_4$), filtered and the volatiles evaporated. The crude residue was purified by flash chromatography on silica gel (EtOAc: petroleum ether 1:1) to give the expected product as a colourless oil (9) (573 mg, 62%). ^1H NMR (d_6 DMSO) δ 0.00 (s, 6H, 2 x CH_3), 0.78 (s, 9H, tBu), 2.01 (s, 3H, SCH_3), 2.19-1.97 (m, 2H, 2 x H2'), 2.25 (s, 3H, $COCH_3$), 3.67 (dd, 1H, *J* = 11.5 Hz, H-5'), 3.78 (dd, 1H, *J* = 11.5, 3.3 Hz, H-5'), 4.06-4.05 (m, 1H, H-4'), 4.17 (d, 2H, *J* = 5.1 Hz, N-C*H*2), 4.30-4.28 (m, 1H, H-3'), 4.63 (s, 2H; CH_2-S), 5.94 (t, 1H, *J* = 6.5 Hz, H-1'), 8.17 (s, 1H, H-6), 9.32 (s, 1H, NHCO), 9.91 (t, 1H, *J* = 5.4 Hz, N*H*CH_2); MS (ES): m/z (%) $(MH)^+$ 593.



(10)

4-*N*-Acetyl-3'-*O*-(azidomethyl)-5'-*O*-(*tert*-butyldimethyleilyl)-5-[3-(2,2,2-trifluoroacetamide)-prop-1-ynyl]-2'-deoxycytidine (10).

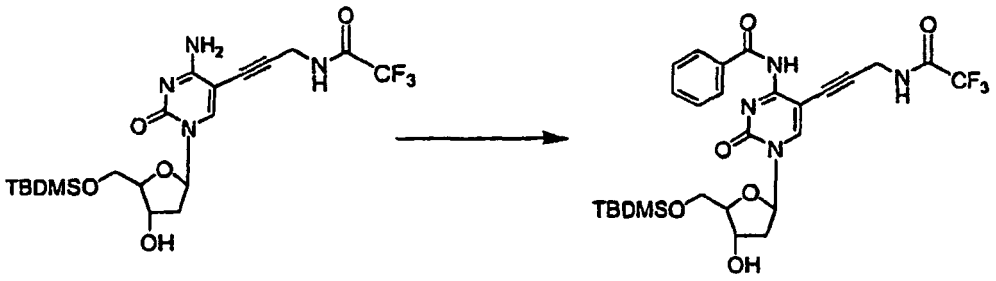
[0131] To a solution of the starting material (9) (470 mg, 0.85 mmol) in dicloromethane (DCM) (8 ml) under N_2 atmosphere and cooled to 0 °C, was added cyclohexene (430 μl, 4.27 mmol) followed by SO_2Cl_2 (1 M in DCM, 1.0 ml, 1.02 mmol). The solution was stirred for 30 minutes at 0 °C, and the volatiles were evaporated. Residue immediately dissolved in DMF (8 ml) stirred under N_2 and sodium azide (275 mg, 4.27 mmol) slowly added. After 18 h, the crude product was evaporated to dryness, dissolved in EtOAc (30 ml) and washed with Na_2CO_3 (3 x 5 ml). The combined organic layer was kept separately. A second extraction of the product from the aqueous layer was performed with DCM (3 x 10 ml). All the combined organic layers were dried ($MgSO_4$), filtered and the volatiles evaporated under reduced pressure to give an oil identified as the expected product (10) (471 mg, 94 % yield). This was used without any further purification. ^1H NMR (d_6 DMSO) δ 0.11 (s, 3H, CH_3), 0.11 (s, 3H, CH_3), 0.88 (s, 9H, t_{Bu}), 2.16-2.25 (m, 1H, H-2'), 2.35 (s, 3H, $COCH_3$), 2.47-2.58 (m, 1H, H-2'), 3.79 (dd, *J* = 11.6, 3.2 Hz, 1H, H-5'), 3.90 (dd, *J* = 11.6, 3.0 Hz, 1H, H-5'), 4.17-4.19 (m, 1H, H-4'), 4.28 (s, 2H, NCH_2), 4.32-4.35 (m, 1H, H-3'), 4.89 (dd, *J* = 14.4, 6.0 Hz, 2H, CH_2-N_3), 6.05 (t, *J* = 6.4 Hz, 1H, H-1'), 8.25 (s, 1H, H-6), 9.46 (br s, 1H, NH*H*), 10.01 (br s, 1H, N*H*H).



(11) (12)

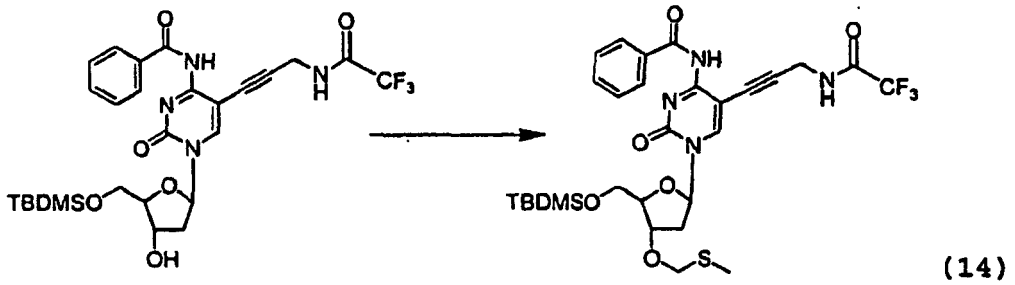
4-*N*-Acetyl-3'-*O*-(azidomethyl)-5-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxycytidine and 3'-*O*-(Azidomethyl)-5-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxycytidine (11).

[0132] To a solution of the starting material (11) (440 mg, 0.75 mmol) in THF (20 ml) at 0 °C and N_2 atmosphere, was added TBAF in THF 1.0 M (0.82 ml, 0.82 mmol). After 1.5 h, the volatiles were evaporated under reduced pressure and the residue purified by flash chromatography on silica gel (EtOAc:petroleum ether 8:2 to EtOAc 100 % to EtOAc:MeOH 8:2). Two compounds were isolated and identified as above described. The first eluted 4-N-Acetyl (11), (53 mg, 15 %) and, the second one 4-NH_2 (12) (271 mg, 84 %). Compound *4-N-Acetyl* (11): ^{1}H NMR (d_6 DMSO) δ 1.98 (s, 3H, CH_3CO), 2.14-2.20 (m, 2H, H*H*-2'), 3.48-3.55 (m, 1H, H-5'), 3.57-3.63 (m, 1H, H-5'), 3.96-4.00 (m, 1H, H-4'), 4.19 (d, *J* = 5.3 Hz, 2H, CH_2-NH), 4.23-4.28 (m, 1H, H-3'), 4.77 (s, 2H, CH_2-N_3), 5.2 (t, 1H, *J* = 5.1 Hz, 5'-OH), 5.95 (t, *J* = 6.2 Hz, 1H, H-1'), 8.43 (s, 1H, H-6), 9.34 (s, 1H, CON*H*), 9.95 (t, *J* = 5.3 Hz, 1H, $NHCH_2$).
Compound 4-NH_2 (12): ^{1}H NMR (d_6 DMSO) δ 1.98-2.07(2H, C*HH*-2'), 3.50-3.63 (m, 2H, C*HH*-5'), 3.96-4.00 (m, 1H, H-4'), 4.09 (d, *J* = 5.3 Hz, 2H, CH_2-NH), 4.24-4.28 (m, 1H, H-3'), 4.76 (s, 2H, CH_2-N_3), 5.13 (t, *J* = 5.3 Hz, 1H, 5'-OH), 5.91 (br s, 1H, NHH), 6.11 (t, *J* = 6.4 Hz, 1H, H-1'), 8.20 (t, *J* = 5.3 Hz, 1H, NCH_2), 8.45 (s, 1H, H-6), 11.04 (br s, 1H, NRH).



4-*N*-Henzoyl-5'-*O*-(*tert*-butyldimethylsilyl)-5-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxycytidine (13).

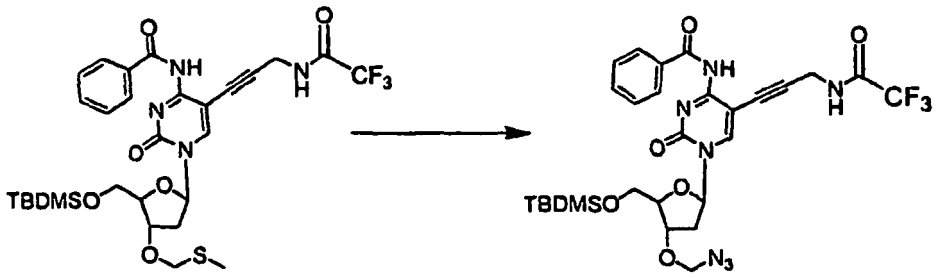
[0133] The starting material (8) (10g, 20.43 mmol) was azeotroped in dry pyridine (2 x 100 ml) then dissolved in dry pyridine (160 ml) under N_2 atmosphere. Chlorotrimethylsilane (10 ml, 79.07 mmol) added drop wise to the solution and stirred for 2 hours at room temperature. Benzoyl chloride (2.6 ml, 22.40 mmol) was then added to solution and stirred for one further hour. The reaction mixture was cooled to 0°C, distilled water (50 ml) added slowly to the solution and stirred for 30 minutes. Pyridine and water were evaporated from mixture under high vacuum to yield a brown gel that was portioned between 100 ml of sat. aq. $NaHCO_3$ (100 ml) solution DCM. The organic phase was separated and the aqueous phase extracted with a further (2 x 100 ml) of DCM. The organic layers were combined, dried ($MgSO_4$), filtered and the volatiles evaporated under reduced pressure. The resulting brown oil was purified by flash chromatography on silica gel (DCM:MeOH 99:1 to 95:5) to yield a light yellow crystalline solid (13) (8.92 g, 74%). ^{1}H NMR (d_6 DMSO): δ 0.00 (s, 6H, CH_3), 0.78 (s, 9H, tBu), 1.94 (m, 1H, H-2'), 2.27 (m, 1H, H-2'), 3.64 (d, 1H, *J* = 11.6 Hz, H-5'), 3.75 (d, 1H, *J* = 11.6 Hz, H-5'), 3.91 (m, 1H, H-4'), 4.09 (br m, 3H, CH_2NH, H-3'), 5.24 (s, 1H, 3'-OH), 6.00 (m, 1H, H-1'), 7.39 (m, 2H, Ph), 7.52 (m, 2H, Ph), 7.86 (m, 1H, Ph), 8.0 (s, 1H, H-6), 9.79 (t, 1H, *J* = 5.4 Hz, N*H*CH_2), 12.67 (br s, 1H, NH). Mass (+ve electrospray) calcd for $C_{27}H_{33}F_3N_4O_6Si$ 594.67, found 595.



4-*N*-Benzoyl-5'-*O*-(*tert*-butyldimethylsilyl)-3'-*O*-methylthiomethyl-5-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxycytidine (14).

[0134] The starting material (13) (2.85 g, 4.79 mmol) was dissolved in dry DMSO (40 ml) under N_2 atmosphere. Acetic

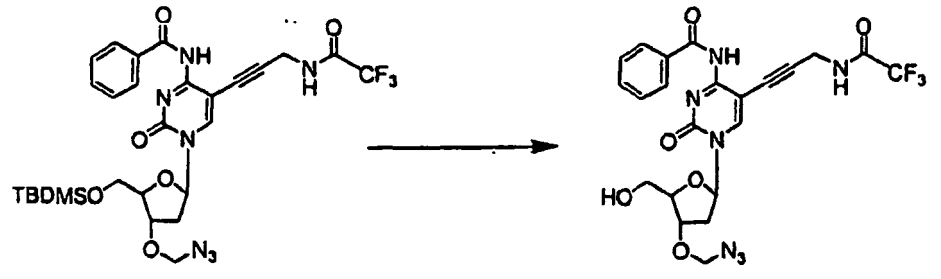
acid (2.7 ml, 47.9 mmol) and acetic anhydride (14.4 ml, 143.7 mmol) were added sequentially and slowly to the starting material, which was then stirred for 18 h at room temperature. Saturated $NaHCO_3$ (150 ml) solution was carefully added to the reaction mixture. The aqueous layer was extracted with EtOAc (3 x 150 ml). The organic layers were combined, dried ($MgSO_4$), filtered and evaporated to yield an orange liquid that was subsequently azeotroped with toluene (4 x 150 ml) until material solidified. Crude residue purified on silica gel (petroleum ether:EtOAc 3:1 to 2:1) to yield a yellow crystalline solid (14) (1.58 g, 50%). ^{1}H NMR (d_6 DMSO): δ 0.00 (s, 6H, CH_3), 0.78 (s, 9H, tBu), 1.99 (s, 3H, CH_3), 2.09 (m, 1H, H-2'), 2.28 (m, 1H, H-2'), 3.66 (d, 1H, *J* = 11.5, 2.9 Hz, H-5'), 3.74 (dd, 1H, *J* = 11.3, 2.9 Hz, H-5'), 3.99 (m, 1H, H-4'), 4.09 (m, 1H, CH_2NH), 4.29 (m, 1H, H-3'), 4.61 (s, 2H, CH_2S), 6.00 (m, 1H, H-1'), 7.37 (m, 2H, Ph), 7.50 (m, 2H, Ph), 7.80 (d, 1H, *J* = 7.55 Hz, H_{Ar}), 7.97 (s, 1H, H-6), 9.79 (br t, 1H, N*H*CH_2), 12.64 (br s, 1H, NH). Mass (-ve electrospray) calcd for $C_{29}H_{37}F_3N_4O_6SSi$ 654.79, found 653.2.



(15)

4-*N*-Benzoyl-5'-*O*-(*tert*-butyldimethylsilyl)-3'-*O*-azidomethyl-5-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxycytidine (15).

[0135] The starting material **(14)** (1.65 g, 2.99 mmol) was dissolved in DCM (18 ml) and cooled to 0°C. Cyclohexene (1.5 ml, 14.95 mmol) and SO_2Cl_2 (0.72 ml, 8.97 mmol) were added and stirred 1 h in ice bath. TLC indicated starting material still to be present whereupon a further aliquot of SO_2Cl_2 (0.24 ml) was added and the mixture stirred for 1 h at 0°C. Volatiles were removed by evaporation to yield a light brown solid that was redissolved in 18 ml of dry DMF (18 ml) under N_2. Sodium azide (0.97 g, 14.95 mmol) was then added to the solution and stirred for 2.5 h at room temperature. The reaction mixture was passed through a pad of silica and eluted with EtOAc and the volatiles removed by high vacuum evaporation. The resulting brown gel was purified by flash chromatography (petroleum ether:EtOAc 4:1 to 2:1) to yield the desired product as a white crystalline solid (15) (0.9 g, 55%). ^{1}H NMR (d_6 DMSO): δ 0.00 (s, 6H, CH_3), 0.78 (s, 9H, tBu), 2.16 (m, 1H, H-2'), 2.22 (m, 1H, H-2'), 3.70 (d, 1H, *J* = 11.5 Hz, H-5'), 3.75 (d, 1H, *J* = 11.3 Hz, H-5'), 4.01 (m, 1H, H-4'), 4.10 (m, 1H, CH_2NH), 4.23 (m, 1H, H-3'), 4.76 (s, 2H, CH_2S), 5.99 (m, 1H, H-1'), 7.37 (m, 2H, Ph), 7.50 (m, 2H, Ph), 7.81 (d, 1H, *J* = 7.4 Hz, Ph), 7.95 (s, 1H, H-6), 9.78 (br s, 1H, N*H*CH_2), 12.64 (br s, 1H, NH). Mass (-ve electrospray) calcd for $C_{28}H_{34}F_3N_7O_6Si$ 649.71, found 648.2

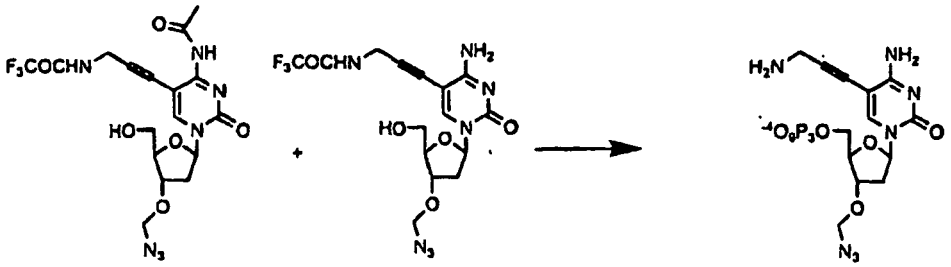


(16)

4-*N*-Benzoyl-3'-*O*-azidomethyl-5-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxycytidine (16).

[0136] The starting material (15) (140 mg, 0.22 mmol) was dissolved in THF (7.5 ml). TBAF (1M soln. in THF, 0.25 ml) was added slowly and stirred for 2 h at room temperature. Volatile material removed under reduced pressure to yield a brown gel that was purified by flash chromatography (EtOAc:DCM 7:3) to yield the desired product (16) as a light coloured crystalline solid (0.9 g, 76%). ^{1}H NMR (d_6 DMSO): δ 2.16 (m, 1H, H-2'), 2.22 (m, 1H, H-2'), 3.70 (d, 1H, *J* =

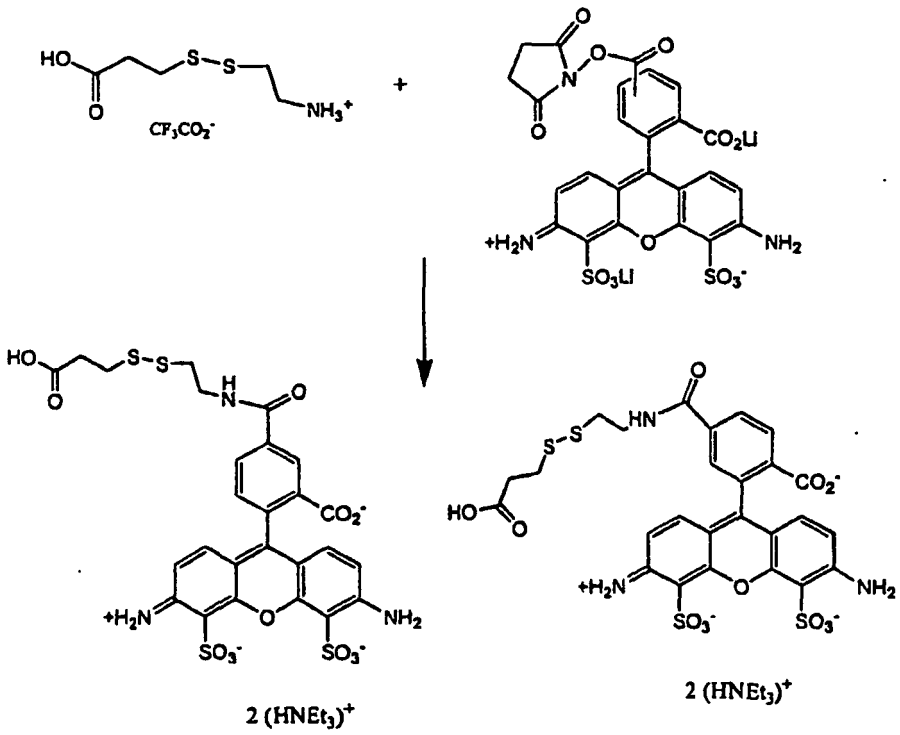
11.5 Hz, H-5'), 3.75 (d, 1H, *J* = 11.3 Hz, H-5'), 4.01 (m, 1H, H-4'), 4.10 (m, 1H, CH_2NH), 4.23 (m, 1H, H-3'), 4.76 (s, 2H, CH_2S), 5.32 (s, 1H, 5' OH), 5.99 (m, 1H, H-1'), 7.37 (m, 2H, Ph), 7.50 (m, 2H, Ph), 7.81 (d, 1H, *J* = 7.35 Hz, Ph), 7.95 (s, 1H, H-6), 9.78 (br s, 1H, N*H*CH_2), 12.64 (br s, 1H, NH). Mass (-ve electrospray) calcd for $C_{22}H_{20}F_3N_7O_6$ 535.44, found 534.



(17)

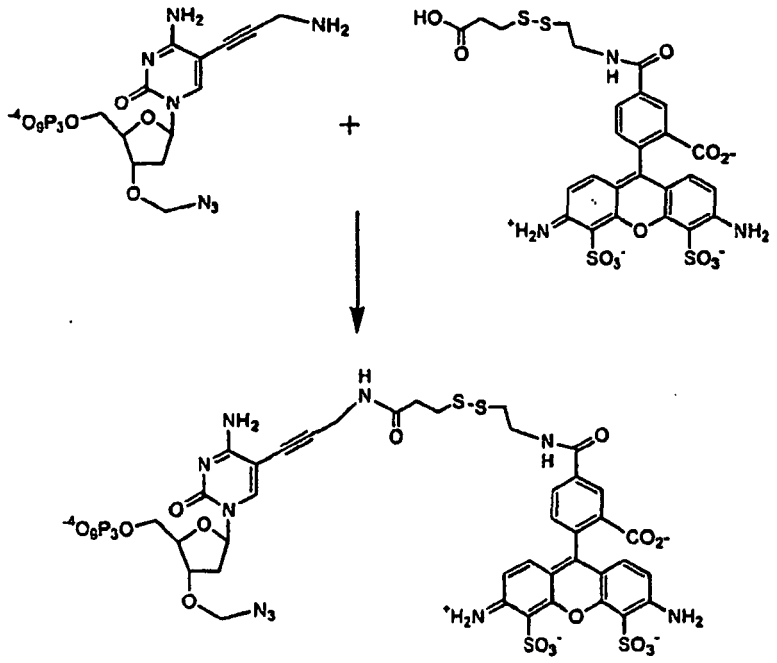
5-(3-Amino-prop-1-ynyl)-3'-*O*-azidomethyl-2'-deoxycytidine 5'-*O*-nucleoside triphosphate (17).

[0137] To a solution of (11) and (12) (290 mg, 0.67 mmol) and proton sponge (175 mg, 0.82 mmol) (both previously dried under P_2O_5 for at least 24 h) in $PO(OMe)_3$ (600 μl), at 0 °C under Argon atmosphere, was slowly added $POCl_3$ (freshly distilled) (82 μl, 0.88 mmol). The solution was vigorously stirred for 3 h at 0 °C and then quenched by addition of tetra-tributylammonium diphosphate (0.5 M) in DMF (5.2 ml, 2.60 mmol), followed by *n*Bu_3N (1.23 ml, 5.20 mmol) and triethylammonium bicarbonate (TEAB) 0.1 M (20 ml). After 1 h at room temperature aqueous ammonia solution (p 0.88, 20 ml) was added to the mixture. Solution stirred at room temperature for 15 h, volatiles evaporated under reduced pressure and the residue was purified by MPLC with a gradient of TEAB from 0.05M to 0.7M. The expected triphosphate was eluted from the column at approx. 0.60 M TEAB. A second purification was done by HPLC in a Zorbax SB-C18 column (21.2 mm i.d. x 25 cm) eluted with 0.1M TEAB (pump A) and 30% CH_3CN in 0.1M TEAB (pump B) using a gradient as follows: 0-5 min 5% B, Φ.2 ml; 5-25 min 80% B, Φ.8 ml; 25-27 min 95 %B, Φ.8 ml; 27-30 min 95 %B, Φ.8 ml; 30-32 min 5 %B, Φ.8 ml; 32-35 min 95 %B, Φ.2 ml, affording the product described above with a r_t(17): 20.8 (14.5 μmols, 2.5% yield); ^{31}p NMR (D_2O, 162 MHz) δ -5.59 (d, *J* = 20.1 Hz, P_x), -10.25 (d, *J* = 19.3 Hz, 1P, P_α), -20.96 (t, *J* = 19.5 Hz, 1P, P_β); ^{1}H NMR (D_2O) δ 2.47-2.54 (m, 1H, H-2'), 2.20-2.27 (m, 1H, H-2'), 3.88 (s, 2H, CH_2N), 4.04-4.12 (m, 1H, RH-5'), 4.16-4.22 (m, 1H, H*H*-5'), 4.24-4.30 (m, 1H, H-4'), 4.44-4.48 (m, 1H, H-3'), 6.13 (t, *J* = 6.3 Hz, 1H, H-1'), 8.35 (s, 1H, H-6); MS (ES): m/z (%) (M-H) 574 (73%), 494 (100 %).



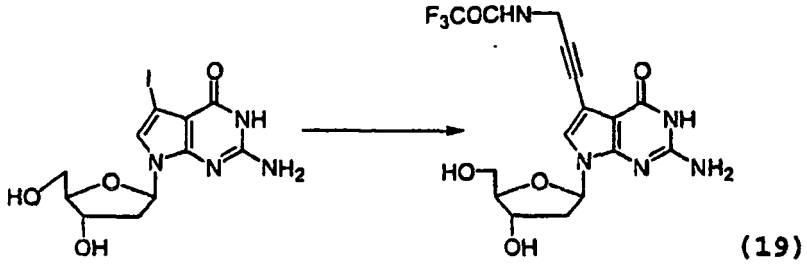
Alexa488 disulfide linker.

[0138] Commercial available Alexa Fluor 488-NHS (35 mg, 54 μmol) was dissolved in DMF (700 μL) and, to ensure full activation, 4-DMAP (7 mg, 59 μmol) and *N*,*N*'-disuccinimidyl carbonate (15 mg, 59 μmol) were sequentially added. After 15 min on complete activation, a solution of the starting disulfide (32.0 mg, 108 μmol) in DMF (300 μL) containing diisopropylethylamine (4 μL) was added over the solution of the activated dye. Further addition of diisopropylethylamine (20 μL) to the final mixture was done, ultrasonicated for 5 min and reacted for 18 h at room temperature in the darkness. The volatiles were evaporated under reduced pressure and the crude residue was first purified passing it through a short ion exchange resin Sephadex -DEAE A-25 (40-120 μ) column, first eluted with TEAB 0.1 M (25 ml) then 1.0 M TEAB (75 ml). The latest containing the two final compounds was concentrated and the residue was HPLC purified in a Zorbax SB-C18 column (21.2 mm i.d. x 25 cm) eluted with 0.1M TEAB (pump A) and CH_3CN (pump B) using a gradient as follows: 0-2 min 2% B, Φ.2 ml; 2-4 min 2% B, Φ.8 ml; 4-15 min 23 %B, Φ.8 ml; 15-24 min 23 %B, Φ.8 ml; 24-26 min 95 %B, Φ.8 ml; 26-28 min 95 %B, Φ.8 ml, 28-30 min 2 %B, Φ.8 ml, 30-33 min 2 %B, Φ.2 ml affording both compounds detailed above with t_r: 19.0 (left regioisomer) and t_r: 19.5 (right regioisomer). Both regioisomers were respectively passed through a dowex ion exchange resin column, affording respectively 16.2 μmol and 10.0 μmol, 62% total yield (based in commercial available Alexa Fluor 488-NHS of 76% purity); ε_{493} = 71,000 cm^{-1} M^{-1} in H_2O. ^1H NMR (D_2O) (left regioisomer) δ 2.51 (t, *J* = 6.8 Hz, 2H, CH_2), 2.66 (t, *J* = 6.8 Hz, 2H, CH_2), 2.71 (t, *J* = 5.8 Hz, 2H, CH_2), 3.43 (t, *J* = 5.8 Hz, 2H, CH_2), 6.64 (d, *J* = 9.2 Hz, 2H, H_{Ar}), 6.77 (d, *J* = 9.2 Hz, 2H, H_{Ar}), 7.46 (s, 1H, H_{Ar}), 7.90 (dd, *J* = 8.1 and 1.5 Hz, 1H, H_{Ar}), 8.20 (d, *J* = 8.1 Hz, 1H, H_{Ar}). ^1H NMR (D_2O) (right regioisomer) δ 2.67 (t, *J* = 6.8 Hz, 2H, CH_2), 2.82 (t, *J* = 6.8 Hz, 2H, CH_2), 2. 93 (t, *J* = 6.1 Hz, 2H, CH_2), 3.68 (t, *J* = 6.1 Hz, 2H, CH_2), 6.72 (d, *J* = 9.3 Hz, 2H, H_{Ar}), 6.90 (d, *J* = 9.3 Hz, 2H, H_{Ar}), 7.32 (d, *J* = 7.9 Hz, 1H, H_{Ar}), 8.03 (dd, *J* = 7.9, 1.7 Hz, 1H, H_{Ar}), 8.50 (d, *J* = 1.8 Hz, 1H, H_{Ar}) Mass (-ve. electrospray) calcd for $C_{26}H_{23}N_3O_{12}S_4$ 697.02, found 692 (M-H), 347 [M/2].



(18)

[0139] To a solution of Alexa Fluor 488 disulfide linker (3.4 μmol, 2.37mg) in DMF (200 μL) was added 4-DMAP (0.75 mg, 5.1 μmol) and *N,N*-disuccinimidyl carbonate (1.70 mg, 5.1 μmol). The mixture was stirred for 15 to full activation of the acid, then it was added into the solution of the nucleotide **(17)** (3.45 mg, 6.0 μmol) in DMF (0.3 ml) containing nBu$_3$N (40 μL) at 0 °C. The mixture was sonicated for 3 min and then continuously stirred for 16 h in the absence of light. The volatiles were evaporated under reduced pressure and the residue was firstly purified by filtration through a short ion exchange resin Sephadex -DEAE A-25 column, first eluted with TEAB 0.1 M (50 ml) removing the unreacted dye-linker, then 1.0 M TEAB (100 ml) to collect the expected product **(18).** After concentration and the residue was HPLC purified in a Zorbax SB-C18 column (21.2 mm i.d. x 25 cm) eluted with 0.1M TEAB (pump A) and CH_3CN (pump B) using a gradient as follows: 0-2 min 2% B, Φ.2 ml; 2-4 min 2% B, Φ.8 ml; 4-15 min 23 %B, Φ.8 ml; 15-24 min 23 %B, Φ.8 ml; 24-26 min 95 %B, Φ.8 ml; 26-28 min 95 %B, Φ.8 ml, 28-30 min 2 %B, Φ.8 ml, 30-33 min 2 %B, Φ.2 ml affording the product detailed above with a r_t**(18):** 19.8 (0.26 μmols, 12% yield based on UV measurement); λ_{max} = 493 nm, ∈ 71,000 cm^{-1} M^{-1} in H_2O); ^{31}P NMR (D_2O, 162 MHz) δ -5.06 (d, J = 20.6 Hz, 1P, P_x), -10.25 (d, J = 19.3 Hz, 1P, P_α), -21.21 (t, J = 19.5 Hz, 1P, P_β); ^{1}H NMR (D_2O) δ 2.09-2.17 (m, 1H, *H*H-2'), 2.43-2.50 (m, 1H, H*H*-2'), 2.61 (t, J = 6.8 Hz, 2H, H_2C-S), 2.83 (2H, S-CH_2), 3.68 (t, J = 6.0 Hz, 2H, ArCONCH_2), 4.06 (s, 2H, CH_2N), 4.08-4.17 (m, 4H, *HH*-5'), 4.25-4.29 (m, 1H, H-4'), 4.46-4.50 (m, 1H, H-3'), 6.09 (t, J = 6.4 Hz, 1H, H-1'), 6.88 (d, J = 9.1 Hz, 1H, H_{Ar}), 6.89 (d, J = 9.3 Hz, 1H, H_{Ar}), 7.15 (d, J = 9.3 Hz, 1H, H_{Ar}), 7.17 (d, J = 9.1 Hz, 1H, H_{Ar}), 7.64 (br s, 1H, H_{Ar}), 8.00-7.94 (m, 2H, H_{Ar}), 8.04 (s, 1H, H-6); MS (ES): m/z (%) (M-H) 1253 (46%), (M-H+ Na)$^-$ 1275 (100 %).

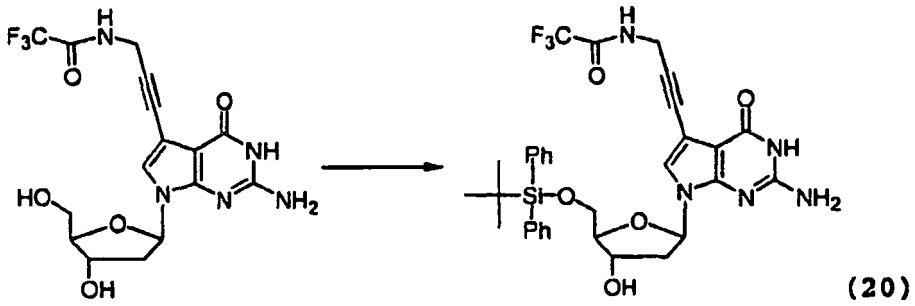


(19)

7-Deaza-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxyguanosine (19).

[0140] Under N_2, a suspension of 7-deaza-7-iodo-guanosine (2 g, 2.75 mmol), $Pd(PPh_3)_4$ (582 mg, 0.55 mmol), CuI (210 mg, 1.1 mmol), Et_3N (1.52 ml, 11 mmol) and the propagylamine (2.5 g, 16.5 mmol) in DMF (40 ml) was stirred at room temperature for 15 h under N_2. The reaction was protected from light with aluminium foil. After TLC indicating the full consumption of starting material, the reaction mixture was concentrated. The residue was diluted with MeOH (20

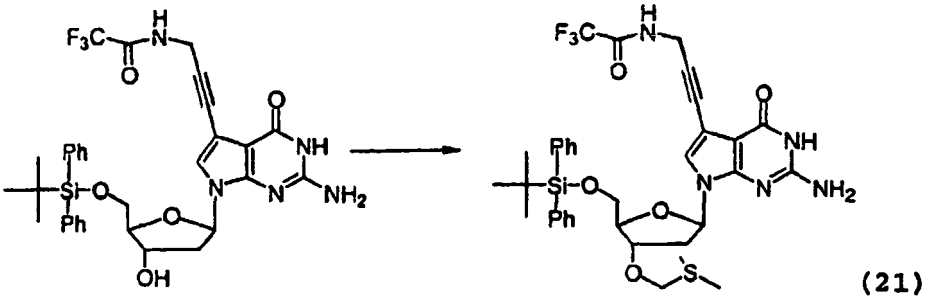
ml) and treated with dowex-HCO_3^-. The mixture was stirring for 30 min and filtered. The solution was concentrated and purified by silica gel chromatography (petroleum ether:EtOAc 50:50 to petroleum ether: EtOAc:MeOH 40: 40: 20), giving (19) as a yellow powder (2.1 g, 92%). ^{1}H NMR (d_6 DMSO) δ 2.07-2.11 (m, 1H, H-2'), 2.31-2.33 (m, 1H, H-2'), 3.49-3.53 (m, 2H, H-5'), 3.77 (br s, 1H, H-4'), 4.25 (d, J = 4.3 Hz, 2H, $\equiv CCH_2$), 4.30 (br s, 1H, H-3'), 4.95 (t, J = 5.2 Hz, 1H, 5'-OH), 5.25 (d, J = 3.4 Hz, 1H, 3'-OH), 6.27-6.31 (m, 1H, H-1'), 6.37 (s, 2H, NH_2), 7.31 (s, 1H, H-8), 10.10 (br s, 1H, $NHCOCF_3$), 10.55 (s, 1H, NH). Mass (-ve electrospray) calcd for $C_{16}H_{16}F_3N_5O_5$ 415, found 414.



(20)

5'-*O*-(*tert*-Butyldiphenyl)-7-deaza-7-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxyguanosine (20).

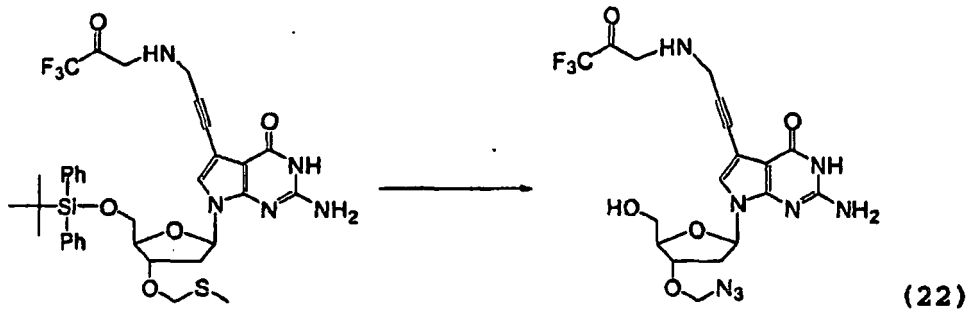
[0141] A solution of (19) (2.4 g, 5.8 mmol) in pyridine (50 ml) was treated with tert-butyldiphenylsilyl chloride (TBDPSCl) (1.65 ml, 6.3 mmol) drop wise at 0 °C. The reaction mixture was then warmed to room temperature. After 4h, another portion of TBDPSCl (260 μL, 1 mmol) was added. The reaction was monitored by TLC, until full consumption of the starting material. The reaction was quenched with MeOH (-5 ml) and evaporated to dryness. The residue was dissolved in DCM and aq. sat. $NaHCO_3$ was added. The aqueous layer was extracted with DCM three times. The combined organic extracts were dried ($MgSO_4$) and concentrated under vacuum. Purification by chromatography on silica (EtOAc to EtOAc: MeOH 85:15) gave (20) a yellow foam (3.1 g, 82%). ^{1}H NMR (d_6 DMSO) δ 1.07 (s, 9H, CH_3), 2.19-2.23 (m, 1H, H-2'), 2.38-2.43 (m, 1H, H-2'), 3.73-3.93 (m, 2H, H-5'), 4.29 (d, J = 5.0 Hz, 2H, CH_2N), 4.42-4.43 (m, 1H, H-3'), 5.41 (br s, 1H, OH), 6.37 (t, J = 6.5 Hz, H-1'), 6.45 (br s, 2H, NH_2), 7.24-7.71 (m, 11H, H-8, H_{Ar}), 10.12 (t, J = 3.6 Hz, 1H, NH), 10.62 (s, 1H, H-3). Mass (+ve electrospray) calcd for $C_{32}H_{34}F_3N_5O_5Si$ 653, found 654.



(21)

5'-*O*-(*tert*-Butyldiphenyl)-7-deaza-3'-*O*-methylthiolmethyl-7-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxy-guanosine (21).

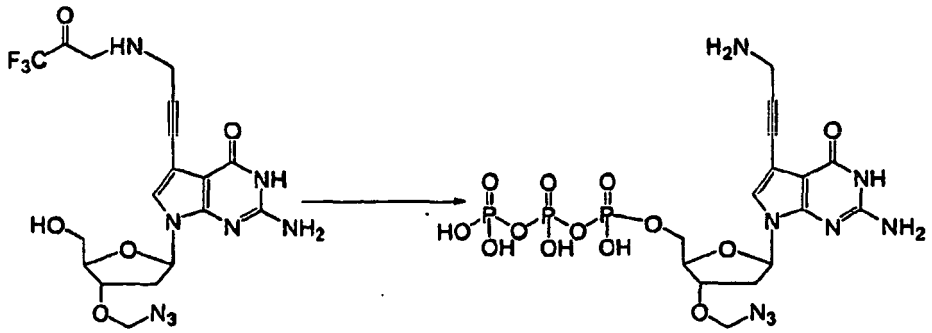
[0142] A solution of **(20)** (1.97 g, 3.0 mmol) in DMSO (15 ml) was treated with Ac_2O (8.5 ml, 90 mmol), and AcOH (2.4 ml, 42 mmol) and stirred at room temperature for 15 h, then 2 h at 40°C. The reaction mixture was diluted with EtOAc (200 ml) and stirred with sat, aq. $NaHCO_3$ (200 ml) for 1 h. The aqueous layer was washed with EtOAc twice. The organic layer was combined, dried ($MgSO_4$) and concentrated under vacuum. Purification by chromatography on silica (EtOAc:Hexane 1:1 to EtOAc:Hexane:MeOH 10:10:1) gave **(21)** as a yellow foam (1.3 g, 60%). ^{1}H NMR ($CDCl_3$) δ 1.04 (s, 9H, CH_3), 2.08 (s, 3H, SCH_3), 2.19-2.35 (m, 2H, H-2), 3.67-3.71 (m, 2H, H-5'), 3.97-3.99 (m, 2H, H-4', H-3'), 4.23 (br s, 2H, CH_2N), 4.58 (s, 2H; CH_2S), 6.31 (dd, J = 5.7, 7.9 Hz, H-1'), 7.19-7.62 (m, 11H, H8, H_{Ar}). Mass (+ve electrospray) calcd for $C_{34}H_{38}F_3N_5O_5SSi$ 713, found: 714.



(22)

3'-*O*-Azidomethyl-7-deaza-7-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxyguanosine (22).

[0143] To a solution of **(21)** (1.3 mg, 1.8 mmol), cyclohexene (0.91 ml, 9 mmol) in CH_2Cl_2 (10 ml) in 4 °C, sulfurylchloride (1M in CH_2Cl_2) (1.1 ml, 1.1 mmol) was added drop wise under N_2. After 30 min., TLC indicated the full consumption of the nucleoside **(22).** After evaporation to remove the solvent, the residue was then subjected to high vacuum for 20 min, and then treated with NaN_3 (585 mmol, 9 mmol) and DMF (10 ml). The resulted suspension was stirred under room temperature for 2h. Extraction with CH_2Cl_2/NaCl (10%) gave a yellow gum, which was treated with TBAF in THF (1 M, 3 ml) and THF (3 ml) at room temperature for 20 min. Evaporation to remove solvents, extraction with EtOAc/sat. aq. $NaHCO_3$, followed by purification by chromatography on silica (EtOAc to EtOAc:MeOH 9:1) gave (22) as a yellow foam (420 mg, 50%). ^{1}H NMR (d_6 DMSO): δ 2.36-2.42 (m, 1H, H-2'), 2.49-2.55 (m, 1H, H-2'), 3.57-3.59 (m, 2H, H-5'), 3.97-4.00 (m, 1H, H-4'), 4.29 (m, 2H, CH_2N), 4.46-4.48 (m, 1H, H-3'), 4.92-4.96 (m, 2H, CH_2N_3), 5.14 (t, J = 5.4 Hz, 1H, 5'-OH), 5.96-6.00 (dd, J = 5.7, 8.7 Hz, 1H, H-1'), 6.46 (br s, 2H, NH_2), 7.39 (s, 1H, H-6), 10.14 (s, 1H, NH), 10.63 (s, 1H, H-3).

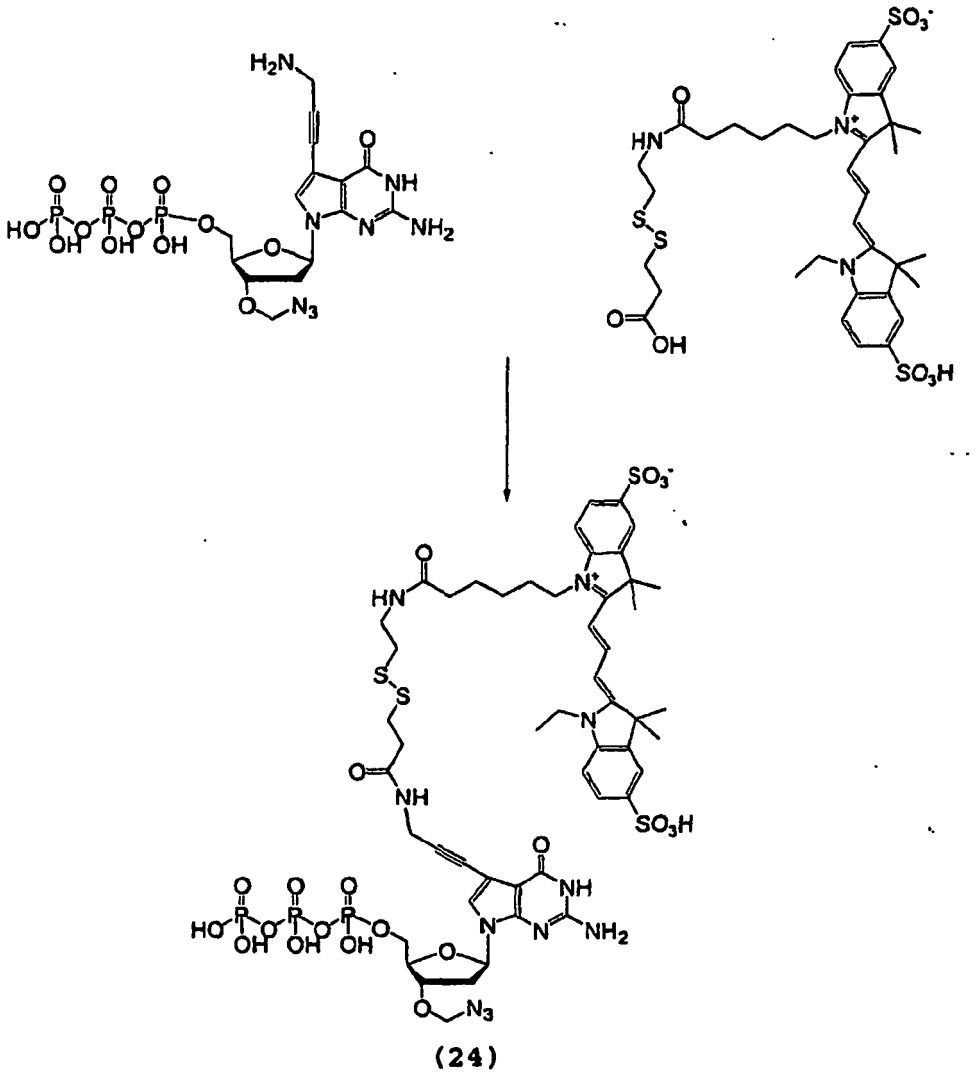


(23)

3'-*O*-Azidomethyl-7-deaza-7-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxyguanosine 5'-*O*-nucleoside triphosphate (23).

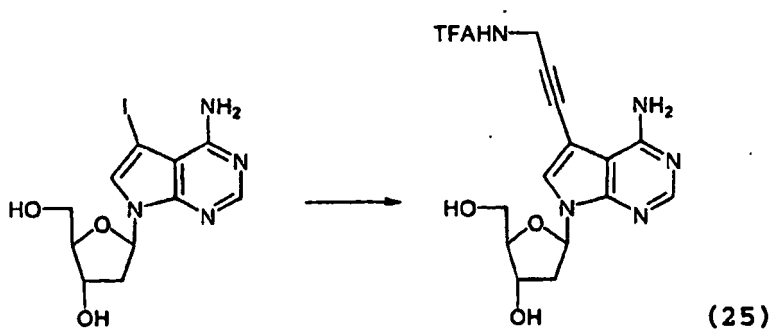
[0144] Tetrasodium diphosphate decahydrate (1.5 g, 3.4 mmol) was dissolved in water (34 ml) and the solution was applied to a column of dowex 50 in the H^+ form. The column was washed with water. The eluent dropped directly into a cooled (ice bath) and stirred solution of tri-n-butyl amine (1.6 ml, 6.8 mmol) in EtOH (14 ml). The column was washed until the pH of the eluent increased to 6. The aqueous ethanol solution was evaporated to dryness and then co-evaporated twice with ethanol and twice with anhydrous DMF. The residue was dissolved in DMF (6.7 ml). The pale yellow solution was stored over 4Å molecular sieves. The nucleoside **(22)** and proton sponge was dried over P_2O_5 under vacuum overnight. A solution of **(22)** (104 mg, 0.22 mmol) and proton sponge (71 mg, 0.33 mmol) in trimethylphosphate (0.4 ml) was stirred with 4Å molecular sieves for 1 h. Freshly distilled $POCl_3$ (25μ l, 0.26 mmol) was added and the solution was stirred at 4°C for 2h. The mixture was slowly warmed up to room temperature and bis (tri-*n*-butyl ammonium) pyrophosphate (1.76 ml, 0.88 mmol) and anhydrous tri-n-butyl amine (0.42 ml, 1.76 mmol) were added. After 5 min, the reaction was quenched with 0.1 M TEAB (triethylammonium bicarbonate) buffer (15 ml) and stirred for 3 h. The water was removed under reduced pressure and the resulting residue dissolved in concentrated ammonia (p 0.88, 10 ml) and stirred at room temperature for 16 h. The reaction mixture was then evaporated to dryness. The residue was dissolved in water and the solution applied to a DEAE-Sephadex A-25 column. MPLC was performed with a linear gradient of 2 L each of 0.05 M and 1 M TEAB. The triphosphate was eluted between 0.7 M and 0.8 M buffer. Fractions containing the product were combined and evaporated to dryness. The residue was dissolved in water and further purified by HPLC. t_r(23) = 20.5

min (Zorbax C18 preparative column, gradient: 5% to 35% B in 30 min, buffer A 0.1M TEAB, buffer B MeCN). The product was isolated as a white foam (225 O.D., 29.6 μmol, 13.4%, ε_{260} = 7,600). ^{1}H NMR (D_2O) δ 2.43-2.5 (m, 2H, H-2'), 3.85 (m, 2H, CH_2N), 3.97-4.07 (m, 2H, H-5'), 4.25 (br s, 1H, H-4'), 4.57 (br s, 1H, H-3'), 4.74-4.78 (m, 2H, CH_2N_3), 6.26-6.29 (m, 1H, H-1'), 7.41 (s, 1H, H-8). ^{31}P-NMR (D_2O) δ -8.6 (m, 1P, P_γ), -10. 1 (d, *J* =19.4 Hz, 1P, P_α), -21.8 (t, *J* = 19.4 Hz, 1P, P_β). Mass (-ve electrospray) calcd for $C_{15}H_{21}N_8O_{13}P_3$ 614, found 613.



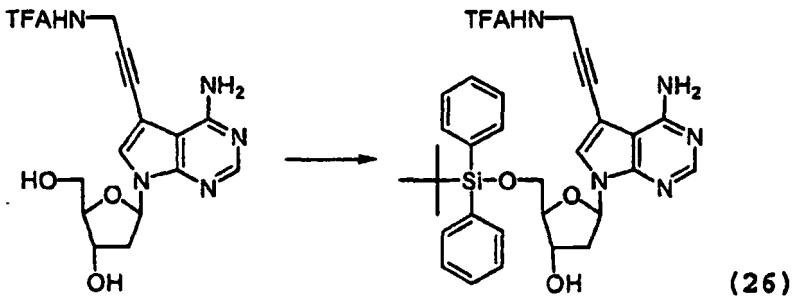
(24)

[0145] A mixture of disulphide linkered-Cy3 (2.5 μmol), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC) (0.95 mg, 5 μmol), 1-hydroxybenzotriazole (HOBt) (0.68 mg, 5 μmol) and N-methyl-morpholine (0.55 μL, 5 μmol) in DMF (0.9 ml) was stirred at room temperature for 1 h. A solution of (23) (44 O.D., 3.75 μmol) in 0.1 ml water was added to the reaction mixture at 4°C, and left at room temperature for 3 h. The reaction was quenched with TEAB buffer (0.1M, 10 ml) and loaded on a DEAE Sephadex column (2 x 5 cm). The column was first eluted with 0.1 M TEAB buffer (100 ml) and then 1 M TEAB buffer (100 ml). The desired triphosphate product was eluted out with 1 M TEAB buffer. Concentrating the fraction containing the product and applied to HPLC. t_r**(24)** = 23.8 min (Zorbax C18 preparative column, gradient: 5% to 55% B in 30 min, buffer A 0.1M TEAB, buffer B MeCN). The product was isolated as a red foam (0.5 μmol, 20%, ε_{max} = 150,000). ^{1}H NMR (D_2O) δ 1.17-1.71 (m, 20H, 4 x CH_2, 4 x CH_3), .. 2.07-2.15 (m, 1H, H-2'), 2.21-2.30 (m, 1H, H-2'), 2.52-2.58 (m, 2H, CH_2), 2.66-2.68 (m, 2H, CH_2), 2.72-2.76 (m, 2H, CH_2), 3.08-3.19 (m, 2H, CH_2), 3.81-3.93 (m, 6H, CH_2, H-5'), 4.08-4.16 (m, 1H, H-4'), 4.45-4.47 (m, 1H, H-3'), 4.70-4.79 (m, 2H, CH_2N_3), 6.05-6.08 (m, 2H, H_{Ar}), 6.15-6.18 (m, 1H, H-1'), 7.11 (s, 1H, H-8), 7.09-7.18 (m, 2H, CH), 7.63-7.72 (m, 4H, H_{Ar}), 8.27-8.29 (m, 1H, CH). ^{31}P NMR (D_2O) δ -4.7 (m, 1P, P_γ), - 9.8 (m, 1P, P_α), -19.7 (m, 1P, P_β). Mass (-ve electrospray) calcd for $C_{51}H_{66}N_{11}O_{21}P_3S_4$1389.25, found 1388 (M-H), 694 [M-2H], 462 [M-3H].



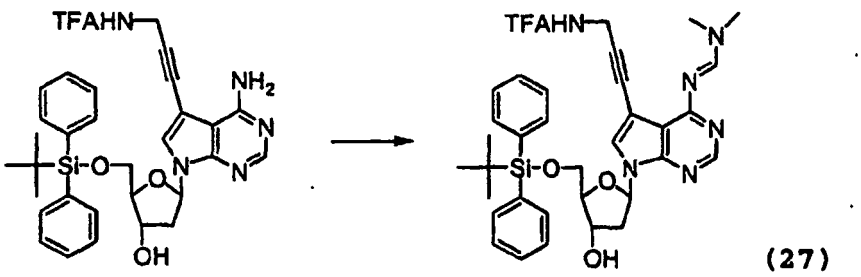
7-Deaza-7-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxyadenosine (25).

[0146] To a suspension of 7-deaza-7-iodo-2'-deoxyadenosine (1 g, 2.65 mmol) and CuI (100 mg, 0.53 mmol) in dry DMF (20 ml) was added triethylamine (740 μl, 5.3 mmol). After stirring for 5 min trifluoro-N-prop-2-ynyl-acetamide (1.2 g, 7.95 mmol) and $Pd(PPh_3)_4$ (308 mg, 0.26 mmol) were added to the mixture and the reaction was stirred at room temperature in the dark for 16 h. MeOH (40 ml) and bicarbonate dowex was added to the reaction mixture and stirred for 45 min. The mixture was filtered. The filtrate washed with MeOH and the solvent was removed under vacuum. The crude mixture was purified by chromatography on silica (EtOAc to EtOAc: MeOH 95:20) to give slightly yellow powder (25) (1.0 g, 95 %). ^{1}H NMR (d_6 DMSO) δ 2.11-2.19 (m, 1H, H-2'), 2.40-2.46 (m, 1H, H-2'), 3.44-3.58 (m, 2H, H-5'), 3.80 (m, 1H, H-4'), 4.29 (m, 3H, H-3', CH_2N), 5.07 (t, J = 5.5 Hz, 1H, OH), 5.26 (d, J = 4.0 Hz, 1H, OH), 6.45 (dd, *J* = 6.1, 8.1 Hz, 1H, H-1'), 7.74 (s, 1H, H-8), 8.09 (s, 1H, H-2), 10.09 (t, J = 5.3 Hz, 1H, NH).



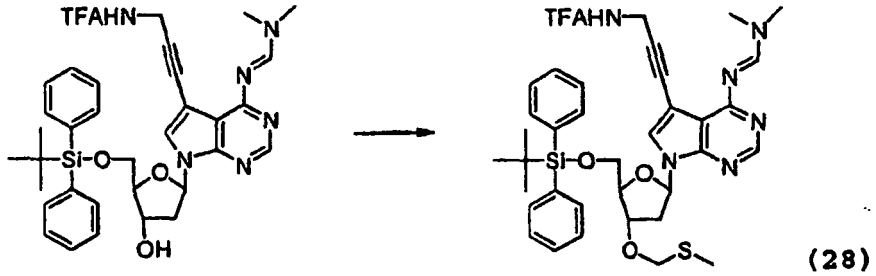
5'-*O*-(*tert*-Butyldiphenylsilyl)-7-deaza-7-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxyadenosine (26).

[0147] The nucleoside (25) (1.13 g, 2.82 mmol) was coevaporated twice in dry pyridine (2 x 10ml) and dissolved in dry pyridine (18 ml). To this solution was added t-butyldiphenylsilylchloride (748 μl, 2.87 mmol) in small portions at 0°C. The reaction mixture was let to warm up at room temperature and left stirring overnight. The reaction was quenched with sat. aq. NaCl solution. EtOAc (25 ml) was added to reaction mixture and the aqueous layer was extracted with EtOAc three times. After drying the combined organic extracts ($MgSO_4$) the solvent was removed under vacuum. Purification by chromatography on silica (DCM then EtOAc to EtOAc: MeOH 85:15) gave (26) as a slightly yellow powder (1.76 g, 97 %). ^{1}H NMR (d_6 DMSO) δ 1.03 (s, 9H, tBu), 2.25-2.32 (m, 1H, H-2'), 2.06-2.47 (m, 1H, H-2'), 3.71-3.90 (m, 2H, H-5'), 3.90-3.96 (m, 1H, H-4'), 4.32 (m, 2H, CH_2N), 4.46 (m, 1H, H-3'), 5.42 (br s, 1H, OH), 6.53 (t, *J* = 6.7 Hz, 1H, H-1'), 7.38-7.64 (m, 11H, H-8 and H_{Ar}), 8.16 (s, 1H, H-2), 10.12 (t, J = 5.3 Hz, 1H, NH).



[0148] 5'-O-(tert-Butyldiphenylsilyl)-7-deaza-4-N,N-dimethylformadin-7-[3-(2,2,2-trifluoroacetamido)-prop-1-

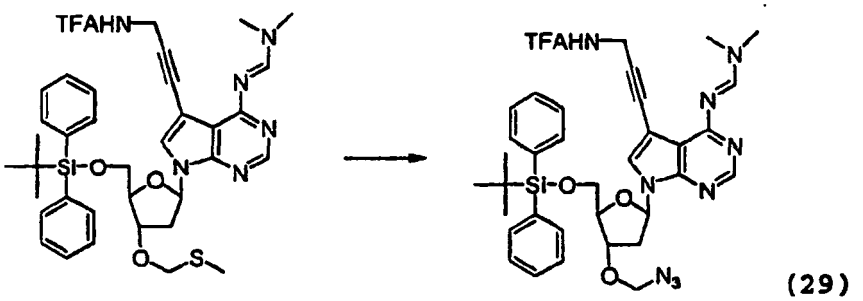
ynyl]-2'-deoxyadenosine (27). A solution of the nucleoside (26) (831 mg, 1.30 mmol) was dissolved in a mixture of MeOH:N,N-dimethylacetal (30 ml: 3ml) and stirred at 40°C. The reaction monitored by TLC, was complete after 1 h. The solvent was removed under vacuum. Purification by chromatography on silica (EtOAc: MeOH 95:5) gave (27) as a slightly brown powder (777 mg, 86 %). ^{1}H NMR (d_6 DMSO) δ 0. 99 (s, 9H, tBu), 2.22-2.29 (m, 1H, H-2'), 2.50-2.59 (m, 1H, H-2'), 3.13 (s. 3H, CH_3), 3.18 (s. 3H, CH_3), 3.68-3.87 (m, 2H, H-5'), 3.88-3.92 (m, 1H, H-4'), 4.25 (m, 2H, CH_2N), 4.43 (m, 1H, H-3'), 6.56 (t, J = 6.6 Hz, 1H, H-1'), 7.36-7.65 (m, 10H, H_{Ar}), 7.71 (s, 1H, H-8), 8.33 (s, 1H, CH), 8.8 (s, 1H, H-2), 10.12 (t, J = 5.3 Hz, 1H, NH).



(28)

5'-*O*-(*tert*-Butyldiphenylsilyl)-7-deaza-4-*N,N*-dimethylformadin-3'-*O*-methylthiomethoxy-7-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxyadenosine (28).

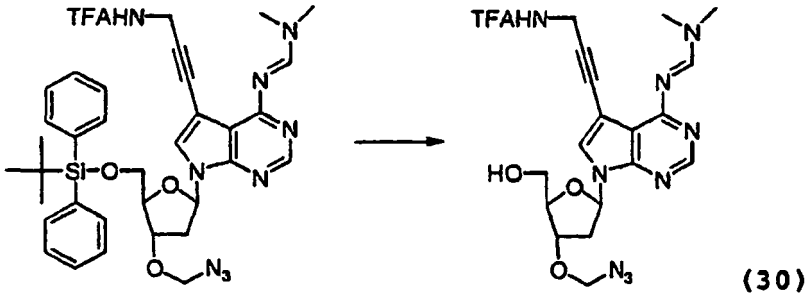
[0149] To a solution of (27) (623 mg, 0.89 mmol) in dry DMSO (8 ml) was added acetic acid (775 µl, 13.35 mmol) and acetic anhydride (2.54ml, 26.7 mmol). The mixture was stirred overnight at room temperature. The reaction was then poured into EtOAc and sat. aq. $NaHCO_3$ (1:1) solution and stirred vigorously. The organic layer was washed one more time with sat. aq. $NaHCO_3$ and dried over $MgSO_4$. After removing the solvent under reduced pressure, the product (28) was purified by chromatography on silica (EtOAc: petroleum ether 1:2, then EtOAc) yielding (28) (350 mg, 52 %). ^{1}H NMR (d_6 DMSO): δ, 1.0 (s, 9H, tBu), 2.09 (s, 3H, SCH_3), 2.41-2.48 (m, 1H, H-2'), 2.64-2.72 (m, 1H, H-2'), 3.12 (s, 3H, CH_3), 3.17 (s, 3H, CH_3), 3.66-3.89 (m, 2H, H-5'), 4.04 (m, 1H, H-4'), 4.26 (m, J = 5.6 Hz, 2H, CH_2), 4.67 (m, 1H, H-3'), 4.74 (br s, 2H, CH_2), 6.49 (t, J = 6.1, 8.1 Hz, 1H, H-1'), 7.37-7.48 (m, 5H, H_{Ar}), 7.58-7.67 (m, 5H, H_{Ar}), 7.76 (s, 1H, H-8), 8.30 (s, 1H, CH), 8.79 (s, 1H, H-2), 10.05 (t, J = 5.6 Hz, 1H, NH).



(29)

3'-*O*-Azidomethyl-5'-*O*-(*tert*-butyldiphenylsilyl)-7-deaza-4-*N,N*-dimethylformadin-7-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxyadenosine (29).

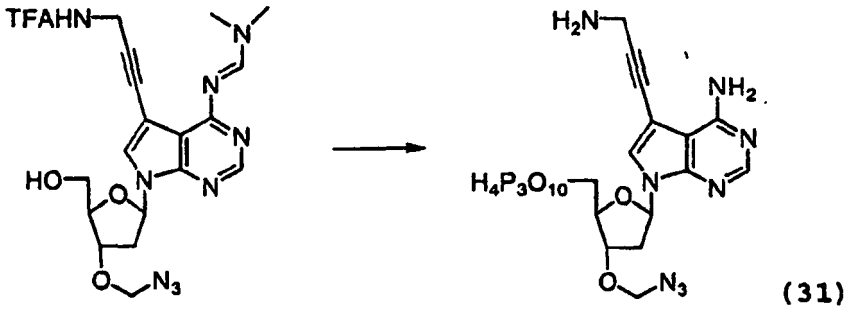
[0150] To a solution of (28) (200 mg, 0.26 mmol) and cyclohexene (0.135 ml, 1.3 mmol) in dry CH_2Cl_2 (5 ml) at 0 °C, sulfurylchoride (32 µl, 0.39 mmol) was added under N_2. After 10 min, TLC indicated the full consumption of the nucleoside (28). The solvent was evaporated and the residue was subjected to high vacuum for 20 min. It was then redissolved in dry DMF (3 ml), cooled to 0°C and treated with NaN_3 (86 mg, 1.3 mmol). The resulting suspension was stirred under room temperature for 3h. The reaction was partitioned between EtOAc and water. The aqueous phases were extracted with EtOAc. The combined organic extracts were combined and dried over $MgSO_4$. After removing the solvent under reduced pressure, the mixture was purified by chromatography on silica (EtOAc) yielding an oil (29) (155 mg, 80 %), ^{1}H NMR (d_6 DMSO): δ 0.99 (s, 9H, tBu), 2.45-2.50 (m, 1H, H-2'), 2.69-2.78 (m, 1H, H-2'), 3.12 (s, 3H, CH_3), 3.17 (s, 3H, CH_3), 3.67-3.88 (m, 2H, H-5'), 4.06 (m, 1H, H-4'), 4.25 (m, 2H, CH_2), 4.61 (m, 1H, H-3'), 4.84-4.97 (m, 2H, CH_2), 6.58 (t, J = 6.6 Hz, 1H, H-1'), 7.35-7.47 (m, 5H, H_{Ar}), 7.58-7.65 (m, 5H, H_{Ar}), 7.77 (s, 1H, H-8), 8.30 (s, 1H, CH), 8.79 (s, 1H, H-2), 10.05 (br s, 1H, NH).



(30)

3'-*O*-Azidomethyl-7-deaza-4-*N,N*-dimethylformadin-7-[3-(2,2,2-trifluoroacetamido)-prop-1-ynyl]-2'-deoxyadenosine (30).

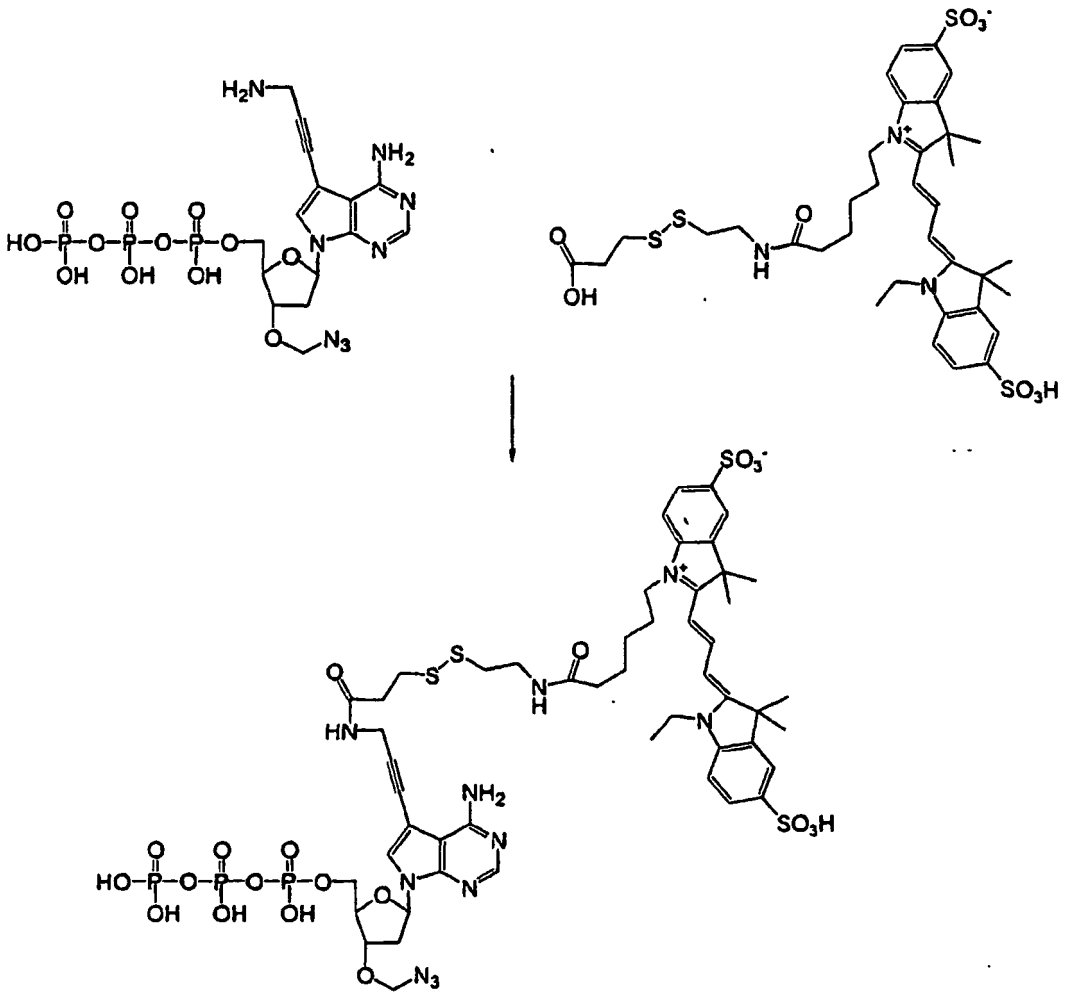
[0151] A solution of (29) (155 mg, 0.207 mmol) in solution in tetrahydrofuran (THF) (3 ml) was treated with TBAF (1 M in THF, 228 μl) at 0°C. The ice-bath was then removed and the reaction mixture stirred at room temperature. After 2 h-TLC indicated the full consumption of the nucleoside. The solvent was removed.. Purification by chromatography on silica (EtOAc:MeOH 95:5) gave (30) (86 mg, 82 %) as a pale brown oil. ^{1}H NMR (d_6 DMSO) δ 2.40-2.48 (dd, *J* = 8.1, 13.6 Hz, 1H, H-2'), 2.59-2.68 (dd, *J* = 8.3, 14 Hz, 1H, H-2'), 3.12 (s, 3H, CH_3), 3.17 (s, 3H, CH_3), 3.52-3.62 (m, 2H, H-5'), 4.02 (m, 1H, H-4'), 4.28 (d, *J* = 5.6 Hz, 2H, CH_2NH), 4.47 (m, 1H, H-3'), 4.89 (s, 2H, CH_2N_3), 5.19 (t, *J* = 5.6 Hz, 1H, OH), 6.49 (dd, *J* = 8.1, 8.7 Hz, 1H, H-1'), 7.88 (s, 1H, H-8), 8.34 (s, 1H, CH), 8.80 (s, 1H, H-2), 10.08 (s, 1H, NH).



(31)

7-(3-Aminoprop-1-ynyl)-3'-*O*-azidomethyl -7-deaza-2'-deoxyadenosine 5'-*O*-nucleoside triphosphate (31).

[0152] The nucleoside (30) and proton sponge was dried over P_2O_5 under vacuum overnight. A solution of (30) (150 mg, 0.294 mmol) and proton sponge (126 mg, 0.588 mmol) in trimethylphosphate (980 μl) was stirred with 4Å molecular sieves for 1 h. Freshly distilled $POCl_3$ (36 μl, 0.388 mmol) was added and the solution was stirred at 4°C for 2h. The mixture was slowly warmed up to room temperature and bis (tri-n-butyl ammonium) pyrophosphate 0.5 M solution in DMF (2.35 ml, 1.17 mmol) and anhydrous tri-n-butyl amine (560 μl, 2.35 mmol) was added. After 5 min, the reaction was quenched with 0.1 M TEAB (triethylammonium bicarbonate) buffer (15 ml) and stirred for 3h. The water was removed under reduced pressure and the resulting residue dissolved in concentrated ammonia (p 0.88, 15 ml) and stirred at room temperature for 16 h. The reaction mixture was then evaporated to dryness. The residue was dissolved in water and the solution applied to a DEAE-Sephadex A-25 column. MPLC was performed with a linear gradient of 0.05 M to 1 M TEAB. Fractions containing the product were combined and evaporated to dryness. The residue was dissolved in water and further purified by HPLC. HPLC: t_r(31) : 19.94 min (Zorbax C18 preparative column, gradient: 5% to 35% B in 20 min, buffer A 0.1M TEAB, buffer B MeCN). The product (31) was isolated as a white foam (17.5 μmol, 5.9%, ε_{280} = 15000). ^{1}H NMR (D_2O) δ 2.67-2.84 (2m, 2H, H-2'), 4.14 (m, 2H, CH_2NH), 4.17-4.36 (m, 2H, H-5'), 4.52 (br s, H-4'), 6.73 (t, J = 6.6 Hz, H-1'), 8.06 (s, 1H, H-8), 8.19 (s, 1H, H-2). ^{31}P NMR (D_2O) δ -5.07 (d, J = 21.8 Hz, 1P, P_γ), -10.19 (d, J =19.8 Hz, 1P, P_α), -21.32 (t, J = 19.8 Hz, 1P, P_β). Mass (-ve electrospray) calcd for $C_{15}H_{21}N_8O_{12}P_3$ 598.05, found 596.



(32)

[0153] To the Cy3 disulphide linker (1.3 μmol) in solution in DMF (450 μl) is added at 0°C 50 μl of a mixture of 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride, 1-hydroxybenzotriazole hydrate and N-methylmorpholine (26 μM each) in DMF. The reaction mixture wais stirred at room temperature for 1 h. The reaction was monitored by TLC (MeOH:CH_2Cl_2 3:7) until all the dye linker was consumed. Then DMF (400 μl) was added at 0°C, followed by the nucleotide (31) (1.2 μmol) in solution in water (100 μl) and the reaction mixture and stirred at room temperature overnight. TLC (MeOH:CH_2Cl_2 4:6) showed complete consumption of the activated ester and a dark red spot appeared on the baseline. The reaction was quenched with TEAB buffer (0.1M, 10 ml) and loaded on a DEAE Sephadex column (2 x 5 cm). The column was first eluted with 0.1 M TEAB buffer (100 ml) to wash off organic residues and then 1 M TEAB buffer (100 ml). The desired triphosphate (32) was eluted out with 1 M TEAB buffer. The fraction containing the product were combined, evaporated and purified by HPLC. HPLC conditions: t_r(32): 22.44 min (Zorbax C18 preparative column, gradient: 5% to 35% B in 20 min, buffer A 0.1M TEAB, buffer B MeCN). The product was isolated as dark pink solid (0.15 μmol, 12.5%, ε_{550} = 150000). ^{1}H NMR (D_2O) δ 2.03 (t, 2H, CH_2), 2.25 (m, 1H, H-2'), 2.43 (m, 1H, H-2'), 2.50 (m, 2H, CH_2), 2.66 (m, 2H, CH_2), 3.79 (m, 2H CH_2), 3.99 (m, 4H, CH_2N, H-5'), 4.18 (br s, 1H, H-4'), 6.02, 6.17 (2d, *J* = 13.64 Hz, 2H, H_{Ar}), 6.30 (dd, *J* = 6.06, 8.58 Hz, H-1'), 7.08, 7.22 (2d, 2H, 2 x =CH), 7.58-7.82 (m, 5H, H_{Ar}, H-2, H-8), 8.29 (m, =CH). ^{31}P NMR (D_2O) δ -4.83 (m, 1P, P_γ), -10.06 (m, 1P, P_α), -20.72 (m, 1P, P_β).

Enzyme Incorporation of 3'-Azidomethyl dNTPs

[0154] To a 100 nM DNA primer/template (primer previously labelled with P32 and T4 polynucleotide kinase) in Tris-HCl pH 8.8 50 mM, Tween-20 0.01%, and $MgSO_4$ 4 mM, add 2 μM compound 6 and 100 nM polymerase *(Thermococcus sp*. 9°N exo $^-$Y409V A485L supplied by New England Biolabs). The template consists of a run of 10 adenine bases to show the effect of the block. The reaction is heated to 65 C for 10 mins. To show complete blocking, a chase is performed

with the four native, unblocked nucleoside triphosphates. Quantitative incorporation of a single azidomethyl blocked dTTP can be observed and thus the azidomethyl group can be seen to act as an effective block to further incorporation.

[0155] By attaching a hairpin DNA (covalently attached self complementary primer/template) to a streptavidin bead The reaction can be performed over multiple cycles as shown in figures 5 and 6.

Preparation of the streptavidin beads

[0156] Remove the storage buffer and wash the beads 3 times with TE buffer (Tris-HCl pH 8, 10 mM and EDTA, 1 mM). Resuspend in B & W buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA and 2.0 M NaCl), add biotinylated ^{32}P labelled hairpin DNA with appropriate overhanging template sequence. Allow to stand at room temperature for 15 minutes. Remove buffer and wash beads 3 times TE buffer.

Incorporation of the fully functional nucleoside triphosphate (FFN)

[0157] To a solution of Tris-HCl pH 8.8 50 mM, Tween-20 0.01%, $MgSO_4$ 4 mM, $MnCl_2$ 0.4 mM (except cycle 1, 0.2 mM), add 2 μM FFN and 100 nM polymerase. This solution is then added to the beads and mixed thoroughly and incubated at 65°C for 10-15 minutes. The reaction mixture is removed and the beads washed 3 times with TE buffer.

Deblocking step

[0158] Tris-(2-carboxyethyl)phosphines trisodium salt (TCEP) (0.1M) is added to the beads and mixed thoroughly. The mixture was then incubated at 65°C for 15 minutes. The deblocking solution is removed and the beads washed 3 times with TE buffer.

Capping step

[0159] Iodoacetamide (431 mM) in 0.1 mM phosphate pH 6.5 is added to the beads and mixed thoroughly, this is then left at room temperature for 5 minutes. The capping solution is removed and the beads washed 3 times with TE buffer.

Repeat as required

[0160] The reaction products can be analysed by placing the bead solution in the well of a standard 12% polyacrylamide DNA sequencing gel in 40% formamide loading buffer. Running the gel under denaturing conditions causes the DNA to be released from the beads and onto the gel. The DNA band shifts are affected by both the presence of dye and the addition of extra nucleotides and thus the cleavage of the dye (and block) with the phosphine cause a mobility shift on the gel.

[0161] Two cycles of incorporation with compounds 18 (C), 24 (G) and 32 (A) and six cycles with compound 6 can be seen in figures Fig. 5 and Fig. 6.

Claims

1. A modified nucleotide molecule comprising a purine or pyrimidine base and a ribose or deoxyribose sugar moiety having a removable 3'-OH blocking group covalently attached thereto, such that the 3' carbon atom has attached a group of the structure

   -O-Z

   wherein Z is any of $-C(R')_2-N(R'')_2$'$C(R')_2-N(H)R''$, and $-C(R')_2-N_3$,
   wherein each R'' is or is part of a removable protecting group;
   each R' is independently a hydrogen atom, an alkyl, substituted alkyl, arylalkyl, alkenyl, alkynyl, aryl, heteroaryl, heterocyclic, acyl, cyano, alkoxy, aryloxy, heteroaryloxy or amido group, or a detectable label attached through a linking group; or $(R')_2$ represents an alkylidene group of formula $=C(R''')_2$ wherein each R''' may be the same or different and is selected from the group comprising hydrogen and halogen atoms and alkyl groups; and
   wherein said molecule may be reacted to yield an intermediate in which each R'' is exchanged for H, which intermediate dissociates under aqueous conditions to afford a molecule with a free 3'OH.

2. A molecule according to claim 1 wherein R' is an alkyl or substituted alkyl.

3. A molecule according to claim 1 or claim 2 wherein - Z is of formula -C(R')-N_3.

4. A molecule according to any one of claims 1 to 3 wherein Z is an azidomethyl group.

5. A molecule according to claim 1 or claim 2 wherein R" is a benzyl or substituted benzyl group.

6. A molecule according to any preceding claim wherein said base is linked to a detectable label via a cleavable linker or a non-cleavable linker.

7. A molecule according to claim 6 wherein said linker is cleavable.

8. A molecule according to any one of claims 1 to 5 wherein a detectable label is linked to the molecule through the blocking group by a cleavable or non-cleavable linker.

9. A molecule according to any one of claims 6 to 8 wherein said detectable label is a fluorophore.

10. A molecule according to any one of claims 6 to 9 wherein said linker is acid labile, photolabile or contains a disulfide linkage.

11. A modified nucleotide molecule as claimed in any one of claims 1 to 10 which comprises one or more ^{32}P atoms in its phosphate portion.

12. A method of controlling the incorporation of a nucleotide as defined in any one of claims 6 to 10 and complementary to a second nucleotide in a target single-stranded polynucleotide in a synthesis or sequencing reaction comprising incorporating into the growing complementary polynucleotide said nucleotide, the incorporation of said nucleotide preventing or blocking introduction of subsequent nucleoside or nucleotide molecules into said growing complementary polynucleotide.

13. The method of claim 12, wherein the incorporation of said first nucleotide is accomplished by a terminal transferase or polymerase or a reverse transcriptase.

14. The method of claim 13 wherein the polymerase is a *Thermococcus sp*

15. The method of claim 14 wherein the *Thermococcus sp* is 9°N or a single mutant or double mutant thereof.

16. The method of claim 15 wherein the double mutant is -Y409V A485L.

17. A method for determining the sequence of a target single-stranded polynucleotide, comprising monitoring the sequential incorporation of complementary nucleotides, wherein at least one incorporation is of a nucleotide as defined in any one of claims 6 to 10 and wherein the identity of the nucleotide incorporated is determined by detecting the label linked to the base, and the blocking group and said label are removed prior to introduction of the next complementary nucleotide.

18. The method according to claim 17 wherein the label of the nucleotide and the blocking group are removed in a single chemical treatment step.

19. The method according to claim 17, comprising:

(a) providing a plurality of different nucleotides wherein said plurality of different nucleotides are either as defined in any one of claims 6 to 10 and wherein the detectable label linked to each type of nucleotide can be distinguished upon detection from the detectable label used for other types of nucleotides;
(b) incorporating the nucleotide into the complement of the target single-stranded polynucleotide;
(c) detecting the label of the nucleotide of (b), thereby determining the type of nucleotide incorporated;
(d) removing the label of the nucleotide of (b) and the blocking group; and
(e) optionally repeating steps (b)-(d) one or more times;

thereby determining the sequence of a target single-stranded polynucleotide.

20. The method according to claim 19, wherein each of the nucleotides are brought into contact with the target sequentially, with removal of non-incorporated nucleotides prior to addition of the next nucleotide, and wherein detection and removal of the label and the blocking group is carried out either after addition of each nucleotide, or after addition of all four nucleotides.

21. The method according to claim 19, wherein each of the nucleotides are brought into contact with the target together simultaneously, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and the blocking group.

22. The method according to claim 19, comprising a first step and a second step, wherein in the first step, a first composition comprising two of the four nucleotides is brought into contact with the target and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label, and wherein in the second step, a second composition comprising the two nucleotides not included in the first composition is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent tc removal of the label and blocking group, and wherein the first and second steps are optionally repeated one or more times.

23. The method according to claim 19, comprising a first step and a second step, wherein in the first step, a composition comprising one of the four nucleotides is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent tc removal of the label and blocking group and wherein in the second step, a second composition comprising the three nucleotides not included in the first composition is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group and wherein the first step and the second step are optionally repeated one or more times.

24. The method according to claim 19, comprising a first step and a second step, wherein in the first step, a first composition comprising three of the four nucleotides is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group and wherein in the second step, a composition comprising the nucleotide not included in the first composition is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group and wherein the first step and the second step are optionally repeated one or more times.

25. A kit, comprising:

    (a) a plurality of different nucleotides wherein said plurality of different nucleotides are either as defined in any one of claims 6 to 10; and
    (b) packaging materials therefor.

26. A kit according to claim 25, wherein the detectable label in each nucleotide can be distinguished upon detection from the detectable label used for any of the other three types of nucleotide.

27. The kit of claim 25 or 26, further comprising an enzyme and buffers appropriate for the action of the enzyme.

28. Use of a nucleotide as defined in any one of claims 1 to 11 in a Sanger or a Sanger-type sequencing method.

29. An oligonucleotide comprising a modified nucleotide of claims 1-11.

30. A nucleotide triphosphate comprising a modified nucleotide of claims 1-11.

Patentansprüche

1. Modifiziertes Nucleotidmolekül, umfassend eine Purin- oder Pyrimidinbase und eine Ribose- oder Desoxyribose-Zuckereinheit mit einer kovalent daran gebundenen, entfernbaren, 3'-OH-blockierenden Gruppe, so dass an dem 3'-Kohlenstoffatom eine Gruppe der Struktur

    -O-Z

gebunden ist,

wobei es sich bei Z um eines von $-C(R')_2-N(R'')_2$, $-C(R')_2-N(H)R''$, und $-C(R')_2-N_3$ handelt,
wobei es sich bei jedem R" um eine entfernbare Schutzgruppe oder einen Teil davon handelt;
es sich bei jedem R' unabhängig um ein Wasserstoffatom, eine Alkyl-, substituierte Alkyl-, Arylalkyl-, Alkenyl-, Alkinyl-, Aryl-, Heteroaryl-, heterozyklische, Acyl-, Cyano-, Alkoxy-, Aryloxy-, Heteroaryloxy- oder Amido-Gruppe oder eine durch eine verknüpfende Gruppe gebundene nachweisbare Markierung handelt; oder $(R')_2$ eine Alkylidengruppe der Formel $=C(R''')_2$ darstellt, wobei jeder R'" gleich oder unterschiedlich sein kann und aus der Gruppe ausgewählt ist, umfassend Wasserstoff- und Halogenatome und Alkylgruppen; und
wobei das Molekül umgesetzt werden kann, um ein Zwischenprodukt zu ergeben, bei welchem jeder R" gegen H ausgetauscht ist, wobei dieses Zwischenprodukt unter wässrigen Bedingungen dissoziiert, um ein Molekül mit einem freien 3'-OH zu hervorzubringen.

2. Molekül gemäß Anspruch 1, wobei es sich bei R' um ein Alkyl oder substituiertes Alkyl handelt.

3. Molekül gemäß Anspruch 1 oder Anspruch 2, wobei -Z die Formel $-C(R')-N_3$ aufweist.

4. Molekül gemäß einem der Ansprüche 1 bis 3, wobei es sich bei Z um eine Azidomethylgruppe handelt.

5. Molekül gemäß Anspruch 1 oder Anspruch 2, wobei es sich bei R" um eine Benzyl- oder substituierte Benzylgruppe handelt.

6. Molekül gemäß einem der vorhergehenden Ansprüche, wobei die Base mittels eines spaltbaren Linkers oder eines nicht spaltbaren Linkers mit einer nachweisbaren Markierung verknüpft ist.

7. Molekül gemäß Anspruch 6, wobei der Linker spaltbar ist.

8. Molekül gemäß einem der Ansprüche 1 bis 5, wobei eine nachweisbare Markierung durch die blockierende Gruppe mit einem spaltbaren oder nicht spaltbaren Linker mit dem Molekül verknüpft ist.

9. Molekül gemäß einem der Ansprüche 6 bis 8, wobei es sich bei der nachweisbaren Markierung um ein Fluorophor handelt.

10. Molekül gemäß einem der Ansprüche 6 bis 9, wobei der Linker säurelabil oder photolabil ist oder eine Disulfidverknüpfung enthält.

11. Modifiziertes Nucleotidmolekül wie in einem der Ansprüche 1 bis 10 beansprucht, welches eines oder mehrere ^{32}P-Atome in seinem Phosphatanteil umfasst.

12. Verfahren zum Kontrollieren des Einbaus eines wie in einem der Ansprüche 6 bis 10 definierten und zu einem zweiten Nucleotid in einem einzelsträngigen Ziel-Polynucleotid komplementären Nucleotids bei einer Synthese- oder Sequenzierreaktion, umfassend das Einbauen des Nucleotids in das wachsende komplementäre Polynucleotid, wobei der Einbau des Nucleotids die Einführung darauffolgender Nucleosid- oder Nucleotidmoleküle in das wachsende komplementäre Polynucleotid verhindert oder blockiert.

13. Verfahren nach Anspruch 12, wobei der Einbau des ersten Nucleotids durch eine terminale Transferase oder Polymerase oder eine reverse Transkriptase erreicht wird.

14. Verfahren nach Anspruch 13, wobei es sich bei der Polymerase um eine *Thermococcus sp* handelt.

15. Verfahren nach Anspruch 14, wobei es sich bei der *Thermococcus sp* um 9°N oder eine Einzelmutante oder Doppelmutante davon handelt.

16. Verfahren nach Anspruch 15, wobei es sich bei der Doppelmutante um -Y409V A485L handelt.

17. Verfahren zum Bestimmen der Sequenz eines einzelsträngigen Ziel-Polynucleotids, umfassend das Überwachen des aufeinanderfolgenden Einbaus komplementärer Nucleotide, wobei es sich bei mindestens einem Einbau um den eines wie in einem der Ansprüche 6 bis 10 definierten Nucleotids handelt und wobei die Identität des eingebauten

Nucleotids durch Nachweisen der mit der Base verknüpften Markierung bestimmt wird und die blockierende Gruppe und die Markierung vor der Einführung des nächsten komplementären Nucleotids entfernt werden.

18. Verfahren gemäß Anspruch 17, wobei die Markierung des Nucleotids und die blockierende Gruppe in einem einzigen chemischen Behandlungsschritt entfernt werden.

19. Verfahren gemäß Anspruch 17, umfassend:

(a) Bereitstellen einer Vielzahl unterschiedlicher Nucleotide, wobei die Vielzahl unterschiedlicher Nucleotide jeweils wie in einem der Ansprüche 6 bis 10 definiert sind und wobei die nachweisbare Markierung, die mit jeder Nucleotidart verknüpft ist, von der nachweisbaren Markierung, die für andere Nucleotidarten verwendet wird, beim Nachweis unterschieden werden kann;
(b) Einbauen des Nucleotides in das Komplement des einzelsträngigen Ziel-Polynucleotids;
(c) Nachweisen der Markierung des Nucleotids nach (b), wodurch die Art des eingebauten Nucleotids bestimmt wird;
(d) Entfernen der Markierung des Nucleotides nach (b) und der blockierenden Gruppe und
(e) gegebenenfalls ein- oder mehrmaliges Wiederholen der Schritte (b) - (d),

wodurch die Sequenz eines einzelsträngigen Ziel-Polynucleotids bestimmt wird.

20. Verfahren gemäß Anspruch 19, wobei die Nucleotide jeweils nacheinander mit dem Ziel in Kontakt gebracht werden, mit Entfernung der nicht eingebauten Nucleotide vor der Zugabe des nächsten Nucleotids, und wobei Nachweis und Entfernung der Markierung und der blockierenden Gruppe entweder nach Zugabe von jedem Nucleotid oder nach Zugabe von allen vier Nucleotiden ausgeführt wird.

21. Verfahren gemäß Anspruch 19, wobei die Nucleotide jeweils gleichzeitig zusammen mit dem Ziel in Kontakt gebracht werden und nicht eingebaute Nucleotide vor dem Nachweis und anschließend an die Entfernung der Markierung und der blockierenden Gruppe entfernt werden.

22. Verfahren gemäß Anspruch 19, umfassend einen ersten Schritt und einen zweiten Schritt, wobei im ersten Schritt eine erste Zusammensetzung, umfassend zwei der vier Nucleotide, mit dem Ziel in Kontakt gebracht wird und nicht eingebaute Nucleotide vor dem Nachweis und anschließend an die Entfernung der Markierung entfernt werden, und wobei in dem zweiten Schritt eine zweite Zusammensetzung, umfassend die beiden Nucleotide, die in der ersten Zusammensetzung nicht eingeschlossen waren, mit dem Ziel in Kontakt gebracht wird und nicht eingebaute Nucleotide vor dem Nachweis und anschließend an die Entfernung der Markierung und der blockierenden Gruppe entfernt werden, und wobei der erste und der zweite Schritt gegebenenfalls einmal oder mehrmals wiederholt werden.

23. Verfahren gemäß Anspruch 19, umfassend einen ersten Schritt und einen zweiten Schritt, wobei im ersten Schritt eine Zusammensetzung, umfassend eines der vier Nucleotide, mit dem Ziel in Kontakt gebracht wird und nicht eingebaute Nucleotide vor dem Nachweis und anschließend an die Entfernung der Markierung und der blockierenden Gruppe entfernt werden, und wobei in dem zweiten Schritt eine zweite Zusammensetzung, umfassend die drei Nucleotide, die in der ersten Zusammensetzung nicht eingeschlossen waren, mit dem Ziel in Kontakt gebracht wird und nicht eingebaute Nucleotide vor dem Nachweis und anschließend an die Entfernung der Markierung und der blockierenden Gruppe entfernt werden, und wobei der erste Schritt und der zweite Schritt gegebenenfalls einmal oder mehrmals wiederholt werden.

24. Verfahren gemäß Anspruch 19, umfassend einen ersten Schritt und einen zweiten Schritt, wobei im ersten Schritt eine erste Zusammensetzung, umfassend drei der vier Nucleotide, mit dem Ziel in Kontakt gebracht wird und nicht eingebaute Nucleotide vor dem Nachweis und anschließend an die Entfernung der Markierung und der blockierenden Gruppe entfernt werden, und wobei in dem zweiten Schritt eine Zusammensetzung, umfassend das Nucleotid, das in der ersten Zusammensetzung nicht eingeschlossen war, mit dem Ziel in Kontakt gebracht wird und nicht eingebaute Nucleotide vor dem Nachweis und anschließend an die Entfernung der Markierung und der blockierenden Gruppe entfernt werden, und wobei der erste Schritt und der zweite Schritt gegebenenfalls einmal oder mehrmals wiederholt werden.

25. Kit, umfassend

(a) eine Vielzahl unterschiedlicher Nucleotide, wobei es sich bei der Vielzahl unterschiedlicher Nucleotide jeweils

um die in einem der Ansprüche 6 bis 10 definierten handelt, und
(b) Verpackungsmaterialien dafür.

26. Kit gemäß Anspruch 25, wobei die nachweisbare Markierung bei jedem Nucleotid beim Nachweis von der nachweisbaren Markierung unterschieden werden kann, die für eine der anderen drei Nucleotidarten verwendet wird.

27. Kit nach Anspruch 25 oder 26, ferner umfassend ein Enzym und für die Wirkung des Enzyms geeignete Puffer.

28. Verwendung eines Nucleotids, wie in einem der Ansprüche 1 bis 11 definiert, bei einem Sanger-Sequenzierverfahren oder einem Sequenzierverfahren der Sanger-Art.

29. Oligonucleotid, umfassend ein modifiziertes Nucleotid nach den Ansprüchen 1 - 11.

30. Nucleotidtriphosphat, umfassend ein modifiziertes Nucleotid nach den Ansprüchen 1 - 11.

Revendications

1. Molécule de nucléotide modifiée comprenant une base purine ou pyrimidine et un groupement sucre ribose ou désoxyribose ayant un groupe de blocage de 3'-OH retirable fixé de manière covalente à celle-ci, de sorte que l'atome de carbone 3' a un groupe fixé de structure

-O-Z

où Z est l'un quelconque de $-C(R')_2-N(R'')_2$, $-C(R')_2-N(H)R''$ et $-C(R')_2-N_3$,
où chaque R" est ou fait partie d'un groupe protecteur retirable ;
chaque R' est indépendamment un atome d'hydrogène, un groupe alkyle, alkyle substitué, arylalkyle, alcényle, alcynyle, aryle, hétéro-aryle, hétérocyclique, acyle, cyano, alcoxy, aryloxy, hétéroaryloxy ou amido, ou un marqueur détectable fixé par le biais d'un groupe de liaison ; ou bien $(R')_2$ représente un groupe alkylidène de formule $=C(R''')_2$ où chaque R"' peut être identique ou différent et est choisi dans le groupe comprenant les atomes d'hydrogène et d'halogène et les groupes alkyle ; et
où ladite molécule peut être mise à réagir pour produire un intermédiaire dans lequel chaque R" est échangé pour H, lequel intermédiaire se dissocie dans des conditions aqueuses pour donner une molécule avec un 3'OH libre.

2. Molécule selon la revendication 1 où R' est un alkyle ou alkyle substitué.

3. Molécule selon la revendication 1 ou la revendication 2 où -Z est de formule $-C(R')-N_3$.

4. Molécule selon l'une quelconque des revendications 1 à 3 où Z est un groupe azidométhyle.

5. Molécule selon la revendication 1 ou la revendication 2 où R" est un groupe benzyle ou benzyle substitué.

6. Molécule selon l'une quelconque des revendications précédentes où ladite base est liée à un marqueur détectable via un lieur clivable ou un lieur non clivable.

7. Molécule selon la revendication 6 où ledit lieur est clivable.

8. Molécule selon l'une quelconque des revendications 1 à 5 où un marqueur détectable est lié à la molécule par le biais du groupe de blocage par un lieur clivable ou non clivable.

9. Molécule selon l'une quelconque des revendications 6 à 8 où ledit marqueur détectable est un fluorophore.

10. Molécule selon l'une quelconque des revendications 6 à 9 où ledit lieur est labile en milieu acide, photolabile ou contient une liaison disulfure.

11. Molécule de nucléotide modifiée selon l'une quelconque des revendications 1 à 10 qui comprend un ou plusieurs atomes ^{32}P dans sa partie phosphate.

12. Procédé de régulation de l'incorporation d'un nucléotide selon l'une quelconque des revendications 6 à 10 et complémentaire d'un second nucléotide dans un polynucléotide simple brin cible dans une réaction de synthèse ou de séquençage comprenant l'incorporation dans le polynucléotide complémentaire en croissance dudit nucléotide, l'incorporation dudit nucléotide empêchant ou bloquant l'introduction de molécules de nucléoside ou de nucléotide subséquentes dans ledit polynucléotide complémentaire en croissance.

13. Procédé selon la revendication 12 où l'incorporation dudit premier nucléotide est accomplie par une terminal transférase ou une polymérase ou une transcriptase inverse.

14. Procédé selon la revendication 13 où la polymérase est une *Thermococcus sp.*

15. Procédé selon la revendication 14 où la *Thermococcus sp* est 9°N ou un mutant unique ou un double mutant de celle-ci.

16. Procédé selon la revendication 15 où le double mutant est -Y409V A485L.

17. Procédé pour déterminer la séquence d'un polynucléotide simple brin cible, comprenant le suivi de l'incorporation séquentielle de nucléotides complémentaires, où au moins une incorporation est d'un nucléotide selon l'une quelconque des revendications 6 à 10 et où l'identité du nucléotide incorporé est déterminée par détection du marqueur lié à la base, et le groupe de blocage et ledit marqueur sont retirés avant l'introduction du nucléotide complémentaire suivant.

18. Procédé selon la revendication 17 où le marqueur du nucléotide et le groupe de blocage sont retirés dans une seule étape de traitement chimique.

19. Procédé selon la revendication 17, comprenant :

(a) la fourniture d'une pluralité de nucléotides différents où ladite pluralité de nucléotides différents est définie dans l'une quelconque des revendications 6 à 10 et
où le marqueur détectable lié à chaque type de nucléotide peut être distingué lors de la détection du marqueur détectable utilisé pour d'autres types de nucléotides ;
(b) l'incorporation du nucléotide dans le complément du polynucléotide simple brin cible ;
(c) la détection du marqueur du nucléotide de (b), pour déterminer le type de nucléotide incorporé ;
(d) le retrait du marqueur du nucléotide de (b) et du groupe de blocage ; et
(e) éventuellement la répétition des étapes (b) - (d) une ou plusieurs fois ;

pour déterminer la séquence d'un polynucléotide simple brin cible.

20. Procédé selon la revendication 19, où chacun des nucléotides est mis en contact avec la cible successivement, avec retrait des nucléotides non incorporés avant l'addition du nucléotide suivant, et où la détection et le retrait du marqueur et du groupe de blocage sont réalisés après addition de chaque nucléotide ou après addition des quatre nucléotides.

21. Procédé selon la revendication 19, où chacun des nucléotides est mis en contact avec la cible en même temps simultanément, et les nucléotides non incorporés sont retirés avant la détection et après le retrait du marqueur et du groupe de blocage.

22. Procédé selon la revendication 19, comprenant une première étape et une seconde étape, où, dans la première étape, une première composition comprenant deux des quatre nucléotides est mise en contact avec la cible et les nucléotides non incorporés sont retirés avant la détection et après le retrait du marqueur, et où dans la seconde étape, une seconde composition comprenant les deux nucléotides non inclus dans la première composition est mise en contact avec la cible, et les nucléotides non incorporés sont retirés avant la détection et après le retrait du marqueur et du groupe de blocage, et où les première et seconde étapes sont éventuellement répétées une ou plusieurs fois.

23. Procédé selon la revendication 19, comprenant une première étape et une seconde étape, où dans la première étape, une composition comprenant l'un des quatre nucléotides est mise en contact avec la cible, et les nucléotides non incorporés sont retirés avant la détection et après le retrait du marqueur et du groupe de blocage et où dans

la seconde étape, une seconde composition comprenant les trois nucléotides non inclus dans la première composition est mise en contact avec la cible, et les nucléotides non incorporés sont retirés avant la détection et après le retrait du marqueur et du groupe de blocage et où la première étape et la seconde étape sont éventuellement répétées une ou plusieurs fois.

24. Procédé selon la revendication 19, comprenant une première étape et une seconde étape, où dans la première étape, une première composition comprenant trois des quatre nucléotides est mise en contact avec la cible, et les nucléotides non incorporés sont retirés avant la détection et après le retrait du marqueur et du groupe de blocage et où dans la seconde étape, une composition comprenant le nucléotide non Inclus dans la première composition est mise en contact avec la cible, et les nucléotides non incorporés sont retirés avant la détection et après le retrait du marqueur et du groupe de blocage et où la première étape et la seconde étape sont éventuellement répétées une ou plusieurs fois.

25. Kit, comprenant :

(a) une pluralité de nucléotides différents où ladite pluralité de nucléotides différents est telle que définie dans l'une quelconque des revendications 6 à 10 ; et
(b) des matériaux d'empaquetage pour ceux-ci.

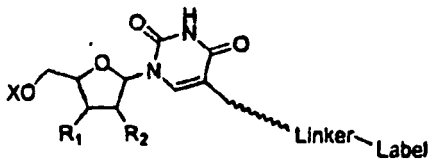
26. Kit selon la revendication 25, où le marqueur détectable dans chaque nucléotide peut être distingué lors de la détection du marqueur détectable utilisé pour l'un quelconque des trois autres types de nucléotides.

27. Kit selon la revendication 25 ou 26, comprenant en outre une enzyme et des tampons appropriés pour l'action de l'enzyme.

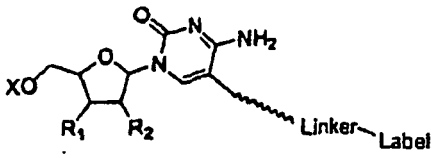
28. Utilisation d'un nucléotide selon l'une quelconque des revendications 1 à 11 dans un procédé de séquençage de Sanger ou de type Sanger.

29. Oligonucléotide comprenant un nucléotide modifié selon les revendications 1-11.

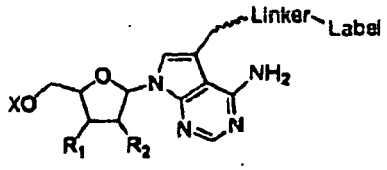
30. Nucléotide triphosphate comprenant un nucléotide modifié selon les revendications 1-11.



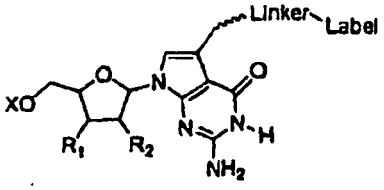
Uridine C5-linker



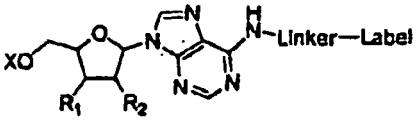
Cytidine C5-linker



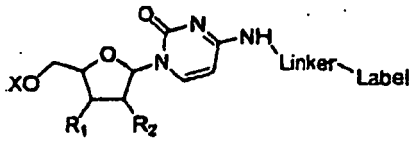
N7 Deazaadenosine C7-linker



N7 Deazaguanosine C7-linker



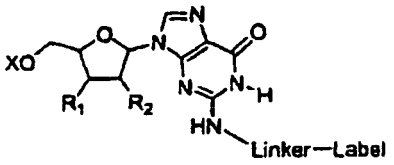
Adenosine N6-linker



Cytidine N4-linker

where R_1 and R_2, which may be the same or different, are each selected from H, OH, or any group which can be transformed into an OH. Suitable groups for R_1 and R_2 are described in Figure 3 ~~and~~ CLAIM 1

X = H, phosphate, diphosphate or triphosphate



Guanosine N2-linker

Fig. 1

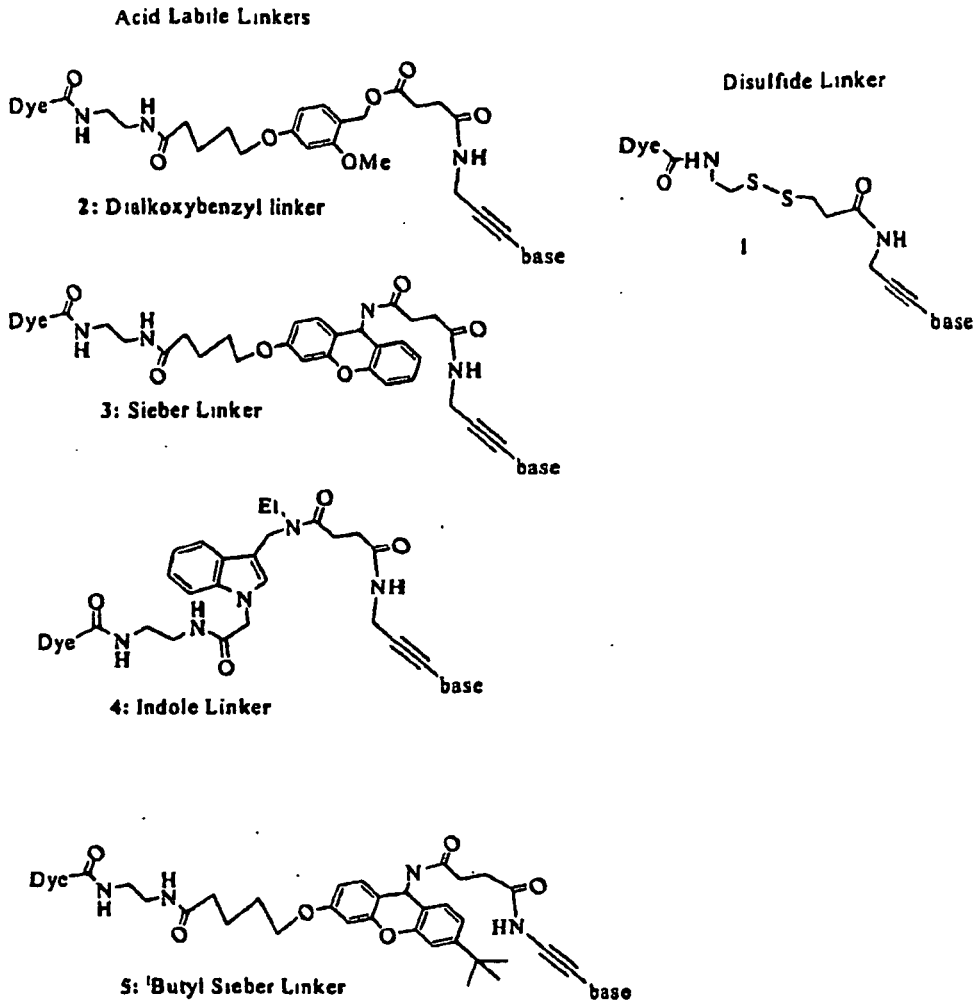
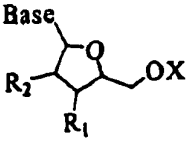
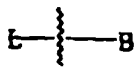

Acid Labile Linkers
Dye
OMe
NH
base
2: Dialkoxybenzyl linker
Disulfide Linker
Dye
S
S
NH
base
1
Dye
NH
base
3: Sieber Linker
Et
Dye
NH
base
4: Indole Linker
Dye
HN
base
5: tButyl Sieber Linker

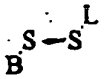
Fig. 2

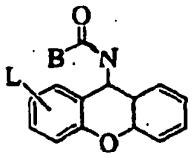
Label ~~Cleavable linker~~~~Base

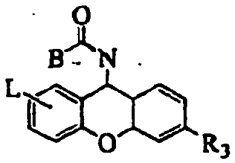


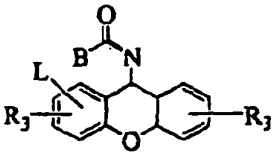


Cleavable linkers may include:







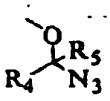


R3 represents one or more substituents independently selected from alkyl, alkoxy, amino or halogen

Alternatively, cleavable linkers may be constructed from any labile functionality used on the 3'-block

where R_1 and R_2, which may be the same or different, are each selected from H, OH, or any group which can be transformed into an OH,

R_1 and R_2 groups may include

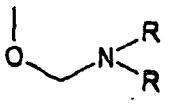


where R_4 is H or alkyl, R_5 is H or alkyl and R_6 is alkyl, cycloalkyl, alkenyl, cycloalkenyl or benzyl

and X is H, phosphate, diphosphate or triphosphate

Fig. 3

Protected hemiaminals:



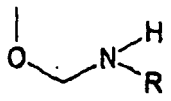


Fig. 4

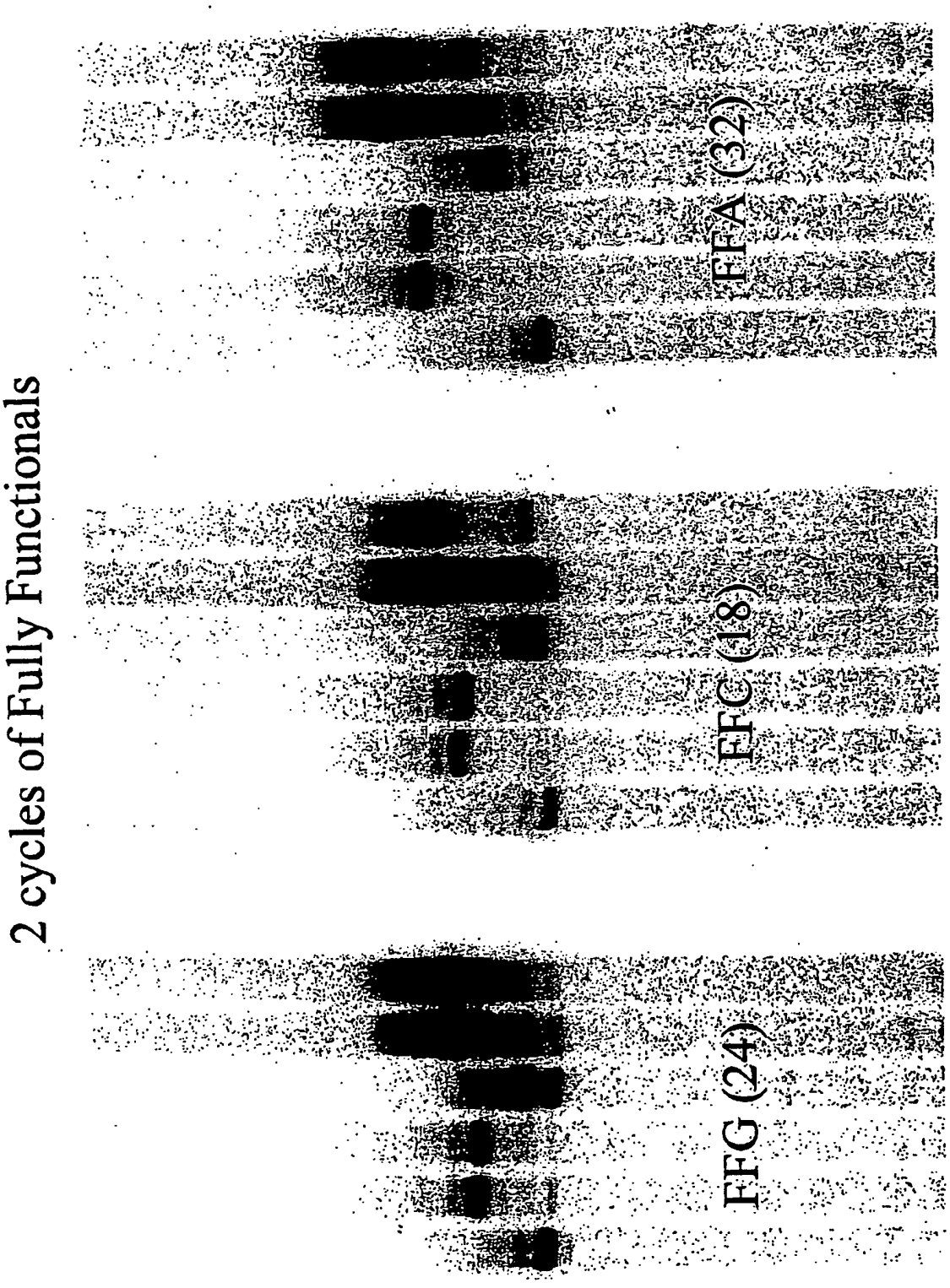


Fig. 5

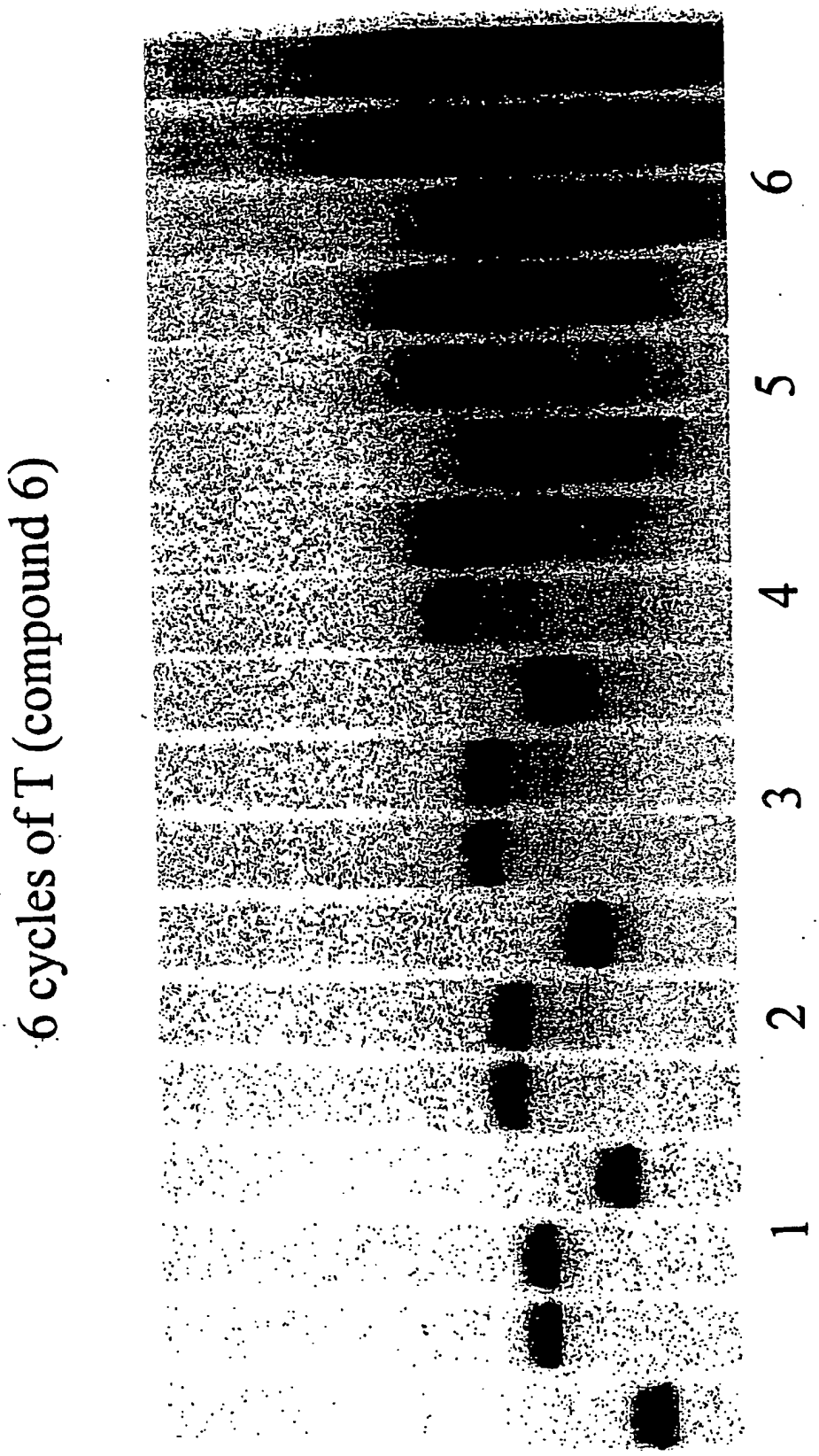


Fig. 6

Fig. 7

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO 0229003 A **[0008] [0010] [0011]**
- WO 03048387 A **[0030]**
- US 5302509 A **[0075]**
- WO 0006770 A **[0107]**

Non-patent literature cited in the description

- **FODOR et al.** *Trends Biotech.,* 1994, vol. 12, 19-26 **[0003]**
- **STIMPSON et al.** *Proc. Natl. Acad. Sci. USA,* 1995, vol. 92, 6379-6383 **[0003]**
- **METZKER et al.** *Nucleic Acids Research,* 1994, vol. 22 (20), 4259-4267 **[0007]**
- **KAMAL et al.** *Tet. Let,* 1999, vol. 40, 371-372 **[0008]**
- **SCHEIT.** Nucleotide Analogs. John Wiley & Son, 1980 **[0067]**
- **UHLMAN et al.** Chemical Reviews. 1990, vol. 90, 543-584 **[0067]**
- **WELCH et al.** *Chem. Eur. J.,* 1999, vol. 5 (3), 951-960 **[0076]**
- **ZHU et al.** *Cytometry,* 1997, vol. 28, 206-211 **[0076]**
- **PROBER et al.** *Science,* 1987, vol. 238, 336-341 **[0076]**
- **CONNELL et al.** *BioTechniques,* 1987, vol. 5 (4), 342-38.4 **[0076]**
- **ANSORGE et al.** *Nucl. Acids Res.,* 1987, vol. 15 (11), 4593-4602 **[0076]**
- **SMITH et al.** *Nature,* 1986, vol. 321, 674 **[0076]**
- *Tet. Let.,* 2000, vol. 46, 8867-8871 **[0077]**
- *J. Amer. Chem. Soc.,* 2001, vol. 123, 8101-8108 **[0077]**
- *Empodocles et al., Nature,* 1999, vol. 399, 126-130 **[0078]**
- **REICHERT et al.** *Anal. Chem.,* 2000, vol. 72, 6025-6029 **[0078]**
- **LACOSTE et al.** *Proc. Natl. Acad. Sci USA,* 2000, vol. 97 (17), 9461-9466 **[0078]**
- **GREENE ; WUTS.** Protective Groups in Organic Synthesis. John Wiley & Sons **[0090]**
- **GUILLIER et al.** *Chem. Rev.,* 2000, vol. 100, 2092-2157 **[0090]**
- **LEE et al.** *J. Org. Chem.,* 1999, vol. 64, 3454-3460 **[0097]**
- **BURGESS et al.** *J. Org. Chem.,* 1997, vol. 62, 5165-5168 **[0100]**
- *Proc. Natl. Acad. Sci. USA,* 1996, 5281-5285 **[0106]**
- *Nucleic Acids Research,* 1999, 2454-2553 **[0106]**
- *Acids Research,* 2002, 605-613 **[0106]**
- **MOYER et al.** *Laser Focus World,* 1993, vol. 29, 10 **[0114]**