SPECIFICATION

NOVEL ARTIFICIAL BASE PAIRS AND USES THEREOF

TECHNICAL FIELD

[0001] The present application claims priority to Japanese Patent Application Nos. 2005-226492 (filed on August 4, 2005) and 2006-182062 (filed on June 30, 2006), the entire contents of which are incorporated herein by reference.

[0002] The present invention relates to nucleic acids based on novel artificial base pairing and uses thereof.

BACKGROUND ART

[0003] In nucleic acids (DNA, RNA) which are biological macromolecules, enormous amounts of genetic information essential for vital activities are recorded as sequences composed of combinations of only 4 different bases. Such a nucleic acid allows self-replication using itself as a template by the action of DNA polymerases, and further undergoes processes of RNA polymerase-mediated transcription and ribosome-mediated translation to ensure the transmission of genetic information from DNA to DNA, from DNA to RNA, and/or from RNA to protein. These replication and transmission events of genetic information enabled exclusive base-pairings (A:T/U, G:C). In addition, nucleic acids can form a variety of higher-order structures and hence exert various functions. By way of example, it is one of the indications that a large number of novel nucleic acids having aptamer and/or ribozyme functions have been generated by *in vitro* selection techniques.

[0004] However, unlike proteins which are composed of 20 types of amino acids, the chemical and physical diversity of nucleic acids is limited by the fact that there are only 4 different bases (2 base pairs) in natural nucleic acids. For example, functional RNAs (e.g., tRNA, rRNA, mRNA) found in living organisms utilize various modified bases to stabilize their own structures and/or RNA-RNA and RNA-protein interactions. Thus, it will be very advantageous to expand the repertory of new bases (base pairs) in developing novel functional nucleic acids.

[0005] With the aim of further expansion of nucleic acid functions, attempts have been made to design nucleosides or nucleotides having unnatural bases. There are two possible approaches for introducing modified bases (or unnatural bases) into nucleic acids: 1) direct introduction by chemical synthesis; and 2) introduction catalyzed by DNA and RNA polymerase enzymes. In the case of 1), there is a need to solve some problems associated with chemical synthesis, such as the stability of amidite units and the presence of protecting groups appropriate for base moieties. If these problems are solved, various unnatural bases can be introduced in a site-selective manner. However, the nucleic acids thus obtained are difficult to be amplified and it is also difficult to synthesize long-chain nucleic acids. In the case of 2), if the enzymes recognize substrates to cause replication and transcription between artificial base pairs in a complementary manner, nucleic acids containing such artificial base pairs can be amplified and prepared. However,

such substrates and base pairs (unnatural nucleotides) are still under development.

[0006] Research background of unnatural artificial base pairs

In natural double-stranded DNA, the "exclusive" A-T and G-C base pairs are formed through specific hydrogen bonding. Studies of unnatural base pairs have been known for combinations based on hydrogen bonding between bases and/or combinations based on the hydrophobicity of bases, but no unnatural base pair has been found that can compete with natural base pairs in all steps of replication, transcription and translation. Under these circumstances, an unnatural base pair capable of competing with natural base pairs in at least one step of replication, transcription and translation will have a specific utility.

[0007] Recent studies of unnatural non-hydrogen-bonded base pairs have shown that hydrogen bonding between paired bases is not absolutely required for efficient and adequate incorporation of nucleotides during replication; rather, shape complementarity between bases plays an important role during replication (Non-patent Documents 5-6). Thus, non-hydrogen-bonded base pairs are potential candidates for expansion of the genetic alphabet and code to create novel biotechnologies (Non-patent Documents 7-9).

[0008] On the other hand, no report has been issued on a non-hydrogen-bonded base pair that is recognized by RNA polymerase with high selectivity and efficiency during transcription. More specifically, the inventors of the

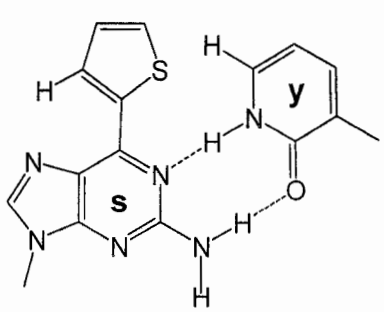
present invention have developed a (s-y) base pair through interaction of hydrogen bonding between a 2-amino-6-(2-thienyl)-9H-purin-9-yl group (s) and a 2-oxo-(1H)pyridin-3-yl group (y), and have succeeded in *in vitro* synthesis of proteins site-specifically containing an unnatural amino acid(s) (Patent Document 1, Non-patent Document 14).

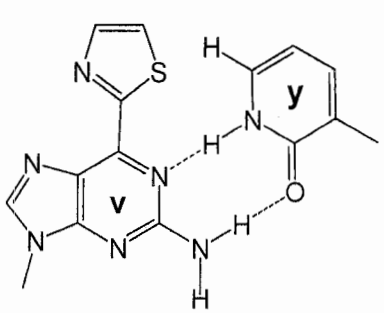
[0009] Moreover, Patent Document 3 discloses a (v-y) base pair between a 2-amino-6-(2-thiazolyl)purin-9-yl group (v) and a 2-oxo-(1H)pyridin-3-yl group (y) (Patent Document 3).

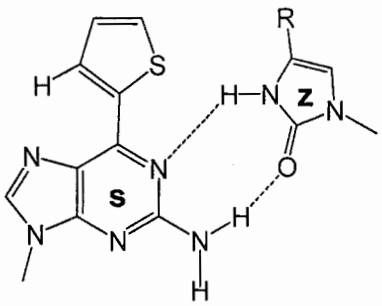
[0010] As the nucleic acid base pairs of s, (s-z) (wherein z represents a 2-oxo-1,3-dihydroimidazol-1-yl group) and (s-5-substituted y) other than the four natural bases and (s-y) has been reported prior to the present invention. These unnatural (s-y), (s-z) and (s-5-substituted y) base pairs all involve hydrogen bonding interaction (Patent Document 2, Non-patent Documents 13-15).

[0011] However, among these unnatural base pairs, (s-y) and (s-5-substituted y) are less selective in a transcription step where s is inserted into RNA using y as a DNA template, while (s-z) is highly selective, but does not necessarily provide high yield.

[0012] [Formula 1]

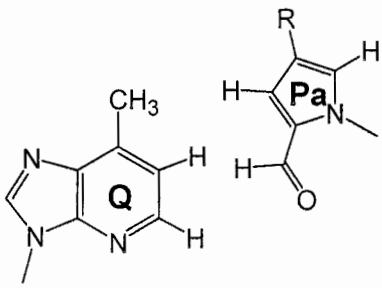






[0013] On the other hand, a 2-formyl-1H-pyrrol-1-yl group (Pa) has been reported to be able to form non-hydrogen-bonded base pairs not only with the four natural bases, but also with 9-methylimidazo[(4,5)-b]pyridine (Q) (Pa-Q) (Non-patent Document 9).

[0014] [Formula 2]



[0015] However, the above non-hydrogen-bonded base pairs are reported only for translation by DNA polymerase or DNA synthesis by reverse transcriptase. There is no knowledge about their recognition by RNA polymerase. A single subunit of T7-like RNA polymerase resembles DNA polymerase both in structure and mechanism (Non-patent Documents 16-17), but recent structural analyses of T7 RNA polymerase complexes have clarified differences between RNA polymerase and DNA polymerase (Non-patent Documents 17-18). In the case of

transcription, an incoming substrate forms a base pair with a base in its template in the "open" conformation, and the formed base pair is maintained during transition from "open" to "closed" conformation (Non-patent Document 18). However, in the case of replication, base pairing starts after transition to the "closed" conformation. This suggests that hydrogen bonding between paired bases is more important in transcription than in replication, which in turn leads to a question of whether non-hydrogen-bonded base pairs could be functional during transcription.

[0016] If unnatural base-mediated specific transcription is achieved, it will be possible to design novel RNA molecules having promoted functions and to expand the genetic code (Non-patent Documents 1-4).

[0017] Introduction of fluorescent probes into RNA

In response to recent growing interest in RNA therapy, introduction of fluorescent probes into RNA have become an important technology for fluorescently labeling RNA and for analyzing the complex higher order structure of RNA. Fluorescent probes previously used for the latter structural analysis include those using a nucleotide having a fluorescent base (e.g., 2-aminopurine). In this regard, the 2-amino-6-(2-thienyl)-9H-purin-9-yl group (s) mentioned above is a fluorescent base, and a nucleotide containing this group as a base has strong fluorescence properties. Thus, if a nucleotide containing s as a base can be introduced at any site in RNA, it will be possible to prepare a useful fluorescent probe.

[0018] However, known techniques used for introducing such a nucleotide having a fluorescent base into RNA through transcription are not necessarily sufficient in terms of yield. It has therefore been substantially impossible to introduce a nucleotide having a fluorescent base (e.g., s) into RNA in a site-specific manner through transcription, thus making it difficult to analyze a higher order structure of long-chain RNA.

[0019] Once a technique has been developed for introducing a fluorescent probe into RNA through transcription, it will be possible to study structural dynamics of RNA in a solution. Moreover, such a technique would be also helpful in developing RNA-based drugs. For these reasons, it is important to establish a technique that allows introduction of a fluorescent probe into long-chain RNA, which has thus far been regarded as impossible. This would enable the development of new techniques for use in *in vivo* structural analysis of functional RNA and in structural analysis of RNA complexes with other molecules.

[0020] There is a demand for the development of a novel artificial base pair for use in introducing, into a nucleic acid, a nucleotide having an unnatural base which can compete with natural base pairs in all steps of replication, transcription and translation and which imparts fluorescence properties to a nucleic acid.

Patent Document 1: WO2001/005801

Patent Document 2: WO2004/007713

Patent Document 3: WO2005/026187

Non-patent Document 1: Benner, S. A., Burgstaller, P., Battersby, T. R. & Jurczyk, S. in The RNA World (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 163-181 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999).

Non-patent Document 2: Bergstrom, D. E. Orthogonal base pairs continue to evolve. Chem. Biol. 11, 18-20 (2004).

Non-patent Document 3: Wang, L. & Schultz, P. G. Expanding the genetic code. Chem. Commun. 1-11 (2002).

Non-patent Document 4: Hendrickson, T. L., de Crecy-Lagard, V. & Schimmel, P. Incorporation of nonnatural amino acids into proteins. Annu. Rev. Biochem. 73, 147-176 (2004).

Non-patent Document 5: Morales, J. C. & Kool, E. T. Efficient replication between non-hydrogen-bonded nucleoside shape analogs. Nat. Struct. Biol. 5, 950-954 (1998).

Non-patent Document 6: Kool, E. T. Hydrogen bonding, base stacking, and steric effects in DNA replication. Annu. Rev. Biophys. Biomol. Struct. 30, 1-22 (2001).

Non-patent Document 7: McMinn, D. L. et al. Efforts toward expansion of the genetic alphabet: DNA polymerase recognition of a highly stable, self-pairing hydrophobic base. J. Am. Chem. Soc. 121, 11585-11586 (1999).

Non-patent Document 8: Wu,Y. et al. Efforts toward expansion of the genetic alphabet: optimization of interbase hydrophobic interactions. J. Am. Chem. Soc. 122, 7621-7632 (2000).

Non-patent Document 9: Mitsui, T. et al. An unnatural hydrophobic base pair with shape complementarity between

pyrrole-2-carbaldehyde and 9-methylimidazo[(4,5)-b]pyridine. J. Am. Chem. Soc. 125, 5298-5307 (2003).

Non-patent Document 10: Piccirilli, J. A., Krauch, T., Moroney, S. E. & Benner, S. A. Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. Nature 343, 33-37 (1990).

Non-patent Document 11: Switzer, C. Y., Moroney, S. E. & Benner, S. A. Enzymatic recognition of the base pair between isocytidine and isoguanosine. Biochemistry 32, 10489-10496 (1993).

Non-patent Document 12: Tor, Y. & Dervan, P. B. Site-specific enzymatic incorporation of an unnatural base, N^6-(6-aminohexyl)isoguanosine, into RNA. J. Am. Chem. Soc. 115, 4461-4467 (1993).

Non-patent Document 13: Ohtsuki, T. et al. Unnatural base pairs for specific transcription. Proc. Natl. Acad. Sci. U.S.A. 98, 4922-4925 (2001).

Non-patent Document 14: Hirao, I. et al. An unnatural base pair for incorporating amino acid analogs into proteins. Nat. Biotechnol. 20, 177-182 (2002).

Non-patent Document 15: Hirao, I. et al. A two-unnatural-base-pair system toward the expansion of the genetic code. J. Am. Chem. Soc. 126, 113298-113305 (2004).

Non-patent Document 16: Cheetham, G. M., Jeruzalmi, D. & Steitz, T. A. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. Nature 399, 80-83 (1999).

Non-patent Document 17: Yin, Y. W. & Steitz, T. A. The

structural mechanism of translocation and helicase activity in T7 RNA polymerase. Cell 116, 393-404 (2004).

Non-patent Document 18: Temiakov, D. et al. Structural basis for substrate selection by T7 RNA polymerase. Cell 116, 381-391 (2004).

Non-patent Document 19: Hirao, I., Fujiwara, T., Kimoto, M. & Yokoyama, S. Unnatural base pairs between 2- and 6-substituted purines and 2-oxo(1H)pyridine for expansion of the genetic alphabet. Bioorg. Med. Chem. Lett. 14, 4887-4890 (2004).

Non-patent Document 20: Jovine, L., Djordjevic, S. & Rhodes, D. The crystal structure of yeast phenylalanine tRNA at 2.0 Å resolution: cleavage by Mg^{2+} in 15-year old crystals. J. Mol. Biol. 301, 401-414 (2000).

Non-patent Document 21: Law, S. M., Eritja, R., Goodman, M. F. & Breslauer, K. J. Spectroscopic and calorimetric characterizations of DNA duplexes containing 2-aminopurine. Biochemistry 35, 12329-12337 (1996).

Non-patent Document 22: Holz, B., Klimasauskas, S., Serva, S. & Weinhold, E. 2-Aminopurine as a fluorescent probe for DNA base flipping by methyltransferases. Nucleic Acids Res. 26, 1076-1083 (1998).

Non-patent Document 23: Rist, M. J. & Marino, J. P. Association of an RNA kissing complex analyzed using 2-aminopurine fluorescence. Nucleic Acids Res. 29, 2401-2408 (2001).

Non-patent Document 24: Bain, J. D., Switzer, C., Chamberlin, A. R. & Benner, S. A. Ribosome-mediated

incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. Nature 356, 537-539 (1992).

Non-patent Document 25: Ludwig, J. & Eckstein, F. Rapid and efficient synthesis of nucleoside 5′-O-(1-thiotriphosphates), 5′-triphosphates and 2′,3′-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. J. Org. Chem. 54, 631-635 (1989).

Non-patent Document 26: Frugier, M., Helm, M., Felden, B., Giege, R. & Florentz, C. Sequences outside recognition sets are not neutral for tRNA aminoacylation. J. Biol. Chem. 273, 11605-11610 (1998).

Non-patent Document 27: Kao, C., Rudisser, S. & Zheng, M. A simple and efficient method to transcribe RNAs with reduced 3′ heterogeneity. Methods 23, 201-205 (2001).

Non-patent Document 28: Mitsui, T., Kimoto, M., Sato, A., Yokoyama, S. & Hirao, I., Bioorg. Med. Chem. Lett., 13, 4515-4518 (2003)

Non-patent Document 29: Fujiwara, T., Kimoto, M., Sugiyama, H., Hirao, I. & Yokoyama, S., Bioorg. Med. Chem. Lett., 11, 2221-2223 (2001)

Non-patent Document 30: Mitsui, T., Kimoto, M., Harada, Y., Yokoyama, S. & Hirao, I., J. Am. Chem. Soc., 127, 8652-8658 (2005)

DISCLOSURE OF THE INVENTION

PROBLEMS TO BE SOLVED BY THE INVENTION

[0021] The present invention aims to provide a nucleic acid based on novel artificial base pairing. More

specifically, in the nucleic acid of the present invention, a nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base forms a base pair with a nucleotide having a 6-substituted 9H-purin-9-yl group as a base.

[0022] Without being limited thereto, in the nucleic acid of the present invention, the above substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group is preferably selected from the group consisting of:

A1) a 2-formyl-1H-pyrrol-1-yl group;

A2) a 2-formyl-4-(1-propyn-1-yl)-1H-pyrrol-1-yl group;

A3) a 2-formyl-4-methyl-1H-pyrrol-1-yl group; and

A4) a 2-formyl-4-ethynyl-1H-pyrrol-1-yl group,

and the above 6-substituted 9H-purin-9-yl group is preferably selected from the group consisting of:

B1) a 2-amino-6-(2-thienyl)-9H-purin-9-yl group;

B2) a 6-(2-thienyl)-9H-purin-9-yl group;

B3) a 2-amino-6-(4-methyl-2-thienyl)-9H-purin-9-yl group;

B4) a 6-(4-methyl-2-thienyl)-9H-purin-9-yl group;

B5) a 2-amino-6-(5-methyl-2-thienyl)-9H-purin-9-yl group;

B6) a 6-(5-methyl-2-thienyl)-9H-purin-9-yl group;

B7) a 2-amino-6-(2-thiazolyl)-9H-purin-9-yl group;

B8) a 6-(2-thiazolyl)-9H-purin-9-yl group;

B9) a 2-amino-6-(4-methyl-2-thiazolyl)-9H-purin-9-yl group;

B10) a 6-(4-methyl-2-thiazolyl)-9H-purin-9-yl group;

B11) a 2-amino-6-(5-methyl-2-thiazolyl)-9H-purin-9-yl group; and

B12) a 6-(5-methyl-2-thiazolyl)-9H-purin-9-yl group.

[0023] The nucleic acid of the present invention preferably forms a base pair(s) in the step of transcription, reverse transcription, replication or translation.

[0024] The present invention also provides a method for preparing a nucleic acid incorporating a nucleotide having a 6-substituted 9H-purin-9-yl group as a base. The preparation method of the present invention comprises effecting transcription, reverse transcription or replication by using, as a template, a nucleic acid incorporating a nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base, so that the nucleotide having a 6-substituted 9H-purin-9-yl group is incorporated at a site complementary to the nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group.

[0025] The present invention further aims to provide such a nucleic acid incorporating a nucleotide having a 6-substituted 9H-purin-9-yl group as a base, which is prepared by the above method of the present invention.

[0026] The present invention further aims to provide a method for analyzing the stereostructure of a local region in the nucleic acid of the present invention, wherein the local region includes a nucleotide having a 6-substituted 9H-purin-9-yl group as a base. The analysis method of the present invention comprises measuring the fluorescence intensity from the base in the nucleotide having a 6-substituted 9H-purin-9-

yl group as a base, under different environmental conditions such as varying temperatures or ion (e.g., magnesium) concentrations.

[0027] In one embodiment of the analysis method of the present invention, a determination is made that the nucleotide having a 6-substituted 9H-purin-9-yl group as a base stacks with its conformationally neighboring nucleotides at *in vivo* temperature if the fluorescence intensity from the base in the nucleotide is substantially increased with increase in temperature, or alternatively, a determination is made that the nucleotide is exposed outside at *in vivo* temperature if the fluorescence intensity is substantially decreased or not increased with increase in temperature.

[0028] The present invention also aims to provide a method for detecting the formation of a nucleic acid duplex. The detection method of the present invention comprises:

I) inducing hybridization between the following nucleic acids:

i) a nucleic acid incorporating a nucleotide having a 6-substituted 9H-purin-9-yl group as a base, which is prepared by the inventive method for preparing a nucleic acid, and

ii) a nucleic acid incorporating a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is attached either directly or through a linker to a fluorescent dye selected from the group consisting of 5-carboxyfluorescein (5-FAM), 6-carboxyfluorescein (6-FAM), 5-carboxytetramethylrhodamine (5-TAMRA), 6-carboxytetramethylrhodamine (6-TAMRA), 5-

dimethylaminonaphthalene-1-sulfonic acid (DANSYL), 5-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (5-HEX), 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (6-HEX), 5-carboxy-2',4,7,7'-tetrachlorofluorescein (5-TET), 6-carboxy-2',4,7,7'-tetrachlorofluorescein (6-TET), 5-carboxy-X-rhodamine (5-ROX) and 6-carboxy-X-rhodamine (6-ROX); and

II) measuring a change in the fluorescence spectrum.

[0029] The present invention further aims to provide a method for preparing a nucleic acid incorporating the following nucleotides on the same strand:

i) a nucleotide having a 6-substituted 9H-purin-9-yl group as a base; and

ii) a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is attached either directly or through a linker to a fluorescent dye selected from the group consisting of 5-carboxyfluorescein (5-FAM), 6-carboxyfluorescein (6-FAM), 5-carboxytetramethylrhodamine (5-TAMRA), 6-carboxytetramethylrhodamine (6-TAMRA), 5-dimethylaminonaphthalene-1-sulfonic acid (DANSYL), 5-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (5-HEX), 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (6-HEX), 5-carboxy-2',4,7,7'-tetrachlorofluorescein (5-TET), 6-carboxy-2',4,7,7'-tetrachlorofluorescein (6-TET), 5-carboxy-X-rhodamine (5-ROX) and 6-carboxy-X-rhodamine (6-ROX). The preparation method of the present invention comprises effecting transcription, reverse transcription or replication by using, as a template, a nucleic acid incorporating the

following nucleotides:

iii) a nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base; and

iv) a nucleotide having a 6-substituted 9H-purin-9-yl group as a base,

so that the nucleotide of i) is incorporated at a site complementary to the nucleotide of iii), while the nucleotide of ii) is incorporated at a site complementary to the nucleotide of iv).

[0030] The present invention further provides such a nucleic acid prepared by the above method.

[0031] The present invention also provides a method for detecting a stem-loop (hairpin) structure formation of the nucleic acid of the present invention incorporating, on the same strand, a nucleotide having a 6-substituted 9H-purin-9-yl group as a base and a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base. The detection method of the present invention comprises:

1) inducing hybridization between the nucleotide of i) and the nucleotide of ii) in the nucleic acid of the present invention; and

2) measuring a change in the fluorescence spectrum.

MEANS FOR SOLVING THE PROBLEMS

[0032] To solve the problems stated above, the inventors of the present invention have studied base pairs constructed by combining several unnatural bases which had been developed by themselves. As a result, the inventors have found that a combination between 2-amino-6-(2-thienyl)purin-9-yl (s) (Non-

patent Documents 13 and 14) and 2-formyl-1H-pyrrol-1-yl (Pa) (Non-patent Document 9) (Figure 1A) is highly selective and efficient during transcription. Namely, the inventors have found that when this Pa nucleotide is inserted into a DNA template and used for RNA transcription reaction, s is selectively introduced in high yield into the generated RNA at a site complementary to Pa in the template DNA. The inventors have thereby arrived at the present invention.

[0033] By the action of T7 RNA polymerase, the substrate s can be inserted in a site-specific manner into RNA generated from a DNA template at a site complementary to Pa in the template. The purine base s has substituents allowing hydrogen bonding (i.e., a hydrogen acceptor at the 1-position and a hydrogen donor amino group at the 2-position). However, Pa does not have such a hydrophilic group on the pairing surface. The inventors of the present invention have found that oxygen in the formyl group of the Pa base does not serve as a hydrogen acceptor during base pairing because it is directed toward the minor groove of double-stranded DNA and interacts with polymerase (Non-patent Document 9). In the present invention, the s-Pa pair has no significant hydrogen bonding interaction, but Pa would have shape complementarity with s because it is a 5-membered ring base.

[0034] Nucleic acids of the present invention based on artificial base pairing

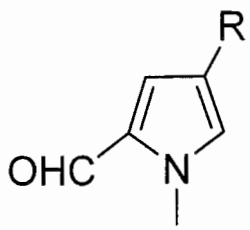
Thus, in one embodiment, the present invention provides a nucleic acid in which a nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base forms a

base pair with a nucleotide having a 6-substituted 9H-purin-9-yl group as a base.

[0035] As used herein, the term "nucleoside" is intended to mean a glycoside compound formed through glycosidic linking between a nucleic acid base and a reducing group of a sugar. It should be noted that the term "nucleic acid base" is intended to encompass adenine, guanine, cytosine, thymine, uracil, and also derivatives thereof. The type of "derivative" is not limited in any way. Specific examples include bases equivalent to a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group and bases equivalent to a 6-substituted 9H-purin-9-yl group. The term "nucleotide" refers to a compound in which the sugar moiety of the above nucleoside forms an ester with phosphoric acid, more preferably a mono-, di- or tri-phosphate ester. The sugar moiety of such a nucleoside or nucleotide may be ribofuranosyl, 2'-deoxyribofuranosyl, or 2'-substituted ribofuranosyl having a substituent (e.g., halogen) at the 2'-position. Likewise, the phosphoric acid moiety may be thiophosphoric acid. Namely, the sugar and phosphoric acid moieties may be in the same form as found in known nucleosides, nucleotides, or derivatives thereof. A ribonucleotide whose sugar moiety is ribofuranosyl can be used as a component of RNA, while a deoxyribonucleotide whose sugar moiety is deoxyribofuranosyl can be used as a component of DNA.

[0036] The "substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group" of the present invention has a structure represented by the following general formula:

[0037] [Formula 3]



[0038] wherein R at the 4-position of the pyrrole ring is hydrogen or may be substituted with a substituent selected from a substituted or unsubstituted C_1-C_3 alkyl group, a substituted or unsubstituted C_2-C_3 alkenyl group, or a substituted or unsubstituted C_2-C_3 alkynyl group. Such an alkyl, alkenyl or alkynyl group may further be substituted with one or more groups independently selected from the group consisting of a lower alkyl group, a halogen group, a hydroxyl group, an amino group, an alkylamino group and an aromatic heterocyclic ring. The substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group of the present invention is herein referred to as "Pa" or "Pa analog," depending on the context.

[0039] Without being limited thereto, the substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group mentioned above may be selected from the group consisting of:

A1) a 2-formyl-1H-pyrrol-1-yl group (Pa);

A2) a 2-formyl-4-(1-propyn-1-yl)-1H-pyrrol-1-yl group;

A3) a 2-formyl-4-methyl-1H-pyrrol-1-yl group; and

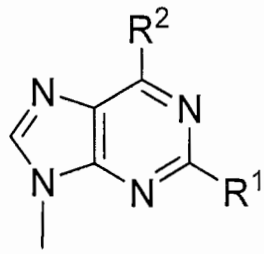
A4) a 2-formyl-4-ethynyl-1H-pyrrol-1-yl group. More preferred is a 2-formyl-1H-pyrrol-1-yl group.

[0040] A nucleoside or nucleotide having the "substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group" ("Pa" or "Pa

analog") of the present invention may be synthesized in a known manner. Starting materials (e.g., pyrrole-2-carboxaldehyde) can be purchased from, for example, Aldrich [1003-29-8] or Merck [807574]. Likewise, Pa derivatives may be synthesized by being derived from Pa, in principle. For example, a derivative having propyne introduced at the 4-position of Pa can be found in Bioorg. Med. Chem. Lett., 13, p. 4515-4518 (2003) (Non-patent Document 28).

[0041] The "6-substituted 9H-purin-9-yl group" of the present invention has a structure represented by the following general formula:

[0042] [Formula 4]



[0043] [wherein

R^1 is hydrogen or an amino group, and

R^2 is a substituted or unsubstituted 2-thienyl or 2-thiazolyl group].

The thienyl or thiazolyl group listed as R^2 may be unsubstituted or may be substituted at the 4- and/or 5-position(s) with one or more groups independently selected from the group consisting of a methyl group, an amino group, a nitro group and a hydroxy group. Among members of the 6-substituted 9H-purin-9-yl group of the present invention, those in which R^2 is a substituted or unsubstituted 2-thienyl group are herein referred to as "s" or "s analog," depending

on the context. Among members of the 6-substituted 9H-purin-9-yl group of the present invention, those in which R^2 is a substituted or unsubstituted 2-thiazolyl group are herein referred to as "v" or "v analog," depending on the context.

[0044] Without being limited thereto, the 6-substituted 9H-purin-9-yl group mentioned above may be selected from the group consisting of:

B1) a 2-amino-6-(2-thienyl)-9H-purin-9-yl group (s);

B2) a 6-(2-thienyl)-9H-purin-9-yl group (s');

B3) a 2-amino-6-(4-methyl-2-thienyl)-9H-purin-9-yl group;

B4) a 6-(4-methyl-2-thienyl)-9H-purin-9-yl group;

B5) a 2-amino-6-(5-methyl-2-thienyl)-9H-purin-9-yl group;

B6) a 6-(5-methyl-2-thienyl)-9H-purin-9-yl group;

B7) a 2-amino-6-(2-thiazolyl)-9H-purin-9-yl group (v);

B8) a 6-(2-thiazolyl)-9H-purin-9-yl group;

B9) a 2-amino-6-(4-methyl-2-thiazolyl)-9H-purin-9-yl group;

B10) a 6-(4-methyl-2-thiazolyl)-9H-purin-9-yl group;

B11) a 2-amino-6-(5-methyl-2-thiazolyl)-9H-purin-9-yl group; and

B12) a 6-(5-methyl-2-thiazolyl)-9H-purin-9-yl group.

More preferred is a 2-amino-6-(2-thienyl)-9H-purin-9-yl group.

[0045] Among the above bases, B1) is "s" and B2) to B6) are "s analogs." Likewise, B7) is "v" and B8) to B12) are "v analogs."

[0046] In an even more preferred embodiment, the above substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group is a 2-formyl-1H-pyrrol-1-yl group, and the above 6-substituted 9H-purin-9-yl group is a 2-amino-6-(2-thienyl)-9H-purin-9-yl group.

[0047] The "6-substituted 9H-purin-9-yl group" of the present invention and a nucleoside or nucleotide containing the same may be synthesized in a known manner. More specifically, Example 1 described later discloses procedures for synthesis of 2-amino-6-(2-thienyl)-9-(1-β-D-ribofuranosyl)purine 5'-triphosphate (sTP) from 2-N-phenoxyacetyl-6-(2-thienyl)-9-(2,3-di-O-acetyl-1-β-D-ribofuranosyl)purine, by way of example.

[0048] Synthesis of s can be found, for example, in the second paragraph on the second page of Bioorg. Med. Chem. Lett., 11, p. 2221-2223 (2001) (Non-patent Document 29). Likewise, synthesis of v is disclosed, for example, in Paragraphs 0026-0027 and Figures 5-6 of Patent Document 3, or alternatively, in the right paragraph on the second page of J. Am. Chem. Soc., 127, p. 8652-8658 (2005) (Non-patent Document 30).

[0049] In addition to s or v, nucleosides or nucleotides having other members of the "6-substituted 9H-purin-9-yl group" of the present invention may be synthesized in the same manner as used for synthesis of s or v. For example, o having furan attached thereto may also be synthesized in the same manner as used for synthesis of s (Non-patent Document 29).

[0050] The present invention provides a nucleic acid in which a nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base forms a base pair with a nucleotide having a 6-substituted 9H-purin-9-yl group as a base. As used herein, the term "nucleic acid" is intended to mean a molecule of a nucleic acid strand in which more than one nucleotide is linked in the direction of 5'→3'. The nucleic acid of the present invention encompasses single-stranded or double-stranded RNA or DNA. The double-stranded nucleic acid may be DNA/DNA, RNA/RNA, or DNA/RNA. DNA also includes cDNA obtained by reverse transcription using RNA as a template. Alternatively, the nucleic acid may form a triplex, a quadruplex, etc.

[0051] With the aim of further expansion of nucleic acid functions, the inventors of the present invention have attempted to design nucleosides or nucleotides having unnatural bases. Embodiments of newly developed artificial base pairs include a base pair between a nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base (e.g., Pa) and a nucleotide having a 6-substituted 9H-purin-9-yl group as a base (e.g., s) (Figure 1). This enables in particular the specific incorporation of s into RNA during T7 RNA polymerase-mediated transcription using a DNA template containing Pa.

[0052] Although there is no significant hydrogen bonding interaction between s and Pa (non-hydrogen-bonded s-Pa pair), the efficiency and selectivity of s-Pa pairing during transcription is as high as that of natural base pairing.

This s-Pa pair shows a higher efficiency than the previously developed hydrophilic s-z pair. The complementary shapes of s and Pa are fitted to each other, but their shapes differ from those of natural purines and pyrimidines. This specific stereochemical fitting would eliminate non-canonical pairing with natural bases, thereby resulting in high selectivity between s and Pa during transcription. In this way, shape complementarity plays an important role in specific base pairing during transcription and replication.

[0053] In the present invention, a nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base and a nucleotide having a 6-substituted 9H-purin-9-yl group as a base are present in two separate nucleic acid strands and can form a duplex through base pairing. Alternatively, these nucleotides may be present in the same single-stranded nucleic acid. In this case, such a single strand may form a loop structure through base pairing.

[0054] The nucleotide of the present invention which has a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group or a 6-substituted 9H-purin-9-yl group as a base can be incorporated into nucleic acids such as DNA or RNA through transcription, replication or reverse transcription reaction. Alternatively, the nucleotide of the present invention may be incorporated into DNA or RNA through chemical synthesis, as in the case of nucleosides or nucleotides having natural bases.

[0055] These transcription, replication and reverse transcription reactions may be accomplished according to known

techniques. Without being limited thereto, for example, it is possible to use T7 RNA polymerase (Takara or other suppliers) for transcription, Klenow fragment (KF) for replication, and AMV Reverse Transcriptase XL (AMV-RT, Life Science) for reverse transcription. In order to avoid removal of the nucleotide of the present invention during the reaction, the replication may also be accomplished, for example, by using Taq DNA polymerase (Takara Taq™) lacking 3′→5′ exonuclease activity to effect PCR amplification of template DNA with a primer containing a nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base.

[0056] Without being limited thereto, in one embodiment, the nucleic acid of the present invention forms a base pair(s) in the step of transcription, reverse transcription, replication or translation of the nucleic acid. In a case where the nucleic acid of the present invention forms a base pair(s) in the transcription step, the above nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base may be a part of DNA, while the above nucleotide having a 6-substituted 9H-purin-9-yl group as a base may be a part of RNA.

[0057] Method for preparing a nucleic acid

The present invention also aims to provide a method for preparing a nucleic acid incorporating a nucleotide having a 6-substituted 9H-purin-9-yl group as a base. The preparation method of the present invention comprises effecting transcription, reverse transcription or replication by using, as a template, a nucleic acid incorporating a nucleotide

having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base, so that the nucleotide having a 6-substituted 9H-purin-9-yl group is incorporated at a site complementary to the nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group.

[0058] The nucleotide having a 2-formyl-1H-pyrrol-1-yl group and the nucleotide having a 6-substituted 9H-purin-9-yl group are as defined herein above in the section "Nucleic acids of the present invention based on artificial base pairing."

[0059] As shown in Examples 1 and 2 described later, even in a case where two or more nucleotides having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base are located adjacent to each other in a template, transcription reaction can proceed, so that a nucleotide having a 6-substituted 9H-purin-9-yl group as a base is incorporated at a complementary site. Thus, the method of the present invention enables the preparation of conventionally unavailable DNA and RNA in which two or more unnatural s bases are located adjacent to each other. Thus, in one embodiment of the preparation method of the present invention, the template has two or more nucleotides having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base.

[0060] The present invention further aims to provide such a nucleic acid incorporating a nucleotide having a 6-substituted 9H-purin-9-yl group as a base, which is prepared by the above method of the present invention.

[0061] The nucleic acid incorporating the nucleotide(s) of the present invention may be used as tRNA, mRNA, antisense DNA or RNA, a ribozyme or an aptamer. The term "antisense DNA or RNA" refers to DNA or RNA capable of inhibiting the expression of a specific gene. It was named to mean that such DNA or RNA is complementary to the full-length or partial sequence of a target gene sequence (sense strand). Antisense DNA or RNA may be used as a tool for artificial regulation of gene expression. Because of containing unnatural bases, such antisense DNA or RNA incorporating the nucleotide(s) of the present invention can be designed to have a different complementarity to a target when compared to the case of using natural bases only. The term "ribozyme" is a generic name for catalysts composed of RNA. The term "aptamer" refers to an *in vitro*-selected nucleic acid having the ability to bind to a specific molecule such as a protein.

The nucleic acid (DNA or RNA) (e.g., mRNA, synthetic RNA) incorporating the nucleotide(s) of the present invention may also encode all or part of a protein or peptide. The nucleic acid of the present invention may be used, e.g., as a gene fragment or a probe. The present invention also encompasses the following embodiments: partial or complete replacement of native genes by the nucleic acids of the present invention; addition of one or more nucleotides of the present invention to native genes; or combinations thereof.

[0062] Furthermore, the nucleic acids of the present invention incorporating nucleotides having unnatural bases may also be used in RNA interference (RNAi). RNA interference

is a phenomenon in which double-stranded RNA (dsRNA) induces mRNA degradation in a sequence-specific manner and hence inhibits gene expression. In a typical example of RNA interference, dsRNA is processed by Dicer belonging to the RNaseIII family into siRNA (short interfering RNA) of approximately 21 to 23 bases in length, which has a 3'-terminal overhang of approximately 2 bases. siRNA is associated into an siRNA-protein complex called RISC and induces mRNA degradation in a sequence-specific manner. RNA interference is shown to be a phenomenon conserved among a wide range of organism species including mammals (e.g., human, mouse), nematodes, plants, drosophila and fungi. The nucleic acids of the present invention incorporating nucleotides having unnatural bases can be used as siRNA in RNA interference or as a part of mRNA to be degraded.

[0063] Moreover, it is also possible to design a new codon containing the nucleotide of the present invention. As described above, even in a case where two or more nucleotides having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base (e.g., Pa) are located adjacent to each other in a template, transcription reaction can proceed, so that a nucleotide having a 6-substituted 9H-purin-9-yl group as a base (e.g., s) is incorporated at a complementary site. Thus, the method of the present invention enables the preparation of conventionally unavailable DNA and RNA in which two or more unnatural s bases are located adjacent to each other. In turn, it is also possible to design a codon containing three s (sss), those containing two s (e.g., ssA, Gss, sGs) and

those containing one s (e.g., sAG, CsT, AGs). Such a new codon may encode either a natural amino acid or an unnatural amino acid. In this way, the present invention not only provides novel unnatural artificial bases, but also enables the design of entirely new genetic codes by designing new codons containing the nucleotides of the present invention, thus providing a world of new genetic codes.

[0064] Further, by designing tRNA systems corresponding to the new codons of the present invention, it is also possible to design a new protein synthesis system in which numerous amino acids can be used. Amino acids which can be used are not limited in any way as long as they can be used in the protein-synthesizing enzyme system in ribosomes. Thus, the present invention provides a new protein synthesis system using the above codons of the present invention. According to the protein synthesis system of the present invention, when a nucleic acid corresponding to a codon at a desired site is efficiently replaced by the nucleic acid of the present invention or when the nucleic acid of the present invention is introduced at a desired site in a codon, it is possible to produce a protein containing a desired unnatural amino acid(s).

[0065] Method for analyzing stereostructure

A nucleotide having a 6-substituted 9H-purin-9-yl group as a base (e.g., s) is a fluorescent base, and a nucleotide containing the same as a base has strong fluorescence properties.

[0066] More specifically, purine derivatives modified to

have an expanded electron conjugated system by attaching an aromatic heterocyclic ring or the like at the 6-position are more likely to develop fluorescence properties. Moreover, without being limited thereto, it is previously known that 2-aminopurine has fluorescence properties. Thus, without being limited thereto, purine derivatives having an aromatic heterocyclic ring at the 6-position and an amino group at the 2-position particularly have strong fluorescence properties. In particular, s, s' and v are preferred fluorescent bases. In contrast to conventional techniques, if the fluorescent dye-labeled fifth base can be introduced by transcription at a specific site in a nucleic acid through artificial base pairing, it significantly facilitates the labeling of nucleic acids and hence the analysis and detection of nucleic acids. The present invention achieves these goals by using an artificial base pair, Pa-s.

[0067] The inventors of the present invention fluorescently labeled yeast $tRNA^{Phe}$ in a site-specific manner by using the fluorescent artificial base s as a probe to obtain characteristic fluorescent profiles, depending on the labeling sites, temperature and Mg^{2+} concentration (Example 3). Moreover, the inventors compared these fluorescent profiles with the UV profiles of the same s-labeled yeast $tRNA^{Phe}$. As a result, the following findings were obtained.

[0068] First, the melting temperature (Tm value) of each tRNA transcript containing s, obtained from changes in its UV absorption spectrum, was as high as the melting temperature (Tm value) of the natural tRNA transcript, obtained from

changes in its UV absorption spectrum. This suggests that the replacement of s at these positions does not significantly destabilize the global tRNA structure.

[0069] Second, each tRNA containing s at a specific position showed characteristic fluorescence intensity changes reflecting the local structural features. Namely, in the higher order structure of tRNA, the fluorescence intensity was weak when s stacked with its neighboring bases, whereas the fluorescence intensity increased with decrease in the degree of stacking. This discovery indicates that when s is introduced at a specific position of RNA and then measured, e.g., for temperature-induced changes in its fluorescence intensity, it is possible to know the stacking state of s in the higher order structure. Thus, as an application of s-containing RNA, there is provided a novel method for analyzing the higher order structure of the RNA.

[0070] More specifically, the s fluorescent profiles of yeast $tRNA^{Phe}$ in the presence of Mg^{2+} (2 and 5 mM) were clearly divided into two groups. The s fluorescence intensity emitted by Group 1 (tRNA 16s, 17s and 47s) at low temperature was 1.7- to 3.4-fold larger than that of Group 2 (RNA 36s, 57s and 59s). The large intensity emitted by Group 1 decreased when the temperature increased, whereas the intensity of Group 2 increased with increase in temperature. These results suggest that the s bases in Group 1 are exposed to the solution at physiological temperature, and hence the non-specific interaction between s and other bases is increased upon denaturation of the folded structure with increasing

temperature. In contrast, the bases in Group 2 stack at low temperature with their neighboring bases in the folded structure, and the base stacking is gradually denatured with increasing temperature. This presumption is entirely consistent with the conformations of the original bases at each site in the crystal structure of tRNA (Figure 4) (Non-patent Document 20). Moreover, the fluorescence intensity of Group 1 is dramatically increased by addition of Mg^{2+} at physiological temperature. This suggests that tRNA forms an active L-shaped structure in the presence of Mg^{2+}, whereby the bases at these positions are kept outside.

[0071] Third, the Tm values obtained from the fluorescent profiles reflected the stability of each local structure in the tRNA structure, in comparison with the Tm values obtained from the UV melting profiles.

[0072] By way of example, the Tm value obtained from the fluorescent profile of tRNA 36s was higher than that obtained from the UV profile. This suggests that the anticodon stem-loop has a higher stability when compared to the stability of the entire tRNA structure. In contrast, the low stability of tRNA 16s and 17s (low Tm values obtained from the fluorescent profiles) suggests that the partial structure including the D-loop may be a fragile region within the tRNA structure. Thus, when the Tm value of each s-labeled product obtained from its fluorescent profile is compared to that obtained from the UV melting profile, it is possible to determine a degree of stability for each s-labeled local structure.

[0073] In this way, site-specific fluorescent probing by

using s labels provides a promising tool for studying local structural features of three-dimensional RNA molecules.

[0074] Thus, the present invention aims to provide a method for analyzing the stereostructure of a local region in the nucleic acid of the present invention, wherein the local region includes a nucleotide having a 6-substituted 9H-purin-9-yl group as a base. The analysis method of the present invention comprises measuring the fluorescence intensity from the base in the nucleotide having a 6-substituted 9H-purin-9-yl group as a base, under different environmental conditions such as varying temperatures or ion concentrations (e.g., magnesium).

[0075] In one embodiment of the analysis method of the present invention, a determination is made that the nucleotide having a 6-substituted 9H-purin-9-yl group as a base stacks with its conformationally neighboring nucleotides at *in vivo* temperature if the fluorescence intensity from the base in the nucleotide is substantially increased with increase in temperature, or alternatively, a determination is made that the nucleotide is exposed outside at *in vivo* temperature if the fluorescence intensity is substantially decreased or not increased with increase in temperature.

[0076] The term "*in vivo* temperature" is intended to mean a body temperature when each organism is alive under natural conditions. This temperature varies depending on the type of organisms and ranges from about 35.0°C to about 39.0°C for humans, about 36.7°C to about 39.3°C for cattle, about 35.0°C to about 39.0°C for mice, about 37.9°C to about 39.9°C for

dogs, and about 38.7°C to about 39.8°C for pigs, by way of example.

[0077] The expression "nucleotide stacks with its conformationally neighboring nucleotides" is more specifically intended to mean a state where two or more bases stack with each other, preferably within about 5Å (stacking).

[0078] The expression "nucleotide is exposed outside" is more specifically intended to mean a state where the base moiety of the nucleotide does not stack with any other base (preferably located at a distance of about 5Å or more) or a state where the base moiety is sufficiently apart from a quencher base.

[0079] The fluorescence intensity is preferably measured at a specific wavelength where the maximum fluorescence properties are obtained depending on the type of the above 6-substituted 9H-purin-9-yl group. For example, the B1) 2-amino-6-(2-thienyl)-9H-purin-9-yl group (s) shows a fluorescence wavelength with its center at 434 nm when excited at a maximum excitation wavelength of 352 nm. Similarly, the B2) 6-(2-thienyl)-9H-purin-9-yl group (s') shows a fluorescence wavelength with its center at 381 nm when excited at 325 nm, while the B7) 2-amino-6-(2-thiazolyl)-9H-purin-9-yl group (v) shows a fluorescence wavelength with its center at 461 nm when excited at 363 nm.

[0080] The expression "fluorescence intensity is substantially increased with increase in temperature" is intended to mean that the fluorescence intensity can be confirmed to increase (e.g., by 1.1-fold or more) when the

temperature is increased from 20°C up to 90°C, by way of example. A greater increase in the fluorescence intensity indicates that the fluorescent base stacks with a larger number of conformationally neighboring nucleotides and/or in closer proximity at *in vivo* temperature.

[0081] The expression "if the fluorescence intensity is substantially decreased or not increased with increase in temperature" is intended to mean that the fluorescence intensity can be confirmed to remain unchanged or to decrease (e.g., by 10% or more) when the temperature is increased from 20°C up to 90°C, by way of example. A greater decrease in the fluorescence intensity indicates that the fluorescent base exists in more isolation and hence the nucleic acid structure around this base is less likely to take a specific structure at *in vivo* temperature.

[0082] Method for detecting the formation of a nucleic acid duplex

The present invention further provides a method for detecting the formation of a nucleic acid duplex, which is another use embodiment of the nucleic acid incorporating a nucleotide having a 6-substituted 9H-purin-9-yl group as a base, which is prepared by the inventive method for preparing a nucleic acid. The detection method of the present invention comprises:

I) inducing hybridization between the following nucleic acids:

i) a nucleic acid incorporating a nucleotide having a 6-substituted 9H-purin-9-yl group as a base, which is prepared

by the inventive method for preparing a nucleic acid, and

ii) a nucleic acid incorporating a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is attached either directly or through a linker to a fluorescent dye selected from the group consisting of 5-carboxyfluorescein (5-FAM), 6-carboxyfluorescein (6-FAM), 5-carboxytetramethylrhodamine (5-TAMRA), 6-carboxytetramethylrhodamine (6-TAMRA), 5-dimethylaminonaphthalene-1-sulfonic acid (DANSYL), 5-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (5-HEX), 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (6-HEX), 5-carboxy-2',4,7,7'-tetrachlorofluorescein (5-TET), 6-carboxy-2',4,7,7'-tetrachlorofluorescein (6-TET), 5-carboxy-X-rhodamine (5-ROX) and 6-carboxy-X-rhodamine (6-ROX); and

II) measuring a change in the fluorescence spectrum.

[0083] The inventive method for detecting the formation of a nucleic acid duplex is based on fluorescence spectral changes induced by hybridization between a nucleic acid incorporating a nucleotide having a 6-substituted 9H-purin-9-yl group as a base and a nucleic acid incorporating a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is substituted with a fluorescent dye. As described in Patent Documents 1 and 2, Non-patent Document 30 or elsewhere, a 6-substituted 9H-purin-9-yl group (e.g., s or v) can form base pairing with a 2-oxo(1H)pyridin-3-yl group (y). As to synthesis of nucleosides or nucleotides having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, details are given in other sections.

[0084] When the fluorescence from the 6-substituted 9H-purin-9-yl group and the fluorescence from the 5-substituted-2-oxo(1H)pyridin-3-yl group substituted with a fluorescent dye having different properties are present in the respective two nucleic acid strands, the nucleic acid duplex formation will cause fluorescence resonance energy transfer (FRET) if these fluorescences are in physical proximity to each other, thus producing a change in the fluorescence spectrum. The detection method of the present invention detects the formation of a nucleic acid duplex by measuring a FRET-induced change in the fluorescence spectrum.

[0085] The term "fluorescence resonance energy transfer (FRET)" refers to a phenomenon in which the excitation energy is transferred by resonance from one fluorescent molecule to another molecule. A molecule providing energy is referred to as a donor, and a molecule receiving the energy is referred to as an acceptor. Once FRET has occurred, a donor losing its energy returns to the ground state, while an acceptor receiving the energy enters an excited state. Thus, the fluorescence of the donor will become weak and, if the acceptor is a fluorescent molecule, its fluorescence will be observed. If the acceptor is a quencher molecule, the fluorescence observed for the donor alone will not be observed as a result of FRET. Standard techniques for FRET-mediated protein detection and nucleic acid detection are known, including a method for detecting a target protein by introducing two FRET-inducing dyes into RNA aptamers (Jhaveri et al., 2000) and a method for detecting a complementary

nucleic acid strand with a hairpin structure (Tyagi et al., 1996). For other information on FRET, see Walter et al., 2001 and Klostermeier et al., 2001.

[0086] To cause FRET, the following three requirements should be satisfied. i) The donor's fluorescence spectrum and the acceptor's absorption spectrum should overlap with each other. The overlapping area of their spectra is desirably larger, but they do not necessarily have to completely overlap with each other. The donor's fluorescence spectrum and the acceptor's absorption spectrum preferably overlap in 30% or more, more preferably in 50% or more range. ii) The donor and acceptor should be in a closer physical proximity to each other. The distance at which FRET occurs with 50% probability is recognized to be 3 nm to 6 nm, and the efficiency of FRET sensitively varies in response to a change in this distance. For example, once the nucleic acids have formed a duplex, the nucleotide having a fluorescent 6-substituted 9H-purin-9-yl group as a base is approached by the nucleotide having, as a base, a 5-substituted-2-oxo(1H)pyridin-3-yl group substituted with a fluorescent dye having different properties. If the distance between these nucleotides is preferably 10 nm or less, more preferably 6 nm or less, and most preferably 3 nm or less, it is possible to detect a change in the fluorescence spectrum. iii) The donor and acceptor should be in proper relative orientation to each other.

[0087] In one embodiment of the detection method of the present invention, the nucleic acid incorporating a

nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base is a nucleic acid incorporating a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is attached to 5-carboxyfluorescein (5-FAM) or 6-carboxyfluorescein (6-FAM) through a linker.

[0088] As described above, for example, s has a maximum excitation wavelength of 353 nm and shows a wavelength with its center at 434 nm when excited at a maximum excitation wavelength of 352 nm. On the other hand, FAM has an absorption peak wavelength of 493 nm and a fluorescence peak wavelength of 522 nm. When excited at a wavelength of 353 nm, an s-containing nucleotide will show a wavelength with its center at 434 nm. In contrast, a FAM-substituted nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base (e.g., FAM-y) absorbs the fluorescence from s upon nearing s, and will show a fluorescence spectrum having a fluorescence peak wavelength of 522 nm. Namely, if excitation with a wavelength of 353 nm causes a greater decrease in the 434 nm fluorescence spectrum and a greater increase in the 522 nm fluorescence spectrum, it can be detected that FAM-y is in closer proximity to s, i.e., a nucleic acid duplex is formed through hybridization between s-containing nucleic acid and FAM-y-containing nucleic acid.

[0089] Nucleosides or nucleotides having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base

In nucleosides or nucleotides having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, a fluorescent dye is attached either directly or through a linker to the 5-

position. These labeled nucleosides or nucleotides are detailed in a preceding patent application filed by the inventors of the present invention (Japanese Patent Application No. 2004-324271, filed on November 8, 2004, not yet published).

[0090] As a fluorescent dye, any known dye may be used and is preferably selected from the group consisting of 5-carboxyfluorescein (5-FAM), 6-carboxyfluorescein (6-FAM), 5-carboxytetramethylrhodamine (5-TAMRA), 6-carboxytetramethylrhodamine (6-TAMRA), 5-dimethylaminonaphthalene-1-sulfonic acid (DANSYL), 5-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (5-HEX), 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (6-HEX), 5-carboxy-2',4,7,7'-tetrachlorofluorescein (5-TET), 6-carboxy-2',4,7,7'-tetrachlorofluorescein (6-TET), 5-carboxy-X-rhodamine (5-ROX) and 6-carboxy-X-rhodamine (6-ROX). In general, fluorescein and rhodamine are expressed in both open-ring and spiro forms. These fluorescent dyes are detailed in, e.g., Tuschl et al., 1994; Misra et al., 2004; Gibson et al., 1996; and Chehab et al., 1989.

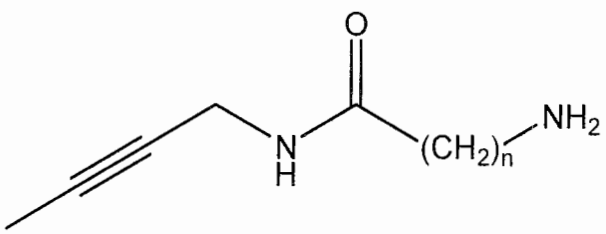
[0091] For example, FAM has an absorption peak wavelength of 493 nm and a fluorescence peak wavelength of 522 nm. Likewise, TAMRA has an absorption peak wavelength of 553 nm and a fluorescence peak wavelength of 578 nm. DANSYL has an absorption peak wavelength of 335 nm and a fluorescence peak wavelength of 518 nm. HEX has an absorption peak wavelength of 535 nm and a fluorescence peak wavelength of 556 nm. TET has an absorption peak wavelength of 521 nm and a

fluorescence peak wavelength of 536 nm. 5-ROX has an absorption peak wavelength of 567 nm and a fluorescence peak wavelength of 591 nm. 6-ROX has an absorption peak wavelength of 570 nm and a fluorescence peak wavelength of 590 nm.

[0092] Since such a nucleoside or nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base has a fluorescent molecule as a substituent at the 5-position, a nucleic acid may be detected in a manner dependent on the type of fluorescent molecules. Thus, it can be used as a labeled nucleic acid probe to detect substances interacting with the nucleic acid. Moreover, since these individual fluorescent dyes have fluorescent colors different from each other, they can also be used in multiple staining.

[0093] The fluorescent dye may be attached either directly or through a linker to the 5-position of the 2-oxo(1H)pyridin-3-yl group. The type of linker is not limited in any way and may be determined as appropriate by those skilled in the art. Without being limited thereto, the linker is preferably selected from the group consisting of chemical formulae I to III shown below:

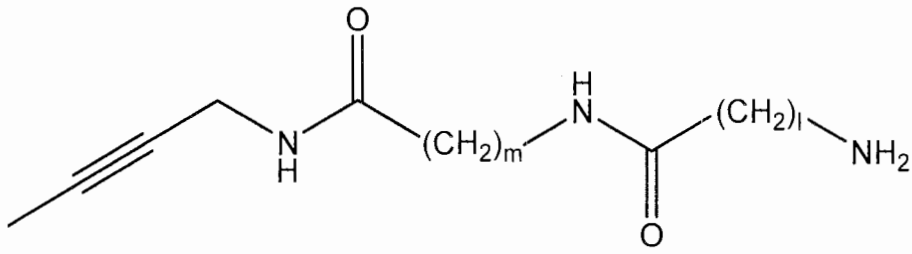
[0094] [Formula 5]



I

[0095] [wherein n is selected from integers of 1 to 5];

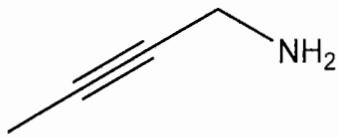
[0096] [Formula 6]



II

[0097] [wherein m and l are each independently selected from integers of 1 to 5]; and

[0098] [Formula 7]



III.

[0099] Nucleosides or nucleotides having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base may be synthesized in any manner. 3-(β-D-Ribofuranosyl)-pyridin-2(1H)-one may first be introduced with a substituent at its 5-position and then with triphosphate. Alternatively, 3-(β-D-ribofuranosyl)-pyridin-2(1H)-one may first be introduced with triphosphate and then with a substituent. In particular, when a large group is introduced, a group such as an aminopropynyl group may first be introduced at the 5-position, to which an activated substituent may then be attached.

[0100] Alternatively, such a nucleotide having a 2-oxo(1H)pyridin-3-yl group as a base may be modified with a fluorescent dye not only during synthesis of the nucleotide per se, but also after synthesis of a nucleic acid containing the nucleotide through transcription or the like.

[0101] A 5-substituted-2-oxo(1H)pyridin-3-yl group forms two hydrogen bonds with a 6-substituted 9H-purin-9-yl group. The 5-substituted-2-oxo(1H)pyridin-3-yl group cannot form any base pair with natural purine bases A (adenine) and G (guanine) in terms of its stereostructure. Likewise, the 6-substituted 9H-purin-9-yl group cannot form any base pair with natural T (thymine), U (uracil) and C (cytosine) due to steric hindrance. Thus, the 6-substituted 9H-purin-9-yl group (e.g., s or v) can specifically form a base pair with the 5-substituted-2-oxo(1H)pyridin-3-yl group.

[0102] A nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base can be incorporated into nucleic acids such as DNA or RNA through transcription, replication or reverse transcription reaction. More specifically, when a nucleic acid containing a nucleotide having a 6-substituted 9H-purin-9-yl group as a base, for example, s- or v-containing template DNA is used to effect transcription reaction with T7 RNA polymerase, a substrate of the 2-oxo(1H)pyridin-3-yl group (e.g., yTP) is incorporated into RNA in a site-specific manner, opposite s or v in the template DNA. Since y can be chemically modified at its 5-position, various functional y derivatives can be incorporated into RNA at any specific site.

[0103] Alternatively, such a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base may also be incorporated into DNA or RNA through chemical synthesis, as in the case of nucleosides or nucleotides having natural bases.

[0104] Method for preparing a nucleic acid incorporating

two artificial bases on the same strand

The present invention further aims to provide a method for preparing a nucleic acid incorporating the following nucleotides on the same strand:

i) a nucleotide having a 6-substituted 9H-purin-9-yl group as a base; and

ii) a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is attached either directly or through a linker to a fluorescent dye selected from the group consisting of 5-carboxyfluorescein (5-FAM), 6-carboxyfluorescein (6-FAM), 5-carboxytetramethylrhodamine (5-TAMRA), 6-carboxytetramethylrhodamine (6-TAMRA), 5-dimethylaminonaphthalene-1-sulfonic acid (DANSYL), 5-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (5-HEX), 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (6-HEX), 5-carboxy-2',4,7,7'-tetrachlorofluorescein (5-TET), 6-carboxy-2',4,7,7'-tetrachlorofluorescein (6-TET), 5-carboxy-X-rhodamine (5-ROX) and 6-carboxy-X-rhodamine (6-ROX). The preparation method of the present invention comprises effecting transcription, reverse transcription or replication by using, as a template, a nucleic acid incorporating the following nucleotides:

iii) a nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base; and

iv) a nucleotide having a 6-substituted 9H-purin-9-yl group as a base,

so that the nucleotide of i) is incorporated at a site

complementary to the nucleotide of iii), while the nucleotide of ii) is incorporated at a site complementary to the nucleotide of iv).

[0105] The above preparation method of the present invention is based on incorporation of a nucleotide having a 6-substituted 9H-purin-9-yl group as a base when a nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base is used as a template, in combination with incorporation of a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base when a nucleotide having a 6-substituted 9H-purin-9-yl group as a base is used as a template.

[0106] The present invention further provides such a nucleic acid prepared by the above method, which incorporates the following nucleotides on the same strand:

i) a nucleotide having a 6-substituted 9H-purin-9-yl group as a base; and

ii) a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is attached either directly or through a linker to a fluorescent dye selected from the group consisting of 5-carboxyfluorescein (5-FAM), 6-carboxyfluorescein (6-FAM), 5-carboxytetramethylrhodamine (5-TAMRA), 6-carboxytetramethylrhodamine (6-TAMRA), 5-dimethylaminonaphthalene-1-sulfonic acid (DANSYL), 5-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (5-HEX), 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (6-HEX), 5-carboxy-2',4,7,7'-tetrachlorofluorescein (5-TET), 6-carboxy-

2',4,7,7'-tetrachlorofluorescein (6-TET), 5-carboxy-X-rhodamine (5-ROX) and 6-carboxy-X-rhodamine (6-ROX).

[0107] The above nucleic acid of the present invention may be used as tRNA, mRNA, antisense DNA or RNA, a ribozyme, or an aptamer. Moreover, it may also be used in RNA interference (RNAi).

[0108] In Example 4 described later, Pa- and v-containing DNA was used as a template to effect transcription, thereby preparing RNA incorporating both s and y or both s and FAM-y on the same strand.

[0109] Method for detecting a stem-loop (hairpin) structure formation of nucleic acids

The present invention also provides a method for detecting a stem-loop (hairpin) structure formation of the nucleic acid of the present invention incorporating, on the same strand, a nucleotide having a 6-substituted 9H-purin-9-yl group as a base and a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base. The detection method of the present invention comprises:

1) inducing hybridization between the nucleotide of i) and the nucleotide of ii) in the nucleic acid of the present invention; and

2) measuring a change in the fluorescence spectrum.

[0110] The method for detecting a stem-loop (hairpin) structure formation of the nucleic acid of the present invention is based on FERT as described herein above in the section "Method for detecting the formation of a nucleic acid duplex." Namely, when the fluorescence from the 6-substituted

9H-purin-9-yl group and the fluorescence from the 5-substituted-2-oxo(1H)pyridin-3-yl group substituted with a fluorescent dye having different properties are present "in the same molecule," the stem-loop (hairpin) formation will cause FRET if these fluorescences are in physical proximity to each other, thus producing a change in the fluorescence spectrum. The detection method of the present invention detects the formation of a stem-loop (hairpin) structure in a nucleic acid by measuring a FRET-induced change in the fluorescence spectrum.

[0111] A FRET-induced change in the fluorescence spectrum to detect the formation of a stem-loop (hairpin) structure in a nucleic acid may be observed in the same manner as used for detecting a change in the fluorescence spectrum for two separate nucleic acid strands which form a duplex.

[0112] In one preferred embodiment of the present invention, the nucleic acid of the present invention which incorporates a nucleotide having a 6-substituted 9H-purin-9-yl group as a base and a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base on the same strand makes a stem-loop (hairpin) structure, by the self-complementation between s and FAM-y in its molecule. In this way, by measuring the FRET efficiency between s and FAM-y as probes in the nucleic acid molecule, the higher order structure (e.g., stem-loop structure) of a nucleic acid can be deduced.

[0113] Moreover, if a nucleic acid molecule is designed to form a stem-loop (hairpin) structure only when another target molecule is attached thereto, it is possible to create a

system for detecting the target molecule based on fluorescence spectral changes. Thus, the present invention is available for use in systems for real-time detection of intermolecular interactions, systems for real-time detection of intramolecular conformational changes, molecular beacons, or intermolecular distance estimation based on FRET efficiency, etc. Such a target molecule includes DNA, RNA, and a protein.

[0114] In Example 4 described later, FRET between s and FAM-y was studied for RNA incorporating s and FAM-y on the same strand. In Example 4 described later, when s was excited at 350 nm, FRET was observed and the fluorescence from FAM-y appeared around 521 nm in both cases where a base pair was formed between s and FAM-y through stem-loop formation and where s was apart from FAM-y by addition of a complementary RNA strand to form a double-stranded structure between these molecules. However, when the s fluorescence intensity of a fragment introduced with s and FAM-y (i.e., the intensity around 440 nm in the spectra represented by solid lines in Figures 12b and 12c) was compared to the s fluorescence intensity of a control fragment introduced with s and y (i.e., the intensity around 440 nm in the spectra represented by broken lines in Figures 12b and 12c), the s fluorescence intensity was found to significantly decrease in a case where a stem-loop structure was formed. The distance between s and FAM-y in this stem-loop structure is 8-10 Å, while the distance between s and FAM-y is increased to 65-67 Å in a double-stranded structure. Thus, a closer distance between s

and FAM-y provides an increased FRET efficiency between them.

[0115] Nucleic acids incorporating two or more adjacent nucleotides having a 6-substituted 9H-purin-9-yl group as a base

Nucleic acids incorporating two or more adjacent nucleotides having a 6-substituted 9H-purin-9-yl group as a base also fall within the scope of the present invention.

[0116] In the above nucleic acids, since two or more nucleotides having a 6-substituted 9H-purin-9-yl group as a base are located adjacent to each other on the same strand, self-quenching occurs between these adjacent artificial bases to cause a significant decrease in the fluorescence intensity of the 9H-purin-9-yl groups, as compared to a case where nucleotides having a 9H-purin-9-yl group as a base are not located adjacent to each other (see Figures 13c and 13d). On the other hand, even where self-quenching occurs between 9H-purin-9-yl groups, hybridization with a 5-substituted-2-oxo(1H)pyridin-3-yl group whose 5-position is substituted with a fluorescent dye permits FRET between a 9H-purin-9-yl group and the fluorescent dye (see Figures 14c and 14d). Moreover, in the nucleic acids incorporating two or more adjacent nucleotides having a 6-substituted 9H-purin-9-yl group as a base, the fluorescence intensity of the 9H-purin-9-yl groups is decreased, i.e., the background fluorescence is low, thus allowing efficient observation of FRET with a fluorescent dye (see Figure 15). Namely, the FRET system of the present invention which uses a self-quenched probe containing two or more adjacent nucleotides having a 6-

substituted 9H-purin-9-yl group as a base can keep the background of the fluorescent donor at a low level and is therefore particularly useful in allowing detection of complementary strands containing a fluorescent acceptor even in the presence of excessive donor probe.

[0117] Such a nucleic acid incorporating two or more adjacent nucleotides having a 6-substituted 9H-purin-9-yl group as a base is available for use in the above method for detecting the formation of a nucleic acid duplex or the above method for detecting a stem-loop or hairpin structure formation of a nucleic acid.

[0118] In such a nucleic acid incorporating two or more adjacent nucleotides having a 6-substituted 9H-purin-9-yl group as a base, there is no particular limitation on the number of adjacent nucleotides having a 6-substituted 9H-purin-9-yl group as a base as long as it is two or more. The number of these adjacent nucleotides is preferably 2 to 10, 2 to 5, 2 to 3, and most preferably 2.

[0119] Such a nucleic acid incorporating two or more adjacent nucleotides having a 6-substituted 9H-purin-9-yl group as a base may be prepared, but not limited to, by the inventive method for preparing a nucleic acid or through known chemical synthesis procedures.

EFFECTS OF THE INVENTION

[0120] According to the present invention, the discovery of new base pairing between unnatural bases Pa and s has suggested that shape complementarity between paired bases is also important during transcription (Non-patent Documents 5-

6). This enables the incorporation of a nucleotide having a 6-substituted 9H-purin-9-yl group as a base during transcription, although such incorporation has previously been difficult to achieve during transcription.

[0121] The method of the present invention achieves significantly higher yields than other methods previously known for obtaining RNA having s at a desired position(s): (1) chemical RNA synthesis; (2) transcription reaction using template DNA having a y nucleotide(s); and (3) transcription reaction using template DNA having a nucleotide(s) having a 2-oxo-1,3-dihydro-imidazol-1-yl group (z).

[0122] Conventional fluorescent probes such as commonly-used 2-aminopurine nucleotides (Non-patent Documents 21-23) cannot be incorporated through transcription into RNA in a site-specific manner, and hence they are particularly difficult to use in long-chain RNA molecules. Namely, in the prior art methods, a nucleotide derivative of 2-aminopurine is introduced into RNA through chemical synthesis. However, long-chain RNA (50 nucleotides or longer) is difficult to synthesize through chemical RNA synthesis, and the operations required are also complicated and time-consuming. Thus, the prior art methods were of limited application.

[0123] In contrast, the present invention enables the site-specific incorporation of a nucleotide having a fluorescent 6-substituted 9H-purin-9-yl group as a base into RNA. Although DNA serving as a template is required to be chemically synthesized in the present invention, chemical synthesis of DNA allows synthesis of longer molecules (around

150 nucleotides) when compared to RNA synthesis, and the chemically synthesized DNA molecules can also be linked together by the action of enzymes. This particularly enables the analysis of long-chain RNA, which is impossible to achieve by conventional fluorescence probing techniques. This may provide a powerful tool for studying the dynamics of local conformational changes at a defined position within a large RNA molecule. In Example 3 of the present invention, a local higher order structure (stacking) within tRNA was actually analyzed.

[0124] Moreover, RNA molecules have now been found to have many new functions (e.g., RNA interference) in the body, and the development of drugs based on such phenomena has also proceeded. However, not many techniques exist for studying *in vivo* behavior of these RNA molecules. Thus, this fluorescence probing technique provides a good method for *in vivo* analysis of RNA and for studying *in vivo* kinetics of RNA drugs.

[0125] Furthermore, the s-Pa base pair newly found by the present invention enables the site-specific incorporation of amino acid analogs into proteins through translation, and hence can be put to practical use in expanding the genetic code (Non-patent Document 15).

BRIEF DESCRIPTION OF THE DRAWINGS

[0126] [Figure 1] Figure 1 shows unnatural s-Pa (A) and s-z (B) base pairs.

[Figure 2] Figure 2 shows an experimental scheme for T7 transcription mediated by s-Pa base pairing, along with the results of gel electrophoresis on the transcripts.

[0127] Figure 2A: Experimental scheme

Figure 2B: Gel electrophoresis of transcripts obtained by using template DNA containing one or two Pa or z bases in the presence of natural NTPs (1 mM) and sTP (1 mM). The transcripts were labeled with [γ-^{32}P]GTP at their 5'-end. The relative yield of each transcript was determined in comparison with the yield of a transcript from a template composed only of natural bases, and each yield was averaged from 3 to 4 data sets.

[Figure 3] Figure 3 shows 2D-TLC analysis of the labeled ribonucleoside 3'-monophosphates obtained from nuclease digestion of the transcripts (17-mer). The transcripts were internally labeled with [α-^{32}P]NTP, and two-dimensional TLC was obtained from the 17-mer fragments transcribed in the presence of 1 mM sTP. $N_1N_2N_3$ in template DNA is PaAC in the middle panel, while $N_1N_2N_3$ is CAC in the right panel.

[Figure 4] Figure 4 shows the introduction sites for s in yeast $tRNA^{Phe}$, along with the structure of the tRNA (Non-patent Document 20). Figure 4 shows the secondary structure of the original tRNA transcript. The positions replaced by s are circled. Each wavy line represents base-base interaction in the tertiary structure. Although the boxed G-C pair was changed from the original C-G pair, this mutation does not significantly alter the original tRNA structure (Non-patent Document 26).

[Figure 5] Figure 5 shows the introduction sites for s in yeast $tRNA^{Phe}$, along with the structure of the tRNA (Non-patent Document 20). Figure 5 shows the tertiary structure of

the original tRNA transcript.

[Figure 6] Figures 6A to 6C each show a magnified view of a region covering the introduction sites for s in yeast $tRNA^{Phe}$. The dark-colored bases are at sites for replacement by s and stack with light-colored bases. Spheres (yellow) each represent Mg^{2+}. Although bases shown in the figures are modified bases, the bases in parentheses correspond to those in the original transcript.

[Figure 7] Figure 7A shows the excitation and fluorescence spectra of s ribonucleoside (1 μM). The solid line with solid circles represents the fluorescence spectrum with excitation at 352 nm. The wavy line with open circles represents the fluorescence spectrum with excitation at 434 nm.

[0128] Figures 7B and 7C show the fluorescence spectra of tRNA 47s (1 μM) and tRNA 57s (1 μM), respectively, when excited at 352 nm. Each fluorescence spectrum was measured at 20°C in 50 mM sodium cacodylate (pH 7.2), 50 mM KCl and 0.1 mM EDTA. In (B), the solid line with solid circles, the wavy line with open circles and the solid line with crosses represent the fluorescence spectra of tRNA47s in the presence of 0 mM Mg^{2+}, 2 mM Mg^{2+} and 5 mM Mg^{2+}, respectively. In (C), the solid line with solid circles, the wavy line with open circles and the solid line with crosses represent the fluorescence spectra of tRNA57s in the presence of 0 mM Mg^{2+}, 2 mM Mg^{2+} and 5 mM Mg^{2+}, respectively.

[Figure 8] Figure 8 shows (A) the temperature dependence of the fluorescence intensity of s ribonucleoside, and (B) an

example showing the fluorescence intensity of tRNA 36s corrected relative to s ribonucleoside. The fluorescence intensity of each tRNA transcript was corrected according to the equation: Yct = Yt/(Rt/R20), wherein Yct represents the corrected intensity of each tRNA at t°C, Yt represents the measured fluorescence intensity of each tRNA at t°C, Rt represents the observed fluorescence intensity of s ribonucleoside at t°C, and R20 represents the observed fluorescence intensity of s ribonucleoside at 20°C.

[Figure 9] Figure 9 shows the melting curves of tRNAs containing s at various positions, obtained from changes in their fluorescence intensity at 434 nm (solid lines) and obtained from changes in their UV absorption at 260 nm (wavy lines). The thick (black), middle (blue) and thin (red) solid lines represent melting curves obtained from changes in the fluorescence intensity at 434 nm in the presence of 0 mM Mg^{2+}, 2 mM Mg^{2+} and 5 mM Mg^{2+}, respectively. The thick (black), middle (blue) and thin (red) wavy lines represent melting curves obtained from changes in the UV absorption at 260 nm in the presence of 0 mM Mg^{2+}, 2 mM Mg^{2+} and 5 mM Mg^{2+}, respectively.

[Figure 10] Figure 10 shows procedures for site-specific introduction of y and s into RNA through transcription using v-y and s-Pa base pairing, along with the results of polyacrylamide electrophoresis on the transcripts. In Figure 10a, P in the sequences represents an artificial base Pa.

[Figure 11] Figure 11 shows the results of nucleotide composition analysis on the transcripts (Bhp2-1) prepared

through transcription using v-y and s-Pa base pairing. Quantification of each spot in Figure 11b is summarized in Table 4.

[Figure 12] Figure 12 shows the results of fluorimetry on RNA (Bhp2-1) introduced with artificial bases s and y or FAM-y. Figure 12b shows the fluorescence spectrum of Bhp2-1 which forms a duplex with its complementary RNA. Figure 12c shows the results of Bhp2-1 which remains in a single-stranded state to thereby form a hairpin structure. In each of Figures 12b and 12c, the solid line represents the results of Bhp2-1 having FAM-y at position 13 and s at position 36, while the wavy line represents the results of Bhp2-1 having y at position 13 and s at position 36.

[Figure 13] Figure 13a shows the fluorescence (flu, solid line) and absorption (abs, broken line) spectra of s deoxyribonucleoside (thin line) and FAM-yTP (thick line) in phosphate buffer (pH 7.0). Figure 13b shows a scheme for FRET experiment. Solid circles and open circles represent their low fluorescence state and their high fluorescence state, respectively, when unnatural base s or v was excited (λ_{ex}). Figures 13c and 13d show the steady-state fluorescence emission spectra of single-stranded DNAs containing s base(s) (c: DNA35s1, DNA35s2a and DNA35s2b) and single-stranded DNAs containing v base(s) (c: DNA35v1, DNA35v2a and DNA35v2b).

[Figure 14] Figure 14 shows the steady-state fluorescence emission spectra of DNA/RNA duplexes. 35-mer DNAs (a: DNA35s1; b: DNA35v1; c: DNA35s2a; d: DNA35v2a; e: DNA35s2b; and f: DNA35v2b) were each hybridized with 17-mer

RNA containing FAM-y (open circles and solid circles) or FAM-hx-y (open triangles or solid triangles). The open circles and open triangles represent the spectra at 20°C, while the solid circles and solid triangles represent the spectra at 70°C.

[Figure 15] Figure 15 shows the steady-state fluorescence emission spectra of DNA/RNA duplexes when using excessive DNA probes containing artificial bases. 35-mer DNAs (1 μM) (a: DNA35s1; b: DNA35v1; c: DNA35s2a; and d: DNA35v2a) were each hybridized with 17-mer RNA (100 nM) containing FAM-y (open circles and solid circles) or FAM-hx-y (open triangles and solid circles). The open circles and open triangles represent the spectra at 20°C, while the solid circles and solid triangles represent the spectra at 70°C.

EXAMPLES

[0129] The present invention will now be further described in the following examples, which are not intended to limit the technical scope of the invention. Based on the detailed description, various changes and modifications will be apparent to those skilled in the art, and such changes and modifications fall within the technical scope of the invention.

[0130] Example 1: T7 transcription through s-Pa base pairing

In this example, T7 transcription was performed through s-Pa base pairing.

[0131] 1) Reagents, solvents and others

All reagents and solvents were purchased from usual suppliers and were used without further purification.

Electrospray-ionization mass spectra (ESI-MS) were recorded on a Waters ZMD 4000 LC/MS system. DNA templates were chemically synthesized using phosphoramides of Pa (Non-patent Document 9) and natural bases in an automated DNA synthesizer (model 392, PerkinElmer Applied Biosystems, Foster City, CA). Gel electrophoresis was used for purification of oligonucleotides.

[0132] 2) Synthesis of 2-amino-6-(2-thienyl)-9-(1-β-D-ribofuranosyl)purine 5'-triphosphate (sTP)

2-N-Phenoxyacetyl-6-(2-thienyl)-9-(2,3-di-O-acetyl-1-β-D-ribofuranosyl)purine (57 mg, 0.1 mmol) (Non-patent Document 29) was dissolved in pyridine and then evaporated in vacuo to remove the solvent together with water contained therein. The residue was dissolved in pyridine (100 μl) and dioxane (300 μl), followed by addition of a 1 M dioxane solution (100 μl) of 2-chloro-4H-1,2,3-benzodioxaphosphorin-4-one (0.11 mmol) (Non-patent Document 25). After 10 minutes, this solution was quickly mixed with a 0.5 M DMF solution (300 μl) of bis(tributylammonium)pyrophosphoric acid (0.15 mmol) and with tri-n-butylamine (100 μl). The reaction mixture was stirred at room temperature for 10 minutes, followed by addition of a 1% iodine solution (in pyridine/water (98:2, v/v), 2 ml). After 15 minutes, the reaction solution was supplemented with a 5% aqueous $NaHSO_3$ solution (150 μl) and then with water (5 ml).

[0133] The solution was stirred at room temperature for 30 minutes, followed by addition of concentrated aqueous ammonia (20 ml). Ammonolysis was performed at 55°C for 12 hours. The

solution was concentrated in vacuo, and the product was purified by DEAE sephadex (A-25) column chromatography (eluted with a linear gradient of 50 mM to 1 M TEAB) and then by C18-HPLC (eluted with a gradient of 0% to 30% CH_3CN in 100 mM triethylammonium acetate using a Gilson HPLC purification system with a Synchropak RPP (Eichorom Technologies)) to obtain sTP.

[0134] [Formula 8]

^1H NMR (270 MHz, D_2O) δ 1.13 (t, 27H, J=7.3 Hz), 3.05 (q, 18H, J=7.3 and 14.8 Hz), 4.19 (m, 2H), 4.32 (m, 1H), 4.56 (m, 1H), 4.83 (m, 1H), 5.94 (d, 1H, J=5.9 Hz), 7.19 (dd, 1H, J=4.0 and 4.9 Hz), 7.66 (d, 1H, J=4.9 Hz), 8.16 (d, 1H, J=4.0 Hz), 8.33 (s, 1H).

^{31}P NMR (109 MHz, D_2O) δ -22.38 (t, 1H, J=19.5 and 20.1 Hz), -10.68 (d, 1H, J=19.5 Hz), -9.24 (d, 1H, J=20.1 Hz).

ESI-MS for $C_{14}H_{17}N_5O_{13}P_3S$ (M-H): calcd, 587.98; found, 587.70.

[0135] 3) Preparation of templates for T7 transcription

Chemically synthesized DNA templates (10 μM 35-mer coding strand and 21-mer non-coding strand for 17-mer RNA synthesis) were each annealed in 10 mM Tris-HCl (pH 7.6) buffer containing 10 mM NaCl by heating at 95°C and then slow cooling to 4°C.

[0136] 4) T7 transcription (17-mer RNA)

Transcription was performed in a 40 mM Tris-HCl (pH 8.0) buffer solution (20 μl) containing 24 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT and 0.01% Triton X-100, in the presence of 1 mM natural NTPs or 1 mM sTP, 2 μCi [γ-^{32}P]GTP, 2 μM

template and 50 units of T7 RNA polymerase (Takara, Kyoto). The use of [γ-^{32}P]GTP allowed determination of yields because the transcripts were labeled only at their 5'-end. After incubation at 37°C for 3 hours, a dye solution (20 μl) containing 10 M urea and 0.05% BPB was added to stop the reaction. The mixtures were each heated at 75°C for 3 hours and the products were analyzed on a 20% polyacrylamide-7 M urea gel.

[0137] 5) Results

In this example, transcription was accomplished by using s (sTP) as a substrate and by using DNA templates containing one or two Pa bases such that the unnatural base(s) is/are located at a complementary position(s) corresponding to positions 13 to 15 of the transcript. Figure 2A shows the experimental scheme. In Figure 2A, the sequences of 10 μM 35-mer coding strand and 21-mer non-coding strand for 17-mer RNA synthesis, as well as the nucleotide sequence of the 17-mer RNA transcript are as shown below.

[0138] Template complementary strand (SEQ ID NO: 1)

5'-ataatacgactcactataggg-3'

Template strand (SEQ ID NO: 2)

3'-tattatgctgagtgatatccctcgaagggannntc-5'

17-mer RNA transcript (SEQ ID NO: 3)

5'-gggagcuucccunnnag-3'

The inventors of the present invention also studied additional DNA templates containing imidazophosphorin-2-one (z) (Figure 1B) as controls (Figure 1B); z is a good template base for site-specific insertion of s into RNA through T7

transcription, but its transcription efficiency is low when compared to natural transcription. Although the s-z pair may form two hydrogen bonds between bases, its configuration is similar to that of the s-Pa pair.

[0139] These templates were used to effect transcription for 3 hours, and the transcripts whose 5′-end was labeled with [γ-^{32}P]GTP were analyzed on a gel. The results obtained are shown in Figure 2B. More specifically, Figure 2B shows the results of gel electrophoresis on the products transcribed using the templates containing one or two Pa or z bases in the presence of natural NTPs (1 mM, N = A, G, C, U) and sTP (1 mM). The transcripts are labeled with [γ-^{32}P]GTP at their 5′-end. The relative yield of each transcript was determined in comparison with the yield of a transcript from a template composed only of natural bases, which was set to 100%. Each yield was averaged from 3 to 4 data sets.

[0140] As shown in Figure 2B, the full-length (17-mer) transcript (containing one s base) obtained from template DNA containing one Pa base was found to have a relative yield of 92% (Figure 2B, Lane 1, N_1=Pa), which was higher than the yield of transcription from template DNA containing one z base (35%, Figure 2B, Lane 4), and was as high as the yield of transcription using a template composed only of natural bases (N_1=C) and NTPs (Figure 2B, Lane 8). In the case of transcription using a template containing two Pa bases, the relative yield was low when compared to natural transcription (Figure 2B, Lane 8), but full-length products were obtained (Figure 2B, Lanes 2 and 3). In contrast, in the case of

transcription using a template containing two z bases, no full-length transcript was obtained (Figure 2B, Lanes 5 and 6).

[0141] Example 2: Composition analysis of nucleotides in T7 transcripts (17-mer RNAs)

In this example, to study the selectivity of s-Pa base pair during transcription, composition analysis was performed on nucleotides in the full-length (17-mer) transcripts obtained by transcription from Pa, z and natural templates. The transcripts were each internally labeled with any of [α-^{32}P]UTP, [α-^{32}P]ATP or [α-^{32}P]GTP, and these transcripts were completely digested with RNase T_2 into a nucleoside 3′-phosphate form. The resulting labeled nucleotides were analyzed by 2D-TLC (Figure 3), and the spots were quantified for each nucleoside 3′-monophosphate (Table 1).

[0142] More specifically, transcription was performed in 40 mM Tris-HCl (pH 8.0) buffer (20 μl) containing 24 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT and 0.01% Triton X-100, in the presence of 10 mM GMP, 1 mM natural NTPs (1 mM, N = A, G, C, U), 0, 1 or 3 mM sTP, 2 μCi [γ-^{32}P]UTP, [γ-^{32}P]ATP or [γ-^{32}P]GTP (Amersham), 2 μM template DNA, and 50 units of T7 RNA polymerase (Takara) (Non-patent Documents 14 and 15). After incubation for 3 hours at 37°C, a dye solution was added to stop the transcription.

[0143] This mixture was heated at 75°C for 3 hours and then electrophoresed on a 15% polyacrylamide-7 M urea gel. The full-length product was eluted from the gel and 0.05 A_{260} units of *E. coli* tRNA was then added thereto, followed by

ethanol precipitation. The transcript was digested at 37°C for 2 hours with 0.075 U/μl RNase T_2 in 15 mM sodium acetate buffer (pH 4.5). By this treatment, the transcript was completely digested into nucleoside 3'-monophosphates by the action of RNase T_2. The digestion products were analyzed by 2D-TLC using a Merck HPTLC plate (100 × 100 mm, Merck, Darmstadt, Germany). The developing solvents used for the first and second dimensions were isobutyric acid/NH_4OH/H_2O (66:1:33 v/v/v) and isopropyl alcohol/HCl/H_2O (70:15:15 v/v/v), respectively. The products on the TLC plate were analyzed with a bio-imaging analyzer. Each spot was quantified by averaging from 3 to 9 data sets.

[0144] The results obtained are shown in Figure 3. Further, the quantification results for each nucleoside 3'-monophosphate are summarized in Table 1.

[0145] Table 1: Nucleotide composition analysis of T7 transcripts

[0146] [Table 1]

Entry No.	Template $N_1N_2N_3$	[α-^{32}P] NTP	sTP (mM)	Composition of nucleotides incorporated at 5' neighbor of U, A or G[a]				
				Ap	Gp	Cp	Up	sp
1	**Pa**AC	UTP	1	0.01[b] [0][c] (0.01)[d]	n.d. [0] (<0.01)	1.99 [2] (0.05)	1.02 [1] (0.03)	0.97 [1] (0.03)
2	**z**AC	UTP	1	0.01 [0] (<0.01)	0.01 [0] (0.01)	1.99 [2] (0.02)	1.02 [1] (0.05)	0.97 [1] (0.05)
3	CAC	UTP	0	0.01 [0] (<0.01)	0.97 [1] (0.03)	2.02 [2] (0.02)	1.01 [1] (0.03)	n.d.[e] [0] (-)
4	CAC	UTP	1	0.01 [0] (<0.01)	0.98 [1] (0.02)	2.03 [2] (0.03)	0.99 [1] (0.04)	n.d. [0] (-)
5	TAC	UTP	0	1.00 [1] (0.03)	0.01 [0] (<0.01)	2.02 [2] (0.01)	0.97 [1] (0.04)	n.d. [0] (-)
6	TAC	UTP	1	1.00 [1] (0.03)	0.01 [0] (<0.01)	2.01[2] (0.01)	0.97 [1] (0.03)	0.01 [0] (<0.01)
7	**Pa**A**Pa**	UTP	1	0.01 [0] (0.01)	0.01 [0] (<0.01)	2.03 [2] (0.03)	1.00 [1] (0.02)	0.95 [1] (0.03)
8	**Pa**A**Pa**	UTP	3	0.01 [0] (0.01)	0.01 [0] (0.01)	2.01 [2] (0.01)	0.99 [1] (0.01)	0.98 [1] (0.01)
9	**Pa**A**Pa**	ATP	1	0.05 [0] (<0.01)	1.06 [1] (0.07)	n.d. [0] (-)	0.01 [0] (<0.01)	0.87 [1] (0.07)
10	**Pa**A**Pa**	ATP	3	0.04 [0] (0.01)	1.02 [1] (0.06)	n.d. [0] (-)	0.01 [0] (<0.01)	0.93 [1] (0.07)
11	CAC	ATP	1	0.03 [0] (<0.01)	1.95 [2] (0.01)	n.d. [0] (-)	0.02 [0] (0.01)	n.d. [0] (-)
12	CAC	ATP	3	0.03 [0] (<0.01)	1.94 [2] (0.01)	n.d. [0] (-)	0.02 [0] (0.01)	n.d. [0] (-)
13	**Pa**Pa**C**	GTP	1	1.99 [2] (0.05)	1.15 [1] (0.05)	0.02 [0] (0.02)	n.d. [0] (-)	0.84 [1] (0.03)
14	**Pa**Pa**C**	GTP	3	1.94 [2] (0.03)	1.13 [1] (0.07)	0.02 [0] (0.01)	0.01 [0] (0.02)	0.89 [1] (0.10)
15	CAC	GTP	1	1.98 [2] (0.06)	1.01 [1] (0.03)	0.01 [1] (0.01)	1.98 [2] (0.08)	0.01 [1] (0.01)
16	CAC	GTP	3	1.99 [2] (0.04)	0.99 [1] (0.03)	0.01 [0] (0.01)	1.99 [2] (0.08)	0.02 [1] (0.01)

[0147] a: Composition of nucleotides incorporated at the 5'-side of U (Entry Nos. 1-8), A (Entry Nos. 9-12) or G (Entry Nos. 13-16) in the transcripts shown in Figure 2

b: The values were determined by using the following equation.

(Radioactivity of each nucleotide)/[All nucleotides (3'-monophosphates)] × (Total number of nucleotides at 5'-side of [α-^{32}P]NTP]

c: The theoretical number of each nucleotide is shown in square brackets.

[0148] d: The standard deviation is shown in parentheses.

[0149] e: Not detected

The results shown in Figure 3 and Table 1 confirmed that s was incorporated into RNA with high selectivity when using the template containing one Pa base. In Figure 3, only when using the Pa-containing template and labeling with [γ-^{32}P]UTP, the labeled s-nucleoside 3'-monophosphate was observed on 2D-TLC. Moreover, as shown in Table 1, the selectivity (97%) of s incorporation into RNA opposite Pa in the template DNA (Table 1, Entry No. 1) was as high as the selectivity through s-z base pairing (Table 1, Entry No. 2) or the selectivity of transcription through natural base pairing (Table 1, Entry Nos. 3 and 5), and there was no misincorporation of sTP opposite natural bases in the template (Table 1, Entry Nos. 4 and 6). In the case of transcription using a template containing two Pa bases, the selectivity of s incorporation at the first position was sufficiently high (95%), while the selectivity of s incorporation at the second position was about 84-87%. However, this selectivity was improved somewhat (89-93%) by increasing the sTP concentration (3 mM). Thus, Pa is superior to z as a template base for use in site-specific incorporation of s into RNA, and this result suggests that non-hydrogen-bonded base pairs may also be functional during transcription.

[0150] Example 3: Preparation of 16s, 17s, 36s, 47s, 57s and 59s tRNAs and their fluorescence spectra and UV melting curves

1) Preparation of 16s, 17s, 36s, 47s, 57s and 59s tRNAs

To prove the usefulness of the s base as a fluorescent probe, the inventors of the present invention incorporated s at a specific site in yeast $tRNA^{Phe}$ and then analyzed the fluorescence properties of s at each site in the tRNA under various conditions to thereby study the higher order structure of the tRNA. Since the fluorescence of the s base is quenched by stacking with its neighboring bases, information is obtained on the local structure around s in RNA molecules.

[0151] The inventors of the present invention introduced s at the following 6 sites in the tRNA, separately: position 16 or 17 in the D-loop (tRNA 16s or 17s), position 36 in the anticodon (tRNA 36s), position 47 in the extra-loop (tRNA 47s), and position 57 or 59 in the TΨC-loop (tRNA 57s or 59s). s at each site is introduced at a site which is not used for base pairing with other bases in the tRNA (Figure 4).

[0152] The sequences of each template for preparing s-containing yeast $tRNA^{Phe}$ molecules (5 μM 94-mer template strand DNA and its complementary strand DNA for RNA synthesis) are as shown in Table 2 below (SEQ ID NOs: 4-11).

[0153] [Table 2]

1. Sequences of DNA templates for s-containing tRNA transcripts.

Non-template strand (94-mer) for all tRNA transcripts

ATAATACGACTCACTATAGGGGATTTAGCTCAGTTGGGAGAGCGCCAGACTGAAG ATCTGGAGGTCCTGTGTTCGATCCACAGAATTCCCACCA (SEQ ID NO: 4)

Template strand (94-mer) for the original tRNA transcript (Tm=2'-OMe-T, Gm=2'-OMe-G)

TmGmGTGGGAATTCTGTGGATCGAACACAGGACCTCCAGATCTTCAGTCTGGCG
CTCTCCCAACTGAGCTAAATCCCCTATAGTGAGTCGTATTAT (SEQ ID NO: 5)

Template strand (94-mer) for tRNA 16**s** (Tm=2'-OMe-T, Gm=2'-OMe-G)
TmGmGTGGGAATTCTGTGGATCGAACACAGGACCTCCAGATCTTCAGTCTGGCG
CTCTCCCA**Pa**CTGAGCTAAATCCCCTATAGTGAGTCGTATTAT (SEQ ID NO: 6)

Template strand (94-mer) for tRNA 17**s** (Tm=2'-OMe-T, Gm=2'-OMe-G)
TmGmGTGGGAATTCTGTGGATCGAACACAGGACCTCCAGATCTTCAGTCTGGCG
CTCTCCC**Pa**ACTGAGCTAAATCCCCTATAGTGAGTCGTATTAT (SEQ ID NO: 7)

Template strand (94-mer) for tRNA 36**s** (Tm=2'-OMe-T, Gm=2'-OMe-G)
TmGmGTGGGAATTCTGTGGATCGAACACAGGACCTCCAGATC**Pa**TCAGTCTGGC
GCTCTCCCAACTGAGCTAAATCCCCTATAGTGAGTCGTATTAT (SEQ ID NO: 8)

Template strand (94-mer) for tRNA 47**s** (Tm=2'-OMe-T, Gm=2'-OMe-G)
TmGmGTGGGAATTCTGTGGATCGAACACAGG**Pa**CCTCCAGATCTTCAGTCTGGC
GCTCTCCCAACTGAGCTAAATCCCCTATAGTGAGTCGTATTAT (SEQ ID NO: 9)

Template strand (94-mer) for tRNA 57**s** (Tm=2'-OMe-T, Gm=2'-OMe-G)
TmGmGTGGGAATTCTGTGGAT**Pa**GAACACAGGACCTCCAGATCTTCAGTCTGGC
GCTCTCCCAACTGAGCTAAATCCCCTATAGTGAGTCGTATTAT (SEQ ID NO: 10)

Template strand (94-mer) for tRNA 59**s** (Tm=2'-OMe-T, Gm=2'-OMe-G)
TmGmGTGGGAATTCTGTGG**Pa**TCGAACACAGGACCTCCAGATCTTCAGTCTGGC
GCTCTCCCAACTGAGCTAAATCCCCTATAGTGAGTCGTATTAT (SEQ ID NO: 11)

[0154] Sequences of 5 μM 94-mer template DNA and its

complementary strand DNA for RNA synthesis

In these sequences, the C2-G71 base pair is replaced by G2-C71 (Non-patent Document 26). Two nucleosides (G and T) at the 5'-end of each template strand DNA are replaced by 2'-O-methylribonucleosides to thereby prevent template-independent production of transcripts which are longer than the desired product (Non-patent Document 27).

[0155] Preparation of 94-mer template DNA and its complementary strand DNA for RNA synthesis was accomplished in the same manner as described in Example 1, the section "3) Preparation of templates for T7 transcription."

[0156] Figures 4 to 6 show the introduction sites for s in yeast $tRNA^{Phe}$, along with tRNA structures (Non-patent Document 20). Figure 4 shows the secondary structure of the original tRNA. The bases replaced by s are circled. Each wavy line represents base-base interaction. The boxed G-C base pair corresponds to the C-G base pair in the original tRNA, and this replacement does not significantly alter the higher order structure of the tRNA (Non-patent Document 26).

[0157] Figure 5 shows the tertiary structure of the natural tRNA, while Figures 6A to 6C each show a magnified view of each site where s is incorporated. The dark-colored bases are replaced by s and stack with light-colored bases. Spheres (yellow) each represent Mg^{2+} in the tRNA. Although bases shown in the figures are those modified in the natural tRNA, the bases in parentheses represent those in the transcripts.

[0158] tRNA transcripts incorporating s at the respective

sites were each prepared with T7 RNA polymerase using Pa-containing template DNA in the presence of 1 mM sTP and natural NTPs. More specifically, in 40 mM Tris-HCl (pH 8.0) buffer containing 24 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT and 0.01% Triton X-100, transcription was first performed in the presence of 10 mM GMP, 1 mM natural NTPs, 1 mM sTP, 0.5 μM template DNA and 2.5 U/μl T7 RNA polymerase. After incubation at 37°C for 6 hours, 1.75 volumes of a dye solution was added to stop the transcription. This solution was heated at 75°C for 3 minutes and then electrophoresed on a 10% polyacrylamide-7 M urea gel. The full-length product was eluted from the gel and precipitated with ethanol. The product was dissolved in 450 μl of 10 mM EDTA (pH 8) and incubated at 75°C for 5 minutes.

[0159] By using a Microcon YM-10 filter (Amicon), the buffer solution was then replaced by a buffer for Tm measurement (containing 50 mM sodium cacodylate (pH 7.2) and 50 mM KCl). The amount of tRNA was determined by absorbance at 260 nm. For fluorescence and UV melting measurement, a solution was prepared to contain 1 μM tRNA in Tm buffer containing 0.1 mM EDTA, 2 mM $MgCl_2$ or 5 mM $MgCl_2$.

[0160] 2) Fluorescence spectra and UV melting curves

A fluorescence spectrum and a UV melting curve were recorded for each tRNA containing s at a specific site, in the range of 20°C to 90°C at a heating rate of 0.5°C/minute by using an FP-6500 spectrofluorometer (JASCO) and a UV-2450 spectrophotometer (SHIMADZU), respectively.

[0161] The s base has two excitation maxima (at 299 and

352 nm). Figure 7A shows the excitation and fluorescence spectra of s ribonucleoside (1 μM). As is apparent from Figure 7A, s shows a fluorescence spectrum with its center at 434 nm, and the quantum yield of the fluorescence was 0.41 at pH 7.0. Figures 7B and 7C show the fluorescence spectra of tRNA 47s (1 μM) and tRNA 57s (1 μM), respectively, when excited at 352 nm. Each fluorescence spectrum was measured at 20°C in 50 mM sodium cacodylate (pH 7.2), 50 mM KCl and 0.1 mM EDTA.

[0162] In this example, in a solution containing 50 mM sodium cacodylate and 50 mM KCl together with 0.1 mM EDTA or 2 mM or 5 mM $MgCl_2$ instead of EDTA, s (excited at 352 nm with a 3 nm spectral bandwidth) was measured for its fluorescence intensity at 434 nm and UV absorption at 260 nm to study the thermodynamic melting profiles of each tRNA transcript (1 μM).

[0163] Quenching of the s fluorescence by collision with solvent molecules is more likely to occur at a higher temperature (Figure 8A). For this reason, temperature-dependent changes in the fluorescence intensity of each tRNA transcript were normalized by s nucleoside monomer. The melting temperature of each tRNA determined from temperature-dependent changes in its fluorescence intensity was calculated using IGOR Pro software (WaveMetrics. Inc.). More specifically, Figure 8 shows (A) temperature-dependent changes in the fluorescence intensity of s ribonucleoside, and (B) the temperature dependence of the fluorescence intensity of tRNA 36s corrected in consideration of

temperature-dependent changes in the fluorescence intensity of s ribonucleoside. The fluorescence intensity of each tRNA transcript was corrected according to the equation: Yct = Yt/(Rt/R20), wherein Yct represents the corrected intensity of each tRNA at t°C, Yt represents the fluorescence intensity of each tRNA actually measured at t°C, Rt represents the fluorescence intensity of s ribonucleoside at t°C, and R20 represents the fluorescence intensity of s ribonucleoside at 20°C.

[0164] 3) Results

Figure 9 shows the melting curves of 16s, 17s, 36s, 47s, 57s and 59s tRNAs obtained from temperature-dependent changes in their fluorescence intensity at 434 nm (solid lines) and obtained from changes in their UV absorption at 260 nm (wavy lines).

[0165] First, the UV melting temperature of each tRNA transcript containing s (64.6-66.4°C in 2 mM $MgCl_2$) was as high as that of the natural tRNA transcript (65.5°C in 2 mM $MgCl_2$). This suggests that the replacement of s at these positions does not significantly destabilize the tRNA structure.

[0166] Second, each tRNA containing s at a specific position showed characteristic fluorescence intensity changes reflecting the local structural features. The fluorescence properties of s in the presence of Mg^{2+} (2 and 5 mM) are clearly divided into two groups: one includes tRNA 16s, 17s and 47s (Group 1), and the other includes tRNA 36s, 57s and 59s (Group 2). In the presence of 2 mM or 5 mM Mg^{2+}, the

fluorescence intensity of s in Group 1 at low temperature was 1.7- to 3.4-fold larger than that of Group 2. This large fluorescence intensity of s in Group 1 decreased when the temperature increased, whereas the fluorescence intensity of s in Group 2 increased with increase in temperature.

[0167] These results suggest that the s base at position 16, 17 or 47 is exposed to the solution at physiological temperature, and hence the non-specific interaction between s and other bases is increased upon denaturation of tRNA with increasing temperature. In contrast, the s base at position 36, 57 or 59 stacks at low temperature with its neighboring bases in tRNA folded into an L-shaped structure, and the tRNA is gradually denatured with increasing temperature to thereby break the stacking structure with the bases adjacent to s. Such a presumption is entirely consistent with the conformations of the original bases at each site in the L-shaped structure of tRNA, as analyzed by the X-ray crystal structure analysis (Figure 4) (Non-patent Document 20). The fluorescence intensity of tRNA 16s, 17s and 47s is dramatically increased by addition of Mg^{2+} at physiological temperature. This suggests that tRNA forms an L-shaped structure upon binding Mg^{2+} to the tRNA, whereby the bases at these positions are exposed to the solution.

[0168] Third, the Tm values obtained from the fluorescent profiles reflect the stability of each local structure in the tRNA structure, in comparison with the Tm values obtained from the UV melting profiles. Table 3 shows melting temperatures (Tm) obtained from the fluorescent and UV

profiles of yeast $tRNA^{Phe}$ containing s at a specific site.

[0169] Table 3: Melting temperatures (Tm) obtained from fluorescent and UV profiles of yeast $tRNA^{Phe}$ containing s at a specific site

[0170] [Table 3]

	Tm (°C) obtained from fluorescent profiles			Tm (°C) obtained from UV profiles		
	0 mM Mg^{2+}	2 mM Mg^{2+}	5 mM Mg^{2+}	0 mM Mg^{2+}	2 mM Mg^{2+}	5 mM Mg^{2+}
Original tRNA	-	-	-	56.1	65.5	70.1
tRNA 16**s**	n.c.	58.0	64.0	55.6	64.6	69.1
tRNA 17**s**	n.c.	59.6	65.5	55.7	65.6	69.1
tRNA 36**s**	57.0	69.5	74.0	55.2	65.2	69.7
tRNA 47**s**	n.c.	61.1	66.0	55.3	65.2	68.8
tRNA 57**s**	54.5	63.6	68.5	55.8	65.1	69.6
tRNA 59**s**	58.1	67.1	71.5	55.8	66.4	70.3

[0171] The fluorescent and UV profiles were obtained by using each tRNA transcript (1 μM) in 50 mM sodium cacodylate (pH 7.2) and 50 mM KCl with or without $MgCl_2$. Each melting temperature was calculated using IgorPro software (WaveMetrics. Inc.). n.c.: Not calculated.

[0172] By way of example, the Tm value obtained from the fluorescent profile of tRNA 36s (69.5°C in 2 mM $MgCl_2$) was higher than that obtained from the UV profile (65.2°C in 2 mM $MgCl_2$). This suggests that the anticodon stem-loop has a higher stability when compared to the stability of the entire tRNA structure. In contrast, the low stability of tRNA 16s (58.0°C in 2 mM $MgCl_2$) and 17s (59.6°C in 2 mM $MgCl_2$) suggests that the partial structure including the D-loop may be a fragile region within the tRNA structure.

[0173] Example 4: FRET experiment on RNA strands incorporating artificial bases s and FAM-y

The use of s-Pa base pairing during transcription allows efficient introduction of a fluorescent artificial base s into RNA. When this s-Pa base pair is combined with the v-y base pair previously developed by the inventors of the present invention, both s and y or both s and 5-modified y can be introduced into a single strand of RNA at any positions. In particular, the use of y introduced with FAM allows the observation of FRET between s and FAM-y.

[0174] In this example, template DNAs for transcription containing Pa and v at various sites were chemically synthesized, and transcription was effected using these templates and the respective substrates of s and y or FAM-y (sTP, yTP, FAM-yTP).

[0175] T7 RNA polymerase-mediated transcription reaction (Analysis of RNA 39-mer site-specifically incorporating both s and y)

Double-stranded template DNA (5 μM) was annealed in 10 mM Tris-HCl buffer (pH 7.6) containing 10 mM NaCl (temperature conditions: heating at 95°C for 3 minutes, followed by slow cooling to 4°C at a rate of -0.1°C/second). T7 transcription reaction was accomplished in a reaction solution (20 μl in total) containing 40 mM Tris-HCl buffer (pH 8.0), 8 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT, 0.01% Triton X-100, 2 μCi [γ-^{32}P]GTP, 1 mM NTPs, 0-1 mM sTP, 0-1 mM yTP, 0.5 μM template DNA, and 50 units of T7 RNA polymerase (Takara). After reaction at 37°C for 3 hours, an electrophoresis dye solution

containing 10 M urea was added in an equal volume (20 μl) to stop the reaction. This solution was heated at 75°C for 3 minutes and then electrophoresed on a 15% polyacrylamide-7 M urea gel. The transcripts were analyzed with a bio-imaging analyzer (Figure 10).

[0176] Nucleotide composition analysis for products of T7 RNA polymerase-mediated transcription reaction

T7 RNA polymerase-mediated transcription reaction was accomplished in a reaction solution containing 40 mM Tris-HCl buffer (pH 8.0), 8 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT, 0.01% Triton X-100, 2 μCi [α-^{32}P]ATP or 2 μCi [α-^{32}P]GTP or 2 μCi [α-^{32}P]CTP or 2 μCi [α-^{32}P]UTP, 1 mM NTPs, 0-1 mM sTP, 0-1 mM yTP, 0.5 μM template DNA, and 50 units of T7 RNA polymerase (Takara). The transcript was purified by electrophoresis on a 15% polyacrylamide-7 M urea gel, eluted from the gel and then collected by ethanol precipitation in the presence of tRNA (0.05 OD). This sample was dissolved in 10 μl sterilized water, and the resulting solution (8.5 μl) was provided for complete digestion with RNase T_2. To the solution (8.5 μl), RNase T_2 (1.5 μl, 0.5 U/μl) was added and reacted at 37°C for 2 hours. A part of this solution was spotted on a TLC plate (Merck, 10 cm × 10 cm) and developed two-dimensionally. The developing solutions used for the first and second dimensions were isobutyric acid:concentrated aqueous ammonia:water = 66:1:33 (v/v/v) and 2-propanol:hydrochloric acid:water = 65:10:25 (v/v/v), respectively. The developed spots were detected and analyzed with a bio-imaging analyzer (Figure 11, Table 4).

[0177] T7 RNA polymerase-mediated transcription reaction

(Preparation of RNA 39-mer site-specifically incorporating both s and y or both s and FAM-y)

T7 transcription reaction was accomplished in a reaction solution (50 μl in total) containing 40 mM Tris-HCl buffer (pH 8.0), 8 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT, 0.01% Triton X-100, 1 mM NTPs, 1 mM sTP, 1 mM yTP or 1 mM FAM-yTP, 0.5 μM template DNA, and 125 units of T7 RNA polymerase (Takara). After reaction at 37°C for 6 hours, a DNase I solution was added and incubated at 37°C for 15 minutes to digest the template DNA. After an electrophoresis dye solution containing 10 M urea was added in an equal volume (50 μl) to stop the reaction, the reaction solution was heated at 75°C for 3 minutes and then electrophoresed on a 15% polyacrylamide-7 M urea gel to purify the desired transcript 39-mer. After elution with sterilized water from the gel, the transcript was collected by ethanol precipitation and measured for UV absorption to determine the concentration of the desired transcript RNA 39-mer.

[0178] Fluorescence spectrum measurement

The RNA 39-mer site-specifically incorporating both s and y or both s and FAM-y (0.2 μM) was used to measure a fluorescence spectrum at 20°C in 10 mM sodium phosphate buffer (pH 7.0), 100 mM NaCl and 0.1 mM EDTA. In the annealing operation for the RNA sample before measurement, a solution containing the RNA 39-mer (0.2 μM) as well as a solution containing the RNA 39-mer and a complementary RNA 42-mer (0.4 μM) having a sequence complementary to the 39-mer were each

heated at 75°C for 3 minutes and then slowly cooled to 4°C at a rate of -0.1°C/3 seconds. The annealed samples were stored at 4°C. For use in measurement, each sample was transferred to a measuring cell and allowed to stand at 20°C for 2 minutes or longer, followed by starting spectral scanning (360-650 nm). For fluorimetry (excitation wavelength: 350 nm), a 3 × 3 mm cell was used, and the following settings were used to measure each spectrum: response 0.5 s, scanning speed 2000 nm/minute, sensitivity fixed at PMT 500V in manual mode, and slit bandwidth fixed at 5 nm (Figure 12).

[0179] Results

Figure 10 shows strategies for site-specific introduction of y and s into RNA through transcription using v-y and s-Pa base pairing, along with the results of polyacrylamide electrophoresis on the transcripts.

[0180] In more detail, Figure 10a shows the sequences of template DNAs and conditions for transcription reaction (in the sequences, Pa is represented by P). The transcripts (Bhp2-1, Bhp2-2, Bhp2-3) were studied by polyacrylamide gel electrophoresis (Figure 10b), indicating that transcription in the presence of both sTP and yTP resulted in higher amounts of the transcripts. In turn, to study whether s and y were selectively incorporated into RNA, substrates having radioactive phosphorus introduced at the α-position, i.e., [α-^{32}P]ATP, [α-^{32}P]GTP, [α-^{32}P]CTP and [α-^{32}P]UTP were each used separately for transcription, and the radioactively labeled transcripts were hydrolyzed with RNase T_2 into nucleoside 3'-monophosphates, followed by two-dimensional TLC to analyze

nucleoside 3′-monophosphates whose phosphoric acid was radioactively labeled, thereby examining the nucleotide composition of the transcripts (Figure 11).

[0181] In this technique, upon incorporation of a radioactive substrate, a nucleotide located at the 5′-side of the substrate can be detected on two-dimensional TLC. For example, Figure 11b shows the results of two-dimensional TLC on Bhp2-1. In Bhp2-1, y is located at the 5′-side of A and s is located at the 5′-side of G. If y and s are selectively incorporated into RNA opposite v and Pa in the template, respectively, a spot corresponding to yp (i.e., nucleoside 3′-monophosphate of y) will appear on two-dimensional TLC of the transcript labeled with [α-^{32}P]ATP, while a spot corresponding to sp (i.e., nucleoside 3′-monophosphate of s) will appear in the case of the transcript labeled with [α-^{32}P]GTP. In Figure 11b, the intended spot actually appears on each TLC. Moreover, the transcript labeled with [α-^{32}P]CTP or [α-^{32}P]UTP shows no spot corresponding to either yp or sp, indicating that none of these artificial bases is misincorporated opposite natural bases in the template. Further, each spot on two-dimensional TLC was quantified; the results obtained are summarized in Table 4.

[0182] [Table 4]

Nucleotide composition analysis of RNA (Bhp2-1) incorporating s and y

Entry No.	Template	[α-^{32}P] NTP	yTP (mM)	sTP (mM)	Composition of nucleotides incorporated as 5′ neighbor of A or G or C or U[a]					
					Ap	Gp	Cp	Up	sp	yp
1	2-1	ATP	0.5	1	4.90[b] [5][c] (0.03)[d]	2.01 [2] (0.03)	1.01 [1] (0.01)	3.03 [3] (0.04)	**0.03 [0] (0.02)**	**1.01 [1] (0.02)**
2	2-1	ATP	1	1	4.94 [5] (0.06)	2.00 [2] (0.02)	1.01 [1] (0.02)	2.96 [3] (0.05)	**0.03 [0] (0.01)**	**1.05 [1] (0.01)**
3	2-1	GTP	0.5	1	2.05 [2] (0.02)	1.03 [1] (0.02)	1.01 [1] (0.01)	2.95 [3] (0.02)	**0.92 [1] (0.02)**	**0.03 [0] (0.01)**
4	2-1	GTP	1	1	2.00 [2] (0.02)	1.05 [1] (0.04)	1.01 [1] (0.01)	2.95 [3] (0.03)	**0.92 [1] (0.03)**	**0.06 [0] (0.01)**
5	2-1	CTP	0.5	1	1.95 [2] (0.02)	1.99 [2] (0.03)	1.94 [2] (0.02)	1.06 [1] (0.02)	0.05 [0] (0.01)	0.01 [0] (<0.01)
6	2-1	CTP	1	1	1.94 [2] (0.01)	2.01 [2] (0.05)	1.95 [2] (0.01)	1.05 [1] (0.03)	0.04 [0] (0.02)	0.01 [0] (<0.01)
7	2-1	UTP	0.5	1	3.02 [3] (0.06)	1.94 [2] (0.05)	3.00 [3] (0.01)	0.02 [0] (0.01)	0.02 [0] (0.01)	n.d.[e] [0] (-)
8	2-1	UTP	1	1	3.05 [3] (0.05)	1.90 [2] (0.09)	3.01 [3] (0.06)	0.02 [0] (<0.01)	0.02 [0] (0.01)	n.d. [0] (-)
9	cont	ATP	0.5	1	4.94 [5] (0.06)	2.05 [2] (0.06)	1.01 [1] (0.03)	4.90 [5] (0.05)	**0.02 [0] (0.01)**	**0.07 [0] (0.03)**
10	cont	ATP	1	1	4.94 [5] (0.03)	2.05 [2] (0.02)	1.05 [1] (0.02)	4.79 [5] (0.05)	**0.02 [0] (0.01)**	**0.15 [0] (0.01)**
11	cont	ATP	0	0	4.95 [5] (0.03)	2.03 [2] (0.01)	1.03 [1] (0.01)	4.99 [5] (0.05)	n.d. [-] (-)	n.d. [-] (-)
12	cont	GTP	0.5	1	2.97 [3] (0.06)	1.05 [1] (0.03)	1.01 [1] (0.01)	2.92 [3] (0.04)	**0.01 [0] (<0.01)**	**0.03 [0] (<0.01)**
13	cont	GTP	1	1	2.97 [3] (0.03)	1.05 [1] (0.01)	1.03 [1] (0.02)	2.89 [3] (0.03)	**0.01 [0] (<0.01)**	**0.06 [0] (<0.01)**
14	cont	GTP	0	0	2.98 [3] (0.04)	1.04 [1] (0.01)	1.02 [1] (0.01)	2.96 [3] (0.04)	n.d. [-] (-)	n.d. [-] (-)

[0183] a: Composition of nucleotides incorporated at 5′-neighbor of A (Entry Nos. 1, 2 and 9-11), G (Entry Nos. 3, 4 and 12-14), C (Entry Nos. 5 and 6) or U (Entry Nos. 7 and 8) shown in Figure 11

b: The values were determined by using the following equation.

(Radioactivity of each nucleotide)/[All nucleotides (3′-monophosphates)] × (Total number of nucleotides at 5′-neighbor of [α-^{32}P]NTP]

c: The theoretical number of each nucleotide is shown in

square brackets.

[0184] d: The standard deviation is shown in parentheses.

[0185] e: Not detected

The results are shown for the cases where yTP was used in an amount of 0.5 mM or 1 mM during transcription. The results of Entry Nos. 1-8 are all well consistent with the theoretical values (the numerical values in brackets). In Entry Nos. 9-14, which show the results obtained for using template DNA composed only of natural bases, Entry No. 10 indicated that a slight amount of yp was incorporated opposite natural template bases when using 1 mM yTP. Thus, it is preferable to set the amount of yTP used during transcription to 0.5 mM. These results indicate that when transcription is effected using s-Pa base paring in combination with v-y base pairing, it is possible to selectively introduce s and y into a single strand of RNA.

[0186] For further study, the substrate of y was replaced by a substrate modified to have FAM at the 5-position of y (FAM-yTP). This substrate and sTP were used for transcription of Bhp2-1, and the transcript was purified by gel electrophoresis and studied for its fluorescence properties. Bhp2-1 alone forms a hairpin structure, in which s and FAM-y are located at positions allowing base pairing, whereby FRET will be observed between s and FAM-y. Upon addition of a complementary strand of this RNA, a double-stranded structure is formed to give a distance between s and FAM-y, thereby reducing the efficiency of FRET. The fluorescence spectrum obtained for RNA of this double-stranded structure is shown in

Figure 12b, while the fluorescence spectrum obtained for the hairpin-type RNA is shown in Figure 12c (solid lines). Each fluorescence spectrum is compared to the spectrum obtained for RNA incorporating s and y (broken line: this RNA shows no FRET event and hence serves as a control showing the fluorescence spectrum of s). When s was excited at 350 nm, FRET was observed and the fluorescence from FAM-y appeared around 521 nm in both cases. However, the hairpin-type RNA shows a stronger fluorescence intensity of FAM-y and also causes significant quenching of s over the control (broken line).

[0187] This result indicated that FRET was observed in correlation with the distance between s and FAM-y. Thus, when s and FAM-y are introduced at various sites in RNA and measured for the efficiency of FRET between them, the distance between s and FAM-y can be estimated and is helpful in RNA structural analysis. Although conventional analyses such as X-ray crystal structure analysis and NMR structural analysis provided only information about static RNA structures, the technique used in this example makes it possible to know the dynamics of conformational changes in RNA in a solution.

[0188] Example 5: FRET experiment on artificial base pairing between s or v and FAM-y in nucleic acid duplexes

In DNA/RNA duplexes, the FRET experiment was performed between a fluorescent artificial base 2-amino-6-(2-thienyl)purin-9-yl or 2-amino-6-(thiazolyl)purin-9-yl group (hereinafter referred to as s or v, respectively) and a fluorophore-labeled 2-oxo-(1H)pyridin-3-yl group (hereinafter referred to as y). The emission spectrum of s or v overlapped

with the absorption spectrum of a nucleotide derivative which was FAM-labeled 5-(3-amino-1-propynyl)-y (FAM-yTP) or FAM-labeled 5-[3-(6-aminohexanamido)-1-propynyl]-y (FAM-hx-yTP) (Figure 13a). Thus, s or v in DNA fragments serves as a fluorescent donor, while FAM-y or FAM-hx-y in their complementary RNA fragments serves as a fluorescent acceptor.

[0189] Synthesis of DNA containing an artificial base s or v

DNA containing an artificial base s or v was synthesized as described, for example, in WO 01/05801 or WO 2005/026187 by using a cyanoethylphosphoroamidite derivative of a nucleotide containing an artificial base s or v in accordance with the DNA chemical synthesis method well known in the art. The sequence of the chemically synthesized 35-mer DNA is as follows: 5'-CA$N_1N_2N_3$CTCGGGATTCCCTATAGTGAGTCGTATTAT-3' (SEQ ID NO: 20; DNA in which $N_1N_2N_3$ is CTs or CTv is referred to as DNA35s1 or DNA35v1, respectively; DNA in which $N_1N_2N_3$ is Css or Cvv is referred to as DNA35s2a or DNA35v2a, respectively; and DNA in which $N_1N_2N_3$ is sTs or vTv is referred to as DNA35s2b or DNA35v2b, respectively). The resulting DNA was purified by gel electrophoresis.

[0190] Preparation of RNA incorporating FAM-y or FAM-hx-y

17-mer RNA incorporating FAM-y or FAM-hx-y at the N' position (5'-GGGAAUCCCGAGN'AGUG-3'; SEQ ID NO: 21) was prepared through T7 transcription using T7 RNA polymerase and a substrate of FAM-y or FAM-hx-y (FAM-yTP or FAM-hx-yTP). The template (composed of DNA35v1 as a coding strand and 5'-ATAATACGACTCACTATAGGG-3' (SEQ ID NO: 22) as a non-coding

strand, 10 μM each) was annealed in a buffer solution containing 10 mM Tris-HCl (pH 7.6) and 10 mM NaCl by heating at 95°C and then slow cooling to 4°C. Transcription reaction was accomplished in a buffer solution containing 40 mM Tris-HCl (pH 8.0), 24 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT and 0.01% Triton X-100, in the presence of 1 mM natural NTPs, 1 mM FAM-yTP or FAM-hx-yTP, 2 μM template and 2.5 units/μL T7 RNA polymerase (Takara). After incubation at 37°C for 3 hours, a dye solution containing 10 M urea and 0.05% BPB was added in an equal volume to stop the reaction. The reaction mixture was heated at 75°C for 3 minutes and the 17-mer transcript was purified by gel electrophoresis.

[0191] Fluorescence spectrum measurement

Fluorescence spectra were measured in a buffer containing 10 mM sodium phosphate (pH 7.0), 100 mM NaCl and 0.1 mM EDTA. Emission spectra were recorded with an excitation wavelength of 350 nm for s and 360 nm for v, respectively. The emission spectral bandpass was set to 5 nm.

[0192] The 17-mer RNA site-specifically incorporating FAM-y or FAM-hx-y was hybridized with 35-mer DNA site-specifically containing one or two s or v bases (DNA35s1, 35s2a, 35s2b, 35v1, 35v2a or 35v2b) to measure emission spectra for the DNA/RNA duplex at different temperatures (20°C and 70°C).

[0193] Results

Resonance energy transfer was observed between a low fluorescence state of s or v and a high fluorescence state of FAM-y or FAM-hx-y. FRET occurred when DNA35s1 or DNA35v1 formed a duplex at 20°C with the 17-mer RNA site-specifically

incorporating FAM-y or FAM-hx-y, whereas FRET was significantly reduced when the duplex was denatured at 70°C (Figures 14a and 14b). In addition, FRET occurred efficiently in a duplex formed between DNA35s2a or DNA35v2a and the RNA site-specifically incorporating FAM-y or FAM-hx-y (Figures 14c and 14d).

[0194] Interestingly, the fluorescence intensity of single-stranded DNA35s2a or DNA35v2a containing two adjacent artificial bases was significantly decreased, whereas there was no significant decrease in the fluorescence intensity of single-stranded DNA35s2b or DNA35v2b containing two artificial bases partitioned by one T base (Figures 13c and 13d). Such a decrease observed in the fluorescence intensity of single-stranded DNA35s2a or DNA35v2a is caused by self-quenching between two adjacent artificial bases.

[0195] Since a strong FAM fluorescence from the 17-mer RNA was also observed during formation of a duplex with DNA35s2a or DNA35v2a in which self-quenching occurred between two adjacent artificial bases, this FRET system using a self-quenched probe facilitates the use of excessive probe against a target strand. When 10 molar equivalents of DNA35s1 or DNA35v1 containing one artificial base was used against the 17-mer RNA, resonance energy transfer was completely buried in the background fluorescence of DNA35s1 or DNA35v1 (Figures 15a and 15b). In contrast, when 10 molar equivalents of DNA35s2a or DNA35v2a containing two adjacent artificial bases was used against the 17-mer RNA, FRET was clearly observed between ss- or vv-containing DNA and FAM-containing RNA

(Figures 15c and 15d). Thus, such a FRET system using a self-quenched probe containing two adjacent artificial bases allows observation of a FRET phenomenon while reducing the background of a fluorescent donor, and hence enables the detection of complementary strands containing a fluorescent acceptor even in the presence of excessive donor probe. Moreover, since excessive self-quenched DNA probe facilitates the formation of a duplex with its complementary fluorescent acceptor strand, the FRET efficiency can be increased.

CLAIMS

1. A method for detecting the formation of a nucleic acid duplex, which comprise:

I) inducing hybridization between the following nucleic acids:

i) a nucleic acid incorporating a nucleotide having a 6-substituted 9H-purin-9-yl group as a base, which is prepared by the method according to any one of claims 6 to 8, and

ii) a nucleic acid incorporating a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is attached either directly or through a linker to a fluorescent dye selected from the group consisting of 5-carboxyfluorescein (5-FAM), 6-carboxyfluorescein (6-FAM), 5-carboxytetramethylrhodamine (5-TAMRA), 6-carboxytetramethylrhodamine (6-TAMRA), 5-dimethylaminonaphthalene-1-sulfonic acid (DANSYL), 5-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (5-HEX), 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (6-HEX), 5-carboxy-2',4,7,7'-tetrachlorofluorescein (5-TET), 6-carboxy-2',4,7,7'-tetrachlorofluorescein (6-TET), 5-carboxy-X-rhodamine (5-ROX) and 6-carboxy-X-rhodamine (6-ROX); and

II) measuring a change in the fluorescence spectrum.

2. The method according to claim 1, wherein the 6-substituted 9H-purin-9-yl group in I)i) is selected from

the group consisting of:

B1) a 2-amino-6-(2-thienyl)-9H-purin-9-yl group;

B2) a 6-(2-thienyl)-9H-purin-9-yl group;

B3) a 2-amino-6-(4-methyl-2-thienyl)-9H-purin-9-yl group;

B4) a 6-(4-methyl-2-thienyl)-9H-purin-9-yl group;

B5) a 2-amino-6-(5-methyl-2-thienyl)-9H-purin-9-yl group;

B6) a 6-(5-methyl-2-thienyl)-9H-purin-9-yl group;

B7) a 2-amino-6-(2-thiazolyl)-9H-purin-9-yl group;

B8) a 6-(2-thiazolyl)-9H-purin-9-yl group;

B9) a 2-amino-6-(4-methyl-2-thiazolyl)-9H-purin-9-yl group;

B10) a 6-(4-methyl-2-thiazolyl)-9H-purin-9-yl group;

B11) a 2-amino-6-(5-methyl-2-thiazolyl)-9H-purin-9-yl group; and

B12) a 6-(5-methyl-2-thiazolyl)-9H-purin-9-yl group.

3. The method according to claim 1 or 2, wherein the nucleic acid in I)ii) is a nucleic acid incorporating a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is attached either directly or through a linker to 5-carboxyfluorescein (5-FAM) or 6-carboxyfluorescein (6-FAM).

4. A method for preparing a nucleic acid incorporating the following nucleotides on the same strand:

i) a nucleotide having a 6-substituted 9H-purin-9-yl group as a base; and

ii) a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is attached either directly or through a linker to a fluorescent dye selected from the group consisting of 5-carboxyfluorescein (5-FAM), 6-carboxyfluorescein (6-FAM), 5-carboxytetramethylrhodamine (5-TAMRA), 6-carboxytetramethylrhodamine (6-TAMRA), 5-dimethylaminonaphthalene-1-sulfonic acid (DANSYL), 5-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (5-HEX), 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (6-HEX), 5-carboxy-2',4,7,7'-tetrachlorofluorescein (5-TET), 6-carboxy-2',4,7,7'-tetrachlorofluorescein (6-TET), 5-carboxy-X-rhodamine (5-ROX) and 6-carboxy-X-rhodamine (6-ROX),

wherein the method comprises effecting transcription, reverse transcription or replication by using, as a template, a nucleic acid incorporating the following nucleotides:

iii) a nucleotide having a substituted or unsubstituted 2-formyl-1H-pyrrol-1-yl group as a base; and

iv) a nucleotide having a 6-substituted 9H-purin-9-yl group as a base,

so that the nucleotide of i) is incorporated at a site complementary to the nucleotide of iii), while the nucleotide of ii) is incorporated at a site complementary to the nucleotide of iv).

5. A nucleic acid incorporating the following

nucleotides on the same strand, which is prepared by the method according to claim 4:

i) a nucleotide having a 6-substituted 9H-purin-9-yl group as a base; and

ii) a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is attached either directly or through a linker to a fluorescent dye selected from the group consisting of 5-carboxyfluorescein (5-FAM), 6-carboxyfluorescein (6-FAM), 5-carboxytetramethylrhodamine (5-TAMRA), 6-carboxytetramethylrhodamine (6-TAMRA), 5-dimethylaminonaphthalene-1-sulfonic acid (DANSYL), 5-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (5-HEX), 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (6-HEX), 5-carboxy-2',4,7,7'-tetrachlorofluorescein (5-TET), 6-carboxy-2',4,7,7'-tetrachlorofluorescein (6-TET), 5-carboxy-X-rhodamine (5-ROX) and 6-carboxy-X-rhodamine (6-ROX).

6. The nucleic acid according to claim 5, which is tRNA, mRNA, antisense DNA or RNA, a ribozyme, an aptamer or siRNA.

7. A method for detecting a stem-loop or hairpin structure formation of a nucleic acid, which comprises:

1) inducing hybridization between the nucleotide of i) and the nucleotide of ii) in the nucleic acid according to claim 5 or 6; and

2) measuring a change in the fluorescence spectrum.

8. The method according to claim 7, wherein the stem-loop or hairpin structure in the nucleic acid is formed upon binding between the nucleic acid according to claim 5 or 6 and a target molecule.

9. A method for detecting the formation of a nucleic acid duplex, which comprises:

I) inducing hybridization between the following nucleic acids:

i) a nucleic acid incorporating two or more adjacent nucleotides having a 6-substituted 9H-purin-9-yl group as a base, and

ii) a nucleic acid incorporating a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is attached either directly or through a linker to a fluorescent dye selected from the group consisting of 5-carboxyfluorescein (5-FAM), 6-carboxyfluorescein (6-FAM), 5-carboxytetramethylrhodamine (5-TAMRA), 6-carboxytetramethylrhodamine (6-TAMRA), 5-dimethylaminonaphthalene-1-sulfonic acid (DANSYL), 5-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (5-HEX), 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (6-HEX), 5-carboxy-2',4,7,7'-tetrachlorofluorescein (5-TET), 6-carboxy-2',4,7,7'-tetrachlorofluorescein (6-TET), 5-carboxy-X-rhodamine (5-ROX) and 6-carboxy-X-rhodamine (6-ROX); and

II) measuring a change in the fluorescence spectrum.

10. The method according to claim 9, wherein the 6-

substituted 9H-purin-9-yl group in I)i) is selected from the group consisting of:

B1) a 2-amino-6-(2-thienyl)-9H-purin-9-yl group;

B2) a 6-(2-thienyl)-9H-purin-9-yl group;

B3) a 2-amino-6-(4-methyl-2-thienyl)-9H-purin-9-yl group;

B4) a 6-(4-methyl-2-thienyl)-9H-purin-9-yl group;

B5) a 2-amino-6-(5-methyl-2-thienyl)-9H-purin-9-yl group;

B6) a 6-(5-methyl-2-thienyl)-9H-purin-9-yl group;

B7) a 2-amino-6-(2-thiazolyl)-9H-purin-9-yl group;

B8) a 6-(2-thiazolyl)-9H-purin-9-yl group;

B9) a 2-amino-6-(4-methyl-2-thiazolyl)-9H-purin-9-yl group;

B10) a 6-(4-methyl-2-thiazolyl)-9H-purin-9-yl group;

B11) a 2-amino-6-(5-methyl-2-thiazolyl)-9H-purin-9-yl group; and

B12) a 6-(5-methyl-2-thiazolyl)-9H-purin-9-yl group.

11. The method according to claim 9 or 10, wherein the nucleic acid in I)ii) is a nucleic acid incorporating a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is attached either directly or through a linker to 5-carboxyfluorescein (5-FAM) or 6-carboxyfluorescein (6-FAM).

12. A method for detecting a stem-loop or hairpin structure formation of a nucleic acid incorporating the following nucleotides on the same strand:

i) two or more adjacent nucleotides having a 6-substituted 9H-purin-9-yl group as a base; and

ii) a nucleotide having a 5-substituted-2-oxo(1H)pyridin-3-yl group as a base, whose 5-position is attached either directly or through a linker to a fluorescent dye selected from the group consisting of 5-carboxyfluorescein (5-FAM), 6-carboxyfluorescein (6-FAM), 5-carboxytetramethylrhodamine (5-TAMRA), 6-carboxytetramethylrhodamine (6-TAMRA), 5-dimethylaminonaphthalene-1-sulfonic acid (DANSYL), 5-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (5-HEX), 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (6-HEX), 5-carboxy-2',4,7,7'-tetrachlorofluorescein (5-TET), 6-carboxy-2',4,7,7'-tetrachlorofluorescein (6-TET), 5-carboxy-X-rhodamine (5-ROX) and 6-carboxy-X-rhodamine (6-ROX),

wherein said method comprises:

1) inducing hybridization between the nucleotide of i) and the nucleotide of ii) in the nucleic acid; and

2)measuring a change in the fluorescence spectrum.

SEQEUNCE LISTING

<110> Riken

<120> Novel artificial base pairs and uses thereof

<130> YCT1186

<150> JP 2005-226492
<151> 2005-08-04
<150> JP 2006-182062
<151> 2006-06-30

<160> 22

<210> 1
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> Synthesized template strand for transcription

<400> 1
ataatacgac tcactatagg g 21

<210> 2
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<221> misc_feature
<222> 3..5
<223> nnn means any of the followings:
cytosine - adenine - 2-formyl-1H-pyrrol-1-yl
cytosine - 2-formyl-1H-pyrrol-1-yl - 2-formyl-1H-pyrrol-1-yl
2-formyl-1H-pyrrol-1-yl - adenine - 2-formyl-1H-pyrrol-1-yl
cytosine - adenine - 2-oxo-1,3-dihydroimidazol-1-yl
cytosine - 2-oxo-1,3-dihydroimidazol-1-yl - 2-oxo-1,3-dihydroimidazol-1-yl
2-oxo-1,3-dihydroimidazol-1-yl - adenine - 2-oxo-1,3-dihydroimidazol-1-yl
cytosine - adenine - cytosine

<220>
<223> Synthesized template strand for transcription

<400> 2
ctnnnaggga agctccctat agtgagtcgt attat 35

<210> 3
<211> 17
<212> RNA
<213> Artificial sequence

<220>
<221> misc_feature
<222> 13..15
<223> nnn means any of the followings:
2-amino-6-(2-thienyl)-purin-9-yl - thymine - guanine
2-amino-6-(2-thienyl)-purin-9-yl - 2-amino-6-(2-thienyl)-purin-9-yl - guanine
2-amino-6-(2-thienyl)-purin-9-yl - thymine - 2-amino-6-(2-thienyl)-purin-9-yl
guanine - thymine - guanine

<220>
<223> Product by transcription

<400> 3
gggagcuucc cunnnag 17

<210> 4
<211> 94
<212> DNA
<213> Artificial sequence

<220>
<223> Synthesized non-template strand for transcription

<400> 4
ataatacgac tcactatagg ggatttagct cagttgggag agcgccagac tgaagatctg
gaggtcctgt gttcgatcca cagaattccc acca 94

<210> 5
<211> 94
<212> DNA
<213> Artificial sequence

<220>
<221> modified_base
<222> 1
<223> tm

<220>
<221> modified_base
<222> 2
<223> gm

<220>
<223> Synthesized template strand for transcription

<400> 5
nngtgggaat tctgtggatc gaacacagga cctccagatc ttcagtctgg cgctctccca
actgagctaa atcccctata gtgagtcgta ttat 94

<210> 6
<211> 94
<212> DNA
<213> Artificial sequence

<220>
<221> modified_base
<222> 1
<223> tm

<220>
<221> modified_base
<222> 2
<223> gm

<220>
<221> misc_feature
<222> 61
<223> n means 2-formyl-1H-pyrrol-1-yl

<220>
<223> Synthesized template strand for transcription

<400> 6
nngtgggaat tctgtggatc gaacacagga cctccagatc ttcagtctgg cgctctccca
nctgagctaa atcccctata gtgagtcgta ttat 94

<210> 7
<211> 94
<212> DNA
<213> Artificial sequence

<220>
<221> modified_base
<222> 1
<223> tm

<220>
<221> modified_base
<222> 2
<223> gm

<220>
<221> misc_feature
<222> 60
<223> n means 2-formyl-1H-pyrrol-1-yl

<220>
<223> Synthesized template strand for transcription

<400> 7
nngtgggaat tctgtggatc gaacacagga cctccagatc ttcagtctgg cgctctcccn
actgagctaa atcccctata gtgagtcgta ttat 94

<210> 8
<211> 94
<212> DNA
<213> Artificial sequence

<220>
<221> modified_base
<222> 1
<223> tm

<220>
<221> modified_base
<222> 2
<223> gm

<220>
<221> misc_feature
<222>
<223> n means 2-formyl-1H-pyrrol-1-yl

<220> 41
<223> Synthesized template strand for transcription

<400> 8
nngtgggaat tctgtggatc gaacacagga cctccagatc ntcagtctgg cgctctccca
actgagctaa atcccctata gtgagtcgta ttat 94

<210> 9
<211> 94
<212> DNA
<213> Artificial sequence

<220>
<221> modified_base
<222> 1
<223> tm

<220>
<221> modified_base
<222> 2
<223> gm

<220>
<221> misc_feature
<222> 30
<223> n means 2-formyl-1H-pyrrol-1-yl

<220>
<223> Synthesized template strand for transcription

<400> 9
nngtgggaat tctgtggatc gaacacaggn cctccagatc ttcagtctgg cgctctccca
actgagctaa atcccctata gtgagtcgta ttat 94

<210> 10
<211> 94
<212> DNA
<213> Artificial sequence

<220>
<221> modified_base
<222> 1
<223> tm

<220>
<221> modified_base
<222> 2
<223> gm

<220>
<221> misc_feature

<222> 20
<223> n means 2-formyl-1H-pyrrol-1-yl

<220>
<223> Synthesized template strand for transcription

<400> 10
nngtgggaat tctgtggatn gaacacagga cctccagatc ttcagtctgg cgctctccca
actgagctaa atcccctata gtgagtcgta ttat 94

<210> 11
<211> 94
<212> DNA
<213> Artificial sequence

<220>
<221> modified_base
<222> 1
<223> tm

<220>
<221> modified_base
<222> 2
<223> gm

<220>
<221> misc_feature
<222> 18
<223> n means 2-formyl-1H-pyrrol-1-yl

<220>
<223> Synthesized template strand for transcription

<400> 11
nngtgggaat tctgtggntc gaacacagga cctccagatc ttcagtctgg cgctctccca
actgagctaa atcccctata gtgagtcgta ttat 94

<210> 12
<211> 76
<212> RNA
<213> Yeast

<220>
<223> Yeast tRNA Phe

<400> 12
ggggauuuag cucaguuggg agagcgccag acugaagauc uggagguccu guguucgauc
cacagaauuc ccacca 76

<210> 13
<211> 62
<212> DNA
<213> Artificial sequence

<220>
<223> Synthesized non-template strand for transcription

<400> 13
agaatttaat acgactcact atagggaatc ccgagtagtg ctactaataa aatgcactag
tg 62

<210> 14
<211> 62
<212> DNA
<213> Artificial sequence

<220>
<221> misc_feature
<222> 4
<223> n means 2-formyl-1H-pyrrol-1-yl

<220>
<221> misc_feature
<222> 27
<223> n means 2-amino-6-(2-thiazolyl)-purin-9-yl

<220>
<223> Synthesized template strand for transcription

<400> 14
cacnagtgca ttttattagt agcactnctc gggattccct atagtgagtc gtattaaatt
ct 62

<210> 15
<211> 62
<212> DNA
<213> Artificial sequence

<220>
<221> misc_feature
<222> 7
<223> n means 2-formyl-1H-pyrrol-1-yl

<220>
<221> misc_feature
<222> 24
<223> n means 2-amino-6-(2-thiazolyl)-purin-9-yl

<220>
<223> Synthesized template strand for transcription

<400> 14
cactagngca ttttattagt agcnctactc gggattccct atagtgagtc gtattaaatt
ct 62

<210> 16
<211> 62
<212> DNA
<213> Artificial sequence

<220>
<221> misc_feature
<222> 11
<223> n means 2-formyl-1H-pyrrol-1-yl

<220>
<221> misc_feature
<222> 21
<223> n means 2-amino-6-(2-thiazolyl)-purin-9-yl

<220>
<223> Synthesized template strand for transcription

<400> 16
cactagtgca ntttattagt ngcactactc gggattccct atagtgagtc gtattaaatt
ct 62

<210> 17
<211> 39
<212> RNA
<213> Artificial sequence

<220>
<221> misc_feature
<222> 13, 16, 19
<223> n means 2-oxo-(1H)pyridin-3-yl or thymine

<220>
<221> misc_feature
<222> 29, 33, 36
<223> n means 2-amino-6-(2-thienyl)-purin-9-yl or adenine

<220>
<223> Product by transcription

<400> 17
gggaaucccg agnagngcna cuaauaaanu gcncungug 39

<210> 18
<211> 39
<212> RNA
<213> Artificial sequence

<220>
<221> misc_feature
<222> 13
<223> n means 2-oxo-(1H)pyridin-3-yl

<220>
<221> misc_feature
<222> 36
<223> n means 2-amino-6-(2-thienyl)-purin-9-yl

<220>
<223> Product by transcription

<400> 18
gggaaucccg agnagugcua cuaauaaaau gcacungug 39

<210> 19
<211> 42
<212> RNA
<213> Artificial sequence

<220>
<223> Synthesized complementary strand

<400> 19
gggcacuagu gcauuuuauu aguagcacua cucgggauuc cc 42

<210> 20
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<221> misc_feature
<222> 3..5
<223> nnn means any of the followings:
cytosine - thymidine - [2-amino-6-(2-thienyl)purine-9-yl];
cytosine - thymidine - [2-amino-6-(2-thiazolyl)purine-9-yl];
cytosine - [2-amino-6-(2-thienyl)purine-9-yl] - [2-amino-6-(2-thienyl)purine-9-yl];
cytosine - [2-amino-6-(2-thiazolyl)purine-9-yl] - [2-amino-6-(2-thiazolyl)purine-9-yl];
[2-amino-6-(2-thienyl)purine-9-yl] - thymidine - [2-amino-6-(2-thienyl)purine-9-yl];
[2-amino-6-(2-thiazolyl)purine-9-yl] - thymidine - [2-amino-6-(2-thiazolyl)purine-9-yl];

<400> 20
cannnctcgg gattccctat agtgagtcgt attat 35

<210> 21
<211> 17
<212> RNA
<213> Artificial sequence

<220>
<221> misc_feature
<222> 13
<223> n means 2-oxo-(1H)pyridin-3-yl

<400> 21
gggaaucccg agnagug 17

<210> 22
<211> 21
<212> DNA
<213> Artificial sequence

<400> 22
ataatacgac tcactatagg g 21