PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number:	WO 92/16180
A61K	A2	(43) International Publication Date:	1 October 1992 (01.10.92)

(21) International Application Number: PCT/US92/02037

(22) International Filing Date: 12 March 1992 (12.03.92)

(30) Priority data:

668,549	13 March 1991 (13.03.91)	US
754,146	3 September 1991 (03.09.91)	US

(60) Parent Application or Grant
(63) Related by Continuation
US 754,146 (CON)
Filed on 3 September 1991 (03.09.91)

(71) Applicants *(for all designated States except US):* SISKA DIAGNOSTICS, INC. [US/US]; 505 Coast Boulevard South, La Jolla, CA 92037 (US). THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 300 Lakeside Drive, 22nd Floor, Oakland, CA 94612-3550 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only) :* GINGERAS, Thomas, Raymond [US/US]; 1528 Juniper Hill Drive, Encinitas, CA 92024 (US). BARRINGER, Kevin, J. [US/US]; 7988 Montara Avenue, San Diego, CA 92126 (US). RICHMAN, Douglas, D. [US/US]; 9551 La Jolla Farms Road, La Jolla, CA 92037 (US). PRODANOVICH, Patricia, C. [US/US]; 5638 Gaines Street, San Diego, CA 92110 (US). DAVIS, Geneva, Ruth [US/US]; 3083 East Fox Run Way, San Diego, CA 92111 (US).

(74) Agents: WATT, Phillip, H. et al.; Fitch, Even, Tabin & Flannery, Room 900, 135 South LaSalle Street, Chicago, IL 60603 (US).

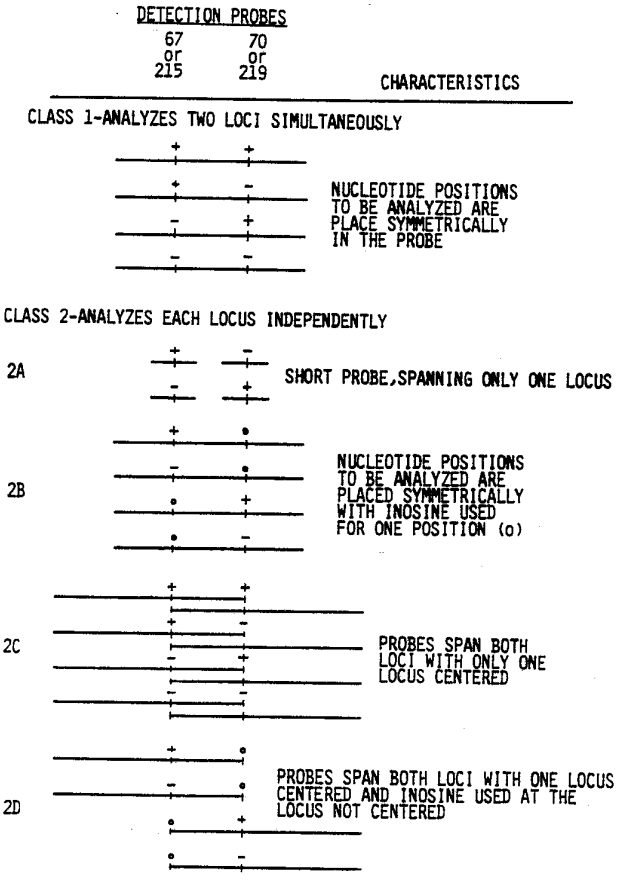
(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), MC (European patent), NL (European patent), SE (European patent), US.

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: DETECTION OF 3'-AZIDO-3'-DEOXYTHYMIDINE RESISTANCE

(57) Abstract

An assay for detecting the genotype of mutant and wild-type alleles at multiple loci associated with resistance to the drug AZT utilizes a series of probes which anneal to an amplified region of the HIV-1 pol gene. The assay format which involves amplification of a rare nucleic acid followed by capture on a solid phase, and detection of the captured sequences by a labelled probe, is applicable to assays for resistance to other drugs involving point mutations.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FI**	Finland	**ML**	Mali
AU	Australia	**FR**	France	**MN**	Mongolia
BB	Barbados	**GA**	Gabon	**MR**	Mauritania
BE	Belgium	**GB**	United Kingdom	**MW**	Malawi
BF	Burkina Faso	**GN**	Guinea	**NL**	Netherlands
BG	Bulgaria	**GR**	Greece	**NO**	Norway
BJ	Benin	**HU**	Hungary	**PL**	Poland
BR	Brazil	**IE**	Ireland	**RO**	Romania
CA	Canada	**IT**	Italy	**RU**	Russian Federation
CF	Central African Republic	**JP**	Japan	**SD**	Sudan
CG	Congo	**KP**	Democratic People's Republic of Korea	**SE**	Sweden
CH	Switzerland			**SN**	Senegal
CI	Côte d'Ivoire	**KR**	Republic of Korea	**SU**	Soviet Union
CM	Cameroon	**LI**	Liechtenstein	**TD**	Chad
CS	Czechoslovakia	**LK**	Sri Lanka	**TG**	Togo
DE	Germany	**LU**	Luxembourg	**US**	United States of America
DK	Denmark	**MC**	Monaco		
ES	Spain	**MG**	Madagascar		

DETECTION OF 3'-AZIDO-3'-DEOXYTHYMIDINE RESISTANCE

BACKGROUND OF THE INVENTION

Human Immunodeficiency Virus (HIV-1) is the causative agent of Acquired Immune Deficiency Syndrome (AIDS) or AIDS-related complex (ARC), an infectious disease characterized by depletion of certain subpopulations of T-lymphocytes from the immune system. The virus enters these T-cells via a specific receptor integral to the CD4 T-cell specific antigen. Upon entry, the virus directs synthesis of reverse transcriptase from the pol gene, thereby enabling the RNA virus to generate a duplex DNA copy capable of integrating the host cell genome, from which further virus production is directed.

By impairing the immune system, the HIV-1 virus leaves the host individual open to a variety of opportunistic bacterial, fungal, and viral disease agents, and a majority of AIDS patients succomb to the complications of one or more such opportunistic infections. In the past several years, there have been great efforts to identify drugs, utilized singly or in combination, capable of enhancing human host resistance to these agents. For example, U.S. Patent No. 4,925,831 describes administration of the combinaton of an aminoalkyl naphthalenediol and one of a series of nucleotide analogs. EP 0 378 643 discloses a therapeutic composition comprising a soluble T4 protein and either azidothymidine or a glucoside inhibitor. U.S. Patent Nos. 4,866,036 and 4,866,035 disclose use of the combination of dipeptidyl glucofuranose compounds and an antiviral agent such as anidothymidine, ansamycin, ribavirin, dioxycytidine, or foscarnet.

The antiviral compound, azidothymidine (AZT), an analog of thymidine, is currently the only clinically approved successful drug in treating AIDS and ARC. AZT administered to AIDS patients produces an increase in CD4 positive cells and usually a decrease in detectable serum levels of the viral p24 protein, although these manifestations are not perfectly correlated. [Richman, et

al., Am. J. Med. ,85:208 (1988)]. The mechanism of AZT action is incompletely understood, although invitro studies suggest that is may be an effective low-level chain terminator in viral replication, and exerts an inhibitory effect on the polymerization reaction. This effect involves a specific interaction between the reverse transcriptase and the azido- group since studies of AZT resistant mutant enzymes shows a similar resistance to 3'-azido-2',3'-dideoxyuridine, but not to 2',3'-deoxycytidine, 2',3'-dideoxy-2',3'-didehydrothymidine, or phosphonoformate. (Larder, et al., Science, 343:1731 (1989).

A major long-term drawback to AZT therapy is a tendency for resistance to the drug to develop. This is found to occur by degrees over a period of prolonged use, and involves certain mutations in the pol gene, consistent with the invitro studies referred to above (Larder and Kemp, Science, 246:1155). Monitoring the emergence of these mutations associated with AZT resistance may have predictive value in planning long-term drug strategies.

SUMMARY OF THE INVENTION

In accordance with present invention, an assay is provided for detecting the genotype of a mutant or wild type marker, which are distinquished by one or more base changes in the chromosomal DNA at one or more loci resulting in an amino acid substitution in the protein encoded by the gene in question. In the first step in such assay, a nucleic acid fraction is extracted from cells expected to express the mutant or wild-type marker. The nucleic acid fraction is then amplified by nucleic acid amplification techniques such as polymerase chain reaction (PCR) or self-sustained sequence replication (3SR).

Primers for such amplification are selected

which span the target marker region. The amplified nucleic acids are then separated by capturing same on a capture means, such as a homologous sequence having a sequence distinct from the marker sequence affixed to a support means, and washing away the non-homologous nucleic acids. Finally, a probe is constructed which hybridizes either to the mutant or wild-type marker regions. These probes have reporter means such as a radioactive substance or fluorophor to permit detecting the hybridization of the probes to the captured sequences.

In a preferred embodiment, the assay target is a sequence of the HIV-1 pol gene which contains mutations or the corresponding wild-type alleles at the 67, 70, 215 and 219 amino acid positions. The assay is performed by first extracting the nucleic acid fraction from cells obtained from patients with a varying degree of AZT resistance, amplifying a nucleic acid sequence in the nucleic acid fraction spanning the AZT resistant mutant-containing region from about nucleotide 2330 to 2787 of the pol gene utilizing either PCR or 3SR, probing the sequence so amplified with probes specific for mutant AZT resistant and wild-type alleles, and detecting the hybridication of the probes to their amplified targets.

The probe sequences which are necessary for the differential determination of the correct AZT resistance genotype comprise a first family of probe sequences for detecting in solution hybridization assays the mutant or wild-type genotype of AZT resistance in the HIV-1 pol gene comprising a first family of probe sequences having a nucleotide sequence of 24 to 30 nucleotides substantially homologous to the region encompassing nucleotides 2330 to 2340 of the pol gene, selected from the group consisting of a probe specific for the mutant allele at nucleotide position 2340 comprising a nucleotide sequence containing an inosine nucleotide at

position 2330 and a guanidine nucleotide at position 2340, said sequence extending from the inosine positioned substantially 2-4 nucleotides from the 5'-terminus to a nucleotide located 3' of the said guanidine nucleotide such that said guanidine nucleotide is located substantially midpoint in the sequence, a probe specific for the mutant allele at nucleotide position 2330 comprising a nucleotide sequence containing an adenosine nucleotide at position 2330 and an inosine nucleotide at position 2340, the sequence extending from a nucleotide located substantially 12 to 15 nucleotides 5' of said adenosine nucleotide to the inosine nucleotide located 3-6 nucleotides from the 3'-terminus, a probe specific for the wild-type genotype comprising a nucleotide sequence containing a guanidine at position 2330 and an adenosine nucleotide at position 2340, said sequence extending from said inosine positioned substantially 3-7 nucleotides from the 5'-terminus to a nucleotide located 3' of said adenosine nucleotide such that said adenosine nucleotide is located substantially midpoint in said sequence; and a second family of probe sequences having a nucleotide sequence of 24 to 30 nucleotides substantially homologous to the region encompassing nucleotides 2774 to 2787 of the pol gene, selected from the group consising of a probe specific for the double base mutant allele at positions 2774 and 2775 comprising a nucleotide sequence containing a thymidine nucleotide at the 5'-terminus, a thymidine or adenosine nucleotide at the penultimate 5' position, said sequence extending from the terminal 5' thymidine nucleotide to a nucleotide located 3' of nucleotide 2787 such that said nucleotide 2787 is located spacedly at substantially the midpoint in the sequence, a probe specific for the mutant allele at position 2787 comprising a nucleotide sequence containing a cytosine nucleotide at position 2787, said sequence being substantially symmetrical about the midpoint nucleotide

between nucleotides 2774 and 2787, a probe specific for the double base mutant allele at positions 2774 and 2775 in combination with the mutant allele at position 2787 comprising a nucleotide sequence containing a thymidine nucleotide at the 5'-terminus, a thymidine or adenosine nucleotide at the penultimate 5' position, and a cytosine nucleotide at position 2787, this sequence extending from the terminal 5' thymidine nucleotide to a nucleotide located 3' of nucleotide 2787 such that said nucleotide 2787 is located spacedly at substantially the midpoint in said sequence, and a probe specific for the wild-type genotype comprising a nucleotide sequence containing an adenosine nucleotide at position 2774, a cytosine nucleotide at position 2776, and a adenosine nucleotide at position 2787, said sequence being substantially symmetrical about the midpoint nucleotide between nucleotides 2774 and 2787.

From the foregoing, it is apparent that the families of novel probes herein disclosed make it possible to determine the genotype of AZT resistance at each of the four major mutant loci. It is a further object of the present invention to use the assay results to predict the course of mutation to full AZT resistance in order to plan and implement appropriate therapy in the treatment of AIDS.

DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention provides an assay for detecting the genotype of a mutant or wild-type marker utilizing a nucleic acid probe-based system. The model assay detects the genotype of AZT resistance which results from point mutations at 4 known sites. Progression to substantially total resistance is associated with the accumulation of all four mutations. The assay method is also applicable to other drug resistance markers, or to other disease states associated

with point mutations at defined loci.

In carrying out the present assay, the nucleic acid fraction from a tissue source must be extracted. Most commonly, a good nucleic acid preparation can be obtained utilizing the phenol/chloroform extraction method described by Maniatas, et al. Molecular Cloning: A Laboratory Manual, 1982. A method of nucleic acid extraction utilizing chaotropic agents can also be employed, as described in Pellegrimo, et al., Biotechniques, v. 5, 1987.

Once the nucleic acids have been isolated and purified, primers are added and an amplification reaction targeting a defined nucleic region is carried out. In general, any method of amplification known in the art will be satisfactory, provided that the interprimer sequence is faithfully replicated. Polymerase chain reaction (PCR) as described in U.S. Patent No. 4,683,202 (Mullis), is perhaps the best-known system, although in the present assay, self-sustaining sequence replication (3SR) is preferred. For the details of 3SR amplification, please refer to Gingeras, et al., Ann. Biol. Clin. 48:498 (1990).

In the preferred embodiment, the amplified sequences are first captured on solid support means by a capture means comprising a nucleic acid sequence homologous to a region of the amplified pol gene distinct from the probe target region. The support means may be any solid material known in the assay art to effect separations of target sequence from nonspecific contaminants. Preferred are beads 1-200 microns in diameter made of polyacrylamide or polystyrene.

The preferred primer sequences for amplifying the region containing the mutational positions conferring AZT resistance are set forth in Table 1. Primer sequences are selected which reproducibly yield greater than 5 logs10 of amplification. Also, it is important to

TABLE 1

POL REGION PRIMERS FOR 3SR AMPLIFICATION

SEQ ID. NO.	NUMBER	REGION	SENSE	T7	SEQUENCE
1	90-126	67/70	ANTI	+	AAT TTA ATA CGA CTC ACT ATA GGG A TT GTA CTG ATA TCT AAT CCC TGG TGT CTC A
2	89-404	67/70	SENSE	-	AA GTT AAA CAA TGG CCA TTG ACA GAA GAA A
3	90-46	67/70	SENSE	-	GT AGC ATG ACA AAA ATC TTA GAG CC
4	90-47	215/219	SENSE	-	T CCA CAG GGA TGG AAA GGA TCA CCA GCA A
5	89-388	215/219	ANTI	+	AAT TTA ATA CGA CTC ACT ATA GGG A TTT TTC TGG CAG CAC TAT AGG CTG TAC TGT
6	89-391	215/219	ANTI	+	AAT TTA ATA CGA CTC ACT ATA GGG A TTT CCC CAC TAA CTT CTG TAT GTC ATT GAC A'

note that primers utilized in 3SR require at one member of the primer pair to contain a T7 RNA polymerase promoter sequence.

The probe sequences of the present invention which detect mutations, either together or singly, at the amino acid 67 and 70 positions are given in Table 2. Preferred probe sequences which differentially detect mutations, either together or singly, at the amino acid 215 and 219 positions are set forth in Table 3. The third to the left column of each of the probe tables gives the genotype of the 67/70 region for each of the mutant and wild-type possibilities. Similarly, column 3 of Table 3 provides the genotype of 215/219 region for each such mutant and wild-type possibility.

The strategy for selecting probe sequences is based on the rationale for designing the following classes, as follows: Class 1 probes are oligonucleotides that monitor both loci of a specific region simultaneously. Class 2 detection probes are oligonucleotides that are specific for a single locus. Class 2A oligonucleotides include shorter probes whose sequences encompass only one of the amino acid codons implicated in AZT resistance. Class 2B probes are similar to class 1 probes, but contain a nucleotide analog, inosine, at one of the two mutant positions. The use of inosine, which partially base pairs equally well with all other nucleotides, neutralizes mismatches and thus enables these probes to be used for one specific locus. For example, a probe whose genotype at the 67/70 region is +/- hybridizes specifically to a +/- target under stringent conditions. Replacement of the mutant nucleotide in this probe at amino acid 70 position with an inosine, changing the genotype from +/- to +/o, allows the probe to hybridize equally well with either a +/+ or +/- target thus enabling detection of any target that is wild-type at position 67. Detection probes for the

TABLE 2

POL 67/70 REGION DETECTION PROBE SUMMARY

SEQ. ID NO.	PROBE	TYPE	SIZE	CLASS	DIAGRAM
7	90-248	C.D.*	24		TGT ACA GAA ATG GAA AAG GAA GGG
8	90-36	+/+	27	1	G AAA AAA GAC AGT ACT AAA TGG AGA AA
9	90-48	-/-	27	1	G AAA AAA AAC AGT ACT AGA TGG AGA AA
10	90-49	+/-	27	1	G AAA AAA GAC AGT ACT AGA TGG AGA AA
11	90-50	-/+	27	1	G AAA AAA AAC AGT ACT AAA TGG AGA AA
12	89-416	70-	15	2A	AGT ACT AGA TGG AGA
13	89-417	70+	15	2A	AGT ACT AAA TGG AGA
14	89-481	67-	18	2A	AG AAA AAA AAC AGT ACT A
15	89-480	67+	18	2A	AG AAA AAA GAC AGT ACT A
16	90-449	0/+	27	2B	G AAA AAA IAC AGT ACT AAA TGG AGA AA
17	90-450	0/-	27	2B	G AAA AAA IAC AGT ACT AGA TGG AGA AA
18	90-451	+/0	27	2B	G AAA AAA GAC AGT ACT AIA TGG AGA AA
19	90-452	-/0	27	2B	G AAA AAA AAC AGT ACT AIA TGG AGA AA
20	90-153	±/+	23	2C	TA AAG AAA AAA GAC AGT ACT AAA
21	90-154	+/±	25	2C	AA GAC AGT ACT AAA TGG AGA AAA (T/C)T
22	90-155	+/∓	25	2C	AA GAC AGT ACT AGA TGG AGA AAA (T/C)T
23	90-156	∓/+	23	2C	TA AAG AAA AAA AAC AGT ACT AAA

TABLE 2 (continued)

SEQ. ID. NO.	PROBE	TYPE	SIZE	CLASS	DIAGRAM
24	90-601	±/-	27	2C	CT ATA AAG AAA AAA GAC AGT ACT AGA T
25	90-602	±/+	27	2C	CT ATA AAG AAA AAA GAC AGT ACT AAA T
26	90-603	+/=	26	2C	AA GAC AGT ACT AGA TGG AGA AAA ITA
27	90-604	=/+	28	2C	GCT ATA AAG AAA AAA AAC AGT ACT AAA T
28	90-605	=/-	28	2C	GCT ATA AAG AAA AAA AAC AGT ACT AGA T
29	90-606	+/±	29	2C	AA AAA GAC AGT ACT AAA TGG AGA AAA ITA
30	90-607	-/±	29	2C	AA AAA AAC AGT ACT AAA TGG AGA AAA ITA
31	90-608	-/=	26	2C	AA AAC AGT ACT AGA TGG AGA AAA ITA
32	90-498	+/0	23	2D	TA AAG AAA AAA GAC AGT ACT AIA
33	90-499	-/0	23	2D	TA AAG AAA AAA AAC AGT ACT AIA
34	90-500	0/-	25	2D	AA IAC AGT ACT AGA TGG AGA AAA IT
35	90-501	0/+	25	2D	AA IAC AGT ACT AAA TGG AGA AAA IT
36	90-553	0/-	26	2D	AA IAC AGT ACT AGA TGG AGA AAA ITA
37	90-554	+/0	27	2D	CT ATA AAG AAA AAA GAC AGT ACT AIA T
38	90-555	-/0	28	2D	GCT ATA AAG AAA AAA AAC AGT ACT AIA T
39	90-556	0/+	29	2D	AA AAA IAC AGT ACT AAA TGG AGA AAA ITA
40	90-580	+/0	30	2D	CT ATA AAG AAA AAA GAC AGT ACT AIA TGG A
41	90-616	-/0	31	2D	GCT ATA AAG AAA AAA AAC AGT ACT AIA TGG A
42	90-665	-/0	30	2D	CT ATA AAG AAA AAA AAC AGT ACT AIA TGG A
43	90-666	-/0	29	2D	T ATA AAG AAA AAA AAC AGT ACT AIA TGG A

67+ G
67- A
70+ A
70- G

TABLE 3

POL 215/219 REGION DETECTION PROBE SUMMARY

SEQ. ID. NO.	PROBE	TYPE	SIZE	CLASS	DIAGRAM
44	89-535	C.D.	30		ATG GGT TAT GAA CTC CAT CCT GAT AAA TGG
45	90-302	-/+	26	1	G GGA ITT T(A/T)C ACA CCA GAC AAA AAAC
46	90-303	+/+	26	1	G GGA ITT ACC ACA CCA GAC AAA AAAC
47	90-304	-/-	26	1	G GGA ITT T(A/T)C ACA CCA GAC CAA AAA C
48	90-305	+/-	26	1	G GGA ITT ACC ACA CCA GAC CAA AAA C
49	89-547	215+	14	2A	GGA (C/T)TT ACC ACA CC
50	89-548	215-	14	2A	GGA (C/T)TT T(T/A)C ACA CC
51	89-483	219+	18	2A	A CCA GAC AAA AAA CAT CA
52	89-482	219-	18	2A	A CCA GAC CAA AAA CAT CA
53	90-446	+/0	26	2B	G GGA ITT ACC ACA CCA GAC IAA AAA C
54	90-447	0/+	26	2B	G GGA ITT IIC ACA CCA GAC AAA AAA C
55	90-448	0/-	26	2B	G GGA ITT IIC ACA CCA GAC CAA AAA C
56	90-453	-/0	26	2B	G GGA ITT T(A/T)C ACA CCA GAC IAA AAA C
57	90-505	=/+F	28	2C	G AGG TGG GGA ITT TTC ACA CCA GAC AAA
58	90-506	=/+Y	28	2C	G AGG TGG GGA ITT TAC ACA CCA GAC AAA
59	90-507	±/+	28	2C	G AGG TGG GGA ITT ACC ACA CCA GAC AAA
60	90-568	±/-	28	2C	G AGG TGG GGA ITT ACC ACA CCA GAC CAA
61	90-569	=/-Y	28	2C	G AGG TGG GGA ITT TAC ACA CCA GAC CAA

TABLE 3 (continued)

SEQ. ID. NO.	PROBE	TYPE	SIZE	CLASS	DIAGRAM
62	90-570	=/-F	28	2C	G AGG TGG GGA ITT TTC ACA CCA GAC CAA
63	90-510	+/±	25	2C	ACC ACA CCA GAC AAA AAA CAT CAG A
64	90-511	+/=	25	2C	ACC ACA CCA GAC CAA AAA CAT CAG A
65	90-566	-/±F	25	2C	TTC ACA CCA GAC AAA AAA CAT CAG A
66	90-567	+/±Y	25	2C	TAC ACA CCA GAC AAA AAA CAT CAG A
67	90-565	-/=F	25	2C	TTC ACA CCA GAC CAA AAA CAT CAG A
68	90-564	-/=Y	25	2C	TAC ACA CCA GAC CAA AAA CAT CAG A
69	90-502	=/0F	28	2D	G AGG TGG GGA ITT TTC ACA CCA GAC IAA
70	90-503	-/0Y	28	2D	G AGG TGG GGA ITT TAC ACA CCA GAC IAA
71	90-504	+/0	28	2D	G AGG TGG GGA ITT ACC ACA CCA GAC IAA
72	90-508	0/±	25	2D	IIC ACA CCA GAC AAA AAA CAT CAG A
73	90-509	0/=	25	2D	IIC ACA CCA GAC CAA AAA CAT CAG A

215+ ACC
215- TTC (F)
TAC (Y)

219+ AAA
219- CAA

mutation at amino acid 215 are synthesized with a degeneracy at the 215 position to compensate for the possibility of two different mutant sequences. Oligonucleotides have been synthesized which specifically detect each mutant sequence. Additionally, an inosine residue at the non-AZT associated degeneracy at the first base of the codon for the amino acid 214 has been maintained.

Class 2C probes span both loci with one locus centered and the other near the end of the probe. This asymmetric alignment of mutant sites would be predicted to enable the probe to function in a manner analogous to the inosine-containing probes described above. This prediction is based upon the assumption that a mismatch toward either end of a probe would have little effect upon the stability of the duplex. Therefore, probes of this structure would be specific for the locus which is centered in the sequence. Class 2D are similar to 2C probes but contain inosine at the mutation sites located near the ends of the probes, thus further minimizing the end-of-probe mismatches. The probe construction strategy is further illustrated in figure 1, which gives the characteristics of the various probe classes.

Further advantages of the present assay and probes will be apparent from the following examples.

EXAMPLE 1

General Materials and Methods

Reverse transcriptase (from avian myeloblastosis virus), ribonuclease H, T7 RNA polymerase, restriction enzymes HindIII, EcoRI and Asp718, BglII the large fragment (Klenow) of E. coli. DNA polymerase I, T4 DNA ligase and calf intestinal alkaline phosphatase were all purchased from commercial vendors. Oligonucleotides for primers and probes were synthesized using phosphoramidite chemistry on Applied Biosystems Model

380A DNA synthesizer using reagents from the same manufacturer. Oligonucleotides were purified by HPLC using a C8 column. Trisacryl Oligobeadstm were synthesized as described previously in Davis, et. al., J. Infect. Dis. 162:13-20 (1990).

Plasmids containing one of the four mutations implicated in the generation of resistance to AZT were constructed by site-directed mutagenesis method similar to that reported by Zollar, et. al, Methods in Enzymology, 100:468-500. Briefly, 20 pmol of an oligonucleotide containing the desired mutation were annealed to 1 pmol of a single strand M13mp18 clone containing an approximately 1700 nt Asp718-BglII insert from pARV. This fragment contains the portion of the pol gene spanning the amino acid 67, 70, 215 and 219. The primers were extended using the Klenow fragment of E. coli polymerase I overnight at room temperature. These extension reactions were then diluted 40-fold, and used to transform JM103. Replica filters from plates containing the transformed bacteria were made with nitrocellulose filters (Millipore, HATF). Phage DNA on the filters were denatured, hybridized with 32P-labeled oligonucleotide probes specific for the desired mutations and exposed to film. Phage mini-preps were prepared from several positive plaques and these were used in a second round of screening as described in Maniatas, et. al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, 1982. The presence of the desired mutations were confirmed by sequencing using a Sequeanse Version II sequencing Kit (United States Biochemicals). RF M13 DNAs containing the point mutations described above were cleaved with the restriction endonucleased HindIII and EcoRI each of the inserts were isolated from preparative agarose gels. The inserts were ligated using units of T4 DNA ligase into the plasmids pT7/T3alpha-18 and pT7/3alpha-19 which had been linearized with the

enzymes EcoRI and HindIII and treated with calf intestinal alkaline phosphatase. Ligation products were then used to transform CaCl2-competent E. coli strain MC1061 cells. The plasmid RTMC/3 (gift from B. Larder) contained all four point mutations. A 2.6 kb EcoRI-HindIII fragment encompassing all four mutations was isolated from the RTMC/3 (3) and ligated into the transcription vectors described above.

The transcription vectors containing one or more of the mutant amino acid codons are listed in Table 1. Sense or antisense RNA transcripts were made from these plasmids using 1 to 10 ug of linearized plasmid and conditions described by the manufacturer.

3SR reactions were carried out as described previously in Guatolli, et. al., PNAS, 87:1874 (1990) except that 10 mM KCl was used instead of 20 mM NACl, BSA was omitted and the incubations were carried out at 42°C. instead of 37°C.. Only the antisense primer oligonucleotide contains a T7 RNA polymerase promoter sequence. When DNA targets were amplified, 3SR reactions were denatured at 100°C. for 5 minutes, annealed at 42°C. for 2 minutes and 10 U of reverse transcriptase were added. Reactions were incubated at 42°C. for 10 minutes, the denaturation/annealing steps repeated, then 3SR reaction initiated.

The 3SR amplification of the 67-70 amino acid region was carried out separately from the 215-219 amino acid region. Oligonucleotides an their positions used as primers in the amplification of each of the pol regions are shown in Figure 2. PCR and Southern hybridization procedures used in these experiments have previously been described in Richman, et. al. (1991).

EXAMPLE 2

Differential Bead-based Sandwich Hybridization (BBSH)

The detection of 3SR amplification products and

the determination of the genotype of the 3SR RNA was accomplished by use of a differential BBSH procedure. Similar to the BBSH protocol described previously (Guatelli, et. al., Proc. Natl. Acad. Sci. USA.87:1874:1878 and Davis, et. al., J. Infect. Dis. 162:13-20. (1990)), differential BBSH assays used 2 ml microcolumns (Isolab, Inc.) which contain 25 mg of Trisacryl Oligobeadstm (10). To these beads, 100 femtomoles of 32P-labeled detection oligonucleotide, and 10-2 to 10-3 aliquots of the 3SR product in a total volume of 30 ul were added. These hybridization reactions were heated at 65°C. for five minutes after which 30 ul of solution hybridization mix (10 x standard sodium phosphate/EDTA, 20% (w/V) dextran sulfate, 0.2% SDS, prewarmed to 65°C.) was added, mixed, transferred to the appropriate temperature and allowed to hybridize for two hours, then washed six times, 1 ml per wash, with standard sodium citrate (SSC) buffer.

Detection oligonucleotides in the 67/70 or 215/219 regions designed to hybridize to specific combinations of genotypes are shown in Figure 2. The probes for the 215/219 region contain an inosine residue at the first nucleotide at the amino acid 214 codon to compensate for a sequence heterogeneity between several HIV-1 isolates. Additionally, probes that contain a mutant 215 locus are degenerate at the second nucleotide of the amino acid 215 codon. This position has been shown to have two possible amino acid substitutions occur.

The assay to determine the genotype of the 3SR amplified material is a thermal melt/batch elution assay. The 3SR amplification products from each mutant region was analyzed by four specific detection probes corresponding to each possible genotype and a control detection oligonucleotide. All samples were analyzed in duplicate. Each amplification reaction was first

subjected to low stringency analyses consisting of a 42°C. hybridization step followed by 2 x SSC (where 1 x equals 0.15 M NaCl, 15 MM sodium citrate) was at 42°C.. This was followed by a second high stringency differential hybridization. For the amino acid 67/70 region, the differential BBSH reactions are hybridized at 42°C., then washed at 50°C. with 0.75 to 1.0 x SSC washes (see Table 6A). The amino acid 215/219 containing BBSH reactions are hybridized at 55°C. followed by washes with 0.5 x SSC at 55°C.

The radioactivity of the pooled washes and the probe remaining on the beads was determined by Cerenkov counting, and the amount of product captured calculated by the percent of the total cpm that remained hybridized to the capture oligo-linked trisacryl beads.

Direct Sequencing of 3SR Amplified Product

Reverse transcriptase sequencing of 3SR products was performed as described in Guatelli, et. al., Proc. Natl. Acad. Sci. USA.87:1874-1878 (1990). Primers used for sequencing in the 67/70 region are 89-415 (5'-CAAAAATTGGGCCTG-3' SEQ. ID. NO. 74), 89-419 (5'-AGAACTCAAGACTTCTGGGAAGTTC-3' SEQ. ID. NO. 75), 89-441 (5'-AATCCATACAATACTCCAGTATTTGC-3' SEQ. ID. NO. 76), 90-416 (5'-GGGAAAATTTCAAAAATTGGGC-3' SEQ. ID. NO. 77), 90-417 (5'-ACAGAAATGGAAAAGGAAGGG-3' SEQ. ID. NO. 78), AND 90-249 (5'-GAAAAAATAAAAGCATTAGTAGA-3' SEQ. ID. NO. 79).

Primers used for sequencing amplified products containing the 215/219 locus are 89-534 (5'-AGGATCTGACTTAGAAATAGGGCAGCA-3' SEQ. ID. NO. 80), 90-414 (5'-TTGTATGTAGGATCTGACTTAG-3' SEQ. ID. NO. 81), 90-415 (5'-ACATGGATGATTTGTATGTAGG-3' SEQ. ID. NO. 92), 90-37 (5'-TATCTATCAATACATGGATGATTTGTATGT-3' SEQ. ID. NO. 83), and 90-46

(5'-GTAGCATGACAAAAATCTTAGAGCC-3' SEQ. ID. NO. 84).

EXAMPLE 3

The specific detection of wild type and mutant HIV-1 pol gene sequences was initially demonstrated by amplifying 0.1 attamoles (104 copies) of RNA and DNA targets of defined sequence. Table 4 lists the control plasmids and RNA transcription vectors used to measure the differential hybridization capabilities of the BBSH assay. These DNA and RNA targets were used in 3SR reactions designed to amplify separately each of the two closely spaced regions of the pol gene, each containing two of the mutations to be monitored. The 3SR RNA products were analyzed by BBSH employing two sets of four oligonucleotide probes (26-27 mers) capable of distinguishing between wild type and mutant sequences present at either the 67-70 or 215-219 amino acid region (Figure 2). The use of shorter oligonucleotide probes (15-20 bases) designed to analyze separately the sequence at each of the codons failed to hybridize to the 3SR products. This failure to hybridize was caused by competing RNA secondary structures in these regions of the pol gene (data not shown). Consequently, longer oligonucleotides capable of greater hybridization stability were designed and employed to identify the genotype of the target nucleic acid. Each oligonucleotide detection probe was designed to monitor two of the specified codons simultaneously. Consequently, it was possible to detect the four combinations of mutants possible in regions 67-70 and 215-219 (Figure 2).

Analyses of the 3SR products generated from the control targets indicated that the differential BBSH assay could detect and distinguish all possible mutations in the 67-70 and the 215-219 regions (Table 7) by greater than a 6:1 ratio. Increased stringency of the post-bead

hybridization washes (0.5 x SSC) was required to accurately identify the genotype of all targets for the 215-219 region (Table 6B). The probes used in the 215-219 region contained inosine at the first base of codon 214 and probes 90-304 and 90-302 were composed of mixed population probes to account for the possibility of either mutation at position 215. Interestingly, the phenylalamine (pPol 215P-218) and tyrosine (pPol 215T-18) mutant codons present at the 215 position were detected with almost equal efficiency.

Differential elution can be accomplished by block elution at a specific high or low stringency salt wash. Alternatively, by introducing a gradient of decreasing salt concentration, elution of the detection probe at an optimal stringency can be detected. The protocol for sequential elution (block) is illustrated in Figure 3.

The HIV-1 viruses isolated from the PBMCs of seven patients who had received AZT therapy for 5 to 26 months were analyzed first by the separate 3SR amplifications of the pol gene segments encoding the 67-70 and 215-219 regions (Figure 1). The use of the isothermal 3SR amplification method permitted the selective amplification of HIV-1 RNA present in the viral stocks produced by the coculture of the patient PBMC with uninfected MT2 cells. The results of the differential BBSH analyses of each sample was verified by the nucleotide sequencing of the 3SR product.

The differential BBSH analyses of the 3SR RNA products of the 67-70 and 215-219 regions (Table 5A) required the use of only 1% or less of each of the amplification reactions. Results of the BBSH analyses of the 67-70 region permitted the straightforward determination of the genotype of each sample with a discrimination of greater than 10:1 in all cases except samples I312-6 (wt/wt) and I429-1 (wt/wt) (Table 3A).

The use of more stringent wash conditions for sample I312-6 resulted in loss of any detectable signal. Consequently, the less stringent wash conditions were required to determine the genotype of this sample. This observation is consistent with the presence of other sequences within the region surveyed by the detection probes. Such sequence variation was confirmed by sequence analysis of this sample which showed a change in the third base of codon 67 and a change in the first base of codon 68 (Figure 6A). Sample I429-1 also exhibited a requirement for lower wash stringency in the BBSH assay to determine the genotype of the HIV-1 pol gene in this sample. The sequence of the 67-70 region of the pol gene of this virus showed no sequence differences in the region covered by the detection probe. However, sequence differences were observed in the region surveyed by the oligonucleotide used to capture the target on the bead support. These sequence variations appear to have led to a lower hybridization efficiency of the complementary 90-36 (wt/wt) probe. The probes 89-381, 89-419 and 89-441 hybridized 40 bases away from the 67 and 70 codons and provided a measure of the amounts of 3SR product in each of the differential BBSH assays.

The determination of the genotype in the 215-219 region of these samples required greater wash stringencies than was used for the 67-70 region owing to the higher G/C content (42% compared to 33%) of this genomic segment. Except for samples I312-6, the signal to noise ratios for all other samples were greater than 7:1 (Table 6B). In addition to the two single base mutations within codons 215 and 219 in sample I312-6 (mut/mut), mutations also occur in codons 214 and 211 which destablized the efficiency of the complementary probe 90-312 resulting in a lower signal to noise ratio.

Correlation of 3SR/Differential BBSH Genotypes Results with PCR/Southern Hybridization Results and Viral Phenotypes

The samples from seven patients analyzed by the 3SR/differential BBSH assay were also analyzed using PCR/differential Souther hybridization method. Viral isolates were obtained directly from the PBMCs and viral pools. Table 6 correlates the results obtained by 1) each set of amplification-mediated hybridization methods (3SR and PCR), 2) the nucleotide sequence analyses of the 3SR products and 3) the AZT-susceptibility assays of the viral isolates obtained from each sample.

The genotypic results obtained by the 3SR/differential BBSH and nucleotide sequence analyses are in agreement for all codons except for the 67 and 70 codons of samples I381-4 and J821-1. For these samples the 3SR/differential BBSH assay detects the predominant mutant forms of the mixed virus population detected by the sequence analyses. The results obtained by the 3SR/differential BBSH and PCR/Souther hybridization analyses for the viral isolates are substantially in agreement except for samples G685-2 and I312-6. The mixed population of viruses detected by the PCR/Southern hybridization assay for sample J821-1 is composed of predominantly mutant virus which agrees with both the 3SR/differential BBSH and sequence analyses. The differences in the genotypes for residue 67 of sample G691-2 and I312-2 are attributable to sequence variation at adjacent codons 66 and 68, respectively.

The discrepancy observed for sample G685-6 may be attributable to the fact two different passages of this viral isolate were analyzed by the PCR/differential Southern hybridization and 3SR/differential BBSH assays. Such variation between viral passages suggest similar discrepancies should be observed between PBMC and coculture isolates. Such discrepancies not involving

mixed viral populations are observed in at least one of these codon genotypes for two of the six samples (Table 7).

Finally, samples G685-6 and J821-1 possess the same genotype (wt/mut/mut/wt) at each of the four monitored codons but differ by a factor of 3 in the level of sensitivity to AZT. Such variation in phenotype virus isolates having the same or similar genotypes suggests that mutations in the pol gene outside the four codons surveyed in this study are influential in the observed levels of AZT resistance. However, it may be noteworthy that except for isolate I429-1, codon altering mutations have been observed in all of the isolates in regions other than the four codons monitored by the BBSH assay. Of these additional mutations, the four viral isolates exhibiting AZT resistance all possess a arginine to lysine mutation at codon 211.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Gingeras, Thomas R
Barringer, Kevin J
Richman, Douglas D
Prodanovich, Patricia C
Davis, Geneva R

(ii) TITLE OF INVENTION: DETECTION OF 3'-AZIDO-3'-DEOXYTHYMIDINE RESISTANCE

(iii) NUMBER OF SEQUENCES: 84

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Baxter Diagnostics Inc.
(B) STREET: One Baxter Parkway, DF2-2E
(C) CITY: Deerfield
(D) STATE: Illinois
(E) COUNTRY: USA
(F) ZIP: 60015

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 7/754,146
(B) FILING DATE: 03-SEP-1991
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Barta, Kent
(B) REGISTRATION NUMBER: 29,042
(C) REFERENCE/DOCKET NUMBER: LSRL-4121 CONT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 708/948/3308
(B) TELEFAX: 708/948-2642

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 55 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AATTTAATAC GACTCACTAT AGGGATTGTA CTGATATCTA ATCCCTGGTG TCTCA 55

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AAGTTAAACA ATGGCCATTG ACAGAAGAAA 30

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GTAGCATGAC AAAAATCTTA GAGCC 25

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TCCACAGGGA TGGAAAGGAT CACCAGCAA 29

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 55 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AATTTAATAC GACTCACTAT AGGGATTTTT CTGGCAGCAC TATAGGCTGT ACTGT 55

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AATTTAATAC GACTCACTAT AGGGATTTCC CCACTAACTT CTGTATGTCA TTGACA 56

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TGTACAGAAA TGGAAAAGGA AGGG 24

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GAAAAAAGAC AGTACTAAAT GGAGAAA 27

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GAAAAAAAAC AGTACTAGAT GGAGAAA 27

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GAAAAAAGAC AGTACTAGAT GGAGAAA 27

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAAAAAAAAC AGTACTAAAT GGAGAAA 27

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AGTACTAGAT GGAGA 15

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGTACTAAAT GGAGA 15

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGAAAAAAAA CAGTACTA 18

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AGAAAAAAGA CAGTACTA 18

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 8
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GAAAAAAIAC AGTACTAAAT GGAGAAA 27

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 8
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GAAAAAAIAC AGTACTAGAT GGAGAAA 27

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 18
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GAAAAAAGAC AGTACTATAT GGAGAAA 27

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 18
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GAAAAAAAAC AGTACTATAT GGAGAAA 27

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TAAAGAAAAA AGACAGTACT AAA 23

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AAGACAGTAC TAAATGGAGA AAAYT 25

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

AAGACAGTAC TAGATGGAGA AAAYT 25

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TAAAGAAAAA AAACAGTACT AAA 23

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CTATAAAGAA AAAAGACAGT ACTAGAT 27

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CTATAAAGAA AAAAGACAGT ACTAAAT 27

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 24
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AAGACAGTAC TAGATGGAGA AAATTA 26

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCTATAAAGA AAAAAAACAG TACTAAAT 28

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCTATAAAGA AAAAAAACAG TACTAGAT 28

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 27
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

AAAAAGACAG TACTAAATGG AGAAAAITA 29

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 27
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

AAAAAAACAG TACTAAATGG AGAAAAITA 29

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 24
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

AAAACAGTAC TAGATGGAGA AAATTA 26

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 22
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TAAAGAAAAA AGACAGTACT AIA 23

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 22
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

TAAAGAAAAA AAACAGTACT AIA 23

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 3, 24

(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AAIACAGIAC TAGATGGAGA AAAIT 25

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 3, 24
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

AAIACAGIAC TAAATGGAGA AAAIT 25

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 3, 24
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

AAIACAGIAC TAGATGGAGA AAAITA 26

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 25
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CTATAAAGAA AAAAGACAGT ACTATAT 27

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 26
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GCTATAAAGA AAAAAAACAG TACTATAT 28

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 6, 27
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

AAAAAIACAG TACTAAATGG AGAAAAITA 29

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 25
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CTATAAAGAA AAAAGACAGT ACTAIATGGA 30

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 26
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCTATAAAGA AAAAAAACAG TACTAIATGG A 31

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 25
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CTATAAAGAA AAAAAACAGT ACTAIATGGA 30

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 24
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

TATAAAGAAA AAAAACAGTA CTATATGGA 29

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

ATGGGTTATG AACTCCATCC TGATAAATGG 30

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 5
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GGGATTTTWC ACACCAGACA AAAAAC 26

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 5
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GGGAITTACC ACACCAGACA AAAAAC 26

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 5
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GGGAITTIWC ACACCAGACC AAAAAC 26

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 5
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GGGAITTACC ACACCAGACC AAAAAC 26

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

GGAYTTACCA CACC 14

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GGAYTTTWCA CACC 14

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

ACCAGACAAA AAACATCA 18

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

ACCAGACCAA AAACATCA 18

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 5, 20
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GGGAITTACC ACACCAGACI AAAAAC 26

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 5, 8, 9
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GGGAITTIIC ACACCAGACA AAAAAC 26

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 5, 8, 9
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55

GGGAITTTIC ACACCAGACC AAAAAC 26

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 5, 20
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GGGAITTTYC ACACCAGACI AAAAAC 26

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 11, 15
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GAGGTGGGGA ITTTICACAC CAGACAAA 28

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 11
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GAGGIGGGGA ITTTACACAC CAGACAAA 28

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 11
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

GAGGIGGGGA ITTACCACAC CAGACAAA 28

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 11
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

GAGGIGGGGA ITTACCACAC CAGACCAA 28

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base

(B) LOCATION: 11
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GAGGIGGGGA ITTTACACAC CAGACCAA 28

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 11
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GAGGIGGGGA ITTTTCACAC CAGACCAA 28

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

ACCACACCAG ACAAAAAACA TCAGA 25

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

ACCACACCAG ACCAAAAACA TCAGA 25

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

TTCACACCAG ACAAAAAACA TCAGA 25

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

TACACACCAG ACAAAAAACA TCAGA 25

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

TTCACACCAG ACCAAAAACA TCAGA 25

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

TACACACCAG ACCAAAAACA TCAGA 25

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 11, 26
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

GAGGTGGGGA ITTTTCACAC CAGACTAA 28

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 11, 26
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

GAGGTGGGGA ITTTACACAC CAGACTAA 28

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 11, 26
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

GAGGIGGGGA IITACCACAC CAGACIAA 28

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1, 2
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

IICACACCAG ACAAAAAACA TCAGA 25

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1, 2
(C) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

IICACACCAG ACCAAAAACA TCAGA 25

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

CAAAAATTGG GCCTG 15

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

AGAACTCAAG ACTTCTGGGA AGTTC 25

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

AATCCATACA ATACTCCAGT ATTTGC 26

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GGGAAAATTT CAAAAATTGG GC 22

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

ACAGAAATGG AAAAGGAAGG G 21

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

GAAAAAATAA AAGCATTAGT AGA 23

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

AGGATCTGAC TTAGAAATAG GGCAGCA 27

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

TTGTATGTAG GATCTGACTT AG 22

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

ACATGGATGA TTTGTATGTA GG 22

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

TATCTATCAA TACATGGATG ATTTGTATGT 30

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

GTAGCATGAC AAAAATCTTA GAGCC 25

CLAIMS:

1. An assay for detecting the genotype of a mutant or wild-type marker comprising

extracting the nucleic acid fraction from cells expressing said marker

amplifiying an RNA sequence in said fraction spanning the said marker by self-sustaining sequence replication

capturing the RNA sequence so amplified by capture means affixed to support means

probing the said captured sequences with probes specific for the said mutant or wild-type marker, and

detecting the hybridization of said probes to said captured sequences.

2. The assay of claim 1 wherein said capture means is a capture nucleic acid sequence homologous to a portion of said amplified RNA sequence distinct from said marker.

3. The assay of claim 1 wherein said support means comprises microbeads.

4. An assay for detecting the genotype of AZT resistance in a patient infected with HIV-1 comprising

extracting the nucleic acid fraction from infected cells

amplifying a nucleic acid sequence in said fraction spanning the AZT resistant mutant-containing region from nucleotide 2330 to 2787 of the HIV-1 pol gene

probing the sequence so amplified with probes specific for mutant AZT resistant and wild-type alleles, and

detecting the hybridization of said probes to their specific amplified targets.

5. The assay of claim 1 wherein the method of amplifying a nucleic acid sequence is self-sustaining sequence replication.

6. Probe sequences for detecting in solution hybridization assays the mutant or wild-type genotype of

AZT resistance in the HIV-1 pol gene comprising

a first family of probe sequences having a nucleotide sequence of 24 to 30 nucleotides substantially homologous to the region encompassing nucleotides 2330 to 2340 of said pol gene, selected from the group consisting of

a probe specific for the mutant allele at nucleotide position 2340 comprising a nucleotide sequence containing an inosine nucleotide at position 2330 and a guanidine nucleotide at position 2340, said sequence extending from said inosine positioned substantially 2-4 nucleotides from the 5'-terminus to a nucleotide located 3' of the said guanidine nucleotide such that said guanidine nucleotide is located substantially midpoint in said sequence,

a probe specific for the mutant allele at nucleotide position 2330 comprising a nucleotide sequence containing an adenosine nucleotide at position 2330 and an inosine nucleotide at position 2340, said sequence extending from a nucleotide located substantially 12 to 15 nucleotides 5' of said adenosine nucleotide to the said inosine nucleotide located 3-6 nucleotides from the 3'-terminus,

a probe specific for the wild-type genotype comprising a nucleotide sequence containing a guanidine at position 2330 and an adenosine nucleotide at position 2340, said sequence extending from said inosine positioned substantially 3-7 nucleotides from the 5'-terminus to a nucleotide located 3' of said adenosine nucleotide such that said adenosine nucleotide is located substantially midpoint in said sequence; and

a second family of probe sequences having a nucleotide sequence of 24 to 30 nucleotides substantially homologous to the region encompassing nucleotides 2774 to 2787 of said pol gene, selected from the group consising of

a probe specific for the double base mutant

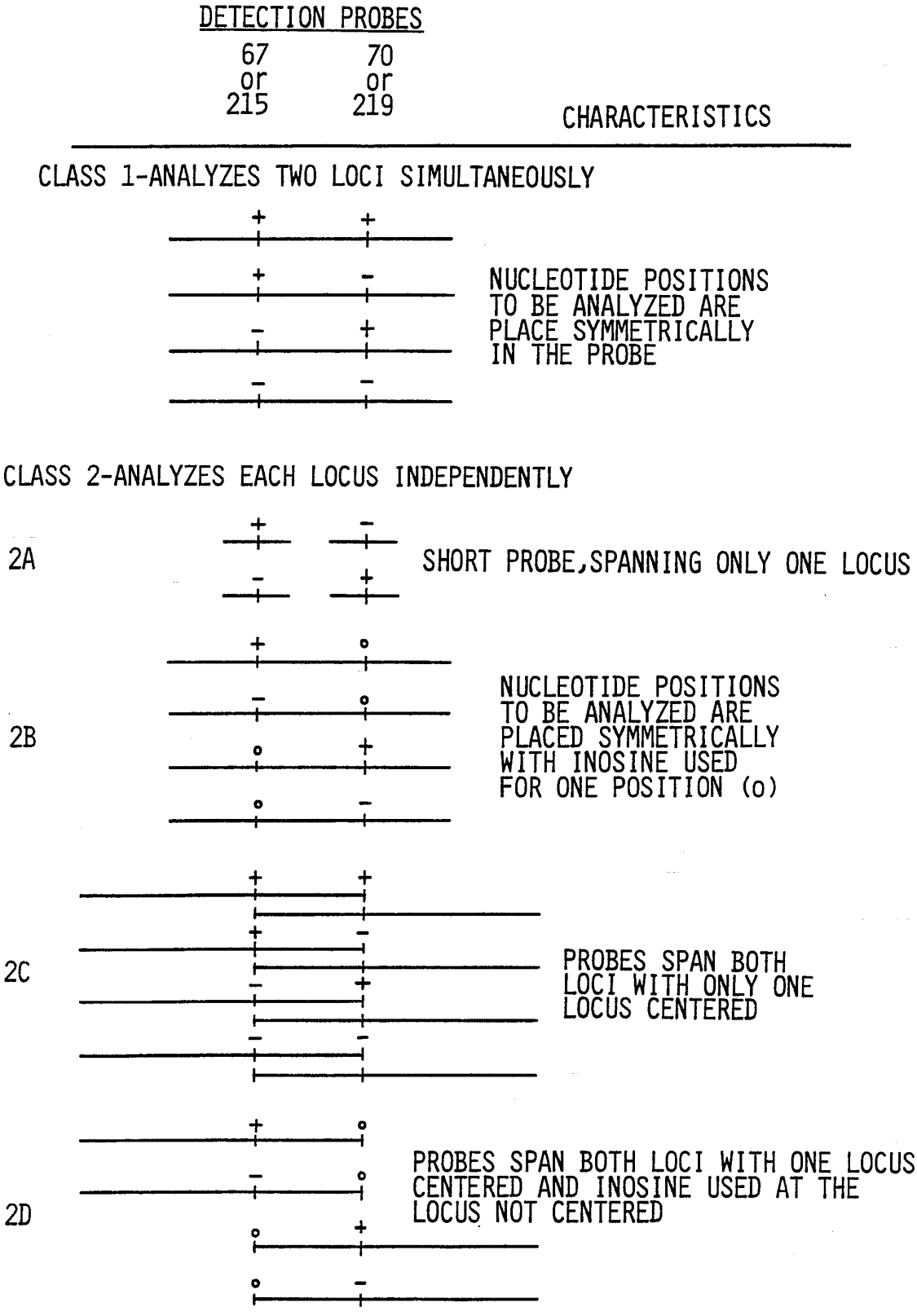
allele at positions 2774 and 2775 comprising a nucleotide sequence containing a thymidine nucleotide at the 5'-terminus, a thymidine or adenosine nucleotide at the penultimate 5' position, said sequence extending from said terminal 5' thymidine nucleotide to a nucleotide located 3' of nucleotide 2787 such that said nucleotide 2787 is located spacedly at substantially the midpoint in said sequence,

a probe specific for the mutant allele at position 2787 comprising a nucleotide sequence containing a cytosine nucleotide at position 2787, said sequence being substantially symmetrical about the midpoint nucleotide between nucleotides 2774 and 2787,

a probe specific for the double base mutant allele at positions 2774 and 2775 in combination with the mutant allele at position 2787 comprising a nucleotide sequence containing a thymidine nucleotide at the 5'-terminus, a thymidine or adenosine nucleotide at the penultimate 5' position, and a cytosine nucleotide at position 2787, said sequence extending from said terminal 5' thymidine nucleotide to a nucleotide located 3' of nucleotide 2787 such that said nucleotide 2787 is located spacedly at substantially the midpoint in said sequence, and

a probe specific for the wild-type genotype comprising a nucleotide sequence containing an adenosine nucleotide at position 2774, a cytosine nucleotide at position 2776, and a adenosine nucleotide at position 2787, said sequence being substantially symmetrical about the midpoint nucleotide between nucleotides 2774 and 2787.

FIG. 1


DETECTION PROBES
67 or 215
70 or 219
CHARACTERISTICS
CLASS 1-ANALYZES TWO LOCI SIMULTANEOUSLY
NUCLEOTIDE POSITIONS TO BE ANALYZED ARE PLACE SYMMETRICALLY IN THE PROBE
CLASS 2-ANALYZES EACH LOCUS INDEPENDENTLY
2A
SHORT PROBE, SPANNING ONLY ONE LOCUS
2B
NUCLEOTIDE POSITIONS TO BE ANALYZED ARE PLACED SYMMETRICALLY WITH INOSINE USED FOR ONE POSITION (o)
2C
PROBES SPAN BOTH LOCI WITH ONLY ONE LOCUS CENTERED
2D
PROBES SPAN BOTH LOCI WITH ONE LOCUS CENTERED AND INOSINE USED AT THE LOCUS NOT CENTERED

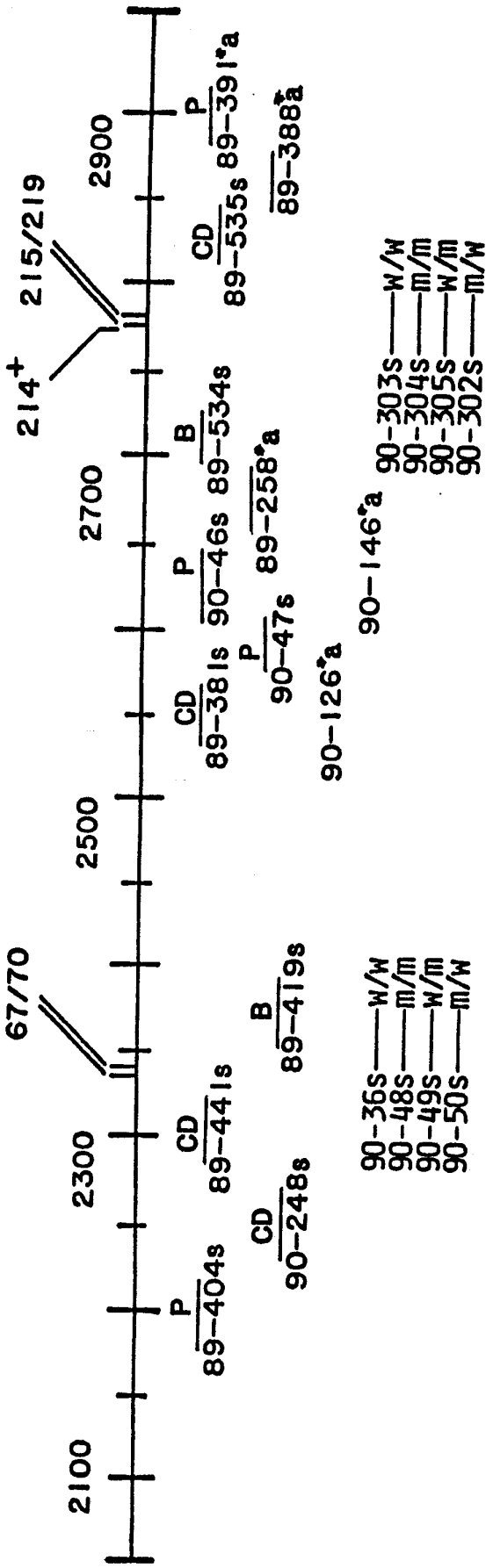
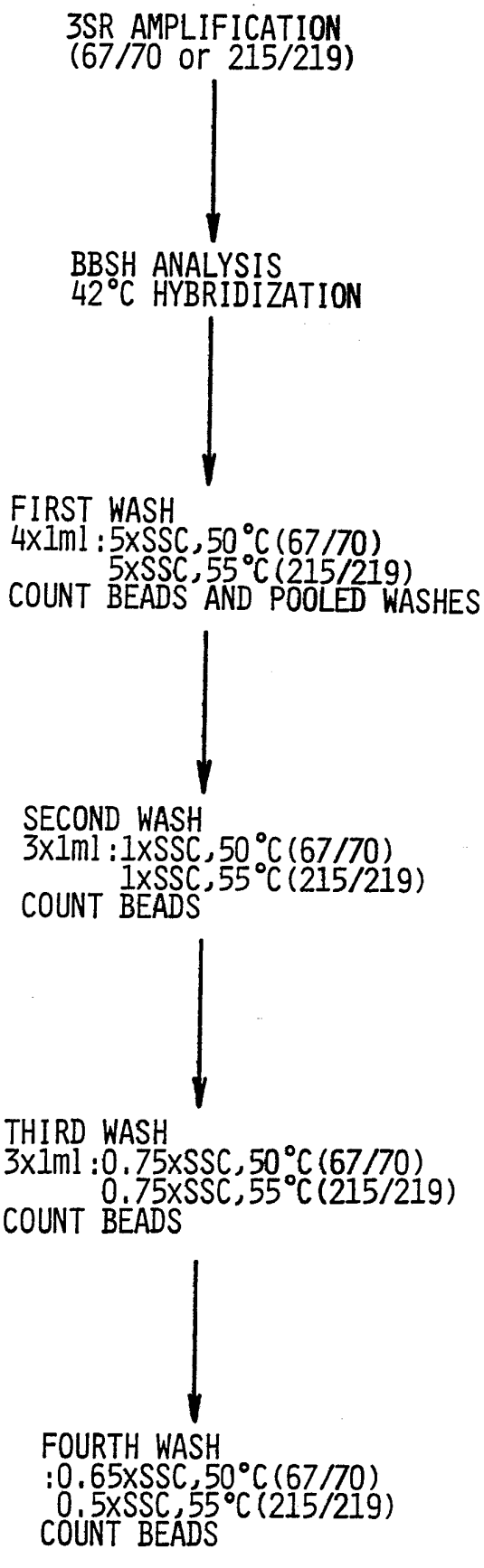

2900
2700
2500
2300
2100
215/219
214+
67/70
P 89-391*a
89-388*a
CD 89-535s
B 89-534s
P 90-46s
89-258*a
CD 89-381s
P 90-47s
90-126*a
90-146*a
B 89-419s
CD 89-441s
CD 90-248s
P 89-404s
90-303s — W/W
90-304s — m/m
90-305s — W/m
90-302s — m/W
90-36s — W/W
90-48s — m/m
90-49s — W/m
90-50s — m/W

FIG. 2

FIG. 3


3SR AMPLIFICATION
(67/70 or 215/219)
BBSH ANALYSIS
42°C HYBRIDIZATION
FIRST WASH
4x1ml:5xSSC,50°C(67/70)
5xSSC,55°C(215/219)
COUNT BEADS AND POOLED WASHES
SECOND WASH
3x1ml:1xSSC,50°C(67/70)
1xSSC,55°C(215/219)
COUNT BEADS
THIRD WASH
3x1ml:0.75xSSC,50°C(67/70)
0.75xSSC,55°C(215/219)
COUNT BEADS
FOURTH WASH
:0.65xSSC,50°C(67/70)
0.5xSSC,55°C(215/219)
COUNT BEADS