



(86) Date de dépôt PCT/PCT Filing Date: 2003/12/05  
 (87) Date publication PCT/PCT Publication Date: 2004/06/24  
 (45) Date de délivrance/Issue Date: 2014/07/08  
 (85) Entrée phase nationale/National Entry: 2005/06/03  
 (86) N° demande PCT/PCT Application No.: US 2003/038505  
 (87) N° publication PCT/PCT Publication No.: 2004/053141  
 (30) Priorités/Priorities: 2002/12/06 (US60/431,319);  
 2002/12/18 (US10/323,438); 2003/09/12 (US10/660,998)

(51) Cl.Int./Int.Cl. *C07H 21/04* (2006.01),  
*C12N 15/09* (2006.01), *C12P 19/34* (2006.01),  
*C12Q 1/44* (2006.01), *C12Q 1/68* (2006.01)

(72) Inventeurs/Inventors:  
 ECKER, DAVID J., US;  
 CROOKE, STANLEY T., US;  
 BLYN, LAWRENCE B., US;  
 HALL, THOMAS A., US;  
 JIANG, YUN, US;  
 GRIFFEY, RICHARD H., US;

(73) Propriétaire/Owner:  
 IBIS BIOSCIENCES, INC., US

(74) Agent: NORTON ROSE FULBRIGHT CANADA

(54) Titre : PROCÉDES D'ANALYSE RAPIDE MEDICO-LEGALE D'UN ADN MITOCHONDRIEN ET CARACTERISATION DE L'HETEROPLASMIE DE L'ADN MITOCHONDRIEN  
 (54) Title: METHODS FOR RAPID FORENSIC ANALYSIS OF MITOCHONDRIAL DNA AND CHARACTERIZATION OF MITOCHONDRIAL DNA HETEROPLASMY

(57) **Abrégé/Abstract:**

The present invention provides methods for rapid forensic analysis of mitochondrial DNA and methods for characterizing heteroplasmy of mitochondrial DNA, which can be used to assess the progression of mitochondrial diseases.



(72) **Inventeurs(suite)/Inventors(continued)**: SAMPATH, RANGARAJAN, US; HOFSTADLER, STEVEN A., US; MCNEIL, JOHN, US; HANNIS, JAMES C., US; WHITE, NEILL, US; SAMANT, VIVEK, US; ESHOO, MARK W., US; DRADER, JARED J., US

(74) **Agent(suite/continued)**: LLP/S.E.N.C.R.L., S.R.L.

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

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
24 June 2004 (24.06.2004)

PCT

(10) International Publication Number  
**WO 2004/053141 A3**

- (51) International Patent Classification<sup>7</sup>: C12Q 1/68,  
C12P 19/34, C07H 21/04
- (21) International Application Number:  
PCT/US2003/038505
- (22) International Filing Date: 5 December 2003 (05.12.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/431,319 6 December 2002 (06.12.2002) US  
10/323,438 18 December 2002 (18.12.2002) US  
10/660,998 12 September 2003 (12.09.2003) US
- (71) Applicant (for all designated States except US): **ISIS PHARMACEUTICALS, INC.** [US/US]; 1896 Rutherford Road, Carlsbad, CA 92008 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **ECKER, David, J.** [US/US]; 1041 Saxony Road, Encinitas, CA 92024 (US). **CROOKE, Stanley, T.** [US/US]; 3211 Piragua Street, Carlsbad, CA 92009 (US). **BLYN, Lawrence, B.** [US/US]; 27892 Perales, Mission Viejo, CA 92692 (US). **HALL, Thomas, A.** [US/US]; 5239 Wohlford Street, Oceanside, CA 92056 (US). **JIANG, Yun** [CN/US]; 6574 Robinea Drive, Carlsbad, CA 92009 (US). **GRIFFEY, Richard, H.** [US/US]; 360 Barsby Street, Vista, CA 92084 (US). **SAMPATH, Rangarajan** [IN/US]; 12223 Mannix Road, San Diego, CA 92129 (US). **HOFSTADLER, Steven, A.** [US/US]; 5014 Viewridge Way, Oceanside, CA 92056 (US). **MCNEIL, John** [US/US]; 6389 Castejon Drive, La Jolla, CA 92037 (US). **HANNIS, James, C.** [US/US]; 983 El Pajodo Place, Vista, CA 92084 (US). **WHITE, Neill** [US/US]; 121 Athena Street, Encinitas, CA 92024 (US).
- (74) Agent: **LEGAARD, Paul, K.**; Cozen O'Connor, 1900 Market Street, Philadelphia, PA 19103 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**  
— with international search report  
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
- (88) Date of publication of the international search report:  
23 June 2005
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS FOR RAPID FORENSIC ANALYSIS OF MITOCHONDRIAL DNA AND CHARACTERIZATION OF MITOCHONDRIAL DNA HETEROPLASMY

(57) Abstract: The present invention provides methods for rapid forensic analysis of mitochondrial DNA and methods for characterizing heteroplasmy of mitochondrial DNA, which can be used to assess the progression of mitochondrial diseases.



WO 2004/053141 A3

## METHODS FOR RAPID FORENSIC ANALYSIS OF MITOCHONDRIAL DNA AND CHARACTERIZATION OF MITOCHONDRIAL DNA HETEROPLASMY

### FIELD OF THE INVENTION

5 This invention relates to the field of mitochondrial DNA analysis. The invention enables the rapid and accurate identification of individuals and eukaryotic organisms by forensics methods as well as characterization of mitochondrial DNA heteroplasmy and prediction of onset of mitochondrial diseases.

### 10 BACKGROUND OF THE INVENTION

Mitochondrial DNA (mtDNA) is found in eukaryotes and differs from nuclear DNA in its location, its sequence, its quantity in the cell, and its mode of inheritance. The nucleus of the cell contains two sets of 23 chromosomes, one paternal set and one maternal set. However, cells may contain hundreds to thousands of mitochondria, each of which may contain several copies  
15 of mtDNA. Nuclear DNA has many more bases than mtDNA, but mtDNA is present in many more copies than nuclear DNA. This characteristic of mtDNA is useful in situations where the amount of DNA in a sample is very limited. Typical sources of DNA recovered from crime scenes include hair, bones, teeth, and body fluids such as saliva, semen, and blood.

In humans, mitochondrial DNA is inherited strictly from the mother (Case J. T. and  
20 Wallace, D.C., *Somatic Cell Genetics*, 1981, 7, 103–108; Giles, R. E. *et al.* *Proc. Natl. Acad. Sci.* 1980, 77, 6715–6719; Hutchison, C.A. *et al.* *Nature*, 1974, 251, 536–538). Thus, the mtDNA sequences obtained from maternally related individuals, such as a brother and a sister or a mother and a daughter, will exactly match each other in the absence of a mutation. This characteristic of mtDNA is advantageous in missing persons cases as reference mtDNA samples can be supplied  
25 by any maternal relative of the missing individual (Ginther, C. *et al.* *Nature Genetics*, 1992, 2, 135–138; Holland, M. M. *et al.* *Journal of Forensic Sciences*, 1993, 38, 542–553; Stoneking, M. *et al.* *American Journal of Human Genetics*, 1991, 48, 370–382).

The human mtDNA genome is approximately 16,569 bases in length and has two general regions: the coding region and the control region. The coding region is responsible for  
30 the production of various biological molecules involved in the process of energy production in the cell. The control region is responsible for regulation of the mtDNA molecule. Two regions of mtDNA within the control region have been found to be highly polymorphic, or variable, within the human population (Greenberg, B. D. *et al.* *Gene*, 1983, 21, 33–49). These two regions are termed “hypervariable Region I” (HVR1), which has an approximate length of 342 base pairs

(bp), and “hypervariable Region II” (HVR2), which has an approximate length of 268 bp. Forensic mtDNA examinations are performed using these two regions because of the high degree of variability found among individuals.

Approximately 610 bp of mtDNA are currently sequenced in forensic mtDNA analysis. Recording and comparing mtDNA sequences would be difficult and potentially confusing if all of the bases were listed. Thus, mtDNA sequence information is recorded by listing only the differences with respect to a reference DNA sequence. By convention, human mtDNA sequences are described using the first complete published mtDNA sequence as a reference (Anderson, S. *et al.*, *Nature*, **1981**, 290, 457–465). This sequence is commonly referred to as the Anderson sequence. It is also called the Cambridge reference sequence or the Oxford sequence. Each base pair in this sequence is assigned a number. Deviations from this reference sequence are recorded as the number of the position demonstrating a difference and a letter designation of the different base. For example, a transition from A to G at Position 263 would be recorded as 263 G. If deletions or insertions of bases are present in the mtDNA, these differences are denoted as well.

In the United States, there are seven laboratories currently conducting forensic mtDNA examinations: the FBI Laboratory; Laboratory Corporation of America (LabCorp) in Research Triangle Park, North Carolina; Mitotyping Technologies in State College, Pennsylvania; the Bode Technology Group (BTG) in Springfield, Virginia; the Armed Forces DNA Identification Laboratory (AFDIL) in Rockville, Maryland; BioSynthesis, Inc. in Lewisville, Texas; and Reliagene in New Orleans, Louisiana.

Mitochondrial DNA analyses have been admitted in criminal proceedings from these laboratories in the following states as of April 1999: Alabama, Arkansas, Florida, Indiana, Illinois, Maryland, Michigan, New Mexico, North Carolina, Pennsylvania, South Carolina, Tennessee, Texas, and Washington. Mitochondrial DNA has also been admitted and used in criminal trials in Australia, the United Kingdom, and several other European countries.

Since 1996, the number of individuals performing mitochondrial DNA analysis at the FBI Laboratory has grown from 4 to 12, with more personnel expected in the near future. Over 150 mitochondrial DNA cases have been completed by the FBI Laboratory as of March 1999, and dozens more await analysis. Forensic courses are being taught by the FBI Laboratory personnel and other groups to educate forensic scientists in the procedures and interpretation of mtDNA sequencing. More and more individuals are learning about the value of mtDNA sequencing for obtaining useful information from evidentiary samples that are small, degraded, or both. Mitochondrial DNA sequencing is becoming known not only as an exclusionary tool but also as a complementary technique for use with other human identification procedures.

Mitochondrial DNA analysis will continue to be a powerful tool for law enforcement officials in the years to come as other applications are developed, validated, and applied to forensic evidence.

Presently, the forensic analysis of mtDNA is rigorous and labor-intensive. Currently, only 1-2 cases per month per analyst can be performed. Several molecular biological techniques are combined to obtain a mtDNA sequence from a sample. The steps of the mtDNA analysis process include primary visual analysis, sample preparation, DNA extraction, polymerase chain reaction (PCR) amplification, postamplification quantification of the DNA, automated DNA sequencing, and data analysis. Another complicating factor in the forensic analysis of mtDNA is the occurrence of heteroplasmy wherein the pool of mtDNAs in a given cell is heterogeneous due to mutations in individual mtDNAs. There are two forms of heteroplasmy found in mtDNA. Sequence heteroplasmy (also known as point heteroplasmy) is the occurrence of more than one base at a particular position or positions in the mtDNA sequence. Length heteroplasmy is the occurrence of more than one length of a stretch of the same base in a mtDNA sequence as a result of insertion of nucleotide residues.

Heteroplasmy is a problem for forensic investigators since a sample from a crime scene can differ from a sample from a suspect by one base pair and this difference may be interpreted as sufficient evidence to eliminate that individual as the suspect. Hair samples from a single individual can contain heteroplasmic mutations at vastly different concentrations and even the root and shaft of a single hair can differ. The detection methods currently available to molecular biologists cannot detect low levels of heteroplasmy. Furthermore, if present, length heteroplasmy will adversely affect sequencing runs by resulting in an out-of-frame sequence that cannot be interpreted.

Mass spectrometry provides detailed information about the molecules being analyzed, including high mass accuracy. It is also a process that can be easily automated. Low-resolution MS may be unreliable when used to detect some known agents, if their spectral lines are sufficiently weak or sufficiently close to those from other living organisms in the sample. DNA chips with specific probes can only determine the presence or absence of specifically anticipated organisms. Because there are hundreds of thousands of species of benign bacteria, some very similar in sequence to threat organisms, even arrays with 10,000 probes lack the breadth needed to detect a particular organism.

Antibodies face more severe diversity limitations than arrays. If antibodies are designed against highly conserved targets to increase diversity, the false alarm problem will dominate,

again because threat organisms are very similar to benign ones. Antibodies are only capable of detecting known agents in relatively uncluttered environments.

Several groups have described detection of PCR products using high resolution electrospray ionization-Fourier transform-ion cyclotron resonance mass spectrometry (ESI-FT-  
5 ICR MS). Accurate measurement of exact mass combined with knowledge of the number of at least one nucleotide allowed calculation of the total base composition for PCR duplex products of approximately 100 base pairs. (Aaserud *et al.*, *J. Am. Soc. Mass Spec.*, 1996, 7, 1266-1269; Muddiman *et al.*, *Anal. Chem.*, 1997, 69, 1543-1549; Wunschel *et al.*, *Anal. Chem.*, 1998, 70, 1203-1207; Muddiman *et al.*, *Rev. Anal. Chem.*, 1998, 17, 1-68). Electrospray ionization-Fourier  
10 transform-ion cyclotron resistance (ESI-FT-ICR) MS may be used to determine the mass of double-stranded, 500 base-pair PCR products via the average molecular mass (Hurst *et al.*, *Rapid Commun. Mass Spec.* 1996, 10, 377-382). The use of matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry for characterization of PCR products has been described. (Muddiman *et al.*, *Rapid Commun. Mass Spec.*, 1999, 13, 1201-1204). However, the  
15 degradation of DNAs over about 75 nucleotides observed with MALDI limited the utility of this method.

U.S. Patent No. 5,849,492 describes a method for retrieval of phylogenetically informative DNA sequences which comprise searching for a highly divergent segment of genomic DNA surrounded by two highly conserved segments, designing the universal primers  
20 for PCR amplification of the highly divergent region, amplifying the genomic DNA by PCR technique using universal primers, and then sequencing the gene to determine the identity of the organism.

U.S. Patent No. 5,965,363 discloses methods for screening nucleic acids for polymorphisms by analyzing amplified target nucleic acids using mass spectrometric techniques  
25 and to procedures for improving mass resolution and mass accuracy of these methods.

WO 99/14375 describes methods, PCR primers and kits for use in analyzing preselected DNA tandem nucleotide repeat alleles by mass spectrometry.

WO 98/12355 discloses methods of determining the mass of a target nucleic acid by mass spectrometric analysis, by cleaving the target nucleic acid to reduce its length, making the  
30 target single-stranded and using MS to determine the mass of the single-stranded shortened target. Also disclosed are methods of preparing a double-stranded target nucleic acid for MS analysis comprising amplification of the target nucleic acid, binding one of the strands to a solid support, releasing the second strand and then releasing the first strand which is then analyzed by MS. Kits for target nucleic acid preparation are also provided.

PCT WO97/33000 discloses methods for detecting mutations in a target nucleic acid nonrandomly fragmenting the target into a set of single-stranded nonrandom length fragments and determining their masses by MS.

U.S. Patent No. 5,605,798 describes a fast and highly accurate mass spectrometer-based process for detecting the presence of a particular nucleic acid in a biological sample for diagnostic purposes.

WO 98/21066 describes processes for determining the sequence of a particular target nucleic acid by mass spectrometry. Processes for detecting a target nucleic acid present in a biological sample by PCR amplification and mass spectrometry detection are disclosed, as are methods for detecting a target nucleic acid in a sample by amplifying the target with primers that contain restriction sites and tags, extending and cleaving the amplified nucleic acid, and detecting the presence of extended product, wherein the presence of a DNA fragment of a mass different from wild-type is indicative of a mutation. Methods of sequencing a nucleic acid via mass spectrometry methods are also described.

WO 97/37041, WO 99/31278 and U.S. Patent No. 5,547,835 describe methods of sequencing nucleic acids using mass spectrometry. U.S. Patent Nos. 5,622,824, 5,872,003 and 5,691,141 describe methods, systems and kits for exonuclease-mediated mass spectrometric sequencing.

Thus, there is a need for a method for bioagent detection and identification which is both specific and rapid, and in which no nucleic acid sequencing is required. The present invention addresses this need.

## SUMMARY OF THE INVENTION

In one aspect, the present invention is directed to method of identifying an individual by obtaining mitochondrial DNA from the individual, amplifying the mitochondrial DNA with intelligent primers to obtain at least one amplification product, determining the molecular mass of the amplification product and comparing the molecular mass with a database of molecular masses calculated from known sequences of mitochondrial DNAs indexed to known individuals, wherein a match between the molecular mass of the amplification product and the calculated molecular mass of a known sequence in the database identifies the individual.



In one embodiment, there is provided a forensic method of mitochondrial DNA analysis comprising the steps of: providing a forensic evidence sample; amplifying two or more segments of mitochondrial DNA obtained from the forensic evidence sample to obtain two or more amplification products; determining the molecular masses of the two or more amplification products by mass spectrometry, without sequencing the two or more amplification products; and comparing the molecular masses of the two or more amplification products with at least one database comprising a plurality of known molecular masses from the two or more segments of mitochondrial DNA from a plurality of subjects thereby reaching a forensic conclusion.

In another embodiment, there is provided a forensic method for tracking the geographic location of a subject comprising the steps of : providing a forensic evidence sample containing mitochondrial DNA obtained from a geographic location; amplifying two or more segments of mitochondrial DNA obtained from the forensic evidence sample to obtain two or more amplification products; determining the molecular masses of the two or more amplification products by mass spectrometry, without sequencing the two or more amplification products; and comparing the molecular masses of the two or more amplification products with at least one database comprising a plurality of known molecular masses from the two or more segments of mitochondrial DNA from a plurality of subjects thereby indicating at least transient presence of the subject at the geographic location.

In another aspect, the present invention is also directed to methods of characterizing the heteroplasmy of a sample of mitochondrial DNA by amplifying the mitochondrial DNA with intelligent primers to obtain a plurality of amplification products, determining the molecular masses and relative abundances of the plurality of amplification products, thereby characterizing the heteroplasmy. Furthermore, the present invention is directed to using these methods to characterize the heteroplasmy of a plurality of samples of mitochondrial DNA taken from an individual at different points of the lifetime of the individual to investigate the rate of naturally occurring mutations in mitochondrial DNA. These methods can also be used to initiate a prediction of the rate of onset of mitochondrial disease.

In one embodiment, there is provided a method of characterizing the heteroplasmy of two or more segments of mitochondrial DNA of a subject comprising the steps of: providing a sample from the subject; amplifying the two or more segments of mitochondrial DNA from the sample with two or more primer pairs to obtain a plurality of amplification products; determining the

molecular masses of the plurality of amplification products of the mitochondrial DNA by mass spectrometry, without sequencing the plurality of amplification products; and determining the base compositions of the plurality of amplification products thereby characterizing the heteroplasmy.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

Figures 1A-1H and Figure 2 are consensus diagrams that show examples of conserved regions from 16S rRNA (Fig. 1A-1, 1A-2, 1A-3, 1A-4, and 1A-5), 23S rRNA (3'-half, Fig. 1B, 1C, and 1D; 5'-half, Fig. 1E-F), 23S rRNA Domain I (Fig. 1G), 23S rRNA Domain IV (Fig. 1H) and 16S rRNA Domain III (Fig. 2) which are suitable for use in the present invention. Lines with arrows are examples of regions to which intelligent primer pairs for PCR are designed. The label for each primer pair represents the starting and ending base number of the amplified region on the consensus diagram. Bases in capital letters are greater than 95% conserved; bases in lower case letters are 90-95% conserved, filled circles are 80-90% conserved; and open circles are less than 80% conserved. The label for each primer pair represents the starting and ending base number of the amplified region on the consensus diagram. The nucleotide sequence of the 16S rRNA consensus sequence is SEQ ID NO:3 and the nucleotide sequence of the 23S rRNA consensus sequence is SEQ ID NO:4.

Figure 2 shows a typical primer amplified region from the 16S rRNA Domain III shown in Figure 1A-1.

Figure 3 is a schematic diagram showing conserved regions in RNase P. Bases in capital letters are greater than 90% conserved; bases in lower case letters are 80-90% conserved; filled circles designate bases which are 70-80% conserved; and open circles designate bases that are less than 70% conserved.

Figure 4 is a schematic diagram of base composition signature determination using nucleotide analog "tags" to determine base composition signatures.

Figure 5 shows the deconvoluted mass spectra of a *Bacillus anthracis* region with and without the mass tag phosphorothioate A (A\*). The two spectra differ in that the measured molecular weight of the mass tag-containing sequence is greater than the unmodified sequence.

Figure 6 shows base composition signature (BCS) spectra from PCR products from *Staphylococcus aureus* (*S. aureus* 16S\_1337F) and *Bacillus anthracis* (*B. anthr.* 16S\_1337F), amplified using the same primers. The two strands differ by only two (AT-->CG) substitutions and are clearly distinguished on the basis of their BCS.

5 Figure 7 shows that a single difference between two sequences (A14 in *B. anthracis* vs. A15 in *B. cereus*) can be easily detected using ESI-TOF mass spectrometry.

Figure 8 is an ESI-TOF of *Bacillus anthracis* spore coat protein sspE 56mer plus calibrant. The signals unambiguously identify *B. anthracis* versus other *Bacillus* species.

10 Figure 9 is an ESI-TOF of a *B. anthracis* synthetic 16S\_1228 duplex (reverse and forward strands). The technique easily distinguishes between the forward and reverse strands.

Figure 10 is an ESI-FTICR-MS of a synthetic *B. anthracis* 16S\_1337 46 base pair duplex.

15 Figure 11 is an ESI-TOF-MS of a 56mer oligonucleotide (3 scans) from the *B. anthracis* saspB gene with an internal mass standard. The internal mass standards are designated by asterisks.

Figure 12 is an ESI-TOF-MS of an internal standard with 5 mM TBA-TFA buffer showing that charge stripping with tributylammonium trifluoroacetate reduces the most abundant charge state from  $[M-8H+]$ 8- to  $[M-3H+]$ 3-.

20 Figure 13 is a portion of a secondary structure defining database according to one embodiment of the present invention, where two examples of selected sequences are displayed graphically thereunder.

Figure 14 is a three dimensional graph demonstrating the grouping of sample molecular weight according to species.

25 Figure 15 is a three dimensional graph demonstrating the grouping of sample molecular weights according to species of virus and mammal infected.

Figure 16 is a three dimensional graph demonstrating the grouping of sample molecular weights according to species of virus, and animal-origin of infectious agent.

30 Figure 17 is a figure depicting how the triangulation method of the present invention provides for the identification of an unknown bioagent without prior knowledge of the unknown agent. The use of different primer sets to distinguish and identify the unknown is also depicted as primer sets I, II and III within this figure. A three dimensional graph depicts all of bioagent space (170), including the unknown bioagent, which after use of primer set I (171) according to a method according to the present invention further differentiates and classifies bioagents according to major classifications (176) which, upon further analysis using primer set II (172)

differentiates the unknown agent (177) from other, known agents (173) and finally, the use of a third primer set (175) further specifies subgroups within the family of the unknown (174).

Figure 18 shows: a) a representative ESI-FTICR mass spectrum of a restriction digest of a 986 bp region of the 16S ribosomal gene from *E. coli* K12 digested with a mixture of *BstNI*, *BsmFI*, *BfaI*, and *NcoI*; b) a deconvoluted representation (neutral mass) of the above spectrum showing the base compositions derived from accurate mass measurements of each fragment; and c) a representative reconstructed restriction map showing complete base composition coverage for nucleotides 1-856. The *NcoI* did not cut .

Figure 19 indicates the process of mtDNA analysis. After amplification by PCR (210), the PCR products were subjected to restriction digests (220) with *RsaI* for HVR1 and a combination of *HpaII*, *HpyCH4IV*, *PacI* and *EaeI* for HVR2 in order to obtain amplicon segments suitable for analysis by FTICR-MS (240). The data were processed to obtain mass data for each amplicon segment (250) which were then compared to the masses calculated for theoretical digests from the FBI mtDNA database by a scoring scheme (260).

Figure 20A indicates predicted and actual mass data with scoring parameters for length heteroplasmy (HVR1-1-outer-variants 1 and 2) in the digest segment from position 94 to 145(variant 1)/146(variant 2) are shown.

Figure 20B indicates that, whereas sequencing fails to resolve the variants due to the length heteroplasmy, mass determination detects multiple species simultaneously and also indicates abundance ratios. In this case, the ratio of variant 1 to variant 2 (short to long alleles) is 1:3.

## DESCRIPTION OF EMBODIMENTS

The present invention provides, *inter alia*, methods for detection and identification of bioagents in an unbiased manner using "bioagent identifying amplicons." "Intelligent primers" are selected to hybridize to conserved sequence regions of nucleic acids derived from a bioagent and which bracket variable sequence regions to yield a bioagent identifying amplicon which can be amplified and which is amenable to molecular mass determination. The molecular mass then provides a means to uniquely identify the bioagent without a requirement for prior knowledge of the possible identity of the bioagent. The molecular mass or corresponding "base composition signature" (BCS) of the amplification product is then matched against a database of molecular masses or base composition signatures. Furthermore, the method can be applied to rapid parallel "multiplex" analyses, the results of which can be employed in a triangulation identification

strategy. The present method provides rapid throughput and does not require nucleic acid sequencing of the amplified target sequence for bioagent detection and identification.

In the context of this invention, a "bioagent" is any organism, cell, or virus, living or dead, or a nucleic acid derived from such an organism, cell or virus. Examples of bioagents include, but are not limited, to cells, including but not limited to, cells, including but not limited to human clinical samples, bacterial cells and other pathogens) viruses, fungi, and protists, parasites, and pathogenicity markers (including but not limited to: pathogenicity islands, antibiotic resistance genes, virulence factors, toxin genes and other bioregulating compounds). Samples may be alive or dead or in a vegetative state (for example, vegetative bacteria or spores) and may be encapsulated or bioengineered. In the context of this invention, a "pathogen" is a bioagent which causes a disease or disorder.

Despite enormous biological diversity, all forms of life on earth share sets of essential, common features in their genomes. Bacteria, for example have highly conserved sequences in a variety of locations on their genomes. Most notable is the universally conserved region of the ribosome. but there are also conserved elements in other non-coding RNAs, including RNase P and the signal recognition particle (SRP) among others. Bacteria have a common set of absolutely required genes. About 250 genes are present in all bacterial species (*Proc. Natl. Acad. Sci. U.S.A.*, **1996**, *93*, 10268; *Science*, **1995**, *270*, 397), including tiny genomes like *Mycoplasma*, *Ureaplasma* and *Rickettsia*. These genes encode proteins involved in translation, replication, recombination and repair, transcription, nucleotide metabolism, amino acid metabolism, lipid metabolism, energy generation, uptake, secretion and the like. Examples of these proteins are DNA polymerase III beta, elongation factor TU, heat shock protein groEL, RNA polymerase beta, phosphoglycerate kinase, NADH dehydrogenase, DNA ligase, DNA topoisomerase and elongation factor G. Operons can also be targeted using the present method. One example of an operon is the bfp operon from enteropathogenic *E. coli*. Multiple core chromosomal genes can be used to classify bacteria at a genus or genus species level to determine if an organism has threat potential. The methods can also be used to detect pathogenicity markers (plasmid or chromosomal) and antibiotic resistance genes to confirm the threat potential of an organism and to direct countermeasures.

Since genetic data provide the underlying basis for identification of bioagents by the methods of the present invention, it is necessary to select segments of nucleic acids which ideally provide enough variability to distinguish each individual bioagent and whose molecular mass is amenable to molecular mass determination. In one embodiment of the present invention, at least one polynucleotide segment is amplified to facilitate detection and analysis in the process of

-10-

identifying the bioagent. Thus, the nucleic acid segments which provide enough variability to distinguish each individual bioagent and whose molecular masses are amenable to molecular mass determination are herein described as "bioagent identifying amplicons." The term "amplicon" as used herein, refers to a segment of a polynucleotide which is amplified in an  
5 amplification reaction.

As used herein, "intelligent primers" are primers that are designed to bind to highly conserved sequence regions of a bioagent identifying amplicon that flank an intervening variable region and yield amplification products which ideally provide enough variability to distinguish each individual bioagent, and which are amenable to molecular mass analysis. By the term  
10 "highly conserved," it is meant that the sequence regions exhibit between about 80-100%, or between about 90-100%, or between about 95-100% identity. The molecular mass of a given amplification product provides a means of identifying the bioagent from which it was obtained, due to the variability of the variable region. Thus design of intelligent primers requires selection of a variable region with appropriate variability to resolve the identity of a given bioagent.  
15 Bioagent identifying amplicons are ideally specific to the identity of the bioagent. A plurality of bioagent identifying amplicons selected in parallel for distinct bioagents which contain the same conserved sequences for hybridization of the same pair of intelligent primers are herein defined as "correlative bioagent identifying amplicons."

In one embodiment, the bioagent identifying amplicon is a portion of a ribosomal RNA  
20 (rRNA) gene sequence. With the complete sequences of many of the smallest microbial genomes now available, it is possible to identify a set of genes that defines "minimal life" and identify composition signatures that uniquely identify each gene and organism. Genes that encode core life functions such as DNA replication, transcription, ribosome structure, translation, and transport are distributed broadly in the bacterial genome and are suitable regions for selection of  
25 bioagent identifying amplicons. Ribosomal RNA (rRNA) genes comprise regions that provide useful base composition signatures. Like many genes involved in core life functions, rRNA genes contain sequences that are extraordinarily conserved across bacterial domains interspersed with regions of high variability that are more specific to each species. The variable regions can be utilized to build a database of base composition signatures. The strategy involves creating a  
30 structure-based alignment of sequences of the small (16S) and the large (23S) subunits of the rRNA genes. For example, there are currently over 13,000 sequences in the ribosomal RNA database that has been created and maintained by Robin Gutell, University of Texas at Austin, and is publicly available on the Institute for Cellular and Molecular Biology web page on the world wide web of the Internet.

There is also a publicly

available rRNA database created and maintained by the University of Antwerp, Belgium on the world wide web of the Internet.

These databases have been analyzed to determine regions that are useful as bioagent identifying amplicons. The characteristics of such regions include: a) between about 80 and 5 100%, or greater than about 95% identity among species of the particular bioagent of interest, of upstream and downstream nucleotide sequences which serve as sequence amplification primer sites; b) an intervening variable region which exhibits no greater than about 5% identity among species; and c) a separation of between about 30 and 1000 nucleotides, or no more than about 50-250 nucleotides, or no more than about 60-100 nucleotides, between the conserved regions.

10 As a non-limiting example, for identification of *Bacillus* species, the conserved sequence regions of the chosen bioagent identifying amplicon must be highly conserved among all *Bacillus* species while the variable region of the bioagent identifying amplicon is sufficiently variable such that the molecular masses of the amplification products of all species of *Bacillus* are distinguishable.

15 Bioagent identifying amplicons amenable to molecular mass determination are either of a length, size or mass compatible with the particular mode of molecular mass determination or compatible with a means of providing a predictable fragmentation pattern in order to obtain predictable fragments of a length compatible with the particular mode of molecular mass determination. Such means of providing a predictable fragmentation pattern of an amplification 20 product include, but are not limited to, cleavage with restriction enzymes or cleavage primers, for example.

Identification of bioagents can be accomplished at different levels using intelligent primers suited to resolution of each individual level of identification. "Broad range survey" intelligent primers are designed with the objective of identifying a bioagent as a member of a 25 particular division of bioagents. A "bioagent division" is defined as group of bioagents above the species level and includes but is not limited to: orders, families, classes, clades, genera or other such groupings of bioagents above the species level. As a non-limiting example, members of the *Bacillus/Clostridia* group or gamma-proteobacteria group may be identified as such by employing broad range survey intelligent primers such as primers which target 16S or 23S 30 ribosomal RNA.

In some embodiments, broad range survey intelligent primers are capable of identification of bioagents at the species level. One main advantage of the detection methods of the present invention is that the broad range survey intelligent primers need not be specific for a particular bacterial species, or even genus, such as *Bacillus* or *Streptomyces*. Instead, the primers

-12-

recognize highly conserved regions across hundreds of bacterial species including, but not limited to, the species described herein. Thus, the same broad range survey intelligent primer pair can be used to identify any desired bacterium because it will bind to the conserved regions that flank a variable region specific to a single species, or common to several bacterial species, allowing unbiased nucleic acid amplification of the intervening sequence and determination of its molecular weight and base composition. For example, the 16S\_971-1062, 16S\_1228-1310 and 16S\_1100-1188 regions are 98-99% conserved in about 900 species of bacteria (16S=16S rRNA, numbers indicate nucleotide position). In one embodiment of the present invention, primers used in the present method bind to one or more of these regions or portions thereof.

Due to their overall conservation, the flanking rRNA primer sequences serve as good intelligent primer binding sites to amplify the nucleic acid region of interest for most, if not all, bacterial species. The intervening region between the sets of primers varies in length and/or composition, and thus provides a unique base composition signature. Examples of intelligent primers that amplify regions of the 16S and 23S rRNA are shown in Figures 1A-1H. A typical primer amplified region in 16S rRNA is shown in Figure 2. The arrows represent primers that bind to highly conserved regions which flank a variable region in 16S rRNA domain III. The amplified region is the stem-loop structure under "1100-1188." It is advantageous to design the broad range survey intelligent primers to minimize the number of primers required for the analysis, and to allow detection of multiple members of a bioagent division using a single pair of primers. The advantage of using broad range survey intelligent primers is that once a bioagent is broadly identified, the process of further identification at species and sub-species levels is facilitated by directing the choice of additional intelligent primers.

"Division-wide" intelligent primers are designed with an objective of identifying a bioagent at the species level. As a non-limiting example, a *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis* can be distinguished from each other using division-wide intelligent primers. Division-wide intelligent primers are not always required for identification at the species level because broad range survey intelligent primers may provide sufficient identification resolution to accomplishing this identification objective.

"Drill-down" intelligent primers are designed with an objective of identifying a sub-species characteristic of a bioagent. A "sub-species characteristic" is defined as a property imparted to a bioagent at the sub-species level of identification as a result of the presence or absence of a particular segment of nucleic acid. Such sub-species characteristics include, but are not limited to, strains, sub-types, pathogenicity markers such as antibiotic resistance genes,



pathogenicity islands, toxin genes and virulence factors. Identification of such sub-species characteristics is often critical for determining proper clinical treatment of pathogen infections.

#### *Chemical Modifications of Intelligent Primers*

Ideally, intelligent primer hybridization sites are highly conserved in order to facilitate  
5 the hybridization of the primer. In cases where primer hybridization is less efficient due to lower levels of conservation of sequence, intelligent primers can be chemically modified to improve the efficiency of hybridization.

For example, because any variation (due to codon wobble in the 3<sup>rd</sup> position) in these conserved regions among species is likely to occur in the third position of a DNA triplet,  
10 oligonucleotide primers can be designed such that the nucleotide corresponding to this position is a base which can bind to more than one nucleotide, referred to herein as a "universal base." For example, under this "wobble" pairing, inosine (I) binds to U, C or A; guanine (G) binds to U or C, and uridine (U) binds to U or C. Other examples of universal bases include nitroindoles such as 5-nitroindole or 3-nitropyrrole (Loakes *et al.*, *Nucleosides and Nucleotides*, 1995, 14, 1001-  
15 1003), the degenerate nucleotides dP or dK (Hill *et al.*), an acyclic nucleoside analog containing 5-nitroindazole (Van Aerschot *et al.*, *Nucleosides and Nucleotides*, 1995, 14, 1053-1056) or the purine analog 1-(2-deoxy- $\beta$ -D-ribofuranosyl)-imidazole-4-carboxamide (Sala *et al.*, *Nucl. Acids Res.*, 1996, 24, 3302-3306).

In another embodiment of the invention, to compensate for the somewhat weaker  
20 binding by the "wobble" base, the oligonucleotide primers are designed such that the first and second positions of each triplet are occupied by nucleotide analogs which bind with greater affinity than the unmodified nucleotide. Examples of these analogs include, but are not limited to, 2,6-diaminopurine which binds to thymine, propyne T which binds to adenine and propyne C and phenoxazines, including G-clamp, which binds to G. Propynylated pyrimidines are described  
25 in U.S. Patent Nos. 5,645,985, 5,830,653 and 5,484,908, each of which is commonly owned. Propynylated primers are claimed in U.S Patent Application Publication No. 20040265802 which is also commonly owned. Phenoxazines are described in U.S. Patent Nos. 5,502,177, 5,763,588, and 6,005,096. G-clamps are described in U.S. Patent Nos. 6,007,992 and 6,028,183.

30

A theoretically ideal bioagent detector would identify, quantify, and report the complete nucleic acid sequence of every bioagent that reached the sensor. The complete sequence of the nucleic acid component of a pathogen would provide all relevant information about the threat, including its identity and the presence of drug-resistance or pathogenicity markers. This ideal has

not yet been achieved. However, the present invention provides a straightforward strategy for obtaining information with the same practical value based on analysis of bioagent identifying amplicons by molecular mass determination.

In some cases, a molecular mass of a given bioagent identifying amplicon alone does not provide enough resolution to unambiguously identify a given bioagent. For example, the molecular mass of the bioagent identifying amplicon obtained using the intelligent primer pair “16S\_971” would be 55622 Da for both *E. coli* and *Salmonella typhimurium*. However, if additional intelligent primers are employed to analyze additional bioagent identifying amplicons, a “triangulation identification” process is enabled. For example, the “16S\_1100” intelligent primer pair yields molecular masses of 55009 and 55005 Da for *E. coli* and *Salmonella typhimurium*, respectively. Furthermore, the “23S\_855” intelligent primer pair yields molecular masses of 42656 and 42698 Da for *E. coli* and *Salmonella typhimurium*, respectively. In this basic example, the second and third intelligent primer pairs provided the additional “fingerprinting” capability or resolution to distinguish between the two bioagents.

In another embodiment, the triangulation identification process is pursued by measuring signals from a plurality of bioagent identifying amplicons selected within multiple core genes. This process is used to reduce false negative and false positive signals, and enable reconstruction of the origin of hybrid or otherwise engineered bioagents. In this process, after identification of multiple core genes, alignments are created from nucleic acid sequence databases. The alignments are then analyzed for regions of conservation and variation, and bioagent identifying amplicons are selected to distinguish bioagents based on specific genomic differences. For example, identification of the three part toxin genes typical of *B. anthracis* (Bowen *et al.*, *J. Appl. Microbiol.*, **1999**, *87*, 270-278) in the absence of the expected signatures from the *B. anthracis* genome would suggest a genetic engineering event.

The triangulation identification process can be pursued by characterization of bioagent identifying amplicons in a massively parallel fashion using the polymerase chain reaction (PCR), such as multiplex PCR, and mass spectrometric (MS) methods. Sufficient quantities of nucleic acids should be present for detection of bioagents by MS. A wide variety of techniques for preparing large amounts of purified nucleic acids or fragments thereof are well known to those of skill in the art. PCR requires one or more pairs of oligonucleotide primers that bind to regions which flank the target sequence(s) to be amplified. These primers prime synthesis of a different strand of DNA, with synthesis occurring in the direction of one primer towards the other primer. The primers, DNA to be amplified, a thermostable DNA polymerase (e.g. *Taq* polymerase), the four deoxynucleotide triphosphates, and a buffer are combined to initiate DNA synthesis. The

-15-

solution is denatured by heating, then cooled to allow annealing of newly added primer, followed by another round of DNA synthesis. This process is typically repeated for about 30 cycles, resulting in amplification of the target sequence.

Although the use of PCR is suitable, other nucleic acid amplification techniques may also be used, including ligase chain reaction (LCR) and strand displacement amplification (SDA). The high-resolution MS technique allows separation of bioagent spectral lines from background spectral lines in highly cluttered environments.

In another embodiment, the detection scheme for the PCR products generated from the bioagent(s) incorporates at least three features. First, the technique simultaneously detects and differentiates multiple (generally about 6-10) PCR products. Second, the technique provides a molecular mass that uniquely identifies the bioagent from the possible primer sites. Finally, the detection technique is rapid, allowing multiple PCR reactions to be run in parallel.

Mass spectrometry (MS)-based detection of PCR products provides a means for determination of BCS which has several advantages. MS is intrinsically a parallel detection scheme without the need for radioactive or fluorescent labels, since every amplification product is identified by its molecular mass. The current state of the art in mass spectrometry is such that less than femtomole quantities of material can be readily analyzed to afford information about the molecular contents of the sample. An accurate assessment of the molecular mass of the material can be quickly obtained, irrespective of whether the molecular weight of the sample is several hundred, or in excess of one hundred thousand atomic mass units (amu) or Daltons. Intact molecular ions can be generated from amplification products using one of a variety of ionization techniques to convert the sample to gas phase. These ionization methods include, but are not limited to, electrospray ionization (ESI), matrix-assisted laser desorption ionization (MALDI) and fast atom bombardment (FAB). For example, MALDI of nucleic acids, along with examples of matrices for use in MALDI of nucleic acids, are described in WO 98/54751 (Genetrace, Inc.).

In some embodiments, large DNAs and RNAs, or large amplification products therefrom, can be digested with restriction endonucleases prior to ionization. Thus, for example, an amplification product that was 10 kDa could be digested with a series of restriction endonucleases to produce a panel of, for example, 100 Da fragments. Restriction endonucleases and their sites of action are well known to the skilled artisan. In this manner, mass spectrometry can be performed for the purposes of restriction mapping.

Upon ionization, several peaks are observed from one sample due to the formation of ions with different charges. Averaging the multiple readings of molecular mass obtained from a

single mass spectrum affords an estimate of molecular mass of the bioagent. Electrospray ionization mass spectrometry (ESI-MS) is particularly useful for very high molecular weight polymers such as proteins and nucleic acids having molecular weights greater than 10 kDa, since it yields a distribution of multiply-charged molecules of the sample without causing a significant amount of fragmentation.

The mass detectors used in the methods of the present invention include, but are not limited to, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), ion trap, quadrupole, magnetic sector, time of flight (TOF), Q-TOF, and triple quadrupole.

In general, the mass spectrometric techniques which can be used in the present invention include, but are not limited to, tandem mass spectrometry, infrared multiphoton dissociation and pyrolytic gas chromatography mass spectrometry (PGC-MS). In one embodiment of the invention, the bioagent detection system operates continually in bioagent detection mode using pyrolytic GC-MS without PCR for rapid detection of increases in biomass (for example, increases in fecal contamination of drinking water or of germ warfare agents). To achieve minimal latency, a continuous sample stream flows directly into the PGC-MS combustion chamber. When an increase in biomass is detected, a PCR process is automatically initiated. Bioagent presence produces elevated levels of large molecular fragments from, for example, about 100-7,000 Da which are observed in the PGC-MS spectrum. The observed mass spectrum is compared to a threshold level and when levels of biomass are determined to exceed a predetermined threshold, the bioagent classification process described hereinabove (combining PCR and MS, such as FT-ICR MS) is initiated. Optionally, alarms or other processes (halting ventilation flow, physical isolation) are also initiated by this detected biomass level.

The accurate measurement of molecular mass for large DNAs is limited by the adduction of cations from the PCR reaction to each strand, resolution of the isotopic peaks from natural abundance  $^{13}\text{C}$  and  $^{15}\text{N}$  isotopes, and assignment of the charge state for any ion. The cations are removed by in-line dialysis using a flow-through chip that brings the solution containing the PCR products into contact with a solution containing ammonium acetate in the presence of an electric field gradient orthogonal to the flow. The latter two problems are addressed by operating with a resolving power of  $>100,000$  and by incorporating isotopically depleted nucleotide triphosphates into the DNA. The resolving power of the instrument is also a consideration. At a resolving power of 10,000, the modeled signal from the  $[\text{M}-14\text{H}]^{14-}$  charge state of an 84mer PCR product is poorly characterized and assignment of the charge state or exact mass is impossible. At a resolving power of 33,000, the peaks from the individual isotopic components are visible. At a resolving power of 100,000, the isotopic peaks are resolved to the

baseline and assignment of the charge state for the ion is straightforward. The [<sup>13</sup>C,<sup>15</sup>N]-depleted triphosphates are obtained, for example, by growing microorganisms on depleted media and harvesting the nucleotides (Batey *et al.*, *Nucl. Acids Res.*, 1992, 20, 4515-4523).

While mass measurements of intact nucleic acid regions are believed to be adequate to  
5 determine most bioagents, tandem mass spectrometry (MS<sup>n</sup>) techniques may provide more definitive information pertaining to molecular identity or sequence. Tandem MS involves the coupled use of two or more stages of mass analysis where both the separation and detection steps are based on mass spectrometry. The first stage is used to select an ion or component of a sample from which further structural information is to be obtained. The selected ion is then fragmented  
10 using, e.g., blackbody irradiation, infrared multiphoton dissociation, or collisional activation. For example, ions generated by electrospray ionization (ESI) can be fragmented using IR multiphoton dissociation. This activation leads to dissociation of glycosidic bonds and the phosphate backbone, producing two series of fragment ions, called the *w*-series (having an intact 3' terminus and a 5' phosphate following internal cleavage) and the *a*-Base series (having an  
15 intact 5' terminus and a 3' furan).

The second stage of mass analysis is then used to detect and measure the mass of these resulting fragments of product ions. Such ion selection followed by fragmentation routines can be performed multiple times so as to essentially completely dissect the molecular sequence of a sample.

20 If there are two or more targets of similar molecular mass, or if a single amplification reaction results in a product which has the same mass as two or more bioagent reference standards, they can be distinguished by using mass-modifying "tags." In this embodiment of the invention, a nucleotide analog or "tag" is incorporated during amplification (e.g., a 5-(trifluoromethyl) deoxythymidine triphosphate) which has a different molecular weight than the  
25 unmodified base so as to improve distinction of masses. Such tags are described in, for example, PCT WO97/33000. This further limits the number of possible base compositions consistent with any mass. For example, 5-(trifluoromethyl)deoxythymidine triphosphate can be used in place of dTTP in a separate nucleic acid amplification reaction. Measurement of the mass shift between a conventional amplification  
30 product and the tagged product is used to quantitate the number of thymidine nucleotides in each of the single strands. Because the strands are complementary, the number of adenosine nucleotides in each strand is also determined.

In another amplification reaction, the number of G and C residues in each strand is determined using, for example, the cytidine analog 5-methylcytosine (5-meC) or propyne C. The

-18-

combination of the A/T reaction and G/C reaction, followed by molecular weight determination, provides a unique base composition. This method is summarized in Figure 4 and Table 1.

Table 1

Mass tag	Double strand sequence	Single strand Sequence	Total mass this strand	Base info this strand	Base info other strand	Total base comp. Top strand	Total base comp. Bottom strand
T*.mass (T*-T) = x	T*ACGT*ACGT* * AT*GCAT*GCA	T*ACGT*ACGT*	3x	3T	3A	3T 2A 2C 2G	3A 2T 2G 2C
		AT*GCAT*GCA	2x	2T	2A		
C*.mass (C*-C) = y	TAC*GTAC*GT ATGC*ATGC*A	TAC*GTAC*GT	2x	2C	2G		
		ATGC*ATGC*A	2x	2C	2G		

5 The mass tag phosphorothioate A (A\*) was used to distinguish a *Bacillus anthracis* cluster. The *B. anthracis* (A<sub>14</sub>G<sub>9</sub>C<sub>14</sub>T<sub>9</sub>) had an average MW of 14072.26, and the *B. anthracis* (A<sub>1</sub>A\*<sub>13</sub>G<sub>9</sub>C<sub>14</sub>T<sub>9</sub>) had an average molecular weight of 14281.11 and the phosphorothioate A had an average molecular weight of +16.06 as determined by ESI-TOF MS. The deconvoluted spectra are shown in Figure 5.

10 In another example, assume the measured molecular masses of each strand are 30,000.115Da and 31,000.115 Da respectively, and the measured number of dT and dA residues are (30,28) and (28,30). If the molecular mass is accurate to 100 ppm, there are 7 possible combinations of dG+dC possible for each strand. However, if the measured molecular mass is accurate to 10 ppm, there are only 2 combinations of dG+dC, and at 1 ppm accuracy there is  
15 only one possible base composition for each strand.

Signals from the mass spectrometer may be input to a maximum-likelihood detection and classification algorithm such as is widely used in radar signal processing. The detection processing uses matched filtering of BCS observed in mass-basecount space and allows for detection and subtraction of signatures from known, harmless organisms, and for detection of  
20 unknown bioagent threats. Comparison of newly observed bioagents to known bioagents is also possible, for estimation of threat level, by comparing their BCS to those of known organisms and

to known forms of pathogenicity enhancement, such as insertion of antibiotic resistance genes or toxin genes.

Processing may end with a Bayesian classifier using log likelihood ratios developed from the observed signals and average background levels. The program emphasizes performance  
5 predictions culminating in probability-of-detection versus probability-of-false-alarm plots for conditions involving complex backgrounds of naturally occurring organisms and environmental contaminants. Matched filters consist of a priori expectations of signal values given the set of primers used for each of the bioagents. A genomic sequence database (e.g. GenBank) is used to define the mass basecount matched filters. The database contains known threat agents and benign  
10 background organisms. The latter is used to estimate and subtract the signature produced by the background organisms. A maximum likelihood detection of known background organisms is implemented using matched filters and a running-sum estimate of the noise covariance. Background signal strengths are estimated and used along with the matched filters to form signatures which are then subtracted. the maximum likelihood process is applied to this “cleaned  
15 up” data in a similar manner employing matched filters for the organisms and a running-sum estimate of the noise-covariance for the cleaned up data.

Although the molecular mass of amplification products obtained using intelligent primers provides a means for identification of bioagents, conversion of molecular mass data to a base composition signature is useful for certain analyses. As used herein, a “base composition  
20 signature” (BCS) is the exact base composition determined from the molecular mass of a bioagent identifying amplicon. In one embodiment, a BCS provides an index of a specific gene in a specific organism.

Base compositions, like sequences, vary slightly from isolate to isolate within species. It is possible to manage this diversity by building “base composition probability clouds” around the  
25 composition constraints for each species. This permits identification of organisms in a fashion similar to sequence analysis. A “pseudo four-dimensional plot” can be used to visualize the concept of base composition probability clouds (Figure 18). Optimal primer design requires optimal choice of bioagent identifying amplicons and maximizes the separation between the base composition signatures of individual bioagents. Areas where clouds overlap indicate regions that  
30 may result in a misclassification, a problem which is overcome by selecting primers that provide information from different bioagent identifying amplicons, ideally maximizing the separation of base compositions. Thus, one aspect of the utility of an analysis of base composition probability clouds is that it provides a means for screening primer sets in order to avoid potential misclassifications of BCS and bioagent identity. Another aspect of the utility of base

composition probability clouds is that they provide a means for predicting the identity of a bioagent whose exact measured BCS was not previously observed and/or indexed in a BCS database due to evolutionary transitions in its nucleic acid sequence.

It is important to note that, in contrast to probe-based techniques, mass spectrometry  
5 determination of base composition does not require prior knowledge of the composition in order to make the measurement, only to interpret the results. In this regard, the present invention provides bioagent classifying information similar to DNA sequencing and phylogenetic analysis at a level sufficient to detect and identify a given bioagent. Furthermore, the process of  
10 determination of a previously unknown BCS for a given bioagent (for example, in a case where sequence information is unavailable) has downstream utility by providing additional bioagent indexing information with which to populate BCS databases. The process of future bioagent identification is thus greatly improved as more BCS indexes become available in the BCS databases.

Another embodiment of the present invention is a method of surveying bioagent  
15 samples that enables detection and identification of all bacteria for which sequence information is available using a set of twelve broad-range intelligent PCR primers. Six of the twelve primers are "broad range survey primers" herein defined as primers targeted to broad divisions of bacteria (for example, the *Bacillus/Clostridia* group or gamma-proteobacteria). The other six primers of the group of twelve primers are "division-wide" primers herein defined as primers  
20 which provide more focused coverage and higher resolution. This method enables identification of nearly 100% of known bacteria at the species level. A further example of this embodiment of the present invention is a method herein designated "survey/drill-down" wherein a subspecies characteristic for detected bioagents is obtained using additional primers. Examples of such a subspecies characteristic include but are not limited to: antibiotic resistance, pathogenicity  
25 island, virulence factor, strain type, sub-species type, and clade group. Using the survey/drill-down method, bioagent detection, confirmation and a subspecies characteristic can be provided within hours. Moreover, the survey/drill-down method can be focused to identify bioengineering events such as the insertion of a toxin gene into a bacterial species that does not normally make the toxin.

30 The present methods allow extremely rapid and accurate detection and identification of bioagents compared to existing methods. Furthermore, this rapid detection and identification is possible even when sample material is impure. The methods leverage ongoing biomedical research in virulence, pathogenicity, drug resistance and genome sequencing into a method which provides greatly improved sensitivity, specificity and reliability compared to existing



methods, with lower rates of false positives. Thus, the methods are useful in a wide variety of fields, including, but not limited to, those fields discussed below.

In other embodiments of the invention, the methods disclosed herein can be used for forensics. As used herein, "forensics" is the study of evidence discovered at a crime or accident scene and used in a court of law. "Forensic science" is any science used for the purposes of the law, in particular the criminal justice system, and therefore provides impartial scientific evidence for use in the courts of law, and in a criminal investigation and trial. Forensic science is a multidisciplinary subject, drawing principally from chemistry and biology, but also from physics, geology, psychology and social science, for example.

The process of human identification is a common objective of forensics investigations. For example, there exists a need for rapid identification of humans wherein human remains and/or biological samples are analyzed. Such remains or samples may be associated with war-related casualties, aircraft crashes, and acts of terrorism, for example. Analysis of mtDNA enables a rule-in/rule-out identification process for persons for whom DNA profiles from a maternal relative are available. Human identification by analysis of mtDNA can also be applied to human remains and/or biological samples obtained from crime scenes.

Nucleic acid segments which provide enough variability to distinguish each individual bioagent and whose molecular masses are amenable to molecular mass determination are herein described as "bioagent identifying amplicons." The bioagent identifying amplicons used in the present invention for analysis of mitochondrial DNA are defined as "mitochondrial DNA identifying amplicons."

Forensic scientists generally use two highly variable regions of human mtDNA for analysis. These regions are designated "hypervariable regions 1 and 2" (HVR1 and HVR2 – which contain 341 and 267 base pairs respectively). These hypervariable regions, or portions thereof, provide one non-limiting example of mitochondrial DNA identifying amplicons.

A mtDNA analysis begins when total genomic DNA is extracted from biological material, such as a tooth, blood sample, or hair. The polymerase chain reaction (PCR) is then used to amplify, or create many copies of, the two hypervariable portions of the non-coding region of the mtDNA molecule, using flanking primers. Care is taken to eliminate the introduction of exogenous DNA during both the extraction and amplification steps via methods such as the use of pre-packaged sterile equipment and reagents, aerosol-resistant barrier pipette tips, gloves, masks, and lab coats, separation of pre- and post-amplification areas in the lab using dedicated reagents for each, ultraviolet irradiation of equipment, and autoclaving of tubes and reagent stocks. In casework, questioned samples are always processed before known samples and

they are processed in different laboratory rooms. When adequate amounts of PCR product are amplified to provide all the necessary information about the two hypervariable regions, sequencing reactions are performed. These chemical reactions use each PCR product as a template to create a new complementary strand of DNA in which some of the nucleotide residues  
5 that make up the DNA sequence are labeled with dye. The strands created in this stage are then separated according to size by an automated sequencing machine that uses a laser to "read" the sequence, or order, of the nucleotide bases. Where possible, the sequences of both hypervariable regions are determined on both strands of the double-stranded DNA molecule, with sufficient redundancy to confirm the nucleotide substitutions that characterize that particular sample. At  
10 least two forensic analysts independently assemble the sequence and then compare it to a standard, commonly used, reference sequence. The entire process is then repeated with a known sample, such as blood or saliva collected from a known individual. The sequences from both samples, about 780 bases long each, are compared to determine if they match. The analysts assess the results of the analysis and determine if any portions of it need to be repeated. Finally,  
15 in the event of an inclusion or match, the SWGDAM mtDNA database, which is maintained by the FBI, is searched for the mitochondrial sequence that has been observed for the samples. The analysts can then report the number of observations of this type based on the nucleotide positions that have been read. A written report can be provided to the submitting agency.

In one embodiment of the present invention, the methods disclosed herein for rapid  
20 identification of bioagents using base composition signatures are employed for analysis of human mtDNA. The advantages provided by this embodiment of the present invention include, but are not limited to, efficiency of mass determination of amplicons over sequence determination, and the ability to resolve mixtures of mtDNA amplicons arising from heteroplasmy. Such mixtures invariably cause sequencing failures.

25 In another embodiment of the present invention, the methods disclosed herein for mtDNA analysis can be used to identify the presence of heteroplasmic variants and to determine their relative abundances. As used herein, "mitochondrial diseases" are defined as diseases arising from defects in mitochondrial function which often arise as a result of mutations and heteroplasmy. If the defect is in the mitochondrial rather than the nuclear genome unusual  
30 patterns of inheritance can be observed. This embodiment can be used to determine rates of naturally occurring mutations contributing to heteroplasmy and to predict the onset of mitochondrial diseases arising from heteroplasmy. Examples of mitochondrial diseases include, but are not limited to: Alpers Disease, Barth syndrome, Beta-oxidation Defects, Carnitine-Acyl-Carnitine Deficiency, Carnitine Deficiency, Co-Enzyme Q10 Deficiency, Complex I Deficiency,

Complex II Deficiency, Complex III Deficiency, Complex IV Deficiency, Complex V  
Deficiency, COX Deficiency, CPEO, CPT I Deficiency, CPT II Deficiency, Glutaric Aciduria  
Type II, KSS, Lactic Acidosis, LCAD, LCHAD, Leigh Disease or Syndrome, LHON, Lethal  
Infantile Cardiomyopathy, Luft Disease, MAD, MCA, MELAS, MERRF, Mitochondrial  
5 Cytopathy, Mitochondrial DNA Depletion, Mitochondrial Encephalopathy, Mitochondrial  
Myopathy, MNGIE, NARP, Pearson Syndrome, Pyruvate Carboxylase Deficiency, Pyruvate  
Dehydrogenase Deficiency, Respiratory Chain, SCAD, SCHAD, VLCAD, and the like.

In another embodiment of the present invention, the methods disclosed herein can be  
10 used to rapidly determine the identity of a fungus or a protist by analysis of its mtDNA.

In addition, epidemiologists, for example, can use the present methods to determine the  
geographic origin of a particular strain of a protist or fungus. For example, a particular strain of  
bacteria or virus may have a sequence difference that is associated with a particular area of a  
country or the world and identification of such a sequence difference can lead to the  
15 identification of the geographic origin and epidemiological tracking of the spread of the  
particular disease, disorder or condition associated with the detected protist or fungus. In  
addition, carriers of particular DNA or diseases, such as mammals, non-mammals, birds, insects,  
and plants, can be tracked by screening their mtDNA. Diseases, such as malaria, can be tracked  
by screening the mtDNA of commensals such as mosquitoes.

20 The present method can also be used to detect single nucleotide polymorphisms (SNPs),  
or multiple nucleotide polymorphisms, rapidly and accurately. A SNP is defined as a single base  
pair site in the genome that is different from one individual to another. The difference can be  
expressed either as a deletion, an insertion or a substitution, and is frequently linked to a disease  
state. Because they occur every 100-1000 base pairs, SNPs are the most frequently bound type of  
25 genetic marker in the human genome.

For example, sickle cell anemia results from an A-T transition, which encodes a valine  
rather than a glutamic acid residue. Oligonucleotide primers may be designed such that they bind  
to sequences that flank a SNP site, followed by nucleotide amplification and mass determination  
of the amplified product. Because the molecular masses of the resulting product from an  
30 individual who does not have sickle cell anemia is different from that of the product from an  
individual who has the disease, the method can be used to distinguish the two individuals. Thus,  
the method can be used to detect any known SNP in an individual and thus diagnose or  
determine increased susceptibility to a disease or condition.

In one embodiment, blood is drawn from an individual and peripheral blood mononuclear cells (PBMC) are isolated and simultaneously tested, preferably in a high-throughput screening method, for one or more SNPs using appropriate primers based on the known sequences which flank the SNP region. The National Center for Biotechnology Information maintains a publicly available database of SNPs on the world wide web of the Internet.

The method of the present invention can also be used for blood typing. The gene encoding A, B or O blood type can differ by four single nucleotide polymorphisms. If the gene contains the sequence CGTGGTGACCCTT (SEQ ID NO:5), antigen A results. If the gene contains the sequence CGTCGTCACCGCTA (SEQ ID NO:6) antigen B results. If the gene contains the sequence CGTGGT-ACCCCTT (SEQ ID NO:7), blood group O results (“-” indicates a deletion). These sequences can be distinguished by designing a single primer pair which flanks these regions, followed by amplification and mass determination.

While the present invention has been described with specificity in accordance with certain of its embodiments, the following examples serve only to illustrate the invention and are not intended to limit the same.

## EXAMPLES

### Example 1: Nucleic Acid Isolation and PCR

In one embodiment, nucleic acid is isolated from the organisms and amplified by PCR using standard methods prior to BCS determination by mass spectrometry. Nucleic acid is isolated, for example, by detergent lysis of bacterial cells, centrifugation and ethanol precipitation. Nucleic acid isolation methods are described in, for example, *Current Protocols in Molecular Biology* (Ausubel et al.) and *Molecular Cloning; A Laboratory Manual* (Sambrook et al.). The nucleic acid is then amplified using standard methodology, such as PCR, with primers which bind to conserved regions of the nucleic acid which contain an intervening variable sequence as described below.

*General Genomic DNA Sample Prep Protocol:* Raw samples are filtered using Supor-200 0.2 µm membrane syringe filters (VWR International) . Samples are transferred to 1.5 ml eppendorf tubes pre-filled with 0.45 g of 0.7 mm Zirconia beads followed by the addition of 350 µl of ATL buffer (Qiagen, Valencia, CA). The samples are subjected to bead beating for 10 minutes at a frequency of 19 1/s in a Retsch Vibration Mill (Retsch). After centrifugation, samples are transferred to an S-block plate (Qiagen) and DNA isolation is completed with a BioRobot 8000 nucleic acid isolation robot (Qiagen).

*Swab Sample Protocol:* Allegiance S/P brand culture swabs and collection/transport system are used to collect samples. After drying, swabs are placed in 17x100 mm culture tubes (VWR International) and the genomic nucleic acid isolation is carried out automatically with a Qiagen Mdx robot and the Qiagen QIAamp DNA Blood BioRobot Mdx genomic preparation kit  
5 (Qiagen, Valencia, CA).

## **Example 2: Mass spectrometry**

*FTICR Instrumentation:* The FTICR instrument is based on a 7 tesla actively shielded superconducting magnet and modified Bruker Daltonics Apex II 70e ion optics and vacuum  
10 chamber. The spectrometer is interfaced to a LEAP PAL autosampler and a custom fluidics control system for high throughput screening applications. Samples are analyzed directly from 96-well or 384-well microtiter plates at a rate of about 1 sample/minute. The Bruker data-acquisition platform is supplemented with a lab-built ancillary NT datastation which controls the autosampler and contains an arbitrary waveform generator capable of generating complex rf-  
15 excite waveforms (frequency sweeps, filtered noise, stored waveform inverse Fourier transform (SWIFT), etc.) for sophisticated tandem MS experiments. For oligonucleotides in the 20-30-mer regime typical performance characteristics include mass resolving power in excess of 100,000 (FWHM), low ppm mass measurement errors, and an operable  $m/z$  range between 50 and 5000  $m/z$ .

*Modified ESI Source:* In sample-limited analyses, analyte solutions are delivered at 150  
20 nL/minute to a 30 mm i.d. fused-silica ESI emitter mounted on a 3-D micromanipulator. The ESI ion optics consists of a heated metal capillary, an rf-only hexapole, a skimmer cone, and an auxiliary gate electrode. The 6.2 cm rf-only hexapole is comprised of 1 mm diameter rods and is operated at a voltage of 380 Vpp at a frequency of 5 MHz. A lab-built electro-mechanical shutter  
25 can be employed to prevent the electrospray plume from entering the inlet capillary unless triggered to the "open" position via a TTL pulse from the data station. When in the "closed" position, a stable electrospray plume is maintained between the ESI emitter and the face of the shutter. The back face of the shutter arm contains an elastomeric seal that can be positioned to form a vacuum seal with the inlet capillary. When the seal is removed, a 1 mm gap between the  
30 shutter blade and the capillary inlet allows constant pressure in the external ion reservoir regardless of whether the shutter is in the open or closed position. When the shutter is triggered, a "time slice" of ions is allowed to enter the inlet capillary and is subsequently accumulated in the external ion reservoir. The rapid response time of the ion shutter (< 25 ms) provides

reproducible, user defined intervals during which ions can be injected into and accumulated in the external ion reservoir.

*Apparatus for Infrared Multiphoton Dissociation:* A 25 watt CW CO<sub>2</sub> laser operating at 10.6 μm has been interfaced to the spectrometer to enable infrared multiphoton dissociation (IRMPD) for oligonucleotide sequencing and other tandem MS applications. An aluminum optical bench is positioned approximately 1.5 m from the actively shielded superconducting magnet such that the laser beam is aligned with the central axis of the magnet. Using standard IR-compatible mirrors and kinematic mirror mounts, the unfocused 3 mm laser beam is aligned to traverse directly through the 3.5 mm holes in the trapping electrodes of the FTICR trapped ion cell and longitudinally traverse the hexapole region of the external ion guide finally impinging on the skimmer cone. This scheme allows IRMPD to be conducted in an *m/z* selective manner in the trapped ion cell (e.g. following a SWIFT isolation of the species of interest), or in a broadband mode in the high pressure region of the external ion reservoir where collisions with neutral molecules stabilize IRMPD-generated metastable fragment ions resulting in increased fragment ion yield and sequence coverage.

### Example 3: Identification of Bioagents

Table 2 shows a small cross section of a database of calculated molecular masses for over 9 primer sets and approximately 30 organisms. The primer sets were derived from rRNA alignment. Examples of regions from rRNA consensus alignments are shown in Figures 1A-1C. Lines with arrows are examples of regions to which intelligent primer pairs for PCR are designed. The primer pairs are >95% conserved in the bacterial sequence database (currently over 10,000 organisms). The intervening regions are variable in length and/or composition, thus providing the base composition "signature" (BCS) for each organism. Primer pairs were chosen so the total length of the amplified region is less than about 80-90 nucleotides. The label for each primer pair represents the starting and ending base number of the amplified region on the consensus diagram.

Included in the short bacterial database cross-section in Table 2 are many well known pathogens/biowarfare agents (shown in bold/red typeface) such as *Bacillus anthracis* or *Yersinia pestis* as well as some of the bacterial organisms found commonly in the natural environment such as *Streptomyces*. Even closely related organisms can be distinguished from each other by the appropriate choice of primers. For instance, two low G+C organisms, *Bacillus anthracis* and *Staph aureus*, can be distinguished from each other by using the primer pair defined by 16S\_1337 or 23S\_855 (ΔM of 4 Da).

**Table 2: Cross Section Of A Database Of Calculated Molecular Masses<sup>1</sup>**

Primer Regions ---->	16S_971	16S_1100	16S_1337	16S_1294	16S_1228	23S_1021	23S_855	23S_193	23S_115
Bug Name									
Acinetobacter calcoaceticus	55619.1	55004	28446.7	35854.9	51295.4	30299	42654	39557.5	54999
<b>Bacillus anthracis</b>	<b>55005</b>	<b>54388</b>	<b>28448</b>	<b>35238</b>	<b>51296</b>	<b>30295</b>	<b>42651</b>	<b>39560</b>	<b>56850</b>
Bacillus cereus	55622.1	54387.9	28447.6	35854.9	51296.4	30295	42651	39560.5	56850.3
Bordetella bronchiseptica	56857.3	51300.4	28446.7	35857.9	51307.4	30299	42653	39559.5	51920.5
Borrelia burgdorferi	56231.2	55621.1	28440.7	35852.9	51295.4	30297	42029.9	38941.4	52524.6
<b>Brucella abortus</b>	<b>58098</b>	<b>55011</b>	<b>28448</b>	<b>35854</b>	<b>50683</b>				
Campylobacter jejuni	58088.5	54386.9	29061.8	35856.9	50674.3	30294	42032.9	39558.5	45732.5
<b>Chlamydia pneumoniae</b>	<b>55000</b>	<b>55007</b>	<b>29063</b>	<b>35855</b>	<b>50676</b>	<b>30295</b>	<b>42036</b>	<b>38941</b>	<b>56230</b>
<b>Clostridium botulinum</b>	<b>55006</b>	<b>53767</b>	<b>28445</b>	<b>35855</b>	<b>51291</b>	<b>30300</b>	<b>42656</b>	<b>39562</b>	<b>54999</b>
Clostridium difficile	56855.3	54386.9	28444.7	35853.9	51296.4	30294	41417.8	39556.5	55612.2
Enterococcus faecalis	55620.1	54387.9	28447.6	35858.9	51296.4	30297	42652	39559.5	56849.3
<b>Escherichia coli</b>	<b>55622</b>	<b>55009</b>	<b>28445</b>	<b>35857</b>	<b>51301</b>	<b>30301</b>	<b>42656</b>	<b>39562</b>	<b>54999</b>
<b>Francisella tularensis</b>	<b>53769</b>	<b>54385</b>	<b>28445</b>	<b>35856</b>	<b>51298</b>				
Haemophilus influenzae	55620.1	55006	28444.7	35855.9	51298.4	30298	42656	39560.5	55613.1
Klebsiella pneumoniae	55622.1	55008	28442.7	35856.9	51297.4	30300	42655	39562.5	55000
<b>Legionella pneumophila</b>	<b>55618</b>	<b>55626</b>	<b>28446</b>	<b>35857</b>	<b>51303</b>				
Mycobacterium avium	54390.9	55631.1	29064.8	35858.9	51915.5	30298	42656	38942.4	56241.2
Mycobacterium leprae	54389.9	55629.1	29064.8	35860.9	51917.5	30298	42656	39559.5	56240.2
Mycobacterium tuberculosis	54390.9	55629.1	29064.8	35860.9	51301.4	30299	42656	39560.5	56243.2
Mycoplasma genitalium	53143.7	45115.4	29061.8	35854.9	50671.3	30294	43264.1	39558.5	56842.4
Mycoplasma pneumoniae	53143.7	45118.4	29061.8	35854.9	50673.3	30294	43264.1	39559.5	56843.4
Neisseria gonorrhoeae	55627.1	54389.9	28445.7	35855.9	51302.4	30300	42649	39561.5	55000
<b>Pseudomonas aeruginosa</b>	<b>55623</b>	<b>55010</b>	<b>28443</b>	<b>35858</b>	<b>51301</b>	<b>30298</b>	<b>43272</b>	<b>39558</b>	<b>55619</b>
<b>Rickettsia prowazekii</b>	<b>58093</b>	<b>55621</b>	<b>28448</b>	<b>35853</b>	<b>50677</b>	<b>30293</b>	<b>42650</b>	<b>39559</b>	<b>53139</b>
<b>Rickettsia rickettsii</b>	<b>58094</b>	<b>55623</b>	<b>28448</b>	<b>35853</b>	<b>50679</b>	<b>30293</b>	<b>42648</b>	<b>39559</b>	<b>53755</b>
<b>Salmonella typhimurium</b>	<b>55622</b>	<b>55005</b>	<b>28445</b>	<b>35857</b>	<b>51301</b>	<b>30301</b>	<b>42658</b>		
<b>Shigella dysenteriae</b>	<b>55623</b>	<b>55009</b>	<b>28444</b>	<b>35857</b>	<b>51301</b>				
Staphylococcus aureus	56854.3	54386.9	28443.7	35852.9	51294.4	30298	42655	39559.5	57466.4
Streptomyces	54389.9	59341.6	29063.8	35858.9	51300.4			39563.5	56864.3
Treponema pallidum	56245.2	55631.1	28445.7	35851.9	51297.4	30299	42034.9	38939.4	57473.4
<b>Vibrio cholerae</b>	<b>55625</b>	<b>55626</b>	<b>28443</b>	<b>35857</b>	<b>52536</b>	<b>29063</b>	<b>30303</b>	<b>35241</b>	<b>50675</b>
Vibrio parahaemolyticus	54384.9	55626.1	28444.7	34620.7	50064.2				
<b>Yersinia pestis</b>	<b>55620</b>	<b>55626</b>	<b>28443</b>	<b>35857</b>	<b>51299</b>				

<sup>1</sup>Molecular mass distribution of PCR amplified regions for a selection of organisms (rows) across various primer pairs (columns). Pathogens are shown in **bold**. Empty cells indicate 5 presently incomplete or missing data.

Figure 6 shows the use of ESI-FT-ICR MS for measurement of exact mass. The spectra from 46mer PCR products originating at position 1337 of the 16S rRNA from *S. aureus* (upper) and *B. anthracis* (lower) are shown. These data are from the region of the spectrum containing signals from the  $[M-8H]^8-$  charge states of the respective 5'-3' strands. The two strands differ 10 by two (AT→CG) substitutions, and have measured masses of 14206.396 and 14208.373 + 0.010 Da, respectively. The possible base compositions derived from the masses of the forward and reverse strands for the *B. anthracis* products are listed in Table 3.

**Table 3: Possible base composition for *B. anthracis* products**

Calc. Mass	Error	Base Comp.
14208.2935	0.079520	A1 G17 C10 T18
14208.3160	0.056980	A1 G20 C15 T10
14208.3386	0.034440	A1 G23 C20 T2
14208.3074	0.065560	A6 G11 C3 T26

14208.3300	0.043020	A6 G14 C8 T18
14208.3525	0.020480	A6 G17 C13 T10
14208.3751	0.002060	A6 G20 C18 T2
14208.3439	0.029060	A11 G8 C1 T26
14208.3665	0.006520	A11 G11 C6 T18
<b>14208.3890</b>	<b>0.016020</b>	<b>A11 G14 C11 T10</b>
14208.4116	0.038560	A11 G17 C16 T2
14208.4030	0.029980	A16 G8 C4 T18
14208.4255	0.052520	A16 G11 C9 T10
14208.4481	0.075060	A16 G14 C14 T2
14208.4395	0.066480	A21 G5 C2 T18
14208.4620	0.089020	A21 G8 C7 T10
14079.2624	0.080600	A0 G14 C13 T19
14079.2849	0.058060	A0 G17 C18 T11
14079.3075	0.035520	A0 G20 C23 T3
14079.2538	0.089180	A5 G5 C1 T35
14079.2764	0.066640	A5 G8 C6 T27
14079.2989	0.044100	A5 G11 C11 T19
14079.3214	0.021560	A5 G14 C16 T11
14079.3440	0.000980	A5 G17 C21 T3
14079.3129	0.030140	A10 G5 C4 T27
14079.3354	0.007600	A10 G8 C9 T19
<b>14079.3579</b>	<b>0.014940</b>	<b>A10 G11 C14 T11</b>
14079.3805	0.037480	A10 G14 C19 T3
14079.3494	0.006360	A15 G2 C2 T27
14079.3719	0.028900	A15 G5 C7 T19
14079.3944	0.051440	A15 G8 C12 T11
14079.4170	0.073980	A15 G11 C17 T3
14079.4084	0.065400	A20 G2 C5 T19
14079.4309	0.087940	A20 G5 C10 T13



Among the 16 compositions for the forward strand and the 18 compositions for the reverse strand that were calculated, only one pair (shown in **bold**) are complementary, corresponding to the actual base compositions of the *B. anthracis* PCR products.

#### 5 **Example 4: BCS of Region from *Bacillus anthracis* and *Bacillus cereus***

A conserved *Bacillus* region from *B. anthracis* (A<sub>14</sub>G<sub>9</sub>C<sub>14</sub>T<sub>9</sub>) and *B. cereus* (A<sub>15</sub>G<sub>9</sub>C<sub>13</sub>T<sub>9</sub>) having a C to A base change was synthesized and subjected to ESI-TOF MS. The results are shown in Figure 7 in which the two regions are clearly distinguished using the method of the present invention (MW=14072.26 vs. 14096.29).

10

#### **Example 5: Identification of additional bioagents**

In other examples of the present invention, the pathogen *Vibrio cholera* can be distinguished from *Vibrio parahemolyticus* with  $\Delta M > 600$  Da using one of three 16S primer sets shown in Table 2 (16S\_971, 16S\_1228 or 16S\_1294) as shown in Table 4. The two mycoplasma species in the list (*M. genitalium* and *M. pneumoniae*) can also be distinguished from each other, as can the three mycobacteriae. While the direct mass measurements of amplified products can identify and distinguish a large number of organisms, measurement of the base composition signature provides dramatically enhanced resolving power for closely related organisms. In cases such as *Bacillus anthracis* and *Bacillus cereus* that are virtually indistinguishable from each other based solely on mass differences, compositional analysis or fragmentation patterns are used to resolve the differences. The single base difference between the two organisms yields different fragmentation patterns, and despite the presence of the ambiguous/unidentified base N at position 20 in *B. anthracis*, the two organisms can be identified.

Tables 4a-b show examples of primer pairs from Table 1 which distinguish pathogens from background.

**Table 4a**

<b>Organism name</b>	<b>23S_855</b>	<b>16S_1337</b>	<b>23S_1021</b>
<i>Bacillus anthracis</i>	42650.98	28447.65	30294.98
<i>Staphylococcus aureus</i>	42654.97	28443.67	30297.96

25

-30-

Table 4b

Organism name	16S_971	16S_1294	16S_1228
<i>Vibrio cholerae</i>	55625.09	35856.87	52535.59
<i>Vibrio parahaemolyticus</i>	54384.91	34620.67	50064.19

Table 5 shows the expected molecular weight and base composition of region 16S\_1100-1188 in *Mycobacterium avium* and *Streptomyces sp.*

5

Table 5

Region	Organism name	Length	Molecular weight	Base comp.
16S_1100-1188	<i>Mycobacterium avium</i>	82	25624.1728	A <sub>16</sub> G <sub>32</sub> C <sub>18</sub> T <sub>16</sub>
16S_1100-1188	<i>Streptomyces sp.</i>	96	29904.871	A <sub>17</sub> G <sub>38</sub> C <sub>27</sub> T <sub>14</sub>

Table 6 shows base composition (single strand) results for 16S\_1100-1188 primer amplification reactions different species of bacteria. Species which are repeated in the table (e.g., *Clostridium botulinum*) are different strains which have different base compositions in the 10 16S\_1100-1188 region.

Table 6

Organism name	Base comp.	Organism name	Base comp.
<i>Mycobacterium avium</i>	A <sub>16</sub> G <sub>32</sub> C <sub>18</sub> T <sub>16</sub>	<i>Vibrio cholerae</i>	A <sub>23</sub> G <sub>30</sub> C <sub>21</sub> T <sub>16</sub>
<i>Streptomyces sp.</i>	A <sub>17</sub> G <sub>38</sub> C <sub>27</sub> T <sub>14</sub>	<i>Aeromonas hydrophila</i>	A <sub>23</sub> G <sub>31</sub> C <sub>21</sub> T <sub>15</sub>
<i>Ureaplasma urealyticum</i>	A <sub>18</sub> G <sub>30</sub> C <sub>17</sub> T <sub>17</sub>	<i>Aeromonas salmonicida</i>	A <sub>23</sub> G <sub>31</sub> C <sub>21</sub> T <sub>15</sub>
<i>Streptomyces sp.</i>	A <sub>19</sub> G <sub>36</sub> C <sub>24</sub> T <sub>18</sub>	<i>Mycoplasma genitalium</i>	A <sub>24</sub> G <sub>19</sub> C <sub>12</sub> T <sub>18</sub>
<i>Mycobacterium leprae</i>	A <sub>20</sub> G <sub>32</sub> C <sub>22</sub> T <sub>16</sub>	<i>Clostridium botulinum</i>	A <sub>24</sub> G <sub>25</sub> C <sub>18</sub> T <sub>20</sub>
<i>M. tuberculosis</i>	A <sub>20</sub> G <sub>33</sub> C <sub>21</sub> T <sub>16</sub>	<i>Bordetella bronchiseptica</i>	A <sub>24</sub> G <sub>26</sub> C <sub>19</sub> T <sub>14</sub>
<i>Nocardia asteroides</i>	A <sub>20</sub> G <sub>33</sub> C <sub>21</sub> T <sub>16</sub>	<i>Francisella tularensis</i>	A <sub>24</sub> G <sub>26</sub> C <sub>19</sub> T <sub>19</sub>
<i>Fusobacterium necroforum</i>	A <sub>21</sub> G <sub>26</sub> C <sub>22</sub> T <sub>18</sub>	<i>Bacillus anthracis</i>	A <sub>24</sub> G <sub>26</sub> C <sub>20</sub> T <sub>18</sub>
<i>Listeria monocytogenes</i>	A <sub>21</sub> G <sub>27</sub> C <sub>19</sub> T <sub>19</sub>	<i>Campylobacter jejuni</i>	A <sub>24</sub> G <sub>26</sub> C <sub>20</sub> T <sub>18</sub>
<i>Clostridium botulinum</i>	A <sub>21</sub> G <sub>27</sub> C <sub>19</sub> T <sub>21</sub>	<i>Staphylococcus aureus</i>	A <sub>24</sub> G <sub>26</sub> C <sub>20</sub> T <sub>18</sub>
<i>Neisseria gonorrhoeae</i>	A <sub>21</sub> G <sub>28</sub> C <sub>21</sub> T <sub>18</sub>	<i>Helicobacter pylori</i>	A <sub>24</sub> G <sub>26</sub> C <sub>20</sub> T <sub>19</sub>
<i>Bartonella quintana</i>	A <sub>21</sub> G <sub>30</sub> C <sub>22</sub> T <sub>16</sub>	<i>Helicobacter pylori</i>	A <sub>24</sub> G <sub>26</sub> C <sub>21</sub> T <sub>18</sub>
<i>Enterococcus faecalis</i>	A <sub>22</sub> G <sub>27</sub> C <sub>20</sub> T <sub>19</sub>	<i>Moraxella catarrhalis</i>	A <sub>24</sub> G <sub>26</sub> C <sub>23</sub> T <sub>16</sub>
<i>Bacillus megaterium</i>	A <sub>22</sub> G <sub>28</sub> C <sub>20</sub> T <sub>18</sub>	<i>Haemophilus influenzae Rd</i>	A <sub>24</sub> G <sub>28</sub> C <sub>20</sub> T <sub>17</sub>

<i>Bacillus subtilis</i>	A <sub>22</sub> G <sub>28</sub> C <sub>21</sub> T <sub>17</sub>	<i>Chlamydia trachomatis</i>	A <sub>24</sub> G <sub>28</sub> C <sub>21</sub> T <sub>16</sub>
<i>Pseudomonas aeruginosa</i>	A <sub>22</sub> G <sub>29</sub> C <sub>23</sub> T <sub>15</sub>	<i>Chlamydophila pneumoniae</i>	A <sub>24</sub> G <sub>28</sub> C <sub>21</sub> T <sub>16</sub>
<i>Legionella pneumophila</i>	A <sub>22</sub> G <sub>32</sub> C <sub>20</sub> T <sub>16</sub>	<i>C. pneumonia AR39</i>	A <sub>24</sub> G <sub>28</sub> C <sub>21</sub> T <sub>16</sub>
<i>Mycoplasma pneumoniae</i>	A <sub>23</sub> G <sub>20</sub> C <sub>14</sub> T <sub>16</sub>	<i>Pseudomonas putida</i>	A <sub>24</sub> G <sub>29</sub> C <sub>21</sub> T <sub>16</sub>
<i>Clostridium botulinum</i>	A <sub>23</sub> G <sub>26</sub> C <sub>20</sub> T <sub>19</sub>	<i>Proteus vulgaris</i>	A <sub>24</sub> G <sub>30</sub> C <sub>21</sub> T <sub>15</sub>
<i>Enterococcus faecium</i>	A <sub>23</sub> G <sub>26</sub> C <sub>21</sub> T <sub>18</sub>	<i>Yersinia pestis</i>	A <sub>24</sub> G <sub>30</sub> C <sub>21</sub> T <sub>15</sub>
<i>Acinetobacter calcoaceti</i>	A <sub>23</sub> G <sub>26</sub> C <sub>21</sub> T <sub>19</sub>	<i>Yersinia pseudotuberculosis</i>	A <sub>24</sub> G <sub>30</sub> C <sub>21</sub> T <sub>15</sub>
<i>Leptospira borgpeterseni</i>	A <sub>23</sub> G <sub>26</sub> C <sub>24</sub> T <sub>15</sub>	<i>Clostridium botulinum</i>	A <sub>25</sub> G <sub>24</sub> C <sub>18</sub> T <sub>21</sub>
<i>Leptospira interrogans</i>	A <sub>23</sub> G <sub>26</sub> C <sub>24</sub> T <sub>15</sub>	<i>Clostridium tetani</i>	A <sub>25</sub> G <sub>25</sub> C <sub>18</sub> T <sub>20</sub>
<i>Clostridium perfringens</i>	A <sub>23</sub> G <sub>27</sub> C <sub>19</sub> T <sub>19</sub>	<i>Francisella tularensis</i>	A <sub>25</sub> G <sub>25</sub> C <sub>19</sub> T <sub>19</sub>
<b><i>Bacillus anthracis</i></b>	<b>A<sub>23</sub>G<sub>27</sub>C<sub>20</sub>T<sub>18</sub></b>	<i>Acinetobacter calcoacetic</i>	A <sub>25</sub> G <sub>26</sub> C <sub>20</sub> T <sub>19</sub>
<b><i>Bacillus cereus</i></b>	<b>A<sub>23</sub>G<sub>27</sub>C<sub>20</sub>T<sub>18</sub></b>	<i>Bacteriodes fragilis</i>	A <sub>25</sub> G <sub>27</sub> C <sub>16</sub> T <sub>22</sub>
<b><i>Bacillus thuringiensis</i></b>	<b>A<sub>23</sub>G<sub>27</sub>C<sub>20</sub>T<sub>18</sub></b>	<i>Chlamydophila psittaci</i>	A <sub>25</sub> G <sub>27</sub> C <sub>21</sub> T <sub>16</sub>
<i>Aeromonas hydrophila</i>	A <sub>23</sub> G <sub>29</sub> C <sub>21</sub> T <sub>16</sub>	<i>Borrelia burgdorferi</i>	A <sub>25</sub> G <sub>29</sub> C <sub>17</sub> T <sub>19</sub>
<i>Escherichia coli</i>	A <sub>23</sub> G <sub>29</sub> C <sub>21</sub> T <sub>16</sub>	<i>Streptobacillus monilifor</i>	A <sub>26</sub> G <sub>26</sub> C <sub>20</sub> T <sub>16</sub>
<i>Pseudomonas putida</i>	A <sub>23</sub> G <sub>29</sub> C <sub>21</sub> T <sub>17</sub>	<i>Rickettsia prowazekii</i>	A <sub>26</sub> G <sub>28</sub> C <sub>18</sub> T <sub>18</sub>
<b><i>Escherichia coli</i></b>	<b>A<sub>23</sub>G<sub>29</sub>C<sub>22</sub>T<sub>15</sub></b>	<i>Rickettsia rickettsii</i>	A <sub>26</sub> G <sub>28</sub> C <sub>20</sub> T <sub>16</sub>
<b><i>Shigella dysenteriae</i></b>	<b>A<sub>23</sub>G<sub>29</sub>C<sub>22</sub>T<sub>15</sub></b>	<i>Mycoplasma mycoides</i>	A <sub>28</sub> G <sub>23</sub> C <sub>16</sub> T <sub>20</sub>

The same organism having different base compositions are different strains. Groups of organisms which are highlighted or in italics have the same base compositions in the amplified region. Some of these organisms can be distinguished using multiple primers. For example, 5 *Bacillus anthracis* can be distinguished from *Bacillus cereus* and *Bacillus thuringiensis* using the primer 16S\_971-1062 (Table 7). Other primer pairs which produce unique base composition signatures are shown in Table 6 (bold). Clusters containing very similar threat and ubiquitous non-threat organisms (e.g. *anthracis* cluster) are distinguished at high resolution with focused sets of primer pairs. The known biowarfare agents in Table 6 are *Bacillus anthracis*, *Yersinia* 10 *pestis*, *Francisella tularensis* and *Rickettsia prowazekii*.

Table 7

Organism	16S_971-1062	16S_1228-1310	16S_1100-1188
<i>Aeromonas hydrophila</i>	A <sub>21</sub> G <sub>29</sub> C <sub>22</sub> T <sub>20</sub>	A <sub>22</sub> G <sub>27</sub> C <sub>21</sub> T <sub>13</sub>	A <sub>23</sub> G <sub>31</sub> C <sub>21</sub> T <sub>15</sub>
<i>Aeromonas salmonicida</i>	A <sub>21</sub> G <sub>29</sub> C <sub>22</sub> T <sub>20</sub>	A <sub>22</sub> G <sub>27</sub> C <sub>21</sub> T <sub>13</sub>	A <sub>23</sub> G <sub>31</sub> C <sub>21</sub> T <sub>15</sub>
<i>Bacillus anthracis</i>	<b>A<sub>21</sub>G<sub>27</sub>C<sub>22</sub>T<sub>22</sub></b>	A <sub>24</sub> G <sub>22</sub> C <sub>19</sub> T <sub>18</sub>	A <sub>23</sub> G <sub>27</sub> C <sub>20</sub> T <sub>18</sub>
<i>Bacillus cereus</i>	A <sub>22</sub> G <sub>27</sub> C <sub>21</sub> T <sub>22</sub>	A <sub>24</sub> G <sub>22</sub> C <sub>19</sub> T <sub>18</sub>	A <sub>23</sub> G <sub>27</sub> C <sub>20</sub> T <sub>18</sub>
<i>Bacillus thuringiensis</i>	A <sub>22</sub> G <sub>27</sub> C <sub>21</sub> T <sub>22</sub>	A <sub>24</sub> G <sub>22</sub> C <sub>19</sub> T <sub>18</sub>	A <sub>23</sub> G <sub>27</sub> C <sub>20</sub> T <sub>18</sub>
<i>Chlamydia trachomatis</i>	<b>A<sub>22</sub>G<sub>26</sub>C<sub>20</sub>T<sub>23</sub></b>	<b>A<sub>24</sub>G<sub>23</sub>C<sub>19</sub>T<sub>16</sub></b>	A <sub>24</sub> G <sub>28</sub> C <sub>21</sub> T <sub>16</sub>
<i>Chlamydia pneumoniae AR39</i>	A <sub>26</sub> G <sub>23</sub> C <sub>20</sub> T <sub>22</sub>	A <sub>26</sub> G <sub>22</sub> C <sub>16</sub> T <sub>18</sub>	A <sub>24</sub> G <sub>28</sub> C <sub>21</sub> T <sub>16</sub>
<i>Leptospira borgpetersenii</i>	A <sub>22</sub> G <sub>26</sub> C <sub>20</sub> T <sub>21</sub>	A <sub>22</sub> G <sub>25</sub> C <sub>21</sub> T <sub>15</sub>	A <sub>23</sub> G <sub>26</sub> C <sub>24</sub> T <sub>15</sub>
<i>Leptospira interrogans</i>	A <sub>22</sub> G <sub>26</sub> C <sub>20</sub> T <sub>21</sub>	A <sub>22</sub> G <sub>25</sub> C <sub>21</sub> T <sub>15</sub>	A <sub>23</sub> G <sub>26</sub> C <sub>24</sub> T <sub>15</sub>
<i>Mycoplasma genitalium</i>	A <sub>28</sub> G <sub>23</sub> C <sub>15</sub> T <sub>22</sub>	<b>A<sub>30</sub>G<sub>18</sub>C<sub>15</sub>T<sub>19</sub></b>	<b>A<sub>24</sub>G<sub>19</sub>C<sub>12</sub>T<sub>18</sub></b>
<i>Mycoplasma pneumoniae</i>	A <sub>28</sub> G <sub>23</sub> C <sub>15</sub> T <sub>22</sub>	<b>A<sub>27</sub>G<sub>19</sub>C<sub>16</sub>T<sub>20</sub></b>	<b>A<sub>23</sub>G<sub>20</sub>C<sub>14</sub>T<sub>16</sub></b>
<i>Escherichia coli</i>	<b>A<sub>22</sub>G<sub>28</sub>C<sub>20</sub>T<sub>22</sub></b>	A <sub>24</sub> G <sub>25</sub> C <sub>21</sub> T <sub>13</sub>	A <sub>23</sub> G <sub>29</sub> C <sub>22</sub> T <sub>15</sub>
<i>Shigella dysenteriae</i>	<b>A<sub>22</sub>G<sub>28</sub>C<sub>21</sub>T<sub>21</sub></b>	A <sub>24</sub> G <sub>25</sub> C <sub>21</sub> T <sub>13</sub>	A <sub>23</sub> G <sub>29</sub> C <sub>22</sub> T <sub>15</sub>
<i>Proteus vulgaris</i>	<b>A<sub>23</sub>G<sub>26</sub>C<sub>22</sub>T<sub>21</sub></b>	<b>A<sub>26</sub>G<sub>24</sub>C<sub>19</sub>T<sub>14</sub></b>	A <sub>24</sub> G <sub>30</sub> C <sub>21</sub> T <sub>15</sub>
<i>Yersinia pestis</i>	A <sub>24</sub> G <sub>25</sub> C <sub>21</sub> T <sub>22</sub>	A <sub>25</sub> G <sub>24</sub> C <sub>20</sub> T <sub>14</sub>	A <sub>24</sub> G <sub>30</sub> C <sub>21</sub> T <sub>15</sub>
<i>Yersinia pseudotuberculosis</i>	A <sub>24</sub> G <sub>25</sub> C <sub>21</sub> T <sub>22</sub>	A <sub>25</sub> G <sub>24</sub> C <sub>20</sub> T <sub>14</sub>	A <sub>24</sub> G <sub>30</sub> C <sub>21</sub> T <sub>15</sub>
<i>Francisella tularensis</i>	<b>A<sub>20</sub>G<sub>25</sub>C<sub>21</sub>T<sub>23</sub></b>	<b>A<sub>23</sub>G<sub>26</sub>C<sub>17</sub>T<sub>17</sub></b>	<b>A<sub>24</sub>G<sub>26</sub>C<sub>19</sub>T<sub>19</sub></b>
<i>Rickettsia prowazekii</i>	<b>A<sub>21</sub>G<sub>26</sub>C<sub>24</sub>T<sub>25</sub></b>	<b>A<sub>24</sub>G<sub>23</sub>C<sub>16</sub>T<sub>19</sub></b>	<b>A<sub>26</sub>G<sub>28</sub>C<sub>18</sub>T<sub>18</sub></b>
<i>Rickettsia rickettsii</i>	<b>A<sub>21</sub>G<sub>26</sub>C<sub>25</sub>T<sub>24</sub></b>	<b>A<sub>24</sub>G<sub>24</sub>C<sub>17</sub>T<sub>17</sub></b>	<b>A<sub>26</sub>G<sub>28</sub>C<sub>20</sub>T<sub>16</sub></b>

The sequence of *B. anthracis* and *B. cereus* in region 16S\_971 is shown below. Shown in bold is the single base difference between the two species which can be detected using the 5 methods of the present invention. *B. anthracis* has an ambiguous base at position 20.

*B.anthraxis*\_16S\_971

GCGAAGAACCUUACCAGGUMUUGACAUCCUCUGACAACCCUAGAGAUAGGGCUUC  
UCCUUCGGGAGCAGAGUGACAGGUGGUGCAUGGUU (SEQ ID NO:1)

10 *B.cereus*\_16S\_971

GCGAAGAACCUUACCAGGUCUUGACAUCCUCUGAAAACCCUAGAGAUAGGGCUUC  
UCCUUCGGGAGCAGAGUGACAGGUGGUGCAUGGUU (SEQ ID NO:2)

**Example 6: ESI-TOF MS of sspE 56-mer Plus Calibrant**

The mass measurement accuracy that can be obtained using an internal mass standard in the ESI-MS study of PCR products is shown in Fig.8. The mass standard was a 20-mer phosphorothioate oligonucleotide added to a solution containing a 56-mer PCR product from the 5 *B. anthracis* spore coat protein sspE. The mass of the expected PCR product distinguishes *B. anthracis* from other species of Bacillus such as *B. thuringiensis* and *B. cereus*.

**Example 7: *B. anthracis* ESI-TOF Synthetic 16S\_1228 Duplex**

An ESI-TOF MS spectrum was obtained from an aqueous solution containing 5  $\mu\text{M}$  10 each of synthetic analogs of the expected forward and reverse PCR products from the nucleotide 1228 region of the *B. anthracis* 16S rRNA gene. The results (Fig. 9) show that the molecular weights of the forward and reverse strands can be accurately determined and easily distinguish the two strands. The  $[\text{M}-21\text{H}^+]^{21-}$  and  $[\text{M}-20\text{H}^+]^{20-}$  charge states are shown.

**15 Example 8: ESI-FTICR-MS of Synthetic *B. anthracis* 16S\_1337 46 Base Pair Duplex**

An ESI-FTICR-MS spectrum was obtained from an aqueous solution containing 5  $\mu\text{M}$  each of synthetic analogs of the expected forward and reverse PCR products from the nucleotide 1337 region of the *B. anthracis* 16S rRNA gene. The results (Fig. 10) show that the molecular weights of the strands can be distinguished by this method. The  $[\text{M}-16\text{H}^+]^{16-}$  through  $[\text{M}-20$  10 $\text{H}^+]^{10-}$  charge states are shown. The insert highlights the resolution that can be realized on the FTICR-MS instrument, which allows the charge state of the ion to be determined from the mass difference between peaks differing by a single  $^{13}\text{C}$  substitution.

**25 Example 9: ESI-TOF MS of 56-mer Oligonucleotide from saspB Gene of *B. anthracis* with Internal Mass Standard**

ESI-TOF MS spectra were obtained on a synthetic 56-mer oligonucleotide (5  $\mu\text{M}$ ) from the saspB gene of *B. anthracis* containing an internal mass standard at an ESI of 1.7  $\mu\text{L}/\text{min}$  as a function of sample consumption. The results (Fig. 11) show that the signal to noise is improved as more scans are summed, and that the standard and the product are visible after only 100 scans. 30

**Example 10: ESI-TOF MS of an Internal Standard with Tributylammonium (TBA)-trifluoroacetate (TFA) Buffer**

An ESI-TOF-MS spectrum of a 20-mer phosphorothioate mass standard was obtained following addition of 5 mM TBA-TFA buffer to the solution. This buffer strips charge from the oligonucleotide and shifts the most abundant charge state from  $[M-8H^+]^{8-}$  to  $[M-3H^+]^{3-}$  (Fig. 12).

#### 5 **Example 11: Master Database Comparison**

The molecular masses obtained through Examples 1-10 are compared to molecular masses of known bioagents stored in a master database to obtain a high probability matching molecular mass.

#### 10 **Example 12: Master Data Base Interrogation over the Internet**

The same procedure as in Example 11 is followed except that the local computer did not store the Master database. The Master database is interrogated over an internet connection, searching for a molecular mass match.

#### 15 **Example 13: Master Database Updating**

The same procedure as in example 11 is followed except the local computer is connected to the internet and has the ability to store a master database locally. The local computer system periodically, or at the user's discretion, interrogates the Master database, synchronizing the local master database with the global Master database. This provides the  
20 current molecular mass information to both the local database as well as to the global Master database. This further provides more of a globalized knowledge base.

#### **Example 14: Global Database Updating**

The same procedure as in example 13 is followed except there are numerous such local  
25 stations throughout the world. The synchronization of each database adds to the diversity of information and diversity of the molecular masses of known bioagents.

#### **Example 15: Biochemical Processing of Large Amplification Products for Analysis by Mass Spectrometry**

30 In the example illustrated in Figure 18, a primer pair which amplifies a 986 bp region of the 16S ribosomal gene in *E. coli* (K12) was digested with a mixture of 4 restriction enzymes: *BstNI*, *BsmFI*, *BfaI*, and *NcoI*. Figure 18(a) illustrates the complexity of the resulting ESI-FTICR mass spectrum which contains multiple charge states of multiple restriction fragments. Upon mass deconvolution to neutral mass, the spectrum is significantly simplified and discrete

oligonucleotide pairs are evident (Figure 18(b)). When base compositions are derived from the masses of the restriction fragments, perfect agreement is observed for the known sequence of nucleotides 1-856 (Figure 18(c)); the batch of *NcoI* enzyme used in this experiment was inactive and resulted in a missed cleavage site and a 197-mer fragment went undetected as it is outside  
5 the mass range of the mass spectrometer under the conditions employed. Interestingly however, both a forward and reverse strand were detected for each fragment measured (solid and dotted lines in, respectively) within 2 ppm of the predicted molecular weights resulting in unambiguous determination of the base composition of 788 nucleotides of the 985 nucleotides in the amplicon. The coverage map offers redundant coverage as both 5' to 3' and 3' to 5' fragments are detected  
10 for fragments covering the first 856 nucleotides of the amplicon.

This approach is in many ways analogous to those widely used in MS-based proteomics studies in which large intact proteins are digested with trypsin, or other proteolytic enzyme(s), and the identity of the protein is derived by comparing the measured masses of the tryptic peptides with theoretical digests. A unique feature of this approach is that the precise mass  
15 measurements of the complementary strands of each digest product allow one to derive a de novo base composition for each fragment, which can in turn be "stitched together" to derive a complete base composition for the larger amplicon. An important distinction between this approach and a gel-based restriction mapping strategy is that, in addition to determination of the length of each fragment, an unambiguous base composition of each restriction fragment is  
20 derived. Thus, a single base substitution within a fragment (which would not be resolved on a gel) is readily observed using this approach. Because this study was performed on a 7 Tesla ESI-FTICR mass spectrometer, better than 2 ppm mass measurement accuracy was obtained for all fragments. Interestingly, calculation of the mass measurement accuracy required to derive unambiguous base compositions from the complementary fragments indicates that the highest  
25 mass measurement accuracy actually required is only 15 ppm for the 139 bp fragment (nucleotides 525-663). Most of the fragments were in the 50-70 bp size-range which would require mass accuracy of only ~50 ppm for unambiguous base composition determination. This level of performance is achievable on other more compact, less expensive MS platforms such as the ESI-TOF suggesting that the methods developed here could be widely deployed in a variety  
30 of diagnostic and human forensic arenas.

This example illustrates an alternative approach to derive base compositions from larger PCR products. Because the amplicons of interest cover many strain variants, for some of which complete sequences are not known, each amplicon can be digested under several different enzymatic conditions to ensure that a diagnostically informative region of the amplicon is not

obscured by a "blind spot" which arises from a mutation in a restriction site. The extent of redundancy required to confidently map the base composition of amplicons from different markers, and determine which set of restriction enzymes should be employed and how they are most effectively used as mixtures can be determined. These parameters will be dictated by the extent to which the area of interest is conserved across the amplified region, the compatibility of the various restriction enzymes with respect to digestion protocol (buffer, temperature, time) and the degree of coverage required to discriminate one amplicon from another.

### Example 16: Analysis of 10 Human Blood Mitochondrial DNA Samples Provided by the FBI

Ten different samples of human DNA provided by the FBI were subjected to rapid mtDNA analysis by the method of the present invention. Intelligent primers (SEQ ID NOs: 8-17 in Table 8) were selected to amplify portions of HVR1 and HVR2. Additional intelligent primers were designed to mtDNA regions other than HVR1 and HVR2 (SEQ ID NOs: 18-43). The primers described below are generally 10-50 nucleotides in length, 15-35 nucleotides in length, or 18-30 nucleotides in length.

**Table 8: Intelligent Primer Pairs for Analysis of mtDNA**

Primer Pair Name	Forward Primer Sequence	Forward SEQ ID NO:	Reverse Primer Sequence	Reverse SEQ ID NO:
HMTHV2_AND RSN_76_353 TMOD	TCACGCGATAGCATTGCG	8	TGGTTTGGCAGAGATGTGTTA AGT	9
HMTHV2_AND RSN_29_429 TMOD	TCTCACGGGAGCTCTCCATGC	10	TCTGTTAAAAGTGCATACCGCC A	11
HMTHV1_AND RSN_16065_ 16410 TMOD	TGACTCACCCATCAACAACCGC	12	TGAGGATGGTGGTCAAGGGAC	13
HMTHV1_AND RSN_16065_ 16354 TMOD	TGACTCACCCATCAACAACCGC	14	TGGATTTGACTGTAATGTGCTA	15
HMTHV1_AND RSN_16064_ 16359	TGACTCACCCATCAACAACCGC	16	TGAAGGGATTTGACTGTAATGT GCTATG	17
HMT_ASN_16 036_522	GAAGCAGATTTGGGTACCACC	18	GTGTGTGTGCTGGGTAGGATG	19
HMT_ASN_81 62_8916	TACGGTCAATGCTCTGAAATCT GTGG	20	TGGTAAGAAGTGGGCTAGGGCA TT	21
HMT_ASN_12 438_13189	TTATGTAAAATCCATTGTCGCA TCCACC	22	TGGTGATAGCGCCTAAGCATAG TG	23
HMT_ASN_14 629_15353	TCCATTACTAAACCCACACTC AACAG	24	TTTCGTGCAAGAATAGGAGGTG GAG	25
HMT_ASN_94 35_10188	TAAGGCCTTCGATACGGGATAA TCCTA	26	TAGGGTCGAAGCCGCACTCG	27
HMT_ASN_10	TACTCCAATGCTAAACTAATC	28	TGTGAGGCGTATTATACCATAG	29



-37-

753_11500	GTCCCAAC		CCG	
HMT_ASN_15 369_16006	TCCTAGGAATCACCTCCCATTC CGA	30	TAGAATCTTAGCTTTGGGTGCT AATGGTG	31
HMT_ASN_13 461_14206	TGGCAGCCTAGCATTAGCAGGA ATA	32	TGGCTGAACATTGTTTGTGGT GT	33
HMT_ASN_34 52_4210	TCGCTGACGCCATAAACTCTT CAC	34	TAAGTAATGCTAGGGTGAGTGG TAGGAAG	35
HMT_ASN_77 34_8493	TAACTAATACTAACATCTCAGA CGCTCAGGA	36	TTTATGGGCTTTGGTGAGGGAG GTA	37
HMT_ASN_63 09_7058	TACTCCCACCCTGGAGCCTC	38	TGCTCCTATTGATAGGACATAG TGGAAGTG	39
HMT_ASN_76 44_8371	TTATCACCTTTCATGATCACGC CCT	40	TGGCATTTCACTGTAAAGAGGT GTTGG	41
HMT_ASN_26 26_3377	TGTATGAATGGCTCCACGAGGG T	42	TCGGTAAGCATTAGGAATGCCA TTGC	43

The process of the analysis is shown in Figure 19. After amplification by PCR (210), the PCR products were subjected to restriction digests (220) with *RsaI* for HVR1 and a combination of *HpaII*, *HpyCH4IV*, *PacI* and *EaeI* for HVR2 in order to obtain amplicon segments suitable for analysis by FTICR-MS (230). The data were processed to obtain mass data for each amplicon segment (240) which were then compared to the masses calculated for theoretical digests from the FBI mtDNA database by a scoring scheme (250). Digestion pattern matches were scored by the sum of (i) the percentage of expected complete digest fragments observed, (ii) the percentage of fragments with a "floating" percentage of potential incomplete digest fragments (to increase sensitivity for incomplete digestion – these are assigned lower weight), (iii) the percentage of the sequence covered by matched masses, (iv) the number of mass peaks accounted for in the theoretical database digest, and (v) the weighted score for matched peaks, weighted by their observed abundance. HVR1 and HVR2 scores were combined and all database entries were sorted by high score. Even in the absence of an exact match in the database, the majority of entries can be ruled out by observing a much lower match score than the maximum score. One with relevant skill in the art will recognize that development of such scoring procedures is can be accomplished without undue experimentation.

The results of analysis of sample 1 are shown in Figures 20A and 20B. In this example, the utility of mass determination of amplicon digest segments is indicated. In Figure 20A, predicted and actual mass data with scoring parameters for length heteroplasmy (HV1-1-outer-variants 1 and 2) in the digest segment from position 94 to 145(variant 1)/146(variant 2) are shown. Figure 20B indicates that, whereas sequencing fails to resolve the variants due to the length heteroplasmy, mass determination detects multiple species simultaneously and indicates abundance ratios. In this case, the ratio of variant 1 to variant 2 (short to long alleles) is 1:3.

-38-

Thus, in addition to efficiency of characterization of individual digested amplicon fragments, the relative abundances of heteroplasmic variants can be determined.

Of the 10 samples analyzed by the present methods, 9 samples were verified as being consistent with members of the FBI database. The remaining sample could not be analyzed due  
5 to a failure of PCR to produce an amplification product.

The claims should be given the broadest interpretation consistent with the description as a whole.

DOCSTOR-#1144799-v1-Sequence\_Listing.TXT  
SEQUENCE LISTING

<110> ISIS Pharmaceuticals, Inc.

<120> Methods For Rapid Forensic Analysis Of Mitochondrial DNA  
And Characterization Of Mitochondrial DNA Heteroplasmy

<130> 13330-4CA  
<140> 2,510,007  
<141> 2003-12-05

<150> 10/323,438  
<151> 2002-12-18

<150> 10/660,998  
<151> 2003-09-12

<150> 60/431,319  
<151> 2002-12-06

<160> 43

<170> PatentIn version 3.0

<210> 1  
<211> 90  
<212> RNA  
<213> Bacillus anthracis

<220>  
<221> misc\_feature  
<222> (20)..(20)  
<223> N = A, U, G or C

<400> 1  
gcgaagaacc uuaccaggun uugacauccu cugacaaccc uagagauagg gcuucuccuu 60  
cgggagcaga gugacaggug gugcaugguu 90

<210> 2  
<211> 90  
<212> RNA  
<213> Bacillus cereus

<400> 2  
gcgaagaacc uuaccagguc uugacauccu cugaaaaccc uagagauagg gcuucuccuu 60  
cgggagcaga gugacaggug gugcaugguu 90

<210> 3  
<211> 1542  
<212> RNA  
<213> Artificial Sequence

<220>  
<221> misc\_feature  
<223> 16S rRNA consensus sequence

<220>  
<221> misc\_feature  
<222> (1)..(7)

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (19)..(19)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (28)..(30)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (43)..(45)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (48)..(48)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (50)..(50)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (52)..(52)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (66)..(66)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (69)..(100)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (103)..(103)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (107)..(108)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (121)..(122)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (124)..(124)

<223> N= A, U, G or C

<220>

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<221> misc\_feature  
<222> (126)..(129)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (131)..(132)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (134)..(134)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (137)..(145)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (148)..(148)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (150)..(150)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (152)..(158)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (163)..(169)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (177)..(178)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (181)..(194)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (199)..(226)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (229)..(237)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (239)..(242)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (245)..(245)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (248)..(248)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (250)..(250)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (257)..(258)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (264)..(264)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (268)..(269)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (276)..(276)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (278)..(280)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (283)..(286)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (291)..(291)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (293)..(294)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (303)..(304)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<222> (306)..(307)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (309)..(309)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (316)..(316)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (320)..(320)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (328)..(328)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (333)..(333)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (337)..(337)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (359)..(360)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (366)..(366)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (369)..(371)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (378)..(379)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (381)..(381)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (384)..(385)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (390)..(392)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (396)..(396)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (398)..(399)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (407)..(409)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (412)..(412)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (415)..(415)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (418)..(419)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (421)..(423)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (425)..(425)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (427)..(427)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (433)..(435)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (438)..(438)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (440)..(446)



DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (449)..(449)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (452)..(479)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (484)..(485)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (488)..(494)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (496)..(497)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (501)..(503)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (508)..(508)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (513)..(513)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (538)..(538)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (542)..(543)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (546)..(546)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (553)..(555)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<221> misc\_feature  
<222> (560)..(560)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (562)..(562)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (564)..(564)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (576)..(576)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (578)..(580)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (582)..(582)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (586)..(586)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (589)..(596)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (599)..(603)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (606)..(606)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (610)..(616)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (620)..(620)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (624)..(633)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (635)..(641)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (644)..(650)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (653)..(653)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (657)..(662)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (665)..(665)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (668)..(673)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (679)..(682)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (689)..(689)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (694)..(694)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (698)..(698)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (701)..(701)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (705)..(705)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<222> (708)..(709)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (711)..(711)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (713)..(713)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (717)..(717)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (721)..(722)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (724)..(724)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (733)..(738)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (743)..(748)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (755)..(755)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (758)..(758)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (760)..(763)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (771)..(771)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (776)..(776)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (780)..(780)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (808)..(808)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (811)..(812)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (819)..(819)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (822)..(826)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (828)..(831)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (833)..(835)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (837)..(859)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (861)..(863)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (868)..(870)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (874)..(878)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (895)..(896)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (903)..(904)

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (906)..(906)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (916)..(916)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (929)..(929)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (932)..(932)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (941)..(941)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (943)..(943)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (948)..(948)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (955)..(955)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (965)..(965)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (967)..(968)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (974)..(974)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (976)..(976)

<223> N= A, U, G or C

<220>

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<221> misc\_feature  
<222> (986)..(990)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (998)..(1012)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1015)..(1015)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1017)..(1043)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1051)..(1051)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1059)..(1059)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1075)..(1076)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1082)..(1082)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1100)..(1100)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1115)..(1123)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1127)..(1127)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1129)..(1129)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1131)..(1131)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (1133)..(1141)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1143)..(1143)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1145)..(1145)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1150)..(1156)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1163)..(1165)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1167)..(1168)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1171)..(1173)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1183)..(1183)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1189)..(1189)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1198)..(1198)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1201)..(1201)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1207)..(1207)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature



## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<222> (1214)..(1214)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1216)..(1219)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1225)..(1225)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1231)..(1231)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1233)..(1233)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1243)..(1247)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1251)..(1252)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1254)..(1254)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1256)..(1257)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1260)..(1260)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1262)..(1265)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1267)..(1268)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1270)..(1274)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (1278)..(1278)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1281)..(1281)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1283)..(1286)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1290)..(1294)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1297)..(1298)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1302)..(1302)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1308)..(1308)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1310)..(1313)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1324)..(1327)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1329)..(1329)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1335)..(1336)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1340)..(1340)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1354)..(1356)

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1362)..(1362)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1364)..(1364)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1366)..(1368)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1383)..(1383)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1388)..(1388)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1409)..(1411)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1414)..(1414)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1416)..(1417)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1420)..(1428)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1431)..(1432)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1436)..(1447)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1449)..(1454)

<223> N= A, U, G or C

<220>

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<221> misc\_feature  
<222> (1456)..(1465)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1467)..(1467)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1469)..(1469)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1472)..(1481)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1484)..(1484)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1489)..(1491)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1508)..(1508)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1511)..(1511)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1514)..(1516)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1520)..(1521)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1524)..(1524)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1527)..(1527)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1542)..(1542)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<400> 3  
 nnnnnnnnaga guuugaucnu ggcucagnnn gaacgcuggc ggnnngcnun anacaugcaa 60  
 gucgancgnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn agnggcnnac gggugaguaa 120  
 nncnunnnaa nnuncnannn nnnnnggnan annnnnnnga aannnnnnnu aauaccnau 180  
 nnnnnnnnnn nnnnaaagnn nnnnnnnnnn nnnnnnnnnn nnnnnngann nnnnnngnn 240  
 nnaunagnun guuggunngg uaanggcnaa ccaagncnnn gannnnuagc ngncugaga 300  
 ggngnncng ccacanuggn acugaganac ggnccanacu ccuacgggag gcagcagunn 360  
 ggaauunnn ncaauggnng naanncugan nnagcnannc cgcgugnnng anganggnnu 420  
 nngnnguaa aannncunnn nnnnnngang annnnnnnnn nnnnnnnnnn nnnnnnnnnu 480  
 gacnnuannn nnnnannaag nncggcnaa cuncgugcca gcagccgagg uauuacgnag 540  
 gnngcnagcg uunnnccgan unanugggag uaaagngnnn gnaggnggnn nnnnnngunn 600  
 nngunaaan nnnnnngcun aacnnnnnnn nncnnnnnnn nacnnnnnnn cungagnnnn 660  
 nnagggnnn nngaaunnn nnguguagng gugnaauncg naganauung nangaanacc 720  
 nnungcgaag gcnnnnnnncu ggnnnnnnnac ugacncunan nncgaaagc nugggnagcn 780  
 aacaggauua gauaccuggg uaguaccangc nuuaaacgnu gnnnnnunnn ngnnngnnnn 840  
 nnnnnnnnnn nnnnnnnnna nnaaacgnnn uaannnncc gccuggggag uacgnncgca 900  
 agnnunaaac ucaaangaau ugacggggnc cngcacaagc ngnggagnau guggnuuaau 960  
 ucgangnnac gcgnanaacc uuaccnnnnn uugacaunnn nnnnnnnnnn nnganannnn 1020  
 nnnnnnnnnn nnnnnnnnnn nnnacaggug nugcauggnu gucgucagcu cgugnguga 1080  
 gnuguugggu uaagucccgn aacgagcgca accnnnnnnn nnguuncna ncnnnnnnnn 1140  
 ngngnacucn nnnnnnacug ccnnngnaa nnggaggaa ggnggggang acgucaanuc 1200  
 nucaugnccc uangnnngg ggcuncacac nuncuacaau ggnnnnnaca nngngngcn 1260  
 annnngnnan nnnnagcnaa ncnnnnaaan nnnnucnag uncggaungh nncugcaac 1320  
 ucgnnnncnu gaagnnggan ucgcuaguaa ucgnnnauca gnangnnncg gugaauacgu 1380  
 ucncgggncu uguacacacc gcccgucann ncangnnagn nnnnnnncc nnaagnnnnn 1440  
 nnnnnnncnn nnnngnnnnn nnnnncnang gnngnnnnnn nganugggnn naagucguaa 1500  
 caagguancc nuannngaan nugnggnugg aucaccuccu un 1542

<210> 4  
 <211> 2904  
 <212> RNA  
 <213> Artificial Sequence

<220>  
 <221> misc\_feature  
 <223> 23S rRNA consensus sequence

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (1)..(4)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (8)..(12)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (16)..(16)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (18)..(22)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (34)..(34)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (38)..(43)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (46)..(46)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (50)..(50)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (57)..(57)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (59)..(65)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (67)..(68)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (70)..(72)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<222> (74)..(75)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (77)..(79)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (82)..(83)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (86)..(87)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (89)..(96)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (98)..(102)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (104)..(104)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (107)..(109)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (111)..(111)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (113)..(113)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (125)..(125)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (131)..(148)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (150)..(177)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (179)..(181)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (184)..(188)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (192)..(192)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (203)..(203)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (208)..(212)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (218)..(218)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (224)..(225)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (228)..(231)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (236)..(236)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (238)..(241)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (246)..(246)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (257)..(259)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (261)..(261)



## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (263)..(264)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (267)..(267)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (269)..(293)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (295)..(297)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (301)..(305)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (309)..(309)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (313)..(321)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (323)..(325)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (329)..(329)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (331)..(331)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (333)..(334)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (337)..(337)

<223> N= A, U, G or C

<220>

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<221> misc\_feature  
<222> (341)..(344)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (348)..(370)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (375)..(377)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (379)..(382)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (384)..(384)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (387)..(387)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (389)..(390)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (392)..(395)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (398)..(399)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (403)..(405)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (407)..(410)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (416)..(421)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (425)..(425)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (435)..(441)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (446)..(446)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (451)..(451)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (453)..(453)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (455)..(456)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (462)..(462)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (467)..(467)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (475)..(475)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (482)..(482)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (487)..(491)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (493)..(493)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (504)..(504)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<222> (507)..(508)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (518)..(522)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (524)..(524)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (527)..(527)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (530)..(532)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (535)..(537)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (540)..(553)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (557)..(558)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (563)..(563)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (571)..(571)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (573)..(574)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (578)..(580)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (582)..(582)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (584)..(584)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (587)..(587)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (590)..(593)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (595)..(599)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (602)..(602)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (605)..(605)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (610)..(618)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (620)..(620)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (623)..(623)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (626)..(626)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (629)..(629)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (634)..(634)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (640)..(642)

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (645)..(646)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (648)..(648)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (652)..(654)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (658)..(662)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (664)..(667)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (672)..(672)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (677)..(677)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (679)..(681)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (686)..(686)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (690)..(692)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (696)..(697)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (702)..(702)

<223> N= A, U, G or C

<220>

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<221> misc\_feature  
<222> (708)..(712)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (717)..(717)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (719)..(723)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (730)..(730)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (737)..(744)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (753)..(758)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (765)..(766)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (771)..(772)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (774)..(774)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (776)..(776)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (779)..(779)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (784)..(785)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (787)..(787)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (790)..(790)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (792)..(792)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (796)..(798)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (800)..(801)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (815)..(816)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (822)..(825)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (832)..(835)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (838)..(838)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (840)..(854)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (857)..(857)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (870)..(879)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (882)..(894)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature



## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<222> (898)..(899)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (901)..(908)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (914)..(914)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (920)..(920)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (923)..(938)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (940)..(940)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (943)..(944)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (946)..(947)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (949)..(951)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (953)..(953)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (955)..(955)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (957)..(957)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (961)..(962)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (964)..(964)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (966)..(968)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (971)..(972)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (974)..(974)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (979)..(979)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (984)..(984)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (991)..(991)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (993)..(994)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (996)..(998)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1004)..(1004)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1008)..(1008)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1011)..(1018)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1026)..(1026)

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1030)..(1030)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1033)..(1033)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1037)..(1042)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1044)..(1045)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1047)..(1047)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1051)..(1053)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1058)..(1058)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1078)..(1078)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1080)..(1080)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1083)..(1083)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1089)..(1090)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (1097)..(1097)

<223> N= A, U, G or C

<220>

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<221> misc\_feature  
<222> (1106)..(1107)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1110)..(1110)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1113)..(1119)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1124)..(1124)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1127)..(1128)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1131)..(1131)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1134)..(1134)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1139)..(1139)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1144)..(1151)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1157)..(1162)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1164)..(1185)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1191)..(1192)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1199)..(1211)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (1216)..(1222)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1224)..(1225)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1227)..(1233)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1238)..(1246)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1251)..(1251)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1253)..(1253)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1257)..(1258)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1260)..(1261)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1264)..(1264)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1269)..(1269)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1273)..(1280)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1285)..(1285)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<222> (1287)..(1288)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1290)..(1294)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1296)..(1296)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1300)..(1300)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1302)..(1304)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1306)..(1306)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1311)..(1311)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1316)..(1321)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1323)..(1323)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1325)..(1325)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1327)..(1328)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1331)..(1336)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1341)..(1341)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (1347)..(1349)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1356)..(1357)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1361)..(1361)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1363)..(1363)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1366)..(1366)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1368)..(1368)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1370)..(1371)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1375)..(1376)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1382)..(1383)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1385)..(1387)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1391)..(1392)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1400)..(1402)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1405)..(1425)

DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1430)..(1435)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1437)..(1454)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1457)..(1564)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1566)..(1567)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1573)..(1599)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1606)..(1607)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1622)..(1622)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1624)..(1627)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1629)..(1630)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1634)..(1634)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1636)..(1637)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1639)..(1640)

&lt;223&gt; N= A, U, G or C

&lt;220&gt;



## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<221> misc\_feature  
<222> (1644)..(1644)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1646)..(1648)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1650)..(1653)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1656)..(1663)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1672)..(1673)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1679)..(1679)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1681)..(1684)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1690)..(1690)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1697)..(1697)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1699)..(1699)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1704)..(1707)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1709)..(1749)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1751)..(1754)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (1756)..(1758)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1760)..(1762)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1764)..(1770)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1772)..(1772)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1781)..(1782)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1793)..(1794)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1796)..(1797)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1801)..(1801)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1804)..(1805)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1808)..(1808)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1812)..(1813)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1816)..(1816)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<222> (1822)..(1822)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1825)..(1826)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1831)..(1831)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1839)..(1839)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1844)..(1845)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1855)..(1856)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1858)..(1866)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1868)..(1872)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1874)..(1884)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1886)..(1888)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1895)..(1896)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1899)..(1899)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1908)..(1909)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (1921)..(1922)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1963)..(1963)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1971)..(1971)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1974)..(1974)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1976)..(1976)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1979)..(1979)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1982)..(1989)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (1997)..(2005)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2007)..(2007)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2009)..(2009)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2011)..(2011)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2015)..(2015)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2018)..(2019)

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2021)..(2021)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2023)..(2026)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2029)..(2029)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2037)..(2040)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2042)..(2042)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2044)..(2044)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2048)..(2052)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2067)..(2068)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2070)..(2070)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2072)..(2072)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2080)..(2081)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2083)..(2085)

<223> N= A, U, G or C

<220>

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<221> misc\_feature  
<222> (2087)..(2089)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2091)..(2091)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2094)..(2108)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2112)..(2113)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2116)..(2116)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2123)..(2123)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2128)..(2128)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2130)..(2132)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2135)..(2142)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2145)..(2146)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2149)..(2155)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2160)..(2160)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2162)..(2166)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (2169)..(2170)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2175)..(2175)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2178)..(2178)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2181)..(2194)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2201)..(2211)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2213)..(2213)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2215)..(2223)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2228)..(2228)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2231)..(2233)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2235)..(2236)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2240)..(2240)  
<223> = A, U, G or C

<220>  
<221> misc\_feature  
<222> (2246)..(2246)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<222> (2258)..(2259)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2265)..(2265)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2269)..(2270)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2281)..(2281)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2283)..(2284)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2286)..(2286)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2292)..(2294)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2297)..(2297)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2299)..(2302)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2305)..(2306)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2309)..(2310)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2314)..(2321)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2325)..(2326)  
<223> N= A, U, G or C



## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (2329)..(2330)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2332)..(2332)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2334)..(2334)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2338)..(2340)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2343)..(2343)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2345)..(2345)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2350)..(2351)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2354)..(2357)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2360)..(2363)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2371)..(2373)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2380)..(2381)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2384)..(2386)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2398)..(2398)

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2402)..(2407)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2414)..(2414)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2418)..(2418)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2437)..(2437)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2441)..(2441)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2443)..(2443)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2458)..(2458)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2461)..(2464)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2474)..(2474)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2477)..(2477)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2486)..(2489)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2513)..(2513)

<223> N= A, U, G or C

<220>

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<221> misc\_feature  
<222> (2516)..(2516)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2530)..(2530)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2533)..(2534)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2547)..(2548)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2560)..(2561)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2568)..(2568)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2571)..(2571)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2575)..(2575)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2586)..(2586)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2588)..(2588)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2606)..(2606)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2617)..(2617)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2619)..(2620)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (2622)..(2622)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2624)..(2624)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2626)..(2626)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2628)..(2630)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2633)..(2635)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2640)..(2642)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2644)..(2646)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2649)..(2650)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2652)..(2652)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2670)..(2674)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2677)..(2678)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2680)..(2680)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<222> (2682)..(2682)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2689)..(2691)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2693)..(2693)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2699)..(2701)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2706)..(2708)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2712)..(2713)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2716)..(2716)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2718)..(2719)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2726)..(2727)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2729)..(2730)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2733)..(2736)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2742)..(2743)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2750)..(2750)  
<223> N= A, U, G or C

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
<221> misc\_feature  
<222> (2760)..(2762)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2766)..(2766)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2768)..(2770)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2772)..(2775)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2779)..(2780)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2783)..(2785)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2788)..(2788)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2790)..(2809)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2812)..(2814)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2816)..(2820)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2824)..(2825)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2827)..(2830)  
<223> N= A, U, G or C

<220>  
<221> misc\_feature  
<222> (2833)..(2833)

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2840)..(2842)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2844)..(2846)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2849)..(2849)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2853)..(2856)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2858)..(2859)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2861)..(2864)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2866)..(2867)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2870)..(2872)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2875)..(2877)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2885)..(2888)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2890)..(2895)

<223> N= A, U, G or C

<220>

<221> misc\_feature

<222> (2899)..(2904)

<223> N= A, U, G or C

<400> 4

DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

nnnnaagnnn nnaagngnnn nngguggaug ccunggcnnn nnnagncgan gaaggangnn	60
nnnnncnncn nnanncnng gnnagnngnn nnnnnncnnn nnanccnng nunuccgaau	120
gggnaacc nnnnnnnnnn nnnnnnnnan nnnnnnnnnn nnnnnnnnnn nnnnnngnn	180
nacnnnnga anugaaacau cunaguannn nnaggaanag aaannaannn ngauuncnnn	240
nguagngcg agcgaannng nannagncnn nnnnnnnnnn nnnnnnnnnn nnnannngaa	300
nnnnnuggna agnnnnnnnn nannngguna nanncngua nnnnaaannn nnnnnnnnnn	360
nnnnnnnnnn aguannncnn nncncgngnn annngunng aannngnnnn gaccannnnn	420
naagncuaaa uacunnnnnn ngaccnauag ngnannagua cngugangga aaggngaaa	480
gnaccnnnnn nanggagug aanagncncc ugaaaccnnn nncnuanaan nngunnnagn	540
nnnnnnnnnn nnnuganngc gunccuuug nannaugnnn cngnganuun nnnunnnng	600
cnagnuuaan nnnnnnnngn agncgnagng aaancgagun nnaanngngc gnnnagunnn	660
nngnnnnaga cncgaancnn ngugancuan nnaugnncag gnugaagnnn nnguaanann	720
nnnuggaggn ccgaacnnnn nnnnguugaa aannnnngg augannugug nnungngng	780
aaanncnaan cnaacnnngn nauagcuggu ucucnncgaa annnuuuag gnnngcnun	840
nnnnnnnnnn nnnngnggu agagcacugn nnnnnnnng gnnnnnnnnn nnnnuacna	900
nnnnnnnnaa acuncgaaun ccnnnnnnnn nnnnnnnngn agnnanncn nngngnuaa	960
nnuncnngu nnanagggn acaaccaga ncncnncnua agnccnaa nnnnnnnnua	1020
aguggnaaan gangugnann nncnnanaca nnnaggangu uggcuuagaa gcagccancn	1080
uunaaagann gcguaanagc ucacunnucn agnnnnnnng cgcngannau nuancggngc	1140
uaannnnnnn nccgaannnn nngnnnnnnn nnnnnnnnnn nnnnngguag nngagcgunn	1200
nnnnnnnnnn ngaagnnnnn nngnnannnn nnnuggannn nnnnnnagug ngnaugnngn	1260
naunaguanc gannnnnnnn gugananncn nnnncnccgn annncnaagg nuuccnnnnn	1320
nangnunnuc nnnnnnggu nagucgnnc cuaagnngag ncnganangn nuagnngaug	1380
gnnannnggu nnauauuccn nnacnnnnnn nnnnnnnnnn nnnnngacgn nnnnngnnnn	1440
nnnnnnnnnn nnnnggnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	1500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	1560
nnnncngaa aannnnnnnn nnnnnnnnnn nnnnnnnnnc guaccnnaaa ccgacacagg	1620
ungnnngnn gagnanncn agngnnngn nnaannnnn nnaaggaac unngcaaanu	1680
nnnccguan cuucggnana aggnnnncnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	1740
nnnnnnnnng nnnnannnan nngnnnnnnn cnacuguuua nnaaaaacac agnncnnugc	1800
naanncgnaa gnganguau anggnugac nccugcccng ugcngaagg uuaanngnn	1860
nnnnngnnn nngnnnnnnn nnnnannnaa gccnguna acggcggng uaacuuaac	1920



## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

```

nnuccuaagg uagcgaaauu ccuugucggg uaaguuccga ccngcacgaa nggngnaang 1980
annnnnnnnc ugucucnnnn nnnnncncng ngaanuunna nunnnnguna agaugcnnnn 2040
uncncgcnnn nngacggaaa gaccccnnngn ancuuuacun nannnunna nugnnnnnnn 2100
nnnnnnnnug unnagnauag gunggagncn nngannnnnn nncggnagnn nnnnnggagn 2160
cnnnnnugn auacnacncu nnnnnnnnnn nnnnucuaac nnnnnnnnnn nancnnnnn 2220
nnngacanug nnnghngggg aguuunacug gggcggunnc cuccnaaann guaacggagg 2280
ngnncnaagg unncunann nnggnnggn aucnnnnnnn nagunnaann gnanaagnnn 2340
gcnunacugn nagnnnnacn nncgagcag nncgaaagn nggnnuagu gauccggngg 2400
unnnnnnugg aagngccnuc gcucaacgga uaaaagnuac ncnggggaua acaggcunau 2460
nnncccaag aguncanauc gacggnnng uuuggcaccu cgaugucggc ucnucncauc 2520
cuggggcugn agnngguccc aaggunngg cuguucgcn nuuaaagngg nacnggagcu 2580
ggguunanaa cgucgugaga caguunguc ccuaucngn gngngngnn gannnuugan 2640
nngnnnugn cnuaguacga gaggaccgn nngnacnan cncuggugn ncnguugunn 2700
ngccannngc anngcngnu agcuannun ggnnngaua anngcugaan gcaucuaagn 2760
nngaancnn cnnnagann agnnucncn nnnnnnnnnn nnnnnnnna gnnncnnnn 2820
agannnnn gungauaggn nngnnnugna agnnngna nnnnunagn nnacnnuac 2880
uaaunnnncn nnnnncuunn nnnn 2904

```

```

<210> 5
<211> 13
<212> DNA
<213> Artificial Sequence

```

```

<220>
<221> misc_feature
<223> Primer

```

```

<400> 5
cgtggtgacc ctt 13

```

```

<210> 6
<211> 14
<212> DNA
<213> Artificial Sequence

```

```

<220>
<221> misc_feature
<223> Primer

```

```

<400> 6
cgtcgtcacc gcta 14

```

```

<210> 7

```

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<211> 13  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <221> misc\_feature  
 <223> Primer  
  
 <400> 7  
 cgtggtacc ctt 13  
  
 <210> 8  
 <211> 18  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 8  
 tcacgcgata gcattgcg 18  
  
 <210> 9  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 9  
 tctcacggga gctctccatg c 21  
  
 <210> 10  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 10  
 tgactcaccc atcaacaacc gc 22  
  
 <210> 11  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 11  
 tgactcaccc atcaacaacc gc 22  
  
 <210> 12  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<400> 12  
 tgactcaccc atcaacaacc gc 22  
  
 <210> 13  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 13  
 gaagcagatt tgggtaccac c 21  
  
 <210> 14  
 <211> 26  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 14  
 tacggtcaat gctctgaaat ctgtgg 26  
  
 <210> 15  
 <211> 28  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 15  
 ttatgtaaaa tccattgtcg catccacc 28  
  
 <210> 16  
 <211> 27  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 16  
 tccattact aaaccacac tcaacag 27  
  
 <210> 17  
 <211> 27  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 17  
 taaggccttc gatacgggat aatccta 27  
  
 <210> 18  
 <211> 30  
 <212> DNA  
 <213> Artificial Sequence

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<220>  
 <223> PCR Primer  
  
 <400> 18  
 tactccaatg ctaaaactaa tcgtcccaac 30  
  
 <210> 19  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 19  
 tcctaggaat cacctcccat tccga 25  
  
 <210> 20  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 20  
 tggcagccta gcattagcag gaata 25  
  
 <210> 21  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 21  
 tcgctgacgc cataaaactc ttcac 25  
  
 <210> 22  
 <211> 31  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 22  
 taactaatac taacatctca gacgctcagg a 31  
  
 <210> 23  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 23  
 tactcccacc ctggagcctc 20  
  
 <210> 24

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<211> 25  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 24  
 ttatcacctt tcatgatcac gccct 25  
  
 <210> 25  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 25  
 tgtatgaatg gctccacgag ggt 23  
  
 <210> 26  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 26  
 tggtttgga gagatgtgtt taagt 25  
  
 <210> 27  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 27  
 tctgttaaaa gtgcataccg cca 23  
  
 <210> 28  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 28  
 tgaggatggt ggtcaaggga c 21  
  
 <210> 29  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> PCR Primer  
  
 <400> 29

DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

tggatttgac tgtaatgtgc ta 22  
 <210> 30  
 <211> 28  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> PCR Primer  
 <400> 30  
 tgaagggatt tgactgtaat gtgctatg 28  
 <210> 31  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> PCR Primer  
 <400> 31  
 gtgtgtgtgc tgggtaggat g 21  
 <210> 32  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> PCR Primer  
 <400> 32  
 tggtagaag tgggctaggg catt 24  
 <210> 33  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> PCR Primer  
 <400> 33  
 tggatagc gcctaagcat agtg 24  
 <210> 34  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> PCR Primer  
 <400> 34  
 tttcgtgcaa gaataggagg tggag 25  
 <210> 35  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220>

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<223> PCR Primer

<400> 35  
tagggtcgaa gccgcactcg 20

<210> 36  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> PCR Primer

<400> 36  
tgtgaggcgt attataccat agccg 25

<210> 37  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> PCR Primer

<400> 37  
tagaatctta gctttgggtg ctaatggtg 29

<210> 38  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> PCR Primer

<400> 38  
tggctgaaca ttgtttggtg gtgt 24

<210> 39  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> PCR Primer

<400> 39  
taagtaatgc tagggtgagt ggtaggaag 29

<210> 40  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> PCR Primer

<400> 40  
tttatgggct ttggtgaggg aggta 25

<210> 41  
<211> 30  
<212> DNA

## DOCSTOR-#1144799-v1-Sequence\_Listing.TXT

<213> Artificial Sequence

<220>

<223> PCR Primer

<400> 41

tgctcctatt gataggacat agtggaagtg

30

<210> 42

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR Primer

<400> 42

tggcatttca ctgtaaagag gtggttg

27

<210> 43

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR Primer

<400> 43

tcggaagca ttaggaatgc cattgc

26



## IN THE CLAIMS

1. A forensic method of mitochondrial DNA analysis comprising the steps of:  
providing a forensic evidence sample;  
amplifying two or more segments of mitochondrial DNA obtained from said forensic evidence sample to obtain two or more amplification products;  
determining the molecular masses of said two or more amplification products by mass spectrometry, without sequencing said two or more amplification products; and  
comparing said molecular masses of said two or more amplification products with at least one database comprising a plurality of known molecular masses from said two or more segments of mitochondrial DNA from a plurality of subjects thereby reaching a forensic conclusion.
2. A forensic method of claim 1 wherein the sample is from an animal.
3. A forensic method of claim 2 wherein the animal is a human.
4. A forensic method of claim 1 wherein the sample is from a nonhuman eukaryotic organism, a fungus, a parasite, or a protozoan.
5. A forensic method of claim 1 further comprising the step of calculating base compositions of said two or more amplification products from said molecular masses without sequencing said amplification products.
6. A forensic method of claim 5 further comprising the step of comparing said base compositions of said two or more amplification products with at least one database comprising a plurality of known base compositions from said two or more segments of mitochondrial DNA from a plurality of subjects, thereby reaching a forensic conclusion.
7. A forensic method of claim 5 wherein the sample is from an animal.

8. A forensic method of claim 7 wherein the animal is a human.
9. A forensic method of claim 5 wherein the sample is from a nonhuman eukaryotic organism, a fungus, a parasite, or a protozoan.
10. A forensic method of claim 1 wherein the molecular masses of the amplification products are determined by ESI-FTICR mass spectrometry.
11. A forensic method of claim 1 wherein the molecular masses of the amplification products are determined by ESI-TOF mass spectrometry.
12. A forensic method of mitochondrial DNA analysis of claim 1 comprising the steps of:  
digesting said two or more amplification products with one or more restriction enzymes to produce a plurality of restriction fragments before said mass spectrometry; and  
determining the molecular masses of said restriction fragments by mass spectrometry without sequencing said restriction fragments.
13. A forensic method of claim 12 further comprising comparing said molecular masses of said restriction fragments with at least one database comprising a plurality of known molecular masses of said restriction fragments of said two or more segments of mitochondrial DNA from a plurality of subjects, thereby reaching a forensic conclusion.
14. A forensic method of claim 12 wherein the restriction enzymes are any combination of *RsaI*, *HpaII*, *HpyCH4IV*, *PacI*, and *EaeI*.
15. A forensic method of claim 13 wherein the sample is from an animal.
16. A forensic method of claim 15 wherein the animal is a human.
17. A forensic method of claim 13 wherein the sample is from a nonhuman eukaryotic organism, a fungus, a parasite, or a protozoan.

18. A forensic method of claim 1 wherein said two or more segments of mitochondrial DNA comprise a portion of a hypervariable region of mitochondrial DNA.

19. A forensic method of claim 18 wherein the hypervariable region comprises HVR1 or HVR2.

20. A forensic method for tracking the geographic location of a subject comprising the steps of:

providing a forensic evidence sample containing mitochondrial DNA obtained from a geographic location;

amplifying two or more segments of mitochondrial DNA obtained from said forensic evidence sample to obtain two or more amplification products;

determining the molecular masses of said two or more amplification products by mass spectrometry, without sequencing said two or more amplification products; and

comparing said molecular masses of said two or more amplification products with at least one database comprising a plurality of known molecular masses from said two or more segments of mitochondrial DNA from a plurality of subjects thereby indicating at least transient presence of the subject at the geographic location.

21. A method of characterizing the heteroplasmy of two or more segments of mitochondrial DNA of a subject comprising the steps of:

providing a sample from said subject;

amplifying said two or more segments of mitochondrial DNA from said sample with two or more primer pairs to obtain a plurality of amplification products;

determining the molecular masses of said plurality of amplification products of the mitochondrial DNA by mass spectrometry, without sequencing said plurality of amplification products; and

determining the base compositions of the plurality of amplification products thereby characterizing the heteroplasmy.

22. A method of claim 21 further comprising obtaining a plurality of amplification products from a plurality of samples of mitochondrial DNA from said subject wherein the characterization of heteroplasmy indicates the rate of naturally occurring mutations in said mitochondrial DNA.

23. A method of claim 22 further comprising comparing said heteroplasmy in said two or more segments of mitochondrial DNA from said sample with at least one database comprising a plurality of base compositions from said two or more segments of mitochondrial DNA from a plurality of subjects with one or more mitochondrial diseases, wherein said comparing correlates said heteroplasmy with the onset of said one or more mitochondrial diseases in a subject.

24. The method of claim 23 wherein the mitochondrial disease is Alpers Disease, Barth syndrome, Beta-oxidation Defects, Carnitine-Acyl-Carnitine Deficiency, Carnitine Deficiency, Co-Enzyme Q10 Deficiency, Complex I Deficiency, Complex II Deficiency, Complex III Deficiency, Complex IV Deficiency, Complex V Deficiency, COX Deficiency, CPEO, CPT I Deficiency, CPT II Deficiency, Glutaric Aciduria Type II, KSS, Lactic Acidosis, LCAD, LCHAD, Leigh Disease or Syndrome, LHON, Lethal Infantile Cardiomyopathy, Luft Disease, MAD, MCA, MELAS, MERRF, Mitochondrial Cytopathy, Mitochondrial DNA Depletion, Mitochondrial Encephalopathy, Mitochondrial Myopathy, MNGIE, NARP, Pearson Syndrome, Pyruvate Carboxylase Deficiency, Pyruvate Dehydrogenase Deficiency, Respiratory Chain, SCAD, SCHAD, or VLCAD.

25. A method of claim 1 wherein said amplifying comprises amplification with at least one oligonucleotide primer selected from the group consisting of SEQ ID NO: 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, and 43.

26. The method of claim 25 wherein said at least one oligonucleotide primer comprises at least one modified nucleobase.

27. The method of claim 26 wherein said modified nucleobase is 5-propynyl-uridine or 5-propynyl-cytidine.

28. A forensic method of claim 1 wherein said two or more amplification products are generated from two hypervariable portions of the noncoding region of mitochondrial DNA using flanking primers.

29. A forensic method of claim 1 wherein said two or more segments of mitochondrial DNA comprise the entire mitochondrial DNA of said subject.

30. A forensic method of claim 1 wherein said forensic conclusion comprises identification of at least one subject from whom said forensic evidence sample is obtained by comparing said molecular masses of said two or more amplification products with said plurality of known molecular masses in said at least one database.

31. A forensic method of claim 30 wherein said forensic conclusion is the identification of a criminal.

32. The forensic method of claim 30 wherein said forensic conclusion is the identification of a crime victim.

33. A forensic method of claim 1 further comprising determining the relative amounts of said two or more amplification products from the abundance of mass spectral peaks corresponding to said two or more amplification products.

34. A forensic method of claim 1 wherein said forensic conclusion further comprises determining the movement of at least one subject from whom said forensic evidence sample is obtained by mitochondrial DNA analysis of a plurality of forensic evidence samples obtained from a plurality of locations.

35. A forensic method of claim 1 wherein said at least one database is a Federal Bureau of Investigation mitochondrial DNA database.

36. A forensic method of claim 1 wherein said forensic conclusion comprises identification of a missing person, detection and identification of a known bioagent, detection and identification of an unknown bioagent, elimination of an individual as a crime suspect, identification of an individual as a crime suspect, identification of a location as a crime scene, identification of a location as an accident scene, identification of evidence useful in a court of law, identification of evidence useful in a criminal investigation, or identification of one or more biological samples from a crime scene.
37. A forensic method of mitochondrial DNA analysis of claim 5 comprising the steps of :  
digesting said two or more amplification products with one or more restriction enzymes to produce a plurality of restriction fragments before said mass spectrometry;  
determining the molecular masses of said restriction fragments by mass spectrometry without sequencing said restriction fragments: and  
calculating base compositions of said restriction fragments from said molecular masses.
38. A forensic method of claim 37 further comprising comparing said base compositions of said restriction fragments with at least one database comprising a plurality of known base compositions of said restriction fragments of said two or more segments of mitochondrial DNA from a plurality of subjects, thereby reaching a forensic conclusion.
39. A forensic method of claim 38 wherein the restriction enzymes are any combination of *RsaI*, *HpaII*, *HpyCH4IV*, *PaeI*, and *EaeI*.
40. A forensic method of claim 37 wherein the sample is from an animal.
41. A forensic method of claim 40 wherein the animal is a human.
42. A forensic method of claim 37 wherein the sample is from a nonhuman eukaryotic organism, a fungus, a parasite, or a protozoan.

43. A forensic method of claim 5 wherein said two or more segments of mitochondrial DNA comprise a portion of a hypervariable region of mitochondrial DNA.

44. A forensic method of claim 43 wherein the hypervariable region comprises HVR1 or HVR2.

45. A forensic method of claim 5 wherein said two or more amplification products are generated from two hypervariable portions of the noncoding region of mitochondrial DNA using flanking primers.

46. A forensic method of claim 5 wherein said two or more segments of mitochondrial DNA comprise the entire mitochondrial DNA of said subject.

47. A forensic method of claim 6 wherein said forensic conclusion comprises identification of at least one subject from whom said forensic evidence sample is obtained by comparing said molecular masses of said two or more amplification products with said plurality of known molecular masses in said at least one database.

48. A forensic method of claim 5 wherein said forensic conclusion is the identification of a criminal.

49. The forensic method of claim 5 wherein said forensic conclusion is the identification of a crime victim.

50. A forensic method of claim 5 further comprising determining the relative amounts of said two or more amplification products from the abundance of mass spectral peaks corresponding to said two or more amplification products.

51. A forensic method of claim 5 wherein said forensic conclusion further comprises determining the movement of at least one subject from whom said forensic evidence sample is

obtained by mitochondrial DNA analysis of a plurality of forensic evidence samples obtained from a plurality of locations.

52. A forensic method of claim 5 wherein the molecular masses of the amplification products are determined by ESI-FTICR mass spectrometry.

53. A forensic method of claim 6 wherein said at least one database is a Federal Bureau of Investigation mitochondrial DNA database.

54. A forensic method of claim 5 wherein said forensic conclusion comprises identification of a missing person, detection and identification of a known bioagent, detection and identification of an unknown bioagent, elimination of an individual as a crime suspect, identification of an individual as a crime suspect, identification of a location as a crime scene, identification of a location as an accident scene, identification of evidence useful in a court of law, identification of evidence useful in a criminal investigation, or identification of one or more biological samples from a crime scene.

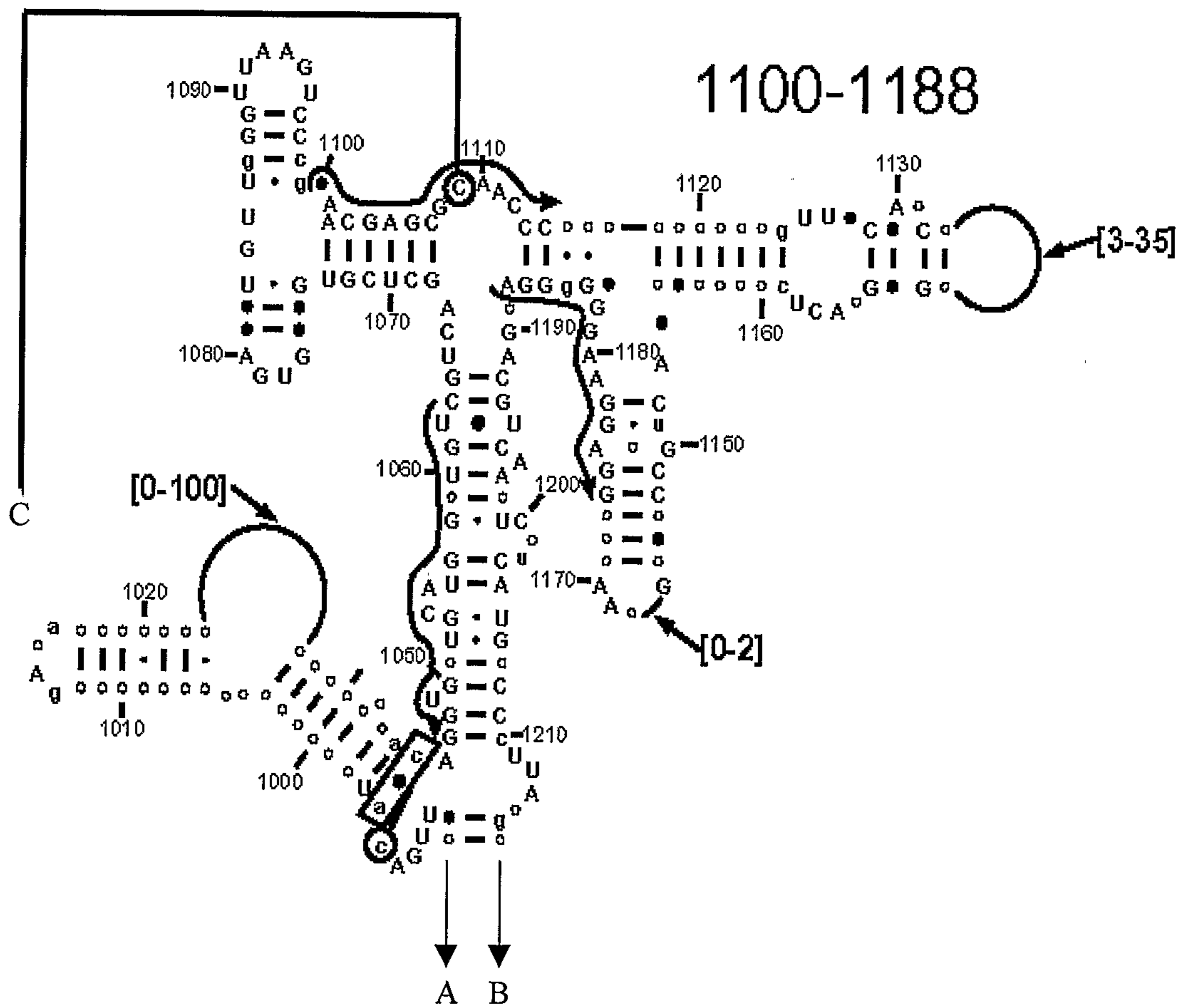
55. A method of claim 21 wherein said heteroplasmy is length heteroplasmy, nucleotide polymorphism heteroplasmy, or both length heteroplasmy and nucleotide polymorphism heteroplasmy.

56. A method of claim 21 wherein said mass spectrometry is electrospray Fourier transform ion cyclotron resonance mass spectrometry.

57. A method of claim 21 wherein said sample from said subject is a forensic evidence sample.



### FIG. 1A-1



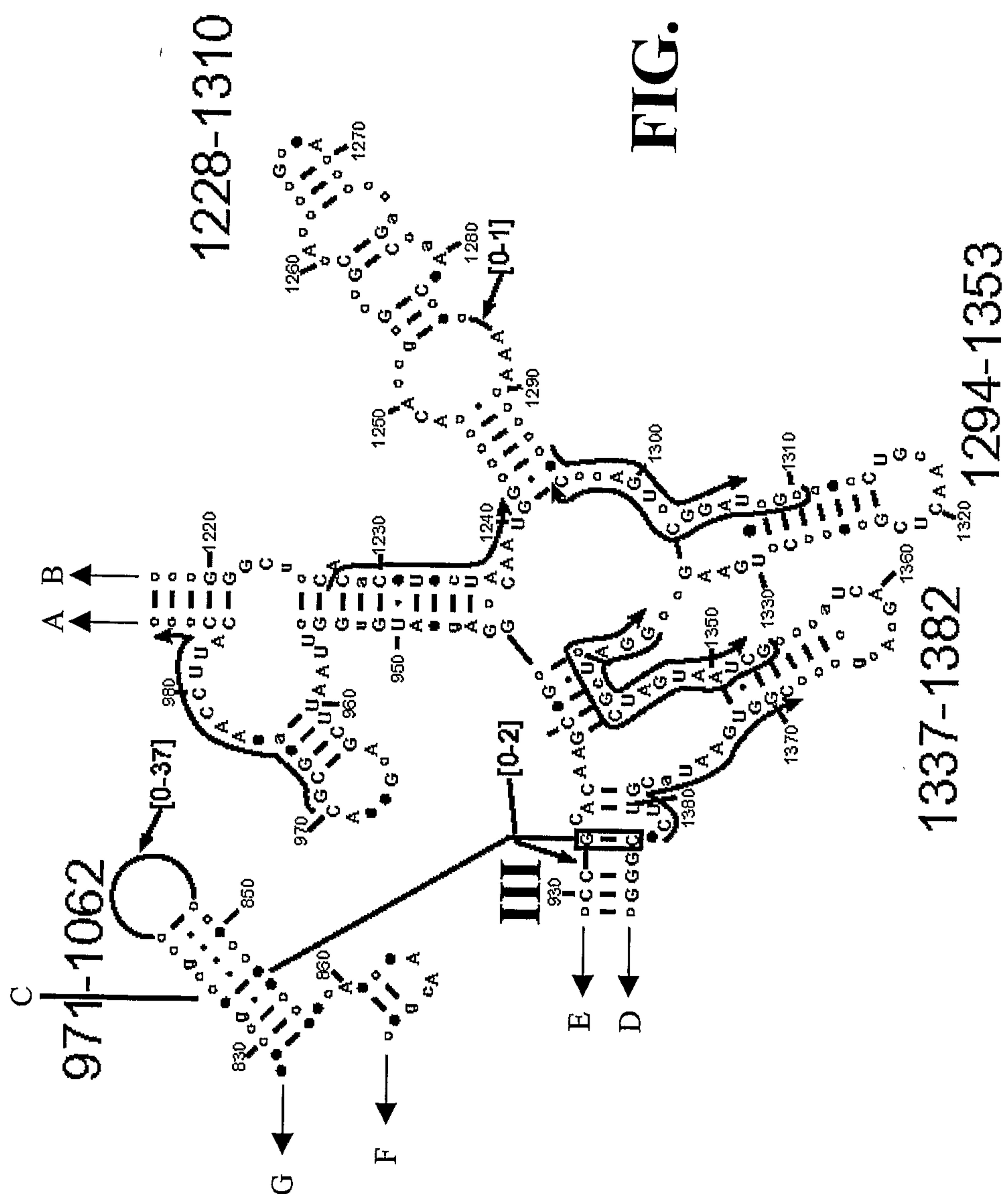


FIG. 1A-2

FIG. 1A-3

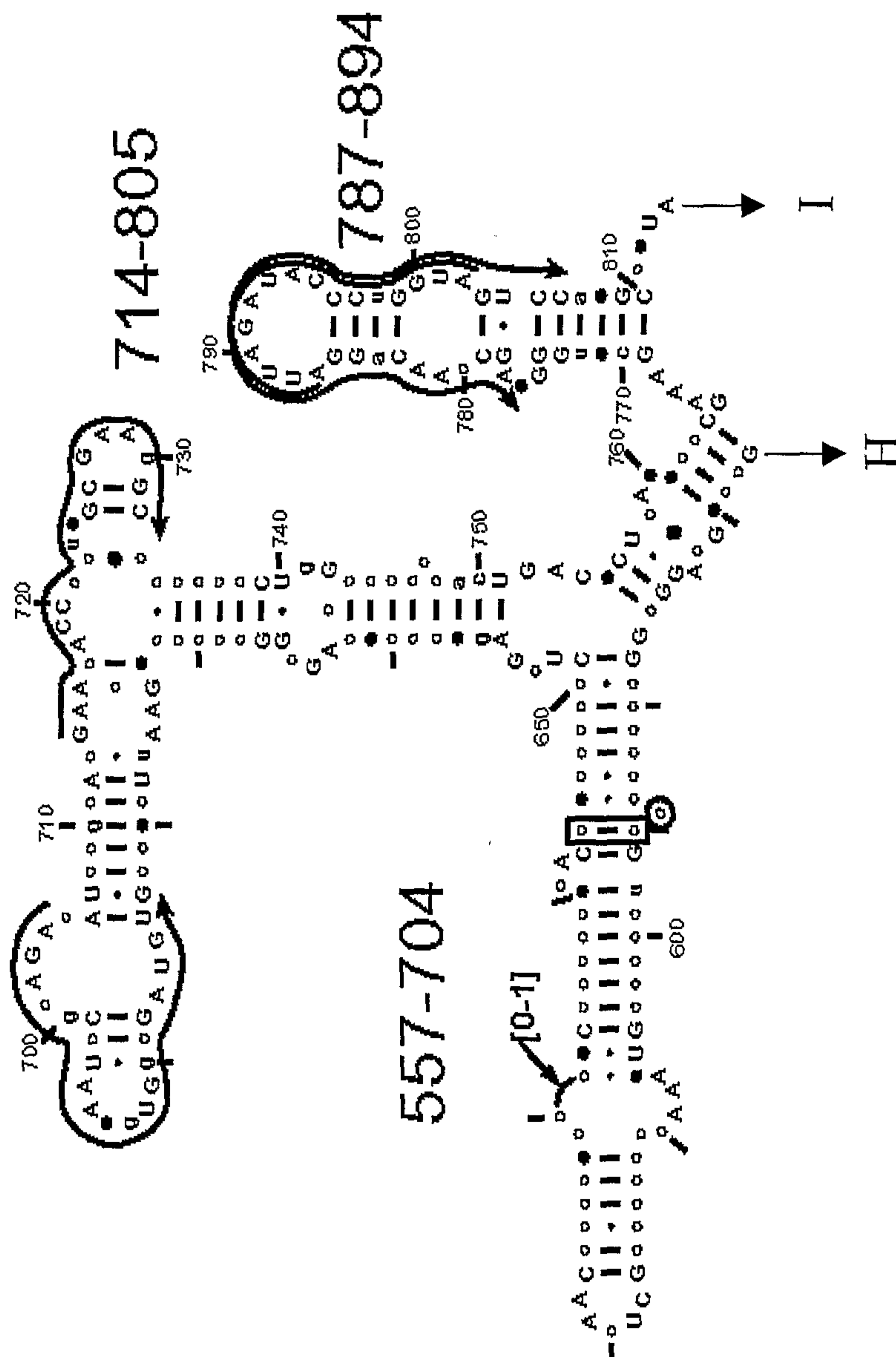
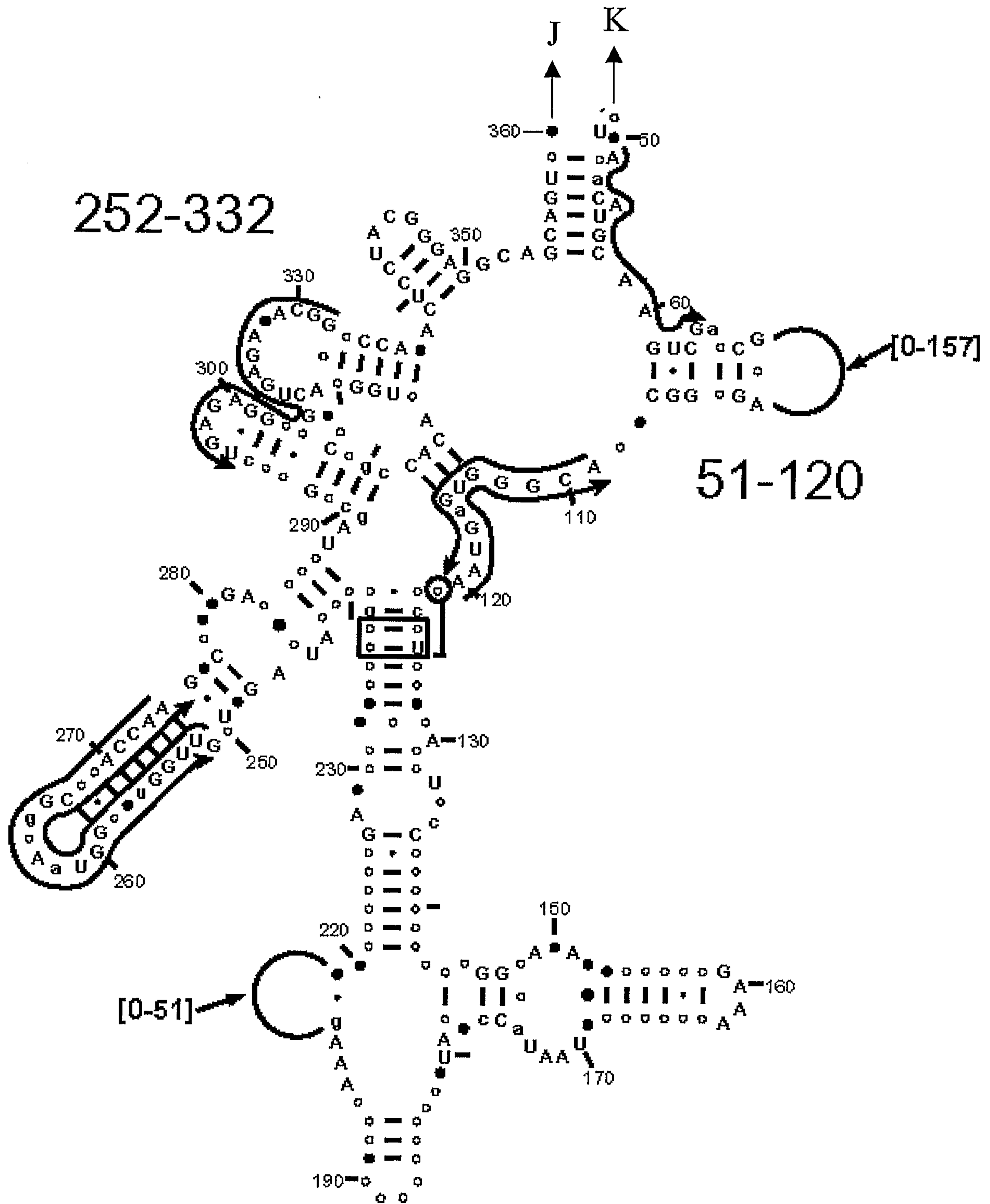


FIG. 1A-4



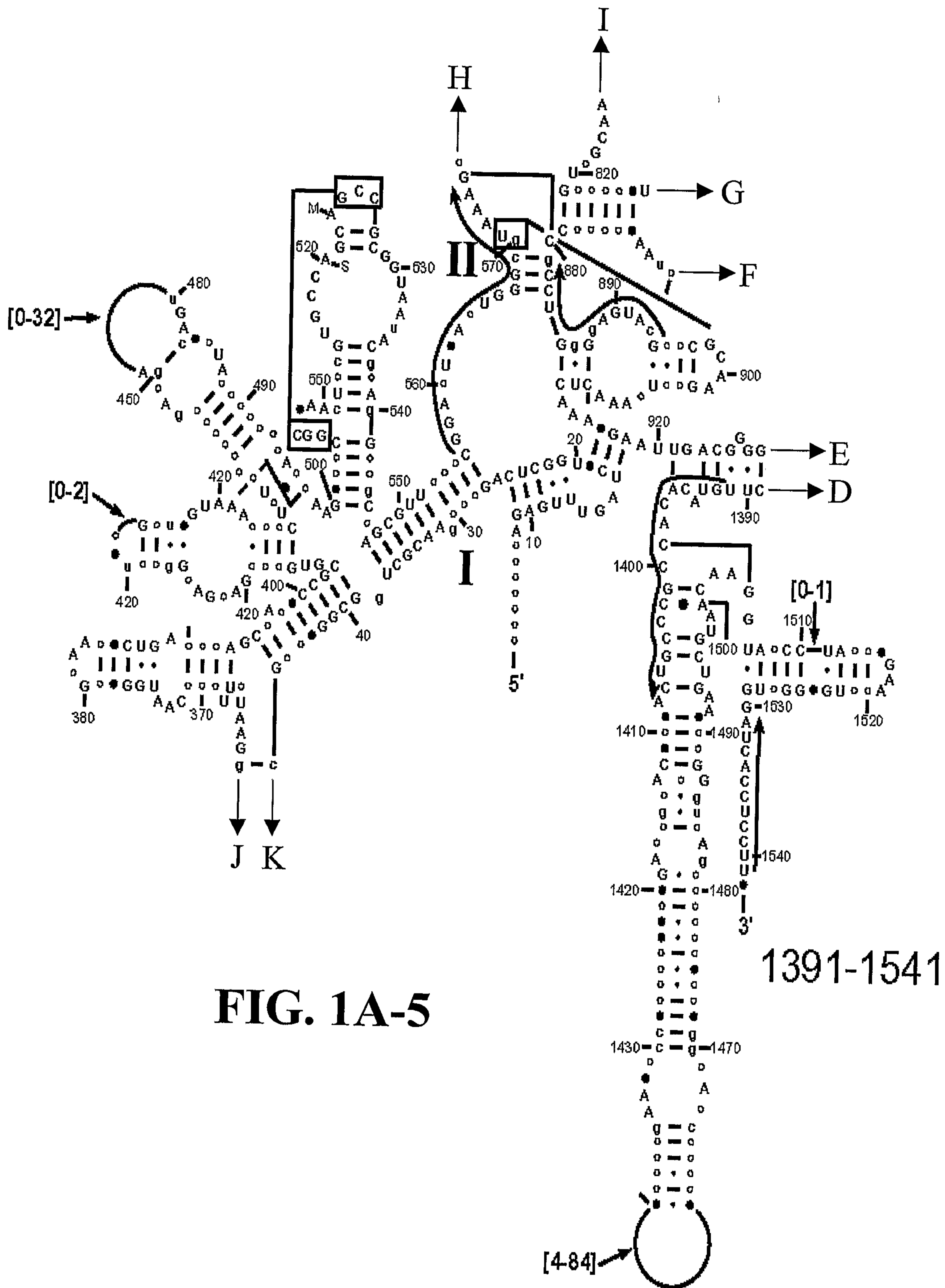


FIG. 1A-5

FIG. 1B

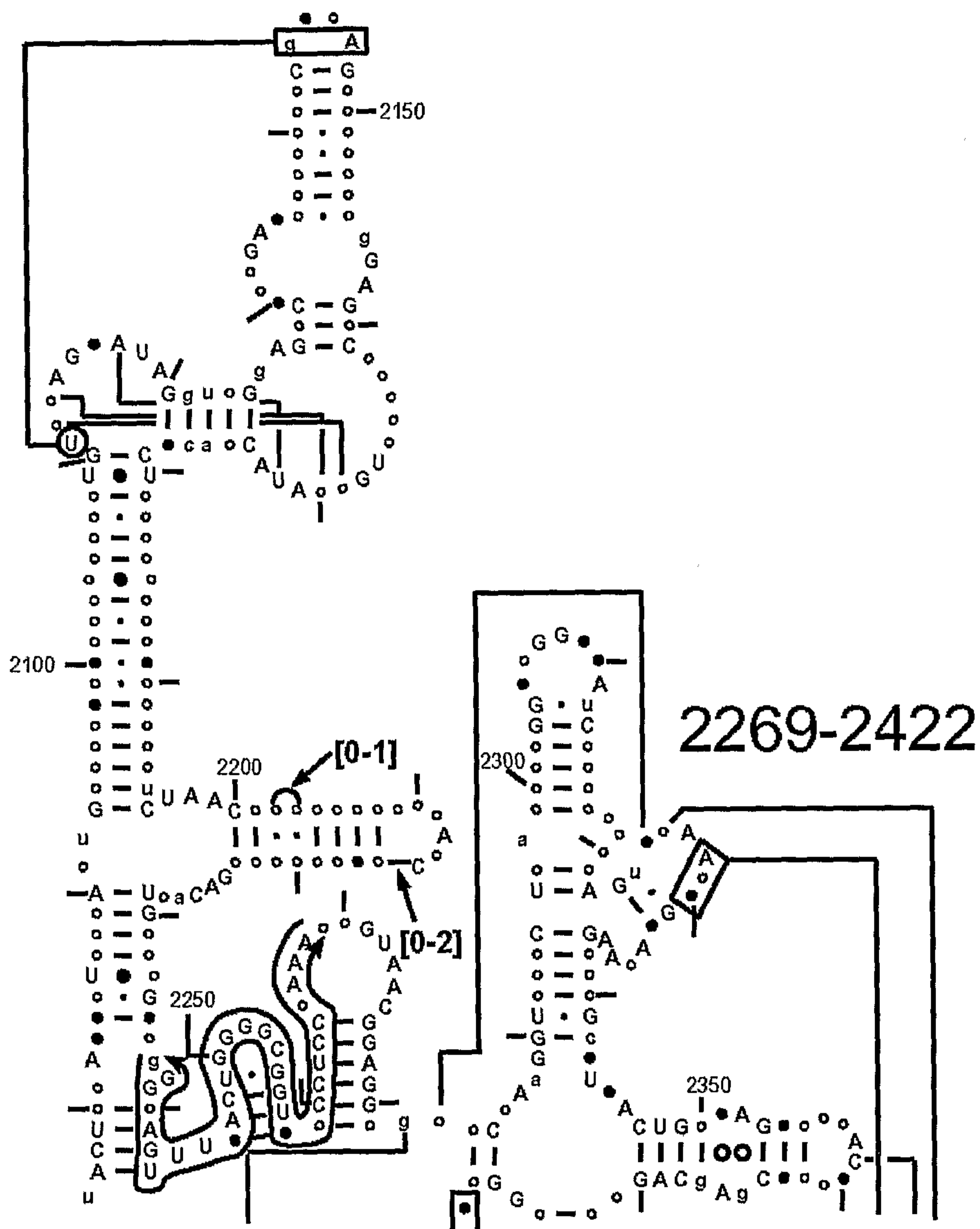


FIG. 1C

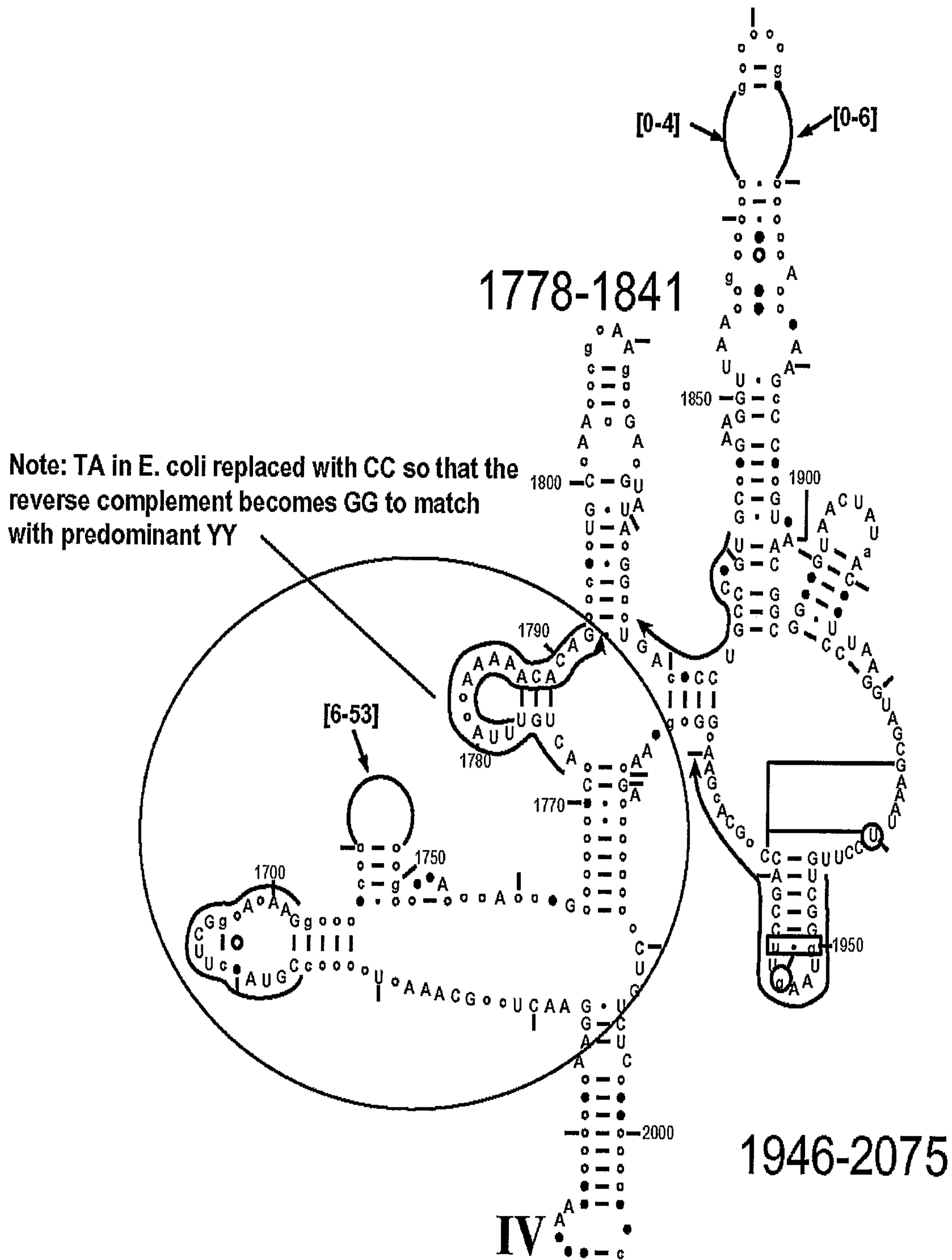


FIG. 1D

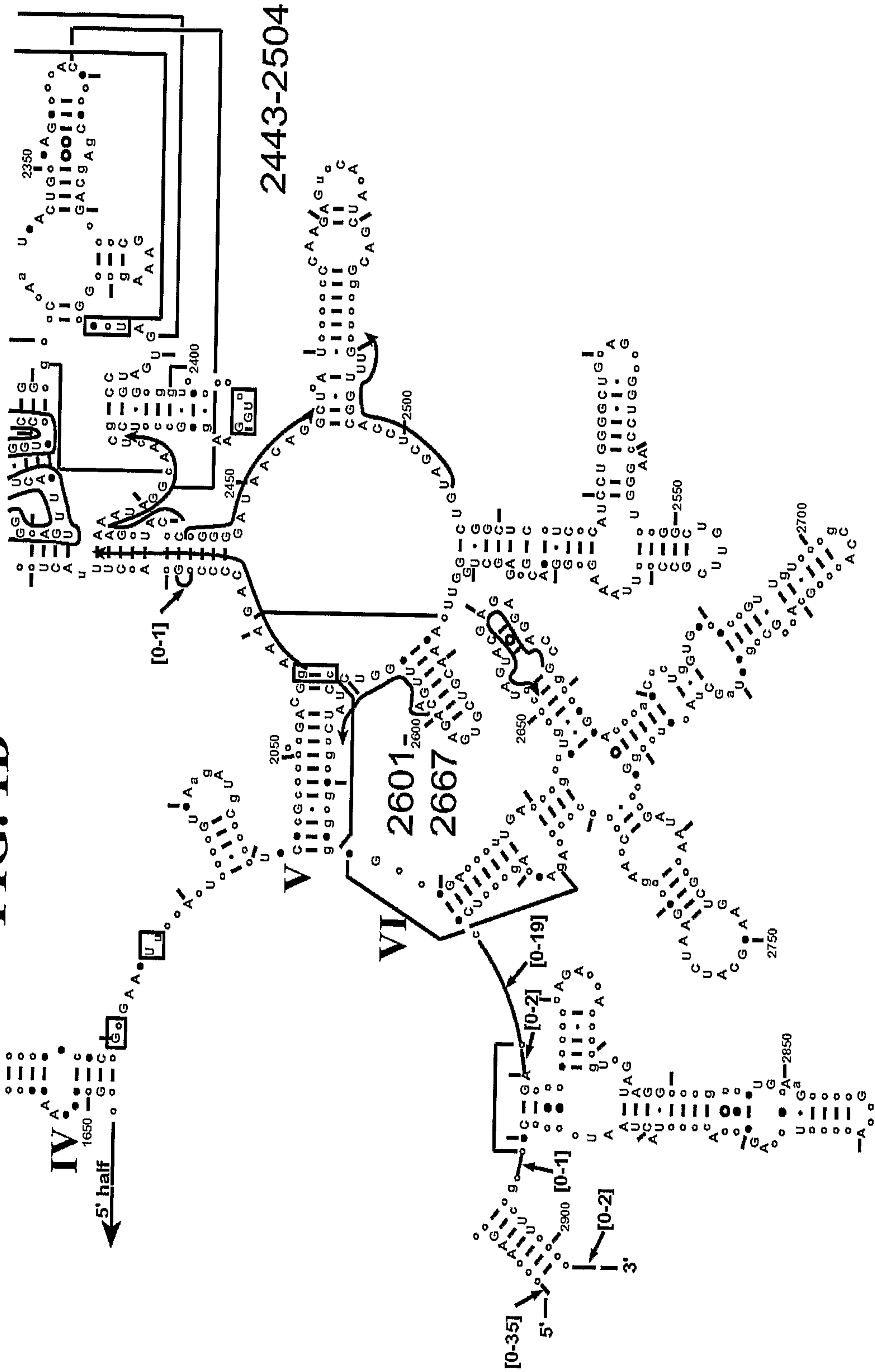




FIG. 1E

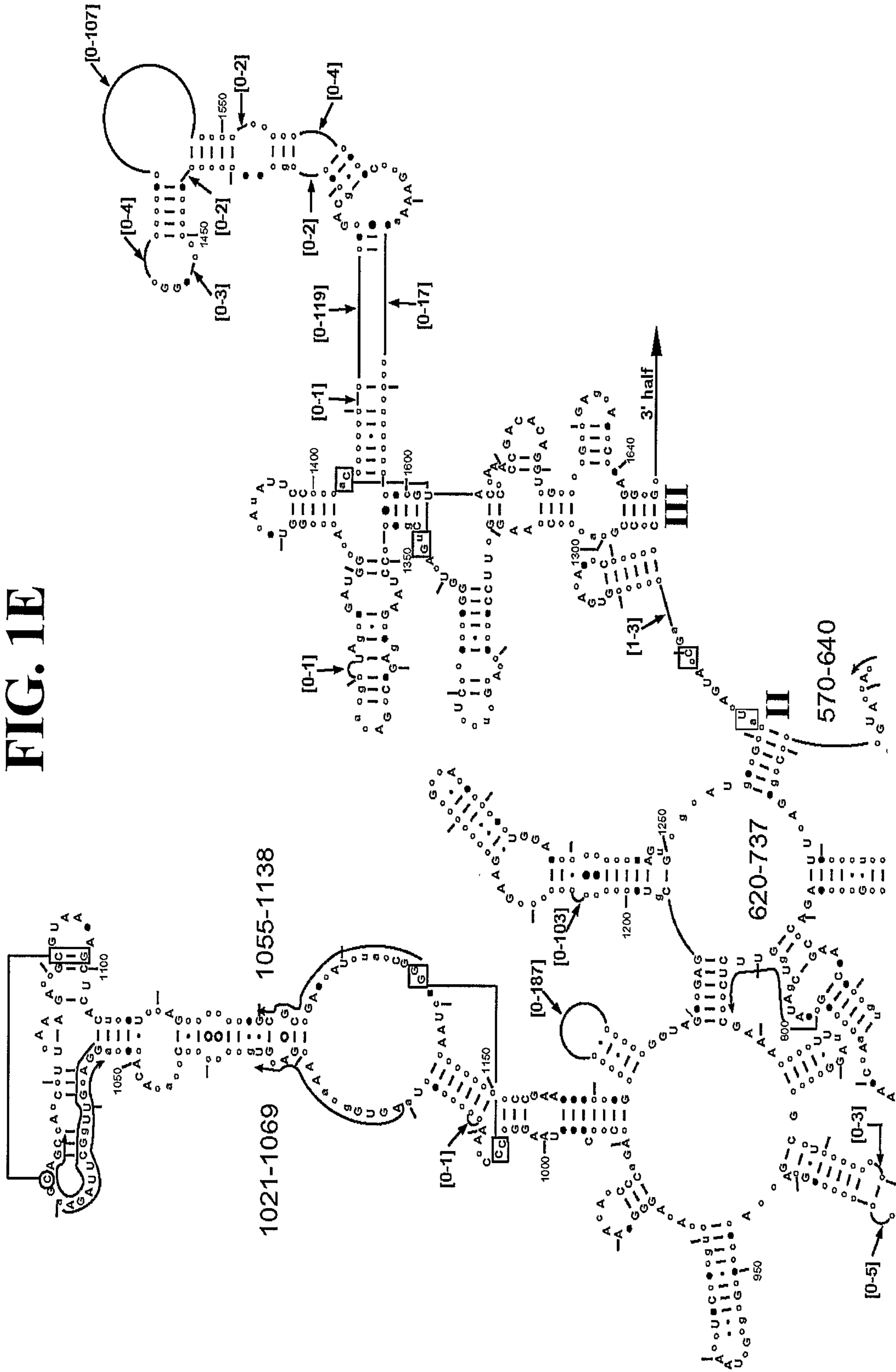


FIG. 1F

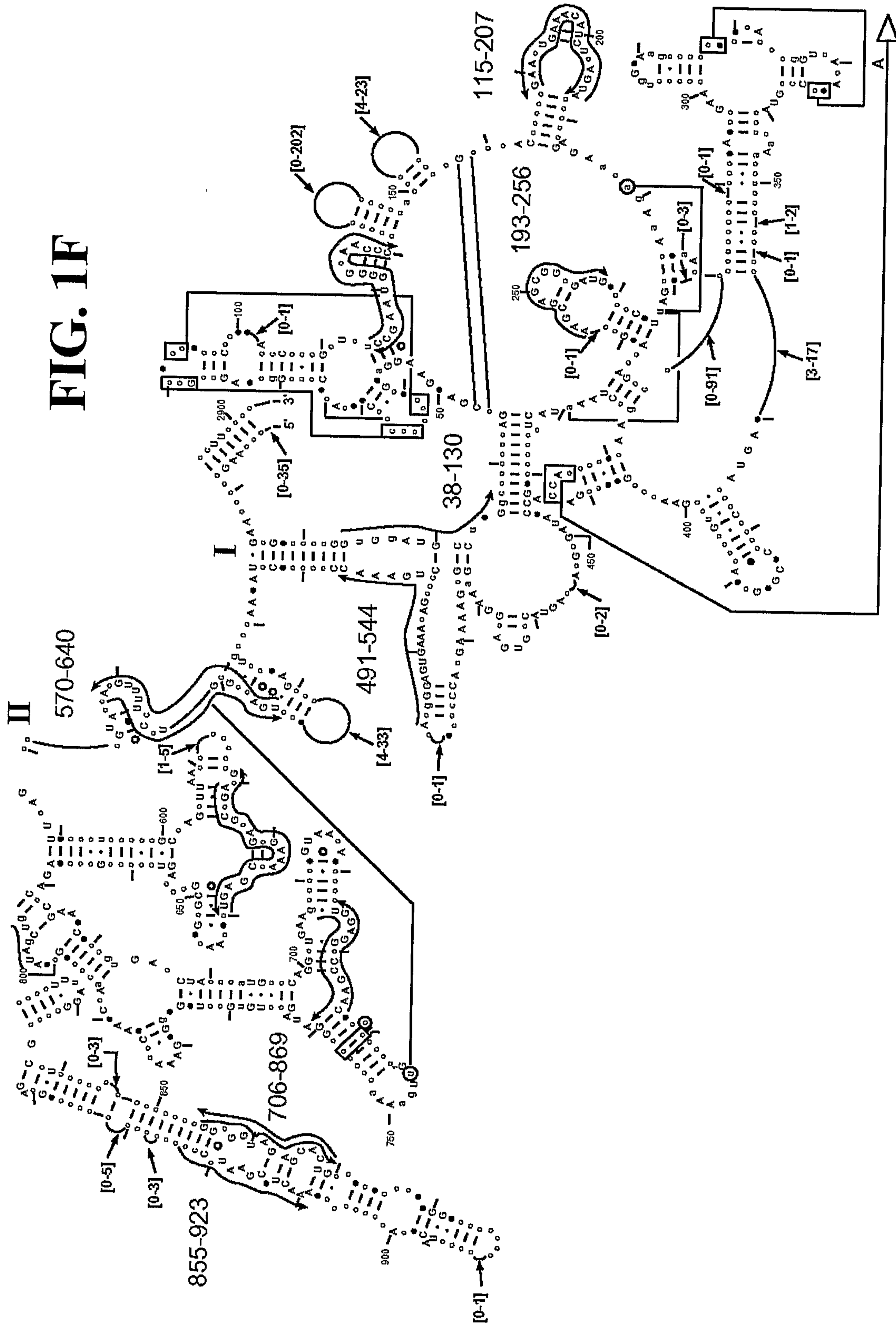


FIG. 1G

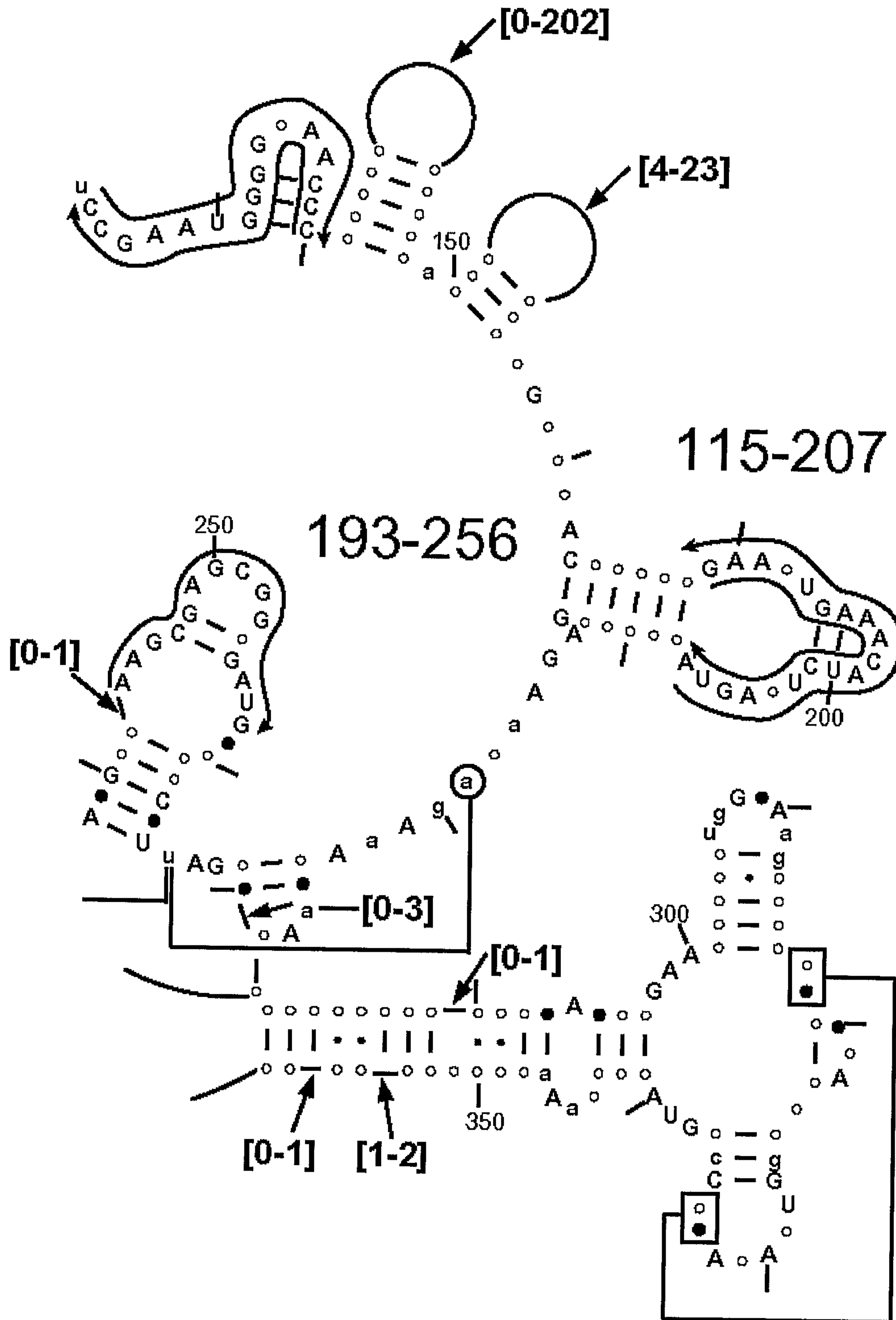


FIG. 1H

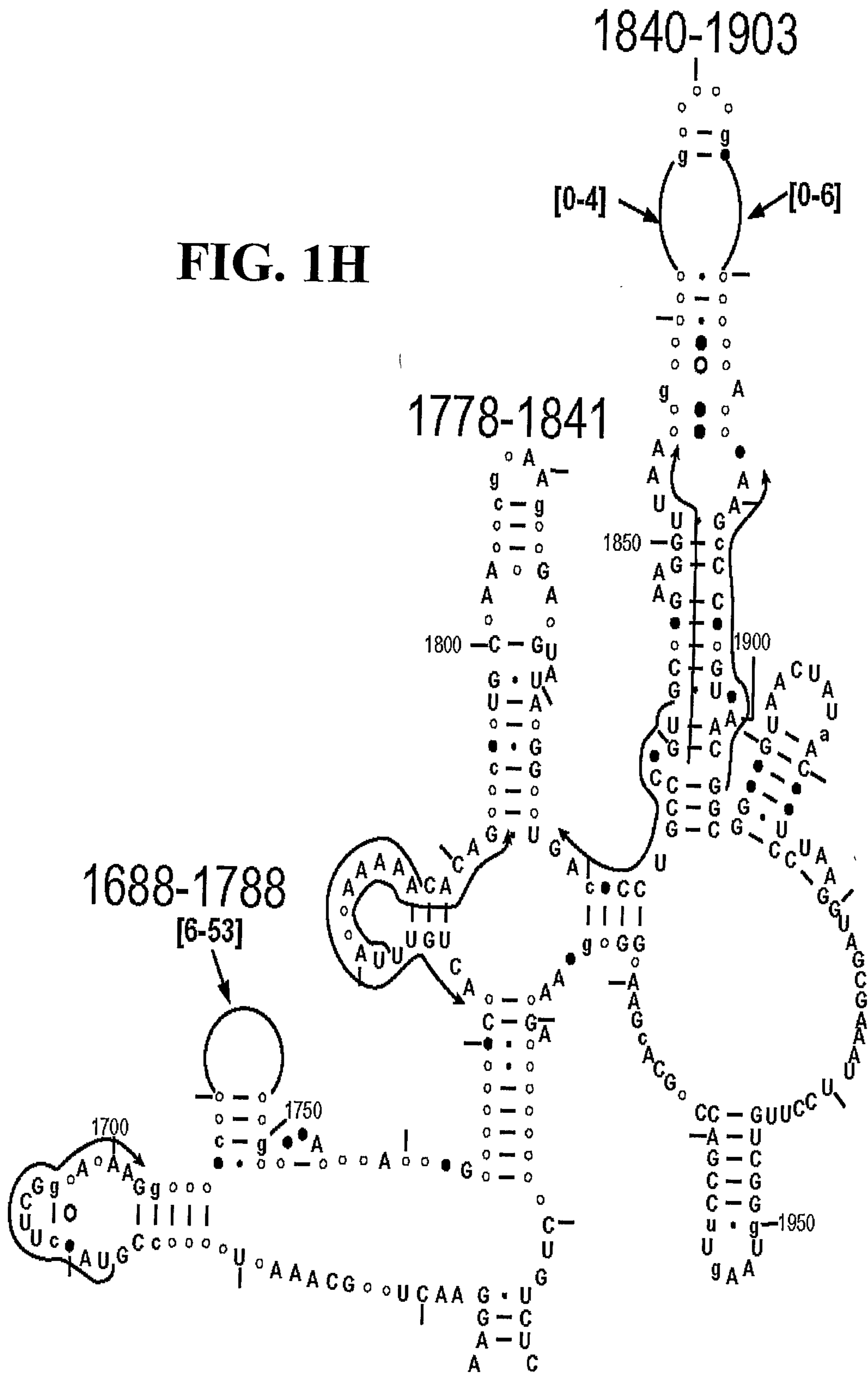


FIG. 2

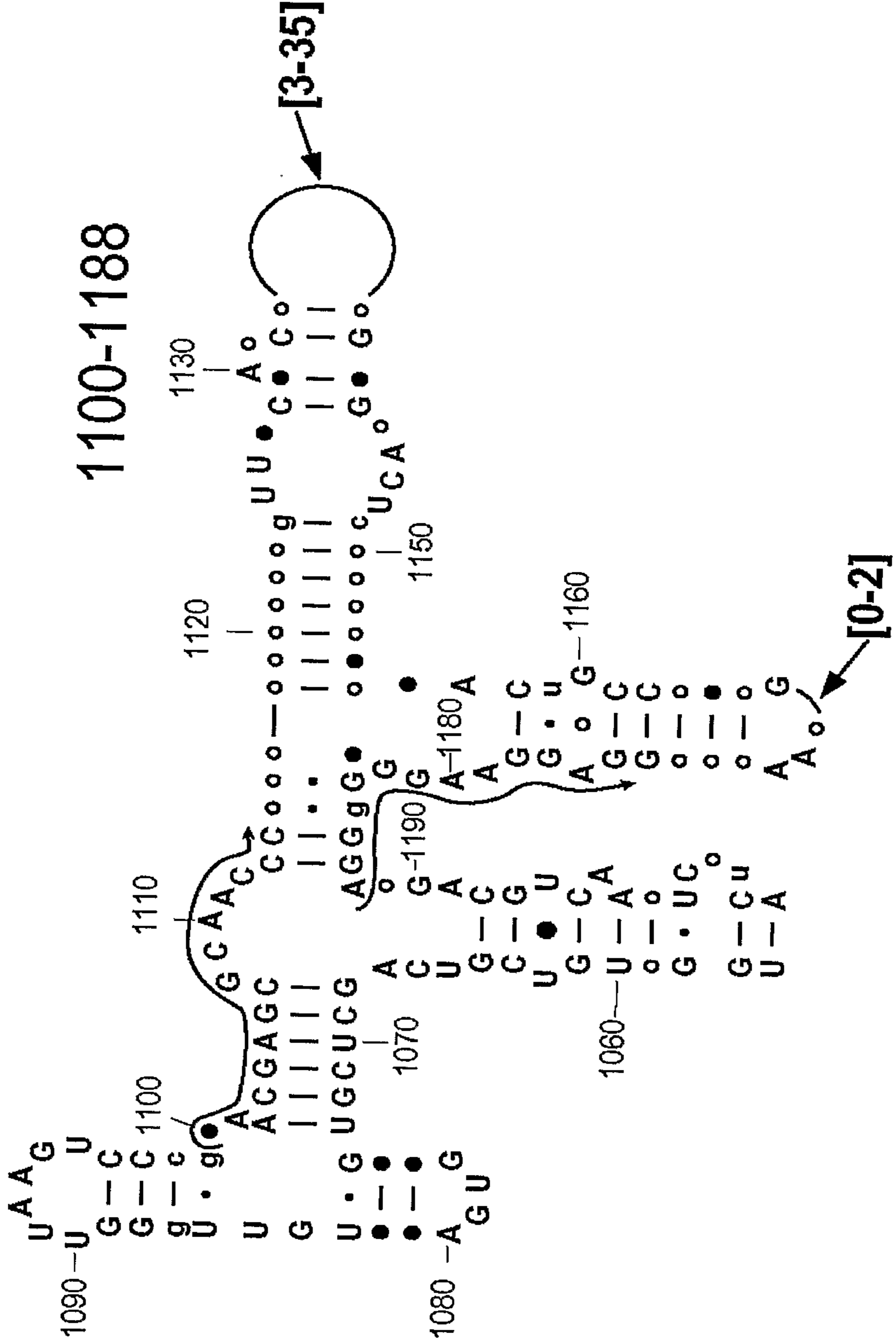


FIG. 3

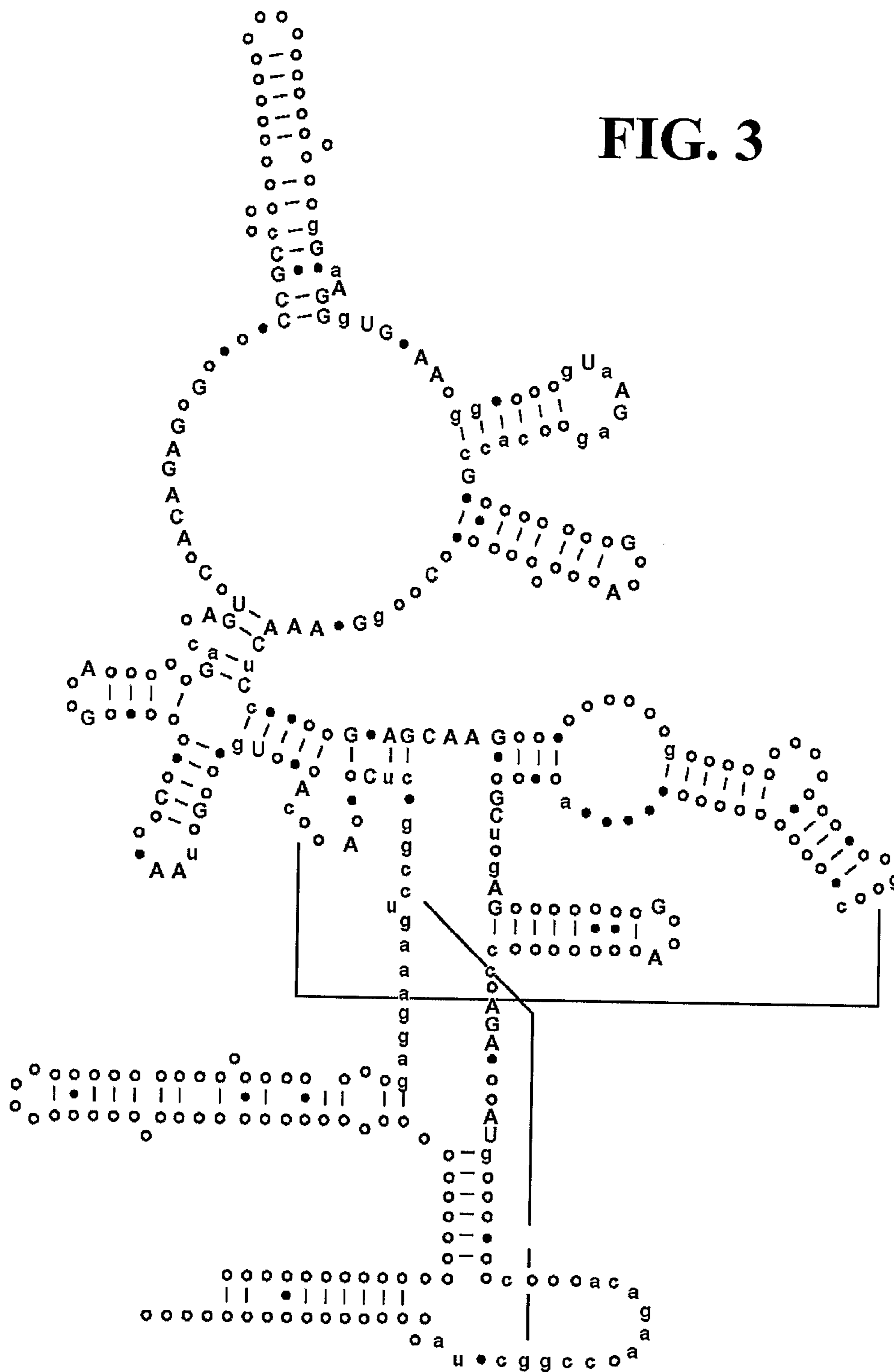
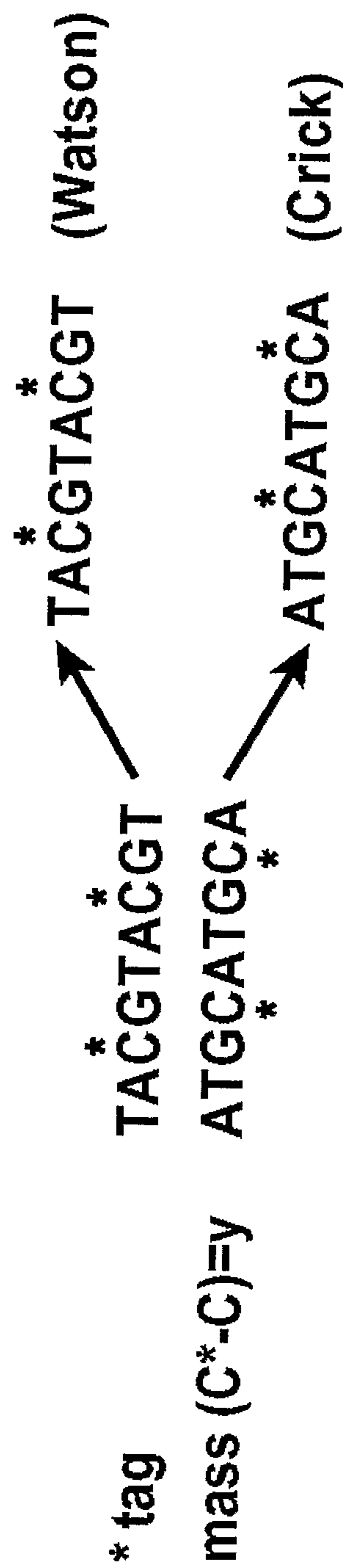
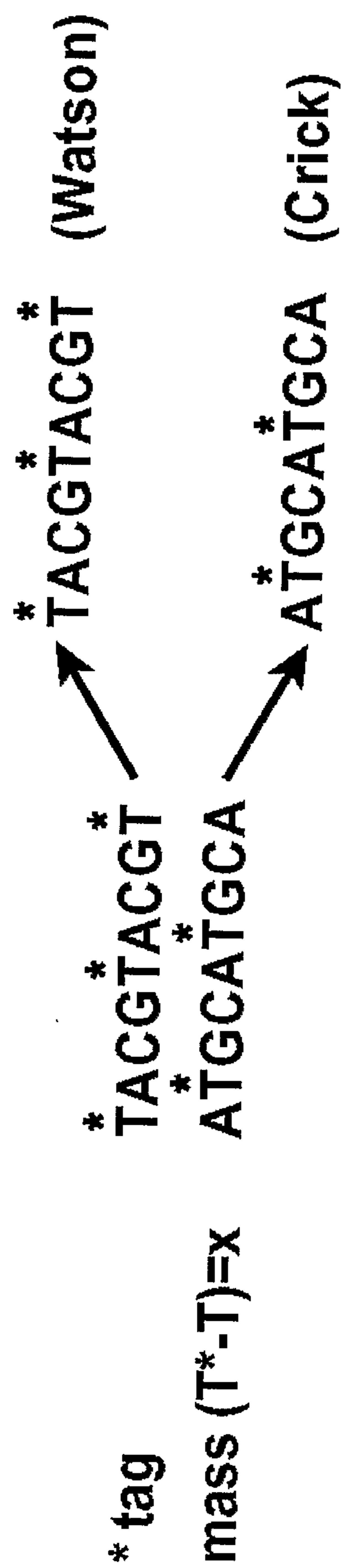
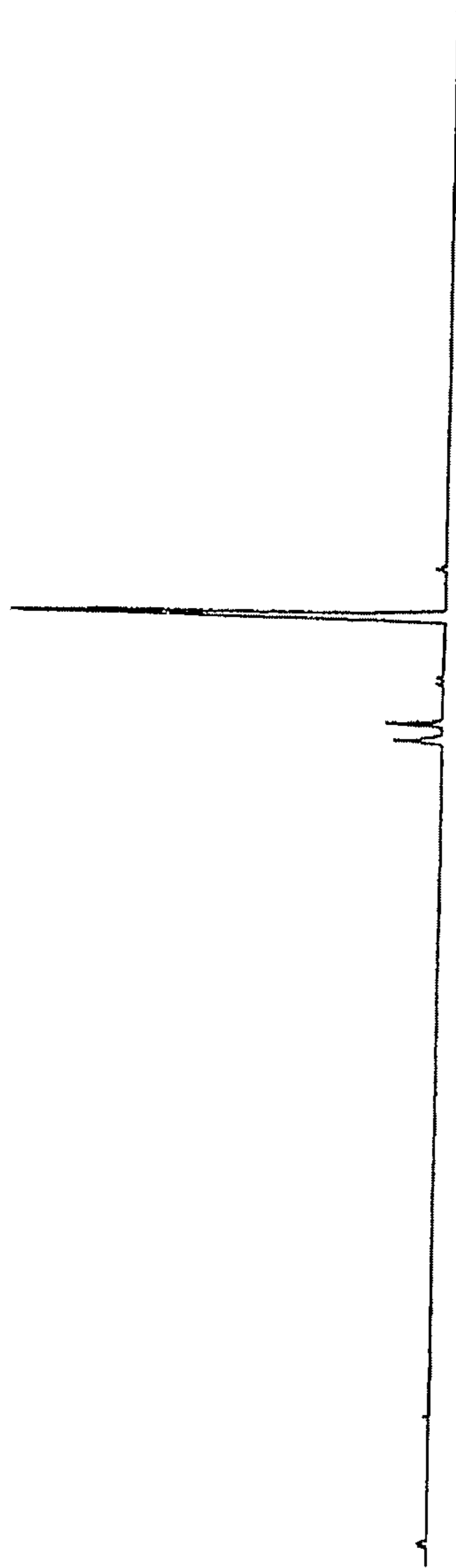


FIG. 4

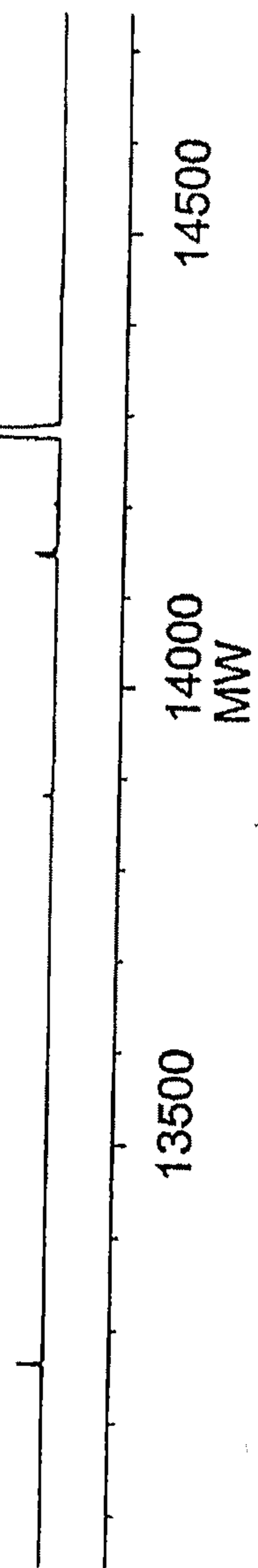


**FIG. 5**

*B. anthracis* ( $A_{14}G_9C_{14}T_9$ )  $MW_{meas} = 14072.2$



*B. anthracis*\* ( $A_1A^*_{13}G_9C_{14}T_9$ )  $MW_{meas} = 14280.9$





**FIG. 6**

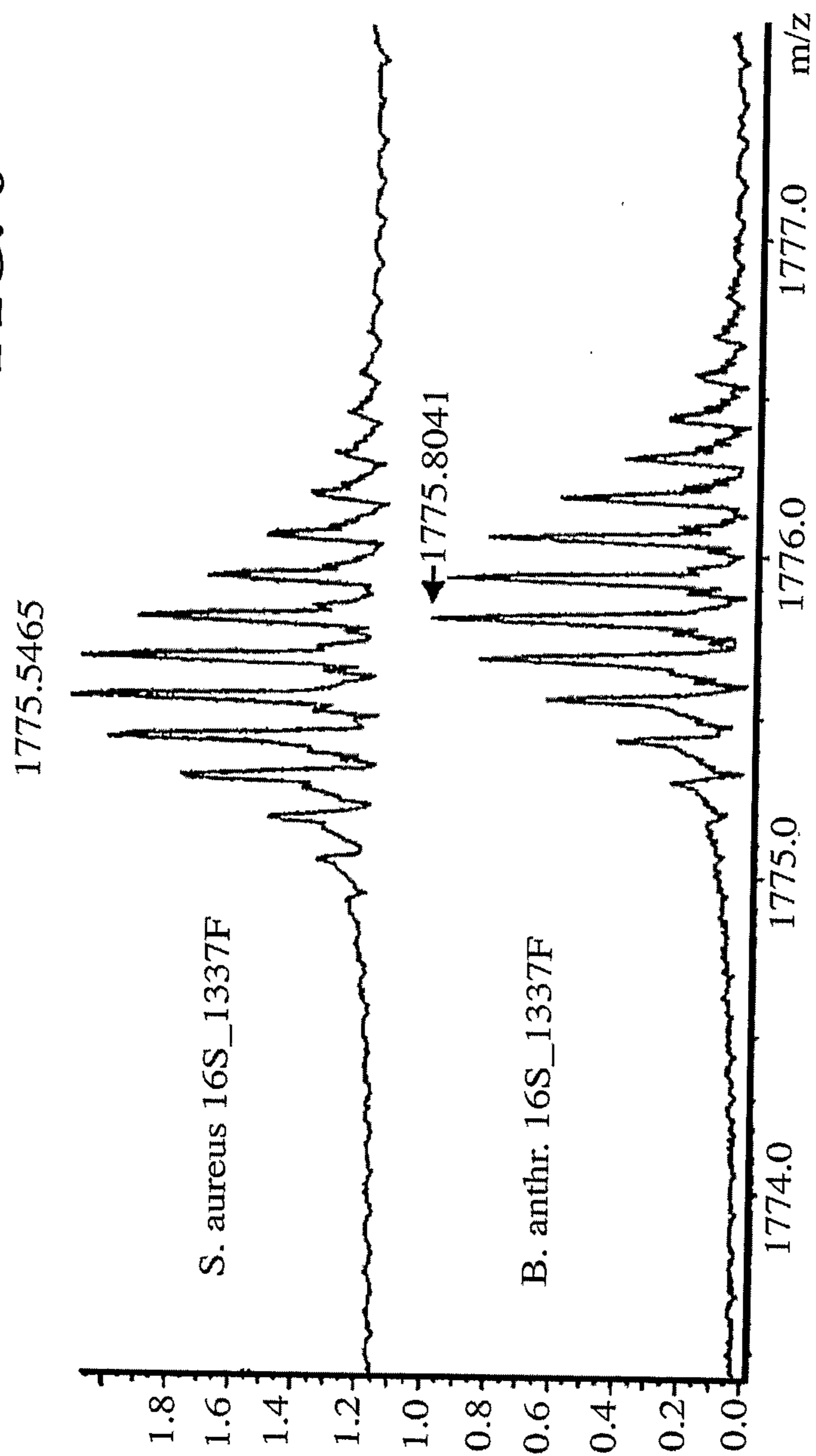
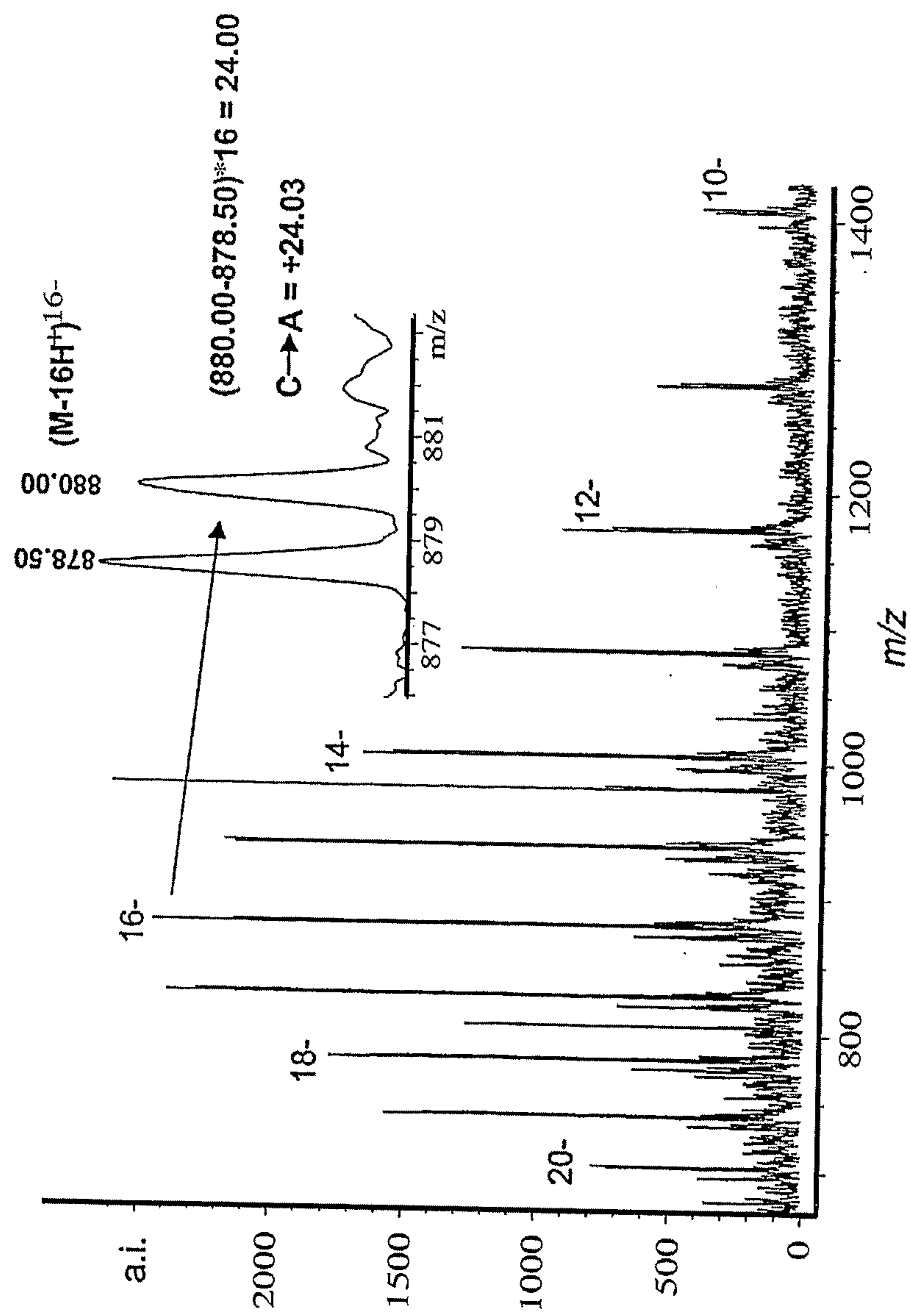
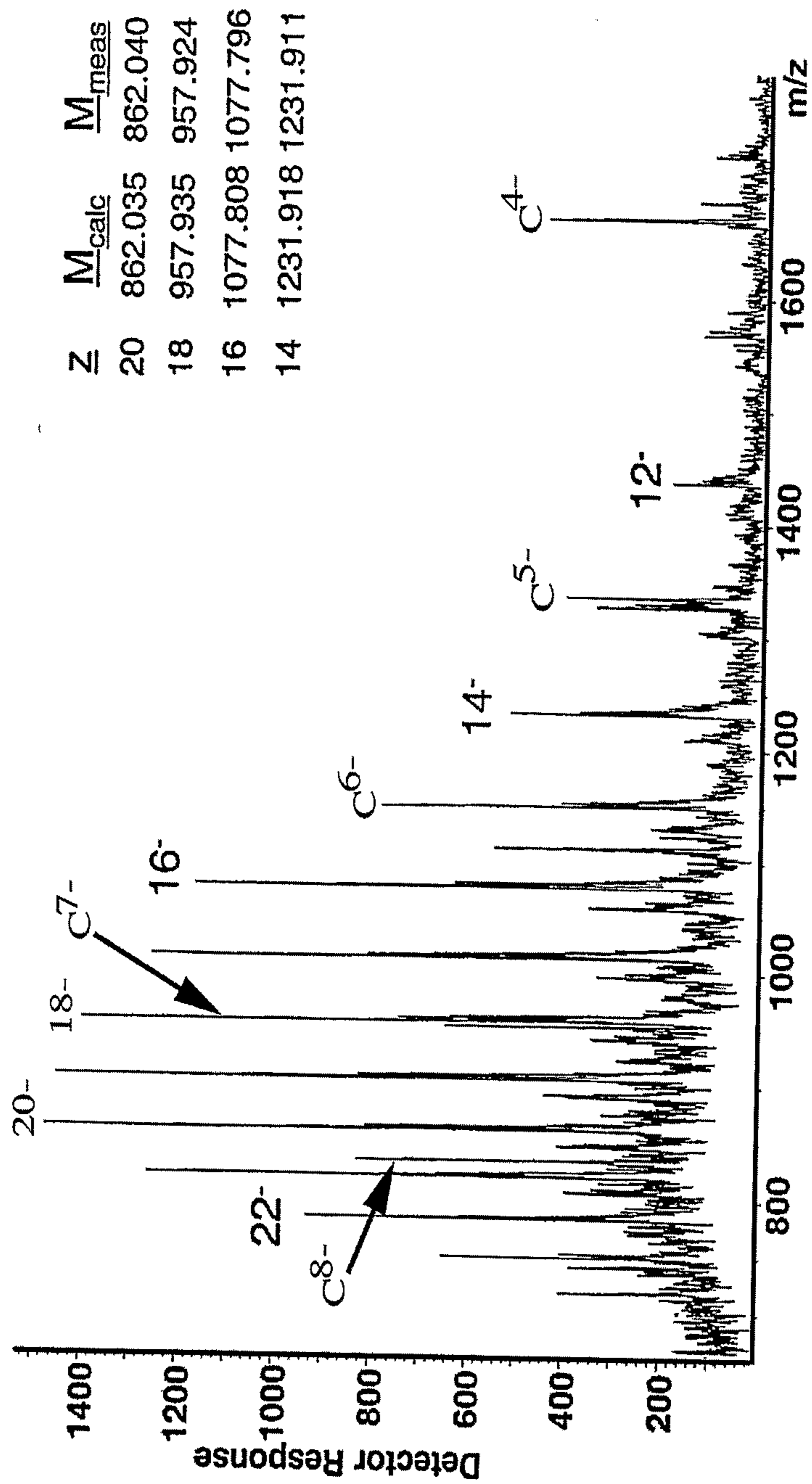


FIG. 7



**FIG. 8**

ESI-TOF MS of sspE 56mer + Calibrant



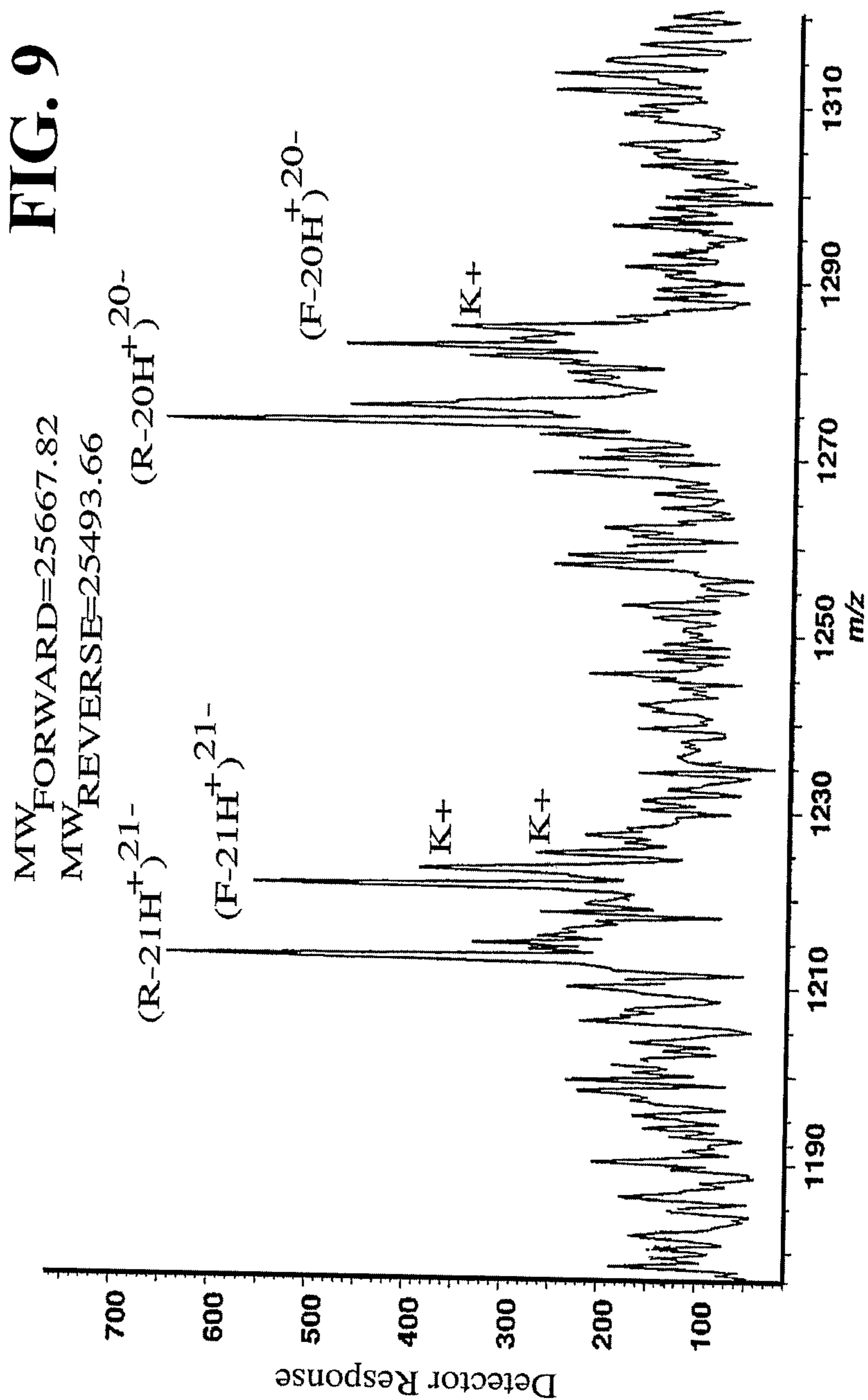


FIG. 10

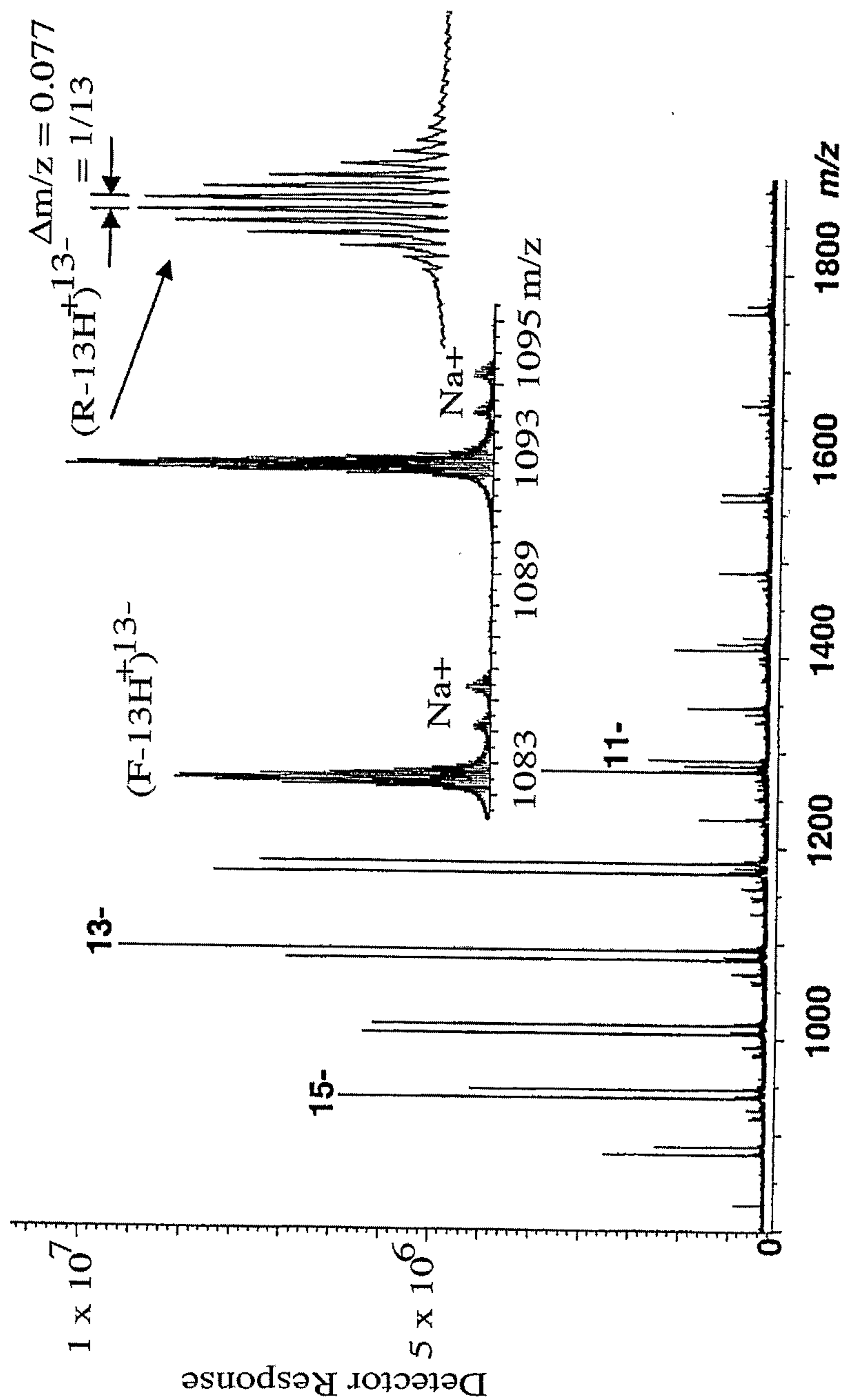
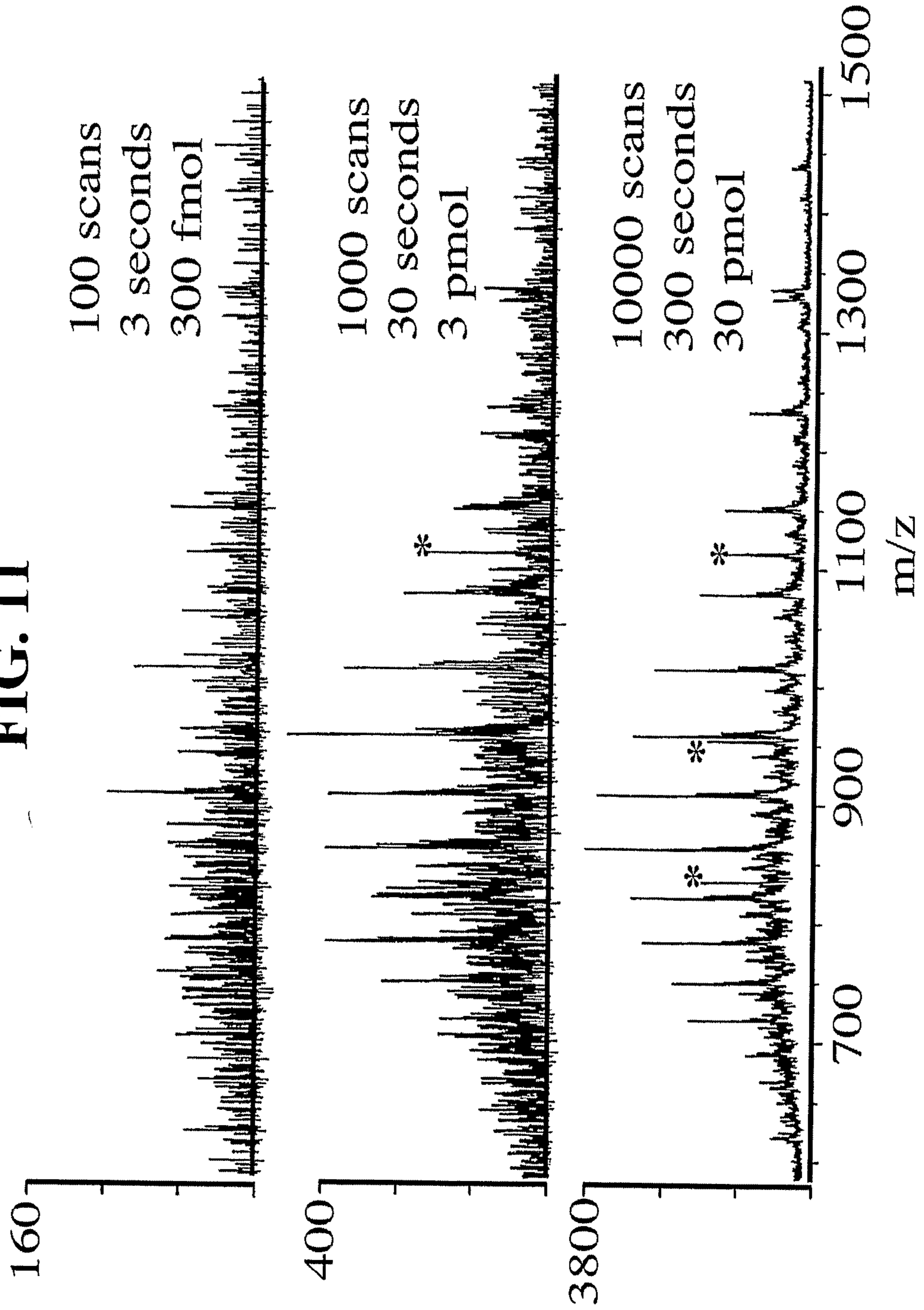
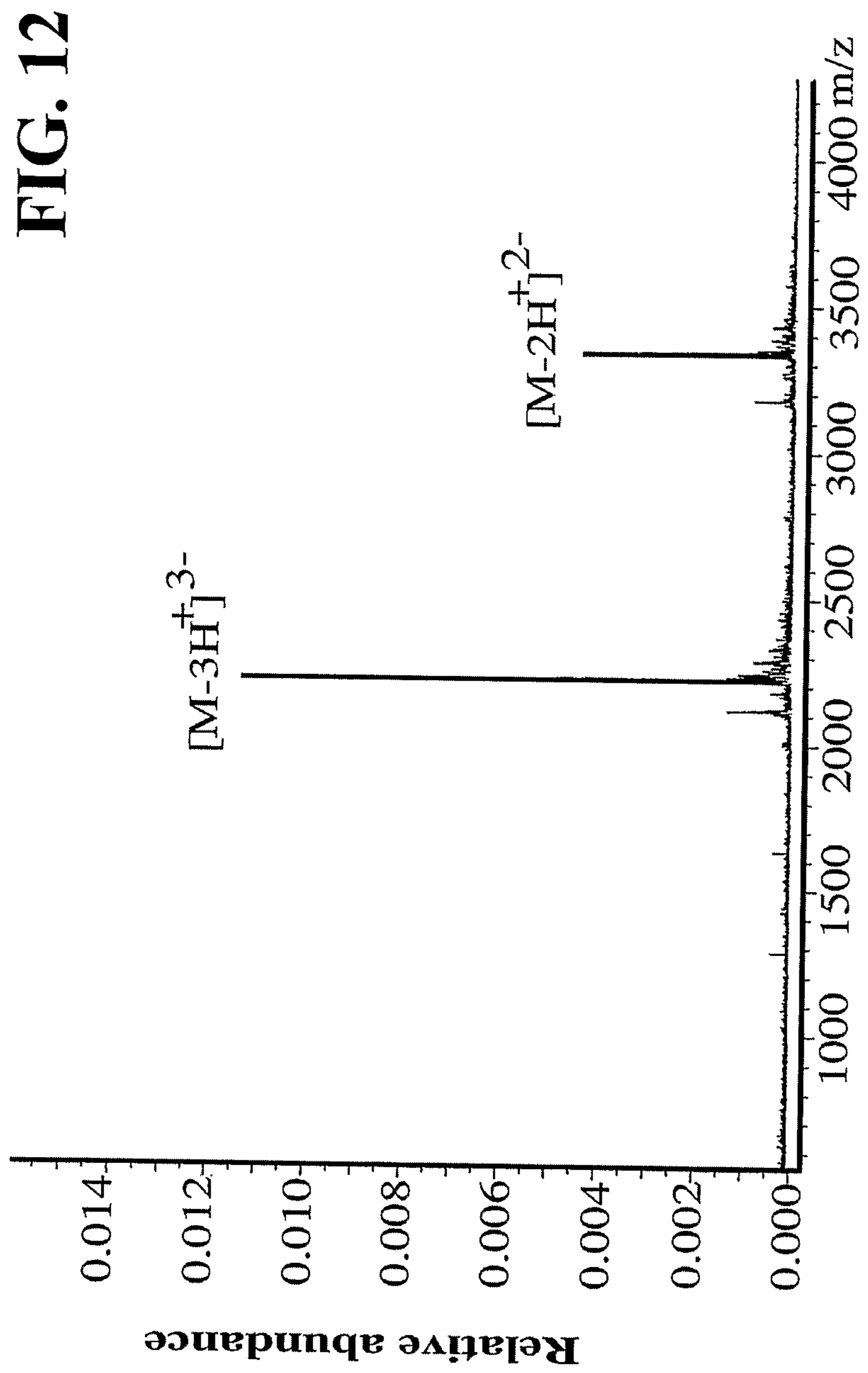


FIG. 11

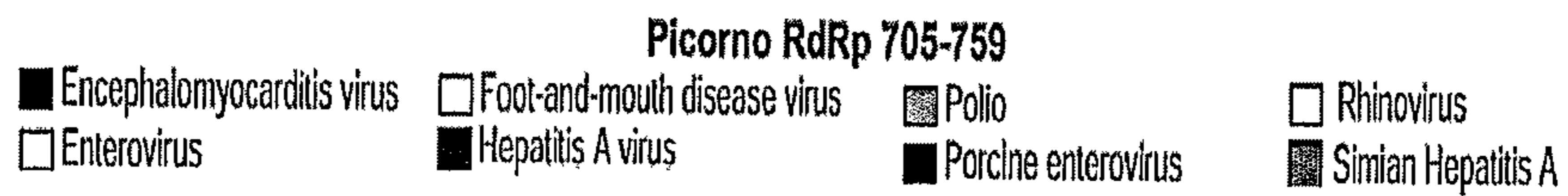
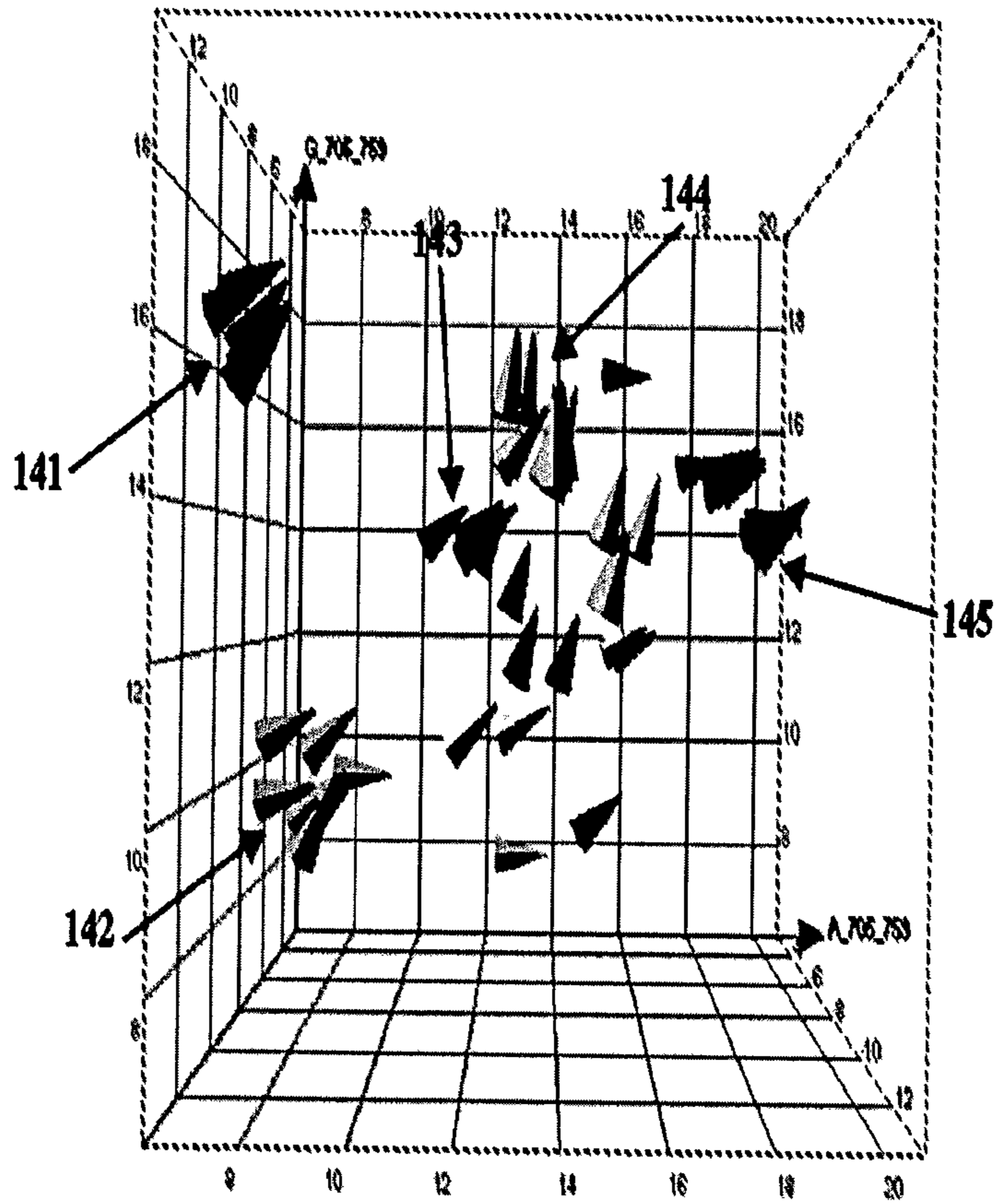




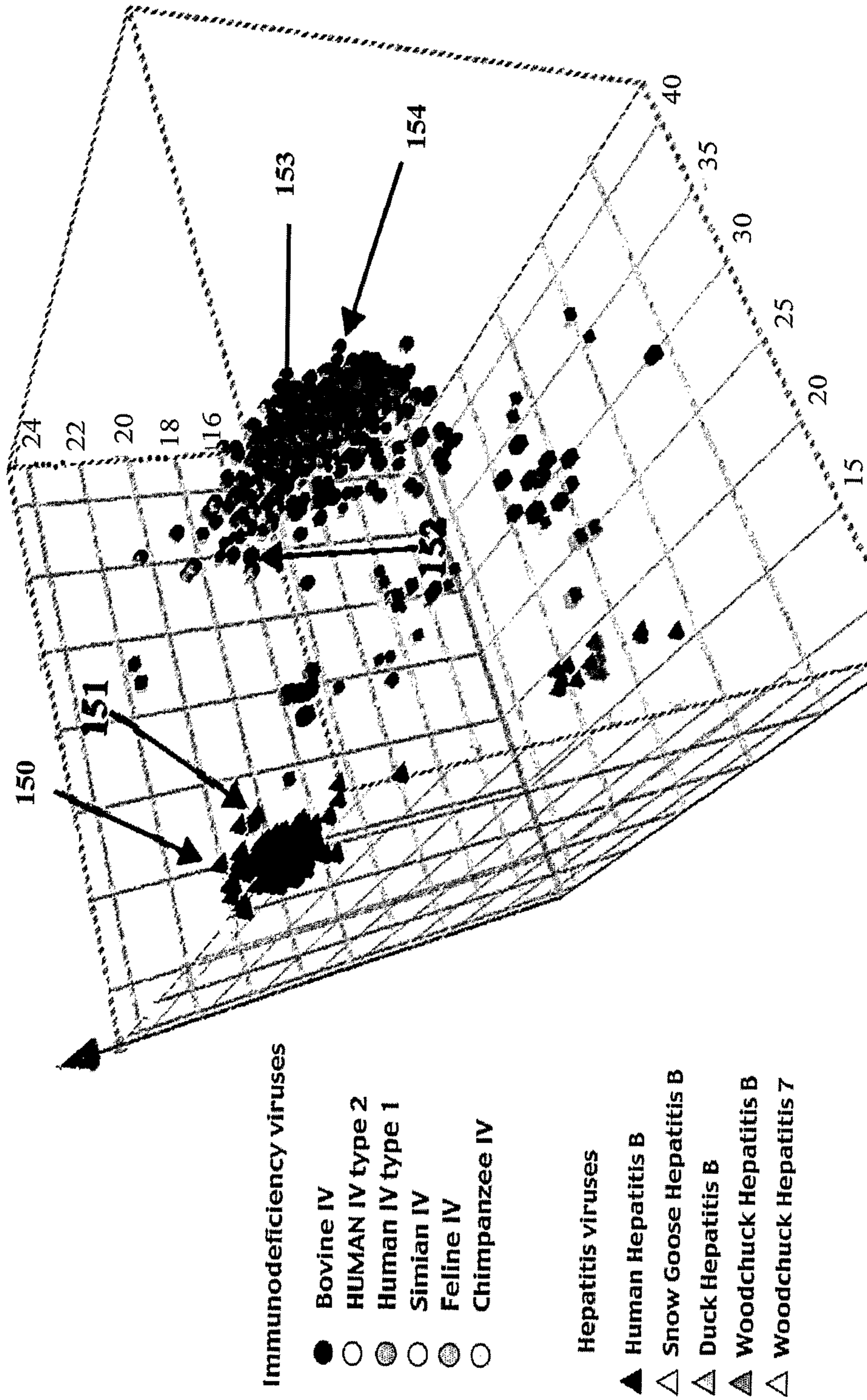




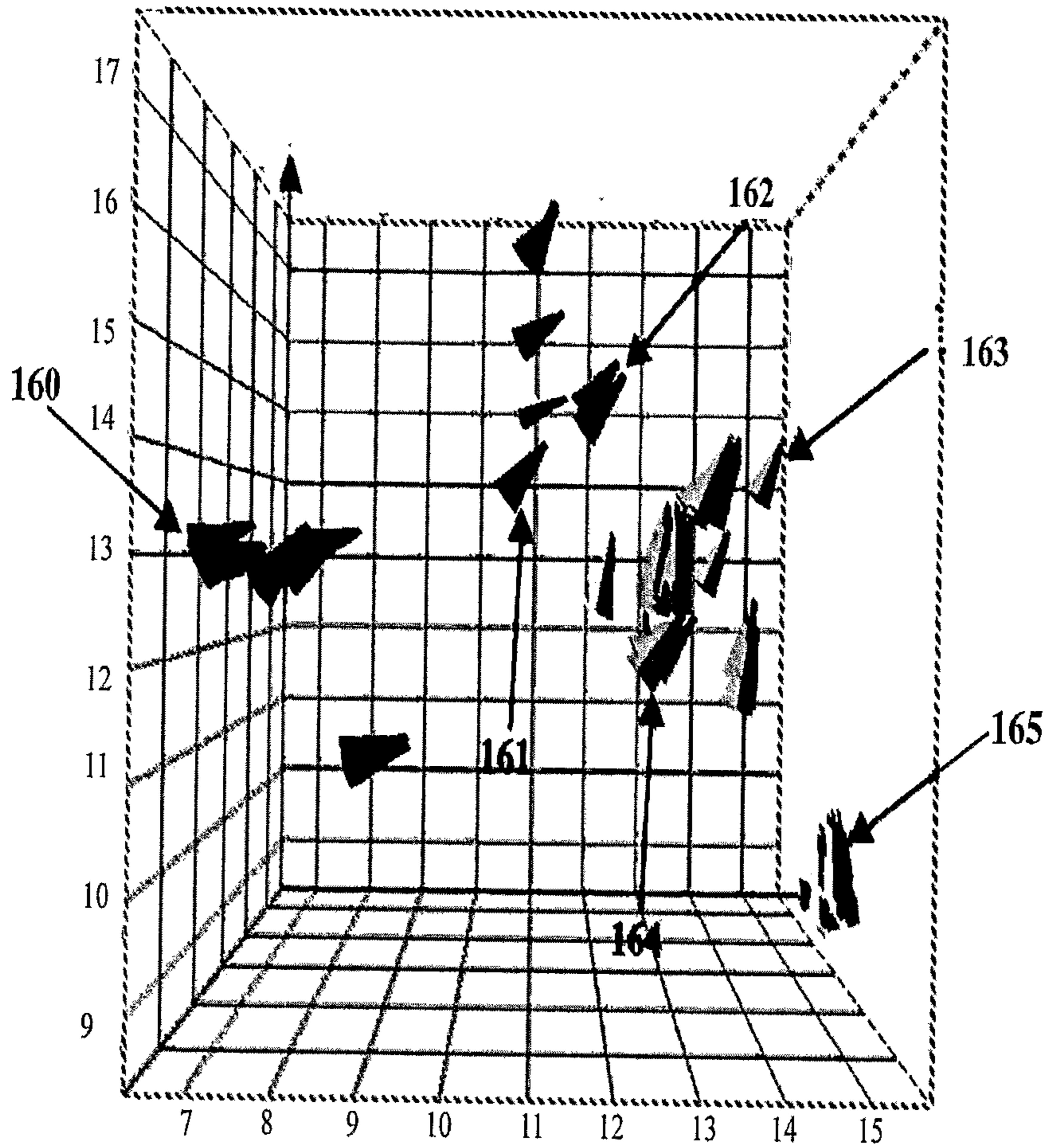
**FIG. 14**



**FIG. 15**



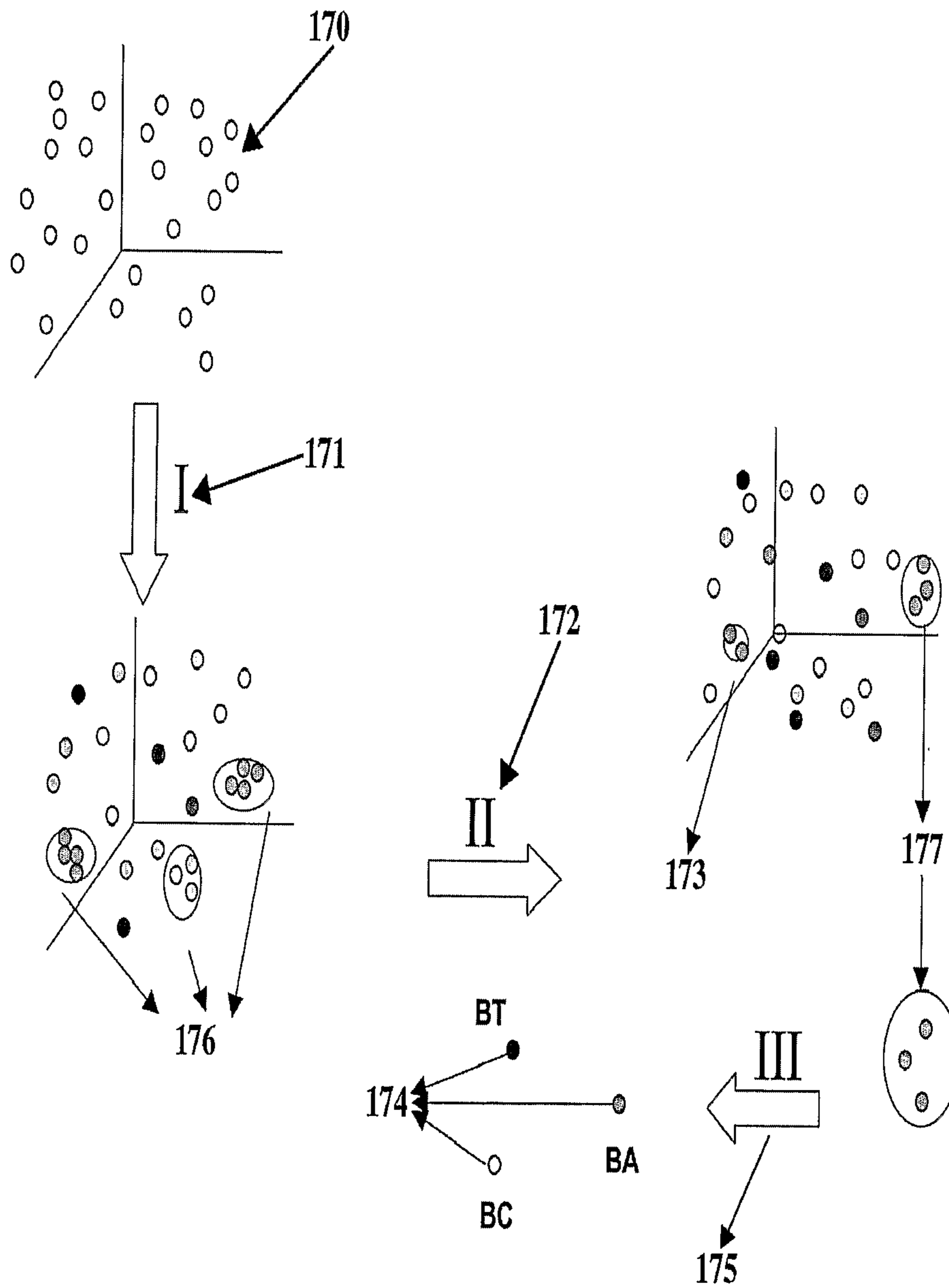
**FIG. 16**



**Flavi RdRp 2453-2493**

- |                     |                                    |                                 |
|---------------------|------------------------------------|---------------------------------|
| ■ Dengue virus type | □ Japanese encephalitis virus      | ■ Tick-borne encephalitis virus |
| □ Dengue virus type | ■ Kunjin virus                     | ■ West Nile virus               |
| ■ Dengue virus type | □ Murray valley encephalitis virus | □ Yellow fever virus            |

FIG. 17





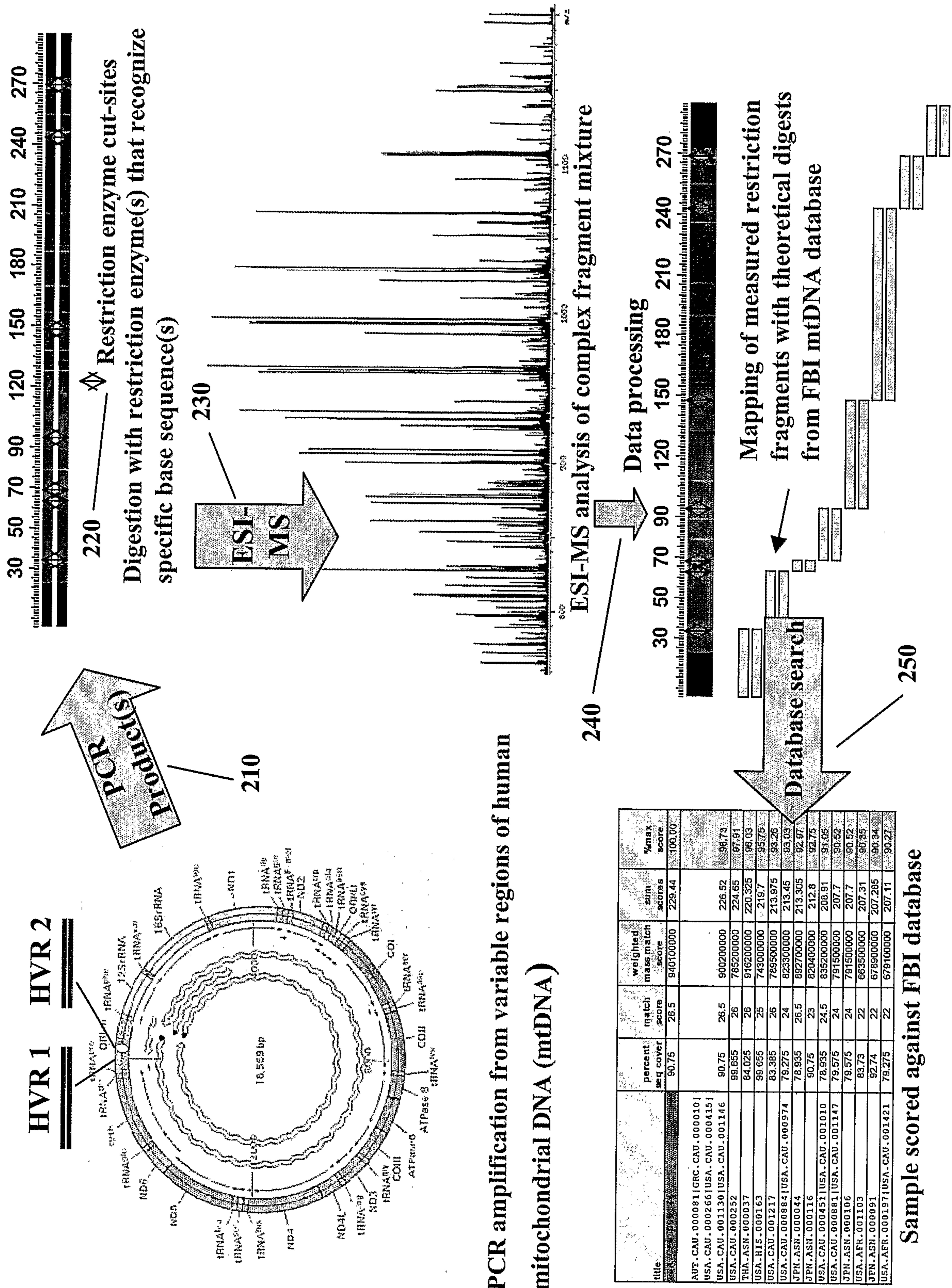
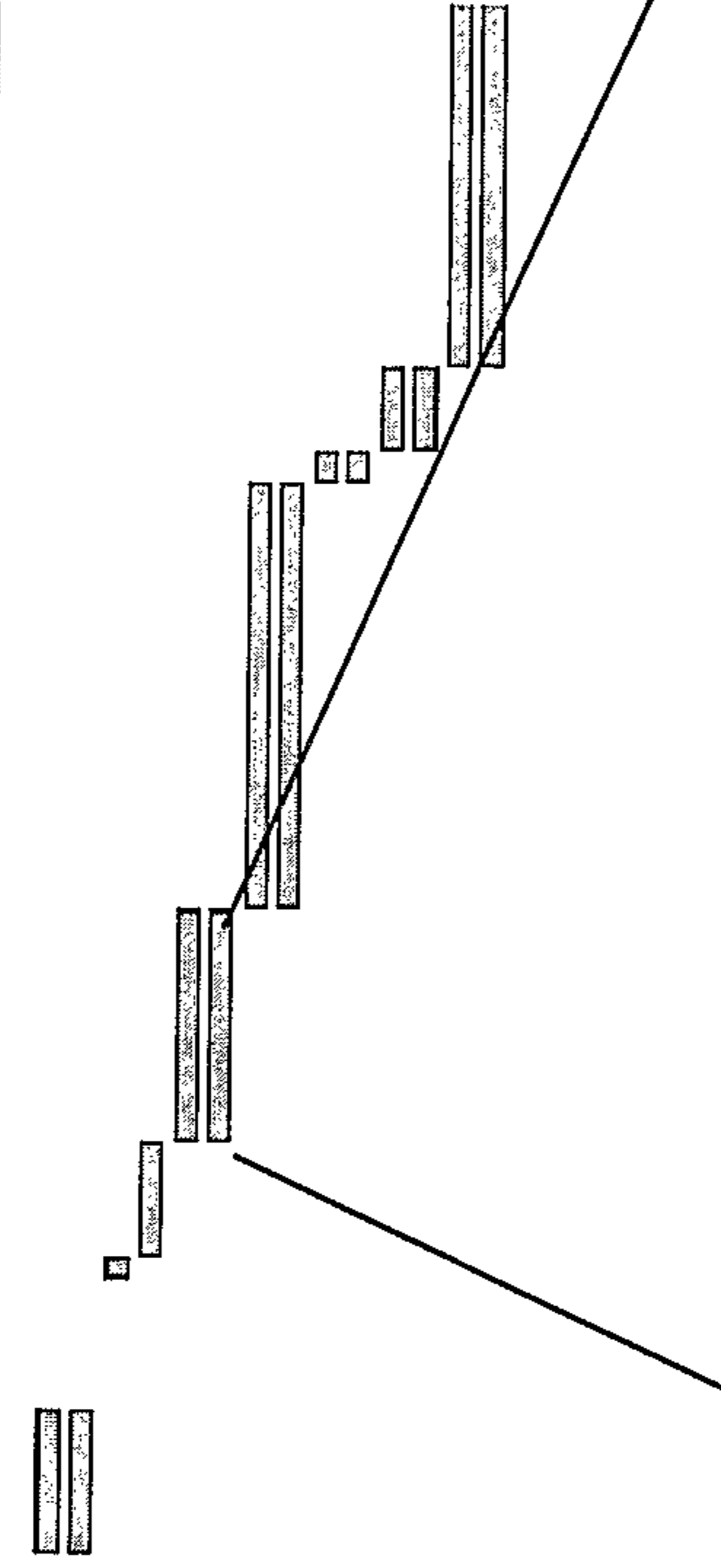
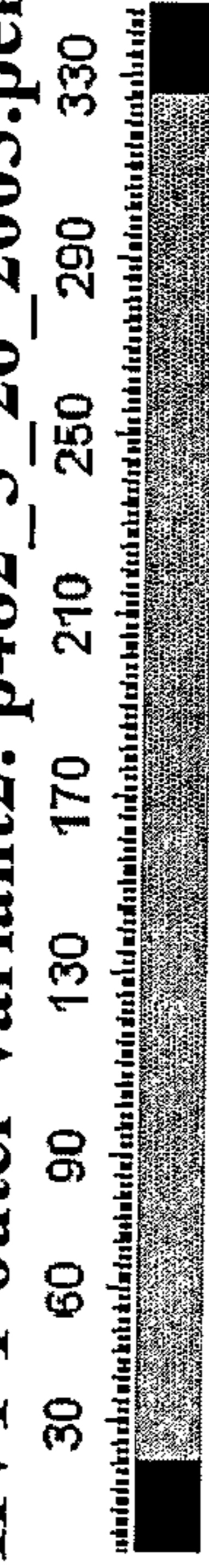


Figure 19

# Sample 1: Heteroplasmy (mixture of lengths) is detrimental to sequencing approaches but does not adversely affect MS measurements

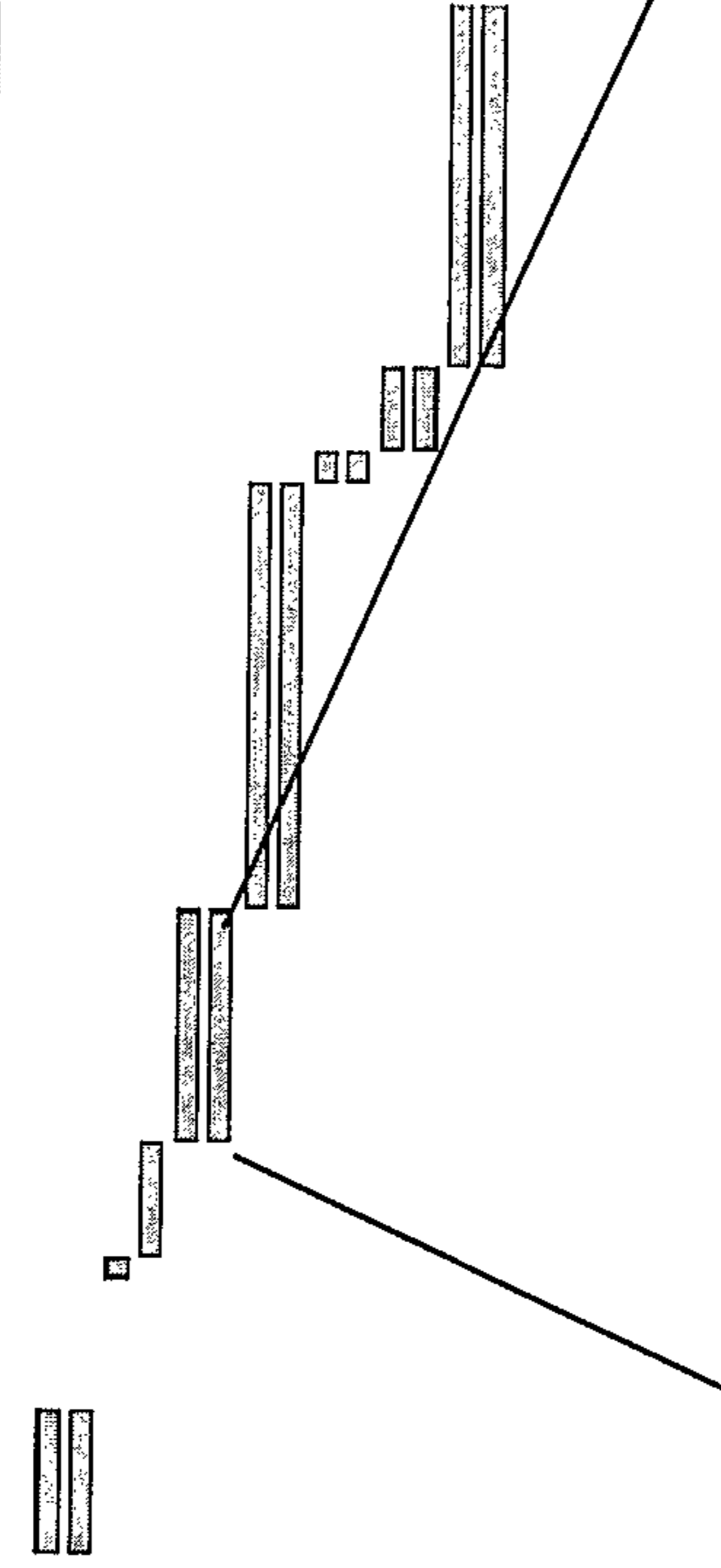
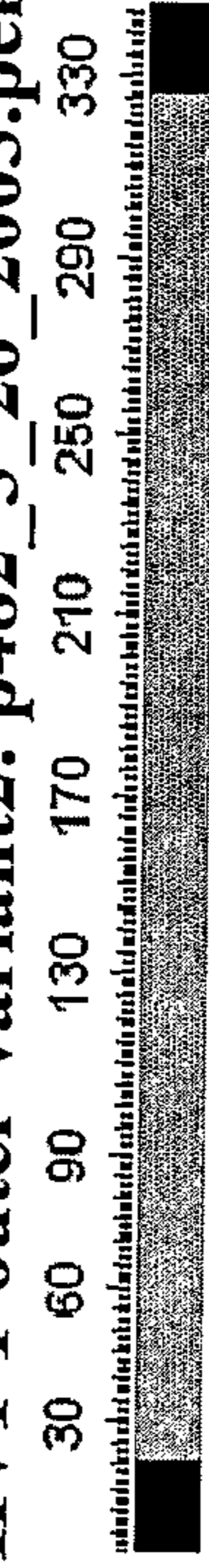
HV1-1-outer-variant1: p482\_3\_26\_2003.pek, scan HV1-1-outer-variant2: p482\_3\_26\_2003.pek, scan 1



Scan number: 1  
 Position: 94 to 145  
 Length: 52  
 Number of matched peaks: 1  
 Fragment seq: ACATAAAAACCCAAATCCACATCAAAACCCCAATCCATGCTTACAAAGCAAGT  
 Average Predicted fragment mass: 15753.66266 ± 0.00000 daltons  
 Average matched peak mass: 15753.63000 ± 0.00000 daltons  
 Total peak abundance: 4.01E+06  
 Average peak abundance: 4.01E+06 ± 0.00E+00  
 Maximum peak abundance: 4.01E+06  
 Weighted average match error: 2.07327 ± 0.00000 ppm  
 Raw average match error: 2.07327 ± 0.00000 ppm  
 Weighted average peak fit: 0.05900 ± 0.00000  
 No ambiguous assignments in group

(opposite strand also observed)

acataaaaaaaaatccacatcaaaaaaaaaaacatgcttacaagcaagt  
 acataaaaaaaaaatccacatcaaaaaaaaaaacatgcttacaagcaagt



Scan number: 1  
 Position: 94 to 146  
 Length: 53  
 Number of matched peaks: 1  
 Fragment seq: ACATAAAACCCAAATCCACATCAAAACCCCAATCCATGCTTACAAAGCAAGT  
 Average Predicted fragment mass: 16042.70903 ± 0.00000 daltons  
 Average matched peak mass: 16042.68860 ± 0.00000 daltons  
 Total peak abundance: 9.97E+06  
 Average peak abundance: 9.97E+06 ± 0.00E+00  
 Maximum peak abundance: 9.97E+06  
 Weighted average match error: 1.26121 ± 0.00000 ppm  
 Raw average match error: 1.26121 ± 0.00000 ppm  
 Weighted average peak fit: 0.02100 ± 0.00000  
 No ambiguous assignments in group

(opposite strand also observed)

acataaaaaaaaatccacatcaaaaaaaaaaacatgcttacaagcaagt  
 acataaaaaaaaaatccacatcaaaaaaaaaaacatgcttacaagcaagt

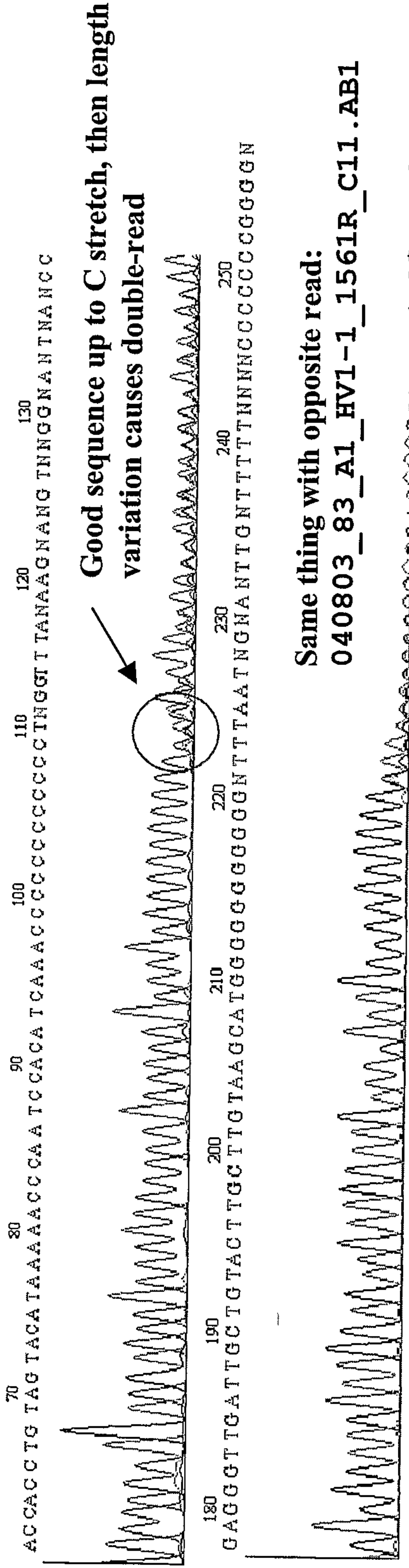
Heteroplasmic region

Figure 20A

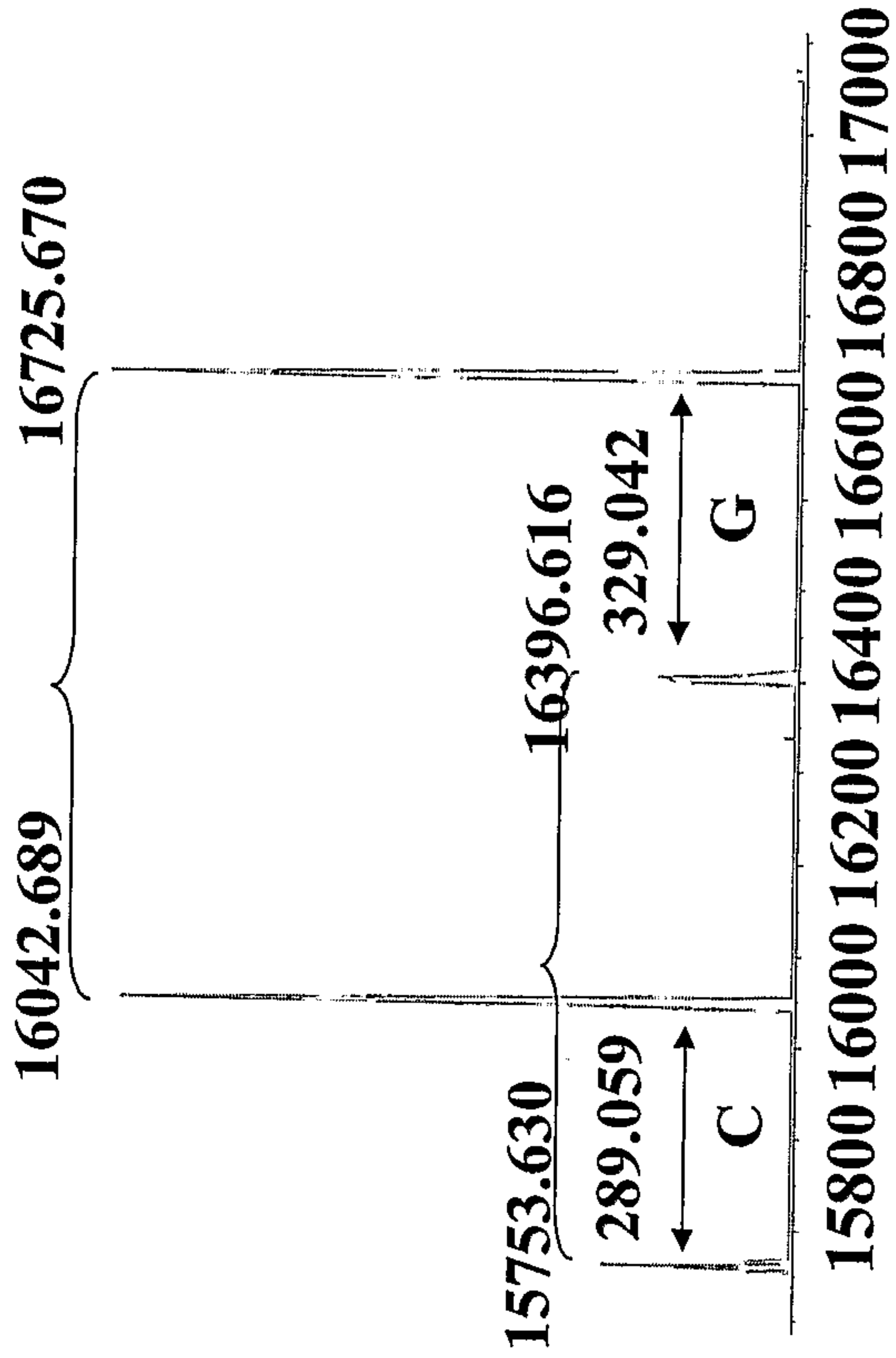
Figure 20B

**MS Approach Succeeds Where Conventional Sequencing Fails**

Sequences differ in length by one base in the 'C' stretch. Both variants appear in digest data. Sequencing profile also points to a length variation in this region: Example trace: 040803\_81\_A1\_HV1-1\_1560F\_A11.AB1



Same thing with opposite read:  
040803\_83\_A1\_HV1-1\_1561R\_C11.AB1



MS detects multiple species simultaneously while sequencing requires pure sample for maximum information content

Ratio of short to long alleles is 1:3