OPIC
Office de la propriété intellectuelle du Canada

CIPO
Canadian Intellectual Property Office

(12) (19) (CA) **Demande-Application**

(21) (A1) **2,302,433**
(86) 1998/08/28
(87) 1999/03/04

(72) VITEK, P. MICHAEL, US
(72) MITSUDA, NORIAKI, US
(72) ROSES, ALLEN D., US
(71) DUKE UNIVERSITY, US

(51) Int.Cl.[7] C07H 21/04, C12N 15/00, C12N 5/00
(30) 1997/08/29 (08/920,422) US

(54) **PROMOTEUR DU GENE DE PRESENILINE-1**

(54) **PRESENILIN-1 GENE PROMOTER**

(57) Cette invention concerne une molécule d'ADN isolée qui dirige la transcription à spécificité neuronale d'un segment d'ADN hétérologue situé en aval (un segment promoteur) dans une cellule de mamifère. La molécule d'ADN isolée contient un segment de promoteur provenant de l'ADN de préséniline-1 génomique de souris, ou une séquence d'ADN qui s'hybride à un tel ADN et dirige la transcription à spécificité neuronale d'un segment d'ADN hétérologue situé en aval dans une cellule de mamifère. Cette invention concerne également des structures d'ADN comprenant de tels segments de promoteur ainsi que les diverses utilisations de ces dernières.

(57) An isolated DNA molecule that directs neuron-specific transcription of a downstream heterologous DNA segment (a promoter segment) in a mammalian cell is disclosed. The isolated DNA molecule contains a promoter segment from the mouse genomic Presenilin-1 DNA, or a DNA sequence that hybridize to such DNA and directs neuron-specific transcription of a downstream heterologous DNA segment in a mammalian cell. DNA constructs comprising such promoter segments and various uses thereof are also disclosed.

CORRECTED VERSION*

CA 02302433 2000-02-28

05-01-99 A09:30 IN

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C07H 21/04, C12N 5/00, 15/00

A1

(11) International Publication Number: WO 99/10368

(43) International Publication Date: 4 March 1999 (04.03.99)

(21) International Application Number: PCT/US98/17905

(22) International Filing Date: 28 August 1998 (28.08.98)

(30) Priority Data:
08/920,422 29 August 1997 (29.08.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 08/920,422 (CON)
Filed on 29 August 1997 (29.08.97)

(71) Applicant *(for all designated States except US)*: DUKE UNIVERSITY [US/US]; 230 North Building, Research Drive, Box 90083, Durham, NC 27708–0083 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: VITEK, P., Michael [US/US]; 205 Park Knoll Lane, Apex, NC 27502 (US). MITSUDA, Noriaki [JP/US]; 1411 Snowcrest Trail, Durham, NC 27707 (US). ROSES, Allen, D. [US/US]; 10 Streanley Court, Durham, NC 27707 (US).

(74) Agents: SIBLEY, Kenneth, D. et al.; Myers, Bigel, Sibley, & Sajovec, P.A., P.O. Box 37428, Raleigh, NC 27627 (US).

(81) Designated States: AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model), GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, 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 the receipt of amendments.

(54) Title: PRESENILIN–1 GENE PROMOTER

(57) Abstract

An isolated DNA molecule that directs neuron–specific transcription of a downstream heterologous DNA segment (a promoter segment) in a mammalian cell is disclosed. The isolated DNA molecule contains a promoter segment from the mouse genomic Presenilin–1 DNA, or a DNA sequence that hybridize to such DNA and directs neuron–specific transcription of a downstream heterologous DNA segment in a mammalian cell. DNA constructs comprising such promoter segments and various uses thereof are also disclosed.

*(Referred to in PCT Gazette No. 20/1999, Section II)

PRESENILIN-1 GENE PROMOTER

This invention was made with Government support under Grant RO1 AG-13839 from the National Institutes of health. The Government has certain rights to this invention.

Field of the Invention

This invention relates to genomic Presinilin-1 gene constructs, and the regulatory regions thereof, and recombinant DNA constructs employing the same.

Background of the Invention

Alzheimer's disease (AD) is a devastating neurological disorder and the most common cause of dementia. The genetics of this disorder suggest that multiple genes are involved. To date, mutations in four genes have been found to be associated with Alzheimer's disease phenotypes including the Amyloid Precursor Protein (APP) gene on chromosome 21 (Citron, M. et al., *Nature* **360**, 672-674 (1992); Suzuki, N. et al., *Science* **264**, 1336-1340 (1994)), the Apolipoprotein-E (APOE) gene on chromosome 19 (Corder, E. H. et al., *Science* **261**, 921-923 (1993); Corder, E. H. et al., *Nat. Genet.* 7, 180-184 (1994); Strittmatter, W. J. et al., *Proc. Natl. Acad. Sci. U. S. A.* **90**, 1977-81 (1993)), the Presenilin-1 (PS-1) gene on chromosome 14 (Sherrington, R. et al., *Nature* **375**, 754-760 (1995)) and the Presenilin-2 (PS-2) gene on chromosome 1 (Levy-Lahad, E.et al., *Science* **269**, 973-977 (1995)). An unknown gene on chromosome 12 appears to associate with a large percentage of late-onset AD patients (Stephanson, J. *J. Am. Med. Asoc.* **277**, 775 (1997)). The majority of familial Alzheimer's disease cases are associated with mutations in the PS-1. To date, over 30 independent mutations in the PS-1 gene have been described in unrelated Alzheimer's

families displaying an early-age-of-onset phenotype. Most of these mutations are missense mutations that result in single amino acid changes (Wasco, W. et al., *Nat. med.* **1**, 848 (1995); Alzheimer's Disease Collaborative Group, *Nat. Genet.* **11**, 219-222 (1995); Campion, D. et al., *Hum. Mol. Genet.* **4**, 2373-2377 (1995); Cruts, M.et al., *Hum. Mol. Genet.* **4**, 2363-2371 (1995); Boteva, K. et al., *Lancet* **347**, 130-131 (1996); Rossor, M. et al., *Lancet* **347**, 1560 (1996); Kamino, K. et al., *Neurosci. Let.* **208**, 195-198 (1996)).

Deletions found in Exons 4 and 9 cause additional mutations as do several truncations of the RNA transcripts arising by differential splicing (Perez-Tur, J. et al., *Neuroreport.* **7**, 297-301 (1995).). Although clustering of these mutations within the protein suggests the location of functionally important domains, the exact function of Presenilin proteins is a matter of active investigation.

One approach to find gene function is to study the regulation of PS-1 gene expression. Using *in situ* hybridization, we and others demonstrate that PS-1 mRNA is most highly expressed in neurons of the brain (Koracs, D. M. et al., *Nat. Med.* **2**, 224-229 (1996)). Immunohistochemistry revealed that the PS-1 protein was abundant in neurons, but was also associated with amyloid plaques and some glial cell types (Scheuner, D. et al., *Nat. Med.* **2**, 864-870 (1996); Lah, J. et al., *J. Neurosci.* **17**, 1971-1980 (1997)). In contrast, Sherrington et al. reported that PS-1 mRNA is widely expressed in a variety of organs throughout the body (*Nature* **375**, 754-760 (1995)). This raises the question as to why mutations in the PS-1 gene product appear to confer a disease state in familial Alzheimer's patients without apparent effect on their peripheral organs. The situation is further compounded because PS-1 mRNA and protein levels from FAD patients and age-matched healthy controls have not been reported, leaving open the possibility that aberrant regulation of PS-1 gene expression further contributes to the disease state.

Mutations in the PS-1 gene's promoter and non-protein encoding regions are not known and reports on the gene's wild-type sequence are lacking. Similarly, no functional analysis of the gene's ability to promote transcription have been reported. Combined with recent reports that PS-1 knockout mice are embryonic lethal (Shen, J. et al., *Cell* **89**, 629-639 (1997)), knowledge of the PS-1 gene sequence and its transcriptional regulation should be important clues that help to identify PS-1 function in both normal and diseased states.

Summary of the Invention

We herein described a complete sequence of the mouse Presenilin-1 gene. This sequence has shown that there are two independent transcription start sites. Functional testing of the DNA regions surrounding these start sites showed that they both were apparently controlled by a single, major promoter that includes the +1 position of Exon 1A. This promoter was also quite interesting because it is mostly active in neuron-like cells. Further characterization can now progress to a complete description of those positive and negative DNA elements and transcription factors which function to control Presenilin-1 gene expression.

A first aspect of the present invention is, accordingly, an isolated DNA molecule that directs neuron-specific transcription of a downstream heterologous DNA segment in a mammalian cell, said isolated DNA molecule having a sequence selected from the group consisting of:

(a) the sequence spanning position –327 to position –206 of the mouse genomic Presenilin-1 gene (with position +1 indicating the transcription start site of exon 1A);

(b) the sequence spanning position –449 to position +1171 of the mouse genomic Presenilin-1 gene;

(c) the sequence spanning position -9 to position +16, of the mouse genomic Presenilin-1 gene (the sequence AGGCCGGAAGTTGCGACACCGGTGA **(SEQ ID NO:1)**); and

(d) DNA sequences that hybridize to isolated DNA having a sequence of (a), (b), or (c) above, and which direct neuron specific transcription of a downstream heterologous DNA segment in a mammalian cell.

By "neuron specific" is meant any level of specificity, so long as the downstream heterologous DNA is preferentially transcribed or expressed in a neuron. By "neuron" is meant an excitable cell that sends a signal, as is well known in the art (*e.g.*, neurons of the brain cortex).

A second aspect of the present invention is a DNA construct comprising an expression cassette, which contains, in the 5' to 3' direction, a promoter segment consisting

of a DNA sequence as given above, and a heterologous DNA segment positioned downstream from said promoter segment and operatively associated therewith.

A third aspect of the present invention is a nerve cell containing a DNA construct as described above.

A fourth aspect of the present invention is a method of making a transgenic non-human animal. The method comprises transforming an animal cell with an expression cassette as described above, and then regenerating an animal from the transformed animal cell.

A fifth aspect of the present invention is a transgenic non-human animal, wherein some or all of the cells of the animal containing a heterologous expression cassette as described above.

The foregoing and other objects and aspects of the present invention are explained in detail in the drawings herein and the specification set forth below.

Brief Description of the Drawings

Figure 1. ***Structure of Three Different Presenilin-1 Transcripts from Mouse Brain.*** The cloned products of 5'-RACE of mouse brain cDNA and DNA sequencing revealed the presence of three independent transcripts (A, B and C) which appear to derive from two unique transcription start sites marked by the vertical arrows. The distance between two transcription start sites is 410 Bp. The sizes of Exon 1A, 1B and 1C are 141 Bp, 371 Bp and 139 Bp, respectively.

Figure 2. ***Cloning and Sequencing Strategy Elucidates the Mouse Presenilin-1 Gene's Exon-Intron Structure.***

Figure 2A. ***Screening Strategy:*** "Screening-A" utilized a fragment of the mouse PS-1 cDNA as Probe A (filled box) to identify lambda phage clones of the mouse PS-1 genomic DNA (represented as double lines). "Screening-B" utilized PCR primers to identify a P1 clone of the mouse PS-1 gene, P1-10809, as represented by the hatched horizontal box.

Figure 2B. ***Sequencing Strategy:*** Lambda phage clones and P1-10809 were restricted and subcloned into pBluescript II KS(+) vector. Thick lines correspond to

individual plasmid subclones from corresponding regions of PS-1 genomic DNA found in P1-10809. Double arrows represent PCR products from P1-10809 template which were sequenced directly. Restriction endonucleases abbreviated as: H, Hind III; E, EcoR I; N, Not I; X, Xho I.

Figure 2C. ***Exon-Intron Structure of the Mouse PS-1 Gene:*** Exons are boxed and double lines represent Introns. Filled boxes and open boxes correspond to the protein coding and untranslated regions, respectively. The translation start codon ATG begins at position +11,420, the translation termination codon TAG is at +45,627 and the putative polyadenylation signal (AATTAA) is at position +46,612.

Figure 3. ***Comparison of the Mouse and Human Presenilin-1 Promoters.*** Mouse PS-1 transcription begins with "G" at position +1 of Exon 1A and human PS-1 transcription begins with "A" (Data not shown.). By DNA sequence similarity searching with BLAST network service available from National Center Biotechnology Information, regions of mouse/human homology are found around the transcription iniation sites for both genes. Consensus binding sites for the transcription factors ETS1 and SP1 are underlined and are conserved in both mouse and human genes.

Figure 4. ***Nucleotide Sequence of the Mouse PS-1 Promoter Region.*** The sequence of mouse PS-1 gene flanking the two transcription initiation sites, marked with vertical arrows, is displayed. Some restriction endonuclease sites are underlined and various promoter elements are boxed and labeled. Exon 1A and Exon 1B are double-underlined.

Figure 5. ***Mouse Presenlin-1 Promoter-Reporter Constructs and their Relative Luciferase Activity (%RLA).***

Figure 5A. ***Structural Organization of PS-1 Promoter:*** Top line represents the region of the PS-1 gene which was analyzed for promoter activity where boxes for Exon 1A and Exon 1B are labeled. Open boxes represent genomic DNA fragments corresponding to the mouse PS-1 gene (top line) which were cloned upstream of the Firefly luciferase reporter gene in the plasmid pGL3-Basic (Promega). Open boxes are labeled on the left with the name of the promoter-reporter plasmid as LUC # and with a nucleotide number of the 5' end of the fragment based on +1 being the "G" at the beginning of Exon 1A. Letters above the open box refer to a restriction enzyme cleavage site. Numbers to the immediate right of the

open box denote the 3' end of the promoter fragment. The numbers on the left hand side are the percentage of Relative Luciferase Activity (%RLA) calculated as described in Materials and Methods followed by the number of times that construct has been transfected into cells and its activity measured which is in parentheses.

Figure 5B. ***Fine Structure Map of the PS-1 Promoter and Promoter-Reporter Construct Activity Strategy.*** Top line represents the region of the PS-1 gene with putative promoter elements, Exon 1A, Exon 1B and restriction enzyme site positions labeled. Open boxes of promoter-reporter constructs are as in (A). Letters above the open box refer to the end of the promoter fragment shown in Figure 3.

Figure 6. ***Cell-Type Specific PS-1 Promoter Activity.*** PS-1 promoter-reporter constructs LUC 29, LUC 27, LUC 4, LUC 3 and LUC 1 were transiently transfected into Neuro2a-neuroblastoma (N2a), undifferentiated P19 (P19), retinoic acid differentiated neuron-like P19 (P19N), dimethylsulfoxide differentiated muscle-like P19 (P19M) and NIH/3T3-fibroblast cells. An SV40-promoter driving Firefly luciferase in pGL3-Basic plasmid (Promega) and pRL-TK (Thymidine Kinase promoter driving Renilla luciferase gene) was also transfected into each cell line as external and internal controls, respectively. After measuring luciferase activity from all combinations of plasmids, the Index of Relative Luciferase Activity (IRLA) was calculated as RLA/ RLA_{SV40}, where RLA_{SV40} is the ratio of Firefly luciferase signal in the external control divided by the Renilla luciferase signal in the internal control, as to compare the activity of different promoter fragments in different cell lines. Plasmid LUC 29 transfected into N2a cells showed the greatest IRLA value which we defined as 100% activity.

Detailed Description of the Invention

Nucleotide sequences are presented herein by single strand only, in the 5' to 3' direction, from left to right. Nucleotides are represented herein in the manner recommended by the IUPAC-IUB Biochemical Nomenclature Commission.

Specific examples of neuron specific promoters of the present invention include, but are not limited to, DNA molecules that comprise mouse Presenilin-1 gene promoter segments from position:

-440 to +91;

-352 to +91;

-327 to +91;

-276 to +91;

-261 to +91;

-215 to +91;

-192 to +91;

-124 to +91;

-87 to +91; and

-32 to +91.

From position:

-276 to +519;

-276 to +206;

-276 to +148; and

-276 to +41.

And from position:

-87 to +41;

-9 to +16; and

-327 to +206.

The corresponding fragments from the Presenilin-1 gene of other mammalian species, such as rat, cat, dog, monkey, or human, may also be employed to carry out the present invention., as discussed in greater detail below.

Promoters of the present invention may be of any animal species of origin, but are preferably of mammalian origin (*e.g.,* mouse, rat, cat, dog, monkey, human). Promoters used to carry out the present invention are, in general, substantially homologous to the mouse segments described above. As used herein, such regions are "substantially homologous" when their DNA binding required for promoter function are homologous to eh various mouse promoter segments described herein. In general, such regions are at least 75%, and more preferably 80%, 85%, 90%, or even 95% homologous to the various mouse promoter segments described herein. It will be apparent that other sequence fragments of the regions

flanking the foregoing sequences, or minor additions, deletions or substitutions thereto, can be prepared which will also carry out the function of a Presenilin-1 gene promoter. They may also be identified by their binding to specific transcription factor proteins and/or their function to promote transcription.

Other DNA segments such as naturally occurring DNA segments or mammalian DNA segments, which encode neuron-specific gene promoters can be identified by their binding, or hybridization, to fragments as described above. Hybridization conditions which will permit such DNA sequence to hybridize to a DNA sequence as given herein are known in the art. For example, hybridization of such sequences to the DNA disclosed herein may be carried out in 25% formamide, 5X SSC, 5X Denhardt's solution, with 100 μg/ml of single stranded DNA and 5% dextran sulfate at 42° C, with wash conditions of 25% formamide, 5X SSC, 0.1% SDS at 42° C for 15 minutes, to allow hybridization of sequences of about 60% homology. More stringent conditions are represented by a wash stringency of 0.3M NaCl, 0.03M sodium citrate, 0.1% SDS at 60° C or even 70° C using a standard hybridization assay (*See* Sambrook et al., Molecular Cloning, A Laboratory Manual (2d Ed. 1989)(Cold Spring Harbor Laboratory)).

The term "operatively associated" as used herein refers to DNA sequences contained within a single DNA molecule that are associated so that the function of one is affected by the other. Thus, a promoter is operatively associated with a gene (or other DNA of interest) when it is capable of affecting the expression of that gene (i.e., the gene is under the transcriptional control of the promoter). The promoter is said to be "upstream" from the gene (or other DNA of interest), which is in turn said to be "downstream" from the promoter.

DNA constructs, or "expression cassettes," of the present invention include, 5' to 3' in the direction of transcription, a promoter of the present invention, a heterologous DNA segment operatively associated with the promoter, and, optionally, transcriptional and translational termination regions such as a termination signal and a polyadenylation signal. All of these regulatory regions should be capable of operating in the transformed cells. The 3' termination region may be derived from the same gene as the transcriptional initiation

region or from a different gene. The expression cassette may be provided in a DNA construct that also has at least one replication system.

As used herein, the term "heterologous gene" or "heterologous DNA segment" means a gene (or DNA segment) which is used to transform a cell by genetic engineering techniques, and which may not occur naturally in the cell. Structural genes are those portions of a genes which comprise a DNA segment coding for a protein, polypeptide, or portion thereof, possibly including a ribosome binding site and/or a translational start codon, but lacking a promoter. The term can also refer to copies of a structural gene naturally found within a cell but artificially introduced. Structural genes may encode a protein not normally found in the cell type into which the gene is introduced or in combination with the promoter to which it is operationally associated. As used herein, the term heterologous DNA segment also includes DNA segments coding for non-protein products, such as ribozymes or anti-sense RNAs (*see, e.g.,* U.S. Patent No. 4,801,540).

The various fragments comprising the various constructs, expression cassettes, markers and the like, may be introduced consecutively by restriction enzyme cleavage of an appropriate replication system and insertion of the particular construct or fragment into the available site. After ligation and cloning, the DNA construct may be isolated for further manipulation. All of these techniques are amply exemplified in the literature. *See, e.g.,* Maniatis et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 1982).

Promoters and the various constructs of the present invention have a variety of different uses. Nerve cells grown in culture may be transformed with constructs of the present invention and the heterologous DNA expressed therein to produce a protein or a peptide, and the protein or peptide then collected for subsequent use (for example, the protein or peptide could encode an antigen, which is used directly in diagnostic assays, or which is injected into an animal to produce antibodies thereto, which antibodies are used in diagnostic assays). Transgenic animals may be produced with the constructs of the invention, as discussed in greater detail below. The promoters may be used in vectors that comprise gene therapy vectors (*e.g.,* viral vectors such as herpesvirus vectors and RNA viruses such as retroviruse (wherein the promoter segment and the heterologous segment exist as an RNA transcript of

the DNA which is inserted into the host cell as a DNA transcript, or in the case of retroviruses as a provirus)) where the heterologous DNA encodes a therapeutic agent (*e.g.*, ApoE2 or ApoE3; nerve growth factor, ciliary neurotrophic factor, etc.); and where preferential expression of the heterologous DNA in nerve cells is desired . Numerous other uses for promoters of the instant invention will be readily apparent to those skilled in the art.

As noted above, a method of making a transgenic animal is also an aspect of the present invention. The method can be carried out on any suitable animal subject, but is preferably carried out with non-human mammals. Murine species or rodents (*e.g.*, mice, rats) are particularly preferred.

The method comprises transforming an animal cell with an expression cassette as described above, in an animal transformation vector, and then regenerating a transgenic animal from the transformed animal cell. The transformation step may be carried out by any suitable means, as discussed in detail below, and the regeneration step may also be carried out by any suitable means, as also discussed in detail below. Where chimeric animals are produced by the process, animals in which all cells (*e.g.*, including both somatic cells and germ cells) are transformed (and in which the expression cassette described above is stably integrated into the genome of the cells) may be regenerated from chimeric animals having transformed germ cells, as is known in the art.

The production of transgenic animals can be carried out by any suitable technique, such as pronuclear microinjection, infection of embryos with retroviruses, embryonic stem cell-mediated techniques, transfer of entire chromosomal segments and gamete transfection in conjunction with *in vitro* fertilization, etc. *See generally* Charles River Laboratories, *Transgenic Animal Science: Principles and Methods* (Summer 1991).

Transgenic animals that carry an expression cassette of the invention can be produced by the genetic transformation of zygotes, as described in T. Wagner et al., U.S. Patent No. 4,873,191 (applicant intends that the disclosure of all U.S. Patent References cited herein be incorporated herein by reference).

In another technique, a pluripotent embryonic stem cell from the species to be transformed may be derived, the expression cassette inserted into the stem cell, and one or more of the stem cells inserted into an early embryo such as a blastocyst of the animal to be

transformed, and the animal raised to birth in a suitable female host (*e.g.*, M. Evans, PCT Application WO90/03432). Methods of producing transgenic animals by subjecting a mixture of DNA and the embryo to an electric discharge are described in U.S. Patent No. 5,567,607 to X. Zhao et al. Mammalian expression vectors are described in U.S. Patent No. 5,627,033 to J. Smith et al.

Animals of the present invention are useful as laboratory models for studying the function of the Presenilin-1 gene, for studying the etiology of Alzheimer's disease, and for studying the activity of various drugs and drug candidates in treating Alzheimer's disease. In such animals the endogeneous Presenilin-1 gene may be active or inactive. The endogeneous Presenilin-1 gene may be inactivated by means of a "knock-out" of the Presenilin-1 gene in accordance with known techniques, such as homologous recombination. *See, e.g.*, O. Smithes, *Nature* **317**, 320 (1985).

In the Examples set forth below, the abbreviation used are: Bp, base pairs; PEA-3, polyoma virus enhancer activator-3; PS-1, presenilin-1; 5'-RACE, rapid amplification of 5'-cDNA ends; N2a, Neuro2a cells; P19N, neuron-like differentiated P19 cells; P19M, muscle-like differentiated P19 cells; RLA, relative luciferase activity; IRLA, Index of relative luciferase activity.

EXPERIMENTAL PROCEDURES

Isolation and characterization of genomic clones -- Labeled oligonucleotides and PCR products of the mouse PS-1 cDNA were used as probes to screen mouse libraries for genomic PS-1 clones. Based on the mouse PS-1 cDNA sequence (Genbank Accession #L42177), an upstream primer of sequence

5'- CGGAGAGAGAAGGAACCAAC-3' **(SEQ ID NO: 2)**

and a downstream primer of sequence

5'-TCAGCTCTTCGTCTTCCTCCTCATC-3' **(SEQ ID NO: 3)**

were used with Quick Clone Mouse Brain cDNA (Clontech) as template to amplify a portion of the mouse PS-1 cDNA by polymerase chain reaction (PCR). Amplification reactions were performed in 100 µl volume containing 1xPCR buffer II (Perkin Elmer), $MgCl_2$ (1.5mM), dATP, dGTP, dCTP and dTTP (0.2 mM each, Perkin Elmer), DNA primers at 0.5 µM, 1 µl cDNA template (0.1 ng) and Ampli-Taq DNA polymerase (5 Units, Perkin Elmer). The reaction cycle was 95°C for 1 minute, 50°C for 1 minute and 72°C for 2 minutes for a total of 30 cycles. This PCR product was gel purified and labelled with alpha-^{32}P-dCTP and a Random Primers DNA Labelling System (Gibco). Labelled probe was used to conventially screen a mouse strain 129/SVJ genomic library in Lambda Fix-II vector (Stratagene) as described (Sambrook, J. Fritsch, E. F., and Maniatis, T, Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).). Screening identified four independent phage clones designated Ph-1, Ph-2, Ph-3 and Ph-4 (**Figure 2**). Following digestion with Not I and/or EcoRI, restriction enzyme fragments were subcloned into pBluescript-II-KS(+) phagemid vector (Stratagene) using a DNA ligation kit (Stratagene). DNA sequence was determined using an Applied BioSystems model 373A automated DNA sequencer with dye terminator chemistry and protocols recommended by the manufacturer. Additional oligonucleotide probes from the 5'-untranslated region of the mouse PS-1 cDNA were labeled with ^{32}P-ATP and T4-polynucleotide kinase and used to

identify plasmid subclones by hybridization. Based on the partial sequence of phage clone Ph-2, the PCR primers 1C-US-for

GATCACAGTCTAGGTTGCTGGTGTG **(SEQ ID NO: 4)**

and 1C-US-rev

TGGGGCAAGGG ACACAAATAAG **(SEQ ID NO: 5)**

were used to further screen a mouse ES-129/SVJ genomic library in a P1 vector (Genome Systems Inc.) by PCR. Of the three P1 clones identified, P1-10809 was digested with EcoRI or Hind III and these restriction enzyme fragments subcloned and sequenced as described above.

5'-Rapid Amplification of cDNA Ends or RACE: -- The 5' end of PS-1 cDNA was identified using mouse-brain Marathon-Ready cDNA (male BALB/c, 9-11 weeks of age, Clontech). Briefly, a 50 µl PCR reaction containing 0.2 µM of a PS-1-specific reverse primer:

TGGCTCAGGGTTGTCAAGTC **(SEQ ID NO: 6)**

0.2 µM of the Clontech AP1 adaptor primer:

CCATCCTAATACGACTCACTATAGGGC **(SEQ ID NO: 7)**

2.5 ng Marathon-Ready cDNA, 1xPCR buffer (Gibco), $MgCl_2$ (1.5 mM), DMSO (5%), dATP, dGTP, dCTP and dTTP (0.2 mM each), and Taq DNA polymerase (5 Units, Gibco) was used with a reaction cycle of 95°C for 45 seconds, 55°C for 30 seconds and 72°C for 90 seconds for a total of 30 cycles in the first amplification step. The 100 µl second PCR amplification step contained 0.5 µM of a mouse PS-1-specific reverse primer, 151-130-reverse:

CAAACCTCTTGGGATTCTTTC **(SEQ ID NO: 8)**

and 0.5 μM of the nested Clontech adaptor primer AP2

ACTCACTATAGGGCTCGAGCGGC **(SEQ ID NO: 9)**

and 0.01 μl of the first PCR amplification, 1xPCR buffer (Gibco), $MgCl_2$ (1.5mM), DMSO (5%), dATP, dGTP, dCTP and dTTP (0.2 mM each) and Taq DNA polymerase (5 Units, Gibco) with the same cycling parameters as in the first amplification step. In some second PCR reactions, the PS-1 specific reverse primer 101-80-reverse:

AAGACCTCGAAGGGCTGCTGTC **(SEQ ID NO: 10)**

was used. RACE amplification products were electrophoresed on 2% agarose gels run in TAE, visualized with ethidium bromide and ultraviolet light, extracted from the gel matrix with a Wizard PCR Preps DNA Purification System (Promega), ligated into a pGEM-T vector (Promega) and transformed into competent DH5-alpha bacterial cells (Gibco). Ampicillin resistant colonies were characterized by restriction enzyme digestion, PCR amplification with a variety of primer combinations and DNA sequencing as above.

Computation of Sequence Similarities:---Comparison of the mouse PS-1 promoter with other eukaryotic promoter sequences were performed using the BLAST network service and the Eukaryotic Promoter database (EPD) Release 45 available from the National Center for Biotechnology Information.

Construction of PS-1 Promoter-Firefly Luciferase Reporters: -- Mouse genomic DNA fragments containing portions of the putative PS-1 promoter were subcloned into the promoterless pGL3-basic vector (Promega) upstream of the firefly luciferase gene. Based on the sequence of genomic DNA, PCR primers were designed to incorporate Xho I sites into the forward primers and Hind III sites into the reverse primers. These primers corresponding to different locations in the genomic DNA (see **Figure 2**) were used to PCR amplify

fragments as above which were purified with a Wizard Purification System (Promega), digested with the appropriate restriction enzymes and repurified with the Wizard kit. Cleaved PCR products were ligated into pGL3 plasmid cleaved with the same restriction enzymes, transformed into competent bacteria and clones containing plasmids with inserts verified by DNA sequencing.

Eucaryotic Cell Culture and Transfection:---Mouse Neuro2a-neuroblastoma cells, mouse P19-embryonal carcinoma and mouse NIH/3T3 fibroblast cells were obtained from the American Type Culture Collection (ATCC). Neuro2a cells were routinely propagated in Minimal Essential Medium with Earle's salt (Gibco) plus 10% fetal calf serum (Hyclone) plus 0.1 mM non-essential amono acids (Gibco). P19 cells were routinely propagated in alpha-MEM (Gibco) plus 2.5% fetal calf serum plus 7.5% bovine serum (Hyclone). Differentiation of P19 cells to neuron-like cells followed treatment with 0.5 μM trans-retinoic acid (Jones-Villeneuve, E. et al., *Mol. Cell. Biol.* 3, 2271-2279 (1983)). Differentiation of P19 cells to muscle-like cells followed treatment with 1% dimethylsulfoxide (Edwards, M. et al., *Mol. Cell. Biol.* 3, 2280-2286 (1983).). NIH/3T3 cells were routinely propagated in DMEM (Gibco) plus 10% fetal calf serum.

For transient transfection, Neuro2a, P19, retinoic acid treated-P19, DMSO-treated P19 and NIH/3T3 cells were plated in 6 well tissue culture dishes at 9×10^4 cells per well and allowed to recover for one day. Cells containing PS-1-promoter-reporter constructs were then co-transfected with 0.3 pmole of one of the Promoter-Firefly Luciferase plasmid constucts, pGL3 Basic vector or pGL3 Promoter plasmid (Promega), which contains an SV40 promoter upstream of the firefly luciferase gene, and 0.3 pmole of pRL-TK plasmid (Promega), which contains an herpes simplex virus thymidine kinase promoter upstream of the Renilla luciferase gene, using the Lipofectin procedure (Gibco) as described in the manufacture's protocol.

Relative Luciferase Activity Measures: -- Transfected cells were cultured for 24 hours, washed twice with 2 ml of Ca^{2+} and Mg^{2+} free PBS and lysed with Passive Lysis Buffer (Promega). Firefly luciferase and Renilla (sea pansy) luciferase activities were measured sequentially using a Dual-Luciferase Reporter Assay System (Promega) and a model TD-20E Luminometer (Turner Design). After measuring the Firefly luciferase signal

(LA_F) and the Renilla (sea pansy) luciferase signal (LA_R), the Relative Luciferase Activity (RLA) was calculated as: $RLA = LA_F / LA_R$, where relative RLA was caculated as a percentage, ie. $\%RLA = RLA / (RLA)_{max}$. To compare the Relative Luciferase Activity in one cell line with another, an Index of Relative Luciferase Activity was calculated as: IRLA = RLA/ RLA_{SV40} where RLA_{SV40} is the ratio of Firefly luciferase signal with an SV40 promoter in pGL3 divided by the Renilla luciferase signal in pRL-TK.

RESULTS

RACE Detects Multiple Transcipts: -- As a prelude to cloning the PS-1 promoter, the exact 5' end of mouse brain PS-1 mRNA was identified by the rapid amplification of cDNA ends (RACE) technique. 5'-RACE with the anti-sense oligonucleotide "101-80-reverse" found in exon 2 of mouse PS-1, gives a major broad band of 210 bp's and a minor band of 430 bp's from single-stranded cDNA templates complementary to mouse brain mRNA (Marathon Ready cDNA, Clontech, data not shown). Each of these bands was isolated from agarose gels, subcloned into the pGEM-T vector (Promega) and sequenced. Sequencing revealed the presence of three different PS-1 transcripts which appear to derive from 2 unique transcriptional start sites (**Figure 1**). This information suggests that the PS-1 gene may contain 2 promoters and that differential splicing generates multiple transcripts.

Isolation of the Mouse PS-1 Gene: -- A mouse genomic DNA library in Lambda-FIX II was screened with Probe A, a ^{32}P-labeled PCR probe corresponding to Exons 2, 3 and 4 of the murine PS-1 cDNA clone (**Figure 2A**). Of the positively hybridizing phage clones, four were selected for restriction mapping with EcoRI and Not I as shown **in Figure 2B**. Only one phage clone, Ph-2 **(Figure 2B)** hybridized to oligonucleotides from the 5'-untranslated region of the mouse PS-1 cDNA. Primers from the phage arms allowed sequencing into the genomic DNA insert. The insert's sequence allowed the PCR primer pair "1C-US-forward and 1C-US-reverse" to be chosen and used to identify a P1 clone of the mouse PS-1 genomic DNA as shown in **Figure 2A**. Clone P1-10809 was identified through a positive PCR reaction product with these primers and hybridization to the PS-1 cDNA fragment probe A (**Figure 2**). P1-10809 was then restricted, mapped and its entire sequence sublconed into multiple pBluescript-II-KS-(+) plasmid vectors as shown by the thick lines in Figure 2B.

Each subclone was sequenced on an Applied Biosystems 373A automated DNA sequencing machine using protocols supplied by the manufacturer.

Characterization of the PS-1 Gene's Exon-Intron Structure: -- The sequence of almost 50 kBp of the P1-10809 clone was compared to the mouse PS-1 cDNA sequence and regions of homology aligned with the MacVector DNA analysis program (IBI, New Haven, CT). The first nucleotide on the 5' end of the RNA transcript is usually designated as nucleotide +1 of exon 1 of a gene. In our case, PS-1 appears to have two different 5' end-sequences which are associated with three different length RNA transcripts which we named Transcript-A, Transcipt-B and Transcript-C. The alignment of Transcript-A with genomic sequence shows that a "G", designated conventionally as position +1 in Exon 1, corresponds to the transcription start site. The Presenilin gene's Exon 1A extends from position +1 to +141 which is spliced to Exon 2 whose 5' end begins at position +11,210 to give Transcript-A. The alignment of Transcript-B with genomic DNA shows that a "C" at position +411 corresponds to the alternative transcription start site. We define this second start site as beginning in Exon 1B which extends from position +411 to +781 and is spliced to Exon 2 (position +11,210) to give Transcript-B. The alignment of Transcript-C with genomic DNA shows the same "C" at position +411 as the alternative transcription start site. We define Exon 1C as extending from position +411 to +549 which is spliced to Exon 2 at position +11,210 to give Transcript-C. As shown in **Figure 2C** and summarized in **Table 1,** Exon 1A, Exon 1B, Exon 1C and Exon 2, together with a portion of Exon 3, comprise the 5' untranslated regions of PS-1 RNA transcripts.

Table 1. ***Numbering Scheme for the Mouse PS-1 Gene's Exon-Intron Structure.***[1]

	Size	Position
Exon 1A	141 bp	1 – 141
Exon 1B	371 bp	411 – 781
Exon 1C	139 bp	411 – 549
Exon 2	67 bp	11210 – 11276
Exon 3	140 bp	11367 – 11506
Exon 4	251 bp	23849 – 24099
Exon 5	142 bp	26311 – 26452
Exon 6	68 bp	34557 – 34624
Exon 7	221 bp	36060 – 36280
Exon 8	99 bp	39773 – 39871
Exon 9	87 bp	40245 – 40331
Exon 10	174 bp	42082 – 42255
Exon 11	119 bp	43217 – 43335
Exon 12		45459 --
A of ATG		11420
T of TAG		45627
AATTAA		46612

[1]The positions of the 5' end and the 3' end of each Exon were counted from the transcription start site of Exon 1A being defined as position +1. Note that this differs from the numbering in SEQ ID NO: 17.

The protein encoding portions of the gene begin with the ATG codon at position +11,420 where translation initiates in Exon 3 followed by exons encoding the remainder of the protein until stopping at a TAG codon (position +45,627) in Exon 12. 983 bp downstream from this TAG stop codon lies the putative polyadenylation signal AATTAA at position +46,612. Interestingly, the intron between Exon 1 and Exon 2 is about 10 kBp, between Exon 3 and Exon 4 is about 12 kBp and between Exon 5 and Exon 6 is about 8 kBp. The entire mouse PS-1 gene sequence is set forth as **SEQ ID NO: 17** herein, but with exon 1A begining at nucleotide 2234. Note in **SEQ ID NO: 17** that the A's of the start codons for all the transcripts, *i.e.*, A,B, and C are at the same position, 13653 in exon #, T's of the stop codons for all the transcripts are at the same position, 47860 in exon 12, and the polyadenylation signals for all the transcripts are at the same position, 48845, So the translations for all the transcripts are the same.

Characterization of the Mouse Presenilin-1 Promoter: -- 2300 bp of DNA sequence located upstream of the +1 initiation of transcription site for Transcript A was compared to its human PS-1 genomic DNA counterpart (**Figure 3**). The region of maximal similarity with human extends from postions -39 to +117 in the mouse sequence. This region is rich in Guanosine (G) and Cytosine (C) residues and contains the sequence motifs GCCGGAAGT (**SEQ ID NO: 11**) resembling an Ets 1-3 element (Fisher, R. et al., *Oncogene* **6**, 2249-2254 (1991)) and a GGGCGGG (**SEQ ID NO: 12**) motif resembling an Sp-1 hexanucleotide element commonly found in the promoters of other genes. The mouse sequences upstream of this region do not share similarity with the human sequence nor do they contain the most common eucaryotic promoter element, a TATA box (**Figure 4**). Instead, this unique mouse sequence contains two CAAATA motifs at postions -365 and -281 which resemble CAAT boxes found in other eucaryotic promoters. This unique region also contains an Ap-2 binding element at position -80 (CCCAGCCC) (**SEQ ID NO: 13**) and a sequence similar to a heat shock inducible element at position -220 (CTCGAATCGCAG) (**SEQ ID NO: 14**). Putative Sp-1 hexanucleotide binding sites with the sequence GGGCGG (**SEQ ID NO: 15**) or CCGCCC (**SEQ ID NO: 16**) are found downstream from the +1 site of transcription initiation with Exon 1A containing one Sp1 motif and Intron 1A containing five motifs. Also downstream of the Cap (transcription initiation) site are two additional Ap-2 sites and another Ets 1-3 motif.

To test whether these elements function to promote transcription, we employed a Dual-Luciferase Reporter Assay System (Promega). In general, we assayed the promoter activity of DNAs flanking the transcription initiation site of PS-1 by inserting these DNA fragments in front of a basic, promoterless firefly Luciferase reporter gene in plasmid pGL3. Constant amounts of pGL3 containing PS-1 promoter fragments and of pRL-TK plasmid containing a Herpes Simplex Virus Thymidine Kinase (HSV-TK) promoter driving expression of sea pansy Luciferase were co-transfected into a constant number of cells. After 24 hours, lysates of transfected cells were sequentially assayed for firefly luciferase (LA_F) and sea pansy luciferase activity (LA_R) so that a ratio of firefly to sea pansy activities could be calculated for each PS-1 promoter fragment as its Relative-Luciferase-Activity or RLA. Of all the fragments tested, plasmid LUC 29 with the fragment -327 to +206 showed the

greatest ratio of firefly to sea pansy activity (**Figure 5** and **Table 2**) which we defined as 100% activity. Larger fragments in LUC 1 (-2232 to +1436), LUC 3 (-499 to +1171) and LUC 16 (-276 to +519) display only a small percentage of the LUC 29 activity suggesting the presence of negative elements that apparently reduce their activities. Interestingly, the high activity of LUC 29 is not found in its flanking fragments such as LUC 2 (-2232 to -496) and LUC 23 (+188 to +519) which both lack significant promoter activity. The LUC 23 result is particularly interesting because the alternative transcription start site begins at position +411 of Exon 1B/Exon 1C and apparently lacks meaningful promoter activity.

Table 2. ***Neuron-Preferred Activity of Total and Core Promoter Regions of the Mouse PS-1 Gene.***[2]

Cell line	N2a	P19N	P19	P19M	NIH/3T3
Control	2.1	10.6	5.0	35.9	1.1
Luc29(total)	36.8	114.1	21.2	114.9	0.3
Luc27(core)	3.6	11.5	3.7	19.4	0.1
(IRLA)total	17.8	10.8	4.3	3.2	0.3
(IRLA)core	1.7	1.1	0.7	0.5	0.1
Total/core	10.2	9.9	5.8	5.9	3.4

[2]LUC29 and LUC27 were transiently transfected into the cell lines to measure the activities of Total- and Core-promoter, respectively. An SV40-promoter driving Firefly luciferase in pGL3-Basic plasmid (Promega) was also transfected as a control. An index of relative luciferase activity (IRLA) was calculated for the Total- and the Core-promoter as IRLA = RLA/ RLA_{SV40}.

To more accurately define the minimal or core regions conferring promoter activity, we studied the -327 to +206 region of the PS-1 gene in greater detail. Sequence comparision showed this region to contain a CAAT box (-281), a heat shock element (-220), an AP2 site (-80), a PEA-3 site (-53), an Ets 1-3 site (-7) and Sp1 sites (+25, +119 and +161). To find which of these elements and/or new elements were functionally active, we performed resection experiments to test smaller fragments within this region for promoter activity. Since LUC 24 and LUC 23 lacked significant activity, we initially focussed on the fragments from -440 to +91 as shown in **Figure 4**. The CAAT box at -281 plays an active role in the PS-1 promoter because LUC 8 (-261 to +91) has less activity than LUC 6 (-327 to +91) which contains this CAAT box. A negative element must reside upstream of this CAAT box

because the activity of LUC 4 (-440 to +91) is about half that of LUC 6 (-327 to +91). The Heat Shock element at -220 may not play a role in PS-1 promoter activity as fragments containing (LUC 8, -261 to +91) and lacking (LUC 10, -192 to +91) this element have similar activites. The AP2 site at -80 and/or the PEA-3 site at -53 appear to play positive roles in PS-1 promoter function as LUC 12 (-87 to +91) has about four fold more activity than LUC 13 (-32 to +91) which lacks these sites. Similarly, the Ets 1-3 site at position -7 plays a positive role as judged by the RLA activity of LUC 14 (-9 to +91) at 7.9% and LUC 15 (+42 to +91) at 0.7%. While the Sp1 site at position +161 does not appear to contribute when LUC 17 (-276 to +206) is compared to LUC 18 (-276 to +148), the Sp1 sites at +25 and +119 appear to be very active in the PS-1 promoter as negative and positive elements, respectively, when LUC 26 (-9 to +41) is compared to LUC 27 (-9 to +16) and LUC 7 (-276 to +91) is compared to LUC 18 (-276 to +148).

Based on these experiments, we tested whether the region from -87 to +41 could contain the core promoter activity in two ways. First, LUC 25 (-87 to +41) had an RLA promoter activity of 28%. Second, the deletion of this region to give LUC 30 (delete -87 to +41 from -327 to +206) decreased activity from 100% (LUC 29) to 0.2% (LUC 30). Taken together, these results strongly suggest that the Ap2, PEA-3, Ets 1-3 and Sp1 elements comprise the major functional elements of the PS-1 promoter in the region -87 to +41.

Cell-Specific Transcription: -- Using *in situ* hybridization to human brain slices, we found that PS-1 RNA was most abundant in neurons and below the limits of detection in other brain cells. This result suggested that the PS-1 promoter may preferentially function in neurons. To test this idea further, we compared the activity of the promoter-fragment/reporter plasmids LUC 1, LUC 3, LUC 4, LUC 27 and LUC 29 in different cell types. As reported above, the mouse Neuro-2A cell line of neuroectodermal lineage supports more RLA promoter activity from LUC 29 and LUC 4 than from LUC 27, LUC 3 and LUC 1 (**Figure 6** and **Table 2**). In contrast, the mouse NIH/3T3 fibroblast cell line supports only minimal promoter activity with each of these promoter/reporter constructs (LUC 29, LUC 27, LUC 4, LUC 3 or LUC 1). To further test the idea that the PS-1 promoter activity is great in neurons, we transfected the mouse embryonal carcinoma cell line P19 with these reporter constructs. P19 cells are uniquely differentiated by retinoic acid treatment into a neuron-like

phenotype (Jones-Villeneuve, E. et al., *Mol. Cell. Biol.* 3, 2271-2279 (1983)) or by dimethylsulfoxide treatment into a muscle-like phenotype (Edwards, M. et al., *Mol. Cell. Biol.* 3, 2280-2286 (1983).). Retinoic acid-treated P19 cells support as much as 2.5 fold more relative luciferase activity from plasmid LUC 29 compared to untreated P19 cells. Untreated P19 cells support as much as 1.3 fold more relative luciferase activity compared to dimethylsulfoxide-treated P19 cells

Discussion

From promoter to poly-adenylation signal, the full sequence of the mouse Presenilin-1 gene and its exon-intron structure set the stage to describe some of its unique functions. In contrast to the reported PS-1 cDNA sequence, 5' RACE (Rapid Amplification of cDNA Ends) surprised us by amplifying three different mRNA transcripts which share two unique transcription start sites. Sequence analysis showed that Transcript-A begins with Exon 1A while Transcript-B and Transcript-C begin with Exon 1B. Exon 1C is a fragment of Exon 1B sharing its 5' end at position +411, but only extending to position +549. This example of alternative splicing in Exon 1B versus Exon 1C to yield multiple RNA transcripts is well known in the field and has been described for Exon 9 in the human PS-1 gene (Perez-Tur, J. et al., *Neuroreport.* 7, 297-301 (1995)). Two distinct transcription start sites, however, have been reported for only a few genes including human catechol-O-methyl transferase (Tenhunen, J. et al., *Eur. J. Biochem.* **223**, 1049-1059 (1994)), mouse Neurotrophin-3 (Leingartner, A. et al., *Eur. J. Neurosci.* **223,** 1149-1159 (1994)) and rat aromatic L-amino acid decarboxylase (Albert, V. et al., *Proc. Natl. Acad. Sci. U. S. A.* **89**, 2053-12057 (1992)). In these cases, each transcriptional start site was associated with a distinct promoter so that a stoichiometry of one promoter per transcription start site was observed.

Our characterization of promoter activities for the PS-1 gene, however, revealed a much different picture. Using a promoter-fragment coupled to the Firefly luciferase reporter with sea pansy Renilla luciferase as an internal standard, we found that the -327 to +206 fragment (LUC 29) contains most of the PS-1 promoter activity. The known sequence motifs which apparently contain this activity are a CAAT box (-281), an Ap2 site (-80), a PEA-3 site (-53), an Ets 1-3 site (-7) and an Sp1 site (+25). While this region overlaps some of Exon

1A, deletion of the -87 to +41 region in LUC 30 reduces promoter activity by 50 fold. For the alternative transcription start site at position +410 in Exon 1B and Exon 1C, we tested LUC 23 (+118 to +519) containing Sp1, Ap2 and Ets 1-3 sites and found it to display about 1% of the activity surrounding the Exon 1A promoter. These results suggest to us that the region surrounding the +1 position of Exon 1A may promote the expression of Transcript-A, Transcript-B and Transcript-C. Alternatively, a weak promoter controlling transcription initiation at position +410 in Exon 1B/Exon 1C may amount to only 1% of the transcription initiation at position +1. By cloning all of the products of the 5' RACE into plasmid vectors and counting the number of clones carrying Exon 1C, we estimate the abundance of Transcript-C to approach 30% of all of the PS-1 transcripts (data not shown) further supporting the idea that the major promoter at +1 functions to control transcription initiation from both the +1 and the +410 sites. Quantitative measurement of Transcript-A, Transcript-B and Transcript-C levels will help to further resolve this issue. The high homology between human and mouse promoters, combined with our description of multiple start sites and alternative splicing for the mouse PS-1 gene, reasonably suggests how the human PS-1 promoter may function (**Figure 6**).

Recently, PS-1 was reported to be expressed predominantly in neurons of the central nervous system (Kovacs, D. et al., *Nat. Med.* **2**, 224-229 (1996)). This result matches our own data that PS-1 RNA, by *in situ* hybridization, is strongly expressed in neurons and at undetectable levels in other cell types. Similarly, several immunohistochemical studies report primarily neuronal localization of PS-1 protein with weak staining of amyloid plaques and some glia surrounding those plaques. On the other hand, Sherrington et al. showed that Northern blots of RNA from different organs all hybridized to a PS-1 cDNA probe suggesting that PS-1 RNA is ubiquitously expressed. At present, these results can not be easily reconciled.

Our data shows preferential promoter activity in neuron-like cells supporting a cell-type-specific pattern of PS-1 expression. We find the greatest amount of PS-1 promoter activity in the mouse Neuro2a-neuroblastoma cell line, followed by the P19 embryonal carcinoma cell line and almost no activity in the mouse NIH/3T3 fibroblast cell line (**Table 2**). To further confirm this finding, we employed the P19 mouse embryonal carcinoma cell

line because of its unique ability to be differentiated into a muscle-like phenotype (aka. P19-DMSO-muscle) following dimethylsulfoxide treatment or into a neuron-like phenotype (aka. P19-RA-neuron) following all trans-retinoic acid treatment (Jones-Villeneuve, E. et al., *Mol. Cell. Biol.* **3**, 2271-2279 (1983); Edwards, M. et al., *Mol. Cell. Biol.* **3**, 2280-2286 (1983)). If our hypothesis that PS-1 promoter activity is preferred in neuron-like cells, then we would predict that P19 cells differentiated with retinoic acid into neuron-like cells would display more PS-1 promoter activity than P19 cells differentiated with dimethylsulfoxide into muscle-like cells. As clearly shown in **Figure 5** and **Table 2**, P19-RA-neuron cells display the most PS-1 promoter activity followed by untreated P19 cells and the least activity in P19-DMSO-mucscle cells. These results combined with the Neuro2a and NIH/3T3 results indicate a clear pattern of PS-1 promoter activity which is preferred in neurons.

The mechanisms controlling neuron-specific promoter activity are poorly understood. The most direct mechanism would be for a positive regulator, that is only present in neuronal cells, to singularly activate the neuron-specific promoter. Alternatively, a negative regulator, that is only present in non-neuronal cells, could globally repress the neuron-specific promoter in all but neuronal cells. Depending upon the exact DNA elements within the promoter, some combination of positive and negative controls of transcriptional activity might also yield neuron-preferred promoter function. Going beyond our characterization of the regions conferring PS-1 promoter activity in Neuro2a cells, we may now look at the data to suggest which of the DNA elements might confer neuron-preferred promoter function. The region showing the highest activity in Neuro2a neuron-like cells extends from -329 to +206 (LUC 29) and contains a CAAT box at -281, a heat-shock inducible element at -218, an Ap2 site at -80, a PEA-3 site at -53, an Ets 1-3 site at -7 and an Sp1 site at +25. The CAAT box probably supports about a third of the positive control of neuron-specific activity as its deletion reduces promoter activity by about a third when comparing LUC 6 (-327 to +91) with LUC 8 (-261 to +91, **Figure 4**). The heat shock element probably does not contribute to neuron-specific activity under normal conditions (non-stressed) as its deletion does not affect promoter activity when comparing LUC 8 (-261 to +91) to LUC 10 (-192 to +91, **Figure 4**). Based on the four fold greater activity of LUC 12 (-87 to +91) compared to LUC 13 (-32 to +91), it appears that both the Ap2 site and the PEA-3 site are good candidates for the positive

control of neuron-specific promoter function. Ap2 sites are reported to be most frequently found in promoters active in cells of neural crest lineage and several examples exist of their involvement with neuron-specific activity (Sato, T. et al., *J. Biol. Chem.* **270**, 10314-10322 (1995); Petersohn, D. et al., *J. Biol. Chem.* **270**, 24361-24369 (1995); Chin, L. et al., *J. Biol. Chem.* **269**, 18507-18513 (1994)). In contrast, the five fold less activity of LUC 26 (-9 to +41) compared to LUC 27 (-9 to +16) implicates the Sp1 site at +25 as a negative regulator of neuron-specific promoter function. These same data could also be interpreted as the Ets 1-3 site having a positive function, possibly as part of a core promoter element from -9 to +16. Direct measurement of LUC 27 (-9 to +16) shows that Neuro2a and P19-RA-neuron cells have more activity than do P19-DMSO-muscle or NIH/3T3 non-neuronal cells supporting the idea that this 25 bp region contributes to neuron-preferred promoter activity. The major transcription start site at position +1 is located in this proposed core promoter element. The ETS-1 transcription factor prefers binding to the Ets 1-3 binding site found in this core by a ratio of five to one over the PEA-3 binding site (Fisher, R. et al., *Oncogene* **6**, 2249-2254 (1991)). This finding is particularly interesting as the ETS-1 transcription factor is thought to be specific for B cells and resting T cells of the immune system and not been previously described for neuronal cells. Sp1 binding sites appear to be ubiquitously distributed in all promoters of all cell types and their ability to function as negative elements appears to be novel.

The foregoing is illustrative of the present invention, and is not to be construed as limiting thereof. The invention is defined by the following claims, with equivalents of the claims to be included therein.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Vitek, Michael P.
Mitsuda, Noriaki
Roses, Allen D.

(ii) TITLE OF INVENTION: Presenilin-1 Gene Promoter

(iii) NUMBER OF SEQUENCES: 17

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Kenneth D. Sibley; Myers Bigel Sibley & Sajovec
(B) STREET: P.O. Box 37248
(C) CITY: Raleigh
(D) STATE: North Carolina
(E) COUNTRY: USA
(F) ZIP: 27627

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Sibley, Kenneth D.
(B) REGISTRATION NUMBER: 31,665
(C) REFERENCE/DOCKET NUMBER: 5405-200

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (919) 854-1400
(B) TELEFAX: (919) 854-1401

(2) INFORMATION FOR SEQ ID NO:1:

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

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

AGGCCGGAAG TTGCGACACC GGTGA 25

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

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

CGGAGAGAGA AGGAACAAAC 20

(2) INFORMATION FOR SEQ ID NO:3:

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

(ii) MOLECULE TYPE: cDNA

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

TCAGCTCTTC GTCTTCCTCC TCATC 25

(2) INFORMATION FOR SEQ ID NO:4:

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

(ii) MOLECULE TYPE: cDNA

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

GATCACAGTC TAGGTTGCTG GTGTG 25

(2) INFORMATION FOR SEQ ID NO:5:

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

(ii) MOLECULE TYPE: cDNA

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

TGGGGCAAGG GACACAAATA AG 22

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

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

TGGCTCAGGG TTGTCAAGTC 20

(2) INFORMATION FOR SEQ ID NO:7:

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

(ii) MOLECULE TYPE: cDNA

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

CCATCCTAAT ACGACTCACT ATAGGGC 27

(2) INFORMATION FOR SEQ ID NO:8:

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

(ii) MOLECULE TYPE: cDNA

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

CAAACCTCTT GGGATTCTTT C 21

(2) INFORMATION FOR SEQ ID NO:9:

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

(ii) MOLECULE TYPE: cDNA

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

ACTCACTATA GGGCTCGAGC GGC 23

(2) INFORMATION FOR SEQ ID NO:10:

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

(ii) MOLECULE TYPE: cDNA

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

AAGACCTCGA AGGGCTGCTG TC 22

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

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

GCCGGAAGT 9

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

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

GGGCGGG 7

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

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

CCCAGCCC 8

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

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

CTCGAATCGC AG 12

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

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

GGGCGG 6

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

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

CCGCCC 6

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48974 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 2234..2374
(D) OTHER INFORMATION: /product= "Exon 1A; initial exon of ps1 transcript A"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 13443..13509
(D) OTHER INFORMATION: /product= "exon 2"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 13600..13739
(D) OTHER INFORMATION: /product= "exon 3"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 26082..26332
(D) OTHER INFORMATION: /product= "exon 4"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 28544..28685
(D) OTHER INFORMATION: /product= "exon 5"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 36788..36855
(D) OTHER INFORMATION: /product= "exon 6"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 38293..38513
(D) OTHER INFORMATION: /product= "exon 7"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 42006..42104
(D) OTHER INFORMATION: /product= "exon 8"

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 42478..42564
(D) OTHER INFORMATION: /product= "exon 9"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 44315..44488
(D) OTHER INFORMATION: /product= "exon 10"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 45450..45568
(D) OTHER INFORMATION: /product= "exon 11"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 47692..48974
(D) OTHER INFORMATION: /product= "exon 12"

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 13653..13739

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 26082..26332

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 28544..28685

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 36788..36855

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 38293..38513

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 42006..42104

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 42478..42564

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 44315..44488

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 45450..45568

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 47692..48974

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 2644..2782
(D) OTHER INFORMATION: /product= "exon 1c; initial exon of PS1 transcript C"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 2644..3014
(D) OTHER INFORMATION: /product= "exon 1b; initial exon of PS1 transcript B"

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

GAATTCTGGT CATATGCAAA TATTACTTAC CTTGGAGTCA TTGTTAGGAA AACCTGGCCT 60

GATTGACTAG GTCCTGGCTT CTAGTGACTG CATTCCTAGT CTTGCCTTGC AGGCTCCTCT 120

CCTAAGTGAC CTAGTCAATC AGGCCAAGTC ATTTCCTGCT TCACTCATTA GTTGCTACTG 180

GGTTTTCTCA TCTTCCTATT GCCAGGTTTT GAATAATGGG GAAAATAAAT GCTACACTGC 240

AGAAATTCTC ACTATTCTTT GGCACTGTTT GGTCATTCTT AGGTTTTAGT GTTTCTAGAC 300

CCTCAATGCT ATATTACATT GAAAAAATAG AAAATTATTT AGTATGTGCC TTTCAAATGT 360

CTCAGATCTT ATTCACATAG TTAGAAGTGT TGTGTGCCTT GATATTTCCG TGGGTGCCCG 420

AGTGGGATAC ATTAGCTTCT CTTGGCACTC TATAAACATT GTCTGTCTAT TTGATATTCA 480

TTGGTGGTTA ATGTTTAATC ACCTTTGCCC TTCCCTGGTG TCAGTCTTCC CATGTCATTT 540

ATGTAATATT TGTGTTTATA TCCCTAAAAA GGAATATAGA AAACAGAAGT ATCTGTAGTC 600

TACTCAGTTC CTGTCCCCCC TCCCCTTTGA GGGTTTACTT TTACAAGAAT CTATTGGAAC 660

ACTGAGTGTA AATACAACTG CCATTTTTGT CCATTAGACA GTTCTCTTGC TCATAGCTTT 720

CATTCATGAG TATTTTTTTT AAGATTTATT TATTTATTAT ATGTAAGTAC ACTGTAGTTG 780

TCTTCAGACG CACCAGAAGA GGGCGTCAGA TCTCATGGTT GTGAGCCGTC ATGTGGTTGC 840

TGGGATTTGA ACTCAGGACT TTTGGAAGAG TAGTCAGTGG TCTTAACCAC TGAGCCATCT 900

CACCAGCCCC ATTCATGAGT TTTATAACCT AGGTGAACTA AGATTTGTTG TAGGACAGTG 960

AGATTTTTTT ATATATATAC CAGTACTCCA CGAACATCTT TGCACAAAAA TGTTTGTCGA 1020

AATTGCTTCT CTGTATAGTT TACAGCATTT GCCATTCTAC CAAAAGATCG CAATTAATGT 1080

AGTTTGCAAA TGGCTTTTAG TCTATGCTTG TTTAGCAAGT GTATTGAGCT TGTTCTGTAC 1140

TGTCTACAAT TTTCTGCCTT CTTCCCATAT AAGTAAATGA TTGAGGGTAT AAGTAAATGA 1200

TTGAGGGAAA ACTACCCAAG GTTATGCTAT GTCTCCTCCC AGTTGCCATT CCGCTGGCGG 1260

AAAAGACTAA AGTATCAAGA TCTTAAGCAT GAAAATAGTG AAGTCTGAGT TTATTGAAGC 1320

AGGGTTGAGA ATGTTGGGAG AAAATGGAGA TGGATGCTCT ACTATTTAGA ACGAGGCCTT 1380

CTGGGAAGTA GAAAAAAGGA AGTTTGACCC AGTGTATCTT AGTTTGACCT GTAATTTGGT 1440

AGGAATCCCA GTCCGATTTC AACTCACTGT CCTGCCCCAT CTCCTCCTGA GTGTTATGAT 1500

TTCAGATACA ACACCATCCC GATGTATTTG TCTCTGTTGG GAGTAGAATC TAGTTCCTCG 1560

TGTGTGCTCC TTAAGCACAT TAGCCGTTTT CTGGTTATTT GGGATTATGA ATAATTTCTC 1620

TGTTGTTTTT TGCCCAAGTG AAATGTGTGG TGTGCTGTTG ATAATTGTCT CTCCGATTAA 1680

GGAAAATCTG AGGGATGTAA AAATCAAGAT AGGAAACATT CTTTGCTTTC TAGAAGCTTG 1740

CCATCTGGGA AGACTTTCAG ACTCGGAACC TAGACTAATA GGATATTTCT CAGGTTCTGT 1800

TTACCCACCA ATCGCTTTGG TTTATTGAGA GTACAAACAG TAAATTTTAT CCGTTTTGAG 1860

GAACTGTTCA AATAAATACT GTGGAGAAAT GGGAAGTGTT GGATTTAAAA TGTCACTACA 1920

AAACAAGGCA CGGTGTCCCC ACACATGGAA ACCAAATAAA TAGGTTCAAG CCCATCCTTT 1980

CCTACAAGGT ATGTTTGATA CCGGCAGGGA ACGCTCGAAT CGCAGTCTCA ACCAAAAACA 2040

AGGGAAAATG TCACTTGTAG ACTGGAAGAA CGCTAGACGC GCCTCAAACC CTAGAGAGGC 2100

CTCAGGTCGC GCACATCCTT ACATCTATGC GAGTGGATTA GGCCAGCTCC AGCCCCAGCC 2160

CTCGTGGCCT GCGCGCGCCA CCGGAAGCTC CGTCCCCTCC CGGGTCTAGG GGCCAACGTC 2220

GCCGAGGCCG GAAGTTGCGA CACCGGTGAG ACCTCTAGGG CGGGGCCTAG GACGACCTGC 2280

TCCGTGGGCC GCGAGTATTC GTCGGAAACA AAACAGCGGC AGCTGAGGCG GAAACCTAGG 2340

CTGCGAGCCG GCCGCCCGGG CGCGGAGAGA GAAGGTGCGT GCCCAGGGTG TGCGGGGCGG 2400

AGGGTGTCTC TGCCGGTCGT GTTCACCGTC GCCTGCCTGC CGGGGGTCCG GGCGGGCCTG 2460

TGTCTCCGAG GGCCGCGCTG CGGGCGTCTC TAGGGATGAG GGGCGGGGTC CAGGCGGGCG 2520

GAGATCGAGG AACCCCGCGT GGGAAACGGG GTGAAGCCGG TTTCTCGGAA CCCAGCCGGG 2580

GCCAGACTGA GAGCAGCCTT CTCCGAGCTT TGGTACCCCG GAAGTGCTGG CTTCCCCGGG 2640

CGGCCGGGAG CAGATGGCTG GCATCAGGGG TGGCCTCTCG ATCAGAGTGG AGCTAGAGAT 2700

AGAGGAAGCG CCCTAGGCTG GGTCGCCTTG AGCAACTGGT GAAACTCTGC GTCTGGTGCC 2760

CCGAGTGTGT CATAGTCCAG AAGTGAGTGA GTGGCACTCG GGGCTAACTT CTCAGTGTGT 2820

CTGTTGGCAG GCGAAGAGTC GTATGGCGCT TGTTCTTTCT TCAGAACTGT CACGATGCAT 2880

GTGTCTGACG CTTGTAGGCG CCTTTAGTGT TTGCTAGTTC TTCCCTTTCA CGTTTTCAAA 2940

TGTGGCACTG CCGCGAACAG AGGCTCAGGC ACATCCGCCT GGTAGTGGAG GACGGAGTGC 3000

ATGAGACAAT TGAGGTAACT TTTTACATAA CTAGGAGAAA TGAGAGTTCC AGTAACAGAA 3060

ACGTAGAAGA AACGTGAGCA GGACAGGCGA TCTGTTCAAC TAACGCATAA GATTGACTGG 3120

GTTCAAGTTT GGAGAATGAG AGAGAAATAA GGAAGTTAGG CAGTGGTTTG GAAATAGTAC 3180

TTTTGAAGTG ACAGCGGGTG AAGATGGCCT GAGGTCAGAG CTGTAAACTG TAAACATGGG 3240

TCACTTAAGT AGATTTGCTG TTTGATGTTG GAGCAGAGAA CTGGGTGGAC ACTTAGGGAA 3300

CACAACAAAG AGCCAGGTAG AAGGAGGACA ACGAGGCTGC TGAGGAGGAC GAATGCAGAA 3360

ACCACAATAG TGCAGTGTCA CAGAAGCTGA GAAACACGGA AGCTTCAGGG AGCATTGTCA 3420

GCGATAGGAA TGAAAGGCCT GGTGTTGTTT TAGAAGGATC TCTGTACCAT TGTACCCTTG 3480

AAAGATGAGC GTGAACCACA AATACTAGTT TTAAATGTTC CCACAGACAC TTTTTTAAAA 3540

AATTTTTTAG AGGTTTATTT ATTTATTATA TGTGAGTACA TTGTAGCTGT CTTCAGACAC 3600

CCCAGAATAG GGCATCAGAT TTCATTACGG ATGTTTGTGA GCCACAATGT GGTTGCTGGA 3660

ATTTGAATTC AGGACCTTCA GAAGAGCAGT CAGTGCTCTT AACAGCTGAG CCATCTCTCC 3720

AGCCCCACAG ACACATTTTT AAAGGTAACA AGAAGCAGTT GAAATTAGTC TTAATGATTT 3780

GTTAATCCAG TGTAGCTAAC AGTTTCATTT TTAACAGGAA ATCCGTGTCA AATTAACTTG 3840

TCCTCTCTAG ATAGTTAGGA GGCTCTTCAG TGACCAGAAG CAACCCCTGA CTACAGTATC 3900

CCCAGCACCT GGCACAGTGT TTCCTTGTCT AAAAGAAACC TGATTCCTAT TTGCTGCTCC 3960

AAGAATCCCT ATTTGGTTTT CAGATGTCTT TTCCAGTTCC TTGTTTGAGT GAAAGCAGAA 4020

AGAATTTATG CAAGTTAATT AGATGTAAAT ATTCGTTTTT AATTTAAAAA TACTTGTATT 4080

CCATCCATTT CCACAGAAGG TAGAGACATT GACTCACAAT GAAAATAACT TCATTACAGA 4140

ATTATTAAAA TAACTTTGGG TTGAGTTTTT ATAATTTCCA GGTATAGGCG GTGCTTTAGG 4200

AGTCCCAGCT ACTTGGAAGG CTGAGCTGGA AAATCACGGC CAGCCTGGAT AACAGAGATG 4260

GTGAATGACT CGAGAGTAGA GCGAATTCAT CCACCTGATC TAACCTAGGA TGGCCTGTAG 4320

TCTGTTTGAC AGTACAGATA GAGACCATGT TGAGTCTTAT TCTGTGGGTG TGATATGTAT 4380

GTGTGTGCAC CTCTACTTTA TATGCAAAGC TGAAGTCTTT GCCTATATGC TCACAAATTC 4440

AGAGGCCAGA GGAGGACAGG GGGAGCGTTT CTCCACTGTT TATTGCCTTA CTTAAAAAAC 4500

AAAAATGTTA TTTTTGTGTG CAGGGTTGGG GGGGCTTCCA TAGTAATATT CGGAAGTCTG 4560

AAGTTAGCTC TCTCTTCCCA CCTTTATATT AGTACTGGGT ACTGAGCATA GGTCTCCGGG 4620

TATGTGGGAC AAGTACCTTT ACTGTCTGAG CCATCTCGAG AGTCTCCCCC CACCCCCACC 4680

CCCACCCCCG CCCCCTGGGC CTTTTTATTT AATTTTTGTT GTTGTTTTGC TTGGTTTTGT 4740

GAGACAGGGA CTCTCTGTGT AGCCCTGGCT TTCCTGGAAC TCACTATGTA AACCAGGCTG 4800

GCCTGAAGTT CACAGAGATC TGCCACCACA CCCAGCTTAA TTTTTAAAAA CTTAATTGTG 4860

TGTGTGGGAG CACGTGCGCC ATGGCACACA GATGTTATAA CACACATGAT CACTTTATGG 4920

AATCCTTTTT GTTTACGATA TGTGGGCGCT GGGGATTAAA TTCTGCTTGT CAGACGCGGG 4980

CTGCACCATG AGCACCAATT TTATTTTATA AGGCACAGTC TCTTTATTGA GCCGAAGCTG 5040

TTTCAGCTAG GTGGGCTGTC TGTCAGTCTC CTGGGAGCCT CTTGTCTCCA CCTCCCAGTG 5100

TTAGGACTGT AGGCGCAGGA ACTATGCCTG GCTTGGATGT GATGCTGGGA ATTTGAGCTT 5160

TGCTCCTTTT GACAGCAAGT GCTTTTACCC ACTGAGCTAA CTCCCTAGAT CCTACTGTCT 5220

TGATATACTT TGAGACTGTT GGTGTGTTTC ACCATAGACC TAAATGGATA TTAATTACTC 5280

TTGTGATTCT AAGGCTCTCG GATGTACTAT TGAGCATATA CTGCCTTTAC AGAATCCAAA 5340

AGAAAATCCA TATTCTAAAT ACAGTTAATC CTGAAGGTTT GAAATAAATT ATGAGCTGAT 5400

TTTAGCTCAT GTTCATTGTG TCTTGTGAGC CCGCCAGTTT TACCATTGCT TTGAAATAAG 5460

TTGCAGCAAT AAGTCTCCAT AGCAGAGCTC CAAACCTCAC AGCTTAGAGC ACTTAAGTTA 5520

ACTCAGGGCT CTGCTTCAGG GTGGAATGTG TGGCTCAGCT GGGACCATTT TGATTTGAGC 5580

ATCTTCTCCT GGGGCCCAGG TTGAGGGGAC AAAAACTATC AAACCACTCT GCTGACTTCT 5640

CAAGGCAAGG ACAGCTGCAC AAGCATGGCT CACATATATC TGCCAGTGTC CATCAGCTAA 5700

AATTGCATTG ATAAAGGCAC AGTATAGGAG CTCACATAAA GACTTAAAAT AGGTGATGGC 5760

TGTTACCCAG AAAGAGACCA AAGAGCAATC TCAGGGAATT GGTTCTAGGA CTACTAGTTT 5820

CCAAAATCTG AGGATGTCCC AGTCCTTACA TTTAACAGTA GTATTTGCGG GTAAGAGTGT 5880

AGTCATGTTC AGAGACCTTG CTTGATTCCT AACACTGAAG AAAAAATATG AACCAAAAAT 5940

AAATATATTA TTTATAAATC CTCCGGTGTA CTTTAAATCA TCTCTCAACT ATAATACCCA 6000

CTAATGTAAA TGCTGCGTAA ATAATCCTTA CATTGTGTTG ATTATGAAGA ACAATGAAAT 6060

CTACGTGCCA AATGTTTTCA GTCTACAGTT GATCGACTCC AAGGATCCTG AATTAGAAGA 6120

TATGAAGGGT AACCGTTCTT CCCTTGGGGA GACTGGTGGA AGGAGAGGAT ATTTACTGAA 6180

TAATAATTGT CTCTACCACA GATGGTTCGC AGAGGTGGAG TGGGATGGAA AGAGGGAAAA 6240

AGCCACCCTC AACCGGCTGC ATCACTGCCC TTGCTCCTTT TAGTTCTGTT ACTTTCAGAA 6300

CATGTCTTGT AGGGCTGGGG GAATGCTTCG CTCGCTCTCT CTGCCTTATG GATCATGGGT 6360

CTGTCTTGGG TATTCCTAAG TCAAGTTCTG GAATACAGTT GAGGGTTTGC TTACATTCAC 6420

TGTGGTTATG AGAAGCCCAG GTAGAACAGG GTTTGATGGT ACNCACTTGT AATCCTGGCA 6480

CTCAGGAGTT GGAGTCAGGA NGACCTCAAG CACAAGCATC GTGAGCAAGA CCTGTCTCCA 6540

TAAATGCAGA AACACCAGGG CGCACCTTTA ACAACAGTGC ACAGGAGGAA GAAGTACGAG 6600

GACCTCTGAG TTTGAAGCCA GCGTGATCTT CATAGCAAAT TGCTGCCCAG CCATGGTTAC 6660

ATTCTGAGAC CCTGTCCCAA ACAGACAGAC AGCCAACGAG AGAGATGGCA CTTAGTGGCA 6720

GGACACTTTC CAAGCATCTT TGCTTCCCAC CAACCAAAAT AAGTAAATCA AAATTCAAAC 6780

TCTAATCAGA GTTTGAGTTC CCTATCAGGG AGATAAGTAG GAAAAACTGC GTGGGGCGCG 6840

GGAGGGTGTG GAGGGTCCTG GAAGAGAGGA ACAGCCTGTT TGTGCAGTCT GCCTGTCAGC 6900

TGTGATACCT GGCCCCAGCT TTCATGATAA AGCTTTAGTT TGTTTGACAG GCAACCTTGA 6960

GATCATGGGG CTTTCGGCAG ACTCAAAGGG CGGAGAGATG AGGGAGTAAA ACAAGAAAAG 7020

CTGGGTCTTG AGAACTCTTG TCCTCTGAGA GTCAGAACGC CACTTGGTGG GATGAAGTGC 7080

CATGTTACCA TGAATATGTC ACCCACCTCC CAAACTTCCA TCATGACGTC TTTAAGTCTG 7140

CAGCTCAGAA CAGAGCCCAC TATTCTGCCC CTCTGCCTTT TCTTGTAGGC CCCTCAGCTC 7200

TCTCTTTGGA GTATGTGCCT TGCTAATAAA CTTCTGCCTA GACTACATGT CTGTGTCTAG 7260

TCTGAGTCTT CTTCTTCCCC ACTTCAAGAC AGGAAATGGA GTCNTCACTT CAGGCCATGG 7320

CGGCAGCAGC CTGATTGGTA GGAAAATAAA TCAAGAGCTT GAAAACGTTA TTTATTTATT 7380

TATTTATTTC CTGCCTCACA CCTTGAAAGC TTTTTAGCAC TATAGCCATT TATGTAGAGC 7440

AAAAAAATAA TTTTGGTAGT TTTTAAAGTG GTAGAAGCTG GGTAAGTGGC TTAGGGACAT 7500

TCCGCCACAG GCTCTTTGTG GTGATAAAGC CATGCTTATG TCTCACTGAA GATGTTTTGT 7560

AGTGTACCAG TTTTATTTAC ATTTGTCTCA AGAGTTAGAA AATGAGTTCA GTGTACAGTG 7620

AGATGGGAGA AAGTACCTGT TGGCTTTTTT GGGGGGTGAG AGAGTTCGAG ACAGGGTTTC 7680

TCTGTGTAGC CCTGGCTGTC CTGGAACTCA CTCTGTAGAC CAGGCTGGCC TCAAACTCAG 7740

GAATCCGCCT GCCTCTGCCT TCCAAGTGCT GGGATTAAAG ATGTGCGCCA CCACTGCCCG 7800

GCTAGAAAGT ACCTTTTTGT GTTTGTTTTT TTTTTCTTTC CTTTAAAATT GATTAAATTT 7860

CAATTAGGTC TCTTGTAGCT CAGGCTAGCT TTCAGATTTA ACTTTGTAGG CAAAGATGAT 7920

TTTGAAGTTC TGAATTTTCC TGCCTAACAT TCTTTCTAAG TGTTGGAATT AACAAGTGTT 7980

TCAAACCCAT ACTGGTTTGT GTTTAAGGAG ATTGAAGCTA GGGCTTGGTG TGTGCTAAAC 8040

AAGTACTCTA TGAACCAAGC TACATCTCCA ATGCTGTGGG TGCTTTAGTG AATGGGAATT 8100

TCTTTTAGGA ACCTATTCTG AAACTTTAAA CATTGCTCCT GAGATGTTTT AGAAGTTTTG 8160

GTGTTGAGAG GAGGTTTTCT AGCTGGATGG ATTTTCTTTT TTCCTCCGCT ACGTGGGTGT 8220

GTGTAACATG GGTGTGTGCT GGAGGCTGTA GAGTCTTTCT GGTTCACTCT CCACTTCTTT 8280

CACTGAGCTA TGGGTTTAGT CAAACCCAGA ACTCACTAAT ATAGCTTGTC TGGCCAGCCA 8340

GCTTGCTTTG GGGAATTCCC AGTTTCTGCC TCTGTGTTCT GGAATTAGAG ACAAGCCACC 8400

AGAGCCACCC AGTATTTAGG TACGTTCTGA GGAATCGTAT ATCCAGTCCT CACATTGCAT 8460

AACCACTGAG CCATCTCTCC AGACCCTGCA TCGGCTTCTC TGTGGAGCAC ATCTAAAGCT 8520

GTGAGATTAC TCATAGGCTG TGGCTTTTCC TTCCCTTGTA TCTCTGCTTT TGGTTTCCCC 8580

AATGCTAAGG ATGGATCCCA GGGTCTGCAC ATGCTGTGCA AGCACTCTGT CACAGGGCTG 8640

CATCCCAGCT CCTAGTCTTT GTCTGATAGA AATTGCCATT GCTCATCAAA ACAATGACTT 8700

CTTTTTCCTT TGCCCGACTG CCCAAATTGT TGTGCTTGGA GCAAGCACAA AGCAGAGGGT 8760

AAAGAAGCAT TTATCAGATC AGAACTGATA GGAAAGTTCT AGTAAAGAAA ATACATTGTG 8820

GGCAGAAACA GGTGGCGTCC TTAAAAGGGA ATGGCGTCTG TTTATAGCAG TCAAGAGACA 8880

CATTGCCTGG CCTTACCTTT TGAGCTATTT GTATGAGGTC CAATTTGATT AAGTAAGAGG 8940

TAGTCTGCTC ATTGTCTTGG GTAAATGTAT AGATTGGCCC TGTCAGGTAA GGTCCTCCGG 9000

ACATTGGGAG TTAGTCTTAT TGTGTATATG AGGTTAGCAG CACAGCTGCA TAGCTAGGAT 9060

CTGATTGTTA GTGAGCATTT TGGTTTTTTA CTGTTGTCCC TTTTGGCCTC CTCATCTCCA 9120

CACAAGTGTC TGAAAGAGAG GAGCTAAATA TCAGGAAGGT GTTGCGGTGT GTCCGTCTTC 9180

AGTGGCTGTC CAAAAAGAAA AGCACCAAAA TCTATATCCC TTTCCCTCCT CCCTCTGCCA 9240

TCCAGGTCTC TAGTCAGCTA CTGGCTTCTG GTGTTCATAG CTCTCAGCTG AAAATGTTAT 9300

CAACTTTGCA AGTGTCTTCT GCCATGCTGC CTGGGCAGTG GGGTATTGAC CTTTTCCAGT 9360

GATGGTTTTG GCTGTTGTTA AATCCGAATC AACTCCGTTT GGTCTACCCC CTTTGTCCCA 9420

GGAGTGTGTG CTTGTTCCAT GACCTGGCCA TGGCCTGCTT CTGCTTCGGG GCTTATTCCT 9480

TCTGATCATA TTCCATCAGG CTCTTCCTGG GATCGTGTGA TTTCTCTGGG TGAGCAGAGG 9540

ACCCCGGCTG AGTGTCCTCA CCACCTGGAA TCTGGATTGG GGGTGTTACT CATTTGCCCC 9600

TTAACCCTGC ACAGGCACAT TGCCTCACCT TGATGTTTAT GACATTCGAA CGCAGAGTCT 9660

TCTAGAGAAA AAAATTTTGG GGGGGGGGGG GAGGGGTTTG AAAGGTAAGG AGAAAACTCC 9720

CTGNAACCGG GGATAGGGGA TTCAAAAGAG AGTTGTCCCT TTTATTTATT TTCAGACATT 9780

TTTGTAGATT TATGTGTATA AGTGTTTTGC CTGTGTGCAT AAATGTACAC ATCTGTGTGC 9840

CTGGTGCCTG TGGAGGTCAG AAGAAGACAC CAGATGCCCT GAAACAGGAG TTATGAATGA 9900

TTATGAGCCA CCATGTGGGT GCTGGGTAGC AAACCTGGGT CCTCTGCAAG AGCAGCAAGT 9960

GCTCTGTAAC CACTGAGCCA ANTGCAGCCC CNTCCCCCAA CCCCACCCCC ACCCCCAGGG 10020

AACTGTATTA GGGTTTCTAC TGCTGCGGAC AAAAACACAA TGACCGAAAA GCAAGTTGGG 10080

GAAAAAGGGT TTATTTGATT ATACTGCCAG ATCATACTCC ATCGTTGGAG GAAGTCAGGA 10140

CAGGAAGTCA AGCAGGGCTG GAATCTGGAG GCAGGAGCTG ATGCAGAGGC CATGGAGGGA 10200

TGCTGCTTAC TGGCTTGCTC CCAGAACCCA GGACCACCAG CTCAGGGATG GTACCACCCA 10260

CCTTGGGCTG GGCTCTCCCC TGTTGATAAT AAATTGAGAA AATGTCTTAC AACTCAATCT 10320

CATGGAGGAA TTTCTTCAAC TGAGGCTCCT TCCTCTCTGA TGACTCTAGC TTTTGTTGAC 10380

ACACACACCA CAAAACCAGC CAGTACAGGA ATAAAATTTT GGGGGTGTGT GTGGTTTGTT 10440

TGGGTTTTGT TGTTGTTGTT GTTGTTGTTT TTAGACAGGA TTTCTCTGTG TAGTCCTGGC 10500

TGTCCTAAAC TCGGTCTGTA GCCCAGACTG ATCCTAAAAC TCAGAGATCT GTCCTGCCTA 10560

TGCTACTGAG TGCTGGGATT AAGGGGGTGT GTCACTATTG CCCAGCTCCC ATGGGAACTT 10620

TTTTAAAGAG TCAAACAGAA TAATGATTAT GAAAAGAACT TTGAAAATAG CATTGGGAGC 10680

ATGCTAAAGT GATGGGTGTG TTATTGGTGG GTCTTAGAGG ATTTGTCCCA GAACGCCCCA 10740

GTGTGGTTCC TTATAATGGC TAAACAGGCT ATGTGAGCCT GAAAGAAACG ACTTAAAGGA 10800

CACAGTCCCA CCCATTGTAA CTTACCTGTT AAAATAAATA CTTGGACTGT CTGTGCAGAA 10860

TCTCAAGTTT CCTTATATAG TGGGAGCATT GATGTTAAAG ATCTCGTTAC TTTGTAAGCA 10920

GTAGTTTTGT AGTAATAAAA CGAGTGTAAT ATAACCTAGT CGGCATTCCT GTAAGNTCCT 10980

AGAGGCAGGG ACGTTGTTTT GCAACAGAGT CCTGTATGTG GGTTAGGATC TTTTGATGCC 11040

AGACAGGGTA AGATGGTTGA CTTTCCTTGT ATTTTGAGAC AGGGTCTCAT TGTGAGGCGG 11100

CTAGCTGCTG TGCGTGCTTG GATTAAAGGC ATGTGGCACC ACGCTCAGCC ACTGATTTTC 11160

TTGTAAAATG GATGGGTAAA GTGAGCATAA TACGTTACTT AAGGTGTTCC CACTTTCTGT 11220

TTCAATTTTG CATGGAAGGA AGGAGAAAGA TGAATATGAA AGACAGTTTA CAGTGGTCTC 11280

AGCCCTGTGC TCACCTTTGT CCTGCCCTTC TCCAGGGCCG GGTGTGGTGG CATATGCCTT 11340

TAATCCCAGC ACTCGGGAGG CAGAGGCAGG CGGATTTCTG AGTTCGAGGC CAGCCTGGTC 11400

TACAAAGTGA ATTCCAGGAT AGCCAGGGGC TACACAGAGA AACCCTGTCT CGACACCCTT 11460

CCCCCCCCCC ACCACCACTA NCCCCAGTTG GCAGGATGCT GGAACAGACA GGTATGGTGG 11520

TTGATCACTT GGAGCACTCA CTGTAAAATG AGTTCCTTAA GTTCAACCTC TCAGAGACCT 11580

GAGAAGGACA AATTTTACCT GAGTAAGCAG GAATTGTAAT CTAAGTAGCC TAAGTAACAA 11640

TTAAAATGAC AGACTTACCC ATTGCCCAGA CAGCCACTCC ACCTGAAAAC AGAAGTAGTC 11700

TTTGCGTGCC CCTCTGGCCC TCTGTGGAGA AGGAACATGA GATAGACCAA CCTAATCTTG 11760

TCTTAGGCAC CATGGGACAA TTAGCTCTGG TGTCTTTGTC CACATTTACA GCAGTGTCCT 11820

GACAGGGGTC AAGGTACTGG GACAAATATG GTTGGTGTAG CCACAATCCA GGTGGAGTCC 11880

TTCATTGTTG TGCCCATATC TTGGAGACCT TGGAGCAGGG CACTGCTATA TTTTGGAAAT 11940

CTCTGTCGTG ATATATTTAA GCATCTTAAA TGCCATATTC AGCAGGTCTC TGAAGTGCCT 12000

GAGGACCGTC TGTCTATCAG AAGTATATCT GAACAGACAA GCTTTGCTTG TTTCTCCTTA 12060

TCCATCTTAG GCTTAGCCTT GAAAATATAG ACAGATGGGT AGATAAAGTT TGAGGTTATC 12120

TGCTCTACAC TGTACCTGTG ATGTTAGAAT AAAGTACCTT TTAAAGAGTG AAGACATAGA 12180

GCATAAGTTT CAGAATACAA AACCCGTTTT AACAAGTGTT AACATGTAAT GTCTCCAAAA 12240

TTAATACCAA GTGTGTTATT ATATGTTTCA AGGTTTGGTT GTCAGGGCCA CCCAGCAAGG 12300

TATGTTGTGT TTGTTTGGTT GGTTTGATTT GGTTTAGTTT GGTTGGTTTT GTTTTATTTT 12360

GTTTTCTTGA ACCCATAACA GGGTTATCTG CCAGCCATGT TCTTGTCCTT AATCAAGATG 12420

GTGGTGAAGC CATGGTTCCA CCCTCTTTAT CCCATGGTAA AGATGGCAAC CACTTCGTAA 12480

TCTTTCTGGC AGCGGGTGAT TTAAGTAGCA GTCTGCTCCC TACCATCACT GTTGCCAGTG 12540

AGGCAGCACA TGGACCTGGG GGCTGTGGGT TGGGGGCGGG GCTAGGTGGC ACCTGGAGCT 12600

GTGTGTGGGG AGGTGGTCGG CTGCTGAAAC TGTTGTGGTC AAGCGGGCAG GTGAGAAGGC 12660

GGTTCACAAG AATTGGAATT ATGGAGTTCC GACCAATTTG GATAAAGCTT TCAATTTTAG 12720

ATGCGGTGGC ACCCATTTTT TTAGGCTTTG CCGGGGACCA AAGGCTCAAA TGTACATTGT 12780

CTGCACATAA GCCTGGGGTT ATACAGTTGG GGTTGGGGGA TTAGCCTCAG GTCAAATGTT 12840

TTTGAACCCA AGGCTCCCCC AGAGGCTGGC TTCTGACAAT TTGTTTTCCA AGGCTGCTGG 12900

GAGGTGGTTA CACCTGGAGT CCAAAGTCTG TCGCATCAGG GGCAGAGAAA GATGGGGGAA 12960

AGGAGTGGTC CCCTATTCTC ACGTGGCAGG GGCCTCAACT CCTCCAGAGT CAGGACACCA 13020

GATGATAAGT TTTTCTCCAC TGGTGAAGTG AGCCAATTCA AGCCAGAGTA AAGTCAATTA 13080

AGGCAGGTTT ATTAGGAAGA AGTTCTCAGG TGGGCTCACT GGTCCTAAAG TGGGGAAGGA 13140

GACAGGGGAG GAGGGAGAGA AAGAGAAAGA ACATGTGCAC ACAGGGAGAG AGAAAATGGA 13200

GAGGGGGCAG AAAGAGACCA AAATGTTTGG ATTATGTAGG AGAGCCTCTT GGGGACCGGA 13260

AGCCCGTGTA GGGACTGAGG AATGCTGGGA GAACCTGGAG GTCAGTTCCT GTCTTGGTAG 13320

ATAAAATATG CATATGAAAT ATGCACCTCA GCTGTTTGTC CCGGGTCTGA ATCCCAACAG 13380

GTACAAGTGG ATTTGACTAA TTACGAGTGG CTTGATAAAA TTCTGCTTTA TACTTAATCC 13440

AGGAACCAAC ACAAGACAGC AGCCCTTCGA GGTCTTTAGG CAGCTTGGAG GAGAACACAT 13500

GAGAGAAAGG TTTGTTCTTC CAGAATGAAA TTCGTGAGAG TTTAAAAACA AATGAGCCGT 13560

AGCCACCTCA AAGCATAAAG TCTTGTTTTC CCTTTTCAGA ATCCCAAGAG GTTTTGTTTT 13620

CTTTGAGAAG GTATTTCTGT CCAGCTGCTC CAATGACAGA GATACCTGCA CCTTTGTCCT 13680

ACTTCCAGAA TGCCCAGATG TCTGAGGACA GCCACTCCAG CAGCGCCATC CGGAGCCAGG 13740

TACAGTGTCA GTGCCGGCTG CTGCTGAGGC CGGGGCAGCT TTGCTCTACT TTGCCGAATG 13800

TCTTTTCTTG TTAAACATTA AGCTGAGGGG AAATGTCACA GCCTTGGTGC TAGGGGACCC 13860

CAAACTGCAC TTGTTTTACT CTGTTCTGCG GGGGGACTTA ACCTCATGTC ACCCCAACTG 13920

AAACTTTGCC CATCTGCATG ATATGTAAGG ACCGCATGCT CCACGTTATC TGTCTTGAAT 13980

GTACTCTTCT CTAGGAAGAA GTGAGTGCAT TGTTTTATGT TTTCTAGTGC TGGCCAGTTT 14040

CTGACAGTGC TCGGTAAGTA TTTAGTGACG GTGATTATAC AATGCAAGGT AAACTGCTGG 14100

GAGCTCAGGA GAGGGGTGGG ATACAGGGTA TATACACAAA CTCTTCACAG ACCACACACA 14160

CTAAATCTGT CTGTTGTAAC CCTGTCAGCT AAGCACTACT GACTGGAGGC AGACTCCACA 14220

GCAGGCCTAG GAGCTGTCTA GGGAATCGGT AGAATTTTCT GAGGTACAAC CAAGATGGAG 14280

GGAAACCTTG ACAAGGCCAG TCTCAGGTCT GCAAGCCTGC CTTGTGTAGT TTGTACTTGC 14340

TAGTGAAAGC TGTCATTAAA GGCAAGTCAC AAGTGAGGAT TAGGGGAGCC TGGAGAGATG 14400

GCTCAGGGGG TCAAAGCAAT TCCTGCTCTT CCAGAGGACC CAAGATAGAT CCATATGAGG 14460

TGGCTCATAA CTGCCTGTTA ACTCCAGCTC CAGAGGACCT CTTCTAGTCT CTGAGGGGTG 14520

TGTGTGTGTG TGTGTGTGTG TGTGTGCGCA CGTGCGTGCG CGCGCGTGTA TGCATCAAGA 14580

TTGAGCCTTA AGACTTTCTT TATACTCTGA AGATTTAGAG TTCCTCACTT TCCCATCAGT 14640

AAATTTTAGA TAATTTACTG AATAATCTCA GTACGCTCCT ACCTTAAATG TTTTGGGAGG 14700

ACCAGGAGTG GGAGGAGCCT GTAGCTAGAA TGTAAGACTT TGCATTTTAA AACTGTTATA 14760

ATCCTGAACA AGTAGGGTCA AGATTAAAGT GGTTACTCCT ACCACAAGTA TAATAAATAG 14820

AAATCATTTT GGAAAAGACT AACCAAATCG TTCTTAGTTG GCAGTCTCAT GTAAGTTTAA 14880

TCACCATTAT TTGTTGGTAG GTGACATAAT GAGGAGACCA TGTAAACCCT CTTTTGTTCC 14940

ACAGATCTCT TGCCTTTACC ATATCCTTGT AACCTCTTAA ACAGAGGCAA CCACAGGCTT 15000

CAGGAGTCCA GAAAATAGTG TCTTATCTTG GGCTGCTATG TATTAATCTT ACTAATGGTA 15060

AAGGGACTTG ATGTTTGATA GGTGATGTTT TTTAAAACCC AAAGATTTCT AAAACTGTTA 15120

TTGTACAATG TTCTAAGCAT CTAAAGCAGC GGTTCTCAAC GTTGCTAACG CTGCAACCCT 15180

TTAATACAGC TCTTCATGTC ATGGTGGCCC CAATCATAAC ATTATTTTCA TTTCTACTTC 15240

CTAACTGTAA TTCTGTTGCT GTTATGAATC ATAACATAAA TATCGGATAT GCAGGATCTC 15300

TGATGTGTGA CCCTTGTGAA AAGGTTGTCC AACCCCCAAA GGGGTCGCGA CCCCCAGCTT 15360

GAGAACCACC GATCTAGAGT GTAGCAAGCA AATACCCAAC TTTATTCAAG TTGGAGACTG 15420

TGTTATTTGG CATATAAGGC AATTTTTTTT CACCCCGCCC CGGCCCTATA AGGCAATTTT 15480

TATTAAGATT TTTCTGTGTC GGTGACTTTT TAAAAAAGAT TTATTATGTA TGTGAGTACA 15540

CTGTTGTTAT CTTCCGACAT ACCGGAAGAG GGCACAGGAT GTCATTACAG ATGATTGTGA 15600

GCCACCATGT GGTTGGTTGC TGGGATTTGA ACTCAGGACC TCTGGAAGAG CAGTCAGTGC 15660

CCTTAACCGC TGAGCCATCT CTCCAGCCCC AATGTTTCTG TGTTTTAAAA GATTTTCTAA 15720

TCAGGAATGG AAGTGATGAT GCATGACTTT AATCTTCATA CTCAGGCAAC AGAGGCAGGT 15780

GGATTCTGTG AGTCTGAGGC CAGGTTACAT GTTTGTCTGT ACGAGCATCT CATCACTAGA 15840

TGGAGTGAGA TCTGATGTGG GTGCTGGGAC TGAACACTGG TGAGAGTTCT TAACCCCTGA 15900

GCTTCTCTGC AGCCACAAGG GGCGTGCAGT CTGCCTGTGG GTAATTACTG AGAACGAGTG 15960

CGGTGCACTG AGGACTTCAG TGTACGTCGC TGCTGTCGGG ACTTCTTAAT GAAAAGCTCA 16020

CTGCTTGCTC TTTTTTAATT TTTTTCAGTG TTCTGGTAAC CATGGCTGTC CTGGAGCCTT 16080

TCATGTAGCC CACGTTGGCC TTTATCTTAC CATCATCCTC TTACCTCAGT CTCCCATGTG 16140

TTCAGATAAA GGCAGAGCTA CCATGCCTGT GAAAAACTTC TGCCTTTAAT TTCTCTTAAA 16200

ATAACACTAT GTCTAAACCA TTTTAGTAGT TAATAAGATT TTTAATTTTG GAAACCATGT 16260

ATTTTTAGTA TCAACTCATT TTAACAGTTT TTAAACACTT TTGTTTTCTC TTTGGTTTTT 16320

CGAGACAGGG TTTCTCTATG TAGCCCTGGC TGTCCTGGAA CTCACTCTGT AGTCCAGGCT 16380

GGCCATGAAC TCAGAAATCC AACCTGCCTC TGCCTCCCAA GTGCTGGGAT TAAAGGCGTG 16440

CGCCAACACT ACCCGACAAC ACTTTTTTCT TTTTTGTTTT CTAAGATCGG TTCTCTTTAT 16500

GTAGCCCTGG CTATACTAGA ACTCAAAGCA TATTTTACAA CTCCTTTAAA AAGATACTTA 16560

AACTTTATTT TGTTTATTTT TAAAAGAGAG TCCATCCTAC TGTGTAGACT GGACTCTGGC 16620

CCACCGGCCT TAGCTTCCCA ATGCTGGCGT GATGATCGAC TGCCACTAGG CCCAGTTTGT 16680

TTTGATGGGG CCTTGTTGAA TTGCCCAGAC TGGCCTCAAT TTCAAGATCT TCCTGCCTCA 16740

ACCAGAAAGT ATTAGCATGT ACTACTGTAC CTCGCTGACA GTCTTTAAGA ATGGCTCCAT 16800

TGTGGAATAA TTGCCATATG GTTAACTATA CATAGAGGAT TCAATGTGAT GACTTTTTTT 16860

TTTTTTTTAA TTTTCAAGAT AAGGTTTCTC TGTATAACAG TCCTATCTGT CCTGGAATTC 16920

ACTCTGTAGA CTAGGCTGGC CTCGAACTCA CAGAGACCCA CCTGCCTCCC AAGCACTGGG 16980

ATTAAAAGGT GTGCACCACC ACTTCTGGCT CAATGCGATG ACTTATAGCA TGCATAAATA 17040

CTCGTAAAAA TCCATAAAAT TCTTTGTAGT GAAGAAAAGA GCAGGCCTCA TTACCCCCAA 17100

GTTGCCCATG TTCCCTTTGT GCTGTCCCCA TGCCTTTCCC CACCCCTAGC AACCACCCTC 17160

AGGTCCTTTT TGTCACCTCC CTCTGGGGTT TACTCCTTCC AGAAAGTTTA TATAAATAGA 17220

GCCATGCTGG GCTGGAGAGA TGGCTCAGCC ATCAAAGGCG AGGCTCACTC AAAATTAAAT 17280

AGAATCTTAC AGACTTTTTT TGTCTGGCTT CTTTAATTAG CATAATTATT TTGAAATTGA 17340

TGGGTGTTAT GTCAGTAGTC TTGTTTTATT TCTGAGTTGT AGCCGGTTAT ATGCTTATGC 17400

CATAGTTCTT TGGCCAGTTC CTTGTTTCAA TATTTGGGGG AGGTTTCAAT TTGACATATT 17460

ACAAATATTC AGTTACAAAA CTTCAAATGG TCATGTGCTC ATTTATTTCG GGTAAATGCC 17520

TAAGAGTAGA AAATAACCCA GTGTTGCAGG CTGTGGTGAG CTTCTTGAGA TAATACCACT 17580

TTTCCAAATT GGTTGAATAA CTGCATTCTT GGATTCTTGA TATGCCTCGC AGCCACCCAC 17640

CTCGTCAGCC CTCAAGGGGT CAGTCCCAAG CTTTCAAACC TTTTAGAATG TATGTGCTAG 17700

CATTTCTTAG TCTTTAATTT GCATATCTCT AATGAGCAAT GGTGTTGAAT GTTTTTGTTT 17760

ATCACGCTGA TGAACCTTCT CCAGTGAGAT GTCTGGTGCA GTGTGCTGCC CATTTTCAAA 17820

TTGTGAGGAC TCTTTATGTT CTGGATATGT AAGTTATAAT ACATATACCT TGCAAATCAT 17880

TGAATATCAG TGAAGTCCTT TTTGTCATTT TTTTTCTCTT GGAAATGTGG GGTTTCTTTT 17940

TTTGTGAAGT TATCTATGAA ATCTTATCCT GTTTTTTAAG TTTCTTTCCA AATCTTTTTT 18000

TTTTTTTTTG GTAGTTATAT GCTTTATAAT TTATCTGTTA TTGGGTATCA TGACACACAC 18060

TTGTTATCTC AACTTCTTAT TTGGGAAGCT GGGGCAAACC CAGCTCACAG CCCAGGCTCA 18120

AAGACAACTT TCATTTAGAT CTGTGGTCTG TTTTAAGTTG ATTTTGTGTA TGTAACACAA 18180

TATGGGTCAG AGTTTGTTTC TCTGTGGATG TCCTTGATCA GCATTGCGCT TGAAGAAGAC 18240

TGTTGTTTGC AGTGAAGTGT CTTTGTGCCC TGGCACACAC TGGTTCACAG ATCTGGCCTC 18300

GATTCTGTCC CACTGATCAG TTAGTGCTTG TACCAGTGAC ACATTCCTGC TATTACTGTA 18360

ATTGTATATT AGGTCTTGGT ATCACTGTTT AACAAGGTGG TGGTTATTTT GGTTTTGGTT 18420

CTTAATAAAA TAGTTGGGCC TGGTCGTACA CATCTTTAGT CCCAACAGCA GGAGGCAGGA 18480

AGCAGATCTC TGAGTTCAAG GTCAGCTGGT CTATTTACTT TGTATATTTA AGTCCTAAGT 18540

GGTTTTGCCA ATCACTTTGC AAACCTTTCT TTTAAAAGGC CAACTGGCAG TTTGTGATTG 18600

CATTGATTCC CTTCCTATGT TGTGTAGAGG ATTTTTGGAA CTGATATTTN CTTTTTTTTT 18660

TTTTTTTGGT TTTTCGAGAC AGGGTTTCTC TGTATAGTCC TGGCTGTCCT GGAACTCACT 18720

TTGTAGACCA GGCTGGCCTT GAACTCAGAA ATCTGCGTGC CTCTGCCTCC CGAGTGCTGG 18780

GATTAAAGGC ATGCGCCACC ACACCCGGCG GAACTGATAT TTTCATAATG CTGGTCTTAT 18840

TCCATGAATA GTATAATTCT ATATTTATTT AGATCTTTTA ACTAAAATTC TTGATTGTGG 18900

TAAATAACAC AAAGTTTAAA GTTTAAAGTT TAAAGTTACC ATCTCATCCA TTTTGTTTCT 18960

GTTTTTGTTG TTGTTGTTGG ATTTTGTTTG TTTTTGTTTT TCAAGACAGC ATTTCTCTGT 19020

AGTCTTGGCT GCCCTGGAAC CCATTCTATA GACCAGGCTG GCCTCAAACT TACAGAGATC 19080

CTCTGCCTTC CCATTGCCAG GATTAAAGGC ACATGCCATC ACTGCCTGAC TCATCATAAC 19140

CAGTTTTTAA CTCATAGTTT AATATTAACC ACATTCACAG TGTGCTCTAG AACCATTGAA 19200

CTGCCCTGCC ATCTGCACGA TGAGTTATCT TTTCTGGTTC TTAGTGTGAC ACTCTAGATC 19260

TTGCCTTCTC CTGGTTTGCT TACTTCGCTC AGTGCCGTGC ACTCAAGAGT CACCCATGTC 19320

ATAGTGTGAA CAGGATTTCT TGTCAACATT TTTATTTATT TATTTATTTT TGGTTTTTCG 19380

AGACAGGGTT TCTCTGTATA GTCCTGGCTG TCCGGGAACT CACTTTGTAG ACCAGCCTGG 19440

CCTCAGAACT CAGAAATCTG CCTGCCTCTG CCTCCCGAGT GCTGGGATTA AAGGCGTGCG 19500

CCACCACACC CGGCTCTTGT CACATTGTTT TCCTTCCTAC TCCCTCCTGT CTTCTTTCCA 19560

ACATAGCCAG GACTACAGGT GTGCACACCG TGTCTGCCTA GGAGTTCTTT AAGGCTGAGG 19620

AATTCTCTAT TGCATATTTA TAATGCTTTT TTTTTTTTGG TATCCATTGT CAATGGATGT 19680

TTGAGTTGTT GGACTTCTGC CACCTGGCTA TAAAGGATAG TGTGCTTCTG TGAGTATGGC 19740

TATGCACACT TCTCTTTGAA ATCCTCTTTG CAATTAANTA ATTTAGTAAA TACACTCAAA 19800

GTGGGATCAC TANATCATAT GCTCTTTTTA ATTATTTTGA GCAGGGTCTT GGTTTGCCCC 19860

AAACTTAATG TCCTTCTGCC TCAATGTCCT GAGTTGCTGG GAATGCAGGT TGTGCCACCA 19920

TGGCTAACCC CACCTACCTC AGGGAATGTT GCTCTCCTTA GTGGTTGTAT CATTTTATGT 19980

ACCAGAAAAC AAAGTATGAG GGCCACAGTT TCTTTACATC ACCACTTGTC ATTTCCTCTT 20040

TATTTTCCTA GAGCTGCGTC TCTAACTTGG CAAGCCCTGT AGACTGAGCT ACATCCAGCC 20100

AGTCTTTGGT AGGGGTGGTT TTGATGGTGG CTATCCTAGC GTGAGATAAG CTTTGATATG 20160

TCCCAGAGAG AATCTGCCAA GAGTTTCTTG CATTTGTAGA TTCCCTGTTC TTTGCTGCTT 20220

GTTCCTTTGA TGCACTTTGT GGGTGGTGTA GTTGTTTGTT TTGCTTTTGT TCCTTAATTT 20280

CTCTTAGCAC CGTTTTTGTA ATTTCGTGAG GACAAGACAT TTGGGACAGG ATCCCTAGAC 20340

TATCTTGGTT CTGTAATTCA GTGGGAAGCC TCATAGACCA CAGGAAAATG ATAGTGTTCA 20400

CTATTACAGT TGATTGCAGT GAAAGCATAC AGATTATTGC CAGGCGTGGT GGCACATTCC 20460

TTTAATCCCA GCACTTGGGA GGCAGAGGCA GGTGGATTTC TGAGTTCGAG GCCAGTCTGG 20520

TCTACAGAAT GAGTTCCAGG ACAGCCAAGT ATACACAAAG AAACCCTATC TTGAGAAACC 20580

AAAAAAAAAA AAAAAAAAAT ACAGATTATC ATTGGTAATA GAAAAAGGCA CTTCAGGCAG 20640

GCCAGAGGGC ACCTGCTTGA GATTCCAGTG GATTTATGCA AACAGTGCTT AATTCTCCTG 20700

GCTACAGCAT GTGACAACAC CCTTGCAGTA TTTTCAACCA GGGGAACAAA CTCAAGCCTC 20760

AATGGCCAAG GTTTTTACTG TTAGCGGTCT ATACAAACAT GGTTGGCCTT TGTTTTCTAC 20820

CCCCTTTAGG GGCCAAGCTA ACTCTGACCT CCTCCGAGAT CACAGTCTAG GTTGCTGGTG 20880

TGTCCCAAGG CATCTGTAAA CAGCCCAATA GGCTACTCCA AGGGCTTGCA GACCCCTCCT 20940

GGGGTCTGAG CAGCGCAGAA GCTTCTTGGC AGGGGCCGGG TCTAGGAGAC CCTAAGACAC 21000

AGAGTTCGTA GCTCTACTGC TCAGTGGGAC CGTAAATCTT TCTAACCACC CGGCTTGGCT 21060

TTCTGATGTC TTTGCGGTAA TGCCCAAAAG GAATCACATC CCAGGTGTTC CTTTACCACC 21120

TGTTTATTCT CTGCAGAGTG GTCCCAGACT GTAGTAATTG TGCTTATTTG TGTCCCTTGC 21180

CCCAGAGTAC TCAGGAGTCC AGACCTTGTC TGTCTCACTT CACTGCTCAG TTCCTGGCTC 21240

CCAGTGCCCT GTGTTCAAGT GACTGAAATT CAGCTCCGCC CCCCCCCAAC TCCCCACCTC 21300

CCACCAACAC ACACTGTAGC AAATGCAGTA ATCAGTACTT AGAAACTGAA ACAAGTTTAT 21360

GTAACCAGGG AATGGCAAAA GACACGAATA TAGTTTAATT TTTAAGTTTT TTTTATATGT 21420

ATGAGAACAG GCAGTTCTGA GCTACCATGA AAGTTCTAGG AACTAAGCCT GTGTCCTATG 21480

CAAGAGCCCT AAGTGTTCAG TAAAAATTAA AATTACACTG AGACCTTGTA TCTCACCTAC 21540

CAGATTGGTA AGAAATCCAG GCATTAGCTC CCATGATTGT GAAGTTAACA GTTTTCTACA 21600

CTGGCAAAAC TCTAAACTGG CGCAATCTCT CAGAAATGAA GGCAGGTTGG CAAATGTAAT 21660

AGAATTTCAA ATATACCTTG ATTCAGCAGT TATTCCTCAG AGAACTTGCT CTACAACTGC 21720

ATTGGGTAAG TATATACGTG TCGCTAAGTG ACTAAGGAGG CCTGCCCTGC AGTGTTGTTT 21780

GTAAGGCAAA AGGTTGGAAG CAAATTAGAT TTCCACTGTT AGGGATCAGG CAAATAAGCA 21840

CAGTGAATTT CTACAGTGGC TGTGTGACTG CCTAGAAAGC TAAGAAATAG AAACATGAAG 21900

AGACAGGCAC AGGGACACAG GACCTATAGT CCCAGCTCCT TCAGAGGCTG AGACAGGATT 21960

GGGCATAAGT TCAGGACTAA CTTGGAGGAC AGAGTAGCAA GCCTGTCTCA AGTTGAAAAA 22020

AGAAAAAGAA GATGTCTTTG TTTTGGTGTG TAAGAACCTT TGAGTGCAGG AGCAAAGCAC 22080

AGCCTTGTGT AGCGGGGGCG GTGGGTTCCA GCTCGGGTCA GTAGCAATCA GAAGTTGGGT 22140

TTTACGAGAC AGCAGCTCAG GAGCAGGGAG ACCTTAAACC TTCAGAAGAA AATGAGGAAC 22200

TGTTGTACAC TTTTTCATAC TTTCTTTCTT TAAAATCATT TAACCTTTAA AATAAAGTTG 22260

TACCTTTTTA AAGTTGGAAA CACAAAGAAT AAGTAAGGGG TAAGAGAAAG AGGTAGTGAG 22320

TGTGAGACCT GTGTGAGAGA AAACCACTCG CTGCTAAGGA GGCCACCCGC TCCTAGACCA 22380

TTTTTTGCTT AGGACAGGGC AGGGTGGGGT CATGCTCTGT ACAAACCAGA ATATCCAGTG 22440

TAGGAGCGAT CTATTCTGGC CTAATTAACT GTTCATACAG ACAGATGCCA GGGCTGTGGC 22500

ACCCAGAAAG TCACACCCAT ATTTAGTTCA TGGTAAACAT TAATTCATCA TTTAGTTAGA 22560

ATTGTATTGC TAAACGCAGA CATCACAGTT TTAGCTGTTG ACATGATCTC TTCCTTCTCT 22620

GGGAAAATAG AAGCCTAGTT TATTTCTGGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT 22680

GTGTGTGTGT ATGGGTGTGC TTTAGTCTTT ATGACTTTAT ATTGCTCACA GGATAAAGTC 22740

TCAAGTCCTT GATAGGGCAA ATAAAGTTTT TTATACCTGT CCCTAAAAGG ACCTTTTTAA 22800

AAGAGCAGAG TTTCTAGAGC CAGGACCCTT TTCTGTGGGA CTGCAGTTGG AGGCAATGCT 22860

TTACCTCTGA GCTACATACC CTTTTTAACA TTTTATTCTG AGACCAGGGA CTAGCACATT 22920

GTCTTCAAAC TTGTACTTTT TCCTGCCTCA GCTCCTGAAT TCTGAGAACT GTAGGTGTGC 22980

CCGCCACACC CAGCTCAGAG TCCAGAGCTT ACTGATGTCA CTATTGCTAC TGCAGAGTGT 23040

CTGGGATACA GTAGAAGTCT CCTGACTACT TGTGGGACCG TGCCTTAGGT ACTTTTCTGT 23100

TGCTGTGATA AAACACCATG ACCAAGGCAA ATTATAGGAG GGCTTTATTT GGGTTCATTC 23160

TTACAGGGGC AGCATAAACG ATCACCATCA CCCATCATGT CAGGGAAGTT ATGGCAGCAG 23220

ACAGGCAGAG TGGCTGAAGC AGGAAGTAAT AGAGACAGAC AGACAGACAT GCAAACACAC 23280

ACACACACAC ACACACAGCA TTGGGAATGG AGAGGCTTTT GAAATCACAA AGCCTGCGAC 23340

CACACCTCTT AAACCTCCCA AATAGCATCA TCAGTTGAGG ACTGGATATG CAAACATTTA 23400

AGCCTATGGG GAGCATTTCC ACCCCCAAAA ATGGGTCTGC TGCAATTGAA ATTAAAAAAA 23460

CCCTTACCCC CCTCCGATTC TCCCAAGCTC CCATACTGTT TGGGAAATGA TAAATTAAGG 23520

AAAACATTTT TCTGGAAGCA GGCTGTCCTG GGTTCAAACC CTCTGACATT CACTAGCTAT 23580

GTCTATTTCT TCACCTGTTA AGGGATGGCA ATAGAGTGAA AAATCAAAGG ATATGCACAT 23640

GTCATAGTCC CTGTTTTATA GAAATAATTT TATGCTGTTA CTGAATGTGG AGTGTATTAG 23700

TATTGTGTGC GTGTGTGCGC ACGCGCTCAT GTGTCTGGAG GTCCGAGGAC AGCGTTGTGG 23760

AGTCTTTCTC CTCCCACCTT TATATGGGTT TCAGAGATGA AGCTGGGGTT GTCAGACTTG 23820

GACCGATGGG TGGCAGGCAC CGTTATTTGC TGAGCCATGT GCCCATGGAG TGTGCTGCTA 23880

ATGAGTGGCG GTAAGACTCC ACAGTAATAG TCACACTCTG TAGTTGAAAG CTGGGCAGTG 23940

GTGGCACACG CCTTTAATCC CAACAATTGG GAGGCAAAGG CAGGTGAATT TTTGAGTTTG 24000

AGGCCAGCCT GGTCTACAGA GTGAGTTCTA GGACAGCCAG GGCTATACAG AGAAACCCTG 24060

TTTTGGAAAA CCAAACCAAA ACCAAAAAAA TAAGAGAGAG AGAGATTGAT TTGAAGGCAG 24120

GTTGTGTTTT CAATCACCTT CCTGAAAGAA ATTAGCTTAA TCAGGTACTG TTTACCAACA 24180

GTAACTATTA AACACTGGAA GTTGGCTTTC CTTCCATCAT GTGGTTTCAG GGAATTGATT 24240

GTTAGGTTTG CCAACAAGTC CCTTTGCCCT CTGAGCCGTC TCACCAGCTC TGTTACTGTT 24300

CTTTTTATTC ATTTTACCAT TGGGNAAATT TGTTTTAACA AATGAATTTA AAAATACCAA 24360

AATAAATTTA AAAAATGAGA AAAAGCCAGG GGCCAATGAG ATGGATCATT GGATAAGGTC 24420

ACATGCCTTC ACATATCATC ATATGTACCC AAGTATCTGT ACACACACAC ACACACACAC 24480

ACACACCACC ACCACCACCA CCACCACCAC CACCACCACC ACCACCCACA CATACACTTA 24540

AAAACACAGT AAAGCCGGGA ATTGAGACCT AGTTTGGTAA TGGAAAAGCT CTTGCTGAGC 24600

ATGTATGGAG CTCTGGGGTC AGTTTTTAGC ACGGGGACAG GAAGCTAAGC ACAGCAAAAA 24660

TGAGCATAAC AGCTCCTGTG TGAGCTGCTG AAACACGTGA CGGGGTCAGC CCTGCAAAGC 24720

CAAAGCTGCA GGACTTCATG ACACAGGGCG ACAACAGGAT GTTTAAGAGG ACTCCCAGTG 24780

AGCGTCCAGT GTTGATAGTG TAGCAGAAGC CAGAGGCCTC GAGCCAGACA AAGACCGTGC 24840

AATGAACATT TGCAAGTGAA GCTGTTGGGG CAAAGGGGTA TATATACTGT GGGACACTCC 24900

ACAGCGTCCC TGGGAGAGTC TTTGTTTTAT TTTCTTTTGC AGGAGTTAGG GTCCCTCTGG 24960

GTAGTTCTGG CTGGCCTGGA ACTCACTACA AACAGGCTGA CATTAAACTC CCAAATGTCA 25020

GGATTAAAGA TGTGGCTACC ATGCCTGGCT GGGAGATTTT TTTTTTTTTT AACTAAGCAA 25080

TTTTAAAAAG TGATTGTAGT CTTAAAACAA TTTATTTATT TTATGCATAT GAATACCATG 25140

TAGCTGTCTT CAGACACATC AGAAGAGGGC ATCAGATCCC ATTACAGATG GTTGTGAGCA 25200

ACCATGTGGT TGCTGGGAAC TGAGCTCAGG AGAGCAGTCA GTGCTTTTAA CCACTGAGCC 25260

ATCTCTCCAG CCCCATAATT GTAGTCTTAC ATGGTGTTGT GAGAAAGAAT ATGGAGAGAT 25320

ATCTTGAACA CCACTATATG CTTTTTAAAG TACTATTAAT AATGAGTCCC AGCACTCGGG 25380

AGGCAGAGGC AGGTAGATCT TGGAGTTCGA GGCCAGCCTG GTCTACAGAA TGAGACTATA 25440

GCCAGGACAG CTTGGAATAC ATGAGAAACC TTGTCTGGGA AAATTAAAAT AACAACAACA 25500

ATACTAAAAT CAGGTAGGGA GTACTTCATG AATTTGAAAA GATGAGATTA CTAAGTTTGT 25560

AGCTGCATGA AGAACATTTG GAGGCGGAAA GTTGGCTCTG TGGTGGTTAA GAGCACTTGC 25620

TCTTCCAGAG GACCCAAGTT CAATTCCCAA TTCCTACATG GCAGCTTACA ACCATCTTTA 25680

ACTCCAGTTC CAGGGGATCA TCTGATGCCC TCTTCTGGCC TCTGTAGACA TAGGTGTATG 25740

TGTGGTGCAT AGAAATGTAT GGAGGAAAAA CACCTATACA CAAAATAAAG AAATCATAAA 25800

GAAGAAGTGT TGGTACAAGT ATATAGTATT TAAGCCAAGC CTCCAAAGGG GGGAATGGTT 25860

GCCTGAGGAG GAGAATAGAA GGAGGGGTGA AGTGAGCCAG GACCCAGGTA AGCGGGAGGG 25920

AGGGCCTGGC TTCCTCGGGG ACAGCAGTAA ATAATAATCC AGACATTAGC ACTGTCTGTA 25980

AGGAGTCCTG GTGGAGCCGT TACCACACAT CCCAGGTGTA TGTCTTCCAG GATGCCTTGT 26040

GCTGAGAGTC CGGCTTACTG AATGCATTTC TCGTTCTTTA GAATGACAGC CAAGAACGGC 26100

AGCAGCAGCA TGACAGGCAG AGACTTGACA ACCCTGAGCC AATATCTAAT GGGCGGCCCC 26160

AGAGTAACTC AAGACAGGTG GTGGAACAAG ATGAGGAGGA AGACGAAGAG CTGACATTGA 26220

AATATGGAGC CAAGCATGTC ATCATGCTCT TTGTCCCCGT GACCCTCTGC ATGGTCGTCG 26280

TCGTGGCCAC CATCAAATCA GTCAGCTTCT ATACCCGGAA GGACGGTCAG CTGTATGTAT 26340

AAGTGCTTCG TTCTCTGGGC TGGTGTGGCT TTTTCCTCGT AGCTTGTTGT TATCCTCTCA 26400

AAGATCTGTG GACCATCTTC CATCTTTCCT GATGGCCAGA GGCAGTGGGG ATGATGGCAG 26460

TGATGATGTG ATGTCTTGGT TTTGTTGTTG TTGTTGTTTT AGTTCTAACC TTTTGAATTC 26520

TTACTTGTGA GCCAGCAGAC ATGGTAGCAG GGCCTGAAAT CCTAGCACTT AGCAAATGGC 26580

TCATGGAAGT TTAGAGGTTT AGGAACCATG GCTTTGGGGG ACCACCATGT GGTGATTGAG 26640

ACACAAGCAT GATGGCATAT GCCTAATCAC TTGGCAGTGA GACGGGTCAG TCAGGAACTC 26700

AAGTCCATCT TGTGCTGATG CAGATTCCAG GCCAGGCTGA GCGACACAGC AAAGCAACCA 26760

CATTCCTATG GTCCCTATCT TGGCTTGTGT GTTAAGTCTT TAAAAACAAC TGCCTTCATC 26820

TACTGGGAGT AGATCCTAGT ATTTCTCATA ACAGCCTGTT TTCACTGTTT CTCAGGCTAG 26880

TCTGGAATCC ACTGTGTAGT CTTGGCTAGT CTCAACTTTC TGATCCTGCC TCAGCCTCCT 26940

GAGTGTGAGG ATTACAGGTG TGTGCCACCA TACCCTGTTT TACGAGACTG TTATAGAAAC 27000

TCAAATGACA TTCTCTTTCT CCCACAGAGC AGTGCAGGGG TAGGCAGTCC AGATTGGTAT 27060

CCGGGTCTTT TATTCAACAA AGCATGTGTG CTCCTCTCTT CAGCTTAGGT CCTAGATTTC 27120

CTTAGGGTCC AGACTAGAGC CTCAGTTTTC ATTTGACAGC AGGAGGGAGT TGAGAAGGGG 27180

ACATAGGCAT CATGTAGCCG AGAGCACTCC CTGCAGGAAC TACCTCTATA GCTTGTTCAG 27240

AGTTCCTTGG TCACTATGCA GTCTCATGGC CACACCTGCT GTAAGGGCAG AAGTAAGGTC 27300

ATTGTGCTGG AAGCCTTGGC ACTTGGAATT AAGGCATTTT TATGTGGGGC TGTCGAGATG 27360

GCTCAGCAGT TAAGAGCACT GACTGCTCCT CTAGAGGTCC TGAGTTCAAT TCCCAGCAAC 27420

CACATGGTGG CTCACATCCT TCTGTAATGG GATCTGATGC ACTCTTCTGG TGTGTCTGAA 27480

GACAGCTACA GTGTACTCAT ATAAATAAAA TAAATATTTT TAAAAAAGGT ATTTTTATGT 27540

AAATGAAATA CGAGAAGACA ACTCCAATAA TGGTCACCCT AGAGATAAAA AACCATTGCT 27600

TAAACCCTGT GACCTCTCTC TAATAACTGT TGTCCAGCCT TAGTAGTTAC TGTGGCAGCA 27660

TCTGGTTATT TAACAGAGTC TATTAAGCTA TGGAGAAAAC CTTGGGTGAG ATCATTATTT 27720

CAGAACAAAG CATCCACTAG ATGGAGCCAG TGCCTACCTG TGAGTTCATC TCTTCATAGT 27780

CTATTTCCTC TTACTCAGAA GACGTTATTT TAGAGGATAG CTATTTTTCT TAAGCTAAAA 27840

ACTATTGGGT ACAGGGTTGG AGAAATGGTT TAGCAGTTAA GAGGGTGCAC CACTCTTGTG 27900

GATGTACTGA GGGGTCCATA ACACCCACGT CAGAGAGCTC ATAACCAGCT TCAGGGATTG 27960

GGTGCCTCTG CCCTTCATGA ACATCTGCCC ACATTCACGC AATGCACACA CACACAATTT 28020

AAAATAAAAA TATATGCCTA CATTTAAAAA TATTATTTGA GTGCTGCTTC ACCCCTCCTC 28080

CCTCTCTTTA GCCCTCACCT TTCTTACTCT CTAGCCCTCC TTCTCTCTCT CCCTTCCCCT 28140

CTCTCTCCAC ATGGCCATGG CCGGCCTCTC TCTCTCTTTC TACCTTCTCT CTTTCCCCTG 28200

CATTTCTACA ACAAAGCTCT AAAACCATTT AAAAAAAAAT ATTACCTAGA GCTAGGGTGG 28260

TACAGACCAC AATCTCCACA CTTGGAATTC AAAGGCAGGA AGATCACGTG TTCAAGTACA 28320

GCCTGGGCTA CATGGTGAGT TCCAAGCTAG CTTGGGGTTT CTTATTAGAT TCTGTGTCAA 28380

AAACTAACCA ACCAAATAAA AATATCCTTG GTTGACTGAT GACCTGCAAA AGGCACCTTC 28440

TAAAAGACTC ATTAAACCGC AGACCTCACA GTGAATTACA GTTTTAAGAG TTGTGCTGTT 28500

TGACATATGC ATTAACACTC TGGTGTTTTG TTTCCCTCTG TAGAATCTAC ACCCCATTCA 28560

CAGAAGACAC TGAGACTGTA GGCCAAAGAG CCCTGCACTC GATCCTGAAT GCGGCCATCA 28620

TGATCAGTGT CATTGTCATT ATGACCATCC TCCTGGTGGT CCTGTATAAA TACAGGTGCT 28680

ACAAGGTGAG CACGAGGCAG TCTGCTGTCC GCCTTGACCT GCCTATGGGT TGTCCTTGTG 28740

TGTGTGGGAG GGGAAGGGTG GGAGAGTTGG GGGCGGGGGA GATGTGTATG CAGGCCTGTG 28800

CATGTCAGAG TGTGCGTGTG CAGTTGGTAN GACCACTTTG TGGAATGCAT CTTTTCCTTC 28860

CATCTTTGCT TGGGTTTTGG ATATCAAATC TCCAAGCTTG TAAGCGTCAC GTTCTGAGCC 28920

CCTTCACAGC CTCCTTGTCT TGACTTGGTC TAACAAAGAG GAAGTGTCTT ATATTTTGGA 28980

TCTAAGTTAT TTTTCCTTTT ATTCTGAAGC ATGTTGAAAG TACCAGAAGA CATACATGGA 29040

AAACAGGTGT TTTCATGTAA GGTCACATGT TTTCAGTATT TTGATTTATA GTGTGCCTAT 29100

GGCATTCCAA AGACTTTTGT ATATATTGTC CTTAACAAAT AGTGCTATAG AATTTAAGTG 29160

CCCCATAAGC CACCTTCAGA TGCTGCTACT GAACGCAGTA GTGACTGGCC CCTCACTTGA 29220

GAACCAGCAT CATTTGCTTA GTCTCTCAGA GCAGCCCTGC TTCTTACCTT GGTCAATGGG 29280

CACAGGGAGA AGTTGCAGCA GAGATGCAGG CCCAGCCTGT CCCCTCTCTC CCTCTCCTCC 29340

CCTCTCTTCT CCCTCTCTTC TCCCCCCTCC TTCCCCCTCC TCCTCCCCTC CCTCCCTCCC 29400

TCCCTCCCTC CCTCAGTAAG CAACTGCGTC CTGGACCTCA GCCACCCAGC ACATGCTTCA 29460

TACTCAAACC TCAGAAAATC AACAAATCAA TACTGGTCAT GGAATCTTCA GGTCGTCCTC 29520

AGTAGTAGGA ACAGTAAATG TTAACTCTCC GTATTGAAGT TTACTCTCTG TCTACATTAT 29580

TTTCTCCTAG ATTTTTCTTT GGGGTTTGTT TTGTTTTGTT TTGTTTTGTT TTTTGAGACA 29640

GGGTTTCTCT GTGTAGCCCT GGCTGGCCTC AAACTCACTC TGTAGACCAG GCTGGTCTCA 29700

AACTCAAAAG ATTCACCTGC CTCTGCCTCC CCAGCGTTGA GATTAAAGAT CTGCGCCGCC 29760

ACCACCTGAC AAATAGCGTG TTGTGCAGCG GTTTTCTCCG CACGTCTCAC TGCACCTGTG 29820

TCTGTCTGCT TGCCATCCCT CCCTCCCTCC CTCCCTCAGT AAGCAACTGC GTCCTGGACC 29880

TCAGCCACCC AGCACATGCT TCATACTCAA ACCTCAGAAA ATCAACAAAT CAATACTGGT 29940

CATGGAATCT TCAGGTCGTC CTCAGTAGTA GGAACAGTAA ATGTTAACTC TCCGTATTGA 30000

AGTTTACTCT CTGTCTACAT TATTTTCTCC TAGATTTTTC TTTGGGGTTT GTTTTGTTTT 30060

GTTTTGTTTT GTTTTTTGAG ACAGGGTTTC TCTGTGTAGC CCCTGGCTGG CCTCAAACTC 30120

ACTCTGTAGA CCAGGCTGGT CTCAAACTCA AAAGATTCAC CTGCCTCTGG CCTCCCCAGC 30180

GTTGAGATTA AAGATCTGCG CCGCCACCAC CTGACAAATA GCGTGTTGTG CAGCGGTTTT 30240

CTCCGCACGT CTCACTGCAC CTGTGTCTGT CTGCTTGCCA GTCTTTACCT CCTGGTCCCT 30300

TTCTTGGCCT TAGTGGACAC ATCCTGGACC TCAGCTGATT AGGGACCAGG CTTGTTCTTT 30360

TGCCTCCTTT TACGATTTTA TTGGGGGTAG AATTCGCGTT CAGAATTTGG CCTTGATTGA 30420

CTGACAGGCA TTATGTAAAA TTCCTCTCAA ACCTAGTCCT TCTTTCCTAA AGGCCATGGT 30480

CCTTGGCGCC CCCATGGAAT CTGTTTATCT GCTTGGCAAT ACTTCCCTAG TTCTATCCGC 30540

TGCCTTTGCA AACTTGGTTT TTCTGGATCC TCTCGCCATT GGCTTGTTAG TGTGCGTTCT 30600

GACCTTGGTT CCTTTCCTCC TCTGCAGTTC CCCTCAGGCT CACAACCCAG TCCCAGACTC 30660

TGGAGGCTTT CCCCTCCTTT GTTTCTCAAA AAGCCCTGAT TAGGGCTGGA GAAATGGTTT 30720

ATCCGTAGGA GCACTGGCTG CTCTTGCAAA GGACATGAAA CCGATGAATT TAGTTTCCTA 30780

CTCCCACGTC TTGGCTCATA TTCGTTCCTC ACTCTGTTTC CAGGGGATCC AACAACAACC 30840

TCTTTTGGCT CCTGGGGTGG CAGGCACAAA TGGCACTTAC TGCATACATA TACACACATG 30900

CACAGGGAAA CCATGCATAC ACATAAGGAA AAAAAAAAAC CTTTGAAACA AAATCACTTA 30960

CTCTCTGGAT TAATTTTGTA CCCTCAAATC TAACCTTAAA TTTTTTTAAA ATTAAGTTTA 31020

TATAGTGTAT GTCATGTGGA CATGTATGTT CCATGCTTCA TATGTAAAAT TCAAAGGTGA 31080

TTTTGCAGGA TTCATTTCTC TCCTTCCATG TTGTGTGACA AACTTTCCCT GCAGCACACA 31140

TCTGCTCGCC AAAAAAAATG AGCGTACCAC TCAGTAGCAA ATTAAAAACA ACTGACCCGG 31200

GGTCTGACAA GCGAGCCAGT GAGGCAAAGG ACAAACTTTT TCCTTTCCTT TTTTCCTTTT 31260

TGGAAACAGG GTTTCTCTGT GTACCCCTGG CTGTCCTCGA ACTCACTCTG TAAACCAGGC 31320

TGGCCTCAAA CTCAAAAATC TGCCTGCCTC TGCCTCCTGA GTGCTGAAAT TAAAGGCGTG 31380

CGCTAACACT GTCTGGCAAG GATATACTTA CACTGAAAGG TTGCTTACAA GGAGCAGGGA 31440

TGATTCAAAA GCAGCTCTGT CCCTGAGAGG CCACTCCAAA AAAAAAAAAA AAAAGCTGCC 31500

ACCCTGGAGC TCCCTGCATG GCTGTGGGCA GCTCCCAGCT ACCTTTCTGC ATTTGGGCTG 31560

CTAGAATCTC CACCCCAAGC GGTCGTTTGC TCCTTTGTAT GTTGTTGCCG ACTGGGCCGC 31620

AGGAATCTTG TGAGGTTCTG TTCCCTCAGC TATGTGGCCT TTGATTTCGT TTGAGGCTTC 31680

TAAAGGGAAT GTCTCAGATC AGAGGAAAAT AAATGTCATC CAAGAGGACC TGCCTCAGTT 31740

GTCACGCCTG CCAGCAAGTG CCTTCACCAA CAAAGCTATC TTGCTGTCCT GAAGACAACT 31800

```
TTGATTCAGT GCTTTCAGCC TAAAAGTCAT TGCTTATTGG ACCCATGGTT TGTTGGTTGA   31860

TTTATTAATT CTACTTTGAT ATCTACAAAC ACATATCCTC CATCCCCTTA TCTTCAATAG   31920

TCTGCTTTGC TAACTTTTTT TTAAACTTTA TTTACTTATT TATTTAATTC ATTTTACAGT   31980

CTAATTGCTG CCCTCCCCTC ACACAGTCCC CTTCCCCCTT CTCCTCTGAG AGGGTGGAGG   32040

CCCCCATTGG GTCTCCCCCC ACCCTATCAC ATCACGTCTT TGAGGTCTAG GTGCAACCTC   32100

TCCCACTGAG GCCAGACAGT GCAGCCCAGT GAAGGGAACA TATTTCACAG ATAGGCAACA   32160

GCTTTGGGGA CATCTCCTGC TCCAGTTTTT GGGGTACCCA CATGAAGACC AAGCTGCACA   32220

TCTGCTGCCT GGGTGCAGGG GCCTGCTTTG CTGGCTTTGA TGTTGCTCCT GTTCATCCAT   32280

TGTCTTCTTT CTTACTGGTC ACTTCCTTTC CCAAGTTCCT CTAGGCCAGT CATCCCCAGG   32340

CTAGCCTTTC TACTTCTTTA GCTTGATTTC ACGTGCCACA AAAACAAAGG TACCCCAGTG   32400

TCTTGGGGCT TTTTTAATTG TTGTTTGTTT TGAGATAGAA CCTCATTGCA TATTCTGGCT   32460

GTCCTAGAGG TCAATGTGTA GACCAAGCTA GCCTCAGAAT CATGGATTAA AGGCAATAAT   32520

CAGCCATGCC TTGTCCTTTT AATGAAAGTG GTAATGGTCC CTCCCATCAC CCTCTTTCCG   32580

TAATTAAGGT ATATTTTCAA ATTCCTTGCA TTTTTTACGG TTTGTTAATA CTGTTAATTG   32640

AAATTTAGTA GTTATAGTTT TCATTTCTAG AATTTATATT TGTTTGTTTT AAAGATGGTA   32700

GATCATTTTG GGGATAATTT TTTTGTCCAG ATTTAATTTT TTTCCTATTA GTTTCCTGTT   32760

TTTATGATAT CTTAAGTTTT GGGGAAGGAT AATTCTATTG ACTTTTACTC ATGGTGATTT   32820

ATACCGTGAA TGAGTGATTT CTGGTCATTT TGAGTTTATA TCTTTGAAGA GTTTATTAGA   32880

GTCATTTTGC ATGCACGTGA GCAGTTGAAG ATTGGCACAT GCCAAAAACT GCCTTTGACC   32940

CTCATGCCAC CTGACCTGAT TTCTATTTTC CTTTAAAATT TTTTTTTTTT TTTAGGGAAA   33000

AGAGGGCTTC TAAAAATTTT TCCTTTTCTT TGTCACTTCT AACTGTATTT GAGATTGTAT   33060

TCAACTATTA AGATTATTGT CTTCTAAAAA GGAACAGGGC GGCCATTTCA AAATAAAACA   33120

TTCCTATGAG GGGGGATTGA GGATGATCAT AGGAAAAATA AAAAAAATAT ATGCAAGGGA   33180

CAGGTATGGG AGGGCTGCAG GTCAATACCA CACCAGCATC AATTGTATTT CTCATACCCT   33240

TCTTATACTA CAGTACCGAG GGAAGCAAAC CCATAGGCTA ACAAAAACAA TTGCTGAAAA   33300

ATACACAGGT GACATCTATT CTCATCCAAA ACCCTCCCTG TTCCTTCCAC CAAGGTCAAT   33360

AAAATCTTCA CCCCCTTGCC TTGAGCCTCA ATTGCACCTC CTCTTATTGT TCCATGCCTC   33420
```

AGCAGGCCAG GAAATCTGCC ATCAGACCCT CACCTGCCTT GATTCTGCCT GGCAGGCAGA 33480

CTCACAAACA CACACACACA CTCACTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT 33540

CTCCCCCCTC TCCCTCTCTC CTCTGCTTCC AACATAGAAA ATTTTAGAGA ATCTGTCCAC 33600

TGTATTTTTA TAAATACAGA GAGATGGGTT CATTTTAGTT CAGCGTTTTC AGGGTTAGCA 33660

AAAATAATGT CTATTGTTGT GGGCTACAGG GTATTTCTCA GTGGTACTTG CATGCATTCT 33720

TAGCTGCCCT CACCTAACAG GCACACACAC ACAGCCCTAT CCTGGCCAGA GTGGGGTCTC 33780

AGCGCCTCCC ATCTTTGTCT TTTCTACAGC TGCTATGCCC CTCAAATCCT GGGTTGTCTT 33840

CCTACTTTGG CACCAACTGT TGTTTTGACC TTCTGTCTCA CCGGTGGTTA GAGTGCACAT 33900

GTGAATGGGA TGTGTGTGTT GGGATGCAGT GTTGTTTTTT GTTTGTTTGT TTGTTTGGGT 33960

TTTTTAAATA TATTTTTGTT TTGTTTTCAA GTTAATGCTG AACTCATAGA ATGGGTTGGG 34020

AAGTGATGTA CCCTCCTGGT CTCTTTTCTT CATTTACACC TAAGAATTGG AACATATGTG 34080

TTGTCTTGCT GTTCGTGTTT ATTGAGTCTT CTTTCACATA AACAGGTTTT TCTAGTATTT 34140

ATATATTGTA TAATTTGGGT TATATCTTGG GCACTTTGTT TTCAGTTTTT AAAAATCAAG 34200

TATTTTATTT ATTACTAGTT TCTGTCAAGA GTTCAGGCCA GGAAATTTCA TAGTGAAACT 34260

CTATTTAGAA AACAAGAGTC TGGAGAGATG GTCAGCAGTT AAGAGCACTG GCTGCTTTTT 34320

CAGAGTCATG AGTTCAATTC CCAGTACTCA CATGGAGGCT CACAACCATC TGTAACTACA 34380

GTTTAGGAAA ATTTGACACC NTCTTCTGGC CTCTGCCCCT GACTGACTAA CAAACACACA 34440

TACACACACA GTGCATAAAT ATACATGTAG ACAAAACACC CATACACATA AAATACCTTT 34500

TTAAAATAGA ACAAGAGTCA CGGCCACATA CGCTTCCTGG GAGATAGAGC AGTGAACTCG 34560

AATGGCTTCC AGATAGGATT TCTGGTTTCT TGGGACTTTT CTGTCTTTGT TCTTCCAGCT 34620

AGACACTGAG ACTGTTTCCT GGTGTCAGGG CCAAAAGCTG GGAAGTCAGG GAGCTACACA 34680

AGACTCGTCT CATACCCTGA TGAACACCGG CTCCCGTCAG AATGAAGATT TTGTTCTCCC 34740

TAAGTGCAGG GCTCCAGCCT CCTCTTTCCT ACGACTTGGG GGTGGCTGGT GGCTGGCTTA 34800

AGAGAGACCA GTGGATCTCC CCTTGTCTTG CTAGTGCCAG CTCTGGCAGG CCTTGCATGG 34860

AAGTGAGCAG AAAGGAAGGG TTAAAGGCTA CTGAAGCCAC GCCACTTGCA CTTTCCTAGC 34920

TCCCTGCCTT CTCTCGTAGT CCGCAGCTAG CTGAGGTTTG AGTTCTTCCA GGTGCAGCTG 34980

TAAACAGAGA AGAGACCGTA GAACGTGCTT AATTCTGTCA GCTAGACCTG GCCTTCACAT 35040

TTGGGTCAGT TCTGTACATG TTTTAATTAT GCTAAATTTG AAAACATCTG GAATGTTGGT 35100

ACCCAACTCT ATTTCTACCC TCGTTCATCT GTGCTGTTTC GTGTTGTTTT TTAATGAGTC 35160

TTAGGAAGCA CAGCCAATTC CATACAAGAA TGTTGTTTAC TTTGACTTCC AGCATGAAGT 35220

AAAGCGAGCT TTTGAGCCAG ACATTCAAAA GCAGCTTGCT GGGACCTGTT TGCAAGCAAG 35280

CCCTGAGGGT TTCTCCCAGC AGTGTTTGAT TGTACAGTGT CTGCAGGAAG GGCGTGGGGG 35340

CCCTGACAGG GACAGACACA GAAGAGCCAG TTGGTTGTGA TGCGTTTGTG ACAGATGATC 35400

TGTAACTTTA AAGGCATGAG GTTTCTAGCA CGGATGGCTC AGGTTTAAAG AGTGCTCAGT 35460

GCTGGCAGCA GAGCTCCTGA CTCTGGCCTT TGGGGAGCCG TGGGAGCTGC TAAGCTGGAG 35520

ACTTGAGAAG GCTTGCTTGC TGTTGATTGT CCAGACTAGG GGTGCTCCTT AAGGCTTTGA 35580

CAATCATAGC TGACCAGTCT GAACTGGAAA GTTTTACATC CTTATTACAG GCAAATTACA 35640

ATATAAAGAA TGAAACAACA TGGTAAGTCA GAATGTCCAG TGCCCATAGA GACCAGACTC 35700

TCCTGAGAAC TGTTCAGTGG GGGTTCGAAA GAACAGCATA TGCTCTTAAC TACTGAGCCC 35760

TCGTTTTTTT CTAACAGTAA AGACTTTTCC TACTTTTTTA TTTTGACATC ATGATAGTTC 35820

CNAAAATTTA AAGTAGAAGA TAGCTGTTCT ATTAAAGCCT ACACACACAC ACACACACAC 35880

ACACACACAC ACACACACAC ACACACACAA ACACACATAT ATATAAATTA TACAGAAATT 35940

ATTTTAAAAG CCTGTCATAT AACAGATTCA TTTGGTGCTA CCCAAAACTT ACGATACAAA 36000

TAAACAACTA CAAAGGGAAC CATCACTAAA CTAAGACACA CAGGACCTAT GCTTCATTTC 36060

AGGGGGAACC CCAAATAACC TTTATCATGT TAGAGGAGCC ATGTCCATTT GCCATGTTTT 36120

GATTACATAC ATCTCTCCAA ACCCTTTATT GCTCTTTTTG TCCATATTTT TTTCTTTCTA 36180

TTTATGGGCC AGAGTAAATA ATAAAAATTA TCTTTATTGA TAACTGTTTT TTGACAAGTG 36240

AATGCCAAAT GATGTCTTTT GTTTATTTTT AAAAAGAACT TGGGACTGGT GAGAGCTCAG 36300

CAGGTAAGAG CACTGACTGC TTTTCCAAAG GTCCCGAGTA TGTGTATATG TCTTTCTGTT 36360

TCATTCTCTG AGTGCACCCA GGATTCTCAC ATAAGTATGT GCTGATCCCC AAAGCTGCAT 36420

TTGTTCTGAC ATTTAAAAGG AGAACATGGA AGCCAGGCAG TGGTGCTGCT CACCTTTAAT 36480

CCCAGCGCTC AGTAGGCAGA GGTAGGCGGA TTTCTGAGTT CAAGGCCAGC CTGGTCTACA 36540

GAGTGAGTTC CAAATCAGCC AAAGCTACCC AGAGAAACCC TGTCTCAGAA AAAGAAAAAG 36600

TAGACGATGG ACATGGGCTA GAACCAGTAA TGTAGCAGTT GCCAGCGTGT GAAAGGCTGG 36660

TGTACCCCAG CCTCGACGAA GGACAACTAG AAAGCAGACT GTAGCACATA CTGTTTCTCA 36720

TGTTAAAAGT CTCCTTGTTT CTAAACTGTG GGATCTGTTG GTTACATTTG AGTTGTCTCT 36780

CTTCTAGGTC ATCCACGCCT GGCTTATTAT TTCATCTCTG TTGTTGCTGT TCTTTTTTTC 36840

GTTCATTTAC TTAGGGTAAG TAATACAGGG TTTTGGTCTC TCTCTCCATT AACTCTGACA 36900

TCCACCACTC CCTTAGAGCC ATGTACCTTG CTGATGAATT AGTCTAAGTA ATCATTTTTT 36960

CTACATACTG TATCTTAACA GGTACATAAA AGATGAGTTA GTGTTGGTGA TGGGGACCAT 37020

GTGCCACAGT ATGCATGTGG CAGTCAGAGG CTAAGTGGTG GCTGTCAGCA GGGTTTTCCC 37080

TAGAGAGAAG TATGGGGGAG GGGAGAGGAG ACTTGTTACG AGTCTTTATG TAAATGTATA 37140

CGTACCTAGC TAAGAGTTGA GGTTAGTTTC ATTTTAGAAT ACCTGAAATT CTTAGTATAA 37200

TATACATTGC AACCTCATTC AGAAACTTCA ACACTTTTGC TTTATTTTAG AATTTTAAAA 37260

AAGGAGAAAA ACCATCTTCC TAGACAGCCC CATGTCCCAC AGTTTGTGCT TGAGTATATT 37320

CTCTCTCCTT TGCTACTTAG GGTGTCTACT ATAGTTACAA TCTTGTAGAT GATAGAGTTT 37380

GTGCTCCAGA AGGTAAAACA GTTTTCTTTA TATAGAAAGT TAACCAGGGT CTGAGGAGAT 37440

GACTAAGATG ACAGTGTTTG CTGACCTGAA GTCACCCATG AAGACCTGAG TTTGGATGCT 37500

CAGAACCTAC ATAGAAGCAG ACCAGTGCAG GGACAATTTC CTGTAACTCT AGCAATGGGA 37560

GTCAGTGACA GATAGATCCC AGGGACTTGC TGGCCAGCCA AGCTAGCCAA AACAGCAAGC 37620

ATCAAGTTTT GTGTCAAAAC AATAAGGTGA ATGTCAACTT CTGGCTTGTA CATGTGCTCT 37680

TACAGGCACA CTCACCCACA GAGAGCACAC ATGGTGCCAG GCAGCGGGCT AGCAGCAGCA 37740

GATCCATGCT GGAGAGCATG CTCAGTGCAC ACGAGGCCTT GAGGTCCATC CCAGCTTCAC 37800

ACAGACAAGT CTAGTCATCC TTCTAGAGAG TGATAAATCT CCTGCTGTAT TCTTAGACAC 37860

TGTTCTCATC AGTGTCTGAT GATGTTATAT ATAGTATTAT AATGTTATTA TGAAATAATA 37920

TTGTTTATGT TCTCACTATT CATTTTTTTT CTAATTTCAG ACCTTTAAAG TAGAATTTCA 37980

TAAAATACAG TATGATAGTG TAGGGGATAT GATTTTCTTT TTGTAGTTTT TGTTTTTATT 38040

TTTACAGGTG TGACCTGTGA TAATATGTGC TCATTTATGT GAATAATGTA ATACCTGTTA 38100

GAATAAGTTC TGCTGGAGGA AAATGTGTTA TTTAAGAGCA AGGTGCTACA TAAATAACAG 38160

CACAGCATAG GTCATAGGAA AAGCAAATAT GGAATGTGAA TAGTTAATAT TTCAGAATCT 38220

TCGCTCTAAA CAATGTGTTC ATGAGAGTTA TTGCACATAA GTCACTTCAA ATCCTTGCGT 38280

GATTGTTTTC AGGGAAGTAT TTAAGACCTA CAATGTCGCC GTGGACTACG TTACAGTAGC 38340

ACTCCTAATC TGGAATTTTG GTGTGGTCGG GATGATTGCC ATCCACTGGA AAGGCCCCCT 38400

TCGACTGCAG CAGGCGTATC TCATTATGAT CAGTGCCCTC ATGGCCCTGG TATTTATCAA 38460

GTACCTCCCC GAATGGACCG CATGGCTCAT CTTGGCTGTG ATTTCAGTAT ATGGTAAAGC 38520

TCAAGACTGA CACTTTGTCC ATCACAGACT CACTCACTGG TGTGCTTTCC TTCCTCTTCT 38580

AGTCCTCTTG ATTTAAGGAA ACTCTTAATG CTTATCTCCC GTAGGCTCCA GTGGTTTTCC 38640

TGTTACACTG CTAAGGATTG CAAGGCAAGT GCTGTGGCCT GAGCCCCAGC GCAGCCTGAG 38700

AGTGAGAGTT TGTAGTTCCT TGGGACTCTG AGCTGGGTTG ATGGGAGGAG GCCCTGCTCA 38760

GAAGGCTTAG GTTCAGAACT CTTCTACCCA TATGTGAGGG GCTGGCTTTA GTCCCCAGAC 38820

CCCAGGAAAC AAAGCAGCAA AGCCACCAGC CGCCACCAGC AGTTCCCCTG AATTTTGTAA 38880

GTTGGGAACA AGTCTCCAGG CTTGTAAGTG TTCAGCGAGA CGGTGGATAC TGAGATGTGC 38940

TAGCAGCAGG CTCACACTGT AAACTGTTCA GCTCATGTTT GAAACTGTAT TCTATTTTGG 39000

GGGAAGGGCT TGTGTCCTGA CTGAGTGTAG AGGTCAGAGG ACTGCTTCAG CAGTGAGTGC 39060

TCCCCTCCCT CAGGGGGTCC TGGGGACCAA GTGAACGGAG AAGGGCAGGC ATGGTGGCCC 39120

GCATGGCTAC TGAGCTGTCT CACTGCCCTC AGTTAATCTT CTCCCTCCAG AGTCTTACTT 39180

TATTTATGTG TAGTGCATTT TGCCTGTAAG TGTGTCTGTG CACTGCCTGC ATTTCCTTGT 39240

GCCTGCAGAG GCCAGAGAAG GGGCTCATAT CTTCTCAATT CATTTTTAAG ATCAGAATCA 39300

TATAGCAACG TGAGTGCAAT CACCCCTTTA AAGATTAACC TCATCCTGAC GTGTGTGTAC 39360

ATGTGCGACT CAGCACCGGG AGGTAGAGGC AGGAGGACCA GAGTTTCAGG GTTATCTATG 39420

ACTGATGTAG TAAGTTTCAG ACCAGCCTAG GCCACTTTAA ACCCTTTTTG AAAAAAAAAA 39480

TGTTTTTTAA CTAAATAAGT AGTAAACTAT TAAGGGTAGT AATGTCCTCC TGTAATCCCA 39540

GCACTTGGGA TTACAGCTGA AGGCAGAAGA ATTAGAAGTT CAGTCCAGGC TATGTTTTGA 39600

GAGCCTCTCT GGAGAGAAAA GCAGTTCTGA GAGTGAGGTT TACATTTGTA AAAGAATTCC 39660

AGTGCCTGTT TGTTTATTGC ATTTAAATTC TACAGAAAGA CCTGTGTGTA GACTCCATAA 39720

GTGTTGTGAA GGTCCTCAGT AAAATCCTAT TATTTTAGCT GGGCAATGGT GGCTGGTGCA 39780

CACCTTTAAT CCCAGCACTT GGAAAGCAGA GGCAGGTGGA TATCCATGAT TTACAGAGCA 39840

AGTTCCAGGC CTGCCAGGGC TACACAGAGA AACTGTCTGG AATAAAACCA AAAACAACAA 39900

CAATAAAAAT AAATTTAAAA AAATTTAAAC ATACTTTTTA ACAGCTGCTG TCCTGAAAAG 39960

TCACAATAAT ACTCTAGCCA TGGTAACGGG GCCTGGCTTT ATCCTCCAGC TGTTGCGCCA 40020

CATGAAAGCC AAAAAGGAGT CTGTTGTCAG GGACGAGTCA GTGGCCACTG TTTTACAGAG 40080

CATCTCCTTA CACTGAGGCG CGTGGTGACT GGAACTGGAA ACTCGTCCCC CTTGAGCTGA 40140

TGGAGCAAAG CCTGTGGCGC TAGAACCGGG CCAGTTCTGT TGTCCACTTG AGCAACTGCT 40200

TAAAGTCAGG GGTTAGAAAA GTCACCTCAA CGCTCACTTC ATGGCAAACT TCATTCTTGG 40260

TGTTTTGTCA GTCGCTGAGA GTAGGAGTAC CATTGTAAGC TAGAGAGACT CAGCCAGAAA 40320

AGGACTTATC ACCAAGCCTG ACAACCTGAC AAGCCCAGGA GTCACAAGGC AATAGGAGAG 40380

AACCAATTCC TGCAAGTTTC GTGTACACAC ACCCTTGTGC AGGCAGGCAC AATAAATAAT 40440

ATAATTTACA ATCCCGCCCC GGGAGTCGGC AAGGTAGTTA TCATAGGGCT TTTGTTGCTA 40500

AGAAGCTTGA AGGTTACAGA GGTAAATTAT TTGCCTGGTC GCTTGCATGT GGCTTGACCT 40560

AAGAGTCTGA ATTCACTCCA TGAGTTACTC TCTCCACGAC CTTCAGCCTC TGCTCACAGC 40620

ACAGGGGAGA CCCCGCAGCT TCAGGTCTTC TGAGTGTGGA CGTGCTGGGC ACCCAGTGTT 40680

GACATAGTGA CTCTCAACCT GTTCCTTACA AGGTCTCAAA ATGGTCTATC AGAAGTGTTC 40740

TGAAGCCTTT TACGTAAATT AATTTTCATT CATTTTAATT GTGCTAAAAA TAACTGAATC 40800

TATTCATTTC AAGATTCTAT AAGGGGAAAC ATGTCATAAA ACCAAAATAG GTACAGGATG 40860

TGGCAGCTTT AAAACAGCTG AAGATCTCAT TAGTTGCAAG CACAGGAAAG CTTGCTTTTC 40920

TCACTTGAAC ATGTAAATGG AAGTGTAGAG GAGTGTTACC ACTGTGCTGT TTTTACGTTC 40980

TGCGGGGAGT GACTGAAGGC CTGAAGGAAG TGACGTCAGC TGTCAAGTCT CCCATTCCTG 41040

TTTTACCTAT ACCTGTTTGA TATTTTTAAA TGTTTGAGGT TTTCTGCATT TCTGTAACAT 41100

AGAAAATGAA AGACAAACTG TAGTTTATGG GGGGAAATGG TGTTAGAACC AGACATAATT 41160

TAGGCCTATC AGGGCTTCTC TCAGATCAGT TTTTATCACC ACCAGATGTT CTGAGACCTA 41220

AATATTTGTG AAATACCTTA ATGACTAGAA ACTGAAGAAA TATTAAATTT TGTCATATAC 41280

AGAGCTATCA TTTTATTTAA GTGTGTTGCT CTATTATGGC AAATTCCTTC GTTATGTCTA 41340

CTTGCTTTGT AGTGCTAGGG ATTGAGTGCA GGTCTCAGGG CTGTTCAAGT AAGTGCCGAA 41400

GTAGCTCACC CCTCCTGCTG TCTGTAAAGC AGGACACTCT CCAGCCAGAG CAGCACACAC 41460

ACACCCTCCT GCTGTCTGTA AAGCAGGACA CTCTCCAGCC AGAGCAGCAC ACATACACCC 41520

TCCTGCTGTC TGTAAAGCAG GACACTCACT CCTCCACTCT GCTCTTGGTA CTACGAAAAG 41580

TTGAATATTT GGAATGGTGT CTATTTGTAA TTTATAGGTG AATTTAAAAA ACAAAAAACC 41640

TCCCTGCCAT TTCTGCTTGT TCCACAATGA GAACTATTCT AGAATAGTTC TGTCTTGCTG 41700

TCTCTGCCTC TTTCCTTCAG GCTAAGAAGA AACATGTCTA GACCTCAGTA ATCACACACT 41760

GTAGATTGAG CACACAGTGG TACGGCCAGA GTATAATTTA CACAGGTCTT CTATGGGTTC 41820

TGTTTTATTT TGTTTAGGTT GCTTTTTCTC TTGGCTGCTA AACTGCAAAT ATTTTCATAT 41880

TTACCCNACA CTTAAAAAAA AAAATTAGAC TTGTTTCTTC AGGTAAGTCC TCCCCTGCGT 41940

ATATCCACTT ACACAATCTA GTCAGTATAC TTTATTGGGG TCTTTTTTTA TTTTTCTCTT 42000

TCTAGATTTG GTGGCTGTTT TATGTCCCAA AGGCCCACTT CGTATGCTGG TTGAAACAGC 42060

TCAGGAAAGA AATGAGACTC TCTTTCCAGC TCTTATCTAT TCCTGTAAGT ATGCTGGAGC 42120

AATGCTGTGT TAGTGCATTC AGTGTTAAGG TGCCCGTGGA GGTCAGAAGT CAGAAGAGAG 42180

CCTTGGATCC CTGGAACTGG ATTTATAGAT GGTTCTGAGT CACCATATGG GTGCTGGGAC 42240

TCAAACCCAA GTCCTCTGCG AGAGCAGCCA CTGCATTCAT GGCTCAACCC GTAAGCCATC 42300

TCGCCAGCCC CAGTACATTG TTTTATGTTT GTTTCTTAAC CTGCAATCTT GTATGGTTTT 42360

AGTCCTGAGT TACTTCTGCA CAGAGACCAT TGTTTTCTGT AGAGAAGTGA GCCTCTTGCT 42420

TGCTAGAGCA GTGTCCTGGG TTTAGGAACT CACCTGCAAC TTTGTCTTTC TCAACAGCAA 42480

CAATGGTGTG GTTGGTGAAT ATGGCTGAAG GAGACCCAGA AGCCCAAAGG AGGGTACCCA 42540

AGAACCCCAA GTATAACACA CAAAGTGAGG TGGAGCTTGT TTATAAATGC CCTAGTTTTG 42600

CAGGTCCATT GTCATTTGTA GATAAATAAT TGAAATGATG AAGGAGTAAT TTTAACCTTC 42660

GAAGTCTTTG GTCACACTCT CCTTGCTGAC GCTCAGGGCC AGAGCACTTG TCTGAGAAGC 42720

CCTGGATGTA GGGAGAATCG TGATAAGGAA AGATATCACG AGAGTCTCCT AGAGAAACAG 42780

GCTCTCCCTA GTAAGTACAC GTAGGACTAT TCACGGAGCT GAAATCCCCC CACACACACA 42840

CACACTGTGA GTTCAAGACT AACCCTGACT GTGTGGGGAG ATCTTGTCAC AGAAAGGGGT 42900

GGGAGTGCAA GGGTCTCTTG TCCAGTATGT CCAAAATAAC TGAAAACATT AGGTGGGAGT 42960

GCAGTCAAGG TTACACTAGG AAGGTCAGTA NTATTTCAGT GGCCTGGGTA ATGGTGACAA 43020

CCAAGGACTG TGATACTGCA GAAGGCACTA AAGAAAACAT CAGAAGCTCT GTCAGAAAGA 43080

AGTGAAGAAC AGGGTGGTGT AAAACTGAAG CAGAAAGGAG CTTTCTTAAA AAGACAACCT 43140

GGGCAGGTAT AGACAGGACG ACCTCTGGGG CTTGCTGACC ANCCGCTCTA ACTGAACCAT 44820

TCCAGTTCCG TGTGCTATGC CAGCTCCAAA ACTAAGCTAG AGAATGATTC AAGAAGACAC 44880

CAGGCATCGA CGTTAGGCAT CGACGTTTGG CCTCTAGGCA CCGTGGACAC ACATAGCACA 44940

CCCCCTCGGT GCCCTTCTTT TAATACATGT ACACAAAAGA GAGAGAGTGT GGTGAATTAA 45000

AATGATAAGC AGATAATCTG AACAGCGCCA AAGCAGGAAG GTGGGCATTG AGATTCTAGG 45060

GTCACACAGA ATCTAAAGCA GAGAGAAGGT AAAGAAGGCA ATCTTTCTGC GGCCTGAACT 45120

GTACACACCA CTTACTGACT TCTCTTTCAG TGCGGGACAA AGTGTCACCG AGCTGCCCAG 45180

GCAGGCCTCA CACTCCTGGT CTTAAACTAT CCTCCTCCTC AGACTTCTGA GTAGCTCAGA 45240

GTGCAGGCTC ACACCACCCT GCCCAGCTTG GGATCATAGC TTTGAATTAT GAAATCACAG 45300

AGTTGGGTGT GGTGGTACAC ACTTGTAGCC CTAGGGCTTG AGTGGCTGAG GCAGGTTCAC 45360

AGGTTCAAGG CCAGCCTGAG CTACATAACA AGACCTTGTC TGGAAAAAAA AATAAGAAAG 45420

AATGACAAAC TTATTCTTTC TTTTTTTAGG AGGAGTAAAA CTTGGACTGG GAGATTTCAT 45480

TTTCTACAGT GTTCTGGTTG GTAAGGCCTC AGCAACCGCC AGTGGAGACT GGAACACAAC 45540

CATAGCCTGC TTTGTAGCCA TACTGATCGT AAGTATATGC TGGCAAAAAC TGGTCAGGAA 45600

ACTGATTCTG CATACCTTTA TGGTAATTTT CTTTTAAGTG TGTTTGTACA CAGATACTGT 45660

GGCAGTCGTG TGACAGAGAG CATTTTTCCT TCCACTTGAG GCAGGGTCTC TCAGAGAACA 45720

ACCTTTTGGG AGCTGGGGCT TTCCTTCCAC CTGTGGGAGG GTGTCTCGTG TTTTTACCAT 45780

GAGCCACCGT GCACTCTGGC CTATTCTAAT TCCCCATCTC TCCATAGGAC ACTAGGATTG 45840

CAGGTGGGCG CCCACCACAT CCAGCTTCTT ACATGACTTC CAGGGGTCAA ACTCAGGTGT 45900

CGTGTTTGGG TGGCAAGTGC TTTTTACCTG CCGGGCCATC TTTCTAGCTC AGAGAATTCT 45960

CTTTTAATTC CAGGCAGGCT AAGAAGCTTT ATAACGCTTT TTTAAAATGG ATAATCAGAG 46020

CCTGCCTGAG TATCTCTCCA CTCTGGCCAC ACATTAAAAT CAGCCGATGC TTCCAGCCTA 46080

GCTGTGTCTT TTCTGCGGAA TGTATTGCTG AGGATGAGCG AACAGGAGCA CCAGGTGCCT 46140

GCTGTATTCT CCTCCTGCAC CAGCCTGGTC TCAGCCCTGT GCTCACCTTT GCCCCGCCCT 46200

TGGGTAGTTA GTTGTGAGGA GTTCTTAAGA GGTCTCCAGG AAACTGTGAG TACTTCTCAT 46260

TAGAAACCAC TGTGAGCCAG GCGGTAGTGG CACACACCTT TGATCCCAGC ACTTGGGAAG 46320

TAGAGGCAGG TGGATCTCTG ATGAGCCTGG TTTACAGAGT GGGTTCCAGG ACAGCCAGGC 46380

ACTATTCACA GAAATTCTGT CTCGAAAAAA AAGGAGGGGT GGGGATGAAA TGAACAGCTT 46440

TGAGAGTCCA GCTGTATAGG GCTCAACAAA GGCTGTCATA GTCCACGGGC TGAGTCAGTG 46500

AATGGCCAGT GCCTTTACAT CTGATGCACA GCCTGCACCA TGCTTGCTTG GGACAGCAGG 46560

TGTGGAGTGA GAGGGGCAAG CCCAGTGCAT AAGCCCACAT TCTGTCTTGC TTATGATGGC 46620

TGTGGCTTTC AAACATGACT TAGAGCACTG GATGTGTTGG CGTGCACCTT TTAATCCTGG 46680

CATTCTGGAG GCAAAGGCAG GTGGATCTCT GAGTTCAAGA CCAACCTTGT TAAGTTCTAG 46740

GCTACATAGT GAGAGCCCCA TCTTAAAAAC AAAAACCTGA CTTACTAATA ACAAATGAAC 46800

TATTTCTTAT TGCTTGGAGT TGGGGTATAT AGTTCAGCTG TGAGCGTGCC TGCATAGTCC 46860

AAAGCCCTGC ATTCAATCTC CAAGCACATA GTGCAGGCAC GGGGCCACAT CCCAACACTA 46920

CGGAGATGAA AGGAGAAGGG TCCGAAGTTC AAGGTCCTCC TTAGCTATAT GGGGGATCTG 46980

AAGCCAGCTT GTCTCAAAAA AAAGGTTTTT TTTCCCCTTT AATCCAGGCA TTATGGTGCA 47040

TACTTACCTC CTGGCACTTA GGAGGCTGAG GGAGAAAGAT ACAGAGTTTG AAGCCAGTCT 47100

GGGTTGTGTG TATATAAATG AAGAAAAAAA ACAACCTCTG CATTTTTTTT TAAAATATTT 47160

ATTTATTACA TGTAAATATA CTGTAGCTGT CTCCCGACAC ACCACAAGAG AGAGTCAGAT 47220

CTCATTATGG ATGGTTGTGA GCCACCATGT GGTTGCCGGG AATTGAACTC AGGACTTCAG 47280

AAGAGCAGTC AGTGCTTCTT AACCGCTGAG CCATCTCTCC AGCCCCACCA CTGCATTTTT 47340

GTAGAGGGCC ACACTCCCAT GCACAGGCCC TGGGCTCTGT CTTCAGCACA GATTTTTTCT 47400

TGTTTTTCCA TTGCATCTCC TCCTGTCTTC TCTTTCCCGT CTCCCTCGAG CCAGCTTTCG 47460

CTGTTGTCCA GCCTGGTCTC AGGTTCTGGC CTCAGACAGT CCTGCCTCAG CCTCCTGAGG 47520

AGCTGCTGCG TGACCGTCAG GCCCAGGGCC ACAAGCTACG CTTTTCTCTC GTGACTCCAG 47580

AGTTCATAAG ATGTTGTTGT AAAAGTTAGG CCTCAAGGAG AGTATTCCCT GAGCGTGGTG 47640

CTGATGCTGC GTGGCGTGGG GATGGTCCTG AGTGCGCCTC TGTCTCCACA GGGCCTGTGC 47700

CTTACATTAC TCCTGCTCGC CATTTTCAAG AAAGCGTTGC CAGCCCTCCC CATCTCCATC 47760

ACCTTCGGGC TCGTGTTCTA CTTCGCCACG GATTACCTTG TGCAGCCCTT CATGGACCAA 47820

CTTGCATTCC ATCAGTTTTA TATCTAGCCT TTCTGCAGTT AGAACATGGA TGTTTCTTCT 47880

TTGATTATCA AAAACACAAA AACAGAGAGC AAGCCCGAGG AGGAGACTGG TGACTTTCCT 47940

GTGTCCTCAG CTAACAAAGG CAGGACTCCA GCTGGACTTC TGCAGCTTCC TTCCGAGTCT 48000

CCCTAGCCAC CCGCACTACT GGACTGTGGA AGGAAGCGTC TACAGAGAAC GGTTTCCAAC 48060

ATCCATCGCT GCAGCAGACG GTGTCCCTCA GTGACTTGAG AGACAAGGAC AAGGAAATGT 48120

GCTGGGCCAA GGAGCTGCCG TGCTCTGCTA GCTTTGACCC GTGGGCATGG AGATTTACCC 48180

GCACTGTGAA CTCTCTAAAG GTTAAACAAA GTGAGGTGAA CCAAACAGAG CTGCCATCTT 48240

CCACACCATG TTGGAAATAA AACACGTCCT AGCTGAACCC TTACTGTCCA GGAAGTTCCG 48300

TGTGGAAGTG GCACTGGGCC GGGCCTCCCT CTCAGGCTCC TTTGCTGCCC ACTTGTAAGT 48360

TTAAATAAGG ACACCGCCCT ACACAAACCT CACCCTGTCA CATCAGTGAC TCTGACCACT 48420

TTTGTTCTCA AACTCTCTCA CTATTATCTG TGGTTGCCGT TTCTTCCCAA GGCCAGCCTG 48480

GACGAATTTG GGGTTGCTCT ATCCTGAGAG TTGTAACCTC AACTTCCAAA GTTTATATTT 48540

TCTTGAAATG ATGGATCTAT TGCTCAACAG TCCCTGTCAT CCTTAAGTGA CTTCTGGGTT 48600

TCCCACAAAT TCCTCACTTT TAGACACACT CTAAGCTTAC TTCTGGCCTG NATGCTTCCT 48660

CTCCCTGTCT CTCCCTTGCC CCACAGCGGT TCCCTGACAG CAGACAAGGC AGCTCTGGGA 48720

GGTAGCTAGT ATCCAATAAC CCAGGGGTTT CCTCATGTGA TGCAAATACT ACGTGTCCAA 48780

CCAATCAGTG CTGTCAACGG GCTGCCATAG CTCCTTCGAT GGCAAATAGG ATGTGTGCCC 48840

AAAGAATTAA AGCGATGAGT GGCTGGTGAA GCGCTCTCTG TCCTGGCTGG TTTTTTGTCA 48900

CCGCAATCCA CTGCCAAAGG CAGAATTTAA ACTAGGGGCT GAAGGATAGC TCCGAGTAAT 48960

GTCCAATGCT GGGA 48974

That which is claimed is:

1. An isolated DNA molecule that directs neuron-specific transcription of a downstream heterologous DNA segment in a mammalian cell, said isolated DNA molecule having a sequence selected from the group consisting of:

(a) the sequence spanning position –327 to position –206 of the mouse genomic Presenilin-1 gene;

(b) the sequence spanning position –449 to position +1171 of the mouse genomic Presenilin-1 gene;

(c) the sequence spanning position -9 to position +16, of the mouse genomic Presenilin-1 gene; and

(d) DNA sequences that hybridize to isolated DNA having a sequence of (a), (b), or (c) above, and which direct neuron specific transcription of a downstream heterologous DNA segment in a mammalian cell.

2. An isolated DNA according to claim 1, said DNA having the sequence of a mammalian Presinilin-1 gene promoter from position:

-440 to +91;

-352 to +91;

-327 to +91;

-276 to +91;

-261 to +91;

-215 to +91;

-192 to +91;

-124 to +91;

-87 to +91; and

-32 to +91.

3. An isolated DNA according to claim 1, said DNA having the sequence of a mammalian Presenilin-1 gene promoter from position:

-276 to +519;

-276 to +206;

-276 to +148; and

-276 to +41.

4. An isolated DNA according to claim 1, said DNA having the sequence of a mammalian Presenilin-1 gene promoter from position:

-87 to +41;

-9 to +16; and

-327 to +206.

5. An isolated DNA according to claim 1, said DNA have the sequence of the mouse Presenilin-1 gene promoter from position -327 to +206.

6. A DNA construct comprising an expression cassette, which construct comprises, in the 5' to 3' direction, a promoter segment consisting of a DNA according to claim 1, and a heterologous DNA segment positioned downstream from said promoter segment and operatively associated therewith.

7. A DNA construct according to claim 6, wherein said construct comprises a plasmid.

8. A DNA construct according to claim 6, wherein said heterologous DNA segment encodes a protein.

9. A cell containing a DNA construct according to claim 6.

10. A gene transfer vector comprising a DNA construct according to claim 6.

11. A transgenic non-human mammal comprising cells that contain a DNA construct comprising an expression cassette, which construct comprises, in the 5' to 3' direction, a

neuron-specific promoter segment and a heterologous DNA segment positioned downstream from said promoter segment and operatively associated therewith;

said neuron-specific promoter having a sequence selected from the group consisting of:

(a) the sequence spanning position –327 to position –206 of the mouse genomic Presenilin-1 gene;

(b) the sequence spanning position –449 to position +1171 of the mouse genomic Presenilin-1 gene;

(c) the sequence spanning position -9 to position +16, of the mouse genomic Presenilin-1 gene; and

(d) DNA sequences that hybridize to isolated DNA having a sequence of (a), (b), or (c) above, and which direct neuron specific transcription of a downstream heterologous DNA segment in a mammalian cell.

12. A transgenic non-human mammal according to claim 11, wherein said mammal is a mouse.

13. A transgenic non human mammal according to claim 11, wherein said expression cassette is stably integrated into the genome of said cells containing the same.

14. A transgenic non-human mammal according to claim 11, wherein said expression cassette is carried by the somatic and germ cells thereof.

15. A transgenic non-human mammal according to claim 11, said heterologous DNA segment encoding a protein.

16. A transgenic-non-human mammal according to claim 11, wherein said heterologous DNA segment is transcribed in the cells thereof.

17. A transgenic-non-human mammal according to claim 11, wherein said heterologous DNA segment is preferentially transcribed in nerve cells thereof.

FIG. 1

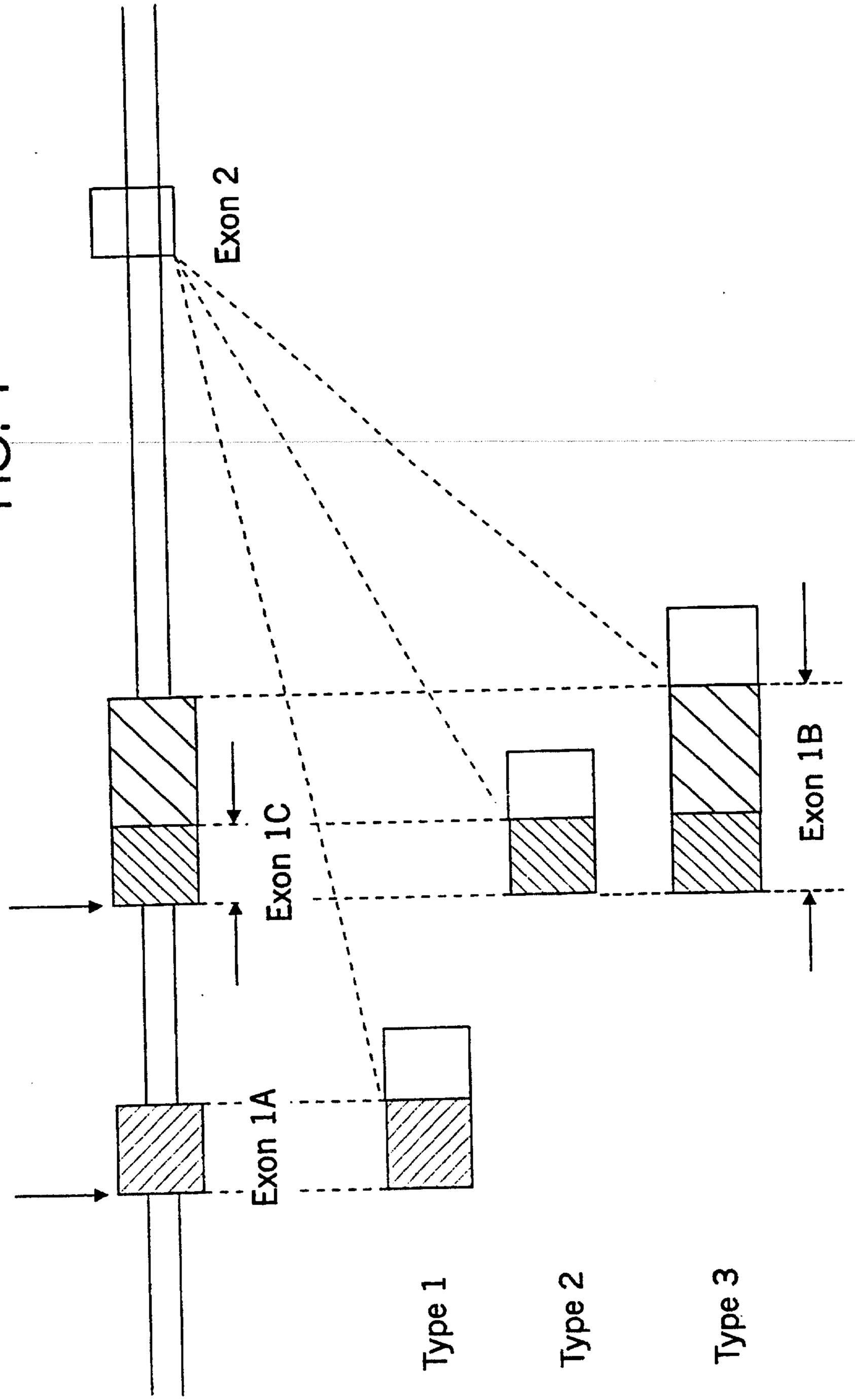
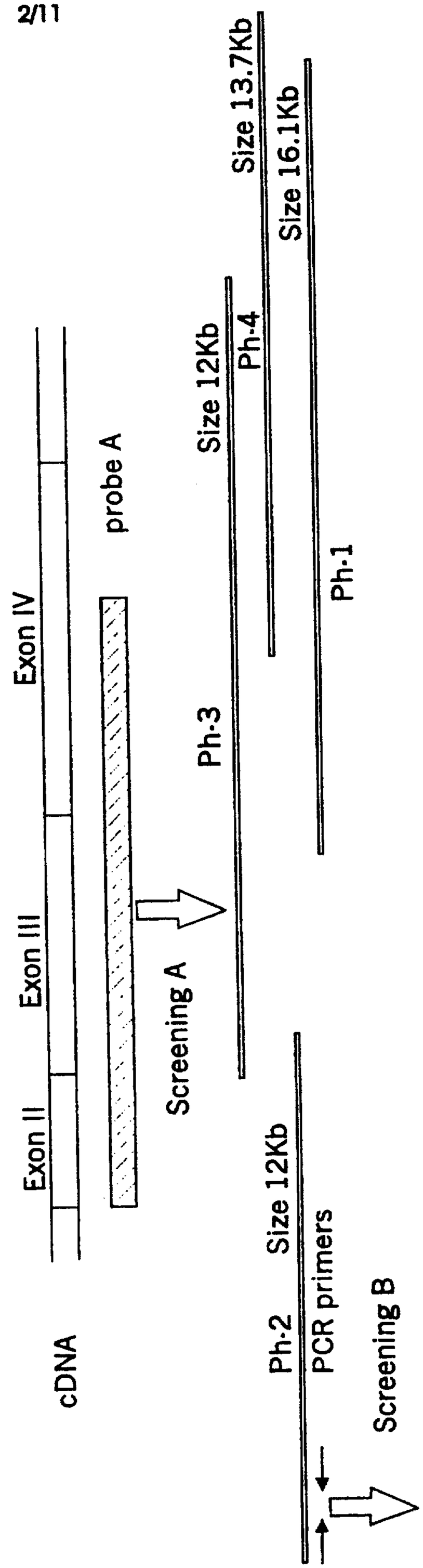


FIG. 2A


Size 13.7Kb
Size 16.1Kb
Exon IV
probe A
Size 12Kb
Ph-4
Ph-1
Exon II
Exon III
Screening A
Ph-3
Size 12Kb
cDNA
Ph-2
PCR primers
Screening B

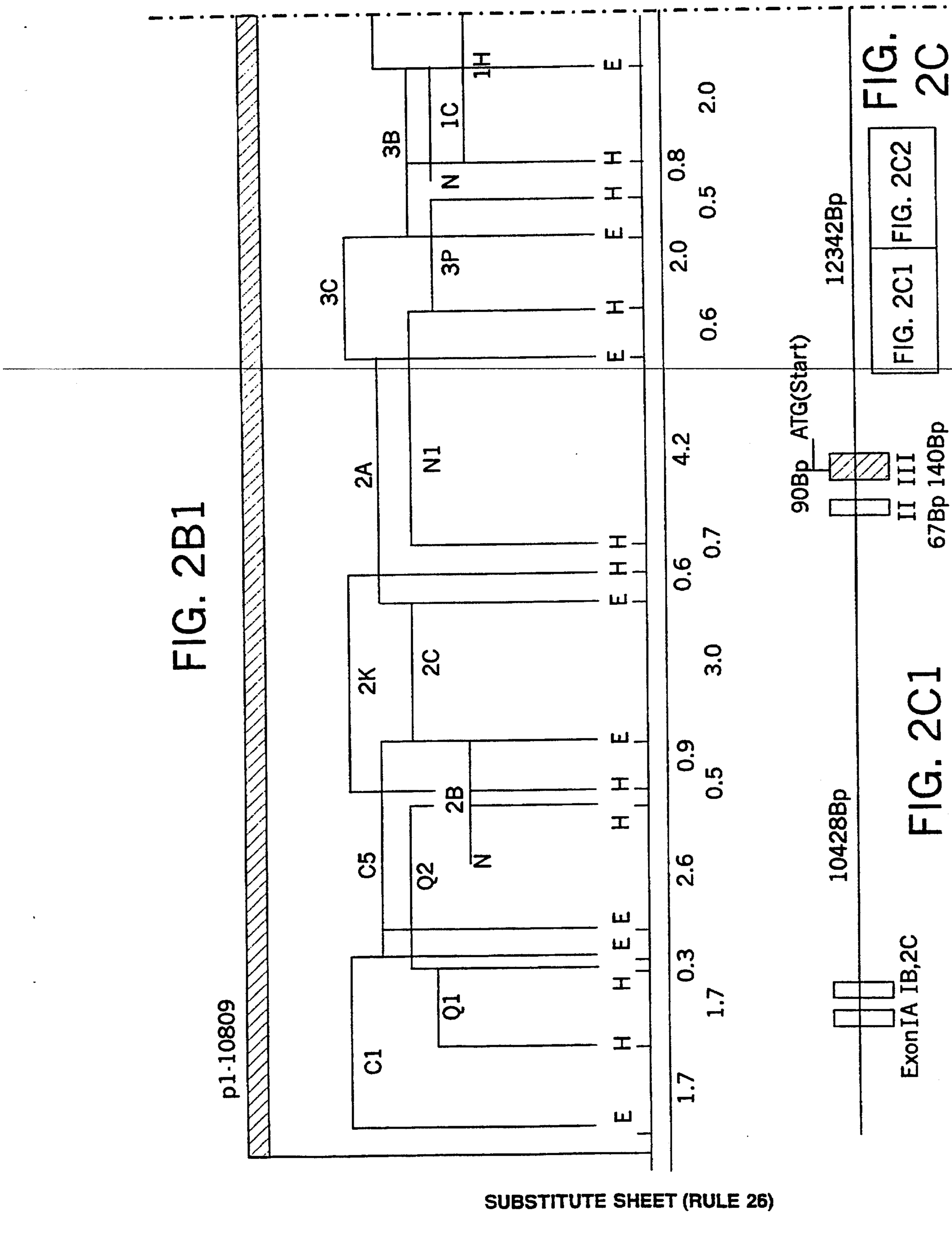


FIG. 2B1
FIG. 2C
FIG. 2C1
FIG. 2C2
p1-10809
C1
Q1
C5
Q2
2K
2C
2B
N
2A
N1
3C
3P
3B
1C
1H
E
H
N
2.0
0.8
0.5
0.6
4.2
0.6
0.7
3.0
0.9
0.5
2.6
0.3
1.7
1.7
12342Bp
90Bp ATG(Start)
II III
67Bp 140Bp
10428Bp
ExonIA IB,2C

FIG. 2B

FIG. 2B1	FIG. 2B2

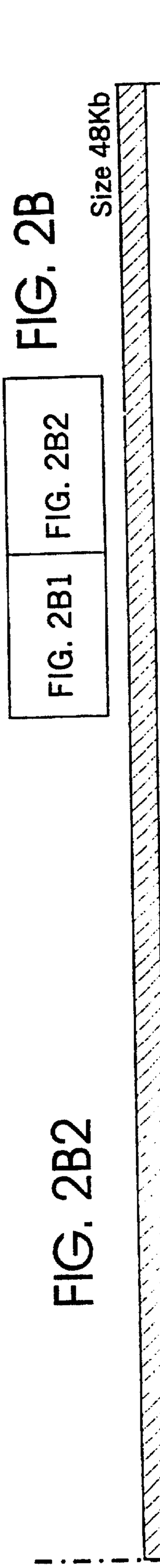


FIG. 2C2

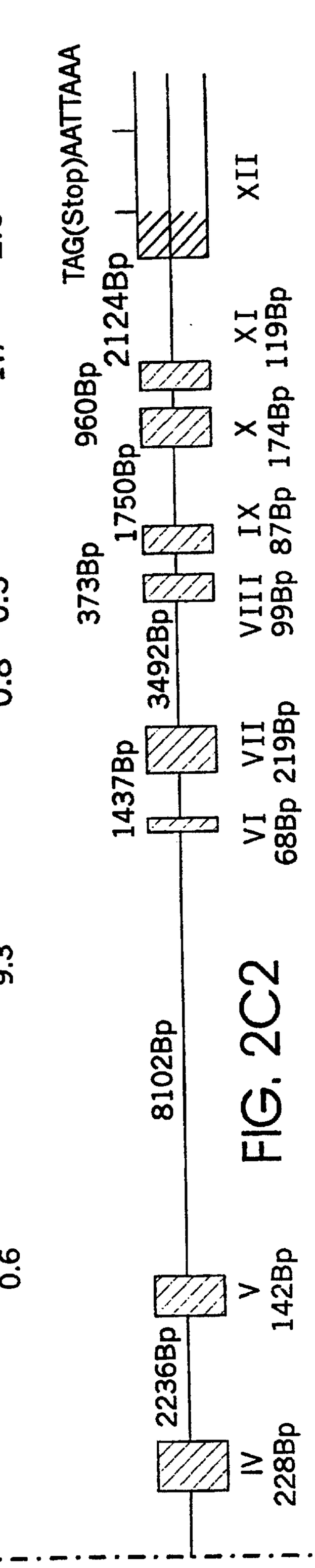


FIG. 3

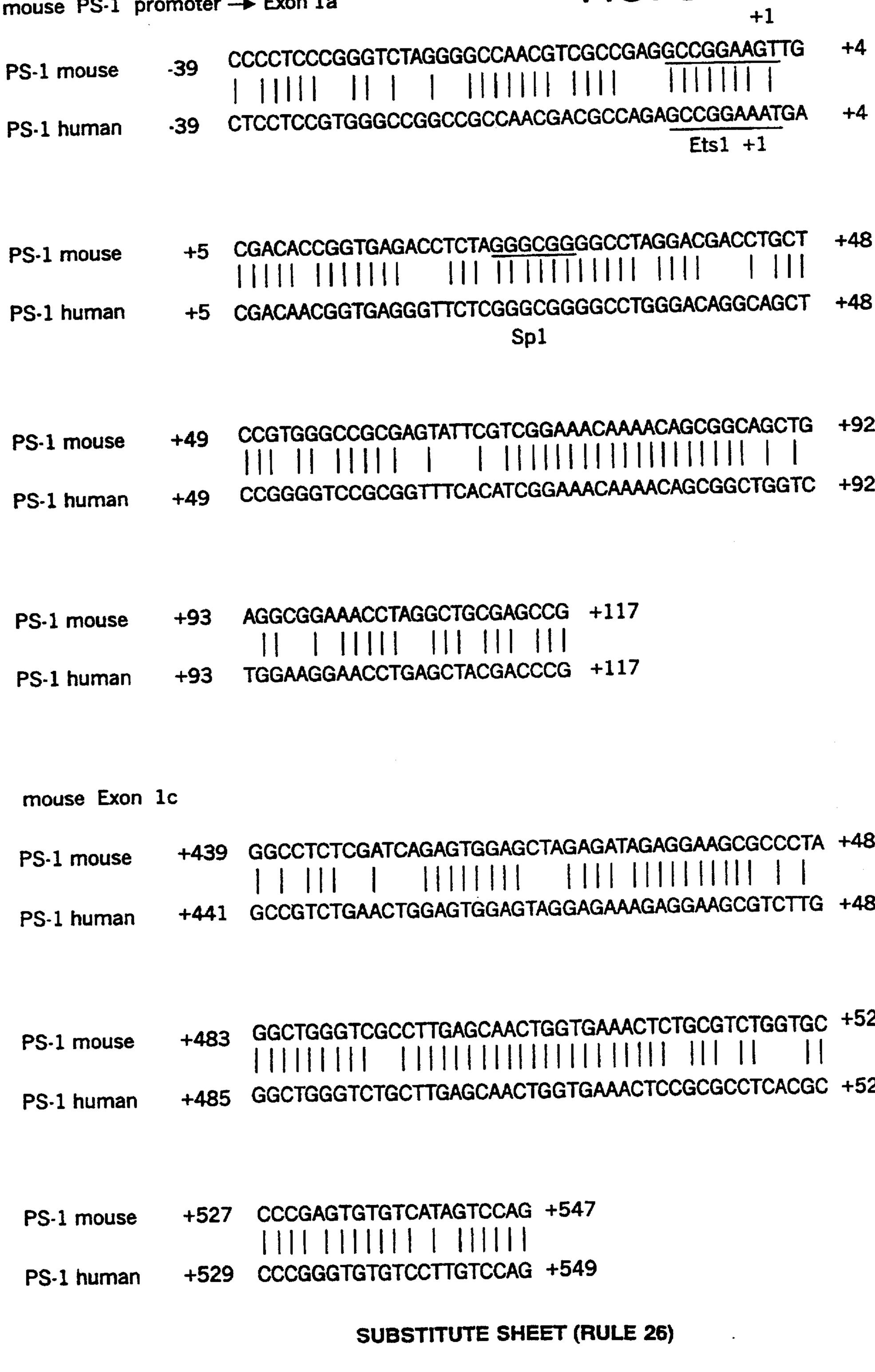
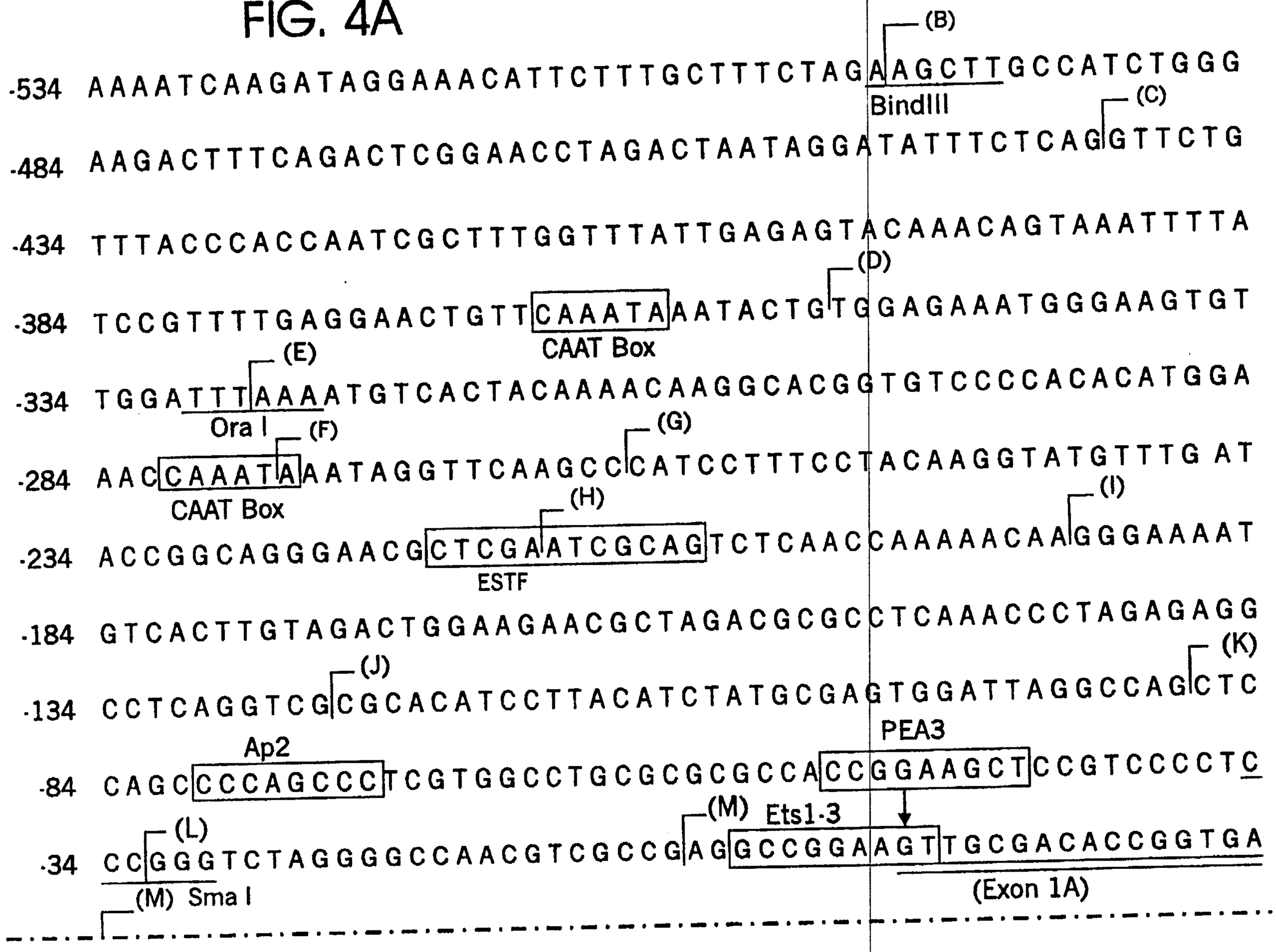

mouse PS-1 promoter → Exon 1a
+1
PS-1 mouse -39 CCCCTCCCGGGTCTAGGGGCCAACGTCGCCGAGGCCGGAAGTTG +4
PS-1 human -39 CTCCTCCGTGGGCCGGCCGCCAACGACGCCAGAGCCGGAAATGA +4
Ets1 +1
PS-1 mouse +5 CGACACCGGTGAGACCTCTAGGGCGGGGCCTAGGACGACCTGCT +48
PS-1 human +5 CGACAACGGTGAGGGTTCTCGGGCGGGGCCTGGGACAGGCAGCT +48
Sp1
PS-1 mouse +49 CCGTGGGCCGCGAGTATTCGTCGGAAACAAAACAGCGGCAGCTG +92
PS-1 human +49 CCGGGGTCCGCGGTTTCACATCGGAAACAAAACAGCGGCTGGTC +92
PS-1 mouse +93 AGGCGGAAACCTAGGCTGCGAGCCG +117
PS-1 human +93 TGGAAGGAACCTGAGCTACGACCCG +117
mouse Exon 1c
PS-1 mouse +439 GGCCTCTCGATCAGAGTGGAGCTAGAGATAGAGGAAGCGCCCTA +482
PS-1 human +441 GCCGTCTGAACTGGAGTGGAGTAGGAGAAAGAGGAAGCGTCTTG +484
PS-1 mouse +483 GGCTGGGTCGCCTTGAGCAACTGGTGAAACTCTGCGTCTGGTGC +526
PS-1 human +485 GGCTGGGTCTGCTTGAGCAACTGGTGAAACTCCGCGCCTCACGC +528
PS-1 mouse +527 CCCGAGTGTGTCATAGTCCAG +547
PS-1 human +529 CCCGGGTGTGTCCTTGTCCAG +549

FIG. 4A



SUBSTITUTE SHEET (RULE 26)

SUBSTITUTE SHEET (RULE 26)

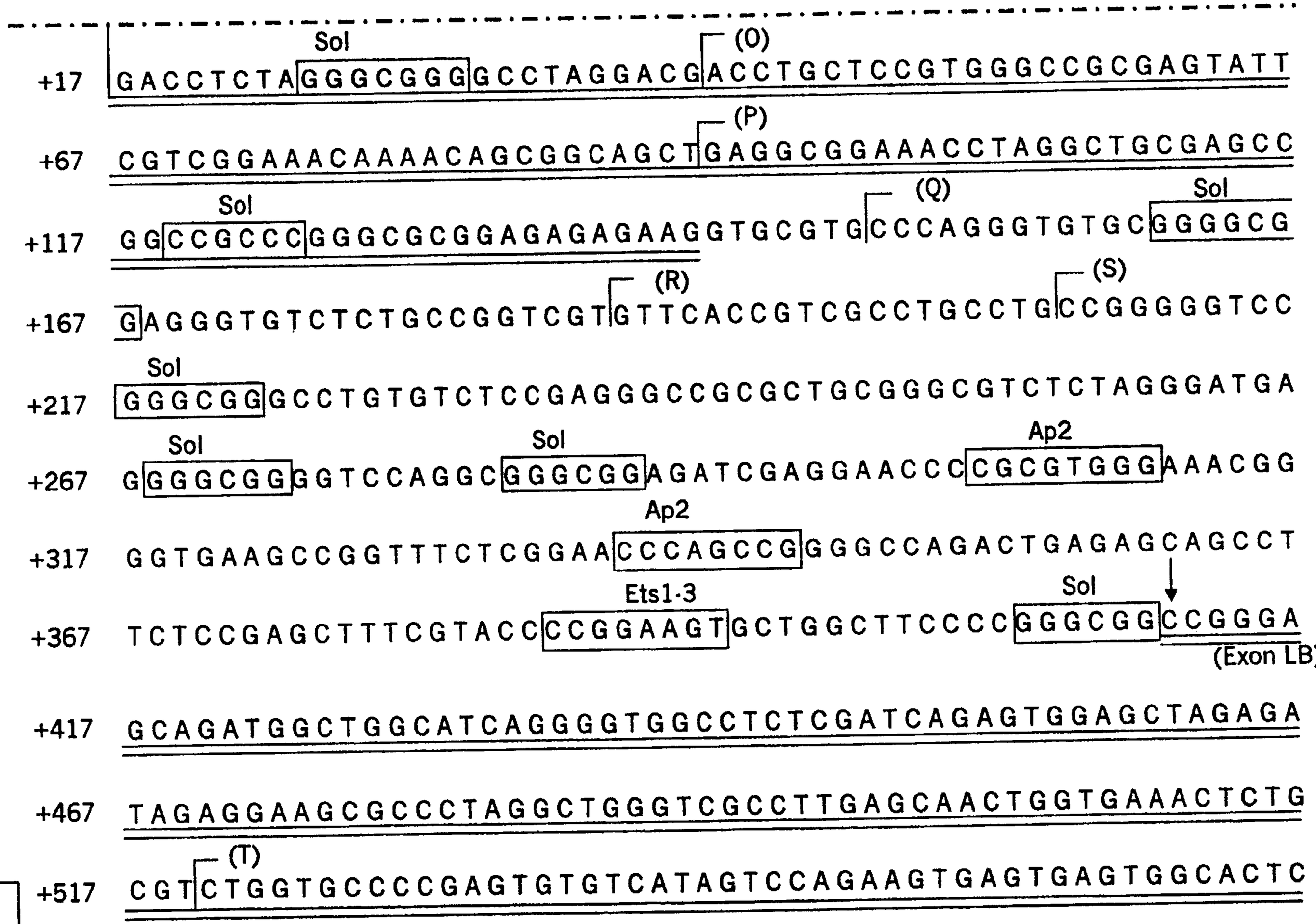


FIG. 4A

FIG. 4B

FIG. 4

FIG. 4B

FIG. 5A

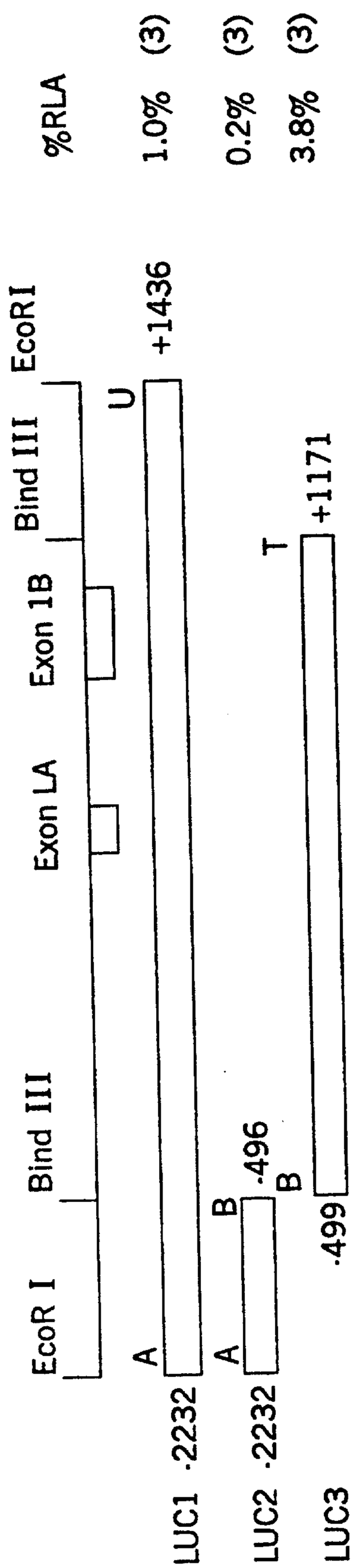

%RLA
1.0%
0.2%
3.8%
EcoRI
Bind III
Exon 1B
U
+1436
T
+1171
Exon LA
Bind III
EcoR I
A
B
.496
.499
.2232
LUC1
LUC2
LUC3

FIG. 5B

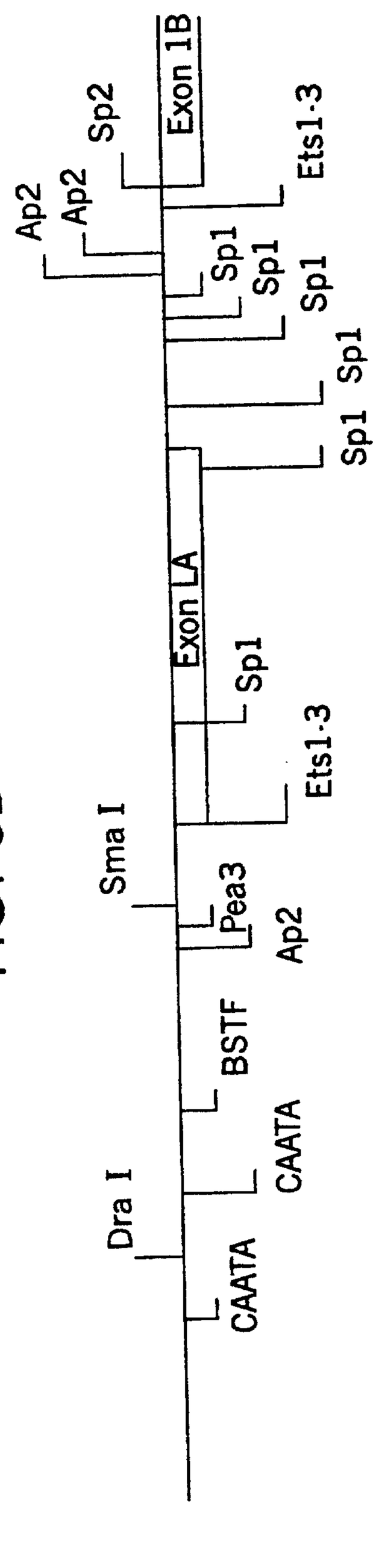

Ap2
Ap2
Sp2
Exon 1B
Sp1
Sp1
Sp1
Ets1-3
Sp1 Sp1
Exon LA
Sp1
Sma I
Pea3
Ap2
Ets1-3
BSTF
Dra I
CAATA
CAATA

FIG. 5C

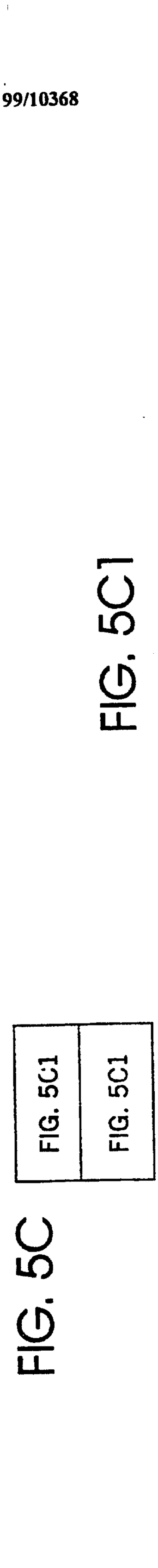


FIG. 5C1

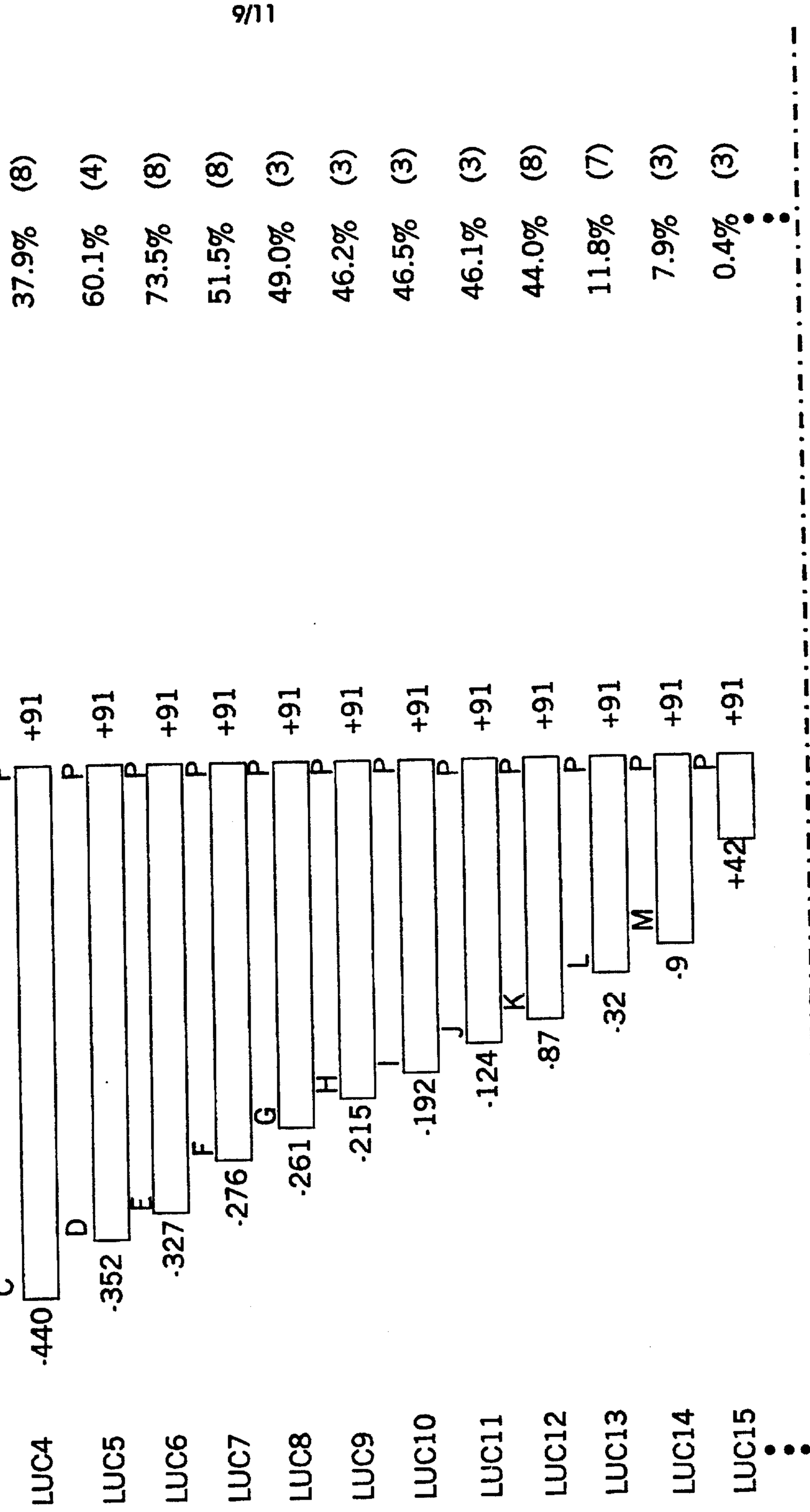


FIG. 5C2

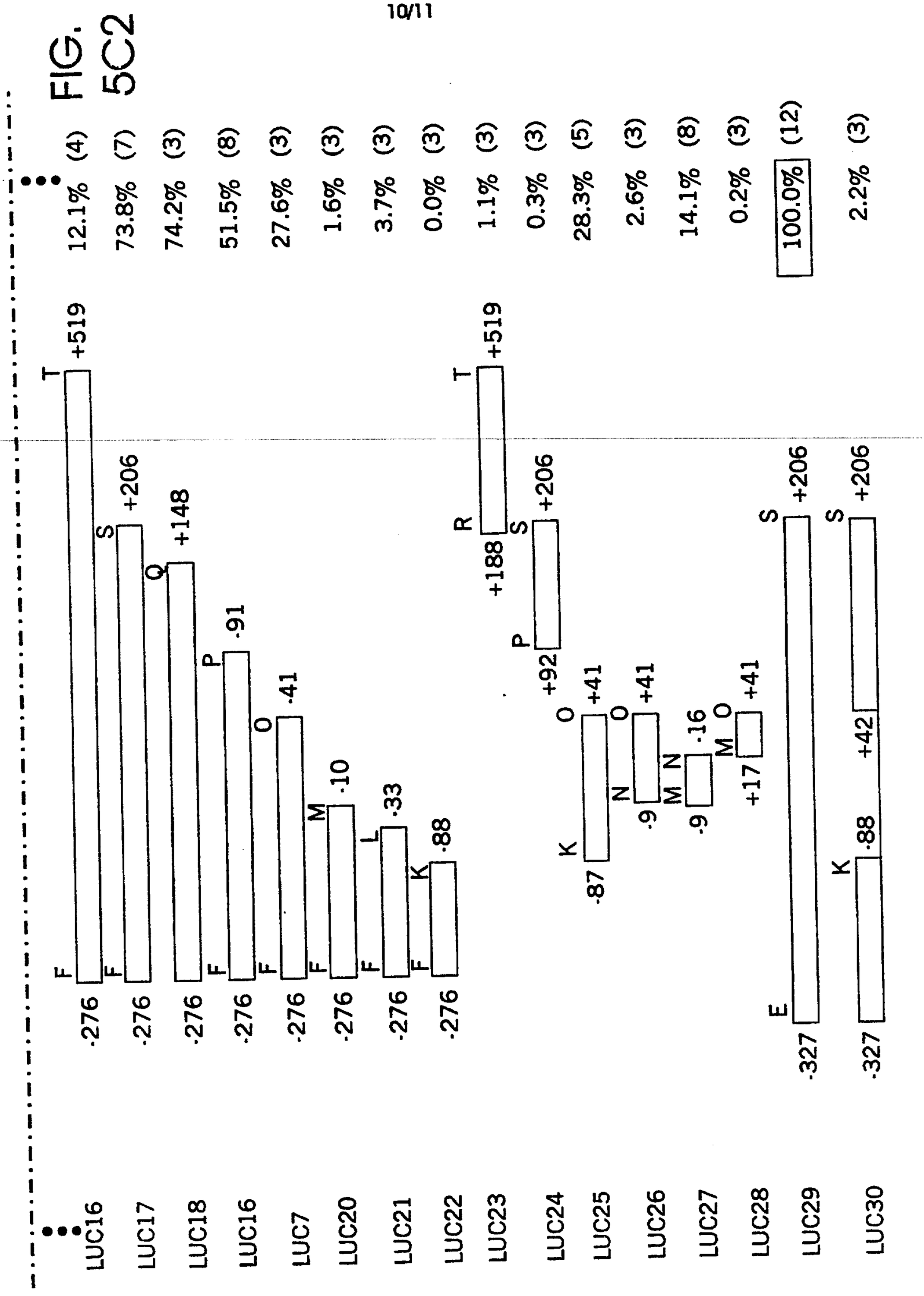
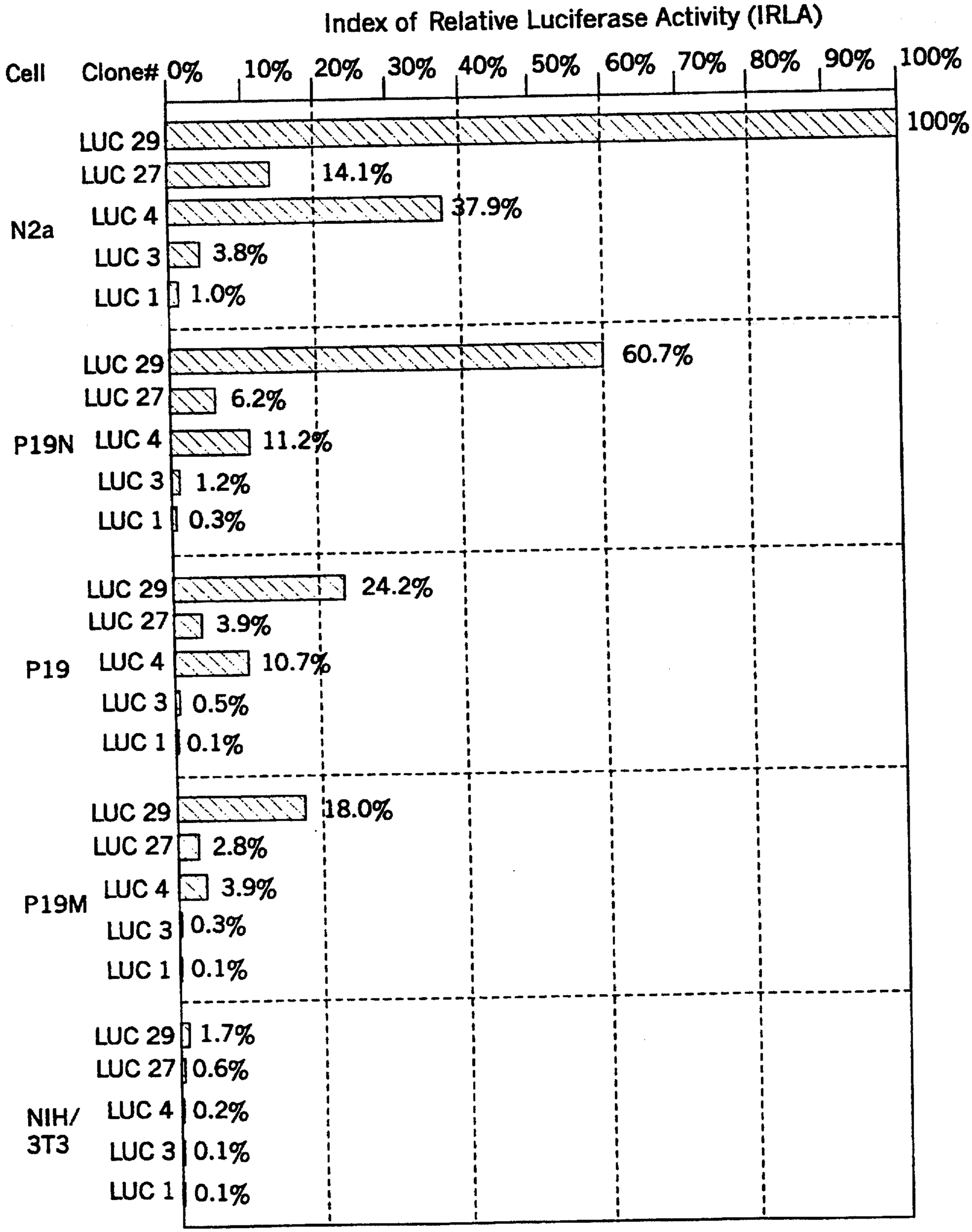

LUC16
LUC17
LUC18
LUC16
LUC7
LUC20
LUC21
LUC22
LUC23
LUC24
LUC25
LUC26
LUC27
LUC28
LUC29
LUC30
12.1% (4)
73.8% (7)
74.2% (3)
51.5% (8)
27.6% (3)
1.6% (3)
3.7% (3)
0.0% (3)
1.1% (3)
0.3% (3)
28.3% (5)
2.6% (3)
14.1% (8)
0.2% (3)
100.0% (12)
2.2% (3)
T +519
S +206
Q +148
P 91
F .276
O .41
M .10
L .33
K .88
R +188
P 92
O +41
K .87
N +41
.9
M N .16
M O +17
S +206
E .327
K .88
+42
.327

FIG. 6


Index of Relative Luciferase Activity (IRLA)
Cell
Clone#
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
N2a
LUC 29
100%
LUC 27
14.1%
LUC 4
37.9%
LUC 3
3.8%
LUC 1
1.0%
P19N
LUC 29
60.7%
LUC 27
6.2%
LUC 4
11.2%
LUC 3
1.2%
LUC 1
0.3%
P19
LUC 29
24.2%
LUC 27
3.9%
LUC 4
10.7%
LUC 3
0.5%
LUC 1
0.1%
P19M
LUC 29
18.0%
LUC 27
2.8%
LUC 4
3.9%
LUC 3
0.3%
LUC 1
0.1%
NIH/ 3T3
LUC 29
1.7%
LUC 27
0.6%
LUC 4
0.2%
LUC 3
0.1%
LUC 1
0.1%