



(86) Date de dépôt PCT/PCT Filing Date: 2001/11/16
(87) Date publication PCT/PCT Publication Date: 2002/09/06
(45) Date de délivrance/Issue Date: 2014/07/08
(85) Entrée phase nationale/National Entry: 2003/05/16
(86) N° demande PCT/PCT Application No.: US 2001/043588
(87) N° publication PCT/PCT Publication No.: 2002/068666
(30) Priorité/Priority: 2000/11/17 (US60/249,608)

(51) Cl.Int./Int.Cl. *C12N 15/82* (2006.01),
C12N 9/18 (2006.01), *C12N 9/24* (2006.01)

(72) Inventeurs/Inventors:
DUNN-COLEMAN, NIGEL, US;
LANGDON, TIMOTHY, GB;
MORRIS, PHILLIP, GB

(73) Propriétaires/Owners:
GENENCOR INTERNATIONAL, INC., US;
INSTITUTE OF GRASSLAND AND ENVIRONMENTAL
RESEARCH, GB

(74) Agent: BERESKIN & PARR LLP/S.E.N.C.R.L.,S.R.L.

(54) Titre : MANIPULATION DE LA TENEUR EN ACIDE PHENOLIQUE ET DE LA DIGESTIBILITE DE PAROIS DE
CELLULES VEGETALES PAR L'EXPRESSION CIBLEE DE GENES CODANT POUR DES ENZYMES DEGRADANT
LES PAROIS CELLULAIRES

(54) Title: MANIPULATION OF THE PHENOLIC ACID CONTENT AND DIGESTIBILITY OF PLANT CELL WALLS BY
TARGETED EXPRESSION OF GENES ENCODING CELL WALL DEGRADING ENZYMES

(57) **Abrégé/Abstract:**

Described herein are methods to enhance the production of more highly fermentable carbohydrates in plants, especially forage grasses. The invention provides for transgenic plants transformed with expression vectors containing a DNA sequence encoding ferulic acid esterase I from *Aspergillus*, preferably *A. Niger*. The expression vectors may optionally comprise a DNA sequence encoding xylanase from *Trichoderma*, preferably *T. reesei*. Expression of the enzyme(s) is targeted to specific cellular compartments, in specific cellular compartments, in specific tissues and under specific environmental conditions. Uses of this invention include, but are not limited to, forage with improved digestibility for livestock, and enhanced biomass conversion.



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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
6 September 2002 (06.09.2002)

PCT

(10) International Publication Number
WO 02/068666 A1

(51) International Patent Classification⁷: **C12N 15/82**

(21) International Application Number: PCT/US01/43588

(22) International Filing Date:
16 November 2001 (16.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/249,608 17 November 2000 (17.11.2000) US

(71) Applicant (*for all designated States except US*): **GENENCOR INTERNATIONAL, INC.** [US/US]; 925 Page Mill Road, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **DUNN-COLEMAN, Nigel** [US/US]; 142 Johnson Avenue, Los Gatos, CA 95032 (US). **LANGDON, Timothy** [GB/GB]; Troedyrhiw, Lon Tyllwyd, Llanfarian, Aberystwyth SY23 4UJ (GB). **MORRIS, Phillip** [GB/GB]; 6 Trefaenor, Comins-coch, Aberystwyth SY23 3UB, Cerdigion, Wales (GB).

(74) Agent: **BOYD, Victoria, L.**; Genencor International, Inc., 925 Page Mill Road, Palo Alto, CA 94304-1013 (US).

(81) Designated States (*national*): AE, AL, AM, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ (utility model), DE (utility model), DK (utility model), DM, EE (utility model), ES, FI (utility model), GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: MANIPULATION OF THE PHENOLIC ACID CONTENT AND DIGESTIBILITY OF PLANT CELL WALLS BY TARGETED EXPRESSION OF GENES ENCODING CELL WALL DEGRADING ENZYMES

(57) Abstract: Described herein are methods to enhance the production of more highly fermentable carbohydrates in plants, especially forage grasses. The invention provides for transgenic plants transformed with expression vectors containing a DNA sequence encoding ferulic acid esterase I from *Aspergillus*, preferably *A. Niger*. The expression vectors may optionally comprise a DNA sequence encoding xylanase from *Trichoderma*, preferably *T. reesei*. Expression of the enzyme(s) is targeted to specific cellular compartments, in specific cellular compartments, in specific tissues and under specific environmental conditions. Uses of this invention include, but are not limited to, forage with improved digestibility for livestock, and enhanced biomass conversion.



WO 02/068666 A1

5

10 **MANIPULATION OF THE PHENOLIC ACID CONTENT AND DIGESTIBILITY
OF PLANT CELL WALLS BY TARGETED EXPRESSION OF GENES
ENCODING CELL WALL DEGRADING ENZYMES**

15

**Inventors: Nigel Dunn-Coleman
Timothy Langdon
Phillip Morris**

CROSS-REFERENCE TO RELATED APPLICATIONS

20 Pursuant to 35 U.S.C. §119(e), the present application claims benefit of
and priority to USSN 60/249,608, entitled "MANIPULATION OF THE PHENOLIC
ACID CONTENT AND DIGESTIBILITY OF FORAGE GRASS CELL WALLS BY
TARGETED EXPRESSION OF A FERULIC ACID ESTERASE GENE", filed
November 17, 2000, by Morris et al.

FIELD OF THE INVENTION

25 This invention relates to methods to enhance to availability of fermentable
carbohydrates.

BACKGROUND OF THE INVENTION

30 The present crisis in livestock agriculture has prompted a resurgence of
interest in grass-fed animals. However, while a high-forage diet may be desirable, it
does not currently satisfy the demands of modern animal production. For the animal
to make efficient use of the forage it consumes, the energy demands of the micro-
organisms in the rumen must be met and synchronized with the availability of plant
proteins. Otherwise this lack of synchrony will lead to (a) proteins and other
35 nutrients being poorly utilized in the rumen, (b) loss of nitrogen, in urine and feces
and therefore, the environment and (c) the need to feed excessive amounts of
protein concentrates as supplements to the ruminant diet.

Cellulose and hemicellulose in grass and maize tissues could meet the

energy requirements of the ruminant or provide new feed-stocks for industrial fermentation to ethanol. This potential is not currently realized because the cell walls are lignified and the cell wall polysaccharides highly cross-linked with phenolic residues and lignin, resulting in low rates of plant cell wall digestion in comparison to rates of protein breakdown in ruminants. This is a particular problem for the most important forages in Europe, the ryegrasses *Lolium perenne* and *L. multiflorum* as well as one of the major impediments to the wider use of better adapted species, such as *Festuca arundinacea*, as a forage crop. Increasing the digestibility index of grasses has therefore been a major breeding objective for several decades but progress has been slow due to difficulties in fixing natural variation in the synthetic varieties derived from these outbreeding species (Hayward, *et al.*, TAG 70:48 (1985)).

Removing labile phenolics by chemical treatment with alkali is known to increase the biodegradability and nutritional value of low-quality feed such as cereal straw, and is employed commercially for feed upgrading. Reducing phenolic cross-linking of cell wall carbohydrates is therefore a predictable way of improving the rate of digestion and digestibility of ryegrass. However chemical modification may have other disadvantages. Therefore, genetic modification would be a preferable method of changing the cell wall chemistry of highly digestible varieties. Many in the field are pursuing this approach. An alternative, however, is to use genetic modification to reduce the levels of phenolic acids in the cell walls available for crosslinking either by directly disrupting ester bonds linking phenolics and lignins to cell wall polysaccharides or by preventing excessive ferulation of cell wall carbohydrates prior to their incorporation into the cell wall.

This invention meets this and other needs by using targeted or inducible expression of cell wall degrading enzymes in plants.

SUMMARY OF THE INVENTION

Provided herein are methods for enhancing the availability of fermentable carbohydrates. In one aspect, there is provided an expression cassette
5 comprising a DNA sequence encoding at least one cell wall degrading enzyme. The DNA sequence encoding at least one cell wall degrading enzyme may be operatively linked to a promoter sequence. The promoter may be constitutive or inducible. The expression cassette may further comprise a targeting sequence.

In one embodiment, the cell wall degrading enzyme is selected from the
10 group consisting of ferulic acid esterase, xylanase, xylosidase, cellulase, endoglucanase, and cellbiohydrolase. In a preferred embodiment cell wall degrading enzyme is derived from a fungal source. In a more preferred embodiment, the fungal ferulic acid esterase is an *Aspergillus* ferulic acid esterase, preferably *A. niger*. In another embodiment the xylanase is derived
15 from *Trichoderma*, preferably *T. reesei*.

In another aspect of the invention, there is provided a plant transformed with the expression cassette comprising a DNA sequence encoding at least one cell degrading enzyme. The plant may be selected from the group consisting of Festuca, Lolium, Avena and Zea. In a preferred embodiment the plant is a
20 forage grass. In another embodiment, the plant is maize.

Further provided herein is a method of controlling the level of phenolic acids in plant cell walls of a transgenic plant. The method, in one embodiment, comprises introducing to a plant cell an expression cassette comprising a DNA sequence encoding at least one cell wall degrading enzyme, preferably a ferulic
25 acid esterase.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various
30 changes and modifications within the scope and spirit of the invention will become apparent to one skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 illustrates a restriction map of a DNA fragment containing the gene encoding the 38kd ferulic acid esterase.

Figures 2 A-E illustrate the complete DNA (SEQ. ID NO: 1), with highlighting to point out the signal sequence, intron and various restriction endonuclease sites, and amino acid sequence (SEQ. ID. NO: 2) corresponding to the gene encoding the 38 kD ferulic acid esterase isolated from *Aspergillus niger*.

Figure 3 illustrates the DNA sequence of the gene encoding the 38 kD esterase (SEQ. ID. NO: 1).

Figure 4 illustrates the construction of the intronless ferulic acid esterase isolated from *Aspergillus niger*.

Figure 5 illustrates that the overlapping of PCR products made with primers FAE-I5 and FAE-I3 creates two possible uninterrupted reading frames – the top in the figure below is functional (highlighted serine is at active site), the bottom is inactivated. (SEQ ID NOS: 3, 4, 95 & 96)

Figure 6 illustrates the possible vector constructions useful in the present invention. Various combinations are possible. Although and FAE gene is depicted another cell wall degrading enzyme may be used alone (i.e., instead of) or in conjunction with the FAE gene. Amp = ampicillin resistance gene.

Figure 7 illustrates pCOR105.

Figure 8 illustrates a generic ALE-TER vector. (SEQ ID NOS: 5 & 6)

Figure 9 illustrates the KDEL-COOH ER retention sequences. (SEQ ID NOS: 7 & 8)

Figure 10 illustrates the FAE-LINKER-FRAMESHIFT structure and sequence.

Figure 11 illustrates plant transformation cassettes.

Figure 12 is a table of the vectors used herein.

Figure 13 depicts the barley aleurain vacuolar and apoplast signal sequence. (SEQ ID NOS: 9 & 10)

Figure 14 illustrates the rat sialyl transferase structure and sequence. (SEQ ID NOS: 11 & 12)

Figure 15 illustrates the potato protease inhibitor II (PPI) motif structure and sequence. (SEQ ID NOS: 13 & 14)

Figure 16 illustrates the targeted expression of gfp to different cell compartment. Also shown are schematics of the vectors used.

Figure 17 illustrates the FAE activity in transgenic *Festuca arundinacea* leaves of different ages under ER and APO targeting sequences.

5 Figure 18 illustrates the FAE activity in transgenic *Festuca arundinacea* leaves of different ages under Vac targeting sequence.

Figure 19 illustrates the FAE activity in transgenic *Lolium mutflorum* leaves of different ages.

10 Figure 20 illustrates the FAE activity in transgenic *Lolium mutflorum* leaves under Vac, ER and APO targeting sequences.

Figure 21 illustrates the levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under Vac targeting sequence.

15 Figure 22 illustrates the levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under APO and ER targeting sequence.

Figure 23 illustrates the *in vitro* dry matter digestibility of leaf tissue of mature *Festuca arundinacea* plants expressing FAE under an actin promoter.

20 Figure 24 illustrates the *in vitro* dry matter digestibility of leaf tissue of mature *Lolium mutflorum* plants expressing FAE under an actin promoter.

Figure 25 illustrates the rate of fermentation and cumulative gas production in *Festuca arundinacea* cells.

Figure 26 illustrates the *in vitro* fermentation of *Festuca arundinacea* cell walls from cell cultures expressing recombinant FAE1.

25 Figure 27 illustrates the Time to maximum rate digestion for *Festuca arundinacea* cells.

Figure 28 illustrates the total gas production in *Festuca arundinacea* cells.

Figure 29 illustrates the kinetics of FAE activity by ferulic acid release from cell wall under self digestion in *Festuca arundinacea* and stimulation by xylanase.

30 Figure 30 illustrates the beta-glucoronidase activity under the *Lolium* See1 senescence promoter in leaves of transgenic plants of *Lolium mutflorum*.

Figure 31 illustrates the release of monomeric and dimeric HCAs on self digestion of leaves of vacuolar targeted FAE expressing plants.

Figure 32 is a schematic of the pTP10-1 vector. Also shown is the 5338 bp nucleotide sequence of the vector. (SEQ ID NOS: 15 & 16)

5 Figure 33 is a schematic of the pUA4-4 vector. Also shown is the 5345 bp nucleotide sequence of the vector. (SEQ ID NOS: 17 & 18)

Figure 34 is a schematic of the pTU4 vector. Also shown is the 5337 bp nucleotide sequence of the vector. (SEQ ID NOS: 19 & 20)

10 Figure 35 is a schematic of the pTT5.14 vector. Also shown is the 5395 bp nucleotide sequence of the vector. (SEQ ID NOS: 21 & 22)

Figure 36 is a schematic of the pTP8-5 vector. Also shown is the 5337 bp nucleotide sequence of the vector. (SEQ ID NOS: 23 & 24)

Figure 37 is a schematic of the pTP5-1 vector. Also shown is the 5277 bp nucleotide sequence of the vector. (SEQ ID NOS: 25 & 26)

15 Figure 38 is a schematic of the pTP4a2 vector. Also shown is the 5327 bp nucleotide sequence of the vector. (SEQ ID NOS: 27, 28 & 94)

Figure 39 is a schematic of the pTP3-1 vector. Also shown is the 5338 bp nucleotide sequence of the vector. (SEQ ID NOS: 29 & 30)

20 Figure 40 is a schematic of the pTU5 vector. Also shown is the 5337 bp nucleotide sequence of the vector. (SEQ ID NO: 31)

Figure 41 is a schematic of the pGT6 vector. Also shown is the 4773 bp nucleotide sequence of the vector. (SEQ ID NO: 32)

Figure 42 is a schematic of the pJQ5 vector. Also shown is the 5034 bp nucleotide sequence of the vector. (SEQ ID NO: 33)

25 Figure 43 is a schematic of the pJO6.1 vector. Also shown is the 4950 bp nucleotide sequence of the vector. (SEQ ID NO: 34)

Figure 44 is a schematic of the pJQ4 vector. Also shown is the 4974 bp nucleotide sequence of the vector. (SEQ ID NO: 35)

30 Figure 45 is a schematic of the pPQ10.1 vector. Also shown is the 5164 bp nucleotide sequence of the vector. (SEQ ID NO: 36)

Figure 46 is a schematic of the pJQ3 vector. Also shown is the 4965 bp nucleotide sequence of the vector. (SEQ ID NO: 37)

Figure 47 is a schematic of the pUG4 vector. Also shown is the 5295 bp nucleotide sequence of the vector.

Figure 48 is a schematic of the pUB8.11 vector. Also shown is the 5001 bp nucleotide sequence of the vector.

5 Figure 49 is a schematic of the pTP11-1 vector. Also shown is the 5387 bp nucleotide sequence of the vector.

Figure 50 illustrates the actin promoter and its corresponding nucleotide sequence.

10 Figure 51 illustrates the Aleurain-NPIR delete structure. The corresponding nucleotide sequences are also shown.

Figure 52 illustrates the SEE1 (senescence enhanced) promoter sequence.

Figure 53 illustrates the SEE1 (senescence enhanced) promoter sequence plus the vacuolar aleurain signal/NPIR sequence.

15

DETAILED DESCRIPTION OF THE INVENTION

The invention will now be described in detail by way of reference only using the following definitions and examples.

20

Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton, *et al.*, DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2D ED., John Wiley and Sons, New York (1994), and Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) provide one of skill with a general dictionary of many of the terms used in this invention. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy

25

30

orientation, respectively. Practitioners are particularly directed to Sambrook J. *et al.*, Molecular Cloning: a laboratory manual. 2nd ed. N.Y., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, 1989, 1659 p. ISBN 0-87969-309-6, and Ausubel F.M. *et al.*, Current protocols in molecular biology, 1993. John Wiley and Sons, Inc., for definitions
5 and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary.

The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully
10 defined by reference to the specification as a whole.

Definitions

It should be noted that, as used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to a composition
15 containing "a compound" includes a mixture of two or more compounds. It should also be noted that the term "or" is generally employed in its sense including "and/or" unless the content clearly dictates otherwise.

"Conservatively modified variants" applies to both amino acid sequences and polynucleotides. With respect to particular polynucleotides, conservatively
20 modified variants refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical polynucleotides encode any given protein. For instance, the codons GCA, GCC,
25 GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every polynucleotide herein which encodes a
30 polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a polynucleotide (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a polynucleotide which encodes a polypeptide is implicit in each described

sequence. For purposes of protein expression, there are "sub-optimal codons." These are codons that are not preferred by a particular genus or species. Altering these "sub-optimal codons" to "preferred codons" is a silent mutation in that the amino acid encoded by the codons is the same but one codon is
 5 preferentially expressed by the particular genus, *e.g.*, *Triticum spp.*

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a polynucleotide, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively
 10 modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

- 15 1) Alanine (A), Serine (S), Threonine (T);
 - 2) Aspartic acid (D), Glutamic acid (E);
 - 3) Asparagine (N), Glutamine (Q);
 - 4) Arginine (R), Lysine (K);
 - 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
 - 20 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).
- (*see, e.g.*, Creighton T.E., *Proteins: Structure and Molecular Properties*, New York: W. H. Freeman and Co. (1984)).

"Pyroglutamic acid" is the cyclized internal amide of L-glutamic acid

The phrase "controlling the level of phenolic acids" refers to the
 25 manipulation of phenolic acid expression in plants, particularly plant cell walls. The manipulation can be either positive; *e.g.*, increasing the levels of phenolic acids; negative, *e.g.*, decreasing the level of phenolic acids; or neutral, *e.g.*, changing the relative amounts of specific phenolic acids in the cell walls but keeping the total amount relatively the same. The timing of manipulation can be
 30 during plant growth or after plant growth, *e.g.*, after a plant has been cut or pulled from the ground or ingested. "Plant cell walls" refers to the cell walls of any cell of the plant.

The term "derived" means that a polynucleotide or protein is related to

another polynucleotide or protein. The relations can be one of homology, *e.g.*, nucleotides and proteins from certain species are homologous to similar polynucleotides and proteins of other species; analogy, *e.g.*, proteins perform the same function and therefore are related to each other regardless of organism of origin. The relationship can be a man-made one, *e.g.*, a protein (and a polynucleotide) can be derived from another protein by mutation; or chemical manipulation (peptidomimetics). Furthermore, a protein or a polynucleotide can be derived from an organism if, in the natural state, the protein or polynucleotide is found in one organism but recombinantly produced in another.

The term "exogenous polynucleotide" refers to a polynucleotide which is introduced into the plant by any means other than a sexual cross or sexual reproduction. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, *in planta* techniques, and the like. Such a plant containing the exogenous polynucleotide is referred to here as an R₁ generation transgenic plant. Transgenic plants which arise from sexual cross or by selfing are progeny of such a plant.

The term "isolated polynucleotide molecule" or "isolated protein" refers to a polynucleotide or protein which is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein which is the predominant species present in a preparation is substantially purified. In particular, an isolated *FAE1* gene is separated from open reading frames which flank the gene and encode a protein other than *FAE1*.

A "*FAE1* encoding polynucleotide" is a nucleic acid sequence comprising (or consisting of) a coding region of an *FAE 1* gene or which encodes a *FAE1* polypeptide. *FAE1* polynucleotides can also be identified by their ability to hybridize under low stringency conditions (see below) to nucleic acid probes having a sequence of 8 to 300 bases, preferably a sequence of 80 to 100 bases

in the sequence disclosed in WO 98/14594.

The term "nucleic acid encoding," "nucleic acid sequence encoding" or "polynucleotide encoding" refers to a polynucleotide which directs the expression of a specific protein or peptide. The polynucleotides include both the DNA strand
5 sequence that is transcribed into RNA and the RNA sequence that is translated into protein. The polynucleotides include both full length polynucleotides as well as shorter sequences derived from the full length sequences. It is understood that a particular polynucleotide includes the degenerate codons of the native sequence or sequences which may be introduced to provide codon preference in
10 a specific host cell. The polynucleotide includes both the sense and antisense strands as either individual single strands or in the duplex form.

The term "operably linked" refers to functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

15 The term "plasmid" refers to a circular double stranded DNA molecule which comprises the coding sequence of interest, regulatory elements, a selection marker and optionally an amplification marker. A plasmid can transform prokaryotic cells or transfect eukaryotic cells. An "expression cassette" means a portion of a plasmid (or the entire plasmid) containing the regulatory
20 elements desired for transcription, translation and/or expression and the coding region of a polynucleotide. A plasmid may contain one or more expression cassettes. If multiple expression cassettes are introduced into a plant, they can be introduced simultaneously or at different times. If simultaneous introduction is desired, the expression cassettes can be on one plasmid or more. Typically, an
25 expression cassette comprises a promoter, poly A+ tail, and signal sequences that target the expressed polypeptide to a specific region of a cell or to be secreted, if desired. Examples of signal sequences that "target expression" of ferulic acid esterase include sequences located upstream of the FAE coding sequence. The polynucleotide that encodes the signal sequence is found
30 preferably within the 100 nucleotides "upstream" (in the 5' direction) from the initiation codon (AUG). More preferably, the polynucleotide that encodes the signal sequence is found within the 50 nucleotides upstream from the initiation

codon. Many different cellular organelles are targeted by the signal sequences used in this invention. The organelles include, but are not limited to, vacuoles, Golgi apparatus, endoplasmic reticula, and apoplasts. In addition to upstream signal sequences, the expression cassette of this invention may include a
5 polynucleotide that encodes a signal sequence at the 3' end. These signal sequences include, but are not limited to stop codons and the KDEL sequence. In addition to KDEL, other similar sequences are contemplated by this invention, including but not limited to RDEL. In addition to a KDEL sequence, a signal sequence can include a linker to a KDEL sequence. A linker is an extension of
10 the reading frame of the encoding polynucleotide to the signal sequence. Preferably, the polynucleotide encoding the signal sequence is directly downstream from the coding sequence, more preferably less than 100 base pairs from the stop codon, more preferably less than 20 base pairs from the stop codon.

15 The term "polynucleotide," "polynucleotide" or "nucleic acid sequence" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses polynucleotides containing known analogues of natural nucleotides which have similar binding properties as the reference polynucleotide
20 and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular *FAE1* polynucleotide of this invention also implicitly encompasses conservatively modified variants thereof (*e.g.* degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions
25 may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzler M.A. *et al.*, Enhanced evolutionary PCR using oligonucleotides with inosine at the 3'-terminus. *Nucleic Acids Res.* 1991 Sep 25;19(18):5081; Ohtsuka E. *et al.*, An alternative approach to deoxyoligonucleotides as hybridization probes by insertion of deoxyinosine at ambiguous codon positions. *J Biol Chem.* 1985 Mar 10;260(5):2605-8;
30 Cassol *et al.* Rapid DNA fingerprinting to control for specimen errors in HIV testing by the polymerase chain reaction. *Molecular and Cellular Probes.* 1992 Apr;6: 327-331; and Rossolini G.M. *et al.*, Use of deoxyinosine-containing primers vs degenerate primers for polymerase chain reaction based on ambiguous sequence information. *Mol Cell Probes.* 1994 Apr;8(2):91-8). The term polynucleotide is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

The term "polypeptide," "peptide," and "protein" are used interchangeably and refer to amino acids connected by peptide bonds. Polypeptides can be

entire proteins or portions thereof. For Example, a FAE1 polypeptide may refer to the entire FAE1 protein or fragments of the FAE1 protein. A "ferulic acid esterase with an altered glycosylation site" refers to a FAE protein wherein a mutation has changed the glycosylation pattern of the protein. Mutations that effect such changes are well known in the art and include, but are not limited to, amino acid substitutions, and mutations in the proteins of the Golgi apparatus and endoplasmic reticulum that effect glycosylation of proteins.

The term "promoter" refers to a polynucleotide that directs expression of a coding sequence. A promoter can be constitutive, *i.e.*, relatively independent of the stage of differentiation of the cell in which it is contained or it can be inducible, *i.e.*, induced by specific environmental factors, such as the length of the day, the temperature, *etc.* or a promoter can be tissue-specific, *i.e.*, directing the expression of the coding sequence in cells of a certain tissue type. A "senescence" promoter is an inducible promoter that causes transcription to be initiated upon a certain event relating to age of the organism. A "heat shock promoter" is an inducible promoter that causes transcription to be initiated upon a change in temperature. An example of a heat shock protein promoter is the Soybean Gmhsp promoter. In addition to these inducible promoters, one of skill will realize that other inducible promoters can be used. For example, a wound induced promoter, like LAP. See, US Patent No. 5,962,670.

The term "purified" denotes that a polynucleotide or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the polynucleotide or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

The term "specifically hybridizes" refers to a nucleic acid probe that hybridizes, duplexes or binds to a particular target DNA or RNA sequence when the target sequences are present in a preparation of total cellular DNA or RNA. "Complementary" or "target" nucleic acid sequences refer to those nucleic acid sequences which selectively hybridize to a nucleic acid probe. Proper annealing conditions depend, for example, upon a probe's length, base composition, and the number of mismatches and their position on the probe, and must often be determined empirically. For discussions of nucleic acid probe design and

annealing conditions, see, for example, Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989) ("Sambrook") or CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, F. Ausubel *et al.*, ed. Greene Publishing and Wiley-Interscience, New York (1987) ("Ausubel").

5 The term "stringent conditions" in the context of polynucleotide hybridization experiments such as Southern and northern hybridizations refers to sequence dependent, binding and washing environments. An extensive guide to the hybridization of polynucleotides is found in Tijssen (1993) LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY--HYBRIDIZATION WITH
10 NUCLEIC ACID PROBES part I chapter 2 "overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and
15 pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent hybridization conditions for hybridization of complementary polynucleotides which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formalin
20 with 1 mg of heparin at between 40 and 50°C, preferably 42°C, with the hybridization being carried out overnight. An example of highly stringent wash conditions is 0.15M NaCl at from 70 to 80°C with 72°C being preferable for about 15 minutes. An example of stringent wash conditions is a 0.2x SSC wash at about 60 to 70°C, preferably 65°C for 15 minutes (see, Sambrook, *supra* for a
25 description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1x SSC at 40 to 50°C, preferably 45°C for 15 minutes. An example low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4-6x SSC at 35 to 45°C, with 40°C
30 being preferable, for 15 minutes. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Polynucleotides which do

not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, e.g., when a copy of a polynucleotide is created using the maximum codon degeneracy permitted by the genetic code.

5 The term "transgenic plant" refers to a plant into which exogenous polynucleotides have been introduced and their progeny. Typically, cells of a plant are transformed with the exogenous polynucleotide and a transgenic plant is regenerated from the transformed cells. The regenerated plant is then bred to produce a strain of transgenic plants.

10 "Xylanase" (EC 3.2.1.8) refers to a well described class of glycosyl hydrolases that hydrolyze xylan. Commercial applications of xylanase include the degradation and bleaching of wood pulp for paper making. Xylanase can also be added to animal feed to improve the digestibility of plant matter. Typically, commercial xylanase is derived from fungi. A preferred xylanase is derived from
15 *Trichoderma*.

Preferred Embodiments

Plant cell walls contain a range of alkali-labile ester-linked phenolic acids. In particular, grass cell walls are characterized by the presence of large amounts of esterified ferulic and p-coumaric acids (mainly in their E configurations), linked
20 to arabinoxylans at the C5 of arabinose. These are released as ferulated oligosaccharides (FAX and PAX) by cellulase treatment but *in vivo* provide a substrate for peroxidase-catalyzed cross-linking of cell wall polysaccharides and lignin. The high levels of these phenolic acids and their dimers have a dramatic influence on the mechanical properties, digestibility and rates of digestion of
25 grasses by ruminants.

Previous work has shown that ferulic acid is the predominant p-hydroxycinnamic acid esterified to grass polysaccharide but until recently the only ferulic acid dehydrodimer to have been isolated was 5,5'-diferulic acid. Recently new dehydrodiferulate dimers and cyclobutane-type dimer mixtures have been
30 isolated from plant cell walls (Waldron, *et al.*, *Phytochemical Analysis* 7:305 (1996)). As can be seen in Figure 1, these mixtures are present in large amounts in grass cells. Ether linked ferulic acid-coniferyl alcohol dimers, have also been isolated from cell walls (Jacquet, *et al.*, *Polyphenol Comm. Bordeaux*

pp451 (1996)) establishing for the first time that ferulate esters are oxidatively co-polymerized with lignin precursors which may anchor lignins to cell wall polysaccharides. The yield of these dimers in grass cells indicates that phenolic dehydrodimer cross-linking of cell wall polysaccharides is much more extensive
5 than was previously thought.

An enzyme system has been reported from parsley endomembranes that catalyses the ferulation of endogenous polysaccharide acceptors from feruloyl CoA, pointing to the ER/golgi as the site of polysaccharide esterification and the CoA ester as the physiological co-substrate (Meyer, *et al.*, *FEBS Lett.* **290**:209
10 (1991)). Further evidence for this has been found in water-soluble extracellular polysaccharides excreted in large amounts into the medium by grass cell cultures. This material is highly esterified with ferulic and p-coumaric acid at levels similar to the cell walls of the cultured cells.

Feruloyl esterase activity has been detected in several fungal species
15 including, anaerobic gut fungi, yeasts, actinomycetes, and a few fiber-degrading ruminal bacteria, which enables them to de-esterify arabinoxylans and pectins.

Two ferulic acid esterases (FAE), distinguished on the basis of molecular weight and substrate specificity, have been isolated from *Aspergillus niger* and have been shown to quantitatively hydrolyze ferulic acid and release
20 dehydrodiferulate dimers from plant cell walls. Furthermore, FAE has been observed to act synergistically with xylanase to release ferulic acid from plant cell walls at a higher rate. Recently, a ferulic acid esterase (FAE) gene has been cloned from *Aspergillus niger* (Michelson, *et.al.* European Patent Application No. 9510370.1). The inventors have found the recombinant enzyme releases ferulic
25 acid and diferulate dimers from grass cell walls in a concentration dependent manner and that this enzyme is stable at 30°C pH 5.0 in the presence of substrate and has a half life of 61 h at 30°C in the presence of vacuolar extracts (pH 4.6) of grass cells. This gene was, therefore, a candidate for targeted and indicible expression of FAE in grasses (*e.g.*, *Lolium multiflorum*).

30 The present invention provides for methods of changing the cell wall structure of transgenic plants and therefore, making them more digestible. The method comprises introducing a ferulic acid esterase coding sequence into the

cells of a plant. Operably linked to the coding sequence is a promoter that can be either constitutive or inducible and signal sequences that serve to target expression of the coding sequence in the desired organelle in the desired cell of the plant. The signal sequences can be either or both N terminal or C terminal sequences.

Optionally, a second and/or third coding sequence is introduced into the plant. It is preferred that a fungal xylanase coding sequence be coexpressed with the FAE coding sequence.

This invention also provides for transgenic plants which contain FAE1 coding sequences, leading to more digestible grasses.

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. Basic texts disclosing the general methods of use in this invention include Sambrook, *et al.*, MOLECULAR CLONING, A LABORATORY MANUAL, 2ND ED. (1989); Kriegler, GENE TRANSFER AND EXPRESSION: A LABORATORY MANUAL (1990); and Ausubel *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (1994)).

A. Isolation of Polynucleotides

The isolation of the polynucleotides, e.g., FAE1 and xylanase of the invention may be accomplished by a number of techniques. See, for example, copending Publication No. WO/1998/014594 which describes the isolation of a FAE from *Aspergillus niger*.

For instance, oligonucleotide probes based on the sequences cited here can be used to identify the desired gene in a cDNA or genomic DNA library from a desired species. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g., using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be

packaged into the appropriate vector. To prepare a library of cDNA from a specific cell culture, *e.g.*, *Aspergillus niger*, mRNA is isolated from the culture and a cDNA library containing the gene transcripts is prepared from the mRNA.

The cDNA or genomic library can then be screened using a probe based
5 upon the sequence of a known polynucleotide such as the polynucleotides cited here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. In addition to probes derived from known polynucleotides, degenerate probes may be used. Techniques for making and using degenerate probes are well known in the art
10 and can be found in Sambrook and Ausubel.

Alternatively, the polynucleotides of interest can be amplified from polynucleotide samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of the genes directly from mRNA, from cDNA, from genomic libraries
15 or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone polynucleotides that code for proteins to be expressed, to make polynucleotides to use as probes for detecting the presence of the desired mRNA in samples, for polynucleotide sequencing, or for other purposes.

20 Appropriate primers and probes for identifying ferulic acid esterase-specific genes, as well as xylanase sequences, from fungi and plant tissues are generated from comparisons of the sequences provided herein. For a general overview of PCR see PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), Academic Press, San
25 Diego (1990). Reaction components are typically: 10 mM Tris-HCl, pH 8.3, 50 mM potassium chloride, 1.5 mM magnesium chloride, 0.001% gelatin, 200 μ M dATP, 200 μ M dCTP, 200 μ M dGTP, 200 μ M dTTP, 0.4 μ M primers, and 100 units per mL Taq polymerase. Program: 96°C for 3 min., 30 cycles of 96°C for 45 sec., 50°C for 60 sec., 72°C for 60 sec, followed by 72°C for 5 min.

30 Polynucleotides may also be synthesized by well-known techniques as described in the technical literature. See, *e.g.*, Carruthers, *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and Adams, *et al.*, *J. Am. Chem.*

Soc. **105:661** (1983). Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

5 Suitable sources for the ferulic acid esterase used in this invention include but are not limited to, *Neurospora crassa*, *Aspergillus spp.* and specifically, *Aspergillus niger*. The xylanase used in this invention can be derived from any suitable source including, but not limited to, *Trichoderma reesei* and *Aspergillus spp.*

10

B. Preparation of Recombinant Vectors

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of plant species are well known and described in the technical and scientific literature. See, for example, Weising, *et al.*, *Ann. Rev. Genet.* **22:421-477** (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding the full length FAE1 protein, will preferably be combined with transcriptional and translational initiation and targeting regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant under the desired conditions.

15

20

Promoters can be identified by analyzing the 5' sequences of a desired gene. Sequences characteristic of promoter sequences can be used to identify the promoter. Sequences controlling eukaryotic gene expression have been extensively studied. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually 20 to 30 base pairs upstream of the transcription start site. In most instances the TATA box is required for accurate transcription initiation. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) N G. Messing, *et al.*, in GENETIC ENGINEERING IN PLANTS, pp. 221-227 (Kosage, Meredith and Hollaender, eds. (1983)).

25

30

A number of methods are known to those of skill in the art for identifying and characterizing promoter regions in plant genomic DNA (see, e.g., Jordano, *et al.*, *Plant Cell* **1**:855-866 (1989); Bustos, *et al.*, *Plant Cell* **1**:839-854 (1989); Green, *et al.*, *EMBO J.* **7**:4035-4044 (1988); Meier, *et al.*, *Plant Cell* **3**:309-316 (1991); and Zhang, *et al.*, *Plant Physiology* **110**:1069-1079 (1996)).

In construction of recombinant expression cassettes of the invention, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, the actin and ubiquitin promoters and other transcription initiation regions from various plant genes known to those of skill. A particularly preferred constitutive promoter is the rice actin promoter (see, McElroy, *Plant Cell*, **2**:163 (1990)).

Alternatively, the plant promoter may direct expression of the polynucleotide of the invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control (inducible promoters). Examples of tissue-specific promoters under developmental control include promoters that initiate transcription only in certain tissues, such as leaves, roots or seeds.

In one aspect of the instant invention, expression of FAE occurs after the the plant has been cut, removed from the ground or ingested. Thus an appropriate promoter would be a senescence promoter. For example, *BFN1* has recently been shown to be a nuclease expressed in senescing leaves, Perez-Amador, *et al.*, *Plant Physiol.* **122**:169 (2000). Similarly, SAG12, a cysteine protease is also found in senescing leaves (Noh & Amasino, *Plant Mol. Biol.* **41**:181 (1999). In a preferred embodiment, the promoter from the *gem* gene of *Festuca pratensis* is used to direct expression of FAE in senescing leaves.

In another aspect, the FAE would be expressed upon ingestion by a foraging animal. Exemplary promoters for this aspect would include Soybean Gmhsp 17.5 promoter and the leucine aminopeptidase (LAP) promoter. The

GMhsp promoter is from a heat shock protein gene and initiates expression if the temperature of the environment is increased. In the laboratory, an increase of 15°C for 2 hours is the preferred heat shock. However, in non-laboratory conditions suitable increases in temperature will occur in silos and in the rumen of animals that have ingested the plants of this invention. The LAP promoter initiates the expression of the FAE gene upon wounding of the plant. Such wounding would occur after cutting the plant or after mastication by a foraging animal. Tissue specific promoters that could be used in this invention include promoters of genes that are differentially expressed in the leaves of grasses. An example of a leaf specific promoter is the *rbcS* promoter of tomato (*Proc. Nat'l Acad. Sci. USA* **84**:7104 (1987)). This promoter normally regulates a gene determined to be important in photosynthesis.

For proper polypeptide expression, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural fungal gene, from a variety of other fungal or plant genes, or from T-DNA. These sequences are well known and readily available to those of skill in the art.

In addition to a promoter and poly A+ sequences, the preferred expression vectors of this invention also will contain signal sequences. These are polynucleotides found at the 5' and/or 3' ends of the coding region and serve to target expression of the gene to specific cellular organelles. These signal sequences can be both upstream or downstream of the coding region. Some preferred examples of upstream signal sequences include the barley aleurain sequence (Rogers, *Proc. Nat'l Acad. Sci. USA* **82**:6512 (1985) which targets vacuoles and the *Aspergillus* apoplast signal. This signal sequence targets expression to the apoplast.

In addition to targeting expression to specific organelles, it may be desirable to retain the expressed FAE in the Golgi or endoplasmic reticulum. The well known ER retention signal, KDEL, can be added to the 3' end of the coding polynucleotide.

The vector comprising the expression cassettes (e.g., promoters and/or coding regions) of the invention will typically comprise a marker gene which

confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to hygromycin, kanamycin, G418, bleomycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

5

C. Production of Transgenic Plants

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment or the constructs may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

See Dalton et al. (Co-transformed, diploid *Lolium perenne* (Perennial Ryegrass), *Lolium multiflorum* (Italian Ryegrass) and *Lolium temulentum* (Darnel) plants produced by microprojectile bombardment. *Plant Cell Reports* (1999) 18(9), 721-726) for exemplary methods for culturing and transformation of grasses.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski, *et al.*, *Embo J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm, *et al.*, *Proc. Natl. Acad. Sci. USA* 82:5824 (1985).

Agrobacterium tumefaciens-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for example Horsch, *et al.*, *Science* 233:496-498 (1984), and Fraley, *et al.*, *Proc. Nat'l. Acad. Sci. USA* 80:4803 (1983). US Patent 5,591,616 discloses *Agrobacterium* mediated transformation techniques in monocotyledons.

Ballistic transformation techniques are described in Klein, *et al.*, *Nature*

327:70-73 (1987). In a preferred embodiment, a particle in-flow gun (PIG) is used to transform the plant cells of this invention.

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant that possesses the transformed genotype and thus the desired phenotype such as improved digestibility. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans, *et al.*, PROTOPLASTS ISOLATION AND CULTURE, HANDBOOK OF PLANT CELL CULTURE, pp. 124-176, Macmillan Publishing Company, New York, 1983; and Binding, *REGENERATION OF PLANTS, PLANT PROTOPLASTS*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee, *et al.*, *Ann. Rev. of Plant Phys.* **38**:467-486 (1987).

To determine the presence of or increase of FAE1 activity, an enzymatic assay can be used or an assay to measure increases and decreases in rates of fermentation. These assays are readily available in the literature and those of skill in the art can readily find them.

One of skill will recognize that other assays can be used to detect the presence or absence of FAE1. These assays include but are not limited to; immunoassays and electrophoretic detection assays (either with staining or western blotting).

The polynucleotides of the invention can be used to confer desired traits on essentially any plant. However, the main utility of this invention is in the improved digestibility of forage plants. Thus, it is envisioned the transgenic plants of this invention will include but not be limited to the following genera *Lolium*, *Festuca*, *Triticum*, *Avena*, and *Medicago*. The *FAE1* genes of the invention are particularly useful in the production of transgenic plants in the genus *Lolium*.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be

introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

As mentioned above, the transgenic plants of this invention can be used as a foraging crop for animals, such as cattle, sheep, goats and horses. In addition, the methods of this invention can be used to transform any plant into which FAE expression is desired. For example, it is advantageous to break down cell walls during biomass conversion or during processing of plants for foodstuffs. This invention would help to achieve this goal more effectively and inexpensively.

10

The inventive methods herein may also be used to provide additional enzymes to enhance the availability of fermentable sugars in plants. Plant carbohydrates may be subject to further modification, either exogenously or endogenously, by the action of other enzymes. Such enzymes include, but are not limited to, endoglucanases, xylosidases and/or cellbiohydrolases. These enzymes may be provided either in an expression cassette provided for herein (i.e., endogenous) or applied to the plant cell walls (i.e., exogenous) to enhance the availability of mono- and/or di-saccharides.

Plants other than grasses may find a use in the present invention. For example, corn (or maize) is specifically contemplated to be useful. The grass Festuca is similar to maize in cell wall structure and therefore provides a good model of the ability to enhance fermentable carbohydrates in corn. Other useful plants contemplated for use in the present invention are Festuca, Lolium, Zea, Avena, Sorghum, Millet (tropical cereals), Miscanthus (a grass with potential for use as a biomass energy crop), Cenchrus, Dichanthium, Brachiaria and Paspalum (apomictic tropical range grasses) and Poa (Kentucky bluegrass).

Cell walls of forage grasses makes up 30-80% of forage dry matter representing a major source of energy for ruminants, but less than 50% of this fraction is digested by the animal. Conversion of low-value biomass to sugars and ethanol is also less than optimal due to the carbohydrate unavailability of the

feedstocks, including but not limited to bagasse, rice straw, corn stover and corn fiber.

Ferulic and other hydroxycinnamic acids are ester linked to arabinosyl residues in arabinoxylans, and play a key role in crosslinking xylans to lignin, resulting in less degradable cell walls. Ferulic acid esterase (FAE) can release both monomeric and dimeric ferulic acid (FA) from arabinoxylans making the cell wall more susceptible to further enzymatic attack. Transgenic plants have been produced expressing an FAE gene following microprojectile bombardment of cell cultures. Measurements of the level of FAE activity from different vectors targeting FAE to the vacuole, ER and apoplast under constitutive or inducible (heat shock) promoters shows that at least for constitutive expression of vacuolar targeted FAE, the activity was highest in young leaves and increased along the leaf lamina. We also show that FAE expression results in release of monomeric and dimeric FA from cell walls on cell death and this was enhanced several fold by the addition of xylanase. An effect of FAE expression on the monomeric and dimeric cell wall ester linked ferulate content in comparison to control (non-transformed) plants is seen. Generally, the lower the levels of monomers and, in particular, dimers of hydroxycinnamic acids in leaves, the higher the digestibility and/or availability of complex carbohydrates for conversion.

Senescence is the terminal phase in leaf development and occurs without growth or morphogenesis. Therefore the metabolism/physiology of this stage of the leaf's lifespan can be targeted directly for alteration with minimal detrimental impact on early development. Senescence follows leaf maturity and is associated with the expression of specific genes. These genes and their controlling elements can be exploited to manipulate development, adaptation, productivity and quality traits in crop plants. There seems to be good conservation of senescence physiology across the range of higher plant species and thus these promoters are useful in the present invention.

The following preparations and examples are given to enable those skilled in the art to more clearly understand and practice the present invention. They should not be considered as limiting the scope and/or spirit of the invention, but

merely as being illustrative and representative thereof.

In the experimental disclosure which follows, the following abbreviations apply: eq (equivalents); M (Molar); μ M (micromolar); N (Normal); mol (moles); mmol (millimoles); μ mol (micromoles); nmol (nanomoles); g (grams); mg (milligrams); kg (kilograms); μ g (micrograms); L (liters); ml (milliliters); μ l (microliters); cm (centimeters); mm (millimeters); μ m (micrometers); nm (nanometers); ° C. (degrees Centigrade); h (hours); min (minutes); sec (seconds); msec (milliseconds); Ci (Curies) mCi (milliCuries); μ Ci (microCuries); TLC (thin layer chromatography); Et (ethyl), Me (methyl).

10

Example 1

Preparation of Enzyme Encoding DNA Sequences

A genomic clone for FAE1 (see Figures 1-3) was used as the starting point for the preparation of an intronless FAE1 encoding DNA sequence. The sequence for the genomic clone is given in Figures 2 and 3. Separate fragments for both FAE exons were recovered by PCR from a 5.5kb EcoRI fragment of the genomic clone in pLITMUS28, and 'cDNA' created by overlapping PCR. See Figure 4.

Two 5' primers were used. FAE-S5 which amplifies the entire reading frame (including the Aspergillus signal), and FAE-N5 which amplifies only the mature protein (i.e. has *no* signal). A number of codons are optimised (underlined in primer sequences below). The overlap product may be derived from either FAE-I5 (wild type) or FAE-I3 (conserved Ser changed to Ala) primers, allowing production of enzymatically inactive protein to check toxicity. As shown in Figure 5, overlapping of PCR products made with FAE-I5 and FAE-I3 creates two possible uninterrupted reading frames. If the complement to FAE-I5 serves as the template when recombined then the encoded protein retains the serine moiety and the esterase is functional (highlighted serine is at active site). If the FAE-I3 primer serves as the template the serine is replaced with an alanine and the esterase is inactivated (highlighted alanine in bottom amino acid sequence given in Figure 5).

25
30

Where possible, codon usage has been optimised in constructed reading frames (codon choice based on published barley preferences).

FAE-I5 (SEQ ID NO: 49)

5 GGCGCCGAGGGAGTGGCCGGTCACGGTCAGCGCGTAGTCC 40-mer

FAE-I3 (SEQ ID NO: 50)

CCGGCCACGCCCTCGGCGCCTCCCTGGCGGCACTC 35-mer

FAE-N5 (SEQ ID NO: 51)

CTAAAGCTTACCATGGCGGGCCGCCTCCACGCAGGGCATCTCCGA 44-mer

10 **FAE-S5 (SEQ ID NO: 83)**

CTAAAGCTTAACATGAAGCAGTTCTCCGCCAA 32-mer

FAE-3 (SEQ ID NO: 52)

TCTAAGCTTGCGGGCCGCGACCGGCCAGGTGCATGCCCGCTCGTCATCCC
50-MER

15

Example 2

Preparation of Vectors

Vectors had the general structure shown in Figure 6.

A. Plant transformation vector series

20

Initial expression vectors were based on **pCOR105** [rice actin promoter - McElroy et al. MGG 231:150-160 (1991)] (Figure 7). pCOR105 Not and SstII sites were first destroyed [cut with NotI and SstI, followed by heat inactivation and T4 DNA polymerase treatment in the presence of dNTPs] using standard methods as described
25 in Maniatis et al. or following the manufacturer's instructions for enzymes to simplify subsequent Not cassette manipulation and allow use of unique Sst site (see below).

The *nos* terminator from **pMA406** (Ainley & Key (1990) PMB 14:949-60) was amplified by PCR using primers TER5 and TER3 to generate a fragment with the following sequence (SEQ ID NO: 53):

(Pst1) (Not 1)
 (AGACTGCAGACCATGGCGGCCGCGKAACCACTGAAGGATGAGCTGTAAAG
 AAGCAGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTT
 5 GCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGT
 AATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTA
 GAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGC
 AAAGTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGATA
 AGCTT CTA GATCT (where K=G or T)
 10 (HindIII) (XbaI)

A redundancy in the TER5 primer (GCGKAA) creates fragments having either a stop codon (TAA) or glutamate codon (GAA) in one reading frame. The glutamate codon is in frame with a downstream KDEL motif.

15 The fragment and modified pCOR105 vector were cut with PstI and XbaI, according to manufacturers instructions, relevant fragments gel-purified, ligated with T4 DNA ligase and transformed into *E. coli*. Resulting clones were then sequenced to establish which TER5 alternatives were present.

Initial FAE expression vectors were then constructed from these vectors
 20 by inserting FAE-S5/FAE-3 PCR products (T4 DNA polymerase 'polished' in the presence of dNTPs, purified and digested with NotI, cloned into EcoRV and NotI digested vector) or FAE-N5/FAE-3 PCR products (purified and NotI digested, cloned into NotI digested and calf intestinal alkaline phosphatase treated vector).

The initial pCOR105-*nos* terminator clones were also modified by the
 25 addition of ALE-5/ALE-3 PCR products (encoding wild-type and modified barley aleurain signal peptides, see below for details). The products were 'polished' with T4 DNA polymerase in the presence of dNTPs, purified and cut with NotI, then cloned into EcoRV and NotI digested vectors. Addition of the ALE sequences creates a series of vectors which can express a reading frame inserted at the
 30 NotI or NcoI sites as a fusion to the barley aleurain signal, with or without vacuolar targeting motif, and with or without an ER retention motif. HindIII sites flanking the translation initiation codon and transcriptional terminator allow easy

movement of transcription units between expression vectors providing different promoter sequences. (See Figure 8 depicting the generic ALE-TER vector.)

Vector sequences were confirmed by sequencing. Two artifacts were found. Firstly, the redundant codon in TER5 was found to be AAA in one clone, which was subsequently used as the source of all KDEL fusions (ie peptide sequence is KPLKDEL (SEQ ID NO: 85), rather than EPLKDEL (SEQ ID NO: 86) as designed). See Figure 9. Secondly, an additional base is found at the site of the redundant codon in one clone, creating a frameshifted terminal peptide (ETTEG (SEQ ID NO: 87), Figure 10) which was used as a control in some constructs.

Exploitation of the modular arrangement of signal peptides in the above vector series allowed various combinations of FAE and targeting motifs to be created using standard molecular biology procedures (i.e., restriction digest, purification of relevant fragments and ligation as appropriate). For example, the NotI fragment containing the FAE reading frame was inserted into the NotI site of the frameshifted clone described above to create vector **pTP3.1**. The native *Aspergillus* COOH-terminus was inserted into a FAE-S5/FAE-3 clone as a SphI (T4 DNA polymerase polished) – NcoI fragment from the FAE genomic clone (replacing the NotI (T4 DNA polymerase polished) – NcoI fragment), creating vector **pTP4a2**, which then encodes the entire, unmodified, *Aspergillus* FAE. Replacement of the Sall/XbaI fragment of pTP3.1 with that of pTP4a2 then created **pTP11.1**, which encodes FAE with a native *Aspergillus* COOH-terminus but a barley aleurain N-terminal signal.

Briefly, other vectors made in this series were; **pTP8.5**, the FAE NotI fragment inserted into the NotI site of an ALE-frameshifted COOH-terminus construct, aleurain N-terminus; **pTP5.1**, replacement of the native *Aspergillus* COOH terminus with a KDEL peptide (NotI/XbaI fragment exchange), *Aspergillus* N-terminal signal retained; **pTU4.4**, BamHI fragment of pTP11.1 replaces BamHI fragment of pTP5.1, creates FAE reading frame fused to heterologous N- and C-termini (aleurain signal and KDEL).

Vectors in which the aleurain vacuolar targeting motif NPIR was replaced by NPGR (found to be inactive in some plant assays) were created by replacing an EcoRV/NotI fragment with ALE PCR product which had been cut with AccI

(T4 DNA polymerase polished) and NotI (vectors **pTT5.5** and **pTT5.14**, Aspergillus COOH-terminus). The BamHI fragment of pTT5.5 was used to replace that of pTP5.1 to produce **pTU5**, creating an FAE reading frame fused to heterologous N- and C-termini (NPGR modification of aleurain signal and KDEL). The aleurain signal was also
 5 modified by PCR mutagenesis to remove the vacuolar targeting NPIR motif in its entirety (directed by primer ALECUT, which contains a NotI site to allow exchange of BglII/NotI fragments). NPIR deletion was created in this way in pTP11.1 (creating **pUA4.4**), and in pTP5.1 by exchange of BamHI fragments with pUA4.4 (creating **pUG4**).

10 Finally, PCR mutagenesis, using overlap of fragments generated by primers GLY3 and GLYB, was also used to alter a potential glycosylation site (asparagine codon changed to aspartate, as carried out for example in Chen, H.M., C. Ford & P. J. Reilly (1994) Biochem J 301 275-281 Substitution of asparagine residues in Aspergillus awamori glucoamylase by site-directed mutagenesis to eliminate N-glycosylation and
 15 inactivation by deamidation; see sequence data for exact change, vector **pTP10.1**).

PCR primers

TER-5 (SEQ ID NO: 54)

AGACTGCAGACCATGGCGGCCGCG**KA**ACCACTGAAGGATGAGCTGTAAAGAAGC

20 AGATCGTTCAAACATTTG 72-MER (The KDEL stop codon is underlined.)

TER-NOT (SEQ ID NO: 55)

AAGACTGCAGACCATGGCGG 20-MER

TER-3 (SEQ ID NO: 56)

AGATCTAGAAGCTTATCGATCTAGTAACATAGATGACACC

25 **ALECUT (SEQ ID NO: 57)**

CTAGGCGGCCGCGCGGGAGGAGGCGACGGCGAC

GLYB (SEQ ID NO: 58)

GAGGGTGTATTCGGTATCGAGTTGCAGGTTCGTATC

GLY3 (SEQ ID NO: 59)

30 CTCGATACCCATTACACCCTCACGCCTTTTCA

B. Construction of different promoter vectors

Various promoters were used to optimize expression and to establish
 35 constitutive, heat-shock inducibility and senescence enhancement.

i. Rice actin promoter and 1st intron

Initial vectors (Figures 11 and 12) were constructed from pCOR105 which was subsequently found to contain a 5bp deletion relative to the published sequence which destroys the *AccI* site (GTAGGTAGAC (SEQ ID NO: 60), deleted bases underlined) and may affect splicing at the adjacent 3' site. The original rice actin sequence in this region
 5 (GTAGGTA) (SEQ ID NO: 84) was therefore restored using oligonucleotide NCO-ACT (CTCACCATGGTAAGCTTCTACC TACAAAAAAGCTCCGCA) (SEQ ID NO: 81) by replacing the *Bgl*III/*Hind*III fragment with a PCR product, to produce vector **pPQ10.1**.

A rice repetitive element is present in the upstream region of the actin promoter used in pCOR105; as this may have unpredictable effects on vector expression it was
 10 removed from pPQ10.1 by deletion of the *Kpn*I/*Eco*RI fragment (end-filled with T4 polymerase and ligated following digest, restoring *Eco*RI but not *Kpn*I), to produce vector pGT6. The *Hind*III fragment containing the FAE reading frame and nos terminator of pTP3.1 (see Example 2A) was then inserted into pGT6 to produce construct **pJO6.3**.

15 *ii. Soybean heat-shock promoter*

A soybean heat shock promoter from a 23kD HSP was obtained from pMA406 (Ainley & Key (1990) PMB 14:949-60). This promoter when fused to β -glucuronidase (Jefferson et al 1987 EMBO J 6:3901-3907) had previously been shown to be inducible by a 10°C heat-shock and show stable expression for 24-48 hours (data not shown). β -
 20 glucuronidase fusions are a sensitive and versatile fusion marker in higher plants. The construction of the co-integration HS vectors is given below.

iii. Senescence enhanced expression (See1) promoter from Lolium multiflorum

The promoter and signal sequence (including NPIR motif) of the LSee1 gene
 25 was amplified from *Lolium multiflorum* cv Tribune with oligonucleotides SEE-NCO and SEE-VAC, and cloned as an *Asp*718/*Not*I replacement of the promoter region of vector pTP11.1. Following sequencing to screen for PCR artifacts, one of three identical clones was chosen (pUB8.11).

The See1 promoter from maize has been cloned previously and has

EMBL accession number is AX050343. See WO0070061.

The *Lolium* version of See1 was also cloned previously (Qiang Li (2000) Studies on leaf senescence and its genetic manipulation in *Lolium mutiflorum* PhD Thesis University of Wales, Aberystwyth) and has been shown to be senescence inducible
5 when used to drive both β -glucuronidase and the *Agrobacterium* ipt gene.

An apoplast-targeted derivative was constructed by amplifying the Potato Protease Inhibitor (PPI) motif with primers PPI-AP6 and SEE-ATG, and cloning the product as an NgoMIV/NotI fragment into pUB8.11 (NgoMIV partial digest), to produce vector **pJQ5.2**. This vector has both the senescence induced promoter and the
10 apoplast target sequence with the gene to be expressed inserted downstream of the apoplast sequence.

PCR Primers

SEE-VAC (SEQ ID NO: 62)

15 AACCATGGCGGCCGCGCGCTCGGTGACGGGCGCGGAT

SEE-NCO (SEQ ID NO: 63)

TTCGGTACCATGGCCAGGTATAATTATGG

SEE-ATG (SEQ ID NO: 64)

CTGCGCCGGCGAGATGGMCGTGCACAAGGAG
20

C. Construction of targeting sequences

In order to examine whether or not the localization of the enzyme would have an effect on the phenolic acid content of the cell wall various signal sequences were utilized. The targeting sequences were added either to the N-terminus or to the C-
25 terminus of the gene of interest.

i. N-terminal signal sequences

Six N-terminal signal sequences were utilized:

- (a) The native *Aspergillus* end of FAE, plus excretion signal [apoplast localisation]

30 This is from the original clone and has the peptide sequence:

MKQFSAKHVLAVVVTAGHALAASTQGI (SEQ ID NO: 88).

- (b) The mature *Aspergillus* end, with no excretion signal [cytoplasmic localisation]

Peptide sequence is MAASTQGI (SEQ ID NO: 89) (underlined motif is common to all constructs). Truncation of the signal sequence in (a) above was carried out by PCR with mutagenic primer FAE-N5.

- (c) The barley aleurain signal, including intact NPIR motif [vacuole
5 localisation]

The barley aleurain vacuolar signal sequence (See Figure 13; Swissprot database accession number P05167) was derived entirely from overlapping primers (ALE-5, ALE-3, ALE-CUT ALE-CAP-5 and ALE CAP-3). Following primer annealing at 37°C and extension with T4 DNA polymerase in the presence of dNTPs according to
10 manufacturers instructions, PCR with flanking primers ALE-5 and ALE-3 was carried out. The product was 'polished' with T4 DNA polymerase, purified, digested with NotI and cloned into EcoRV/NotI digested pCOR105-nos terminator vector (see above). ALE-3 contains redundancies so that clones encoding NPIR or NPGR motifs may be
15 recovered. Two versions of the signal, with and without the vacuole targeting motif, were produced, to give putative vacuolar NPIR and apoplast (NPGR) signal sequences.

PCR Primers

ALE-5 (SEQ ID NO: 65)

GGAATTCGTAGACAAGCTTACMATGGCCACGCCGCGTCCT 41-MER

ALE-3 (SEQ ID NO: 66)

TATCCATGGCGGCCGCGCGGTCTGGTGACGGGCGGMYCGGGTTGGAGTCGGCG
AA 55-MER

ALE-CUT (SEQ ID NO: 67)

CTAGGCGGCCGCGCGGGAGGAGGCGACGGCGAC 33-mer

ALECAP-5 (SEQ ID NO: 68)

GCGACGGCGACGGCGGCCGTGGCCAGCACGGCGAGCGCCAGGAGGAGGACGC
GG 54-MER

ALECAP-3 (SEQ ID NO: 69)

TCGCCGTGCGCTCCTCCTCCTCCTTCGCCGACT 33-MER

(d) The barley aleurain signal, mutated to a NPGR motif [cytoplasmic localisation]

(e) The rat sialyl transferase golgi targeting motif [golgi localisation]

A Golgi targeting vector, **pJQ3.2**, was made by inserting a reading frame
 5 encoding the relevant rat sialyl transferase (RST) motif (See Figure 14. RST motif
 shown to function in plants by Boevink P, Oparka K, Cruz SS, Martin B, Betteridge A,
 Hawes C, (1998) PLANT JOURNAL 15 441-447 Stacks on tracks: the plant Golgi
 apparatus traffics on an actin/ER network) into vector pPQ10.1, and replacing the
 EcoRI/NotI promoter/signal fragment of **pJO6.3** with the fragment from this vector.
 10 Briefly, the RST motif was constructed by annealing oligonucleotides RST-F1A, RST-
 F1B, RST-F2A and RST-F2B, and amplifying the product with RST-5AD and RST-3A.
 This product was cloned and sequenced. Clones were found to have a deletion which
 was corrected by PCR with RST-RPT, followed by overlap-PCR and cloning of
 products.

15

PCR primers

RST-5AD (SEQ ID NO: 70)

ACTAAGCTTAAGGAGATATAACAATGATCCACACCAACCTCAA

RST-F1A (SEQ ID NO: 71)

20 TTCCATGATCCACACCAACCTCAAAAAGAAGTTCTCCCTCTTCAT

RST-F1B (SEQ ID NO: 72)

AGAGTGATCACGGCGAAGAGGAGGAAGACGAGGATGAAGAGGGAGAACTTCTTTT

RST-F2A (SEQ ID NO: 73)

TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGCCCTCACCTCCAAGC

25 CAAGGA

RST-F2B (SEQ ID NO: 74)

CATTGGAACCTCCTTGGCTTGGAGGGTG

RST-3A (SEQ ID NO: 75)

AACCATGGCGGCCGCCATTTGGAACCTCCTTGGCT

30 **RST-RPT (SEQ ID NO: 76)**

TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGGCCCTCACCTCCAAGC
CAAGGA

(f) otif [cytoplasmic localisation]

(g) The potato protease inhibitor II (PPI) apoplast motif [apoplast localisation]

5 An apoplast targeting reading frame was designed to encode the relevant potato
protease inhibitor II (PPI) motif (See Figure 15) and cloned into **pJO6.3**, to produce
vector **pJQ4.9**. Briefly, the PPI motif was constructed by annealing oligonucleotides
PPI-AP1, PPI-AP2, PPI-AP3, PPI-AP4, PPI-AP5 and PPI-AP6, and cloning this product
as a HindIII/NotI fragment into vector pPQ10.1; the EcoRI/NotI promoter/signal fragment
10 of pJO6.3 was then replaced with the equivalent fragment from the modified pPQ10.1
vector.

PCR primers

PPI-AP1 (SEQ ID NO: 77)

15 GGAATTCGTAGACAAGCTTACMATGGMCGTGCACAAGGAGGT

PPI-AP2 (SEQ ID NO: 78)

GATCAGGAGGTAGGCWACGAAGTTWACCTCCTTGTGC

PPI-AP3 (SEQ ID NO: 79)

CCTACCTCCTGATCGTSCCTCGGCCTCCTCTTGCTCGT

20 PPI-AP4 (SEQ ID NO: 80)

CCTTGGCGTCCACGTGCTCCATGGCGGAWACGAGCAAGAGGAG

PPI-AP5 (SEQ ID NO: 81)

GTGGACGCCAAGGCCTGCACCCKCGAGTGCGGCAACCTC

PPI-AP6 (SEQ ID NO: 82)

25 GGAATTCGCGGCCCGCCGGGCAGATGCCGAAGCCGAGGTTGCCGCACT

ii. C-terminal end signal sequences

Four C-terminal signal sequences were utilized:

(a) Native Aspergillus end, [CTW] (vacuole and apoplast vectors)

30 This was derived directly from the genomic clone (see Example 1) as a NcoI-SphI
fragment (Sph end filled with T4 polymerase) which replaces the

Nco1-Not1 region of a standard actin -FAE vector (Not1 end filled with T4 DNA polymerase).

- (b) Expression vector linker alone [CTW-PVAAA] (plant optimised C-terminus for vacuole, golgi and apoplast vectors)

5 CTW is the peptide sequence of the Aspergillus FAE COOH end and is here provided by oligo FAE3. In this primer the reading frame is extended to provide the additional amino acids PVAAA which are partially encoded by the Not1 site used for cloning downstream signals see c) and d) below. Some COOH amino acids /motifs may affect compartment targeting, the PVAAA sequences are expected to be neutral in this
10 respect while the native Aspergillus end may not be.

- (c) Linker plus KPLKDEL (SEQ ID NO: 90) [first K is primer artifact, intended to be E] {ER retention vectors}

These sequences are provided by primer TER5 introduced during PCR to generate the nos terminator fragment, and identified by sequencing within a specific
15 clone . KDEL targeting has been demonstrated in plants by Denecke et al. ((1992) EMBO J 11: 2345-2355 Plant and mammalian sorting signals for protein retention in the endoplasmic reticulum contain a conserved epitope).

- (d) Linker plus ETTEG (SEQ ID NO: 87) [frameshift of (c)] (loss of ER retention - vacuole vectors)

20 These sequences are provided by primer TER5 introduced during PCR to generate the nos terminator fragment, and identified by sequencing within a specific clone (see Example 2A).

The KDEL (SEQ ID NO: 97) signal is for ER retention, while others provide controls. A frameshift in the TER5 region [additional A] was used in subsequent
25 constructs to destroy the ER KDEL retention signal.

The linker used in the above C-terminal targeting sequences was PVAAA (SEQ ID NO: 91).

30 *D. Co-integration and co-transformation vectors.*

Co-transformation vectors

A Hygromycin resistance gene driven by a CaMV345S promoter (**pRob5**) (35S-HYG-CMV in pUC18 (modified HYG, derived from pGL2) Bilang et al (1991)

Gene 100:247-50) was used for co-transformation experiments with **pTT3** and **pTP3.1**, **pJQ4.9**, **pJQ3.2**, **pJO6.3**, **pJQ5.2**, **pUB8.1** 1 vectors.

5 *Co-integration vectors*

1. **Actin promoter constructs** - **pTR2.22**, **pTR6.1** , **pTR8.1**, **pTR9.4**, **pTR7.1**, **pTT5.5** and **5.1**.

10 The CAMV35S-hyg region from pAJEB64TCA [a plant expression vector constructed by Andy Bettany at IGER containing CaMV-HYG from pTRA151 (Zheng et al 1991 Plant Physiol 97:832-835) (CaMV35S-HYG-tml terminator as clonable cassette in pUC4) cloned into KpnI site of pCOR105] was added as a HindIII fragment at the KpnI site (T4 polymerase blunt) of pTP4a2, in divergent orientation to FAE to create
15 **pTR2.22**. The FAE/Nos HindIII fragment of this vector was replaced as follows in co-expression vectors . From pTP5.1 for **pTR6.1** , from pTP10.1 to **pTR8.1**, from pTP11.1 to **pTR9.4**. Signal sequences of FAE in pTR2.22 were replaced as HindIII/BglII fragments in **pTR7.1** (fragment from pT09.1). PCR products (ALE5/ALE-G) was digested with Acc1 and T4 polymerase, polished, followed by Not1 digest and cloning
20 into EcoRV/ Not1 digested pTR2.22 to give clones **pTT5.5** and **5.1**.

PCR primer

ALE-G

TATCCATGGCGGCCGCGCGGTCTGGTGACGGGCGGCCCGGGTTGGAGTCGGCG
AA (SEQ ID NO: 92)

25

2. **Actin promoter constructs** -**pUF1**, **pUA1K3**, **pUH4**, **pUH5**, **pUH6**, **pUH7**, **pUH8**, **pUH9**.

30 The HygR gene from pAJEB64TCA, driven by the CaMV promoter, was first cloned as an end-filled HindIII fragment at the end-filled XbaI site of pTP3.1, to give **pHOX3**. For ease of cloning the downstream HindIII site was destroyed to create **pUA1K3** and replacement of the FAE/Nos terminator HindIII fragment in this vector was carried out as follows. From pTP5.1 for **pUF1**, from **pTP11.1** for **pUH4**, from pTP8.5 for
35 **UH5**, from pTT5 for **pUH6**, from pUA4.4 for **pUH7**, from pTU5 for **pUH8** and from pUG4 for **pUH9**.

3. Heat-shock promoter constructs - pUH10, pUH12, pUC5.11.

5 A co-transformation vector in which FAE is expressed from the soybean heat shock promoter was made by first modifying pMA406 to remove the nos terminator (BglII linearised and gel purified, KpnI digested, T4 DNA polymerase polished in the presence of dNTPs and recircularised), and then inserting the FAE HindIII fragment from pTP11.1, creating pTT3.1, which encodes the full aleurain signal and the native
10 *Aspergillus* COOH-terminus.

Following assays of various constructs, co-integration vectors were constructed with FAE and HygR genes arranged in tandem.

The HygR gene from pAJEB-64-TCA, driven by the CaMV promoter, was first cloned as an end-filled HindIII fragment at the end-filled XbaI site of pTP3.1, to give pHOX3 and subsequently excised as a HindIII/SacI fragment (partial SacI
15 digest, relevant sites found in flanking pTP3.1 sequences) which was cloned into the HindIII/SacI sites of pMA406, in tandem orientation (vector **pUH1a20**). FAE sequences were then cloned into the HindIII site of **pUH1a20** downstream of the heat -shock promoter (HindIII fragment from pTU5 for **pUH10**, HindIII fragment from pTT5 for **pUH12**). A pTP3.1 derivative was made by cloning the
20 CaMV/HygR HindIII cassette from pAJEB-64-TCA in tandem orientation downstream of the FAE gene in pTP3.1, inactivating the middle HindIII site by partial digestion and end-filling, and excising the combined FAE/HygR cassette as a single HindIII fragment, which was inserted at the HindIII site in pMA406 to
25 produce **pUC5.11**.

Example 3

Transformation of Plant Cells

Eight to ten weeks old embryogenic *F. arundinacea* and *L. multiflorum*
30 suspension cultures were bombarded either with a single co-integration plasmid DNA vector containing FAE and hyg resistance genes, or with a co-transformation vector containing FAE and with plasmid pROB5 conferring hygromycin resistance (CAMV35S-hpt- nos) using a Particle Inflow Gun (PIG) (Finer et al. (1992) Development of the particle inflow gun for DNA delivery to

plant cells Plant Cell Reports 11:323-328) and 1.5-3.0 μm gold particles as in Dalton *et al* (Dalton et al. (1999) Co-transformed diploid *Lolium perenne* (*Perennial ryegrass*), *Lolium multiflorum* (Italian ryegrass) and *Lolium temulentum* (Darnel) plants produced by microprojectile bombardment. Plant Cell Reports. 5 18: 721-726) and Kuai et al (Regeneration of fertile transgenic tall fescue (*Festuca arundinacea*) plants with a stable highly expressed foreign gene. Plant Cell Tissue and Organ Culture (1999) 58:149-154). Transformants were selected with hygromycin (25 to 50mg /l) over a 10-12 week selection period at 25°C under continuous white fluorescent light ($60 \mu\text{E m}^2 \text{s}^{-1}$) and plants regenerated via 10 somatic embryogenesis as in Dalton *et al* 1999, *supra*. Regenerated plants were screened for FAE activity on transfer to soil and expressing plants grown to maturity in a containment growth room at 18°C under 16h fluorescent lights ($350 \mu\text{E m}^2 \text{s}^{-1}$). Mature plants (6-8 weeks old) were re-assayed for FAE activity and fresh tissue harvested for Southern, Northern and Western analysis, and for self 15 digestion analysis. The remaining tissue was freeze dried and powdered for cell wall structure analysis, in vitro-dry matter digestibility (IVDMD) determinations and for in-vitro gas production determinations of rates of tissue digestion.

Example 4

Targeting of Expression Product

20 To verify that the targeting sequences are effective in delivering the gene the targeting sequences were operably linked to a green fluorescent protein GFP. The vector constructs are shown in Figure 16. Cells were transformed by particle bombardment as in Example 3. Localization of the GFP could be 25 visualized under a microscope 1 day after bombardment (i.e., shooting). See Figure 16.

Example 5

FAE1 activity

30 Plants regenerated from transformed cells showed FAE activity in all plant tissues tested. Cells were transformed as above under the direction of the ER and APO targeting sequences. FAE activity in transformed *Festuca arundinacea*

leaves of different ages was elevated compared to control (untransformed) plants. See Figures 17 and 18.

Similar results were seen with *Lolium mutiflorum* leaves at different ages transformed as above under the direction of vacuolar, ER and APO targeting
5 sequence. See Figures 19 and 20.

FAE expression under a heat shock promoter can also be induced. (Data not shown.)

Thus, we have demonstrated FAE expression in *Festuca* and *Lolium* leaves under constitutive and HS promoters with effective FAE targeting to the
10 vac, ER and apo.

FAE assay

FAE activity was determined in soluble extracts of fresh (or frozen at -70°C) leaves or cell cultures (0.5g) with 0.1M NaAc, pH 5.0 extraction buffer. Extracts were incubated with 24mM EF (ethyl 4-hydroxy-3-methoxycinnamate) or
15 1% FAXX as substrate, at 28°C for 24hrs and FAE activity calculated as the amount of ferulic acid released. FAE activity was also determined by measuring the release of monomeric and dimeric ferulic acid from self-digested leaf or cell culture samples. Fresh, or frozen, leaves or cell cultures (0.5g) were ground in 0.1M NaAc, pH5.0 extraction buffer in the presence and absence of xylanase
20 (1000U **GC140**/sample) without added substrate and incubated at 28°C for 72hrs. Following incubation, and centrifugation, soluble extracts were loaded onto an activated reverse phase C18 μ Nova sep-pak column (Waters), eluted with 100% MeOH and the MeOH sample analysed by HPLC.

25

Example 6

Chemical Analysis of Cell Wall Extracts

Ester bound compounds were extracted from freeze dried powdered leaves or cell cultures (50 -100mg) with NaOH (5ml of 1M) followed by incubation at 25°C for 23hrs under N₂. After centrifugation and acidification of the soluble extract with
30 concentrated HCl, the extracted phenolics were loaded onto an activated reverse phase C18 μ Nova sep-pak column (Waters) and eluted with 100% MeOH. and the MeOH sample analysed by HPLC.

HPLC was carried out with methanol: 5% acetic acid either with a 35-65% MeOH gradient in 15min (FAE assay) or with a 30-70% MeOH gradient in 25 min (monomer and dimer cell wall components) at 2ml/min on a μ Nova Pak C18 8x10 RCM (Waters). Extracts were detected and quantified with a diode array detector
5 (240-400nm Waters 996PDA) monitored at 280nm for aldehydes and 340nm for hydroxycinnamic acids. .

Levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under VAC, and ER and APO targeting sequences are reduced compared to control (untransformed) plants.
10 The results can be seen in Figure 21 and 22, respectively. Thus, we show where this does not result in reduced cell wall phenolics in growing plants with vac targeting but does result in lower phenolics with ER and apo targeting. In addition,

Levels of esterified monomeric and dimeric hydroxycinnamic acids in
15 *Festuca arundinacea* plants expressing FAE are not significantly reduced when FAE is VAC, targeting (Fig 21) which is as predicted for correct vacuolar targeting, but are significantly reduced, as predicted, in some plants when FAE was ER and APO targeted, compared to control (untransformed) plants. The results can be seen in Figure 22.

20

Example 7

In vitro dry matter digestibility. (IVDMD)

The *in vitro* dry matter digestibility (IVDMD) was estimated on 1.0 g dry weight of leaf or cell culture tissue using the pepsin/cellulase method of Jones
25 and Hayward (The effect of pepsin treatment of herbage on the prediction of dry matter digestibility from solubility in fungal cellulase solutions. Journal of the Science of Food and Agriculture (1975) 26:711-718).

We show that the presence of FAE in the plants results in higher digestibility of the leaves. This may be due to internal FAE activity acting on
30 normal cell walls with vacuole located FAE and to both FAE activity and the lower cell wall crosslinking with ER and apo targeted FAE (as also found with cell cultures).

End point digestibility as determined by IVDMD were higher in leaf tissue of some transformed plants of *Festuca* expressing FAE, compared to control (untransformed) plants. Examples are shown where vacuolar, ER or apoplast targeted FAE under a constitutive actin promoter have been effective at increasing IVDMD. Similar results were obtained with in leaves of *Lolium*, but were less pronounced.

The results can be seen in Figures 23 and 24.

Example 8

In vitro gas production measurements

In each experiment, 1.0-g samples of freeze dried powdered leaf tissue or cell culture were fermented in three 165-ml capacity serum bottles according to the pressure transducer technique of Theodorou et al. (Theodorou et al. (1994) A new gas production method using a pressure transducer to determine the fermentation kinetics of ruminant feeds. *Animal Feed Science and Technology* 48: 185-197). Grab samples of rumen-digesta were taken at 8.00 h before the morning feeding from fistulated wethers fed grass hay, and transported to the laboratory in a pre-warmed (39°C) vacuum flask. The microbial inoculum and culture media were prepared as described by Theodorou et al. (1994). Each serum bottle received 10 ml of microbial inoculum, 85 ml of buffer and 4 ml of reducing agent.

At the end of the incubation period, (144h) the contents of each serum bottle were filtered through pre-weighed sintered glass funnels and freeze dried to constant weight. Dry matter loss was calculated as the difference between the dry weight of the sample pre- and post-incubation. Additionally, the concentration of volatile fatty acids (VFA) in the liquid fraction of the culture media at the end of the 144-h incubation period was determined by gas chromatography. A Chrompack™ CP 9000 chromatograph fitted with an automatic sampler (Chrompack™ 911) and a flame-ionisation detector, linked to a Dell PC with A1-450 integration software, was used for VFA quantification.

Gas production data were fitted to the model of France et al. (France, J., Dhanoa, M.S., Theodorou, M.K, Lister, S.J., Davies. D.R. and Isac, D. 1993. A model to interpret gas accumulation profiles associated with *in vitro* degradation of ruminant feeds. *Journal of Theoretical Biology*. 163: 99-111.) using the MLP (Ross,

G.J.S. 1987. *MLP, Maximum Likelihood Program Version 3.08*. Oxford Numerical Algorithms Group) package. The equation is in the form, $Y = A\{1 - e^{[-b(t-T) - c(\sqrt{t}-\sqrt{T})]}\}$ where Y is the cumulative gas production (ml), A is the asymptote (i.e. gas pool), T is lag time, and b (h^{-1}) and c ($\text{h}^{-0.5}$) are decay rate constants. A combined fractional rate (h^{-1}) of gas production (μ) was calculated as, $\mu = b + c/2\sqrt{t}$, where t is the incubation time (h).

It can be seen for *Festuca arundinacea* (denoted as BN in Figure 25) that cell cultures have a higher rate of digestion and cumulative gas production in the presence of FAE and that the addition of an exogenous xylanase further enhance the availability of fermentable carbohydrates. Similar results are found in FAE expressing cultures without added FAE. Fermentation rates are further increased compared with controls by the addition of exogenous FAE or xylanase as these cultures expressing FAE have a reduced cell wall phenolic composition to controls Figures 26-28.

15

Example 9

FAE & xylanase transformed plants

Addition of exogenous xylanase (GC140) greatly increased FAE mediated release of phenolics from *Festuca* and *Lolium* leaves expressing *A. niger* FAE. See Figures 29-31 which show that phenolic release from leaf cell walls is increased in all FAE expressing plants on cell death and this is stimulated by xylanase irrespective of the targeting. Therefore expression of a fungal xylanase in plant cells is tested.

The FAE expression cassette is modified to comprise a fungal xylanase gene (either *T. reesei* or *A. niger*) to yield a FAE-xylanase expression cassette. The FAE-xylanase expression cassette is used to transform plant cells in a manner similar to those described in Example 3. The transformed cells are allowed to grow and are selected on an appropriate medium. The enzymes so expressed increase the availability of fermentable carbohydrates to a greater extent than the FAE expression cassette.

30

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will

be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

SEQUENCE LISTING

<110> Genencor International, Inc. and
Institute of Grassland and Environmental Research

<120> Manipulation of the Phenolic Acid
Content and Digestibility of Plant Cell Walls by Targeted
Expression of Genes Encoding Cell Wall Degrading Enzymes

<130> 11816-62

<140> CA 2,429,347

<141> 2001-11-16

<150> US 60/249,608

<151> 2000-11-17

<160> 97

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 2436

<212> DNA

<213> *Aspergillus niger*

<400> 1

ccatggtggt	gtcgatatcg	gcagtagtct	ttgccgaaac	gttgagggtt	acagtgatct	60
gcgtcggaca	tacttcgggg	aatctacggc	ggaatatcaa	agtcttcgga	atatccatat	120
tgggaaagga	cagaagctcc	ggggtagttt	gatagatgag	ctccggtgta	ttaaatacggg	180
agctgacagg	agtgagcgtc	atgtagacca	tctagtaatg	tcagtcgcgc	gcaatttcgc	240
acatgaaaca	agttgatttc	gggaccccat	tgttacatct	ctcggctaca	gctcgagatg	300
tgcttgccga	gtatacttag	aagccatgcc	agcgtgttgt	tatacgacca	aaagtcaggg	360
aatatgaaac	gatcgtcggg	tatttcttgt	ttttatccta	aattagtctt	ccagtggttt	420
atttaagaga	tagatccctt	cacaaacact	catccaacgg	acttctcata	ccactcattg	480
acataatttc	aaacagctcc	aggcgcattt	agttcaacat	gaagcaattc	tccgccaaac	540
acgtcctcgc	agttgtggtg	actgcagggc	acgccttagc	agcctctacg	caaggcatct	600
ccgaagacct	ctacagccgt	ttagtcgaaa	tggccactat	ctcccaagct	gcctacgccg	660
acctgtgcaa	cattccgtcg	actattatca	agggagagaa	aattttacaat	tctcaaactg	720
acattaacgg	atggatcctc	cgcgacgaca	gcagcaaaga	aataatcacc	gtcttccgtg	780
gcactggtag	tgatacgaat	ctacaactcg	atactaacta	caccctcacg	cctttcgaca	840
ccctaccaca	atgcaacggg	tgtgaagtac	acgggtggata	ttatattgga	tgggtctccg	900
tccaggacca	agtcgagtcg	cttgtcaaac	agcagggttag	ccagtatccg	gactatgcgc	960
tgactgtgac	gggccacagg	tatgccctcg	tgatttcttt	caattaagtg	tataatactc	1020
actaactcta	cgatagtctc	ggagcgtccc	tggcagcact	cactgccgcc	cagctgtctg	1080
cgacatacga	caacatccgc	ctgtacacct	tcggcgaaac	gcgcagcggc	aatcaggcct	1140
tcgcgtcgta	catgaacgat	gccttccaag	cctcgagccc	agatacgacg	cagtatttcc	1200
gggtcactca	tgccaacgac	ggcatcccaa	acctgcccc	ggtggagcag	gggtacgccc	1260
atggcggtgt	agagtactgg	agcgttgatc	cttacagcgc	ccagaacaca	tttgtctgca	1320
ctggggatga	agtgcagtgc	tgtgaggccc	agggcggaca	gggtgtgaat	aatgcgcaca	1380
cgacttatatt	tgggatgacg	agcggagcct	gtacatgggtg	atcagtcatt	tcagcctccc	1440
cgagtgtacc	aggaaagatg	gatgtcctgg	agagggcatg	catgtacgta	tacccgaagc	1500
acactttttc	ggtaaatcag	gacatgtaat	aagttccttc	catgaataga	tatgggttacc	1560
ctcaccataa	gccttgaggt	tgcctttctc	ttttgattgt	gaatatatat	ttaaagtaga	1620
tgacagatat	ctctaaacac	cttatccgct	taaacccatc	atagattgtg	tcacgtgata	1680
gaccccttga	atgatgagcg	aaatgtatca	gtcccgttta	aatcaaacc	tttcagccta	1740
gcacagtcag	aatacaccaa	ccccattcta	aggtagtact	aaatatgaat	acagcctaaa	1800
tgcatcgcta	tatgatccca	taaagaagca	acaacctttc	agatctcggt	ttgcgctgcg	1860
aagagctagc	tctaccatgg	tctcaattat	gagtggagcg	tttagtctcg	tttaagccta	1920
gctatcttat	aaggacaaca	catgtacatg	ggcttacttg	tagagaggta	ggatccccggg	1980
cttcttcaca	tctcgaggag	ttgtctacac	gtcgcgtcca	tgtcataagc	cggtactcga	2040

```

cggtgtcgtg accgtgaccc agaccctgtg tgatagcgtt gagaaggccc tatatttgaa 2100
tttccaatct cagctttacg aagatatgcc catggtggag ggtagtaaa ccgatgatga 2160
tcgtgtgcag catgagatga gaccgtggcc aatcctgttc aaatgccaag acccgctcc 2220
taccacatgt aaggcatccg tcggccgcac gttgaattgt gcaaatgccg agatcataaa 2280
agcggccaca cttccacgtc ggtactggat gggttgcgcg tggccatact gtgttttcca 2340
ttgcgtgggt cgttcgtgtt actgcgacgc agattctgta ggcaaggcgc agggctctct 2400
tctgaggtag aaaacacccc atattaatct gaattc 2436

```

<210> 2
 <211> 281
 <212> PRT
 <213> *Aspergillus niger*

<400> 2

Met	Lys	Gln	Phe	Ser	Ala	Lys	His	Val	Leu	Ala	Val	Val	Val	Thr	Ala	1	5	10	15
Gly	His	Ala	Leu	Ala	Ala	Ser	Thr	Gln	Gly	Ile	Ser	Glu	Asp	Leu	Tyr	20	25	30	
Ser	Arg	Leu	Val	Glu	Met	Ala	Thr	Ile	Ser	Gln	Ala	Ala	Tyr	Ala	Asp	35	40	45	
Leu	Cys	Asn	Ile	Pro	Ser	Thr	Ile	Ile	Lys	Gly	Glu	Lys	Ile	Tyr	Asn	50	55	60	
Ser	Gln	Thr	Asp	Ile	Asn	Gly	Trp	Ile	Leu	Arg	Asp	Asp	Ser	Ser	Lys	65	70	75	80
Glu	Ile	Ile	Thr	Val	Phe	Arg	Gly	Thr	Gly	Ser	Asp	Thr	Asn	Leu	Gln	85	90	95	
Leu	Asp	Thr	Asn	Tyr	Thr	Leu	Thr	Pro	Phe	Asp	Thr	Leu	Pro	Gln	Cys	100	105	110	
Asn	Gly	Cys	Glu	Val	His	Gly	Gly	Tyr	Tyr	Ile	Gly	Trp	Val	Ser	Val	115	120	125	
Gln	Asp	Gln	Val	Glu	Ser	Leu	Val	Lys	Gln	Gln	Val	Ser	Gln	Tyr	Pro	130	135	140	
Asp	Tyr	Ala	Leu	Thr	Val	Thr	Gly	His	Ser	Leu	Gly	Ala	Ser	Leu	Ala	145	150	155	160
Ala	Leu	Thr	Ala	Ala	Gln	Leu	Ser	Ala	Thr	Tyr	Asp	Asn	Ile	Arg	Leu	165	170	175	
Tyr	Thr	Phe	Gly	Glu	Pro	Arg	Ser	Gly	Asn	Gln	Ala	Phe	Ala	Ser	Tyr	180	185	190	
Met	Asn	Asp	Ala	Phe	Gln	Ala	Ser	Ser	Pro	Asp	Thr	Thr	Gln	Tyr	Phe	195	200	205	
Arg	Val	Thr	His	Ala	Asn	Asp	Gly	Ile	Pro	Asn	Leu	Pro	Pro	Val	Glu	210	215	220	
Gln	Gly	Tyr	Ala	His	Gly	Gly	Val	Glu	Tyr	Trp	Ser	Val	Asp	Pro	Tyr	225	230	235	240
Ser	Ala	Gln	Asn	Thr	Phe	Val	Cys	Thr	Gly	Asp	Glu	Val	Gln	Cys	Cys	245	250	255	
Glu	Ala	Gln	Gly	Gly	Gln	Gly	Val	Asn	Asn	Ala	His	Thr	Thr	Tyr	Phe	260	265	270	
Gly	Met	Thr	Ser	Gly	Ala	Cys	Thr	Trp								275	280		

<210> 3
 <211> 40
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> functional PCR product reading frame

<400> 3
 ggactacgcg ctgaccgtga ccggccactc cctcggcgcc

40

<210> 4
 <211> 35
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> inactivated PCR product reading frame

<400> 4
 ccggccacgc cctcggcgcc tccctggcgg cactc 35

<210> 5
 <211> 10
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> retention sequence

<400> 5
 Ala Ala Ala Glu Pro Leu Lys Asp Glu Leu
 1 5 10

<210> 6
 <211> 33
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> retention sequence encoding sequence

<400> 6
 gcggccgcgg aaccactgaa ggatgagctg taa 33

<210> 7
 <211> 15
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> FAE-linker-frameshift sequence

<400> 7
 Gly Ala Cys Thr Trp Pro Val Ala Ala Ala Glu Thr Thr Glu Gly
 1 5 10 15

<210> 8
 <211> 48
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> FAE-linker-frameshift sequence

<400> 8
 ggcgcacatgca cctggccgggt cgcggccgcg gaaaccactg aaggatga 48

<210> 9
 <211> 41
 <212> PRT

<213> Hordeum sp.

<400> 9

```
Met Ala His Ala Arg Val Leu Leu Leu Ala Leu Ala Val Leu Ala Thr
 1             5             10             15
Ala Ala Val Ala Val Ala Ser Ser Ser Ser Phe Ala Asp Ser Asn Pro
                20             25             30
Ile Arg Pro Val Thr Asp Arg Ala Ala
        35             40
```

<210> 10

<211> 134

<212> DNA

<213> Hordeum sp.

<400> 10

```
aagcttacca tggcccacgc ccgcgtcctc ctcttggegc tcgccgtgct ggccacggcc      60
gccgtcgccg tcgcctcctc ctctccttc gccgactcca acccgatccg gcccgtcacc      120
gaccgcgcgg ccgc                                     134
```

<210> 11

<211> 46

<212> PRT

<213> Rattus sp.

<400> 11

```
Met Ile His Thr Asn Leu Lys Lys Lys Phe Ser Leu Phe Ile Leu Val
 1             5             10             15
Phe Leu Leu Phe Ala Val Ile Cys Val Trp Lys Lys Gly Ser Asp Tyr
                20             25             30
Glu Ala Leu Thr Leu Gln Ala Lys Glu Phe Gln Met Ala Ala
        35             40             45
```

<210> 12

<211> 149

<212> DNA

<213> Rattus sp.

<400> 12

```
aagcttacca tgatccacac caacctcaaa aagaagttct ccctcttcat cctcgtcttc      60
ctcctcttcg ccgtgatctg cgtgtggaag aagggtccg actacgaggc cctcaccctc      120
caagccaagg agttccaaat ggcggccgc                                     149
```

<210> 13

<211> 50

<212> PRT

<213> Solanum sp.

<220>

<221> VARIANT

<222> (1)...(50)

<223> Xaa = Any Amino Acid

<400> 13

```
Met Xaa Val His Lys Glu Val Asn Phe Val Ala Tyr Leu Leu Ile Val
 1             5             10             15
Leu Gly Leu Leu Leu Leu Val Ser Ala Met Glu His Val Asp Ala Lys
                20             25             30
Ala Cys Thr Xaa Glu Cys Gly Asn Leu Gly Phe Gly Ile Cys Pro Ala
        35             40             45
Ala Ala
```

50

<210> 14
 <211> 159
 <212> DNA
 <213> Solanum sp.

<400> 14
 aagcttacma tggmcgtgca caaggaggts aacttcgtsg cctacctcct gatcgtsctc 60
 ggcctcctct tgctcgtstc cgccatggag cacgtggacg ccaaggcctg caccckcgag 120
 tgcggcaacc tcggcttcgg catctgcccc gcggccgcc 159

<210> 15
 <211> 5338
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pTP10-1 vector

<400> 15
 aagcttacca tggcccacgc ccgcgtcctc ctcttggcgc tcgcctgctt ggccacggcc 60
 gccgtcgccg tcgcctcctc ctctccttcc gccgactcca acccgatccg gcccgtcacc 120
 gaccgcgcgg ccgcctccac gcagggcacc tccgaagacc tctacagccg tttagtcgaa 180
 atggccacta tctcccaagc tgccctacgc gacctgtgca acattccgtc gactattatc 240
 aagggagaga aaattttacaa ttctcaaact gacattaacg gatggatcct ccgcgacgac 300
 agcagcaaag aaataatcac cgtcttccgt ggcactggta gtgatacgaa tctacaactc 360
 gatactgact acaccctcac gcctttcgac accctaccac aatgcaacgg ttgtgaagta 420
 cacgggtggat attatatattg atgggtctcc gtccaggacc aagtcgagtc gcttgtcaaa 480
 cagcaggtta gccagtatcc ggactacgcg ctgaccgtga ccggccackc cctcggcgcc 540
 tccctggcgg cactcactgc cgcctcagct tctgcgacat acgacaacat ccgcctgtac 600
 accttcggcg aaccgcgcag cggcaatcag gccttcgcgt cgtacatgaa cgatgccttc 660
 caagcctcga gccagatac gacgcagtat ttccgggtca ctcatgccaa cgacggcacc 720
 ccaaacctgc ccccggtgga gcaggggtac gcccatggcg gtgtagagta ctggagcgtt 780
 gatccttaca gcgcccagaa cacatttgtc tgcactgggg atgaagtgca gtgctgtgag 840
 gccagggcg gacaggggtg gaataatgcg cacacgactt attttgggat gacgagcggc 900
 gcatgcacct ggccggctgc ggccgcggaa accactgaag gatgagctgt aaagaagcag 960
 atcgttcaaa cathttggcaa taaagtttct taagattgaa tcctgttgcc ggtcttgcca 1020
 tgattatcat ataatttctg ttgaattacg ttaagcatgt aataattaac atgtaatgca 1080
 tgacgttatt tatgagatgg gtttttatga ttagagtcct gcaattatac atttaatacg 1140
 cgatagaaaa caaaatatag cgcgcaaact aggataaatt atcgcgcgcg gtgtcatcta 1200
 tgttactaga tcgataagct tctagagcgg ccggtggagc tccaattcgc cctatagtga 1260
 gtcgtattac gcgcgctcac tggccgtcgt tttaaacgt cgtgactggg aaaaccctgg 1320
 cgttacccaa cttaatcgcc ttgcagcaca tcccccttcc gccagctggc gtaatagcga 1380
 agaggccccg accgatcgcc cttcccaaca gttgcgcagc ctgaatggcg aatgggacgc 1440
 gccctgtagc ggcgcattaa gcgcggcggg tgtgggtggt acgcgcagcg tgaccgctac 1500
 acttgccagc gccctagcgc ccgctccttt cgctttcttc ccttcctttc tcgccacggt 1560
 cgccggcttt ccccgtaag ctctaaatcg ggggctccct ttaggggttc gatttagtgc 1620
 tttacggcac ctgcaccca aaaaacttga ttaggggtgat ggttcacgta gtgggccatc 1680
 gccctgatag acgggttttcc gccctttgac gttggagtcc acgttcttta atagtggact 1740
 cttgttccaa actggaacaa cactcaaccc tatctcggtc tattcttttg atttataagg 1800
 gattttgccg atttcggcct attgggttaaa aaatgagctg atttaacaaa aatttaacgc 1860
 gaattttaac aaaatatata cgcttacaat ttaggtggca cttttcgggg aaatgtgcgc 1920
 ggaaccccta tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa 1980
 taacctgat aaatgcttca ataataattga aaaaggaaga gtatgagtat tcaacatttc 2040
 cgtgtcgccc ttattccctt ttttgoggca ttttgccctc ctgttttttg tcaccagaa 2100
 acgctgggtga aagtaaaaga tgctgaagat cagttgggtg cacgagtggg ttacatcgaa 2160
 ctggatctca acagcggtaa gatccttgag agttttcgcc ccgaagaacg ttttccaatg 2220
 atgagcactt ttaaagttct gctatgtggc gcgggtattat cccgtattga cgccgggcaa 2280
 gagcaactcg gtcgccgcac acactattct cagaatgact tgggttgagta ctaccagtc 2340
 acagaaaagc atcttacgga tggcatgaca gtaagagaat tatgcagtcg tgccataacc 2400

atgagtgata	acactgcggc	caacttactt	ctgacaacga	tgggaggacc	gaaggagcta	2460
accgcttttt	tgcacaacat	gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	2520
ctgaatgaag	ccataccaaa	cgacgagcgt	gacaccacga	tgcctgtagc	aatggcaaca	2580
acgttgcgca	aactattaac	tggcgaaacta	cttactctag	cttcccggca	acaattaata	2640
gactggatgg	aggcggataa	agttgcagga	ccacttctgc	gctcggccct	tccggctggc	2700
tggtttattg	ctgataaatc	tggagccggt	gagcgtgggt	ctcgcggtat	cattgcagca	2760
ctggggccag	atggtaagcc	ctcccgtatc	gtagttatct	acacgacggg	gagtcaggca	2820
actatggatg	aacgaaatag	acagatcgct	gagataggtg	cctcactgat	taagcattgg	2880
taactgtcag	accaagttta	ctcatatata	cttttagattg	atttaaaact	tcatttttta	2940
tttaaaagga	tctaggtgaa	gacccctttt	gataatctca	tgaccaaaat	cccttaacgt	3000
gagttttcgt	tccactgagc	gtcagacccc	gtagaaaaga	tcaaaggatc	ttcttgagat	3060
cctttttttc	tgcgcgtaat	ctgctgcttg	caaacaaaaa	aaccaccgct	accagcgggtg	3120
gtttgtttgc	cggatcaaga	gctaccaact	ctttttccga	aggtaactgg	cttcagcaga	3180
gcgcagatac	caaatactgt	ccttctagtg	tagccgtagt	taggccacca	cttcaagaac	3240
tctgtagcac	cgcctacata	cctcgctctg	ctaactctgt	taccagtggc	tgctgccagt	3300
ggcgataagt	cgtgtcttac	cgggttggac	tcaagacgat	agttaccgga	taaggcgcag	3360
cggtcgggct	gaacgggggg	ttcgtgcaca	cagcccagct	tggagcgaac	gacctacacc	3420
gaactgagat	acctacagcg	tgagctatga	gaaagcgcca	cgcttcccga	agggagaaaag	3480
gcggacaggt	atccggtaag	cggcagggtc	ggaacaggag	agcgcacgag	ggagcttcca	3540
gggggaaacg	cctggtatct	ttatagtcct	gtcgggtttc	gccacctctg	acttgagcgt	3600
cgatttttgt	gatgctcgtc	agggggggcg	agcctatgga	aaaacgccag	caacgcggcc	3660
tttttacggt	tccctggcct	ttgctggcct	tttgctcaca	tgttctttcc	tgcgttatcc	3720
cctgattctg	tggataaccg	tattaccgcc	tttgagttag	ctgataccgc	tcgccgcagc	3780
cgaacgaccg	agcgcagcga	gtcagtgagc	gaggaaagcg	aagagcgccc	aatacgcaaa	3840
ccgcctctcc	ccgcgcgttg	gccgattcat	taatgcagct	ggcacgacag	gtttcccgcac	3900
tggaaagcgg	gcagtgagcg	caacgcaatt	aatgtgagtt	agctcactca	ttaggcaccc	3960
caggctttac	actttatgct	tccggctcgt	atgttggtgtg	gaattgtgag	cggataacaa	4020
tttcacacag	gaaacagcta	tgaccatgat	tacgccaaagc	gcgcaattaa	ccctcactaa	4080
agggaacaaa	agctgggtac	cgggcccccc	ctcgagggtca	ttcatatgct	tgagaagaga	4140
gtcgggatag	tccaaaataa	aacaaaggta	agattacctg	gtcaaaaagt	aaaacatcag	4200
ttaaaagggtg	gtataagtaa	aatatcggta	ataaaaagggtg	gccccaaagt	aaatttactc	4260
ttttctacta	ttataaaaaat	tgaggatggt	ttgtcggtac	tttgatacgt	cattttttgta	4320
tgaattgggtt	tttaagttta	ttcgcgattt	ggaaatgcat	atctgtatct	gagtcgggtt	4380
ttaagttcgt	tgctttttgta	aatacagagg	gattttgtata	agaaatatct	ttaaaaaacc	4440
catatgctaa	tttgacataa	tttttgagaa	aaatatatat	tcaggcgaat	tccacaatga	4500
acaataataa	gattaaaata	gcttgccccc	gttgccagcga	tgggtatctt	ttctagtaaa	4560
ataaaagata	aacttagact	caaaaacattt	acaaaaacaa	cccctaaagt	cctaaagccc	4620
aaagtgcctat	gcacgatcca	tagcaagccc	agcccaaccc	aacccaaccc	aacccacccc	4680
agtgcagcca	actggcaaat	agtctccacc	cccggcacta	tcaccgtgag	ttgtccgcac	4740
caccgcacgt	ctcgcagcca	aaaaaaaaaa	aagaaagaaa	aaaaagaaaa	agaaaaaacag	4800
cagggtgggtc	cgggtcgtgg	gggcccggaaa	agcagaggagg	atcgcgagca	gcgacgaggc	4860
ccggccctcc	ctccgcttcc	aaagaaacgc	cccccatcgc	cactatatac	ataccccccc	4920
ctctcctccc	atccccccaa	ccctaccacc	accaccacca	ccacctcctc	ccccctcgtc	4980
gccggacgac	gagctcctcc	cccctccccc	tccgcgcgcg	ccggtaacca	ccccgcccct	5040
ctcctctttc	tttctcgttt	tttttttttcg	tctcgggtctc	gatctttggc	cttggttagtt	5100
tgggtggggcg	agagcggctt	cgtcgcgccag	atcgggtgcgc	gggagggggcg	ggatctcgcg	5160
gctggcgtct	ccgggcgtga	gtcgggcccg	atcctcgcgg	ggaatggggc	tctcggatgt	5220
agatcttctt	tctttcttct	ttttgtggta	gaatttgaat	ccctcagcat	tgttcatcgg	5280
tagtttttct	tttcatgatt	tgtgacaaat	gcagcctcgt	gcggagcttt	tttgtagc	5338

<210> 16

<211> 311

<212> PRT

<213> Artificial Sequence

<220>

<223> pTP10-1 vector

<221> VARIANT

<222> (1)...(311)

<223> Xaa = Any Amino Acid

<400> 16

```

Met Ala His Ala Arg Val Leu Leu Leu Ala Leu Ala Val Leu Ala Thr
 1          5          10          15
Ala Ala Val Ala Val Ala Ser Ser Ser Ser Phe Ala Asp Ser Asn Pro
          20          25          30
Ile Arg Pro Val Thr Asp Arg Ala Ala Ala Ser Thr Gln Gly Ile Ser
          35          40          45
Glu Asp Leu Tyr Ser Arg Leu Val Glu Met Ala Thr Ile Ser Gln Ala
          50          55          60
Ala Tyr Ala Asp Leu Cys Asn Ile Pro Ser Thr Ile Ile Lys Gly Glu
65          70          75          80
Lys Ile Tyr Asn Ser Gln Thr Asp Ile Asn Gly Trp Ile Leu Arg Asp
          85          90          95
Asp Ser Ser Lys Glu Ile Ile Thr Val Phe Arg Gly Thr Gly Ser Asp
          100          105          110
Thr Asn Leu Gln Leu Asp Thr Asp Tyr Thr Leu Thr Pro Phe Asp Thr
          115          120          125
Leu Pro Gln Cys Asn Gly Cys Glu Val His Gly Gly Tyr Tyr Ile Gly
          130          135          140
Trp Val Ser Val Gln Asp Gln Val Glu Ser Leu Val Lys Gln Gln Val
145          150          155          160
Ser Gln Tyr Pro Asp Tyr Ala Leu Thr Val Thr Gly His Xaa Leu Gly
          165          170          175
Ala Ser Leu Ala Ala Leu Thr Ala Ala Gln Leu Ser Ala Thr Tyr Asp
          180          185          190
Asn Ile Arg Leu Tyr Thr Phe Gly Glu Pro Arg Ser Gly Asn Gln Ala
          195          200          205
Phe Ala Ser Tyr Met Asn Asp Ala Phe Gln Ala Ser Ser Pro Asp Thr
          210          215          220
Thr Gln Tyr Phe Arg Val Thr His Ala Asn Asp Gly Ile Pro Asn Leu
225          230          235          240
Pro Pro Val Glu Gln Gly Tyr Ala His Gly Gly Val Glu Tyr Trp Ser
          245          250          255
Val Asp Pro Tyr Ser Ala Gln Asn Thr Phe Val Cys Thr Gly Asp Glu
          260          265          270
Val Gln Cys Cys Glu Ala Gln Gly Gly Gln Gly Val Asn Asn Ala His
          275          280          285
Thr Thr Tyr Phe Gly Met Thr Ser Gly Ala Cys Thr Trp Pro Val Ala
          290          295          300
Ala Ala Glu Thr Thr Glu Gly
305          310

```

<210> 17

<211> 5345

<212> DNA

<213> Artificial Sequence

<220>

<223> pUA4-4 vector

<400> 17

```

aagcttacca tggcccacgc ccgcgtcctc ctcttggcgc tcgccgtgct ggccacggcc      60
gccgtcgccg tcgcctcctc ccgcgcggcc gcctccacgc agggcatctc cgaagacctc      120
tacagccgtt tagtcgaaat ggccactatc tccaagctg cctacgccga cctgtgcaac      180
attccgtcga ctattatcaa gggagagaaa atttacaatt ctcaaactga cattaacgga      240
tggatcctcc gcgacgacag cagcaaagaa ataatcacgc tcttcctggt cactggtagt      300
gatacgaatc tacaactcga tactaactac accctcacgc ctttcgacac cctaccacaa      360
tgcaacgggt gtgaagtaca cggtggatat tatattggat gggctctccgt ccaggaccaa      420
gtcgagtcgc ttgtcaaaca gcaggtagc cagtatccgg actacgcgct gaccgtgacc      480
ggccackccc tcggcgctc cctggcgcca ctactgccg cccagctgtc tgcgacatac      540

```

gacaacatcc	gcctgtacac	cttcggcgaa	ccgcgcagcg	gcaatcaggc	cttcgcgctcg	600
tacatgaacg	atgccttcca	agcctcgagc	ccagatacga	cgcagtattt	ccgggtcact	660
catgccaacg	acggcatccc	aaacctgccc	ccgggtggagc	aggggtacgc	ccatggcggt	720
gtagagtact	ggagcgttga	tccttacagc	gcccagaaca	catttgtctg	cactggggat	780
gaagtgcagt	gctgtgaggc	ccagggcgga	caggggtgtga	ataatgcgca	cacgacttat	840
tttgggatga	cgagcggagc	ctgtacatgg	tgatcagtca	tttcagcctc	cccagagtga	900
ccaggaaaga	tggatgtcct	ggagaggggg	ccgcgtaacc	actgaaggat	gagctgtaaa	960
gaagcagatc	gttcaaacat	ttggcaataa	agtttcttaa	gattgaatcc	tggtgccggt	1020
cttgcgatga	ttatcatata	atttctgttg	aattacgtta	agcatgtaat	aattaacatg	1080
taatgcatga	cgttatttat	gagatgggtt	tttatgatta	gagtcgccga	attatacatt	1140
taatacgcga	tagaaaacaa	aatatagcgc	gcaaactagg	ataaattatc	gcgcgcgggtg	1200
tcattctatgt	tactagatcg	ataagcttct	agagcggccg	gtggagctcc	aattcgcct	1260
atagtgcgct	gtattacgcg	cgctcactgg	ccgtcgtttt	acaacgctcg	gactgggaaa	1320
acctggcgt	tacccaactt	aatcgcttg	cagcacatcc	ccctttcgcc	agctggcgta	1380
atagcgaaga	ggcccgccac	gatcgccctt	cccaacagtt	gcgcagcctg	aatggcgaat	1440
gggacgcgcc	ctgtagcggc	gcattaagcg	cggcgggtgt	ggtgggttacg	cgcagcgtga	1500
ccgctacact	tgccagcgcc	ctagcgcccc	ctcctttcgc	tttcttcct	tcctttctcg	1560
ccacgttcgc	cggctttccc	cgtcaagctc	taaatcgggg	gctcccttta	gggttccgat	1620
ttagtgcctt	acggcacctc	gaccccaaaa	aacttgatta	gggtgatggg	tcacgtagtg	1680
ggccatcgcc	ctgatagacg	gtttttcgcc	ctttgacgtt	ggagtccacg	ttctttaata	1740
gtggactctt	gttccaaact	ggaacaacac	tcaaccctat	ctcggctctat	tcttttgatt	1800
tataagggat	tttgccgatt	tcggcctatt	ggttaaaaaa	tgagctgatt	taacaaaaat	1860
ttaacgcgaa	ttttaacaaa	atattaacgc	ttacaattta	ggtggcactt	ttcggggaaa	1920
tgtgcgcgga	acctctat	gtttat	ctaaatacat	tcaaataatgt	atccgctcat	1980
gagacaataa	ccctgataaa	tgtttcaata	atattgaaaa	aggaagagta	tgagtattca	2040
acatttccgt	gtcgccttta	ttcccttttt	tgcggcattt	tgccttcctg	tttttgctca	2100
cccagaaacg	ctgggtgaaag	taaaagatgc	tgaagatcag	ttgggtgcac	gagtgggtta	2160
catcgaactg	gatctcaaca	gcggtaagat	ccttgagagt	tttcgccccg	aagaacgttt	2220
tccaatgatg	agcactttta	aagttctgct	atgtggcgcg	gtattatccc	gtattgacgc	2280
cgggcaagag	caactcggtc	gccgcataca	ctattctcag	aatgacttgg	ttgagtactc	2340
accagtcaca	gaaaagcatc	ttacggatgg	catgacagta	agagaattat	gcagtgcctgc	2400
cataaccatg	agtataaca	ctgcggccaa	cttacttctg	acaacgatcg	gaggaccgaa	2460
ggagctaacc	gcttttttgc	acaacatggg	ggatcatgta	actcgcttg	atcgttgga	2520
accggagctg	aatgaagcca	taccaaacga	cgagcgtgac	accacgatgc	ctgtagcaat	2580
ggcaacaacg	ttgcgcaaac	tattaactgg	cgaactactt	actctagctt	cccggcaaca	2640
attaatagac	tggatggagg	cggataaagt	tgcaggacca	cttctgcgct	cggcccttcc	2700
ggctggctgg	tttattgctg	ataaatctgg	agccgggtgag	cgtgggtctc	gcggtatcat	2760
tgcagcactg	gggccagatg	gtaagccctc	ccgtatcgta	gttatctaca	cgacggggag	2820
tcaggcaact	atggatgaac	gaaatagaca	gatcgctgag	ataggtgcct	cactgattaa	2880
gcattggtaa	ctgtcagacc	aagtttactc	atatatactt	tagattgatt	taaaacttca	2940
tttttaattt	aaaaggatct	aggtgaagat	cctttttgat	aatctcatga	ccaaaatccc	3000
ttaacgtgag	ttttcgttcc	actgagcgtc	agaccccgta	gaaaagatca	aaggatcttc	3060
ttgagatcct	ttttttctgc	gcgtaatctg	ctgcttgcaa	acaaaaaaac	caccgctacc	3120
agcgggtggt	tgtttgccgg	atcaagagct	accaactctt	tttccgaagg	taactggctt	3180
cagcagagcg	cagataccaa	atactgtcct	tctagtgtag	ccgtagttag	gccaccactt	3240
caagaactct	gtagcaccgc	ctacatacct	cgctctgcta	atcctgttac	cagtggctgc	3300
tgccagtggc	gataagtcgt	gtcttaccgg	gttggtactca	agacgatagt	taccggataa	3360
ggcgcagcgg	tcgggctgaa	cgggggggtt	gtgcacacag	cccagcttgg	agcgaacgac	3420
ctacaccgaa	ctgagatacc	tacagcgtga	gctatgagaa	agcggccacgc	ttcccgaagg	3480
gagaaaggcg	gacaggtatc	cggtaagcgg	cagggtcgga	acaggagagc	gcacgaggga	3540
gcttccaggg	ggaaacgcct	ggtatcttta	tagtcctgtc	gggtttcgcc	acctctgact	3600
tgagcgtcga	tttttgtgat	gctcgtcagg	ggggcggagc	ctatggaaaa	acgccagcaa	3660
cgcggccttt	ttacggttcc	tggccttttg	ctggcctttt	gctcacatgt	tctttcctgc	3720
gttatccct	gattctgtgg	ataaccgtat	taccgccttt	gagtgcgctg	ataccgctcg	3780
ccgcagccga	acgaccgagc	gcagcgcgag	agtgcgcgag	gaagcgggaag	agcgcccaat	3840
acgcaaaccg	cctctccccg	cgcgttggcc	gattcattaa	tgacgctggc	acgacagggt	3900
tcccgaactg	aaagcgggca	gtgagcgcga	cgcaattaat	gtgagttagc	tactcatta	3960
ggcaccaccg	gctttacact	ttatgcttcc	ggctcgtatg	ttgtgtggaa	ttgtgagcgg	4020
ataacaattt	cacacaggaa	acagctatga	ccatgattac	gccaaagcgc	caattaaccc	4080
tcactaaagg	gaacaaaagc	tgggtaccgg	gccccccctc	gaggtcattc	atatgcttga	4140
gaagagagtc	gggatagtc	aaaataaaac	aaaggtaaga	ttacctggtc	aaaagtga	4200

```

acatcagtta aaaggtggta taagtaaaat atcggtaata aaaggtggcc caaagtgaaa 4260
tttactcttt tctactatta taaaaattga ggatgttttg tcggtacttt gatacgtcac 4320
ttttgtatga attgggtttt aagttttattc gcgatttgga aatgcataac tgtatttgag 4380
tcgggttttta agttcggttc ttttgtaaat acagagggat ttgtataaga aatatcttta 4440
aaaaacccat atgctaattt gacataattt ttgagaaaaa tatatatcca ggccaattcc 4500
acaatgaaca ataataagat taaaatagct tgcccccggt gcagcgatgg gtattttttc 4560
tagtaaaata aaagataaac ttagactcaa aacatttaca aaaacaaccc ctaaagtcct 4620
aaagcccaaa gtgctatgca cgatccatag caagcccagc ccaacccaac ccaacccaac 4680
ccaccccagt gcagccaact ggcaaatagt ctccaccccc ggccactatca ccgtgagttg 4740
tcgcgaccac cgcacgtctc gcagccaaaa aaaaaaaaaa aaagaaaaaa aagaaaaaga 4800
aaaacagcag gtgggtccgg gtcgtggggg ccggaaaagc gaggaggatc gcgagcagcg 4860
acgaggcccg gccctccctc cgcttccaaa gaaacgcccc ccacgcgccac tatatacata 4920
ccccccctc tcctcccatc cccccaaccc taccaccacc accaccacca cctcctcccc 4980
cctcgctgcc ggacgacgag ctctccccc ctcccccctc gccgcgcgcg gtaaccaccc 5040
cgcccctctc ctctttcttt ctccgttttt tttttcgtct cgggtctcgat ctttggcctt 5100
ggtagtttgg gtgggcgaga gcggcttcgt cgcccagatc ggtgcgcggg aggggcggga 5160
tctcgcggtt ggcgtctccg ggcgtgagtc ggcccggatc ctgcgcggga atggggctct 5220
cggatgtaga tcttctttct ttcttctttt tgtggtagaa tttgaatccc tcagcattgt 5280
tcacggttag tttttctttt catgatttgt gacaaatgca gcctcgtgcg gagctttttt 5340
gtagc 5345

```

<210> 18

<211> 287

<212> PRT

<213> Artificial Sequence

<220>

<223> pUA4-4 vector

<221> VARIANT

<222> (1)...(287)

<223> Xaa = Any Amino Acid

<400> 18

```

Met Ala His Ala Arg Val Leu Leu Leu Ala Leu Ala Val Leu Ala Thr
1          5          10          15
Ala Ala Val Ala Val Ala Ser Ser Arg Ala Ala Ala Ser Thr Gln Gly
20          25          30
Ile Ser Glu Asp Leu Tyr Ser Arg Leu Val Glu Met Ala Thr Ile Ser
35          40          45
Gln Ala Ala Tyr Ala Asp Leu Cys Asn Ile Pro Ser Thr Ile Ile Lys
50          55          60
Gly Glu Lys Ile Tyr Asn Ser Gln Thr Asp Ile Asn Gly Trp Ile Leu
65          70          75          80
Arg Asp Asp Ser Ser Lys Glu Ile Ile Thr Val Phe Arg Gly Thr Gly
85          90          95
Ser Asp Thr Asn Leu Gln Leu Asp Thr Asn Tyr Thr Leu Thr Pro Phe
100          105          110
Asp Thr Leu Pro Gln Cys Asn Gly Cys Glu Val His Gly Gly Tyr Tyr
115          120          125
Ile Gly Trp Val Ser Val Gln Asp Gln Val Glu Ser Leu Val Lys Gln
130          135          140
Gln Val Ser Gln Tyr Pro Asp Tyr Ala Leu Thr Val Thr Gly His Xaa
145          150          155          160
Leu Gly Ala Ser Leu Ala Ala Leu Thr Ala Ala Gln Leu Ser Ala Thr
165          170          175
Tyr Asp Asn Ile Arg Leu Tyr Thr Phe Gly Glu Pro Arg Ser Gly Asn
180          185          190
Gln Ala Phe Ala Ser Tyr Met Asn Asp Ala Phe Gln Ala Ser Ser Pro
195          200          205
Asp Thr Thr Gln Tyr Phe Arg Val Thr His Ala Asn Asp Gly Ile Pro

```

210		215		220
Asn Leu Pro Pro Val Glu Gln Gly Tyr Ala His Gly Gly Val Glu Tyr				
225		230		235
Trp Ser Val Asp Pro Tyr Ser Ala Gln Asn Thr Phe Val Cys Thr Gly				
	245		250	255
Asp Glu Val Gln Cys Cys Glu Ala Gln Gly Gly Gln Gly Val Asn Asn				
	260		265	270
Ala His Thr Thr Tyr Phe Gly Met Thr Ser Gly Ala Cys Thr Trp				
275		280		285

<210> 19

<211> 5337

<212> DNA

<213> Artificial Sequence

<220>

<223> pTU4 vector

<400> 19

aagcttacca	tggcccacgc	ccgcgtcctc	ctcctggcgc	tcgccgtgct	ggccacggcc	60
gccgtcgccg	tcgcctcctc	ctcctccttc	gccgactcca	acccgatccg	gcccgtcacc	120
gaccgcgcgg	ccgcctccac	gcagggcctc	tccgaagacc	tctacagccg	tttagtcgaa	180
atggccacta	tctcccaagc	tgcctacgcc	gacctgtgca	acattccgtc	gactattatc	240
aagggagaga	aaattttaca	ttctcaaact	gacattaacg	gatggatcct	ccgcgacgac	300
agcagcaaag	aaataatcac	cgtcttccgt	ggcactggta	gtgatacgaa	tctacaactc	360
gatactaact	acaccctcac	gcctttcgac	accctaccac	aatgcaacgg	ttgtgaagta	420
cacgggtggat	attatatattg	atgggtctcc	gtccaggacc	aagtcgagtc	gcttgtcaaa	480
cagcaggtta	gccagtatcc	ggactacgcg	ctgaccgtga	ccggccackc	cctcggcgcc	540
tccctggcgg	cactcactgc	cgcccagctg	tctgcgacat	acgacaacat	ccgcctgtac	600
accttcggcg	aaccgcgcag	cggcaatcag	gccttcgcgt	cgtacatgaa	cgatgccttc	660
caagcctcga	gccagatac	gacgcagtat	ttccgggtca	ctcatgccaa	cgacggcctc	720
ccaaacctgc	ccccgggtga	gcaggggtac	gcccattggc	gtgtagagta	ctggagcggt	780
gattccttaca	gcgcccagaa	cacatttgtc	tgcactgggg	atgaagtgca	gtgctgtgag	840
gcccagggcg	gacaggggtg	gaataatgcg	cacacgactt	atcttgggat	gacgagcggc	900
gcattgcacct	ggccgggtcg	ggccgcggaa	ccactgaagg	atgagctgta	aagaagcaga	960
tcgtttcaaac	atctggcaat	aaagtcttct	aagattgaat	cctgcttgccg	gtcttgcgat	1020
gattatcata	taattttctgt	tgaattacgt	taagcatgta	ataattaaca	tgtaatgcat	1080
gacgttatatt	atgagatggg	tttttatgat	tagagtcccg	caattataca	tttaatacgc	1140
gatagaaaac	aaaatatagc	gcgcaaacta	ggataaatta	tcgcgcgcgg	tgtcatctat	1200
gttactagat	cgataagctt	ctagagcggc	cggtggagct	ccaattcgcc	ctatagttag	1260
tcgtattacg	cgcgctcact	ggccgtcggt	ttacaacgtc	gtgactggga	aaaccctggc	1320
gttacccaac	ttaatcgctt	tgcagcacat	ccccctttcg	ccagctggcg	taatagcgaa	1380
gaggccccga	ccgatecgcc	ttcccaacag	ttgcgcagcc	tgaatggcga	atgggacgcg	1440
cctgttagcg	gcgcatttaag	cgccggcgggt	gtgggtggta	cgcgacgcgt	gaccgctaca	1500
cttgccagcg	ccctagcgcc	cgctcctttc	gctttcttcc	cttcctttct	cgccacgttc	1560
gccggctttc	cccgtcaagc	tctaaatcgg	gggtctccct	tagggttccg	atttagtgct	1620
ttacggcacc	tcgaccocaa	aaaacttgat	taggggtgat	gttcacgtag	tgggccatcg	1680
ccctgataga	cggtttttcg	ccctttgacg	ttggagtcca	cgttctttta	tagtggactc	1740
ttgtttccaa	ctggaacaac	actcaaccct	atctcggtct	attcttttga	tttataaggg	1800
atctttgcca	tttcggccta	ttggttaaaa	aatgagctga	tttaacaaaa	atttaacgcg	1860
aatttttaaca	aaatattaac	gcttacaatt	taggtggcac	ttttcgggga	aatgtgcgcg	1920
gaacccttat	ttgtttatatt	ttctaaatac	attcaaatat	gtatccgctc	atgagacaat	1980
aaccctgata	aatgcttcaa	taatattgaa	aaaggaagag	tatgagtatt	caacatttcc	2040
gtgtcgccct	tattcccttt	tttgccggcat	tttgcccttc	tgtttttgct	caccacagaa	2100
cgctggtgaa	agtaaaagat	gctgaagatc	agttgggtgc	acgagtgggt	tacatcgaa	2160
tggatctcaa	cagcggtaag	atccttgaga	gttttcgccc	cgaagaacgt	tttccaatga	2220
tgagcacttt	taaagtcttg	ctatgtggcg	cggtattatc	ccgtattgac	gccgggcaag	2280
agcaactcgg	tcgcccata	cactattctc	agaatgactt	ggttgagtac	tcaccagtca	2340
cagaaaagca	tcttacggat	ggcatgacag	taagagaatt	atgcagtgct	gccataacca	2400
tgagtgataa	cactgcggcc	aacttacttc	tgacaacgat	cggaggaccg	aaggagctaa	2460
ccgctttttt	gcacaacatg	ggggatcatg	taactcgctt	tgatcggttg	gaaccggagc	2520

tgaatgaagc	cataccaaac	gacgagcgtg	acaccacgat	gcctgtagca	atggcaacaa	2580
cgttgcgcaa	actattaact	ggcgaactac	ttactctagc	ttcccggcaa	caattaatag	2640
actggatgga	ggcggataaa	gttgcaggac	cacttctgcg	ctcggccctt	ccggctggct	2700
ggtttattgc	tgataaatct	ggagccgggtg	agcgtgggtc	tgcgggtatc	attgcagcac	2760
tggggccaga	tggtaagccc	tcccgtatcg	tagttatcta	cacgacgggg	agtcaggcaa	2820
ctatggatga	acgaaataga	cagatcgctg	agatagggtc	ctcactgatt	aagcattggg	2880
aactgtcaga	ccaagtttac	tcatatatac	tttagattga	tttaaaactt	catttttaat	2940
ttaaaaggat	ctaggtgaag	atcctttttg	ataatctcat	gaccaaatac	ccttaacgtg	3000
agttttcgtt	ccactgagcg	tcagaccccc	tagaaaagat	caaaggatct	tcttgagatc	3060
ctttttttct	gcgcgtaatc	tgctgcttgc	aaacaaaaaa	accaccgcta	ccagcgggtg	3120
tttgtttgcc	ggatcaagag	ctaccaactc	tttttccgaa	ggtaactggc	ttcagcagag	3180
cgcagatacc	aaatactgtc	cttctagtgt	agccgtagt	aggccaccac	ttcaagaact	3240
ctgtagcacc	gcctacatac	ctcgcctctg	taatcctgtt	accagtggct	gctgccagtg	3300
gcgataagtc	gtgtcttacc	gggttggtg	caagacgata	gttaccggat	aaggcgcagc	3360
ggtcgggctg	aacggggggg	tcgtgcacac	agcccagctt	ggagcgaacg	acctacaccg	3420
aactgagata	cctacagcgt	gagctatgag	aaagcggcac	gcttcccga	gggagaaagg	3480
cggacaggta	tccggtaagc	ggcagggtcg	gaacaggaga	gcgcacgagg	gagcttccag	3540
ggggaaacgc	ctggtatctt	tatagtcctg	tcgggtttcg	ccacctctga	cttgagcgtc	3600
gattttttgtg	atgctcgtca	ggggggcgga	gcctatggaa	aaacgccagc	aacgcggcct	3660
ttttacgggtt	cctggccttt	tgctggcctt	ttgctcacat	gttcttttct	gcgttatccc	3720
ctgattctgt	ggataaccgt	attaccgcct	ttgagtgagc	tgataccgct	cgccgcagcc	3780
gaacgaccga	gcgcagcgag	tcagtgagcg	aggaaagcga	agagcgccca	atacgcaaac	3840
cgcctctccc	cgcgcgttgg	ccgattcatt	aatgcagctg	gcacgacagg	tttcccgcact	3900
ggaaagcggg	cagtgagcgc	aacgcaatta	atgtgagtta	gctcactcat	taggcacccc	3960
aggctttaca	ctttatgctt	ccggctcgta	tggtgtgtgg	aattgtgagc	ggataacaat	4020
ttcacacagg	aaacagctat	gaccatgatt	acgccaagcg	cgcaattaac	cctcactaaa	4080
gggaacaaaa	gctgggtacc	gggccccccc	tcgagggtcat	tcatatgctt	gagaagagag	4140
tcgggatagt	ccaaaataaa	acaaaggtaa	gattacctgg	tcaaaagtga	aaacatcagt	4200
taaaagggtg	tataagtaaa	atatcggtaa	taaaagggtg	cccaaagtga	aatttactct	4260
tttctactat	tataaaaatt	gaggatgttt	tgctcggtact	ttgatacgtc	atttttgtat	4320
gaattgggtt	ttaagtttat	tcgcgatttg	gaaatgcata	tctgtatttg	agtcgggttt	4380
taagttcgtt	gcttttgtaa	atacagaggg	atttgtataa	gaaatatctt	taaaaaaccc	4440
atatgcta	ttgacataat	ttttgagaaa	aatatatatt	caggcgaatt	ccacaatgaa	4500
caataataag	attaaaatag	cttgcccccg	ttgcagcgat	gggtattttt	tctagtataa	4560
taaaagataa	acttagactc	aaaacattta	caaaaacaac	ccctaaagtc	ctaaagccca	4620
aagtgcata	cacgatccat	agcaagccca	gcccacccca	acccaaccca	acccacccca	4680
gtgcagccaa	ctggcaaata	gtctccaccc	ccggcactat	caccgtgagt	tgtccgcacc	4740
accgcacgtc	tcgcagccaa	aaaaaaaaaa	agaaagaaaa	aaaagaaaaa	gaaaaacagc	4800
agggtgggtcc	gggtcgtggg	ggccggaaaa	gcgaggagga	tcgcgagcag	cgacgaggcc	4860
cggccctccc	tcgccttcca	aagaaacgcc	ccccatcgcc	actatataca	tacccccccc	4920
tctcctccca	tccccccaac	cctaccacca	ccaccaccac	cacctcctcc	cccctcgtcg	4980
ccggacgacg	agctcctccc	ccctccccct	ccgcgcgcgc	cggtaacccac	cccgcacctc	5040
tctcttttct	ttctccggtt	tttttttcgt	ctcgggtctcg	atctttggcc	ttggtagttt	5100
gggtgggcga	gagcggcttc	gtcgcccaga	tcgggtgcgcg	ggagggggcg	gatctcgcgg	5160
ctggcgtctc	cgggcgtgag	tcggcccggg	tcctcgcggg	gaatggggct	ctcggatgta	5220
gatcttcttt	ctttcttctt	tttgtggtag	aatttgaatc	cctcagcatt	gttcacgcgt	5280
agtttttctt	ttcatgattt	gtgacaaatg	cagcctcgtg	cggagctttt	ttgtagc	5337

<210> 20

<211> 313

<212> PRT

<213> Artificial Sequence

<220>

<223> pTU4 vector

<221> VARIANT

<222> (1)...(313)

<223> Xaa = Any Amino Acid

<400> 20

```

Met Ala His Ala Arg Val Leu Leu Leu Ala Leu Ala Val Leu Ala Thr
 1          5          10          15
Ala Ala Val Ala Val Ala Ser Ser Ser Ser Phe Ala Asp Ser Asn Pro
          20          25          30
Ile Arg Pro Val Thr Asp Arg Ala Ala Ala Ser Thr Gln Gly Ile Ser
          35          40          45
Glu Asp Leu Tyr Ser Arg Leu Val Glu Met Ala Thr Ile Ser Gln Ala
 50          55          60
Ala Tyr Ala Asp Leu Cys Asn Ile Pro Ser Thr Ile Ile Lys Gly Glu
65          70          75          80
Lys Ile Tyr Asn Ser Gln Thr Asp Ile Asn Gly Trp Ile Leu Arg Asp
          85          90          95
Asp Ser Ser Lys Glu Ile Ile Thr Val Phe Arg Gly Thr Gly Ser Asp
          100          105          110
Thr Asn Leu Gln Leu Asp Thr Asn Tyr Thr Leu Thr Pro Phe Asp Thr
          115          120          125
Leu Pro Gln Cys Asn Gly Cys Glu Val His Gly Gly Tyr Tyr Ile Gly
130          135          140
Trp Val Ser Val Gln Asp Gln Val Glu Ser Leu Val Lys Gln Gln Val
145          150          155          160
Ser Gln Tyr Pro Asp Tyr Ala Leu Thr Val Thr Gly His Xaa Leu Gly
          165          170          175
Ala Ser Leu Ala Ala Leu Thr Ala Ala Gln Leu Ser Ala Thr Tyr Asp
          180          185          190
Asn Ile Arg Leu Tyr Thr Phe Gly Glu Pro Arg Ser Gly Asn Gln Ala
195          200          205
Phe Ala Ser Tyr Met Asn Asp Ala Phe Gln Ala Ser Ser Pro Asp Thr
210          215          220
Thr Gln Tyr Phe Arg Val Thr His Ala Asn Asp Gly Ile Pro Asn Leu
225          230          235          240
Pro Pro Val Glu Gln Gly Tyr Ala His Gly Gly Val Glu Tyr Trp Ser
          245          250          255
Val Asp Pro Tyr Ser Ala Gln Asn Thr Phe Val Cys Thr Gly Asp Glu
          260          265          270
Val Gln Cys Cys Glu Ala Gln Gly Gly Gln Gly Val Asn Asn Ala His
          275          280          285
Thr Thr Tyr Phe Gly Met Thr Ser Gly Ala Cys Thr Trp Pro Val Ala
290          295          300
Ala Ala Glu Pro Leu Lys Asp Glu Leu
305          310

```

<210> 21

<211> 5395

<212> DNA

<213> Artificial Sequence

<220>

<223> pTT5.14 vector

<400> 21

```

cctgacgccg aggatccatg gccacgccc gcgtcctcct cctggcgctc gccgtgctgg      60
ccacggcgcg cgtcgcgcgc gcctcctcct cctccttcgc cgactccaac ccgggcccggc      120
ccgtcaccga ccgcgcgggc gcctccacgc agggcatctc cgaagacctc tacagccggt      180
tagtcgaaat ggccactatc tcccaagctg cctacgcgca cctgtgcaac attccgtcga      240
ctattatcaa gggagagaaa atttacaatt ctcaaactga cattaacgga tggatcctcc      300
gcgacgacag cagcaaagaa ataatcacgc tcttcctgtg cactggtagt gatacgaatc      360
tacaactcga tactaactac accctcacgc ctttcgacac cctaccacaa tgcaacggtt      420
gtgaagtaca cgggtggatat tatattggat ggggtctcgt ccaggaccaa gtcgagtcgc      480
ttgtcaaaca gcaggtttagc cagtatccgg actacgcgct gaccgtgacc ggccackccc      540
tcggcgccct cctggcggca ctactgccg ccagctgtc tgcgacatac gacaacatcc      600
gcctgtacac cttcggcgaa ccgcgcagcg gcaatcaggc cttcgcgctc tacatgaacg      660

```

atgccttcca	agcctcgagc	ccagatacga	cgcagtat	ccgggtcact	catgccaacg	720
acggcatccc	aaacctgccc	ccgggtggagc	aggggtacgc	ccatggcggt	gtagagtact	780
ggagcgttga	tccttacagc	gcccagaaca	catttgtctg	cactggggat	gaagtgcagt	840
gctgtgaggc	ccagggcgga	caggggtgtga	ataatgcgca	cacgacttat	tttgggatga	900
cgagcggagc	ctgtacatgg	tgatcagtc	tttcagcctc	cccagagtga	ccaggaaaga	960
tggatgtcct	ggagaggggg	ccgcgtaacc	actgaaggat	gagctgtaaa	gaagcagatc	1020
gttcaaaca	ttggcaataa	agtttcttaa	gattgaatcc	tgttgccggg	cttgcgatga	1080
ttatcatata	atttctgttg	aattacgtta	agcatgta	aattaacatg	taatgcatga	1140
cgttat	gagatgggtt	tttatgatta	gagtcgccga	attatacatt	taatacgcga	1200
tagaaaacaa	aatatagcgc	gcaaactagg	ataaattatc	gcgcgcgggtg	tcattctatgt	1260
tactagatcg	ataagcttct	agagcggccg	gtggagctcc	aattcgccct	atagtgaatc	1320
gtattacgcg	cgtcactgg	ccgtcgtttt	acaacgtcgt	gactgggaaa	acctggcgt	1380
tacccaactt	aatcgccttg	cagcacatcc	ccctttcgcc	agctggcgta	atagcgaaga	1440
ggcccgcacc	gacgcctt	cccaacagtt	gcgcagcctg	aatggcgaat	gggacgcgcc	1500
ctgtagcggc	gcattaagcg	cggcgggtgt	ggtgggttacg	cgcagcgtga	ccgctacact	1560
tgccagcgc	ctagcgcgcg	ctcctttcgc	tttcttccct	tcctttctcg	ccacgttcgc	1620
cggctttccc	cgtcaagctc	taaatcgggg	gctcccttta	gggttccgat	ttagtgcctt	1680
acggcacctc	gaccccaaaa	aacttgatta	gggtgatggg	tcacgtagtg	ggccatcgcc	1740
ctgatagacg	gtttttcgc	ctttgacgtt	ggagtccacg	ttctttaata	gtggactctt	1800
gttccaaact	ggaacaacac	tcaaccctat	ctcggctctat	tccttttgatt	tataagggat	1860
tttgccgatt	tcggcctatt	ggttaaaaaa	tgagctgatt	taacaaaaat	ttaacgcgaa	1920
ttttaacaaa	atattaacgc	ttacaattta	ggtggcactt	ttcggggaaa	tgtgcgcgga	1980
acctctat	gtttat	ctaaatacat	tcaaataatgt	atccgctcat	gagacaataa	2040
ccttgataaa	tgcttcaata	atattgaaaa	aggaagagta	tgagtattca	acatttccgt	2100
gtcgccttta	ttcccttttt	tgccgcattt	tgcttccctg	tttttgctca	cccagaaacg	2160
ctggtgaaag	taaaagatgc	tgaagatcag	ttgggtgcac	gagtgggtta	catcgaactg	2220
gatctcaaca	gcggtaagat	ccttgagagt	tttcgccccg	aagaacgttt	tccaatgatg	2280
agcactttta	aagttctgct	atgtggcgcg	gtattatccc	gtattgacgc	cgggcaagag	2340
caactcggtc	gccgcataca	ctattctcag	aatgacttgg	ttgagtactc	accagtcaca	2400
gaaaagcatc	ttacggatgg	catgacagta	agagaattat	gcagtgcctgc	cataaccatg	2460
agtgataaca	ctgcggccaa	cttacttctg	acaacgatcg	gaggaccgaa	ggagctaacc	2520
gcttttttgc	acaacatggg	ggatcatgta	actcgccttg	atcgttggga	accggagctg	2580
aatgaagcca	taccaaacga	cgcgcgtgac	accacgatgc	ctgtagcaat	ggcaacaacg	2640
ttgcgcaaac	tattaactgg	cgaactactt	actctagctt	cccggcaaca	attaatagac	2700
tggatggagg	cggataaagt	tgcaggacca	cttctgcgct	cggcccttcc	ggctggctgg	2760
tttattgctg	ataaatctgg	agccggtgag	cgtgggtctc	gcggtatcat	tgacgactg	2820
gggccagatg	gtaagccctc	ccgtatcgta	gttatctaca	cgacggggag	tcaggcaact	2880
atggatgaac	gaaatagaca	gatcgtgag	ataggtgcct	cactgattaa	gcattggtaa	2940
ctgtcagacc	aagtttactc	atatatactt	tagattgatt	taaaacttca	tttttaattt	3000
aaaaggatct	aggtgaagat	cctttttgat	aatctcatga	ccaaaatccc	ttaacgtgag	3060
ttttcgttcc	actgagcgtc	agaccccgta	gaaaagatca	aaggatcttc	ttgagatcct	3120
ttttttctgc	gcgtaatctg	ctgcttgcaa	acaaaaaaac	caccgctacc	agcgggtggtt	3180
tgtttgccgg	atcaagagct	accaactctt	tttccgaagg	taactggctt	cagcagagcg	3240
cagataccaa	atactgtcct	tctagtgtag	ccgtagttag	gccaccactt	caagaactct	3300
gtagcaccgc	ctacatacct	cgtctgcta	atcctgttac	cagtggctgc	tgccagtggc	3360
gataagtcgt	gtcttaccgg	gttggactca	agacgatagt	taccggataa	ggcgcagcgg	3420
tcgggctgaa	cgggggggtt	gtgcacacag	cccagcttgg	agcgaacgac	ctacaccgaa	3480
ctgagatacc	tacagcgtga	gctatgagaa	agcgcacgcg	ttcccgaagg	gagaaaggcg	3540
gacaggtatc	cggtaagcgg	cagggtcgga	acaggagagc	gcacgaggga	gcttccaggg	3600
ggaaacgcct	ggtatcttta	tagtcctgtc	gggtttcgc	acctctgact	tgagcgtcga	3660
tttttgat	gctcgtcagg	ggggcggagc	ctatggaaaa	acgccagcaa	cgcggccttt	3720
ttacggttcc	tggccttttg	ctggcctttt	gtcacatgt	tccttctctgc	gttatccctt	3780
gattctgtgg	ataaccgtat	taccgccttt	gagtgaactg	ataccgctcg	ccgcagccga	3840
acgaccgagc	gcagcagatc	agtgaagcag	gaagcgggaag	agcgcaccaat	acgcaaaccg	3900
cctctccccg	cgcgttggcc	gattcattaa	tgcagctggc	acgacagggt	tcccgaactg	3960
aaagcgggca	gtgagcgcga	cgcaattaat	gtgagttagc	tactcatta	ggcaccaccag	4020
gctttacact	ttatgcttcc	ggctcgtatg	ttgtgtggaa	ttgtgagcgg	ataacaattt	4080
cacacaggaa	acagctatga	ccatgattac	gccaaagcgcg	caattaaccc	tactaaagg	4140
gaacaaaagc	tgggtaccgg	gccccccctc	gaggtcattc	atatgcttga	gaagagagtc	4200
gggatagtcc	aaaataaaac	aaaggtaaga	ttacctggtc	aaaagtga	acatcagtta	4260
aaagggtgga	taagtaaaat	atcggtaata	aaagggtggc	caaagtga	tttactcttt	4320

```

tctactatta taaaaattga ggatgttttg tcggtacttt gatacgtcat ttttgtatga 4380
attgggttttt aagttttattc gcgatttgga aatgcatatc tgtatttgag tcggtttttta 4440
agttcgttgc ttttgtaaat acagagggat ttgtataaga aatatcttta aaaaacccat 4500
atgctaattt gacataattt ttgagaaaaa tatatatcca ggcgaattcc acaatgaaca 4560
ataataagat taaaatagct tgcccccggt gcagcgatgg gtattttttc tagtaaaata 4620
aaagataaac ttagactcaa aacatttaca aaaacaaccc ctaaagtcct aaagcccaaa 4680
gtgctatgca cgatccatag caagcccagc ccaacccaac ccaacccaac ccaccccagt 4740
gcagccaact ggcaaatagt ctccaccccc ggcaactatca ccgtgagttg tccgcaccac 4800
cgcacgtctc gcagccaaaa aaaaaaaaaa aaagaaaaaa aagaaaaaga aaaacagcag 4860
gtgggtccgg gtcggtggggg ccggaaaagc gaggaggatc gcgagcagcg acgaggcccg 4920
gccctccctc cgtttccaaa gaaacgcccc ccatcgccac tatatacata cccccccctc 4980
tcctcccatc cccccaaccc taccaccacc accaccacca cctcctcccc cctcgctgcc 5040
ggacgacgag ctccctcccc ctccccctcc gccgcgcgcg gtaaccaccc cgcccctctc 5100
ctctttcttt ctccgttttt tttttcgtct cgggtctcgat ctttggcctt ggtagtttgg 5160
gtgggcgaga gcggcttcgt cggccagatc ggtgcgcggg aggggcggga tctcgcggtc 5220
ggcgtctccg ggcgtgagtc ggcccggatc ctgcgcggga atggggctct cggatgtaga 5280
tcttctttct ttcttctttt tgtggtagaa tttgaatccc tcagcattgt tcatcggtag 5340
tttttctttt catgatttgt gacaaatgca gcctcgtgcg gagctttttt gtagc 5395

```

<210> 22

<211> 301

<212> PRT

<213> Artificial Sequence

<220>

<223> pTT5.14 vector

<221> VARIANT

<222> (1)...(301)

<223> Xaa = Any Amino Acid

<400> 22

```

Met Ala His Ala Arg Val Leu Leu Leu Ala Leu Ala Val Leu Ala Thr
 1           5           10           15
Ala Ala Val Ala Val Ala Ser Ser Ser Ser Phe Ala Asp Ser Asn Pro
           20           25           30
Gly Arg Pro Val Thr Asp Arg Ala Ala Ala Ser Thr Gln Gly Ile Ser
           35           40           45
Glu Asp Leu Tyr Ser Arg Leu Val Glu Met Ala Thr Ile Ser Gln Ala
           50           55           60
Ala Tyr Ala Asp Leu Cys Asn Ile Pro Ser Thr Ile Ile Lys Gly Glu
65           70           75           80
Lys Ile Tyr Asn Ser Gln Thr Asp Ile Asn Gly Trp Ile Leu Arg Asp
           85           90           95
Asp Ser Ser Lys Glu Ile Ile Thr Val Phe Arg Gly Thr Gly Ser Asp
           100          105          110
Thr Asn Leu Gln Leu Asp Thr Asn Tyr Thr Leu Thr Pro Phe Asp Thr
           115          120          125
Leu Pro Gln Cys Asn Gly Cys Glu Val His Gly Gly Tyr Tyr Ile Gly
           130          135          140
Trp Val Ser Val Gln Asp Gln Val Glu Ser Leu Val Lys Gln Gln Val
145          150          155          160
Ser Gln Tyr Pro Asp Tyr Ala Leu Thr Val Thr Gly His Xaa Leu Gly
           165          170          175
Ala Ser Leu Ala Ala Leu Thr Ala Ala Gln Leu Ser Ala Thr Tyr Asp
           180          185          190
Asn Ile Arg Leu Tyr Thr Phe Gly Glu Pro Arg Ser Gly Asn Gln Ala
           195          200          205
Phe Ala Ser Tyr Met Asn Asp Ala Phe Gln Ala Ser Ser Pro Asp Thr
           210          215          220
Thr Gln Tyr Phe Arg Val Thr His Ala Asn Asp Gly Ile Pro Asn Leu

```

225					230					235					240
Pro	Pro	Val	Glu	Gln	Gly	Tyr	Ala	His	Gly	Gly	Val	Glu	Tyr	Trp	Ser
				245					250					255	
Val	Asp	Pro	Tyr	Ser	Ala	Gln	Asn	Thr	Phe	Val	Cys	Thr	Gly	Asp	Glu
			260					265					270		
Val	Gln	Cys	Cys	Glu	Ala	Gln	Gly	Gly	Gln	Gly	Val	Asn	Asn	Ala	His
		275					280					285			
Thr	Thr	Tyr	Phe	Gly	Met	Thr	Ser	Gly	Ala	Cys	Thr	Trp			
	290					295					300				

<210> 23

<211> 5337

<212> DNA

<213> Artificial Sequence

<220>

<223> pTP8-5 vector

<400> 23

aagcttacca	tggcccacgc	ccgcgtcctc	ctcctggcgc	tgcgcgtgct	ggccacggcc	60
gcgctgcgcg	tgcctcctc	ctcctccttc	gcgactcca	acccgatccg	gcccgtcacc	120
gaccgcgcgg	ccgcctccac	gcagggcatc	tccgaagacc	tctacagccg	tttagtcgaa	180
atggccacta	tctcccaagc	tgcctacgcc	gacctgtgca	acattccgtc	gactattatc	240
aagggagaga	aaattttacaa	ttctcaaact	gacattaacg	gatggatcct	ccgcgacgac	300
agcagcaaag	aaataatcac	cgtcttccgt	ggcactggta	gtgatacgaa	tctacaactc	360
gatactaact	acaccctcac	gcctttcgac	accctaccac	aatgcaacgg	ttgtgaagta	420
cacggtggat	attatatagg	atgggtctcc	gtccaggacc	aagtcgagtc	gcttgtcaaa	480
cagcaggtta	gccagtatcc	ggactacgcg	ctgaccgtga	ccggccackc	cctcggcgcc	540
tcctggcgg	cactcactgc	cgcccagctg	tctgcgacat	acgacaacat	ccgcctgtac	600
accttcggcg	aaccgcgcag	cggcaatcag	gccttcgcgt	cgtacatgaa	cgatgccttc	660
caagcctcga	gcccagatac	gacgcagtat	ttccgggtca	ctcatgccaa	cgacggcatc	720
ccaaacctgc	ccccggtgga	gcaggggtac	gcccattggc	gtgtagagta	ctggagcggt	780
gateccttaca	gcgcccagaa	cacatttgct	tgcactgggg	atgaagtgca	gtgctgtgag	840
gcccagggcg	gacaggggtg	gaataatgcg	cacacgactt	atthttggg	gacgagcggc	900
gcattgcacct	ggccggtcgc	ggccgcgtaa	ccactgaagg	atgagctgta	aagaagcaga	960
tcgttcaaac	atthggcaat	aaagtthctt	aagattgaat	cctgthggcg	gtcttgcgat	1020
gattatcata	taattthctg	tgaattacgt	taagcatgta	ataattaaca	tgtaatgcat	1080
gacgttatth	atgagatggg	thtttatgat	tagagthccg	caattataca	thtaatacgc	1140
gatagaaaac	aaaatatagc	gcgcaaacta	ggataaatta	tcgcgcgcgg	tgtcatctat	1200
gttactagat	cgataagctt	ctagagcggc	cggthggagc	ccaattcgcc	ctatagtgag	1260
tcgtattacg	cgcgctcact	ggccgtcgtt	ttacaacgtc	gtgactggga	aaaccctggc	1320
gttaccacaac	thaatcgcct	tgcagcacat	ccccctthcg	ccagctggcg	taatagcgaa	1380
gaggcccgcg	ccgatcgccc	thcccaacag	thgcgcagcc	tgaatggcga	atgggacgcg	1440
ccctgtagcg	gcgcattaa	cgcggcggg	gtgggtggth	cgcgcagcgt	gaccgctaca	1500
cttgccagcg	ccctagcgcc	cgtctctthc	gctthctthc	cttctthctt	cgccacgttc	1560
gcccggctth	cccgtcaagc	tctaaatcgg	gggtctccct	tagggthccg	atthagtgtc	1620
thacggcacc	tcgaccccaa	aaaacttgat	tagggthgat	gtthacgtag	thggccatcg	1680
ccctgataga	cggthththc	ccctthgacg	thggagthca	cgtthctthaa	tagthggactc	1740
thgtthcaaa	ctggaacaac	actcaaccct	atctcggtct	atthctththga	thtataaggg	1800
atththggcg	thtcggccta	thggthtaaa	aatgagctga	thtaacaaaa	atthaacgcg	1860
aatththaca	aatatthaac	gctthacaatt	taggtggcac	ththtcgggga	aatgtgcgcg	1920
gaacccttat	thgtththth	thctaaatac	atthcaatat	gtatccgctc	atgagacaat	1980
aaccctgata	aatgctthcaa	taatattgaa	aaaggaagag	tatgagtatt	caacattthc	2040
gtgtcgccct	tattccctth	thtgccggcat	thtgccctthc	tgththththg	caccacagaaa	2100
cgtgtgtgaa	agtaaaagat	gctgaagatc	agthgggtgc	acgagthggg	tacatcgaa	2160
tggtatctcaa	cagcggtaag	atcctthgaga	gtththcgccc	cgaagaacgt	thtccaatga	2220
tgagcactth	taagththctg	ctatgtggcg	cggthattatc	ccgtattgac	gcccgggcaag	2280
agcaactcgg	tcgcccata	cactattctc	agaatgactt	ggtthgagta	tcaccagthca	2340
cagaaaagca	thttacggat	ggcatgacag	taagagaatt	atgcagthgct	gccataacca	2400
tgagthgata	cactgcgggc	aactthcttc	tgacaacgat	cggaggaccg	aaggagctaa	2460
ccgctththth	gcacaacatg	ggggatcatg	taactcgcct	tgatcgthtg	gaaccggagc	2520

tgaatgaagc	cataccaaac	gacgagcgtg	acaccacgat	gcctgtagca	atggcaacaa	2580
cgttgcgcaa	actattaact	ggcgaactac	ttactctagc	ttcccggcaa	caattaatag	2640
actggatgga	ggcggataaa	gttgcaggac	cacttctgcg	ctcggccctt	ccggctggct	2700
ggtttattgc	tgataaatct	ggagccggtg	agcgtgggtc	tcgcggtatc	attgcagcac	2760
tggggccaga	tggttaagccc	tcccgtatcg	tagttatcta	cacgacgggg	agtcaggcaa	2820
ctatggatga	acgaaataga	cagatcgctg	agatagggtc	ctcactgatt	aagcattggt	2880
aactgtcaga	ccaagtttac	tcatatatac	tttagattga	tttaaaactt	cattttttaat	2940
ttaaaaggat	ctaggtgaag	atcctttttg	ataatctcat	gaccaaatac	ccttaacgtg	3000
agttttcgtt	ccactgagcg	tcagaccccc	tagaaaagat	caaaggatct	tcttgagatc	3060
ctttttttct	gcgcgtaatc	tgctgcttgc	aaacaaaaaa	accaccgcta	ccagcgggtg	3120
tttgtttgcc	ggatcaagag	ctaccaactc	tttttccgaa	ggtaactggc	ttcagcagag	3180
cgcagatacc	aaatactgtc	cttctagtgt	agccgtagt	aggccaccac	ttcaagaact	3240
ctgtagcacc	gcctacatac	ctcgctctgc	taatcctgtt	accagtggct	gctgccagt	3300
gcgataagtc	gtgtcttacc	gggttggact	caagacgata	gttaccggat	aaggcgcagc	3360
ggtcgggctg	aacggggggg	tcgtgcacac	agcccagctt	ggagcgaacg	acctacaccg	3420
aactgagata	cctacagcgt	gagctatgag	aaagcgccac	gcttcccga	gggagaaagg	3480
cggacaggta	tccggtaagc	ggcagggctc	gaacaggaga	gcgcacgagg	gagcttccag	3540
ggggaaacgc	ctgggtatctt	tatagtcctg	tcgggtttcg	ccacctctga	cttgagcgtc	3600
gattttttgtg	atgctcgtca	ggggggcgga	gcctatggaa	aaacgccagc	aacgcggcct	3660
ttttacgggtt	cctggccttt	tgctggcctt	ttgctcacat	gttctttcct	gcgttatccc	3720
ctgattctgt	ggataaccgt	attaccgcct	ttgagtgagc	tgataccgct	cgccgcagcc	3780
gaacgaccga	gcgcagcgag	tcagtgagcg	aggaagcgga	agagcgccca	atacgcaaac	3840
cgcctctccc	cgcgcgttgg	ccgattcatt	aatgcagctg	gcacgacagg	tttcccgaact	3900
ggaaagcggg	cagtgagcgc	aacgcgaatta	atgtgagtta	gtcactcat	taggcacccc	3960
aggctttaca	ctttatgctt	ccggctcgta	tggtgtgtgg	aattgtgagc	ggataacaat	4020
ttcacacagg	aaacagctat	gaccatgatt	acgccaaagc	cgcaattaac	cctcactaaa	4080
gggaacaaaa	gctgggtacc	gggccccccc	tcgaggctcat	tcatatgctt	gagaagagag	4140
tcgggatagt	ccaaaataaa	acaaaggtaa	gattacctgg	tcaaaagtga	aaacatcagt	4200
taaaagggtg	tataagtaaa	atatcggtaa	taaaagggtg	cccaaagtga	aatttactct	4260
tttctactat	tataaaaatt	gaggatgttt	tgctcggtact	ttgatacgtc	atttttgtat	4320
gaattggttt	ttaagtttat	tcgcgatttg	gaaatgcata	tctgtatttg	agtcgggtttt	4380
taagttcggt	gctttttgtaa	atacagaggg	atttgtataa	gaaatatctt	taaaaaaccc	4440
atatgcta	ttgacataat	ttttgagaaa	aatatatatt	caggcgaatt	ccacaatgaa	4500
caataataag	attaaaatag	cttgcccccg	ttgcagcgat	gggtattttt	tctagtaaaa	4560
taaaagataa	acttagactc	aaaacattta	caaaaacaac	ccctaaagtc	ctaaagccca	4620
aagtgc	cacgatccat	agcaagccca	gcccacccca	acccaaccca	acccacccca	4680
gtgcagccaa	ctggcaaata	gtctccaccc	ccggcactat	caccgtgagt	tgtccgcacc	4740
accgcacgtc	tcgcagccaa	aaaaaaaaaa	agaaagaaaa	aaaagaaaaa	gaaaaacagc	4800
agggtgggtcc	gggtcgtggg	ggccggaaaa	gcgaggagga	tcgcgagcag	cgacgaggcc	4860
cggccctccc	tccgcttcca	aagaaacgcc	ccccatcgcc	actatataca	tccccccccc	4920
tctcctccc	tccccccaac	cctaccacca	ccaccaccac	cacctcctcc	cccctcgctg	4980
ccggacgacg	agctcctccc	ccctccccct	ccgcgcgcgc	cggtaaccac	cccgcacctc	5040
tctctttct	ttctccgttt	tttttttcgt	ctcggtctcg	atctttggcc	ttggtagttt	5100
gggtgggcga	gagcggcttc	gtcgcccaga	tcgggtgcgcg	ggagggggcg	gatctcgcg	5160
ctggcgtctc	cgggcgtgag	tcggccccga	tcctcgcggg	gaatggggct	ctcggtatga	5220
gatcttcttt	ctttcttctt	tttgtggtag	aatttgaatc	cctcagcatt	gttcacgggt	5280
agtttttctt	ttcatgattt	gtgacaaatg	cagcctcgtg	cggagctttt	ttgtagc	5337

<210> 24

<211> 306

<212> PRT

<213> Artificial Sequence

<220>

<223> pTP8-5 vector

<221> VARIANT

<222> (1)...(306)

<223> Xaa = Any Amino Acid

<400> 24

```

Met Ala His Ala Arg Val Leu Leu Leu Ala Leu Ala Val Leu Ala Thr
 1          5          10          15
Ala Ala Val Ala Val Ala Ser Ser Ser Ser Phe Ala Asp Ser Asn Pro
          20          25          30
Ile Arg Pro Val Thr Asp Arg Ala Ala Ala Ser Thr Gln Gly Ile Ser
          35          40          45
Glu Asp Leu Tyr Ser Arg Leu Val Glu Met Ala Thr Ile Ser Gln Ala
 50          55          60
Ala Tyr Ala Asp Leu Cys Asn Ile Pro Ser Thr Ile Ile Lys Gly Glu
65          70          75          80
Lys Ile Tyr Asn Ser Gln Thr Asp Ile Asn Gly Trp Ile Leu Arg Asp
          85          90          95
Asp Ser Ser Lys Glu Ile Ile Thr Val Phe Arg Gly Thr Gly Ser Asp
          100          105          110
Thr Asn Leu Gln Leu Asp Thr Asn Tyr Thr Leu Thr Pro Phe Asp Thr
          115          120          125
Leu Pro Gln Cys Asn Gly Cys Glu Val His Gly Gly Tyr Tyr Ile Gly
130          135          140
Trp Val Ser Val Gln Asp Gln Val Glu Ser Leu Val Lys Gln Gln Val
145          150          155          160
Ser Gln Tyr Pro Asp Tyr Ala Leu Thr Val Thr Gly His Xaa Leu Gly
          165          170          175
Ala Ser Leu Ala Ala Leu Thr Ala Ala Gln Leu Ser Ala Thr Tyr Asp
          180          185          190
Asn Ile Arg Leu Tyr Thr Phe Gly Glu Pro Arg Ser Gly Asn Gln Ala
195          200          205
Phe Ala Ser Tyr Met Asn Asp Ala Phe Gln Ala Ser Ser Pro Asp Thr
210          215          220
Thr Gln Tyr Phe Arg Val Thr His Ala Asn Asp Gly Ile Pro Asn Leu
225          230          235          240
Pro Pro Val Glu Gln Gly Tyr Ala His Gly Gly Val Glu Tyr Trp Ser
          245          250          255
Val Asp Pro Tyr Ser Ala Gln Asn Thr Phe Val Cys Thr Gly Asp Glu
          260          265          270
Val Gln Cys Cys Glu Ala Gln Gly Gly Gln Gly Val Asn Asn Ala His
275          280          285
Thr Thr Tyr Phe Gly Met Thr Ser Gly Ala Cys Thr Trp Pro Val Ala
290          295          300
Ala Ala
305

```

<210> 25

<211> 5277

<212> DNA

<213> Artificial Sequence

<220>

<223> pTP5-1 vector

<400> 25

```

aagcttaaca tgaagcagtt ctccgccaaa cacgtcctcg cagttgtggt gactgcaggg      60
cacgccttag cagcctctac gcaaggcatc tccgaagacc tctacagccg ttagtcgaa      120
atggccacta tctccaagc tgcctacgcc gacctgtgca acattccgtc gactattatc      180
aagggagaga aaatttacaa ttctcaaact gacattaacg gatggatcct ccgcgacgac      240
agcagcaaag aaataatcac cgtcttccgt ggcaactggt gtgatacgaa tctacaactc      300
gatactaact acaccctcac gcctttcgac accctaccac aatgcaacgg ttgtgaagta      360
cacggtggat attatattgg atgggtctcc gtccaggacc aagtcgagtc gcttgtcaaa      420
cagcaggtta gccagtatcc ggactacgcg ctgaccgtga ccggccackc cctcggcgcc      480
tccctggcgg cactcactgc cgcccagctg tctgcgacat acgacaacat ccgcctgtac      540
accttcggcg aaccgcgcag cggcaatcag gccttcgcgt cgtacatgaa cgatgccttc      600
caagcctcga gccagatac gacgcagtat ttccgggtca ctcatgccaa cgacggcatc      660

```

ccaaacctgc	ccccggtgga	gcaggggtac	gcccattggcg	gtgtagagta	ctggagcgtt	720
gattccttaca	gcgcccagaa	cacatttgtc	tgcactgggg	atgaagtgca	gtgctgtgag	780
gcccagggcg	gacaggggtgt	gaataatgcg	cacacgactt	atthttgggat	gacgagcggc	840
gcatgcacct	ggccggtcgc	ggccgcggaa	ccactgaagg	atgagctgta	aagaagcaga	900
tcgttcaaac	atthtggaat	aaagtttctt	aagattgaat	cctgttgccg	gtcttgcgat	960
gattatcata	taatttctgt	tgaattacgt	taagcatgta	ataattaaca	tgtaatgcat	1020
gacgttatth	atgagatggg	thtttatgat	tagagtcccg	caattataca	thtaatacgc	1080
gatagaaaac	aaaatatagc	gcgcaaacta	ggataaatta	tcgcgcgcgg	tgtcatctat	1140
gttactagat	cgataagctt	ctagagcggc	cgggtggagct	ccaattcgcc	ctatagttag	1200
tcgtattacg	cgcgctcact	ggccgctcgt	ttacaacgtc	gtgactggga	aaaccctggc	1260
gttaccaca	thaatcgctt	tgcagcacat	cccccttctg	ccagctggcg	taatagcgaa	1320
gaggcccgca	ccgatcgccc	thcccaacag	ttgcgcagcc	tgaatggcga	atgggacgcg	1380
ccctgtagcg	gcgcattaa	cgcggcgggt	gtggtggtta	cgcgcagcgt	gaccgctaca	1440
cttgccagcg	ccctagcgcc	cgcctccttt	gctthctctc	cttcccttct	cgcacgcttc	1500
gcccgtctth	cccgtcaagc	tctaaatcgg	gggtctccct	taggggtccg	atthagtgtc	1560
thacggcacc	tcgaccccaa	aaaacttgat	taggggtgat	gttcacgtag	tgggcatcgc	1620
ccctgataga	cggthttctg	ccctthgacg	thggagtcca	cgttctthaa	tagtggactc	1680
thgtthccaa	ctggaacaac	actcaaccct	atctcgggtc	atthctthtg	thtataaggg	1740
atthttgcga	thttcgccct	thggthtaaa	aatgagctga	thtaacaaaa	atthaacgcg	1800
aatthtaaca	aaatattaac	gctthacaatt	taggtggcac	thttcgggga	aatgtgcgcg	1860
gaacccttat	thgtthtatth	thctaaatac	atthcaaatat	gtatccgctc	atgagacaat	1920
aaacctgata	aatgctthca	taatatthga	aaaggaagag	tatgagtatt	caacatthtc	1980
gtgtcgccct	tattccctth	thttgcggcat	thttgccttc	tgtthtttgc	caccagaaaa	2040
cgtcgttgaa	agtaaaagat	gctgaagatc	agttgggtgc	acgagtgggt	tacatcgaa	2100
tggatctcaa	cagcggtaag	atccttgaga	gtthtcgccc	cgaagaacgt	thtccaatga	2160
tgagcactth	taaggttctg	ctatgtggcg	cggatattat	ccgtattgac	gcccggcag	2220
agcaactcgg	tcgcccata	cactattctc	agaatgactt	ggttgagtac	tcaccagtca	2280
cagaaaagca	tcttacggat	ggcatgacag	taagagaatt	atgcagtgc	gccataacca	2340
tgagtataaa	cactgcggcc	aaactacttc	tgacaacgat	cggaggaccg	aaggagctaa	2400
ccgctthttth	gcacaacatg	ggggatcatg	taactcgcct	tgatcgttgg	gaaccggagc	2460
tgaatgaagc	cataccaaa	gacgagcgtg	acaccacgat	gcctgtagca	atggcaacaa	2520
cgttgccgca	actattaa	ggcgaactac	thactctagc	thcccggcaa	caattaatag	2580
actggatgga	ggcggataaa	gttgccaggac	cacttctgcg	ctcggccctt	ccggctggct	2640
ggthttattgc	tgataaatct	ggagccgggt	agcgtgggtc	tcgcccgtat	atthcagcac	2700
tggggccaga	tggtaagccc	thcccgtatc	tagthtatct	cacgacgggg	agtcaggcaa	2760
ctatggatga	acgaaataga	cagatcgctg	agataggtgc	ctcactgatt	aagcattgg	2820
aaactgtcaga	ccaagthttac	thcatatata	thtagattga	thtaaaaact	cattthttat	2880
thtaaaaggat	ctaggtgaag	atcctthttg	ataatctcat	gaccaaatac	ccttaacgtg	2940
agthtttcgt	ccactgagcg	tcagaccccg	tagaaaagat	caaaggatct	tcttgagatc	3000
ctthttthct	gcgcgtaatc	tgcgtgcttg	aaacaaaaaa	accaccgcta	ccagcgggtg	3060
thtgthttgc	ggatcaagag	ctaccaactc	thttthccga	ggtaactggc	thcagcagag	3120
cgcagatacc	aaatactgtc	cttctagtgt	agccgtagt	aggccaccac	thcaagaact	3180
ctgtagcacc	gcctacatac	ctcgcctctg	taatcctgt	accagtggct	gctgccagt	3240
gcgataagtc	gtgtcttacc	gggttgga	caagacgata	gttaccggat	aaggcgcagc	3300
ggtcgggctg	aacgggggg	tcgtgcacac	agcccagctt	ggagcgaacg	acctacaccg	3360
aaactgagata	cctacagcgt	gagctatgag	aaagcgcac	gctthcccga	gggagaaagg	3420
cggacaggta	thccggtaa	ggcagggtcg	gaacaggaga	gcgcacgagg	gagctthccag	3480
ggggaaacgc	ctggtatctt	tatagtcctg	tcgggtthtc	ccacctctga	cttgagcgtc	3540
gattthttgt	atgctcgtca	ggggggcgga	gcctatggaa	aaacgcgcag	aacgcggcct	3600
thttacgggt	cctggcctth	tgcgtggcct	thgtctacat	gttctthctt	gcgttatccc	3660
ctgattctgt	ggataaccgt	attaccgcct	thgagttagc	tgataccgct	cgcgcagcc	3720
gaacgaccga	gcgcagcag	tcagtgagcg	aggaagcgg	agagcgccca	atacgcaaac	3780
cgcctctccc	cgcgcgttgg	cagattcatt	aatgcagctg	gcacgacagg	thtcccga	3840
ggaaagcggg	cagttagcgc	aacgcata	atgtgagtta	gctcactcat	taggcacccc	3900
aggctthtaca	ctthtatgct	cggctcgt	tgttggtg	aattgtgagc	ggataaca	3960
thcacacagg	aaacagctat	gacctagatt	acgccaagcg	cgaattaac	cctcactaaa	4020
gggaacaaaa	gctgggtacc	gggccccccc	tcgaggtcat	thcatatgct	gagaagagag	4080
tcgggatagt	ccaaaataaa	acaaaggtaa	gattacctgg	thaaaagtga	aaacatcagt	4140
taaaagggtg	tataagtaaa	atatcggtaa	taaaagggtg	cccaaagtga	aatttactct	4200
thtctactat	tataaaaatt	gaggatgtth	tgtcgggtact	thgatacgtc	atthttgtat	4260
gaattgggtth	thtaagttht	tcgcgatttg	gaaatgcata	tctgtatttg	agtcgggtth	4320

```

taagtctcgtt gcttttgttaa atacagaggg atttgtataa gaaatatctt taaaaaaccc 4380
atatgctaata ttgacataat ttttgagaaa aatatatatt caggcgaatt ccacaatgaa 4440
caataataag attaaaatag cttgcccccg ttgcagcgat gggatatttt tctagtaaaa 4500
taaaagataa acttagactc aaaacattta caaaaacaac ccctaaagtc ctaaagccca 4560
aagtgcctatg cacgatccat agcaagccca gcccaaccca acccaaccca acccacccca 4620
gtgcagccaa ctggcaaata gtctccaccc ccggcactat caccgtgagt tgtccgcacc 4680
accgcacgtc tcgcagccaa aaaaaaaaaa agaaagaaaa aaaagaaaaa gaaaaacagc 4740
aggtgggtcc gggtcgtggg ggccggaaaa gcgaggagga tcgcgagcag cgacgaggcc 4800
cggccctccc tccgcttcca aagaaacgcc ccccatcgcc actatataca tcccccccc 4860
tctcctccca tcccccaac cctaccacca ccaccaccac cacctcctcc cccctcgtg 4920
ccggacgacg agctcctccc cctccccct ccgccgccgc cggtaaccac cccgcccctc 4980
tctcttttct ttctcgttt ttttttctgt ctcggtctcg atctttggcc ttggtagttt 5040
gggtgggcga gagcggcttc gtcgccaga tcggtgcgcg ggaggggagg gatctcgcgg 5100
ctggcgtctc cgggcgtgag tcggcccgga tctcgcggg gaatggggct ctcgcatgta 5160
gatcttcttt ctttcttctt tttgtgtag aatttgaatc cctcagcatt gttcatcggt 5220
agtttttctt ttcattgatt gtgacaaatg cagcctcgtg cggagctttt ttgtagc 5277

```

<210> 26

<211> 293

<212> PRT

<213> Artificial Sequence

<220>

<223> pTP5-1 vector

<221> VARIANT

<222> (1)...(293)

<223> Xaa = Any Amino Acid

<400> 26

```

Met Lys Gln Phe Ser Ala Lys His Val Leu Ala Val Val Val Thr Ala
 1          5          10          15
Gly His Ala Leu Ala Ala Ser Thr Gln Gly Ile Ser Glu Asp Leu Tyr
          20          25          30
Ser Arg Leu Val Glu Met Ala Thr Ile Ser Gln Ala Ala Tyr Ala Asp
          35          40          45
Leu Cys Asn Ile Pro Ser Thr Ile Ile Lys Gly Glu Lys Ile Tyr Asn
          50          55          60
Ser Gln Thr Asp Ile Asn Gly Trp Ile Leu Arg Asp Asp Ser Ser Lys
          65          70          75          80
Glu Ile Ile Thr Val Phe Arg Gly Thr Gly Ser Asp Thr Asn Leu Gln
          85          90          95
Leu Asp Thr Asn Tyr Thr Leu Thr Pro Phe Asp Thr Leu Pro Gln Cys
          100          105          110
Asn Gly Cys Glu Val His Gly Gly Tyr Tyr Ile Gly Trp Val Ser Val
          115          120          125
Gln Asp Gln Val Glu Ser Leu Val Lys Gln Gln Val Ser Gln Tyr Pro
          130          135          140
Asp Tyr Ala Leu Thr Val Thr Gly His Xaa Leu Gly Ala Ser Leu Ala
          145          150          155          160
Ala Leu Thr Ala Ala Gln Leu Ser Ala Thr Tyr Asp Asn Ile Arg Leu
          165          170          175
Tyr Thr Phe Gly Glu Pro Arg Ser Gly Asn Gln Ala Phe Ala Ser Tyr
          180          185          190
Met Asn Asp Ala Phe Gln Ala Ser Ser Pro Asp Thr Thr Gln Tyr Phe
          195          200          205
Arg Val Thr His Ala Asn Asp Gly Ile Pro Asn Leu Pro Pro Val Glu
          210          215          220
Gln Gly Tyr Ala His Gly Gly Val Glu Tyr Trp Ser Val Asp Pro Tyr
          225          230          235          240
Ser Ala Gln Asn Thr Phe Val Cys Thr Gly Asp Glu Val Gln Cys Cys

```

				245						250					255	
Glu	Ala	Gln	Gly	Gly	Gln	Gly	Val	Asn	Asn	Ala	His	Thr	Thr	Tyr	Phe	
			260					265					270			
Gly	Met	Thr	Ser	Gly	Ala	Cys	Thr	Trp	Pro	Val	Ala	Ala	Ala	Glu	Pro	
		275					280					285				
Leu	Lys	Asp	Glu	Leu												
	290															

<210> 27

<211> 5327

<212> DNA

<213> Artificial Sequence

<220>

<223> pTP4a2 vector

<400> 27

gatcctccgc	gacgacagca	gcaaagaaat	aatcacccgc	ttccgtggca	ctggtagtga	60
tacgaatcta	caactcgata	ctaactacac	cctcacgcct	ttcgacaccc	taccacaatg	120
caacggttgt	gaagtacacg	gtggatatta	tattggatgg	gtctccgtcc	aggaccaagt	180
cgagtgcgtt	gtcaaacagc	aggttagcca	gtatccggac	tacgcgctga	ccgtgaccgg	240
ccackccctc	ggcgccctcc	tggcggcact	cactgccgcc	cagctgtctg	cgacatacga	300
caacatccgc	ctgtacacct	tggcggaacc	gcgccagcgg	aatcaggcct	tcgcgtcgtg	360
catgaacgat	gccttccaag	cctcgagccc	agatacgcg	cagtatttcc	gggtcactca	420
tgccaacgac	ggcatcccaa	acctgcccc	gggtggagcag	gggtacgccc	atggcggtgt	480
agagtactgg	agcgttgatc	cttacagcgc	ccagaacaca	tttgtctgca	ctggggatga	540
agtgcagtgc	tgtgaggccc	agggcggaca	gggtgtgaat	aatgcgcaca	cgacttattt	600
tgggatgacg	agcggagcct	gtacatggtg	atcagtcatt	tcagcctccc	cgagtgtacc	660
aggaaagatg	gatgtcctgg	agagggggcc	gcgtaaccc	tgaaggatga	gctgtaaaga	720
agcagatcgt	tcaaacattt	ggcaataaag	tttcttaaga	ttgaatcctg	ttgccggtct	780
tgcgatgatt	atcatataat	ttctgttgaa	ttacgttaag	catgtaataa	ttaacatgta	840
atgcatgacg	ttatttatga	gatgggtttt	tatgattaga	gtcccgcgat	tatacattta	900
atacgcgata	gaaaacaaaa	tatagcgcgc	aaactaggat	aaattatcgc	gcgcggtgtc	960
atctatgtta	ctagatcgat	aagcttctag	agcggccggt	ggagctccaa	ttcgccctat	1020
agtgagtcgt	attacgcgcg	ctcactggcc	gtcgttttac	aacgtcgtga	ctgggaaaac	1080
cctggcggtta	cccaacttaa	tcgccttgca	gcacatcccc	ctttcgccag	ctggcgtaat	1140
agcgaagagg	cccgcaccga	tcgcctttcc	caacagttgc	gcagcctgaa	tggcgaatgg	1200
gacgcgccct	gtagcggcgc	attaagcgcg	gcgggtgtgg	tggttacgcg	cagcgtgacc	1260
gctacacttg	ccagcgcctt	agcgcgcctt	cctttcgcct	tcttcccttc	ctttctcgcc	1320
acgttcgcgc	gctttccccc	tcaagctcta	aatcgggggc	tccctttagg	gttccgattt	1380
agtgcctttac	ggcacctcga	ccccaaaaaa	cttgattagg	gtgatgggtc	acgtagtggg	1440
ccatcgccct	gatagacggt	ttttcgccct	ttgacgttgg	agtcacggtt	ctttaatagt	1500
ggactcttgt	tccaaactgg	aacaacactc	aaccctatct	cggtctatct	ttttgattta	1560
taagggattt	tgccgatttc	ggcctattgg	ttaaaaaatg	agctgattta	acaaaaattt	1620
aacgcgaatt	ttaacaaaat	attaacgctt	acaatttagg	tggcactttt	cggggaaatg	1680
tgcgcggaac	ccctatttgt	ttatttttct	aaatacatct	aaatatgtat	ccgctcatga	1740
gacaataacc	ctgataaatg	cttcaataat	attgaaaaag	gaagagtatg	agtattcaac	1800
atttccgtgt	cgcccttatt	cccttttttg	cggcattttg	ccttccctgt	tttgctcacc	1860
cagaaacgct	ggtgaaagta	aaagatgctg	aagatcagtt	gggtgcacga	gtgggttaca	1920
tcgaaactgga	tctcaacagc	ggtaagatcc	ttgagagttt	tcgccccgaa	gaacgttttc	1980
caatgatgag	cactttttaa	gttctgctat	gtggcgcggt	attatcccgt	attgacgcgc	2040
ggcaagagca	actcggtcgc	cgcatacact	attctcagaa	tgacttggtt	gagtactcac	2100
cagtcacaga	aaagcatctt	acggatggca	tgacagtaag	agaattatgc	agtgtctgcca	2160
taaccatgag	tgataacact	gcggccaaact	tacttctgac	aacgatcgga	ggaccgaagg	2220
agctaaccgc	ttttttgcac	aacatggggg	atcatgtaac	tcgccttgat	cgttgggaac	2280
cggagctgaa	tgaagccata	ccaaacgcgc	agcgtgacac	cacgatgcct	gtagcaatgg	2340
caacaacggt	gcgcaaaact	ttaactggcg	aactacttac	tctagcttcc	cggcaacaat	2400
taatagactg	gatggaggcg	gataaagttg	caggaccact	tctgcgctcg	gcccttccgg	2460
ctggctgggt	tattgctgat	aaatctggag	ccggtgagcg	tgggtctcgc	ggtatcattg	2520
cagcactggg	gccagatggg	aagccctccc	gtatcgtagt	tatctacacg	acggggagtc	2580
aggcaactat	ggatgaacga	aatagacaga	tcgctgagat	aggtgcctca	ctgattaagc	2640

```

attggttaact gtcagaccaa gtttactcat atatacttta gattgattta aaacttcatt 2700
tttaatttaa aaggatctag gtgaagatcc tttttgataa tctcatgacc aaaatccctt 2760
aacgtgagtt ttcgttccac tgagcgtcag accccgtaga aaagatcaaa ggatcttctt 2820
gagatccttt ttttctgcgc gtaatctgct gcttgcaaac aaaaaaacca ccgctaccag 2880
cggtggtttg tttgccggat caagagctac caactctttt tccgaaggta actggcttca 2940
gcagagcgca gataccaaat actgtccttc tagtgtagcc gtagttaggc caccacttca 3000
agaactctgt agcaccgcct acatacctcg ctctgctaata cctgttacca gtggctgctg 3060
ccagtggcga taagtcgtgt cttaccgggt tggactcaag acgatagtta ccggataagg 3120
cgcagcggtc gggctgaacg ggggggttcgt gcacacagcc cagcttgag cgaacgacct 3180
acaccgaact gagataccta cagcgtgagc tatgagaaag cgccacgctt cccgaaggga 3240
gaaaggcgga caggtatccg gtaagcggca gggtcggaac aggagagcgc acgagggagc 3300
ttccaggggg aaacgcctgg tatctttata gtctgtcgg gtttcgccac ctctgacttg 3360
agcgtcgatt tttgtgatgc tcgtcagggg ggccggagcct atggaaaaac gccagcaacg 3420
cggccttttt acggttcctg gccttttgct ggcccttttg tccatgttc tttcctgcgt 3480
tatccctga ttctgtggat aaccgtatta ccgcctttga gtgagctgat accgctcgcc 3540
gcagccgaac gaccgagcgc agcgagtcag tgagcgagga agcggaagag cgcccaatac 3600
gcaaaccgcc tctccccgcg cgttggccga ttcattaatg cagctggcac gacaggtttc 3660
ccgactggaa agcgggcagt gagcgcaacg caattaatgt gagttagctc actcattagg 3720
caccaccaggc tttacacttt atgcttccgg ctctgtatgt gtgtggaatt gtgagcggat 3780
aacaatttca cacaggaaac agctatgacc atgattacgc caagcgcgca attaacctc 3840
actaaaggga acaaaagctg ggtaccgggc ccccccctga ggctattcat atgcttgaga 3900
agagagtcgg gatagtccaa aataaaacaa aggttaagatt acctggtcaa aagtgaaaac 3960
atcagttaaa aggtggtata agtaaaatat cggtaataaa aggtggccca aagtgaaatt 4020
tactcttttc tactattata aaaattgagg atgttttgtc ggtactttga tacgtcattt 4080
ttgtatgaat tggtttttaa gtttattcgc gatttggaat tgcatactctg tatttgagtc 4140
ggtttttaag ttcgttgctt ttgtaaatac agagggattt gtataagaaa tatcttttaa 4200
aaaccatata gctaatttga cataattttt gagaaaaata tatattcagg cgaattccac 4260
aatgaacaat aataagatta aaatagcttg ccccgcttgc agcgatgggt attttttcta 4320
gtaaaataaa agataaaact agactcaaaa catttacaaa aacaaccctt aaagtcctaa 4380
agcccaaagt gctatgcacg atccatagca agcccagccc aaccacaacc aaccacaacc 4440
accccagtgc agccaactgg caaatagtct ccaccccccg cactatcacc gtgagttgtc 4500
cgcaccaccc cagctctcgc agccaaaaaa aaaaaaagaa agaaaaaaa gaaaaagaaa 4560
aacagcaggt gggtcggggt cgtggggggc ggaaaagcga ggaggatcgc gagcagcgac 4620
gaggcccgcc cctccctccg cttccaaaga aacgcccccc atcgccacta tatacatacc 4680
ccccctctc ctcccatccc cccaacccta ccaccaccac caccaccacc tctccccccc 4740
tcgctgccgg acgacgagct cctccccctt cccctccgc cgccgccggt aaccaccccg 4800
cccctctcct ctttctttct ccgttttttt tttcgtctcg gtctcgatct ttggccttgg 4860
tagtttgggt gggcgagagc ggcttcgtcg cccagatcgg tgcgcgggag gggcgggatc 4920
tcgcggctgg cgtctccggg cgtgagtcgg cccggatcct cgcggggaat ggggctctcg 4980
gatgtagatc ttctttcttt cttctttttg tggtagaatt tgaatccctc agcattgttc 5040
atcggtagtt tttcttttca tgatttgtga caaatgcagc ctctgtcgga gcttttttgt 5100
agcaagctta acatgaagca gttctccgcc aaacacgtcc tcgcagttgt ggtgactgca 5160
gggcacgcct tagcagcctc tacgcaaggc atctccgaag acctctacag ccgttttagtc 5220
gaaatggcca ctatctccca agctgcctac gccgacctgt gcaacattcc gtcgactatt 5280
atcaaggag agaaaattta caattctcaa actgacatta acggatg 5327

```

<210> 28

<211> 209

<212> PRT

<213> Artificial Sequence

<220>

<223> pTP4a2 vector

<221> VARIANT

<222> (1)...(209)

<223> Xaa = Any Amino Acid

<400> 28

```

Ile Leu Arg Asp Asp Ser Ser Lys Glu Ile Ile Thr Val Phe Arg Gly
 1             5             10             15

```

Thr Gly Ser Asp Thr Asn Leu Gln Leu Asp Thr Asn Tyr Thr Leu Thr
 20 25 30
 Pro Phe Asp Thr Leu Pro Gln Cys Asn Gly Cys Glu Val His Gly Gly
 35 40 45
 Tyr Tyr Ile Gly Trp Val Ser Val Gln Asp Gln Val Glu Ser Leu Val
 50 55 60
 Lys Gln Gln Val Ser Gln Tyr Pro Asp Tyr Ala Leu Thr Val Thr Gly
 65 70 75 80
 His Xaa Leu Gly Ala Ser Leu Ala Ala Leu Thr Ala Ala Gln Leu Ser
 85 90 95
 Ala Thr Tyr Asp Asn Ile Arg Leu Tyr Thr Phe Gly Glu Pro Arg Ser
 100 105 110
 Gly Asn Gln Ala Phe Ala Ser Tyr Met Asn Asp Ala Phe Gln Ala Ser
 115 120 125
 Ser Pro Asp Thr Thr Gln Tyr Phe Arg Val Thr His Ala Asn Asp Gly
 130 135 140
 Ile Pro Asn Leu Pro Pro Val Glu Gln Gly Tyr Ala His Gly Gly Val
 145 150 155 160
 Glu Tyr Trp Ser Val Asp Pro Tyr Ser Ala Gln Asn Thr Phe Val Cys
 165 170 175
 Thr Gly Asp Glu Val Gln Cys Cys Glu Ala Gln Gly Gly Gln Gly Val
 180 185 190
 Asn Asn Ala His Thr Thr Tyr Phe Gly Met Thr Ser Gly Ala Cys Thr
 195 200 205
 Trp

<210> 29

<211> 5338

<212> DNA

<213> Artificial Sequence

<220>

<223> pTP3-1 vector

<400> 29

```

aagcttacca tggcccacgc ccgcgtcctc ctcttgggcg tcgccgtgct ggccacggcc      60
gccgtcgccg tcgcctcctc ctctcctctc gccgactcca acccgatccg gcccgtcacc      120
gaccgcgcgg ccgcctccac gcagggcatc tccgaagacc tctacagccg tttagtcgaa      180
atggccacta tctcccaagc tgcctacgcc gacctgtgca acattccgtc gactattatc      240
aagggagaga aaattttacaa ttctcaaact gacattaacg gatggatcct ccgcgacgac      300
agcagcaaag aaataatcac cgtcttccgt ggcactggta gtgatacgaa tctacaactc      360
gatactaact acaccctcac gcctttcgac accctaccac aatgcaacgg ttgtgaagta      420
cacggtggat attatattgg atgggtctcc gtccaggacc aagtcgagtc gcttgtcaaa      480
cagcaggtta gccagtatcc ggactacgcg ctgaccgtga ccggccackc cctcggcgcc      540
tccctggcgg cactcactgc cgcccagctg tctgcgacat acgacaacat ccgcctgtac      600
accttcggcg aaccgcgcag cggcaatcag gccttcgcgt cgtacatgaa cgatgccttc      660
caagcctcga gccagatac gacgcagtat ttccgggtca ctcatgccaa cgacggcatc      720
ccaaacctgc ccccggtgga gcaggggtac gcccatggcg gtgtagagta ctggagcgtt      780
gatccttaca gcgcccagaa cacatttgtc tgcactgggg atgaagtgca gtgctgtgag      840
gcccgaggcg gacagggtgt gaataatgcg cacacgactt attttgggat gacgagcggc      900
gcatgcacct ggccggtcgc ggccgcggaa accactgaag gatgagctgt aaagaagcag      960
atcgttcaaa catttggaac taaagtttct taagattgaa tcctgttgcc ggtcttgca      1020
tgattatcat ataatttctg ttgaattacg ttaagcatgt aataattaac atgtaatgca      1080
tgacgttatt tatgagatgg gtttttatga ttagagtccc gcaattatac atttaatacg      1140
cgatagaaaa caaaatatag cgcgcaaact aggataaatt atcgcgcgcg gtgtcatcta      1200
tgttactaga tcgataagct tctagagcgg ccggtggagc tccaattcgc cctatagtga      1260
gtcgtattac gcgcgctcac tggccgtcgt tttaaacgt cgtgactggg aaaaccctgg      1320
cgttacccaa cttaatcgcc ttgcagcaca tccccctttc gccagctggc gtaatagcga      1380
agaggcccgcc accgatcgcc cttcccaaca gttgcgcagc ctgaatggcg aatgggacgc      1440
gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt acgcgcagcg tgaccgctac      1500

```

acttgccagc	gccctagcgc	ccgctccttt	cgctttcttc	ccttcctttc	tcgccacgtt	1560
cgccggcttt	ccccgtcaag	ctctaaatcg	ggggctccct	ttaggggttc	gatttagtgc	1620
tttacggcac	ctcgacccca	aaaaacttga	ttaggggtgat	ggttcacgta	gtggggccatc	1680
gccctgatag	acgggtttttc	gccctttgac	gttggagtc	acgttcttta	atagtggact	1740
cttgttccaa	actggaacaa	cactcaaccc	tatctcggtc	tattcttttg	atttataagg	1800
gattttgccg	atttcggcct	attgggttaa	aaatgagctg	atttaacaaa	aatttaacgc	1860
gaattttaac	aaaatattaa	cgcttacaat	ttaggtggca	cttttcgggg	aaatgtgcgc	1920
ggaaccccta	tttgtttatt	tttctaaata	cattcaaata	tgtatccgct	catgagacaa	1980
taaccctgat	aaatgcttca	ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	2040
cgtgtcgccc	ttattccctt	ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaa	2100
acgctggtga	aagtaaaaga	tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	2160
ctggatctca	acagcggtaa	gaccttgag	agttttcgcc	ccgaagaacg	ttttccaatg	2220
atgagcactt	ttaaagtctt	gctatgtggc	gcggatttat	cccgatttga	cgccgggcaa	2280
gagcaactcg	gtcgccgcct	acactattct	cagaatgact	tggttgagta	ctcaccagtc	2340
acagaaaagc	atcttacgga	tggcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	2400
atgagtata	acactgcggc	caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	2460
accgcttttt	tgcacaacat	gggggatcat	gtaactcgcc	ttgatcggtg	ggaaccggag	2520
ctgaatgaag	ccataccaaa	cgacgagcgt	gacaccacga	tgctgtagc	aatggcaaca	2580
acgttgcgca	aactattaac	tggcgaacta	cttactctag	cttcccggca	acaattaata	2640
gactggatgg	aggcggataa	agttgcagga	ccacttctgc	gctcggccct	tcgggctggc	2700
tggttttattg	ctgataaaatc	tggagccggg	gagcgtgggt	ctcgcgggtat	cattgcagca	2760
ctggggccag	atggtaagcc	ctcccgtatc	gtagttatct	acacgacggg	gagtcaggca	2820
actatggatg	aacgaaatag	acagatcgct	gagatagggtg	cctcactgat	taagcattgg	2880
taactgtcag	accaagttta	ctcatatata	cttttagattg	atttaaaact	tcatttttaa	2940
tttaaaagga	tctaggtgaa	gacctttttt	gataatctca	tgaccaaaat	cccttaacgt	3000
gagttttcgt	tccactgagc	gtcagacccc	gtagaaaaga	tcaaaggatc	ttcttgagat	3060
cctttttttc	tgcgcgtaat	ctgctgcttg	caaacaaaaa	aaccaccgct	accagcggtg	3120
gtttgtttgc	cggatcaaga	gctaccaact	ctttttccga	aggtaactgg	cttcagcaga	3180
gcgcagatac	caaatactgt	ccttctagt	tagccgtagt	taggccacca	cttcaagaac	3240
tctgtagcac	cgcctacata	cctcgtctgt	ctaactcctgt	taccagtggc	tgctgccagt	3300
ggcgataagt	cgtgtcttac	cgggttgagc	tcaagacgat	agttaccgga	taaggcgcag	3360
cggtcgggct	gaacgggggg	ttcgtgcaca	cagcccagct	tggagcgaac	gacctacacc	3420
gaactgagat	acctacagcg	tgagctatga	gaaagcgcca	cgcttcccga	aggagagaaag	3480
gcggacaggt	atccggtaag	cggcaggggc	ggaacaggag	agcgcacgag	ggagcttcca	3540
gggggaaacg	cctggtatct	ttatagtcct	gtcgggtttc	gccacctctg	acttgagcgt	3600
cgatttttgt	gatgctcgtc	agggggggcg	agcctatgga	aaaacgccag	caacgcggcc	3660
tttttacggg	tcctggcctt	ttgctggcct	tttgctcaca	tgttcttttc	tgcgttatcc	3720
cctgattctg	tggataaccg	tattaccgcc	tttgagttag	ctgataccgc	tcgccgcagc	3780
cgaacgaccg	agcgcagcga	gtcagttagc	gaggaaagcg	aagagcgccc	aatacgcaaa	3840
ccgcctctcc	ccgcgcggtg	gccgattcat	taatgcagct	ggcacgacag	gtttcccagc	3900
tggaaagcgg	gcagttagcg	caacgcaatt	aatgtgagtt	agctcactca	ttaggcaccc	3960
caggctttac	actttatgct	tccggctcgt	atgttgtgtg	gaattgtgag	cggataacaa	4020
tttcacacag	gaaacagcta	tgacctatga	tacgccaagc	gcgcaattaa	ccctcactaa	4080
agggaacaaa	agctgggtac	cgggcccccc	ctcgaggtca	ttcatatgct	tgagaagaga	4140
gtcgggatag	tccaaaataa	aacaaaggta	agattacctg	gtcaaaaagt	aaaacatcag	4200
ttaaaagggtg	gtataagtaa	aatatcggtg	ataaaaagggtg	gcccaaagt	aaatttactc	4260
ttttctacta	ttataaaaaat	tgaggatggt	ttgtcgggtac	tttgatacgt	cattttttgta	4320
tgaattgggt	tttaagttta	ttcgcgattt	ggaaatgcat	atctgtatct	gagtcgggtt	4380
ttaagtctgt	tgtttttgta	aatacagagg	gatttgtata	agaaatatct	ttaaaaaacc	4440
catatgctaa	tttgacataa	tttttgagaa	aaatatatat	tcaggcgaat	tcacacatga	4500
acaataataa	gattaaaata	gcttgccccc	gttgacagcga	tgggtatttt	ttctagtaaa	4560
ataaaagata	aacttagact	caaaaacattt	acaaaaacaa	cccctaaagt	cctaaagccc	4620
aaagtgtctat	gcacgatcca	tagcaagccc	agcccaaccc	aacccaaccc	aacccaaccc	4680
agtgcagcca	actggcaaat	agtctccacc	cccggcacta	tcaccgtgag	ttgtccgcac	4740
caccgcacgt	ctcgcagcca	aaaaaaaaaa	aagaaagaaa	aaaaagaaaa	agaaaaacag	4800
cagggtgggtc	cgggtcgtgg	gggcccggaaa	agcggaggag	atcgcgagca	gcgacgaggc	4860
ccggccctcc	ctccgcttcc	aaagaaacgc	cccccatcgc	cactatatac	ataccccccc	4920
ctctcctccc	atccccccaa	ccctaccacc	accaccacca	ccacctctc	ccccctcgt	4980
gccggacgac	gagctcctcc	cccccccccc	tccgcgcgcg	ccggtaacca	ccccgcccct	5040
ctcctctttc	ttctcctgtt	tttttttttcg	tctcgggtctc	gatctttggc	cttggttagtt	5100
tgggtgggcg	agagcggctt	cgtcgcccag	atcgggtgcgc	gggagggggcg	ggatctcgcg	5160

```

gctggcgtct cggggcgtga gtcggcccg atcctcgcg ggaatggggc tctcggtatgt 5220
agatcttctt tctttcttct ttttgtggta gaatttgaat ccctcagcat tgttcatcgg 5280
tagtttttct tttcatgatt tgtgacaaat gcagcctcgt gcggagcttt tttgtagc 5338

```

<210> 30
 <211> 311
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> pTP3-1 vector

<221> VARIANT
 <222> (1)...(311)
 <223> Xaa = Any Amino Acid

<400> 30

Met	Ala	His	Ala	Arg	Val	Leu	Leu	Leu	Ala	Leu	Ala	Val	Leu	Ala	Thr	1	5	10	15
Ala	Ala	Val	Ala	Val	Ala	Ser	Ser	Ser	Ser	Phe	Ala	Asp	Ser	Asn	Pro	20	25	30	
Ile	Arg	Pro	Val	Thr	Asp	Arg	Ala	Ala	Ala	Ser	Thr	Gln	Gly	Ile	Ser	35	40	45	
Glu	Asp	Leu	Tyr	Ser	Arg	Leu	Val	Glu	Met	Ala	Thr	Ile	Ser	Gln	Ala	50	55	60	
Ala	Tyr	Ala	Asp	Leu	Cys	Asn	Ile	Pro	Ser	Thr	Ile	Ile	Lys	Gly	Glu	65	70	75	80
Lys	Ile	Tyr	Asn	Ser	Gln	Thr	Asp	Ile	Asn	Gly	Trp	Ile	Leu	Arg	Asp	85	90	95	
Asp	Ser	Ser	Lys	Glu	Ile	Ile	Thr	Val	Phe	Arg	Gly	Thr	Gly	Ser	Asp	100	105	110	
Thr	Asn	Leu	Gln	Leu	Asp	Thr	Asn	Tyr	Thr	Leu	Thr	Pro	Phe	Asp	Thr	115	120	125	
Leu	Pro	Gln	Cys	Asn	Gly	Cys	Glu	Val	His	Gly	Gly	Tyr	Tyr	Ile	Gly	130	135	140	
Trp	Val	Ser	Val	Gln	Asp	Gln	Val	Glu	Ser	Leu	Val	Lys	Gln	Gln	Val	145	150	155	160
Ser	Gln	Tyr	Pro	Asp	Tyr	Ala	Leu	Thr	Val	Thr	Gly	His	Xaa	Leu	Gly	165	170	175	
Ala	Ser	Leu	Ala	Ala	Leu	Thr	Ala	Ala	Gln	Leu	Ser	Ala	Thr	Tyr	Asp	180	185	190	
Asn	Ile	Arg	Leu	Tyr	Thr	Phe	Gly	Glu	Pro	Arg	Ser	Gly	Asn	Gln	Ala	195	200	205	
Phe	Ala	Ser	Tyr	Met	Asn	Asp	Ala	Phe	Gln	Ala	Ser	Ser	Pro	Asp	Thr	210	215	220	
Thr	Gln	Tyr	Phe	Arg	Val	Thr	His	Ala	Asn	Asp	Gly	Ile	Pro	Asn	Leu	225	230	235	240
Pro	Pro	Val	Glu	Gln	Gly	Tyr	Ala	His	Gly	Gly	Val	Glu	Tyr	Trp	Ser	245	250	255	
Val	Asp	Pro	Tyr	Ser	Ala	Gln	Asn	Thr	Phe	Val	Cys	Thr	Gly	Asp	Glu	260	265	270	
Val	Gln	Cys	Cys	Glu	Ala	Gln	Gly	Gly	Gln	Gly	Val	Asn	Asn	Ala	His	275	280	285	
Thr	Thr	Tyr	Phe	Gly	Met	Thr	Ser	Gly	Ala	Cys	Thr	Trp	Pro	Val	Ala	290	295	300	
Ala	Ala	Glu	Thr	Thr	Glu	Gly										305	310		

<210> 31
 <211> 5337
 <212> DNA

<213> Artificial Sequence

<220>

<223> pTU5 vector

<400> 31

aagcttacca	tggcccacgc	ccgcgtcctc	ctcctggcgc	tcgccgtgct	ggccacggcc	60
gccgtcgccg	tcgcctcctc	ctcctccttc	gccgactcca	acccgggccg	gcccgtcacc	120
gaccgcgcgg	ccgcctccac	gcagggcatc	tccgaagacc	tctacagccg	tttagtcgaa	180
atggccacta	tctcccaagc	tgcctacgcc	gacctgtgca	acattccgtc	gactattatc	240
aagggagaga	aaattttacaa	ttctcaaact	gacattaacg	gatggatcct	ccgcgacgac	300
agcagcaaag	aaataatcac	cgtcttccgt	ggcactggta	gtgatacgaa	tctacaactc	360
gatactaact	acaccctcac	gcctttcgac	accctaccac	aatgcaacgg	ttgtgaagta	420
cacgggtggat	attatattgg	atgggtctcc	gtccaggacc	aagtcgagtc	gcttgtcaaa	480
cagcaggtta	gccagtatcc	ggactacgcg	ctgaccgtga	ccggccactc	cctcggcgcc	540
tccctggcgg	cactcactgc	cgcccagctg	tctgcgacat	acgacaacat	ccgcctgtac	600
accttcggcg	aaccgcgcag	cggcaatcag	gccttcgcgt	cgtacatgaa	cgatgccttc	660
caagcctcga	gccagatac	gacgcagtat	ttccgggtca	ctcatgccaa	cgacggcatc	720
ccaaacctgc	ccccgggtgga	gcaggggtac	gcccattggcg	gtgtagagta	ctggagcggt	780
gacctttaca	gcgcccagaa	cacatttgtc	tgcactgggg	atgaagtgca	gtgctgtgag	840
gccagggcg	gacaggggtg	gaataatgcg	cacacgactt	attttgggat	gacgagcggc	900
gcattgcacct	ggccgggtcgc	ggccgcggaa	ccactgaagg	atgagctgta	aagaagcaga	960
tcgttcaaac	atgtggcaat	aaagtcttct	aagattgaat	cctgttgccg	gtcttgcat	1020
gattatcata	taatttctgt	tgaattacgt	taagcatgta	ataattaaca	tgtaatgcat	1080
gacgttattt	atgagatggg	tttttatgat	tagagtcccg	caattataca	tttaatacgc	1140
gatagaaaac	aaaatatagc	gcgcaaacta	ggataaatta	tcgcgcgcgg	tgatcatctat	1200
gttactagat	cgataagctt	ctagagcggc	cgggtggagct	ccaattcgcc	ctatagttag	1260
tcgtattacg	cgcgctcact	ggccgtcgtt	ttacaacgtc	gtgactggga	aaaccctggc	1320
gttaccacaac	ttaatcgctt	tgcagcacat	ccccctttcg	ccagctggcg	taatagcgaa	1380
gaggcccgca	ccgatcgccc	ttcccaacag	ttgcgcagcc	tgaatggcga	atgggacgcg	1440
ccctgtagcg	gcgcatttaag	cgcggcgggg	gtgggtggta	cgcgcagcgt	gaccgctaca	1500
cttgccagcg	ccctagcgcc	cgtctctttc	gctttcttcc	cttctcttct	cgccacgttc	1560
gccggctttc	cccgtcaagc	tctaaatcgg	gggctccctt	tagggttccg	atttagtgct	1620
ttacggcacc	tcgaccccaa	aaaacttgat	taggggtgatg	gttcacgtag	tgggccatcg	1680
ccctgataga	cggttttttcg	ccctttgacg	ttggagtcca	cgttctttta	tagtggactc	1740
ttgttccaaa	ctggaacaac	actcaaccct	atctcgggtct	attcttttga	tttataaggg	1800
attttgccga	tttcggccta	ttggttaaaa	aatgagctga	tttaacaaaa	atttaacgcg	1860
aattttaaca	aaatattaac	gcttacaatt	taggtggcac	ttttcgggga	aatgtgcgcg	1920
gaacccttat	ttgtttattt	ttctaaatac	attcaaatat	gtatccgctc	atgagacaat	1980
aaccctgata	aatgcttcaa	taatattgaa	aaaggaagag	tatgagtatt	caacatttcc	2040
gtgtcgccct	tattcccttt	tttgcgcat	tttgcccttc	tgtttttgct	caccagaaa	2100
cgtggtgaa	agtaaaagat	gctgaagatc	agttgggtgc	acgagtgggt	tacatcgaa	2160
tggatctcaa	cagcggttaag	atccttgaga	gttttcgccc	cgaagaacgt	tttccaatga	2220
tgagcacttt	taaagtcttg	ctatgtggcg	cggattatc	ccgtattgac	gccgggcaag	2280
agcaactcgg	tcgcgcgata	cactattctc	agaatgactt	ggttgagtac	tcaccagtca	2340
cagaaaagca	tcttacggat	ggcatgacag	taagagaatt	atgcagtgtc	gccataacca	2400
tgagtgataa	cactgcggcc	aacttacttc	tgacaacgat	cggaggaccg	aaggagctaa	2460
ccgctttttt	gcacaacatg	ggggatcatg	taactcgctt	tgatcggttg	gaaccggagc	2520
tgaatgaagc	cataccaaac	gacgagcgtg	acaccacgat	gcctgtagca	atggcaacaa	2580
cgttgcgcaa	actattaact	ggcgaactac	ttactctagc	ttcccggcaa	caattaatag	2640
actggatgga	ggcggataaa	gttgacggac	cacttctgcg	ctcggccctt	ccggctggct	2700
ggttttattgc	tgataaatct	ggagccgggtg	agcgtgggtc	tcgcgggtatc	attgcagcac	2760
tggggccaga	tggtaagccc	tcccgtatcg	tagttatcta	cacgacgggg	agtcaggcaa	2820
ctatggatga	acgaaataga	cagatcgctg	agataggtgc	ctcactgatt	aagcattgggt	2880
aactgtcaga	ccaagtttac	tcatatatac	tttagattga	tttaaaaactt	cattttttaat	2940
ttaaaaggat	ctaggtgaag	atcctttttg	ataatctcat	gaccaaatac	ccttaacgtg	3000
agttttcgtt	ccactgagcg	tcagaccccg	tagaaaagat	caaaggatct	tcttgagatc	3060
ctttttttct	gcgcgtaatc	tgctgcttgc	aaacaaaaaa	accaccgcta	ccagcgggtg	3120
tttggttgcc	ggatcaagag	ctaccaactc	tttttccgaa	ggtaactggc	ttcagcagag	3180
cgcagatacc	aaatactgtc	cttctagtgt	agccgtagtt	aggccaccac	ttcaagaact	3240
ctgtagcacc	gcctacatac	ctcgctctgc	taatcctggt	accagtggct	gctgccagtg	3300

gcgataagtc	gtgtcttacc	gggttggact	caagacgata	gttaccggat	aaggcgcagc	3360
ggtcgggctg	aacggggggt	tcgtgcacac	agcccagctt	ggagcgaacg	acctacaccg	3420
aactgagata	cctacagcgt	gagctatgag	aaagcgccac	gcttcccga	gggagaaagg	3480
cggacaggta	tccggtaagc	ggcagggtcg	gaacaggaga	gcgcacgagg	gagcttccag	3540
ggggaaacgc	ctggtatctt	tatagtcctg	tcgggtttcg	ccacctctga	cttgagcgtc	3600
gattttttgtg	atgctcgtca	ggggggcgga	gcctatggaa	aaacgccagc	aacgcggcct	3660
ttttacgggt	cctggccttt	tgctggcctt	ttgctcacat	gttctttcct	gcgttatccc	3720
ctgattctgt	ggataaccgt	attaccgcct	ttgagtgagc	tgataccgct	cgccgcagcc	3780
gaacgaccga	gcgcagcgag	tcagtgagcg	aggaagcgga	agagcgccca	atacgcaaac	3840
cgcctctccc	cgcgcgttgg	ccgattcatt	aatgcagctg	gcacgacagg	tttcccgaact	3900
ggaaagcggg	cagtgagcgc	aacgcaatta	atgtgagtta	gctcactcat	taggcacccc	3960
aggctttaca	ctttatgctt	ccggctcgt	tggtgtgtgg	aattgtgagc	ggataacaat	4020
ttcacacagg	aaacagctat	gaccatgatt	acgccaaagc	cgcaattaac	cctcactaaa	4080
gggaacaaaa	gctgggtacc	gggccccccc	tcgaggctcat	tcatatgctt	gagaagagag	4140
tcgggatagt	ccaaaataaa	acaaaggtaa	gattacctgg	tcaaaagtga	aaacatcagt	4200
taaaagggtg	tataagtaaa	atatcggtaa	taaaagggtg	cccaaagtga	aatttactct	4260
tttctactat	tataaaaatt	gaggatgttt	tgtcggtaact	ttgatacgtc	atttttgtat	4320
gaattgggtt	ttaagtattat	tcgcgatttg	gaaatgcata	tctgtatttg	agtcgggtttt	4380
taagtctcgt	gctttttgtaa	atacagaggg	atltgtataa	gaaatatctt	taaaaaaccc	4440
atatgcta	ttgacataat	ttttgagaaa	aatatatatt	caggcgaatt	ccacaatgaa	4500
caataataag	attaaaatag	cttgcccccg	ttgcagcgat	gggtattttt	tctagtaaaa	4560
taaaagataa	acttagactc	aaaacattta	caaaaacaac	ccctaaagtc	ctaaagccca	4620
aagtgtctat	cacgatccat	agcaagccca	gcccaaccca	acccaaccca	acccacccca	4680
gtgcagccaa	ctggcaaata	gtctccaccc	ccggcactat	caccgtgagt	tgtccgcacc	4740
accgcacgtc	tcgcagccaa	aaaaaaaaaa	agaaagaaaa	aaaagaaaaa	gaaaaacagc	4800
agggtgggtcc	gggtcgtggg	ggccggaaaa	gcgaggagga	tcgcgagcag	cgacgaggcc	4860
cggccctccc	tccgcttcca	aagaaacgcc	ccccatcgcc	actatataca	tacccccccc	4920
tctcctccca	tcccccaaac	cctaccacca	ccaccaccac	cacctcctcc	cccctcgtcg	4980
ccggacgacg	agctcctccc	ccctccccct	ccgcgcgcgc	cggtaacccac	cccgcacctc	5040
tctctcttct	ttctccgttt	tttttttctg	ctcgggtctcg	atctttggcc	ttggtagttt	5100
gggtggggcga	gagcggcttc	gtcgcccaga	tcgggtgcgcg	ggagggggcgg	gatctcgcg	5160
ctggcgtctc	cgggcgtgag	tcggcccggga	tcctcgcggg	gaatggggct	ctcggtatga	5220
gatcttcttt	ctttcttctt	tttgtggtag	aatttgaatc	cctcagcatt	gttcatcggt	5280
agtttttctt	ttcatgattt	gtgacaaatg	cagcctcgtg	cggagctttt	ttgtagc	5337

<210> 32

<211> 4773

<212> DNA

<213> Artificial Sequence

<220>

<223> pGT6 vector

<400> 32

aagcttacca	tggtgagcaa	gggcgaggag	ctgttcaccg	gggtgggtgcc	catcctgggtc	60
gagctggacg	gcgacgtgaa	cggccacaag	ttcagcgtgt	ccggcgaggg	cgagggcgat	120
gccacctacg	gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	gcccgtgccc	180
tggcccaccc	tcgtgaccac	cttcacctac	ggcgtgcagt	gcttcagccg	ctaccccgac	240
cacatgaagc	agcacgactt	cttcaagtcc	gccatgccc	aaggctacgt	ccaggagcgc	300
accatcttct	tcaaggacga	cggcaactac	aagacccgcg	ccgaggtgaa	gttcgagggc	360
gacaccctgg	tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	cggcaacatc	420
ctggggcaca	agctggagta	caactacaac	agccacaacg	tctatatcat	ggccgacaag	480
cagaagaacg	gcatcaagg	gaacttcaag	atccgccaca	acatcgagga	cggcagcgtg	540
cagctcgccg	accactacca	gcagaacacc	cccatcgggc	acggcccccg	gctgctgccc	600
gacaaccact	acctgagcac	ccagtccgcc	ctgagcaaag	accccaacga	gaagcgcgat	660
cacatgggtcc	tgctggagtt	cgtgaccgcc	gccgggatca	ctcacggcat	ggacgagctg	720
tacaagtaaa	gcggccgccc	gggctgcagg	gaaaccactg	aaggatgagc	tgtaaagaag	780
cagatcgttc	aaacattttg	caataaagtt	tcttaagatt	gaatcctgtt	gccggtcttg	840
cgatgattat	catataattt	ctgttgaatt	acgttaagca	tgtaataatt	aacatgtaat	900
gcatgacgtt	atltatgaga	tgggttttta	tgattagagt	cccgaatta	tacattta	960
acgcgataga	aaacaaaata	tagcgcgcaa	actaggataa	attatcgcg	gcgggtgtcat	1020

ctatgttact	agatcgataa	gcttctagag	cggccggtgg	agctccaatt	cgccctatag	1080
tgagtcgtat	tacgcgcgct	cactggccgt	cgtttttacaa	cgtcgtgact	gggaaaaccc	1140
tggcggttacc	caacttaatc	gccttgcagc	acatccccct	ttcgccagct	ggcgtaatag	1200
cgaagaggcc	cgcaccgac	gcccttccca	acagttgccc	agcctgaatg	gcgaatggga	1260
cgcgcctgt	agcggcgcat	taagcgcggc	gggtgtggtg	gttacgcgca	gcgtgaccgc	1320
tacacttgcc	agcgccttag	cgcgcgctcc	tttcgctttc	ttcccttcct	ttctcgccac	1380
gttcgcccgc	tttccccgtc	aagctctaaa	tcgggggctc	ccttttaggg	tccgatttag	1440
tgcttttacgg	cacctcgacc	ccaaaaaact	tgattagggg	gatggttcac	gtagtggggc	1500
atcgccctga	tagacgggtt	ttcgcccttt	gacgttggag	ttcacgttct	ttaatagtgg	1560
actcttggtc	caaactggaa	caacactcaa	ccctatctcg	gtctattctt	ttgatttata	1620
agggattttg	ccgatttcgg	cctattgggt	aaaaaatgag	ctgattttaac	aaaaatttaa	1680
cgcgaatttt	aacaaaatat	taacgcttac	aatttaggtg	gcacttttcg	gggaaatgtg	1740
cgcggaaccc	ctatttggtt	atctttctaa	atacatctca	atatgtatcc	gctcatgaga	1800
caataaccct	gataaatgct	tcaataatat	tgaaaaagga	agagtatgag	tattcaacat	1860
ttcgtgtcg	cccttattcc	cttttttgcg	gcattttgcc	ttcctgtttt	tgctcaccca	1920
gaaacgctgg	tgaaagtaaa	agatgctgaa	gatcagttgg	gtgcacgagt	gggttacatc	1980
gaactggatc	tcaacagcgg	taagatcctt	gagagttttc	gccccgaaga	acgttttcca	2040
atgatgagca	cttttaaagt	tctgctatgt	ggcgcgggtat	tatcccggtat	tgacgcgggg	2100
caagagcaac	tcggtcgccc	catacactat	tctcagaatg	acttggttga	gtactcacca	2160
gtcacagaaa	agcatcttac	ggatggcatg	acagtaagag	aattatgcag	tgctgccata	2220
accatgagtg	ataacactgc	ggccaaacta	cttctgacaa	cgatcggagg	accgaaggag	2280
ctaaccgctt	ttttgcacaa	catgggggat	catgtaactc	gccttgatcg	ttgggaaccg	2340
gagctgaatg	aagccatacc	aaacgacgag	cgtgacacca	cgatgcctgt	agcaatggca	2400
acaacgttgc	gcaaactatt	aactggcgaa	ctacttactc	tagcttcccg	gcaacaatta	2460
atagactgga	tggaggcgga	taaagtgtga	ggaccacttc	tgcgctcggc	ccttccggct	2520
ggctggttta	ttgctgataa	atctggagcc	ggtagcgtg	ggtctcgcg	tatcattgca	2580
gcactggggc	cagatggtaa	gccctcccg	atcgtagtta	tctacacgac	ggggagtcag	2640
gcaactatgg	atgaacgaaa	tagacagatc	gctgagatag	gtgcctcact	gattaagcat	2700
tggttaactgt	cagaccaagt	ttactcatat	atactttaga	ttgatttaaa	acttcatttt	2760
taattttaaaa	ggatctaggt	gaagatcctt	tttgataatc	tcatgaccaa	aatcccttaa	2820
cgtgagtttt	cgttccactg	agcgtcagac	cccgtagaaa	agatcaaagg	atcttcttga	2880
gacccctttt	ttctgcgcgt	aatctgctgc	ttgcaaacaa	aaaaaccacc	gctaccagcg	2940
gtggttttgt	tgcgggatca	agagctacca	actctttttc	cgaaggtaac	tggcttcagc	3000
agagcgcaga	taccaaatac	tgtccttcta	gtgtagccgt	agttaggcca	ccacttcaag	3060
aactctgtag	caccgcctac	atacctcgct	ctgctaattc	tgttaccagt	ggctgctgcc	3120
agtggcgata	agtcgtgtct	taccgggttg	gactcaagac	gatagttacc	ggataaggcg	3180
cagcggtcgg	gctgaacggg	gggttcgtgc	acacagccca	gcttgagcgc	aacgacctac	3240
accgaactga	gatacctaca	gcgtgagcta	tgagaaagcg	ccacgccttc	cgaagggaga	3300
aaggcggaca	ggtatccgg	aagcggcagg	gtcggaacag	gagagcgcac	gagggagctt	3360
ccagggggaa	acgcctggta	tctttatagt	cctgtcgggt	ttcgccacct	ctgacttgag	3420
cgtcgatttt	tgtgatgctc	gtcagggggg	cggagcctat	ggaaaaacgc	cagcaacgcg	3480
gccttttttac	ggttccctgg	cttttgctgg	ccttttgctc	acatgttctt	tcctgcgtta	3540
tcccttgatt	ctgtggataa	ccgtattacc	gcctttgagt	gagctgatac	cgctcgccgc	3600
agccgaacga	ccgagcgcag	cgagtcagtg	agcgagggaag	cggagagcgc	cccaatacgc	3660
aaaccgcctc	tccccgcgcg	ttggccgatt	cattaatgca	gctggcacga	caggtttccc	3720
gactggaaag	cgggcagtg	gcgcaacgca	attaatgtga	gttagctcac	tcattaggca	3780
ccccaggctt	tacactttat	gcttccggct	cgtatgttgt	gtggaattgt	gagcggataa	3840
caatttcaca	caggaaacag	ctatgaccat	gattacgcca	agcgcgcaat	taaccctcac	3900
taaagggaac	aaaagctgga	attccacaat	gaacaataat	aagattaaaa	tagcttgccc	3960
ccgttgcagc	gatgggtatt	ttttctagta	aaataaaaaga	taaacttaga	ctcaaaacat	4020
ttacaaaaaac	aaccctctaa	gtcctaaagc	ccaaagtgtc	atgcacgac	catagcaagc	4080
ccagcccaac	ccaaccaca	ccaaccacc	ccagtgcagc	caactggcaa	atagtctcca	4140
cccccgccac	tatcaccggt	agttgtccgc	accaccgcac	gtctcgccgc	caaaaaaaaa	4200
aaaagaaaga	aaaaaaagaa	aaagaaaaaac	agcaggtggg	tcggggtcgt	gggggcccga	4260
aaagcagagga	ggatcgcgag	cagcgcagag	gcccggccct	ccctccgctt	ccaaagaaac	4320
gccccccatc	gccactatat	acataccccc	ccctctctct	ccatcccccc	aaccctacca	4380
ccaccaccac	caccacctcc	tccccctcgc	ctgccggacg	acgagctcct	ccccctccc	4440
cctccgcgcg	cgcgggtaac	caccccgccc	ctctctctct	tctttctcgc	tttttttttt	4500
cgtctcggtc	tcgatctttg	gccttggttag	tttgggtggg	cgagagcggc	ttcgtcgccc	4560
agatcggtgc	gcgggagggg	cgggatctcg	cggctggcgt	ctccgggcgt	gagtcggccc	4620
ggatcctcgc	ggggaatggg	gctctcggtg	gtagatcttc	tttctttctt	ctttttgtgg	4680

tagaatttga	atccctcagc	attgttcatc	ggtagttttt	cttttcatga	tttgtgacaa	4740
atgcagcctc	gtgcggagct	tttttgtagg	tag			4773

<210> 33

<211> 5034

<212> DNA

<213> Artificial Sequence

<220>

<223> pJQ5 vector

<400> 33

catggggccag	gtataattat	gggatatctc	aagcaaataa	tcgaaatata	accattggct	60
acaatatctg	agctccgagt	tctgactgca	gtctggatga	cgctgttgt	atctagaact	120
ctagatagca	cagccacagc	acctacagga	gtgcgacact	tgtggactgt	agtagtggtg	180
gagacggagc	tctttcctac	ctcctgacgt	tgccgccgtt	gtccattcca	acggcatcac	240
tctcaaccaa	tcacgcgctc	ccaacaaaat	atcgtccccc	atgtcttggc	ggagagagag	300
tacatacatg	ctgtcgcgcc	gtttttgtct	gaatctcgct	tccactggcc	aatcagctca	360
gctcccggga	gctcactcat	tcaagatccc	atcgtcgtcg	tcacccctgg	cgtcatggga	420
tggaaaagaa	cctccgttgc	tcggatgagt	cagccatata	cccgaacaga	gtactgcaag	480
ataacccaat	tcagattccc	ccaatagaga	aagtatagca	tgctttcggg	ttttgtttgg	540
cttaattgac	tttatttttg	ttggagttga	atgctgattt	gttgtgtaaa	atgcccaacc	600
atctgaatat	cgagacggat	aataggctgg	ctaattaatt	tatagcaaga	ttctgtagtg	660
cacatcgcaa	atatctttct	gggcattaca	gctggaggct	tcatacagct	gaaacactct	720
gcagagcctg	aagcaagtgg	tgaagcgtgg	cgatgagatg	ggtataaaac	ccccggcacc	780
gggacgcgag	ctcccgccta	ccagtaccat	ctcgcctcgc	tccccctgcc	ggacgaccca	840
gtaaaatact	gttgcccact	cgccggcgag	atggmcgtgc	acaaggaggt	saacttcgts	900
gcctacctcc	tgatcgtsct	cggcctcctc	ttgctcgtst	ccgccatgga	gcacgtggac	960
gccaaggcct	gcacccckcga	gtgcggcaac	ctcggcttcg	gcatctgccc	ggcggccgcc	1020
tccacgcagg	gcattctccga	agacctctac	agccgttttag	tcgaaatggc	cactatctcc	1080
caagctgcct	acgccgacct	gtgcaacatt	ccgtcgacta	ttatcaaggg	agagaaaatt	1140
tacaattctc	aaactgacat	taacggatgg	atcctccgcg	acgacagcag	caaagaaata	1200
atcacctgct	tccgtggcac	tggtagtgat	acgaatctac	aactcgatac	taactacacc	1260
ctcacgcctt	tcgacaccct	accacaatgc	aacggttgtg	aagtacacgg	tggatattat	1320
attggatggg	tctccgtcca	ggaccaagtc	gagtcgcttg	tcaaacagca	ggttagccag	1380
tatccggact	acgcgctgac	cgtgaccggc	cackccctcg	gcgcctccct	ggcggcactc	1440
actgccgccc	agctgtctgc	gacatacgac	aacatccgcc	tgtacacctt	cggcgaaccg	1500
cgcagcggca	atcaggcctt	cgcgtcgtac	atgaacgatg	ccttccaagc	ctcgagccca	1560
gatacgacgc	agtatttccg	ggtcactcat	gccaacgacg	gcatcccaaa	cctgcccccg	1620
gtggagcagg	ggtacgcccc	tggcgggtga	gagtactgga	gcgttgatcc	ttacagcgcc	1680
cagaacacat	ttgtctgcac	tggggatgaa	gtgcagtgct	gtgaggccca	gggcggacag	1740
ggtgtgaata	atgcgcacac	gacttatttt	gggatgacga	gcggagcctg	tacatggtga	1800
tcagtcattt	cagcctcccc	gagtgtacca	ggaaagatgg	atgtcctgga	gagggggccg	1860
cgtaaccact	gaaggatgag	ctgtaaagaa	gcagatcggt	caaacatttg	gcaataaagt	1920
ttcttaagat	tgaatcctgt	tgccgggtctt	gcgatgatta	tcatataatt	tctgttgaat	1980
tacgttaagc	atgtaataat	taacatgtaa	tgcatgacgt	tatttatgag	atgggttttt	2040
atgattagag	tcccgcgaatt	atacatttaa	tacgcgatag	aaaacaaaat	atagcgcgca	2100
aactaggata	aattatcgcg	cgcgggtgtca	tctatgttac	tagatcgata	agcttctaga	2160
gcggccgggtg	gagctccaat	tcgccctata	gtgagtcgta	ttacgcgcgc	tcactggccg	2220
tcgtttttaca	acgtcgtgac	tgggaaaacc	ctggcggttac	ccaacttaat	cgccttgcag	2280
cacatccccc	tttcgccagc	tggcgtaata	gcgaagaggc	ccgcaccgat	cgccttccc	2340
aacagttgcg	cagcctgaat	ggcgaatggg	acgcgccttg	tagcggcgca	ttaagcgcgg	2400
cgggtgtggt	ggttacgcgc	agcgtgaccg	ctacacttgc	cagcgcctta	gcgcccgcctc	2460
ctttcgcctt	cttcccttcc	tttctcgcca	cgttcgcggg	ctttccccgt	caagctctaa	2520
atcggggggct	cccttttaggg	ttccgattta	gtgcttttacg	gcacctcgac	cccaaaaaac	2580
ttgattaggg	tgatggttca	cgtagtgggc	catcgccctg	atagacggtt	tttcgccctt	2640
tgacgttgga	gtccacgttc	tttaatatgtg	gactcttggt	ccaaactgga	acaacactca	2700
accctatctc	ggtctattct	tttgatttat	aagggttttt	gccgatttctg	gcctattggt	2760
taaaaaatga	gctgatttaa	caaaaattta	acgcgaattt	taacaaaata	ttaacgctta	2820
caatttaggt	ggcacttttc	ggggaaatgt	gcgcggaacc	cctatttggt	tatttttcta	2880
aatacattca	aatatgtatc	cgtcatgag	acaataaccc	tgataaatgc	ttcaataata	2940

ttgaaaaagg	aagagtatga	gtattcaaca	tttccgtgtc	gcccttattc	cctttttttgc	3000
ggcattttgc	cttcctgttt	ttgctcacc	agaaacgctg	gtgaaagtaa	aagatgctga	3060
agatcagttg	ggtgcacgag	tgggttacat	cgaactggat	ctcaacagcg	gtaagatcct	3120
tgagagtttt	cgccccgaag	aacgttttcc	aatgatgagc	actttttaag	ttctgctatg	3180
tggcgcggtg	ttatcccgtg	ttgacgcggg	gcaagagcaa	ctcggtcgcc	gcatacacta	3240
ttctcagaat	gacttggttg	agtactcacc	agtcacagaa	aagcatctta	cggatggcat	3300
gacagtaaga	gaattatgca	gtgctgccat	aaccatgagt	gataacactg	cggccaactt	3360
acttctgaca	acgatcggag	gaccgaagga	gctaaccgct	tttttgcaca	acatggggga	3420
tcatgtaact	cgccttgatc	gttgggaacc	ggagctgaat	gaagccatac	caaacgacga	3480
gcgtgacacc	acgatgcctg	tagcaatggc	aacaacgctg	cgcaaactat	taactggcga	3540
actacttact	ctagcttccc	ggcaacaatt	aatagactgg	atggaggcgg	ataaagttgc	3600
aggaccactt	ctgcgctcgg	cccttcgggc	tggctggttt	attgctgata	aatctggagc	3660
cggtgagcgt	gggtctcgcg	gtatcattgc	agcactgggg	ccagatggta	agccctccc	3720
tatcgtagtt	atctacacga	cggggagtcg	ggcaactatg	gatgaacgaa	atagacagat	3780
cgtgagata	ggtgcctcac	tgattaagca	ttggtaactg	tcagaccaag	tttactcata	3840
tatactttag	attgatttaa	aacttcattt	ttaatttaaa	aggatctagg	tgaagatcct	3900
ttttgataat	ctcatgacca	aatccctta	acgtgagttt	tcgttccact	gagcgtcaga	3960
ccccgtagaa	aagatcaaag	gatcttcttg	agatcccttt	tttctgcgcg	taatctgctg	4020
cttgcaaaca	aaaaaaccac	cgtaccagc	ggtggtttgt	ttgccggatc	aagagctacc	4080
aactcttttt	ccgaaggtaa	ctggcttcag	cagagcgcag	ataccaaata	ctgtccttct	4140
agtgtagccg	tagttaggcc	accacttcaa	gaactctgta	gcaccgccta	catacctcgc	4200
tctgctaate	ctgttaccag	tggctgctgc	cagtggcgat	aagtcgtgtc	ttaccggggt	4260
ggactcaaga	cgatagttac	cggataaggc	gcagcggtcg	ggctgaacgg	gggggttcgtg	4320
cacacagccc	agcttggagc	gaacgaccta	caccgaactg	agatacctac	agcgtgagct	4380
atgagaaagc	gccacgcttc	ccgaaggagg	aaaggcggac	aggtatccgg	taagcggcag	4440
ggtcggaaca	ggagagcgca	cgaaggagct	tccaggggga	aacgcctggg	atctttatag	4500
tctgtcgggg	tttcgccacc	tctgacttga	gcgtcgattt	ttgtgatgct	cgtcaggggg	4560
gcggagccta	tggaaaaacg	ccagcaacgc	ggccttttta	cggttcctgg	ccttttgctg	4620
gccttttgct	cacatgttct	ttcctgcgtt	atccctgat	tctgtggata	accgtattac	4680
cgcctttgag	tgagctgata	ccgctcgccg	cagccgaacg	accgagcgca	gcgagtcagt	4740
gagcgaggaa	gcggaagagc	gcccataacg	caaaccgcct	ctccccgcgc	gttggccgat	4800
tcattaatgc	agctggcacg	acaggtttcc	cgactggaaa	gcgggcagtg	agcgcaacgc	4860
aattaatgtg	agttagctca	ctcattaggc	accccaggct	ttacacttta	tgcttcgggc	4920
tcgatatgtg	tgtggaattg	tgagcggata	acaatttcac	acaggaaaca	gctatgacca	4980
tgattacgcc	aagcgcgcaa	ttaaccctca	ctaaagggaa	caaagctgg	gtac	5034

<210> 34

<211> 4950

<212> DNA

<213> Artificial Sequence

<220>

<223> pJ06.1 vector

<400> 34

aagcttacca	tggcccacgc	ccgcgtcctc	ctcctggcgc	tcgccgtgct	ggccacggcc	60
gccgtcgccg	tcgcctcctc	ctcctccttc	gccgactcca	accgatccg	gcccgtcacc	120
gaccgcgcgg	ccgcctccac	gcagggcatc	tccgaagacc	tctacagccg	tttagtcgaa	180
atggccacta	tctcccaagc	tgcctacgcc	gacctgtgca	acattccgtc	gactattatc	240
aagggagaga	aaatttacaa	ttctcaaact	gacattaacg	gatggatcct	ccgcgacgac	300
agcagcaaag	aaataatcac	cgtcttccgt	ggcactggta	gtgatacgaa	tctacaactc	360
gatactaact	acaccctcac	gcctttcgac	accctaccac	aatgcaacgg	ttgtgaagta	420
cacgggtggat	attatatgtg	atgggtctcc	gtccaggacc	aagtcgagtc	gcttgtcaaa	480
cagcaggtta	gccagtatcc	ggactacgcg	ctgaccgtga	ccggccackc	cctcggcgcc	540
tccctggcgg	cactcactgc	cgcctcagctg	tctgcgacat	acgacaacat	ccgcctgtac	600
accttcggcg	aaccgcgcag	cggcaatcag	gccttcgcgt	cgtacatgaa	cgatgccttc	660
caagcctcga	gccagatac	gacgcagtat	ttccgggtca	ctcatgccaa	cgacggcatc	720
ccaaacctgc	ccccgggtga	gcaggggtac	gcccattggc	gtgtagagta	ctggagcgtt	780
gatccttaca	gcgcccagaa	cacatttgtc	tgcactgggg	atgaagtgca	gtgctgtgag	840
gcccagggcg	gacaggggtg	gaataatgcg	cacacgactt	atcttgggat	gacgagcggc	900
gcattgcacct	ggccgggtcg	ggccgcggaa	accactgaag	gatgagctgt	aaagaagcag	960

atcgttcaaa	catttgga	taaagtttct	taagattgaa	tcctggtgcc	ggtcttgcca	1020
tgattatcat	ataatttctg	ttgaattacg	ttaagcatgt	aataattaac	atgtaatgca	1080
tgacgttatt	tatgagatgg	gtttttatga	ttagagtccc	gcaattatac	atttaatacg	1140
cgatagaaaa	caaaatatag	cgcgcgcaact	aggataaatt	atcgcgcgcg	gtgtcatcta	1200
tgttactaga	tcgataagct	tctagagcgg	ccggtggagc	tccaattcgc	cctatagtga	1260
gtcgtattac	gcgcgctcac	tggccgctcgt	tttacaacgt	cgtgactggg	aaaaccctgg	1320
cgttacccaa	cttaatcgcc	ttgcagcaca	tcccccttcc	gccagctggc	gtaatagcga	1380
agaggccccg	accgatcgcc	cttcccaaca	gttgccgcagc	ctgaatggcg	aatgggacgc	1440
gccctgtagc	ggcgcattaa	gcgcggcggg	tgtggtgggt	acgcgcagcg	tgaccgctac	1500
acttgccagc	gccctagcgc	ccgctccttt	cgctttcttc	ccttccttcc	tcgccacggt	1560
cgccggcttt	ccccgtcaag	ctctaaatcg	ggggctccct	ttagggttcc	gatttagtgc	1620
tttacggcac	ctcgacccca	aaaaacttga	ttagggtgat	ggttcacgta	gtgggccatc	1680
gccctgatag	acgggtttttc	gccctttgac	gttgagatcc	acgttcttta	atagtggact	1740
cttgttccaa	actggaacaa	cactcaaccc	tatctcggtc	tattcttttg	atttataagg	1800
gattttgccg	atttcggcct	attggttaaa	aaatgagctg	atttaacaaa	aatttaacgc	1860
gaattttaac	aaaatattaa	cgcttacaat	ttagggtggca	cttttcgggg	aaatgtgcgc	1920
ggaaccccta	tttgtttatt	tttctaaata	cattcaaata	tgtatccgct	catgagacaa	1980
taaccctgat	aaatgcttca	ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	2040
cgtgtcgccc	ttattccctt	ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaa	2100
acgctggtga	aagtaaaaga	tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	2160
ctggatctca	acagcggtaa	gaccttgag	agttttcgcc	ccgaagaacg	ttttccaatg	2220
atgagcactt	ttaaagtctt	gctatgtggc	gcggtattat	cccgatttga	cgccgggcaa	2280
gagcaactcg	gtcgccgcct	acactattct	cagaatgact	tggttgagta	ctcaccagtc	2340
acagaaaagc	atcttacgga	tggcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	2400
atgagtgata	acactgcggc	caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	2460
accgcttttt	tgcacaacat	gggggatcat	gtaactcgcc	ttgatcggtg	ggaaccggag	2520
ctgaatgaag	ccataccaaa	cgacgagcgt	gacaccacga	tgctgtagc	aatggcaaca	2580
acgttgcgca	aactattaac	tggcgaacta	cttactctag	cttcccggca	acaattaata	2640
gactggatgg	aggcggataa	agttgcagga	ccacttctgc	gctcggccct	tccggctggc	2700
tggttttattg	ctgataaate	tggagccggg	gagcgtgggt	ctcgcggtat	cattgcagca	2760
ctggggccag	atggtaagcc	ctcccgtatc	gtagttatct	acacgacggg	gagtcaggca	2820
actatggatg	aacgaaatag	acagatcgct	gagataggtg	cctcactgat	taagcattgg	2880
taactgtcag	accaagttta	ctcatatata	cttttagattg	atttaaaact	tcatttttaa	2940
tttaaaagga	tctaggtgaa	gacctttttt	gataatctca	tgaccaaaat	cccttaacgt	3000
gagttttcgt	tccactgagc	gtcagacccc	gtagaaaaga	tcaaaggatc	ttcttgagat	3060
ccttttttttc	tgcgcgtaat	ctgctgcttg	caaacaaaaa	aaccaccgct	accagcgggtg	3120
gtttgtttgc	cggatcaaga	gctaccaact	ctttttccga	aggtaactgg	cttcagcaga	3180
gcgcagatac	caaatactgt	ccttctagt	tagccgtagt	taggccacca	cttcaagaac	3240
tctgtagcac	cgcctacata	cctcgctctg	ctaactcctgt	taccagtggc	tgctgccagt	3300
ggcgataagt	cgtgtcttac	cgggttggac	tcaagacgat	agttaccgga	taaggcgcag	3360
cggtcggggt	gaacgggggg	ttcgtgcaca	cagcccagct	tggagcgaac	gacctacacc	3420
gaactgagat	acctacagcg	tgagctatga	gaaagcgcca	cgcttcccga	agggagaaag	3480
gcggacaggt	atccggtaag	cggcaggggtc	ggaacaggag	agcgcacgag	ggagcttcca	3540
gggggaaacg	cctggtatct	ttatagtcct	gtcgggtttc	gccacctctg	acttgagcgt	3600
cgattttttgt	gatgctcgtc	agggggggcg	agcctatgga	aaaacgccag	caacgcggcc	3660
tttttacggg	tcctggcctt	ttgctggcct	tttgcctaca	tgttctttcc	tgctgttatcc	3720
cctgattctg	tggataaccg	tattaccgcc	tttgagttag	ctgataccgc	tcgccgcagc	3780
cgaacgaccg	agcgcagcga	gtcagttagc	gaggaagcgg	aagagcgccc	aatacgcaaa	3840
ccgcctctcc	ccgcgcgttg	gccgattcat	taatgcagct	ggcacgacag	gtttcccgcac	3900
tggaaagcgg	gcagttagcg	caacgcgaatt	aatgtgagtt	agctcactca	ttaggcaccc	3960
caggcttttac	actttatgct	tccggctcgt	atgttgtgtg	gaattgtgag	cggataacaa	4020
tttcacacag	gaaacagcta	tgaccatgat	tacgccaaagc	gcgcaattaa	ccctcactaa	4080
agggaacaaa	agctggaatt	ccacaatgaa	caataataag	attaaaatag	cttgcccccg	4140
ttgcagcgat	gggtattttt	tctagtaaaa	taaaagataa	acttagactc	aaaacattta	4200
caaaaacaac	ccctaaagtc	ctaaagccca	aagtgttatg	cacgatccat	agcaagccca	4260
gcccacccca	acccaaccca	acccacccca	gtgcagccaa	ctggcaataa	gtctccaccc	4320
ccggcactat	caccgtgagt	tgtccgcacc	accgcacgtc	tcgcagccaa	aaaaaaaaaa	4380
agaaagaaaa	aaaagaaaaa	gaaaaacagc	aggtgggtcc	gggtcgtggg	ggccggaaaa	4440
gcgaggagga	tcgcgagcag	cgacgaggcc	cggccctccc	tccgcttcca	aagaaacgcc	4500
ccccatcgcc	actatataca	tacccccccc	tctcctccca	tccccccaac	cctaccacca	4560
ccaccaccac	cacctcctcc	ccctcgtctg	ccggacgacg	agctcctccc	ccctccccct	4620

cgcgcgcgcg	cggtaaccac	cccgcccctc	tcctctttct	ttctccggtt	tttttttcgt	4680
ctcgggtctg	atctttggcc	ttggtagttt	gggtgggcga	gagcggcttc	gtcgcccaga	4740
tcgggtgcgcg	ggaggggcgg	gatctgcgcg	ctggcgtctc	cgggcgtgag	tcggcccggg	4800
tcctcgcgcgg	gaatggggct	ctcggatgta	gatcttcttt	ctttcttctt	tttgtggtag	4860
aatttgaatc	cctcagcatt	gttcacgcgt	agtttttctt	ttcatgattt	gtgacaaatg	4920
cagcctcgtg	cggagctttt	ttgtaggtag				4950

<210> 35

<211> 4974

<212> DNA

<213> Artificial Sequence

<220>

<223> pJQ4 vector

<400> 35

aattccacaa	tgaacaataa	taagattaaa	atagcttgcc	cccgttgccg	cgatgggtat	60
tttttctagt	aaaataaaag	ataaacttag	actcaaaaaca	tttacaaaaa	caacccttaa	120
agtcctaaag	cccaaagtgc	tatgcacgat	ccatagcaag	cccagcccaa	cccaacccaa	180
cccaaccac	cccagtgcag	ccaactggca	aatagtctcc	accccgggca	ctatcacctg	240
gagttgtccg	caccaccgca	cgtctcgcag	ccaaaaaaa	aaaaagaaag	aaaaaaaaga	300
aaaagaaaaa	cagcaggtgg	gtccgggtcg	tggggggcgg	aaaagcagag	aggatcgcga	360
gcagcgacga	ggcccggccc	tcctcgcgt	tccaaagaaa	cgccccccat	cgccactata	420
tacatacccc	cccctctcct	cccatccccc	caaccctacc	accaccacca	ccaccacctc	480
ctccccctc	gctgccggac	gacgagctcc	tcctccctcc	ccctccgcgc	ccgccggtaa	540
ccaccgccgc	cctctcctct	ttctttctcc	gttttttttt	tcgtctcggg	ctcgatcttt	600
ggccttggtg	gtttgggtgg	gcgagagcgg	cttcgtcgcg	cagatcgggt	cgcgggaggg	660
gcgggatctc	gcggctggcg	tctccgggcg	tgagtcggcc	cggatcctcg	cggggaatgg	720
ggctctcgga	tgtagatctt	ctttctttct	tctttttgtg	gtagaatttg	aatccctcag	780
cattgttcat	cggtagtttt	tcttttcatg	atttgtgaca	aatgcagcct	cgtgcggagc	840
ttttttgtag	gtagaagctt	acmatggmcg	tgcacaagga	ggtsaacttc	gtsgcctacc	900
tcctgatcgt	scctcgccct	ctcttgctcg	tstccgccat	ggagcacgtg	gacgccaagg	960
cctgcaccck	cgagtgcggc	aacctcggtt	tcggcatctg	cccggcggcc	gcctccacgc	1020
agggcatctc	cgaagacctc	tacagccggt	tagtcgaaat	ggccactatc	tcceaagctg	1080
cctacgcgga	cctgtgcaac	attccgtcga	ctattatcaa	gggagagaaa	atttacaatt	1140
ctcaaaactga	cattaacgga	tggatccctc	gcgacgacag	cagcaaagaa	ataatcaccg	1200
tcttcgcgtg	cactggtagt	gatacgaatc	tacaactcga	tactaactac	accctcacgc	1260
ctttcgcacac	cctaccacaa	tgcaacgggt	gtgaagtaca	cgggtggatat	tatattggat	1320
gggtctccgt	ccaggaccaa	gtcagatcgc	ttgtcaaaca	gcagggttagc	cagtatccgg	1380
actacgcgct	gaccgtgacc	ggccackccc	tcggcgcctc	cctggcggca	ctcactgccg	1440
cccagctgtc	tgcgacatac	gacaacatcc	gcctgtacac	cttcggcgaa	ccgcgcagcg	1500
gcaatcaggg	cttcgcgtcg	tacatgaacg	atgccttcca	agcctcgagc	ccagatacga	1560
cgcagtattt	ccgggtcact	catgccaacg	acggcatccc	aaacctgccc	ccggtggagc	1620
aggggtacgc	ccatggcggg	gtagagtact	ggagcgttga	tccttacagc	gcccagaaca	1680
catttgtctg	cactggggat	gaagtgcagt	gctgtgaggc	ccagggcgga	caggggtgtga	1740
ataatgcgca	cacgacttat	tttgggatga	cgagcggcgc	atgcacctgg	ccggtcgcgg	1800
ccgcggaaac	cactgaagga	tgagctgtaa	agaagcagat	cgttcaaaca	tttggcaata	1860
aagtttctta	agattgaatc	ctgttgccgg	tcttgcgatg	attatcatat	aatttctgtt	1920
gaattacgtt	aagcatgtaa	taattaacat	gtaatgcagt	acgttattta	tgagatgggt	1980
ttttatgatt	agagtcgccg	aattatacat	ttaatacgcg	atagaaaaca	aaatatagcg	2040
cgcaaaactag	gataaattat	cgcgcgcggg	gtcatctatg	ttactagatc	gataagcttc	2100
tagagcggcc	ggtggagctc	caattcgccc	tatagttagt	cgtattacgc	gcgctcactg	2160
gccgtcgttt	tacaacgtcg	tgactgggaa	aacctggcg	ttacccaact	taatcgcttt	2220
gcagcacatc	cccctttcgc	cagctggcgt	aatagcgaag	aggcccgcac	cgatcgccct	2280
tcccaacagt	tgcgcagcct	gaatggcgaa	tgggacgcgc	cctgtagcgg	cgcattaagc	2340
gcggcgggtg	tgggtggttac	gcgcagcgtg	accgctacac	ttgccagcgc	cctagcgccc	2400
gctcctttcg	ctttcttccc	ttcctttctc	gccacgttcg	ccggctttcc	ccgtcaagct	2460
ctaaatcggg	ggctcccttt	agggttccga	tttagtgctt	tacggcacct	cgaccccaaa	2520
aaacttgatt	aggggtgatg	ttcacgtagt	gggccatcgc	cctgatagac	ggtttttcgc	2580
cctttgacgt	tggagtcac	gttctttaat	agtggactct	tgttccaaac	tggaaacaaca	2640
ctcaacccta	tctcgggtcta	ttcttttgat	ttataaggga	ttttgccgat	ttcggcctat	2700

tggttaaaaa	atgagctgat	ttaacaaaaa	tttaacgcga	attttaacaa	aatattaacg	2760
cttacaattt	aggtggcact	tttcggggaa	atgtgcgcgg	aacccttatt	tgtttatttt	2820
tctaaataca	ttcaaataatg	tatccgctca	tgagacaata	accctgataa	atgcttcaat	2880
aatattgaaa	aaggaagagt	atgagtatct	aacatttccg	tgtcgcctt	attccctttt	2940
ttgcggcatt	ttgccttcct	gtttttgctc	acccagaaac	gctggtgaaa	gtaaaagatg	3000
ctgaagatca	gttgggtgca	cgagtgggtt	acatcgaaact	ggatctcaac	agcggtaaga	3060
tccttgagag	ttttcgcccc	gaagaacgtt	ttccaatgat	gagcactttt	aaagttctgc	3120
tatgtggcgc	ggtattatcc	cgtattgacg	ccgggcaaga	gcaactcggg	cgccgcatac	3180
actatttctca	gaatgacttg	gttgagtact	caccagtcac	agaaaagcat	cttacggatg	3240
gcatgacagt	aagagaatta	tgcagtgctg	ccataaccat	gagtataaac	actgcggcca	3300
acttacttct	gacaacgac	ggaggaccga	aggagctaac	cgcttttttg	cacaacatgg	3360
gggatcatgt	aactcgcctt	gatcgttggg	aaccggagct	gaatgaagcc	ataccaaacg	3420
acgagcgtga	caccacgatg	cctgtagcaa	tggcaacaac	gttgcgcaaa	ctattaactg	3480
gcgaactact	tactctagct	tcccggcaac	aattaataga	ctggatggag	gcggataaag	3540
ttgcaggacc	acttctgcgc	tcggcccttc	cggctggctg	gtttattgct	gataaatctg	3600
gagccggtga	gcgtgggtct	cgcggtatca	ttgcagcact	ggggccagat	ggtaagccct	3660
cccgtatcgt	agttatctac	acgacgggga	gtcaggcaac	tatggatgaa	cgaaatagac	3720
agatcgctga	gatagggtgcc	tcactgatta	agcattggta	actgtcagac	caagtttact	3780
catatatact	ttagattgat	ttaaaacttc	atttttaatt	taaaaggatc	taggtgaaga	3840
tcctttttga	taatctcatg	acaaaaatcc	cttaacgtga	gttttcgttc	cactgagcgt	3900
cagaccccg	agaaaagatc	aaaggatctt	cttgagatcc	tttttttctg	cgcgtaatct	3960
gctgcttgca	aacaaaaaaa	ccaccgctac	cagcgggtgg	ttgtttgccc	gatcaagagc	4020
taccaactct	ttttccgaag	gtaactggct	tcagcagagc	gcagatacca	aatactgtcc	4080
ttctagtgt	gccgtagtta	ggccaccact	tcaagaactc	tgtagcaccg	cctacatacc	4140
tcgctctgct	aatcctgtta	ccagtggctg	ctgccagtgg	cgataagtcg	tgtcttaccg	4200
ggttggactc	aagacgatag	ttaccggata	aggcgcagcg	gtcgggctga	acgggggggtt	4260
cgtgcacaca	gcccagcttg	gagcgaacga	cctacaccga	actgagatac	ctacagcgtg	4320
agctatgaga	aagcgcacag	cttcccgaag	ggagaaaggc	ggacaggtat	ccggtaagcg	4380
gcagggtcgg	aacaggagag	cgcacgaggg	agcttccagg	gggaaacgcc	tggatatctt	4440
atagtcctgt	cgggtttcgc	cacctctgac	ttgagcgtcg	atttttgtga	tgctcgtcag	4500
gggggcggag	cctatggaaa	aacgccagca	acgcggcctt	tttacggttc	ctggcctttt	4560
gctggccttt	tgtctacatg	ttctttcctg	cgttatcccc	tgattctgtg	gataaccgta	4620
ttaccgcctt	tgagtgagct	gataccgctc	gccgcagccg	aacgaccgag	cgcagcaggt	4680
cagtgagcga	ggaagcggaa	gagcgcacca	tacgcaaacc	gcctctcccc	gcgcgttggc	4740
cgattcatta	atgcagctgg	cacgacaggt	ttcccgaactg	gaaagcgggc	agtgagcgca	4800
acgcaattaa	tgtgagttag	ctcactcatt	aggcacccca	ggctttacac	tttatgcttc	4860
cggctcgtat	gttgtgtgga	atttgtgagcg	gataacaatt	tcacacagga	aacagctatg	4920
accatgatta	cgccaagcgc	gcaattaacc	ctcactaaag	ggaacaaaag	ctgg	4974

<210> 36

<211> 5164

<212> DNA

<213> Artificial Sequence

<220>

<223> pPQ10.1 vector

<400> 36

aagcttacca	tgggtgagcaa	gggcgaggag	ctgttcaccg	gggtgggtgcc	catcctgggtc	60
gagctggacg	gcgacgtgaa	cggccacaag	ttcagcgtgt	ccggcgaggg	cgaggggcgat	120
gccacctacg	gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	gcccgtgccc	180
tggcccaccc	tcgtgaccac	cttcacctac	ggcgtgcagt	gcttcagccg	ctaccccgac	240
cacatgaagc	agcagcactt	cttcaagtcc	gccatgcccc	aaggctacgt	ccaggagcgc	300
accatcttct	tcaaggacga	cggcaactac	aagaccgcg	ccgaggtgaa	gttcgagggc	360
gacaccctgg	tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	cggcaacatc	420
ctggggcaca	agctggagta	caactacaac	agccacaacg	tctatatcat	ggccgacaag	480
cagaagaacg	gcatcaaggt	gaacttcaag	atccgccaca	acatcgagga	cggcagcgtg	540
cagctcgccg	accactacca	gcagaacacc	cccatcgggc	acggccccgt	gctgctgccc	600
gacaaccact	acctgagcac	ccagtccgcc	ctgagcaaag	accccaacga	gaagcgcgat	660
cacatgggtcc	tgctggagtt	cgtgaccgcc	gccgggatca	ctcacggcat	ggacgagctg	720
tacaagtaaa	gcggccgccc	gggctgcagg	gaaaccactg	aaggatgagc	tgtaaagaag	780

cagatcgttc	aaacatttgg	caataaagtt	tcttaagatt	gaatcctggt	gccggtcttg	840
cgatgattat	catataat	ctgttgaatt	acgttaagca	tgtaataatt	aacatgtaat	900
gcacgacgtt	atztatgaga	tgggttttta	tgattagagt	cccgaatta	tacatttaat	960
acgcgataga	aaacaaaata	tagcgcgcaa	actaggataa	attatcgcg	gcggtgtcat	1020
ctatgttact	agatcgataa	gcttctagag	cggccggtgg	agctccaatt	cgccctatag	1080
tgagtcgtat	tacgcgcgct	cactggccgt	cgttttacaa	cgctcgtgact	gggaaaaccc	1140
tggcgttacc	caacttaate	gccttgcagc	acatccccct	ttcgccagct	ggcgtaatag	1200
cgaagaggcc	cgcaccgac	gcccttccca	acagttgcg	agcctgaatg	gcgaatggga	1260
cgcgccctgt	agcggcgcat	taagcgcggc	gggtgtggtg	gttacgcgca	gcgtgaccgc	1320
tacacttgcc	agcgccctag	cgcccgctcc	tttcgctttc	ttcccttcc	ttctcgccac	1380
gttcgcccgg	tttccccgtc	aagctctaaa	tggggggctc	cctttagggt	tccgatttag	1440
tgcctttacg	cacctcgacc	ccaaaaaact	tgattagggt	gatggttcac	gtagtggg	1500
atcgccctga	tagacgggtt	ttcgcccttt	gacgttggag	tccacgttct	ttaatagtgg	1560
actcttgttc	caactggaa	caacactcaa	ccctatctcg	gtctattctt	ttgatttata	1620
agggattttg	ccgatttcgg	cctattgggt	aaaaaatgag	ctgatttaac	aaaaatttaa	1680
cgcgaatttt	aacaaaatat	taacgccttac	aatttaggtg	gcacttttcg	gggaaatgtg	1740
cgcggaaccc	ctatttggtt	atttttctaa	atacattcaa	atatgtatcc	gctcatgaga	1800
caataaccct	gataaatgct	tcaataatat	tgaaaaagga	agagtatgag	tattcaacat	1860
ttcctgtctg	cccttattcc	cttttttgcg	gcattttgcc	ttcctgtttt	tgctcaccca	1920
gaaacgctgg	tgaaagtaaa	agatgctgaa	gatcagttgg	gtgcacgagt	gggttacatc	1980
gaactggatc	tcaacagcgg	taagatcctt	gagagttttc	gccccgaaga	acgttttcca	2040
atgatgagca	cttttaaaagt	tctgctatgt	ggcgcggtat	tatcccgat	tgacgcccgg	2100
caagagcaac	tgggtcgccg	catacactat	tctcagaatg	acttggttga	gtactcacca	2160
gtcacagaaa	agcatcttac	ggatggcatg	acagtaagag	aattatgcag	tgctgccata	2220
accatgagtg	ataaactgc	ggccaactta	cttctgacaa	cgatcggagg	accgaaggag	2280
ctaaccgctt	ttttgcacaa	catgggggat	catgtaactc	gccttgatcg	ttgggaaccg	2340
gagctgaatg	aagccatacc	aaacgacgag	cgtgacacca	cgatgcctgt	agcaatggca	2400
acaacgttgc	gcaaactatt	aactggcgaa	ctacttactc	tagcttccc	gcaacaatta	2460
atagactgga	tggaggcgga	taaagttagc	ggaccacttc	tgcgctcggc	ccttccggct	2520
ggctggttta	ttgctgataa	atctggagcc	ggtgagcgtg	ggtctcgcg	tatcattgca	2580
gcaactatgg	atgaacgaaa	tagacagatc	gctgagatag	gtgcctcact	gattaagcat	2640
tggtaaactgt	cagaccaagt	ttactcatat	atactttaga	ttgatttaaa	acttcatttt	2700
taatttaaaa	ggatctaggt	gaagatcctt	tttgataatc	tcatgaccaa	aatcccttaa	2760
cgtgagtttt	cgttccactg	agcgtcagac	cccgtagaaa	agatcaaagg	atcttcttga	2820
gatccttttt	ttctgcgcgt	aatctgctgc	ttgcaaacaa	aaaaaccacc	gctaccagcg	2880
gtggtttgtt	tgcgggatca	agagctacca	actctttttc	cgaaggtaac	tggtttcagc	2940
agagcgcaga	taccaaatac	tgtccttcta	gtgtagccgt	agttaggcca	ccacttcaag	3000
aactctgtag	caccgcctac	atacctcgct	ctgctaatac	tgttaccagt	ggctgctgcc	3060
agtggcgata	agtctgtctt	taccgggttg	gactcaagac	gatagttacc	ggataaggcg	3120
cagcggtcgg	gctgaacggg	gggttcgtgc	acacagccca	gcttggagcg	aacgacctac	3180
accgaactga	gatacctaca	gcgtgagcta	tgagaaagcg	ccacgcttcc	cgaagggaga	3240
aaggcggaca	ggtatccggg	aagcggcagg	gtcggaaacag	gagagcgcac	gagggagctt	3300
ccagggggaa	acgcctggta	tctttatagt	cctgtcgggt	ttcgccacct	ctgacttgag	3360
cgtcgatttt	tgtgatgctc	gtcagggggg	cggagcctat	ggaaaaacgc	cagcaacgcg	3420
gcctttttac	ggttcctggc	cttttgctgg	ccttttgctc	acatgttctt	tcctgcgtta	3480
tcccttgatt	ctgtggataa	ccgtattacc	gcctttgagt	gagctgatac	cgctcgccgc	3540
agccgaacga	ccgagcgcag	cgagtcagtg	agcgagggaag	cggaagagcg	cccaatacgc	3600
aaaccgcctc	tccccgcgcg	ttggccgatt	cattaatgca	gctggcacga	caggtttccc	3660
gactggaaaag	cgggcagtg	gcgcaacgca	attaatgtga	gttagctcac	tcattaggca	3720
ccccaggctt	tacactttat	gcttccggct	cgtatgttgt	gtggaattgt	gagcggataa	3780
caatttcaca	caggaaacag	ctatgaccat	gattacgcca	agcgcgcaat	taaccctcac	3840
taaagggaac	aaaagctggg	taccgggccc	cccctcgagg	tcattcatat	gcttgagaag	3900
agagtcggga	tagtccaaaa	taaaacaaag	gtaagattac	ctggtcaaaa	gtgaaaacat	3960
cagttaaaaag	gtggtataag	taaaatatcg	gtaataaaaag	gtggcccaaa	gtgaaattta	4020
ctcttttcta	ctattataaa	aattgaggat	gttttgctcg	tactttgata	cgctattttt	4080
gtatgaattg	gttttttaagt	ttattcgcg	tttggaatg	catatctgta	tttgagtcgg	4140
tttttaagtt	cgttgctttt	gtaaatacag	agggatttgt	ataagaaata	tctttaaaaa	4200
acccatatgc	taatttgaca	taatttttga	gaaaaatata	tattcaggcg	aattccacaa	4260
tgaacaataa	taagattaaa	atagcttgcc	cccggttgag	cgatgggtat	tttttctagt	4320
aaaataaaaag	ataaacttag	actcaaaaaca	tttacaacaaa	caacccttaa	agtcctaaag	4380
						4440

cccaaagtgc	tatgcacgat	ccatagcaag	cccagcccaa	cccaacccaa	cccaacccac	4500
cccagtgcag	ccaactggca	aatagtctcc	acccccggca	ctatcacctg	gagttgtccg	4560
caccaccgca	cgtctcgcag	ccaaaaaaaa	aaaaagaaag	aaaaaaaaaga	aaaagaaaaa	4620
cagcaggtgg	gtccgggtcg	tggggggccgg	aaaagcgagg	aggatcgcg	gcagcgacga	4680
ggccccggccc	tccctccgct	tccaaagaaa	cgccccccat	cgccactata	tacatacccc	4740
ccccctctcct	cccatcccc	caaccctacc	accaccacca	ccaccacctc	ctccccctc	4800
gctgccggac	gacgagctcc	tccccctctc	ccctccgccc	ccgccggtaa	ccacccccgc	4860
cctctcctct	ttctttctcc	gttttttttt	togtctcggt	ctcgatcttt	ggccttggt	4920
gtttgggtgg	gcgagagcgg	cttcgtcgcc	cagatcggtg	cgcgggagg	gcgggatctc	4980
gcggctggcg	tctccgggcg	tgagtcggcc	cggatcctcg	cggggaatgg	ggctctcgga	5040
tgtagatctt	ctttctttct	tctttttgtg	gtagaatttg	aatccctcag	cattgttcat	5100
cggtagtttt	tcttttcatg	atttgtgaca	aatgcagcct	cgtgccggagc	ttttttgtag	5160
gtag						5164

<210> 37

<211> 4965

<212> DNA

<213> Artificial Sequence

<220>

<223> pJQ3 vector

<400> 37

aattccacaa	tgaacaataa	taagattaaa	atagcttgcc	cccgttgcag	cgatgggtat	60
tttttctagt	aaaataaaag	ataaacttag	actcaaaaca	tttacaaaaa	caacccttaa	120
agtcctaaag	cccaaagtgc	tatgcacgat	ccatagcaag	cccagcccaa	cccaacccaa	180
cccaacccac	cccagtgcag	ccaactggca	aatagtctcc	acccccggca	ctatcacctg	240
gagttgtccg	caccaccgca	cgtctcgcag	ccaaaaaaaa	aaaaagaaag	aaaaaaaaaga	300
aaaagaaaaa	cagcaggtgg	gtccgggtcg	tggggggccgg	aaaagcgagg	aggatcgcg	360
gcagcgacga	ggccccggccc	tccctccgct	tccaaagaaa	cgccccccat	cgccactata	420
tacatacccc	ccccctctcct	cccatcccc	caaccctacc	accaccacca	ccaccacctc	480
ctccccctc	gctgccggac	gacgagctcc	tccccctctc	ccctccgccc	ccgccggtaa	540
ccacccccgc	cctctcctct	ttctttctcc	gttttttttt	togtctcggt	ctcgatcttt	600
ggccttggt	gtttgggtgg	gcgagagcgg	cttcgtcgcc	cagatcggtg	cgcgggagg	660
gcgggatctc	gcggctggcg	tctccgggcg	tgagtcggcc	cggatcctcg	cggggaatgg	720
ggctctcgga	tgtagatctt	ctttctttct	tctttttgtg	gtagaatttg	aatccctcag	780
cattgttcat	cggtagtttt	tcttttcatg	atttgtgaca	aatgcagcct	cgtgccggagc	840
ttttttgtag	gtagaagctt	accatgatcc	acaccaacct	caaaaagaag	ttctccctct	900
tcctcctcgt	cttcctcctc	ttcgccgtga	tctgcgtgtg	gaagaagggc	tccgactacg	960
aggccctcac	cctccaagcc	aaggagttec	aaatggcggc	cgctccacg	cagggcatct	1020
ccgaagacct	ctacagccgt	ttagtcgaaa	tggccactat	ctcccaagct	gcctacgccg	1080
acctgtgcaa	cattccgtcg	actattatca	agggagagaa	aattttacaat	tctcaaactg	1140
acattaacgg	atggatcctc	cgcgacgaca	gcagcaaaga	aataatcacc	gtcttccgtg	1200
gcactggtag	tgatacgaat	ctacaactcg	atactaacta	caccctcacg	cctttcgaca	1260
ccctaccaca	atgcaacggg	tgtgaagtac	acgggtggata	ttatattgga	tgggtctccg	1320
tccaggacca	agtcgagtcg	cttgtcaaac	agcagggttag	ccagtatccg	gactacgcgc	1380
tgaccgtgac	cggccackcc	ctcggcgcct	ccctggcggc	actcactgcc	gccagctgt	1440
ctgcgacata	cgacaacatc	cgcctgtaca	ccttcggcga	accgcgcagc	ggcaatcagg	1500
ccttcgcgtc	gtacatgaac	gatgccttcc	aagcctcgag	cccagatacg	acgcagtatt	1560
tccgggtcac	tcatgccaac	gacggcatcc	caaacctgcc	cccgggtggag	caggggtacg	1620
cccatggcgg	tgtagagtac	tggagcgttg	atccttacag	cgcccagaac	acatttgtct	1680
gcactgggga	tgaagtgcag	tgctgtgagg	cccagggcgg	acaggggtgtg	aataatgcgc	1740
acacgactta	ttttgggatg	acgagcggcg	catgcacctg	gccggtcgcg	gccgcggaaa	1800
ccactgaagg	atgagctgta	aagaagcaga	togttcaaac	atttggcaat	aaagtctctt	1860
aagattgaat	cctgttgccg	gtcttgcgat	gattatcata	taatttctgt	tgaattacgt	1920
taagcatgta	ataattaaca	tgtaatgcat	gacgttat	atgagatggg	tttttatgat	1980
tagagtccc	caattataca	tttaatacgc	gatagaaaac	aaaatatagc	gcgcaaacta	2040
ggataaatta	tcgcgcgcgg	tgtcatctat	gttactagat	cgataagctt	ctagagcggc	2100
cgggtggagct	ccaattcgcc	ctatagtgag	togtattacg	cgcgctcact	ggccgtcgtt	2160
ttacaacgct	gtgactggga	aaacctggc	gttaccacaac	ttaatcgctt	tgcagcacat	2220
ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	ccgatcgccc	ttcccaacag	2280

ttgcgcagcc	tgaatggcga	atgggacgcg	ccctgtagcg	gcgcattaag	cgcggcgggt	2340
gtggtggtta	cgcgacgcgt	gaccgctaca	cttgccagcg	ccctagcgcc	cgctccttct	2400
gctttcttcc	cttcctttct	cgccacgttc	gccggctttc	cccgtcaagc	tctaaatcgg	2460
gggctccctt	taggggttccg	atttagtgct	ttacggcacc	tcgaccccaa	aaaacttgat	2520
taggggtgatg	gttcacgtag	tggggccatcg	ccctgataga	cggttttttcg	ccctttgacg	2580
ttggagtcca	cgttctttta	tagtggactc	ttgttccaaa	ctggaacaac	actcaaccct	2640
atctcgggtct	attctttttga	tttataaggg	attttgccga	tttcggccta	ttgggttaaaa	2700
aatgagctga	tttaacaaaa	atttaacgcg	aattttaaca	aaatattaac	gcttacaatt	2760
taggtggcac	ttttcgggga	aatgtgcgcg	gaacccctat	ttgtttatatt	ttctaaatac	2820
attcaaatat	gtatccgctc	atgagacaat	aaccctgata	aatgcttcaa	taatattgaa	2880
aaaggaagag	tatgagtatt	caacatttcc	gtgtcgccct	tattcccttt	tttgcggcat	2940
tttgcccttcc	tgtttttgct	cacccagaaa	cgctgggtgaa	agtaaaagat	gctgaagatc	3000
agttgggtgc	acgagtgggt	tacatcgaa	tggatctcaa	cagcggtaag	atccttgaga	3060
gttttcgccc	cgaagaacgt	tttccaatga	tgagcacttt	taaagttctg	ctatgtggcg	3120
cggattatct	ccgtattgac	gccgggcaag	agcaactcgg	tcgccgcata	cactattctc	3180
agaatgactt	ggttgagtac	tcaccagtca	cagaaaagca	tcttacggat	ggcatgacag	3240
taagagaatt	atgcagtgct	gccataacca	tgagtgataa	cactgcggcc	aacttacttc	3300
tgacaacgat	cggaggaccg	aaggagctaa	ccgctttttt	gcacaacatg	ggggatcatg	3360
taactcgcct	tgatcgttgg	gaaccggagc	tgaatgaagc	cataccaaac	gacgagcgtg	3420
acaccacgat	gcctgtagca	atggcaacaa	cgttgcgcaa	actattaact	ggcgaactac	3480
ttactctagc	ttcccggcaa	caattaatag	actggatgga	ggcggataaa	gttgcaggac	3540
cacttctgcg	ctcggccctt	ccggctggct	ggtttattgc	tgataaatct	ggagccgggtg	3600
agcgtgggtc	tcgcgggtatc	attgcagcac	tggggccaga	tggttaagccc	ttccgtatcg	3660
tagttatcta	cacgacgggg	agtcaggcaa	ctatggatga	acgaaataga	cagatcgctg	3720
agataggtgc	ctcactgatt	aagcattggt	aactgtcaga	ccaagtttac	tcatatatac	3780
tttagattga	tttaaaactt	catttttaat	ttaaaaggat	ctaggtgaag	atcctttttg	3840
ataatctcat	gaccaaatac	ccttaacgtg	agttttcgtt	ccactgagcg	tcagaccccg	3900
tagaaaagat	caaaggatct	tcttgagatc	ctttttttct	gcgcgtaatc	tgctgcttgc	3960
aaacaaaaaa	accaccgcta	ccagcgggtg	tttgtttgcc	ggatcaagag	ctaccaactc	4020
tttttccgaa	ggtaactggc	ttcagcagag	cgcagatacc	aaatactgtc	cttctagtgt	4080
agccgtagtt	aggccaccac	ttcaagaact	ctgtagcacc	gcctacatac	ctcgtctctg	4140
taatcctggt	accagtggct	gctgccagtg	gcgataagtc	gtgtcttacc	gggttggact	4200
caagacgata	gttaccggat	aaggcgcagc	ggtcgggctg	aacggggggg	tcgtgcacac	4260
agcccagctt	ggagcgaacg	acctacaccg	aactgagata	cctacagcgt	gagctatgag	4320
aaagcgccac	gcttcccga	gggagaaaag	cggacaggta	tccggtaagc	ggcaggggtcg	4380
gaacaggaga	gcgcacgagg	gagcttccag	ggggaaacgc	ctggtatctt	tatagtcctg	4440
tcgggtttcg	ccacctctga	cttgagcgtc	gattttttgtg	atgctcgtca	ggggggcgga	4500
gcctatggaa	aaacgccagc	aacgcggcct	ttttacgggt	cctggccttt	tgctggcctt	4560
ttgctcacat	gttcttttct	gcgttatccc	ctgattctgt	ggataaccgt	attaccgcct	4620
ttgagtgagc	tgataccgct	cgccgcagcc	gaacgaccga	gcgcagcgag	tcagtgagcg	4680
aggaagcgga	agagcgccca	atacgcaaac	cgctctctcc	cgcgcggttg	ccgattcatt	4740
aatgcagctg	gcacgacagg	tttcccga	ggaaagcggg	cagtgagcgc	aacgcaatta	4800
atgtgagtta	gctcactcat	taggcacccc	aggctttaca	ctttatgctt	ccggctcgta	4860
tggtgtgtgg	aattgtgagc	ggataacaat	ttcacacagg	aaacagctat	gaccatgatt	4920
acgccaagcg	cgcaattaac	cctcactaaa	gggaacaaaa	gctgg		4965

<210> 38

<211> 5295

<212> DNA

<213> Artificial Sequence

<220>

<223> pUG4 vector

<400> 38

aagcttacca	tggccacgc	ccgcgtcctc	ctcctggcgc	tcgccgtgct	ggccacggcc	60
gccgtcgccg	tcgcctcctc	ccgcgcggcc	gcctccacgc	agggcatctc	cgaagacctc	120
tacagccgtt	tagtcgaaat	ggccactatc	tccaagctg	cctacgccga	cctgtgcaac	180
attccgtcga	ctattatcaa	gggagagaaa	atttacaatt	ctcaaactga	cattaacgga	240
tggatcctcc	gcgacgacag	cagcaaagaa	ataatcaccc	tcttccgtgg	cactggtagt	300
gatacgaatc	tacaactcga	tactaactac	accctcacgc	ctttcgacac	cctaccacaa	360

tgcaacgggtt	gtgaagtaca	cggtggatat	tatataggat	gggtctccgt	ccaggaccaa	420
gtcgagtcgc	ttgtcaaaca	gcaggttagc	cagtatccgg	actacgcgct	gaccgtgacc	480
ggccackccc	tcggcgccctc	cctggcggca	ctcactgccg	cccagctgtc	tgcgacatac	540
gacaacatcc	gcctgtacac	cttcggcgaa	ccgcgcagcg	gcaatcaggc	cttcgcgctcg	600
tacatgaacg	atgccttcca	agcctcgagc	ccagatacga	cgcagtatct	ccgggtcact	660
catgccaacg	acggcatccc	aaacctgccc	ccggtggagc	aggggtacgc	ccatggcgggt	720
gtagagtact	ggagcgttga	tccttacagc	gcccagaaca	catttgtctg	cactggggat	780
gaagtgcagt	gctgtgaggc	ccagggcgga	caggggtgtga	ataatgcgca	cacgacttat	840
tttgggatga	cgagcggcgc	atgcacctgg	ccggtcgcgg	ccgcggaacc	actgaaggat	900
gagctgtaaa	gaagcagatc	gttcaaaca	ttggcaataa	agtttcttaa	gattgaatcc	960
tgttgccgggt	cttgcgatga	ttatcatata	atttctgttg	aattacgtta	agcatgtaat	1020
aattaacatg	taatgcatga	cgttatctat	gagatgggtt	tttatgatta	gagtcccgcg	1080
attatacatt	taatacgcga	tagaaaacaa	aatatagcgc	gcaaaactagg	ataaattatc	1140
gcgcgcgggtg	tcattctatgt	tactagatcg	ataagcttct	agagcggccg	gtggagctcc	1200
aattcgccct	atagttagtc	gtattacgcg	cgctcactgg	ccgtcgtttt	acaacgtcgt	1260
gactgggaaa	accctggcgt	tacccaactt	aatcgccctg	cagcacatcc	ccctttcgcc	1320
agctggcgta	atagcgaaga	ggcccgccac	gatcgccctt	cccaacagtt	gcgcagcctg	1380
aatggcggaat	gggacgcgcg	ctgtagcggc	gcattaagcg	cggcgggtgt	ggtgggttacg	1440
cgcagcgtga	ccgctacact	tgccagcgcc	ctagcgcggc	ctcctttcgc	tttcttcctt	1500
tcctttctcg	ccacgttcgc	cggctttccc	cgtcaagctc	taaatcgggg	gctcccttta	1560
gggttccgat	ttagtgcctt	acggcacctc	gaccccaaaa	aacttgatta	gggtgatggt	1620
tcacgtagtg	ggccatcgcc	ctgatagacg	gtttttcgcc	ctttgacgtt	ggagtccacg	1680
ttctttaata	gtggactctt	gttccaaact	ggaacaacac	tcaaccctat	ctcggtctat	1740
tcttttgatt	tataagggat	tttgccgatt	tcggcctatt	ggttaaaaaa	tgagctgatt	1800
taacaaaaat	ttaacgcgaa	ttttaacaaa	atattaacgc	ttacaattta	ggtggcactt	1860
ttcggggaaa	tgtgcgcgga	acccctatct	gtttatcttt	ctaaatacat	tcaaatatgt	1920
atccgctcat	gagacaataa	ccctgataaa	tgcttcaata	atattgaaaa	aggaagagta	1980
tgagtattca	acatttccgt	gtcgccctta	ttcccttttt	tgccgcatct	tgccctcctg	2040
tttttgctca	cccagaaacg	ctggtgaaag	taaaagatgc	tgaagatcag	ttgggtgcac	2100
gagtgggtta	catcgaaactg	gatctcaaca	gcggtaagat	ccttgagagt	tttcgccccg	2160
aagaacgttt	tccaatgatg	agcactttta	aagttctgct	atgtggcgcg	gtattatccc	2220
gtattgacgc	cgggcaagag	caactcgggtc	gccgcataca	ctattctcag	aatgacttgg	2280
ttgagtactc	accagtcaca	gaaaagcatc	ttacggatgg	catgacagta	agagaattat	2340
gcagtgcctg	cataaccatg	agtataaca	ctgcggccaa	cttacttctg	acaacgatcg	2400
gaggaccgaa	ggagctaacc	gcttttttgc	acaacatggg	ggatcatgta	actcgccttg	2460
atcgttggga	accggagctg	aatgaagcca	taccaaacga	cgagcgtgac	accacgatgc	2520
ctgtagcaat	ggcaacaacg	ttgcgcgaac	tattaactgg	cgaactactt	actctagctt	2580
cccggcaaca	attaatagac	tggatggagg	cggataaagt	tgaggacca	cttctgcgct	2640
cggcccttcc	ggctggctgg	tttattgctg	ataaatctgg	agccggtgag	cgtgggtctc	2700
gcggtatcat	tgagcactg	gggcccagatg	gtaagccctc	ccgtatcgta	gttatctaca	2760
cgacggggag	tcaggcaact	atggatgaac	gaaatagaca	gatcgctgag	ataggtgcct	2820
cactgattaa	gcattggtaa	ctgtcagacc	aagtttactc	atatataact	tagattgatt	2880
taaaacttca	tttttaattt	aaaaggatct	aggtgaagat	cctttttgat	aatctcatga	2940
ccaaaatccc	ttaacgtgag	ttttcgcttc	actgagcgtc	agaccccgta	gaaaagatca	3000
aaggatcttc	ttgagatcct	ttttttctgc	gcgtaatctg	ctgcttgcaa	acaaaaaac	3060
caccgctacc	agcgggtggtt	tggttgccgg	atcaagagct	accaactctt	tttccgaagg	3120
taactggctt	cagcagagcg	cagataccaa	atactgtcct	tctagtgtag	ccgtagttag	3180
gccaccactt	caagaactct	gtagcaccgc	ctacatacct	cgtctgtcta	atcctgttac	3240
cagtggctgc	tgccagtggc	gataagtcgt	gtcttaccgg	gttggaactca	agacgatagt	3300
taccggataa	ggcgagcgcg	tcgggctgaa	cgggggggtc	gtgcacacag	cccagcttgg	3360
agcgaacgac	ctacaccgaa	ctgagatacc	tacagcgtga	gctatgagaa	agcgccacgc	3420
ttcccgaagg	gagaaaggcg	gacaggtatc	cggtaagcgg	cagggctcga	acaggagagc	3480
gcacgagggg	gcttccaggg	ggaaacgcct	ggtatcttta	tagtcctgtc	gggtttcgcc	3540
acctctgact	tgagcgtcga	tttttgtgat	gctcgtcagg	ggggcggagc	ctatggaaaa	3600
acgccagcaa	cgcggccctt	ttacgggttc	tgcccttttg	ctggcccttt	gctcacatgt	3660
tctttcctgc	gttatccctt	gattctgtgg	ataaccgtat	taccgccttt	gagtgagctg	3720
ataccgctcg	ccgcagccga	acgaccgagc	gcagcagatc	agtgagcgag	gaagcggaag	3780
agcgcccaat	acgcaaaccg	cctctccccg	cgcgttggcc	gattcattaa	tgagctggc	3840
acgacagggt	tcccgaactg	aaagcgggca	gtgagcgcaa	cgcaattaat	gtgagttagc	3900
tactcatta	ggcaccaccg	gctttacact	ttatgcttcc	ggctcgtatg	ttgtgtggaa	3960
ttgtgagcgg	ataacaattt	cacacaggaa	acagctatga	ccatgattac	gccaagcgcg	4020

```

caattaaccc tcactaaagg gaacaaaagc tgggtaccgg gccccccctc gaggtcattc 4080
atatgcttga gaagagagtc gggatagtc aaaataaaac aaaggtaaga ttacctggtc 4140
aaaagtgaac acatcagtta aaagggtgga taagtaaaat atcggtaata aaagggtggc 4200
caaagtgaac tttactcttt tctactatta taaaaattga ggatgttttg tcgggtacttt 4260
gatacgtcac ttttgtatga attgggtttt aagttttatt gcgatttgga aatgcatatc 4320
tgtatttgag tcgggttttta agttcgttgc ttttgtaaat acagagggat ttgtataaga 4380
aatatcttta aaaaacccat atgctaattt gacataattt ttgagaaaaa tatatatcca 4440
ggcgaattcc acaatgaaca ataataagat taaaatagct tgcccccggt gcagcgatgg 4500
gtattttttt tagtaaaata aaagataaac ttagactcaa aacatttaca aaaacaaccc 4560
ctaaagtcct aaagcccaaa gtgctatgca cgatccatag caagcccagc ccaacccaac 4620
ccaacccaac ccaccccagt gcagccaact ggcaaatagt ctccaccccc ggcactatca 4680
ccgtgagttg tccgcaccac cgcacgtctc gcagccaaaa aaaaaaaaaa aaagaaaaaa 4740
aagaaaaaga aaaacagcag gtgggtccgg gtcgtggggg ccggaaaagc gaggaggatc 4800
gagagcagcg acgaggcccc gccctccctc cgcttccaaa gaaacgcccc ccatcgccac 4860
tatatacata cccccccctc tcttcccatc cccccaaccc taccaccacc accaccacca 4920
cctcctcccc cctcgtgccc ggacgacgag ctcttccccc ctccccctcc gccgcccggc 4980
gtaaccaccc cgccccctct ctctttcttt ctccgttttt tttttcgtct cggctctgat 5040
ctttggcctt ggtagtttgg gtgggcgaga gcggtctcgt cgcccagatc ggtgcgcggg 5100
aggggcgga tctcgcggtt ggcgtctccg ggcgtgagtc ggcccggatc ctcgcgggga 5160
atggggctct cggatgtaga tcttctttct ttcttctttt tgtggtagaa tttgaatccc 5220
tcagcattgt tcacggttag tttttctttt catgatttgt gacaaatgca gcctcgtgcg 5280
gagctttttt gtagc 5295

```

<210> 39

<211> 299

<212> PRT

<213> Artificial Sequence

<220>

<223> pUG4 vector

<221> VARIANT

<222> (1)...(299)

<223> Xaa = Any Amino Acid

<400> 39

```

Met Ala His Ala Arg Val Leu Leu Leu Ala Leu Ala Val Leu Ala Thr
 1             5             10             15
Ala Ala Val Ala Val Ala Ser Ser Arg Ala Ala Ala Ser Thr Gln Gly
 20             25             30
Ile Ser Glu Asp Leu Tyr Ser Arg Leu Val Glu Met Ala Thr Ile Ser
 35             40             45
Gln Ala Ala Tyr Ala Asp Leu Cys Asn Ile Pro Ser Thr Ile Ile Lys
 50             55             60
Gly Glu Lys Ile Tyr Asn Ser Gln Thr Asp Ile Asn Gly Trp Ile Leu
 65             70             75             80
Arg Asp Asp Ser Ser Lys Glu Ile Ile Thr Val Phe Arg Gly Thr Gly
 85             90             95
Ser Asp Thr Asn Leu Gln Leu Asp Thr Asn Tyr Thr Leu Thr Pro Phe
 100            105            110
Asp Thr Leu Pro Gln Cys Asn Gly Cys Glu Val His Gly Gly Tyr Tyr
 115            120            125
Ile Gly Trp Val Ser Val Gln Asp Gln Val Glu Ser Leu Val Lys Gln
 130            135            140
Gln Val Ser Gln Tyr Pro Asp Tyr Ala Leu Thr Val Thr Gly His Xaa
 145            150            155            160
Leu Gly Ala Ser Leu Ala Ala Leu Thr Ala Ala Gln Leu Ser Ala Thr
 165            170            175
Tyr Asp Asn Ile Arg Leu Tyr Thr Phe Gly Glu Pro Arg Ser Gly Asn
 180            185            190
Gln Ala Phe Ala Ser Tyr Met Asn Asp Ala Phe Gln Ala Ser Ser Pro

```

	195		200		205
Asp	Thr Thr Gln Tyr Phe Arg Val Thr His Ala	Asn Asp Gly Ile Pro			
210		215		220	
Asn	Leu Pro Pro Val Glu Gln Gly Tyr Ala His	Gly Gly Val Glu Tyr			
225		230		235	240
Trp	Ser Val Asp Pro Tyr Ser Ala Gln Asn Thr Phe Val Cys Thr Gly				
	245		250		255
Asp	Glu Val Gln Cys Cys Glu Ala Gln Gly Gly Gln Gly Val Asn Asn				
	260		265		270
Ala	His Thr Thr Tyr Phe Gly Met Thr Ser Gly Ala Cys Thr Trp Pro				
	275		280		285
Val	Ala Ala Ala Glu Pro Leu Lys Asp Glu Leu				
290		295			

<210> 40

<211> 5001

<212> DNA

<213> Artificial Sequence

<220>

<223> pUB8.11 vector

<221> misc_feature

<222> (1)...(5001)

<223> n = A,T,C or G

<400> 40

catggggccag	gtataattat	gggatatctc	aagcaaataa	tcgaaatatc	accattggct	60
acaatatctg	agctccgagt	tctgactgca	gtctggatga	cgcgtgttgt	atctagaact	120
ctagatagca	cagccacagc	acctacagga	gtgcgacact	tgtggactgt	agtagtggtg	180
gagacggagc	tctttcctac	ctcctgacgt	tgccgccgtt	gtccattcca	acggcatcac	240
tctcaaccaa	tcacgcgctc	ccaacaaaat	atcgtccccc	atgtcttggc	ggagagagag	300
tacatacatg	ctgtcgcgcc	gttttttgtct	gaatctcgct	tccactggcc	aatcagctca	360
gctcccggga	gctcactcat	tcaagatccc	atcgtcgtcg	tcacccctgg	cgtcatggga	420
tggaaaagaa	cctccgttgc	tcggatgagt	cagccatata	cccgaacaga	gtactgcaag	480
ataacccaat	tcagattccc	ccaatagaga	aagtatagca	tgctttcggg	ttttgtttgg	540
cttaattgac	tttatTTTTg	ttggagttga	atgctgattt	gttggtgtaa	atgccccacc	600
atctgaatat	cgagacggat	aataggctgg	ctaattaatt	tatagcaaga	ttctgtagtg	660
cacatcgcaa	atatctttct	gggcattaca	gctggaggct	tcatacagct	gaaacactct	720
gcagagcctg	aagcaagtgg	tgaagcgtgg	cgatgagatg	ggtataaaaac	ccccggcacc	780
gggacgcgag	ctcccgccct	ccagtaccat	ctcgcctcgc	tccccctgcc	ggacgaccca	840
gtaaaataact	gttgcccact	cgccggcgag	atggcccacg	gccgcatact	cttcttggcg	900
ctcgcctgtc	tggccaccgc	cgcggtggcc	gccgcatact	tggcggactc	caaccgcata	960
cgccccgtca	ccgagcgcgc	ggccgcctcc	acgcagggca	tctccgaaga	cctctacagc	1020
cgttttagtcg	aaatggccac	tatctcccaa	gctgcctacg	ccgacctgtg	caacattccg	1080
tcgactatta	tcaagggaga	gaaaatttac	aattctcaaa	ctgacattaa	cggatggatc	1140
ctccgcgacg	acagcagcaa	agaaataatc	accgtcttcc	gtggcactgg	tagtgatacg	1200
aatctacaac	tcgatactaa	ctacaccctc	acgcctttcg	acaccctacc	acaatgcaac	1260
ggttggtgaag	tacacggtgg	atattatatt	ggatgggtct	ccgtccagga	ccaagtcgag	1320
tcgcttggtca	aacagcaggt	tagccagtat	ccggactacg	cgctgaccgt	gaccggccac	1380
kccctcggcg	cctccctggc	ggcactcact	gccgcccagc	tgtctgcgac	atacgacaac	1440
atccgcctgt	acaccttcgg	cgaaccgcgc	agcggcaatc	aggccttcgc	gtcgtacatg	1500
aacgatgcct	tccaagcctc	gagcccagat	acgacgcagt	atttccgggt	cactcatgcc	1560
aacgacggca	tcccaaacct	gcccccggtg	gagcaggggt	acgcccattg	cgggtgtagag	1620
tactggagcg	ttgatcctta	cagcgcgccag	aacacatttg	tctgcaactg	ggatgaagtg	1680
cagtgtgtgtg	aggcccaggg	cggacagggg	gtgaataatg	cgcacacgac	ttattttggg	1740
atgacgagcg	gagcctgtac	atgggtgatca	gtcattttcag	cctccccgag	tgtaccagga	1800
aagatggatg	tcctggagag	ggggccgcgt	aaccactgaa	ggatgagctg	taaagaagca	1860
gacggttcaa	acatttggca	ataaagtttc	ttaagattga	atcctgttgc	cggctcttgcg	1920
atgattatca	tataatttct	gttgaattac	gttaagcatg	taataattaa	catgtaatgc	1980
atgacgttat	ttatgagatg	ggttttttatg	attagagtcc	cgcaattata	catttaatac	2040

```

gcgatagaaa acaaaatata ggcgcgcaaac taggataaat tatcgcgcgcg ggtgtcatct 2100
atgttactag atcgataagc ttctagagcg gccggtggag ctccaattcg ccctatagtg 2160
agtcgtatta cgcgcgctca ctggccgctcg ttttacaacg tcgtgactgg gaaaaccctg 2220
gcgttaccca acttaatcgc cttgcagcac atcccccttt cgccagctgg cgtaatagcg 2280
aagaggcccc caccgatcgc ccttcccaac agttgcgcag cctgaatggc gaatgggacg 2340
cgccctgtag cggcgcatta agcgcggcgg gtgtgggtgt tacgcgcagc gtgaccgcta 2400
cacttgccag cgccctagcg ccgctcctt tcgctttctt cccttccttt ctgcgccagt 2460
tcgccggctt tccccgtcaa gctctaaatc gggggctccc tttaggggtc cgatttagtg 2520
ctttacggca cctcgacccc aaaaaacttg attaggggtga tggttcacgt agtgggcat 2580
cgccctgata gacgggtttt cgccctttga cgttggagtc cagttctttt aatagtggac 2640
tcttgttcca aactggaaca acactcaacc ctatctcggc ctattctttt gatttataag 2700
ggattttgcc gatttcggcc tattgggtta aaaatgagct gatttaacaa aaatttaacg 2760
cgaattttta caaaatatta acgcttataa tttaggtggc acttttcggg gaaatgtgcg 2820
cggaaccctt atttgtttat ttttctaaat acattcaaat atgtatccgc tcatgagaca 2880
ataaccctga taaatgcttc aataatattg aaaaaggaag agtatgagta ttcaacattt 2940
ccgtgtcgcc cttattccct tttttgcggc attttgcctt cctgtttttg ctcaccaga 3000
aacgctgggt aaagtaaaag atgctgaaga tcagttgggt gcacgagtgg gttacatcga 3060
actggatctc aacagcggta agatccttga gagttttcgc cccgaagaac gttttccaat 3120
gatgagcact tttaaagtgc tgctatgtgg cgcggtatta tcccgattg acgcgggca 3180
agagcaactc ggtcgccgca tacactattc tcagaatgac ttggttgagt actcaccagt 3240
cacagaaaag catcttacgg atggcatgac agtaagagaa ttatgcagtg ctgccataac 3300
catgagtgat aacactgcgg ccaacttact tctgacaacg atcggaggac cgaaggagct 3360
aaccgctttt ttgcacaaca tgggggatca tgtaactcgc cttgatcgtt gggaaccgga 3420
gctgaatgaa gccataccaa acgacgagcg tgacaccacg atgcctgtag caatggcaac 3480
aacgttgcg aaactattaa ctggcgaact acttactcta gcttcccggc aacaattaat 3540
agactggatg gaggcggata aagttgcagg accacttctg cgctcggccc ttccggctgg 3600
ctggtttatt gctgataaat ctggagccgg tgagcgtggg tctcgcggta tcattgcagc 3660
actggggcca gatggtaagc cctcccgat cgtagttatc tacacgacgg ggagtcaggc 3720
aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga ttaagcattg 3780
gtaactgtca gaccaagttt actcatatat actttagatt gatttaaaac ttcattttta 3840
atttaaaagg atctaggtga agatcctttt tgataatctc atgacccaaa tcccttaacg 3900
tgagttttcg ttccactgag cgtcagaccc cgtagaaaag atcaaaggat cttcttgaga 3960
tccttttttt ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc taccagcgtt 4020
ggtttgtttg ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag 4080
agcgcagata ccaaatactg tccttctagt gtagccgtag ttaggccacc acttcaagaa 4140
ctctgtagca ccgcctacat acctcgctct gctaatactg ttaccagtgg ctgctgccag 4200
tggcgataag tcgtgtctta ccgggttgga ctcaagacga tagttaccgg ataaggcgca 4260
gcggtcgggc tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac 4320
cgaactgaga tacctacagc gtgagctatg agaaagcgcc acgcttcccg aagggagaaa 4380
ggcggacagg tatccggtaa gcggcagggc cggaacagga gagcgcacga gggagcttcc 4440
agggggaaac gcctgggtatc tttatagtc tgtegggttt cgccacctct gacttgagcg 4500
tcgatttttg tgatgctcgt cagggggggc gagcctatgg aaaaacgcca gcaacgcggc 4560
ctttttacgg ttcttgccct tttgctggcc ttttgctcac atgttctttc ctgcgttatc 4620
ccctgattct gtggataacc gtattaccgc ctttgagtga gctgataacc ctgcgcgag 4680
ccgaacgacc gagcgcagcg agtcagttag cgaggaagcg gaagagcgcc caatacgcaa 4740
accgcctctc ccgcgcgctt ggccgattca ttaatgcagc tggcacgaca ggtttcccga 4800
ctggaaagcg ggcagtgagc gcaacgcaat taatgtgagt tagctcactc attaggcacc 4860
ccaggcttta cactttatgc ttccggctcg tatgttgtgt ggaattgtga gcggataaca 4920
atttcacaca ggaaacagct atgaccatga ttacgccaag cgcgcaatta accctcacta 4980
aagggaacaa aagctgggta c 5001

```

<210> 41

<211> 5387

<212> DNA

<213> Artificial Sequence

<220>

<223> pTP11-1 vector

<400> 41

```

aagcttacca tggcccacgc ccgcgtcctc ctctggcgcg tcgcccgtgct ggccacggcc 60

```

gccgtcgccg	tcgcctcctc	ctcctccttc	gccgactcca	accgatccg	gcccgtcacc	120
gaccgcgcgg	ccgcctccac	gcagggcatc	tccgaagacc	tctacagccg	tttagtcgaa	180
atggccacta	tctcccaagc	tgcctacgcc	gacctgtgca	acattccgtc	gactattatc	240
aagggagaga	aaatttacia	ttctcaaact	gacattaacg	gatggatcct	ccgcgacgac	300
agcagcaaag	aaataatcac	cgtcttccgt	ggcactggta	gtgatacgaa	tctacaactc	360
gatactaact	acaccctcac	gcctttcgac	accctaccac	aatgcaacgg	ttgtgaagta	420
cacgggtggat	attatatattg	atgggtctcc	gtccaggacc	aagtcgagtc	gcttgtcaaa	480
cagcaggtta	gccagtatcc	ggactacgcg	ctgaccgtga	ccggccackc	cctcggcgcc	540
tccctggcgg	cactcactgc	cgcctcagctg	tctgcgacat	acgacaacat	ccgcctgtac	600
accttcggcg	aaccgcgcag	cggcaatcag	gccttcgcgt	cgtacatgaa	cgatgccttc	660
caagcctcga	gccagatac	gacgcagtat	ttccgggtca	ctcatgccaa	cgacggcatc	720
ccaaacctgc	ccccgggtgga	gcaggggtac	gcccattggcg	gtgtagagta	ctggagcggt	780
gatccttaca	gcgcccagaa	cacatttgtc	tgcactgggg	atgaagtgca	gtgctgtgag	840
gcccgaggcg	gacaggggtg	gaataatgcg	cacacgactt	atcttgggat	gacgagcgga	900
gcctgtacat	ggtgatcagt	catttcagcc	tccccgagtg	taccaggaaa	gatggatgtc	960
ctggagaggg	ggccgcgtaa	ccactgaagg	atgagctgta	aagaagcaga	tcgttcaaac	1020
atctggcaat	aaagtttctt	aagattgaat	cctggttgcg	gtcttgcgat	gattatcata	1080
taatttctgt	tgaattacgt	taagcatgta	ataattaaca	tgtaatgcat	gacgttatctt	1140
atgagatggg	tttttatgat	tagagtcccg	caattataca	tttaatacgc	gatagaaaac	1200
aaaatatagc	gcgcaaacta	ggataaatta	tgcgcgcggg	tgtcatctat	gttactagat	1260
cgataagctt	ctagagcggc	cgggtggagct	ccaattcgcc	ctatagttag	tcgtattacg	1320
cgcgctcact	ggcgcgtcgt	ttacaacgtc	gtgactggga	aaacctggc	gttaccacaac	1380
ttaatcgcc	tgcagcacat	ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	1440
ccgatcgccc	ttcccaacag	ttgcgcagcc	tgaatggcga	atgggacgcg	ccctgtagcg	1500
gcgcattaag	cgcggcgggg	gtggtgggtta	cgcgcagcgt	gaccgctaca	cttgccagcg	1560
ccctagcgcc	cgtccttttc	gctttcttcc	cttcccttct	cgcacgcttc	gccggctttc	1620
cccgtcaagc	tctaaatcgg	gggtccctt	taggggtccg	atcttagtgc	ttacggcacc	1680
tcgaccccaa	aaaacttgat	taggggtgatg	gttcacgtag	tgggcatcg	ccctgataga	1740
cggtttttctg	ccctttgacg	ttggagtcca	cgttctttaa	tagtggactc	ttgttccaaa	1800
ctggaacaac	actcaacctt	atctcggtct	attcttttga	tttataaggg	atcttgccga	1860
tttcggccta	ttggttaaaa	aatgagctga	tttaacaaaa	atttaacgcg	aatttttaaca	1920
aaatattaac	gcttacaatt	taggtggcac	ttttcgggga	aatgtgcgcg	gaacctctat	1980
ttgtttatctt	ttctaaatac	attcaaatac	gtatccgctc	atgagacaat	aaccttgata	2040
aatgcttcaa	taatatgtga	aaaggaagag	tatgagtatt	caacatttcc	gtgtcgccct	2100
tattcccttt	tttgcggcat	tttgccttcc	tgtttttgct	caccagagaa	cgtggtgaa	2160
agtaaaagat	gctgaagatc	agttgggtgc	acgagtgggt	tacatcgaa	tggatctcaa	2220
cagcggtaag	atccttgaga	gttttcgccc	cgaagaacgt	tttccaatga	tgagcacttt	2280
taaagtctctg	ctatgtggcg	cggattatct	ccgtattgac	gccgggcaag	agcaactcgg	2340
tcgcccgcata	cactattctc	agaatgactt	ggttgagtac	tcaccagtca	cagaaaagca	2400
tcttacggat	ggcatgacag	taagagaatt	atgcagtgc	gccataacca	tgagtgataa	2460
cactgcggcc	aacttacttc	tgacaacgat	cggaggaccg	aaggagctaa	ccgctttttt	2520
gcacaacatg	ggggatcatg	taactcgcct	tgatcggttg	gaaccggagc	tgaatgaagc	2580
cataccaaac	gacgagcgtg	acaccacgat	gcctgtagca	atggcaacaa	cgttgcgcaa	2640
actattaact	ggcgaactac	ttactctagc	ttcccggcaa	caattaatag	actggatgga	2700
ggcggataaa	gttgacaggac	cacttctgcg	ctcggccctt	ccggctggct	ggtttattgc	2760
tgataaatct	ggagccgggtg	agcgtgggtc	tgcgggtatc	attgcagcac	tggggccaga	2820
tggttaagccc	tcccgtatcg	tagttatcta	cacgacgggg	agtcaggcaa	ctatggatga	2880
acgaaataga	cagatcgctg	agataggtgc	ctcactgatt	aagcattggg	aactgtcaga	2940
ccaagtttac	tcatatatac	tttagattga	tttaaaaact	cattttttaat	ttaaaaggat	3000
ctaggtgaag	atcctttttg	ataatctcat	gacaaaatc	ccttaacgtg	agttttcgtt	3060
ccactgagcg	tcagaccccc	tagaaaagat	caaaggatct	tcttgagatc	ctttttttct	3120
gcgcgtaatc	tgctgcttgc	aaacaaaaaa	accaccgcta	ccagcgggtg	tttgtttgct	3180
ggatcaagag	ctaccaactc	tttttccgaa	ggtaactggc	ttcagcagag	cgcagatacc	3240
aaatactgtc	cttctagtgt	agccgtagt	agggcaccac	ttcaagaact	ctgtagcacc	3300
gcctacatac	ctcgctctgc	taatcctgtt	accagtggct	gctgccagtg	gcgataagtc	3360
gtgtcttacc	gggttggact	caagacgata	gttaccggat	aaggcgcagc	ggtcgggctg	3420
aacggggggg	tcgtgcacac	agcccagctt	ggagcgaacg	acctacaccg	aactgagata	3480
cctacagcgt	gagctatgag	aaagcgccac	gcttcccgaa	gggagaaagg	cggacaggta	3540
tccggtaagc	ggcagggctg	gaacaggaga	gcgcacgagg	gagcttccag	ggggaaacgc	3600
ctgggtatctt	tatagtcctg	tccgggttctg	ccacctctga	cttgagcgtc	gatttttgtg	3660
atgctcgtca	ggggggcgga	gcctatggaa	aaacgccagc	aacgcggcct	ttttacgggt	3720

```

cctggccttt tgctggcctt ttgctcacat gttcttttct gcgttatccc ctgattctgt 3780
ggataaccgt attaccgcct ttgagtgagc tgataccgct cgccgcagcc gaacgaccga 3840
gcgcgagcag tcagtgagcg aggaagcgga agagcgccca atacgcaaac cgcctctccc 3900
cgcgcgttgg ccgattcatt aatgcagctg gcacgacagg tttcccgcact ggaaagcggg 3960
cagtgagcgc aacgcaatta atgtgagtta gctcactcat taggcacccc aggctttaca 4020
ctttatgctt ccggctcgta tggtgtgtgg aattgtgagc ggataacaat ttcacacagg 4080
aaacagctat gaccatgatt acgccaagcg cgcaattaac cctcactaaa gggaacaaaa 4140
gctgggtacc gggccccccc tcgagggtcat tcatatgctt gagaagagag tcgggatagt 4200
ccaaaataaa acaaaggtaa gattacctgg tcaaaagtga aaacatcagt taaaagggtg 4260
tataagtaaa atatcggtaa taaaagggtg cccaaagtga aatttactct tttctactat 4320
tataaaaatt gaggatgttt tgctcggtact ttgatacgtc atttttgtat gaattgggtt 4380
ttaagtttat tcgcgatttg gaaatgcata tctgtatttg agtcgggttt taagttcggt 4440
gctttttgtaa atacagaggg atttgtataa gaaatatctt taaaaaaccc atatgcta 4500
ttgacataat ttttgagaaa aatatatatt caggcgaatt ccacaatgaa caataataag 4560
attaaaatag cttgcccccg ttgcagcgat ggggtatttt tctagtaaaa taaaagataa 4620
acttagactc aaaacattta caaaaacaac ccctaaagtc ctaaagccca aagtgctatg 4680
cacgatccat agcaagccca gcccaaccca acccaaccca acccacccca gtgcagccaa 4740
ctggcaaata gtctccaccc ccggcactat caccgtgagt tgtccgcacc accgcacgtc 4800
tcgcagccaa aaaaaaaaaa agaaagaaaa aaaagaaaaa gaaaaacagc aggtgggtcc 4860
gggtcgtggg ggccggaaaa gcgaggagga tcgcgagcag cgacgaggcc cggccctccc 4920
tccgcttcca aagaaacgcc ccccatcgcc actatataca tccccccccc tctcctccca 4980
tccccccaac cctaccacca ccaccaccac cacctcctcc cccctcgctg ccggacgacg 5040
agctcctccc ccctccccct ccgcccgcgc cggtaaccac cccgcccctc tcctctttct 5100
ttctccgttt tttttttcgt ctcggtctcg atctttggcc ttggtagttt gggtagggcg 5160
gagcggcttc gtcgcccaga tcggtgcgcg ggagggggcg gatctcgcgg ctggcgtctc 5220
cgggcgtgag tcggcccgga tcctcgcggg gaatggggct ctcggatgta gatcttcttt 5280
ctttcttctt tttgtggtag aatttgaatc cctcagcatt gttcatcggt agtttttctt 5340
ttcatgattt gtgacaaatg cagcctcgtg cggagctttt ttgtagc 5387

```

<210> 42

<211> 301

<212> PRT

<213> Artificial Sequence

<220>

<223> pTP11-1 vector

<221> VARIANT

<222> (1)...(301)

<223> Xaa = Any Amino Acid

<400> 42

```

Met Ala His Ala Arg Val Leu Leu Leu Ala Leu Ala Val Leu Ala Thr
 1             5             10             15
Ala Ala Val Ala Val Ala Ser Ser Ser Ser Phe Ala Asp Ser Asn Pro
      20             25             30
Ile Arg Pro Val Thr Asp Arg Ala Ala Ala Ser Thr Gln Gly Ile Ser
      35             40             45
Glu Asp Leu Tyr Ser Arg Leu Val Glu Met Ala Thr Ile Ser Gln Ala
      50             55             60
Ala Tyr Ala Asp Leu Cys Asn Ile Pro Ser Thr Ile Ile Lys Gly Glu
65             70             75             80
Lys Ile Tyr Asn Ser Gln Thr Asp Ile Asn Gly Trp Ile Leu Arg Asp
      85             90             95
Asp Ser Ser Lys Glu Ile Ile Thr Val Phe Arg Gly Thr Gly Ser Asp
      100            105            110
Thr Asn Leu Gln Leu Asp Thr Asn Tyr Thr Leu Thr Pro Phe Asp Thr
      115            120            125
Leu Pro Gln Cys Asn Gly Cys Glu Val His Gly Gly Tyr Tyr Ile Gly
      130            135            140
Trp Val Ser Val Gln Asp Gln Val Glu Ser Leu Val Lys Gln Gln Val

```

145					150				155				160		
Ser	Gln	Tyr	Pro	Asp	Tyr	Ala	Leu	Thr	Val	Thr	Gly	His	Xaa	Leu	Gly
				165					170					175	
Ala	Ser	Leu	Ala	Ala	Leu	Thr	Ala	Ala	Gln	Leu	Ser	Ala	Thr	Tyr	Asp
			180					185					190		
Asn	Ile	Arg	Leu	Tyr	Thr	Phe	Gly	Glu	Pro	Arg	Ser	Gly	Asn	Gln	Ala
		195					200					205			
Phe	Ala	Ser	Tyr	Met	Asn	Asp	Ala	Phe	Gln	Ala	Ser	Ser	Pro	Asp	Thr
	210					215					220				
Thr	Gln	Tyr	Phe	Arg	Val	Thr	His	Ala	Asn	Asp	Gly	Ile	Pro	Asn	Leu
225					230					235					240
Pro	Pro	Val	Glu	Gln	Gly	Tyr	Ala	His	Gly	Gly	Val	Glu	Tyr	Trp	Ser
			245						250					255	
Val	Asp	Pro	Tyr	Ser	Ala	Gln	Asn	Thr	Phe	Val	Cys	Thr	Gly	Asp	Glu
			260					265					270		
Val	Gln	Cys	Cys	Glu	Ala	Gln	Gly	Gly	Gln	Gly	Val	Asn	Asn	Ala	His
	275						280					285			
Thr	Thr	Tyr	Phe	Gly	Met	Thr	Ser	Gly	Ala	Cys	Thr	Trp			
	290					295					300				

<210> 43

<211> 1259

<212> DNA

<213> Artificial Sequence

<220>

<223> actin promoter

<400> 43

ggtaccgggc	ccccctcga	ggtcattcat	atgcttgaga	agagagtcgg	gatagtccaa	60
aataaaaaca	aggtaagatt	acctgggtcaa	aagtgaaaac	atcagttaaa	aggtggtata	120
agtaaaatat	cggtaataaaa	agggtggccca	aagtgaatt	tactcttttc	tactattata	180
aaaattgagg	atgttttgtc	ggctactttga	tacgtcattt	ttgtatgaat	tgggttttta	240
gtttattcgc	gattttggaaa	tgcataatctg	tattttgagtc	ggttttttaag	ttcgttgctt	300
ttgtaaatac	agaggggattt	gtataagaaa	tatcttttaa	aaacccatat	gctaatttga	360
cataattttt	gagaaaaata	tatattcagg	cgaattccac	aatgaacaat	aataagatta	420
aaatagcttg	cccccgttgc	agcgatgggt	atctttttcta	gtaaaataaa	agataaaactt	480
agactcaaaa	cattttacaaa	aacaaccct	aaagtcctaa	agcccaaagt	gctatgcacg	540
atccatagca	agcccagccc	aaccaaaccc	aaccaaaccc	acccagtgcc	agccaactgg	600
caaatagtct	ccacccccgg	cactatcacc	gtgagttgtc	cgcaccaccg	cacgtctcgc	660
agccaaaaaa	aaaaaaagaa	agaaaaaaa	gaaaaagaaa	aacagcaggt	gggtccgggt	720
cgtggggggc	ggaaaagcga	ggaggatcgc	gagcagcgac	gaggcccggc	cctccctccg	780
cttccaaaga	aacgcccccc	atcgccacta	tatacatacc	ccccctctc	ctcccatccc	840
cccaacccta	ccaccaccac	caccaccacc	tcctcccccc	tcgctgccgg	acgacgagct	900
cctccccct	ccccctccgc	cgccgccggg	aaccaccccg	cccctctcct	ctttctttct	960
ccgttttttt	tttcgtctcg	gtctcgatct	ttggccttgg	tagtttgggt	gggcgagagc	1020
ggcttcgtcg	cccagatcgg	tgcgcgggag	gggcgggatc	tcgcggctgg	cgtctccggg	1080
cgtgagtcgg	cccggatcct	cgcggggaat	ggggctctcg	gatgtagatc	ttctttcttt	1140
cttctttttg	tggtagaatt	tgaatccctc	agcattgttc	atcggtagtt	tttcttttca	1200
tgattttgtg	caaatgcagc	ctcgtgcgga	gcttttttgt	aggtagaagc	ttaccatgg	1259

<210> 44

<211> 27

<212> PRT

<213> Artificial Sequence

<220>

<223> aleurain-NPIR delete structure

<400> 44

Met Ala His Ala Arg Val Leu Leu Leu Ala Leu Ala Val Leu Ala Thr

1 5 10 15
Ala Ala Val Ala Val Ala Ser Ser Arg Ala Ala
20 25

<210> 45
<211> 93
<212> DNA
<213> Artificial Sequence

<220>
<223> aleurain-NPIR delete structure encoding sequence

<400> 45
aagcttacca tggcccacgc ccgcgtcctc ctccctggcgc tcgccgtgct ggccacggcc 60
gccgtcgccg tcgcctcctc ccgcgcggcc gcc 93

<210> 46
<211> 873
<212> DNA
<213> Artificial Sequence

<220>
<223> SEE1 (senescence enhanced) promoter

<400> 46
catggggccag gtataattat gggatatctc aagcaaataa tcgaaatatc accattggct 60
acaatatctg agctccgagt tctgactgca gtctggatga ccgctgttgt atctagaact 120
ctagatagca cagccacagc acctacagga gtgcgacact tgtggactgt agtagtggtg 180
gagacggagc tctttcctac ctccctgacgt tgccgccggt gtccattcca acggcatcac 240
tctcaaccaa tcacgcgctc ccaacaaaat atcgctcccc atgtcttggc ggagagagag 300
tacatacatg ctgtcgcgcc gtttttgtct gaatctcgct tccactggcc aatcagctca 360
gctcccggga gctcactcat tcaagatccc atcgctcgctg tcacccctgg cgtcatggga 420
tggaagaaagaa cctccggttg tcggatgagt cagccatata cccgaacaga gtactgcaag 480
ataacccaat tcagattccc ccaatagaga aagtatagca tgctttcggg ttttgtttgg 540
cttaattgac tttatTTTTTg ttggagttga atgctgattt gttgtgtaaa atgcccaccc 600
atctgaatat cgagacggat aataggctgg ctaattaatt tatagcaaga ttctgtagtg 660
cacatcgcaa atatctttct gggcattaca gctggaggct tcatcagcct gaaacactct 720
gcagagcctg aagcaagtgg tgaagcgtgg cgatgagatg ggtataaaac ccccggcacc 780
gggacgcgag ctcccgccta ccagtaccat ctgcctcgc tccccctgcc ggacgaccca 840
gtaaaataact gttgcccact cgccggcgag atg 873

<210> 47
<211> 39
<212> PRT
<213> Artificial Sequence

<220>
<223> SEE1 promoter plus vacuolar aleurain signal/NPIR
sequence

<400> 47
Met Ala His Gly Arg Ile Leu Phe Leu Ala Leu Ala Val Leu Ala Thr
1 5 10 15
Ala Ala Val Ala Ala Ala Ser Leu Ala Asp Ser Asn Pro Ile Arg Pro
20 25 30
Val Thr Glu Arg Ala Ala Ala
35

<210> 48
<211> 987
<212> DNA

<213> Artificial Sequence

<220>

<223> SEE1 promoter plus vacuolar aleurain signal/NPIR
encoding sequence

<221> misc_feature

<222> (1)...(987)

<223> n = A,T,C or G

<400> 48

catgggcccag	gtataattat	gggatatctc	aagcaaataa	tcgaaatatc	accattggct	60
acaatatctg	agctccgagt	tctgactgca	gtctggatga	cgcggtgttg	atctagaact	120
ctagatagca	cagccacagc	acctacagga	gtgcgacact	tgtggactgt	agtagtggtg	180
gagacggagc	tctttcctac	ctcctgacgt	tgccgccggt	gtccattcca	acggcatcac	240
tctcaaccaa	tcacgcgctc	ccaacaaaat	atcgtccccc	atgtcttggc	ggagagagag	300
tacatacatg	ctgtcgcgcc	gtttttgtct	gaatctcgct	tccactggcc	aatcagctca	360
gctcccggga	gctcactcat	tcaagatccc	atcgtcgtcg	tcacccctgg	cgtcatggga	420
tggaaaagaa	cctccggttg	tcggatgagt	cagccatata	cccgaacaga	gtactgcaag	480
ataacccaat	tcagattccc	ccaatagaga	aagtatagca	tgctttcggg	ttttgtttgg	540
cttaattgac	tttatttttg	ttggagttga	atgctgattt	gttggtgtaa	atgcccacc	600
atctgaatat	cgagacggat	aataggctgg	ctaattaatt	tatagcaaga	ttctgtagtg	660
cacatcgcaa	atatctttct	gggcattaca	gctggaggct	tcatacagct	gaaacactct	720
gcagagcctg	aagcaagtgg	tgaagcgtgg	cgatgagatg	ggtataaaac	ccccggcacc	780
gggacgcgag	ctcccgcta	ccagtaccat	ctcgccctcg	tccccctgcc	ggacgaccca	840
gtaaaatact	gttgcccact	cgccggcgag	atggcccacg	gccgcatact	cttcttggcg	900
ctcgccgtct	tggccaccgc	cgcggtggcc	gccgcatact	tggcggactc	caaccgcgac	960
cggcccgtca	ccgagcgcg	ggccgcc				987

<210> 49

<211> 40

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 49

ggcgccgagg	gagtggccgg	tcacggtcag	cgcgtagtcc	40
------------	------------	------------	------------	----

<210> 50

<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 50

ccggccacgc	cctcggcgcc	tcctggcgcc	cactc	35
------------	------------	------------	-------	----

<210> 51

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 51

ctaaagctta	ccatggcgcc	cgccctccacg	cagggcatct	ccga	44
------------	------------	-------------	------------	------	----

<210> 52
 <211> 50
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 52
 tctaagcttg cggccgcgac cggccaggtg catgcgccgc tcgtcatccc 50

<210> 53
 <211> 325
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> amplified nos terminator sequence from pMA406
 vector

<400> 53
 agactgcaga ccatggcggc cgcgkaacca ctgaaggatg agctgtaaag aagcagatcg 60
 ttcaaacatt tggcaataaa gtttcttaag attgaatcct gttgccggtc ttgcgatgat 120
 tatcatataa tttctgttga attacgttaa gcatgtaata attaacatgt aatgcatgac 180
 gttatttatg agatggggtt ttatgattag agtcccgcga ttatacatTT aatacgcgat 240
 agaaaacaaa atatagcgcg caaactagga taaattatcg cgcgcggtgt catctatgtt 300
 actagatcga taagcttcta gatct 325

<210> 54
 <211> 72
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 54
 agactgcaga ccatggcggc cgcgkaacca ctgaaggatg agctgtaaag aagcagatcg 60
 ttcaaacatt tg 72

<210> 55
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 55
 aagactgcag accatggcgg 20

<210> 56
 <211> 40
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 56

agatctagaa gcttatcgat ctagtaacat agatgacacc 40

<210> 57
 <211> 33
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 57
 ctaggcggcc gcgcgggagg aggcgacggc gac 33

<210> 58
 <211> 36
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 58
 gaggggtgtat tcggtatcga gttgcagggtt cgtatc 36

<210> 59
 <211> 32
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 59
 ctcgataccc attacaccct cagcctttc ga 32

<210> 60
 <211> 10
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Accl site

<400> 60
 gtaggtagac 10

<210> 61
 <211> 39
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> synthetic oligonucleotide

<400> 61
 ctcaccatgg taagcttcta cctacaaaaa agctccgca 39

<210> 62
 <211> 36
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 62
 aaccatggcg gccgcgcgct cggtgacggg ccggat 36

<210> 63
 <211> 29
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 63
 ttcggtacca tggccaggta taattatgg 29

<210> 64
 <211> 31
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 64
 ctgcgccggc gagatggmcg tgcacaagga g 31

<210> 65
 <211> 42
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 65
 ggaattcgta gacaagctta cmatggccca cgcccgcgtc ct 42

<210> 66
 <211> 55
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 66
 tatccatggc ggccgcgcgg tcggtgacgg gccggmycgg gttggagtcg gcgaa 55

<210> 67
 <211> 33
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 67
 ctaggcggcc gcgcgggagg aggcgacggc gac 33

<210> 68
 <211> 54
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 68
 gcgacggcga cggcggccgt ggccagcacg gcgagcgcca ggaggaggac gcgg 54

<210> 69
 <211> 33
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 69
 tcgccgtcgc ctctctctcc tcttcgccg act 33

<210> 70
 <211> 43
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 70
 actaagctta aggagatata acaatgatcc acaccaacct caa 43

<210> 71
 <211> 45
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 71
 ttccatgatc cacaccaacc tcaaaaagaa gttctccctc ttcac 45

<210> 72
 <211> 55
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 72
 agagtgatca cggcgaagag gaggaagacg aggatgaaga gggagaactt ctttt 55

<210> 73
 <211> 61
 <212> DNA
 <213> Artificial Sequence

<220>

<223> primer

<400> 73

tatagatctg cgtgtggaag aagggctccg actacgaggc cctcaccctc caagccaagg	60
a	61

<210> 74

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 74

catttggaac tccttggtt ggagggtg	28
-------------------------------	----

<210> 75

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 75

aaccatggcg gccgccattt ggaactcctt ggct	34
---------------------------------------	----

<210> 76

<211> 61

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 76

tatagatctg cgtgtggaag aagggctccg actacgaggc cctcaccctc caagccaagg	60
a	61

<210> 77

<211> 42

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 77

ggaattcgta gacaagctta cmatggmcgt gcacaaggag gt	42
--	----

<210> 78

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 78

gatcaggagg taggcwacga agttwacctc cttgtgc 37

<210> 79
 <211> 37
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 79
 cctacctcct gatcgtsctc ggcctcctct tgctcgt 37

<210> 80
 <211> 43
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 80
 ccttggcgtc cacgtgctcc atggcggawa cgagcaagag gag 43

<210> 81
 <211> 39
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 81
 gtggacgcca aggcctgcac cckcgagtgc ggcaacctc 39

<210> 82
 <211> 47
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 82
 ggaattcgcg gccgccgggc agatgccgaa gccgaggttg ccgcact 47

<210> 83
 <211> 32
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 83
 ctaaagctta acatgaagca gttctccgcc aa 32

<210> 84
 <211> 8
 <212> DNA
 <213> Oryza sp.

<400> 84
gtaggtag

<210> 85
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> KDEL fusion peptide

<400> 85
Lys Pro Leu Lys Asp Glu Leu
1 5

<210> 86
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> designated KDEL fusion peptide

<400> 86
Glu Pro Leu Lys Asp Glu Leu
1 5

<210> 87
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> frameshifted terminal peptide

<400> 87
Glu Thr Thr Glu Gly
1 5

<210> 88
<211> 27
<212> PRT
<213> Aspergillus niger

<400> 88
Met Lys Gln Phe Ser Ala Lys His Val Leu Ala Val Val Val Thr Ala
1 5 10 15
Gly His Ala Leu Ala Ala Ser Thr Gln Gly Ile
20 25

<210> 89
<211> 9
<212> PRT
<213> Aspergillus niger

<400> 89
Met Ala Ala Ala Ser Thr Gln Gly Ile
1 5

<210> 90

<211> 7
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> ER retention vector

<400> 90
 Lys Pro Leu Lys Asp Glu Leu
 1 5

<210> 91
 <211> 5
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> c-terminal targeting sequence

<400> 91
 Pro Val Ala Ala Ala
 1 5

<210> 92
 <211> 55
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 92
 tatccatggc ggccgcgcgg tcggtgacgg gccggcccgg gttggagtcg gcgaa 55

<210> 93
 <211> 8
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> ferulic esterase end

<400> 93
 Cys Thr Trp Pro Val Ala Ala Ala
 1 5

<210> 94
 <211> 72
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> pTP4a2 vector

<400> 94
 Met Lys Gln Phe Ser Ala Lys His Val Leu Ala Val Val Val Thr Ala
 1 5 10 15
 Gly His Ala Leu Ala Ala Ser Thr Gln Gly Ile Ser Glu Asp Leu Tyr
 20 25 30
 Ser Arg Leu Val Glu Met Ala Thr Ile Ser Gln Ala Ala Tyr Ala Asp
 35 40 45

What is claimed:

1. A transgenic *Festuca* or *Lolium* plant cell comprising an expression cassette comprising a promoter operably linked to a polynucleotide as set forth in SEQ ID NO: 1 or encoding a ferulic acid esterase set forth in SEQ ID NO: 2 and a signal sequence that targets expression of the ferulic acid esterase to the endoplasmic reticulum, vacuole, apoplast or golgi apparatus.
2. The plant cell of claim 1, wherein the polynucleotide is FAE I from *Aspergillus niger*.
3. The plant cell of claim 2, wherein the polynucleotide further comprises a polynucleotide that encodes CTWPVAAA at the 3' end.
4. The plant cell of claim 2 wherein sub-optimal codons are modified to *Triticum spp.* preferred codons.
5. The plant cell of claim 1, wherein the promoter is an inducible promoter.
6. The plant cell of claim 5, wherein the promoter is a senescence promoter.
7. The plant cell of claim 5, wherein the promoter is a heat shock promoter.
8. The plant cell of claim 1, wherein the promoter is a constitutive promoter.

9. The plant cell of claim 1, wherein the signal sequence is upstream of the 5' end of the ferulic acid esterase polynucleotide.

10. The plant cell of claim 1, wherein the signal sequence is the signal sequence of a golgi targeted gene.

11. The plant cell of claim 10, wherein the golgi targeted gene encodes a rat sialyl transferease signal sequence.

12. The plant cell of claim 1, wherein the signal sequence is an apoplast signal sequence.

13. The plant cell of claim 12, wherein the signal sequence is from *Aspergillus niger* ferulic acid esterase.

14. The plant cell of claim 1, wherein the signal sequence is downstream of the 3' end of the ferulic acid esterase polynucleotide.

15. The plant cell of claim 14, wherein the signal sequence encodes a KDEL sequence.

16. The plant cell of claim 1, further comprising a second expression cassette comprising a promoter operably linked to a xylanase encoding polynucleotide.

17. The plant cell of claim 16, wherein the xylanase encoding polynucleotide is from *Trichoderma reesei*.

18. The plant cell of claim 16, wherein the first and second expression cassettes are present on separate plasmids.

19. A method of increasing a level of phenolic acids in plant cell walls of a *Festuca* or *Lolium* transgenic plant relative to a level in corresponding plants for which an expression cassette has not been introduced, the method comprising introducing into the plant an expression cassette comprising a promoter operably linked to a polynucleotide as set forth in SEQ ID NO: 1 or encoding a ferulic acid esterase as set forth in SEQ ID NO: 2 and a signal sequence that targets expression of the ferulic acid esterase to the endoplasmic reticulum, vacuole, apoplast or golgi apparatus.

20. The method of claim 19, wherein the polynucleotide is a FAE 1 gene from *Aspergillus niger*.

21. The method of claim 20, wherein the polynucleotide encodes a sequence that comprises CTWPVAAA at the 3' end.

22. The method of claim 20 wherein sub-optimal codons are modified to *Triticum spp.* preferred codons.

23. The method of claim 19, wherein the promoter is an inducible promoter.

24. The method of claim 19, wherein the promoter is a senescence promoter.

25. The method of claim 23, wherein the promoter is a heat shock protein promoter.

26. The method of claim 19, wherein the promoter is a constitutive promoter.

27. The method of claim 19, wherein the signal sequence is upstream of the 5' end of the polynucleotide encoding a ferulic acid esterase.

28. The method of claim 27, wherein the signal sequence is the signal sequence of a golgi targeted gene.

29. The method of claim 28, wherein the golgi targeted gene encodes a rat sialyl transferease signal sequence.

30. The method of claim 27, wherein the signal sequence is a fungal apoplast signal sequence.

31. The method of claim 30, wherein the signal sequence is from *Aspergillus niger* ferulic acid esterase.

32. The method of claim 19, wherein the signal sequence is downstream of the 3' end of the polynucleotide encoding a ferulic acid esterase.

33. The method of claim 31, wherein the signal sequence encodes a KDEL sequence.

34. The method of claim 19, further comprising simultaneous introduction into the plant of a second expression cassette comprising a promoter operably linked to a polynucleotide encoding a xylanase.

35. The method of claim 34, wherein the second polynucleotide encodes a fungal xylanase.

36. The method of claim 35, wherein the fungal xylanase is from *Trichoderma reesei*.

37. The method of claim 34, wherein the first and second expression cassettes are present on separate plasmids.

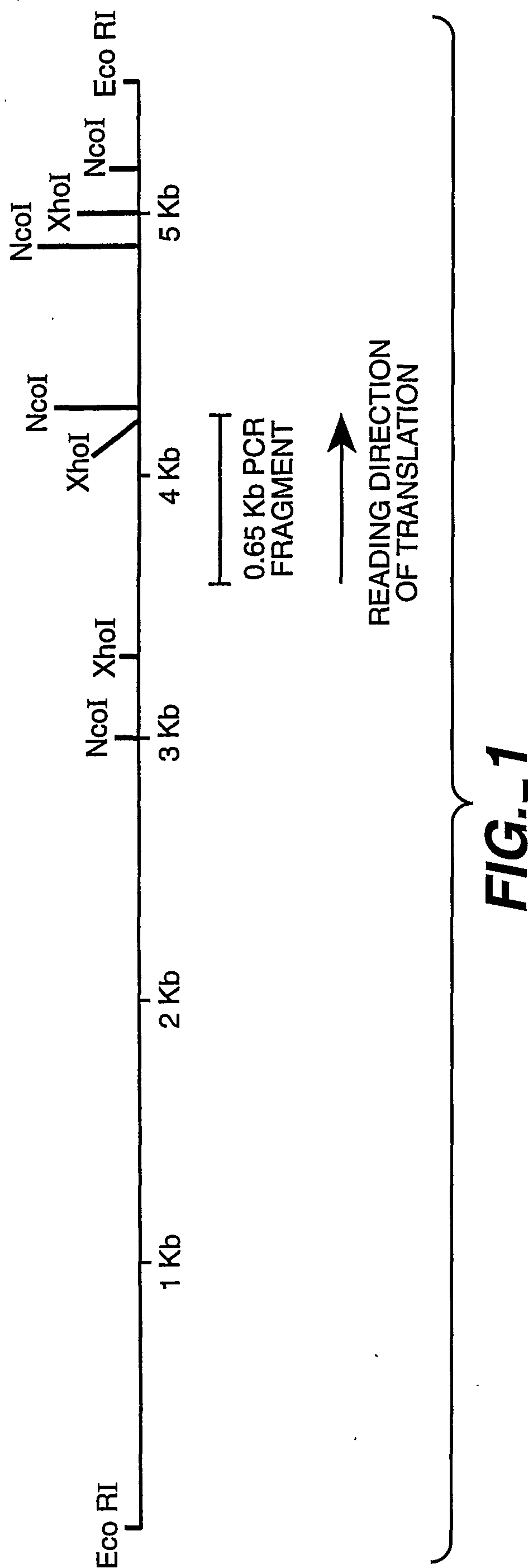


FIG. 2A

FIG. 2B

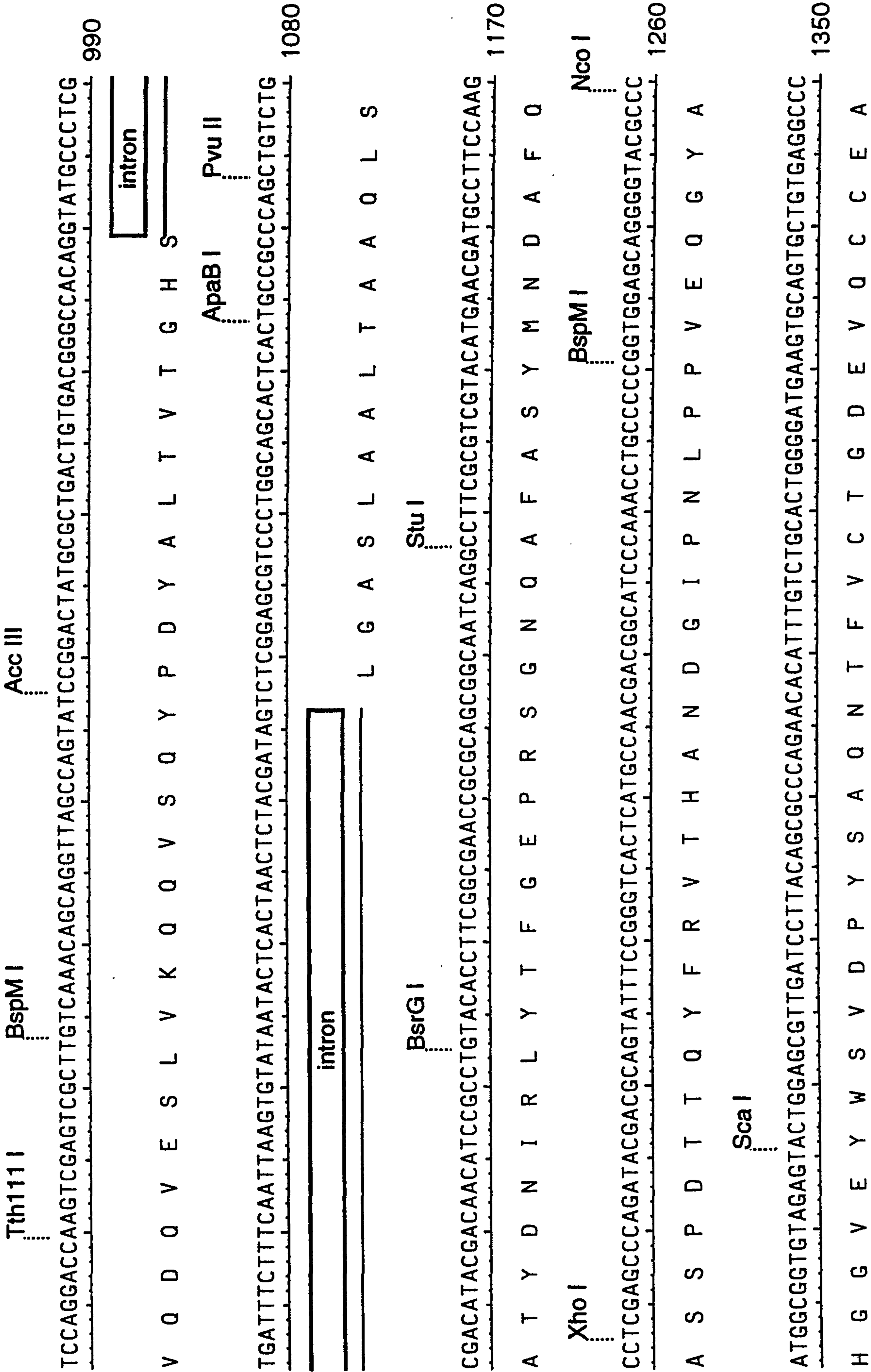


FIG._2C

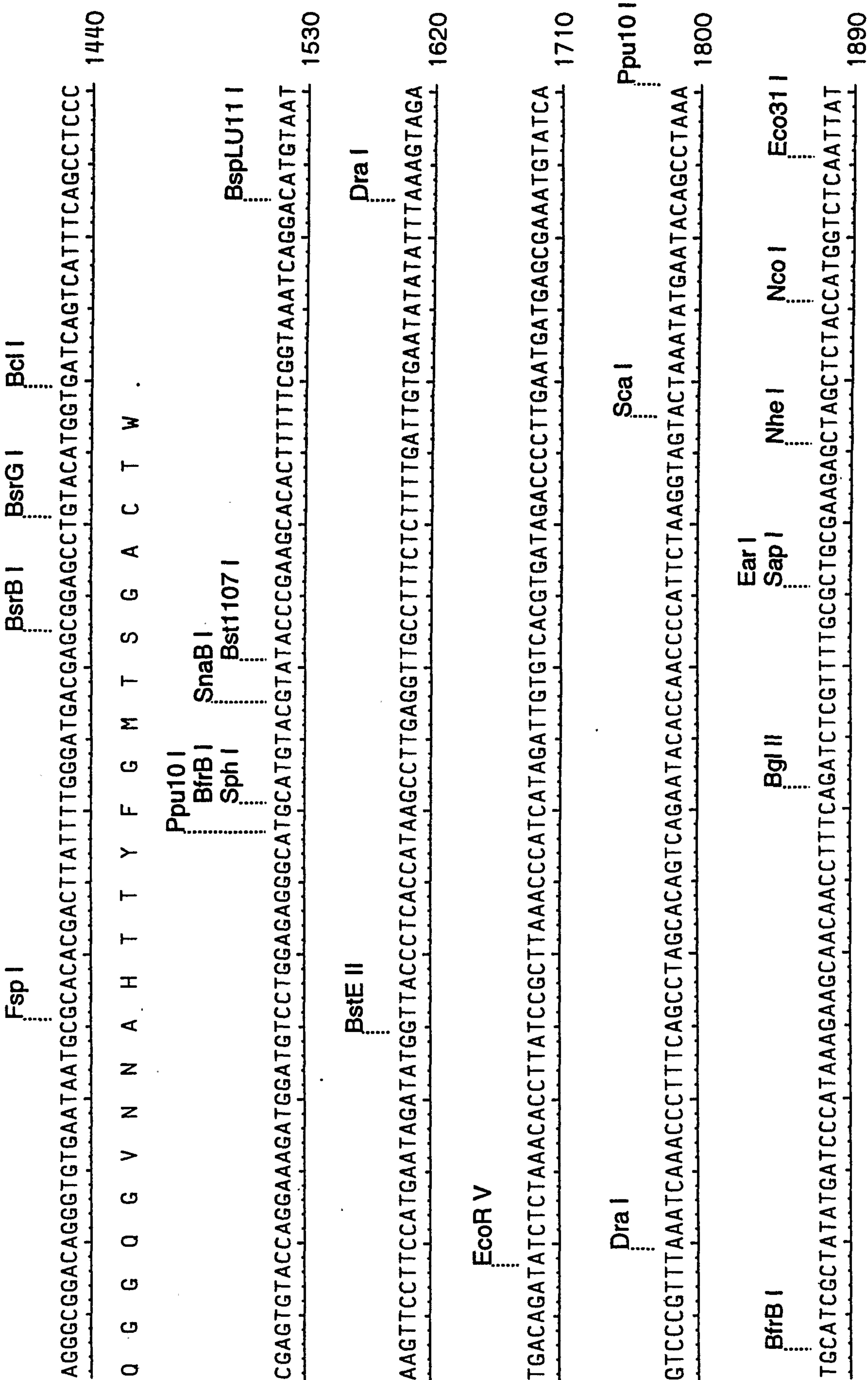


FIG.-2D

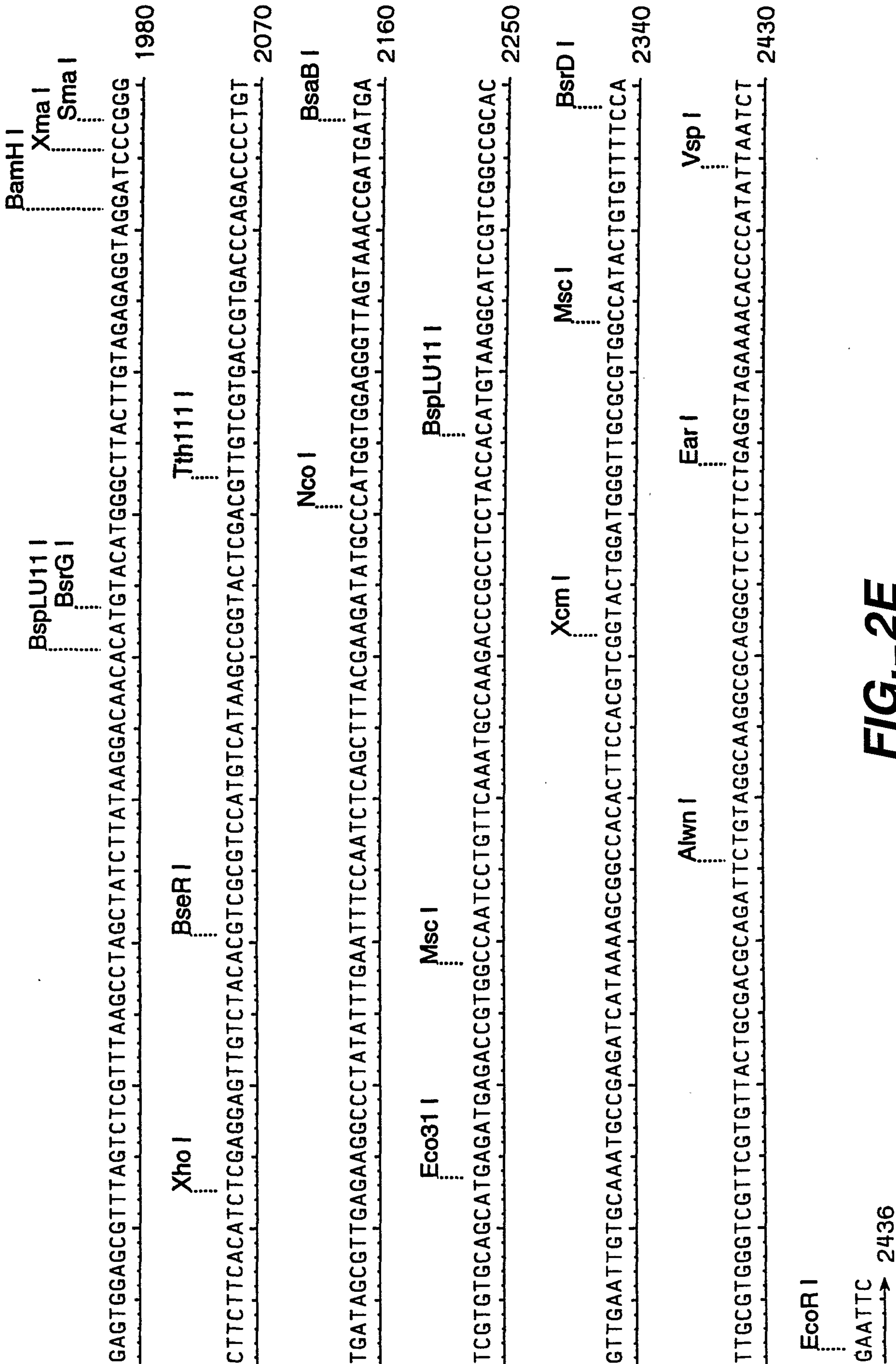


FIG.-2E

CCATGGTGGTGCATATCGGCAGTAGTCTTTGCCGAAACGTTGAGGGTTACAGTGTATCTGCGTCGGACATACTT
CGGGGAATCTACGGCGGAATATCAAAGTCTTCGGAATATCCATATTGGGAAAGGACAGAAAGCTCCGGGGTAGTTT
GATAGATGAGCTCCGGTGTATTAAATCGGGAGCTGACAGGAGTGAAGCTCATGTAGACCATCTAGTAATGTCAGT
CGCGCGCAATTTCGCACATGAACAAGTTGATTTCCGGACCCCATTTGTTACATCTCTCGGCTACAGCTCGAGATG
TGCCCTGCCGAGTACTTAGAAGCCATGCCAGCTTCAAAATTAGTCTTCAGTGGTTTATTAAAGAGATAGATCCCTTCACAAACACT
CATCCAACGGACTTCTCATACCACTCAATTGACATAATTCAACAGCTCCAGGCGCATTTAGTTCAACATGAAGC
AATTCCTCCGCAACACAGTCTCTCGAGTTGTGGTGACTGAGGCAAGCTTAGCAGCTCTACGCAAGGCATCT
CCGAAGACCTCTACAGCCGTTTAGTCCGAATGGCCAATACTCCCAAGCTGCTACGCCGACCTGTGCAACATTC
CGTCGACTATTATCAAGGAGAGAAATTTACAATTCTCAAACTGACATTAAACGGATGGAATCCTCCCGGACGACA
GCAGCAAAAGAAATAATCACCCGCTTCCGTGGCACTGGTAGTACGAATCTACAACCTCGATACATACTAACACCC
TCACGCCCTTCGACACCCCTACCAATAATGCAACGGTTGTGAAGTACACGGTGGATAATTATTTGGATGGGTCTCCG
TCCAGGACCAAGTCGAGTCGCTTGTCAAAACAGCAGGTTAGCCAGTATCCGGACTATGCGCTGACTGTGACGGGCC
ACAGGTATGCCCTCGTGATTCTTCAATTAGTGTAATACTCACTCACTACGATAGTCTCGGAGCGTCCC
TGGCAGCACTCACTGCCGCCAGCTGTCTCGGACATACGACAACATCCGCCCTGTACACCTTCGGCGAACCGCGCA
GCGGCAATCAGGCCCTTCGCGTCGTACATGAACGATGCCCTCGAGCCAGATACGACGAGTATTTC
GGTCACTCATGCCAACGACGCGCATCCCAAACCTGCCCTGGAGCAGGGGTACGCCCATGGCGGTGAGAGT
ACTGGAGCGTTGATCCCTTACAGCGGCCCAACAATTTGTCTGCACTGGGGATGAAGTGCAAGTGTGAGGCC
AGGCGGACAGGGTGTGAATAATGCGCACACGACTTATTTGGGATGACGAGCGGCA TGACATGGTGATCAG
TCATTTTCAGCCCTCCGAGTGTAACAGGAAGATGGATGTCCTGAGAGGCA TGCA TGATACCCGAAGC
ACACTTTTTCGGTAATCAGGACATGTAAAGTTCCATGAATAGATA TGTTACCTCACCATAGCCCTT
GAGGTTGCCCTTCTCTTTGATTGTGAATATATATTAAAGTAGACAGATATCTCTAAACACCTTATCCGCT
TAAACCCATCATAGATTGTGTACGTGATAGACCCCTTGAAATGATGAGCGAAATGTATCAGTCCCGTTTAAATCA
AACCCTTTCAGCCTAGCACAGTCAGAAATACACCAACCCATTCTAAGGTAGTACTAAATATGAATACAGCCCTAA
TGCAATCGCTATATGATCCCATAAAGAAAGCAACCTTTCAGATCTCGTTTTCGCTGCGAAGAGCTAGCTCTAC
CATGGTCTCAATTATGAGTGGAGCGTTTAGTCTCGTTTAAAGCCTAGCTATCTTATAAGGACAAACATGTACATG
GGCTTACTTGTAGAGAGGTAGGATCCGGGCTTCTTCACATCTCGAGAGTTGTCTACACGTCCGTCATGTCA
TAAGCCGGTACTCGACGTTGTGTCGTGACCCAGACCCCTGTTGATAGCGTTGAGAAAGCCCTATATTGAA
TTTCCAATCTCAGCTTTACGAAGATATGCCCATGGTGGAGGTTAGTAAACCGATGATCGTGTGAGCATGA
GATGAGACCGTGCCCAATCCTGTTCAAAATGCCAAGACCCGCTCTACCATGTAAAGGCATCCGTCCGGCAC
GTTGAATTGTGCAAAATGCCGAGATCATAAAGCGGCACATCCACGTCCGTTACTGGATGGGTGCGGCC
ATACTGTGTTTTCATTTGCGTGGTCTGTTACTGCGACGAGATTCTGTAGGCAAGCGCAGGGCTCTCT
TCTGAGGTAGAAACACCCCATATTAAATCTGAATTC

FIG.-3

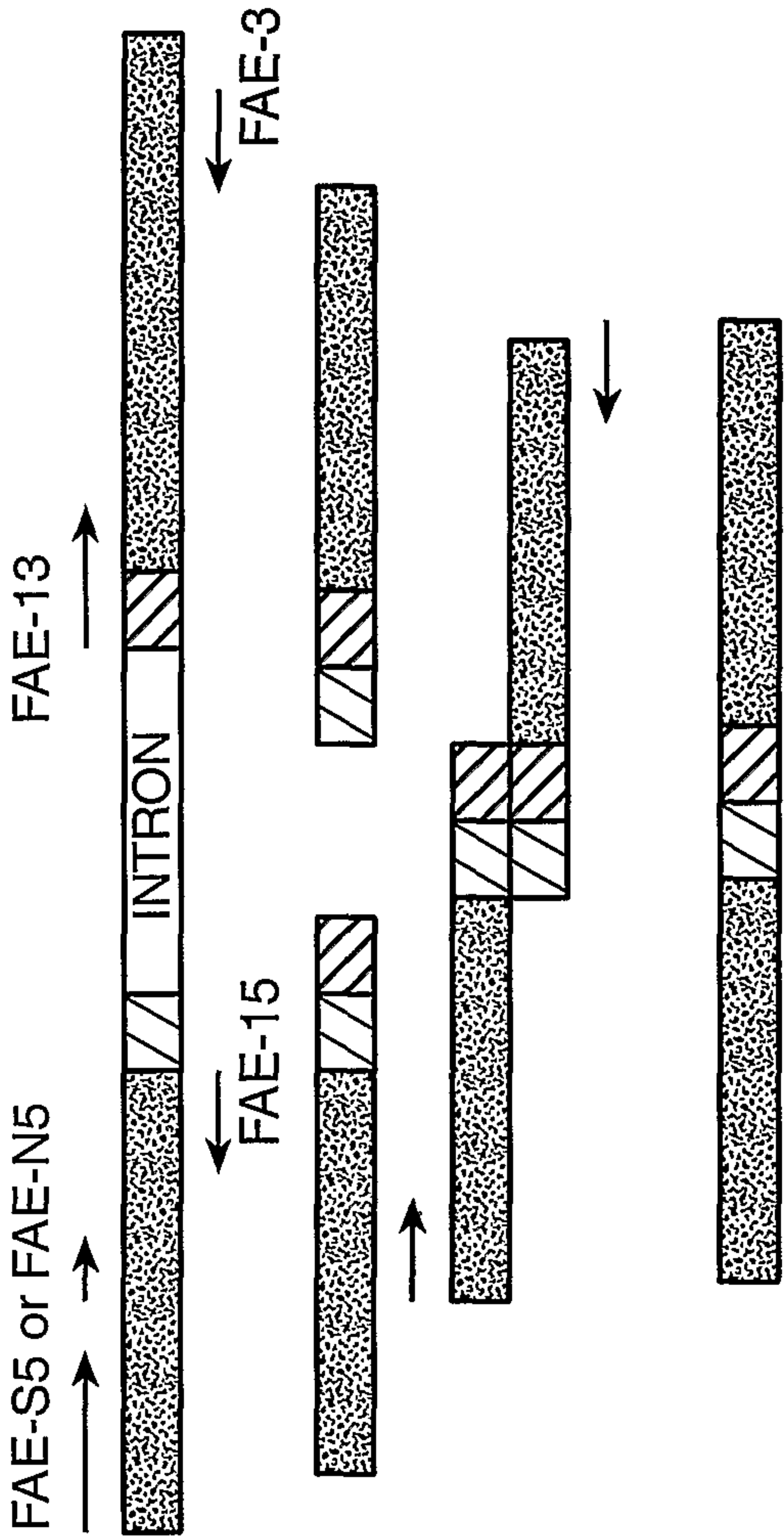


FIG. 4

FAE-I3 CCGGCCACGCCCTCGGGCCTCCCTGGCGGCACTC 35-mer
FAE-I5 GCGGCCGAGGAGTGCGCCGGTCACTCGGCGTAGTCC 40-mer

intron position in original

Y A L T V T G H S L G A S L A A L
GGACTACGCGCTGACCGTGACCGGCCACTCCCTCGGCGCC
CCGGCCACGCCCTCGGGCCTCCCTGGCGGCACTC
Y A L T V T G H A L G A S L A A L

complement, FAE-I5
FAE-I3

FIG. 5

Vector Construction

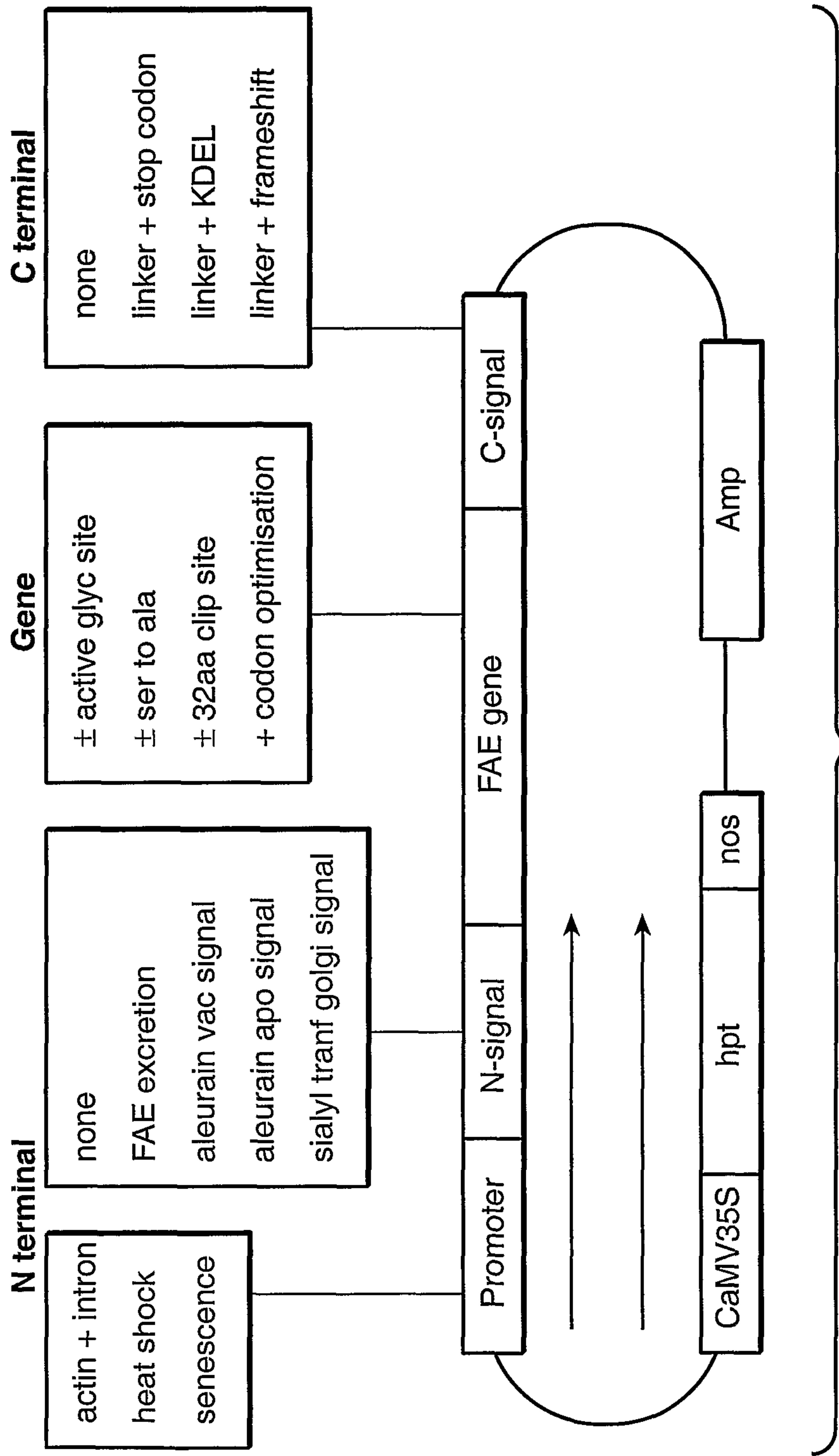
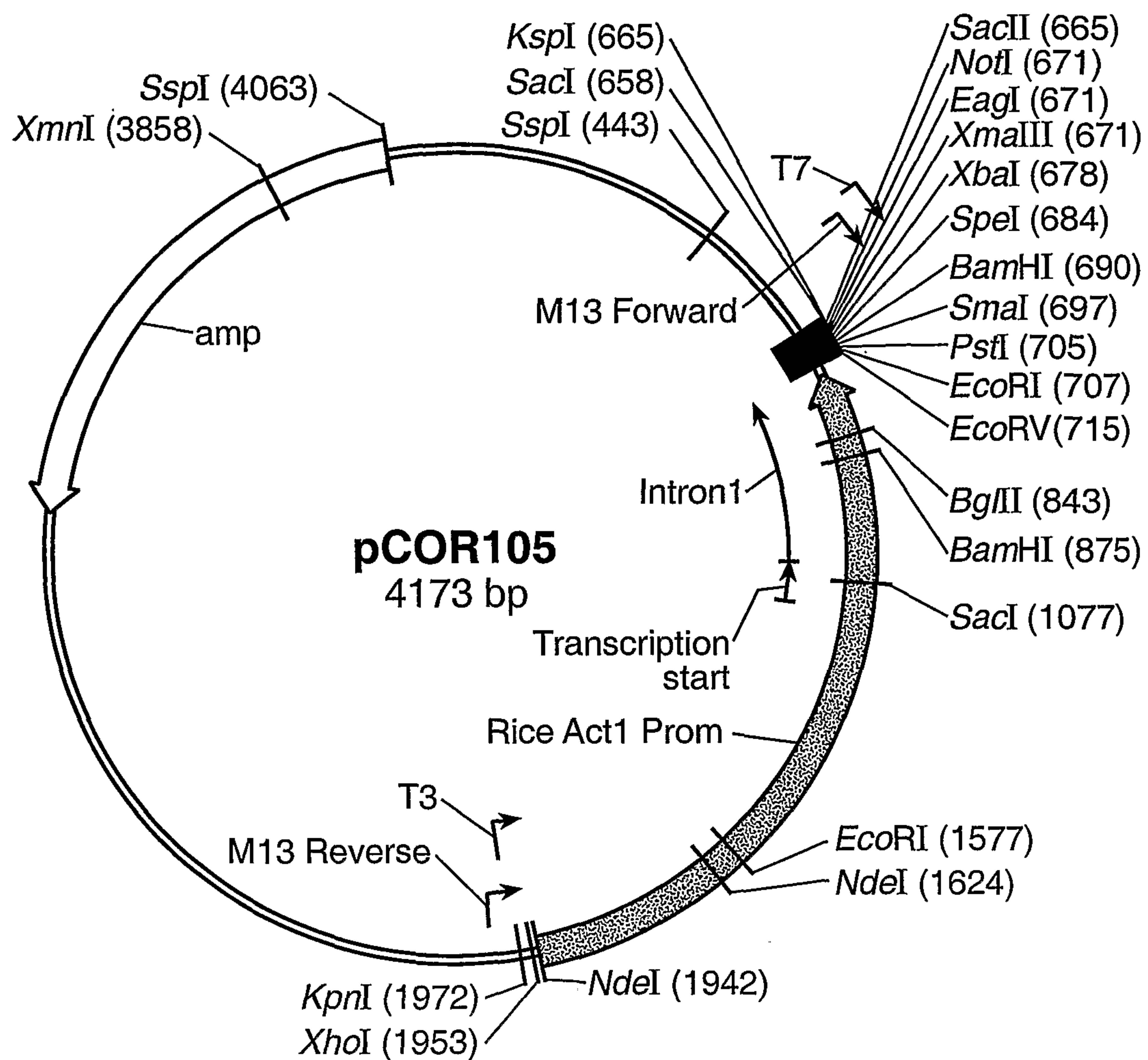
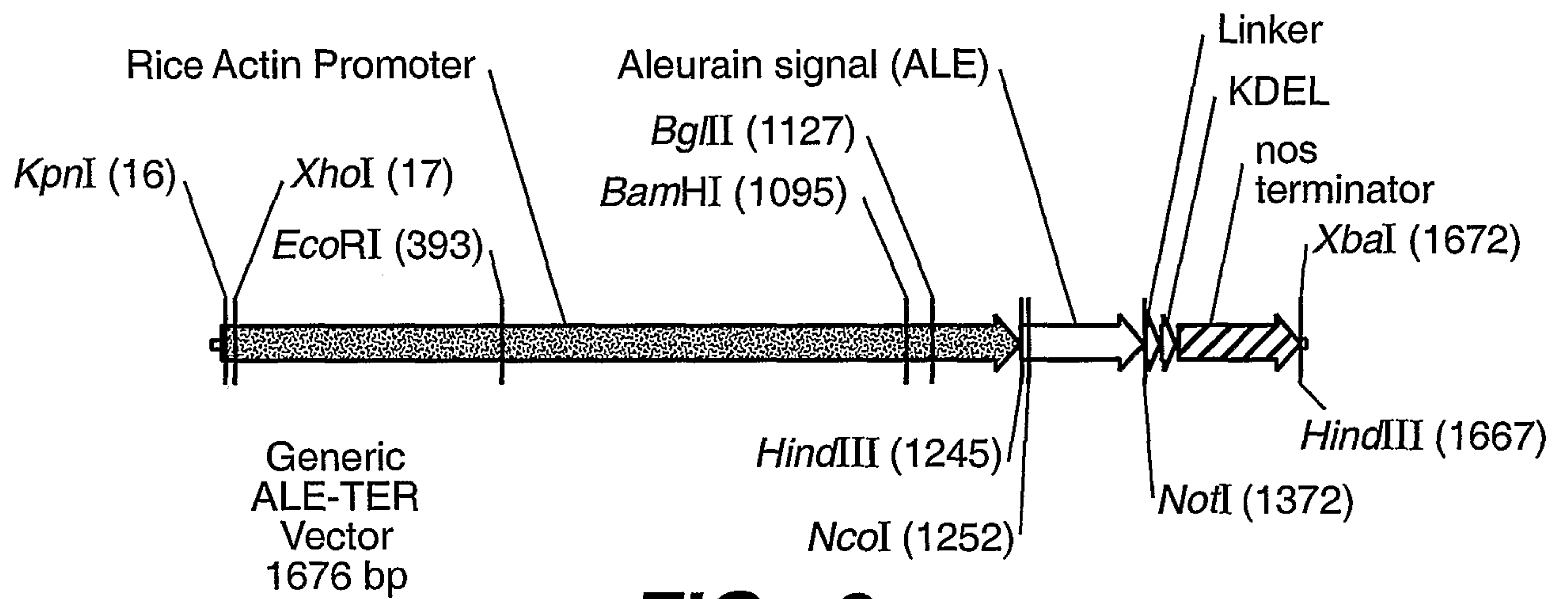


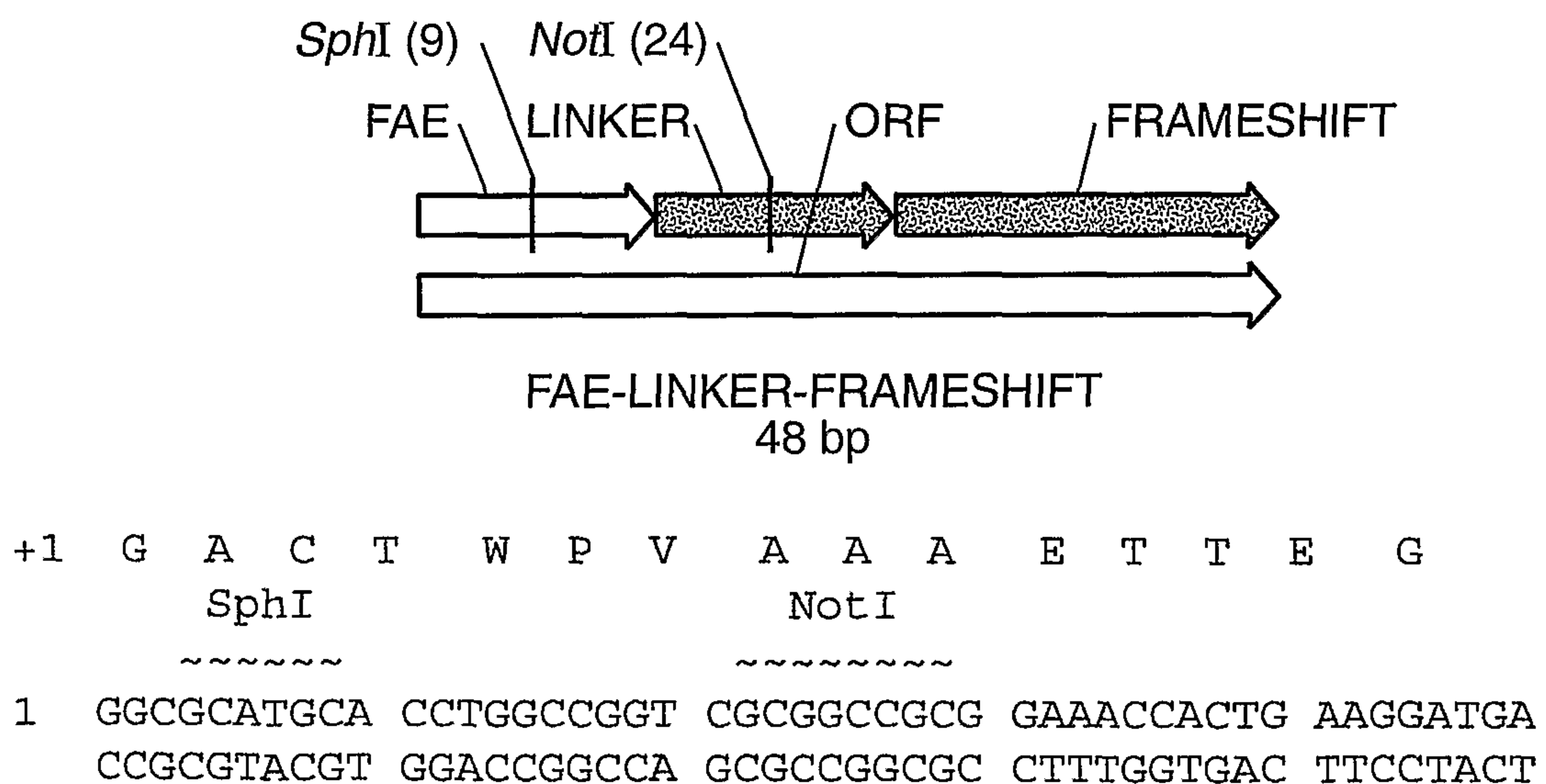
FIG._6

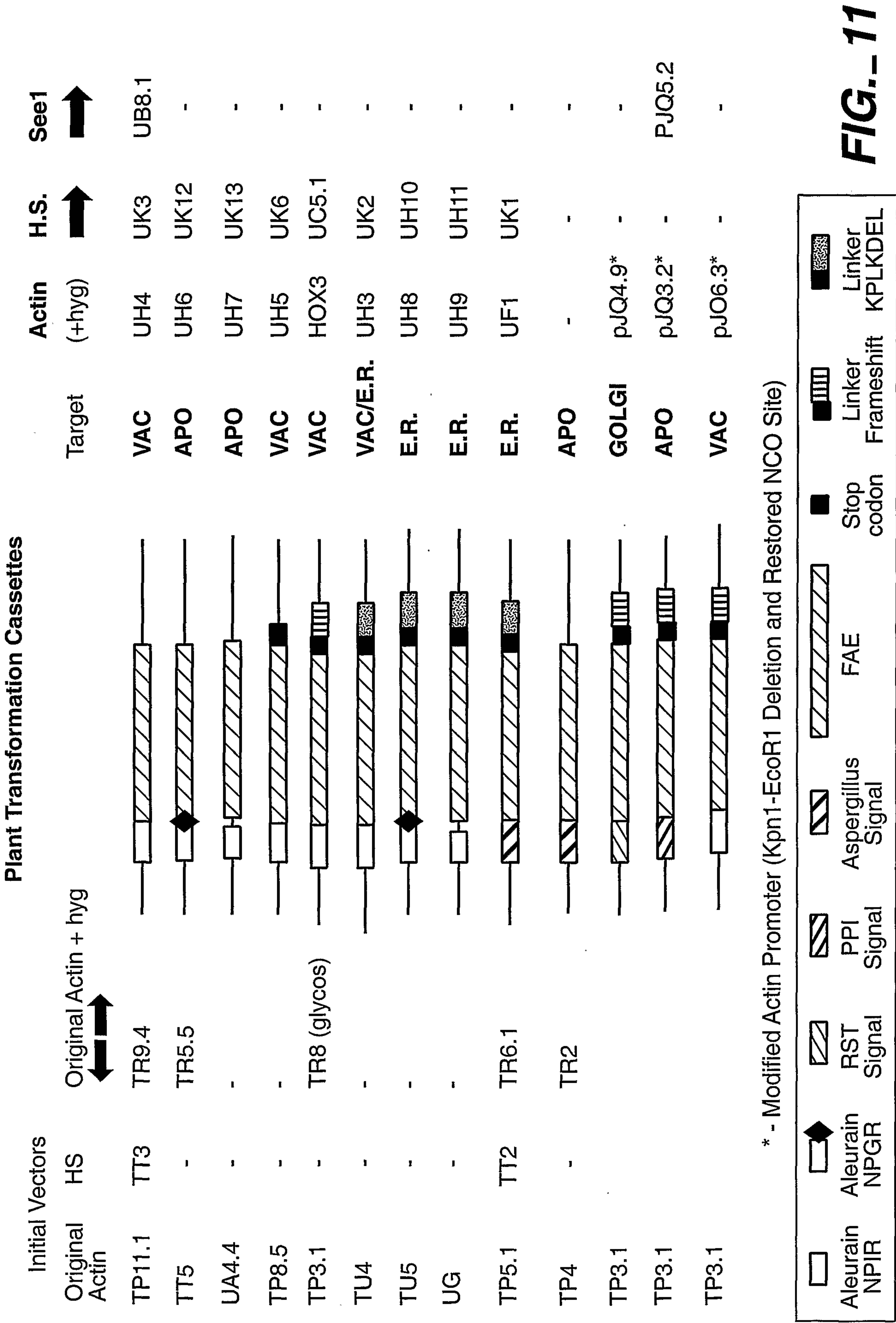
**FIG. 7**

**FIG._8**

KDEL-COOH ER retention sequence

NotI
~~~~~  
A A A K P L K D E L \*  
1 GCGGCCGCGA AACCACTGAA GGATGAGCTG TAA

**FIG.\_9****FAE-LINKER-FRAMESHIFT Structure and Sequence****FIG.\_10**



**Vectors****Original Actin promoter in pCOR105**

|       | <b>Target</b> | <b>Signal sequences</b>               | <b>Vectors</b>               |
|-------|---------------|---------------------------------------|------------------------------|
| (i)   | APO           | - aleurain-NPGR-FAE                   | pUH6, pTT5, TT5.5, pTT5.1    |
|       |               | - aleurain-delNPIR -FAE               | pUH7, pUA4.4,                |
| (ii)  | ER            | - aleurain-NPGR-FAE-linker-KDEL       | pTU5, pUH8,                  |
|       |               | - aleurain-delNPIR-FAE-linker-KDEL    | pUG4, pUH9,                  |
| (iii) | VAC           | - aleurain-NPIR-FAE                   | pTP11.1, pTR9.4, pUH4, pUK3, |
| (iv)  | ER/VAC        | - aleurain-NPIR-FAE-linker-KDEL       | pTU4, pUH3,                  |
| (v)   | VAC           | - aleurain-NPIR-FAE-linker-frameshift | pUA1K3, pTP3.1, pUC5.11      |
| (vi)  | VAC           | - aleurain-NPIR-FAE-linker-stop       | pTP8.5, pUH5                 |
| (vii) | ER            | - Aspergillus signal -FAE-KDEL        | pTP5.1, pTP6.1, pUF1,        |

**Modified actin promoter (Kpn1-EcoR1 deletion and restored NCO site)**

|       |       |                                       |        |
|-------|-------|---------------------------------------|--------|
| (i)   | VAC   | - aleurain-NPIR-FAE-linker-frameshift | pJ06.3 |
| (ii)  | GOLGI | - RST-FAE-linker-frameshift           | pJQ3.2 |
| (iii) | APO   | - PPI-FAE-linker-frameshift           | pJQ4.9 |

**Heat-shock promoter**

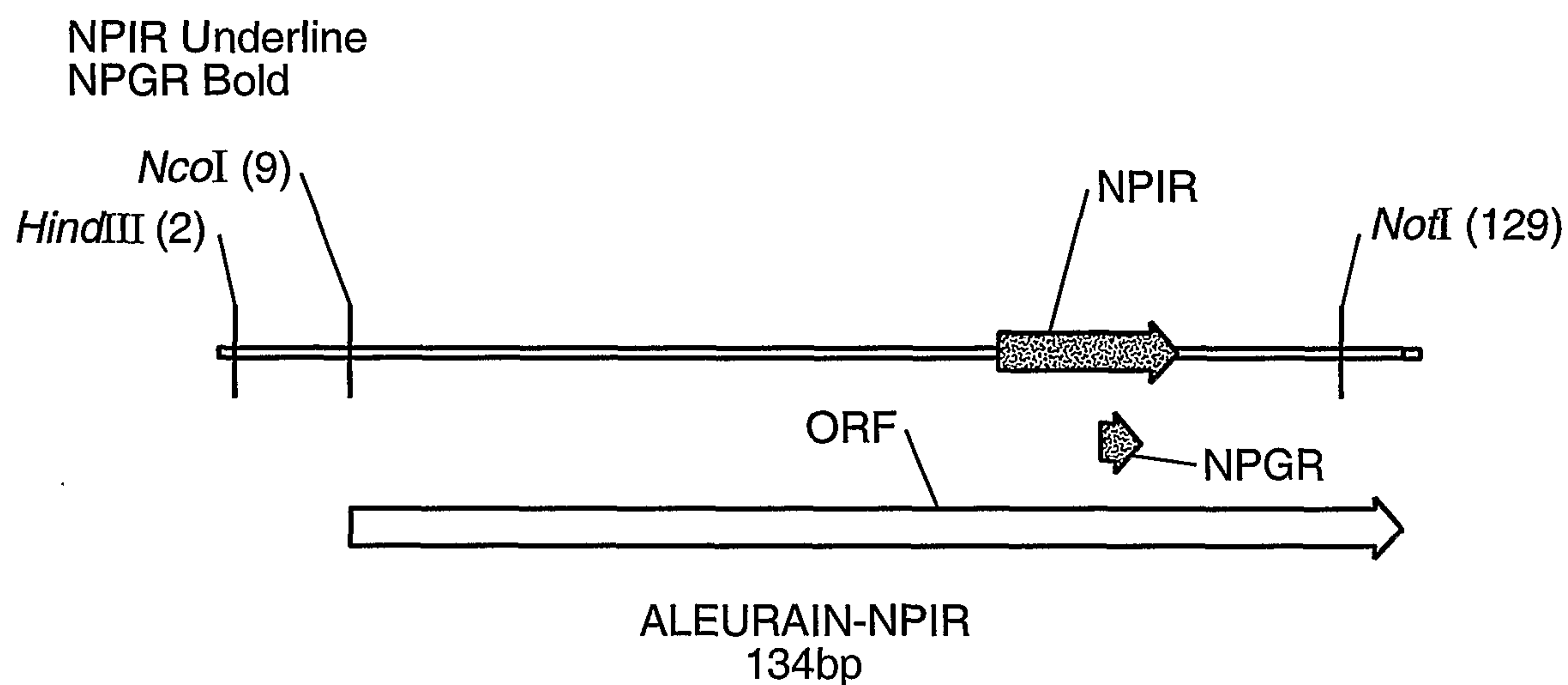
|       |        |                                       |                  |
|-------|--------|---------------------------------------|------------------|
| (i)   | APO    | - aleurain-NPGR-FAE                   | pUH12            |
|       |        | - aleurain-delNPIR-FAE                | pUH13            |
|       |        | - Aspergillus signal-FAE              | pTP4a2, pTR2.22, |
| (ii)  | ER     | - aleurain-NPGR-FAE-linker-KDEL       | pUH10            |
|       |        | - aleurain-delNPIR-FAE-linker-KDEL    | pUH11            |
| (iii) | VAC    | - aleurain-NPIR -FAE                  | pUK3,pTT3        |
| (iv)  | ER/VAC | - aleurain-NPIR-FAE-linker-KDEL       | pUK2             |
| (v)   | VAC    | - aleurain-NPIR-FAE-linker-frameshift | pUC5.11, pHOX3   |
| (vi)  | VAC    | - aleurain-NPIR-FAE-linker-stop       | pUK6             |
| (vii) | ER     | - Aspergillus signal -FAE-KDEL        | pUK1, pTT2       |

**Senescence promoter**

|      |     |                                  |        |
|------|-----|----------------------------------|--------|
| (i)  | APO | - See1-PPI-FAE-linker-frameshift | pJQ5.2 |
| (ii) | VAC | - See1-aleurain-deleted NPIR-FAE | pUB8.1 |

**FIG. 12**

# **ALEURAIN-NPIR (Vacuolar) and NPGR (Apoplast) Structure and Sequence**



```

+1      M  A  H  A  R  V  L  L  L  A  L  A  V  L  A  T  A  A  V  A
HindIII NcoI
~~~~~
1 AAGCTTACCA TGGCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTCGCCG
 TTCGAATGGT ACCGGGTGCG GGCAGCAGGAG GAGGACCGCG AGCGGCACGA CCGGTGCCGG CGGCAGCGGC

+1 V A S S S S F A D S N P I R P V T D R A A
 NotI
                                   ~~~~~
71  TCGCCTCCTC CTCCTCCTTC GCCGACTCCA ACCCGATCCG GCCCGTCACC GACCGCGCGG CCGC
   AGCGGAGGAG GAGGAGGAAG CGGCTGAGGT TGGGCTAGGC CGGGCAGTGG CTGGCGCGCC GGCG
  
```

**FIG. 13**

RAT SIALYL TRANSFERASE Golgi signal sequence

HindIII  
~~~~~  
M I H T N L K K K F S L F I L V F L L F A
1 AAGCTTACCA TGATCCACAC CAACCTCAA AAGAA GTTCT CCTCTTTCAT CCTCGTCTTC CTCCTCTTCG
V I C V W K K G S D Y E A L T L Q A K E F Q M
71 CCGTGATCTG CGTGTGGAAG AAGGGCTCCG ACTACGAGGC CCTCACCCCTC CAAGCCAAGG AGTTCCAAAT
NotI
~~~~~  
A A  
141 GGCGGCCCGC

FIG.\_14

POTATO PROTEASE INHIBITOR II Apoplast signal sequence

HindIII  
~~~~~  
M X V H K E V N F V A Y L L I V L G L L L
1 AAGCTTACMA TGGMCGTGCA CAAGGAGGTS AACTTCGTSG CCTACCTCCT GATCGTSCTC
GGCCTCCTCT
NcoI
~~~~~  
L V S A M E H V D A K A C T X E C G N L  
G F G  
71 TGCTCGTSTC CGCCATGGAG CACGTGGACG CCAAGGCCTG CACCCKCGAG TCGGCAACC  
TCGGCTTCGG  
NotI  
~~~~~  
I C P A A A
141 CATCTGCCCG GCGGCCGC

FIG._15

Targeting Expression of gfp to Different Cell Compartments

Actin Promoter Targeting Vectors

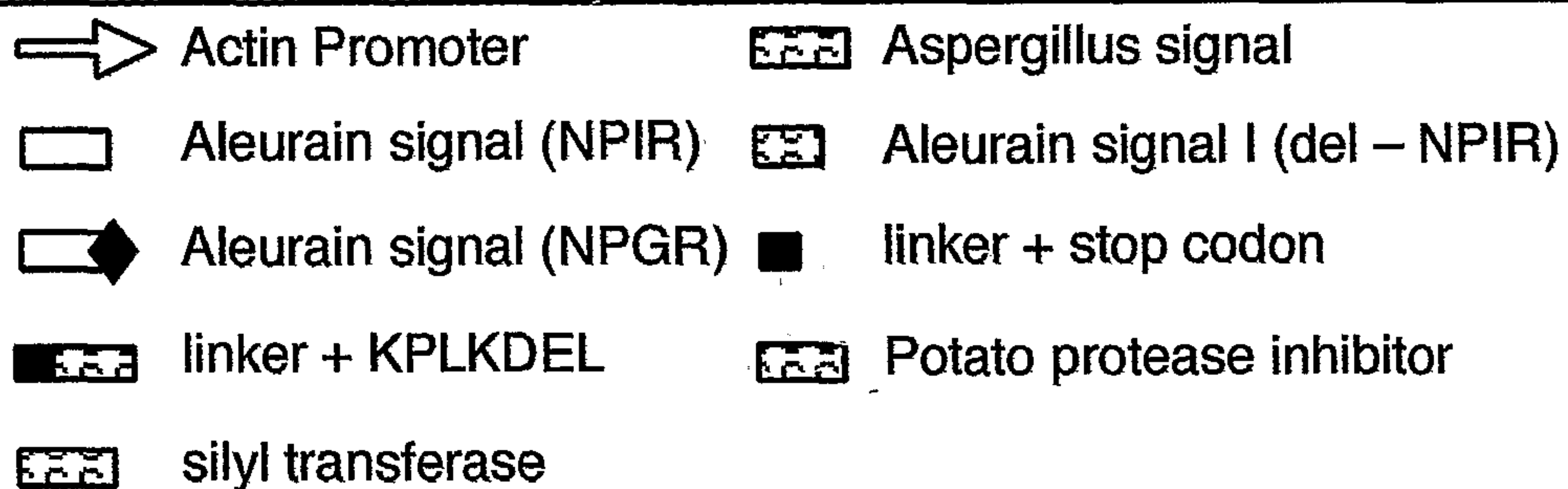
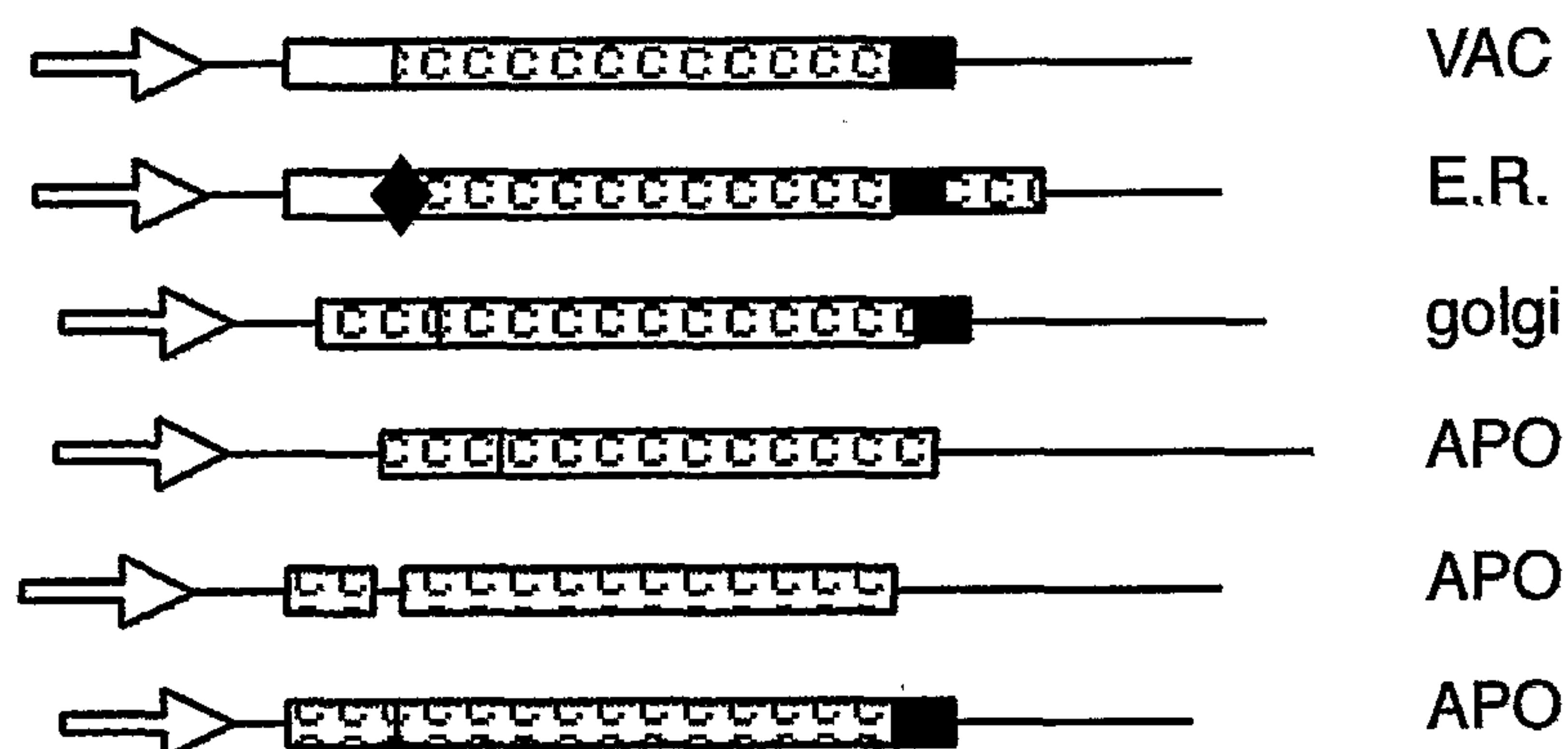


FIG. 16A

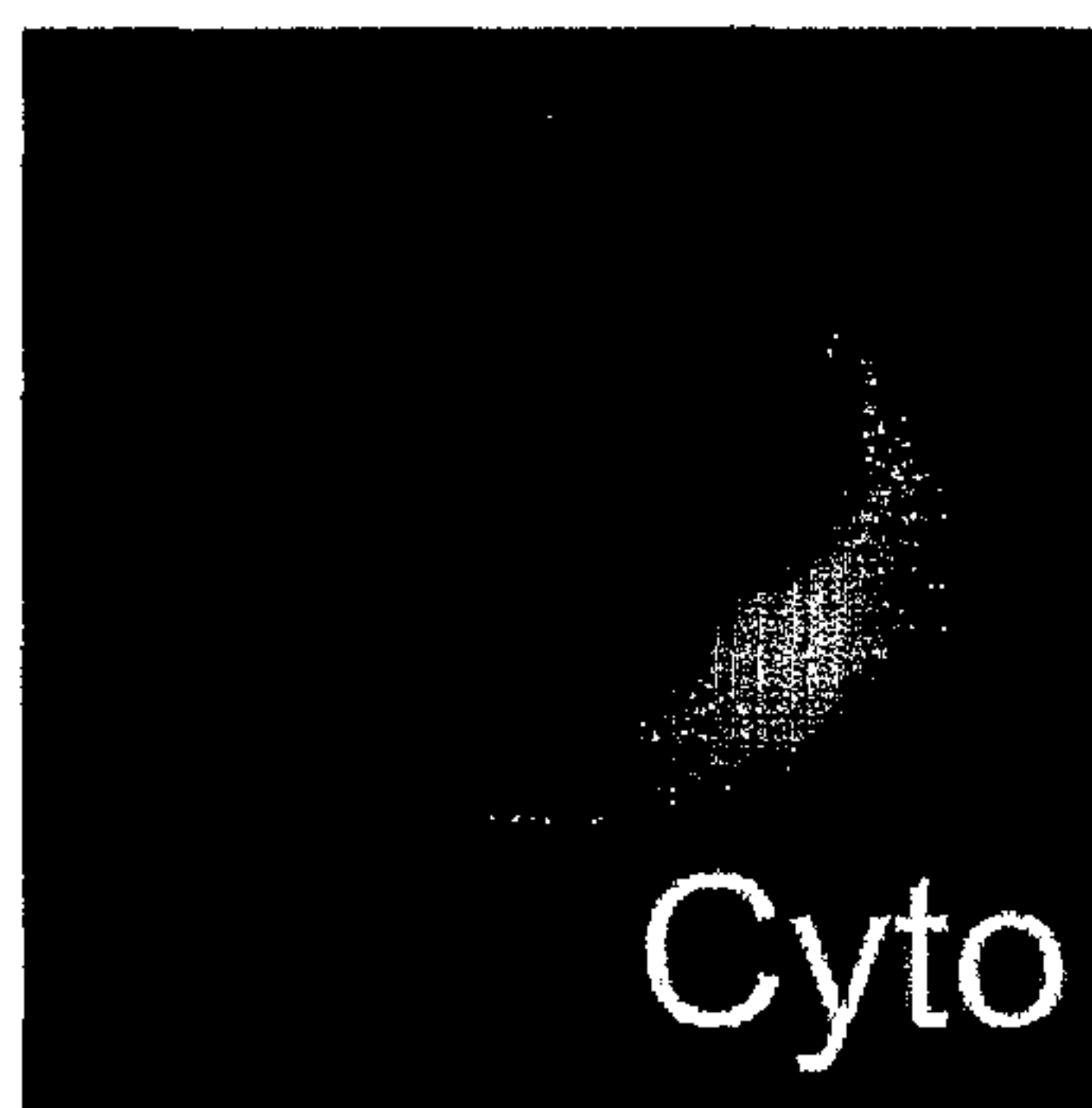


FIG. 16B

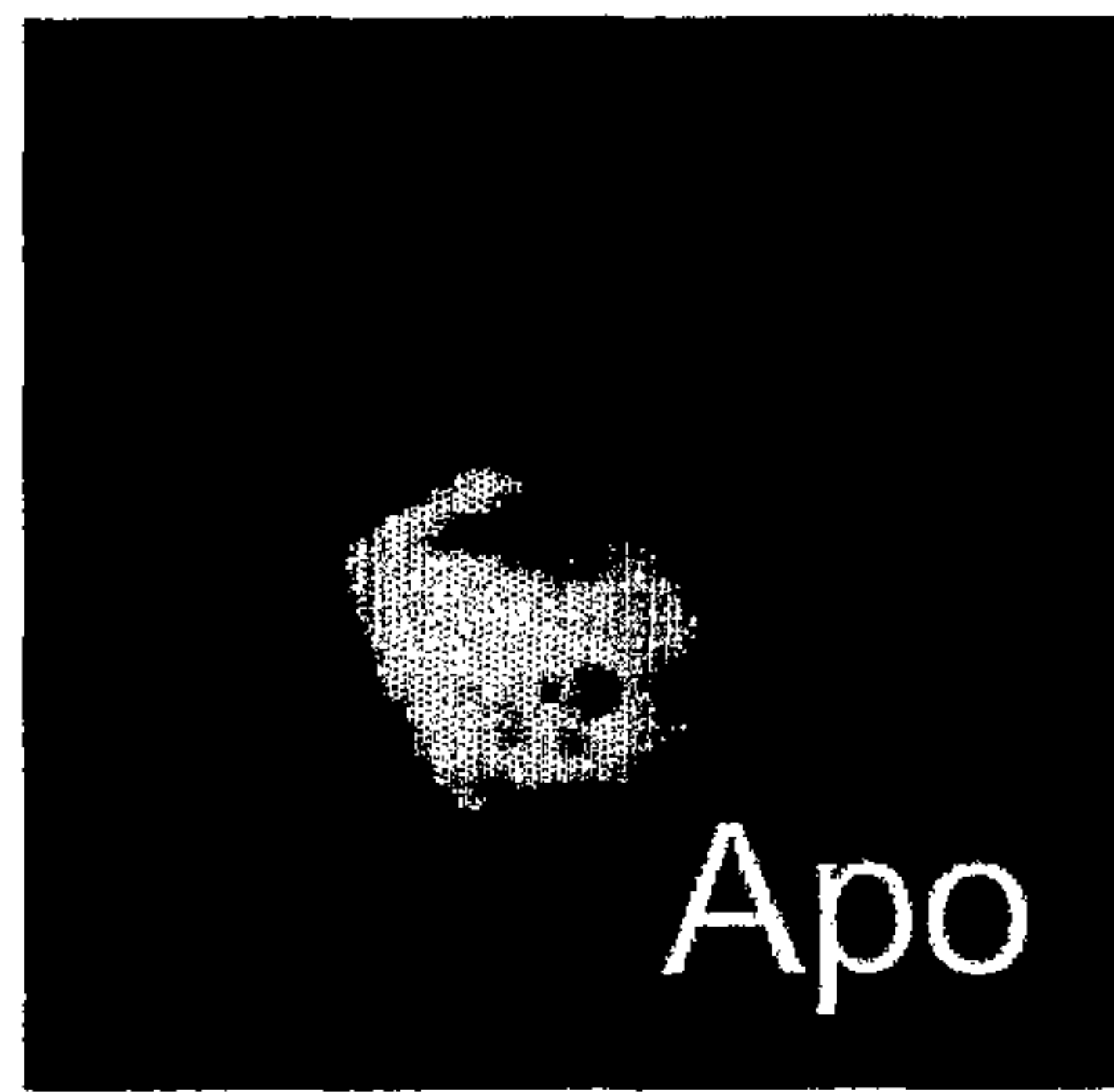


FIG. 16C

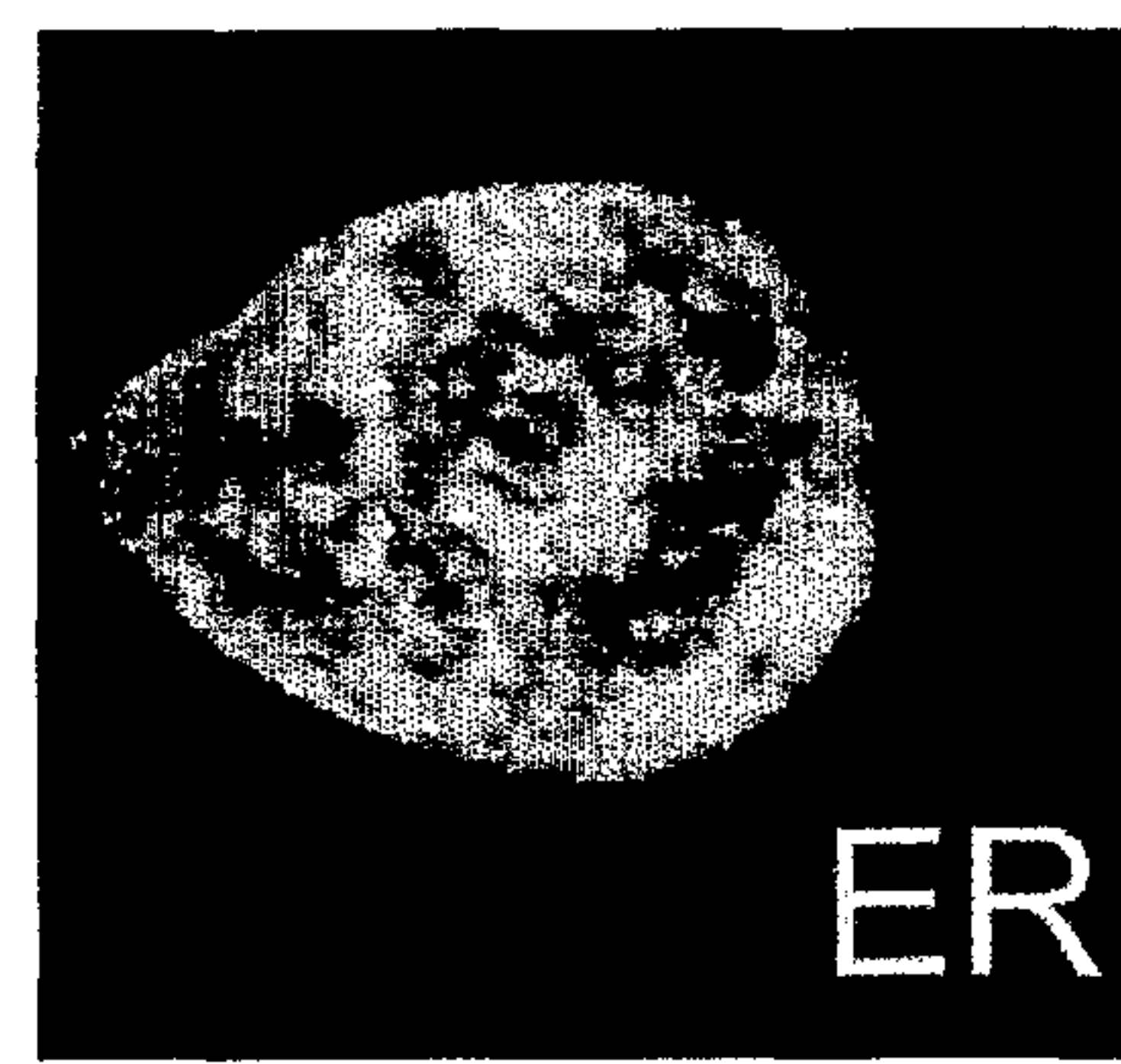


FIG. 16D



FIG. 16E

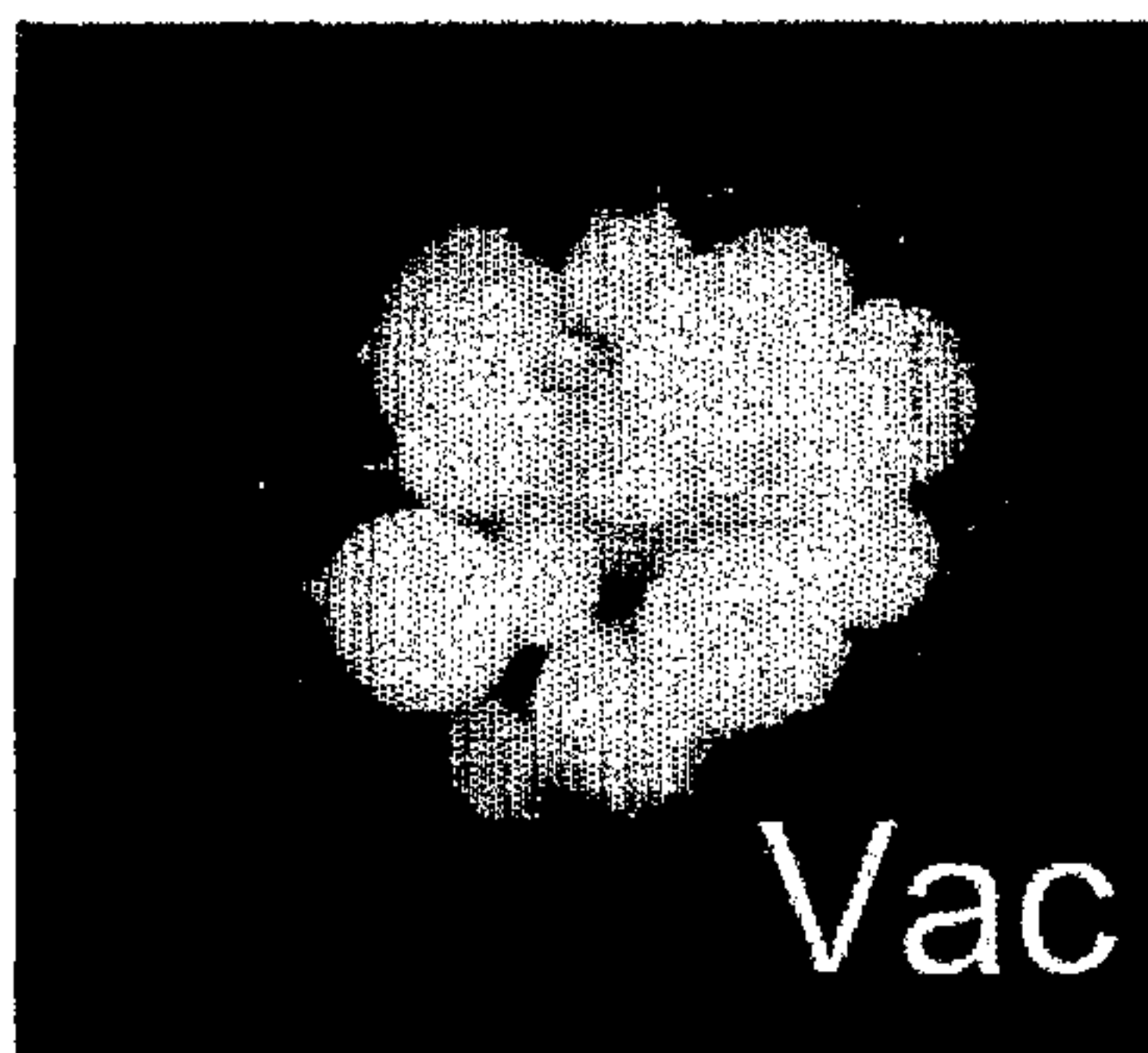


FIG. 16F

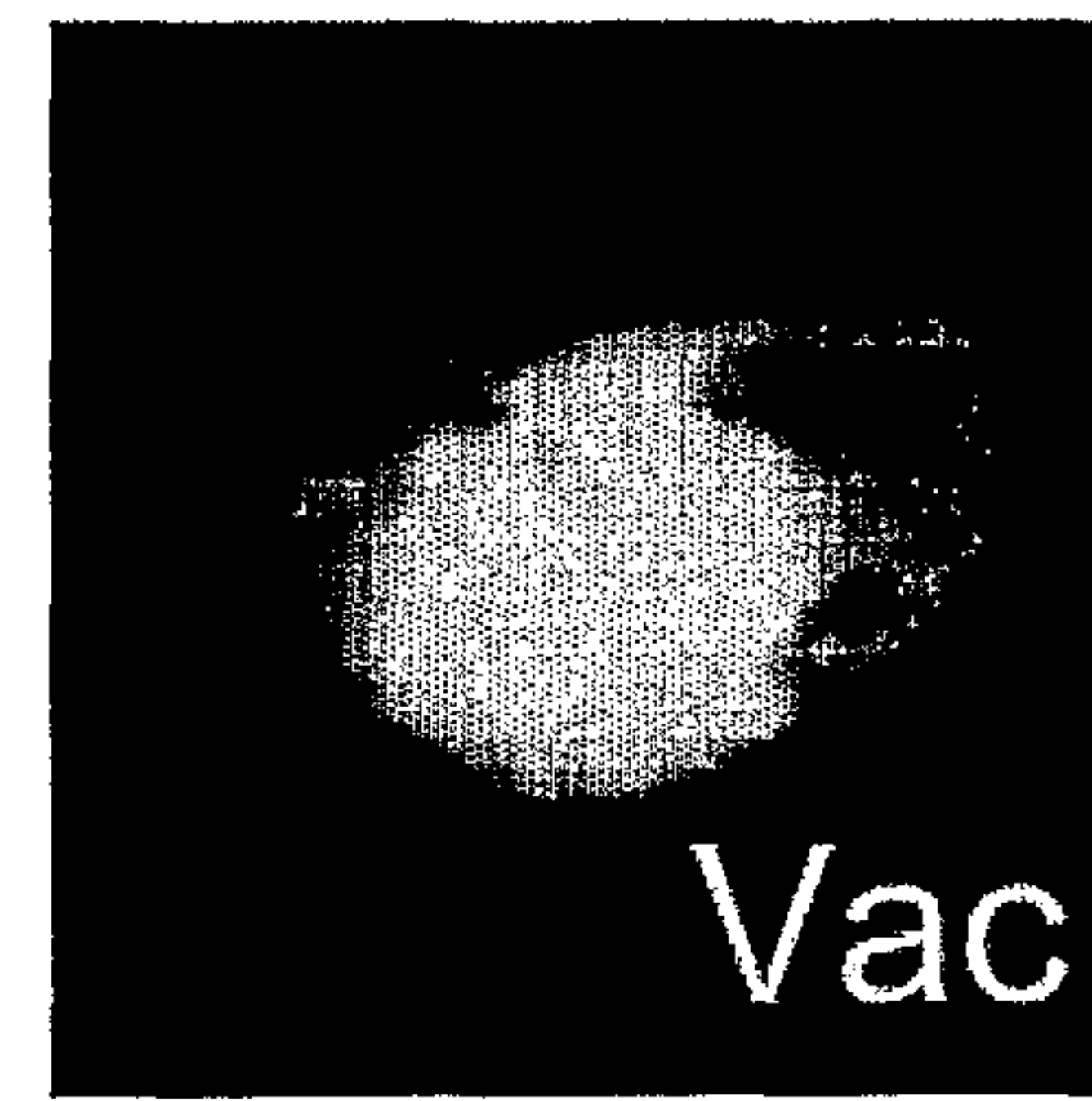


FIG. 16G

**FAE Activity in Transgenic *Festuca arundinacea* Leaves
of Different Ages Under ER and APO Targeting Sequence**

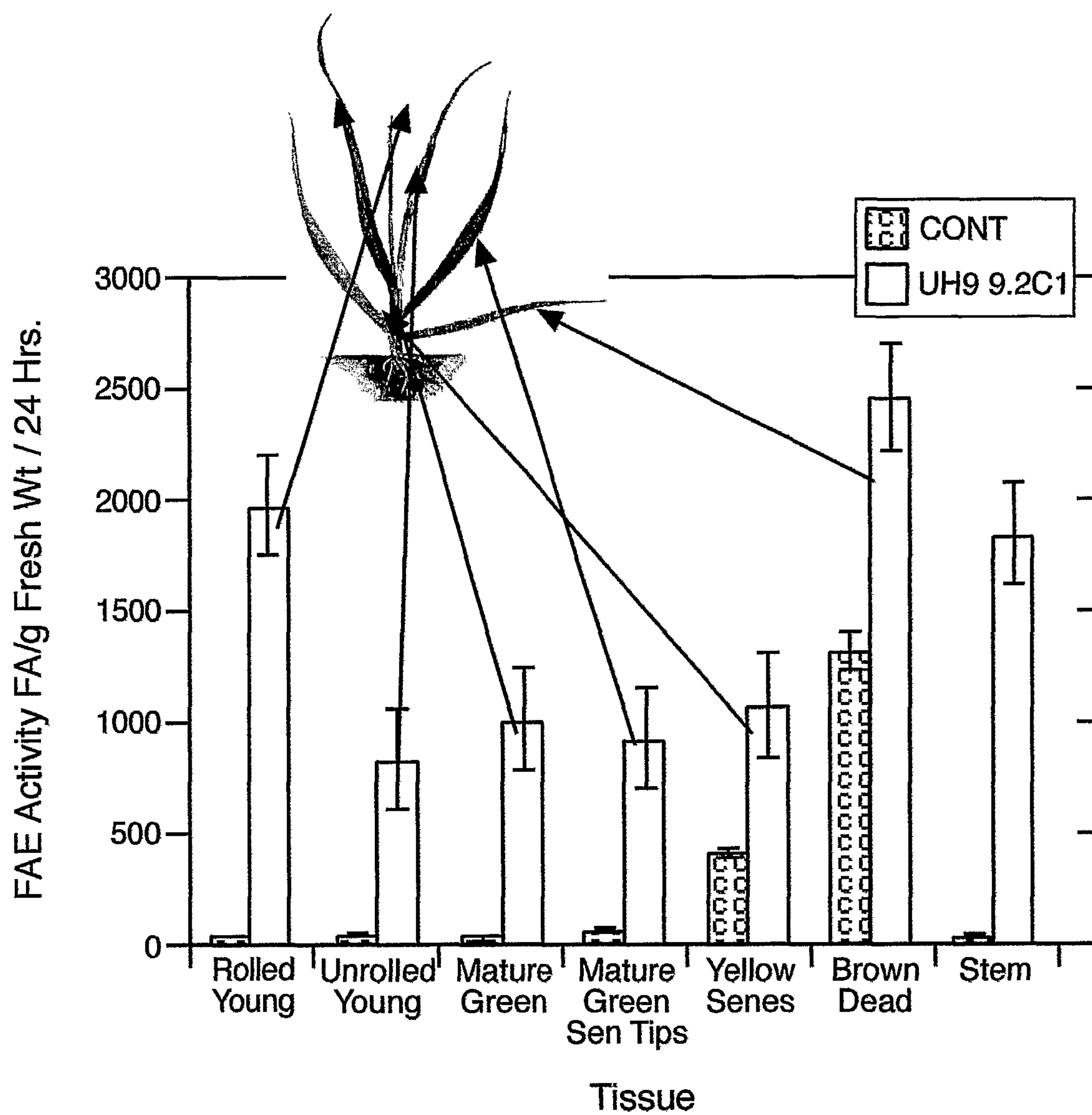


FIG. 17A

**FAE Activity in Transgenic *Festuca arundinacea* Leaves
of Different Ages Under ER and APO Targeting Sequence**

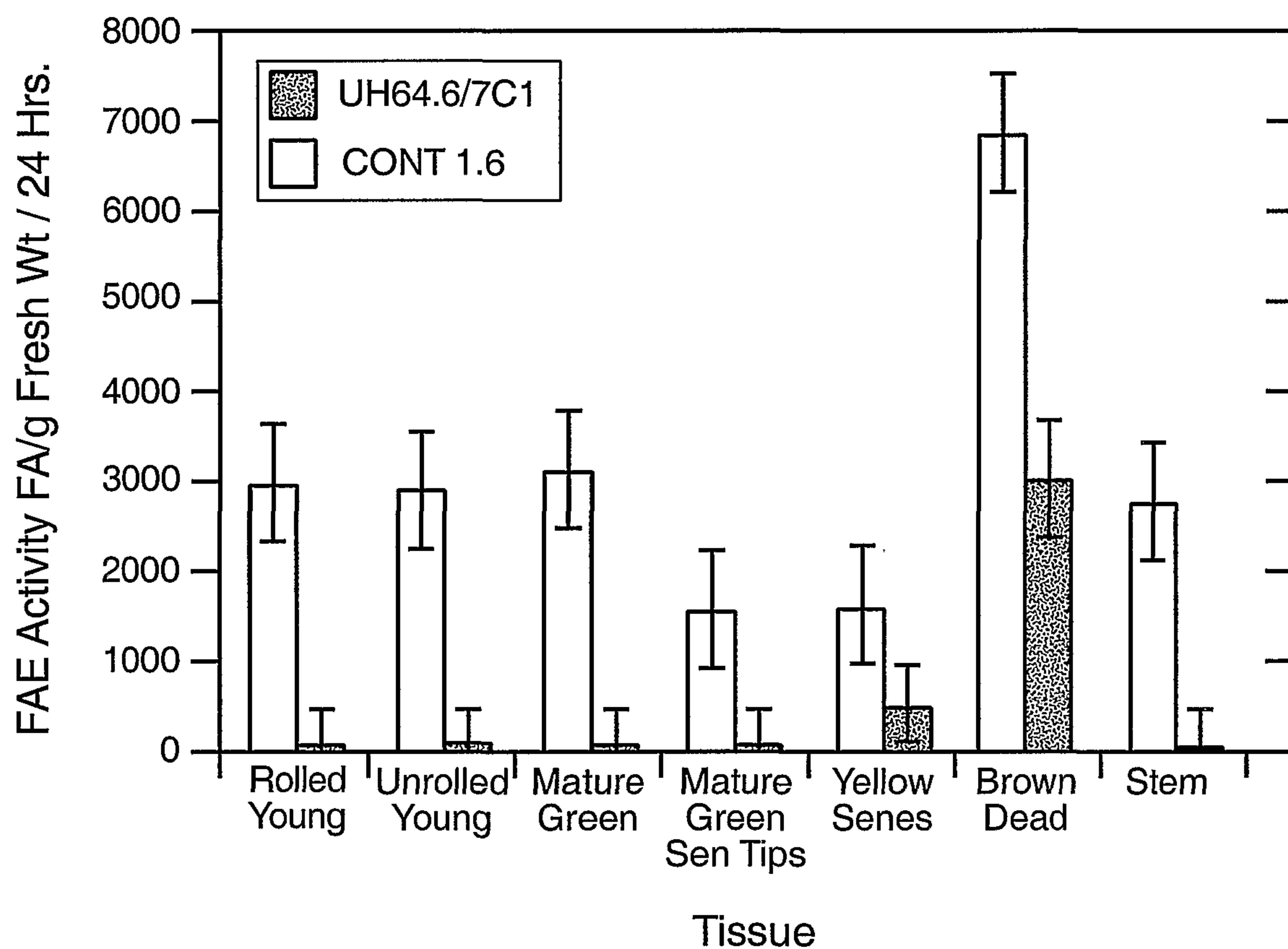
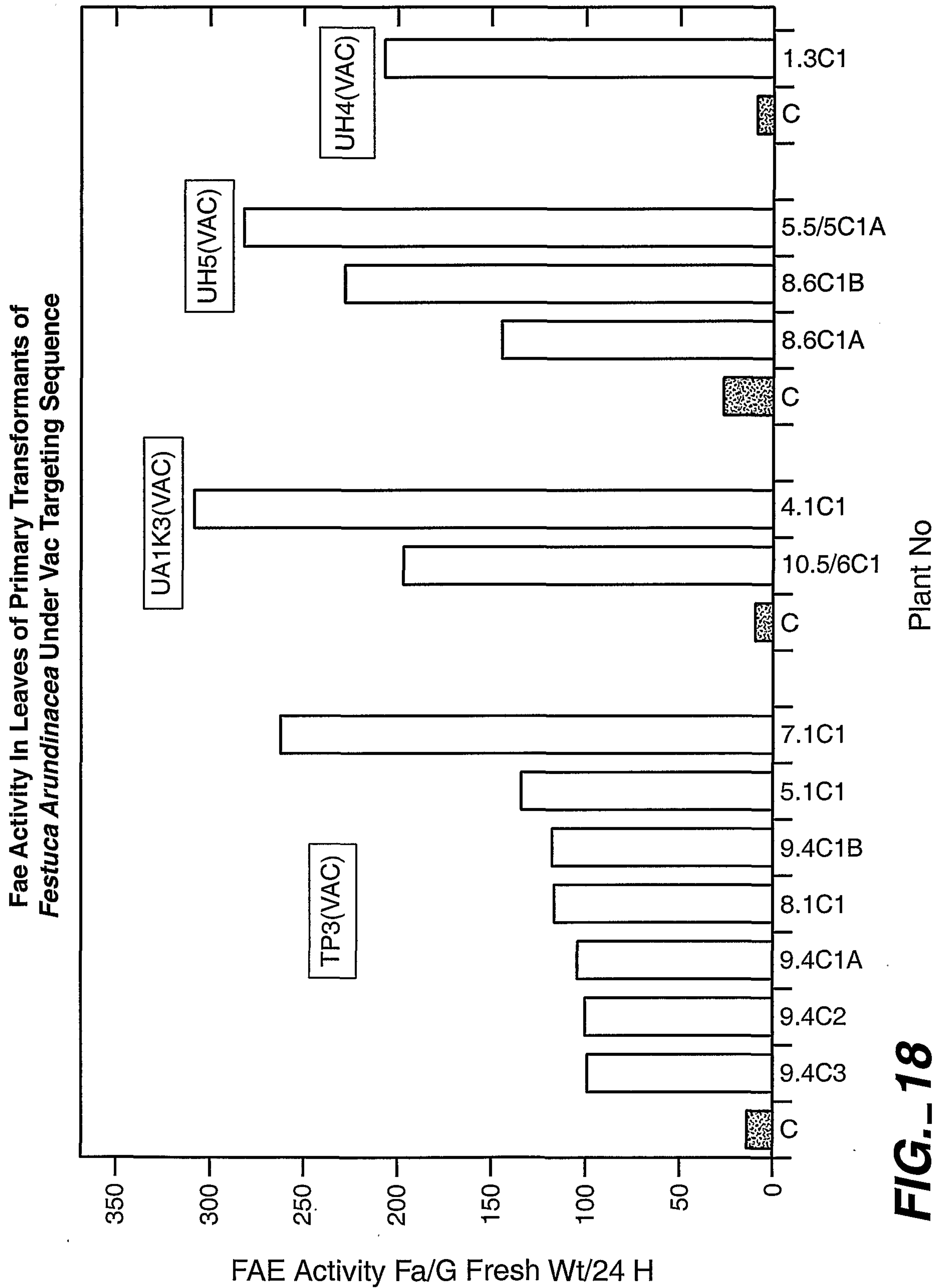
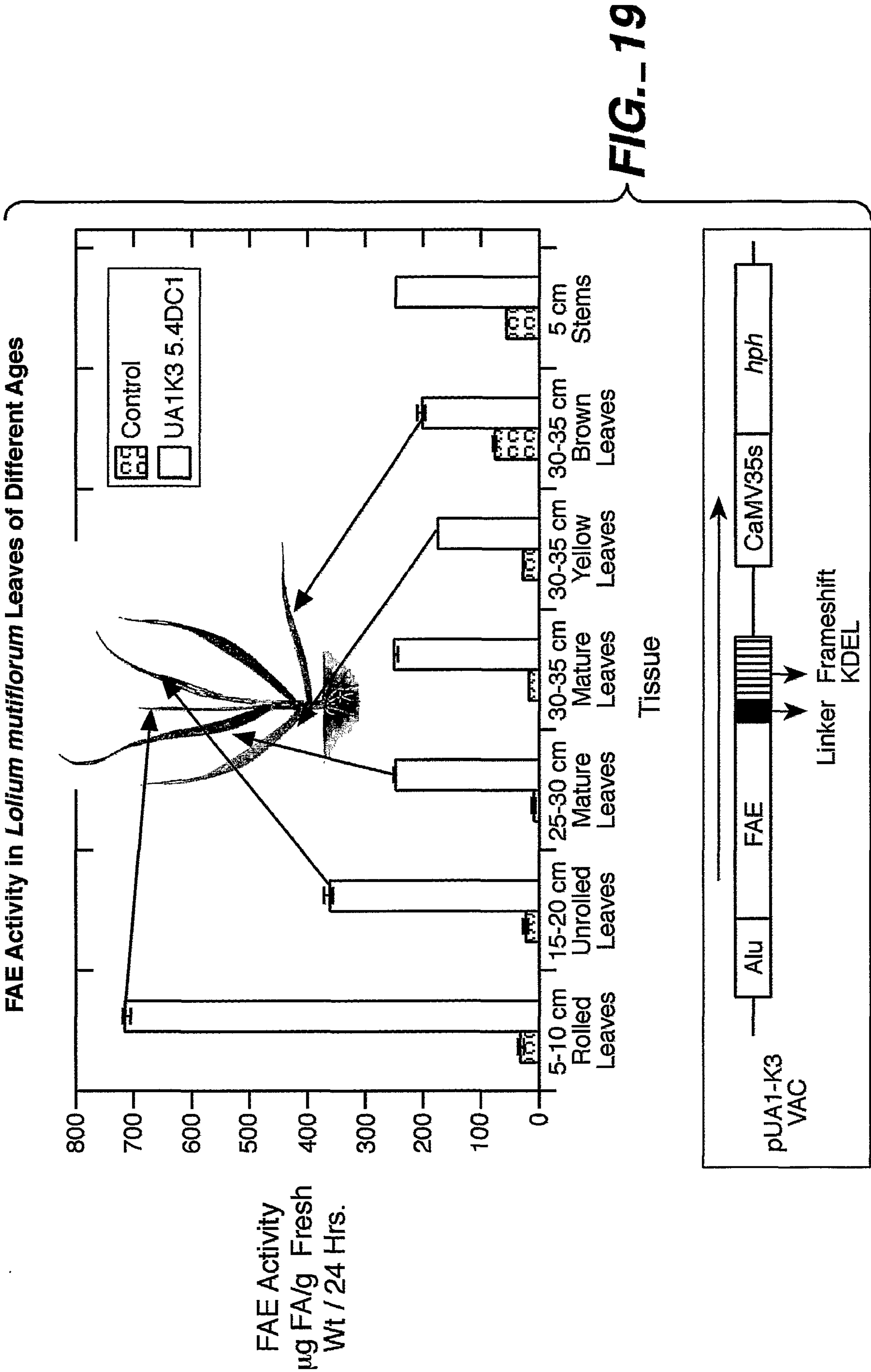


FIG. 17B

**FIG.-18**



FAE Activity in Leaves of Primary Transformants of *Lolium multiflorum* Under VAC APO and ER Targeting Sequence

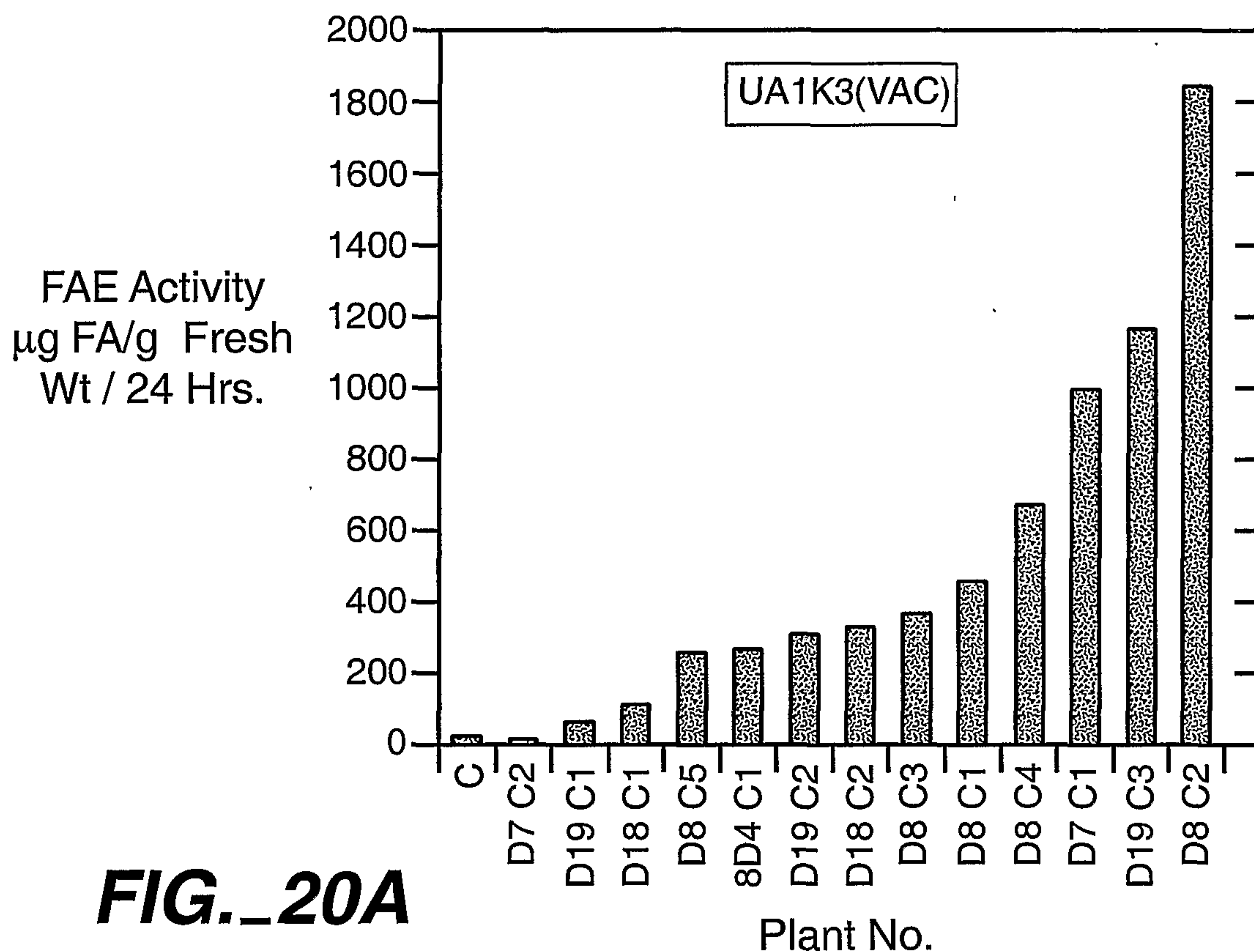


FIG. 20A

FAE Activity in Leaves of Primary Transformants of *Lolium multiflorum* Under VAC APO and ER Targeting Sequence

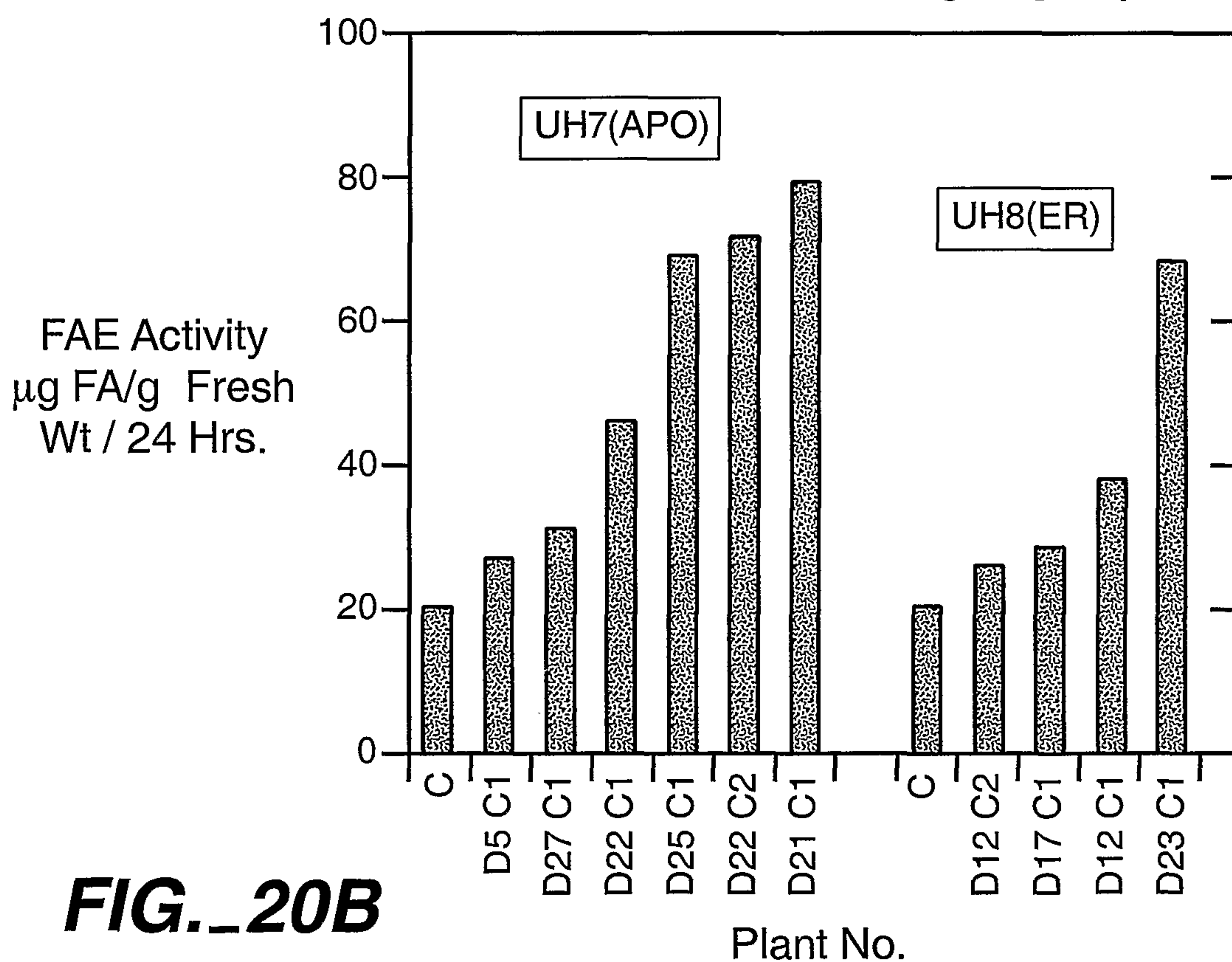


FIG. 20B

Levels of Esterified Monomeric and Dimeric Hydroxycinnamicacids in *Festuca Arundinacea* Plants Expressing FAE Under VAC Targeting Sequence

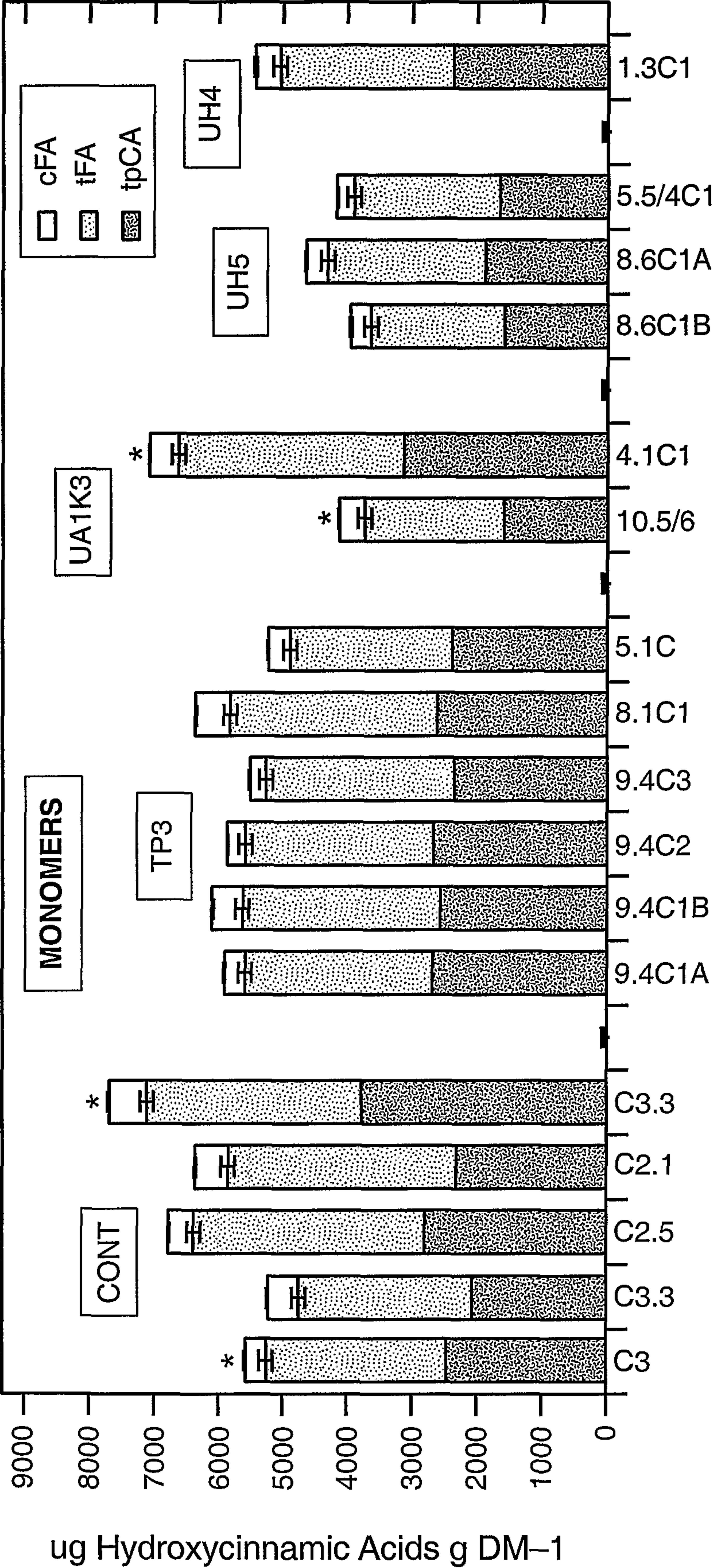


FIG. 21A

Levels of Esterified Monomeric and Dimeric Hydroxycinnamicacids in
Festuca Arundinacea Plants Expressing FAE Under VAC Targeting Sequence

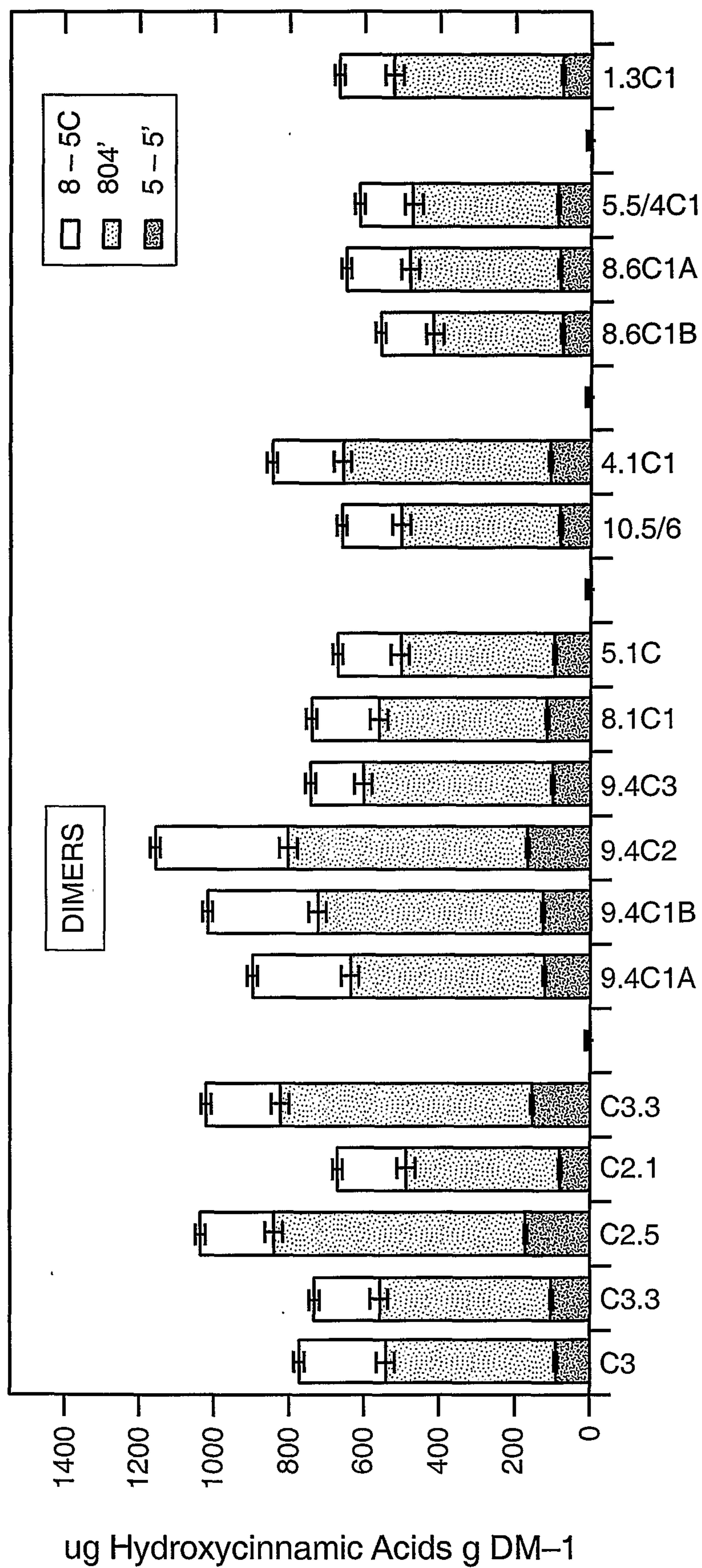


FIG. 21B

Levels of Esterified Monomeric and Dimeric Hydroxycinnamic Acids in
Leaves of *F. a.* Expressing FAE Under ER and APO Targeting Sequence

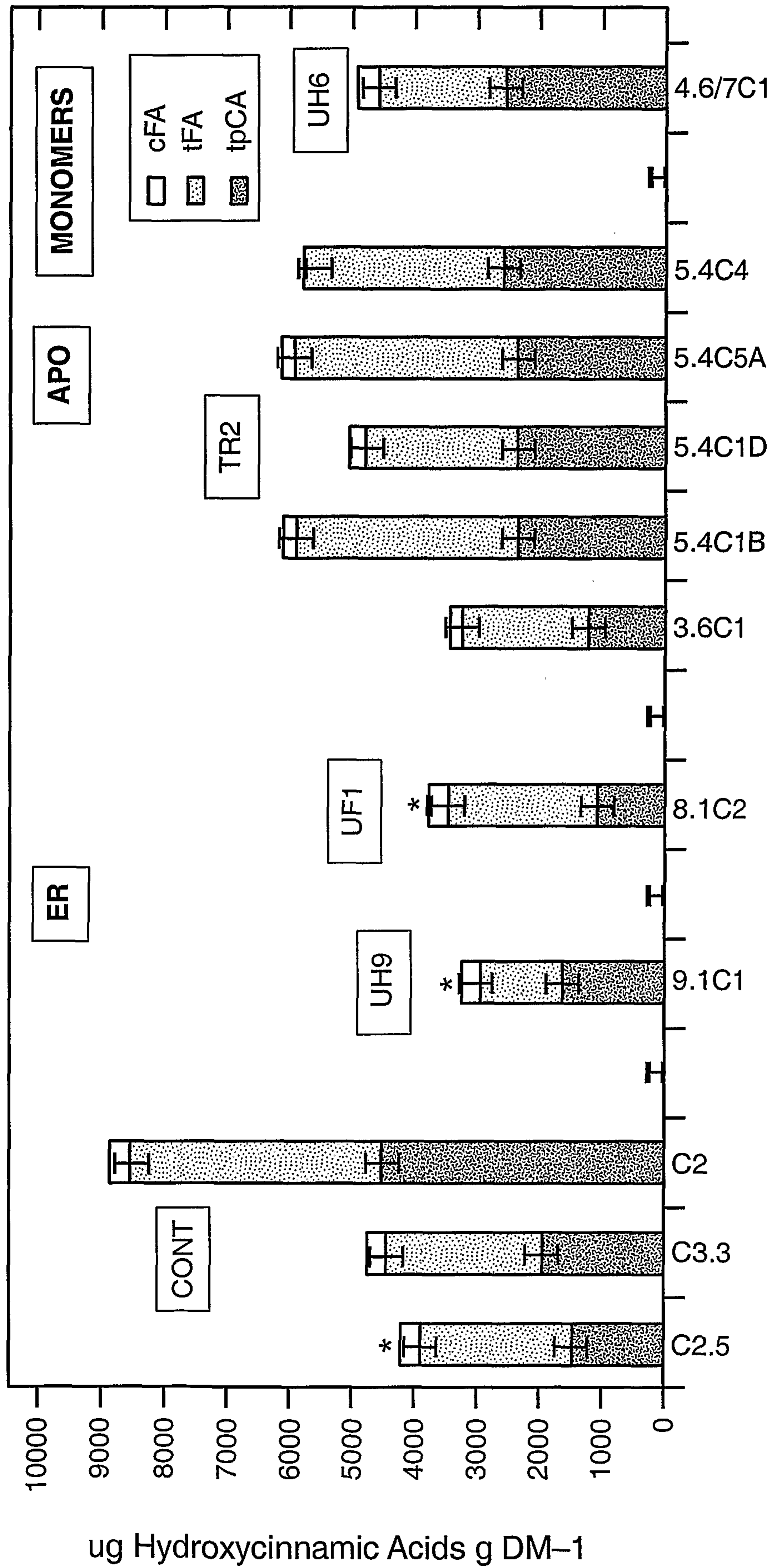


FIG. 22A

Levels of Esterified Monomeric and Dimeric Hydroxycinnamic Acids in
Leaves of *F. a.* Expressing FAE Under ER and APO Targeting Sequence

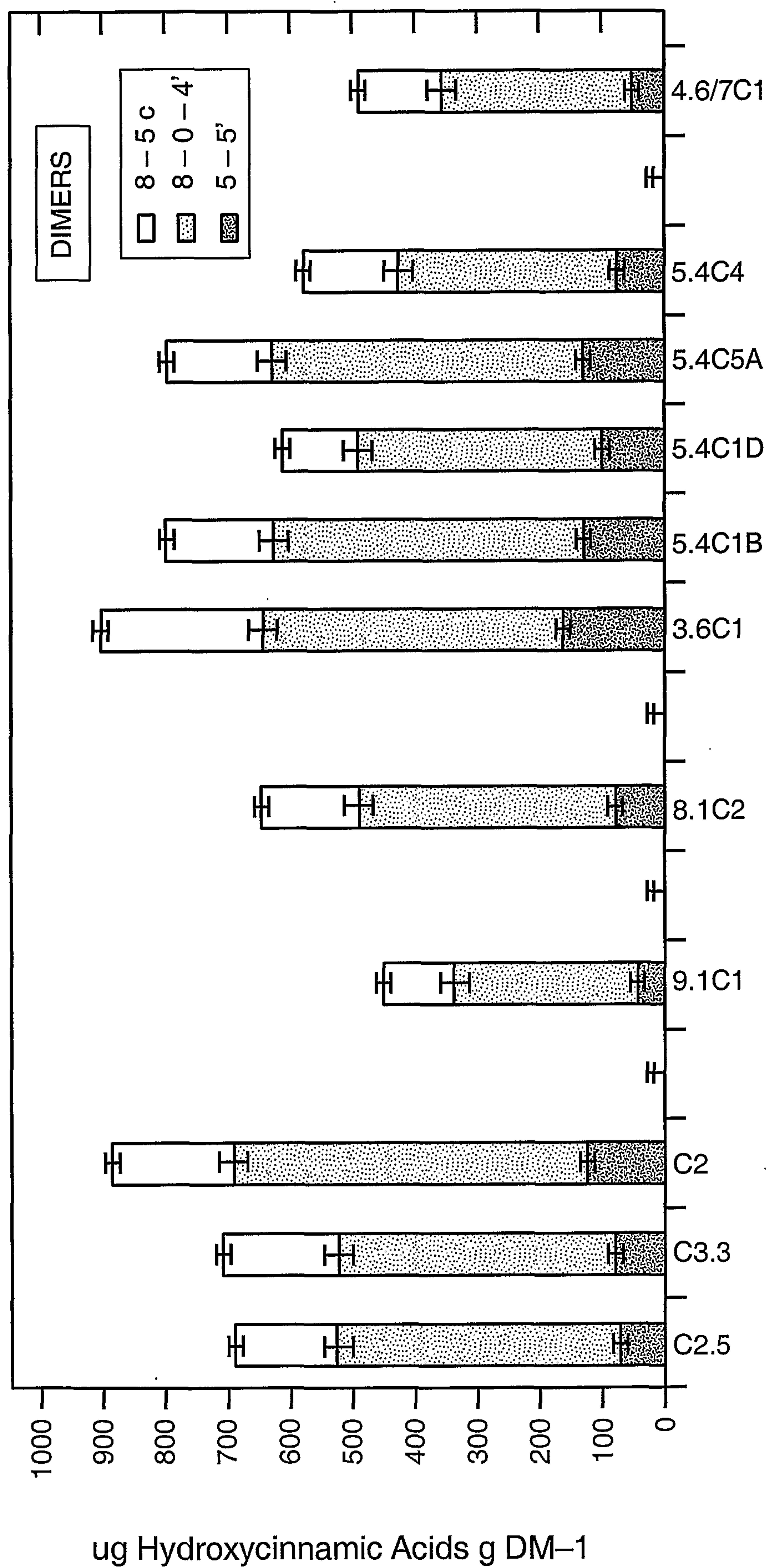
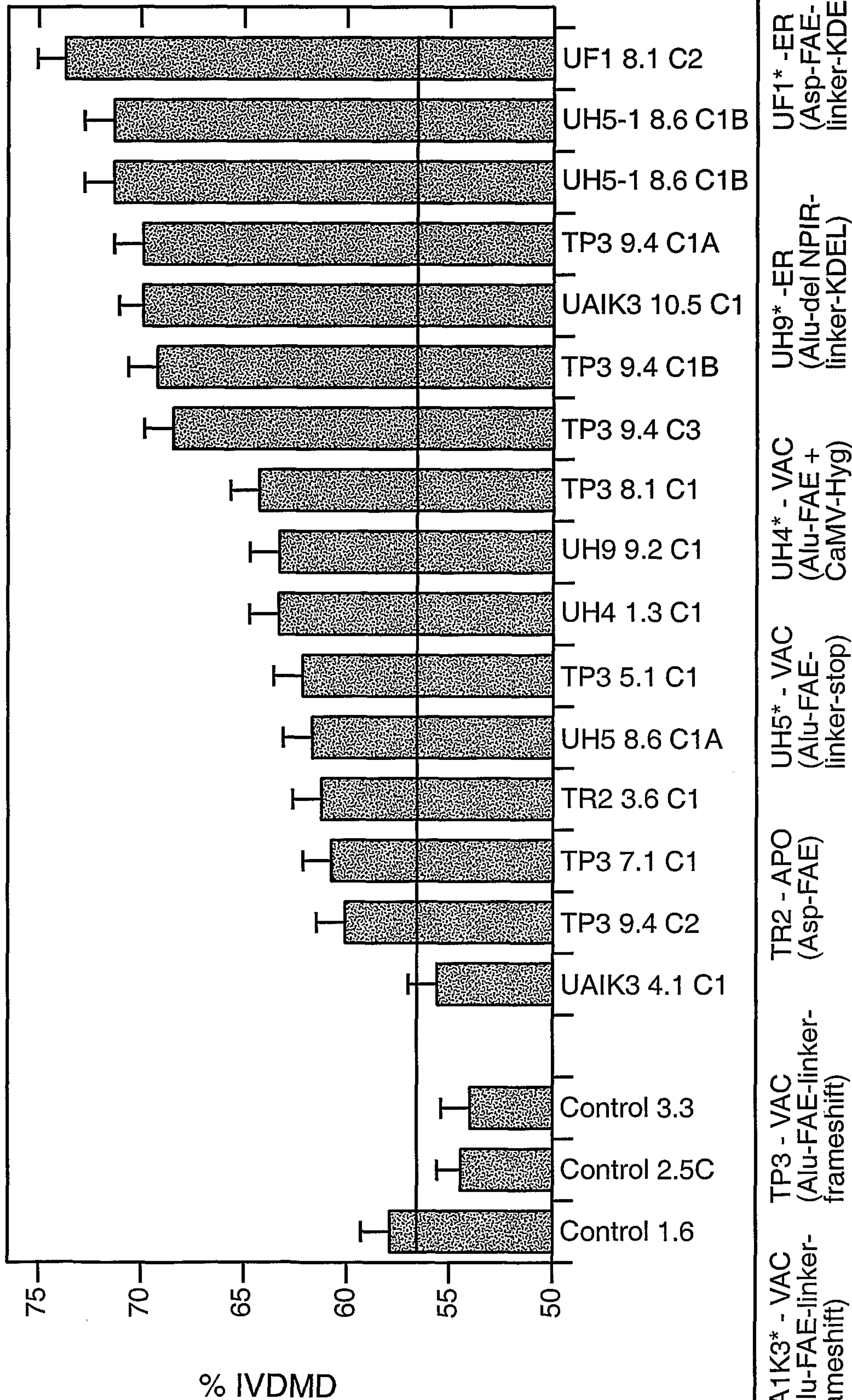


FIG. 22B

**In Vitro Dry Matter Digestibility of Leaf Tissue of Mature
Festuca arundinacea Plants Expressing FAE Under an Actin Promoter**



*Co-integration Vector

FIG. 23

In Vitro Dry Matter Digestibility of Leaf Tissue of Mature
Lolium multiflorum Plants Expressing FAE Under an Actin Promoter

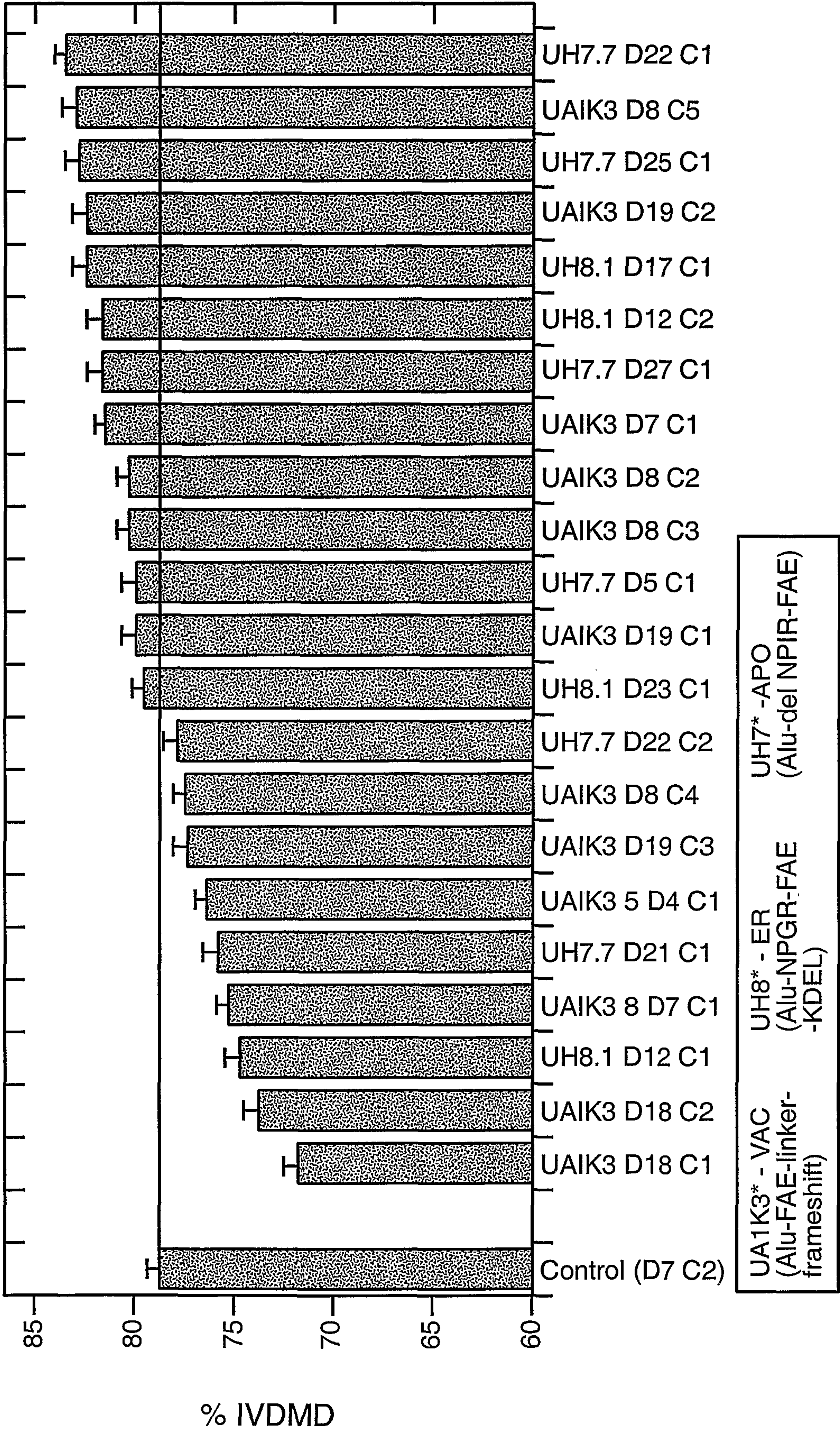


FIG. 24

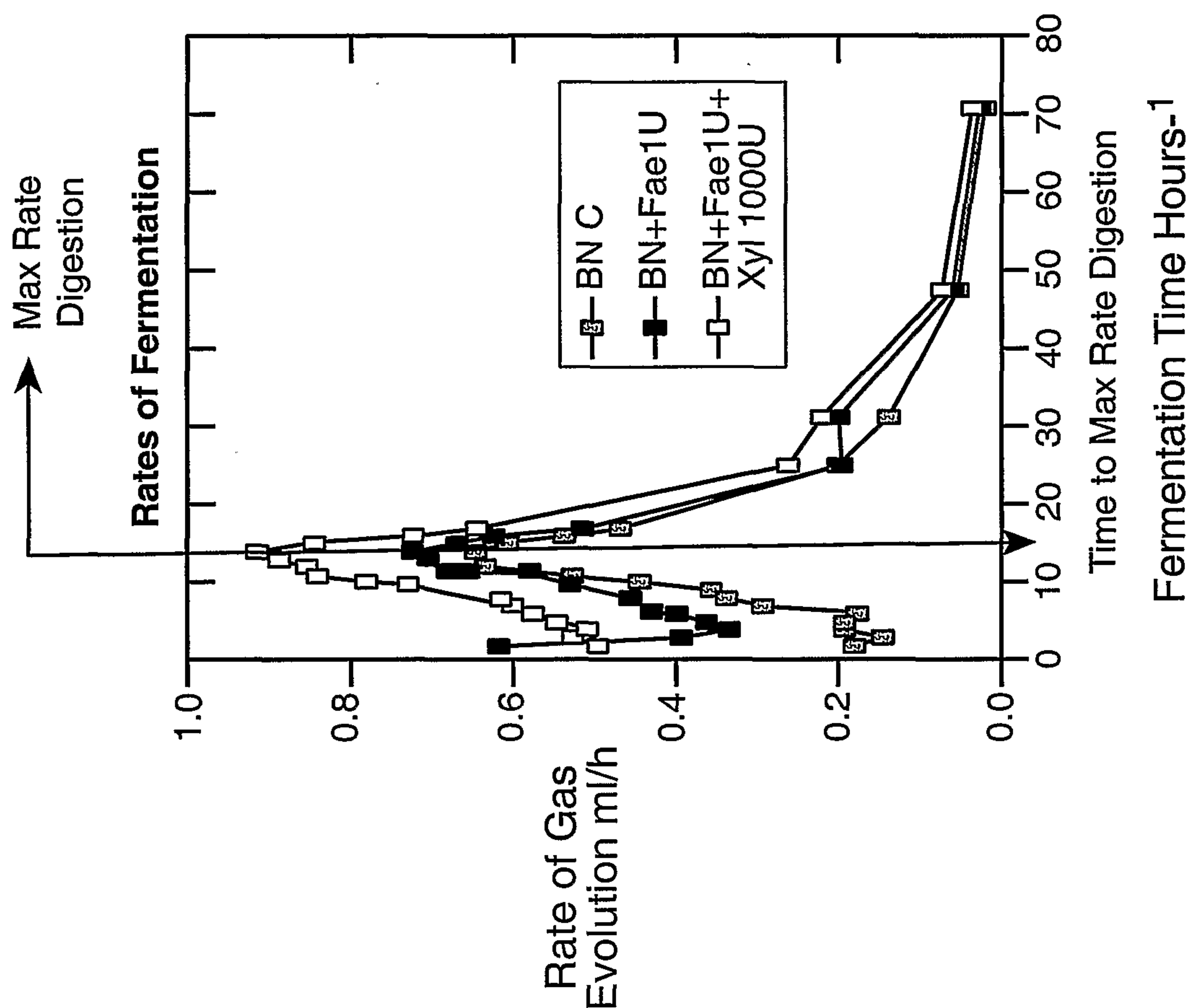


FIG. 25B

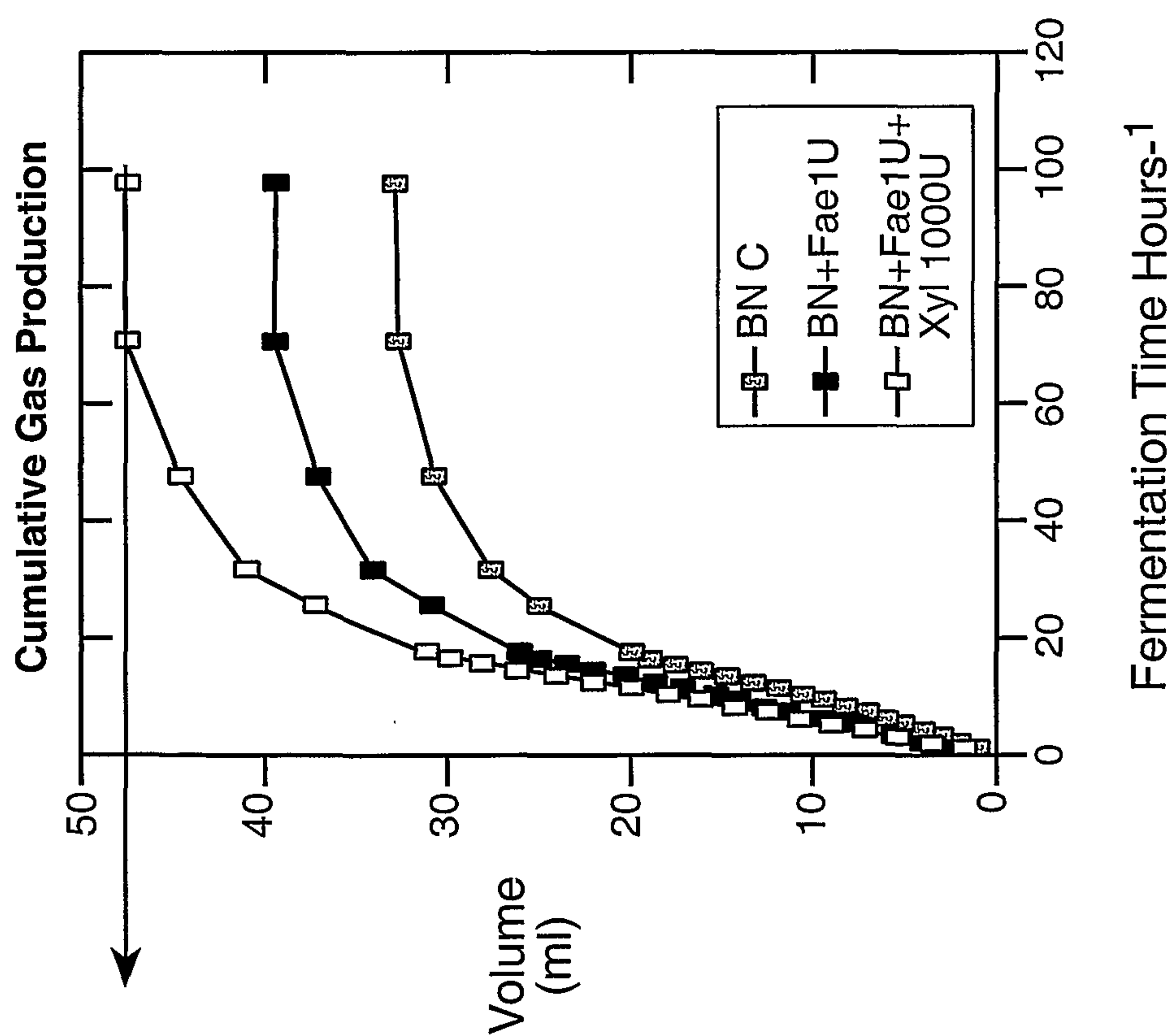


FIG. 25A

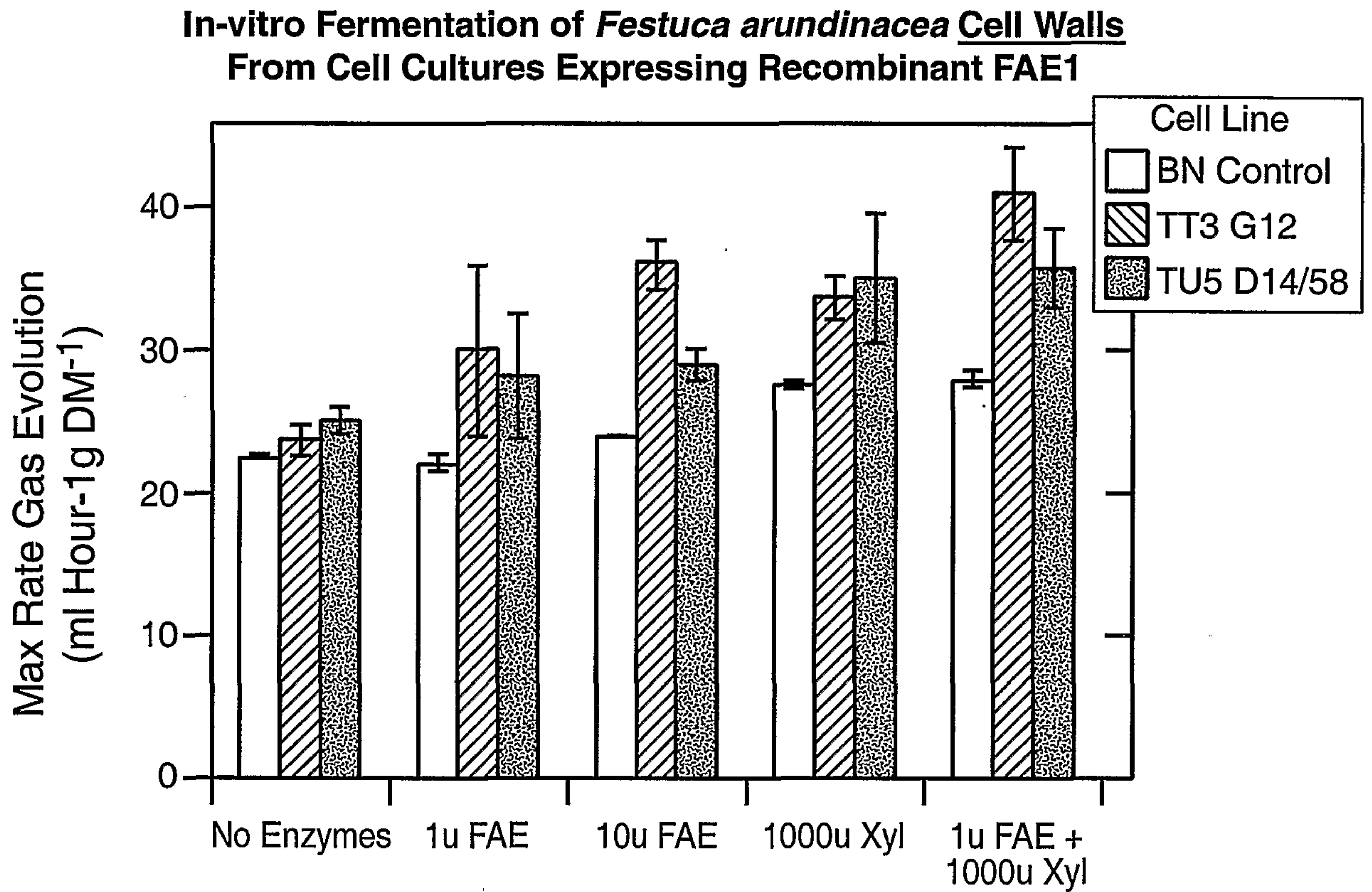


FIG. 26A Maximum Rate of Digestion

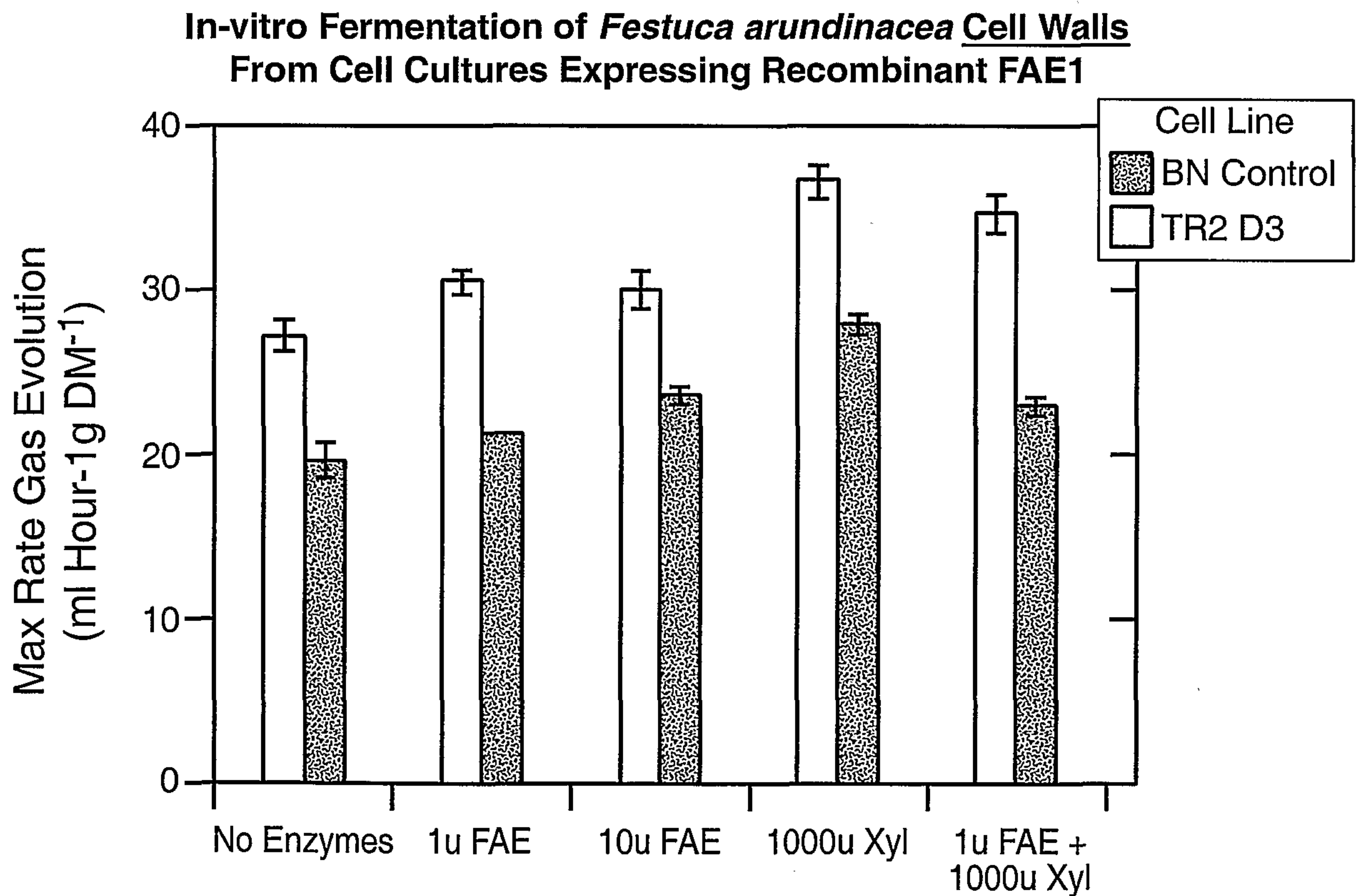
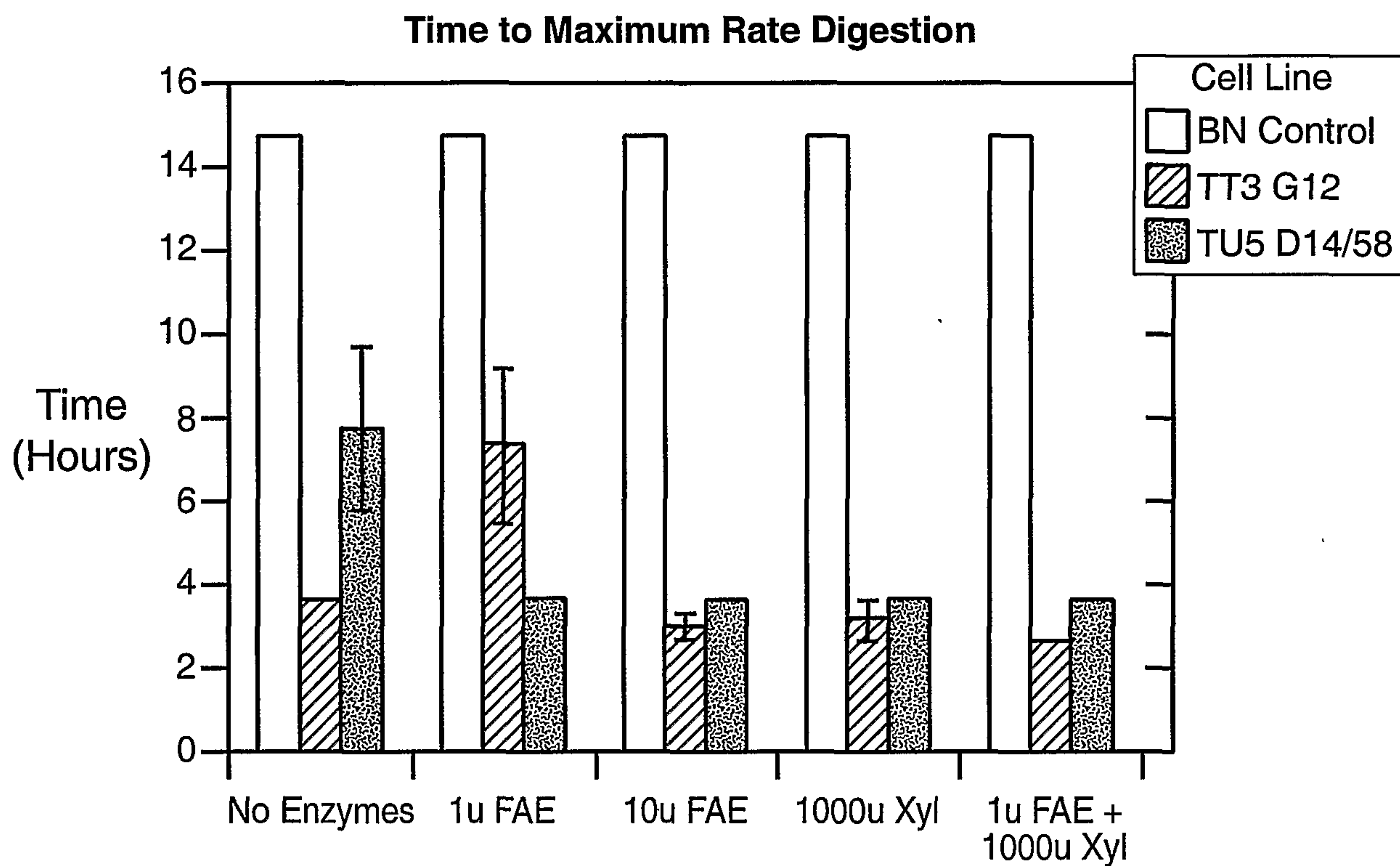
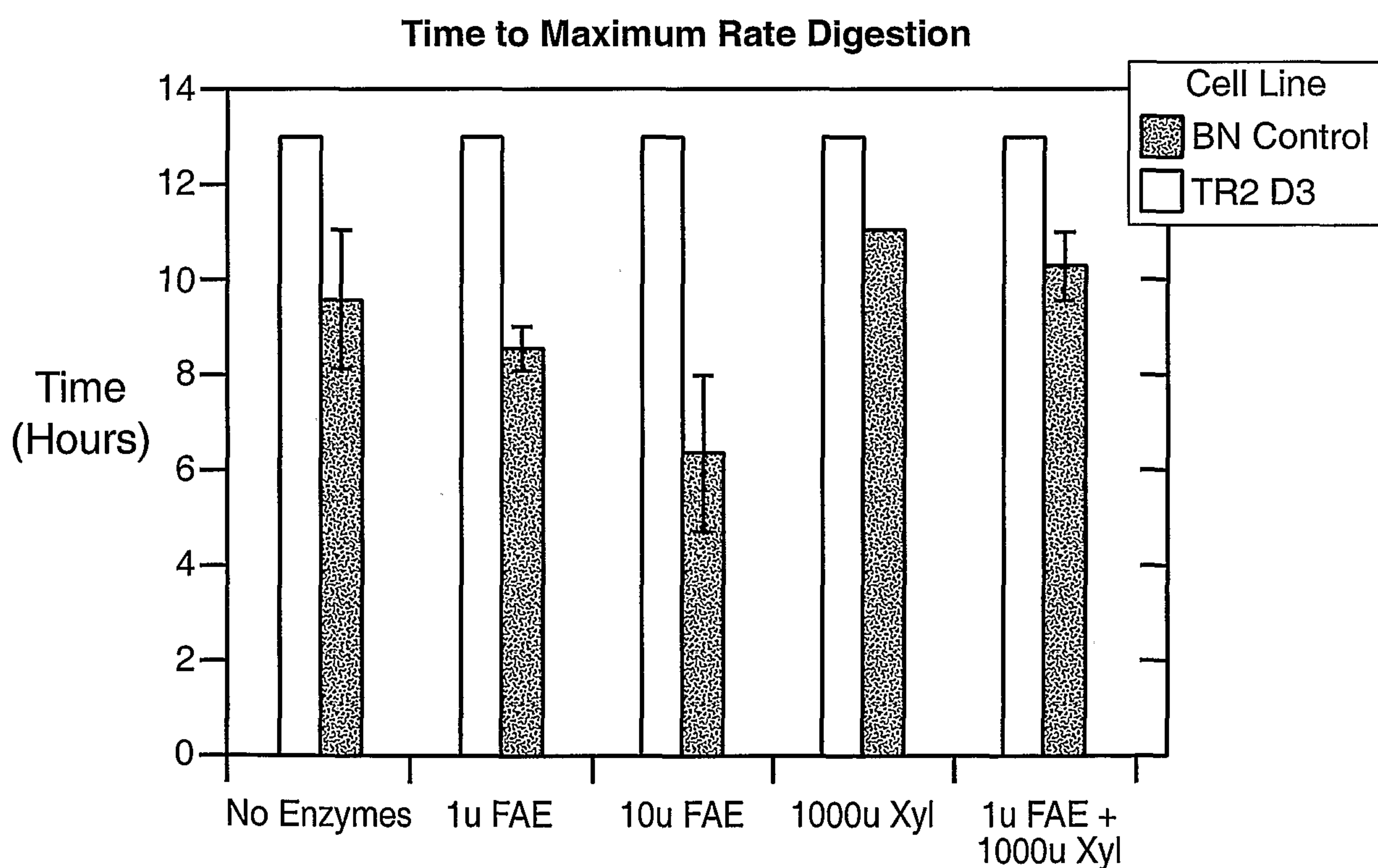
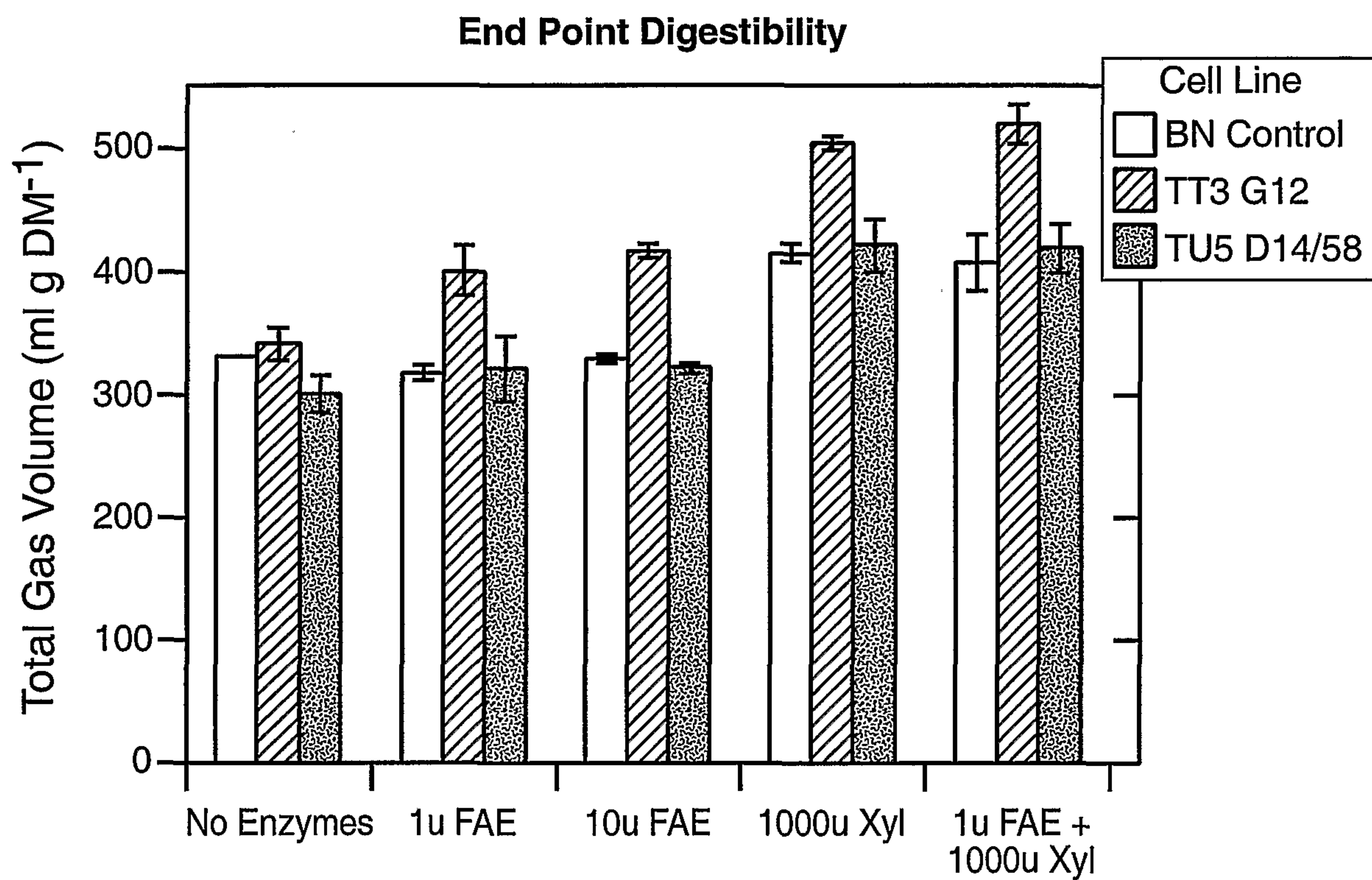
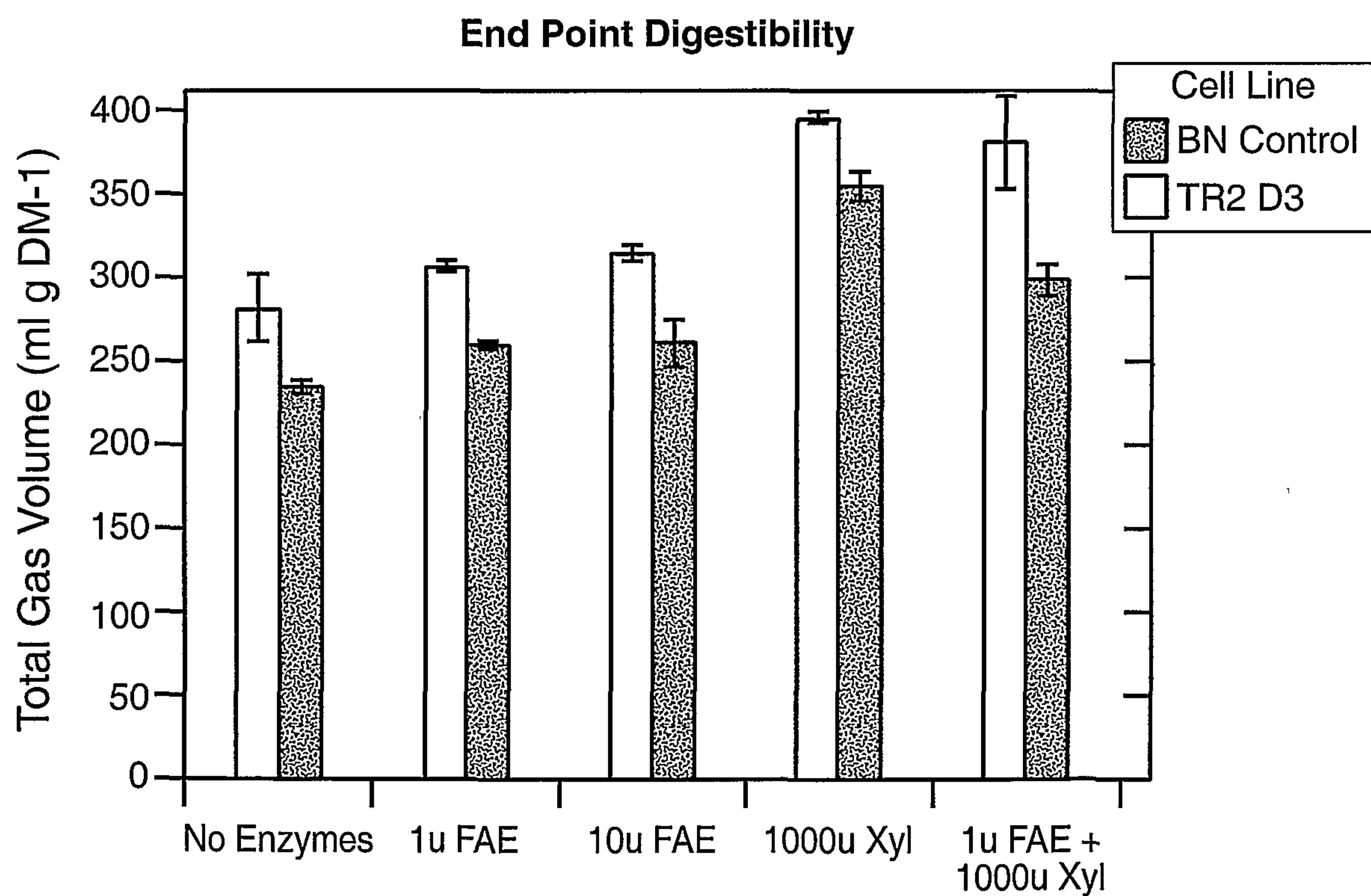
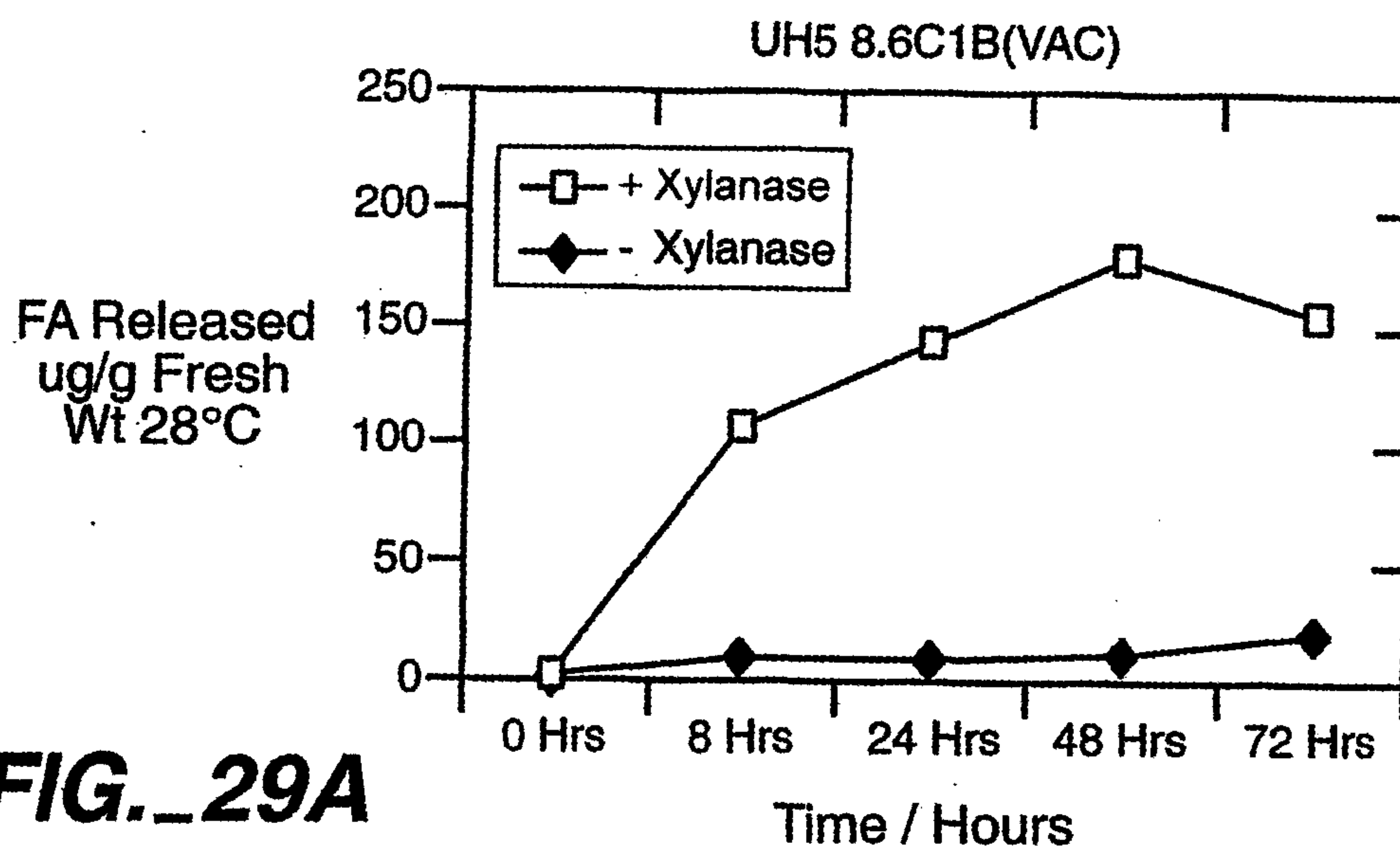


FIG. 26B Maximum Rate of Digestion

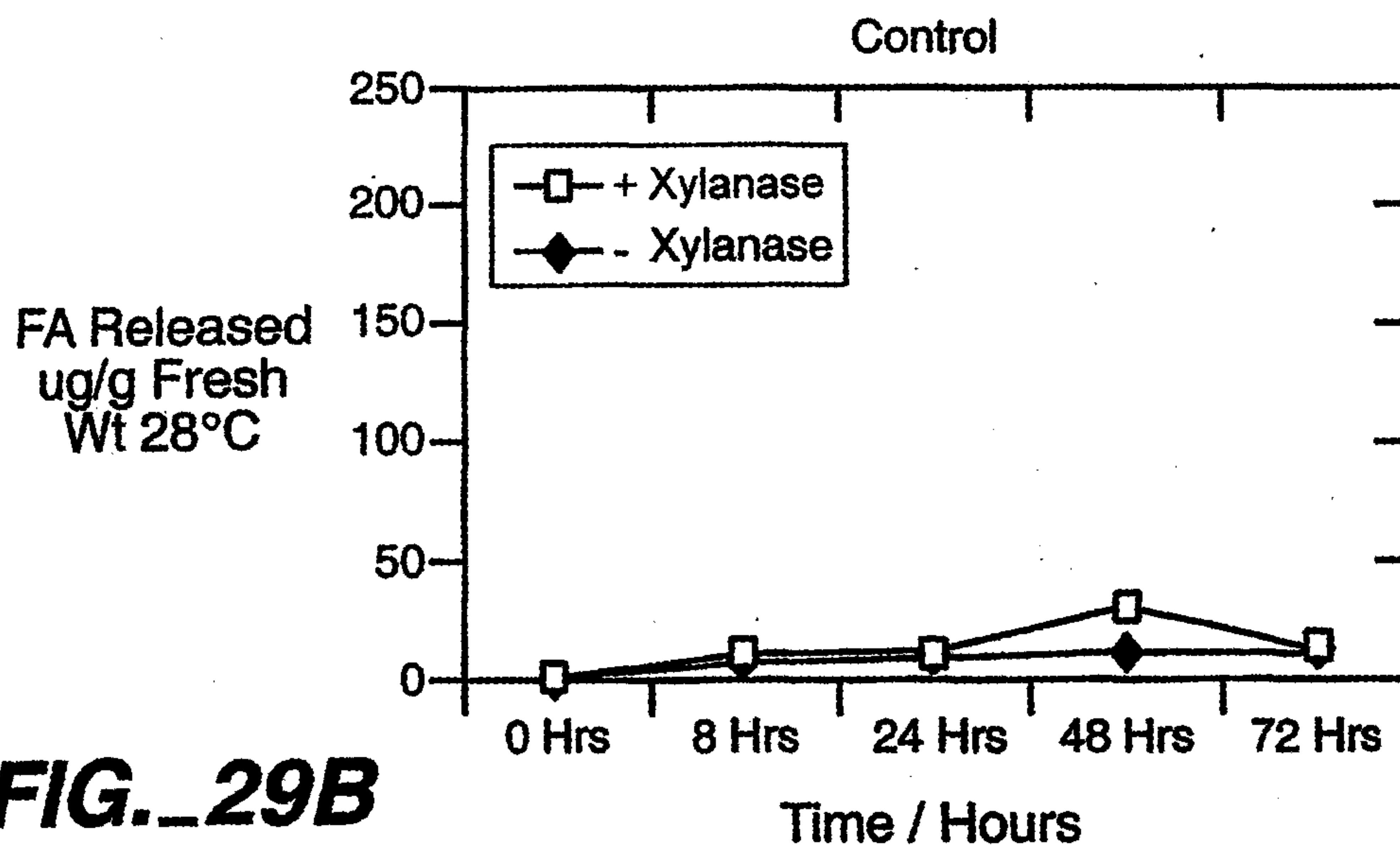
**FIG. 27A****FIG. 27B**

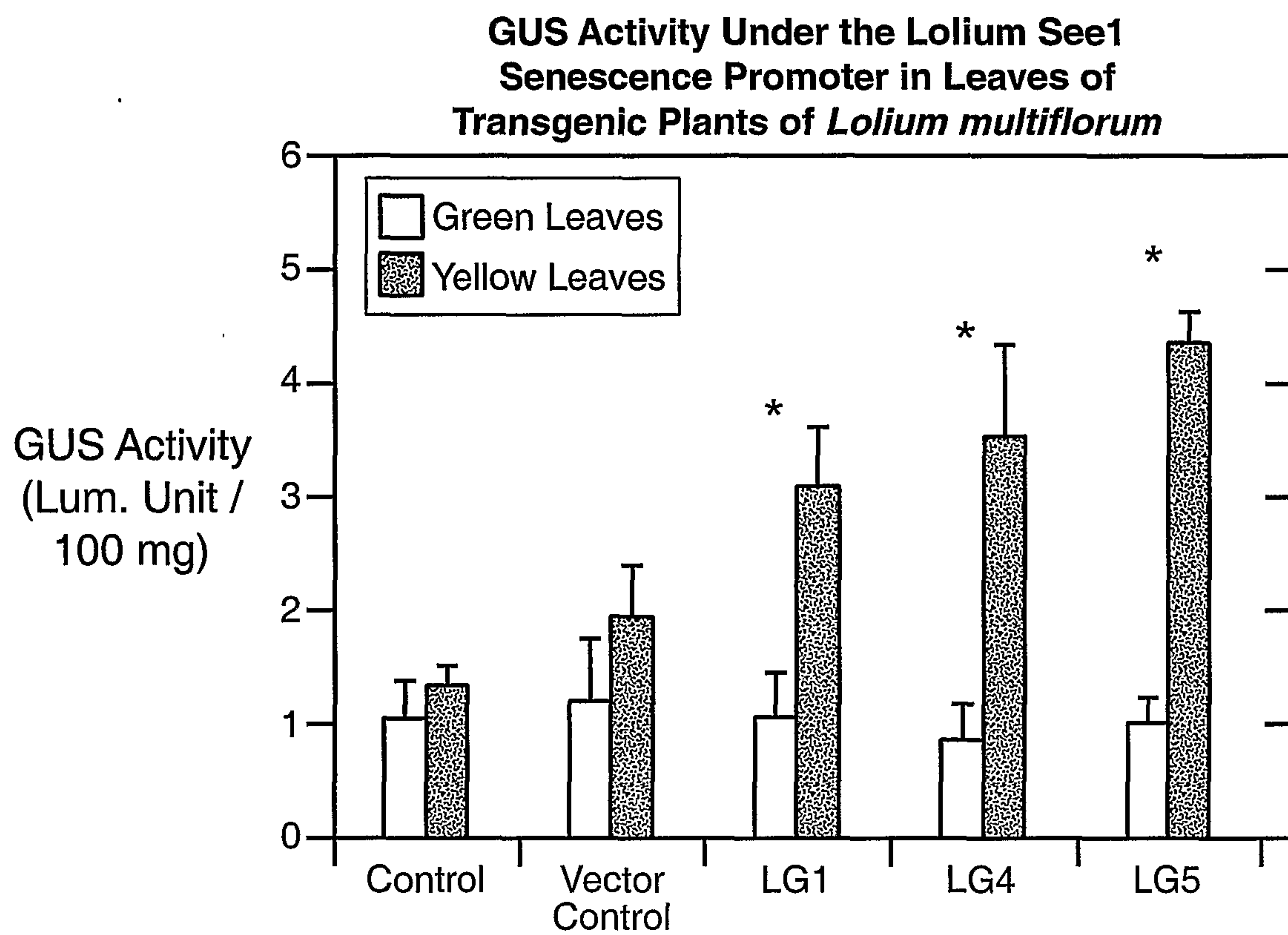
**FIG. 28A****FIG. 28B**

**Kinetics of FAE Activity by Ferulic Acid Release
from Cell Wall under Self Digestion in *Festuca arundinacea*
and Stimulation by Xylanase**

**FIG._29A**

**Kinetics of FAE Activity by Ferulic Acid Release
from Cell Wall under Self Digestion in *Festuca arundinacea*
and Stimulation by Xylanase.**

**FIG._29B**

**FIG._30**

Release of Monomeric and Dimeric HCAs
on Self Digestion of Leaves of Vacuolar
Targeted FAE Expressing Plants

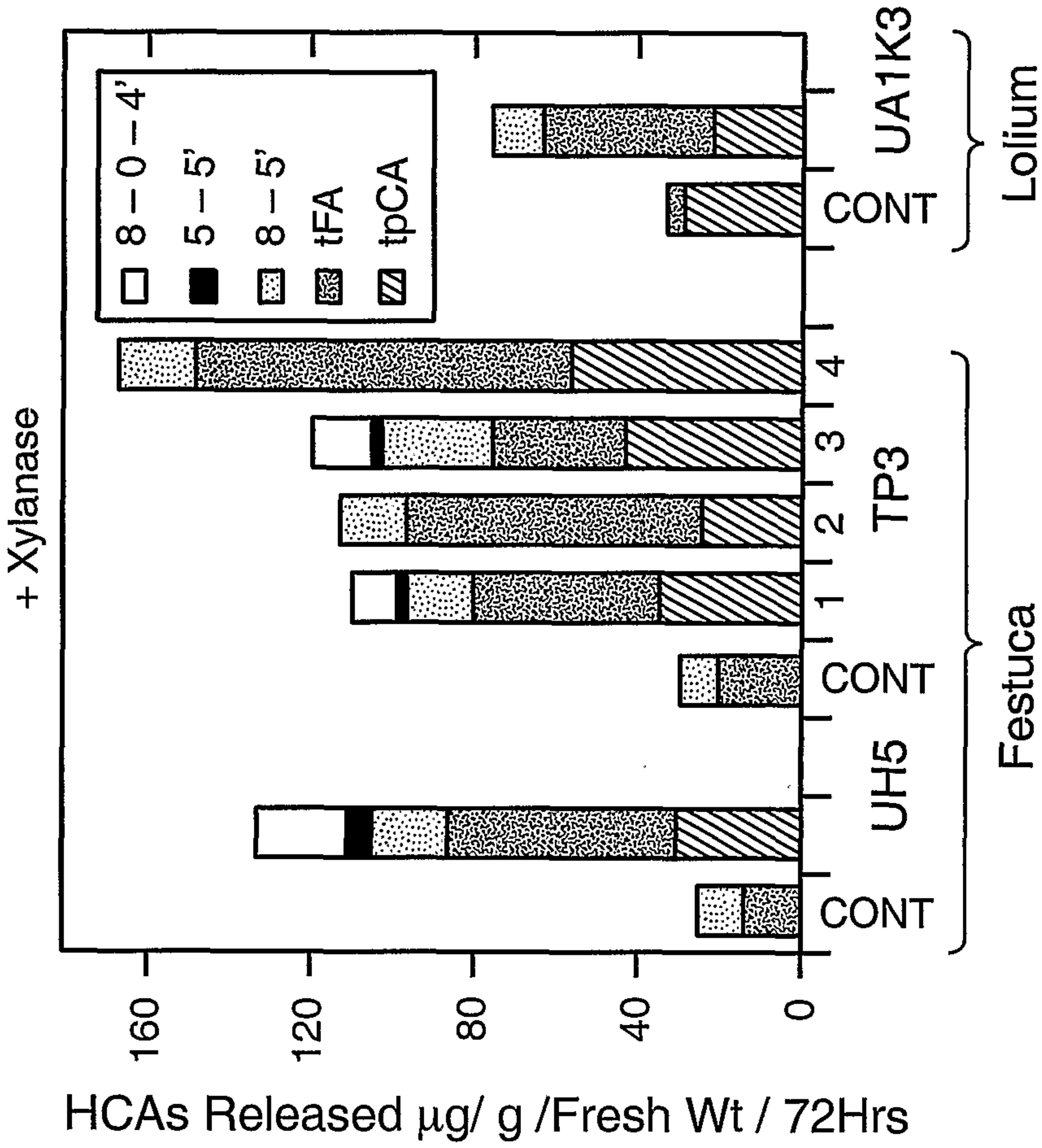


FIG._31B

Release of Monomeric and Dimeric HCAs
on Self Digestion of Leaves of Vacuolar
Targeted FAE Expressing Plants

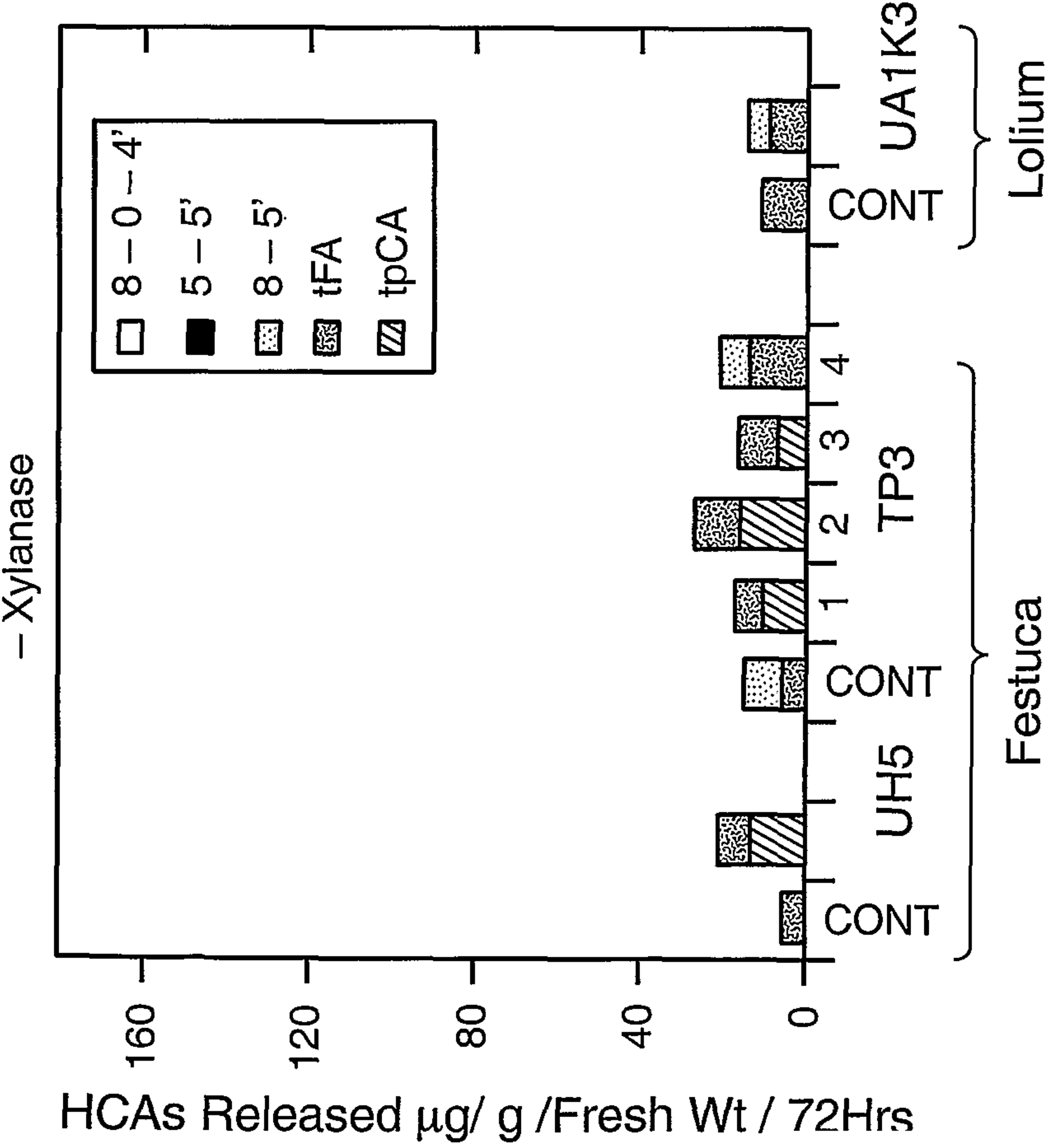


FIG._31A

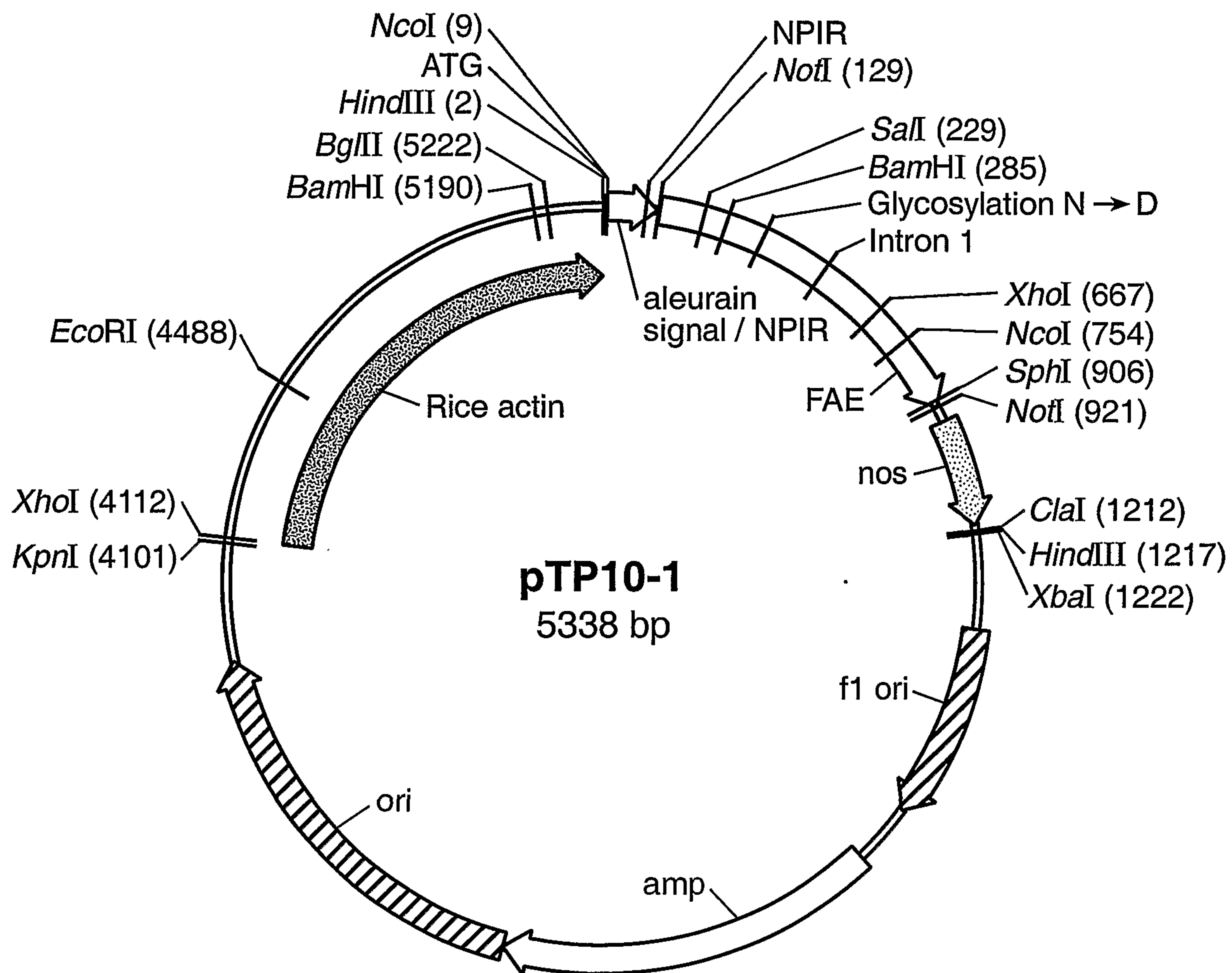
**FIG. 32A**

FIG. 32B

XhoI

~~~~~

A F A S Y M N D A F Q A S S P D T T Q Y F R V T

631 G C C T T C G C G T C G T A C A T G A A C G A T G C C T T C C A A G C C T C G A G C C A G A T A C G A C G C A G T A T T T C C G G G T C A

NcoI

~~~~~

. H A N D G I P N L P P V E Q G Y A H G G V E Y .

701 C T C A T G C C A A C G A C G G C A T C C C A A C C T G C C C C G G T G G A G C A G G G G T A C G C C A T G G C G G T G T A G A G T A

. W S V D P Y S A Q N T F V C T G D E V Q C C E

771 C T G G A G C G T T G A T C C T T A C A G C G C C A G A A C A C A T T T G T C T G C A C T G G G A T G A A G T G C A G T G C T G T G A G

SphI

~~~~

A Q G G Q G V N N A H T T Y F G M T S G A C T W

841 G C C A G G G C G G A C A G G G T G T G A A T A T G C G C A C A G A C T T A T T T G G G A T G A C G A G C G G C G C A T G C A C C T

NotI

~~~~~

. P V A A A E T T E G *

911 G G C C G G T C G C G G C C G G A A A C C A C T G A A G G A T G A G C T G T A A A G A A G C A G A T C G T T C A A C A T T T G G C A A

981 T A A A G T T T C T T A A G A T T G A A T C C T G T T G C C G G T C T T G C G A T G A T T A C A T A A T T T C T G T G A A T T A C G

1051 T T A A G C A T G T A A T A A T T A A C A T G T A A T G C A T G A C G T T A T T T A T G A G A T G G G T T T T T A T G A T T A G A G T C C C

1121 G C A A T T A T A C A T T T A A T A C G C G A T A G A A A A C A A A T A T A G C G C G C A A A C T A G G A T A A A T T A T C G C G C G G

HindIII

~~~~~

Clal

XbaI

~~~~~

1191 G T G T C A T C T A T G T T A C T A G A T C G A T A A G C T T C T A G A G C G G C C G G T G G A G C T C C A A T T C G C C T A T A G T A

1261 G T C G T A T T A C G C G C G T C A C T G G C C G T C G T T T T A C A A C G T T T T A C A C C T G G A A A C C C T G G C G T T A C C C A A

1331 C T T A A T C G C C T T G C A G C A C A T C C C C C T T T C G C C A G C T G G C G T A A T A G C G A A G A G G C C C G C A C C G A T C G C C

1401 C T T C C C A A C A G T T G C G C A G C T G A A T G G C G A A T G G A C G C G C C T G T A G C G C C A T T A A G C G C G C G G G G

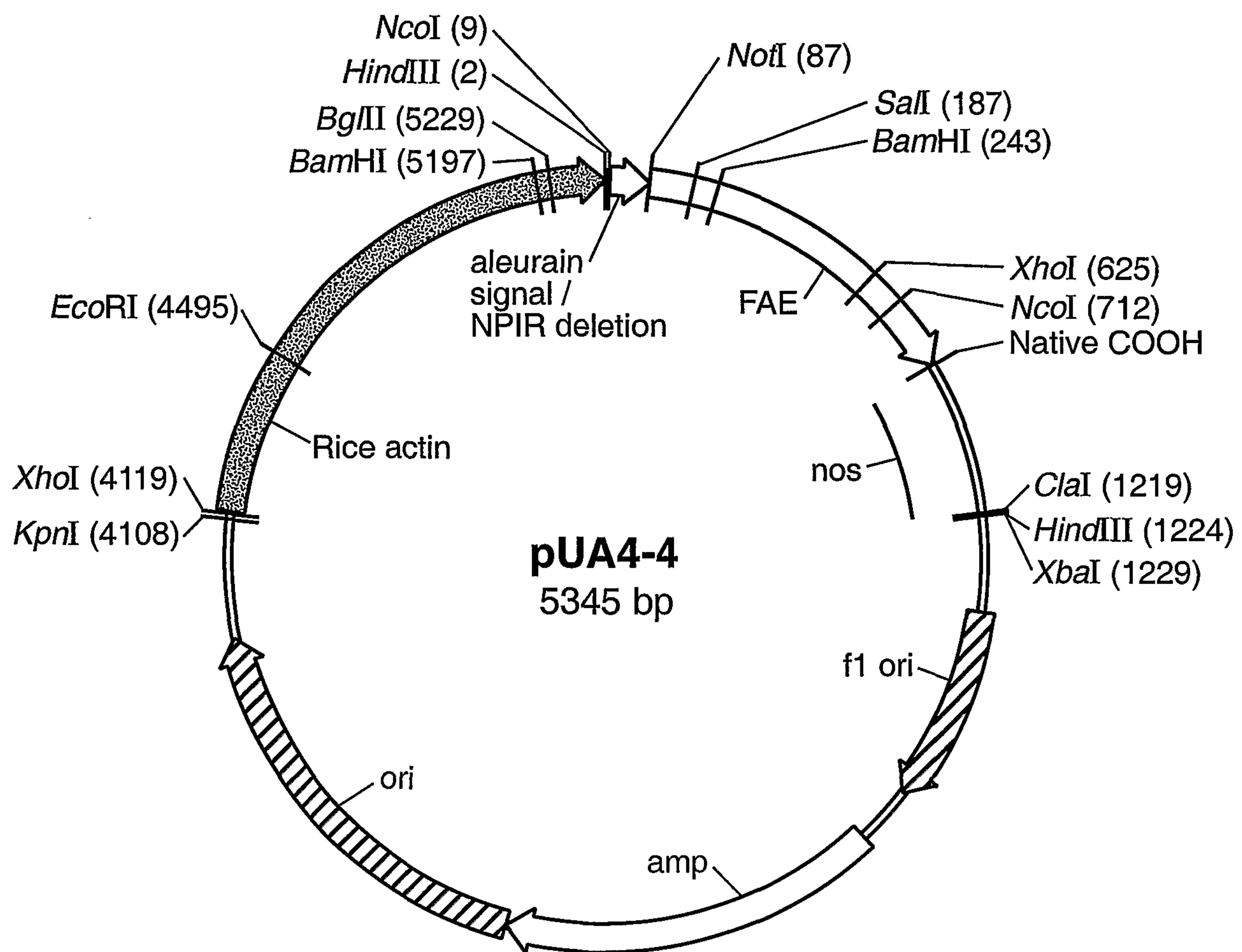
FIG._32C

1471 TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCG CCGCTCCTTT CGCTTTCTTC
1541 CCTTCCTTTC TCGCCACGTT CGCCGGCTTT CCCGTC AAG CTC TAAATCG GGGCTCCCT TTAGGGTTCC
1611 GATTTAGTGC TTACGGCAC CTCGACCCCA AAAACTTGA GTAGGGTAT TTAGGGTAT GGTTCACGTA GTGGCCATC
1681 GCCCTGATAG ACGGTTTTC CACTTTTTC GCTTGTGAC GTTGGAGTCC AC GTTCTTTA ATAGTGACT CTGTTCCAA
1751 ACTGGAACAA CACTCAACCC TATCTCGGC ATTCTTTG TATTTTAAAG GATTTGCCG ATTTCCGCC
1821 ATTGGTTAAA AATGAGCTG AATTAAACAA AATTAAACG CAAATTAAAC AAAATATTAA CGCTTACAAT
1891 TTAGGTGGCA CTTTTCGGG CATGAGACAA TAACTCTGAT AAATGCTTCA ATAATATTGA AAAAGGAAGA GTATGAGTAT
1961 TGTATCCGCT CATGAGACAA CGTGTCGCC TTATTCCTT TTTTGCCCTC ATGTTTTCG CACGAGTGG TTACATCGAA CTGCCAGAA
2031 TCAACATTT CAGTGTCGCC AAGTAAAGA TGCTGAAGAT CAGTTGGTG CAGCAACTCG GTCCGCCAT ACCTATTCT
2101 ACGCTGGTGA GATCCTTGAG GCGGTATTAT CTCACCAATC ACAGAAAGC ATCTTACGGA TGGCATGACA GTAAAGAAAT
2171 ACAGCGGTAA GATCCTTGAG GCGGTATTAT CTCACCAATC ACAGAAAGC ATCTTACGGA TGGCATGACA GTAAAGAAAT
2241 GCTATGTGGC CAGTATGAG GCGGTATTAT CTCACCAATC ACAGAAAGC ATCTTACGGA TGGCATGACA GTAAAGAAAT
2311 CAGAAATGACT TGGTGAAGTA TGGCATAACC ATGAGTGATA ACCTGCGG CAACTTACTT CAACAACGA TCGGAGGACC
2381 TATGCAGTGC GAAGGAGCTA ACCGCTTTT TGCACAACAT GGGGATCAT ACCTGCGG CAACTTACTT CAACAACGA TCGGAGGACC
2451 GAAGGAGCTA ACCGCTTTT TGCACAACAT GGGGATCAT ACCTGCGG CAACTTACTT CAACAACGA TCGGAGGACC
2521 CTGAATGAAG CCATACCAA TGGCATAACC ATGAGTGATA ACCTGCGG CAACTTACTT CAACAACGA TCGGAGGACC
2591 AACTATTAA CAGTATGAG GCGGTATTAT CTCACCAATC ACAGAAAGC ATCTTACGGA TGGCATGACA GTAAAGAAAT
2661 AGTTGCAGGA CACTTCTGC CAGTATGAG GCGGTATTAT CTCACCAATC ACAGAAAGC ATCTTACGGA TGGCATGACA GTAAAGAAAT
2731 GAGCGTGGGT CTCGCGGTAT GAGTCAAGCA ACTATGAGT AACGAAATAG ACAGATCGCT CTTTAAACT TCAATTTTAA
2801 ACACGACGGG GAGTCAAGCA ACTATGAGT AACGAAATAG ACAGATCGCT CTTTAAACT TCAATTTTAA
2871 TAAGCATGG TAACGTGTCAG ACCAAGTTTA CTCATATATA CTTTAGATTG ATTTAAACT CCGTTTTCGT
2941 TTAAAGGA TCTAGGTGAA GTCAGACCC CTCAGACAG CTAAGGATC TCAAGGAAAT TGACCAAAAT CCGTTTTCGT
3011 TCCACTGAGC GTCAGACCC CTCAGACAG CTAAGGATC TCAAGGATC TCAAGGAAAT TGACCAAAAT CCGTTTTCGT
3081 CTGCTGCTTG CAAACAAAAG AGGTAACTGG CTTCAAGAAC ACCACCGCT ACCAGCGGTG GTTTGTTTG GGTATCAAGA GTACCAACT
3151 CTTTTCCTG AGGTAACTGG CTTCAAGAAC ACCACCGCT ACCAGCGGTG GTTTGTTTG GGTATCAAGA GTACCAACT
3221 TAGGCCACCA CTTCAAGAAC ACCACCGCT ACCAGCGGTG GTTTGTTTG GGTATCAAGA GTACCAACT
3291 TGCTGCCAGT GGCATAAGT GGCATAAGT GGCATAAGT GGCATAAGT GGCATAAGT GGCATAAGT GGCATAAGT GGCATAAGT
3361 CCGTCGGCT GAACGGGGG TTCGTGCACA CAGCCAGCT TGGAGCGAAC GACCTACACC GACCTACACC GACCTACACC GACCTACACC
3431 ACCTACAGC TGAGCTATGA GAAAGCGCA CGCTTCCGA AGGAGAAAG GCGGACAGGT ATCCGGTAAG
3501 CGGCAGGTC GAAACAGGAG AGCGACGAG GGAGCTTCCA GGGGAAACG CTTGGTATCT TTAATAGTCTT
3571 GTCGGGTTTC GCCACCTCTG ACTTGAGCGT CGATTTTGT GATGCTCGTC AGGGGGCGG AGCTATGGA
3641 AAAACGCCAG CAACGCGGC TTTTACGGT TCCTGGCTT TGGCTCACA TTGCTCACA TTGCTCACA TTGCTCACA
3711 TGGGTTATCC CCTGATCTG TGGATAACCG TATTACCGC TTTGAGTGAG TTTGAGTGAG TTTGAGTGAG TTTGAGTGAG
3781 CGAACGACCG AGCGACCGA GTCAGTGAG GAGGAAGCG GAGGAAGCG GAGGAAGCG GAGGAAGCG GAGGAAGCG
3851 CCGCGGTTG GCCGATTCAT TAAATGAGT GGCACGAC GGTTCCTGAC TGGAAAGCG TGGAAAGCG TGGAAAGCG

FIG.-32D

| | | | | | | | |
|------|-------------|-------------|-------------|------------|-------------|------------|------------|
| 3921 | CAACGCAATT | AATGTGAGTT | AGCTCACTCA | TTAGGCACCC | CAGGCTTTAC | ACTTTATGCT | TCCGGCTCGT |
| 3991 | ATGTTGTGTG | GAAATGTGAG | CGGATAACAA | TTTCACACAG | GAAACAGCTA | TGACCATGAT | TACGCCAAGC |
| | | | | KpnI | XhoI | | |
| | | | | ~~~~~ | ~~~~~ | | |
| 4061 | GCGCAATTAA | CCCTCACTAA | AGGGAACAAA | AGCTGGGTAC | CGGGCCCCCC | CTCGAGGTCA | TTCATATGCT |
| 4131 | TGAGAAGAGA | GTCGGGATAG | TCCAAAATAA | AACAAAGGTA | AGATTACCTG | GTCAAAAGTG | AAAACATCAG |
| 4201 | TTAAAAGGTG | GTATAAGTAA | AATATCGGTA | ATAAAAGGTG | GCCCAAAGTG | AAATTTACTC | TTTTCTACTA |
| 4271 | TTATATAAAAT | TGAGGATGTT | TTGTCGGTAC | TTTGATACGT | CATTTTTGTA | TGAATTGGTT | TTTAAGTTTA |
| 4341 | TTCCGCGATT | GGAATATGCAT | ATCTGTATTT | GAGTCGGTTT | TTAAAGTTTCT | TGCTTTTGTA | AATACAGAGG |
| 4411 | GATTTGATA | AGAAATATCT | TTAAAAAACC | CATATGCTAA | TTTGACATAA | TTTTTGAGAA | AAATATATAT |
| | | | | ECORI | | | |
| | | | | ~~~~~ | | | |
| 4481 | TCAGGCGAAT | TCCACAAATGA | ACAATAATA | GATTAAAATA | GCTTGCCCCC | GTTCAGCGA | TGGGTATTTT |
| 4551 | TTCTAGTAAA | ATAAAAGATA | AACTTAGACT | CAAAACATTT | ACAAAACAA | CCCCTAAAGT | CCTAAAGCCC |
| 4621 | AAAGTGCTAT | GCACGATCCA | TAGCAAGCCC | AGCCCAACCC | AACCCACCCC | AACCCACCCC | AGTGCAGCCA |
| 4691 | ACTGGCAAA | AGTCTCCACC | CCCGGCACTA | TCACCGTGAG | TTGTCCGCAC | CACCGCACGT | CTCGCAGCCA |
| 4761 | AAAAAATAAA | AAGAAAGAAA | AAAAAGAAA | AGAAAACAG | CAGGTGGGTC | CGGTCTGTGG | GGGCCGGAAA |
| 4831 | AGCGAGGAGG | ATCGCGAGCA | GCGACGAGGC | CCGGCCCTCC | CTCCGCTTCC | AAAGAAACGC | CCCCCATCGC |
| 4901 | CACATATATAC | ATACCCCTCC | CTCTCTCTCC | ATCCCCCAA | CCCTACCAAC | ACCACCAACA | CCACCTCCTC |
| 4971 | CCCCCTCGCT | GCCGGACGAC | GAGCTCTCTC | CCCTCTCCCC | TCCGCCGCCG | CCGGTAACCA | CCCCGCCCTC |
| 5041 | CTCCTCTTTC | TTTCTCCGTT | TTTTTTTTCG | TCTCGGTCTC | GATCTTTGGC | CTTGGTAGTT | TGGGTGGGCG |
| 5111 | AGAGCGGCTT | CGTCGCCCAG | ATCGGTGCGC | GGGAGGGGCG | GGATCTCGCG | GCTGGCGTCT | CCGGGCGTGA |
| | | | | BamHI | BglII | | |
| | | | | ~~~~~ | ~~~~~ | | |
| 5181 | GTCGGCCCCG | ATCCTCGCGG | GGAATGGGGC | TCTCGGATGT | AGATCTTCTT | TCTTCTTCT | TTTTGTGGTA |
| 5251 | GAAATTGAAT | CCCTCAGCAT | TGTTCAATCGG | TAGTTTCTCT | TTTTCATGAT | TGTGACAAAT | GCAGCCTCGT |
| 5321 | GCGGAGCTTT | TTTGTAGC | | | | | |

FIG.-32E

**FIG. 33A**

NCOI
~~~~~
HindIII
~~~~~
M A H A R V L L L A L A V L A T A A V A V
1 AAGCTTACCA TGGCCACGC CCGGTCCTC CTCCTGGCG TCGCCGTGCT GCCACGGCC GCCGTCGCCG
NotI
~~~~~
. A S S R A A A S T Q G I S E D L Y S R L V E M .
71 TCGCCTCCTC CCGCGCGGC GCCTCCACGC AGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAAT
SalI
~~~~~
. A T I S Q A A Y A D L C N I P S T I I K G E K
141 GGCCACTATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC ATTCCGTCGA CTATTATCAA GGGAGAGAAA
BamHI
~~~~~
I Y N S Q T D I N G W I L R D D S S K E I I T V
211 ATTACAATT CTCAAACTGA CATTAACGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA ATAATCACCG
. F R G T G S D T N L Q L D T N Y T L T P F D T .
281 TCTTCCGTGG CACTGGTAGT GATACGAATC TACAACCTCGA TACTAACTAC ACCCTCAGC CTTTCGACAC
. L P Q C N G C E V H G G Y Y I G W V S V Q D Q
351 CCTACCACAA TGAACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA
V E S L V K Q Q V S Q Y P D Y A L T V T G H X L
421 GTCGAGTCGC TTGTCAAACA GCAGGTTAGC CAGTATCCGG ACTACGCGCT GACCGTGACC GGCCACKCCC
. G A S L A A L T A A Q L S A T Y D N I R L Y T .
491 TCGGCGCCTC CCTGGCGGCA CTCACTGCCG CCCAGCTGTC TCGACATAC GACAACATCC GCCTGTACAC
XhoI
~
. F G E P R S G N Q A F A S Y M N D A F Q A S S
561 CTTCGGCGAA CCGCGCAGCG GCAATCAGC CTTGCGGTCG TACATGAACG ATGCCTTCCA AGCCTCGAGC
P D T T Q Y F R V T H A N D G I P N L P P V E Q
631 CCAGATACGA CGCAGTATT CCGGTCACT CATGCCAACG ACGGCATCCC AAACCTGCC CCGGTGGAGC
NCOI
~~~~~
. G Y A H G G V E Y W S V D P Y S A Q N T F V C .

FIG._33B

701 AGGGGTACGC CCAATGGCGGT GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCAGAACA CATTTGCTG
• T G D E V Q C C E A Q G G Q G V N A H T T Y
771 CACTGGGGAT GAAGTGCAGT GCTGTGAGC CCAGGGCGGA CAGGGTGTGA ATAATGCGCA CACGACTTAT
F G M T S G A C T W *
841 TTTGGGATGA CGAGCGGAGC CTGTACATGG TGATCAGTCA TTTCAGCCTC CCCGAGTGTA CCAGGAAAGA
911 TGGATGTCCT GGAGAGGGGG CCGCGTAACC ACTGAAGGAT GAGCTGTAAA GAAGCAGATC GTTCAAACAT
981 TTGGCAATAA AGTTTCTTAA GATTGAATCC TGTGCGCGT TGTGCGATGA TTATCATATA ATTCCTGTTG
1051 AATTACGTTA AGCATGTAAT AATTAAACATG TAATGCATGA CGTTATTAT GAGATGGGTT TTTATGATTA
1121 GAGTCCCGCA ATTATACATT TAATACGCGA TAGAAAACAA AATATAGCGC GCAAAC TAGG ATAAATTATC
HindIII
~~~~~

Clal XbaI  
~~~~~

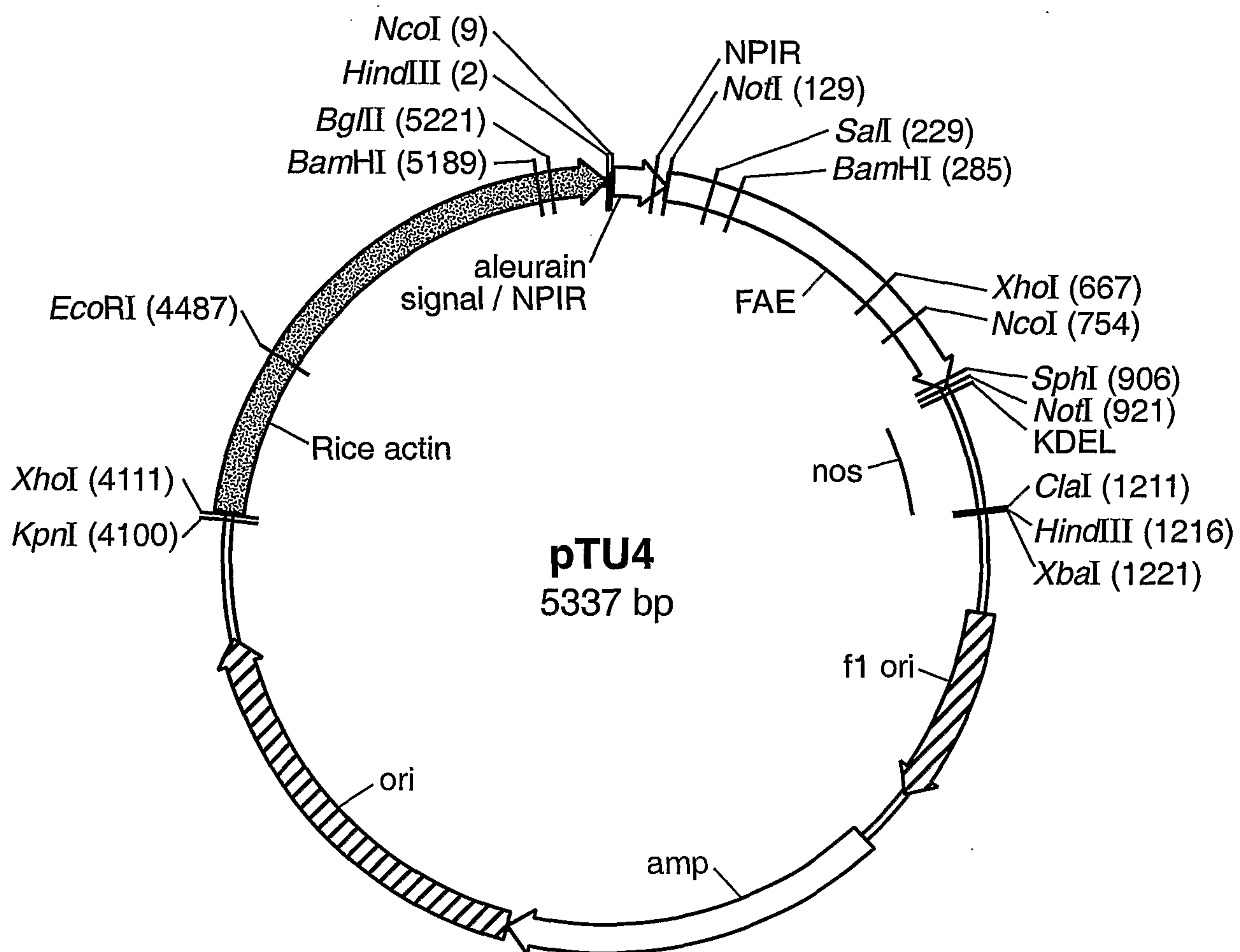
1191 GCGCGCGGTG TCATCTATGT TACTAGATCG ATAAGCTTCT AGAGCGGCGG GTGGAGCTCC AATTCGCCCT
1261 ATAGTGAGTC GTATTACGCG CGCTCACTGG CCGTCGTTTT ACAACGTCGT GACTGGGAAA ACCCTGGCGT
1331 TACCCAACTT AATCGCCTTG CAGCACATCC CCTTTCGCC AGCTGGCGTA ATAGCGAAGA GGCCCGCAC
1401 GATCGCCCTT CCCAACAGTT GCGAGCCTG AATGGCGAAT GGGACGCGC CTGTAGCGG GCATTAAGCG
1471 CGGCGGGTGT GGTGGTTACG CGCAGCGTGA CCGTACACT TGCCAGCGC CTAGCGCCG CTCCTTTCG
1541 TTTCTTCCCT TCCTTTCTCG CCACGTTCCG CGGCTTTCC CGTCAAGCTC TAAATCGGG GCTCCCTTA
1611 GGGTTCCGAT TTAGTGCTTT ACGCACCTC GACCCAAA AACTTGATTA GGTGATGGT TCACGTAGT
1681 GGCCATCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTCCTTAAATA GTGGACTCTT
1751 GTTCCAAACT GGAACAAAC TCAACCTAT CTCGGTCTAT TCTTTTGATT TATAAGGAT TTTGCCGATT
1821 TCGGCCATTT GGTAAAAA TGAGCTGATT TAACAAAAT TTAAACGCAA TTTTAAACAA ATATTAAACG
1891 TTACAATTTA GGTGGCACTT TTCGGGAAA TGTGCGCGGA ACCCTATT GTTTATTTT CTAATAACAT
1961 TCAAAATATGT ATCCGCTCAT GAGACAAATA CCTGATAAA TGCTTCAATA ATATTGAAA AGGAAGAGTA
2031 TGAGTATTCA ACATTCCGT GTCGCCCTTA TTCCCTTTT TCGCGCATTT TGCCTTCCTG TTTTGTCTCA
2101 CCCAGAAACG CTGGTGAAAG TAAAGATCG TGAAGATCAG TTGGGTGCAC GAGTGGGTTA CATCGAACTG
2171 GATCTCAACA GCGGTAAGAT CTTGAGAGT TTTGCGCCCG AAGAACGTTT TCCAATGATG AGCATTTTA
2241 AAGTTCCTG ATGTGGCGCG GTATTATCCC GTATTGACGC CGGGCAAGAG CAACCTCGGTC GCCGCATACA
2311 CTATTCTCAG AATGACTTGG TTGAGTACTC ACCAGTCACA GAAAAGCATC TTACGGATGG CATGACAGTA
2381 AGAGAAATTAT GCAGTGCTGC CATAACCATG AGTGATAACA CTGCGGCCAA CTTACTTCTG ACAACGATCG
2451 GAGGACCGAA GGAGCTAACC GCTTTTTCG ACAACATGGG GGATCATGTA ACTCGCCTTG ATCGTTGGGA
2521 ACCGGAGCTG AATGAAGCCA TACCAAACGA CGAGCGTGAC ACCACGATGC CTGTAGCAAT GGCAACAACG
2591 TTGCGCAAAC TATTAACTGG CGAACTACTT ACTCTAGCTT CCCGGCAACA ATTAATAGAC TGGATGGAGG
2661 CGGATAAAGT TGCAGGACCA CTCTCGCGCT CGGCCCTTCC GGCTGGCTGG TTTATTTGCTG ATAAATCTGG

FIG._33C

FIG. 33D

| | | | | | | | |
|------|-------------|------------|------------|-------------|------------|------------|------------|
| 4901 | CCATCGCCAC | TATATACATA | CCCCCCCCTC | TCCTCCCATC | CCCCAACCC | TACCACCACC | ACCACCACCA |
| 4971 | CCTCCTCCCC | CCTCGCTGCC | GGACGACGAG | CTCCTCCCCC | CTCCCCCTCC | GCCGCCGCCG | GTAACCACCC |
| 5041 | CGCCCCCTCTC | CTCTTTCCTT | CTCCGTTT | TTTTTCGTCT | CGGTCTCGAT | CTTGGCCCTT | GGTAGTTGG |
| 5111 | GTGGGCGAGA | GCGGCTTCGT | CGCCAGATC | GGTGCGCGGG | AGGGCGGGA | TCTCGCGGCT | GGCGTCGCCG |
| | | BamHI | | | BglII | | |
| | | ~~~~~ | | | ~~~~~ | | |
| 5181 | GGCGTGAGTC | GGCCCGGATC | CTCGCGGGA | ATGGGCTCT | CGGATGTAGA | TCTCTTCT | TTCTTCTTT |
| 5251 | TGTGGTAGAA | TTTGAATCCC | TCAGCATTGT | TCAATCGGTAG | TTTTTCTTTT | CATGATTGT | GACAAATGCA |
| 5321 | GCCTCGTGCG | GAGCTTTTTT | GTAGC | | | | |

FIG._33E

**FIG. 34A**

NCOI
~~~~~
HindIII
~~~~~
M A H A R V L L L A L A V L A T A A V A V
1 AAGCTTACCA TGGCCACGC CCGGTCCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTGCGCG
NotI
~~~~~
A S S S F A D S N P I R P V T D R A A A S T
71 TCGCCTCCTC CTCCTCCTTC GCCGACTCCA ACCGATCCG GCCCGTCACC GACCGCGCGG CCGCCTCCAC
Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC
Sali
~~~~~
D L C N I P S T I I K G E K I Y N S Q T D I N G
211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAAAC T GACATTAACG
BamHI
~~~~~
W I L R D D S S K E I I T V F R G T G S D T N
281 GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCACTGGTA GTGATACGAA
L Q L D T N Y T L T P F D T L P Q C N G C E V
351 TCTACAACTC GATACTAACT ACACCCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA
H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S
421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA
Q Y P D Y A L T V T G H X L G A S L A A L T A
491 GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC
A Q L S A T Y D N I R L Y T F G E P R S G N Q
561 CGCCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG
XhoI
~~~~~
A F A S Y M N D A F Q A S S P D T T Q Y F R V T
631 GCCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA
NCOI
~~~~~
H A N D G I P N L P P V E Q G Y A H G G V E Y

FIG.\_34B

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
• W S V D P Y S A Q N T F V C T G D E V Q C C E  
771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCAC TGGG ATGAAGTGCA GTGCTGTGAG  
SphI  
~~~~  
A Q G G Q G V N N A H T T Y F G M T S G A C T W
841 GCCCAGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC GCATGCACCT
NotI
~~~~~  
• P V A A A E P L K D E L \*  
911 GGCCGGTCGC GGCCGCGGAA CCACTGAAGG ATGAGCTGTA AAGAAAGCAGA TCGTTCAAAC ATTGGCAAT  
981 AAGGTTTCTT AAGATTGAAT CCTGTTGCCG GTC TTGCGAT GATTATCATA TAATTTCGT TGAATTACGT  
1051 TAAGCATGTA ATAAATTAACA TGTAAATGCAT GACGTTATTT ATGAGATGGG TTTTATATGAT TAGAGTCCCG  
1121 CAATTATACA TTTAATACGC GATAGAAAAC AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGG  
HindIII  
~~~~~  
Clal XbaI
~~~~~  
1191 TGTCAATCTAT GTTACTAGAT CGATAAGCTT CTAGAGCGGC CCGTGGAGCT CCAATTCGCC CTATAGTGAG  
1261 TCGTATTACG CGCGCTCACT GGCCGTCGTT TTACAACGTC GTGACTGGGA AAACCTTGGC GTTACCCAAC  
1331 TTAAATCGCCT TGCAGCACAT CCCCTTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCC  
1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCTGTAGCG CCCTCCTTTC CGCGCGGGT  
1471 GTGGTGGTTA CGCGCAGCGT GACCGCTACA CTTGCCAGCG CCTAGCGCC CGCTCCTTTC GCTTCTTCC  
1541 CTTCCCTTCT CGCCACGTC GCCGCTTTC CCCGTCAAGC TCTAAATCGG GGCCTCCCTT TAGGTTCCG  
1611 ATTTAGTGCT TTACGGCAC TCGACCCCAA AAAACTTGAT TAGGGTGATG GTTCACGTAG TGGGCCATCG  
1681 CCTGATAGA CGGTTTTCG CCTTTGACG TTGGAGTCCA CGTTC TTAA TAGTGGACTC TTGTTCCAAA  
1751 CTGGAACAAC ACTCAACCCT ATCTCGGTCT ATTC TTTGA TTTATAAGG ATTTTGCCGA TTTCGGCCTA  
1821 TTGGTTAAAA AATGAGCTGA TTAAACAAA AATTTAACA AATAATTAAC GCTTACAATT  
1891 TAGGTGGCAC TTTTCGGGA AATGTGCGG GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT  
1961 GTATCCGCTC ATGAGACAAAT AACCTGATA AATGCTTCAA TAATA TTGA AAAGGAAGAG TATGAGTATT  
2031 CAACATTTCC GTGTCGCCCT TATTCCTTT TTTGCGGCAT TTTGCCCTTC TGT TTTTGCT CACCCAGAAA  
2101 CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT TACATCGAAC TGGATCTCA  
2171 CAGCGGTAAG ATCCTTGAGA GTTTTCGCC CGAAGAACGT TTTC CAATGA TGAGCAC TTT TAAAGTCTG  
2241 CTATGTGGCG CGGTATTATC CCGTATTGAC GCCGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC  
2311 AGAATGACTT GGTGAGTAC TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCA TGACAG TAAAGAGAA  
2381 ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG

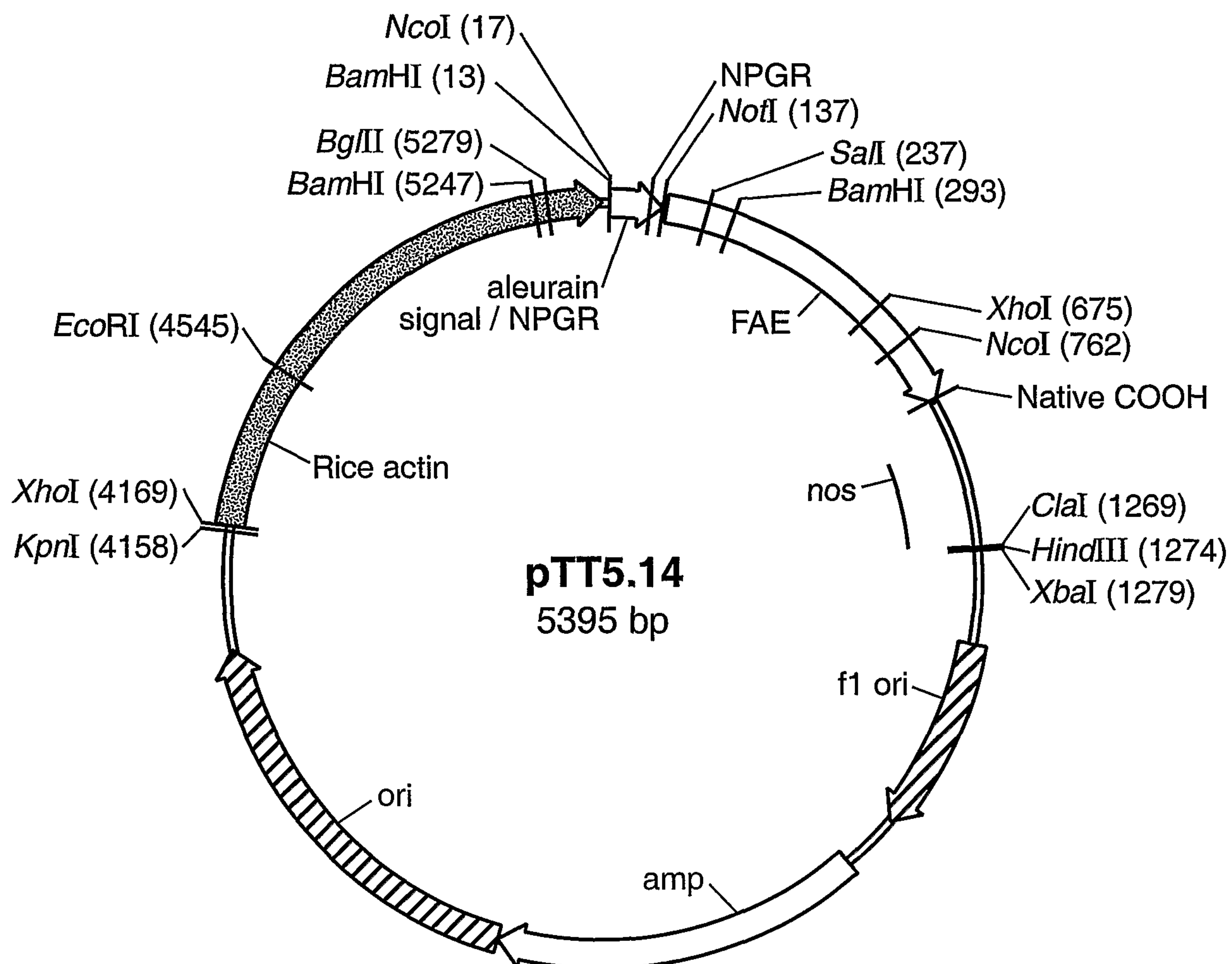
FIG.-34C

|           |            |            |            |             |             |             |            |
|-----------|------------|------------|------------|-------------|-------------|-------------|------------|
| 2451      | AAGGAGCTAA | CCGCTTTT   | GCACAACATG | GGGGATCATG  | TAACTCGCCT  | TGATCGTTGG  | GAACCGGAGC |
| 2521      | TGAATGAAGC | CATACCAAC  | GACGAGCGTG | ACACCACGAT  | GCCGTGTA    | ATGGCAACAA  | CGTTGCCGAA |
| 2591      | ACTATTAAC  | GGCGAACTAC | TTACTCTAGC | TTCCCGGCAA  | CAATTAATAG  | ACTGGATGGA  | GGCGGATAAA |
| 2661      | GTTGCAGGAC | CAC        | TTCTGCG    | CTCGGCTGGCT | GGTTTATTGC  | TGATAAAATCT | GGAGCCGGTG |
| 2731      | AGCGTGGGTC | TCGCGGTATC | ATTGCAGCAC | TGGGCCAGA   | TGGTAAGCCC  | TCCCCTATCG  | TAGTTATCTA |
| 2801      | CACGACGGG  | AGTCAGGCAA | CTATGGATGA | ACGAAATAGA  | CAGATCGCTG  | AGATAGGTGC  | CTCACTGATT |
| 2871      | AAGCATTGGT | AAC        | TGTCAGA    | CTAATCTCAT  | TTTAGATTGA  | TTTAAACCTT  | CATTTTAAAT |
| 2941      | TTAAAGGAT  | CTAGGTGAAG | ATCCTTTTG  | ATAATCTCAT  | GACCAAAATC  | CCTTAACGTG  | AGTTTTCGTT |
| 3011      | CCACTGAGCG | TCAGACCCCG | TAGAAAGAT  | CAAAGGATCT  | TCTTGAGATC  | CCTTTTTCCT  | GCGCGTAATC |
| 3081      | TGCTGCTTGC | AAACAAAAAA | ACCACCGCTA | CCAGCGGTGG  | TTTGTTTGCC  | GGATCAAGAG  | CTACCAACTC |
| 3151      | TTTTTCCGAA | GGTAAC     | TTCAGCAGAG | CGCAGATACC  | AAATACTGTC  | CTTCTAGTGT  | AGCCGTAGTT |
| 3221      | AGGCCACCAC | TTCAAGAACT | CTGTAGCAC  | GCCTACATAC  | CTCGCTCTGC  | TAAATCCTGTT | ACCAGTGGCT |
| 3291      | GCTGCCAGTG | GCGATAAGTC | GTGTCTTACC | GGTTGGACT   | CAAGACGATA  | GTTACC      | GTTACC     |
| 3361      | GGTCGGGCTG | AACGGGGGT  | TGTTGCACAC | AGCCAGCTT   | GGAGCGAACG  | ACCTACACCG  | AACTGAGATA |
| 3431      | CCTACAGCGT | GAGCTATGAG | AAAGGCCAC  | GCTTCCGAA   | GGAGAAAGG   | CGGACAGGTA  | TCCGGTAAGC |
| 3501      | GGCAGGTCG  | GAACAGGAGA | GCGCACGAGG | GAGCTTCCAG  | GGGAAACGC   | CTGGTATCTT  | TATAGTCCTG |
| 3571      | TCGGGTTTCG | CCACCTCTGA | CTTGAGCGTC | GATTTTGTG   | ATGCTCGTCA  | GGGGGGCGGA  | GCCTATGGAA |
| 3641      | AAACGCCAGC | AACGCGGCTT | TTTACGGTT  | CTTGGCTTT   | TGCTGGCTT   | TTGCTCACAT  | GTCTTTCCT  |
| 3711      | GCGTTATCCC | CTGATTCTGT | GGATAACCGT | ATTACCGCTT  | TTGAGTGAGC  | TGATACCGCT  | CGCCGCAGCC |
| 3781      | GAACGACCGA | GCGCAGCGAG | TCAGTGAGCG | AGGAAGCGGA  | AGAGCGCCCA  | ATACGCAAC   | CGCTCTCCC  |
| 3851      | CGCGCGTTGG | CCGATTCAAT | AATGCAGCTG | GCACGACAGG  | TTTCCCGACT  | GGAAAGCGGG  | CAGTGAGCGC |
| 3921      | AACGCAATTA | ATGTGAGTTA | GCTCACTCAT | TAGGCACCC   | AGGCTTTACA  | CTTTATGCTT  | CCGGCTCGTA |
| 3991      | TGTTGTGTGG | AATTGTGAGC | GGATAACAAT | TTCACACAGG  | AAACAGCTAT  | GACCATGATT  | ACGCCAAGCG |
| KpnI XhoI |            |            |            |             |             |             |            |
| 4061      | CGCAATTAA  | CC         | GGAAACAAA  | GCTGGGTACC  | GGGCCCCCCC  | TCGAGGTCA   | TCATATGCTT |
| 4131      | GAGAAAGAG  | TCGGGATAGT | CCAAAATAAA | ACAAAGGTAA  | GATTACCTGG  | TCAAAAGTGA  | AAACATCAGT |
| 4201      | TAAAAGGTGG | TATAAGTAAA | ATATCGGTAA | TAAAAGGTGG  | CCCAAAGTGA  | AATTACTCT   | TTTCTACTAT |
| 4271      | TATAAAAT   | GAGGATGTTT | TGTCGGTACT | TTGATACGTC  | ATTTTGTGAT  | GAA         | TTAAGTTTAT |
| 4341      | TCGCGATTG  | GAAATGCATA | TC         | AGTCGGTTT   | TAAAGTTCGTT | GCTTTTGTA   | ATACAGAGGG |
| 4411      | ATTTGTATA  | GAAATATCTT | TAAAAAACCC | ATATGCTAAT  | TTGACATAAT  | TTTTGAGAAA  | AAATATATAT |
| ECORI     |            |            |            |             |             |             |            |
| 4481      | CAGGCGAAT  | CCACAATGAA | CAATAATAAG | ATTAAAAATAG | CTTGCCCCCG  | TTGCAGCGAT  | GGGTATTTT  |
| 4551      | TCTAGTAAAA | TAAAAGATAA | ACTTAGACTC | AAAACATTTA  | CAAAAACAC   | CCCTAAAGTC  | CTAAAGCCCA |

FIG.-34D

|      |            |            |             |            |            |             |             |
|------|------------|------------|-------------|------------|------------|-------------|-------------|
| 4621 | AAGTGCTATG | CACGATCCAT | AGCAAGCCCA  | GCCCAACCCA | ACCCAACCCA | ACCCACCCCA  | GTGCAGCCAA  |
| 4691 | CTGGCAATA  | GTCGCCACCC | CCGGCACTAT  | CACCGTGAGT | TGTCCGCACC | ACCGCACGTC  | TCGCAGCCAA  |
| 4761 | AAAAAAAAA  | AGAAAGAAA  | AAAAGAAAA   | GAAAAACAGC | AGGTGGGTCC | GGGTCGTGGG  | GGCCGGAAAA  |
| 4831 | GCGAGGAGGA | TCGCGAGCAG | CGACGAGGCC  | CGGCCCTCCC | TCCGCTTCCA | AAGAAACGCC  | CCCCATCGCC  |
| 4901 | ACTATATACA | TACCCCCCCC | TCTCCTCCCA  | TCCCCCCAAC | CCTACCACCA | CCACCACCAC  | CACCTCCTCC  |
| 4971 | CCCCTCGCTG | CCGGACGACG | AGCTCCTCCC  | CCCTCCCCCT | CCGCCGCCGC | CGGTAACCAC  | CCCGCCCCCTC |
| 5041 | TCCTCTTTCT | TTCTCCGTTT | TTTTTTTCGT  | CTCGGTCCTG | ATCTTTGGCC | TTGGTAGTTT  | GGGTGGGCCGA |
| 5111 | GAGCGGCTTC | GTCGCCCAGA | TCGGTGC GCG | GGAGGGGCGG | GATCTCGCGG | CTGGCGTCTC  | CGGGCGTGAG  |
|      |            | BamHI      |             |            | BglII      |             |             |
|      | ~~~~~      |            |             |            | ~~~~~      |             |             |
| 5181 | TCGGCCCCGA | TCCTCGCGGG | GAAATGGGGCT | CTCGGATGTA | GATCTTCTTT | CTTTCCTCTT  | TTTGTGGTAG  |
| 5251 | AATTGGAATC | CCTCAGCATT | GTTCATCGGT  | AGTTTTTCTT | TTCATGATTT | GTGACAAAAAG | CAGCCTCCTG  |
| 5321 | CGGAGCTTTT | TTGTAGC    |             |            |            |             |             |

FIG. 34E

**FIG. 35A**

NcoI
~~~~~
BamHI
~~~~~
M
A
H
A
R
V
L
L
L
A
L
A
V
L
A
T
A
A
.
1
CCTGACGCCG
AGGATCCATG
GCCACGCC
GGTCCTCCT
CCTGGCGCTC
GCCGTGCTGG
CCACGGCCGC
NotI
.
V
A
V
A
S
S
S
F
A
D
S
N
P
G
R
P
V
T
D
R
A
A
71
CGTCGCCGTC
GCCCTCCTCCT
CCTCCTTCGC
CGACTCCAAC
CCGGGCCGGC
CCGTCACCGA
CCGCGCGGCC
NotI
~~
A
S
T
Q
G
I
S
E
D
L
Y
S
R
L
V
E
M
A
T
I
S
Q
A
A
141
GCCTCCACGC
AGGGCATCTC
CGAAGACCTC
TACAGCCGTT
TAGTCGAAAT
GGCCACTATC
TCCCAAGCTG
SalI
~~~~~
.
Y
A
D
L
C
N
I
P
S
T
I
I
K
G
E
K
I
Y
N
S
Q
T
D
.
211
CCTACGCCGA
CCTGTGCAAC
ATTCCGTCGA
CTATTATCAA
GGGAGAGAAA
ATTACAATT
CTCAAACTGA
BamHI
~~~~~
.
I
N
G
W
I
L
R
D
D
S
S
K
E
I
I
T
V
F
R
G
T
G
S
281
CATTAACGGA
TGGATCCTCC
GCGACGACAG
CAGCAAAGAA
ATAATCACCG
TCTTCCGTGG
CACTGGTAGT
D
T
N
L
Q
L
D
T
N
Y
T
L
T
P
F
D
T
L
P
Q
C
N
G
C
351
GATACGAATC
TACAACTCGA
TACTAACTAC
ACCCTCACGC
CTTTCGACAC
CCTACCACAA
TGCAACGGTT
.
E
V
H
G
G
Y
Y
I
G
W
V
S
V
Q
D
Q
V
E
S
L
V
K
Q
.
421
GTGAAGTACA
CGGTGGATAT
TATATTGGAT
GGTCTCCGT
CCAGGACCAA
GTCGAGTCGC
TTGTCAAACA
.
Q
V
S
Q
Y
P
D
Y
A
L
T
V
T
G
H
X
L
G
A
S
L
A
A
491
GCAGGTTAGC
CAGTATCCGG
ACTACGCGCT
GACCGTGACC
GGCCACKCCC
TCGGCGCCTC
CCTGGCGGCA
L
T
A
A
Q
L
S
A
T
Y
D
N
I
R
L
Y
T
F
G
E
P
R
S
G
561
CTCACTGCCG
CCCAGCTGTC
TGCACATAC
GACAAACATCC
GCCTGTACAC
CTTCGGCGAA
CCGCGCAGCG
XhoI
~~~~~
.
N
Q
A
F
A
S
Y
M
N
D
A
F
Q
A
S
S
P
D
T
T
Q
Y
F
.
631
GCAATCAGGC
CTTCGCGTCG
TACATGAACG
ATGCTTCCA
AGCCTCGAGC
CCAGATACGA
CGCAGTATT
NotI
~~~~~
.
R
V
T
H
A
N
D
G
I
P
N
L
P
P
V
E
Q
G
Y
A
H
G
G

FIG.\_35B

701 CCGGGTCACT CATGCCAACG ACGGCATCCC AAACCTGCCC CCGGTGGAGC AGGGGTACGC CCATGGCGGT  
V E Y W S V D P Y S A Q N T F V C T G D E V Q C  
771 GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCAGAACA CATTGTCTG CACTGGGAT GAAGTGCAGT  
. C E A Q G G Q G V N N A H T T Y F G M T S G A .  
841 GCTGTGAGG C CAGGGCGGA CAGGGTGTGA ATAATGCGA CACGACTTAT TTTGGGATGA CGAGCGGAGC  
. C T W \*  
911 CTGTACATGG TGATCAGTCA TTTCAGCCTC CCCGAGTGTA CCAGGAAAGA TGGATGTCCCT GGAGAGGGG  
981 CCGCGTAACC ACTGAAGGAT GAGCTGTAAA GAAGCAGATC GTTCAAACAT TTGGCAATAA AGTTTCTTAA  
1051 GATTGAATCC TGTGCGCGGT CTGCGGATGA TTATCATATA ATTTCTGTTG AATTACGTTA AGCATGTAAT  
1121 AATTAAACATG TAAATGCATGA CGTTATTTAT GAGATGGGTT TTTATGATTA GAGTCCCGCA ATTATACATT  
1191 TAATACGCGA TAGAAAACAA AATATAGCGC GCAAAC TAGG ATAAATTATC GCGCGCGGTG TCATCTATGT

HindIII

Clal

XbaI

~~~~~

~~~~~

1261 TACTAGATCG ATAAGCTTCT AGAGCGGCCG GTGGAGCTCC AATTCGCCCT ATAGTGAGTC GTATTACGCG  
1331 CGCTCACTGG CCGTCGTTT ACAACGTCGT GACTGGGAAA ACCCTGGCGT TACCCAACTT ATCGCCTTG  
1401 CAGCACATCC CCTTTTCGCC AGCTGGCGTA ATAGCGAAGA GGCCCGCAC GCATTAAGCG CGCGGGGTG GTGGTTACG  
1471 GCGCAGCCTG AATGGCGAAT GGACGCGCC CTGAGCGCCG CTCCTTTCGC TTTCTTCCCT TCTTTCCTG  
1541 CGCAGCGTGA CCGCTACACT TGCCAGCGCC TAAATCGGG GCTCCCTTA GGGTTCGGAT TTAGTGCTTT  
1611 CCACGTTTCG CCGCTTTCCC CGTCAAGCTC TAAATCGGG TACGACTCTT GTGCCAAT TCGGCCATTT GTTAAAAA  
1681 ACGGCACCTC GACCCCAAAA AACTTGATTA GGTGATGGT TCACGATAGT GCCATCGCC CTGATAGACG  
1751 GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTCTTTAATA GTGGACTCTT GTGCCAAT TCGGCCATTT GTTAAAAA  
1821 TCAACCCCTAT CTCGGTCTAT TCTTTTGATT TATAAGGGAT TTTGCCGATT TTTCCAAACT GGTAAAAA  
1891 TGAGCTGATT TAACAAAAAT TTAACGCGAA TTTTAACAA ATATTAAACG ATATTATA GTGGCATT  
1961 TTCGGGAAA TGTGCGCGGA ACCCTATT TTTTATT TTTTATACAT CTAAATATGT ATCCGCTCAT  
2031 GAGACAATA CCTGATAA TGCTTCAATA ATATTGAAA AGGAAGAGTA TGAGTATTCA ACATTCCGT  
2101 GTCGCCCTTA TTCCCTTTT TCGGCAATT TGCCTTCTG TTTTTCCTCA CCCAGAAACG CTGGTGAAAG  
2171 TAAAAGATGC TGAAGATCAG TTGGGTGCAC GAGTGGGTTA CATCGAACTG GATCTCAACA GCGTAAGAT  
2241 CCTTGAGAGT TTTCGCCCCG AAGAACGTTT TCCAATGATG AGCACTTTTA AAGTTCGCT ATGTGGCGCG  
2311 GTATTATCCC GTATTGACGC CGGCAAGAG CAACTCGGT CCGGCATACA CTATTCTCAG AATGACTTGG  
2381 TTGAGTACTC ACCAGTCACA GAAAAGCATC TTACGGATGG CATGACAGTA AGAGAAATAT GCAGTGTGC  
2451 CATAACCATG AGTGATAACA CTGCGGCCAA CTTACTTCTG ACAACGATCG GAGGACCGAA GGAGCTAAC  
2521 GCTTTTTCG ACAACATGGG GGATCATGTA ACTCGCCTTG ATCGTTGGA ACCGGAGCTG AATGAAGCCA  
2591 TACCAAACGA CGAGCGTGAC ACCACGATG CTGTAGCAAT GGCAACAACG TTGCGCAAC TATTAACTGG  
2661 CGAACTACTT ACTCTAGCTT CCCGGCAACA ATTAATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA

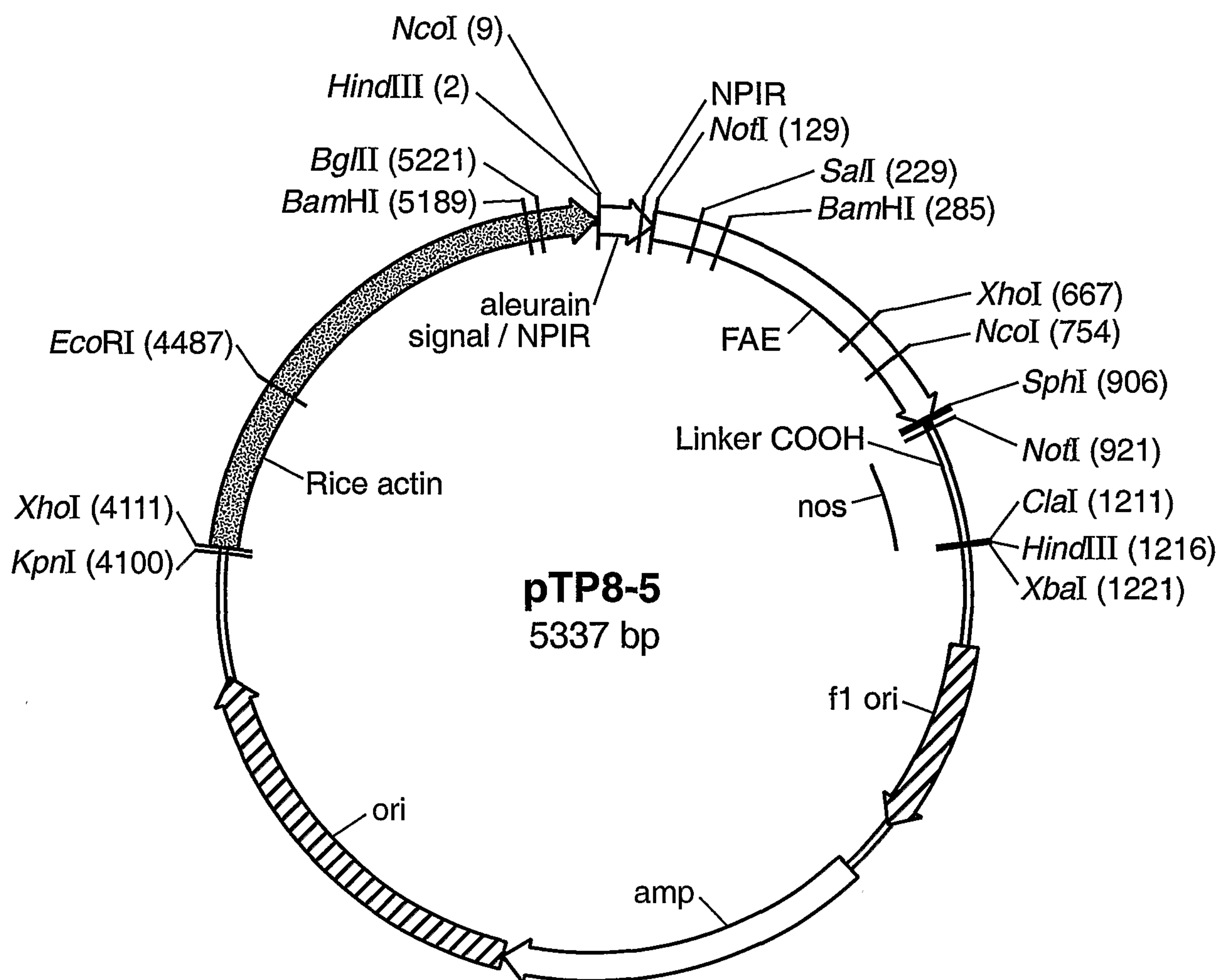
FIG. 35C

|       |             |            |            |             |             |             |             |
|-------|-------------|------------|------------|-------------|-------------|-------------|-------------|
| 2731  | CTTCTGCGCT  | CGGCCCTTCC | GGCTGGCTGG | TTTATTGCTG  | ATAAATCTGG  | AGCCGGTGAG  | CGTGGGTCTC  |
| 2801  | GCGGTATCAT  | TGCAGCACTG | GGCCAGATG  | GTAAGCCCTC  | CCGTATCGTA  | GTTATCTACA  | CGACGGGGAG  |
| 2871  | TCAGGCAACT  | ATGGATGAAC | GAAATAGACA | GATCGCTGAG  | ATAGGTGCCCT | CACTGAATTAA | GCAATTGGTAA |
| 2941  | CTGTCAGACC  | AAGTTTACTC | ATATATACTT | TAGATTGATT  | TAAAACCTCA  | TTTTTTAATT  | AAAAGGATCT  |
| 3011  | AGGTGAAGAT  | CCTTTTGGAT | AATCTCATGA | CCAAAATCCC  | TTAAACGTGAG | TTTTTCGTTCC | ACTGAGCGTC  |
| 3081  | AGACCCCGTA  | GAAAAGATCA | AAGGATCTTC | TTGAGATCCT  | TTTTTTTCTGC | GCGTAAATCTG | CTGCTTGCAA  |
| 3151  | ACAAAAAAAC  | CACCGCTACC | AGCGGTGGTT | TGTTTGCCGG  | ATCAAGAGCT  | ACCAACTCTT  | TTTCCGAAGG  |
| 3221  | TAACCTGGCT  | CAGCAGAGCG | CAGATACCAA | ATACTGTCCT  | TCTAGTGTAG  | CCGTAGTTAG  | GCCACCACTT  |
| 3291  | CAAGAACTCT  | GTAGCACCGC | CTACATACCT | CGTCTGCTA   | ATCCTGTTAC  | CAGTGGCTGC  | TGCCAGTGGC  |
| 3361  | GATAAGTCGT  | GTCTTACCGG | GTTGGACTCA | AGACGATAGT  | TACCGGATAA  | GGCGCAGCGG  | TCGGGCTGAA  |
| 3431  | CGGGGGGTT   | GTGCACACAG | CCAGCTTGG  | AGCGAACGAC  | CTACACCGAA  | CTGAGATACC  | TACAGCGTGA  |
| 3501  | GCTATGAGAA  | AGGCCACGC  | TTCCCGAAAG | GAGAAAGCG   | GACAGGTATC  | CGGTAAGCGG  | CAGGGTCGGA  |
| 3571  | ACAGGAGAGC  | GCACGAGGGA | GCTTCCAGGG | GGAACGCCCT  | GGTATCTTTA  | TAGTCCCTGTC | GGGTTTCGCC  |
| 3641  | ACCTCTGACT  | TGAGCGTCGA | TTTTTGTGAT | GCTCGTCAGG  | GGGCGGAGC   | CTATGGAAAA  | ACGCCAGCAA  |
| 3711  | CGCGGCCCTT  | TTACGGTTCC | TGGCCTTTTG | CTGGCCTTTT  | GCTCACATGT  | TCTTTTCTGC  | GTTATCCCCCT |
| 3781  | GATTCTGTGG  | ATAACCGTAT | TACCGCCTTT | GAGTGAGCTG  | ATACCGCTCG  | CCGCAGCCGA  | ACGACCGAGC  |
| 3851  | GCAGCGAGTC  | AGTGAGCGAG | GAAAGCGAAG | AGCGCCCAAT  | ACGCAAAACCG | CCTCTCCCCG  | CGCGTTGGCC  |
| 3921  | GATTCAATTAA | TGCAGCTGGC | ACGACAGGTT | TCCCGACTGG  | AAAGCGGGCA  | GTGAGCGCAA  | CGCAATTAAAT |
| 3991  | GTGAGTTAGC  | TCACTCATTA | GGCACCCAG  | GCTTTACACT  | TTATGCTTCC  | GGCTCGTATG  | TTGTGTGGAA  |
| 4061  | TTGTGAGCGG  | ATAACAAATT | CACACAGGAA | ACAGCTATGA  | CCATGATTAC  | GCCAAGCGCG  | CAATTAAACCC |
| KpnI  |             |            |            |             |             |             |             |
| ~~~~~ |             |            |            |             |             |             |             |
| 4131  | TCACTAAAGG  | GAACAAAAGC | TGGGTACCGG | GCCCCCCTC   | GAGGTCATTC  | ATATGCTTGA  | GAAGAGAGTC  |
| 4201  | GGGATAGTCC  | AAAATAAAAC | AAAGGTAAGA | TTACCTGGTC  | AAAAGTGAAA  | ACATCAGTTA  | AAAGGTGGTA  |
| 4271  | TAAGTAAAT   | ATCGGTAATA | AAAGGTGGCC | CAAAGTGAAA  | TTTACTCTTT  | TCTACTATTA  | TAAAAATTGA  |
| 4341  | GGATGTTTGG  | TCGGTACTTT | GATACGTCAT | TTTTTGATGA  | ATTGGTTTTT  | AAGTTTATTC  | GCGATTTGGA  |
| 4411  | AATGCATATC  | TGTATTTGAG | TCGGTTTTTA | AGTTCGTTGC  | TTTTTGTAAT  | ACAGAGGGAT  | TTGTATAAGA  |
| EcoRI |             |            |            |             |             |             |             |
| ~~~~~ |             |            |            |             |             |             |             |
| 4481  | AAATATCTTA  | AAAACCCAT  | ATGCTAAATT | GACATAAATT  | TTGAGAAAAA  | TATATATTCA  | GGCGAATTCC  |
| 4551  | ACAAATGAACA | ATAATAAGAT | TAAAATAGCT | TGCCCCCGTT  | GCAGCGATGG  | GTAATTTTTC  | TAGTAAAAATA |
| 4621  | AAAGATAAAC  | TTAGACTCAA | AACATTTACA | AAAACAACCC  | CTAAAGTCCT  | AAAGCCCAA   | GTGCTATGCA  |
| 4691  | CGATCCATAG  | CAAGCCAGC  | CCAACCCAAC | CCAACCCAAC  | CCACCCAGT   | GCAGCCAACT  | GGCAAAATAGT |
| 4761  | CTCCACCCC   | GGCACTATCA | CCGTGAGTTG | TCCGCACCCAC | CGCACGTCCT  | GCAGCCAAAA  | AAAAAAAAG   |
| 4831  | AAAGAAAAAA  | AAGAAAAAGA | AAAACAGCAG | GTGGGTCCGG  | GTCGTGGGG   | CCGGAAAAAGC | GAGGAGGATC  |

**FIG. 35D**

|      |    |      |      |     |      |     |     |    |     |     |     |     |     |     |     |     |     |     |    |     |     |       |          |
|------|----|------|------|-----|------|-----|-----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|-----|-----|-------|----------|
| 4901 | GC | GAG  | CAG  | CG  | AC   | GAG | GCC | CG | GCC | CTC | CGT | TCC | AAA | GAA | ACG | CCC | CCA | TCC | GC | CAC | TAT | ATAC  | ATA      |
| 4971 | CC | CCCC | CC   | TC  | CCCA | ACC | TAC | CA | CA  | CC  | TAC | CA  | CA  | ACC | ACC | ACC | CC  | CT  | CC | CCC | CT  | CG    | CTGCC    |
| 5041 | GG | AC   | GAC  | GAG | CT   | CC  | TC  | CC | CC  | CT  | CC  | GCC | CG  | CCG | GTA | ACC | ACC | CG  | CC | CT  | CT  | TT    | CTTT     |
| 5111 | CT | CCG  | TTTT | TT  | TTT  | TC  | GT  | CT | CG  | AT  | CT  | TG  | GC  | CT  | GT  | AG  | TT  | GG  | AG | AG  | GC  | TT    | CGT      |
| 5181 | CG | CC   | CA   | GAT | GG   | TC  | GC  | CG | GG  | AG  | GG  | CG  | GG  | GA  | TC  | TC  | GC  | CG  | GG | TC  | AG  | TC    | GGGATC   |
|      |    |      |      |     |      |     |     |    |     |     |     |     |     |     |     |     |     |     |    |     |     |       | BamHI    |
|      |    |      |      |     |      |     |     |    |     |     |     |     |     |     |     |     |     |     |    |     |     |       |          |
| 5251 | CT | CG   | CG   | GG  | GA   | AT  | GG  | GC | CT  | CT  | CT  | TT  | CT  | TT  | CT  | TT  | CT  | TT  | TT | TG  | TT  | GA    | ATCCC    |
| 5321 | TC | AG   | CA   | TT  | GT   | TC  | AT  | CG | GT  | AG  | TT  | TT  | CT  | TT  | CT  | TT  | CT  | TT  | CT | GC  | AA  | ATGCA | GAGCTTTT |
| 5391 | GT | AG   |      |     |      |     |     |    |     |     |     |     |     |     |     |     |     |     |    |     |     |       |          |

FIG.\_35E

**FIG.\_36A**

NcoI
 ~~~~~
 HindIII
 ~~~~~
 M A H A R V L L L A L A V L A T A A V A V
 1 AAGCTTACCA TGGCCACGC CCGGTCCTC CTCCTGGCG TCGCCGTGCT GGCCACGGCC GCCGTGCGCG
 NotI
 ~~~~~
 . A S S S F A D S N P I R P V T D R A A A S T .
 71 TCGCCTCCTC CTCCTCCTTC GCCGACTCCA ACCGATCCG GCCGTCACC GACCGCGCGG CCGCCTCCAC
 . Q G I S E D L Y S R L V E M A T I S Q A A Y A
 141 GCAGGGCATC FCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC
 SalI
 ~~~~~
 D L C N I P S T I I K G E K I Y N S Q T D I N G
 211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTTACAA TTCTCAAACT GACATTAACG
 BamHI
 ~~~~~
 . W I L R D D S S K E I I T V F R G T G S D T N .
 281 GATGGATCCT CCGGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCAC TGGA GTGATACGAA
 . L Q L D T N Y T L T P F D T L P Q C N G C E V
 351 TCTACAATC GATACTAAT ACACCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA
 H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S
 421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA
 . Q Y P D Y A L T V T G H X L G A S L A A L T A .
 491 GCCAGTATCC GGAATACGG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC
 . A Q L S A T Y D N I R L Y T F G E P R S G N Q
 561 CGCCAGCTG TCTGCGACAT ACGACAACAT CGCCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG
 XhoI
 ~~~~~
 A F A S Y M N D A F Q A S S P D T T Q Y F R V T
 631 GCCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCAGATAC GACGCAGTAT TTCCGGGTCA
 NcoI
 ~~~~~
 . H A N D G I P N L P P V E Q G Y A H G G V E Y .

FIG._36B

| | | | | | | | |
|------|-------------|-------------|------------|-------------|-------------|--------------|-------------|
| 701 | CTCATGCCAA | CGACGGCATC | CCAAACCTGC | CCCCGGTGGA | GCAGGGGTAC | GCCCATGGCG | GTGTAGAGTA |
| | • W S V | D P Y S | A Q N | T F V | C T G D | E V Q | C C E |
| 771 | CTGGAGCGTT | GATCCTTACA | GCGCCAGAA | CACATTTGTC | TGCACTGGGG | ATGAAAGTGCA | GTGCTGTGAG |
| | | | | | | | SphI |
| | | | | | | | ~~~~ |
| | A Q G G | Q G V | N N A | H T T Y | F G M | T S G | A C T W |
| 841 | GCCAGGGCG | GACAGGGTGT | GAATAATCGG | CACACGACTT | ATTTTGGGAT | GACGAGCGGC | GCATGCACCT |
| | | | | | | | NotI |
| | | | | | | | ~~~~~ |
| | • P V A | A A * | | | | | |
| 911 | GGCCGGTCGC | GGCCGCGTAA | CCACTGAAGG | ATGAGCTGTA | AAGAAGCAGA | TCGTTCAAAC | ATTTGGCAAT |
| 981 | AAAGTTTCTT | AAGATTGAAT | CCTGTGCCG | GTC TTGCCAT | GATTATCATA | TAA TTCTGT | TGAAATTACGT |
| 1051 | TAAGCATGTA | ATAATTAAACA | TGTAATGCAT | GACGTTATTT | ATGAGATGGG | TTTTTATGAT | TAGAGTCCCG |
| 1121 | CAATTATACA | TTTAATACGC | GATAGAAAAC | AAAATATAGC | GCGCAAAC TA | GGATAAATTA | TCGCGCGCGG |
| | | | | | | | HindIII |
| | | | | | | | ~~~~~ |
| | | | | | | | Clal |
| | | | | | | | XbaI |
| | | | | | | | ~~~~~ |
| 1191 | TGTCATCTAT | GTTACTAGAT | CGATAAGCTT | CTAGAGCGGC | CGGTGGAGCT | CCAATTCCGC | CTATAGTGAG |
| 1261 | TCGTATTACG | CGCGTCACT | GGCCGTCGTT | TTACAACGTC | GTGACTGGGA | AAACCCCTGGC | GTTACCCAAC |
| 1331 | TTAATCGCCT | TGCAGCACAT | CCCCCTTTCG | CCAGCTGGCG | TAATAGCGAA | GAGGCCCGCA | CCGATCGCCC |
| 1401 | TTCCCAACAG | TGCGCAGCC | TGAATGGCGA | ATGGGACGCG | CCCTGTAGCG | GCGCATTAAAG | CGCGGCGGGT |
| 1471 | GTGGTGGTTA | CGCGCAGCGT | GACCGCTACA | CTTGCCAGCG | CCCTAGCGCC | CGCTCCCTTT | CGTTTCTTCC |
| 1541 | CTTCCCTTCT | CGCCACGTT | GCCGGCTTTC | CCCGTCAAGC | TCTAAATCGG | GGGCTCCCTT | TAGGGTTCCG |
| 1611 | ATTTAGTGCT | TTACGGCAC | TCGACCCCAA | AAACTTGAT | TAGGGTGATG | GTTCAACGTAG | TGGGCCATCG |
| 1681 | CCCTGATAGA | CGGTTTTCG | CCCTTTGACG | TTGGAGTCCA | CGTTC TTAA | TAGTGGACTC | TTGTTCCAAA |
| 1751 | CTGGAACAAC | ACTCAACCT | ATCTCGGTCT | ATTC TTGA | TTTATAAGG | ATTTGCCGA | TTTCGGCCTA |
| 1821 | TTGGTTAAAA | AATGAGCTGA | TTTAACAAA | ATTTAACGCG | AA TTTAACA | AAATA TTAAAC | GCTTACAATT |
| 1891 | TAGGTGGCAC | TTTTCGGGA | AATGTGCGCG | GAACCCCTAT | TTGTTTATTT | TTCTAAATAC | ATTCAAATAT |
| 1961 | GTATCCGCTC | ATGAGACAA | TAATCTGATA | AAATGCTTCAA | TAATATGAA | AAAGGAAGAG | TATGAGTATT |
| 2031 | CAACATTTC | GTGTCGCCCT | TATTCCTTT | TTTGCGGCAT | TTTGCCCTTC | TGTTTTTGCT | CACCCAGAAA |
| 2101 | CGCTGGTGAA | AGTAAAAGAT | GCTGAAGATC | AGTTGGGTGC | ACGAGTGGGT | TACATCGAAC | TGGA TCTCAA |
| 2171 | CAGCGGTAAG | ATCCTTGAGA | GTTTTCGCC | CGAAGAACGT | TTTCCAATGA | TGAGCATT | TAAAGTTCTG |
| 2241 | CTATGTGGCG | CGGTATTATC | CCGTATTGAC | GCCGGGCAAG | AGCAACTCGG | TCGCCGCATA | CAC TATTCTC |
| 2311 | AGAA TGACTT | GGTTGAGTAC | TCACCAGTCA | CAGAAAAGCA | TCTTACGGAT | GGCATGACAG | TAAGAGAA TT |

FIG.-36C

| | | | | | | | |
|-------|-------------|--------------|-------------|--------------|-------------|-------------|-------------|
| 2381 | ATGCAGTGCT | GCCATAACCA | TGAGTGATAA | CACTGCGGCC | AACTTACTTC | TGACAACGAT | CGGAGGACCG |
| 2451 | AAGGAGCTAA | CCGCTTTT | GCACAACATG | GGGATCATG | TAACTCGCCT | TGATCGTTGG | GAACCGGAGC |
| 2521 | TGAATGAAGC | CATACCAAC | GACGAGCGTG | ACACCACGAT | GCCTGTAGCA | ATGGCAACAA | CGTTGCGCAA |
| 2591 | ACTATTAACT | GGCGAACTAC | TTACTCTAGC | TTCCCCGGCAA | CAATTAAATAG | ACTGGATGGA | GGCGGATAAA |
| 2661 | GTTGCAGGAC | CAC TTC TGCG | CTCGGCCCTT | CCGGCTGGCT | GGTTTATTGC | TGATAAATCT | GGAGCCGGTG |
| 2731 | AGCGTGGGTC | TCGCGGTATC | ATTGCAGCAC | TGGGGCCAGA | TGGTAAGCCC | TCCC GTATCG | TAGTTATCTA |
| 2801 | CACGACGGG | AGTCAGGCAA | CTATGGATGA | ACGAAATAGA | CAGATCGCTG | AGATAGGTGC | CTCACTGATT |
| 2871 | AAGCATTGGT | AAC TGTCAGA | CCAAGTTTAC | TCATAATATAC | TTTAGATTGA | TTTAAAAC TT | CATTTTTAAT |
| 2941 | TTAAAAGGAT | CTAGGTGAAG | ATCCTTTT | ATAATCTCAT | GACCAAAATC | CTTTAACGTG | AGTTTTCGTT |
| 3011 | CCACTGAGCG | TCAGACCCCG | TAGAAAAGAT | CAAAGGATCT | TCTTGAGATC | CTTTT TTTCT | GCGCGTAATC |
| 3081 | TGCTGCTTGC | AAACAAA AAA | ACCACCGCTA | CCAGCGGTGG | TTTGTTTGCC | GGATCAAGAG | CTACCAACTC |
| 3151 | TTTTTCCGAA | GGTAAC TGCG | TTCAGCAGAG | CGCAGATACC | AAATACTGTC | CTTCTAGTGT | AGCCGTAGTT |
| 3221 | AGGCCACCAC | TTCAAGAACT | CTGTAGCAC | GCCTACATAC | CTCGCTCTGC | TAATCCTGTT | ACCAGTGGCT |
| 3291 | GCTGCCAGTG | GCGATAAGTC | GTGTCTTACC | GGGTGGACT | CAAGACGATA | GTTACC GGAT | AAGGCGCAGC |
| 3361 | GGTCGGGCTG | AACGGGGGT | TCGTGCACAC | AGCCAGCTT | GGAGCGAACG | ACCTACACCG | AAC TGAGATA |
| 3431 | CCTACAGCGT | GAGCTATGAG | AAAGCGCCAC | GCTTCCCGAA | GGAGAAAGG | CGGACAGGTA | TCCGGTAAGC |
| 3501 | GGCAGGGTCG | GAACAGGAGA | CGCACGAGG | GAGCTTCCAG | GGGAAAACGC | CTGGTATCTT | TATAGTCCTG |
| 3571 | TCGGGTTCG | CCACCTCTGA | CTTGAGCGTC | GATTTT | ATGCTCGTCA | GGGGGCGGA | GCCTATGGAA |
| 3641 | AAACGCCAGC | AACGCGGCT | TTTTTACGGTT | CCTGGCCTTT | TGCTGGCCTT | TTGCTCACAT | GTTCTTTCCCT |
| 3711 | GCGTTATCCC | CTGATTC TGT | GGATAACCGT | ATTACCGCCT | TTGAGTGAGC | TGATACCGCT | CGCCGCAGCC |
| 3781 | GAACGACCGA | GCGCAGCGAG | TCAGTGAGCG | AGGAAGCGGA | AGAGCGCCA | ATACGCAAAC | CGCCTCTCCC |
| 3851 | CGCGCGTTGG | CCGATTTCATT | AATGCAGCTG | GCACGACAGG | TTTCCCGACT | GGAAGCGGG | CAGTGAGCGC |
| 3921 | AACGCAATTA | ATGTGAGTTA | GCTCACTCAT | TAGGCACCCC | AGGCTTTACA | CTTTATGCTT | CCGGCTCGTA |
| 3991 | TGTTGTGTGG | AATTGTGAGC | GGATAACAAT | TTCACACACAGG | AAACAGCTAT | GACCATGATT | ACGCCAAGCG |
| ~~~~~ | | | | | | | |
| 4061 | CGCAATTAAAC | CCTCACTAAA | GGGAACAAA | GCTGGGTACC | GGGCCCCCCC | TCGAGGTCAT | TCATATGCTT |
| 4131 | GAGAAGAGAG | TCGGGATAGT | CCAAAATAAA | ACAAAGGTAA | GATTACCTGG | TCAAAAGTGA | AAACATCAGT |
| 4201 | TAAAAGGTGG | TATAAGTAAA | ATATCGGTAA | TAAAAGGTGG | CCCAAAGTGA | AATTTACTCT | TTTCTACTAT |
| 4271 | TATAAAAATT | GAGGATGTTT | TGTCGGTACT | TTGATACGTC | ATTTTGTAT | GAATTGGTTT | TTAAGTTTAT |
| 4341 | TCGCGAATTG | GAAATGCATA | TCTGTAATTG | AGTCGGT TTT | TAAGTTCGTT | GC TTTGTAA | ATACAGAGGG |
| 4411 | ATTTGTTATA | GAAATATCTT | TAAAAAACCC | ATATGCTAAT | TTGACATAAT | TTTTTGAGAAA | AATATATATT |
| ~~~~~ | | | | | | | |
| 4481 | CAGGCGAATT | CCACAA TGAA | CAATAATAAG | ATTAAAATAG | CTTGCCCCCG | TTGCAGCGAT | GGGTATTTT |
| ~~~~~ | | | | | | | |

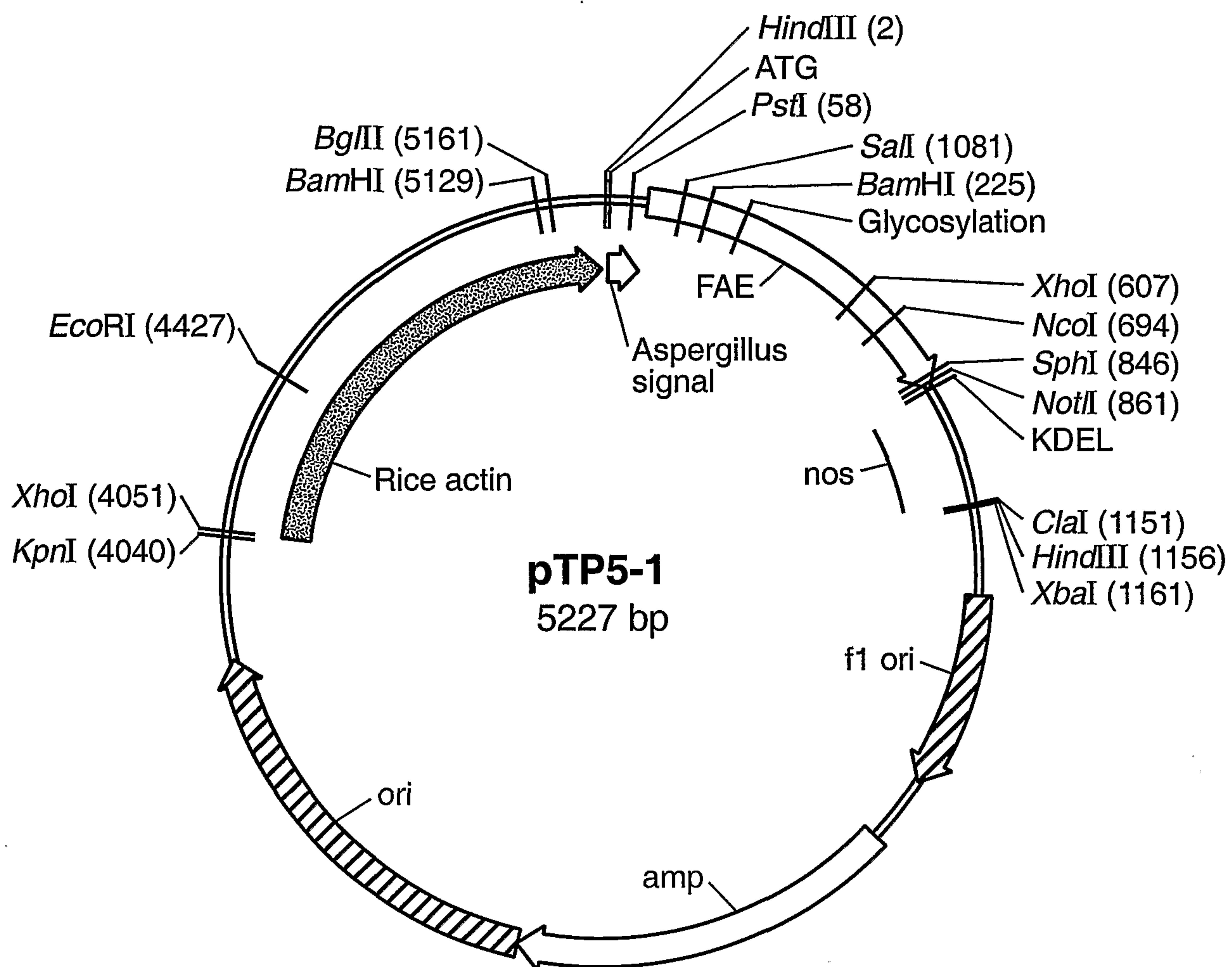
~~~~~  
KpnI  
~~~~~  
~~~~~  
XhoI  
~~~~~

ECORI

FIG._36D

| | | | | | | | |
|------|------------|-------------|-------------|------------|------------|------------|------------|
| 4551 | TCTAGTAAA | TAAAGATAA | ACTTAGACTC | AAACATTTA | CAAAACAAC | CCCTAAAGTC | CTAAAGCCCA |
| 4621 | AAGTGCTATG | CACGATCCAT | AGCAAGCCCA | GCCCAACCCA | ACCCAACCCA | ACCCACCCCA | GTGCAGCCAA |
| 4691 | CTGGCAATA | GTCCTCACCC | CCGGCACTAT | CACCGTGAGT | TGTCGCACC | ACCGCACGTC | TCGCAGCCAA |
| 4761 | AAAAA | AGAAAGAAA | AAAGAAAAA | GAAACACAGC | AGGTGGGTCC | GGTCGTGGG | GGCCGGAAAA |
| 4831 | GCGAGGAGGA | TCGCGAGCAG | CGACGAGGCC | CGGCCCTCCC | TCCGCTTCCA | AAGAAACGCC | CCCCATCGCC |
| 4901 | ACTATATACA | TACCCCCCCC | TCTCCTCCCA | TCCCCCCAAC | CCTACCACCA | CCACCACCAC | CACCTCCTCC |
| 4971 | CCCCTCGCTG | CCGGACGACG | AGCTCCTCCC | CCCTCCCCCT | CCGCCGCCGC | CGGTAAACAC | CCCGCCCCTC |
| 5041 | TCCCTCTTCT | TTCCTCCGTTT | TTTTTTTCGT | CTCGGTCCTG | ATCTTTGGCC | TTGGTAGTTT | GGGTGGGCCG |
| 5111 | GAGCGGCTTC | GTCGCCCAGA | TCGGTGCGCG | GGAGGGGCGG | GATCTCGCGG | CTGGCGTCTC | CGGGCGTGAG |
| | | BamHI | | BglII | | | |
| | ~~~~~ | | | ~~~~~ | | | |
| 5181 | TCGGCCCGGA | TCCTCGCGGG | GAAATGGGGCT | CTCGGATGTA | GATCTTCTTT | CTTTCCTCTT | TTTGTGGTAG |
| 5251 | AATTGAATC | CCTCAGCATT | GTTCAATCGGT | AGTTTTCCTT | TTCATGATTT | GTGACAAATG | CAGCCTCGTG |
| 5321 | CGGAGCTTTT | TTGTAGC | | | | | |

FIG._36E

**FIG. 37A**

HindIII
~~~~~

M  
1  
71

AAGCTTAAACA  
• A S T  
CAGCCTCTAC

TGAAGCAGTT  
S E D L  
GCAAGGCATC

CTCCGCCAAA  
Y S R  
TCCGAAGACC

CACGTCCTCG  
L V E  
TCTACAGCCG

CAGTTGTGGT  
V V V  
TTTAGTCGAA

GACTGCAGGG  
T A G  
ATGGCCACTA

CACGCCCTTAG  
H A L A  
TCTCCCAAGC

PstI  
~~~~~

• A
141

Y A D L C N I P S T I I K G E K I Y N S Q T

TGCCTACGCC
GACCTGTGCA
ACATTCCGTC

GACCTATATC
GACTATTATC
AAGGGAGAGA

AAATTACAA
TTCTCAAAC

TTCTCAAAC

BamHI
~~~~~

D I N G W I L R D D S S K E I I T V F R G T G S

GACATTAACG  
• D T N L Q L D T N Y T L T P F D T L P Q C N G •

GTGATACGAA  
• C E V H G G Y Y I G W V S V Q D Q V E S L V K

TTGTGAAGTA  
Q Q V S Q Y P D Y A L T V T G H X L G A S L A A

CAGCAGGTTA  
• L T A A Q L S A T Y D N I R L Y T F G E P R S •

CACTCACTGC  
CGCCACGCTG  
TCTGCGACAT

ACGACAACAT  
CCGCTGTAC  
ACCTTCGGCG

AACCGCGCAG

XhoI  
~~~~~

• G N Q A F A S Y M N D A F Q A S S P D T T Q Y

CGGCAATCAG
GCCTTCGCGT
CGTACATGAA

CGATGCCCTTC
CAAGCCTCGA
GCCCAGATAC

GACGCAGTAT
NcoI
~~

• F R V T H A N D G I P N L P P V E Q G Y A H G G

631
• V E Y W S V D P Y S A Q N T F V C T G D E V Q •

GTGTAGAGTA
• C C E A Q G G Q G V N N A H T T Y F G M T S G

GTGCTGTGAG
GCCCAGGGCG
GACAGGGTGT

GAATAATGCG
CACACGACTT
ATTTGGGAT

GACGAGCGGC

FIG.-37B

FIG. 37C

| | | | | | | | |
|------|-------------|-------------|-------------|------------|-------------|-------------|------------|
| 2731 | TAGTTATCTA | CACGACGGGG | AGTCAGGCAA | CTATGGATGA | ACGAAATAGA | CAGATCGCTG | AGATAGGTGC |
| 2801 | CTCACTGATT | AAGCATTGGT | AACGTGTCAGA | CCAAGTTTAC | TCATATATAC | TTTAGATTGA | TTTAAAACTT |
| 2871 | CATTTTTAAT | TTAAAAGGAT | CTAGGTGAAG | ATCCTTTTGG | ATAATCTCAT | GACCAAAATC | CCTTAACGTG |
| 2941 | AGTTTTCGTT | CCACTGAGCG | TCAGACCCCG | TAGAAAAGAT | CAAAGGATCT | TCTTGAGATC | CTTTTTTTCT |
| 3011 | GCGCCTAATC | TGCTGCTTGC | AAACAAAAAA | ACCACCGCTA | CCAGCGGTGG | TTTGTTTGCC | GGATCAAGAG |
| 3081 | CTACCAACTC | TTTTCCTCGAA | GGTAACCTGGC | TTCAGCAGAG | CGCAGATACC | AAATACTGTC | CTTCTAGTGT |
| 3151 | AGCCGTAGTT | AGGCCACCAC | TTCAAGAACT | CTGTAGCAC | GCCTACATAC | CTCGCTCTGC | TAATCCTGTT |
| 3221 | ACCAGTGGCT | GCTGCCAGTG | GCGATAAGTC | GTGTCTTACC | GGGTTGGACT | CAAGACGATA | GTTACCGGAT |
| 3291 | AAGGCGCAGC | GGTCGGGCTG | AACGGGGGGT | TCGTGCACAC | AGCCCAGCTT | GGAGCGAAACG | ACCTACACCG |
| 3361 | AACTGAGATA | CCTACAGCGT | GAGCTATGAG | AAAGCGCCAC | GCTTCCCGAA | GGAGAGAAAGG | CGGACAGGTA |
| 3431 | TCCGGTAAGC | GGCAGGGTCG | GAAACAGGAG | GGCACGAGG | GAGCTTCCAG | GGGAAACGC | CTGGTATCTT |
| 3501 | TATAGTCCTG | TCGGGTTCG | CCACCTCTGA | CTTGAGCGTC | GATTTTGTG | ATGCTCGTCA | GGGGGGCGGA |
| 3571 | GCCTATGGAA | AAACGCCAGC | AACGGGCCT | TTTTACGGTT | CCTGGCCTTT | TGCTGGCCTT | TTGCTCACAT |
| 3641 | GTTCTTTCCCT | GCGTTATCCC | CTGATTCGT | GGATAACCGT | ATTACCGCCT | TTGAGTGAGC | TGATACCGCT |
| 3711 | CGCCGCAGCC | GAACGACCGA | CGCAGCGAG | TCAGTGAGCG | AGGAAGCGGA | AGAGCGCCCA | ATACGCAAAC |
| 3781 | CGCCTCTCCC | CGCGGTTGG | CCGATTCAAT | AAATGAGCTG | GCACGACAGG | TTTCCCGACT | GGAAGCGGG |
| 3851 | CAGTGAGCGC | AACGCAATTA | ATGTGAGTTA | GCTCACTCAT | TAGGCACCCC | AGGCTTTACA | CTTTATGCTT |
| 3921 | CCGGCTCGTA | TGTTGTGTGG | AATTGTGAGC | GGATAACAAT | TTCACACACAG | AAACAGCTAT | GACCATGATT |
| | | | | | KpnI | | XhoI |
| 3991 | ACGCCAAGCG | CGCAATTAA | CCTCACTAAA | GGGAACAAAA | GCTGGGTACC | GGGCCCCCCC | TCGAGGTCAT |
| 4061 | TCATATGCTT | GAGAAAGAG | TCGGGATAGT | CCAAAATAAA | ACAAAGGTAA | GATTACCTGG | TCAAAAGTGA |
| 4131 | AAACATCAGT | TAAAAGGTGG | TATAAGTAAA | ATATCGGTAA | TAAAAGGTGG | CCCAAAGTGA | AAATTACTCT |
| 4201 | TTTCTACTAT | TATAAAAATT | GAGGATGTTT | TGTCGGTACT | TTGATACGTC | ATTTTGTGAT | GAAATTGGTT |
| 4271 | TTAAGTTTAT | TCGCGATTGG | GAAATGCATA | TCGTATTTG | AGTCGGTTTT | TAAAGTTCGTT | GCTTTTGTA |
| 4341 | ATACAGAGGG | ATTTGTATA | GAAATATCTT | TAAAAAACCC | ATATGCTAAT | TTGACATAAT | TTTTTGAGAA |
| | | | | ECORI | | | |
| 4411 | AAATATATAT | CAGGCCGAAT | CCACAATGAA | CAATAATAAG | ATTAAAAATAG | CTTGCCCCCG | TTGCAGCGAT |
| 4481 | GGGTATTTTT | TCTAGTAAAA | TAAAAGATAA | ACTTAGACTC | AAAACATTTA | CAAAAACAAC | CCCTAAAGTC |
| 4551 | CTAAAGCCCA | AAGTGCTATG | CACGATCCAT | AGCAAGCCCA | GCCCAACCCA | ACCCAACCCA | ACCCACCCCA |
| 4621 | GTGCAGCCAA | CTGGCAAATA | GTCCTCACCC | CCGGCACTAT | CACCGTGAGT | TGTCGCGACC | ACCGCACGTC |
| 4691 | TCGCAGCCAA | AAAAAATAAA | AGAAAGAAAA | AAAAGAAAA | GAAAAACAGC | AGGTGGGTCC | GGGTGTTGGG |

FIG._37D

| | | | | | | | |
|------|-------------|------------|------------|-------------|------------|------------|-------------|
| 4761 | GGCCGGGAAA | GCGAGGAGGA | TCGCGAGCAG | CGACGAGGCC | CGGCCCTCCC | TCCGCTTCCA | AAGAAACGCC |
| 4831 | CCCCATCGCC | ACTATATACA | TACCCCCCCC | TCCTCTCCCA | TCCCCCAAC | CCTACCACCA | CCACCACCAC |
| 4901 | CACCTCCTCC | CCCCTCGCTG | CCGACGACG | AGCTCCTCCC | CCCTCCCCCT | CCGCCGCCGC | CGGTAACCCAC |
| 4971 | CCCGCCCCTC | TCCTCTTTCT | TTCTCCGTTT | TTTTTTTCGT | CTCGGTCTCG | ATCTTTGGCC | TTGGTAGTTT |
| 5041 | GGGTGGGCCG | GAGCGGCTTC | GTCGCCCAGA | TCGGTGCGCG | GGAGGGCGG | GATCTCGCGG | CTGGCGTCTC |
| | | BamHI | | | | BglII | |
| | | ~~~~~ | | | | ~~~~~ | |
| 5111 | CGGGCGTGAG | TCGGCCCGGA | TCCTCGCGGG | GAATGGGGCT | CTCGGATGTA | GATCTTCTTT | CTTTCCTTCTT |
| 5181 | TTTGTTGGTAG | AATTGGAATC | CCTCAGCATT | GTTCAATCGGT | AGTTTTTCTT | TTCATGATTT | GTGACAAATG |
| 5251 | CAGCCTCGTG | CGGAGCTTTT | TTGTAGC | | | | |

FIG._37E

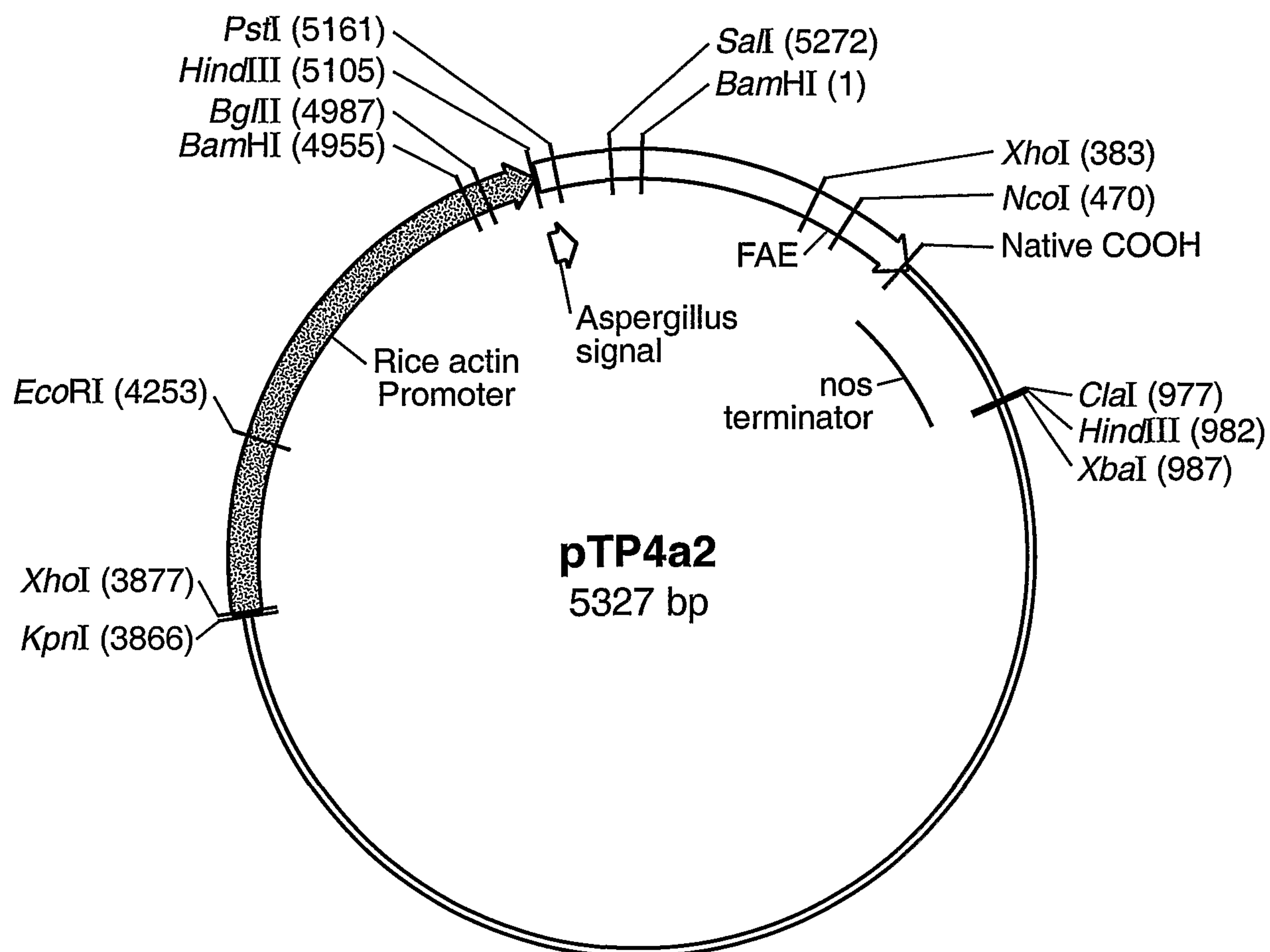
**FIG. 38A**

FIG. 38B

981 AAGCTTCTAG AGCGGCCGGT GGAGCTCCAA TTCGCCCTAT AGTGAGTCGT ATTACGCGCG CTCACTGGCC
1051 GTCGTTTAC AACGTCGTGA CTGGGAAAC CCTGGCGTTA CCAACTTAA CCGCTTGCA TCGCTCCCC
1121 CTTTCGCCAG CTGGCGTAAT AGCGAGAGG CCGCACCGA TCGCCCTTCC TCGAGTTGC CAACAGTTGC GAGCCTGAA
1191 TGGCGAATGG GACGCGCCCT TAGCGGCGC ATTAAGCGCG ATTAAGCGCG GCGGTGTGG GCGGTACGG CAGCGTGACC
1261 GCTACACTTG CCAGCGCCCT AGCGCCGCT CTTTCGCTT CTTCCCTTC TCTTCGCTC CTTTCGCG ACGTTCGCG
1331 GCTTCCCCG TCAAGCTCTA AATCGGGGC TCCCTTTAGG GTTCCGATTT AGTGCTTTAC AGTACCTCGA GGCACCTCGA
1401 CCCCAAAAA CTTGATTAGG GTGATGGTTC ACGTAGTGG CCAATCGCCCT CCATCGCCCT TTTTCGCCCT TTTTCGCCCT
1471 TTGACGTTGG AGTCCACGTT CTTTAATAGT GGACTCTTGT GGACTCTTGT TCCAACCTGG AACAACTC AACCTATCT
1541 CGGTCATTC TTTTGATTTA TAAGGGAATT TAAGGGAATT TGCCGATTTC GGCTATTGG TTAAAAAATG AGCTGATTTA
1611 ACAAAAAATT AACGCGAATT TTAACAAAAT ATTAACGCTT ACAATTTAGG TGGCACTTTT CGGGGAAATG CGGGGAAATG
1681 TCGCGGAAC CCTATTGT TTATTTTCT TAATACATTC AAATATGTAT AAATATGTAT CCGCTCATGA GACAATAACC
1751 CTGATAAATG CTTCAATAAT ATTGAAAAAG GAAGAGTATG AGTATTCAAC AGTATCCGT ATTTCCGTAT CGCCCTTATT
1821 CCTTTTGTG CGGCATTTTG CCTTCCCTGT CTTGCTCAC TTTGCTCAC ATTCTCAGAA ATTTCTGGTT TGACTTGGTT GAGTACTCAC
1891 AAGATCAGTT GGTGCACGA GTGGTTACA TCGAACTGGA TCTCAACAGC CAGAAACGCT CAGAAACGTA AAAGATGCTG TTGAGAGTTT
1961 TCGCCCCGAA GAACGTTTTC CAATGATGAG CACTTTTAAA GTTCTGCTAT GTTCTGGCT GTGGCGCGT ATATCCCGT
2031 ATTGACGCCG GGCAAGAGCA ACTCGTCCG CGCATACACT CGCATACAG AAATCTCAGAA ATTTCTGGTT TGACTTGGTT GAGTACTCAC
2101 CAGTCACAGA AAAGCATCTT ACGGATGGCA TGACAGTAAG AGAATTATGC AGAATTATGC AGTCTGCCA TAACCATGAG
2171 TGATAACACT GCGCCAAC TACTTCTGAC AACGATCGGA AACGATCGGA GGACCGAAG GGACCGAAG GAGTAAACC AGCTAACCG TTTTTCGCAC
2241 AACATGGGG ATCATGTAA TCGCCCTTGAT TAGCAATGG CAACAACGTT GCGCAAACTA TTAACCTGGC CAGGACCAT CAGGACCAT
2311 AGCGTGACAC CACGATGCC TACGATGCTG TAATAGACTG GATGAGGCG GATGAGGCG GATGAGGCG GATGAGGCG TCTGCGCTCG
2381 TCTAGCTTCC CCGCAACAAT TATTGCTGAT TATTGCTGAT AAATCTGGAG CCGGTGAGCG CCGGTGAGCG TGGGTCTCG GGTATCATG
2451 GCCCTTCCGG CAGCATGGT GCCAGATGGT AAGCCCTCCC GTATCCTAGT TATCTACACG ACGGGGAGTC ACGGGGAGTC AGGCAACTAT
2521 CAGCACTGGG GGATGAACGA AATAGACAGA TCCTGAGAT AGGTGCCCTCA CTGATTAGC CTGATTAGC ATTGGTAAC GTCAGACCAA
2591 GGATGAACGA AATAGACAGA TCCTGAGAT AGGTGCCCTCA CTGATTAGC CTGATTAGC ATTGGTAAC GTCAGACCAA
2661 GTTTACTCAT ATATACTTTA GATTGATTTA AAACCTCAT TTTAATTAA AAGGATCTAG AAGGATCTAG GTGAAGATCC
2731 TTTTGTGATA TCTCATGACC AAAATCCCTT AACGTGAGTT TTCGTTCCAC TGAGCGTCAG ACCCGTAGA ACCCGTAGA
2801 AAAGATCAAA GGATCTTCTT GAGATCCCTT TTTTCTGCGC GTAAATCTGCT GTAAATCTGCT GCTTGCAAAC AAAAAAACCA
2871 CCGCTACCAG CCGTGGTTTG TTTGCCGGAT CAAGAGCTAC CAACCTCTTT CCACTCTTT TCCGAAGGTA ACTGGCTTCA ACTGGCTTCA
2941 GCAGAGCGCA GATACCAAAT ACTGTCTTTC TAGTGTAGCC GTAGTTAGCC GTAGTTAGCC GTAGTTAGCC TAAGTCTGT TAAGTCTGT
3011 AGCACCGCCT ACATACCTCG CTCTGCTAAT CCGGTAGCC CCGGTAGCC GTGGCTGCTG GTGGCTGCTG CCAGTGGCGA TAAGTCTGT
3081 CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAG CCGAGCGGT CCGAGCGGT CCGAGCGGT GGGGTGAGC GGGGTGAGC
3151 GCACACAGCC CAGCTTGGAG CGAACGACCT ACACCGAAT GAGATACCTA GAGATACCTA GAGATACCTA CAGCGTGAGC TATGAGAAAG
3221 CGCCACGCTT CCCGAAGGGA GAAAGCGGA CAGGTATCCG GAAAGCGGA GAAAGCGGA GAAAGCGGA GAAAGCGGA AGGAGAGCGC
3291 ACGAGGGAGC TTCCAGGGG AAACGCTGG TATCTTTATA GTCCCTGTCGG GTCCCTGTCGG GTCCCTGTCGG GTCCCTGTCGG
3361 AGCGTCGATT TTTGTGATGC TCCTCAGGG GCGGAGCCT ATGGAAAAAC GCCAGCAAC GCCAGCAAC CGGCCTTT

FIG. 38C

3431 ACGGTTCCTG GCCTTTGGCT GGCCTTTTCG TCACATGTTT TTTCCCTGCGT TATCCCCCTGA TTCCTGTGGAT
3501 AACCGTATTA CCGCCTTTGA GTGAGCTGAT ACCGCTCGCC GCAGCCGAAC GACCGAGCGC AGCGAGTCAG
3571 TGAGCGAGGA AGCGGAAGAG CGCCCAATAC GCAAACCGCC TCTCCCGCGG CGTTGGCCGA TTCAATTAATG
3641 CAGCTGGCAC GACAGGTTTC CCGACTGGAA AGCGGGCAGT GAGCGCAACG CAAATTAATGT GAGTTAGCTC
3711 ACTCATTAGG CACCCAGGC TTACACTTT ATGCTTCCGG CTCGTATGTT GTGTGGAATT GTGAGCGGAT
3781 AACAAATTCA CACAGGAAC AGCTATGACC ATGATTACGC CAAGCGCGCA ATTAACCCCTC ACTAAAGGGA
KpnI XhoI
~~~~~ ~~~~~  
3851 ACAAAAGCTG GTACCGGC GTACATTCAT ATGCTTGAGA AGAGAGTCGG GATAGTCAA  
3921 AATAAAACAA AGGTAAGATT ACCTGGTCAA AAGTGAAAC ATCAGTTAAA AGGTGGTATA AGTAAATAT  
3991 CCGTAATAAA AGGTGGCCA AAGTGAAATT TACTCTTTC TACTATTATA AAAATTGAGG ATGTTTGTG  
4061 GGTACTTTGA TACGTCATTT TTGTATGAAT TGGTTTTAA GTTTATTCGC GATTGGAAA TGCATATCTG  
4131 TATTTGAGTC GGTTTTTAAG TTCGTTGCTT TTGTAAATAC AGAGGGAATT GTATAAGAAA TATCTTTTAA  
ECORI  
~~~~~  
4201 AAACCCATAT GCTAATTTGA CATAATTTT GAGAAAAATA TATAATCAGG CGAATTCAC AATGAACAAT
4271 AATAAGATTA AATAGCTTG CCCCCTTGC AGGATGGGT ATTTTCTTA GTAAATATAA AGATAAACTT
4341 AGACTCAAAA CATTTACAAA AACCAACCCT AAAGTCCTA AAAGTCCAGT AGCCCAAGT CCAATAGTCT CCACCCCGG
4411 AGCCAGCC ACACAACC AACCAACC ACCCAGTG AGCCAACTGG CAAATAGTCT CCACCCCGG
4481 CACTATCAC GTAGTTGTC CGACCAACC CACGTCGCG AGCCAAAAA AAAAAGAGAA AGAAAAAAA
4551 GAAAAAGAAA AACAGCAGGT GGTCCCGGT CTTCCAAAGA AACGCCCTC ATGACATACC CCCCCTCTC
4621 GAGGCCCGGC CTTCCCTCCG CCAACCCCTA CACCAACCAC CACCAACCCT TCCTGCCCGG ACGACGAGCT
4691 CTCCCATCCC CCTCCCTCCG CGCCGCCGGT AACCAACCCT TAGTTTGGGT GGGCGAGAGC GGCCTCGTCG CCCAGATCGG
4761 CCTCCCTCCCT TTTTCGTCCT GTCTCGATCT TTGGCCTTGG
BamHI
~~~~~  
4901 TGCGCGGGAG GGGCGGATC TCGCGGCTGG CGTCTCCGG CGTGAGTCGG CCCGGATCCT CGCGGGGAAT  
BglII  
~~~~~  
4971 GGGGCTCTCG GATGTAGATC TTCCTTCTTT CTCCTTTTG TGGTAGAATT TGAATCCCTC AGCATTTGTC
HindIII
~
5041 ATCGGTAGTT TTTCTTTTCA TGATTGTGA CAAATGCAGC CTCGTGCGGA GCTTTTGTGT AGCAAGCTTA

FIG._38D

PstI

~~~~~

M K Q F S A K H V L A V V T A G H A L A A S .

5111 ACATGAAGCA GTTCTCCGCC AAACACGTC TCGCAGTTGT GGTGACTGCA GGCACGCCT TAGCAGCCTC

. T Q G I S E D L Y S R L V E M A T I S Q A A Y

5181 TACGCAAGGC ATCTCCGAAG ACCTCTACAG CCGTTTAGTC GAAATGGCCA CTATCTCCCA AGCTGCCCTAC

SaI

~~~~~

A D L C N I P S T I I K G E K I Y N S Q T D I N

5251 GCCGACCTGT GCAACATTCC GTCGACTATT ATCAAGGGAG AGAAATTTA CAATTCTCAA ACTGACATTA

B

~

. G W

5321 ACGGATG

FIG._38E

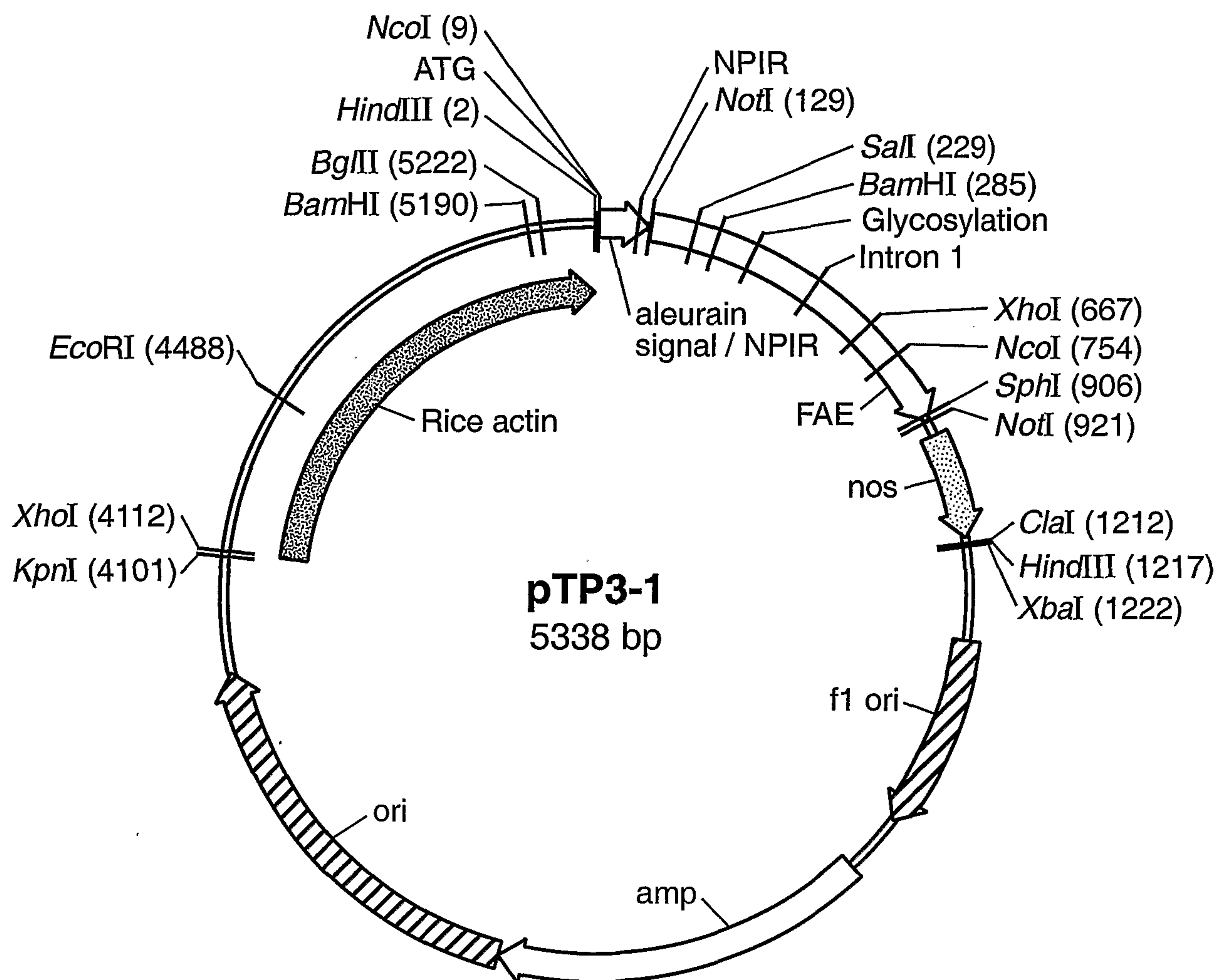
**FIG._39A**

FIG. 39B

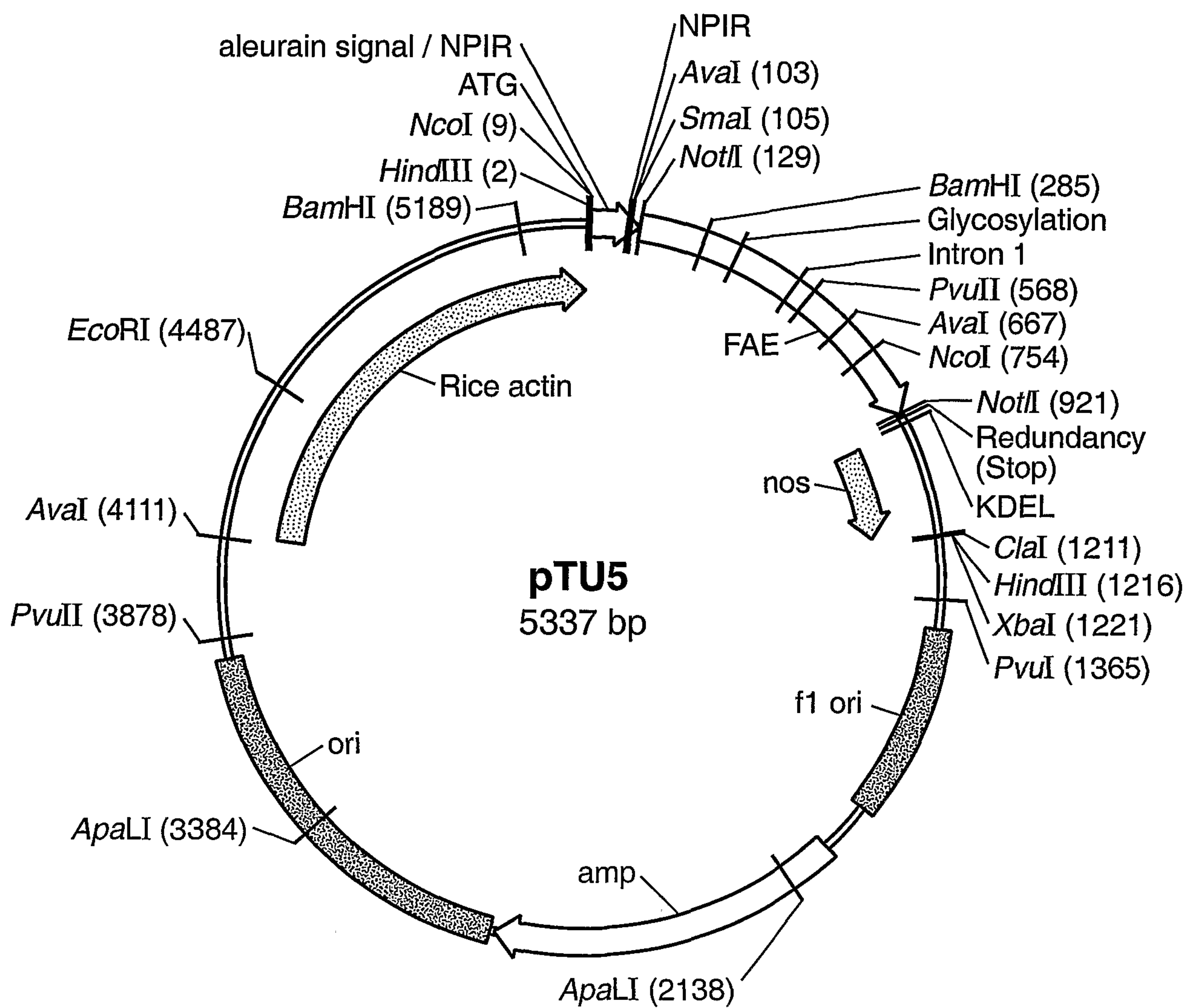
FIG. 39C

1471 TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCG CCGCTCCTTT CGCTTCTTC
1541 CTTTCCTTTC TCGCCACGTT CGCCGGCTTT CCCGTCAG GGTCTCCCT TTAGGGTTCC
1611 GATTAGTGC TTACGGCAC CTCGACCCA AAAAATTGA TTAGGGTAT TTAGGGTAT GGTTCACGTA GTGGCCATC
1681 GCCCTGATAG ACGGTTTTC CACTCAACC TATCTCGGC TATCTTTTG GTTGGAGTCC ACGTCTTTA ATAGTGGACT CTTGTTCCAA
1751 ACTGGAACAA AATGAGCTG AATTAACAA AATTTAAC GAATTTAAC AATTTGCCG GATTTGCCG ATTTCCGCC
1821 ATTGGTTAAA TTAGTGGG CTTTTCGGG AATGTGCG AAATGTGCG AATTTAAC GAATTTAAC AAAATATTAA CGCTTACAAT
1891 TTAGGTGGCA CATGAGACAA TAACCTGTAT AATGCTTCA ATAATATTGA AAAAGGAAGA GTATGAGTAT CATTCAATA
1961 TGTATCCGCT CGTGTGCCC AAGTAAAGA TGGTTGAGT TGCCTAGT CACGAGTGG TTACATCGAA CTGGATCTCA
2031 TCAACATTTC TCGTGTGCCC TATTTCCCTT TATTTCCGCA TTTTGCCTTC CTGTTTTCG CACCCAGAA TCACCCAGAA
2101 ACGCTGGTGA AAGTAAAGA TGGTTGAGT TGCCTAGT CACGAGTGG TTACATCGAA TTACATCGAA CTGGATCTCA
2171 ACAGCGGTAA GATCCTTGAG AGTTTTCGCC CCGAAGAAC CCGAAGAAC GAGCAACTCG ATGAGCACTT TTAAGTTCT
2241 GCTATGTGGC GCGGTATTAT CCCGTATTGA CCGGTGCG CCGGTGCG CCGGTGCG GAGCAACTCG ATGAGCACTT TTAAGTTCT
2311 CAGAAATGACT TGGTTGAGT TGCCTAGT CACCTAGT CACCTAGT CACCTAGT TGGCATGACA TGGCATGACA GTAAAGAAAT
2381 TATGCAGTGC TGCCTAGT TGCCTAGT TGCCTAGT TGCCTAGT TGCCTAGT TGCCTAGT TGCCTAGT TGCCTAGT TGCCTAGT
2451 GAAGGAGCTA ACCGCTTTT TGCCTAGT TGCCTAGT TGCCTAGT TGCCTAGT TGCCTAGT TGCCTAGT TGCCTAGT TGCCTAGT
2521 CTGAATGAAG CCAATACCAA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA
2591 AACTATTAACT TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA TGGCGAATA
2661 AGTTGCAGGA CCACTTCTGC CCACTTCTGC CCACTTCTGC CCACTTCTGC CCACTTCTGC CCACTTCTGC CCACTTCTGC CCACTTCTGC
2731 GAGCGTGGGT CTCGCGGTAT CTCGCGGTAT CTCGCGGTAT CTCGCGGTAT CTCGCGGTAT CTCGCGGTAT CTCGCGGTAT CTCGCGGTAT
2801 ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAAATAG AACGAAATAG AACGAAATAG AACGAAATAG AACGAAATAG AACGAAATAG
2871 TAAGCATTGG TAACGTGTCAG ACCAAGTTTA CTCATATATA CTCATATATA CTCATATATA CTCATATATA CTCATATATA CTCATATATA
2941 TTAAAGGA TCTAGGTGAA GATCCTTTT GATAATCTCA GATAATCTCA GATAATCTCA GATAATCTCA GATAATCTCA GATAATCTCA
3011 TCCACTGAGC GTCAGACCC GTCAGACCC GTCAGACCC GTCAGACCC GTCAGACCC GTCAGACCC GTCAGACCC GTCAGACCC
3081 CTGCTGCTTG CAAACAAAAG AACCACCGCT ACCAGCGGTG ACCAGCGGTG ACCAGCGGTG ACCAGCGGTG ACCAGCGGTG ACCAGCGGTG
3151 CTTTTCCTGA AGGTAACCTG CTTAGAGTAC CTTAGAGTAC CTTAGAGTAC CTTAGAGTAC CTTAGAGTAC CTTAGAGTAC CTTAGAGTAC
3221 TAGGCCACCA CTTCAAGAAC TCTGTAGTAC TCTGTAGTAC TCTGTAGTAC TCTGTAGTAC TCTGTAGTAC TCTGTAGTAC TCTGTAGTAC
3291 TGCTGCCAGT GCGGATAAGT CGTGTCTTAC CGTGTCTTAC CGTGTCTTAC CGTGTCTTAC CGTGTCTTAC CGTGTCTTAC CGTGTCTTAC
3361 CCGTCGGGCT GAACGGGGG TTCGTGCACA CAGCCAGCT CAGCCAGCT CAGCCAGCT CAGCCAGCT CAGCCAGCT CAGCCAGCT CAGCCAGCT
3431 ACCTACAGCG TGAGCTATGA GAAAGCGCA CAGCTTCCGA CAGCTTCCGA CAGCTTCCGA CAGCTTCCGA CAGCTTCCGA CAGCTTCCGA
3501 CGGCAGGGTC GGAACAGGAG AGCGACGAG AGCGACGAG AGCGACGAG AGCGACGAG AGCGACGAG AGCGACGAG AGCGACGAG
3571 GTCGGGTTTC GCCACCTCTG ACTTGAGCGT ACTTGAGCGT ACTTGAGCGT ACTTGAGCGT ACTTGAGCGT ACTTGAGCGT ACTTGAGCGT
3641 AAAACGCCAG CAACGGGCC TTTTACGGT TTTTACGGT TTTTACGGT TTTTACGGT TTTTACGGT TTTTACGGT TTTTACGGT
3711 TGCGTTATCC CCTGATTCTG TGGATAACCG TATTACCGC TATTACCGC TATTACCGC TATTACCGC TATTACCGC TATTACCGC
3781 CGAACGACCG AGCGAGCGA GTCAGTGAG GAGGAAGCG GAGGAAGCG GAGGAAGCG GAGGAAGCG GAGGAAGCG GAGGAAGCG
3851 CCGCGCGTTG GCCGATTCT TAAATGAGT GGCACGAC GGCACGAC GGCACGAC GGCACGAC GGCACGAC GGCACGAC GGCACGAC

FIG. 39D

3921 CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT
3991 ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC
KpnI
XhoI
~~~~~  
4061 GCGCAATTAA CCTCACTAA AGGGAACAAA AGCTGGGTAC CGGGCCCCC CTCGAGGTCA TTCATATGCT  
4131 TGAGAAAGAGA GTCGGGATAG TCCAAAATAA AACAAAGGTA AGATTACCTG GTCAAAAGTG AAAACATCAG  
4201 TTAAAAGGTG GTATAAGTAA AATATCGGTA ATAAAGGTG GCCAAAGTG AAATTACTC TTTTCTACTA  
4271 TTATAAAAAAT TGAGGATGTT TTGTCGGTAC TTTGATACGT CATTTTGTGTA TGAATTGGTT TTTAAGTTTA  
4341 TTCGCCGATTT GGAATGCAAT ATCTGTATTT GAGTCGGTTT TTAAGTTTCT TGCCTTTGTA AATACAGAGG  
4411 GATTGTGATA AGAAATATCT TTAAAAAACC CATATGCTAA TTTGACATAA TTTTGTGAGAA AAATATATAT  
ECORI  
~~~~~  
4481 TCAGGCGGAAT TCCACAATGA ACAATAATAA GATTAAAATA GCTTGCCCCC GTTGCAGCGA TGGGTATTTT
4551 TTCTAGTAAA ATAAAGATA AACTTAGACT CAAACATTT ACAAAACAA CCCCTAAAGT CTTAAAGCCC
4621 AAAGTGCTAT GCACGATCCA TAGCAAGCCC AGCCCAACCC AACCCACCC AACCCACCC AGTGCAGCCA
4691 ACTGGCAAAAT AGTCTCCACC CCCGGCACTA TCACCGTGAG TTGTCCGCAC TGTCCGACGT CTCGCAGCCA
4761 AAAAAAATAA AAGAAAGAAA AAAAGAAA AGAAACAG CAGGTGGGTG CGGTCTGTGG GGGCCGGAAA
4831 AGCGAGGAGG ATCGCGAGCA GCGACGAGG CCGGCCCTCC CTCGGCTTCC AAAGAAACGC CCCCATCGC
4901 CACTATATAC ATACCCCCC CTCTCTCC ATCCCCCAA CCTACCACC ACCACCAACA CCGGTAACA CCGGCCCTC
4971 CCCCCTCGCT GCCGGACGAC GAGCTCCTCC CCCCTCCCC TCCGCCGCCG CCGGTAACA CCGGCCCTC
5041 CTCCCTCTTC TTCTCTCCGT TTTTCTTCG TCTCGGTCTC GATCTTTGGC CTTGGTAGTT TGGGTGGCG
5111 AGAGCGGCTT CGTCGCCAG ATCGGTGCG ATCGTCGCG GGATCTCGG GCTGGCGTCT CCGGGCGTGA
BamHI
BgIII
~~~~~  
5181 GTCGGCCCCG ATCCTCGCG GGAATGGGC TCTCGGATGT AGATCTTCTT TCTTCTTCT TTTTGTGGTA  
5251 GAATTGGAAT CCTCAGCAT TGTTCATCG TAGTTTCTT TTTTCAATG TGTGACAAAT GCAGCCTCGT  
5321 GCGGAGCTTT TTTGTAGC

FIG.- 39E

**FIG. 40A**

HindIII NcoI  
~~~~~

1 AAGCTTACCA TGGCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT
TTCGAATGGT ACCGGGTGCG GGCGCAGGAG GAGGACCGCG AGCGGCACGA

51 GGCCACGGCC GCCGTCGCCG TCGCCTCCTC CTCCTCCTTC GCCGACTCCA
CCGGTGCCGG CGGCAGCGGC AGCGGAGGAG GAGGAGGAAG CGGCTGAGGT

SmaI
~~~~~

AvaI NotI  
~~~~~

101 ACCCGGGCCG GCCCGTCACC GACCGCGCGG CCGCCTCCAC GCAGGGCATC
TGGGCCCCGGC CGGGCAGTGG CTGGCGCGCC GGCGGAGGTG CGTCCCCTAG

151 TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC
AGGCTTCTGG AGATGTCGGC AAATCAGCTT TACCGGTGAT AGAGGGTTCG

201 TGCCTACGCC GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA
ACGGATGCGG CTGGACACGT TGTAAGGCAG CTGATAATAG TTCCCTCTCT

BamHI
~~~~~

251 AAATTTACAA TTCTCAAAC TACATTAACG GATGGATCCT CCGCGACGAC  
TTTAAATGTT AAGAGTTTGA CTGTAATTGC CTACCTAGGA GGCGCTGCTG

301 AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCACCTGGTA GTGATACGAA  
TCGTCGTTTC TTTATTAGTG GCAGAAGGCA CCGTGACCAT CACTATGCTT

351 TCTACAAC TC GATACTAACT ACACCCTCAC GCCTTTTCGAC ACCCTACCAC  
AGATGTTGAG CTATGATTGA TGTGGGAGTG CGGAAAGCTG TGGGATGGTG

401 AATGCAACGG TTGTGAAGTA CACGGTGGAT ATTATATTGG ATGGGTCTCC  
TTACGTTGCC AACACTTCAT GTGCCACCTA TAATATAACC TACCCAGAGG

451 GTCCAGGACC AAGTCGAGTC GCTTGTCAAA CAGCAGGTTA GCCAGTATCC  
CAGGTCCTGG TTCAGCTCAG CGAACAGTTT GTCGTCCAAT CGGTCATAGG

501 GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG  
CCTGATGCGC GACTGGCACT GGCCGGTGMG GGAGCCGCGG AGGGACCGCC

PvuII  
~~~~~

551 CACTCACTGC CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC
GTGAGTGACG GCGGGTCGAC AGACGCTGTA TGCTGTTGTA GGCGGACATG

601 ACCTTCGGCG AACCGCGCAG CGGCAATCAG GCCTTCGCGT CGTACATGAA
TGGAAGCCGC TTGGCGCGTC GCCGTTAGTC CGGAAGCGCA GCATGTACTT

AvaI
~~~~~

651 CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA  
GCTACGGAAG GTTCGGAGCT CGGGTCTATG CTGCGTCATA AAGGCCAGT

**FIG. 40B**

SUBSTITUTE SHEET (RULE 26)

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC  
GAGTACGGTT GCTGCCGTAG GGTTTGGACG GGGGCCACCT CGTCCCCATG

NcoI  
~~~~~

751 GCCCATGGCG GTGTAGAGTA CTGGAGCGTT GATCCTTACA GCGCCAGAA
CGGGTACCGC CACATCTCAT GACCTCGCAA CTAGGAATGT CGCGGGTCTT

801 CACATTTGTC TGCAC TGGGG ATGAAGTGCA GTGCTGTGAG GCCCAGGGCG
GTGTAAACAG ACGTGACCCC TACTTCACGT CACGACACTC CGGGTCCCGC

851 GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC
CTGTCCCACA CTTATTACGC GTGTGCTGAA TAAACCCCTA CTGCTCGCCG

NotI
~~~~~

901 GCATGCACCT GGCCGGTCGC GGCCGCGGAA CCACTGAAGG ATGAGCTGTA  
CGTACGTGGA CCGGCCAGCG CCGGCGCCTT GGTGACTTCC TACTCGACAT

951 AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT AAGATTGAAT  
TTCTTCGTCT AGCAAGTTTG TAAACCGTTA TTTCAAAGAA TTCTAACTTA

1001 CCTGTTGCCG GTCTTGCGAT GATTATCATA TAATTTCTGT TGAATTACGT  
GGACAACGGC CAGAACGCTA CTAATAGTAT ATTAAAGACA ACTTAATGCA

1051 TAAGCATGTA ATAATTAACA TGTAATGCAT GACGTTATTT ATGAGATGGG  
ATTCGTACAT TATTAATTGT ACATTACGTA CTGCAATAAA TACTCTACCC

1101 TTTTATGAT TAGAGTCCCG CAATTATACA TTTAATACGC GATAGAAAAC  
AAAATACTA ATCTCAGGGC GTTAATATGT AAATTATGCG CTATCTTTTG

1151 AAAATATAGC GCGCAAAC TA GGATAAATTA TCGCGCGCGG TGT CATCTAT  
TTTTATATCG CGCGTTTGAT CCTATTTAAT AGCGCGCGCC ACAGTAGATA

XbaI  
~~~~~

ClaI HindIII
~~~~~

1201 GTTACTAGAT CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTCGCC  
CAATGATCTA GCTATTTCGAA GATCTCGCCG GCCACCTCGA GGTTAAGCGG

1251 CTATAGTGAG TCGTATTACG CGCGCTCACT GGCCGTCGTT TTACAACGTC  
GATATCACTC AGCATAATGC GCGCGAGTGA CCGGCAGCAA AATGTTGCAG

1301 GTGACTGGGA AAACCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT  
CACTGACCCT TTTGGGACCG CAATGGGTTG AATTAGCGGA ACGTCGTGTA

PvuII  
~~~~~

1351 CCCCTTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC
GGGGGAAAGC GGTCGACCGC ATTATCGCTT CTCCGGGCGT GGCTAGCGGG

1401 TTCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCCTGTAGCG
AAGGGTTGTC AACGCGTCGG ACTTACCGCT TACCCTGCGC GGGACATCGC

FIG. 40C

SUBSTITUTE SHEET (RULE 26)

| | | | | | |
|----------------|---------------------------|---------------------------|---------------------------|---------------------------|----------------------------|
| 1451 | GCGCATTAAG
CGCGTAATTC | CGCGGCGGGT
GCGCCGCCCA | GTGGTGGTTA
CACCACCAAT | CGCGCAGCGT
GCGCGTCGCA | GACCGCTACA
CTGGCGATGT |
| 1501 | CTTGCCAGCG
GAACGGTCGC | CCCTAGCGCC
GGGATCGCGG | CGCTCCTTTC
GCGAGGAAAG | GCTTTCCTTC
CGAAAGAAGG | CTTCCTTTCT
GAAGGAAAGA |
| 1551 | CGCCACGTTC
GCGGTGCAAG | GCCGGCTTTC
CGGCCGAAAG | CCCGTCAAGC
GGGCAGTTCG | TCTAAATCGG
AGATTTAGCC | GGGCTCCCTT
CCCGAGGGAA |
| 1601 | TAGGGTTCCG
ATCCCAAGGC | ATTTAGTGCT
TAAATCACGA | TTACGGCACC
AATGCCGTGG | TCGACCCCAA
AGCTGGGGTT | AAAAC TTGAT
TTTTGA ACTA |
| 1651 | TAGGGTGATG
ATCCC ACTAC | G TTCACGTAG
CAAGTGCATC | TGGGCCATCG
ACCCGGTAGC | CCCTGATAGA
GGGACTATCT | CGGTTTTTTCG
GCCAAAAAGC |
| 1701 | CCCTTTGACG
GGGAAACTGC | TTGGAGTCCA
AACCTCAGGT | CGTTC TTTAA
GCAAGAAATT | TAGTGGACTC
ATCACCTGAG | TTGTTCCAAA
ACAAGGTTT |
| 1751 | CTGGAACAAC
GACCTTGTTG | ACTCAACCCT
TGAGTTGGGA | ATCTCGGTCT
TAGAGCCAGA | ATTCTTTTGA
TAAGAAA ACT | TTTATAAGGG
AAATATTCCC |
| 1801 | ATTTTGCCGA
TAAAACGGCT | TTTCGGCCTA
AAAGCCGGAT | TTGGTTAAAA
AACCAATTTT | AATGAGCTGA
TTACTCGACT | TTTAACAAAA
AAATTGTTTT |
| 1851 | ATTTAACGCG
TAAATTGCGC | AATTTTAACA
TTAAAATTGT | AAATATTAAC
TTTATAATTG | GCTTACAATT
CGAATGT TAA | TAGGTGGCAC
ATCCACCGTG |
| 1901 | TTTTCGGGGA
AAAAGCCCCT | AATGTGCGCG
TTACACGCGC | GAACCCCTAT
CTTGGGGATA | TTGTTTATTT
AACAAATAAA | TTCTAAATAC
AAGATTTATG |
| 1951 | ATTCAAATAT
TAAGTTTATA | GTATCCGCTC
CATAGGCGAG | ATGAGACAAT
TACTCTGT TA | AACCCTGATA
TTGGGACTAT | AATGCTTCAA
TTACGAAGTT |
| 2001 | TAATATTGAA
ATTATAACTT | AAAGGAAGAG
TTTCCTTCTC | TATGAGTATT
ATACTCATAA | CAACATTTCC
GTTGTAAAGG | GTGTCGCCCT
CACAGCGGGA |
| 2051 | TATTCCCTTT
ATAAGGGAAA | TTTGCGGCAT
AAACGCCGTA | TTTGCCTTCC
AAACGGAAGG | TGTTTTTTGCT
ACAAAAACGA | CACCCAGAAA
GTGGGTCTTT |
| ApaLI
~~~~~ | | | | | |
| 2101 | CGCTGGTGAA
GCGACCACTT | AGTAAAAGAT
TCATTTTCTA | GCTGAAGATC
CGACTTCTAG | AGTTGGGTGC
TCAACCCACG | ACGAGTGGGT
TGCTCACCCA |
| 2151 | TACATCGAAC
ATGTAGCTTG | TGGATCTCAA
ACCTAGAGTT | CAGCGGTAAG
GTCGCCATTC | ATCCTTGAGA
TAGGAACTCT | GTTTTTCGCCC
CAAAAGCGGG |
| 2201 | CGAAGAACGT
GCTTCTTGCA | TTTCCAATGA
AAAGGTTACT | TGAGCACTTT
ACTCGTGAAA | TAAAGTTCTG
ATTTCAAGAC | CTATGTGGCG
GATACACCGC |
| 2251 | CGGTATTATC
GCCATAATAG | CCGTATTGAC
GGCATAACTG | GCCGGGCAAG
CGGCCCGTTC | AGCAACTCGG
TCGTTGAGCC | TCGCCGCATA
AGCGGCGTAT |
| 2301 | CACTATTCTC
GTGATAAGAG | AGAATGACTT
TCTTACTGAA | GGTTGAGTAC
CCA ACTCATG | TCACCAGTCA
AGTGGTCAGT | CAGAAAAGCA
GTCTTTTCGT |

FIG. 40D

SUBSTITUTE SHEET (RULE 26)

| | | | | | |
|------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| 2351 | TCTTACGGAT
AGAATGCCTA | GGCATGACAG
CCGTACTGTC | TAAGAGAATT
ATTCTCTTAA | ATGCAGTGCT
TACGTCACGA | GCCATAACCA
CGGTATTGGT |
| 2401 | TGAGTGATAA
ACTCACTATT | CACTGCGGCC
GTGACGCCGG | AACTTACTTC
TTGAATGAAG | TGACAACGAT
ACTGTTGCTA | CGGAGGACCG
GCCTCCTGGC |
| 2451 | AAGGAGCTAA
TTCCTCGATT | CCGCTTTTTT
GGCGAAAAAA | GCACAACATG
CGTGTTGTAC | GGGGATCATG
CCCCTAGTAC | TAACTCGCCT
ATTGAGCGGA |
| 2501 | TGATCGTTGG
ACTAGCAACC | GAACCGGAGC
CTTGGCCTCG | TGAATGAAGC
ACTTACTTCG | CATACCAAAC
GTATGGTTTG | GACGAGCGTG
CTGCTCGCAC |
| 2551 | ACACCACGAT
TGTGGTGCTA | GCCTGTAGCA
CGGACATCGT | ATGGCAACAA
TACCGTTGTT | CGTTGCGCAA
GCAACGCGTT | ACTATTA ACT
TGATAATTGA |
| 2601 | GGCGAACTAC
CCGCTTGATG | TTACTCTAGC
AATGAGATCG | TTCCCGGCAA
AAGGGCCGTT | CAATTAATAG
GTTAATTATC | ACTGGATGGA
TGACCTACCT |
| 2651 | GGCGGATAAA
CCGCCTATTT | GTTGCAGGAC
CAACGTCCTG | CACTTCTGCG
GTGAAGACGC | CTCGGCCCTT
GAGCCGGGAA | CCGGCTGGCT
GGCCGACCGA |
| 2701 | GGTTTATTGC
CCAAATAACG | TGATAAATCT
ACTATTTAGA | GGAGCCGGTG
CCTCGGCCAC | AGCGTGGGTC
TCGCACCCAG | TCGCGGTATC
AGCGCCATAG |
| 2751 | ATTGCAGCAC
TAACGTCGTG | TGGGGCCAGA
ACCCCGGTCT | TGGTAAGCCC
ACCATTCGGG | TCCCGTATCG
AGGGCATAGC | TAGTTATCTA
ATCAATAGAT |
| 2801 | CACGACGGGG
GTGCTGCCCC | AGTCAGGCAA
TCAGTCCGTT | CTATGGATGA
GATACCTACT | ACGAAATAGA
TGCTTTTATCT | CAGATCGCTG
GTCTAGCGAC |
| 2851 | AGATAGGTGC
TCTATCCACG | CTCACTGATT
GAGTGACTAA | AAGCATTTGGT
TTCGTAACCA | AACTGTCAGA
TTGACAGTCT | CCAAGTTTAC
GGTTCAAATG |
| 2901 | TCATATATAC
AGTATATATG | TTTAGATTGA
AAATCTAACT | TTTAAAACTT
AAATTTTGAA | CATTTTTTAAT
GTAAAAATTA | TTAAAAGGAT
AATTTTCCTA |
| 2951 | CTAGGTGAAG
GATCCACTTC | ATCCTTTTTG
TAGGAAAAAC | ATAATCTCAT
TATTAGAGTA | GACCAAATC
CTGGTTTTAG | CCTTAACGTG
GGAATTGCAC |
| 3001 | AGTTTTTCGTT
TCAAAAGCAA | CCACTGAGCG
GGTGACTCGC | TCAGACCCCG
AGTCTGGGGC | TAGAAAAGAT
ATCTTTTCTA | CAAAGGATCT
GTTTCCTAGA |
| 3051 | TCTTGAGATC
AGA ACTCTAG | CTTTTTTTTCT
GAAAAAAAGA | GCGCGTAATC
CGCGCATTAG | TGCTGCTTGC
ACGACGAACG | AAACAAAAAA
TTTGTTTTTT |
| 3101 | ACCACCGCTA
TGGTGGCGAT | CCAGCGGTGG
GGTCGCCACC | TTTGTTTGCC
AAACAAACGG | GGATCAAGAG
CCTAGTTCTC | CTACCAACTC
GATGGTTGAG |
| 3151 | TTTTTCCGAA
AAAAAGGCTT | GGTAACTGGC
CCATTGACCG | TTCAGCAGAG
AAGTCGTCTC | CGCAGATACC
GCGTCTATGG | AAATACTGTC
TTTATGACAG |
| 3201 | CTTCTAGTGT
GAAGATCACA | AGCCGTAGTT
TCGGCATCAA | AGGCCACCAC
TCCGGTGGTG | TTCAAGAACT
AAGTTCTTGA | CTGTAGCACC
GACATCGTGG |
| 3251 | GCCTACATAC
CGGATGTATG | CTCGCTCTGC
GAGCGAGACG | TAATCCTGTT
ATTAGGACAA | ACCAGTGGCT
TGGTCACCGA | GCTGCCAGTG
CGACGGTCAC |

FIG. 40E

SUBSTITUTE SHEET (RULE 26)

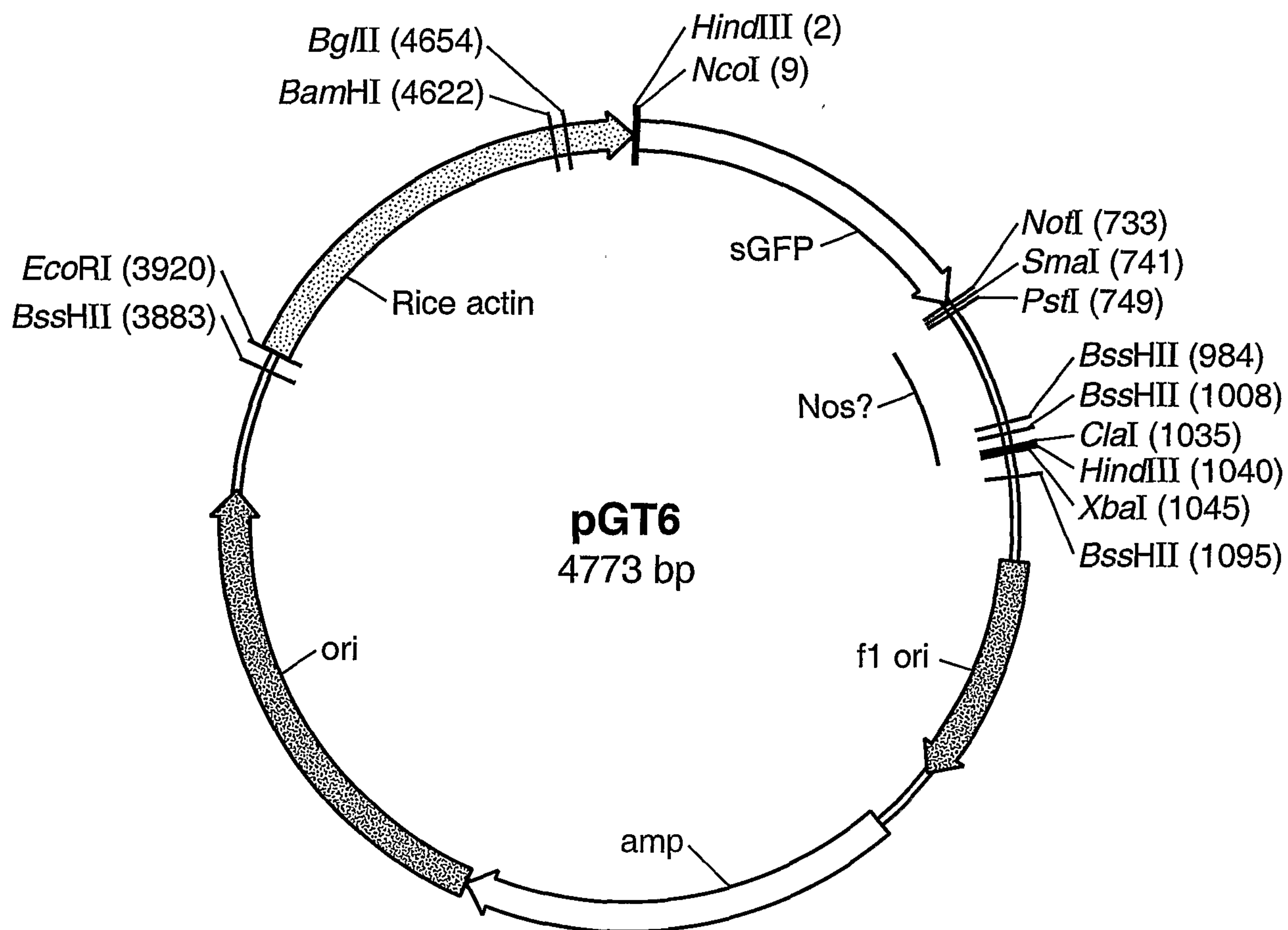
| | | | | | |
|------|------------|------------|------------|------------|-------------|
| 3301 | GCGATAAGTC | GTGTCTTACC | GGGTTGGACT | CAAGACGATA | GTTACCGGAT |
| | CGCTATTCAG | CACAGAATGG | CCCAACCTGA | GTTCTGCTAT | CAATGGCCTA |
| | | | | ApaLI | |
| | | | | ~~~~~ | |
| 3351 | AAGGCGCAGC | GGTCGGGCTG | AACGGGGGGT | TCGTGCACAC | AGCCCAGCTT |
| | TTCCGCGTCG | CCAGCCCGAC | TTGCCCCCCA | AGCACGTGTG | TCGGGTCGAA |
| 3401 | GGAGCGAACG | ACCTACACCG | AACTGAGATA | CCTACAGCGT | GAGCTATGAG |
| | CCTCGCTTGC | TGGATGTGGC | TTGACTCTAT | GGATGTCGCA | CTCGATACTC |
| 3451 | AAAGCGCCAC | GCTTCCCGAA | GGGAGAAAGG | CGGACAGGTA | TCCGGTAAGC |
| | TTTCGCGGTG | CGAAGGGCTT | CCCTCTTTCC | GCCTGTCCAT | AGGCCATTCT |
| 3501 | GGCAGGGTCG | GAACAGGAGA | GCGCACGAGG | GAGCTTCCAG | GGGGAAACGC |
| | CCGTCCCAGC | CTTGTCTCTT | CGCGTGCTCC | CTCGAAGGTC | CCCCTTTGCG |
| 3551 | CTGGTATCTT | TATAGTCCTG | TCGGGTTCG | CCACCTCTGA | CTTGAGCGTC |
| | GACCATAGAA | ATATCAGGAC | AGCCCAAAGC | GGTGGAGACT | GAACTCGCAG |
| 3601 | GATTTTGTG | ATGCTCGTCA | GGGGGGCGGA | GCCTATGGAA | AAACGCCAGC |
| | CTAAAAACAC | TACGAGCAGT | CCCCCGCCT | CGGATACCTT | TTTGCGGTCTG |
| 3651 | AACGCGGCCT | TTTTACGGTT | CCTGGCCTTT | TGCTGGCCTT | TTGCTCACAT |
| | TTGCGCCGGA | AAAATGCCAA | GGACCGGAAA | ACGACCGGAA | AACGAGTGTA |
| 3701 | GTTCTTTCCT | GCGTTATCCC | CTGATTCTGT | GGATAACCGT | ATTACCGCCT |
| | CAAGAAAGGA | CGCAATAGGG | GACTAAGACA | CCTATTGGCA | TAATGGCGGA |
| 3751 | TTGAGTGAGC | TGATACCGCT | CGCCGCAGCC | GAACGACCGA | GCGCAGCGAG |
| | AACTCACTCG | ACTATGGCGA | GCGGCGTCGG | CTTGCTGGCT | CGCGTCGCTC |
| 3801 | TCAGTGAGCG | AGGAAGCGGA | AGAGCGCCCA | ATACGCAAAC | CGCCTCTCCC |
| | AGTCACTCGC | TCCTTCGCCT | TCTCGCGGGT | TATGCGTTTG | GCGGAGAGGG |
| | | | PvuII | | |
| | | | ~~~~~ | | |
| 3851 | CGCGCGTTGG | CCGATTCATT | AATGCAGCTG | GCACGACAGG | TTTCCCGACT |
| | GCGCGCAACC | GGCTAAGTAA | TTACGTCGAC | CGTGCTGTCC | AAAGGGCTGA |
| 3901 | GGAAAGCGGG | CAGTGAGCGC | AACGCAATTA | ATGTGAGTTA | GCTCACTCAT |
| | CCTTTCGCCC | GTCACCTCGC | TTGCGTTAAT | TACACTCAAT | CGAGTGAGTA |
| 3951 | TAGGCACCCC | AGGCTTTACA | CTTTATGCTT | CCGGCTCGTA | TGTTGTGTGG |
| | ATCCGTGGGG | TCCGAAATGT | GAAATACGAA | GGCCGAGCAT | ACAACACACC |
| 4001 | AATTGTGAGC | GGATAACAAT | TTCACACAGG | AAACAGCTAT | GACCATGATT |
| | TTAACACTCG | CCTATTGTTA | AAGTGTGTCC | TTTGTGCGTA | CTGGTACTAA |
| 4051 | ACGCCAAGCG | CGCAATTAAC | CCTCACTAAA | GGGAACAAAA | GCTGGGTACC |
| | TGCGGTTCGC | GCGTTAATTG | GGAGTGATTT | CCCTTGTTTT | CGACCCATGG |
| | | | AvaI | | |
| | | | ~~~~~ | | |
| 4101 | GGGCCCCCCC | TCGAGGTCAT | TCATATGCTT | GAGAAGAGAG | TCGGGATAGT |
| | CCCGGGGGGG | AGCTCCAGTA | AGTATACGAA | CTCTTCTCTC | AGCCCTATCA |

FIG. 40F

SUBSTITUTE SHEET (RULE 26)

| | | | | | |
|------|------------|------------|------------|-------------|-------------|
| 4151 | CCAAAATAAA | ACAAAGGTAA | GATTACCTGG | TCAAAAGTGA | AAACATCAGT |
| | GGTTTTATTT | TGTTTCCATT | CTAATGGACC | AGTTTTTCACT | TTTGTAGTCA |
| 4201 | TAAAAGGTGG | TATAAGTAAA | ATATCGGTAA | TAAAAGGTGG | CCCAAAGTGA |
| | ATTTTCCACC | ATATTCATTT | TATAGCCATT | ATTTTCCACC | GGGTTTCACT |
| 4251 | AATTTACTCT | TTTCTACTAT | TATAAAAATT | GAGGATGTTT | TGTCGGTACT |
| | TTAAATGAGA | AAAGATGATA | ATATTTTTAA | CTCCTACAAA | ACAGCCATGA |
| 4301 | TTGATACGTC | ATTTTTGTAT | GAATTGGTTT | TTAAGTTTAT | TCGCGATTTG |
| | AACTATGCAG | TAAAAACATA | CTTAACCAAA | AATTCAAATA | AGCGCTAAAC |
| 4351 | GAAATGCATA | TCTGTATTTG | AGTCGGTTTT | TAAGTTCGTT | GCTTTTGTA |
| | CTTTACGTAT | AGACATAAAC | TCAGCCAAAA | ATTCAAGCAA | CGAAAACATT |
| 4401 | ATACAGAGGG | ATTTGTATAA | GAAATATCTT | TAAAAAACCC | ATATGCTAAT |
| | TATGTCTCCC | TAAACATATT | CTTTATAGAA | ATTTTTTGGG | TATACGATTA |
| | | | | EcoRI | |
| | | | | ~~~~~ | |
| 4451 | TTGACATAAT | TTTTGAGAAA | AATATATATT | CAGGCGAATT | CCACAATGAA |
| | AACTGTATTA | AAAACCTCTT | TTATATATAA | GTCCGCTTAA | GGTGTTACTT |
| 4501 | CAATAATAAG | ATTAAAATAG | CTTGCCCCCG | TTGCAGCGAT | GGGTATTTTT |
| | GTTATTATTC | TAATTTTATC | GAACGGGGGC | AACGTCGCTA | CCCATAAAAA |
| 4551 | TCTAGTAAAA | TAAAAGATAA | ACTTAGACTC | AAAACATTTA | CAAAAACAAC |
| | AGATCATTTT | ATTTTCTATT | TGAATCTGAG | TTTTGTAAAT | GTTTTTGTTG |
| 4601 | CCCTAAAGTC | CTAAAGCCCA | AAGTGCTATG | CACGATCCAT | AGCAAGCCCA |
| | GGGATTTTCA | GATTTTCGGT | TTCACGATAC | GTGCTAGGTA | TCGTTTCGGT |
| 4651 | GCCCAACCCA | ACCCAACCCA | ACCCACCCCA | GTGCAGCCAA | CTGGCAAATA |
| | CGGGTTGGGT | TGGGTGGGGT | TGGGTGGGGT | CACGTCGGTT | GACCGTTTAT |
| 4701 | GTCTCCACCC | CCGGCACTAT | CACCGTGAGT | TGTCCGCACC | ACCGCACGTC |
| | CAGAGGTGGG | GGCCGTGATA | GTGGCACTCA | ACAGGCGTGG | TGGCGTGCAG |
| 4751 | TCGCAGCCAA | AAAAAAAAAA | AGAAAGAAAA | AAAAGAAAAA | GAAAAACAGC |
| | AGCGTCGGTT | TTTTTTTTTT | TCTTTCTTTT | TTTTCTTTTT | CTTTTTGTCTG |
| 4801 | AGGTGGGTCC | GGGTCGTGGG | GGCCGGAAAA | GCGAGGAGGA | TCGCGAGCAG |
| | TCCACCCAGG | CCCAGCACCC | CCGGCCTTTT | CGCTCCTCCT | AGCGCTCGTC |
| 4851 | CGACGAGGCC | CGGCCCTCCC | TCCGCTTCCA | AAGAAACGCC | CCCCATCGCC |
| | GCTGCTCCGG | GCCGGGAGGG | AGGCGAAGGT | TTCTTTGCGG | GGGGTAGCGG |
| 4901 | ACTATATACA | TACCCCCCCC | TCTCCTCCCA | TCCCCCCAAC | CCTACCACCA |
| | TGATATATGT | ATGGGGGGGG | AGAGGAGGGT | AGGGGGGGTT | GGATGGTGGT |
| 4951 | CCACCACCAC | CACCTCCTCC | CCCCTCGCTG | CCGGACGACG | AGCTCCTCCC |
| | GGTGGTGGTG | GTGGAGGAGG | GGGGAGCGAC | GGCCTGCTGC | TCGAGGAGGG |
| 5001 | CCCTCCCCCT | CCGCCGCCGC | CGGTAACCAC | CCCGCCCCCT | TCCTCTTTCT |
| | GGGAGGGGGA | GGCGGCGGCG | GCCATTGGTG | GGGCGGGGAG | AGGAGAAAGA |

FIG. 40G

**FIG. 41A**

HindIII NcoI
~~~~~ ~~~~~  
1   AAGCTTACCA TGGTGAGCAA GGGCGAGGAG CTGTTCACCG GGGTGGTGCC CATCTGGTC GAGCTGGACG  
     TTCGGAATGGT ACCACTCGTT CCCGCTCCTC GACAAGTGGC CCACACACGG GTAGGACCAG CTCGACCTGC  
  
71   GCGACGTGAA CGGCCACAAG TTCAGCGTGT CCGCGAGGG CGAGGGCGAT GCCACCTACG GCAAGCTGAC  
     CGCTGCACCTT GCCGGTGTTT AAGTCGCACA GGCCGCTCCC GCTCCCCTA CGGTGGATGC CGTTCGACTG  
  
141   CCTGAAAGTTC ATCTGCACCA CCGGCAAGCT GCCCGTGCC TGGCCACCC TCGTGACCAC CTTCACCTAC  
     GGACTTCAAG TAGACGTGGT GGCCGTTTCA CCGGCACGG ACCGGTGGG AGCACTGGTG GAAGTGGATG  
  
211   GGCGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCACGACTT CTTCAAGTCC GCCATGCCCG  
     CCGCACGTCA CGAAGTCGGC GATGGGGCTG GTGTACTTCG TCGTGCTGAA GAA GTTCAGG CCGTACGGGC  
  
281   AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CCGCAACTAC AAGACCCGCG CCGAGGTGAA  
     TTCCGATGCA GGTCCTCGCG TGGTAGAAGA AGTTCCTGCT GCCGTTGATG TTTCTGGGCGC GGCTCCACTT  
  
351   GTTTCGAGGGC GACACCCCTGG TGAACCGCAT CGAGCTGAAG GGCAATCGACT TCAAGGAGGA CGGCAACATC  
     CAAGCTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA AGTTCCTCCT GCCGTTGTAG  
  
421   CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT GGCCGACAAG CAGAAGAACG  
     GACCCCGTGT TCGACCTCAT GTTGATGTTG TCGGTGTTGC AGATATAGTA CCGGCTGTTC GTCTTCTTGC  
  
491   GCATCAAGGT GAAC TTCAAG ATCCGCCACA ACATCGAGGA CCGCAGCGTG CAGCTCGCCG ACCACTACCA  
     CGTAGTTCCA CTGAAGTTC TAGGCGGTGT TGTAGCTCCT GCCGTCGCAC GTCGAGCGGC TGGTGATGGT  
  
561   GCAGAACACC CCCATCGGGC ACGGCCCCGT GTGCTGCCC GACAACCACT ACCTGAGCAC CCAGTCCGCC  
     CGTCTTGTG GGGTAGCCGC TGCCGGGGCA CGACGACGGG CTGTTGGTGA TGGACTCGTG GGT CAGGCGG  
  
631   CTGAGCAAAG ACCCAACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT CGTGACCGCC GCCGGATCA  
     GACTCGTTTC TGGGTTGCT CTTTCGCGCTA GTGTACCAGG ACGACCTCAA GCACTGGCGG CGGCCCTAGT

FIG.\_41B

| SmaI   |            |             |            |              |             |             |             |
|--------|------------|-------------|------------|--------------|-------------|-------------|-------------|
| ~~~~~  |            |             |            |              |             |             |             |
| NotI   |            | PstI        |            |              |             |             |             |
| ~~~~~  |            | ~~~~~       |            |              |             |             |             |
| 701    | CTCACGGCAT | GGACGAGCTG  | TACAAGTAAA | GCGGCCGCC    | GGGCTGCAGG  | GAAACCACTG  | AAGGATGAGC  |
|        | GAGTGCCGTA | CCTGCTCGAC  | ATGTTCAATT | CGCCGGCGGG   | CCCGACGTCC  | CTTTGGTGAC  | TTCCCTACTCG |
| 771    | TGTAAGAAG  | CAGATCGTTC  | AAACATTTGG | CAATAAAGTT   | TCTTAAGATT  | GAATCCTGTT  | GCCGGTCTTG  |
|        | ACATTCTTC  | GCTAGCAAG   | TTTGTAACC  | GTTATTCAA    | AGAAATCTAA  | CTTAGGACAA  | CGGCCAGAAC  |
| 841    | CGATGATTAT | CATATAATT   | CTGTTGAATT | ACGTTAAGCA   | TGTAATAATT  | AACATGTAAT  | GCATGACGTT  |
|        | GCTACTAATA | GTATATTAA   | GACAACTTAA | TGCAATTCTG   | ACATTATTAA  | TTGTACATTA  | CGTACTGCAA  |
| 911    | ATTATGAGA  | TGGGTTTTTA  | TGATTAGAGT | CCGCAATTA    | TACATTTAAT  | ACGCGATAGA  | AAACAAAATA  |
|        | TAAATACTCT | ACCCAAAAAT  | ACTAATCTCA | GGCGTTAAT    | ATGTAAATTA  | TGCGCTATCT  | TTTGTTTAT   |
|        |            |             |            | XbaI         |             |             |             |
|        |            |             |            | ~~~~~        |             |             |             |
| BssHII |            | BssHII      |            | ClaI HindIII |             |             |             |
| ~~~~~  |            | ~~~~~       |            | ~~~~~        |             |             |             |
| 981    | TAGCGCGCAA | ACTAGGATAA  | ATTATCGCGC | GCGGTGTCAT   | CTATGTTACT  | AGATCGATAA  | GCTTCTAGAG  |
|        | ATCGCGCGTT | TGATCCCTATT | TAATAGCGCG | CGCCACAGTA   | GATACAAATGA | TCTAGCTATT  | CGAAGATCTC  |
|        |            |             |            | BssHII       |             |             |             |
|        |            |             |            | ~~~~~        |             |             |             |
| 1051   | CGGCCGGTGG | AGCTCCAATT  | CGCCCTATAG | TGAGTCGTAT   | TACGCGCGCT  | CAC'TGGCCGT | CGT'TTACAA  |
|        | GCCGGCCACC | TCGAGGTTAA  | GCGGGATATC | ACTCAGCATA   | ATGCGCGCGA  | GTGACCGGCA  | GCAAAATGTT  |
| 1121   | CGTCGTGACT | GGGAAAACCC  | TGGCGTTACC | CAACTTAATC   | GCCTTGCAGC  | ACATCCCCCT  | TTCGCCAGCT  |
|        | GCAGCACTGA | CCCTTTTGGG  | ACCGCAATGG | GTTGAAATTAG  | CGGAACGTCG  | TGTAGGGGGA  | AAGCGGTCGA  |
| 1191   | GGCGTAATAG | CGAAGAGGCC  | CGCACCGATC | GCCCTTCCCA   | ACAGTTGCGC  | AGCCTGAATG  | GCGAATGGGA  |
|        | CCGCATTATC | GCTTCTCCGG  | GCGTGGCTAG | CGGGAAGGGT   | TGTCAACGCG  | TCGGACTTAC  | CGCTTACCC   |
| 1261   | CGCGCCCTGT | AGCGGCGCAT  | TAAGCGCGGC | GGGTGTGGTG   | GTTACGCGCA  | GCGTGACCGC  | TACACTTGCC  |
|        | GCGCGGGACA | TCGCCGCGTA  | ATTGCGGCCG | CCACACCCAC   | CAATGCGCGT  | CGCACTGGCG  | ATGTGAACGG  |

FIG. 41C

1331 AGCGCCCTAG CGCCGCTCC TTTCGCTTC TTCCCTTCCT TTCTCGCCAC GTTCGCCGGC TTTCCCCGTC  
TCGCGGGATC GCGGCGAGG AAGCGAAG AAGGGAAGGA AAGAGCGGTG CAAGCGCCG AAAGGGCAG

1401 AAGCTCTAAA TCGGGGGCTC CCTTTAGGGT TCCGATTAG TGCTTTACGG CACCTCGACC CCAAAAAACT  
TTCGAGATT T AGCCCCGAG GGAATCCCA AGGCTAAATC ACGAAATGCC GTGGAGCTGG GGT'TTTTGA

1471 TGATTAGGGT GATGGTTCAC GTAGTGGCC ATCGCCCTGA TAGACGGTTT TTCGCCCTTT GACGTTGGAG  
ACTAATCCCA CTACCAAGTG CATCACCCG TAGCGGACT ATCTGCCAAA AAGCGGAAA CTGCAACCTC

1541 TCCACGTTCT TTAATAGTG ACTCTTGTC CAAACTGGAA CAACACTCAA CCTATCTCG GTCATTCCT  
AGGTGCAAGA AATTATCAC TGAGAACAA GTTTGACCTT GTTGAGATT GGGATAGAG CAGATAAGAA

1611 TTGATTTATA AGGATTTTG CCGATTTCG CCTATTGGT AAAAATGAG CTGATTTAAC AAAAATTAA  
AACTAAATAT TCCCTAAAC GGTAAAGCC GGATAACCAA TTTT'TTACTC GACTAAATTG TTTT'TAAAT

1681 CGCGAATTT AACAAAATAT TAACGCTTAC AATTAGGTG GCACTTTTCG GGGAAATGTG CGCGGAACCC  
GCGCTTAAA TTGTTT'TATA ATTGCGAATG TTAATCCAC CGTGAAAAGC CCTTTTACAC GCCTTTGGG

1751 CTATTTGTTT ATTTTTC'TAA ATACATTCAA ATATGTATCC GCTCATGAGA CAATAACCCCT GATAAATGCT  
GATAAACAAA TAAAAAGATT TATGTAAAGT TATACATAGG CGAGTACTCT GTTATTGGGA CTATTTACGA

1821 TCAATAATAT TGAAAAAGGA AGAGTATGAG TATTCAACAT TTCCGTGTCG CCTTATTCC CTTT'TTGGC  
AGTTAT'TATA ACTTTTTCCT TCTCATACTC ATAAGTTGTA AAGGCACAGC GGGAATAAGG GAAAAAACGC

1891 GCATTTTGCC TTCCCTGTTT TGCTCACCCA GAAACGCTGG TGAAAGTAAA AGATGCTGAA GATCAGTTGG  
CGTAAAACGG AAGGACAAA ACGAGTGGT CTTTGC'GACC ACTTTCATT TCTACGACTT CTAGTCAACC

1961 GTGCACGAGT GGGTTACATC GAACTGGATC TCAACAGCGG TAAGATCCTT GAGAGTTTTC GCCCGAAGA  
CACGTGCTCA CCCAATGTAG CTTGACC'TAG AGTTGTCGCC ATTC'TAGGAA CTCTCAAAAG CGGGCTTCT

2031 ACGTTT'TCCA ATGATGAGCA CTTT'TAAAGT TCTGCTATGT GCGCGGTAT TATCCCGTAT TGACGCCGGG  
TGCAAAAGGT TACTACTCGT GAAAATTCA AGACGATACA CCGCGCCATA ATAGGCATA ACTGCGGCC

2101 CAAGAGCAAC TCGGTCGCC CATACACTAT TCTCAGAA'TG ACTTGGTTGA GTACTCACC GTACAGAAA  
GTTCTCGTTG AGCCAGCGC GTATGTGATA AGAGTCTTAC TGAACCACT CATGAGTGGT CAGTGTCTT

FIG.-41D

2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA ACCATGAGTG ATAACACTGC  
TCGTAGAAATG CCTACCGTAC TGTCAATCTC TTAATACGTC ACGACGGTAT TGGTACTCAC TATTGTGACG

2241 GGCCAACTTA CTTC TGACAA CGATCGGAGG ACCGAAAGAG CTAACCGCTT TTTTGACAA CATGGGGAT  
CCGGTTGAAT GAAGACTGTT GCTAGCCTCC TGGCTTCCTC GATTGGCGAA AAAACGTGTT GTACCCCTA

2311 CATGTAATC GCCTTGATCG TTGGGAACCG GAGCTGAATG AAGCCATACC AAACGACGAG CGTGACACCA  
GTACATTGAG CGGAAC TAGC AACCTTGGC CTCGACTTAC TTCGGTATGG TTTGCTGCTC GCACTGTGGT

2381 CGATGCCCTGT AGCAATGGCA ACAACGTTGC GCAAACTATT AACTGGCGAA CTACTTACTC TAGCTTCCC  
GCTACGGACA TCGTTACCGT TGTTGCAACG CGTTTGATAA TTGACCGCTT GATGAAATGAG ATCGAAGGC

2451 GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTTC TGCGTCGGC CCTTCCGGCT  
CGTTGTTAAT TATCTGACCT ACCTCCGCCT ATTTCAACGT CCTGGTGAAG ACGCGAGCCG GGAAGGCCGA

2521 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTC TC GCGG TATCATTGCA GCACTGGGC  
CCGACC AAT AACGACTATT TAGACC TC GCGC CCAAGAGGCC ATAGTAACGT CGTGACCCCG

2591 CAGATGGTAA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGAGTCAG GCAACTATGG ATGAACGAAA  
GTC TACCAT CCGGAGGCA TAGCATCAAT AGATGTGCTG CCC TCAGTC CGTTGATACC TACTTGCTTT

2661 TAGACAGATC GCTGAGATAG GTGCC TCACT GATTAAGCAT TGGTAACTGT CAGACCAAGT TTACTCATAT  
ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTGTA ACCATTGACA GTCTGGTTCA AATGAGTATA

2731 ATACTTTAGA TTGATTTAAA ACTTCATTTT TAATTTAAAA GGATCTAGGT GAAGATCCTT TTTGATAATC  
TATGAAATCT AACTAAATTT TGAAGTAAAA ATTAATTTT CCTAGATCCA CTTCTAGGAA AAAC TATTAG

2801 TCATGACCAA AATCCCTTAA CGTGAGTTT CGTTC CACTG AGC GTCAGAC CCCGTAGAAA AGATCAAAGG  
AGTACTGGTT TTAGGGAAAT GCACTCAAAA GCAAGGTGAC TCGCAGTCTG GGCATCTTT TCTAGTTTCC

2871 ATCTTCTTGA GATCCTTTTT TTCTGCGCGT AATCTGCTGC TTGCAACAA AAAAACACC GCTACCAGCG  
TAGAAGAACT CTAGGAAAAA AAGACGCGCA TTAGACGACG AACGTTTGT TTTT TGGT CGATGGTCG

2941 GTGGTTTGT TTGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAA TGGCTTCAGC AGAGCGCAGA  
CACCAAACAA ACGGCC TAGT TCTCGATGGT TGAGAAAAAG GCTTCCATTG ACCGAAAGTCG TCTCGCGTCT

FIG. 41E

3011 TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG AACTCTGTAG CACCGCCTAC  
ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC TTGAGACATC GTGGCGGATG

3081 ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC AGTGGCGATA AGTCGTGTCT TACCGGGTTG  
TATGGAGCGA GACGATTAGG ACAA TGGTCA CCGACGACGG TCACCGCTAT TCAGCACAGA ATGGCCCAAC

3151 GACTCAAGAC GATAGTTACC GGATAAGCG CAGCGGTCGG GCTGAACGGG GGGTTCGTGC ACACAGCCCA  
CTGAGTTCTG CTATCAATGG CCTATTCCGC GTCGCCAGCC CGACTTGCCC CCAAGCACG TGTGTCGGGT

3221 GCTTGGAGCG AACGACCTAC ACCGAACTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC  
CGAACCTCGC TTGCTGGATG TGGCTTGACT CTATGGATGT CGCACTCGAT ACTCTTTCGC GGTGCCGAAGG

3291 CGAAGGGAGA AAGCGGACA GGTA TCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT  
GCTTCCCTCT TTCCGCCCTGT CCATAGGCCA TTCGCCGTCC CAGCCTTGTC CTCTCGCGTG CTCCTCGAA

3361 CCAGGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCCGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT  
GGTCCCCCTT TCGGGACCAT AGAAATATCA GGACAGCCCA AAGCGGTGGA GACTGAAC TC GCAGCTAAAA

3431 TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAACGC CAGCAACGC G CCTTTTAC GGTTCCTGGC  
ACACTACGAG CAGTCCCCC GCCTCGGATA CCTTTTGGC GTCGTTGCG CGGAAAAATG CCAAGGACCG

3501 CTTTGTGCTG CCTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT CTGTGGATAA CCGTATTACC  
GAAAACGACC GGAACACGAG TGTACAAGAA AGGACGCAAT AGGGGACTAA GACACCTATT GGCATAATGG

3571 GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG  
CGGAAACTCA CTCGACTATG GCGAGCGGCG TCGGCTTGCT GGCTCGCGTC GCTCAGTCAC TCGCTCCTTC

3641 CGGAAGAGCG CCCAATACGC AAACCGCCTC TCCCCGCGG TTGGCCGATT CATTAA TGCA GCTGGCACGA  
GCCTTCTCGC GGGTTATGCG TTTGGCGGAG AGGGCGCGC AACCGCTAA GTAATTACGT CGACCGTGCT

3711 CAGGTTTCCC GACTGGAAAG CGGGCAGTGA GCGCAACGCA ATTAATGTGA GTTAGCTCAC TCATTAGGCA  
GTCCAAAGGG CTGACCTTTC GCCCGTCACT CGCGTTGCGT TAATTACACT CAATCGAGTG AGTAATCCGT

3781 CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT GTGGAATTGT GAGCGGATAA CAATTTCACA  
GGGTCCGAA ATGTGAAATA CGAAGGCCGA GCATACAACA CACCTTAACA CTCGCCATT GTTAAAGTGT

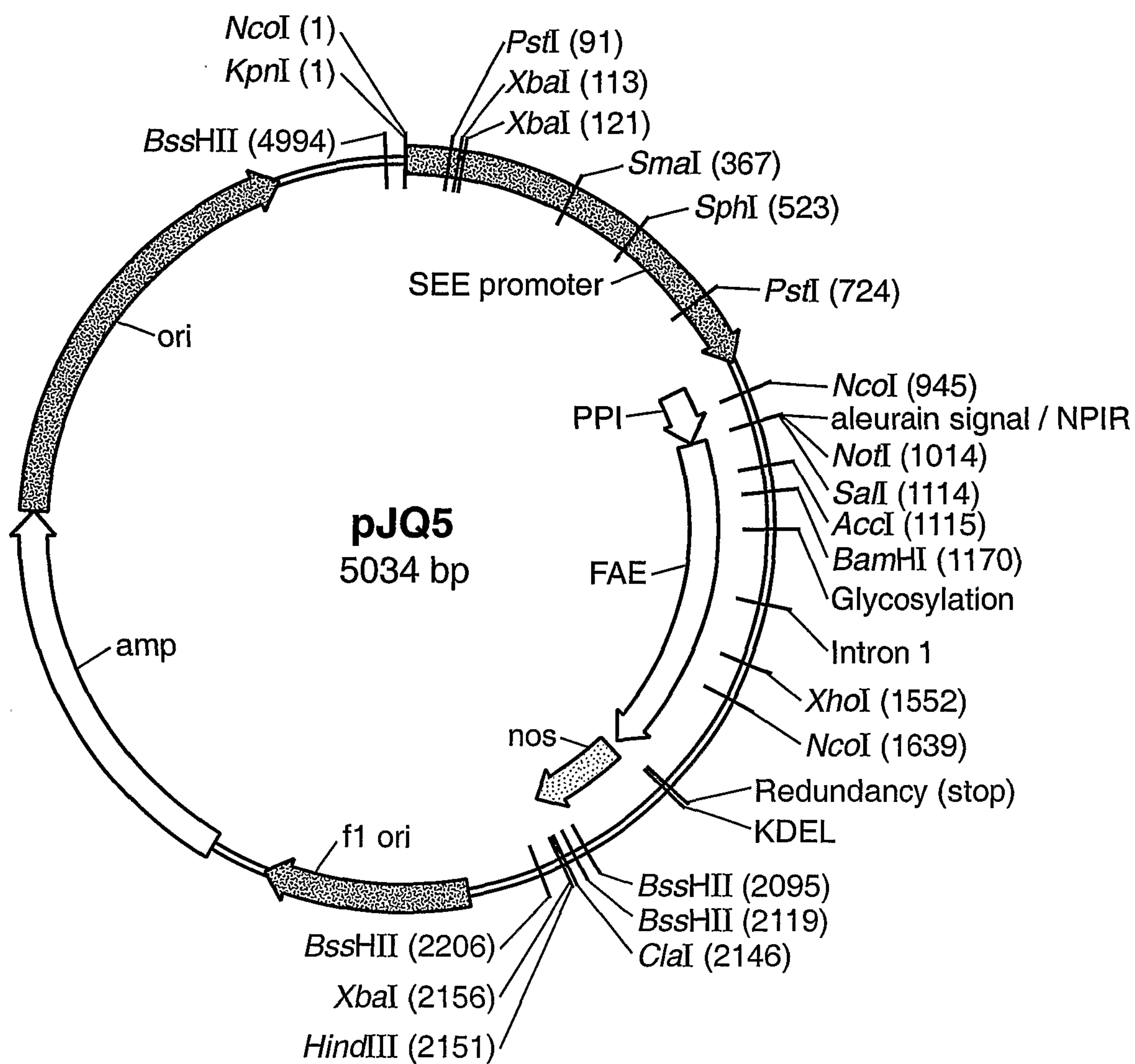
FIG. 41F

|      |            | BssHII      |             | EcoRI       |            |
|------|------------|-------------|-------------|-------------|------------|
|      |            | ~~~~~       |             | ~~~         |            |
| 3851 | CAGGAAACAG | CTATGACCAT  | GATTACGCA   | AGCGCGCAAT  | TAACCTCAC  |
|      | GTCCTTTGTC | GATACTGGTA  | CTAATGCGGT  | TCGCGCGTTA  | ATTGGAGTG  |
|      |            | EcoRI       |             | TATTCCCTTG  |            |
|      |            | ~~~~~       |             | TATTCCCTTG  |            |
| 3921 | ATTCCACAAT | GAAACAATAAT | AAGATTAAA   | TAGCTTGCCC  | CCGTTGCAGC |
|      | TAAGGTGTTA | CTTGTTATTA  | TTCCTAATTT  | ATCGAACGGG  | GGCAACGTCG |
| 3991 | AAATAAAAGA | TAAACTTAGA  | CTCAAAAACAT | TTACAAAAC   | AACCCCTAAA |
|      | TTTATTTTCT | ATTGAATCT   | GAGTTTTGTA  | AATGTTTTTG  | TTGGGGAATT |
| 4061 | ATGCACGATC | CATAGCAAGC  | CCAGCCCAAC  | CCAACCCACC  | CCAGTGCAGC |
|      | TACGTGCTAG | GTATCGTTTG  | GGTCGGGTTG  | GGTTGGGTTG  | GGTCACGTCG |
| 4131 | ATAGTCTCCA | CCCCCGGCAC  | TATCACCGTG  | AGTTGTCCGC  | ACCACCGCAC |
|      | TATCAGAGGT | GGGGCCCGTG  | ATAGTGGCAC  | TCAACAGGCG  | TGGTGGCGTG |
| 4201 | AAAAGAAAGA | AAAAAAAGAA  | AAAGAAAAAC  | AGCAGGTGGG  | TCCGGGTCGT |
|      | TTTTCCTTCT | TTTTTTTCTT  | TTTCTTTTGT  | TCGTCCACCC  | AGGCCAGCA  |
| 4271 | GGATCGCGAG | CAGCGACGAG  | GCCCCGCCCT  | CCCTCCGCTT  | CCAAAGAAAC |
|      | CCTAGCGCTC | GTCGCTGCTC  | CGGGCCGGGA  | GGAGGCGGAA  | GGTTCTTTTG |
| 4341 | ACATACCCCC | CCCTCTCCTC  | CCATCCCCCC  | AACCCCTACCA | CCACCACCAC |
|      | TGTATGGGGG | GGGAGAGGAG  | GGTAGGGGGG  | TTGGGATGGT  | GGTGGTGGTG |
| 4411 | CTGCCGGACG | ACGAGCTCCT  | CCCCCCTCCC  | CCCTCCGCCG  | CGCCGGTAAC |
|      | GACGGCCTGC | TGCTCGAGGA  | GGGGGAGGGG  | GGAGCGGGCG  | GCGGCATTG  |
| 4481 | TCCTTCTCCG | TTTTTTTCTT  | CGTCTCGGTC  | TCGATCTTTG  | GCCTTGGTAG |
|      | AGAAAGAGGC | AAAAAAAGAA  | GCAGAGCCAG  | AGCTAGAAAC  | CGGAACCATC |
| 4551 | TTTCGTCGCC | AGATCGGTGC  | GCGGGAGGGG  | CGGGATCTCG  | CGGCTGGCGT |
|      | AAGCAGCGGG | TCTAGCCACG  | CGCCCTCCCC  | GCCCTAGAGC  | GCCGACCGCA |

FIG. 41G

|      | BamHI                                                                       | BglII                 |
|------|-----------------------------------------------------------------------------|-----------------------|
|      | ~~~~~                                                                       | ~~~~~                 |
| 4621 | GGATCCCTCGC GGGGAATGGG GCTCTCGGAT GTAGATCTTC TTCTTTTCTT CTTTGTGG TAGAATTGA  | CTTTTGTGG TAGAATTGA   |
|      | CCTAGGAGCG CCCCTTACCC CGAGAGCCTA CATCTAGAAG AAAGAAAGAA GAAAACACC ATCTTAAACT | GAAAACACC ATCTTAAACT  |
| 4691 | ATCCCTCAGC ATTGTTCATC GGTA GTTTT CTTTTCATGA TTGTGACAA ATGCAGCCTC GTGCGGAGCT | ATGCAGCCTC GTGCGGAGCT |
|      | TAGGGAGTCG TAACAAGTAG CCATCAAAA GAAAGTACT AAACACTGTT TACGTCGGAG CACGCCTCGA  | TACGTCGGAG CACGCCTCGA |
| 4761 | TTTTTGTAGG TAG                                                              |                       |
|      | AAAAACATCC ATC                                                              |                       |

FIG. 41H

**FIG. 42A**

NcoI  
~~~~~  
KpnI
~

1
CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
GTACCCCGGTC CATATTAAATA CCTATATAGAG TTCGTTTATT AGCTTTATAG TGGTAACCGA TGTATATAGAC

PstI
~~~~~

71  
AGCTCCGAGT TCTGACTGCA GTCCTGGATGA CGCGTGTGTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC  
TCGAGGCTCA AGACTGACGT CAGACCTACT GCGCACAAACA TAGATCTTGA GATCTATCGT GTCGGTGTCTG

XbaI XbaI  
~~~~~ ~~~~~

141
ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
TGGATGTCCCT CACGCTGTGA ACACCTGACA TCATCACAAAC CTCTGCCCTCG AGAAAGGATG GAGGACTGCA

SmaI
~~~~~

211  
TGCCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGGGCTC CCAACAAAAT ATCGTCCCCC  
ACGGCGGCAA CAGGTAAGGT TGCCGTAGTG AGAGTTGGTT AGTGCGCGAG GGTGTGTTTA TAGCAGGGGG

SphI  
~~~~~

281
ATGTCCTTGGC GGAGAGAGAG TACATACATG CTGTCGGGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
TACAGAACCG CCTCTCTCTC ATGTATGTAC GACAGCGCGG CAAAACAGA CTTAGAGCGA AGTGACCCG

SmaI
~~~~~

351  
AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCTG TCACCCCTGG CGTCATGGGA  
TTAGTCGAGT CGAGGGCCCT CGAGTGAGTA AGTCTTAGGG TAGCAGCAGC AGTGGGGACC GCAGTACCCT

SphI  
~~~~~

421
TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
ACCTTTTCTT GGAGGCAACG AGCCTACTCA GTCGGTATAG GGGCTTGTCT CATGACGTTT TATTGGGTTA

SphI
~~~~~

491  
TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTGG CTTAATTGAC TTTATTTTGG  
AGTCTAAGGG GGTATCTCTT TTCATATCGT ACGAAAGCCC AAAACAAACC GAATTAACTG AAATAAAAC

SphI  
~~~~~

561
TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAAATAT CGAGACGGAT AATAGGCTGG
AACCCTCAACT TACGACTAAA CAACACATTT TACGGGTTGG TAGACTTATA GCTCTGCCCTA TTATCCGACC

FIG. 42B

SUBSTITUTE SHEET (RULE 26)

| | | | | | | | |
|------|--------------------------------|------------|-------------|-------------|-------------|------------|------------|
| 631 | CTAATTAAATT | TATAGCAAGA | TTCGTGATGT | CACATCGCAA | ATATCTTTCT | GGGCATTACA | GCTGGAGGCT |
| | GATTAATTAA | ATATCGTTCT | AAGACATCAC | GTGTAGCGTT | TATAGAAAGA | CCCGTAATGT | CGACCTCCGA |
| | PstI
~~~~~ | | | | | | |
| 701 | TCATCAGCCT | GAAACACTCT | GCAGAGCCTG | AAGCAAGTGG | TGAAGCGTGG | CGATGAGATG | GGTATAAAAC |
| | AGTAGTCGGA | CTTTGTGAGA | CGTCTCGGAC | TTTCGTTTACC | ACTTCGCACC | GCTACTCTAC | CCATATTTTG |
| 771 | CCCCGGCACC | GGGACGCGAG | CTCCCCGCCTA | CCAGTACCAT | CTCGCCTCGC | TCCCCCTGCC | GGACGACCCA |
| | GGGGCCGTGG | CCCTGCGCTC | GAGGGCGGAT | GGTCATGGTA | GAGCGGAGCG | AGGGGGACGG | CCTGCTGGGT |
| 841 | GTAATAATACT | GTTGCCCACT | CGCCGGCGAG | ATGGMCGTGC | ACAAGGAGGT | SAACTTCGTS | GCCTACCTCC |
| | CATTTTATGA | CAACGGGTGA | GCGGCCGCTC | TACCKGCACG | TGTTCTCTCCA | STTGAAGCAS | CGGATGGAGG |
| | NcoI
~~~~~ | | | | | | |
| 911 | TGATCGTSCT | CGGCCTCCTC | TTGCTCGTST | CCGCCATGGA | GCACGTGGAC | GCCAAGGCCT | GCACCKCKGA |
| | ACTAGCASGA | GCCGGAGGAG | AACGAGCASA | GGCGGTACCT | CGTGCACTG | CGGTTCCGGA | CGTGGGMGCT |
| | NotI
~~~~~ | | | | | | |
| 981 | GTGCGGCAAC | CTCGGCTTCG | GCATCTGCC | GGCGGCCGCC | TCCACGCAGG | GCATCTCCGA | AGACCTCTAC |
| | CACGCCGTTG | GAGCCGAAGC | CGTAGACGGG | CCGCCGGCGG | AGGTGCGTCC | CGTAGAGGCT | TCTGGAGATG |
| | SalI
~~~~~
AccI
~~~~~ | | | | | | |
| 1051 | AGCCGTTTAG | TCGAAATGGC | CACTATCTCC | CAAGCTGCCT | ACGCCGACCT | GTGCAACATT | CCGTCGACTA |
| | TCGGCAAATC | AGCTTTTACC | GTGATAGAGG | GTTCCGACGGA | TGCGGCTGGA | CACGTTGTAA | GGCAGCTGAT |
| | BamHI
~~~~~ | | | | | | |
| 1121 | TTATCAAGGG | AGAGAAAATT | TACAAATCTC | AAACTGACAT | TAACGGATGG | ATCCTCCGCG | ACGACAGCAG |
| | AATAGTTCCC | TCTCTTTTAA | ATGTTAAGAG | TTTGACTGTA | ATTGCCCTACC | TAGGAGGCGC | TGCTGTCTGC |

FIG. 42C

1191 CAAAGAAATA ATCACCGTCT TCCGTGGCAC TGGTAGTGAT ACGAATCTAC AACTCGATAC TAACTACACC
GTTTCTTTAT TAGTGGCAGA AGGCACCGTG ACCATCACTA TGCTTAGATG TTGAGCTATG ATTGATGTGG

1261 CTCACGCCCTT TCGACACCCCT ACCACAATGC AACGGTTGTG AAGTACACGG TGGATATTAT ATTGGATGGG
GAGTGC GGAA AGCTGTGGGA TGGTGTTACG TTGCCAACAC TTCAATGTGCC ACCTATAATA TAACCTACCC

1331 TCTCCGTCCA GGACCAAGTC GAGTCGCTTG TCAAAACAGCA GGTTAGCCAG TATCCGGACT ACGCGCTGAC
AGAGGCAGGT CCTGGTTCAG CTCAGCGAAC AGTTTGTCGT CCAATCGGTC ATAGGCCCTGA TGC CGACTG

1401 CGTGACCGGC CACKCCCTCG GCGCCTCCCT GCGGGCACTC ACTGCCGCC AGCTGTCTGC GACATACGAC
GCAC TGCGCG GTGMGGGAGC CCGCGAGGA CCGCCGTGAG TGACGGCGGG TCGACAGACG CTGTA TGCTG

1471 AACATCCGCC TGTACACCTT CGCGAAACCG CGCAGCGGCA ATCAGGCCTT CGCGTCGTAC ATGAACGATG
TTGTAGGCGG ACATGTGGAA GCGCCTTGGC GCGTCGCCGT TAGTCCGGAA GCGCAGCATG TACTTGCTAC

XhoI
~~~~~

1541 CCTTCCAAGC CTCGAGCCCA GATACGACGC AGTATTTCCG GGTCACTCAT GCCAACGACG GCATCCCAA  
GGAAGGTTTCG GAGCTCGGGT CTATGCTGCG TCATAAAGGC CCAGTGAGTA CGGTTGCTGC CGTAGGGTTT

NcoI  
~~~~~

1611 CCTGCCCCCG GTGGAGCAGG GGTACGCCCA TGGCGGTGTA GAGTACTGGA GCGTTGATCC TTACAGCGCC
GGACGGGGGC CACCTCGTCC CCATGCGGGT ACCGCCACAT CTCATGACCT CGCAACTAGG AATGTCGCGG

1681 CAGAACACAT TTGTCTGCAC TGGGGATGAA GTGCAGTGCT GTGAGGCCCA GGGCGGACAG GGTGTGAATA
GTC TTGTGTA AACAGACGTG ACCCTACTT CACGTCACGA CACTCCGGGT CCCGCCCTGC CCACACTTAT

1751 ATGCGCACAC GACTTATTTT GGGATGACGA GCGGAGCCTG TACATGGTGA TCAGTCA TTT CAGCCTCCCC
TACGCGTG TG CTGAATAAAA CCTACTGCT CGCCTCGGAC ATGTACC ACT AGTCAGTAAA GTCGGAGGGG

1821 GAGTG TACCA GGAAAGATGG ATGTCTTGA GAGGGGCGG CGTAACCACT GAAGGATGAG CTGTAAAGAA
CTCACATGGT CCTTCTACC TACAGGACCT CTCCCCCGG GCATTGGTGA CTTCCCTACTC GACATTTCTT

FIG.- 42D

| | | | | | | | |
|------|------------|------------|-------------|------------|------------|------------|-------------|
| 1891 | GCAGATCGTT | CAAACATTTG | GCAATAAAGT | TTCTTAAGAT | TGAATCCTGT | TGCCGGTCTT | GCATGATTA |
| | CGTCTAGCAA | GTTTGTAAC | CGTTATTCA | AAGAAATCTA | ACTTAGGACA | ACGGCCAGAA | CGCTACTAAT |
| 1961 | TCAATATAAT | TCTGTTGAAT | TACGTTAAGC | ATGTAATAAT | TAACATGTAA | TGCATGACGT | TATTTATGAG |
| | AGTATATTAA | AGACAACTTA | ATGCAATTCTG | TACATTATTA | ATTGTACATT | ACGTACTGCA | ATAAATATCTC |
| 2031 | ATGGGTTTTT | ATGATTAGAG | TCCCGCAATT | ATACATTTAA | TACGCGATAG | AAAACAAAAT | ATAGCGCGCA |
| | TACCCAAAAA | TACTAATCTC | AGGCGTTAA | TATGTAAATT | ATGCGCTATC | TTTTGTTTTA | TATCGCGCGT |
| 2101 | AACTAGGATA | AATATCGCG | CGCGGTGTC | TCTATGTTAC | TAGATCGATA | AGCTTCTAGA | GCGGCCGGTG |
| | TTGATCCTAT | TTAATAGCG | GCGCCACAGT | AGATACAATG | ATCTAGCTAT | TCGAAGATCT | CGCCGGCCAC |
| 2171 | GAGCTCCAAT | TCGCCCTATA | GTGAGTCGTA | TTACGCGCGC | TCACTGGCCG | TCGTTTTACA | ACGTCGTGAC |
| | CTCGAGGTTA | AGCGGGATAT | CACTCAGCAT | AATGCGCGCG | AGTGACCGGC | AGCAAAATGT | TGCAGCACTG |
| 2241 | TGGGAAAACC | CTGGCGTTAC | CCAACTTAAT | CGCCTTGCAG | CACATCCCCC | TTTCGCCAGC | TGGCGTAATA |
| | ACCCTTTGTG | GACCGCAATG | GGTTGAATTA | GCGGAACGTC | GTGTAGGGGG | AAAGCGGTCG | ACCGCATTAT |
| 2311 | GCGAAGAGGC | CCGCACCGAT | CGCCCTTCCC | AACAGTTGCG | CAGCCTGAAT | GGCGAATGGG | ACGCGCCCCTG |
| | CGCTTCTCCG | GGCGTGGCTA | GCGGGAAGGG | TTGTCAACGC | GTCGGACTTA | CCGCTTACCC | TGCGCGGGAC |
| 2381 | TAGCGGCGCA | TTAAGCGCGG | CGGGTGTTGT | GGTTACGCGC | AGCGTGACCG | CTACACTTGC | CAGCGCCCCTA |
| | ATCGCCGCGT | AATTCGCGCC | GCCACACCA | CCAATGCGCG | TCGCACTGGC | GATGTGAACG | GTCGCGGGAT |
| 2451 | GCGCCCGCTC | CTTTCGCTTT | CTTCCCTTCC | TTTCTCGCCA | CGTTCGCCGG | CTTTCCCCGT | CAAGCTCTAA |
| | CGCGGGCGAG | GAAAGCGAAA | GAAGGAAGG | AAAGAGCGGT | GCAAGCGGCC | GAAAGGGGCA | GTTCGAGATT |

BssHII
~~~~~

XbaI  
~~~~~

Clal HindIII
~~~~~

BssHII  
~~~~~

BssHII
~~~~~

FIG.- 42E

2521 ATCGGGGGCT CCTTTAGGG TTCCGATTTA GTGCTTTACG GCACCTCGAC CCCAAAAAAC TTGATTAGGG  
TAGCCCCCGA GGGAAATCCC AAGGCTAAAT CACGAAATGC CGTGGAGCTG GGGTTTTTG AACTAATCCC

2591 TGATGGTTCA CGTAGTGGG CATCGCCCTG ATAGACGGTT TTTCGCCCTT TGACGTTGGA GTCCACGTTT  
ACTACCAAGT GCATCACCG GTAGCGGGAC TATCTGCCAA AAAGCGGAA ACTGCAACCT CAGGTGCAAG

2661 TTTAATAGTG GACTCTTGTT CCAAACTGA ACAACACTCA ACCCTATCTC GGTCTATTCT TTTGATTTAT  
AAATATATC C TGAGAACAA GGTTTGACCT TGTGTGAGT TGGGATAGAG CCAGATAAGA AAACATAATA

2731 AAGGGATTTT GCCGATTTCG GCTATTGGT TAAAAAATGA GCTGATTTAA CAAAAATTTA ACGCGAATTT  
TTCCCTAAAA CGGCTAAAGC CGGATAACCA ATTTTTTACT CGACTAAATT GTTTTTAAAT TCGCCTTAAA

2801 TAACAAAAATA TTAACGCTTA CAATTTAGGT GGCACTTTTC GGGGAAATGT GCGCGGAACC CCTATTGTT  
ATTGTTTTAT AATTGCGAAT GTTAAATCCA CCGTGAAAAG CCCCTTTACA CGCGCCTTGG GGATAAACAA

2871 TATTTTTC TAATACATTCA AATATGTATC CGCTCATGAG ACAATAAACC TGATAAATGC TTCAATAATA  
ATAAAAAGAT TTAATGTAAGT TTATACATAG GCGAGTACTC TGTTATTGGG ACTATTTACG AAGTTATTAT

2941 TTGAAAAAGG AAGAGTATGA GTATTCAACA TTTCGGTGC GCCCTTATTC CCTTTTTTG GGCATTTTGC  
AACTTTTTC TTCTCATACT CATAAGTTGT AAAGGCACAG CGGGAATAAG GGAAAAACG CCGTAAAAACG

3011 CTTCCTGTTT TTGCTCACCC AGAAACGCTG GTGAAAGTAA AAGATGCTGA AGATCAGTTG GGTGCACGAG  
GAAGGACAAA AACGAGTGGG TCTTTGCGAC CACTTTTCATT TTCTACGACT TCTAGTCAAC CCACGTGCTC

3081 TGGGTTACAT CGAACTGGAT CTC AACAGCG GTAAAGATCCT TGAGAGTTT CGCCCCGAAG AACGTTTTC  
ACCCAATGTA GCTTGACCTA GAGTTGTGC GATCTAGGA ACTCTCAAAA GCGGGGCTTC TTGCAAAAGG

3151 AATGATGAGC ACTTTTAAAG TTC TGCTATG TGGCGCGGTA TTATCCCGTA TTGACGCCGG GCAAGAGCAA  
TTACTACTCG TGAAAAATTC AAGACGATAC ACCGCGCCAT AATAGGGCAT AACTGCGGCC CGTTCCTCGT

3221 CTCGGTCGCC GCATACACTA TTC TCAGAA T GACTTGGTTG AGTACTCACC AGTCACAGAA AAGCATCTTA  
GAGCCAGCG CGTATGTGAT AAGAGTCTTA CTGAACCAAC TCATGAGTGG TCAGTGTCTT TTCGTAGAAT

3291 CGGATGGCAT GACAGTAAGA GAATTATGCA GTGCTGCCAT AACCATGAGT GATAAFACTG CGGCCAACTT  
GCCTACCGTA CTGTCAATTCT CTTAATACGT CACGACGGTA TTGGTACTCA CTATTGTGAC GCCGGTTGAA

FIG. 42F

3361 ACTTCTGACA ACGATCGGAG GACCGAAGGA GCTAACCCTT TTTTTCACA ACATGGGGGA TCATGTAACT  
TGAAGACTGT TGCTAGCCTC CTGGCTTCCT CGATTGGCGA AAAAACGTGT TGTACCCCTT AGTACATTGA

3431 CGCCTTGATC GTTGGGAACC GGAGCTGAAT GAAGCCATAC CAAACGACGA GCGTGACACC ACGATGCCCTG  
GGGAACTAG CAACCCTTGG CCTCGACTTA CTTCGGTATG GTTTGCTGCT CGCACTGTGG TGCTACGGAC

3501 TAGCAATGGC AACAACTAT CGAAACTAT TAACTGGCGA ACTACTTACT CTAGCTTCCC GGCAACAATT  
ATCGTTACCG TTGTTGCAAC GCGTTTGATA ATTGACCCTT TGATGAATGA GATCGAAGGG CCGTTGTTAA

3571 AATAGACTGG ATGGAGGCGG ATAAAGTTGC AGGACCACCTT CTGCGCTCGG CCCTTCCGGC TGGCTGGTTT  
TTATCTGACC TACCTCCGCC TATTTCACG TCCTGGTGAA GACGCGAGCC GGGAAAGCCG ACCGACCAAA

3641 ATTGCTGATA AATCTGGAGC CGGTGAGCGT GGTCTCGCG GTATCATTCG AGCACTGGG CCAGATGGTA  
TAACGACTAT TTAGACCCTCG GCCACTCGCA CCCAGAGCGC CATAGTAACG TCGTGACCCC GGTCTACCAT

3711 AGCCCTCCCG TATCGTAGTT ATCTACACGA CGGGAGTCA GGCAACTATG GATGAACGAA ATAGACAGAT  
TCGGGAGGGC ATAGCATCAA TAGATGTGCT GCCCTCAGT CCGTTGATAC CTACTTGCTT TATCTGTCTA

3781 CGCTGAGATA GGTGCCCTCAC TGATTAAGCA TTGGTAACTG TCAGACCAAG TTFACTCATA TATACTTTAG  
GCGACTCTAT CCACGGAGTG ACTAATTGCT AACCATTGAC AGTCTGGTTC AAATGAGTAT ATATGAAATC

3851 ATTGATTTAA AACTTCATTT TTAAATTAAA AGGATCTAGG TGAAGATCCT TTTTGATAAT CTCA TGACCA  
TAACTAAATT TTGAAGTAAA AATTAAATTT TCCTAGATCC ACTTCTAGGA AAAACTATTA GAGTACTGGT

3921 AAATCCCTTA ACGTGAGTTT TCGTTCCACT GAGCGTCAGA CCCCGTAGAA AAGATCAAAG GATCTTCTTG  
TTTAGGGGAA TGCACCTCAA AGCAAGGTGA CTCGCAGTCT GGGGCATCTT TTCTAGTTTC CTAGAAGAAC

3991 AGATCCCTTT TTTC TGCGCG TAATCTGCTG CTTGCAAACA AAAAACCCAC CGCTACCAGC GGTGGTTTGT  
TCTAGGAAA AAAGACGCGC ATTAGACGAC GAACGTTTGT TTTTTTGGTG GCGATGGTCG CCACCAAACA

4061 TTGCCGGATC AAGAGCTACC AACTCTTTT CCGAAGGTAA CTGGCTTCAG CAGAGCGCAG ATACCAAATA  
AACGGCCTAG TTCTCGATGG TTGAGAAAA GCTTCCATT GACCGAAGTC GTCTCGCGTC TATGGTTTAT

4131 CTGTCCCTTCT AGTGAGCCG TAGTTAGGCC ACCACTTCAA GAAC TCTGTA GCACCGCCTA CATACCTCGC  
GACAGGAAGA TCACATCGGC ATCAATCCGG TGGTGAAGTT CTTGAGACAT CGTGGCGGAT GTATGGAGCG

FIG. 42G

4201 TCTGCTAATC CTGTTACCAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGA<sup>CT</sup>CAAGA  
AGACGATTAG GACAAATGGTC ACCGACGACG GTCACCGCTA TTCAGCACAG AATGGCCCAA CCTGAGTCT

4271 CGATAAGTTAC CGGATAAGGC GCAGCGGTCG GGCTGAACGG GGGGTTTCGTG CACACAGCCC AGCTTGGAGC  
GCTATCAATG GCCTATTCCG CGTCGCCAGC CCGACTTGCC CCCCAAGCAC GTGTGTCGGG TCGAACCTCG

4341 GAACGACCTA CACCGAACTG AGATACCTAC AGCGTGAGCT ATGAGAAAGC GCCACGCTTC CCGAAGGGAG  
CTTGCTGGAT GTGGCTTGAC TCTATGGATG TCGCACTCGA TACTCTTTCG CGGTGCGAAG GGC<sup>T</sup>TCCCCTC

4411 AAAGCGGAC AGGTATCCGG TAAGCGGCAG GGTCGGAACA GGAGAGCGCA CGAGGGAGCT TCCAGGGGGA  
TTTCCGCCCTG TCCATAGGCC ATTGCGCGTC CCAGCCTTGT CCTCTCGCGT GCTCCCTCGA AGGTCCCCT

4481 AACGCCTGGT ATCTTTATAG TCCTGTCCGG TTTCCGCCACC TCTGACTTGA GCGTCGATTT TTGTGATGCT  
TTGCGGACCA TAGAAATATC AGGACAGCC AAAGCGGTGG AGACTGAACT CGCAGCTAAA AACACTACGA

4551 CGTCAGGGG GCGGAGCCTA TGGAAAAACG CCAGCAACGC GGCTTTTTA CGGTTCCTGG CCTTTTGCTG  
GCAGTCCCC CGCTCGGAT ACCTTTGTG GGTCTGTGCG CCGGAAAAAT GCCAAGGACC GGAAAAACGAC

4621 GCCTTTTGCT CACATGTTCT TTCTGCGGTT ATCCCCTGAT TCTGTGGATA ACCGTATTAC CGC<sup>TT</sup>TGAG  
CGGAAAAACGA GTGTACAAGA AAGGACGCAA TAGGGGACTA AGACACCTAT TGGCATAATG GCGGAAAC<sup>TC</sup>

4691 TGAGCTGATA CCGCTCGCCG CAGCCGAACG ACCGAGCGCA GCGAGTCAGT GAGCGAGGAA GCGGAAGAGC  
ACTCGACTAT GCGGAGCGGC GTCGGCTTGC TGGCTCGCGT CGCTCAGTCA CTCGCTCCTT CGC<sup>TT</sup>TCTCG

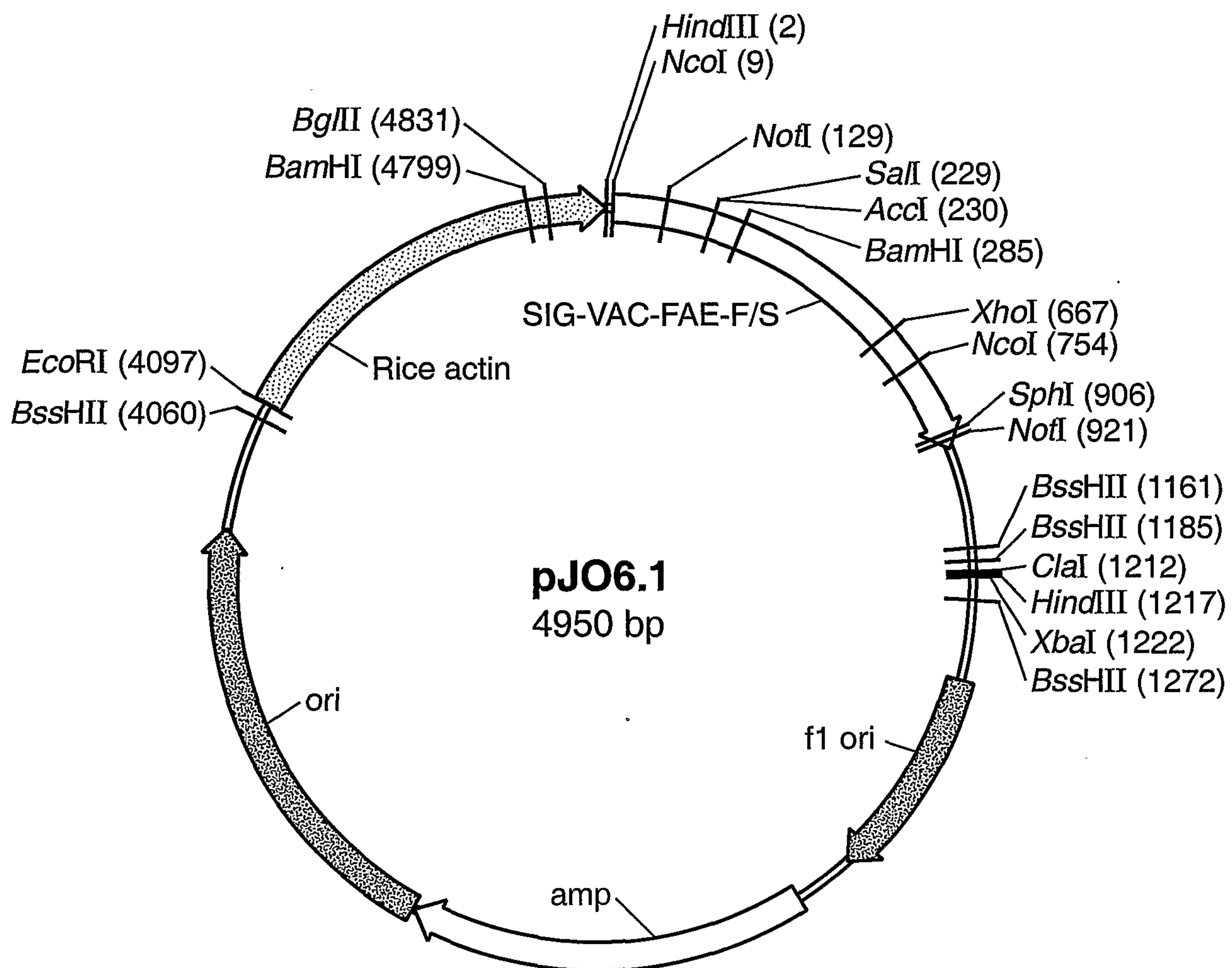
4761 GCCCAATACG CAAACCGCCT CTC<sup>CCC</sup>CGCG GTTGGCCGAT TCATTAAATGC AGCTGGCACG ACAGGTTTCC  
CGGGTTATGC GTTTGGCGGA GAGGGCGCG CAACCGGCTA AGTAAATTACG TCGACCGTGC TGTCCAAGG

4831 CGACTGGAAA GCGGGCAGTG AGCGCAACGC AATTAATGTG AGTTAGCTCA CTCATTAGG ACCCAGGCT  
GCTGACCTTT CGCCCGTCAC TCGCGTTGCG TTAATTACAC TCAATCGAGT GAGTAAATCCG TGGG<sup>GT</sup>TCCGA

4901 TTACACTTTA TGCTTCCGGC TCGTATGTTG TGTGGAATTG TGAGCGGATA ACAA<sup>TT</sup>TCAC ACAGGAAACA  
AATGTGAAAT ACGAAGGCCG AGCATACAAC ACACCTTAAC ACTCGCCTAT TGT<sup>TAA</sup>AGTG TGTCC<sup>TT</sup>TGT

FIG.- 42H

**SUBSTITUTE SHEET (RULE 26)**

**FIG. 43A**

|              |             |            |            |             |            |             |             |           |  |
|--------------|-------------|------------|------------|-------------|------------|-------------|-------------|-----------|--|
| HindIII NcoI |             |            |            |             |            |             |             |           |  |
| ~~~~~ ~~~~~~ |             |            |            |             |            |             |             |           |  |
| 1            | AAGCTTACCA  | TGGCCACGC  | CCGCTCCTC  | CTCCTGGCGC  | TCGCCGTGCT | GGCCACGGCC  | GCCGTCGCCG  | CGGAGCGGC |  |
|              | TTCGAATGGT  | ACCGGTGCG  | GGCGCAGGAG | GAGGACCGCG  | AGCGGCACGA | CCGGTGCCGG  | CGGAGCGGC   |           |  |
|              | NotI        |            |            |             |            |             |             |           |  |
|              | ~~~~~       |            |            |             |            |             |             |           |  |
| 71           | TCGCCCTCCTC | CTCCTCCTTC | GCCGACTCCA | ACCCGATCCG  | GCCCCTCACC | GACCGCGCGG  | CCGCTCCAC   |           |  |
|              | AGCGGAGGAG  | GAGGAGGAAG | CGGCTGAGGT | TGGGCTAGGC  | CGGGCAGTGG | CTGGCGCGCC  | GGCGGAGGTG  |           |  |
| 141          | GCAGGGCATC  | TCCGAAGACC | TCTACAGCCG | TTTAGTCGAA  | ATGGCCACTA | TCTCCCAAGC  | TGCCCTACGCC |           |  |
|              | CGTCCCCTAG  | AGGCTTCTGG | AGATGTCGGC | AAATCAGCTT  | TACCGGTGAT | AGAGGGTTTCG | ACGGATGCCG  |           |  |
|              | SalI        |            |            |             |            |             |             |           |  |
|              | ~~~~~       |            |            |             |            |             |             |           |  |
|              | AccI        |            |            |             |            |             |             |           |  |
|              | ~~~~~       |            |            |             |            |             |             |           |  |
| 211          | GACCTGTGCA  | ACATTCCGTC | GACTATTATC | AAGGAGAGA   | AAATTTACAA | TTCTCAAAC   | GACATTAACG  |           |  |
|              | CTGGACACGT  | TGTAAGGCAG | CTGATAATAG | TTCCCCTCTCT | TTTAAATGTT | AAGAGTTTGA  | CTGTAATTGC  |           |  |
|              | BamHI       |            |            |             |            |             |             |           |  |
|              | ~~~~~       |            |            |             |            |             |             |           |  |
| 281          | GATGGATCCT  | CCGCGACGAC | AGCAGCAAAG | AAATAATCAC  | CGTCTTCCGT | GGCACTGGTA  | GTGATACGAA  |           |  |
|              | CTACCTAGGA  | GGCGCTGCTG | TCGTCGTTTC | TTTATTAGTG  | GCAGAAGGCA | CCGTGACCAT  | CACATATGCTT |           |  |
| 351          | TCTACAACTC  | GATACTAACT | ACACCCTCAC | GCCTTTCGAC  | ACCCTACCAC | AATGCAACGG  | TTGTGAAGTA  |           |  |
|              | AGATGTTGAG  | CTATGATTGA | TGTGGGAGTG | CGGAAAGCTG  | TGGGATGGTG | TTACGTTGCC  | AACACTTCAT  |           |  |
| 421          | CACGGTGGAT  | ATTATATTGG | ATGGGTCCTC | GTCCAGGACC  | AAGTCGAGTC | GCTTGTCAA   | CAGCAGGTTA  |           |  |
|              | GTGCCACCCTA | TAAATATAAC | TACCCAGAGG | CAGGTCCTGG  | TTCAGCTCAG | CGAACAGTTT  | GTGCTCCAAT  |           |  |
| 491          | GCCAGTATCC  | GGACTACGGG | CTGACCGTGA | CCGGCCACKC  | CCTCGGCGCC | TCCCTGGCGG  | CACCTACTGC  |           |  |
|              | CGGTCATAGG  | CCTGATGCGC | GACTGGCACT | GGCCGGTGMG  | GGAGCCGCGG | AGGGACCGCC  | GTGAGTGACG  |           |  |
| 561          | CGCCCAGCTG  | TCTGCGACAT | ACGACAACAT | CCGCCGTGAC  | ACCTTCGGCG | AACCGGCAG   | CGGCAATCAG  |           |  |
|              | GCGGGTCGAC  | AGACGCTGTA | TGCTGTTGTA | GGCGGACATG  | TGGAAGCCGC | TTGGCGCGTC  | GCCGTTAGTC  |           |  |

FIG. 43B

|        |                                                                                                                                                                  |
|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| XhoI   |                                                                                                                                                                  |
| ~~~~~  |                                                                                                                                                                  |
| 631    | GCC TTC GCGT CGTACA TGAA CGATGCC TTC CAAGCC TC GA GCCAGATAC GACGAGTAT TTCCGGGTCA<br>CGGAAGCGCA GCATGTACTT GCTACGGAAG GTTCGGAGCT CGGGTCTATG CTGCGTCATA AAGGCCCAGT |
| NcoI   |                                                                                                                                                                  |
| ~~~~~  |                                                                                                                                                                  |
| 701    | CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA<br>GAGTACGGTT GCTGCCGTAG GGT TTGGACG GGGCCACCT CGTCCCCATG CGGGTACCGC CACATCTCAT     |
| 771    | CTGGAGCGTT GATCCTTACA GCGCCAGAA CACAT TTGTC TGCAC TGGG ATGAAGTGCA GTGCTGTGAG<br>GACCTCGCAA CTAGGAATGT CGCGG GTCTT GTGTAAACAG ACGTGACCC TACTTCACGT CACGACACTC     |
| SphI   |                                                                                                                                                                  |
| ~~~~~  |                                                                                                                                                                  |
| 841    | GCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC GCATGCACCT<br>CGGGTCCCGC CTGTCCACAC CTTATTACGC GTGTGCTGAA TAAACCCCTA CTGCTCGCCG CGTACGTGGA      |
| NotI   |                                                                                                                                                                  |
| ~~~~~  |                                                                                                                                                                  |
| 911    | GGCCGGTCGC GGCCGGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAA CATTTGGCAA<br>CCGGCCAGCG CCGGCGCCTT TGGTGACTTC CTACTCGACA TTTC TTCGT TAGCAA GTT GTAAACC GTT      |
| 981    | TAAAGTTTCT TAAGATTGAA TCCTGT TGCC GTCTTGCGA TGATTATCAT ATAA TTCTG TTGAA TTACG<br>ATTTCAAAGA ATTCTA ACTT AGGACAACGG CCAGAACGCT ACTAATAGTA TATTAAAGAC AACTTAATGC   |
| 1051   | TTAAGCATGT AATAATT AAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTATGA TTAGAGTCCC<br>AATTCGTACA TTATT AATG TACATTACGT ACTGCAATAA ATACTCTACC CAAAAATACT AATCTCAGGG     |
| BssHII |                                                                                                                                                                  |
| ~~~~~  |                                                                                                                                                                  |
| 1121   | GCAATTATAC ATTTAATACG CGATAGAAA CAAAATATAG CGCGCAAACT AGGATAAATT ATCGCGCGCG<br>CGTTAATATG TAAATTATGC GCTATCTTTT GTTTTATATC GCGCGTTTGA TCCTATTTAA TAGCGCGCGC      |

FIG. 43C

1191

GTGTCATCTA TGTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC TCCAATTCGC CCTATAGTGA  
CACAGTAGAT ACAATGATCT AGCTATTCTGA AGATCTCGCC GGCCACCTCG AGGTTAAGCG GGATATCACT

1261

GTCGTATTAC GCGCGCTCAC TGGCCGTCGT TTTACAACGT CGTGACTGGG AAAACCCCTGG CGTTACCCAA  
CAGCATAATG CCGCGAGTG ACCGGCAGCA AAATGTTGCA GCACTGACCC TTTTGGGACC GCAATGGGT

1331

CTTAATCGCC TTGCAGCACA TCCCCCTTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC  
GAATTAGCGG AACGTCGTGT AGGGGAAAG CGGTCGACCG CATTATCGCT TCTCCGGGCG TGGCTAGCGG

1401

CTTCCCAACA GTTGGCAGC CTGAATGGCG AATGGGACGC GCCCTGTAGC GCGGCATTAA GCGCGGCGGG  
GAAGGGTTGT CAACGCGTGG GACTTACCGC TTACCCCTGCG CGGGACATCG CCGCGTAATT CGCGCCGCCC

1471

TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCGC CCGCTCCTTT CGCTTCTTC  
ACACCACCAA TGC GCGTCTGC ACTGGCGATG TGAACGGTCG CGGGATCGCG GCGGAGGAAA GCGAAAGAAG

1541

CCTTCCCTTC TCGCCACGTT CGCCGGCTTT CCCCGTCAAG CTCTAAATCG GGGGCTCCCT TTAGGGTTCC  
GGAAGGAAAG AGCGGTGCAA GCGGCCGAAA GGGCAGTTC GAGATTTAGC CCCCGAGGGA AATCCCAAG

1611

GATTTAGTGC TTTACGGCAC CTCGACCCCA AAAAActTGA TTAGGGTGAT GGTTCACGTA GTGGCCATC  
CTAAATCACG AAATGCCGTG GAGCTGGGT TTTTGTGAAct AATCCCACTA CCAAGTGCAT CACCCGGTAG

1681

GCCCTGATAG ACGGTTTTC GCCCTTTGAC GTTGGAGTCC ACGTTCCTTA ATAGTGACT CTGTTCCTCA  
CGGGACTATC TGCCAAAAG CCGGAAACTG CAACCTCAGG TGCAAGAAAT TATCACCTGA GAACAAGGTT

1751

ACTGGAACAA CACTCAACCC TATCTCGGTC TATTCCTTTG ATTTATAAGG GATTTGCCG ATTCGGCCT  
TGACCTTGTT GTGAGTTGGG ATAGAGCCAG ATAAGAAAC TAAATATTCC CTAAACCGC TAAAGCCGA

1821

ATTGGTTAAA AAATGAGCTG ATTTAACAA AATTTAACG GAATTTTAA AAAATATTAA CGCTTACAAT  
TAACCAATTT TTACTCTGAC TAAATTGTTT TTAATTGCG CTTAAAAATTG TTTTATAATT GCGAATGTTA

FIG. 43D

1891 TTAGGTGGCA CTTTTCGGGG AAATGTGCGC GGAACCCCTA TTTGTTTATT TTCTAAATA CATTCAAATA  
AATCCACCGT GAAAGCCCC TTACACGCG CCTTGGGAT AAACAATAA AAGATTAT GTAAAGTTAT

1961 TGTATCCGCT CATGAGACAA TAACCCTGAT AAATGCTTCA ATAATATTGA AAAAGGAAGA GTATGAGTAT  
ACATAGGCGA GTACTCTGTT ATGGGACTA TTACGGAAGT TATTATAACT TTTCCCTTCT CATACTCATA

2031 TCAACATTTC CGTGTCGCC TTAATCCCTT TTTTGGGCA TTTTGCCCTC CTGTTTTTGC TCACCCAGAA  
AGTTGTAAAG GCACAGCGGG AATAAGGAA AAAACGCCGT AAAACGGAAG GACAAAACG AGTGGGCTT

2101 ACGCTGGTGA AAGTAAAAGA TGCTGAAGAT CAGTTGGGTG CACGAGTGGG TTACATCGAA CTGGATCTCA  
TGCGACCACT TTCAATTTCT ACGACTTCTA GTCAACCCAC GTGCTCACCC AATGTAGCTT GACCTAGAGT

2171 ACAGCGGTAA GATCCTTGAG AGTTTTCGCC CCGAAGAACG TTTTCCAATG ATGAGCACTT TTAAGTTCT  
TGTCGCCATT CTAGGAACTC TCAAAAGCGG GGCTTCTTGC AAAAGGTAC TACTCGTGAA AATTCAAGA

2241 GCTATGTGGC GCGTATTAT CCCGTATTGA CGCCGGGCAA GAGCAACTCG GTCGCCGCAT ACACTATTCT  
CGATACACCG CGCATATAA GGCATAACT GCGGCCGTT CTCGTTGAGC CAGCGCGTA TGTGATAAGA

2311 CAGAA TGACT TGGTTGAGTA CTCACCAGTC ACAGAAAGC ATCTTACGGA TGGCATGACA GTAAGAGAAT  
GTCTTACTGA ACCAACTCAT GAGTGGTCAG TGTCTTTTCG TAGAATGCCT ACCGTACTGT CATCTCTTA

2381 TATGCAGTGC TGCCATAACC ATGAGTGATA ACACTGCGC CAACTTACTT CTGACAACGA TCGAGGACC  
ATACGTACG ACGGTATTGG TACTCACTAT TGTGACGCCG GTTGAATGAA GACTGTTGCT AGCCTCCTGG

2451 GAAGGAGCTA ACCGCTTTT TGCACAACAT GGGGATCAT GTAACTCGCC TTGATCGTTG GGAACCGGAG  
CTTCCCTCGAT TGGCGAAAAA ACGTGTGTA CCCCTAGTA CATTGAGCGG AACTAGCAAC CCTTGGCCTC

2521 CTGAATGAAG CCATACCAA CGACGAGCGT GACACCAGA TGCCGTGAGC AATGGCAACA ACGTTGCGCA  
GACTTACTTC GGTA TGTTT GCTGCTCGCA CTGTGGTGCT ACGGACATCG TTACC GTTGT TGCAACGCGT

2591 AACTATT AAC TGGCGAACTA CTTACTCTAG CTTCCCGCA ACAATTATA GACTGGATGG AGCGGATAA  
TTGATAATTG ACCGCTTGAT GAATGAGATC GAAGGCCGT TGTTAATTAT CTGACCTACC TCCGCCATT

2661 AGTTGCAGGA CCACTTCTGC GCTCGGCCCT TCCGGCTGGC TGGTTTATTG CTGATAAATC TGGAGCCGGT  
TCAACGTCTT GGTGAAGACG CGAGCCGGA AGCCGACCG ACCAAATAAC GACTATTAG ACCTCGGCCA

FIG. 43E

2731 GAGCGTGGGT CTCGCGGTAT CATTGCAGCA CTGGGGCCAG ATGGTAAGCC CTCCCGTATC GTAGTTATCT  
CTCGCACCCA GAGCGCCATA GTAAACGTCTGT GACCCCGGTC TACCAATTCGG GAGGCATAG CATCAATAGA

2801 ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAAATAG ACAGATCGCT GAGATAGGTG CCTCACTGAT  
TGTGCTGCC CTCAGTCCGT TGATACCTAC TTGCTTTATC TGTCTAGCGA CTCATATCCAC GGAGTGACTA

2871 TAAGCATTTG TAACTGTCTAG ACCAAGTTTA CTCATATATA CTTTAGATTG ATTTAAAACT TCATTTTTAA  
ATTCTAACC ATTGACAGTC TGGTTCAAAT GAGTATATAT GAAATCTAAC TAAATTTTGA AGTAAAAAAT

2941 TTTAAAAGGA TCTAGGTGAA GATCCTTTT GATAATCTCA TGACCAAAAT CCCTTAACGT GAGTTTTCGT  
AAATTTTCCT AGATCCACTT CTAGGAAAAA CTATTAGAGT ACTGGTTTTA GGGAAATTGCA CTCAAAAAGCA

3011 TCCACTGAGC GTCAGACCCC GTAGAAAAA GTAGAAAGATC TTCTTGAGAT CTTTTTTTC TCGCGGTAAT  
AGGTGACTCG CAGTCTGGG CATCTTTCT AGTTCCCTAG AAGAACTCTA GGAAAAAAG ACGGCAATTA

3081 CTGCTGCTTG CAAACAAAA AACCACCGCT ACCAGCGGTG GTTTGTTTG CGGATCAAGA GCTACCAACT  
GACGACGAAC GTTTGTTTT TTGGTGGCGA TGGTCGCCAC CAAACAACG GCCTAGTTCT CGATGGTTGA

3151 CTTTTCCTGA AGGTAACCTG CTTCAGCAGA GCGCAGATAC CAAATACTGT CTTCTAGTG TAGCCGTAGT  
GAAAAAGGCT TCCATTGACC GAAAGTCGTCT CGCGTCTATG GTTTATGACA GGAAGATCAC ATCGGCATCA

3221 TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCCTACATA CCTCGCTCTG CTAATCCTGT TACCAGTGGC  
ATCCGGTGGT GAAGTCTTG AGACATCGTG GCGGATGTAT GGAGCGAGAC GATTAGGACA ATGGTCACCG

3291 TGCTGCCAGT GCGGATAAGT CGTGTCTTAC CGGGTTGGAC TCAAGACGAT AGTTACCGGA TAAGGCGCAG  
ACGACGGTCA CCGCTATTCA GCACAGAAATG GCCCAACCTG AGTTCIGCTA TCAATGGCCT ATTCCGCGTC

3361 CGGTCGGGCT GAACGGGGG TTCTGTGCACA CAGCCCAGCT TGGAGCGAAC GACCTACACC GAAC TGAGAT  
GCCAGCCCGA CTTGCCCCCC AAGCACGTGT GTCGGGTCGA ACCTCGCTTG CTGGATGTGG CTTGACTCTA

3431 ACCTACAGCG TGAGCTATGA GAAAGCGCCA CGCTTCCCGA AGGGAGAAAG GCGGACAGGT ATCCGGTAAG  
TGGATGTCGC ACTCGATACT CTTTCGCGGT GCGAAGGGCT TCCCCTCTTC CGCCTGTCCA TAGGCCATTC

3501 CGGCAGGGTC GGAACAGGAG AGCGCACGAG GGAGCTTCCA GGGGAAACG CCTGGTATCT TTATAGTCTT  
GCCGTCCCAG CCTGTCTCTC TCGCGTGCTC CCTCGAAGGT CCCCCTTTC GGACCATAGA AATATCAGGA

FIG. 43F

3571 GTCGGGTTTC GCCACCTCTG ACTTGAGCGT CGATTTTGT GATGCTCGTC AGGGGGGCGG AGCCTATGGA  
CAGCCCAAAG CGGTGGAGAC TGAACTCGCA GCTAAAAACA CTACGAGCAG TCCCCCCGCC TCGGATACCT

3641 AAAACGCCAG CAACGCGGCC TTTTTACGGT TCCTGGCCTT TTGCTGCACA TGTTC TTTC  
TTTTCGGTC GTTGCGCCG GAAAA TGCCA AGGACCGGAA AACGACCGGA AACGAGTGT ACAAGAAAGG

3711 TCGGTTATCC CCTGATTCTG TGGATAACCG TATTACCGCC TTTGAGTGAG CTGATACCGC TCGCCGCAGC  
ACGCAATAGG GGACTAAGAC ACCTATTGGC ATAA TGCGG AACTCAGTC GACTATGGCG AGCGCGTCG

3781 CGAACGACCG AGCGAGCGA GTCAGTGAGC GAGGAAGCGG AAGAGCGCCC AATACGCAA CCGCCTCTCC  
GCTTGCTGGC TCGCGTCGCT CAGTCACTCG CTCC TTCGCC TTCTCGCGG TTATGCGTTT GGCGGAGAGG

3851 CCGCGCGTTG GCCGATTCT TAATGCAGCT GGCACGACAG GTTTC CCGAC TGGAAAGCGG GCAGTGAGCG  
GGCGGCAAC CGGCTAAGTA ATTACGTCGA CCGTGCTGTC CAAAGGCTG ACCTTTCGCC CGTCACTCGC

3921 CAACGCAATT AATGTAGTT AGTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT  
GTTGCGTTAA TTACACTCAA TCGAGTGAGT AATCCGTGGG GTCCGAAATG TGAAATACGA AGGCCGAGCA

BssHII  
~~

3991 ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC  
TACAACACAC CTTAACACTC GCCTATTGTT AAAGTGTGTC CTTTGTGAT ACTGGTACTA ATCGGTTTCG

BssHII  
~~~~

4061 GCGCAATTAA CCTCACTAA AGGGAACAAA AGCTGGAATT CCACAATGAA CAATAATAAG ATTAAAAATAG
CGCGTTAATT GGGAGTGATT TCCCTTGTTT TCGACCTTAA GGTGTTACTT GTTATTATTC TAAATTTATC

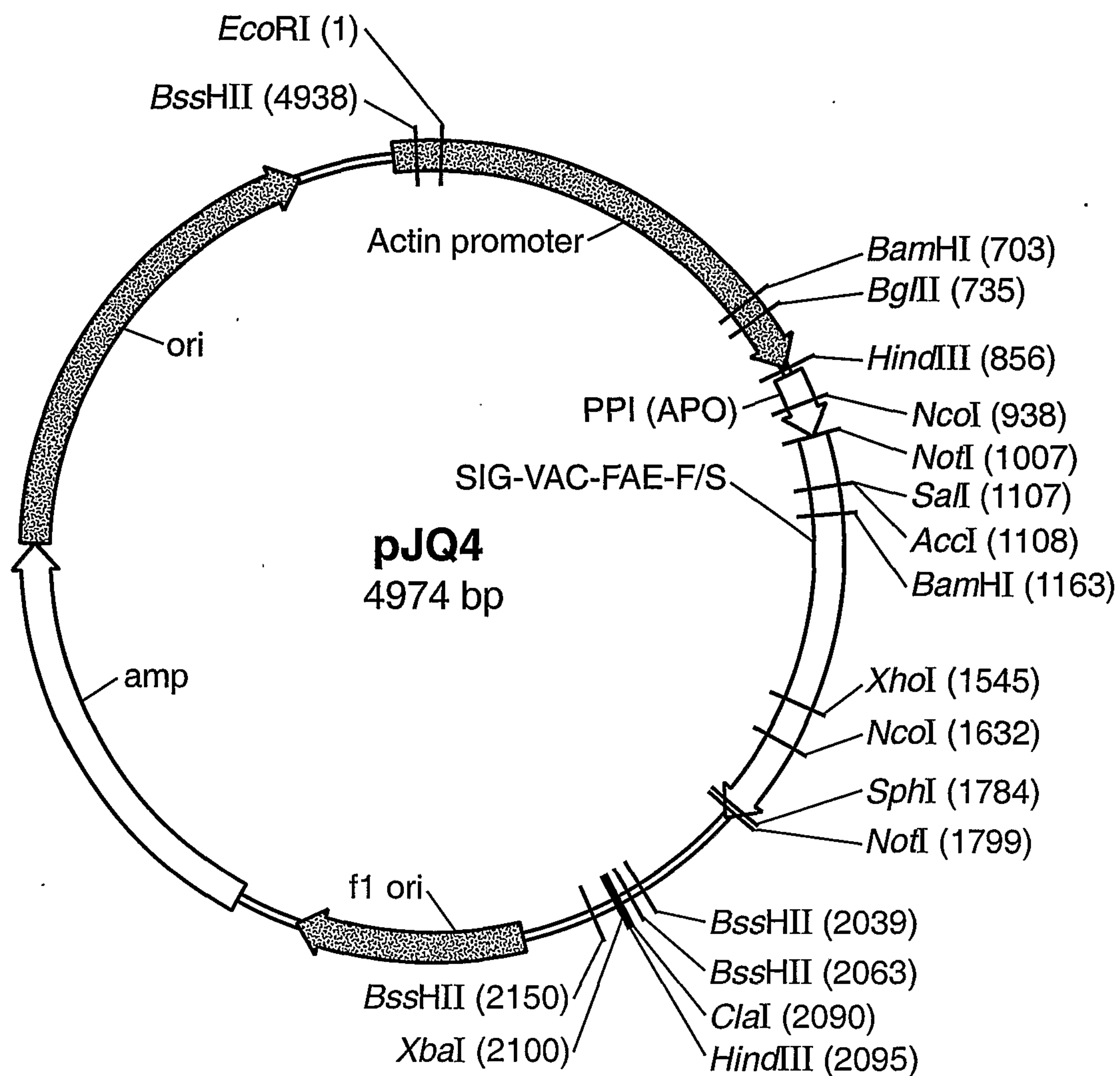
4131 CTTGCCCCCG TTGCAGCGAT GGGTATTTT TCTAGTAAA TAAAAGATAA ACTTAGACTC AAAACATTTA
GAACGGGGC AACGTCGCTA CCCATAAAA AGATCATTTT ATTTTCTATT TGAATCTGAG TTTTGTAAT

4201 CAAAAACAAC CCTAAAGTC CTAAAGCCCA AAGTGCTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA
GTTT TTGTG GGGATTTCAG GATTTCGGGT TTCACGATAC GTGCTAGGTA TCGTTCGGGT CGGGTIGGGT

FIG._ 43G

| | | | | | | | |
|-------|-------------|-------------|------------|------------|------------|-------------|-------------|
| 4271 | ACCCAACCCA | ACCCACCCCA | GTGCAGCCAA | CTGGCAAATA | GTCTCCACCC | CCGGCACTAT | CACCGTGAGT |
| | TGGGTTGGGT | TGGGTGGGT | CACGTCGGTT | GACCGTTTAT | CAGAGGTGGG | GGCCGTGATA | GTGGCACTCA |
| 4341 | TGTCCGCACC | ACCGCACGTC | TCGCAGCCAA | AAAAAAAAAA | AGAAAGAAAA | AAAAAGAAAA | GAAAAACAGC |
| | ACAGGCGTGG | TGGCGTGCAG | AGCGTCGGTT | TTTTTTT | TCCTTCTTT | TTTTCTTTT | CTTTTGTCTG |
| 4411 | AGGTGGGTCC | GGGTCGTGGG | GGCCGGAAAA | CGAGGAGGA | TCGCGAGCAG | CGACGAGGCC | CGGCCCTCCC |
| | TCCACCCAGG | CCAGCACCC | CCGGCCTTT | CGCTCCTCCT | AGCGCTCGTC | GCTGCTCCGG | GCCGGGAGGG |
| 4481 | TCCGCTTCCA | AAGAAACGCC | CCCATCGCC | ACTATATACA | TACCCCCCC | TCTCCTCCCA | TCCCCCCAAC |
| | AGCGGAAGGT | TTCCTTGCGG | GGGTAGCGG | TGATATATGT | ATGGGGGGG | AGAGGAGGGT | AGGGGGTTG |
| 4551 | CCTACCACCA | CCACCACCAC | CACCTCCTCC | CCCCTCGCTG | CCGGACGACG | AGCTCCTCCC | CCCTCCCCCT |
| | GGAATGGTGGT | GGTGGTGGTG | GTGGAGGAGG | GGGAGCGAC | GGCCTGCTGC | TCGAGGAGGG | GGGAGGGGGA |
| 4621 | CCGCCGCCGC | CGGTAACCAC | CCCGCCCTC | TCCTCTTTCT | TTCTCCGTTT | TTTTTTTCGT | CTCGGTCCTG |
| | GGCGCGGCG | GCCATTGGTG | GGCGGGGAG | AGGAGAAAGA | AAGAGGCAAA | AAAAAAGCA | GAGCCAGAGC |
| 4691 | ATCTTTGGCC | TTGGTAGTTT | GGTGGGCGA | GAGCGGCTTC | GTGCCCCAGA | TCGGTGCGCG | GGAGGGCGGG |
| | TAGAAACCGG | AACCATCAAA | CCCACCCGCT | CTCGCCGAAG | CAGCGGGTCT | AGCCACGCGC | CCTCCCCGCC |
| BglII | | | | | | | |
| ~~~~~ | | | | | | | |
| 4761 | GATCTCGCGG | CTGGCGTCTC | CGGGCGTGAG | TCGGCCCCGA | TCCTCGCGGG | GAATGGGGCT | CTCGGATGTA |
| | CTAGAGCGCC | GACCGCAGAG | GCCCGCACTC | AGCCGGGCT | AGGAGCGCCC | CTTACCCCGA | GAGCCTACAT |
| BglII | | | | | | | |
| ~~~~~ | | | | | | | |
| 4831 | GATCTTCTTT | CTTCTTCTTT | TTTGTGGTAG | AATTGAATC | CCTCAGCATT | GTTCAATCGGT | AGTTTTTCTT |
| | CTAGAAGAAA | GAAAGAAAGAA | AAACACCATC | TTAAACTTAG | GGAGTCGTAA | CAAGTAGCCA | TCAAAAAAGAA |
| 4901 | TTCAATGATTT | GTGACAAATG | CAGCCTCGTG | CGGAGCTTTT | TTGTAGGTAG | | |
| | AAGTACTAAA | CACTGTTTAC | GTCGGAGCAC | GCCTCGAAAA | AACATCCATC | | |

FIG. 43H

**FIG._44A**

| | | | | | | | | | |
|-------|-------------|------------|------------|------------|-------------|-------------|-------------|--|--|
| EcoRI | | | | | | | | | |
| ~~~~~ | | | | | | | | | |
| 1 | AATCCACAA | TGAACAATAA | TAAGATTAAA | ATAGCTTGCC | CCCGTTGCAG | CGATGGGTAT | TTTTTCTAGT | | |
| | TTAAGGTGTT | ACTTGTTATT | ATTCTAAATT | TATCGAACGG | GGCAACGTC | GCTACCCATA | AAAAGATCA | | |
| 71 | AAAATAAAAG | ATAAACTTAG | ACTCAAAACA | TTTACAAAAA | CAACCCCTAA | AGTCCTAAAG | CCCAAAGTGC | | |
| | TTTTATTTC | TATTGAATC | TGAGTTTGT | AAATGTTTT | GTTGGGATT | TCAGGATTTC | GGTTTCACG | | |
| 141 | TATGCACGAT | CCATAGCAAG | CCCAGCCCAA | CCCAACCCAC | CCAGTGCAG | CCACTGGCA | | | |
| | ATACGTGCTA | GGTATCGTTC | GGTCCGGTT | GGTTGGGT | GGTCACTC | GGTGACCGT | | | |
| 211 | AATAGTCTCC | ACCCCCGGCA | CTATCACCGT | GAGTTGTCCG | CACCACCGCA | CGTCTCGCAG | CCAAAAAAA | | |
| | TTATCAGAGG | TGGGGCCGT | GATAGTGCA | CTCAACAGGC | GTGGTGGCGT | GCAGAGCGTC | GGTTTTTT | | |
| 281 | AAAAGAAAG | AAAAAAAGA | AAAAGAAAAA | CAGCAGGTGG | GTCCGGGTCCG | TGGGGGCCGG | AAAAGCGAGG | | |
| | TTTTTCTTTC | TTTTTTTCT | TTTTCTTTT | GTCGTCCACC | CAGGCCCAGC | ACCCCGGCC | TTTTTCGCTCC | | |
| 351 | AGGATCGCGA | GCAGCGACGA | GGCCCGGCC | TCCCTCCGCT | TCCAAAGAAA | CGCCCCCAT | CGCCACTATA | | |
| | TCCTAGCGCT | CGTCCTGCT | CCGGCCGGG | AGGAGGCGA | AGGTTTCTTT | CGGGGGGTA | CGGTGATAT | | |
| 421 | TACATACCCC | CCCTCTCCT | CCCATCCCC | CAACCCTACC | ACCACACCA | CCACCACCTC | CTCCCCCTC | | |
| | ATGTATGGG | GGGAGAGGA | GGTAGGGG | GTTGGATGG | TGGTGGTGGT | GGTGGTGGAG | GAGGGGGAG | | |
| 491 | GCTGCCGGAC | GACGAGCTCC | TCCCCCTCC | CCCTCCGCCG | CCGCCGGTAA | CCACCCCGCC | CCTCTCCTCT | | |
| | CGACGGCCTG | CTGCTCGAGG | AGGGGGGAGG | GGAGGCCGC | GGCGCCATT | GGTGGGGCGG | GGAGAGGAGA | | |
| 561 | TTCTTTCTCC | GTTTTTTTTT | TCGTCTCGGT | CTCGATCTTT | GGCCTTGGTA | GTTTGGGTGG | GCAGAGCGG | | |
| | AAGAAAGAGG | CAAAAAAAA | AGCAGAGCCA | GAGCTAGAAA | CCGGAACCAT | CAAAACCCACC | CGCTCTCGCC | | |
| 631 | CTTCGTCGCC | CAGATCGGTG | CGCGGGAGGG | GGGGATCTC | GGGCTGGCG | TCTCCGGGCG | TGAGTCGGCC | | |
| | GAAGCAGCGG | GTCTAGCCAC | CGCCCTCCC | CGCCCTAGAG | CGCCGACCGC | AGAGGCCCGC | ACTCAGCCGG | | |
| BamHI | | | | | | | | | |
| ~~~~~ | | | | | | | | | |
| 701 | CGGATCCCTCG | CGGGGAATGG | GGCTCTCGGA | TGTAGATCTT | CTTCTTTCT | TCTTTTTGTG | GTAGAAATTG | | |
| | GCCTAGGAGC | GCCCTTACC | CCGAGAGCCT | ACATCTAGAA | GAAAGAAAGA | AGAAAAAACAC | CATCTTAAC | | |

FIG. 44B

| | | | | | | | |
|------|--------------------------------|------------|-------------|-------------|-------------|-------------|-------------|
| 771 | AATCCCTCAG | CATTGTTCAT | CGGTAGTTT | TCTTTTCATG | ATTTGTGACA | AATGCAGCCT | CGTGCGGAGC |
| | TTAGGGAGTC | GTAACAAGTA | GCCATCAAAA | AGAAAAGTAC | TAAACACTGT | TTACGTCGGA | GCACGCCCTCG |
| | HindIII
~~~~~ | | | | | | |
| 841 | TTTTTTGTAG | GTAGAAGCTT | ACMATGGMCG | TGCACAAGGA | GGTSAACTTC | GTSGCCCTACC | TCCTGATCGT |
| | AAAAAACATC | CATCTTCGAA | TGKTACCKGC | ACGTGTTTCT | CCASTTGAAG | CASCGGATGG | AGGACTAGCA |
| | NcoI
~~~~~ | | | | | | |
| 911 | SCTCGGCCCTC | CTCTTGCTCG | TSTCCGCCAT | GGAGCAGTG | GACGCCAAGG | CCTGCACCK | CGAGTGCGGC |
| | SGAGCCGGAG | GAGAACGAGC | ASAGCGGTA | CCTCGTGCAC | CTGCGGTTCC | GGACGTGGGM | GCTCACGCCG |
| | NotI
~~~~~ | | | | | | |
| 981 | AACCTCGGCT | TCGGCATCTG | CCCGCGGCC | GCCTCCACGC | AGGGCATCTC | CGAAGACCTC | TACAGCCGTT |
| | TTGGAGCCGA | AGCCGTAGAC | GGCCGCCCG | CGGAGGTGCG | TCCCCTAGAG | GCTTCTGGAG | ATGTCGGCAA |
| | Sall
~~~~~
AccI
~~~~~ | | | | | | |
| 1051 | TAGTCGAAAT | GGCCACTATC | TCCCAAGCTG | CCTACGCCGA | CCTGTGCAAC | ATTCCGTCGA | CTATTATCAA |
| | ATCAGCTTTA | CCGGTGATAG | AGGGTTCGAC | GGATGCCGGCT | GGACACGTTG | TAAGGCAGCT | GATAATAGTT |
| | BamHI
~~~~~ | | | | | | |
| 1121 | GGGAGAGAAA | ATTTACAAAT | CTCAAACTGA | CATTAAACGGA | TGGATCCTCC | GGGACGACAG | CAGCAAAGAA |
| | CCCTCTCTTT | TAAATGTTAA | GAGTTTGACT | GTAATTGCCT | ACCTAGGAGG | CGCTGCTGTC | GTCGTTTCTT |
| 1191 | ATAATCACCG | TCTTCCGTGG | CAC TGGTAGT | GATACGAAATC | TACAAACTCGA | TACTAACTAC | ACCTCACGC |
| | TATTAGTGGC | AGAAGGCACC | GTGACCATCA | CTATGCTTAG | ATGTTGAGCT | ATGATTGATG | TGGGAGTGCG |
| 1261 | CTTTCGACAC | CCTACCACAA | TGCAACGGTT | GTGAAGTACA | CGGTGGATAT | TATATTGGAT | GGGTCTCCGT |
| | GAAAGCTGTG | GGATGGTGTT | ACGTTGCCAA | CAC TTCAATG | GCCACCTATA | ATATAACCTA | CCCAGAGGCA |

FIG.- 44C

| | | | | | | | |
|------|-------------|-------------|-------------|------------|-------------|-------------|-------------|
| 1331 | CCAGGACCAA | GTCGAGTCGC | TTGTCAAACA | GCAGGTTAGC | CAGTATCCGG | ACTACGCGCT | GACCGTGACC |
| | GGTCCTGGTT | CAGCTCAGCG | AACAGTTTGT | CGTCCAATCG | GTCAATAGGC | TGATGCGCGA | CTGGCACTGG |
| 1401 | GGCCACKCCC | TCGGCGCCTC | CCTGGCGGCA | CTCACTGCCG | CCCAGCTGTC | TGCGACATAC | GACAACATCC |
| | CCGGTGMGGG | AGCCGCGGAG | GGACCGCCGT | GAGTGACGGC | GGGTCGACAG | ACGCTGTATG | CTGTTGTAGG |
| 1471 | GCCTGTACAC | CTTCGGCGAA | CCGCGCAGCG | GCAATCAGGC | CTTCGCGTCG | TACATGAACG | ATGCCTTCCA |
| | CGGACATGTG | GAAGCCGCTT | GGCGCGTCGC | CGTTAGTCCG | GAAGCGCAGC | ATGTACTTGC | TACGGAAGGT |
| XhoI | | | | | | | |
| | ~~~~~ | | | | | | |
| 1541 | AGCCTCGAGC | CCAGATACGA | CGCAGTATTT | CCGGGTCACT | CATGCCAACG | ACGGCATCCC | AAACCTGCCC |
| | TCGGAGCTCG | GGTCTATGCT | GGTCAATAA | GGCCAGTGA | GTACGGTTGC | TGCCGTAGGG | TTTGGACGGG |
| NcoI | | | | | | | |
| | | | ~~~~~ | | | | |
| 1611 | CCGGTGGAGC | AGGGGTACGC | CCATGGCCGGT | GTAGAGTACT | GGAGCGTTGA | TCCTTACAGC | GCCAGAACA |
| | GGCCACCCTCG | TCCCCATGCG | GGTACCGCCA | CATCTCATGA | CCTCGCAACT | AGGAATGTCG | CGGGTCTTGT |
| 1681 | CATTGTCTG | CACTGGGGAT | GAAGTGCAGT | GCTGTGAGGC | CCAGGGCGGA | CAGGGTGTGA | ATAATGCGCA |
| | GTAAACAGAC | GTGACCCCTA | CTTCACGTCA | CGACACTCCG | GGTCCC GCCT | GTCCCACACT | TATTACGCCT |
| SphI | | | | | | | |
| | | | ~~~~~ | | | | |
| 1751 | CACGACTTAT | TTTGGGATGA | CGAGCGGCGC | ATGCACCTGG | CCGGTCGCGG | CCGCGGAAC | CAC TGAAGGA |
| | GTGCTGAATA | AAACCC TACT | GCTCGCCGCG | TACGTGGACC | GGCCAGCGCC | GGCGCCTTTG | GTGACTTCCT |
| 1821 | TGAGCTGTAA | AGAAGCAGAT | CGTTCAAACA | TTTGGCAATA | AAGTTTCTTA | AGATTGAATC | CTGTTGCCGG |
| | ACTCGACATT | TCTTCGTCTA | GCAAGTTTGT | AAACCGTTAT | TTCAAAGAAT | TCTAACTTAG | GACAAACGGCC |
| 1891 | TCTTGGGATG | ATTATCATAT | AATTTCTGTT | GAATTACGTT | AAGCATGTAA | TAAATTAACAT | GTAATGCATG |
| | AGAACGCTAC | TAAATAGTATA | TTAAAGACAA | CTTAATGCAA | TTCGTACATT | ATTAATTGTA | CATTACGTAC |
| 1961 | ACGTTATTTA | TGAGATGGGT | TTTTATGATT | AGAGTCCCGC | AATTATACAT | TTAAATACGC | ATAGAAAACA |
| | TGCAATAAAT | ACTCTACCCA | AAAATACTAA | TCTCAGGGCG | TTAATATGTA | AATTATGCGC | TATCTTTTGT |

FIG._ 44D

FIG. 44E

2731 TTTAACGCGA ATTTTAACAA AATATTAACG CTTACAATTT AGGTGGCACT TTTCGGGGAA ATGTGCGCGG
AAATTGCGCT TAAAATTGTT TTATAATTGC GAATGTTAAA TCCACCGTGA AAAGCCCCCTT TACACGCGCC

2801 AACCCCTATT TGTTTATTTT TCTAAATACA TTCAAATATG TATCCGCTCA TGAGACAATA ACCCTGATAA
TTGGGGATAA ACAAAATAAA AGATTTATGT AAGTTTATAC ATAGCGGAGT ACTCTGTTAT TGGGACTATT

2871 ATGCTTCAAT AATATTGAAA AAGGAAGAGT ATGAGTATTC AACATTCCG TGTCCGCCCTT ATTCCCTTTT
TACGAAGTTA TTATAACTTT TTCCCTTCTCA TACTCATAAG TTGTAAGGC ACAGCGGGAA TAAGGGAAAA

2941 TTGCGGCATT TTGCCCTTCCT GTTTTGTCT ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA
AACGCCGTAA AACGGAAGGA CAAAACGAG TGGGTCCTTG CGACCACCTT CATTTTCTAC GACTTCTAGT

3011 GTTGGGTGCA CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC
CAACCCACGT GCTCACCCAA TGTAGCTTGA CCTAGAGTTG TCGCCATCT AGGAACCTCTC AAAAGCGGGG

3081 GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTAATTATCC CGTATTGACG
CTTC TTGCAA AAGTTACTA CTCGTGAAA CTCGTAAGAGT ATACACCGCG CCATAATAGG GCATAACTGC

3151 CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC
GGCCCGTTCT CGTTGAGCCA GCGGCGTATG TGATAAGAGT CTTACTGAAC CAACTCATGA GTGGTCAGTG

3221 AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA TGCAGTGCTG CCATAAACCAT GAGTGATAAC
TC TTTTCGTA GAATGCC TAC CGTACTGTCA TTC TCTTAAT ACGTCACGAC GGTA TTGGTA CTCAC TATTG

3291 ACTGCGGCCA ACTTACTTCT GACAACGATC GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG
TGACGCCGGT TGAATGAAGA CTGTTGCTAG CCTCCTGGCT TCCTCGATTG GCGAAAAAAC GTGTGTACC

3361 GGGATCATGT AACTCGCCTT GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA
CCCTAGTACA TTGAGCGGAA CTAGCAACCC TTGGCCCTCGA CTTACTTCCG TATGGTTTGC TGCTCGCACT

3431 CACCACGATG CCTGTAGCAA TGGCAACAAC GTTGCGCAA CTATTAACTG GCGAACTACT TACTCTAGCT
GTGGTGCTAC GGACATCGTT ACCGTTGTTG CAACGCGTTT GATAATTGAC CGCTTGATGA ATGAGATCGA

3501 TCCCGGCAAC AATTAAAGA CTGGATGGAG GCGGATAAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC
AGGCCCGTTG TTAATTATCT GACCTACCTC GCCTATTTC AACGTCCTGG TGAAGACGCG AGCCGGGAAG

FIG. 44F

3571 CGGCTGGCTG GTTTATTGCT GATAAATCTG GAGCCGGTGA GCGTGGGTCT CGCGGTATCA TTGCAGCACT
GCCGACCGAC CAAATAACGA CTATTTAGAC CTCGGCCACT CGCACCCAGA GCGCCATAGT AACGTCGTGA

3641 GGGGCCAGAT GGTAAGCCCT CCCGTATCGT AGTTATCTAC ACGACGGGGA GTCAGGCAAC TATGGATGAA
CCCCGGTCTA CCAATCGGGA GGCATAGCA TCAATAGATG TGCTGCCCCCT CAGTCCGTTG ATACCTACTT

3711 CGAAATAGAC AGATCGCTGA GATAGGTGCC TCACTGATTA AGCATTGGTA ACTGTCAGAC CAAGTTTACT
GCTTTATCTG TCTAGCGACT CTATCCACGG AGTGACTAAT TCGTAACCAT TGACAGTCTG GTTCAAAATGA

3781 CATATATACT TTAGATTGAT TTTAAACTTC ATTTTTAATT TAAAAGGATC TAGGTGAAGA TCCTTTTTGA
GTATATATGA AATCTAACTA AATTTTGAAG TAAAAATTAA ATTTTCCTAG ATCCACTTCT AGGAAAAACT

3851 TAATCTCATG ACCAAAATCC CTTAAACGTGA GTTTTCGTTC CACTGAGCGT CAGACCCCGT AGAAAAGATC
ATTAGAGTAC TGGTTTTAGG GAATTGCACT CAAAAGCAAG GTGACTCGCA GTC TGGGCA TCTTTTCTAG

3921 AAAGGATCTT CTTGAGATCC TTTTTTCTG CGCGTAATCT GCTGCTTGCA AACAAAAAA CCACCGCTAC
TTTCCCTAGAA GAACTCTAGG AAAAAAGAC GCGCATTAGA CGACGAACGT TTGTTTTTT GGTGGCGATG

3991 CAGCGGTGGT TTGTTTGCCG GATCAAGAGC TACCAACTCT TTTTCCGAAG GTAACTGGCT TCAGCAGAGC
GTCGCCACCA AACAAACGGC CTAGTTCTCG ATGGTTGAGA AAAAGGCTTC CATTGACCGA AGTCGTCTCG

4061 GCAGATACCA AATACTGTCC TTCTAGTGTA GCCGTAGTTA GGCCACCACT TCAAGAACTC TGTAGCACCG
CGTCTATGGT TTATGACAGG AAGATCACAT CGGCATCAAT CCGGTGGTGA AGTTC TTGAG ACATCGTGGC

4131 CCTACATACC TCGCTCTGCT AATCCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTCG TGTCTTACCG
GGATGTATGG AGCGAGACGA TTAGGACAAAT GTTCACCGAC GACGGTCACC GCTATTCAGC ACAGAAATGGC

4201 GGTTTGGACTC AAGACGATAG TTACCGGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGTT CGTGACACA
CCAAACCTGAG TTCTGCTATC AATGGCCTAT TCCGCTCGC CAGCCCGACT TGCCCCCAA GCACGTGTGT

4271 GCCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA AAGGCCACG
CGGGTCGAAC CTCGCTTGCT GGATGTGGCT TGACTCTATG GATGTCGCAC TCGATACTCT TTCGCGGTGC

4341 CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCGG AACAGGAGAG CGCACGAGGG
GAAGGGCTTC CCTCTTTCCG CCTGTCCATA GGCCATTTCG CGTCCCAGCC TTGTCCCTCTC GCGTGTCTCC

FIG. 44G

4411 AGCTTCCAGG GGGAAACGCC TGGTATCTTT ATAGTCCCTGT CGGGTTTCGC CACCTCTGAC TTGAGCGTCG
TCGAAGGTCC CCTTTTGCGG ACCATAGAAA TATCAGGACA GCCCAAAGCG GTGGAGACTG AACTCGCAGC

4481 ATTTTGTGA TGCTCGTCAG GGGGCGGAG CCTATGAAA AACGCCAGCA ACGCGCCTT TTTACGGTTC
TAAAAACACT ACGAGCAGTC CCCCGCCTC GGATACCTTT TTGCGGTCGT TGCGCCGGA AAATGCCAAG

4551 CTGGCCTTT GCTGGCCTT TGTCACATG TTCCTTCCCTG CGTTATCCC TGAATCTGTG GATAACCGTA
GACCGGAAA CGACCGGAAA ACGAGTGTA AAGAAAGGAC GCAATAGGG ACTAAGACAC CTAATTGGCAT

4621 TTACCGCCTT TGAGTGAGCT GATACCGCTC GCCGCAGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA
AATGGCGGAA ACTCACTCGA CTATGGCGAG CGCGTGGCTC TTGCTGGCTC GCGTGGCTCA GTCACCTCGCT

4691 GGAAGCGGAA GAGCGCCCAA TACGCAAACC GCCTCTCCC GCGCGTTGGC CGATTCAATTA ATGCAGCTGG
CCTTCGCCCTT CTCGGGGTT ATCGGTTTGG CGGAGAGGGG CCGGCAACCG GCTAAGTAAT TACGTCGACC

4761 CACGACAGGT TTCCCGACTG GAAAGCGGC AGTGAGCGCA ACGCAATTA TGTGAGTTAG CTCACTCATT
GTGCTGTCCA AAGGGCTGAC CTTTCGCCCG TCACTCGCGT TGCGTTAATT ACACTCAATC GAGTGAGTAA

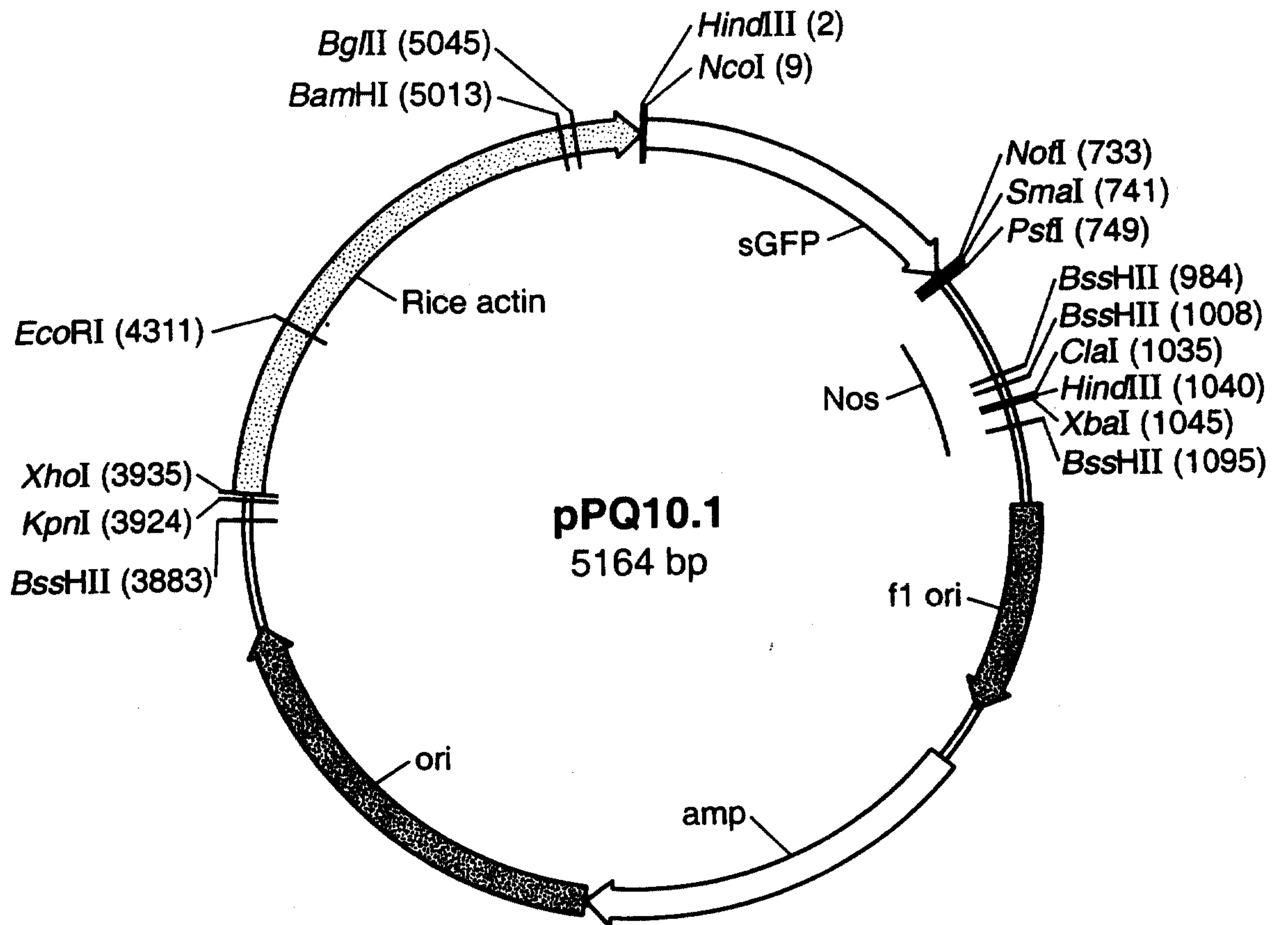
4831 AGGCACCCCA GGCATTACAC TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTGAGCG GATAACAAAT
TCCGTGGGGT CCGAAATGTG AAATACGAAG GCCGAGCATA CAACACACCT TAACACTCGC CTATTGTTAA

4901 TCACACAGGA AACAGCTATG ACCATGATTA CGCCAAGCGC GCAATTAAAC CTCACTAAAG GGAACAAAAG
AGTGTGTCCCT TTGTCGATAC TGGTACTAAT GCGGTTGCGG CGTTAATTGG GAGTGATTTC CCTTGTTTTC

EcoR
4971 CTGG
GACC

BssHII
~~~~~

FIG.- 44H

**FIG. 45A**

HindIII NgoI  
~~~~~ ~~~~~~  
1 AAGCTTACCA TGGTGAGCAA GGGCGAGGAG CTGTTACCG GGGTGGTGCC CATCTGGTC GAGCTGGACG
TTCGAAATGGT ACCACTCGTT CCCGCTCCTC GACAAGTGGC CCCACCACGG GTAGGACCAG CTCGACCCTGC
71 GCGACGTGAA CGGCCACAAG TTCAGCGTGT CCGGCGAGGG CGAGGGCGGAT GCCACCCTACG GCAAGCTGAC
CGCTGCACCTT GCCGGTGTTC AAGTCGCACA GGCCGCTCCC GCTCCCGCTA CGGTGGATGC CGTTCGACTG
141 CCTGAAGTTC ATCTGCACCA CCGGCAAGCT GCCCGTGCCC TGGCCACCC TCGTGACCAC CTTACCTTAC
GGACTTCAAG TAGACGTGGT GGCCGTTCGA CCGGCACGGG ACCGGGTGGG AGCACTGGTG GAAAGTGGATG
211 GGCGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCAGGACTT CTTCAAGTCC GCCATGCCCG
CCGCACGTCA CGAAGTCGGC GATGGGGCTG GTGTACTTCG TCGTGCTGAA GAAGTTCAGG CCGTACGGGC
281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC AAGACCCGCG CCGAGGTGAA
TTCCGATGCA GGTCTTCGCG TGGTAGAAGA AGTTCCTGCT GCCGTTGATG TTCTGGGCGC GGCTCCACTT
351 GTTCGAGGGC GACACCCCTGG TGAACCGCAT CGAGCTGAAG GGCATCGACT TCAAGGAGGA CGGCAACATC
CAAGCTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA AGTTCCCTCCT GCCGTTGTAG
421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT GGCCGACAAG CAGAAGAACG
GACCCCGTGT TCGACCCTCAT GTTGATGTTG TCGGTGTTGC AGATATAGTA CCGGCTGTTT GTCTTCTTGC
491 GCATCAAGGT GAACTTCAAG ATCCGCCACA ACATCGAGGA CGGCAGCGTG CAGCTCGCCG ACCACTACCA
CGTAGTTCCA CTTGAAGTTC TAGGCGGTGT TGTAGCTCCT GCCGTCGCAC GTCGAGCGGC TGGTGATGGT
561 GCAGAACACC CCATCGGGC ACGGCCCCGT GCTGCTGCC GACAACCACT ACCTGAGCAC CCAGTCCGCC
CGTCTTGTGG GGTAGCCGC TGCCGGGGCA CGACGACGGG CTGTTGGTGA TGGACTCGTG GGTACAGCGG
631 CTGAGCAAAG ACCCAACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT CGTGACCGCC GCCGGGATCA
GACTCGTTTC TGGGGTTGCT CTTCCGCGCTA GTGTACCAGG ACGACCTCAA GCACTGGCGG CGGCCCTAGT

FIG. 45B

```

                                SmaI
                                ~~~~~
                                NotI      PstI
                                ~~~~~

701 CTCACGGCAT GGACGAGCTG TACAAGTAA GCGGCCGCC GGGCTGCAGG GAAACCACTG AAGATGAGC
   GAGTGCCGTA CCTGCTCGAC ATGTTCAATT CGCCGGCGGG CCCGACGTCC CTTTGGTGAC TTCCTACTCG

771 TGTAAAGAAG CAGATCGTTC AACATTTGG CAATAAAGTT TCTTAAGATT GAATCCTGTT GCCGGTCTTG
   ACATTTCTTC GTCTAGCAAG TTTGTAAACC GTTATTTCAG AGAATTCTAA CTTAGGACAA CGGCCAGAAC

841 CGATGATTAT CATATAATT CTGTTGAATT ACGTTAAGCA TGTAAATAAT AACATGTAAT GCATGACGTT
   GCTACTAATA GTATATTAA GACAACCTAA TGCAATTTCGT ACATTATTAA TTGTACATTA CGTACTGCAA

911 ATTTATGAGA TGGGTTTTTA TGATTAGAGT CCCGCAATTA TACATTTAAT ACGCGATAGA AAACAAATA
   TAAATACTCT ACCCAAAAT ACTAATCTCA GGGCGTTAAT ATGTAAATA TGCGCTATCT TTTGTTTTAT

                                XbaI
                                ~~~~~

                                ClaI HindIII
                                ~~~~~

981 TAGCGCGCAA ACTAGGATAA ATTATCGCGC GCGGTGTCAT CTATGTACT AGATCGATAA GCTTCTAGAG
   ATCGCGCGTT TGATCCTATT TAATAGCGCG CGCCACAGTA GATACAATGA TCTAGCTATT CGAAGATCTC

                                BssHII
                                ~~~~~

                                BssHII
                                ~~~~~

1051 CGGCCGGTGG AGCTCCAATT CGCCCTATAG TGAGTCGTAT TACGCGCGCT CACTGGCCGT CGTTTACAA
   GCCGGCCACC TCGAGGTTAA GCGGGATATC ACTCAGCATA ATGCGCGCGA GTGACCGGCA GCAAAATGTT

1121 CGTCGTGACT GGGAAACCC TGGCGTTACC CAACTTAATC GCCTTGCAGC ACATCCCCCT TTCGCCAGCT
   GCAGCACTGA CCCTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCTG TGTAGGGGGA AAGCGGTCGA

1191 GGCGTAAATAG CGAAGAGGCC CGCACCGATC GCCCTTCCCA ACAGTTGCGC AGCCTGAATG GCGAATGGGA
   CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAGGGT TGTCACCGCG TCGGACTTAC CGCTTACCCCT
```

FIG.. 45C

IPED 10 JUN 2002

1261 CGCGCCCTGT AGCGGCGCAT TAAGCGCGG GGGTGTGGTG GTTACGCGCA GCGTGACCGC TACACTTGCC
GCGCGGGACA TCGCCGCGTA ATTGCGGCCG CCCACACCAC CAATGCGCGT CGCACTGGCG ATGTGAACGG

1331 AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCCGCG TTTCGCCGTC
TCGCGGGATC GCGGGCGAGG AAGCGAAAG AAGGGAAGGA AAGAGCGGTG CAAGCGGCCG AAAGGGGCAG

1401 AAGCTCTAAA TCGGGGGCTC CCTTTAGGGT TCCGATTAG TCGTTTACGG CACCTCGACC CCAAAAACCT
TTCGAGATTT AGCCCCCGAG GGAATCCCA AGGCTAATC ACGAAATGCC GTGGAGCTGG GGTTTTGA

1471 TGATTAGGT GATGGTTCAC GTAGTGGCC ATCGCCCTGA TAGACGGTTT TTCGCCCTTT GACGTTGGAG
ACTAATCCCA CTACCAAGTG CATCACCCGG TAGCGGGACT ATCTGCCAAA AAGCGGAAA CTGCAACCTC

1541 TCCACGTTCT TTAATAGTGG ACTCTTGTTT CAACTGGAA CAACACTCAA CCTATCTCG GTCTATTCTT
AGGTGCAAGA AATTATCACC TGAGAACAAG GTTTGACCTT GTTGTGAGTT GGGATAGAGC CAGATAAGAA

1611 TTGATTATA AGGATTTTG CCGATTCGG CCTATTGGTT AAAAAATGAG CTGATTTAAC AAAAATTAA
AACTAAATAT TCCCTAAAC GGCTAAAGCC GGATAACCAA TTTTTTTACTC GACTAAATTG TTTTFAAATT

1681 CGCGAATTTT AACAAAATAT TAACGCTTAC AATTAGGTG GCACTTTTCG GGGAAATGTG CGCGGAACCC
GCGCTTAAA TTGTTTATA ATTGCGAATG TTAAATCCAC CGTGAAAAGC CCTTTTACAC GCGCCTTGGG

1751 CTATTTGTTT ATTTTCTAA ATACATTCAA ATATGTATCC GCTCATGAGA CAATAACCTT GATAAATGCT
GATAAACAAA TAAAAGATT TATGTAAGTT TATACATAGG CGAGTACTCT GTTATTGGGA CTATTACGA

1821 TCAATAATAT TGAAAAAGGA AGAGTATGAG TATCAACAT TTCCGTGTG CGCTTATTC CTTTTTTCG
AGTTATTATA ACTTTTTCCT TCTCATCTC ATAAGTTGTA AAGGCACAGC GGGAAATAAG GAAAAAACGC

1891 GCATTTTGCC TTCCGTGTTT TGCTCACCCA GAAACGCTGG TGAAAGTAA AGATGCTGAA GATCAGTTGG
CGTAAACCG AAGGACAAA ACGAGTGGGT CTTTGGGACC ACTTTCATTT TCTACGACTT CTAGTCAACC

1961 GTGCACGAGT GGGTACATC GAACTGGATC TCAACAGCGG TAAGATCCTT GAGAGTTTC GCCCGAAGA
CACGTGCTCA CCCAATGTAG CTTGACCCTAG AGTTGTGCGC ATCTAGGAA CTCTCAAAAG CGGGGCTTCT

2031 ACGTTTCCA ATGATGAGCA CTTTAAAGT TCTGCTATGT GCGCGGTAT TATCCCGTAT TGACGCCGGG
TGCAAAAGGT TACTACTCGT GAAATTTCA AGACGATACA CCGGCCATA ATAGGCATA ACTGCGGCC

FIG.- 45D

10 JUN 2002

2101 CAAGAGCAAC TCGGTCGCCG CATACTAT TCTCAGAAATG ACTTGGTTGA GTACTACCA GTCACAGAAA
GTTCTCGTTG AGCCAGCGGC GTATGTGATA AGAGTCTTAC TGAACCAACT CATGAGTGGT CAGTGTCTTT

2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA ACCATGAGTG ATAACACTGC
TCGTAGAAATG CCTACCGTAC TGTCAATTCTC TTAATACGTC ACGACGGTAT TGGTACTCAC TATTGTGACG

2241 GGCCAACTTA CTTCTGACAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT TTTTGCACAA CATGGGGGAT
CCGGTTGAAT GAAAGACTGTT GCTAGCCTCC TGGCTTCCTC GATGGCGAA AAAACGTGTT GTACCCCTTA

2311 CATGTAACTC GCCTTGATCG TTGGGAACCG GAGCTGAATG AAGCCATACC AAACGACGAG CGTGACACCA
GTACATTGAG CGGAACCTAGC AACCTTGGC CTCGACTTAC TTCGGTATGG TTTGCTGCTC GCACTGTGGT

2381 CGATGCCCTGT AGCAATGGCA ACAACGTTGC GCAAACATAA AACTGGCGAA CTACTTACTC TAGCTTCCCG
GCTACGGACA TCGTTACCGT TGTGCAACG CGTTGATAA TTGACCGCTT GATGAATGAG ATCGAAGGGC

2451 GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACCTC TGCCTCGGC CCTTCCGGCT
CGTTGTTAAT TATCTGACCT ACCTCCGCCT ATTTCAACGT CCTGGTGAAG ACGCGAGCCG GGAAGGCCGA

2521 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCGG TATCATTGCA GCCTGGGGC
CCGACCAAT AACGACTATT TAGACCTCGG CCACTCGCAC CCAGAGCGCC ATAGTAACGT CGTGACCCCG

2591 CAGATGGTAA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGAGTCAAG GCAACTATGG ATGAACGAAA
GTCTACCAAT CCGGAGGCA TAGCATCAAT AGATGTGCTG CCCCTCAGTC CGTTGATACC TACTTGCTTT

2661 TAGACAGATC GCTGAGATAG GTGCCCTCACT GATTAAGCAT TGGTAACTGT CAGACCAAGT TTACTCATAT
ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTCGTA ACCATTGACA GTCTGGTTCA AATGAGTATA

2731 ATACTTTAGA TTGATTTAAA ACTTCATTTT TAAATTAAA GGATCTAGGT GAAGATCCTT TTTGATAATC
TATGAAAATCT AACTAAAATT TGAAGTAAAA ATTAAATTTT CCTAGATCCA CTTCTAGGAA AACTATTAG

2801 TCATGACCAA AATCCCTTAA CGTGAGTTT CGTTCACCTG AGCGTCAGAC CCCGTAGAAA AGATCAAGG
AGTACTGGTT TTAGGGAAAT GCACTCAAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTTT TCTAGTTTCC

2871 ATCTTCTTGA GATCCTTTTT TTCTGCGCGT AATCTGCTGC TTGCAACAA AAAACCAACC GCTACCAGCG
TAGAAGAACT CTAGGAAAAA AAGACGCGCA TTAGACGACG AACGTTTGTT TTTTGGTGG CGATGGTCCG

FIG. 45E

121 / 154

2941 GTGGTTTGTT TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA
CACCAAACAA ACGGCCCTAGT TCTCGATGGT TGAGAAAAAG GCTTCCATTG ACCGAAGTCG TCTCGCGTCT

3011 TACCAAATAC TGTCCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACCTCAAG AACTCTGTAG CACCGCCTAC
ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GTGAAGTTC TTGAGACATC GTGGCGGATG

3081 ATACCTCGCT CTGCTAATCC TGTACCAGT GGCTGCTGCC AGTGGCGATA AGTCGTGTCT TACCGGGTTG
TATGGAGCGA GACGATTAGG ACAATGGTCA CCGACGACGG TCACCGCTAT TCAGCACAGA ATGGCCCCAAC

3151 GACTCAAGAC GATAGTTACC GGATAAGCG CAGCGGTCGG GCTGAACGGG GGGTTCGTGC ACACAGCCCA
CTGAGTTCTG CTATCAATGG CCTATTCCGC GTCGCCAGCC CGACTTGCCC CCCAAGCACG TGTGTCGGGT

3221 GCTTGGAGCG AACGACCTAC ACCGAACCTA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC
CGAACCTCGC TTGCTGGATG TGGCTTGACT CTATGGATGT CGCACTCGAT ACTCTTTCGC GGTGCGAAGG

3291 CGAAGGGAGA AAGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT
GCTTCCCTCT TTCCGCCCTGT CCATAGGCCA TTCGCCGTCC CAGCCTTGTC CTCTCGCGTG CTCCCTCGAA

3361 CCAGGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTGCGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT
GGTCCCCCTT TCGGGACCAT AGAAATATCA GGACAGCCCA AAGCGGTGGA GACTGAATC GCAGCTAAAA

3431 TGTGATGCTC GTCAGGGGG GAGAGCCTAT GGAAAAACGC CAGCAACGCG GCCTTTTAC GGTTCCTGGC
ACACTACGAG CAGTCCCCC GCCTCGGATA CCTTTTTCG GTCGTGCGC CGGAAAAATG CCAAGGACCG

3501 CTTTGTGCTG CCTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCTGAT TCGTGGATAA CCGTATTACC
GAAAACGACC GGAAAAACGAG TGTACAAGAA AGGACGCAAT AGGGGACTAA GACACCTATT GGCATAATGG

3571 GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG
CGGAAACTCA CTCGACTATG GCGAGCGGCG TCGGCTTGCT GGCTCGCGTC GCTCAGTCAC TCGTCCCTTC

3641 CGGAAGAGCG CCCAATACGC AAACCGCCTC TCCCCCGCGG TTGGCCGATT CATTAAATGCA GCTGGCACGA
GCCTTCTCGC GGGTTATGCG TTGCGCGGAG AGGGCGCGC AACCGGCTAA GTAAATTACGT CGACCGTGCT

3711 CAGGTTTCCC GACTGGAAG CGGGCAGTGA GCGCAACGCA ATTAATGTGA GTTAGCTCAC TCATTAGGCA
GTCCAAAGGG CTGACCTTTC GCCCGTCACT CGCGTTGCGT TAATTACACT CAATCGAGTG AGTAATCCGT

FIG. 45F

```
3781  CCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT GTGGAATTGT GAGCGGATAA CAATTTACACA
      GGGTCCGAA ATGTGAAATA CGAAGGCCGA GCATACAACA CACCTTAACA CTCGCCATT CTCTAAAGTGT

      BssHII                               KpnI
      ~~~~~                               ~~~
3851  CAGGAAACAG CTATGACCAT GATTACGCA AGCGCGCAAT TAACCCCTCAC TAAAGGGAAC AAAAGCTGGG
      GTCCTTTGTC GATACTGGTA CTAATGCGGT TCGCGCGTTA ATTGGGAGTG ATTTCCCTTG TTTTCGACCC

      KpnI           XhoI
      ~~~~          ~~~~~
3921  TACCGGGCCC CCCCTCGAGG TCATTCATAT GCTTGAGAAG AGAGTCGGGA TAGTCCAAA TAAACAAAG
      ATGGCCCCGG GGGGAGCTCC AGTAAAGTATA CGAACTCTTC TCTCAGCCCT ATCAGGTTT ATTTTGTTTC

3991  GTAAGATTAC CTGGTCAAAA GTGAAAACAT CAGTTAAAG GTGGTATAAG TAAATATCG GTAATAAAG
      CATCTAATG GACCAGTTTT CACTTTTGTA GTCAATTTTC CACCATAATC ATTTATATAGC CATTAATTTTC

4061  GTGGCCCCAA GTGAAATTTA CTCTTTTCTA CTATTATAAA AATTGAGGAT GTTTGTGCG TACTTTGATA
      CACCGGGTTT CACTTTAAAT GAGAAAAGAT GATAATATT TTAACCTCCTA CAAAACAGCC ATGAAACTAT

4131  CGTCATTTTT GTATGAATTG GTTTTAAAGT TTATTCGCGA TTGGGAAATG CATATCTGTA TTTGAGTCGG
      GCAGTAAAAA CATACTTAAC CAAAATTTCA AATAAGCGCT AAACCTTTAC GTATAGACAT AAACCTCAGCC

4201  TTTTAAAGTT CGTTGCTTTT GTAAATACAG AGGATTGTTG ATAAGAAATA TCTTTAAAA ACCCATATGC
      AAAAATTCAA GCAACGAAAA CATTTATGTC TCCCTAAACA TATTCCTTAT AGAAATTTT TGGGTATACG

      EcoRI
      ~~~~~
4271  TAATTTGACA TAATTTTTGA GAAAAATATA TATCAGCGG AATCCACAA TGAACAATAA TAAGATTAAA
      ATTAAACTGT ATTAAAAACT CTTTTATAT ATAAGTCCGC TTAAGGTGTT ACTTGTTATT ATTCTAATTT

4341  ATAGCTTGCC CCCGTTGCAG CGATGGGTAT TTTTCTAGT AAAATAAAG ATAACTTAG ACTCAAAACA
      TATCGAACGG GGGCAACGTC GCTACCCATA AAAAAGATCA TTTTATTTTC TATTGAATC TGAGTTTGT

4411  TTTACAAAAA CAACCCCTAA AGTCCTAAG CCCAAAGTC TATGCAGAT CCATAGCAAG CCCAGCCCAA
      AAATGTTTTT GTTGGGGATT TCAGGATTC GGGTTTCAG ATACGTGCTA GGATCGTTC GGGTCGGGTT
```

FIG.-45G

JUN 2002

4481 CCCAACCCAA CCCAACCCAC CCCAGTGCAG CCAACTGGCA AATAGTCTCC ACCCCCGGCA CTATCACCGT
GGGTGGGTT GGGTTGGTG GGGTCACGTC GGTGACCGT TTATCAGAGG TGGGGCCGT GATAGTGGCA

4551 GAGTTGTCCG CACCACCGCA CGTCTCGCAG CCAAAAAA AAAAGAAAG AAAAAGAAA
CTCAACAGGC GTGGTGGCGT GCAGAGCGTC GGTTTTTT TTTTCTTTC TTTTCTTTT

4621 CAGCAGGTGG GTCCGGGTCG TGGGGCCGG AAAGCGAGG AGGATCGGA GCAGCGACGA GGCCCGGCC
GTCGTCCACC CAGGCCAGC ACCCCGGCC TTTTCGCTCC TCCTAGCGT CGTCGCTGCT CCGGGCCGG

4691 TCCCTCCGCT TCCAAGAAA CGCCCCCAT CGCCACTATA TACATACCC CCCCTCTCCT CCCATCCCC
AGGAGGCCGA AGTTTCTTT GCGGGGGTA GCGGTGATAT ATGTATGGG GGGGAGAGGA GGTAGGGG

4761 CAACCCCTACC ACCACCACCA CCACCACCTC CTCCCCCCTC GTGCCGGAC GACGAGCTCC TCCCCCTCC
GTTGGGATGG TGGTGGTGGT GGTGGTGAG GAGGGGGAG CGACGGCCTG CTGCTCGAG AGGGGGGAG

4831 CCTCCGCGG CCGCCGGTAA CCACCCCGCC CCTCTCCTCT TTCTTTCTCC GTTTTCTTT TCGTCTCGGT
GGGAGCGGC GCGGCCATT GGTGGGCGG GGAGAGGAGA AAGAAAGAG CAAAAAAA AGCAGAGCCA

4901 CTCGATCTTT GGCCTTGGTA GTTTGGGTGG GCGAGAGCGG CTTCGTCGCC CAGATCGGTG CGCGGGAGGG
GAGCTAGAAA CCGGAACCAT CAACCCACC CGCTCTCGCC GAAGCAGCGG GTCTAGCCAC GCGCCCTCCC

BamHI
~~~~~

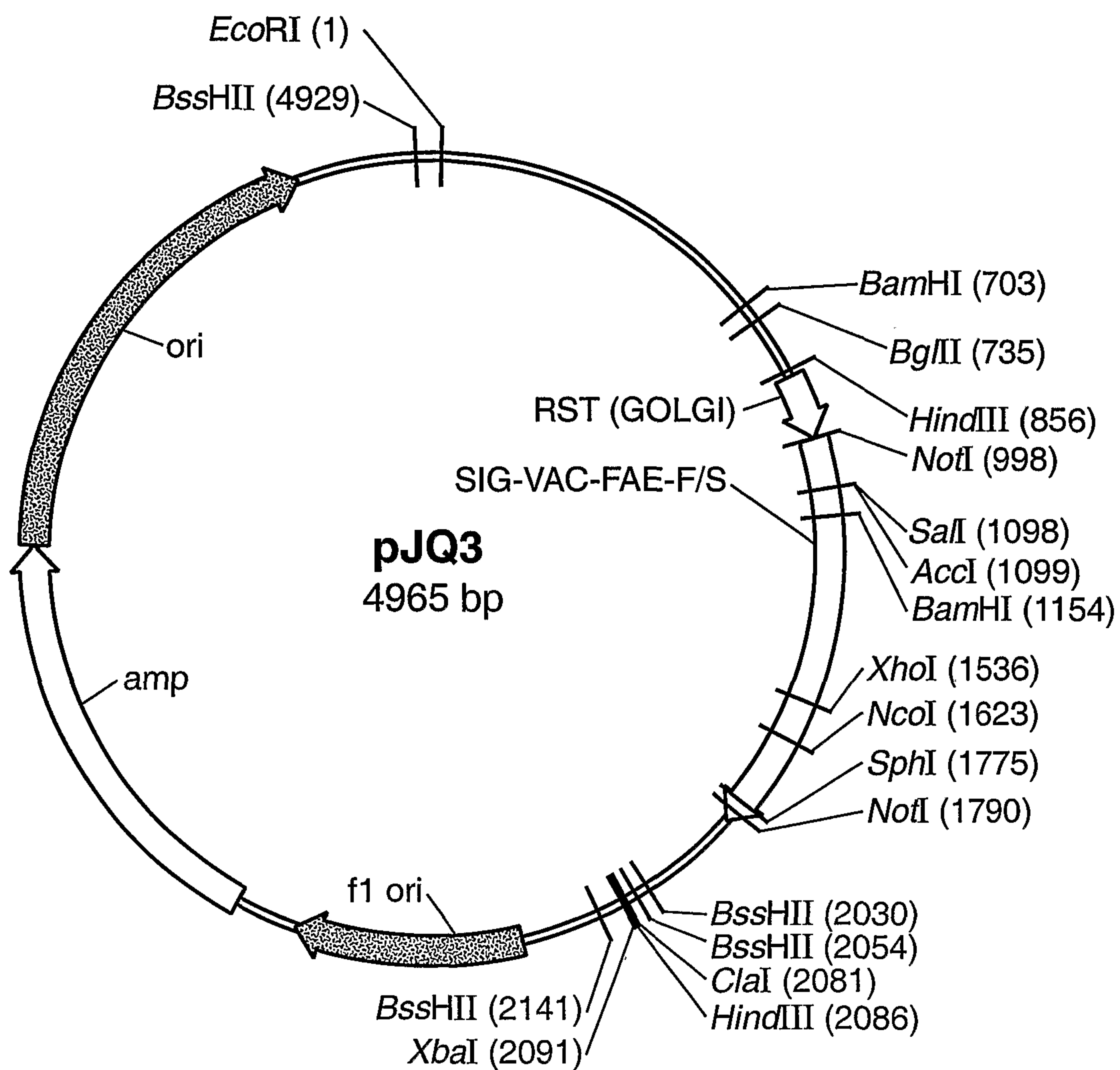
4971 GCGGGATCTC GCGGCTGGCG TCTCCGGGCG TGAGTCGGCC CGGGAATGG GGTCTCGGA  
CGCCCTAGAG CGCCGACCGC AGAGGCCCGC ACTCAGCCGG GCCTAGGAGC GCCCTTACC CCGAGAGCCT

BglII  
~~~~~

5041 TGTAGATCTT CTTTCTTTCT TCTTTTGTG GTAGAAATTG AATCCCTCAG CATGTTCAT CGGTAGTTT
ACATCTAGAA GAAAGAAA AGAAAAAC CATCTTAAC TTAGGGAGTC GTAACAAGTA GCCATCAAAA

5111 TCTTTTCATG ATTTGTGACA AATGCAGCCT CGTGCGGAGC TTTTGTGTAG GTAG
AGAAAAGTAC TAAACACTGT TTACGTCGGA GCACGCCCTG AAAAAACATC CATC

FIG.-45H

**FIG. 46A**

| | |
|-------|---|
| EcoRI | |
| ~~~~~ | |
| 1 | AAATCCACAA TGAACAATAA TAAGATTAAA ATAGCTTGCC CCCGTTGCAG CGATGGGTAT TTTTCTAGT
TTAAGGTGTT ACTTGTTATT ATTCTAATTT TATCGAACGG GGGCAACGTC GCTACCCATA AAAAAGATCA |
| 71 | AAAAATAAAG ATAAACTTAG ACTCAAAACA TTACAAAAA CAACCCCTAA AGTCCATAAG CCCAAAGTGC
TTTTATTTC TATTGAAATC TGAGTTTGT AATGTTTTT GTTGGGATT TCAGGATTTC GGGTTTCACG |
| 141 | TATGCACGAT CCATAGCAAG CCAGCCCAA CCAACCCAC CCAAGTGCAG CCAACTGGCA
ATACGTGCTA GGTATCGTTC GGTFCGGTT GGTTCGGTG GGTACAGTC GGTGACCGT |
| 211 | AAATAGTCTC ACCCCGGCA CTATCACCGT GAGTTGTCCG CACCACCGCA CGTCTCGCAG CCAAAAAAA
TTATCAGAGG TGGGGCCGT GATAGTGCA CTCAACAGGC GTGGTGGCGT GCAGAGCGTC GGTTTTTTT |
| 281 | AAAAAGAAAG AAAAAAGA AAAAGAAAA CAGAGGTGG GTCCGGGTCG TGGGGCCGG AAAAGCGAGG
TTTTTCTTTC TTTTTTTTCT TTTTCTTTT GTCTCCACC CAGGCCAGC ACCCCCGGC TTTTCGCTCC |
| 351 | AGGATCGCGA GCAGCGACGA GGCCCGGCC TCCCTCCGCT TCCAAAGAAA CGCCCCCAT CGCCACTATA
TCCTAGCGCT CGTCGCTGCT CCGGGCCGG AGGAGGCGA AGGTTTCTTT GCGGGGGTA GCGGTGATAT |
| 421 | TACATACCCC CCCCTCTCCT CCATCCCCC CAACCTACC ACCACACCA CCACACCTC CTCCTCCCTC
ATGTATGGG GGGAGAGGA GGTAGGGG GTTGGGATGG TGGTGTGGG GTGTGGGAG GAGGGGGAG |
| 491 | GCTGCCGGAC GACGAGCTCC TCCCCCTCC CCTCCGCCG CCGCCGGTAA CCACCCCGCC CCTCTCCTCT
CGACGGCCTG CTGCTCGAG AGGGGGAGG GGAGGCGGC GCGGCCATT GTGGGGCGG GGAGAGGAGA |
| 561 | TTCTTTCTCC GTTTTTTTT TCGTCTCGGT CTCGATCTTT GGCCTTGGTA GTTTGGGTGG GCGAGAGCGG
AAGAAAGAGG CAAAAAAAAG AGCAGAGCCA GAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC |
| 631 | CTTCGTCCG CAGATCGGTG CCGGGAGGG GCGGATCTC GCGCTGGCG TCTCCGGCG TGAGTCGGCC
GAAGCAGCG GTCTAGCCAC GCGCCTCCC CGCCGACCGC AGAGGCCCGC ACTCAGCCCG |
| BamHI | |
| ~~~~~ | |
| 701 | CGGATCCTCG CGGGAATGG GGTCTCGGA TGATGATCTT CTTTCTTTCT TCTTTTGTG GTAGAATTG
GCCTAGGAG GCCCTTACC CCGAGAGCCT ACATCTAGAA GAAAGAAAG AGAAAAACAC CATCTTAAAC |

BglII
~~~~~

FIG. 46B

**FIG. 46C**

|      |             |            |            |            |            |            |             |
|------|-------------|------------|------------|------------|------------|------------|-------------|
| 1331 | AGTCGAGTCG  | CTTGTCAAAC | AGCAGGTTAG | CCAGTATCCG | GACTACGCGC | TGACCGTGAC | CGGCCACKCC  |
|      | TCAGCTCAGC  | GAACAGTTTG | TCGTCCAATC | GGTCATAGGC | CTGATGCGCG | ACTGGCACTG | GCCGGTGMGG  |
| 1401 | CTCGGCGCCT  | CCCTGGCGGC | ACTCACTGCC | GCCCAGCTGT | CTGCGACATA | CGACAACATC | CGCCTGTACA  |
|      | GAGCCGCGGA  | GGGACCGCCG | TGAGTGACGG | CGGGTCGACA | GACGCTGTAT | GCTGTTGTAG | GCGGACATGT  |
|      |             |            |            |            |            |            | XhoI        |
|      |             |            |            |            |            |            | ~~~~~       |
| 1471 | CCTTCGGCGA  | ACCGCGCAGC | GGCAATCAGG | CCTTCGCGTC | GTACATGAAC | GATGCCTTCC | AAGCCTCGAG  |
|      | GGAAGCCGCT  | TGGCGCGTCG | CCGTTAGTCC | GGAAGCGCAG | CATGTACTTG | CTACGGAAGG | TTCGGAGCTC  |
| 1541 | CCCAGATACG  | ACGCAGTATT | TCCGGGTCAC | TCATGCCAAC | GACGGCATCC | CAAACCTGCC | CCCGGTGGAG  |
|      | GGGTCTATGC  | TGCGTCATAA | AGGCCCAGTG | AGTACGGTTG | CTGCCGTAGG | GTTTGGACGG | GGGCCACCTC  |
|      |             |            |            |            |            |            | NcoI        |
|      |             |            |            |            |            |            | ~~~~~       |
| 1611 | CAGGGGTACG  | CCCATGGCGG | TGTAGAGTAC | TGGAGCGTTG | ATCCTTACAG | CGCCCAGAAC | ACATTTGTCT  |
|      | GTCCCCCATGC | GGGTACCGCC | ACATCTCATG | ACCTCGCAAC | TAGGAATGTC | GCGGGTCTTG | TGTAACACAGA |
| 1681 | GCACTGGGGA  | TGAAGTGCAG | TGCTGTGAGG | CCCAGGGCGG | ACAGGGTGTG | AATAATGCGC | ACACGACTTA  |
|      | CGTGACCCCT  | ACTTCACGTC | ACGACACTCC | GGTCCCCGCC | TGTCCCACAC | TTATTACGCG | TGTGCTGAAT  |
|      |             |            |            |            |            |            | SphI        |
|      |             |            |            |            |            |            | ~~~~~       |
| 1751 | TTTTTGGGATG | ACGAGCGGCG | CATGCACCTG | GCCGGTCGCG | GCCGCGGAAA | CCACTGAAGG | ATGAGCTGTA  |
|      | AAAACCCCTAC | TGCTCGCCGC | GTACGTGGAC | CGGCCAGCGC | CGGCGCCTTT | GGTGACTTCC | TACTCGACAT  |
| 1821 | AAGAAGCAGA  | TCGTTCAAAC | ATTTGGCAAT | AAAGTTTCTT | AAGATTGAAT | CCTGTTGCCG | GTC TTGCGAT |
|      | TTCTTTCGTCT | AGCAAGTTTG | TAAACCGTTA | TTTCAAAGAA | TTCTAACTTA | GGACAACGGC | CAGAACGCTA  |
| 1891 | GATTATCATA  | TAAATTCTGT | TGAATTACGT | TAAGCATGTA | ATAATTAAAC | TGTAATGCAT | GACGTTATTT  |
|      | CTAATAGTAT  | ATTAAAGACA | ACTTAATGCA | ATTCGTACAT | TATTAATTGT | ACATTACGTA | CTGCAATAAA  |

FIG. 46D

|      |              |             |            |            |            |             |             |   |   |  |
|------|--------------|-------------|------------|------------|------------|-------------|-------------|---|---|--|
| 1961 | BssHII       |             |            |            |            |             |             |   |   |  |
|      | ATGAGATGGG   | TTTTTATGAT  | TAGAGTCCCG | CAATTATACA | TTTAATACGC | GATAGAAAC   | AAAAATAGC   | ~ | ~ |  |
|      | TACTCTACCC   | AAAAATACTA  | ATCTCAGGC  | GTTAATATGT | AAATTATGCG | CTATCTTTTG  | TTTTATATCG  |   |   |  |
| 2031 | BssHII       |             |            |            |            |             |             |   |   |  |
|      | ~            | ~           | ~          | ~          | ~          | ~           | ~           | ~ | ~ |  |
|      | GCGCAAACTA   | GGATAAATTA  | TCGCGCGCGG | TGTCATCTAT | GTTACTAGAT | CGATAAGCTT  | CTAGAGCGGC  |   |   |  |
|      | CGCGTTTGAT   | CCTATTTAAT  | AGCGCGCGCC | ACAGTAGATA | CAATGATCTA | GCTATTGAA   | GATCTCGCCG  |   |   |  |
| 2101 | BssHII       |             |            |            |            |             |             |   |   |  |
|      | ~            | ~           | ~          | ~          | ~          | ~           | ~           | ~ | ~ |  |
|      | CGGTGGAGCT   | CCAATTCGCC  | CTATAGTGAG | TCGTATTACG | CGCGCTCACT | GGCCGTCGTT  | TTACAACGTC  |   |   |  |
|      | GCCACCTCGA   | GGTTAAGCGG  | GATATCACTC | AGCATAATGC | GCGCGAGTGA | CCGGCAGCAA  | AATGTTGCAG  |   |   |  |
| 2171 | BssHII       |             |            |            |            |             |             |   |   |  |
|      | ~            | ~           | ~          | ~          | ~          | ~           | ~           | ~ | ~ |  |
|      | GTGACTGGGA   | AAACCCTGGC  | GTTACCCAAC | TTAATCGCCT | TGCAGCACAT | CCCCCTTTTCG | CCAGCTGGCG  |   |   |  |
|      | CAC TGACCCCT | TTTGGGACCG  | CAATGGGTTG | AATTAGCGGA | ACGTCGTGTA | GGGGGAAAGC  | GGTCGACCCG  |   |   |  |
| 2241 | BssHII       |             |            |            |            |             |             |   |   |  |
|      | ~            | ~           | ~          | ~          | ~          | ~           | ~           | ~ | ~ |  |
|      | TAATAGCGAA   | GAGCCCGCA   | CCGATCGCCC | TTCCCAACAG | TTGCGCAGCC | TGAATGGCGA  | ATGGGACGCG  |   |   |  |
|      | ATTATCGCTT   | CTCCGGGCGT  | GGCTAGCGGG | AAGGTTGTC  | AACGCGTCGG | ACTTACCGCT  | TACCC TGCGC |   |   |  |
| 2311 | BssHII       |             |            |            |            |             |             |   |   |  |
|      | ~            | ~           | ~          | ~          | ~          | ~           | ~           | ~ | ~ |  |
|      | CCCTGTAGCG   | GCGCATTAAG  | CGCGGCGGGT | GTGGTGGTTA | CGCGCAGCGT | GACCGCTACA  | CTTGCCAGCG  |   |   |  |
|      | GGGACATCGC   | CGCGTAATTC  | GCGCCGCCCA | CACCACCAAT | GCGCGTCGCA | CTGGCGATGT  | GAACGGTCGC  |   |   |  |
| 2381 | BssHII       |             |            |            |            |             |             |   |   |  |
|      | ~            | ~           | ~          | ~          | ~          | ~           | ~           | ~ | ~ |  |
|      | CCCTAGCGCC   | CGCTCCTTTC  | GCTTTCCTTC | CTTCCCTTCT | CGCCACGTTT | GCCGGCTTTC  | CCCGTCAAGC  |   |   |  |
|      | GGGATCGCGG   | GCGAGGAAAG  | CGAAAGAAGG | GAAGGAAAGA | GCGGTGCAAG | CGGCCGAAAG  | GGCAGTTCG   |   |   |  |
| 2451 | BssHII       |             |            |            |            |             |             |   |   |  |
|      | ~            | ~           | ~          | ~          | ~          | ~           | ~           | ~ | ~ |  |
|      | TCTAAATCGG   | GGGCTCCCTT  | TAGGGTTCCG | ATTAGTGCT  | TTACGGCACC | TCGACCCCAA  | AAAACTTGAT  |   |   |  |
|      | AGATTTAGCC   | CCCGAGGGAA  | ATCCCAAGGC | TAAATCACGA | AATGCCGTGG | AGCTGGGGTT  | TTTTGAAC TA |   |   |  |
| 2521 | BssHII       |             |            |            |            |             |             |   |   |  |
|      | ~            | ~           | ~          | ~          | ~          | ~           | ~           | ~ | ~ |  |
|      | TAGGGTGATG   | GTTACCGTAG  | TGGGCCATCG | CCCTGATAGA | CGGTTTTTCG | CCCTTTGACG  | TTGGAGTCCA  |   |   |  |
|      | ATCCCACTAC   | CAAGTGCA TC | ACCCGGTAGC | GGGACTATCT | GCCAAAAGC  | GGGAAACTGC  | AACCTCAGGT  |   |   |  |

FIG. 46E

2591 CGTTCCTTTAA TAGTGGACTC TTGTTCCAAA CTGGAACAAC ACTCAACCCT ATCTCGGTCT ATTCCTTTGA  
GCAAGAAAT ATCACCTGAG AACAAAGTTT GACCTTGTTG TGAGTTGGGA TAGAGCCAGA TAAGAAAACT

2661 TTTATAAGGG ATTTTGCCGA TTTTCGGCCTA TTGGTTAAAA AATGAGCTGA TTTAACAAAA ATTTAACGCG  
AAATATTCCT TAAACCGGT AAAGCCGGAT AACCAATT TT TACTCGACT AAATTGTTTT TAAATTGCGC

2731 AATTTTAACA AAATATTAC GCTTACAATT TAGGTGGCAC TTTTCGGGA AATGTGCGG GAACCCCTAT  
TTAAAAATTGT TTTATAATTG CGAATGTTAA ATCCACCGTG AAAAGCCCCT TTACACGCGC CTTGGGGATA

2801 TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC ATGAGACAAT AACCTGATA AATGCTTCAA  
AACAAATAAA AAGATTATG TAAGTTTATA CATAGCGAG TACTCTGTTA TTGGGACTAT TTACGAAAGTT

2871 TAATATTGAA AAAGGAAGAG TATGAGTATT CAACATTTC GTGTCGCCCT TATTCCTTT TTTGCGGCAT  
ATTATAACTT TTTCCTTCTC ATACTCAAA GTTGTAAGG CACAGCGGA ATAAGGAAA AAACGCCGTA

2941 TTTGCCTTCC TGTTTTGT TGTGCTTCT CACCCAGAAA CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC  
AAACGGAAGG ACAAAACGA GTGGGTCTTT GCGACCACTT TCAATTTCTA CGACTTCTAG TCAACCCACG

3011 ACGAGTGGGT TACATCGAAC TGGATCTCAA CAGCGTAAG ATCCTTGAGA GTTTTCGCC CGAAGAACGT  
TGCTCACCCA ATGTAGCTTG ACCTAGAGTT GTCGCCATTC TAGGAACTCT CAAAAGCGGG GCTTCTTGCA

3081 TTTCCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG CCGTATTATC CCGTATTGAC GCCGGCAAG  
AAAGTTACT ACTCGTGAAA ATTTCAAGAC GATACACCGC GCCATAATAG GGCATAACTG CGGCCCGTTC

3151 AGCAACTCGG TCGCCGATA CACTATTCTC AGAATGACTT GGTTGAGTAC TCACCAGTCA CAGAAAAGCA  
TCGTTGAGCC AGCGCGTAT GTGATAAGAG TCTTACTGAA CCAACTCATG AGTGGTCAGT GTCTTTTCGT

3221 TCTTACGGAT GGCA TGACAG TAAGAGAATT ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC  
AGAAATGCC TA CCGTACTGTC ATTCTCTTAA TACGTCACGA CCGTATTGGT ACTCACTATT GTGACGCCGG

3291 AACTTACTTC TGACAAACGAT CGGAGGACCG AAGGAGCTAA CCGCTTTTT GCACAAACATG GGGATCATG  
TTGAATGAAG ACTGTTGCTA GCCTCCTGGC TTCCCTCGATT GGCGAAAAA CGTGTTGTAC CCCCTAGTAC

3361 TAACTCGCCT TGATCGTTGG GAACCGGAGC TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT  
ATTGAGCGGA ACTAGCAACC CTTGGCCCTG ACTTACTTCG GTATGGTTTG CTGCTCGCAC TGTGGTGCTA

FIG. 46F

3431 GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAACT GGCGAACTAC TTACTCTAGC TTCCCGGCAA  
CGGACATCGT TACCGTTGTT GCAACGCGTT TGATAATTGA CCGCTTGATG AATGAGATCG AAGGGCCGTT

3501 CAATTAAATAG ACTGGATGGA GCGGATATAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGCTGGCT  
GTTAATTATC TGACCTACCT CCGCCTATT CAACGTCCTG GTGAAGACGC GAGCCGGGA GGCACCGA

3571 GGTTTATTGC TGATAAATCT GGAGCCGGTG AGCGTGGTC TCGCGGTATC ATTGCAGCAC TGGGGCCAGA  
CCAAATAACG ACTATTTAGA CCTCGGCCAC TCGCACCCAG AGCGCCATAG TAACGTCGTG ACCCCGGTCT

3641 TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA  
ACCATTCGGG AGGCATAGC ATCAATAGAT GTGCTGCCCC TCAGTCCGTT GATACCTACT TGCCTTATCT

3711 CAGATCGCTG AGATAGGTGC CTCACTGATT AAGCATTTGGT AACTGTCAGA CCAAGTTTAC TCATATATAC  
GTCTAGCGAC TCTATCCACG GAGTGACTAA TTTCGTAACCA TTGACAGTCT GGTTCAAATG AGTATATATG

3781 TTTAGATTGA TTTAAAACCTT CATTTTTAAT TTAAAAGGAT CTAGGTGAAG ATCCTTTTIG ATAACTCAT  
AAATCTAACT AAATTTTGAA GTAAAAATTA AATTTTCCCTA GATCCACTTC TAGGAAAAAC TATTAGAGTA

3851 GACCAAATC CCTAACGTG AGTTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT  
CTGGTTTTAG GGAATTGCAC TCAAAAAGCAA GTGACTCGC AGTCTGGGC ATCTTTTCTA GTTTCCTAGA

3921 TCTTGAGATC CTTTTTTTCT GCGCGTAATC TGCTGCTTGC AAACAAAAAA ACCACCGCTA CCAGCGGTGG  
AGAACTCTAG GAAAAAAGA CGCGCATTAG ACGACGAAACG TTTGTTTTTT TGGTGGCGAT GGTGCCCACC

3991 TTTGTTTGCC GGATCAAGAG CTACCAACTC TTTTTCGAA GGTAACGTGC TTCAGCAGAG CGCAGATACC  
AAACAAACGG CCTAGTTCTC GATGGTTGAG AAAAAGGCTT CCATTGACCG AAGTCGTCTC GCGTCTATGG

4061 AAATACTGTC CTTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAACT CTGTAGCACC GCTACATAC  
TTTATGACAG GAAGATCACA TCGGCATCAA TCCGGTGGTG AAGTTCCTGA GACATCGTGG CCGATGTATG

4131 CTCGCTCTGC TAAATCCTGTT ACCAGTGGCT GTGCCAGTG GCGATAAGTC GTGTCTTACC GGGTTGGACT  
GAGCGAGACG ATTAGGACAA TGGTCACCGA CGACGGTCAC CGCTATTTCAG CACAGAAATGG CCCAACCTGA

4201 CAAGACGATA GTTACCGGAT AAGGCGCAGC GGTCCGGCTG AACGGGGGT TCGTGCACAC AGCCAGCTT  
GTTCTGCTAT CAAATGGCCTA TTCCGCGTCG CCAGCCCCCA TTGCCCCCCA AGCACGTGTG TCGGGTCGAA

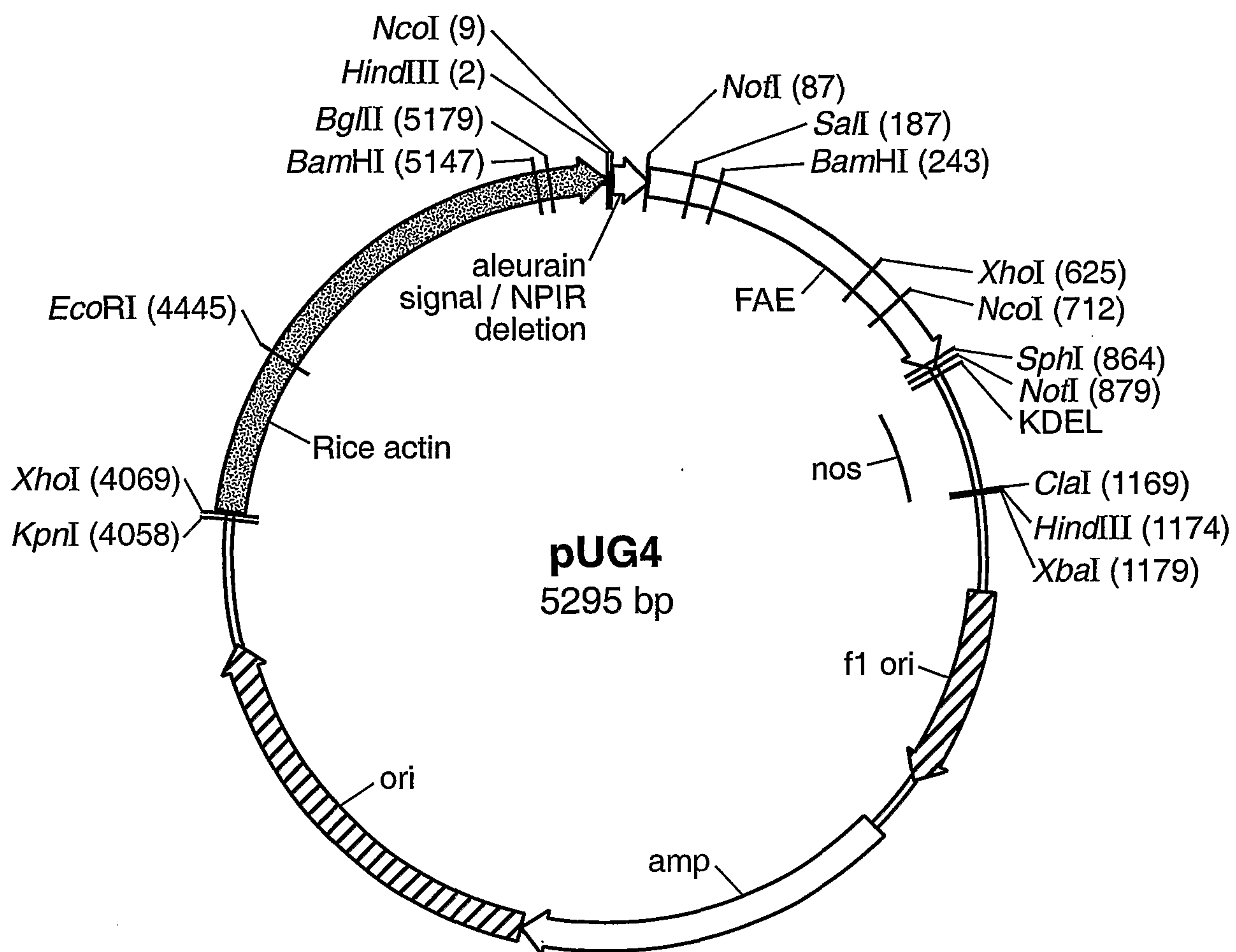
FIG. - 46G

|       |             |            |             |             |             |             |            |
|-------|-------------|------------|-------------|-------------|-------------|-------------|------------|
| 4271  | GGAGCGAACG  | ACCTACACCG | AACTGAGATA  | CCTACAGCGT  | GAGCTATGAG  | AAAGGCCAC   | GCTTCCCGAA |
|       | CCTCGCTTGC  | TGGATGTGGC | TTGACTCTAT  | GGATGTCGCA  | CTCGATACTC  | TTTCGCGGTG  | CGAAGGGCTT |
| 4341  | GGGAGAAAGG  | CGGACAGGTA | TCCGGTAAGC  | GGCAGGGTCG  | GAACAGGAGA  | GCGCACGAGG  | GAGCTTCCAG |
|       | CCCCTCTTCC  | GCCGTGCCAT | AGGCCATTCTG | CCGTCCCAGC  | CTTGTCCTCT  | CGCGTGCTCC  | CTCGAAGGTC |
| 4411  | GGGGAAACGC  | CTGGTATCTT | TATAGTCCTG  | TCGGGTTTCG  | CCACCTCTGA  | CTTGAGCGTC  | GATTTTGTG  |
|       | CCCCTTTGCG  | GACCATAGAA | ATATCAGGAC  | AGCCCAAAGC  | GGTGGAGACT  | GAACTCGCAG  | CTAAAAACAC |
| 4481  | ATGCTCGTCA  | GGGGGGCGGA | GCCTATGGAA  | AAACGCCAGC  | AACGCGGCCT  | TTTTACGGTT  | CCTGGCCTTT |
|       | TACGAGCAGT  | CCCCCCGCCT | CGGATACCTT  | TTTGCGGTCTG | TTGCGCCGGA  | AAAATGCCAA  | GGACCGGAAA |
| 4551  | TGCTGGCCTT  | TTGCTCACAT | GTTCTTTTCT  | GCGTTATCCC  | CTGATTCTGT  | GGATAACCGT  | ATTACCGCCT |
|       | ACGACCGGAA  | AACGAGTGTA | CAAGAAAGGA  | CGCAATAGGG  | GACTAAGACA  | CCTATTGGCA  | TAATGGCGGA |
| 4621  | TTGAGTGAGC  | TGATACCGCT | CGCCGCAGCC  | GAACGACCGA  | GCGCAGCGAG  | TCAGTGAGCG  | AGGAAGCGGA |
|       | AATCACTCG   | ACTATGGCGA | GCGGCGTCGG  | CTTGCTGGCT  | CGCGTCGCTC  | AGTCACTCGC  | TCCTTCCGCT |
| 4691  | AGAGCGCCCA  | ATACGCAAAC | CGCCTCTCCC  | CGCGGTTGG   | CCGATTCAAT  | AATGCAGCTG  | GCACGACAGG |
|       | TCTCGCGGGT  | TATGCGTTTG | GCGGAGAGGG  | GCGCGCAACC  | GGCTAAGTAA  | TTACGTGCGAC | CGTGCTGTCC |
| 4761  | TTTCCCGACT  | GGAAAGCGGG | CAGTGAGCGC  | AACGCAATTA  | ATGTGAGTTA  | GCTCACTCAT  | TAGGCACCCC |
|       | AAAGGGCTGA  | CCTTTCGCCC | GTCACTCGCG  | TTGCGTTAAT  | TACACTCAAT  | CGAGTGAGTA  | ATCCGTGGGG |
| 4831  | AGGC'TTTACA | CTTTATGCTT | CCGGCTCGTA  | TGTTGTGTGG  | AATTGTGAGC  | GGATAACAAT  | TTCACACAGG |
|       | TCCGAAATGT  | GAAATACGAA | GGCCGAGCAT  | ACAACACACC  | TTAAACACTCG | CCTATTGTTA  | AAGTGTGTCC |
| ~~~~~ |             |            |             |             |             |             |            |
| 4901  | AAACAGCTAT  | GACCATGATT | ACGCCAAGCG  | CGCAATTAAAC | CCTCACTAAA  | GGGAACAAAA  | GCTGG      |
|       | TTTGTGCGATA | CTGGTACTAA | TGCGGTTTCG  | GCGTTAATTG  | GGAGTGATTT  | CCCTTGTTTT  | CGACC      |

EcoRI

BssHII

FIG.\_ 46H

**FIG.\_47A**

NcoI
~~~~~
HindIII
~~~~~
1 AAGCTTACCA TGGCCACGC CCGGTCCTC CTCTGGCGC TCGCCGTGCT GCCACGGC GCCGTGCGC
M A H A R V L L L A L A V L A T A A V A V
NotI
~~~~~
71 TCGCCTCCTC CCGCGCGGC GCCTCCACGC AGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAAT
A S S R A A A S T Q G I S E D L Y S R L V E M
Sali
~~~~~
141 GGCCACTATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC ATTCCGTCGA CTATTATCAA GGGAGAGAAA
A T I S Q A A Y A D L C N I P S T I I K G E K
BamHI
~~~~~
211 ATTTACAATT CTCAAACTGA CATTAAAGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA ATAATCACCG
F R G T G S D T N L Q L D T N Y T L T P F D T
281 TCTTCCGTGG CACTGGTAGT GATACGAATC TACAACCTCGA TACTAACTAC ACCCTCACGC CTTTCGACAC
L P Q C N G C E V H G G Y Y I G W V S V Q D Q
351 CCTACCACAA TGCAACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA
V E S L V K Q Q V S Q Y P D Y A L T V T G H X L
421 GTCGAGTCGC TTGTCAAACA GCAGGTTAGC CAGTATCCGG ACTACGCGCT GACCGTGACC GGCCACKCCC
G A S L A A L T A A Q L S A T Y D N I R L Y T
491 TCGGCGCCTC CCTGGCGGCA CTCACTGCCG CCCAGCTGTC TCGACATAC GACAACATCC GCCTGTACAC
XhoI
~~~~~
561 CTTCGGCGAA CCGCGCAGCG GCAATCAGGC CTTGCGGTCG TACATGAACG ATGCCTTCCA AGCCTCGAGC
P D T T Q Y F R V T H A N D G I P N L P P V E Q
631 CCAGATACGA CGCAGTATTT CCGGGTCACT CATGCCAACG ACGGCATCCC AAACCTGCC CCGGTGGAGC
NcoI
~~~~~

FIG.-47B

FIG. 47C

| | | | | | | | |
|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 2591 | ATTAATAGAC | TGGATGGAGG | CGGATAAAGT | TGCAGGACCA | CTTCTGCGCT | CGGCCCTTCC | GGCTGGCTGG |
| 2661 | TTTATTGCTG | ATAAATCTGG | AGCCGGTGAG | CGTGGGTCTC | GCGGTATCAT | TGCAGCACTG | GGGCCAGATG |
| 2731 | GTAAGCCCTC | CCGTATCGTA | GTTATCTACA | CGACGGGAG | TCAGGCAACT | ATGGATGAAC | GAAATAGACA |
| 2801 | GATCGCTGAG | ATAGGTGCCT | CACTGATTAA | GCAATGGTAA | CTGTCAGACC | AAGTTTACTC | ATATATACTT |
| 2871 | TAGATTGATT | TAAAACTTCA | TTTTTAATTT | AAAAGGATCT | AGGTGAAGAT | CCTTTTGTGAT | AATCTCATGA |
| 2941 | CCAAAATCCC | TTAACGTGAG | TTTTTCGTTCC | ACTGAGCGTC | AGACCCCGTA | GAAAAGATCA | AAGGATCTTC |
| 3011 | TTGAGATCCT | TTTTTTCTGC | GCGTAACTCG | CTGCTTGCAA | ACAAAAAAC | CACCGCTACC | AGCGGTGGTT |
| 3081 | TGTTTGCCGG | ATCAAGAGCT | ACCAACTCTT | TTTCCGAAGG | TAACTGGCTT | CAGCAGAGCG | CAGATACCAA |
| 3151 | ATACTGTCCT | TCTAGTGTAG | CCGTAGTTAG | GCCACCACCT | CAAGAACTCT | GTAGCACCGC | CTACATACCT |
| 3221 | CGCTCTGCTA | ATCCTGTTAC | CAGTGGCTGC | TGCCAGTGGC | GATAAGTCGT | GTCTTACCGG | GTTGGACTCA |
| 3291 | AGACGATAGT | TACCGGATAA | GGCGAGCGG | TCGGGCTGAA | CGGGGGGTTC | GTGCACACAG | CCCAGCTTGG |
| 3361 | AGCGAACGAC | CTACACCGAA | CTGAGATACC | TACAGCGTGA | GCTATGAGAA | AGCGCCACGC | TTCCCGAAGG |
| 3431 | GAGAAAGGCG | GACAGGTATC | CGGTAAGCGG | CAGGGTCGGA | ACAGGAGAGC | GCACGAGGGA | GCTTCCAGGG |
| 3501 | GGAAACGCCT | GGTATCTTTA | TAGTCCTGTC | GGGTTTCGCC | ACCTCTGACT | TGAGCGTCCA | TTTTTTGTGAT |
| 3571 | GCTCGTCAGG | GGGCGGAGC | CTATGGAAA | ACGCCAGCAA | CGCGGCCTTT | TTACGGTCC | TGGCCCTTTG |
| 3641 | CTGGCCCTTT | GCTCACATGT | TCTTTCCCTGC | GTTATCCCTT | GATTCGTGG | ATAACCGTAT | TACCGCCCTT |
| 3711 | GAGTGAGCTG | ATACCGCTCG | CCGAGCCGA | ACGACCGAGC | GCAGCGAGTC | AGTGAGCGAG | GAAAGCGGAA |
| 3781 | AGCGCCCAAT | ACGCAAAACCG | CCTCTCCCG | CGCGTTGGCC | GATTCATTAA | TGCAGCTGGC | ACGACAGGTT |
| 3851 | TCCCAGACTGG | AAAGCGGGCA | GTGAGCGCAA | CGCAATTAA | GTGAGTTAGC | TCACCTCATTA | GGCACCCCCAG |
| 3921 | GCTTTTACACT | TTATGCTTCC | GGCTCGTATG | TTGTGTGGAA | TTGTGAGCGG | ATAACAATTT | CACACAGGAA |
| 3991 | ACAGCTATGA | CCATGATTAC | GCCAAGCGCG | CAATTAAACC | TCACATAAAG | GAAACAAAGC | TGGGTACCGG |
| 4061 | GCCCCCCTC | GAGGTCATT | ATATGCTTGA | GAAGAGAGTC | GGGATAGTCC | AAAATAAAAC | AAAGGTAAGA |
| 4131 | TTACCTGGTC | AAAAGTGAAA | ACATCAGTTA | AAAGGTGGTA | TAAGTAAAT | ATCGGTAATA | AAAGGTGGCC |
| 4201 | CAAAGTGAAA | TTTACTCTTT | TCTACTATT | TAAAAATTGA | GGATGTTTTG | TCGGTACTTT | GATACGTCA |
| 4271 | TTTTGTATGA | ATTGGTTTTT | AAGTTTATTC | GCGATTTGGA | AATGCATATC | TGTATTTGAG | TCGGTTTTTA |
| 4341 | AGTTCGTTGC | TTTTTGTAAT | ACAGAGGGAT | TTGTATAAGA | AATATCTTTA | AAAAACCCAT | ATGCTAAATTT |
| 4411 | GACATAAATTT | TTGAGAAAAA | TATATATTCA | GGCGAATTCC | ACAAATGAACA | ATAATAAGAT | TAAAAATAGCT |
| 4481 | TGCCCCCGTT | GCAGCGATGG | GTAATTTTTT | TAGTAAATA | AAAGATAAAC | TTAGACTCAA | AACATTTACA |
| 4551 | AAAACAACCC | CTAAAGTCCT | AAAGCCCAA | GTGCTATGCA | CGATCCATAG | CAAGCCCAGC | CCAACCCAAC |
| 4621 | CCAACCCAAC | CCACCCAGT | GCAGCCAACT | GGCAAAATAGT | CTCCACCCCC | GGCACTATCA | CCGTGAGTTG |

KpnI
~~

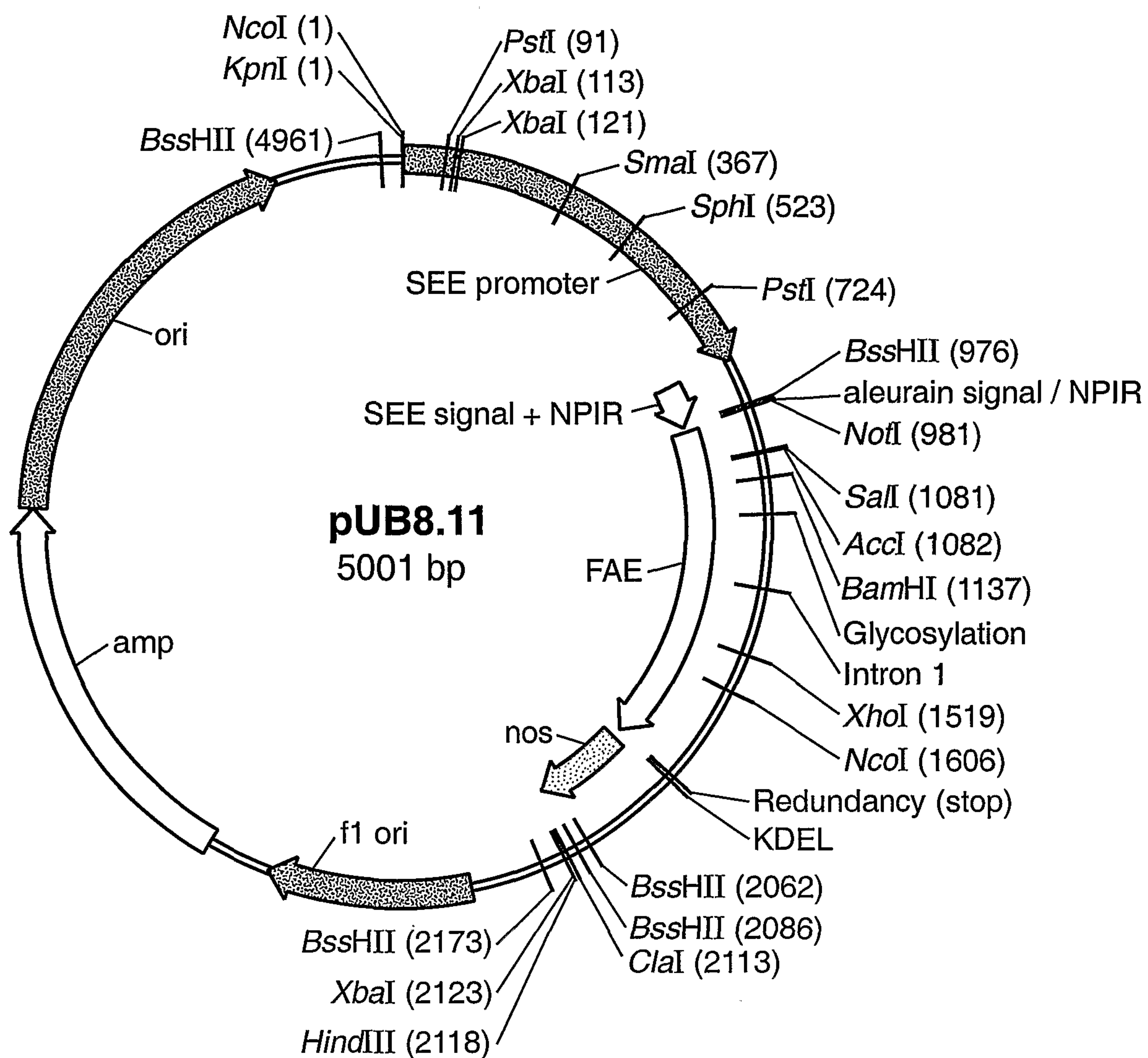
XhoI
~~~~~

ECORI  
~~~~~

FIG.- 47D

| | | | | | | | |
|------|-------------|-------------|------------|------------|-------------|-------------|------------|
| 4691 | TCCGCACCCAC | CGCACGTCCTC | GCAGCCAAAA | AAAAAAAG | AAGAAAAA | AAGAAAAAGA | AAAACAGCAG |
| 4761 | GTGGGTCCGG | GTCGTGGGG | CCGAAAAAGC | GAGGAGGATC | GCGAGCAGCG | ACGAGGCCCG | GCCCTCCCTC |
| 4831 | CGCTTCCAAA | GAAACGCCCC | CCATCGCCAC | TATATACATA | CCCCCCCCTC | TCCTCCCATC | CCCCCAACCC |
| 4901 | TACCACCAAC | ACCACCAACA | CCTCCTCCCC | CCTCGCTGCC | GGACGACGAG | CTCCTCCCCC | CTCCCCCTCC |
| 4971 | GCCGCCGCCG | GTAACCAACC | CGCCCCCTC | CTCTTCTTT | CTCCGTTTTT | TTTTTCGTCT | CGGTCTCGAT |
| 5041 | CTTTGGCCCT | GGTAGTTTGG | GTGGGCGAGA | CGGCTTTCGT | CGCCACAGATC | GGTGC GCGGG | AGGGGCGGGA |
| | | | | BamHI | | | BglII |
| | | | | ~~~~~ | | | |
| 5111 | TCTCGCGGCT | GGCGTCTCCG | GGCGTGAGTC | GGCCCCGATC | CTCGCGGGGA | ATGGGGCTCT | CGGATGTAGA |
| | BglII | | | | | | |
| | ~~~ | | | | | | |
| 5181 | TCTTCTTTCT | TTCTTCTTTT | TGTGGTAGAA | TTTGAATCCC | TCAGCATTGT | TCATCGGTAG | TTTTTCTTTT |
| 5251 | CATGATTTGT | GACAAATGCA | GCCTCGTGCG | GAGCTTTTTT | GTAGC | | |

FIG.-47E

**FIG. 48A**

| | | | |
|-------|--|-------|-------|
| NcoI | | | |
| ~~~~~ | | | |
| KpnI | | | |
| ~ | | | |
| 1 | CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG | | |
| | GTACCCGGGC CATATTAAATA CCTATAGAG TTCGTTTATT AGCTTTATAG TGGTAACCGA TGTATATAGAC | | |
| | | PstI | XbaI |
| | | ~~~~~ | ~~~~~ |
| 71 | AGCTCCGAGT TCTGACTGCA GTCGGGATGA CGCGTGTTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC | | |
| | TCGAGGCTCA AGACTGACGT CAGACCTACT GCGCACAAAC TAGATCTTGA GATCTATCGT GTCGGTGTCTG | | |
| 141 | ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT | | |
| | TGGATGTCCCT CACGCTGTGA ACACCTGACA TCATCACAAC CTCTGCCCTCG AGAAAGGATG GAGGACTGCA | | |
| 211 | TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC | | |
| | ACGGCGGCAA CAGGTAAGGT TGCCGTAGTG AGAGTTGGTT AGTGCGCGAG GGTGTGTTTA TAGCAGGGGG | | |
| 281 | ATGTCCTTGGC GGAGAGAGAG TACATACATG CTGTCGCGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC | | |
| | TACAGAACCG CCTCTCTCTC ATGTATGTAC GACAGCGCGG CAAAACAGA CTTAGAGCGA AGGTGACCGG | | |
| | | SmaI | |
| | | ~~~~~ | |
| 351 | AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCTG TCACCCCTGG CGTCATGGGA | | |
| | TTAGTCGAGT CGAGGGCCCT CGAGTGAGTA AGTCTAGGG TAGCAGCAGC AGTGGGGACC GCAGTACCCT | | |
| 421 | TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT | | |
| | ACCTTTTCTT GGAGGCAACG AGCCTACTCA GTCGGTATAG GGGCTTGTCT CATGACGTTT TATTGGGTTA | | |
| | | SphI | |
| | | ~~~~~ | |
| 491 | TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTGG CTTAATGAC TTTATTTTG | | |
| | AGTCTAAGG GGTATCTCTT TTCATATCGT ACGAAAGCCC AAAACAAACC GAATTAAC TG AATAAAAC | | |
| 561 | TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG | | |
| | AACCTCAACT TACGACTAAA CAACACATTT TACGGGTTGG TAGACTTATA GCTCTGCC TA TTATCCGACC | | |

FIG. 48B

631 CTAATTAAATT TATAGCAAGA TTCGTGATGT CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
GATTAATTAA ATATCGTTCT AAGACATCAC GTGTAGCGTT TATAGAAAGA CCCGTAATGT CGACCTCCGA

PstI
~~~~~

701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAC  
AGTAGTCGGA CTTTGTGAGA CGTCTCGGAC TTCGTTACCC ACTTCGCACC GCTACTCTAC CCATATTTG

771 CCCCGGCACC GGGACGCGAG CTCCTGGCCTA CCAGTACCAT CTCGCCTCGC TCCCCCTGCC GGACGACCCA  
GGGGCCGTGG CCTGCGCTC GAGGGCGGAT GTCATGGTA GAGCGGAGCG AGGGGACGG CCTGCTGGGT

841 GTAAATAACT GTTGCCCACT CGCCGGCGAG ATGGCCACG GCCGATCCT CTTCTTGGCG CTCGCCGTCT  
CATTTATGA CAACGGGTGA GCGCCGCTC TACCGGGTGC CGCGTAGGA GAAGAACCGC GAGCGGCAGA

BsSHII  
~~~~~  
NotI
~~

911 TGGCCACCGC CGCGGTGGCC GCCGCATCNT TGGCGGACTC CAACCCGATC CGGCCCGTCA CCGAGCGCGC
ACCGTGGCG GCGCCACCGG CGCGTAGNA ACCGCTGAG GTTGGGCTAG GCCGGCAGT GGCTCGCGCG

NotI
~~~~~

981 GGCCGCCCTCC ACGCAGGGCA TCTCCGAAGA CCTCTACAGC CGTTTAGTCG AAATGGCCAC TATCTCCAA  
CCGGCGGAGG TGCCTCCCGT AGAGGCTTCT GGAGATGTCG GCAAATCAGC TTACCCGGTG ATAGAGGGTT

SalI  
~~~~~  
AccI
~~~~~

1051 GCTGCCCTACG CCGACCTGTG CAACATTCCG TCGACTATTA TCAAGGAGA GAAATTTAC AATTCCTCAA  
CGACGGATGC GGCTGGACAC GTTGTAAGGC AGTGATAAT AGTTCCCTCT CTTTTAAATG TTAAGAGTTT

FIG. 48C

|      |                                                                                |  |  |
|------|--------------------------------------------------------------------------------|--|--|
|      | BamHI                                                                          |  |  |
|      | ~~~~~                                                                          |  |  |
| 1121 | CTGACATTAA CGGATGGATC CTCGCGGACG ACAGCAGCAA AGAAATAATC ACCGTC TTCC GTGGCACTGG  |  |  |
|      | GACTGTAAATT GCCTACCTAG GAGGCGCTGC TGTCGTCGTT TCTTTATTAG TGGCAGAAAG CACCGTGACC  |  |  |
| 1191 | TAGTGATACG AATCTACAAC TCGATACTAA CTACACCCCTC ACGCCTTTCG ACACCCCTACC ACAATGCAAC |  |  |
|      | ATCACTATGC TTAGATGTTG AGCTATGATT GATGTGGGAG TGC GGAAAGC TGTGGGATGG TGTACGTTG   |  |  |
| 1261 | GGTTGTGAAG TACACGGGTG ATATTATATT GGATGGGTCT CCGTCCAGGA CCAAGTCGAG TCGCTTGTCA   |  |  |
|      | CCAACACTTC ATGTGCCACC TATAATATAA CTTACCCAGA GGCAGGTCCT GTTCAGCTC AGCGAACAGT    |  |  |
| 1331 | AACAGCAGGT TAGCCAGTAT CCGGACTACG CGCTGACCGT GACCGGCCAC KCCCTCGGCG CCTCCCTGGC   |  |  |
|      | TTGTCTGTCCA ATCGGTCATA GCCTGATGC GCGACTGGCA CTGGCCGGTG MGGAGCCGC GGAGGGACCG    |  |  |
| 1401 | GGCACTCACT GCCGCCCAGC TGTCTGCGAC ATACGACAAC ATCCGCCCTGT ACACCTTCGG CGAACCGCGC  |  |  |
|      | CCGTGAGTGA CGCGGGGTCG ACAGACGCTG TATGCTGTTG TAGGCGGACA TGTGGAAGCC GCTTGGCGCG   |  |  |
|      | XhoI                                                                           |  |  |
|      | ~~~~~                                                                          |  |  |
| 1471 | AGCGGCAATC AGGCCTTCGC GTCGTACATG AACGATGCCT TCCAAGCCTC GAGCCAGAT ACGACGCAGT    |  |  |
|      | TCGCCGTTAG TCCGGAAGCG CAGCATGTAC TTGCTACGGA AGGTTCCGAG CTCGGGTCTA TGCTGCGTCA   |  |  |
|      | NcoI                                                                           |  |  |
|      | ~~~~~                                                                          |  |  |
| 1541 | ATTTCGGGT CACTCATGCC AACGACGGCA TCCCAAACCT GCCCCCGGTG GAGCAGGGT ACGCCCATGG     |  |  |
|      | TAAAGGCCCA GTGAGTACGG TTGCTGCCGT AGGTTTGGA CGGGGCCAC CTCGTCCCCA TCGGGGTACC     |  |  |
| 1611 | CGGTGTAGAG TACTGGAGCG TTGATCCCTA CAGCGCCAG AACACATTG TCTGCACTGG GGATGAAGT      |  |  |
|      | GCCACATCTC ATGACCTCGC AACTAGGAAT GTCGCGGGTC TTGTGTAAAC AGACGTGACC CCTACTTCAC   |  |  |
| 1681 | CAGTGCTGTG AGGCCCAGGG CGGACAGGGT GTGAATAATG CGCACACGAC TTA TTTTGGG ATGACGAGCG  |  |  |
|      | GTCACGACAC TCCGGGTCCC GCCTGTCCCA CACTTATTAC GCGTGTGCTG AATAAAACCC TACTGCTCGC   |  |  |
| 1751 | GAGCCTGTAC ATGGTGATCA GTCA TTTTCA CCTCCCCGAG TGTAACAGGA AAGATGGATG TCCTGGAGAG  |  |  |
|      | CTCGGACATG TACCACTAGT CAGTAAAGTC GGAGGGGCTC ACATGGTCCT TTCTACCTAC AGGACCTCTC   |  |  |

FIG. 48D

1821

GGGGCCGCGT AACCACTGAA GGATGAGCTG TAAAGAAGCA GATCGTTCAA ACATTGGCA ATAAAGTTTC

CCCCGGCGCA TTGGTGACTT CCTACTCGAC ATTCTTTCGT CTAGCAAGTT TGTA AACCGT TATTCAAG

1891

TTAAGATTGA ATCCTGTTGC CGGTCTTGC ATGATTATCA TATAATTCT GTTGAAATTAC GTTAAGCATG

AATTCCTAAT TAGGACAACG GCCAGAACG TACTAATAGT ATATTAAAGA CAACCTTAATG CAATTCGTAC

1961

TAATAATTAA CATGTAATGC ATGACGTTAT TTATGAGATG GGTTTTTATG ATTAGAGTCC CGCAATTATA

ATTATTAAAT GTACATTACG TACTGCAATA AATACTCTAC CCAAAAATAC TAATCTCAGG GCGTTAATAT

2031

CATTTAATAC GCGATAGAAA ACAAATATA GCGCGCAAC TAGGATAAAT TATCGCGCGC GGTGTCATCT

GTAAATTATG CGCTATCTTT TGTTTTATAT CGCGCGTTTG ATCCTATTTA ATAGCGCGCG CCACAGTAGA

2101

ATGTTACTAG ATCGATAAGC TTCTAGAGCG GCCGGTGGAG CTCCAATTCG CCTATAGTG AGTCGTATTA

TACAAATGATC TAGCTATTCTG AAGATCTCG CGGCCACCTC GAGGTTAAGC GGGATATCAC TCAGCATAAAT

2171

CGCGCGCTCA CTGGCCGTCG TTTTACAACG TCGTGACTGG GAAAACCCCTG GCGTTACCCA ACTTAATCGC

GCGCGCGAGT GACCGGCAGC AAAATGTTGC AGCACTGACC CTTTITGGGAC CGCAATGGGT TGAATTAGCG

2241

CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC

GAACGTCGTG TAGGGGGAAG GCGGTCGACC GCATTATCGC TTCTCCGGGC GTGGCTAGCG GGAAGGGTTG

2311

AGTTGCGCAG CCTGAATGGC GAATGGGACG CGCCCTGTAG CCGCGCATTA AGCGCGGCGG GTGTGGTGGT

TCAACGCGTC GGACTTACCG CTTACCCCTGC GCGGACATC GCCGCGTAAT TCGCGCCGCC CACACCACCA

2381

TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT TCGCTTTCCT CCCTTCCTTT

ATGCGCGTCG CACTGGCGAT GTGAACGGTC GCGGATCGC GCGGAGGAA AGCGAAAGAA GGAAGGAAA

FIG. 48E

2451 CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC GGGGGCTCCC TTTAGGGTTC CGATTTAGTG  
GAGCGGTGCA AGCGGCCGAA AGGGGCAGTT CGAGATTTAG CCCC CGAGGG AAATCCCAAG GCTAAATCAC

2521 CTTTACGGCA CCTCGACCCC AAAAAACTTG ATTAGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA  
GAAATGCCGT GGAGCTGGGG TTTT TTTGAAC TAAATCCACT ACCAAGTGCA TCACCCGGTA GCGGGACTAT

2591 GACGGTTTTT CGCCCTTTGA CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTCCA AACTGGAACA  
CTGCCAAAAA GCGGAAACT GCAACCTCAG GTGCAAGAAA TTATCACCTG AGAACAAAGT TTGACCTTGT

2661 ACACCTCAACC CTATCTCGGT CTATTCCTTT GATTATAAG GGATTTTGCC GATTTCGGC TATTGGTTAA  
TGTGAGTTGG GATAGAGCCA GATAAGAAA CTAAATATTC CCTAAAACGG CTAAAGCCGG ATAAACCAATT

2731 AAAATGAGCT GATTTAACAA AAATTTAACG CGAATTTTAA CAAAATATTA ACGCTTACAA TTTAGGTGGC  
TTT TACTCGA CTAAATTGTT TTTAAATTGC GCTTAAAAAT GCTTTTATAAT TGCGAATGTT AAATCCACCG

2801 ACTTTTCGGG GAAATGTGG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT ACATTCAAAT ATGTATCCGC  
TGAAAAGCCC CTTTACACGC GCCTTGGGA TAAACAAATA AAAAGATTA TGTAAGTTTA TACATAGGCG

2871 TCATGAGACA ATAACCCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG AGTATGAGTA TTCAACATTT  
AGTACTCTGT TATTGGGACT ATTACGAAG TTATATAAC TTTTTCCTTC TCATACTCAT AAGTTGTAAA

2941 CCGTGTCGCC CTTATTCCCT TTTTTCGGC ATTTTGCCCT CCGTTTTTG CTCACCCAGA AACGCTGGTG  
GGCACAGCGG GAAATAAGGA AAAACGCCG TAAACGGAA GGACAAAAC GAGTGGTCT TTGCGACCAC

3011 AAAGTAAAAG ATGCTGAAGA TCAGTTGGGT GCACGAGTGG GTTACATCGA ACTGGATCTC AACAGCGGTA  
TTTCAATTTC TACGACTTCT AGTCAACCCA CGTGCTCACC CAATGTAGCT TGACCTAGAG TTGTCGCCAT

3081 AGATCCTTGA GAGTTTTCGC CCCGAAGAAC GTTTTCCAAT GATGAGCACT TT TAAAGTTC TGCTATGTGG  
TCTAGGAACT CTCAAAAGCG GGGCTTCTTG CAAAAGGTTA CTACTCGTGA AAATTCAAG ACGATACACC

3151 CGCGGTATTA TCCCGTATTG ACGCCGGCA AGAGCAACTC GGTCGCCGCA TACACTATTC TCAGAATGAC  
GGCCATAAT AGGCATAAC TGCGGCCCGT TCTCGTTGAG CCAGCGCGGT ATGTGATAAG AGTCTTACTG

3221 TTGGTTGAGT ACTCACCAGT CACAGAAAAG CATCTACGG ATGGCATGAC AGTAAGAGAA TTATGCAGTG  
AACCAACTCA TGAGTGGTCA GTGTCTTTC GTAGAAATGCC TACCGTACTG TCATTCTCTT AATACGTCAC

FIG.\_48F

3291 CTGCCATAAC CATGAGTGAT AACACTGCCG CCAACTTACT TCTGACAACG ATCGGAGGAC CGAAGGAGCT  
GACGGTATTG GTACTCACTA TTGTGACGCC GGTGAAATGA AGACTGTTGC TAGCCTCCTG GCTTCCCTCGA

3361 AACCGCTTTT TTGCACAACA TGGGGGATCA TGTAACTCGC CTTGATCGTT GGAACCCGGA GCTGAATGAA  
TTGGCGAAAA AACGTGTTGT ACCCCCTAGT ACATTGAGCG GAAC TAGCAA CCCTTGGCCT CGACTTACTT

3431 GCCATACCAA ACGACGAGCG TGACACCACG ATGCCCTGTAG CAATGGCAAC AACGTTGCGC AAAC TATTAA  
CGGTATGGTT TGCTGCTCGC ACTGTGGTGC TACGGACATC GTTACC GTTGCAACGCG TTTGATAAATT

3501 CTGGCGAACT ACTTACTCTA GCTTCCCGGC AACAA TTAAT AGACTGGATG GAGGCGGATA AAGTTGCAGG  
GACCGCTTGA TGAATGAGAT CGAAGGGCCG TTGT TAAATA TCTGACCTAC CTCCGCCCTAT TTCAACGTCC

3571 ACCACTTCTG CGCTCGGCC TTCCGGCTGG CTGGTTTATT GCTGATAAAT CTGGAGCCGG TGAGCGTGGG  
TGGTGAAGAC GCGAGCCGGG AAGCCGACC GACCAAATAA CGACTATTTA GACCTCGGCC ACTCGCACCC

3641 TCTCGCGGTA TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT CGTAGTTATC TACACGACGG  
AGAGCGCCAT AGTAACGTCG TGACCCCGGT CTACCATTCG GGAGGGCATA GCATCAATAG ATGTGCTGCC

3711 GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGATCGC TGAGATAGGT GCCTCACTGA TTAAGCATTG  
CCTCAGTCCG TTGATACCTA CTTGCTTTAT CTGTCTAGCG ACTCTATCCA CGGAGTGACT AATTCGTAAC

3781 GTAACGTCA GACCAAGTTT ACTCATATAT ACTTTAGATT GATTTAAAC TTCA TTTTA ATTTAAAGG  
CATTGACAGT CTGGTTCAA TGAGTATATA TGAAATCTAA CTAAATTTTG AAGTAAAAAT TAAATTTTCC

3851 ATCTAGGTGA AGATCCTTTT TGATAATCTC ATGACC AAA TCCCTTAACG TGAGTTTTCG TTCCACTGAG  
TAGATCCACT TCTAGGAAA ACTATTAGAG TACTGGTTT AGGGAATTGC ACTCAAAAGC AAGTGACTC

3921 CGTCAGACCC CGTAGAAAAG ATCAAAGGAT CTTCTTGAGA TCC TTTT TTT CTGCGCGTAA TCTGCTGCTT  
GCAGTCTGGG GCATCTTTTC TAGTTTCCCTA GAAGAACTCT AGGAAAAAA GACGCGCATT AGACGACGAA

3991 GCAAACAAA AAACCAACCG TACCA GCGGT GGT TTGTTG CCGGATCAAG AGCTACCAAC TCTTTTCCG  
CGTTTGTTTT TTTGGTGCG ATGGTCGCCA CCAACAAAC GCCCTAGTTC TCGATGGTTG AGAAAAAGGC

4061 AAGGTAAC TGCTTCAGCAG AGCGCAGATA CCAAATAC TG TCC TTTCTAGT GTAGCCGTAG TTAGGCCACC  
TTCCATTGAC CGAAGTCGTC TCGCGTCTAT GGT TTTATGAC AGGAAGATCA CATCGGCATC AATCCGGTGG

FIG. 48G

4131 ACTTCAAGAA CTCGTAGCA CCGCCTACAT ACCTCGTCT GCTAATCCTG TTACCAGTGG CTGCTGCCAG  
TGAAAGTTCTT GAGACATCGT GCGGATGTA TGGAGCGAGA CGATTAGGAC AATGGTCACC GACGACGGTC

4201 TGGCGATAAG TCGTGTCTTA CCGGGTTGGA CTCAAGACGA TAGTTACCGG ATAAGGCGCA GCGGTCGGGC  
ACCGCTATTTC AGCACAGAAAT GGCCCAACCT GAGTTC TGCT ATCAAATGGCC TATTCGCGT CGCCAGCCCCG

4271 TGAACGGGG GTTCGTGCAC ACAGCCACG TTGGAGCGAA CGACCTACAC CGAACTGAGA TACCTACAGC  
ACTTGCCCC CAAGCACGTG TGTGGGTCG AACCTCGCTT GCTGGATGTG GCTTGACTCT ATGGAATGTCG

4341 GTGAGCTATG AGAAAGCGCC ACGCTTCCC AGGGAGAAA GGCGACAGG TATCCGGTAA GCGGCAGGGT  
CACTCGATAC TCTTTCGCGG TCGGAAGGC TTCCCTCTTT CCGCTGTCC ATAGGCCATT CGCCGTCCCA

4411 CGGAACAGGA GAGGCACGA GGGAGCTTC AGGGGAAAC GCCTGGTATC TTTATAGTCC TGTGCGGTTT  
GCCTTGTCCT CTCGCGTGCT CCTCGAAG CCCTCCCTTTG CGGACCATAG AAATATCAGG ACAGCCCAA

4481 CGCCACCTCT GACTTGAGCG TCGATTTTG TGATGCTCGT CAGGGGGCG GAGCCTATGG AAAACGCCA  
GCGGTGGAGA CTGAACTCGC AGCTAAAAAC ACTACGAGCA GTCCCCCCGC CTCGGATACC TTTTTCGGT

4551 GCAACGCGC CTTTTTACGG TTCCTGCCCT TTTGCTGGC TTTTGCTCAC ATGTTCTTTC CTGCGTTATC  
CGTTGCGCCG GAAAAATGCC AAGGACCGGA AAACGACCGG AAAACGAGTG TACAAGAAAG GACGCAATAG

4621 CCCTGATTCT GTGGATAACC GTATTACCG CTTTGAGTGA GCTGATACCG CTCGCCGCAG CCGAACGACC  
GGGACTAAGA CACCTATTGG CATAATGGC GAAACTCACT CGACTATGGC GAGCGCGTC GGCTTGCTGG

4691 GAGCGCAGC AGTCAGTGAG CGAGGAAGC GAAGAGCGC CAATACGCA ACCGCCCTC CCCGCCGTT  
CTCGCGTCG TCAGTCACTC GCTCCTTCG CTTCTCGCG GTTATGCGTT TGGCGGAGAG GGGCGCGCAA

4761 GGCCGATTCA TTAATGCAGC TGGCACGACA GGTTTCCCGA CTGGAAAGCG GGCAGTGAG GCAACGCAAT  
CCGGCTAAGT AATTACGTG ACCGTGCTGT CCAAAGGGCT GACCTTTCG CCGTCACTCG CGTTGCGTTA

4831 TAATGTGAGT TAGCTCACTC ATTAGGCACC CCAGGCTTTA CACTTTATGC TTCCGGCTCG TATGTTGTGT  
ATTACACTCA ATCGAGTGAG TAATCCGTG GTCCGAAAT GTGAAATACG AAGCCGAGC ATACAACACA

FIG. 48H

4901

BssHII

~~~~~

GGAATTGTGA GCGGATAACA ATTCACACA GGAACACAGCT ATGACCATGA TTACGCCAAG CGCGCAATTA
CCTTAACACT CGCCTATTGT TAAAGTGTGT CCTTGTCTGA TACTGGTACT AATGCGGTTC GCGCGTTAAT

4971

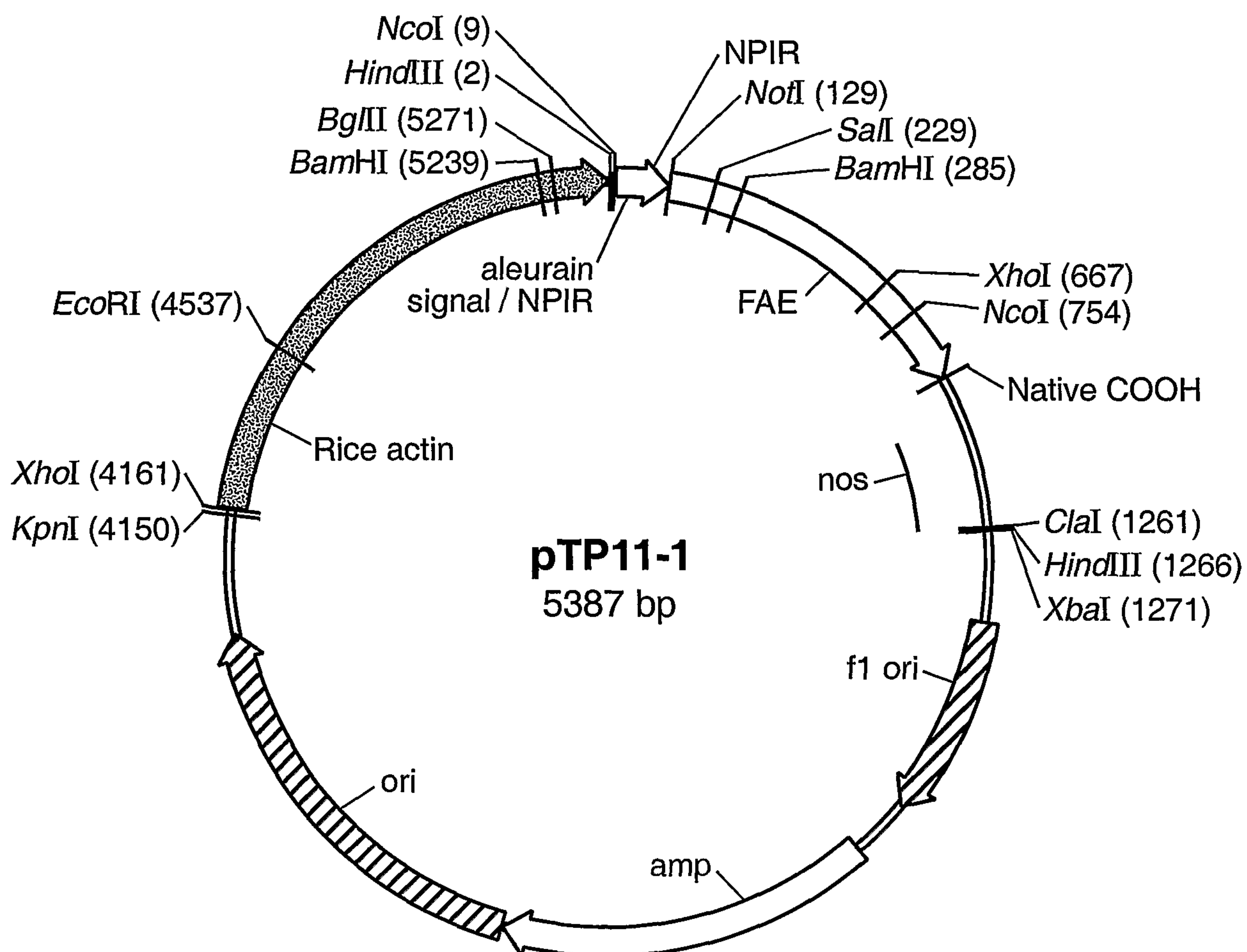
NcoI

KpnI

~~~~~

ACCTTCACTA AAGGGAACAA AAGCTGGGTA C  
TGGGAGTGAT TTCCCTTGTT TTCGACCCAT G

FIG. 48I

**FIG. 49A**

NCOI  
~~~~~  
HindIII
~~~~~  
M A H A R V L L L A L A V L A T A A V A V  
1 AAGCTTACCA TGGCCACGC CCGCGTCCTC CTCCTGGCG TCGCCGTGCT GGCACGGC GCGTCGCCG  
NotI  
~~~~~  
A S S S F A D S N P I R P V T D R A A A S T
71 TCGCCTCCTC CTCCTCCTC GCCACTCA ACCGATCCG GCCGTCACC GACCGCGCG CCGCTCCAC
Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC
SalI
~~~~~  
ACCI  
~~~~~  
D L C N I P S T I I K G E K I Y N S Q T D I N G
211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAAACT GACATTAAACG
BamHI
~~~~~  
W I L R D D S S K E I I T V F R G T G S D T N  
281 GATGGATCCT CCGGACGAC AGCAGCAAAG AATAATCAC CGTCTTCCGT GGACTGGTA GTGATACGAA  
L Q L D T N Y T L T P F D T L P Q C N G C E V  
351 TCTACAACTC GATACTAACT ACACCTCAC GCCTTTCGAC ACCTACCAC AATGCAACGG TTGTGAAGTA  
H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S  
421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA  
Q Y P D Y A L T V T G H X L G A S L A A L T A  
491 GCCAGTATCC GGAATACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC  
A Q L S A T Y D N I R L Y T F G E P R S G N Q  
561 CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG  
XhoI  
~~~~~  
A F A S Y M N D A F Q A S S P D T T Q Y F R V T
631 GCCTTCGGGT CGTACATGAA CGATGCCCTC CAAGCCTCGA GCCAGATAC GACGCAGTAT TTCCGGGTCA

FIG. 49B

FIG. 49C

2451 TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAAACGAT CGGAGGACCG AAGGAGCTAA CCGCTTTTTC
2521 GCACAACATG GGGATCATG TAACTCGCCT TGATCGTTGG GAACCGGAGC TGAATGAAGC CATACCAAC
2591 GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAACT GCGAACTAC
2661 TTACTCTAGC TTCCCGGCAA CAATTAAATAG ACTGGATGGA GCGGATAAA GGTGCAGGAC CACTTCTGCG
2731 CTCGGCCCTT CCGCTGGCT TGGTAAAGCC TCCCGTATCG TAGTTATCTA CACGACGGG AGTCAGGCAA
2801 ATTGCAGCAC TGGGGCCAGA ACGAAATAGA CAGATCGCTG TTTAGATTGA TTTAAAACCTT CATTTTAAAT CTAGGTGAAG
2871 CTATGGATGA ACATAATATAC TTAAGATTGA GACCAAATC CTTAAACGTG AGTTTTCGTT CCACCTGAGCG TCAGACCCCG
2941 CCAAGTTTAC ATCCTTTTIG ATAACTTCAT CAAAGGATCT TCTTGAGATC CTTTTTTCT GCGGTAAATC TGCTGCTTGC AAACAATAAA
3011 ATCCTTTTIG ATAACTTCAT CAAAGGATCT TCTTGAGATC CTTTTTTCT GGATCAAGAG CTACCAACTC TTTTTCGAA GGTAACCTGGC
3081 TAGAAAAGAT CCAAGGATCT CAAAGGATCT TCTTGAGATC CTTTTTTCT GGATCAAGAG CTACCAACTC TTTTTCGAA GGTAACCTGGC
3151 ACCACCGCTA CCAGCGGTGG TTTGTTTGCC TTTGTTTGCC GGATCAAGAG CTACCAACTC TTTTTCGAA GGTAACCTGGC
3221 TTCAGCAGAG CGCAGATACC AAATACTGTC CTTCTAGTGT AGCCGTAGTT AGCCACAC CAGTACAGCTT GAGCTATGAG
3291 CTGTAGCAC CCCTACATAC CTCGCTCTGC TAATCCTGTT ACCAGTGGCT TCCGGTAAGC TCCGGTAAAGC GCGATAAGTC
3361 GTGCTTACC GGTGACTT CAAGACGATA GTTACCAGGAT AAGCGCAGC TACCGGCTG GGTCCAGTG AACGGGGGT
3431 TCGTGACAC AGCCAGCTT GGAGCGAACG CCGACAGGTA TCCGGTAAAGC TCCGGTAAAGC TCCGGTAAAGC GCGATAAGTC
3501 AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CCGACAGGTA TCCGGTAAAGC TCCGGTAAAGC TCCGGTAAAGC GCGATAAGTC
3571 GCGCAGGAG GAGCTTCCAG GATTTTTGTG ATGCTCGTCA GGGGGCGGA GCTATGGAA GCTATGGAA GCTATGGAA GCGATAAGTC
3641 CTTGAGCGTC GATTTTTGTG ATGCTCGTCA GGGGGCGGA GCTATGGAA GCTATGGAA GCTATGGAA GCGATAAGTC
3711 TTTTACGGTT CTTGGCCCTT TGCTGGCCCTT TGCTGCACAT TTGCTGCACAT TTGCTGCACAT TTGCTGCACAT TTGCTGCACAT
3781 GGATAACCGT ATTACCGCTT TTGAGTGAGC TGATACCGCT TGATACCGCT TGATACCGCT TGATACCGCT TGATACCGCT
3851 TCAGTGAGCG AGGAAGCGGA AGAGCGCCA ATACGCAAC ATACGCAAC ATACGCAAC ATACGCAAC ATACGCAAC
3921 AATGCAGCTG GCACGACAGG TTTCCTCGACT GGAAGCGGG CAGTGAGCGC CAGTGAGCGC CAGTGAGCGC CAGTGAGCGC
3991 GCTCACTCAT TAGGCACCC AGGCTTTACA CTTATGCTT CCGCTCGTA CCGCTCGTA CCGCTCGTA CCGCTCGTA
4061 GGATAACAA TTCACACAGG AACAGCTAT GACCATGATT ACGCCAAGCG ACGCCAAGCG ACGCCAAGCG ACGCCAAGCG

KpnI

XhoI

4131 GGGAAACAAA GCTGGGTACC GGGCCCCC TCAGGTCAT TCATATGCTT GAGAAGAGAG TCGGGATAGT
4201 CCAAAATAAA ACAAGGTAA GATTACCTGG TCAAAAGTGA AAACATCAGT TAAAAGGTGG TATAAGTAAA
4271 ATATCGGTAA TAAAAGGTGG CCAAAAGTGA AATTACTCT TTTCTACTAT TATAAAATTT GAGGATGTTT
4341 TGTCCGGTACT TTGATACGTC ATTTTGTAT GAATTGGTTT TTAAGTTTAT TCGCGATTG GAAATGCATA
4411 TCTGTATTG AGTCGGTTTT TAAGTTTCGTT GCTTTTGTAA ATACAGAGGG ATTTGTATAA GAAATATCTT

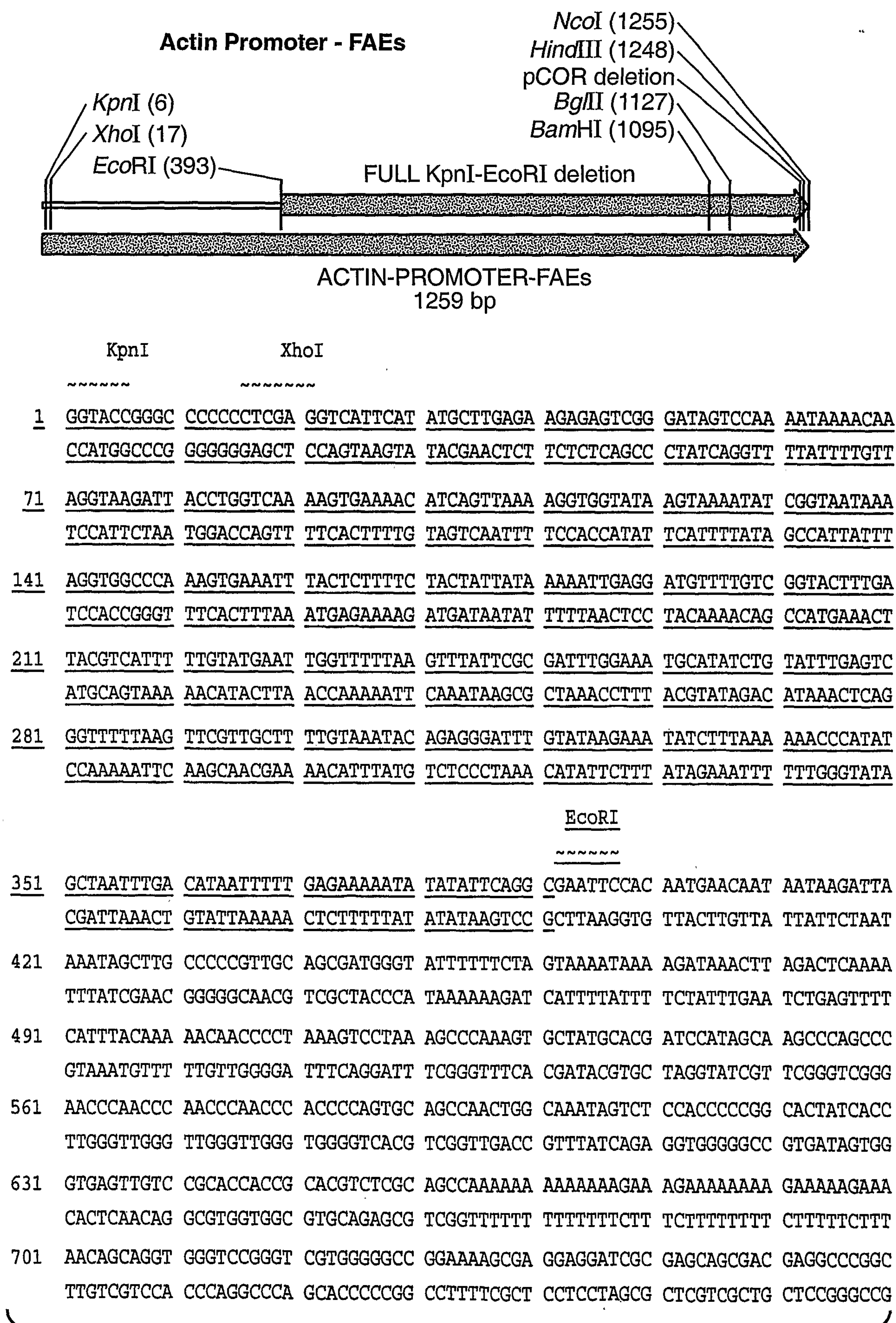
ECORI

4481 TAAAAAACCC ATATGCTAAT TTGACATAAT TTTTGAGAAA AATATATATT CAGGCGAATT CCACAATGAA
4551 CAATAATAAG ATTAATAATAG CTTGCCCCCG TTGCAGCGAT GGGTATTTT TCTAGTAAAA TAAAAGATAA
4621 ACTTAGACTC AAAACATTTA CAAAACAAC CCTAAAGTC CTAAAGCCCA AAGTGCTATG CACGATCCAT

FIG. 49D

4691 AGCAAGCCCA GCCAAACCCA ACCCAACCA ACCCAACCA GTGAGCCAA CTGGCAATA GTCTCCACCC
4761 CCGGCACTAT CACCGTGAGT TGTCGCACAC ACCGCACGTC TCGCAGCCAA AAAAAAATA AGAAAGAAAA
4831 AAAAGAAAA GAAAAACAGC AGGTGGGTCC AGGTGGGTCC GCGCGGAAAA GCGAGGAGGA TCGCGAGCAG
4901 CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCATCGCC ACTATAACA TACCCCCCCC
4971 TCTCCTCCCA TCCCCCACA CCTACCACA CCCTCCTCC CACCTCCTG CCCCTCGCTG CCGACGACG
5041 AGCTCCTCCC CCTCCTCCT CCTCCTCCT CCGTAACCA CCGGCCCTC TCCTCTTTCT TTCTCCGTTT
5111 TTTTTTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT GGTGGGCGA GAGGGCTTC GTCGCCAGA
BamHI
~~~~~  
5181 TCGGTGCGG GGAGGGGCGG GATCTGCGG CTGGCGTCTC CCGGCGTGAG TCGGCCCGA TCCTCGCGG  
BglII  
~~~~~  
5251 GAATGGGGCT CTCGGATGTA GATCTTCTTT CTCTCTCTTT TTTGTGGTAG AATTGAATC CCTCAGCAT
5321 GTTCATCGGT AGTTTTCCTT TTATGATTT TGACAAATG CAGCCTCGTG CCGAGCTTTT TTGTAGC

FIG. 49E

**FIG. 50A**

771 CCTCCCTCCG CTTCCAAAGA AACGCCCCC ATCGCCACTA TATACATACC CCCCCTCTC CTCCCATCCC
GGAGGGAGGC GAAGGTTTCT TTGCGGGGG TAGCGGTGAT ATATGTATGG GGGGGGAGAG GAGGGTAGGG

841 CCCAACCCTA CCACCACCAC CACCACCACC TCCTCCCCC TCGCTGCCGG ACGACGAGCT CCTCCCCCT
GGGTGGGAT GGTGGTGGTG GTGGTGGTGG AGGAGGGGG AGCGACGGCC TGCTGCTCGA GGAGGGGGGA

911 CCCCCTCCGC CGCCGCCGGT AACCACCCCG CCCCTCTCCT CTTTCTTTCT CCGTTTTTTT TTTCGTCTCG
GGGGGAGGCG GCGGCGGCCA TTGGTGGGGC GGGGAGAGGA GAAAGAAAGA GGCAAAAAA AAAGCAGAGC

981 GTCTCGATCT TTGGCCTTGG TAGTTTGGGT GGGCGAGAGC GGCTTCGTCTG CCCAGATCGG TGC GCGGGAG
CAGAGCTAGA AACCGGAACC ATCAAACCCA CCCGCTCTCG CCGAAGCAGC GGGTCTAGCC ACGCGCCCTC

BamHI

~~~~~

1051 GGGCGGGATC TCGCGGCTGG CGTCTCCGGG CGTGAGTCGG CCCGGATCCT CGCGGGGAAT GGGGCTCTCG  
CCCGCCCTAG AGCGCCGACC GCAGAGGCC GCACTCAGCC GGGCCTAGGA GCGCCCCTTA CCGGAGAGC

BglIII

~~~~~

1121 GATGTAGATC TTCTTTCTTT CTTCTTTTGG TGGTAGAATT TGAATCCCTC AGCATTGTTC ATCGGTAGTT
CTACATCTAG AAGAAAGAAA GAAGAAAAC ACCATCTTAA ACTTAGGGAG TCGTAACAAG TAGCCATCAA

HindIII NcoI

~~~~~

1191 TTTCTTTTCA TGATTTGTGA CAAATGCAGC CTCGTGCGGA GCTTTTTTGT **AGGTAGA**AGC TTACCATGG  
AAAGAAAAGT ACTAAACACT GTTTACGTCG GAGCACGCCT CGAAAAACA **TCCATC**TTCG AATGGTACC

KpnI-EcoRI - deletion underlined and restored NCO site in bold in vectors pJQ4.9,  
pJQ3.2 and pJO6.3.

**FIG.\_50B**

ALEURAIN\_deleted NPIR (Apoplast) Structure and Sequence

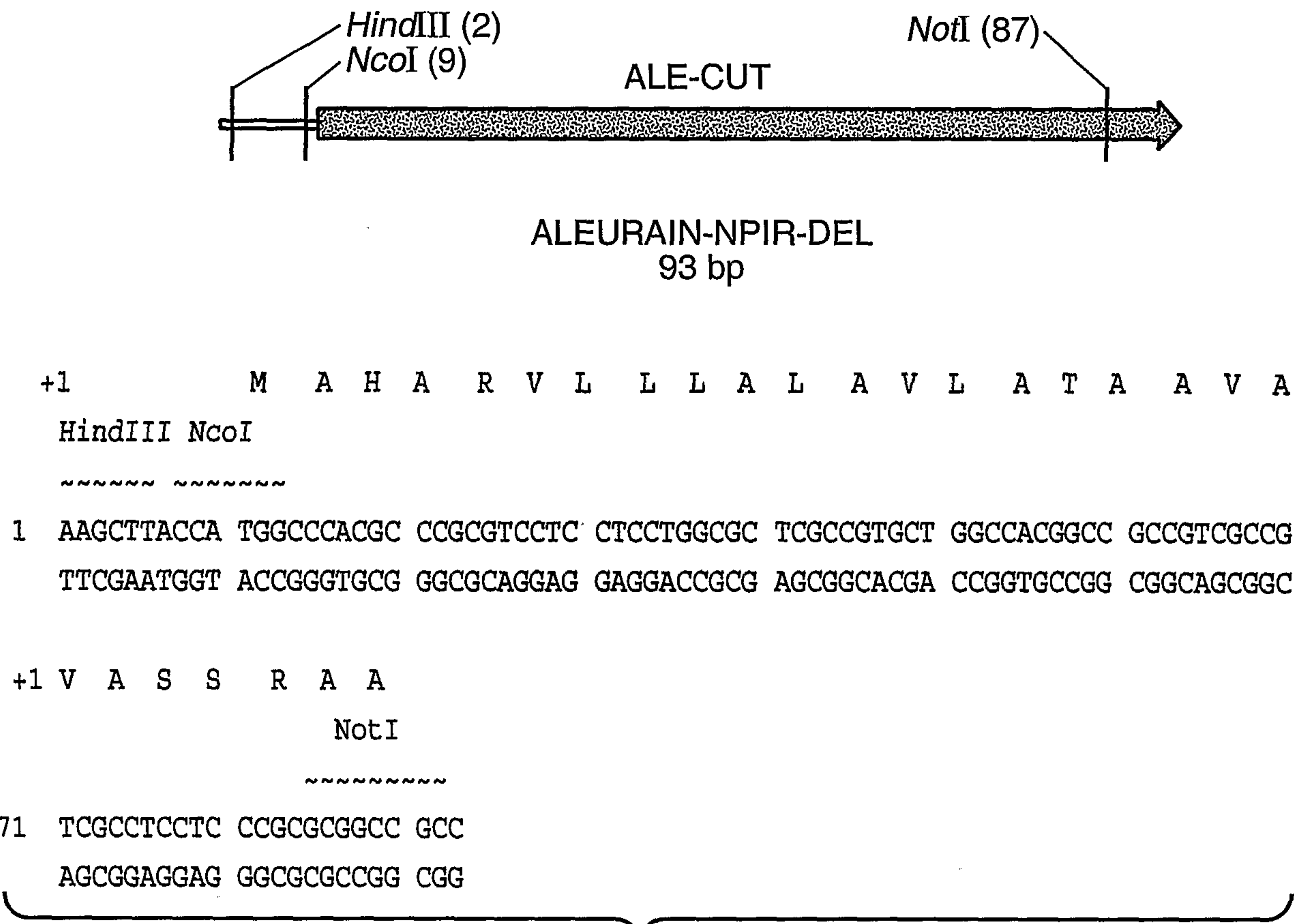


FIG.\_51

**SEE1 ( Senescence enhanced ) PROMOTER sequence**

```
1   CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
      PstI                      XbaI      XbaI
      ~~~~~~                ~~~~~~ ~~~~~~
71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGTG ATCTAGAACT CTAGATAGCA CAGCCACAGC
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTCGCGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
 SmaI
      ~~~~~~
351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
      SphI
      ~~~~~~
491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTTTGG CTTAATTGAC TTTATTTTGG
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG
631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
 PstI
      ~~~~~~
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCggcacc GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCCTCGC TCCCCCTGCC GGACGACCCA
841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATG
```

**FIG. 52****SEE1 ( Senescence enhanced ) PROMOTER plus vacuolar aleurain SIGNAL/NPIR sequence**

```
1   CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
      PstI                      XbaI      XbaI
      ~~~~~~                ~~~~~~ ~~~~~~
71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGTG ATCTAGAACT CTAGATAGCA CAGCCACAGC
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTCGCGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
 SmaI
      ~~~~~~
351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
      SphI
      ~~~~~~
491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTTTGG CTTAATTGAC TTTATTTTGG
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG
631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
 PstI
      ~~~~~~
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCggcacc GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCCTCGC TCCCCCTGCC GGACGACCCA
      M A H G R I L F L A L A V L
841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATGGCCACG GCCGCATCCT CTTCTTGGCG CTCGCCGTCT
      BssHII
      NotI
      . A T A A V A A A S L A D S N P I R P V T E R A .
911 TGGCCACCGC CGCGGTGGCC GCCGCATCNT TGGCGGACTC CAACCCGATC CGGCCCGTCA CCGAGCGCGC
      NotI
      ~~~~~~
 . A A
981 GGCCGCC
```

**FIG. 53**