



(86) Date de dépôt PCT/PCT Filing Date: 2001/11/16
(87) Date publication PCT/PCT Publication Date: 2002/09/06
(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 (60/249,608) US

(51) Cl.Int.⁷/Int.Cl.⁷ C12N 15/82
(71) Demandeur/Applicant:
GENENCOR INTERNATIONAL, INC., US
(72) Inventeurs/Inventors:
DUNN-COLEMAN, NIGEL, US;
LANGDON, TIMOTHY, GB;
MORRIS, PHILLIP, GB
(74) Agent: BERESKIN & PARR

(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 plannts, 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

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

10

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

15

CROSS-REFERENCE TO RELATED APPLICATIONS

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.

20

FIELD OF THE INVENTION

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

25

BACKGROUND OF THE INVENTION

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
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.

30

35

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:__), with
5 highlighting to point out the signal sequence, intron and various restriction endonuclease sites, and amino acid sequence (SEQ. ID. NO:__) 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
10 esterase (SEQ. ID. NO:__).

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
FAE-I3 creates two possible uninterrupted reading frames – the top in the figure below is
15 functional (highlighted serine is at active site), the bottom is inactivated.

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.

20 Figure 7 illustrates pCOR105.

Figure 8 illustrates a generic ALE-TER vector.

Figure 9 illustrates the KDEL-COOH ER retention sequences.

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

25 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.

Figure 14 illustrates the rat sialyl transferase structure and sequence.

30 Figure 15 illustrates the potato protease inhibitor II (PPI) motif structure and sequence.

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.

5 Figure 33 is a schematic of the pUA4-4 vector. Also shown is the 5345 bp nucleotide sequence of the vector.

Figure 34 is a schematic of the pTU4 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

10 Figure 35 is a schematic of the pTT5.14 vector. Also shown is the 5395 bp nucleotide sequence of the vector.

Figure 36 is a schematic of the pTP8-5 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

Figure 37 is a schematic of the pTP5-1 vector. Also shown is the 5277 bp nucleotide sequence of the vector.

15 Figure 38 is a schematic of the pTP4a2 vector. Also shown is the 5327 bp nucleotide sequence of the vector.

Figure 39 is a schematic of the pTP3-1 vector. Also shown is the 5338 bp nucleotide sequence of the vector.

20 Figure 40 is a schematic of the pTU5 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

Figure 41 is a schematic of the pGT6 vector. Also shown is the 4773 bp nucleotide sequence of the vector.

Figure 42 is a schematic of the pJQ5 vector. Also shown is the 5034 bp nucleotide sequence of the vector.

25 Figure 43 is a schematic of the pJO6.1 vector. Also shown is the 4950 bp nucleotide sequence of the vector.

Figure 44 is a schematic of the pJQ4 vector. Also shown is the 4974 bp nucleotide sequence of the vector.

30 Figure 45 is a schematic of the pPQ10.1 vector. Also shown is the 5164 bp nucleotide sequence of the vector.

Figure 46 is a schematic of the pJQ3 vector. Also shown is the 4965 bp nucleotide sequence of the vector.

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. All patents and publications, including all sequences disclosed within such patents and publications, referred
20 to herein are expressly incorporated by reference.

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
25 (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
30 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

orientation, respectively. Practitioners are particularly directed to Sambrook *et al.*, 1989, and Ausubel FM *et al.*, 1993, for definitions 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.

5 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 defined by reference to the specification as a whole.

Definitions

10 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 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
15 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 modified variants refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not
20 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, 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
25 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 polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a polynucleotide (except AUG,
30 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, *Proteins* (1984)).

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

The phrase "controlling the level of phenolic acids" refers to the manipulation of phenolic acid expression in plants, particularly plant cell walls.
 25 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 during plant growth or after plant growth, *e.g.*, after a plant has been cut or pulled
 30 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.

10 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 (Batzer *et al.*, *Nucleic Acid Res.* **19**:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* **260**:2605-2608 (1985); and Cassol *et al.*, 1992; Rossolini *et al.*, *Mol. Cell. Probes* **8**:91-98 (1994)). The term polynucleotide is used interchangeably with
30 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 US application 08/952,445 which describes the isolation of a FAE from *Aspergillus niger*, and copending US application 09/658,772 which describes the isolation of a xylanase from *T. reesei*.

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.

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)).

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: ____)

5 GGCGCCGAGGGAGTGGCCGGTCACGGTCAGCGCGTAGTCC 40-mer

FAE-I3 (SEQ ID NO: ____)

CCGGCCACGCCCTCGGGCGCTCCCTGGCGGACTC 35-mer

FAE-N5 (SEQ ID NO: ____)

CTAAAGCTTACCATGGCGGCCGCCTCCACGCAGGGGCATCTCCGA 44-mer

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

CTAAAGCTTAACATGAAGCAGTTCTCCGCCAA 32-mer

FAE-3 (SEQ ID NO: ____)

TCTAAGCTTGCGGCCGCGACCGGCCAGGTGCATGCGCCGCTCGTCATCCC
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
25 methods as described 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).

30

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:____):

(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, rather than EPLKDEL 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, 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).

5 The aleurain signal was also 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
15 mutagenesis to eliminate N-glycosylation and inactivation by deamidation; see sequence data for exact change, vector **pTP10.1**).

PCR primers

20 **TER-5 (SEQ ID NO:___)**
AGACTGCAGACCATGGCGGCCGCG**KA**ACCACTGAAGGATGAGCTGTAAAG
AAGCAGATCGTTCAAACATTTG 72-MER (The KDEL stop codon is underlined.)
TER-NOT (SEQ ID NO:___)
AAGACTGCAGACCATGGCGG 20-MER
25 **TER-3 (SEQ ID NO:___)**
AGATCTAGAAGCTTATCGATCTAGTAACATAGATGACACC
ALECUT (SEQ ID NO:___)
CTAGGCGGCCGCGCGGGAGGAGGCGACGGCGAC
GLYB (SEQ ID NO:___)
30 GAGGGTGTATTCGGTATCGAGTTGCAGGTTTCGTATC
GLY3 (SEQ ID NO:___)
CTCGATACCCATTACACCCTCACGCCTTTTCA

35 *B. Construction of different promoter vectors*

Various promoters were used to optimize expression and to establish 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, deleted bases underlined) and may affect splicing at the adjacent 3' site. The original rice actin
 5 sequence in this region (GTAGGTAAG) was therefore restored using oligonucleotide **NCO-ACT** (CTCACCATGGTAAGCTTCTACC TACAAAAAAGCTCCGCA) 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
 10 promoter used in **pCOR105**; as this may have unpredictable effects on vector expression it was 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
 15 into pGT6 to produce construct **pJO6.3**.

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 -
 20 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 -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 was amplified from *Lolium multiflorum* cv Tribune with oligonucleotides SEE-NCO
 30 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 multiflorum* PhD Thesis University of Wales, Aberystwyth) and has been shown to be senescence
5 inducible 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
10 the apoplast target sequence with the gene to be expressed inserted downstream of
the apoplast sequence.

PCR Primers

SEE-VAC (SEQ ID NO: ____)

15 AACCATGGCGGCCGCGCGCTCGGTGACGGGCCGGAT

SEE-NCO (SEQ ID NO: ____)

TTCGGTACCATGGCCAGGTATAATTATGG

SEE-ATG (SEQ ID NO: ____)

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-
25 terminus or to the C-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.

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

Peptide sequence is MAAASTQGI (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
10 of dNTPs according to 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 recovered. Two versions of
15 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: ____)

20 GGAATTCGTAGACAAGCTTACMATGGCCCACGCCCGCGTCCT 41-MER

ALE-3 (SEQ ID NO: ____)

TATCCATGGCGGCCGCGCGGTCTGGTGACGGGCGGMYCGGGTTGGAGTC
GGCGAA 55-MER

ALE-CUT (SEQ ID NO: ____)

25 CTAGGCGGCCGCGCGGGAGGAGGCGACGGCGAC 33-mer

ALECAP-5 (SEQ ID NO: ____)

GCGACGGCGACGGCGGCCGTGGCCAGCACGGCGAGCGCCAGGAGGAGG
ACGCGG 54-MER

ALECAP-3 (SEQ ID NO: ____)

30 TCGCCGTCGCCTCCTCCTCCTCCTTCGCCGACT 33-MER

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

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

5 A Golgi targeting vector, **pJQ3.2**, was made by inserting a reading frame 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
10 EcoRI/NotI promoter/signal fragment of **pJO6.3** with the fragment from this vector. 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
15 deletion which was corrected by PCR with RST-RPT, followed by overlap-PCR and cloning of products.

PCR primers

RST-5AD (SEQ ID NO: ____)

ACTAAGCTTAAGGAGATATAACAATGATCCACACCAACCTCAA

20 **RST-F1A (SEQ ID NO: ____)**

TTCCATGATCCACACCAACCTCAAAAAGAAGTTCTCCCTCTTCAT

RST-F1B (SEQ ID NO: ____)

AGAGTGATCACGGCGAAGAGGAGGAAGACGAGGATGAAGAGGGGAGAACTTCT
TTT

25 **RST-F2A (SEQ ID NO: ____)**

TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGCCCTCACCTCCAA
GCCAAGGA

RST-F2B (SEQ ID NO: ____)

CATTGGAACCTCCTTGGCTTGGAGGGTG

30 **RST-3A (SEQ ID NO: ____)**

AACCATGGCGGCCGCCATTGGAACCTCCTTGGCT

RST-RPT (SEQ ID NO: ____)

TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGCCCTCACCTCC
AAGCCAAGGA

(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
10 cloning this product as a HindIII/NotI fragment into vector pPQ10.1; the EcoRI/NotI promoter/signal fragment of pJO6.3 was then replaced with the equivalent fragment from the modified pPQ10.1 vector.

PCR primers

15 PPI-AP1 (SEQ ID NO: ____)

GGAATTCGTAGACAAGCTTACMATGGMCGTGCACAAGGAGGT

PPI-AP2 (SEQ ID NO: ____)

GATCAGGAGGTAGGCWACGAAGTTWACCTCCTTGTGC

PPI-AP3 (SEQ ID NO: ____)

20 CCTACCTCCTGATCGTSCCTCGGCCTCCTCTTGCTCGT

PPI-AP4 (SEQ ID NO: ____)

CCTTGCGTCCACGTGCTCCATGGCGGAWACGAGCAAGAGGAG

PPI-AP5 (SEQ ID NO: ____)

GTGGACGCCAAGGCCTGCACCKCGAGTGCGGCAACCTC

25 PPI-AP6 (SEQ ID NO: ____)

GGAATTCGCGGCCGCGGGCAGATGCCGAAGCCGAGGTTGCCGCACT

ii. C-terminal end signal sequences

Four C-terminal signal sequences were utilized:

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

This was derived directly from the genomic clone (see Example 1) as a Nco1-Sph1 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
10 sequences are expected to be neutral in this respect while the native Aspergillus end may not be.

- (c) Linker plus KPLKDEL [first K is primer artifact, intended to be E] {ER retention vectors)

These sequences are provided by primer TER5 introduced during PCR to
15 generate the nos terminator fragment, and identified by sequencing within a specific 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 [frameshift of (c)] (loss of ER retention - vacuole
20 vectors)

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 signal is for ER retention, while others provide controls. A
25 frameshift in the TER5 region [additional A] was used in subsequent constructs to destroy the ER KDEL retention signal.

The linker used in the above C-terminal targeting sequences was PVAAA.

D. Co-integration and co-transformation vectors.

30

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
15 to FAE to create **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 AccI and T4 polymerase, polished,
20 followed by NotI digest and cloning into EcoRV/ NotI digested pTR2.22 to give clones **pTT5.5** and **5.1**.

PCR primer

ALE-G

TATCCATGGCGGCCGCGCGGTCTGGTGACGGGCGGCCCGGGTTGGAGTC
25 GGCGAA

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
35 was carried out as follows. From pTP5.1 for **pUF1**, from **pTP11.1** for pUH4, from pTP8.5 for **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.

What is claimed:

1. A transgenic plant comprising an expression cassette comprising a promoter operably linked to a ferulic acid esterase encoding polynucleotide.
5
2. The plant of claim 1, wherein the polynucleotide is derived from *Aspergillus niger*.
3. The plant of claim 2, wherein the polynucleotide is FAE ! from *Aspergillus niger*.
- 10 4. The plant of claim 3, wherein the polynucleotide encodes a ferulic acid esterase with an altered glycosylation site.
5. The plant of claim 3, wherein the polynucleotide encodes a ferulic acid esterase with a substitution so that glycosylation is altered.
6. The plant of claim 3, wherein the polynucleotide further
15 comprises a polynucleotide that encodes CTWPVAAA at the 3' end.
7. The plant of claim 3 wherein sub-optimal codons are modified to *Triticum spp.* preferred codons.
8. The plant of claim 1, wherein the introduction of the ferulic acid esterase polynucleotide into the plant is by sexual reproduction.
- 20 9. The plant of claim 1, wherein the promoter is an inducible promoter.
10. The plant of claim 9, wherein the promoter is a senescence promoter.
11. The plant of claim 9, wherein the promoter is a heat shock
25 promoter.
12. The plant of claim 1, wherein the promoter is a constitutive promoter
13. The plant of claim 1, wherein the expression cassette further comprises a polynucleotide sequence that targets expression of the

polynucleotide.

14. The plant of claim 13, wherein the polynucleotide sequence is upstream of the N-terminus of the ferulic acid esterase polynucleotide.

15. The plant of claim 14, wherein the polynucleotide is derived
5 from the signal sequence of a vacuolar targeted gene

16. The plant of claim 15, wherein the targeted gene is a barley aleurain gene.

17. The plant of claim 15, wherein the vacuolar signal sequence of the polynucleotide is modified to produce a endoplasmic reticulum or apoplast
10 signal sequence.

18. The plant of claim 15, wherein the polynucleotide is derived from the signal sequence of a vacuolar targeted senescence gene.

19. The plant of claim 18, wherein the senescence gene is a *Lolium See1* signal sequence.

20. The plant of claim 13, wherein the polynucleotide is derived
15 from the signal sequence of a golgi targeted gene.

21. The plant of claim 20, wherein the targeted gene is a rat sialyl transferease signal sequence.

22. The plant of claim 13, wherein the polynucleotide is derived
20 from the signal sequence of an apoplast signal sequence.

23. The plant of claim 22, wherein the signal sequence is from *Aspergillus niger* ferulic acid esterase.

24. The plant of claim 16, wherein the polynucleotide is derived from *Solanum tuberosum*.

25. The plant of claim 13, wherein the polynucleotide sequence is downstream of the C-terminus of the ferulic acid esterase polynucleotide

26. The plant of claim 25, wherein the polynucleotide sequence is a KDEL sequence.

27. The plant of claim 25, wherein the polynucleotide sequence is a stop codon.

28. The plant of claim 25, wherein the polynucleotide sequence is an extension of the ferulic acid esterase reading frame to provide a linker to
5 KDEL.

29. The plant of claim 1, further comprising introduction into the plant a second expression cassette comprising a promoter operably linked to a xylanase encoding polynucleotide.

30. The plant of claim 29, wherein the xylanase encoding
10 polynucleotide is from *Trichoderma reesei*.

31. The plant of claim 29, wherein the first and second expression cassettes are present on separate plasmids.

32. The transgenic plant of claim 1, selected from the group consisting of *Festuca*, *Lolium*, *Zea* and *Avena*.

33. The transgenic plant of claim 32, wherein the plant is a
15 *Festuca* plant.

34. A method of controlling the level of phenolic acids in plant cell walls of a transgenic plant, the method comprising introducing into the plant an expression cassette comprising a promoter operably linked to a ferulic acid
20 esterase encoding polynucleotide.

35. The method of claim 34, wherein the polynucleotide is derived from *Aspergillus niger*.

36. The method of claim 35, wherein the polynucleotide is a FAE
1 gene from *Aspergillus niger*.

37. The method of claim 36, wherein the polynucleotide encodes
25 the ferulic acid esterase with an altered glycosylation site.

38. The method of claim 36, wherein the polynucleotide encodes the ferulic acid esterase with a substitution such that glycosylation is altered.

39. The method of claim 36, wherein the polynucleotide

comprises CTWPVAAA at the 3' end.

40. The method of claim 36 wherein sub-optimal codons are modified to *Triticum spp.* preferred codons.

41. The method of claim 36, wherein the polynucleotide
5 comprises SEQ ID NO:1.

42. The method of claim 34, wherein the introduction of the ferulic acid esterase polynucleotide into the plant is by transformation of cell cultures.

43. The method of claim 42, wherein the cell cultures are
10 regenerated to plants.

44. The method of claim 34 wherein the ferulic acid esterase polynucleotide is introduced into the plant by sexual reproduction.

45. The method of claim 34, wherein the transgenic plant is a member of a genus selected from the group consisting of *Festuca*, *Lolium*,
15 *Avena* and *Zea*.

46. The method of claim 45, wherein the transgenic plant is a member of the genus *Festuca*.

47. The method of claim 46, wherein the transgenic plant is a *Festuca arundinacea*.

48. The method of claim 34, wherein the promoter is an
20 inducible promoter.

49. The method of claim 48, wherein the promoter is a senescence promoter.

50. The method of claim 48, wherein the promoter is a heat
25 shock protein promoter.

51. The method of claim 34, wherein the promoter is a constitutive promoter.

52. The method of claim 51, wherein the promoter is an actin

promoter.

53. The method of claim 34, wherein the expression cassette further comprises a polynucleotide sequence that targets expression of the polynucleotide.

5 54. The method of claim 53, wherein the polynucleotide sequence is upstream of the N-terminus of the ferulic acid esterase polynucleotide.

55. The method of claim 54, wherein the polynucleotide is derived from the signal sequence of a vacuolar targeted gene.

10 56. The method of claim 55, wherein the targeted gene is a barley aleurain gene.

57. The method of claim 55, wherein the polynucleotide is derived from the signal sequence of a *Lolium* See1 signal sequence.

15 58. The method of claim 55, wherein the vacuolar signal sequence of the polynucleotide is modified to produce a endoplasmic reticulum signal sequence.

59. The method of claim 55, wherein the vacuolar signal sequence of the polynucleotide is modified to produce an apoplast signal sequence.

20 60. The method of claim 54, wherein the polynucleotide is derived from the signal sequence of a golgi targeted gene.

61. The method of claim 60, wherein the targeted gene is a rat sialyl transferease signal sequence.

25 62. The method of claim 59, wherein the polynucleotide is derived from the signal sequence of a fungal apoplast signal sequence.

63. The method of claim 62, wherein the signal sequence is from *Aspergillus niger* ferulic acid esterase.

64. The method of claim 59, wherein the apoplast signal

sequence is from a potato.

65. The method of claim 53, wherein the polynucleotide sequence is downstream of the C-terminus of the ferulic acid esterase polynucleotide

5 66. The method of claim 65, wherein the polynucleotide sequence is a KDEL sequence.

67. The method of claim 65, wherein the polynucleotide sequence is a stop codon.

68. The method of claim 65, wherein the polynucleotide
10 sequence is an extension of the ferulic acid esterase reading frame to provide a linker to KDEL.

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

15 70. The method of claim 69, wherein the second polynucleotide is a fungal xylanase.

71. The method of claim 70, wherein the fungal xylanase is from *Trichoderma reesei*.

72. The method of claim 35, wherein the first and second
20 expression cassettes are present on separate plasmids.

73. The method of claim 1, wherein the first and second expression cassettes are present on separate plasmids.

74. A transgenic plant produced by the method of claim 34.

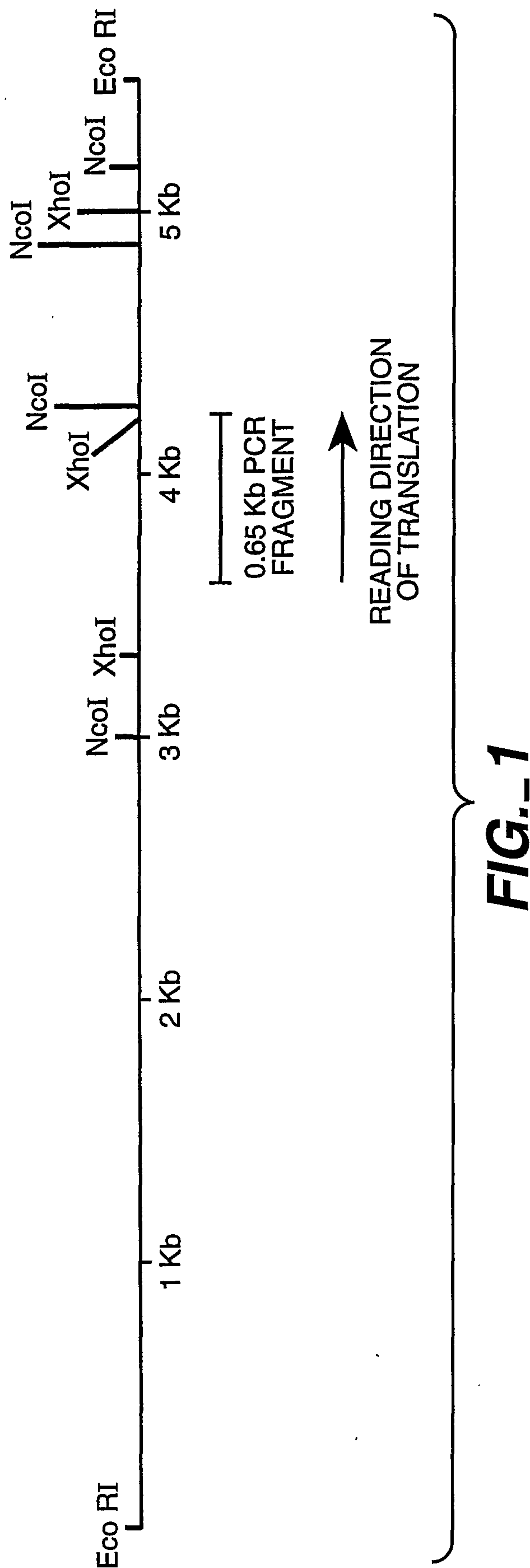


FIG. 2A

FIG. 2B

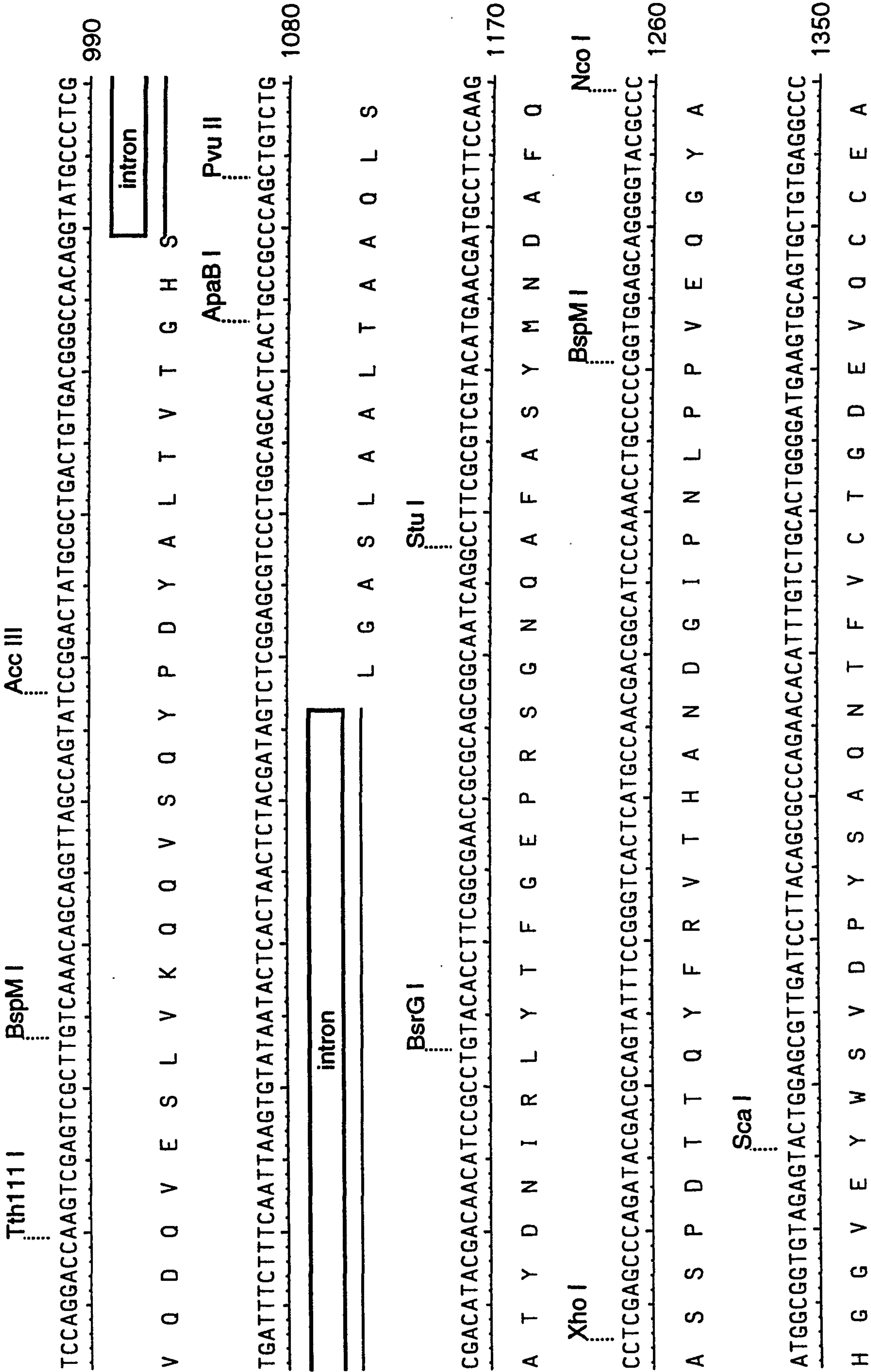


FIG._2C

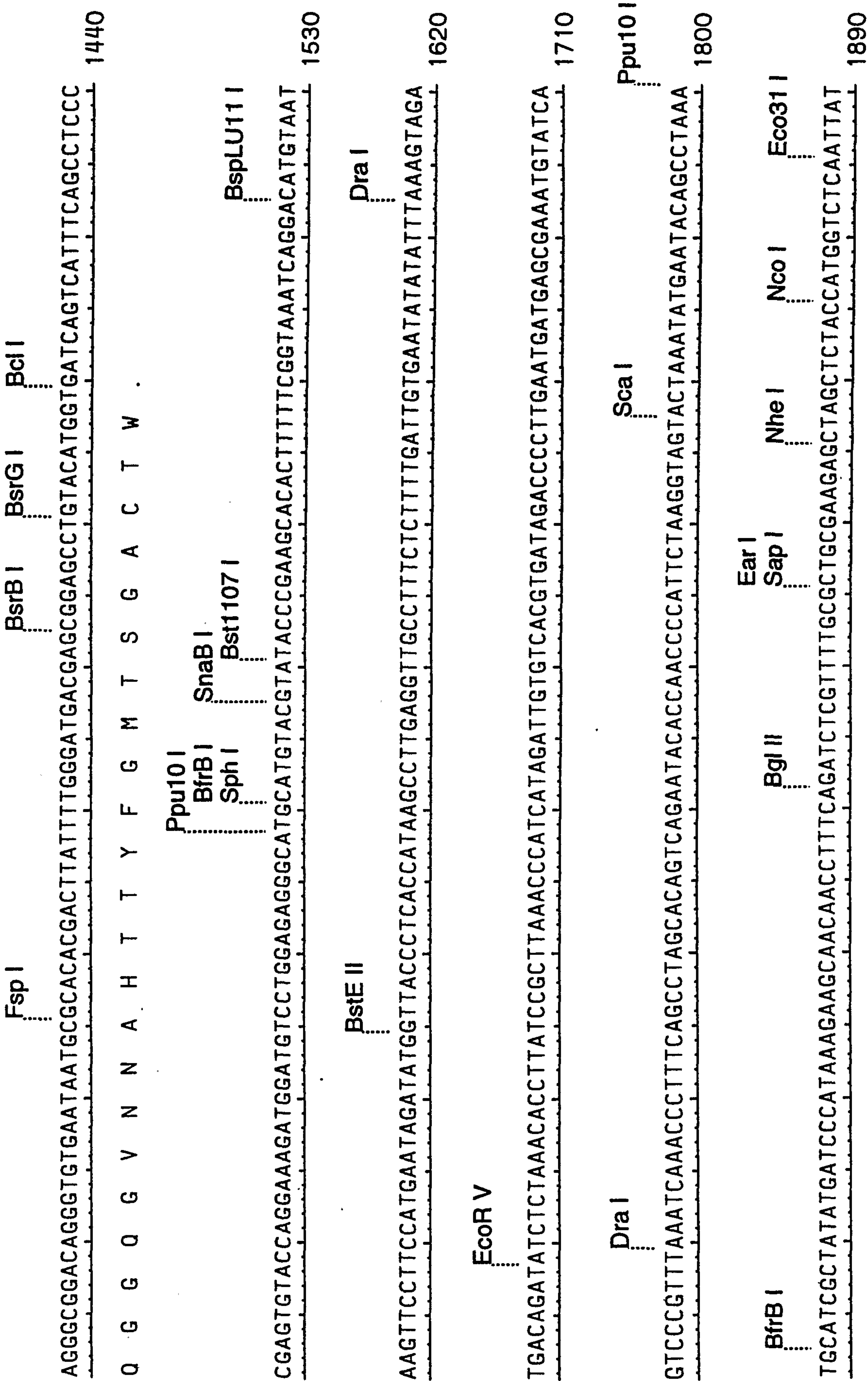


FIG.-2D

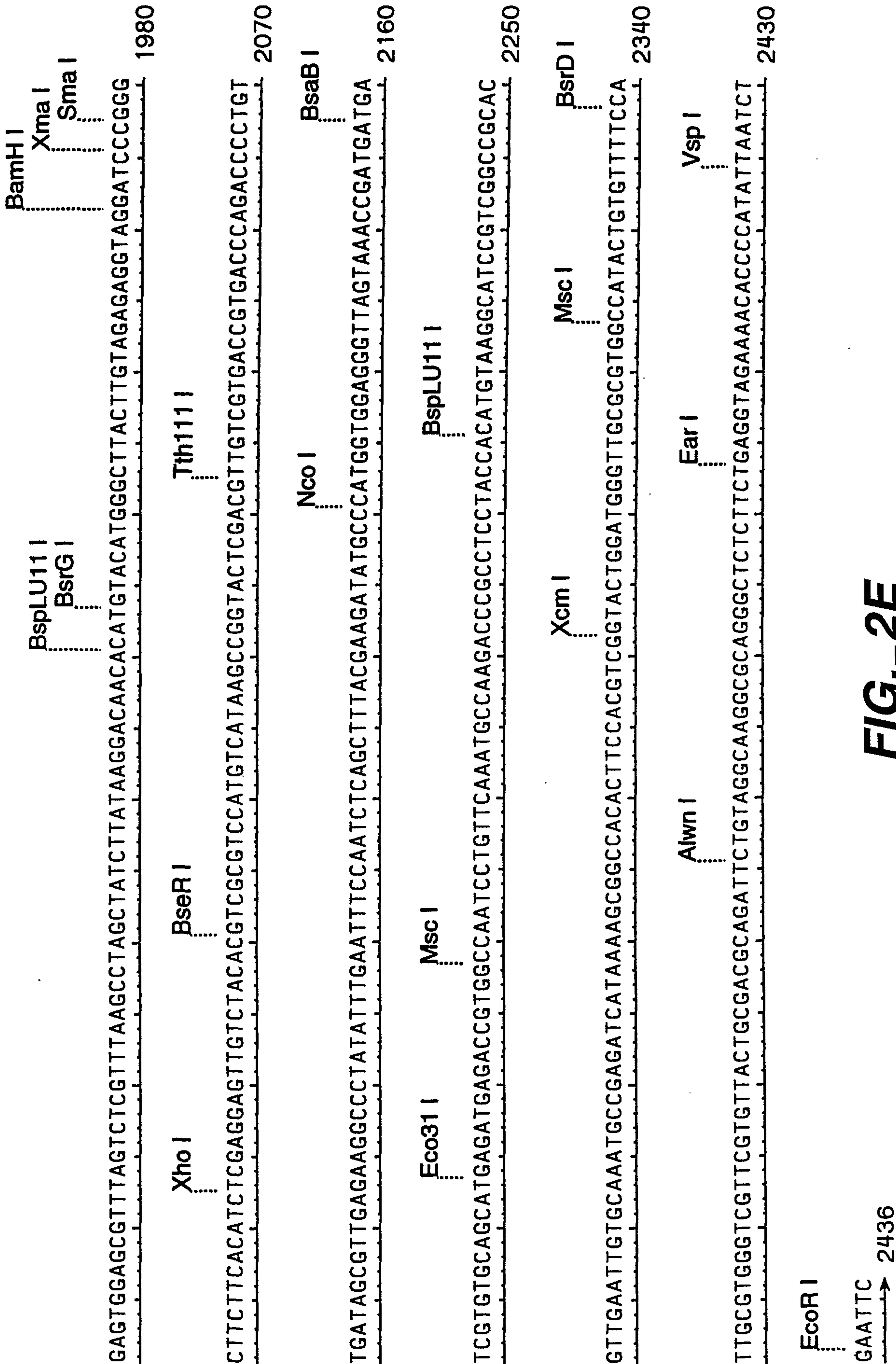


FIG.-2E

CCATGGTGGTGCATATCGGCAGTAGTCTTTGCCGAAACGTTGAGGGTTACAGTGTATCTGCGTCGGACATACTT
CGGGGAATCTACGGCGGAATATCAAAGTCTTCGGAATATCCATATTGGGAAAGGACAGAAAGCTCCGGGGTAGTTT
GATAGATGAGCTCCGGTGTTAATAATCGGGAGCTGACAGGAGTGAAGCTCATGTAGAACCATCTAGTAATGTCAGT
CGCGCGCAATTTCGCACATGAACAAGTTGATTTCCGGACCCCATTTGTTACATCTCTCGGCTACAGCTCGAGATG
TGCCCTGCCGAGTACTTAGAAGCCATGCCAGCTTCAAAATTAGTCTTCAGTGGTTTATTAAAGAGATAGATCCCTTCACAAACACT
CATCCAACGGACTTCTCATACCACTCATTTGACATAATTCAACAGCTCCAGGCGCATTTAGTTCAACATGAAGC
AATTCCTCCGCAACACAGTCTCTCGAGTTGTGGTGACTGAGGCAAGCTTAGCAGCTCTACGCAAGGCATCT
CCGAAGACCTCTACAGCCGTTTAGTCCGAATGGCCAATACTCCCAAGCTGCTACGCCGACCTGTGCAACATTC
CGTCGACTATTATCAAGGAGAGAAATTACAAATTCCTCAAACTGACATTAAACGGATGGAATCCTCCGCGACGACA
GCAGCAAAAGAAATAATCACCCGCTTCCGTGGCACTGGTAGTATACGAATCTACAACCTCGATACTAATAACACC
TCACGCCCTTCGACACCCCTACCAATAATGCAACGGTTGTGAAGTACACGGTGGATAATTATTTGGATGGGTCTCCG
TCCAGGACCAAGTCGAGTCGCTTGTCAAAACAGCAGGTTAGCCAGTATCCGGACTATGCGCTGACTGTGACGGGCC
ACAGGTATGCCCTCGTGATTCTTCAATTAGTGTAATACTCACTCACTACGATAGTCTCGGAGCGTCCC
TGGCAGCACTCACTGCCGCCAGCTGTCTCGGACATACGACAACATCCGCCCTGTACACCTTCGGCGAACCGCGCA
GCGGCAATCAGGCCCTTCGCGTCGTACATGAACGATGCCCTCGAGCCACAGATACGACGAGTATTTC
GGTCACTCATGCCAACGACGCGCATCCCAAACCTGCCCTGGAGCAGGGGTACGCCCATGGCGGTGTAGAGT
ACTGGAGCGTTGATCCCTTACAGCGGCCCAACATTTGTCTGCACTGGGGATGAAGTGCAAGTGTGTGAGGCC
AGGCGGACAGGGTGTGAATAATGCGCACACGACTTATTTGGGATGACGAGCGGCA TGCAATGGTGATCAG
TCATTTTCAGCCTCCCGAGTGTAACAGGAAGATGGATGTCCTGAGAGGCA TGCAATACCCGAAGC
ACACTTTTTCGGTAATCAGGACATGTAAAGTTCCATGAATAGATA TGTTACCTCACCATAGCCTT
GAGGTTGCCCTTCTCTTTGATTGTGAATATATATTAAAGTAGACAGATATCTCTAAACACCTTATCCGCT
TAAACCCATCATAGATTGTGTACGTACGTAGACCCCTTGAAATGATGAGCGAAATGTATCAGTCCCGTTTAAATCA
AACCCTTTCAGCCTAGCACAGTCAGAAATACACCAACCCATTCTAAGGTAGTACTAAATATGAATACAGCCTAAA
TGCAATCGCTATATGATCCCATAAAGAAAGCAACCTTTCAGATCTCGTTTTCGCTGCGAAGAGCTAGCTCTAC
CATGGTCTCAATTATGAGTGGAGCGTTTAGTCTCGTTTAAAGCCTAGCTATCTTATAAGGACAAACATGTACATG
GGCTTACTTGTAGAGAGGTAGGATCCGGGCTTCTTCACATCTCGAGAGTTGTCTACACGTCCGTCATGTCA
TAAGCCGGTACTCGACGTTGTGTCGTGACCCAGACCCCTGTGATAGCGTTGAGAGGCCCTATATTGAA
TTTCCAATCTCAGCTTACGAAGATATGCCCATGGTGGAGGTTAGTAAACCGATGATCGTGTGAGCATGA
GATGAGACCGTGCCCAATCCTGTTCAAAATGCCAAGACCCGCTTACCAATGTAAGGCATCCGTCCGGCCGAC
GTTGAATTGTGCAAAATGCCGAGATCAATAAAGCGGCCACATCCACGTCCGTTACTGGATGGGTGGCC
ATACTGTGTTTTCATTTGCGTGGTCTGTTACTGCGACGAGATTCTGTAGGCAAGCGCAGGGCTCTCT
TCTGAGGTAGAAACACCCCATATTAAATCTGAATTC

FIG.-3

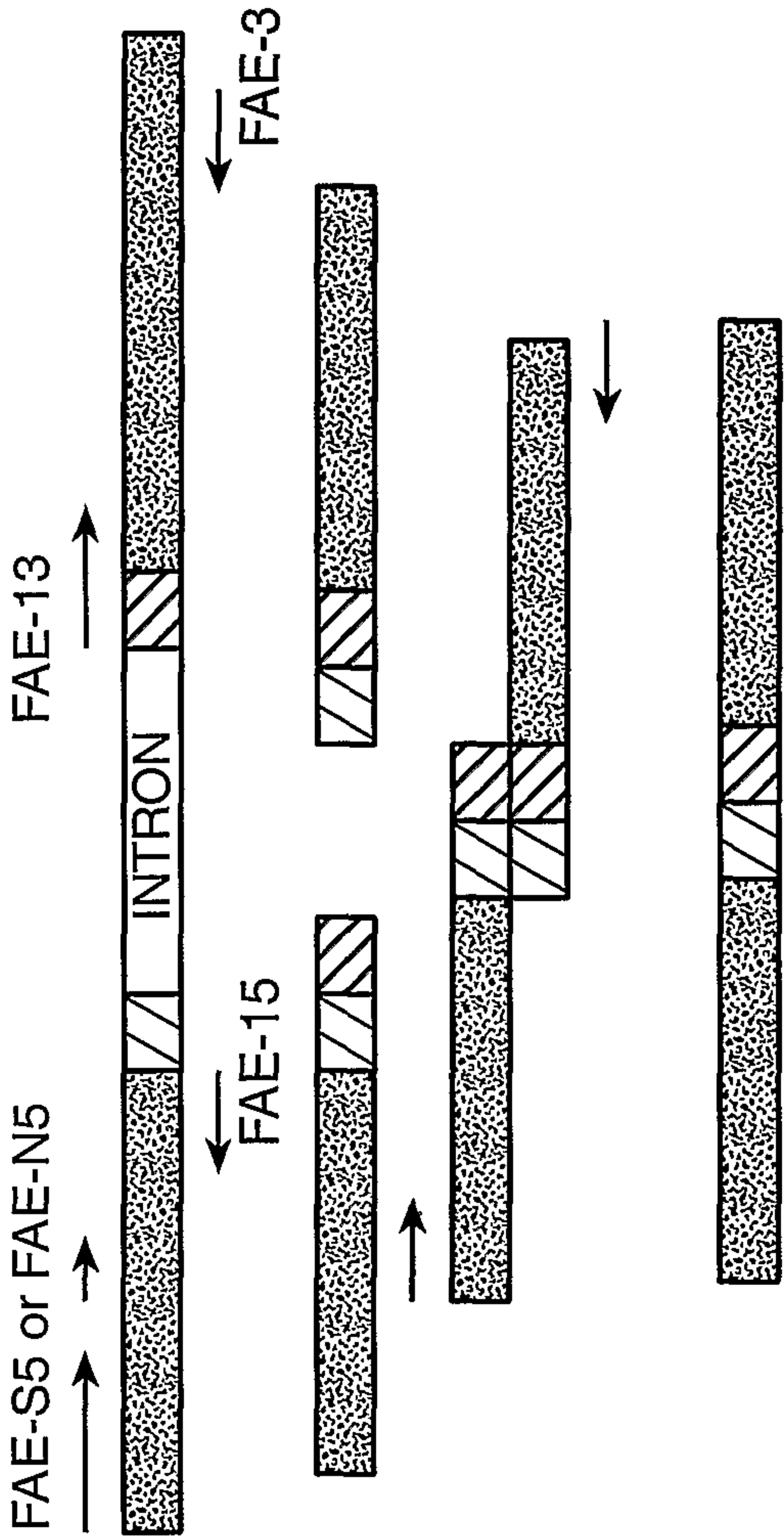


FIG. 4

FAE-I3 CCGGCCACGCCCTCGGGCCCTCCCTGGCGGCACTC 35-mer
FAE-I5 GCGGCCGAGGAGTGCGCCGCTCACGGTCAGCGCGTAGTCC 40-mer

intron position in original

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

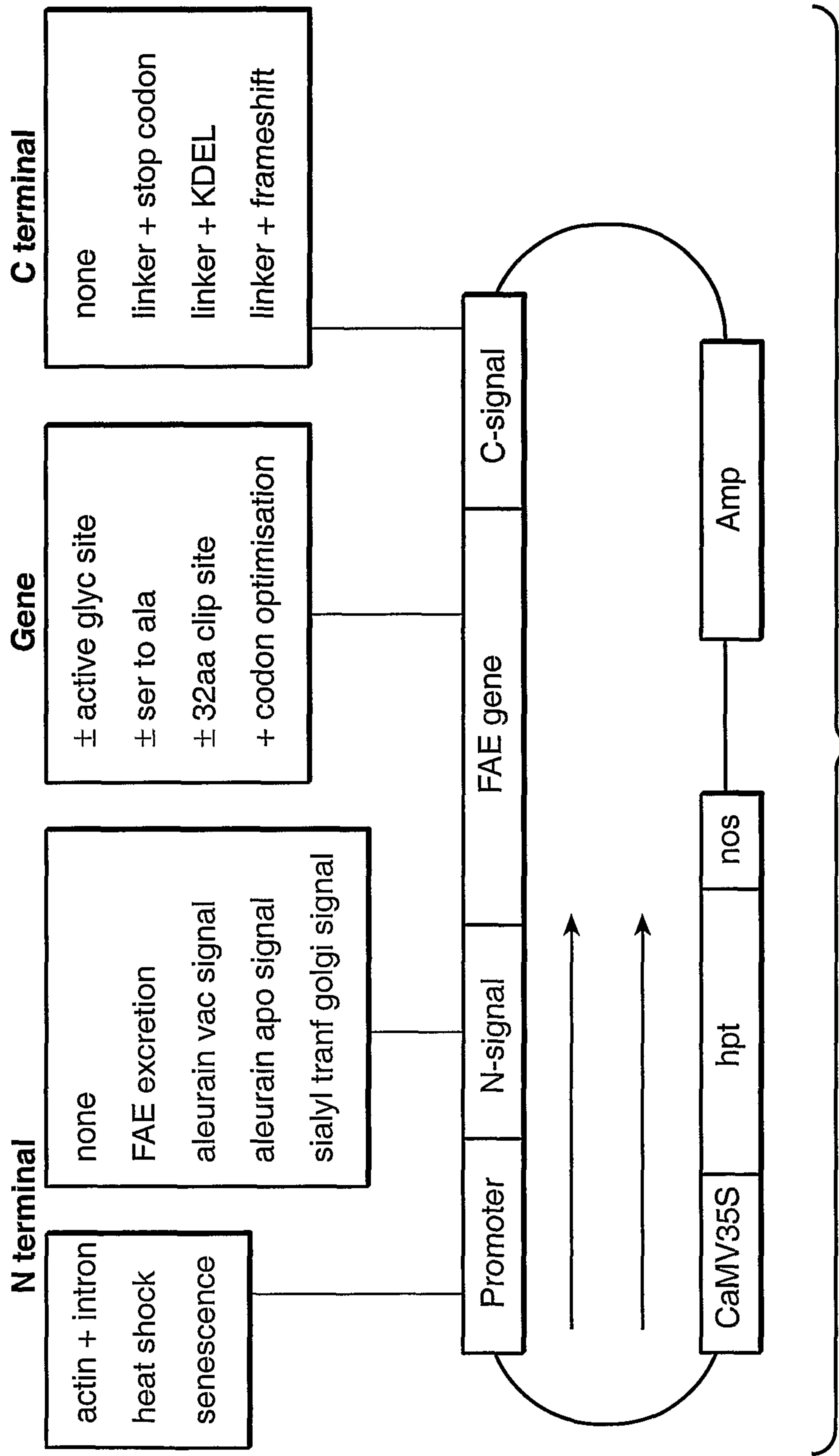
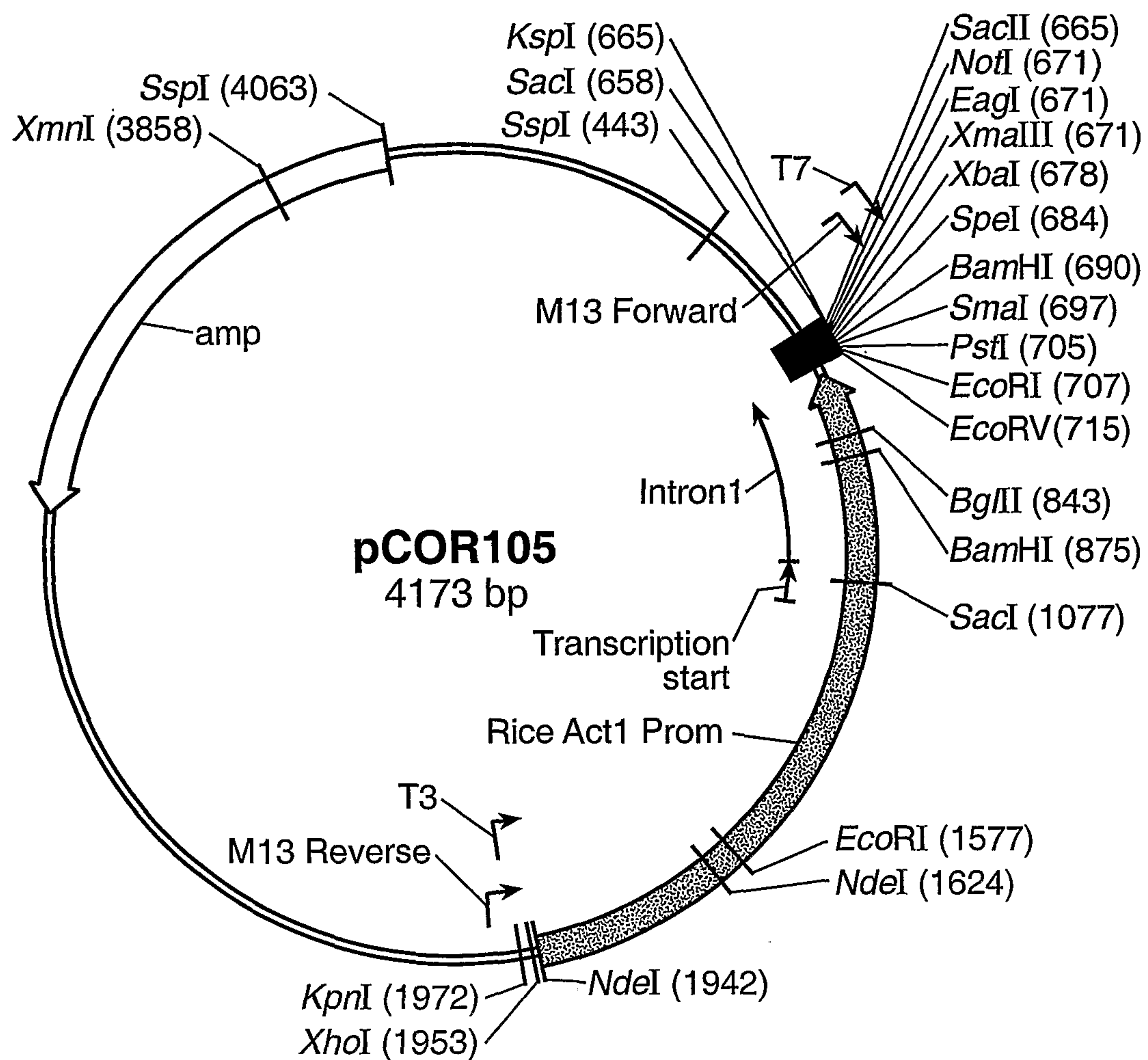
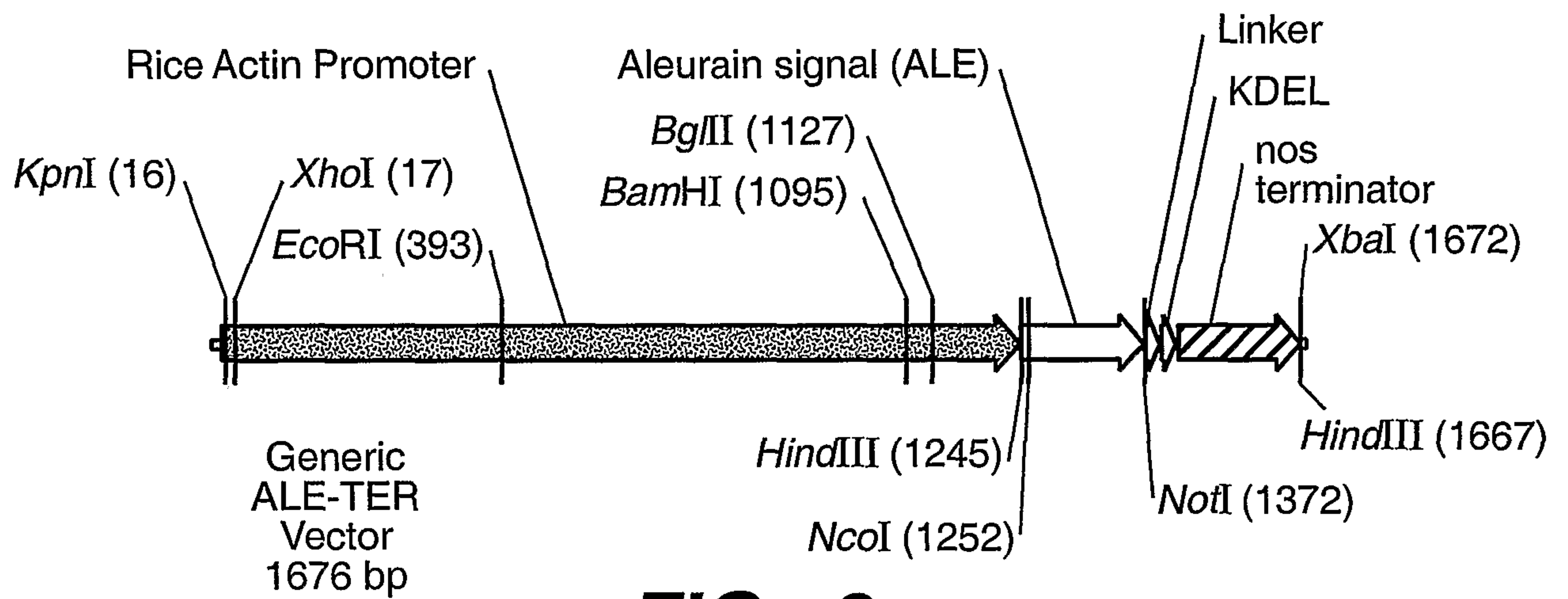


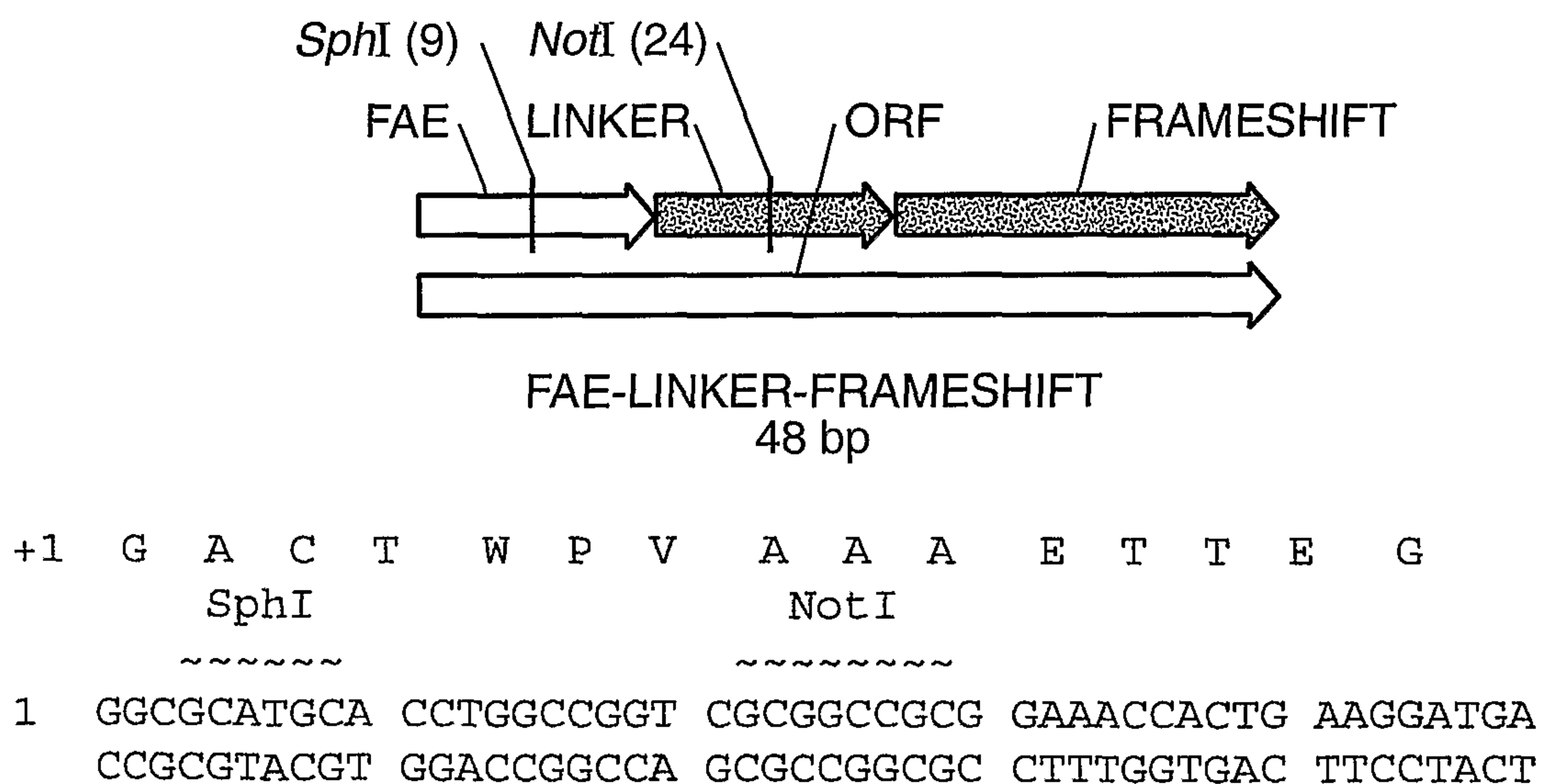
FIG. 6

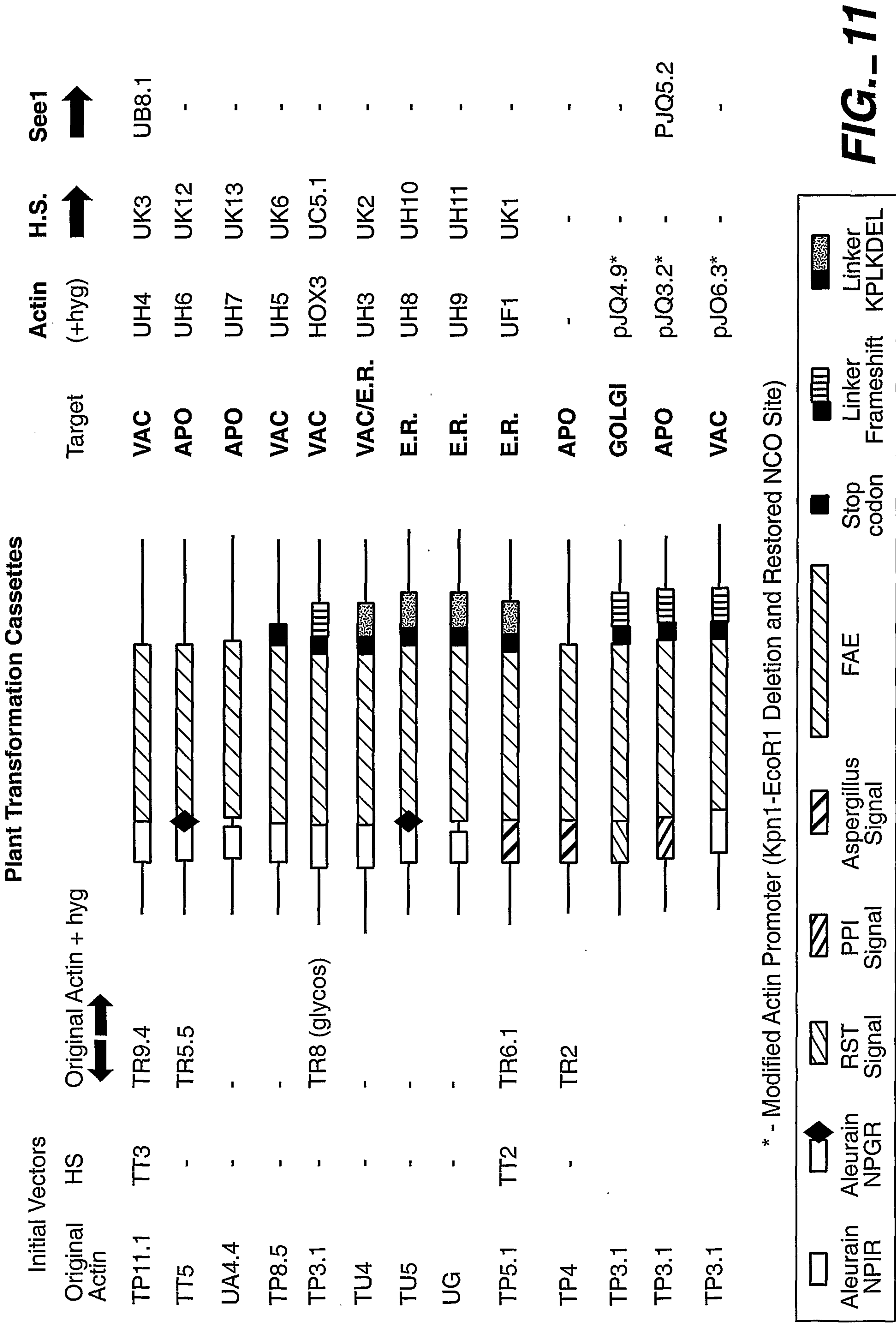
**FIG. 7**

**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**

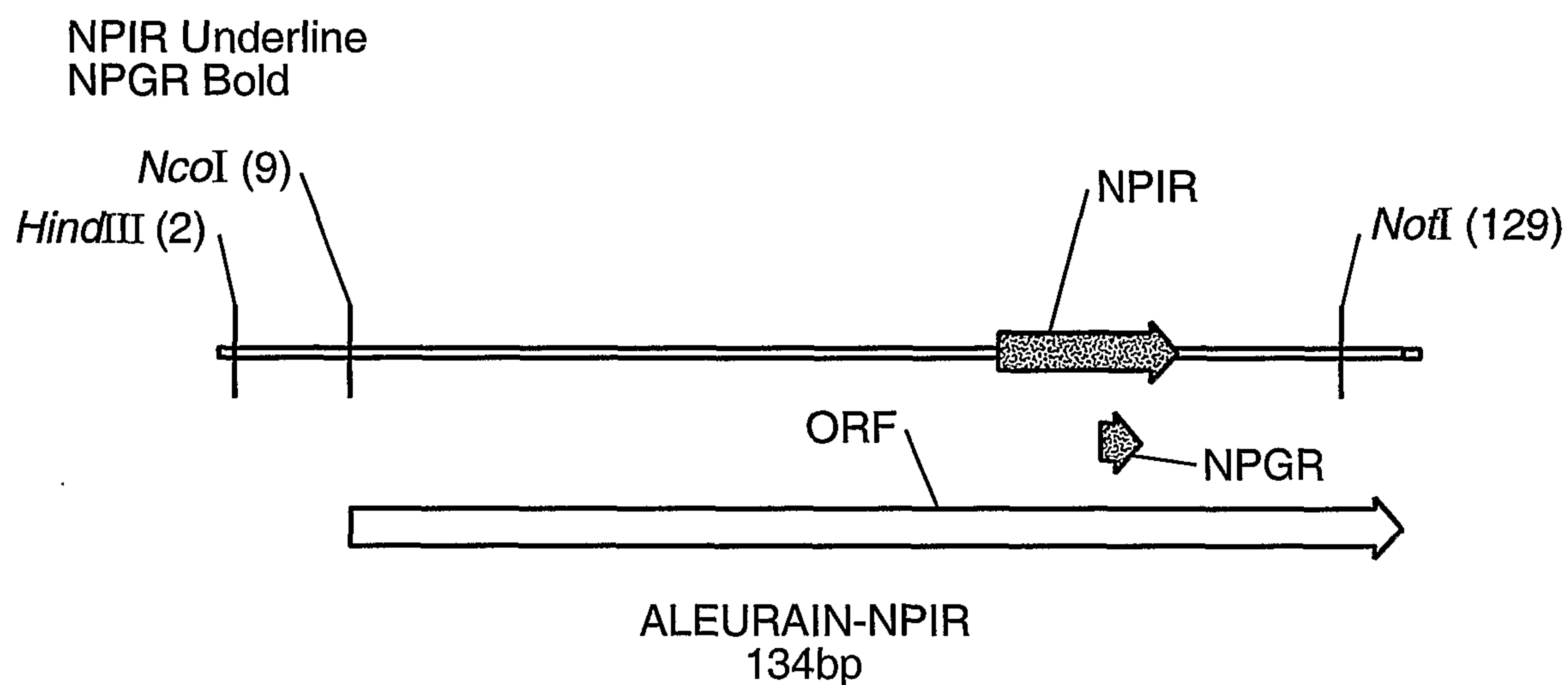
|       |        |                                       |                  |
|-------|--------|---------------------------------------|------------------|
| (i)   | APO    | - aleurain-NPGR-FAE                   | pUH12            |
|       |        | - aleurain-delNPIR-FAE                | pUH13            |
|       |        | - 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 GCGCGTCACC 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 GGCGGCCGC

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

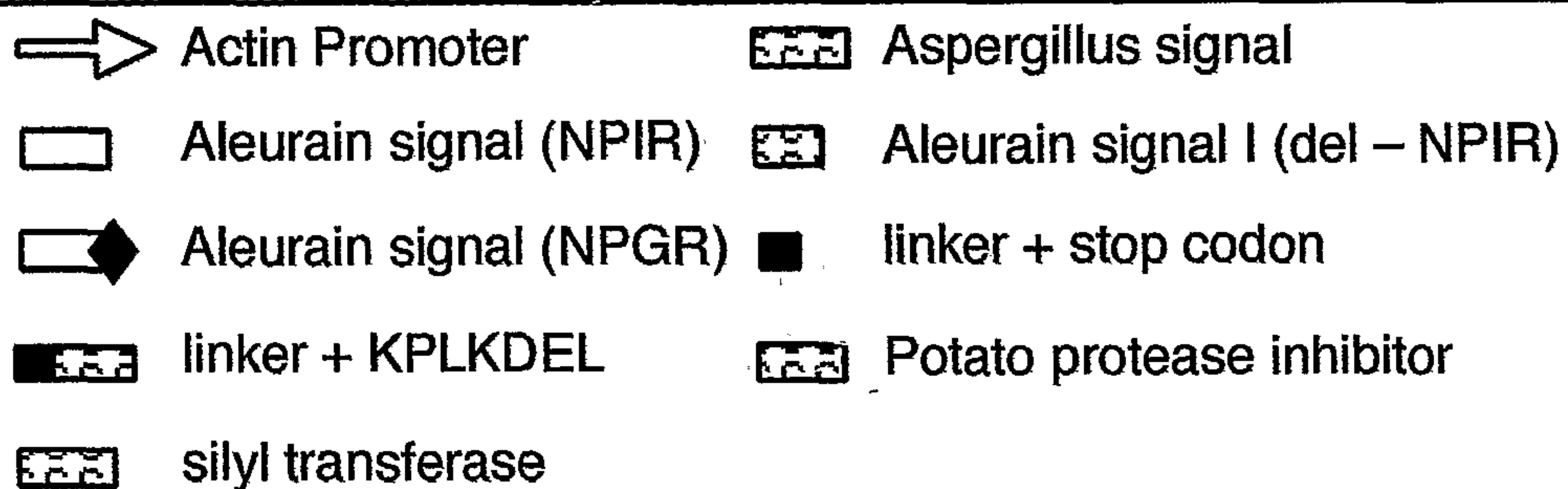
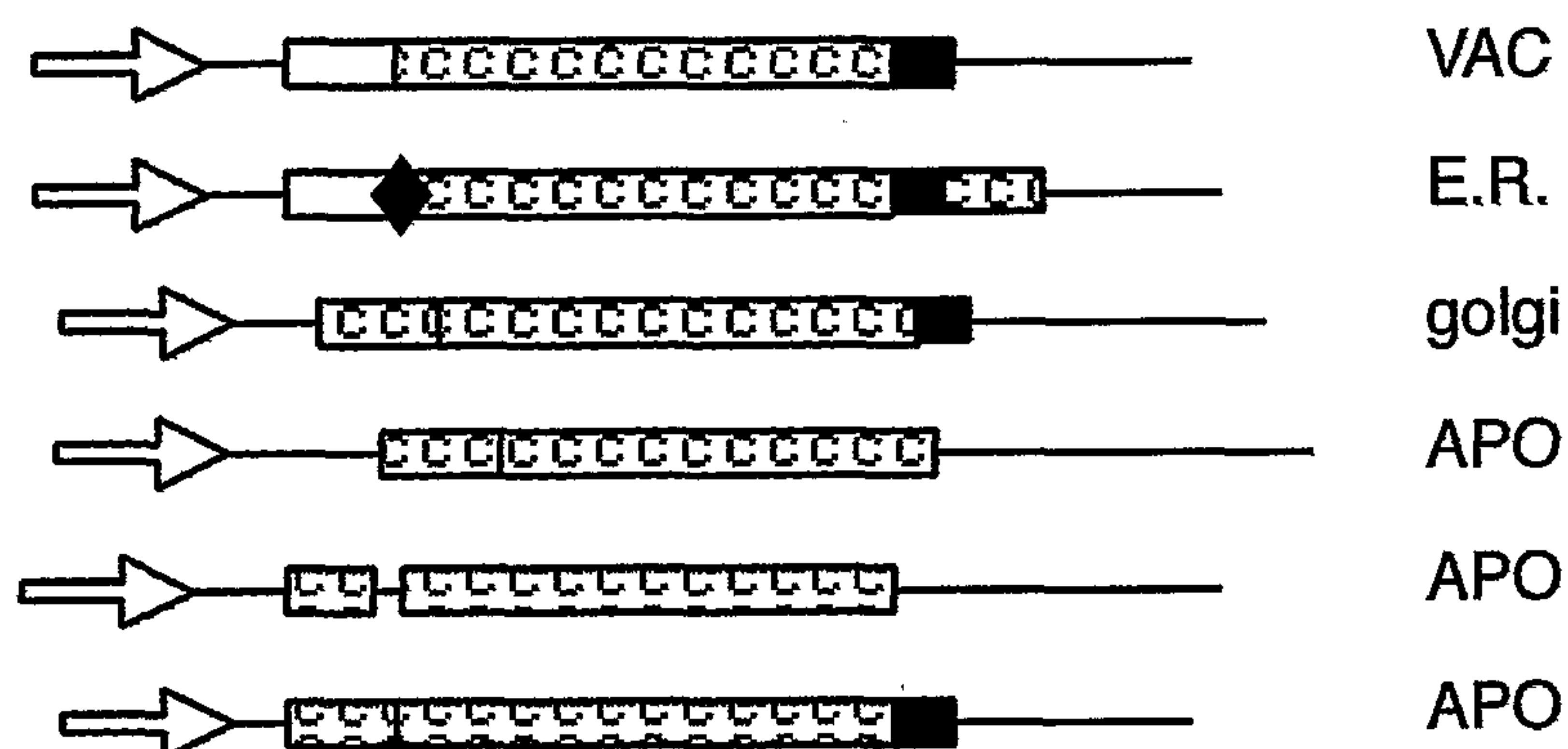


FIG._16A

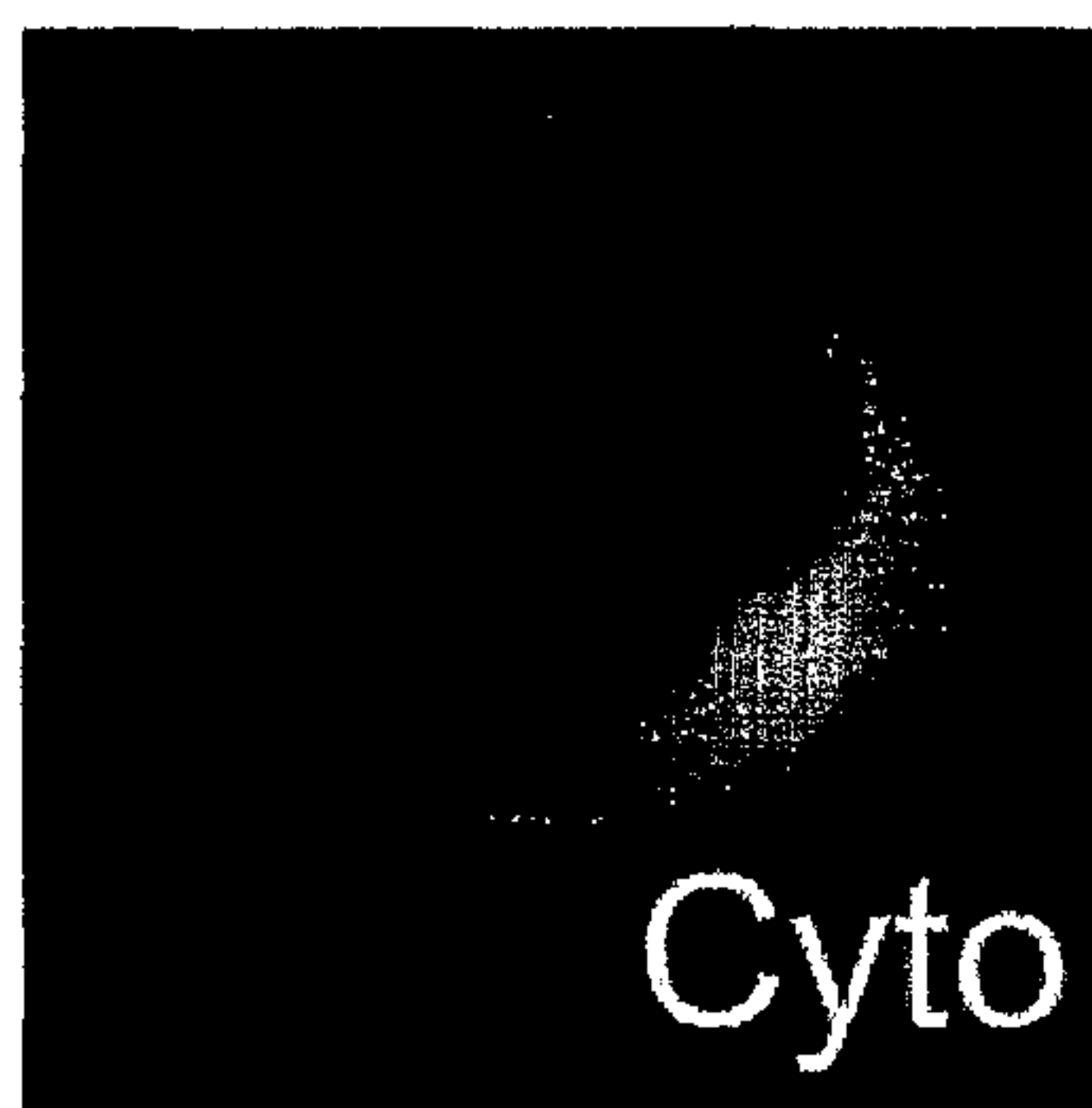


FIG._16B

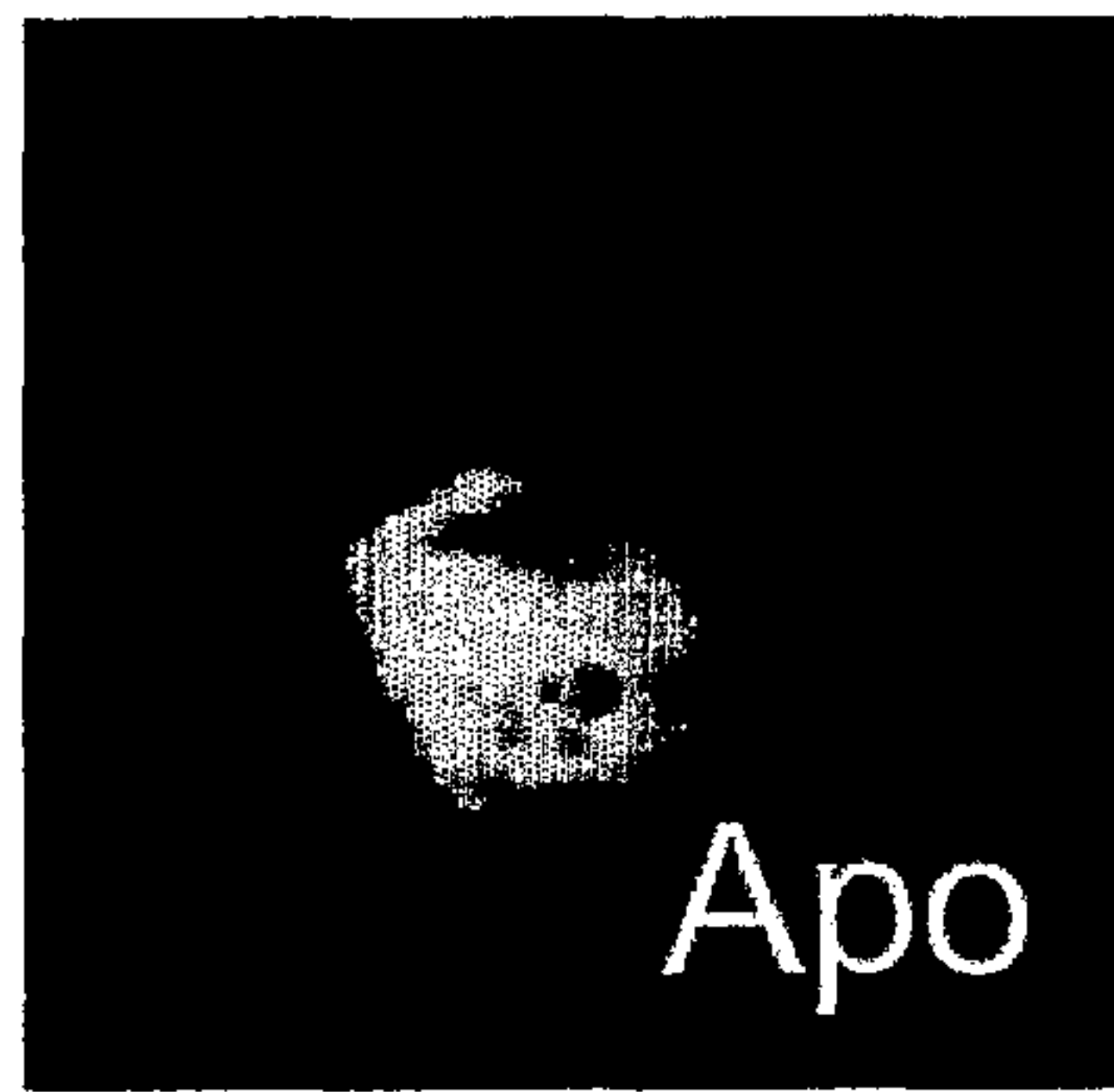


FIG._16C

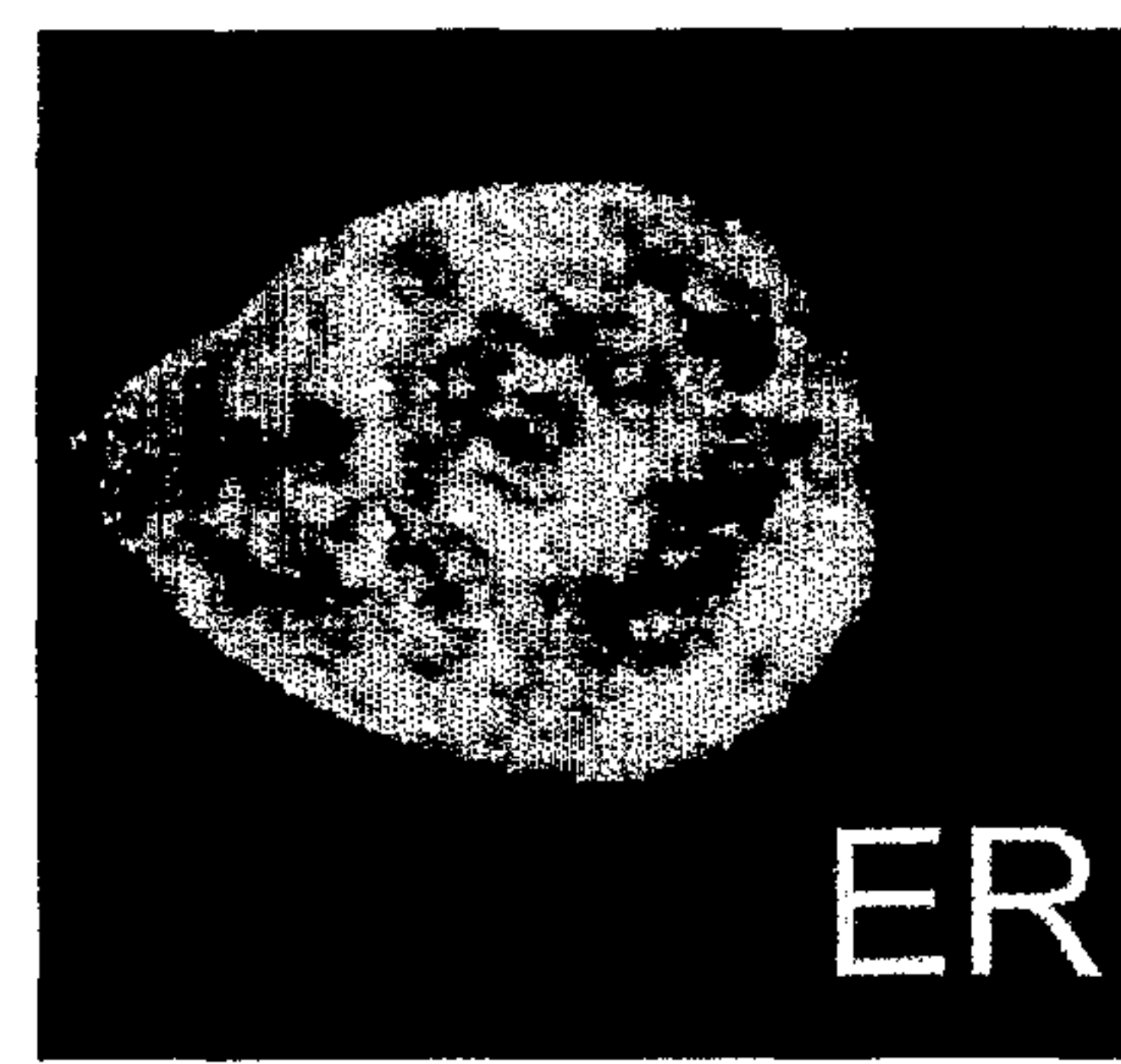


FIG._16D



FIG._16E

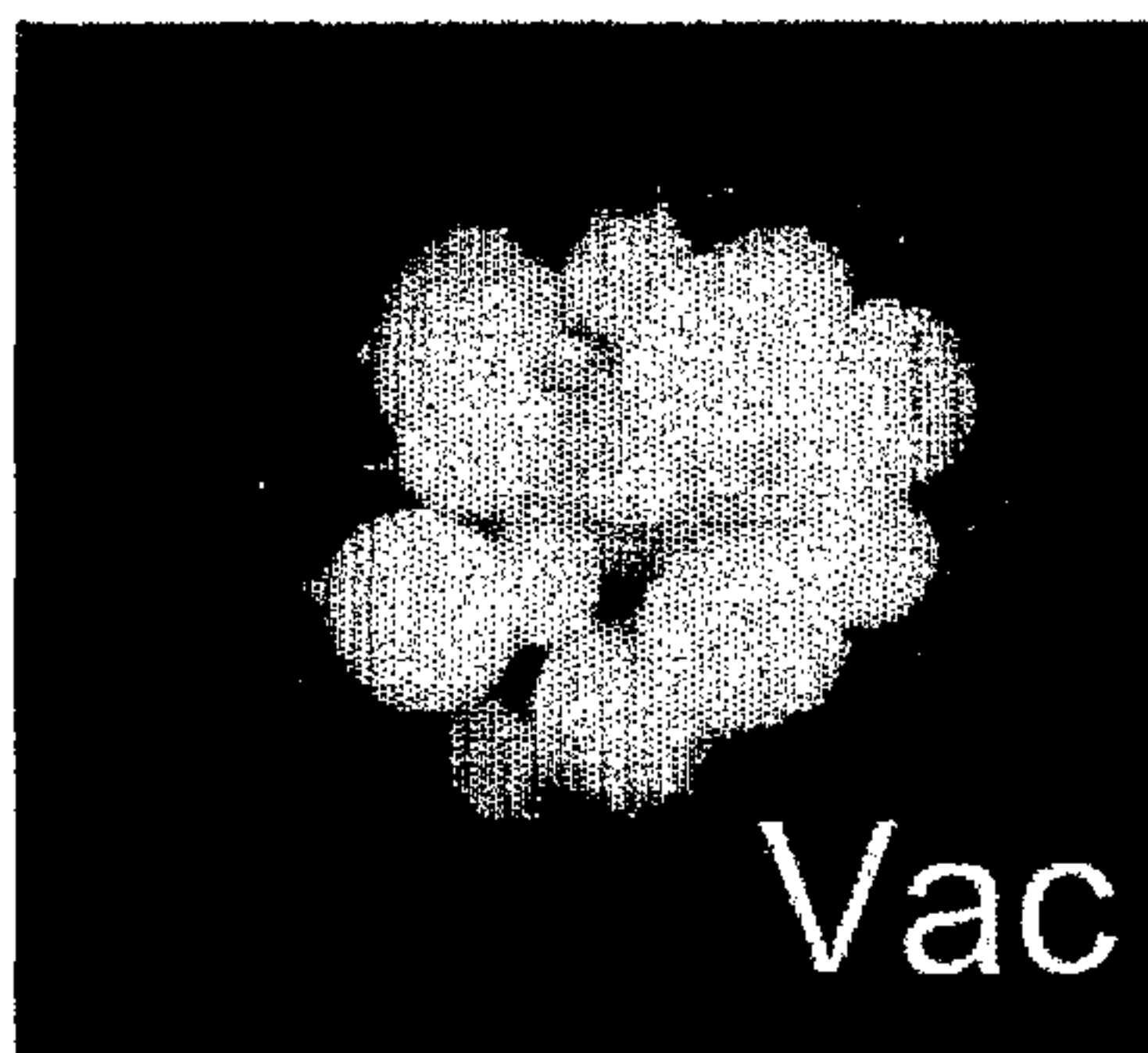


FIG._16F

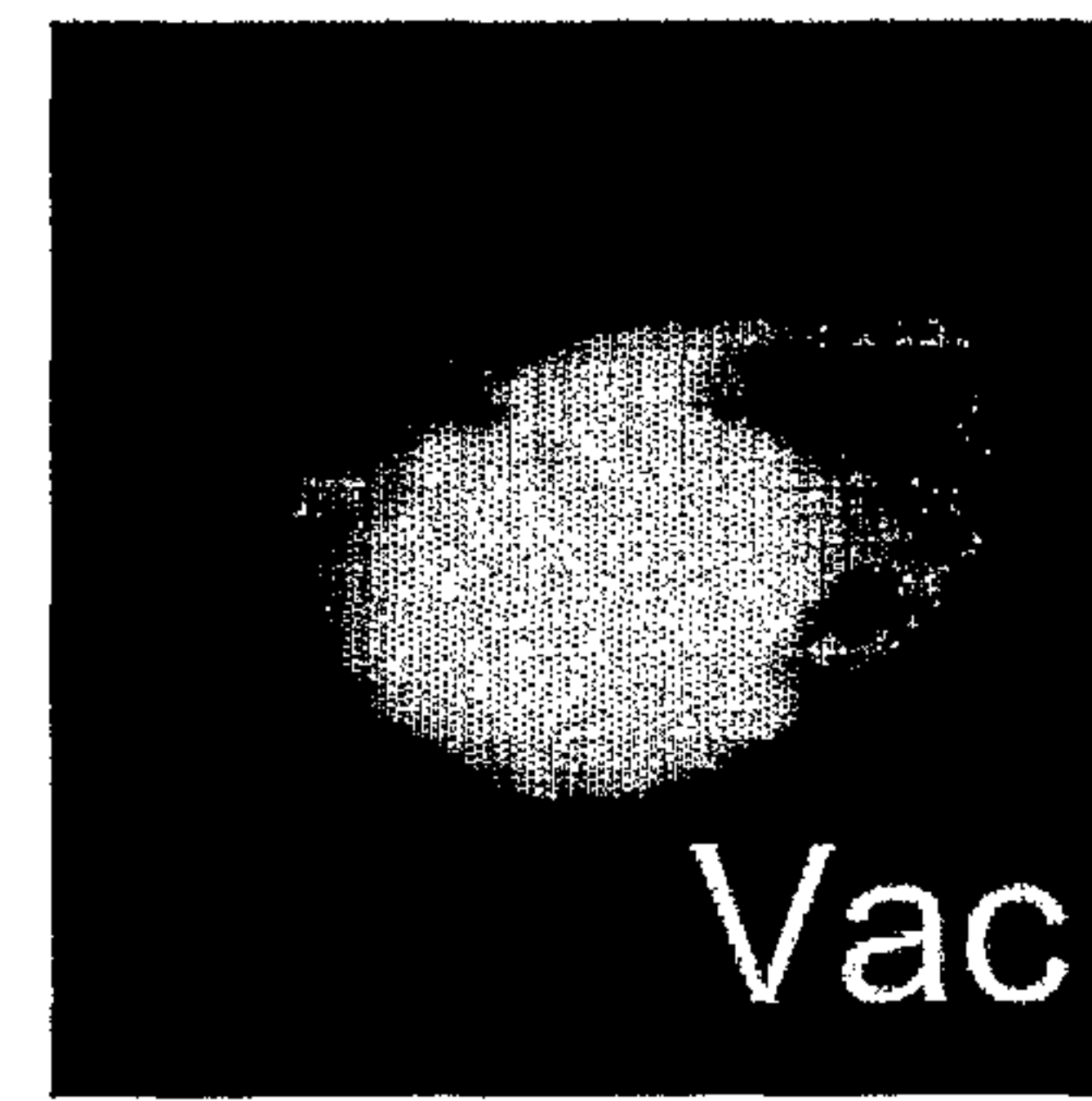


FIG._16G

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

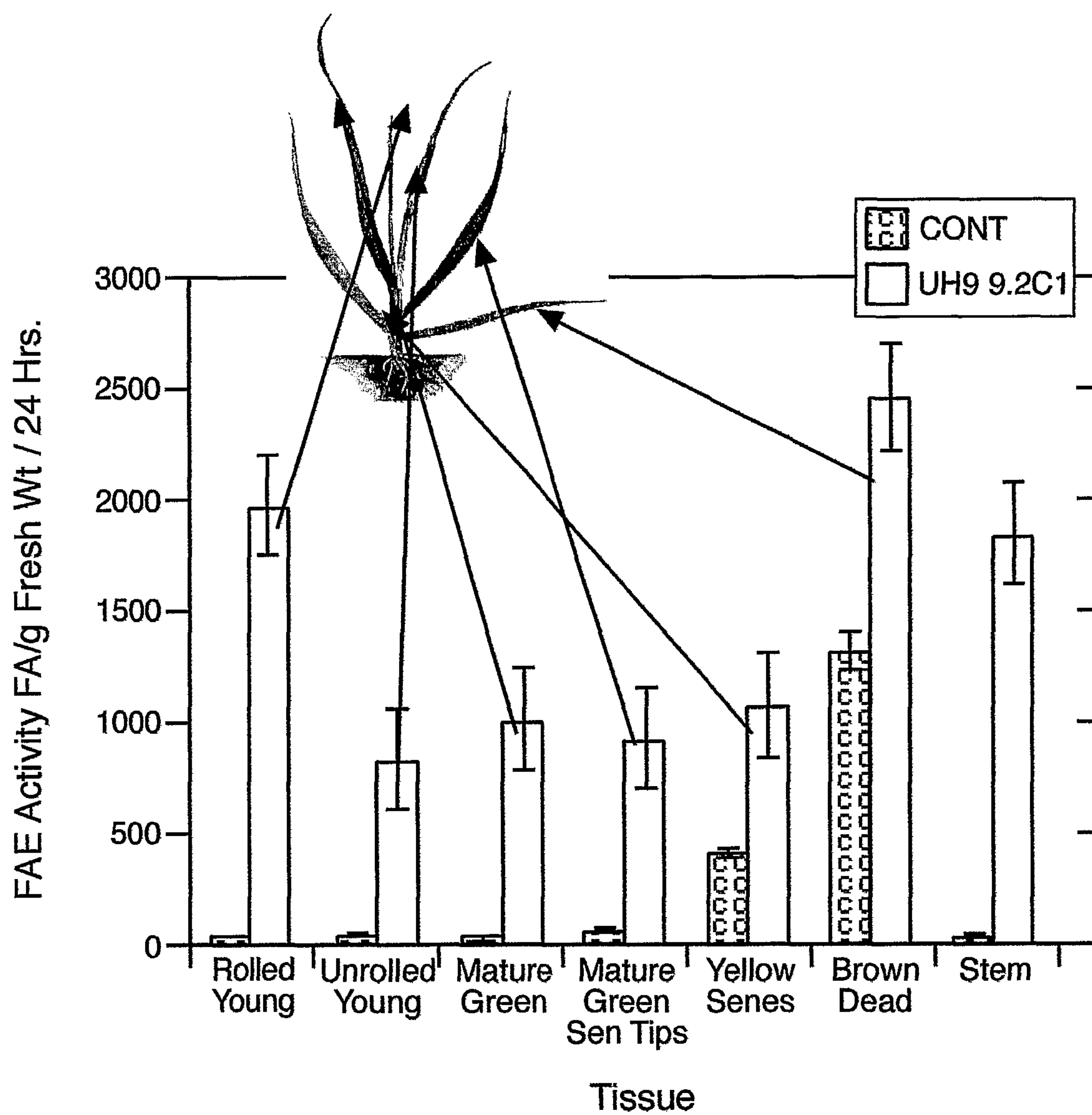


FIG. 17A

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

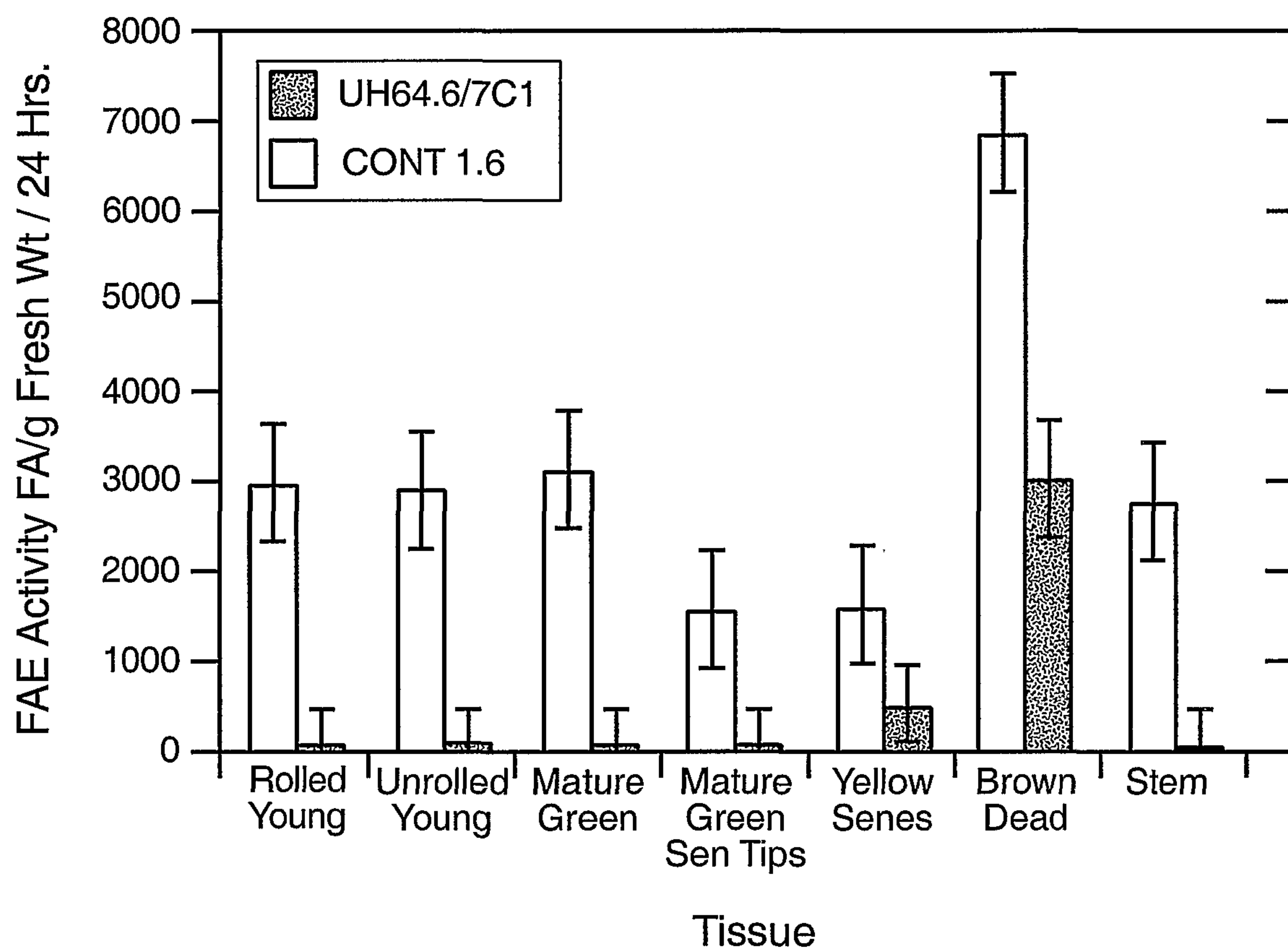
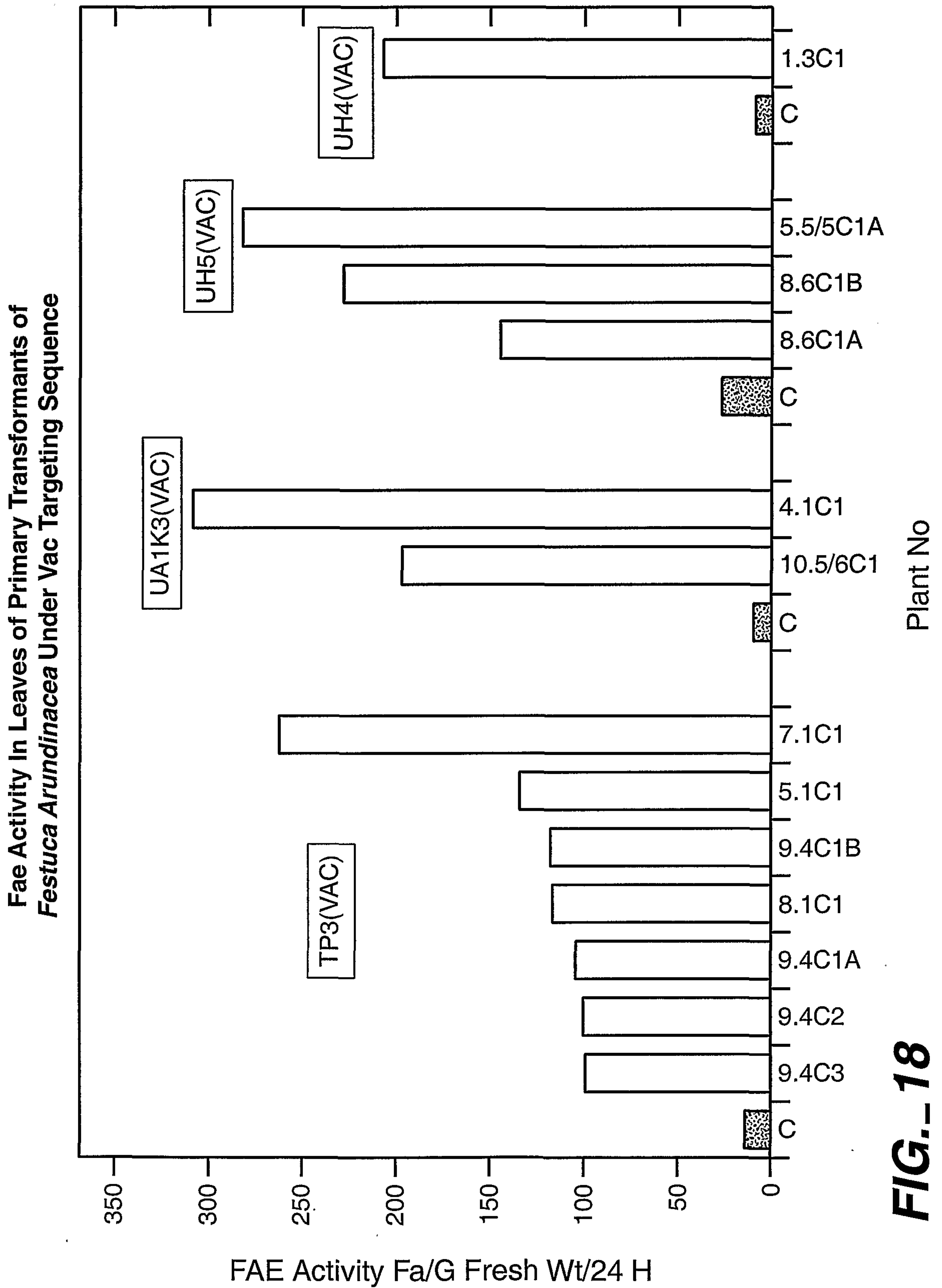
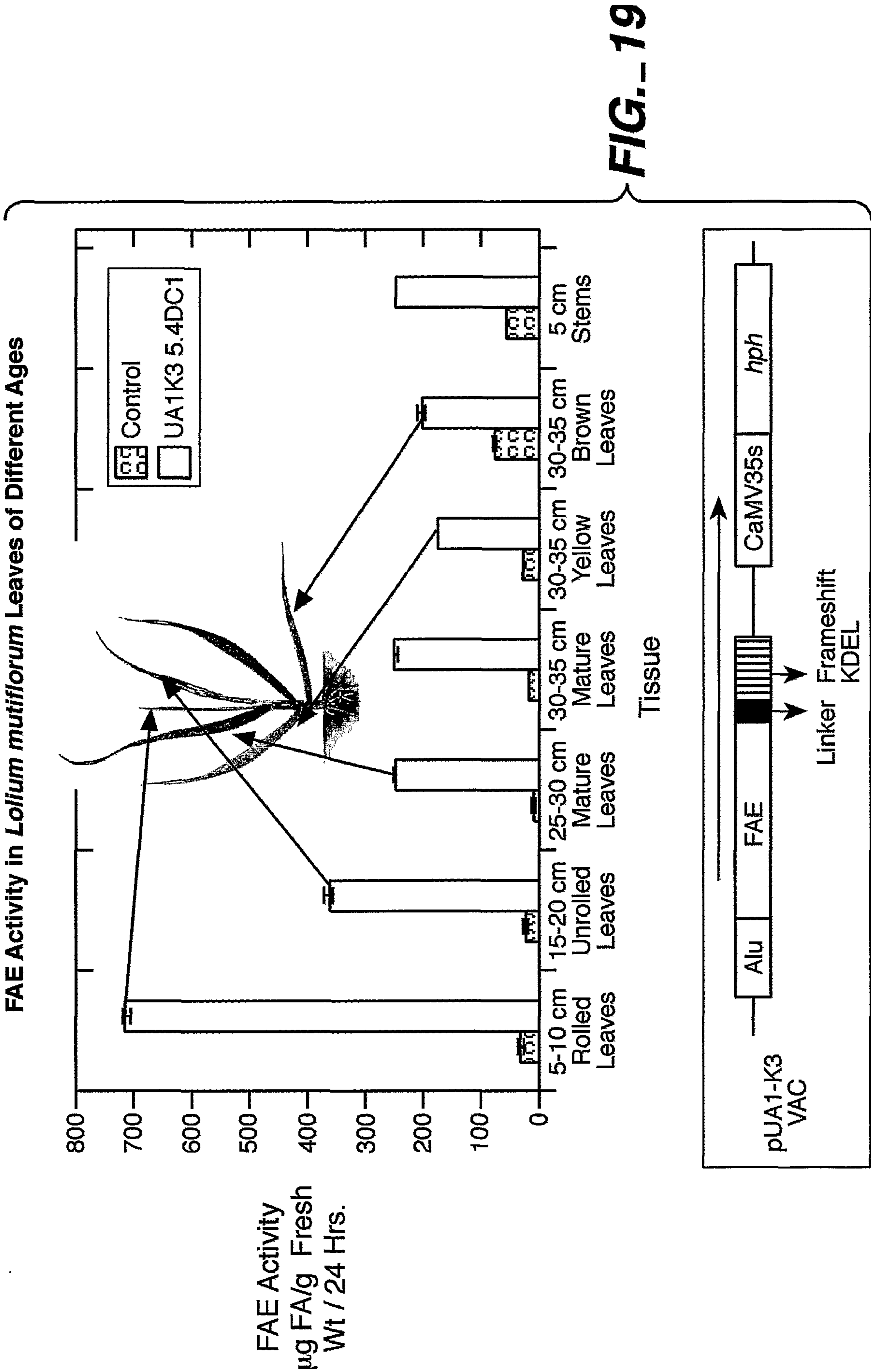
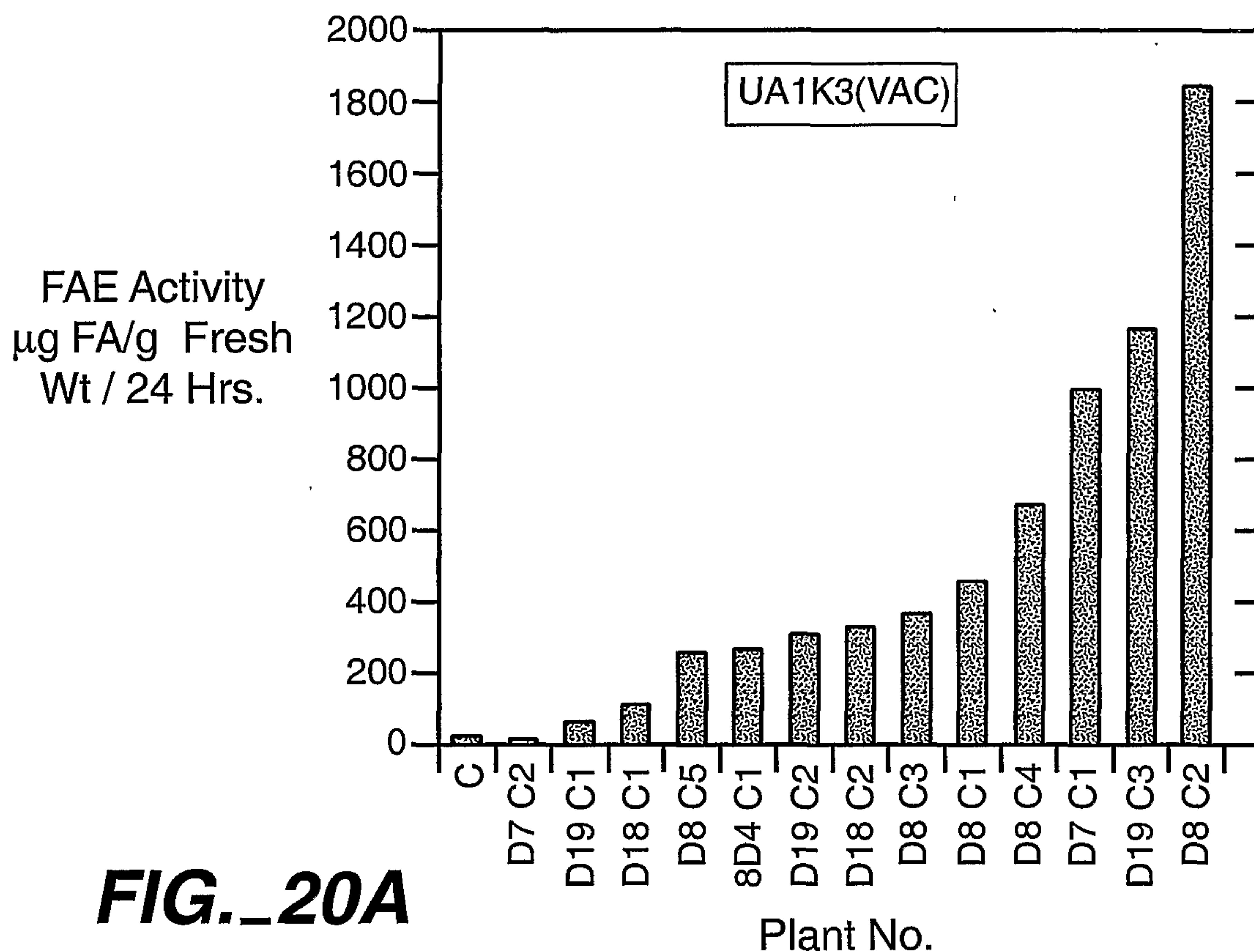


FIG. 17B

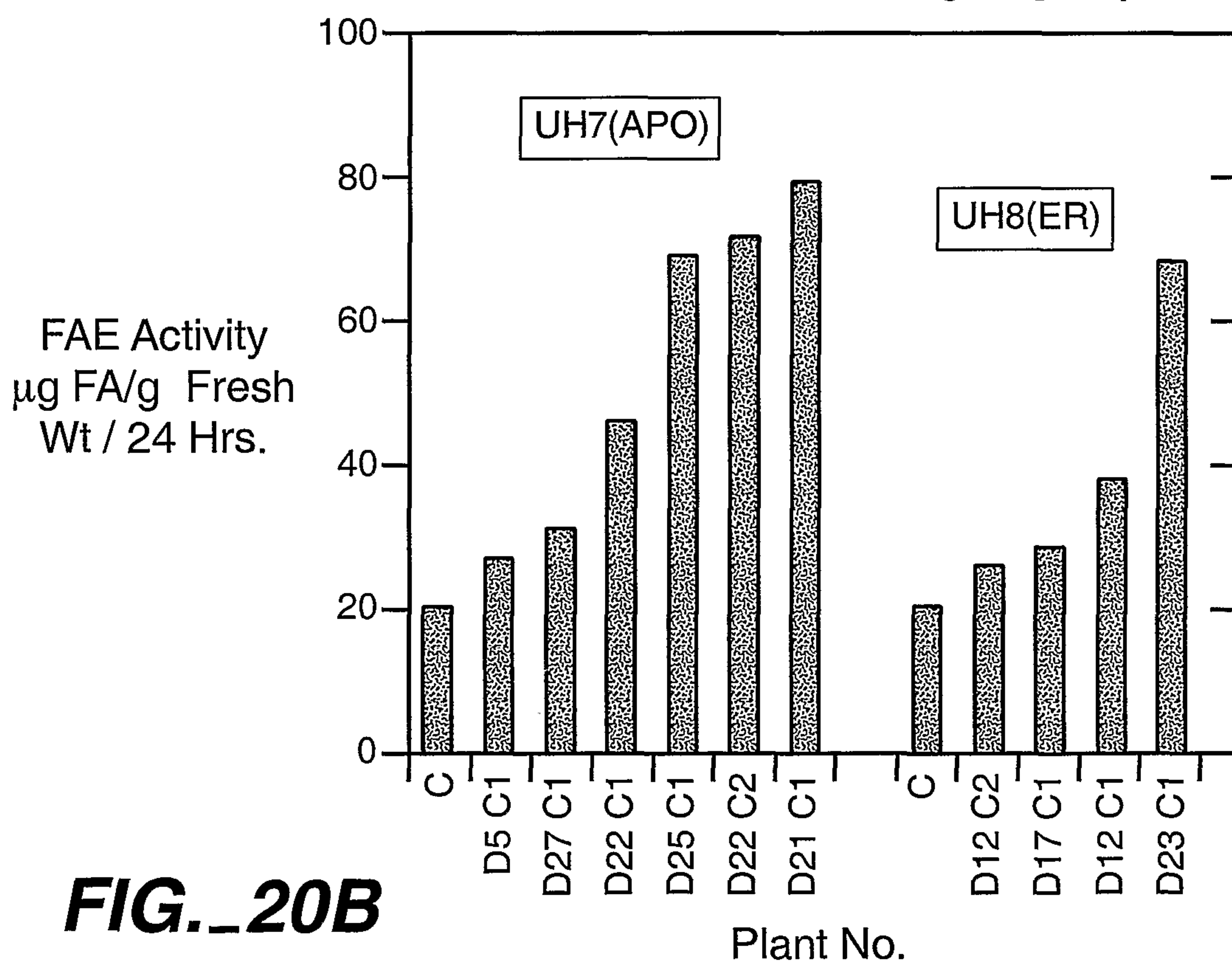
**FIG.-18**



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



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



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

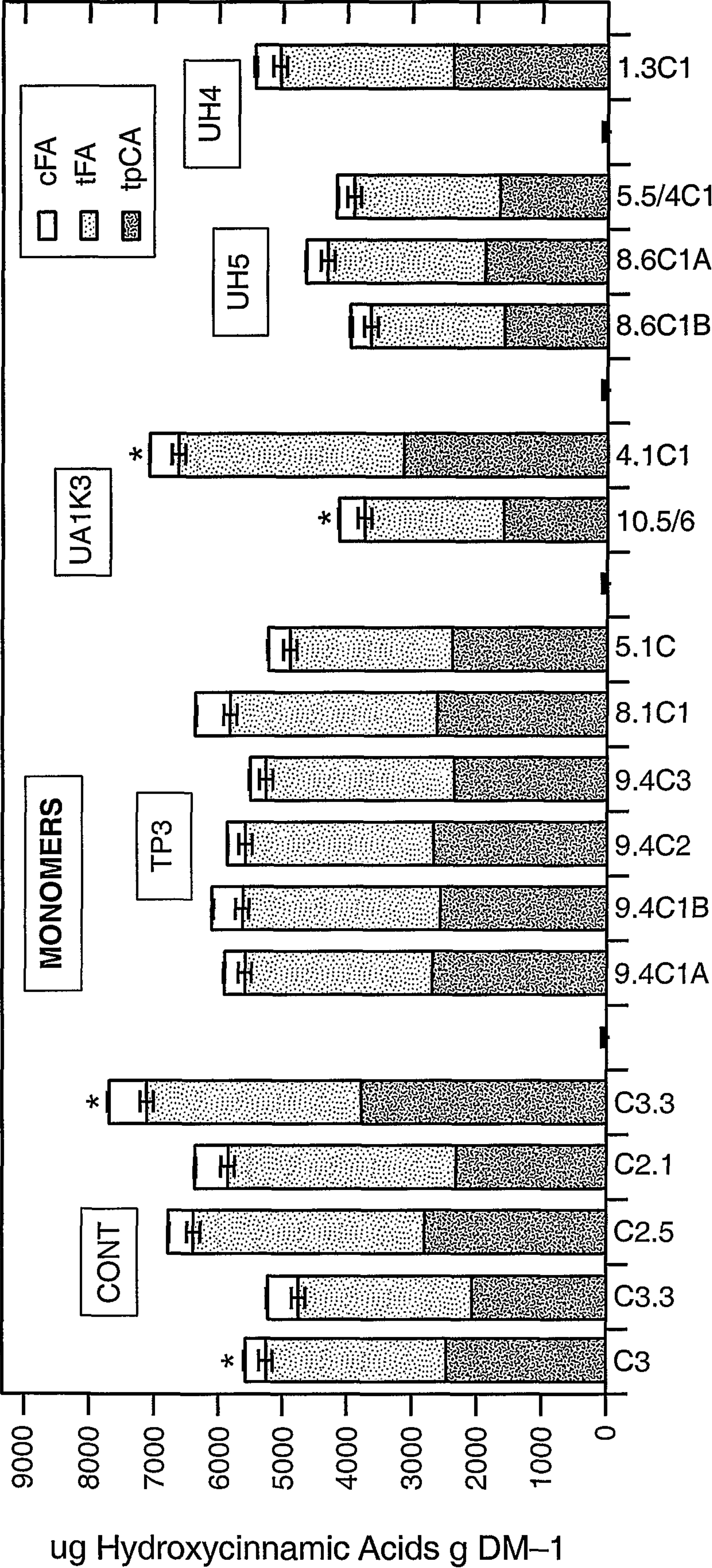


FIG. 21A

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

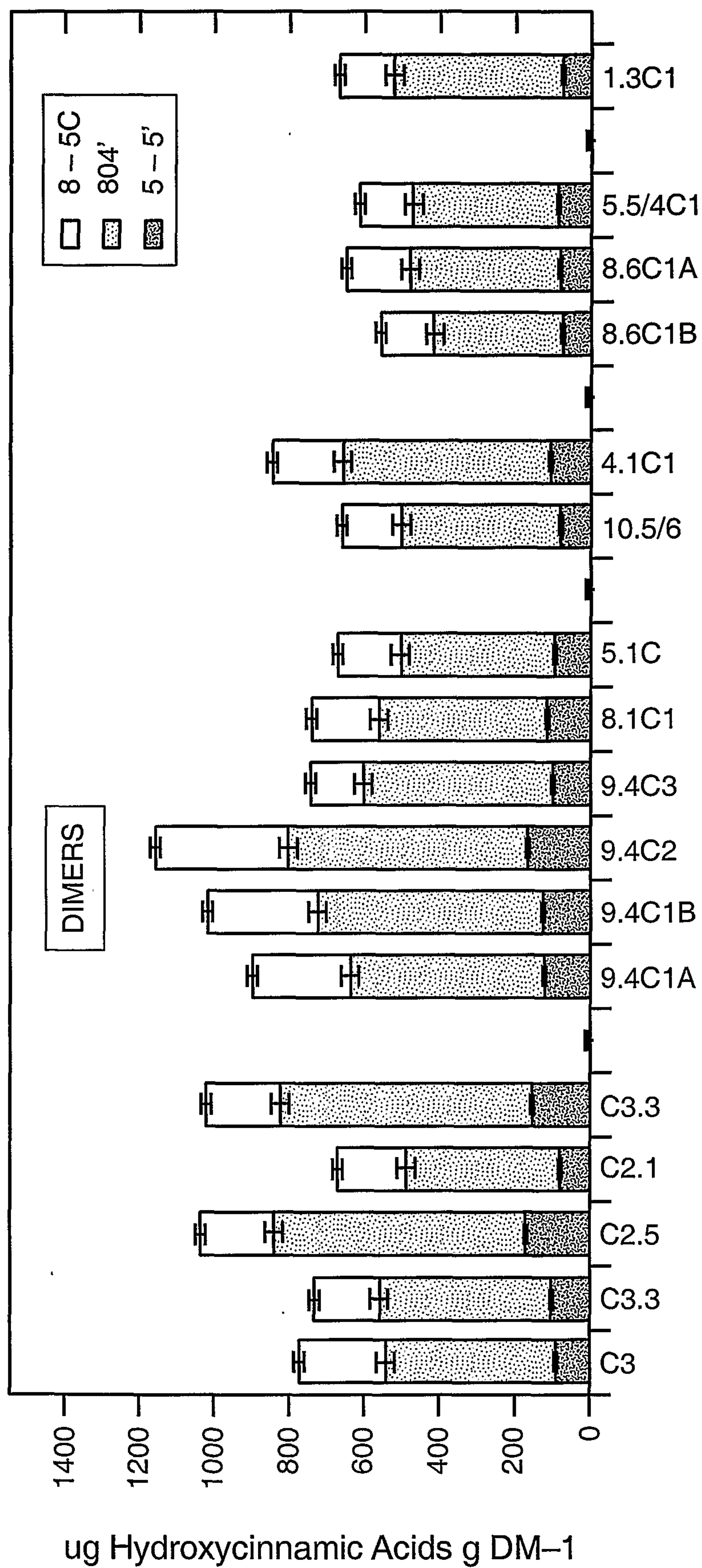


FIG. 21B

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

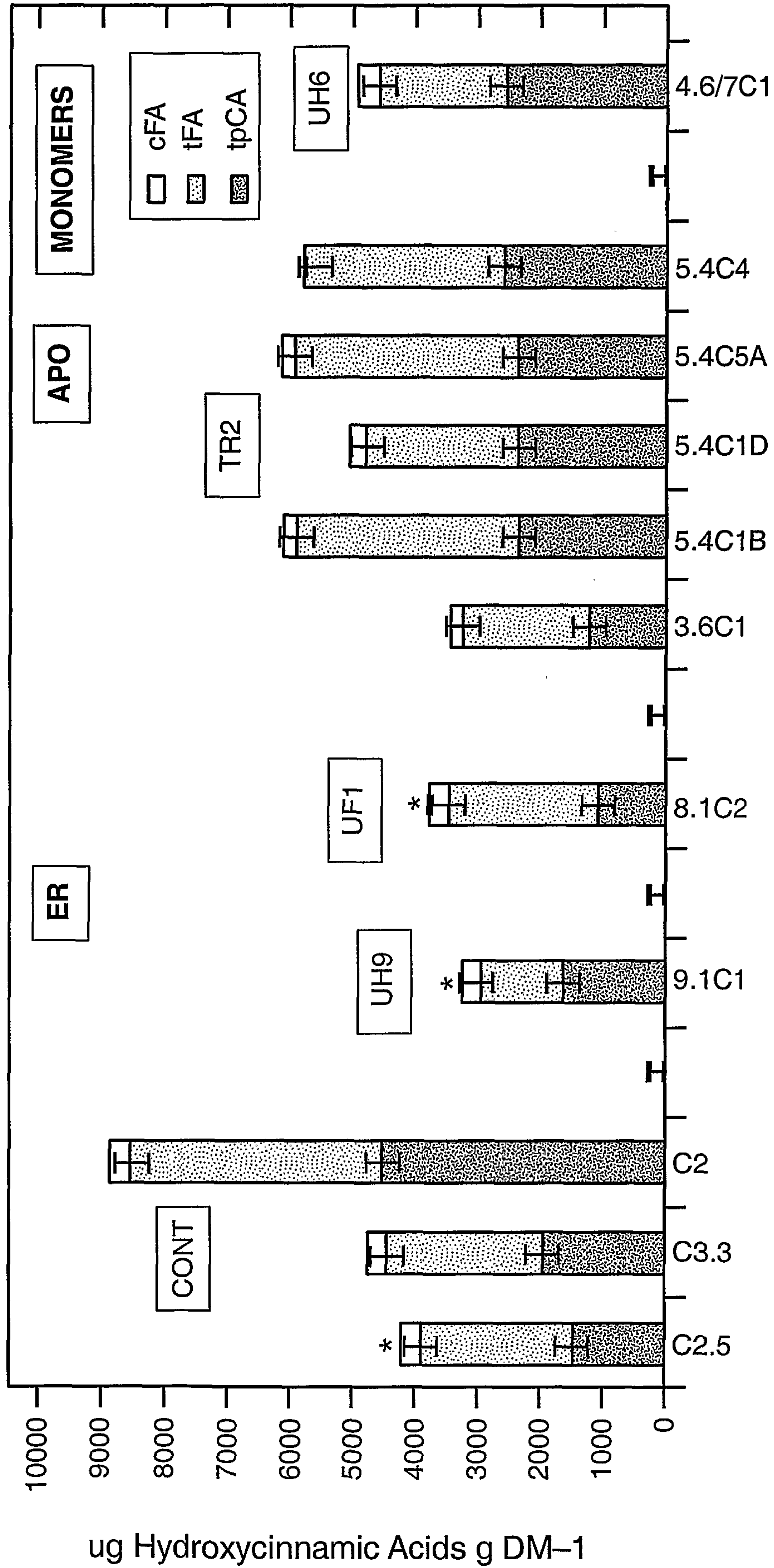


FIG. 22A

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

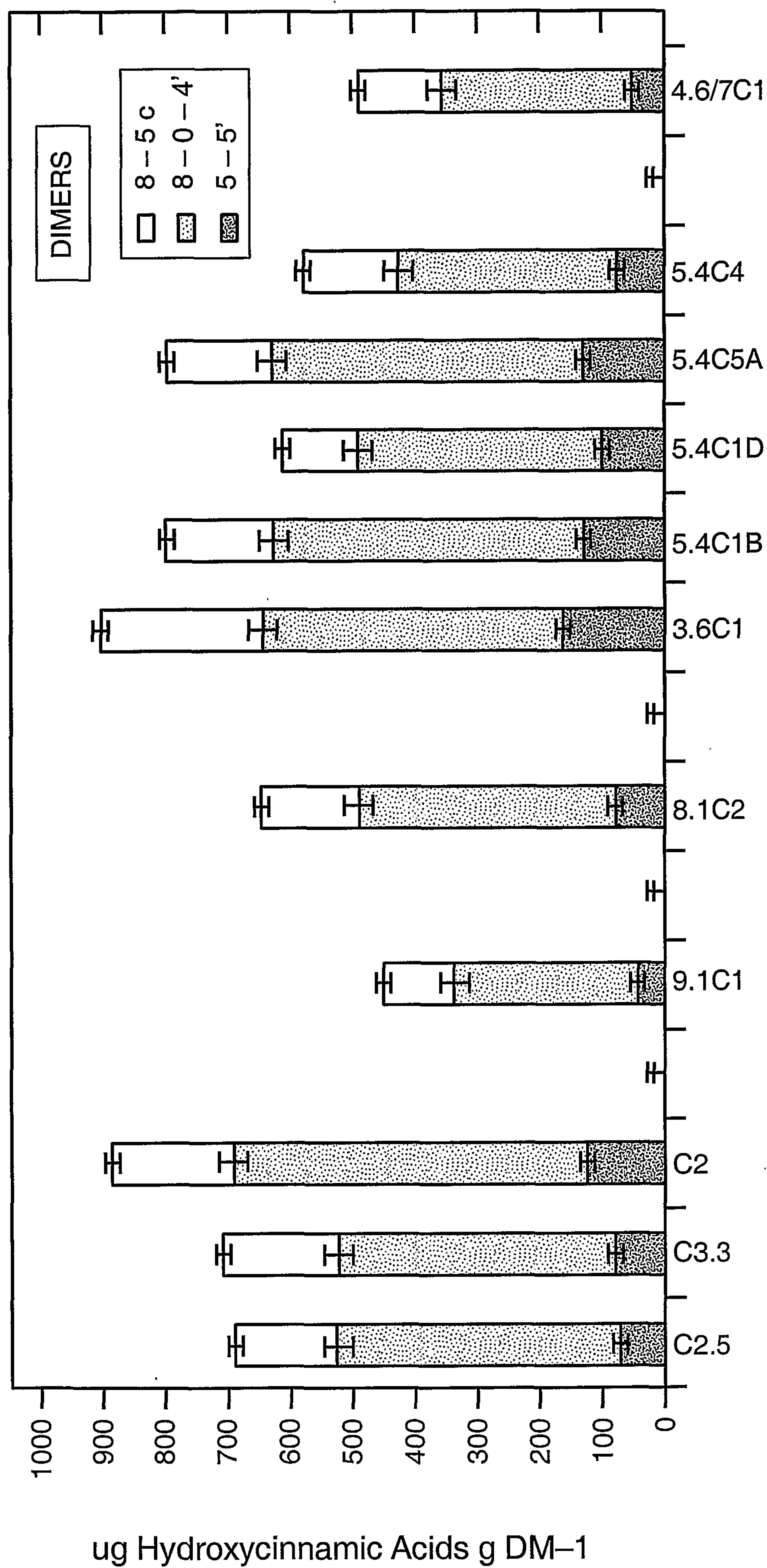
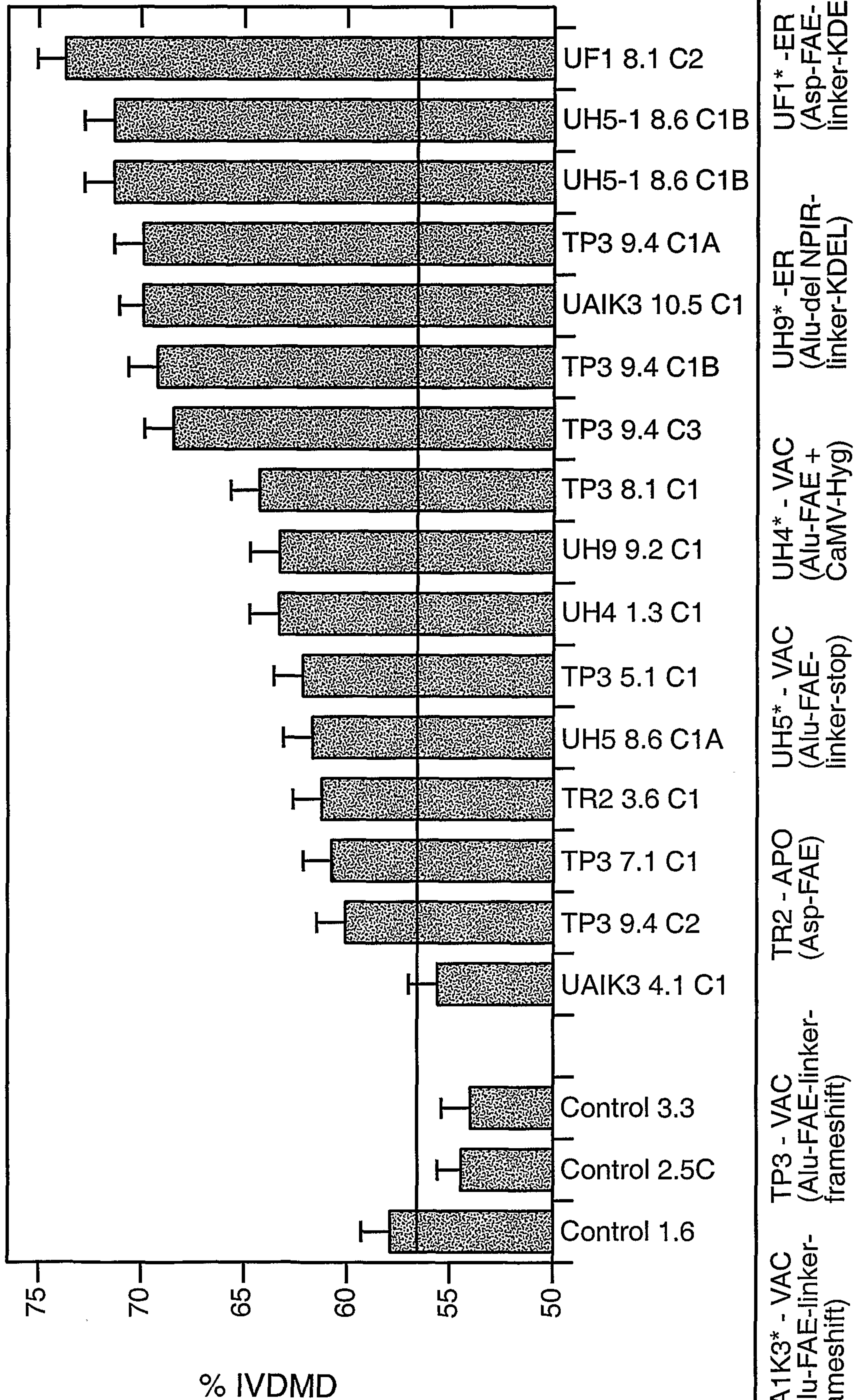


FIG. 22B

**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

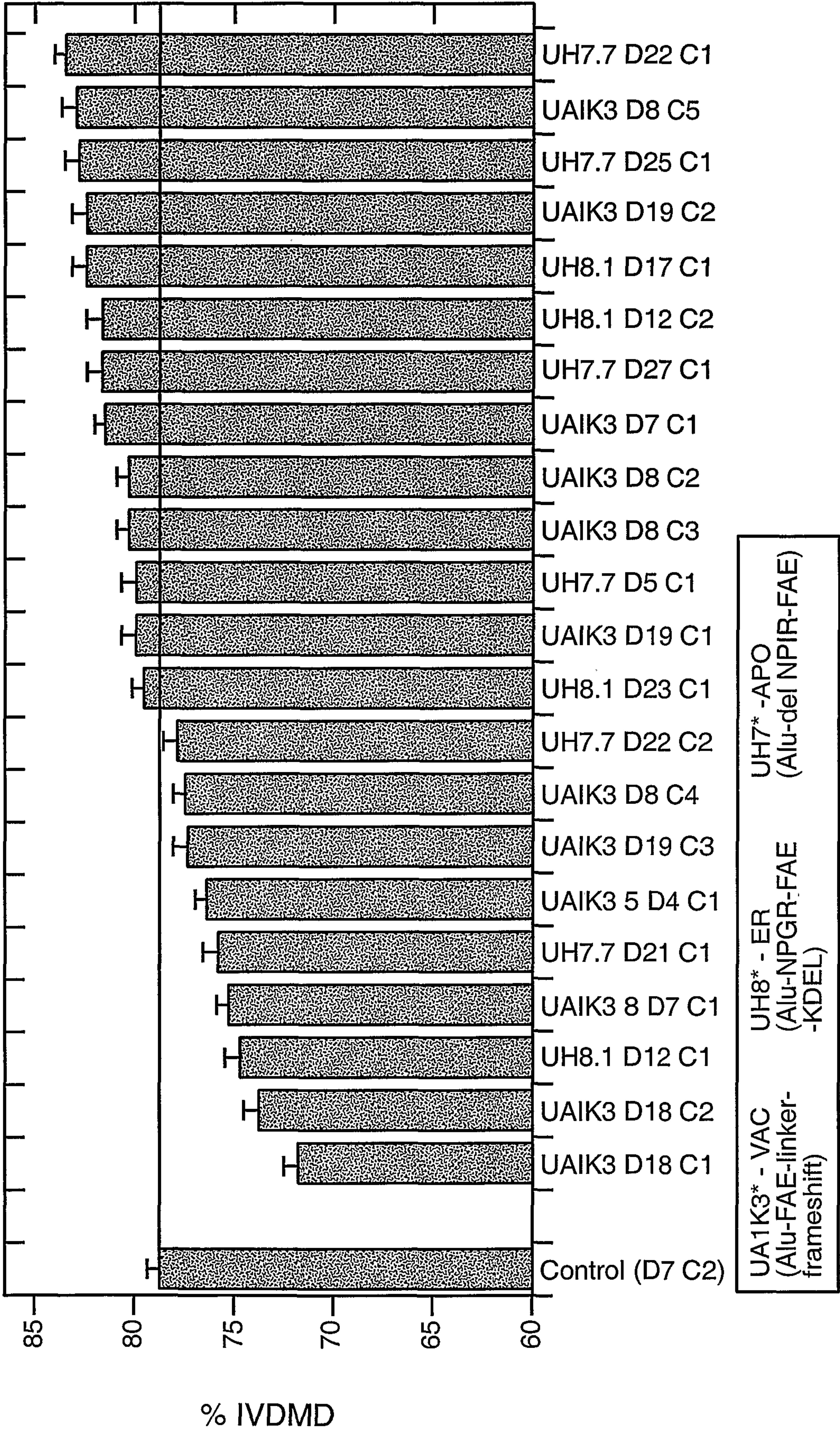


FIG. 24

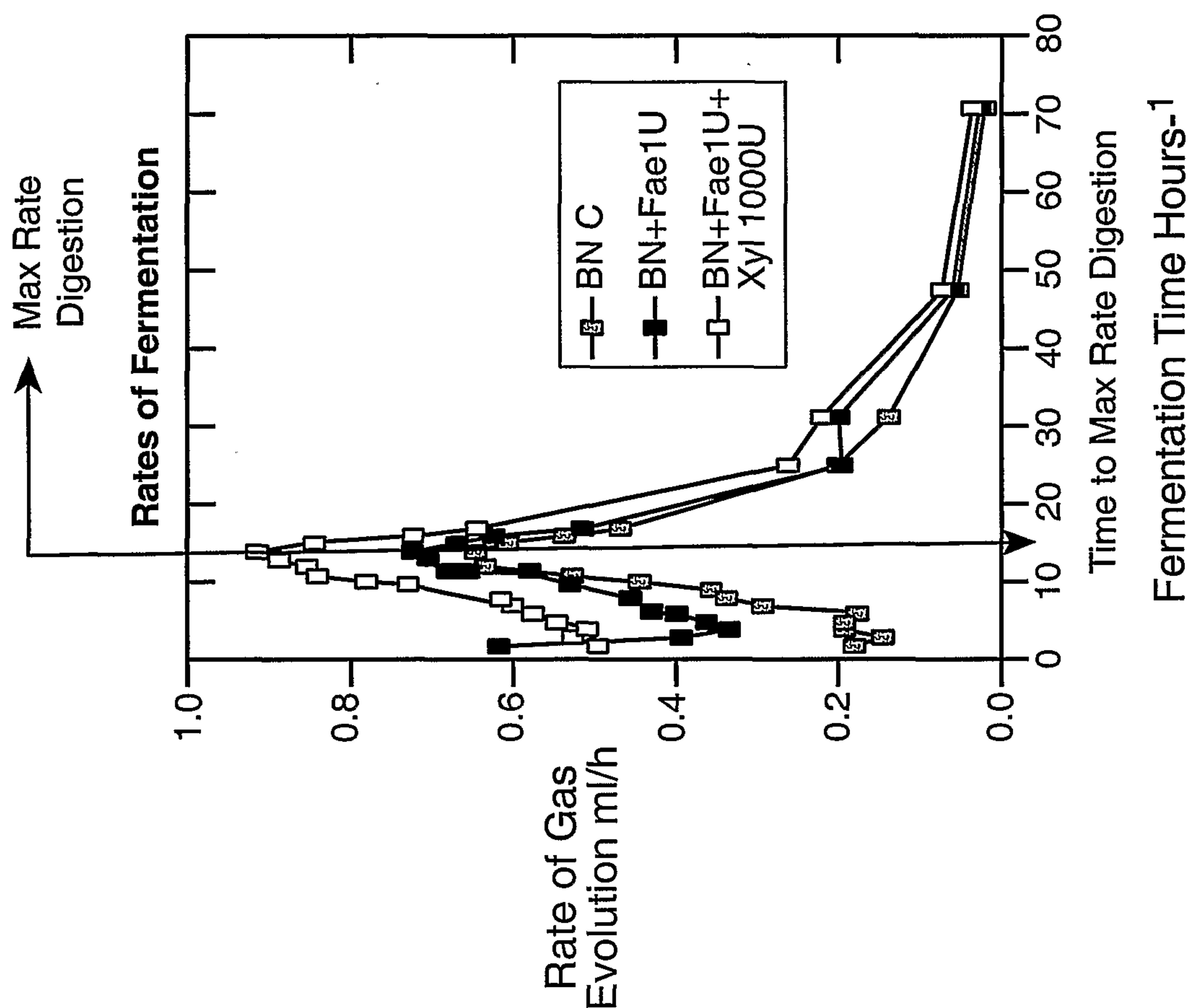


FIG. 25B

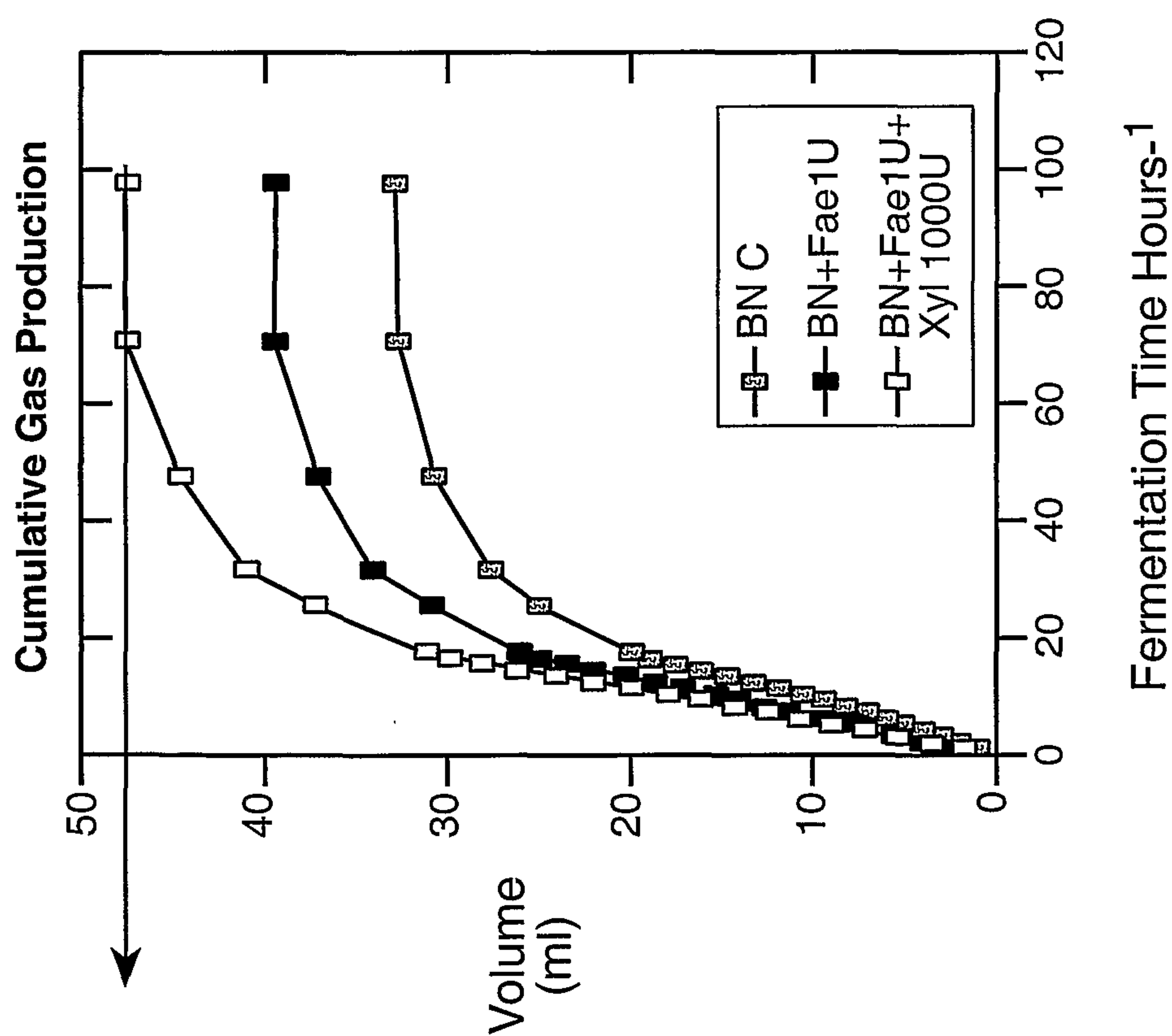


FIG. 25A

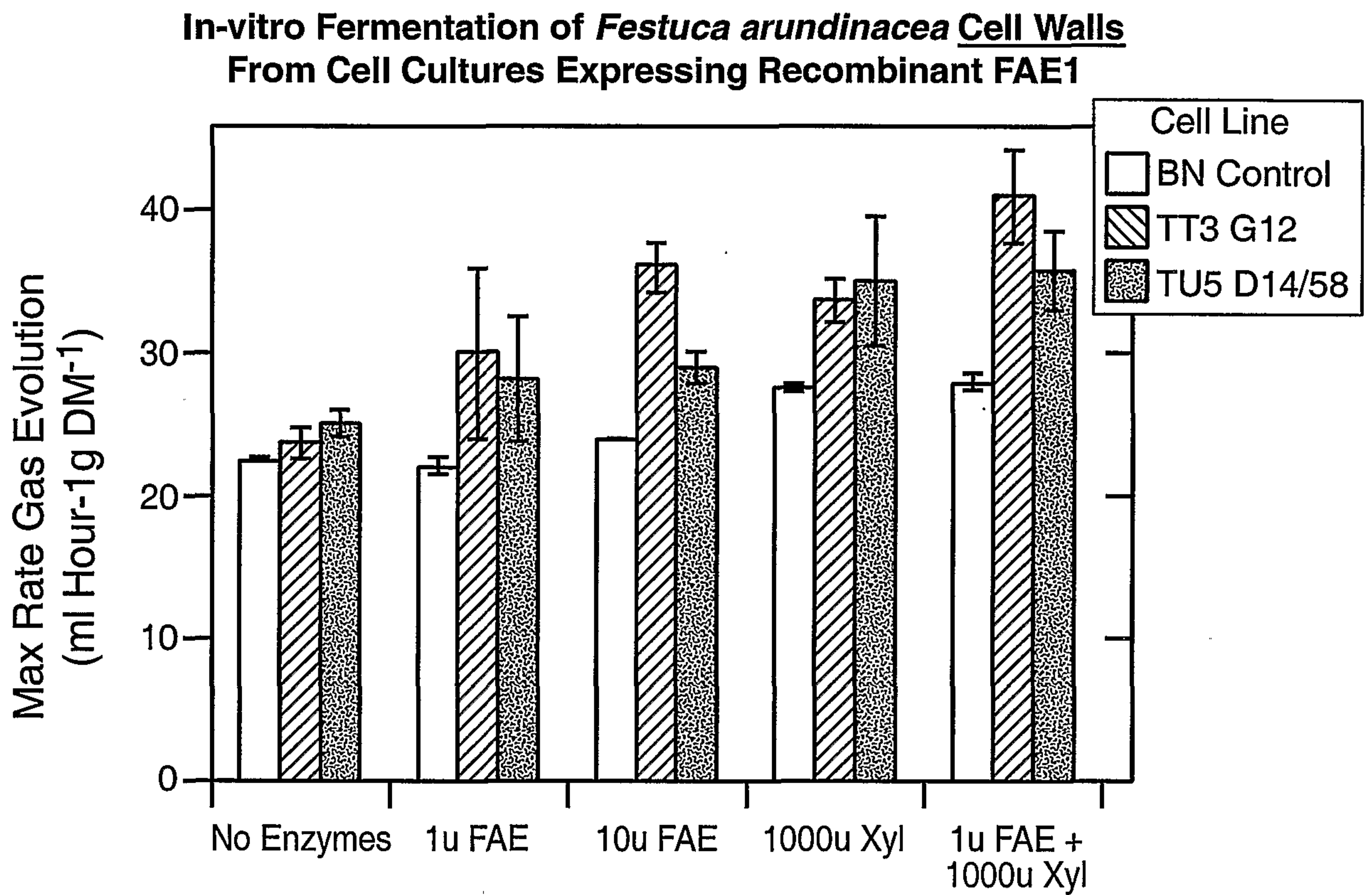


FIG. 26A Maximum Rate of Digestion

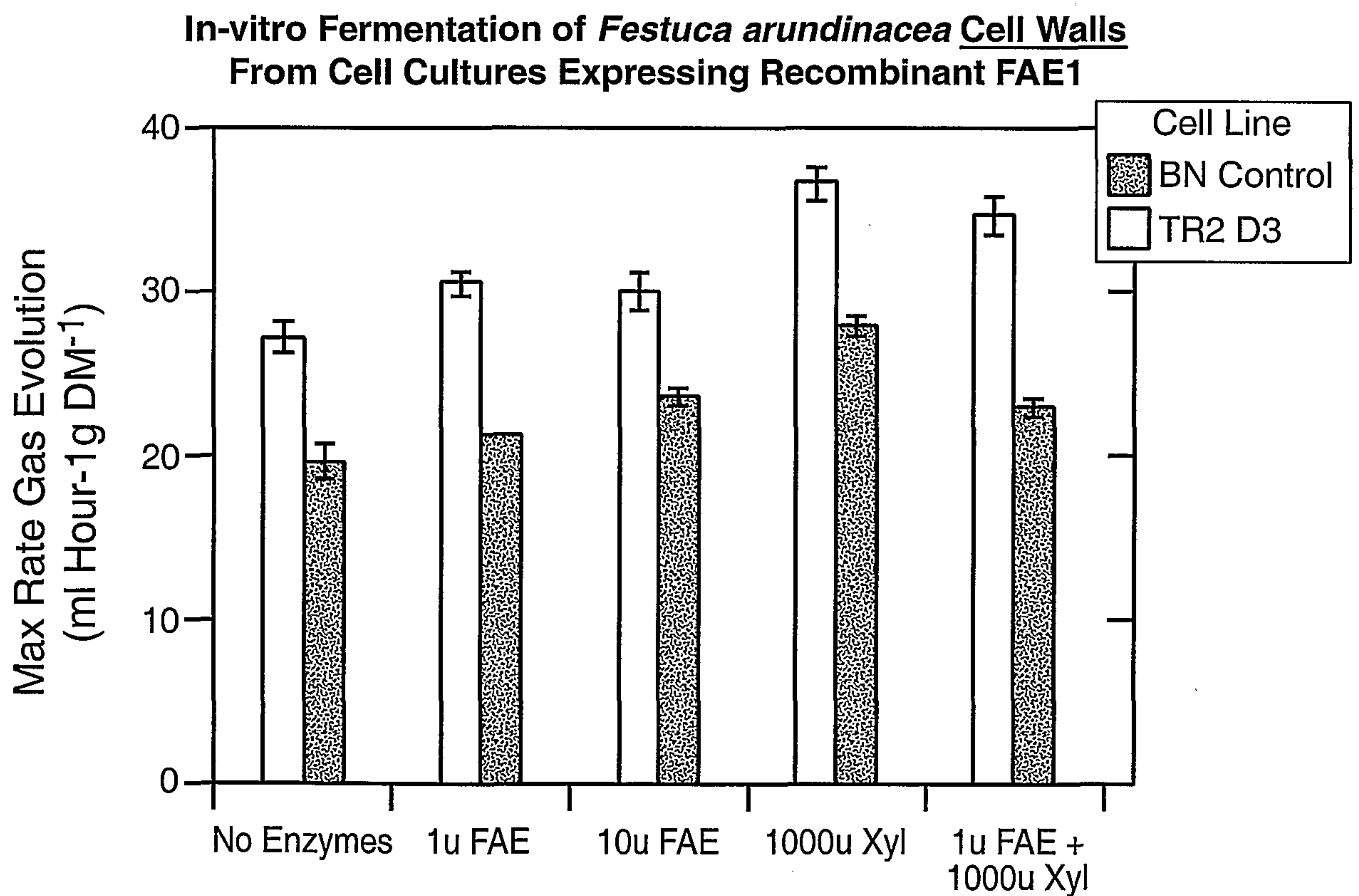
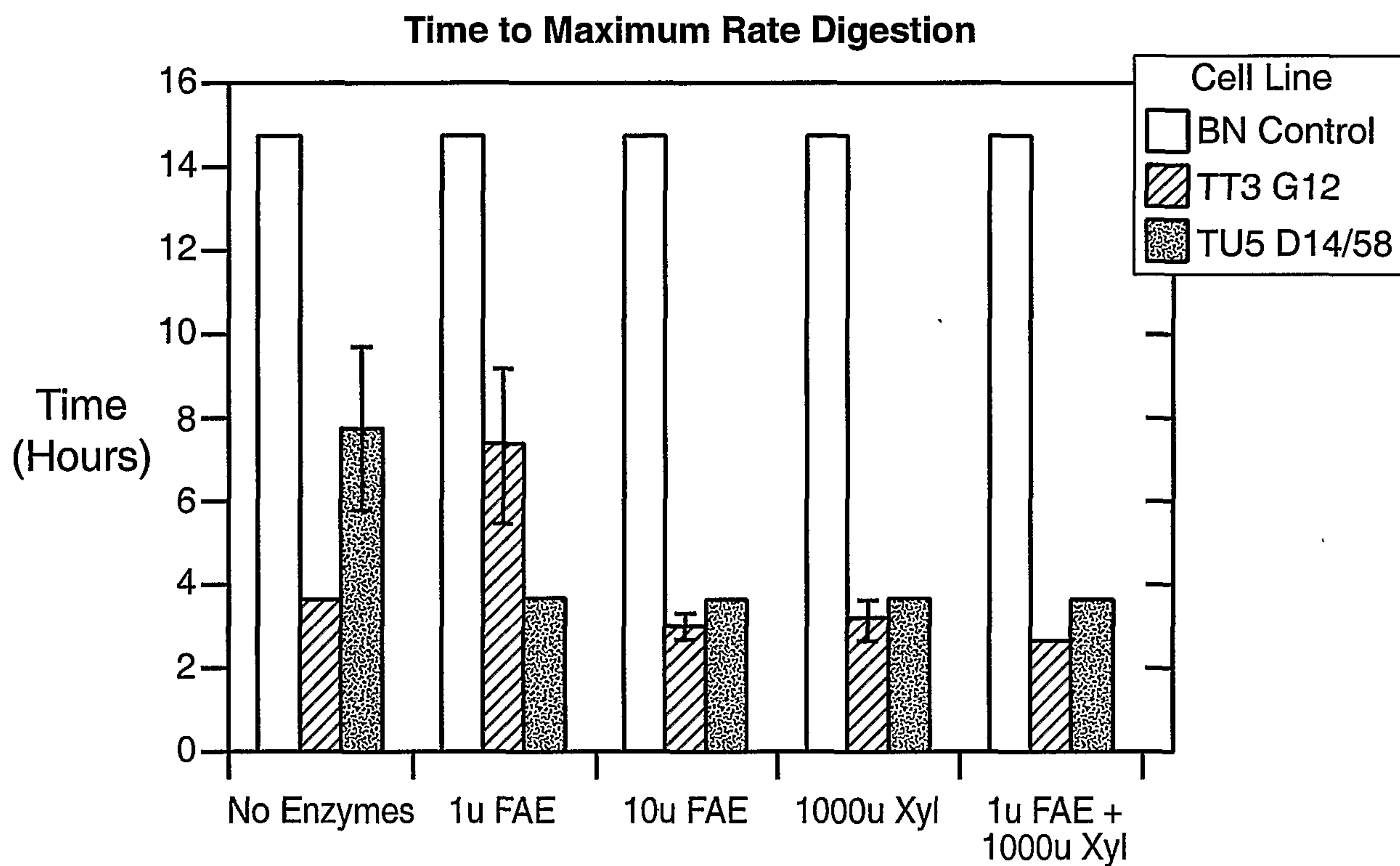
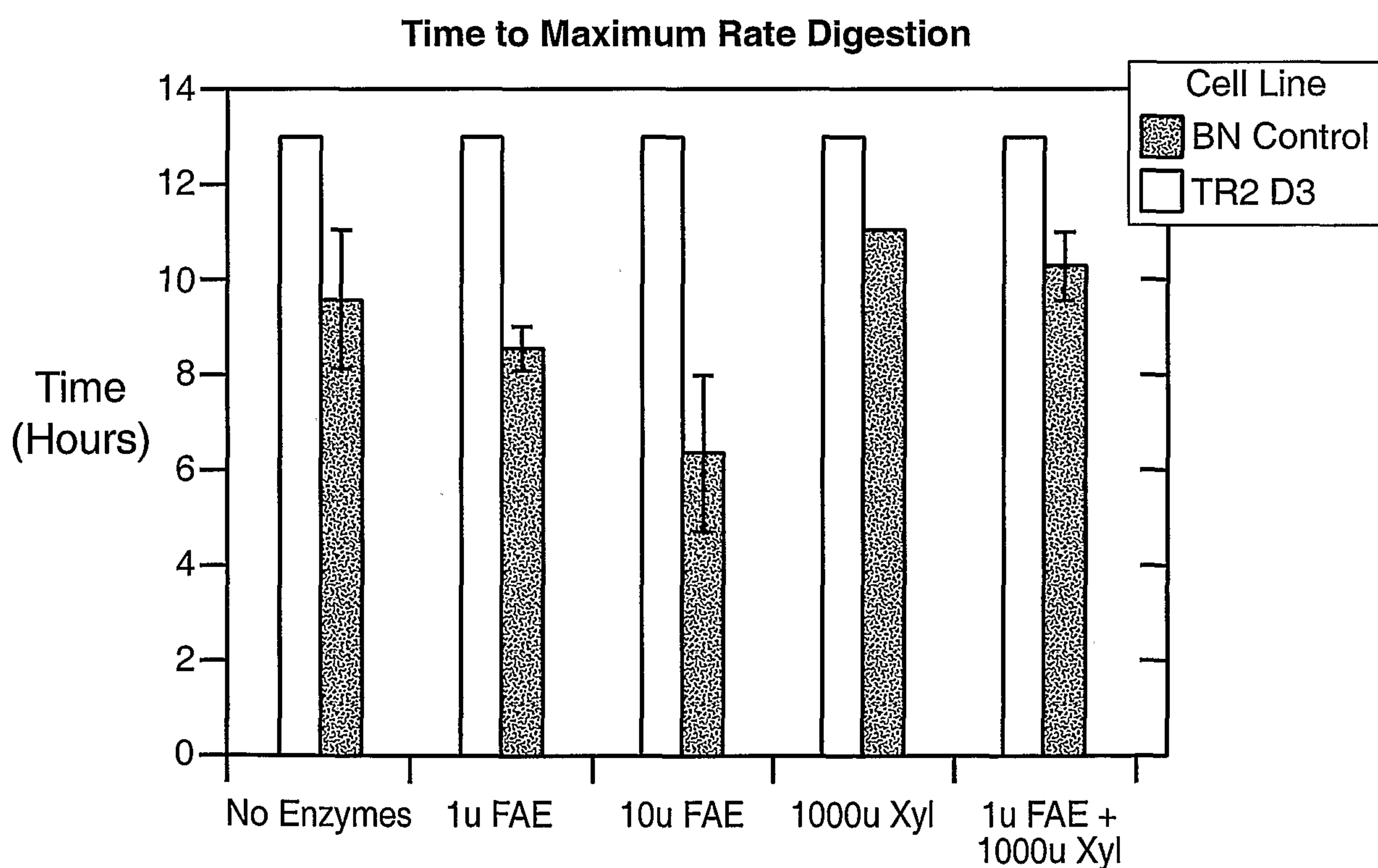
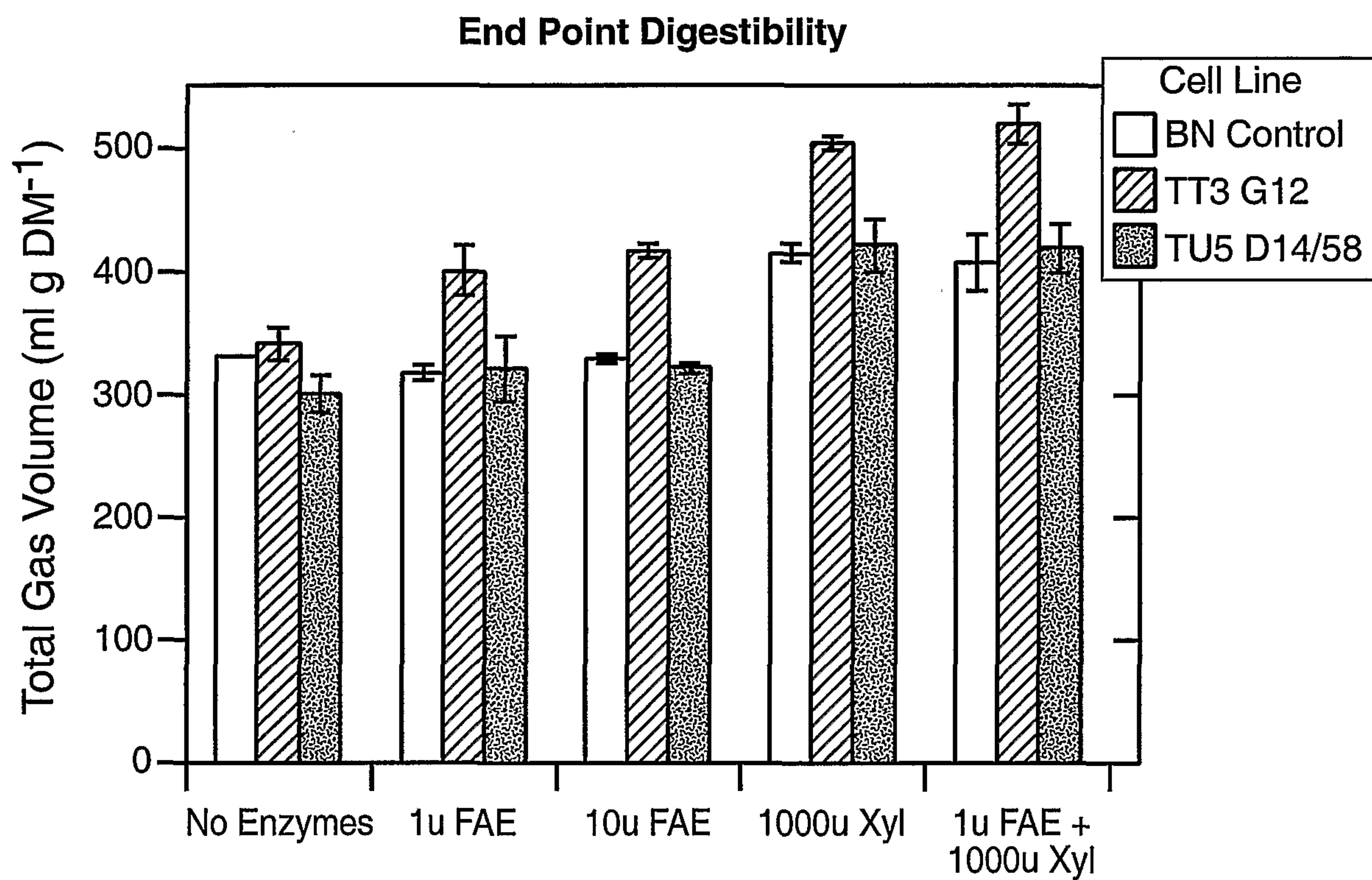
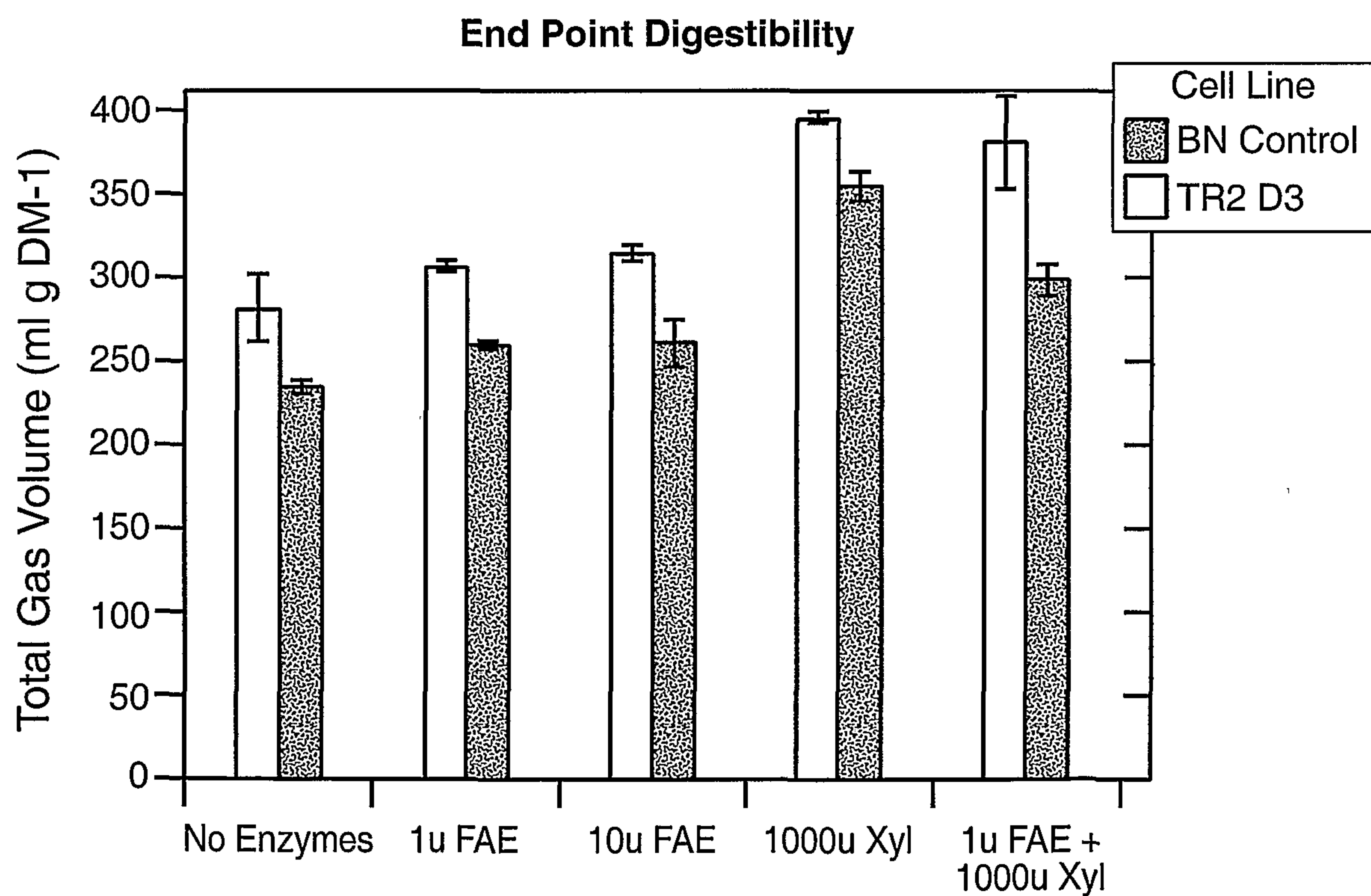
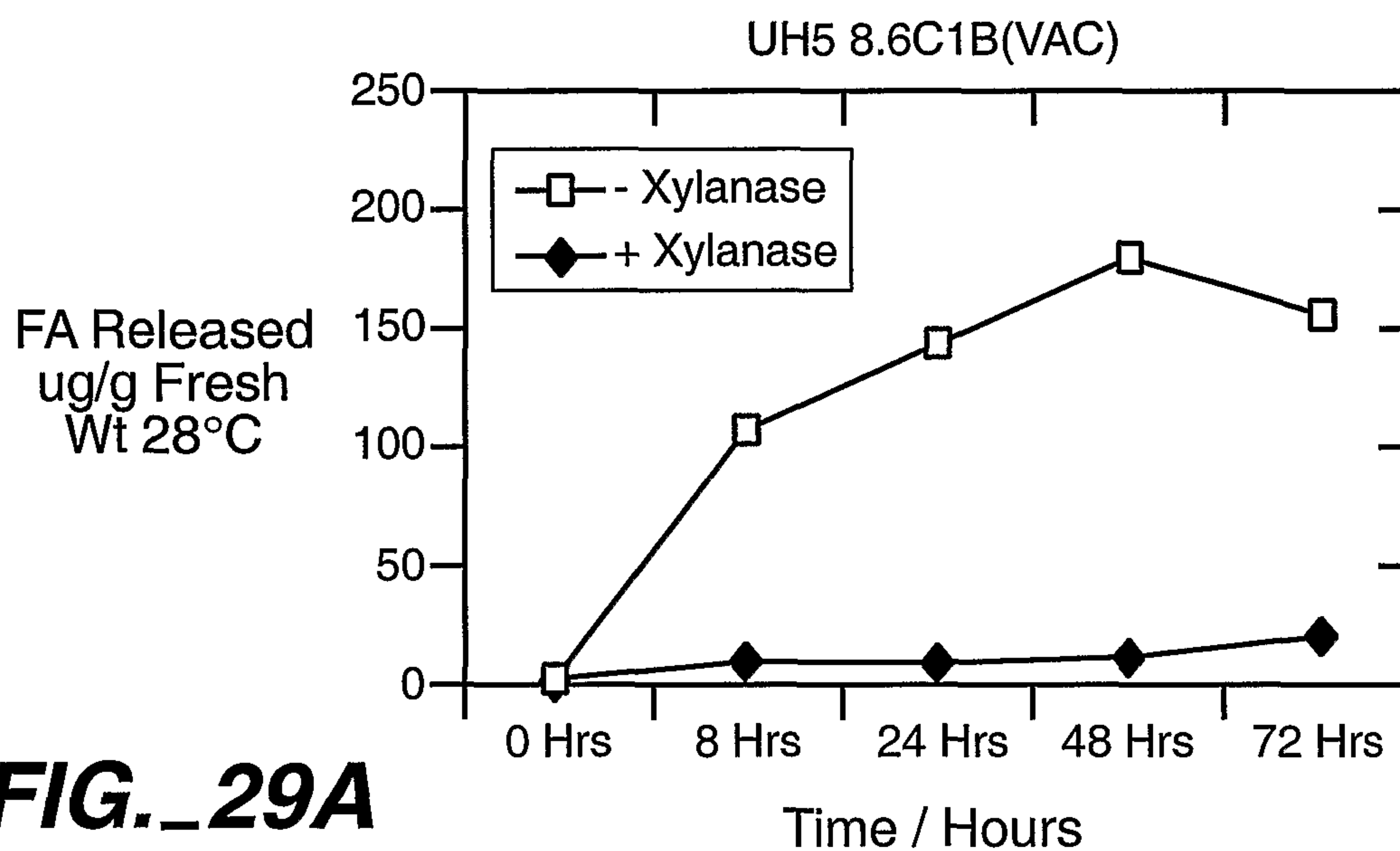


FIG. 26B Maximum Rate of Digestion

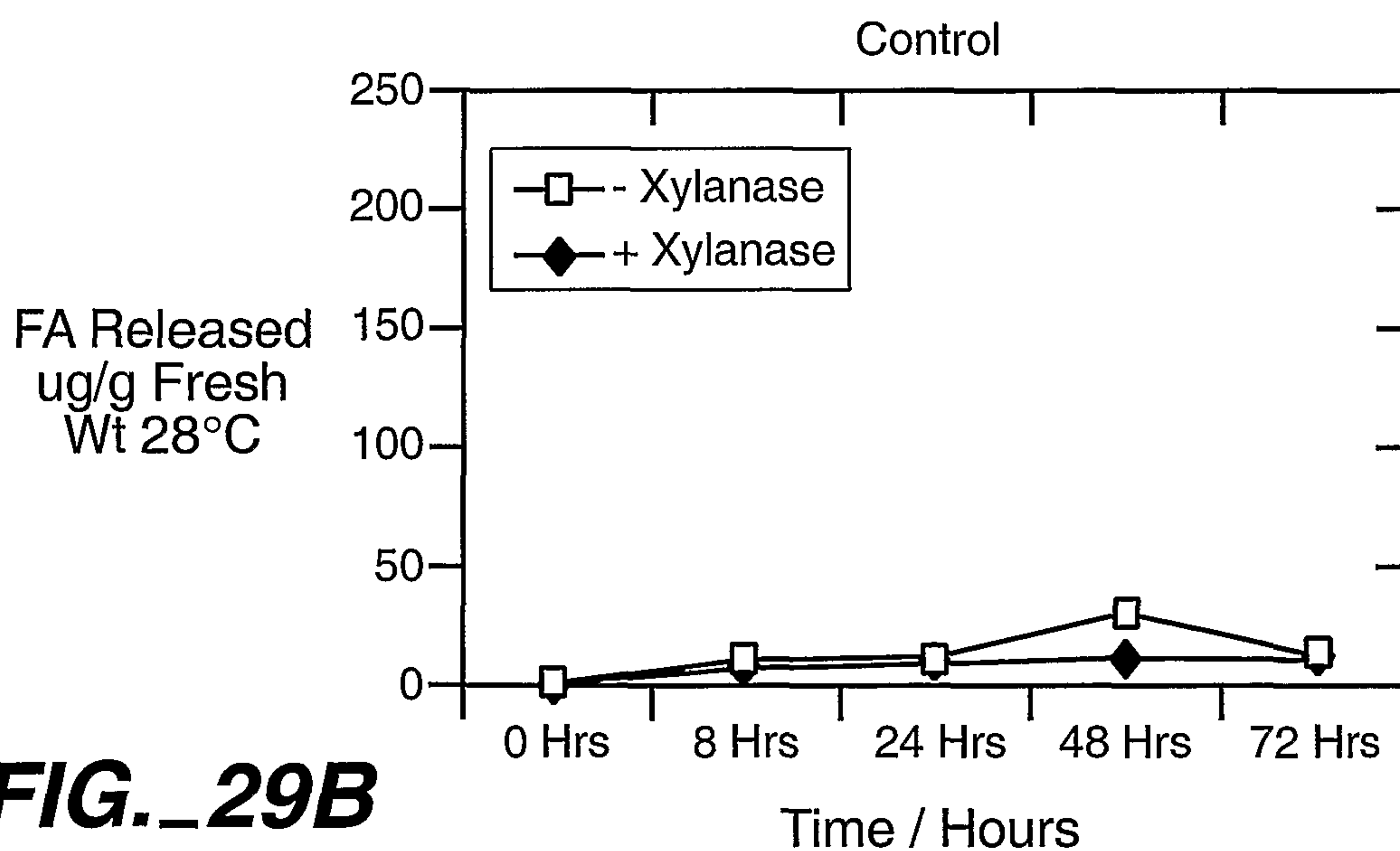
**FIG. 27A****FIG. 27B**

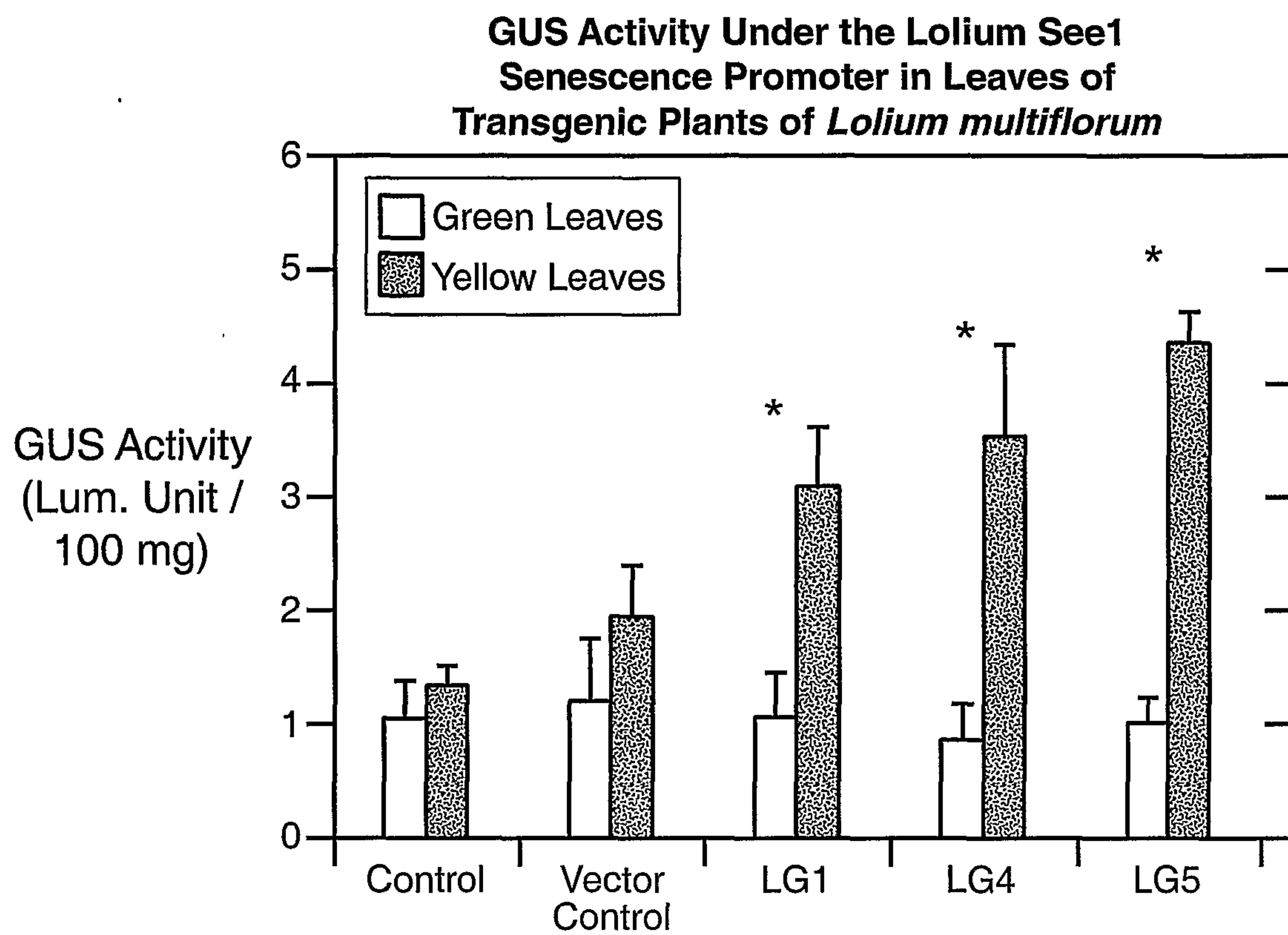
**FIG. 28A****FIG. 28B**

**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

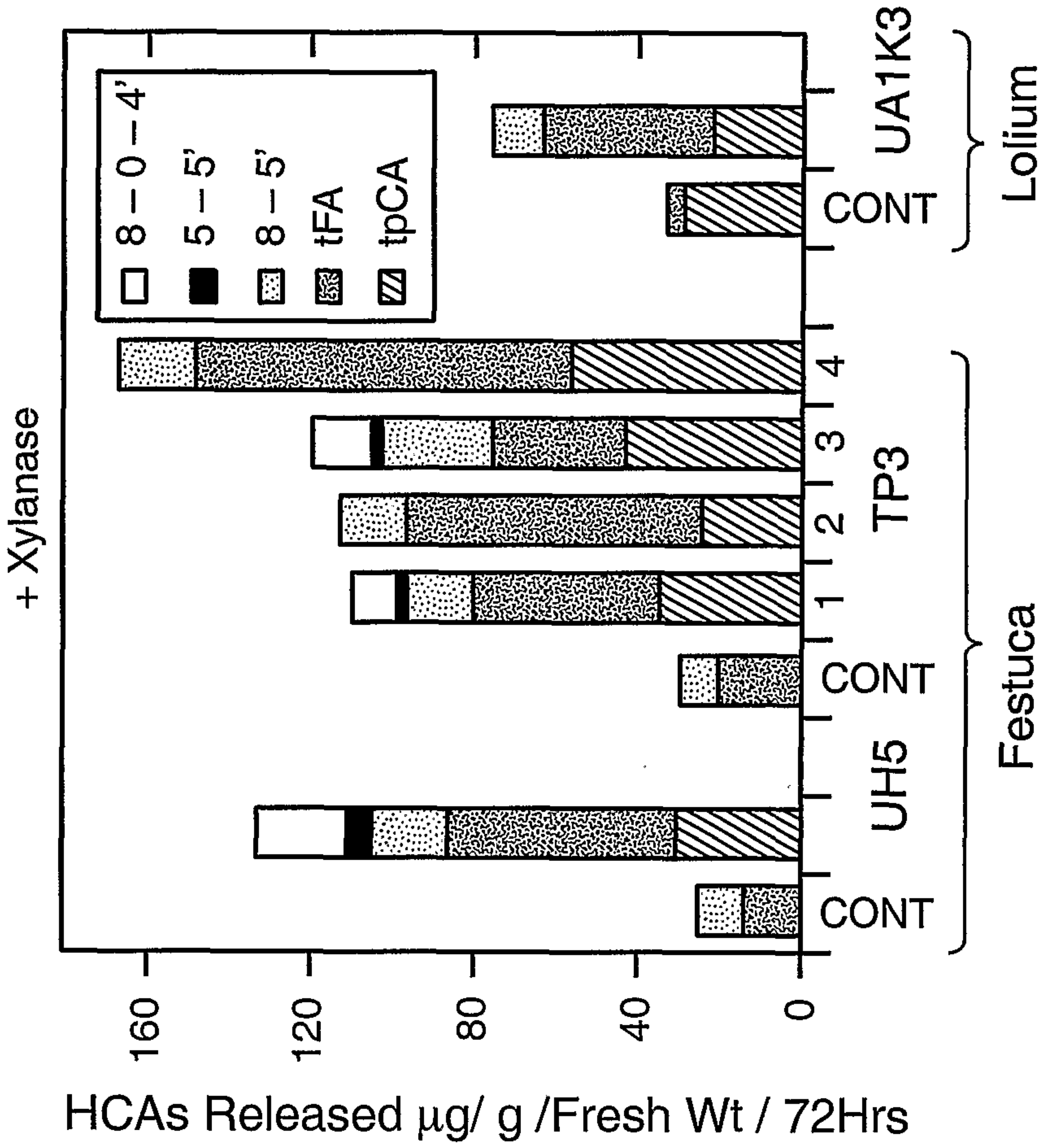


FIG._31B

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

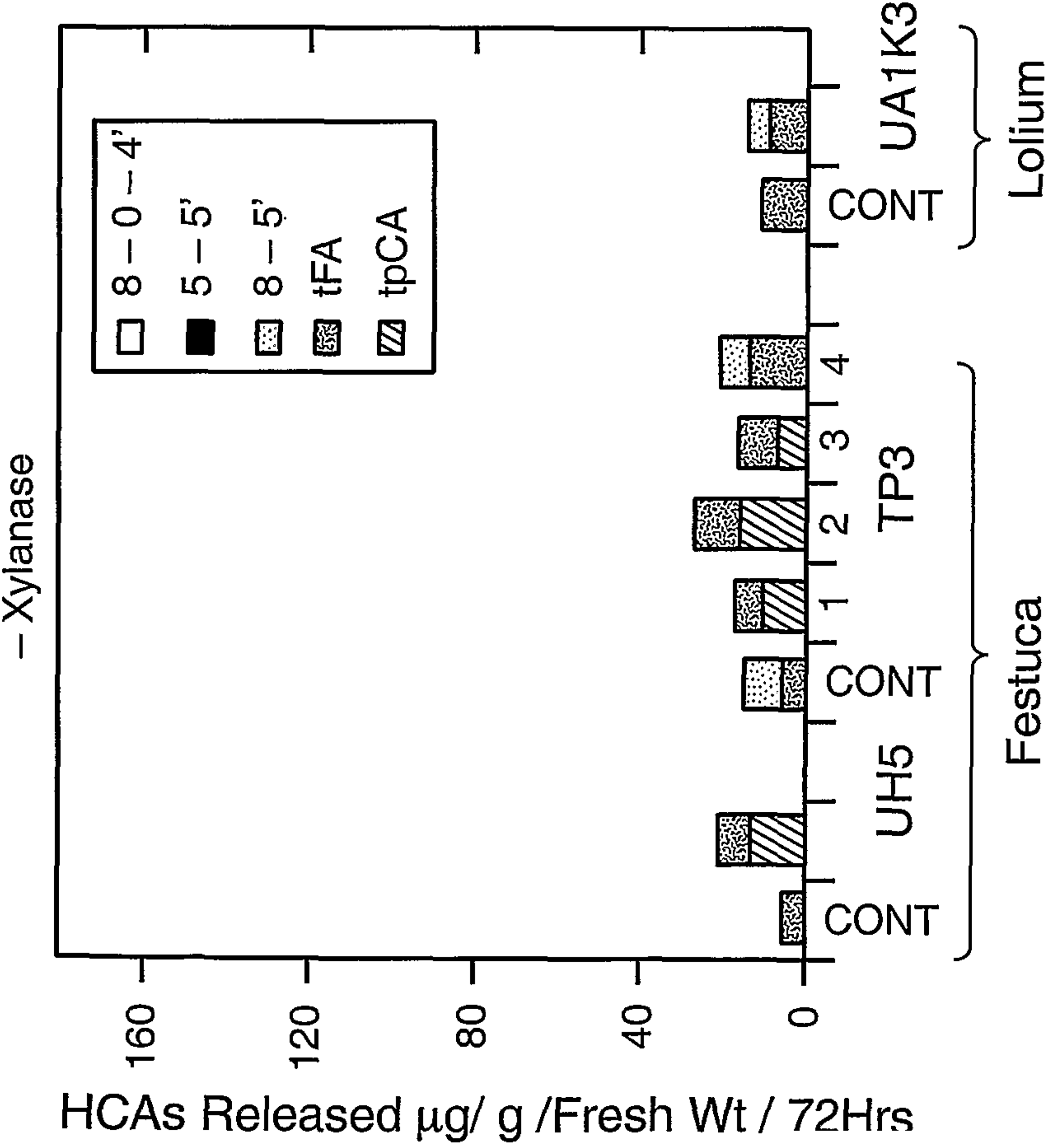


FIG._31A

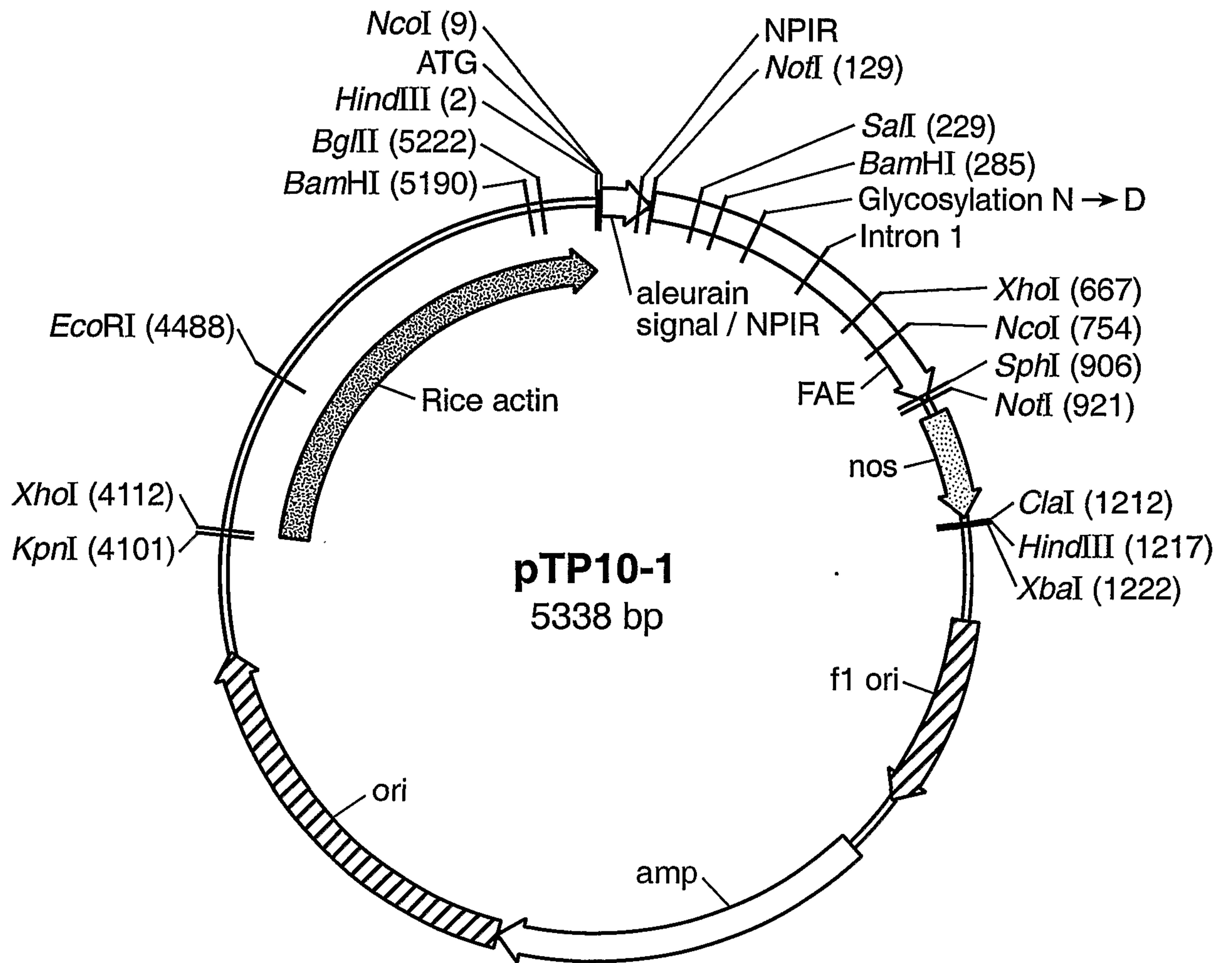
**FIG. 32A**

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

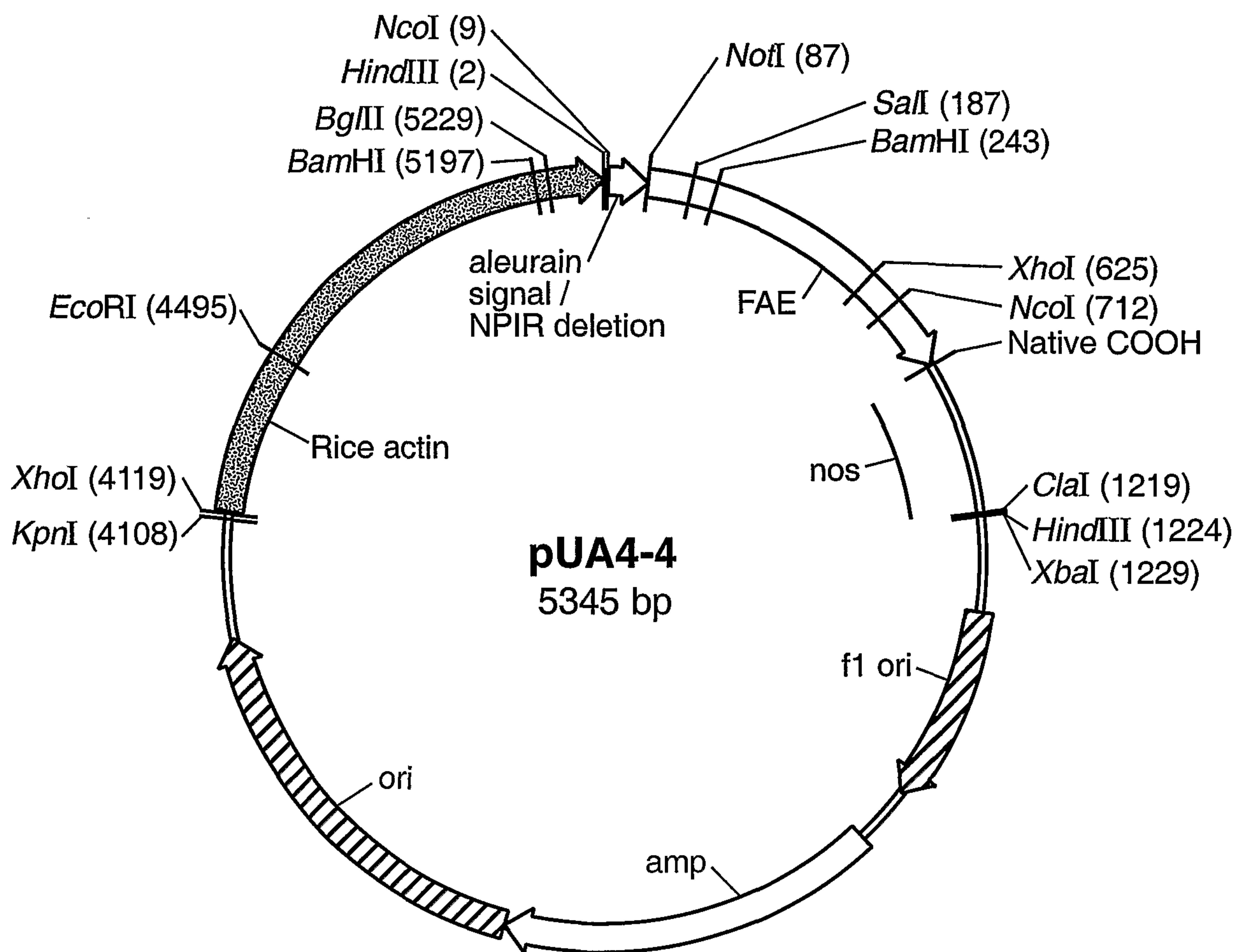
FIG._32C

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 GGTACGTA GTGGCCATC
1681 GCCCTGATAG ACGGTTTTC CACTTTTGAC GCCCTTTGAC GTTGGAGTCC ATTTCTTTG ATTTATAAGG GATTTGCCG ATTTCCGCC
1751 ACTGGAACAA CACTCAACCC TATCTCGGTC TATCTTTTG AATTTAACG GAATTTTAA AAAATATTAA CGCTTACAAT
1821 ATTGGTTAAA AATGAGCTG CTTTTCGGG AAATGTGCG AAATGCCCTA AATAATTGA AAAAGGAAGA GTATGAGTAT
1891 TTAGGTGGCA CTTTTCGGG CATGAGACAA CATGCCCTG TATTTCCCTT TTTTGCCCTC CACGAGTGG TTACATCGAA CTGGATCTCA
1961 TGTATCCGCT CATGAGACAA CATGCCCTG TATTTCCCTT TTTTGCCCTC AATAATTGA AAAAGGAAGA GTATGAGTAT
2031 TCAACATTTT CGTGTGCC CAGTAAAGA TGGTTGAGTA TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
2101 ACGCTGGTGA AAGTAAAGA GATCCTTGAG GCGGTATTAT TGGTTGAGTA TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
2171 ACAGCGGTAA GATCCTTGAG GCGGTATTAT TGGTTGAGTA TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
2241 GCTATGTGGC CAGAAATGACT TATGCAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
2311 CAGAAATGACT TATGCAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
2381 TATGCAGTGC TATGCAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
2451 GAAGGAGCTA CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
2521 CTGAATGAAG CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
2591 AACTATTAACT TATGCAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
2661 AGTTGCAGGA CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
2731 GAGCGTGGGT CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
2801 ACACGACGGG CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
2871 TAAGCATTGG CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
2941 TTAAAGGA CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
3011 TCCACTGAGC CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
3081 CTGCTGCTTG CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
3151 CTTTTCCTG CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
3221 TAGGCCACCA CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
3291 TGCTGCCAGT CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
3361 CCGTCGGGCT CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
3431 ACCTACAGCG CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
3501 CGGCAGGGTC CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
3571 GTCGGGTTTC CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
3641 AAAACGCCAG CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
3711 TGCCTTATCC CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
3781 CGAACGACCG CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC
3851 CCGCGCGTTG CAGGAGTGC TATGCAGTGC TGGCATAACC ACCGCTTTT GAAGGAGCTA CAGGAGTGC

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 | CATTTTGTGA | TGAATTGGTT | TTTAAGTTTA |
| 4341 | TTCCGCCGATT | GGAATATGCAT | ATCTGTATTT | GAGTCGGTTT | TTAAAGTTTCT | TGCTTTTGTG | AATACAGAGG |
| 4411 | GATTTGTTATA | AGAAATATCT | TTAAAAAACC | CATATGCTAA | TTTGACATAA | TTTTTGAGAA | AAATATATAT |
| | | | | ECORI | | | |
| | | | | ~~~~~ | | | |
| 4481 | TCAGGCGAAT | TCCACAAATGA | ACAATAATAA | GATTAAAATA | GCTTGCCCCC | GTTCAGCGGA | TGGGTATTTT |
| 4551 | TTCTAGTAAA | ATAAAAGATA | AACCTAGACT | CAAAACATTT | ACAAAACAA | CCCCTAAAGT | CCTAAAGCCC |
| 4621 | AAAGTGCTAT | GCACGATCCA | TAGCAAGCCC | AGCCCAACCC | AACCCACCCC | AACCCACCCC | AGTGCAGCCA |
| 4691 | ACTGGCAAAT | AGTCTCCACC | CCCGGCACTA | TCACCGTGAG | TTGTCCGCAC | CACCGCACGT | CTCGCAGCCA |
| 4761 | AAAAAATAAA | AAGAAAGAAA | AAAAAGAAA | AGAAAACAG | CAGGTGGGTC | CGGTCTGTGG | GGGCCGGA |
| 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 | TTTGTTAGC | | | | | |

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 TACAACCTCA 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 CAGTATCCG 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 CTTGCGCGAA 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 CGCAGTATTT 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 AGAGCGGCCG GTGGAGCTCC AATTCGCCCT
1261 ATAGTGAGTC GTATTACGCG CGCTCACTGG CCGTCGTTTT ACAACGTCGT GACTGGGAAA ACCCTGGCGT
1331 TACCCAACTT AATCGCCTTG CAGCACATCC CCTTTCGCC AGCTGGCGTA ATAGCGAAGA GGCCCGCAC
1401 GATCGCCCTT CCCAACAGTT GCGAGCCTG AATGGCGAAT GGGACGCGC CTGTAGCGC GCATTAAGCG
1471 CGGCGGGTGT GGTGGTTACG CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTCGC
1541 TTTCTTCCCT TCCTTTCTCG CCACGTTCCG CGGCTTCCC CGTCAAGCTC TAAATCGGG GCTCCCCTTA
1611 GGGTTCCGAT TTAGTGCTTT ACGCACCTC GACCCAAA AACTTGATTA GGTGATGGT TCACGTAGTG
1681 GGCCATCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTCCTTTAATA 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 AAGTTCCTGCT ATGTGGCGCG GTATTATCCC GTATTGACGC CGGGCAAGAG CAACTCGGTC 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

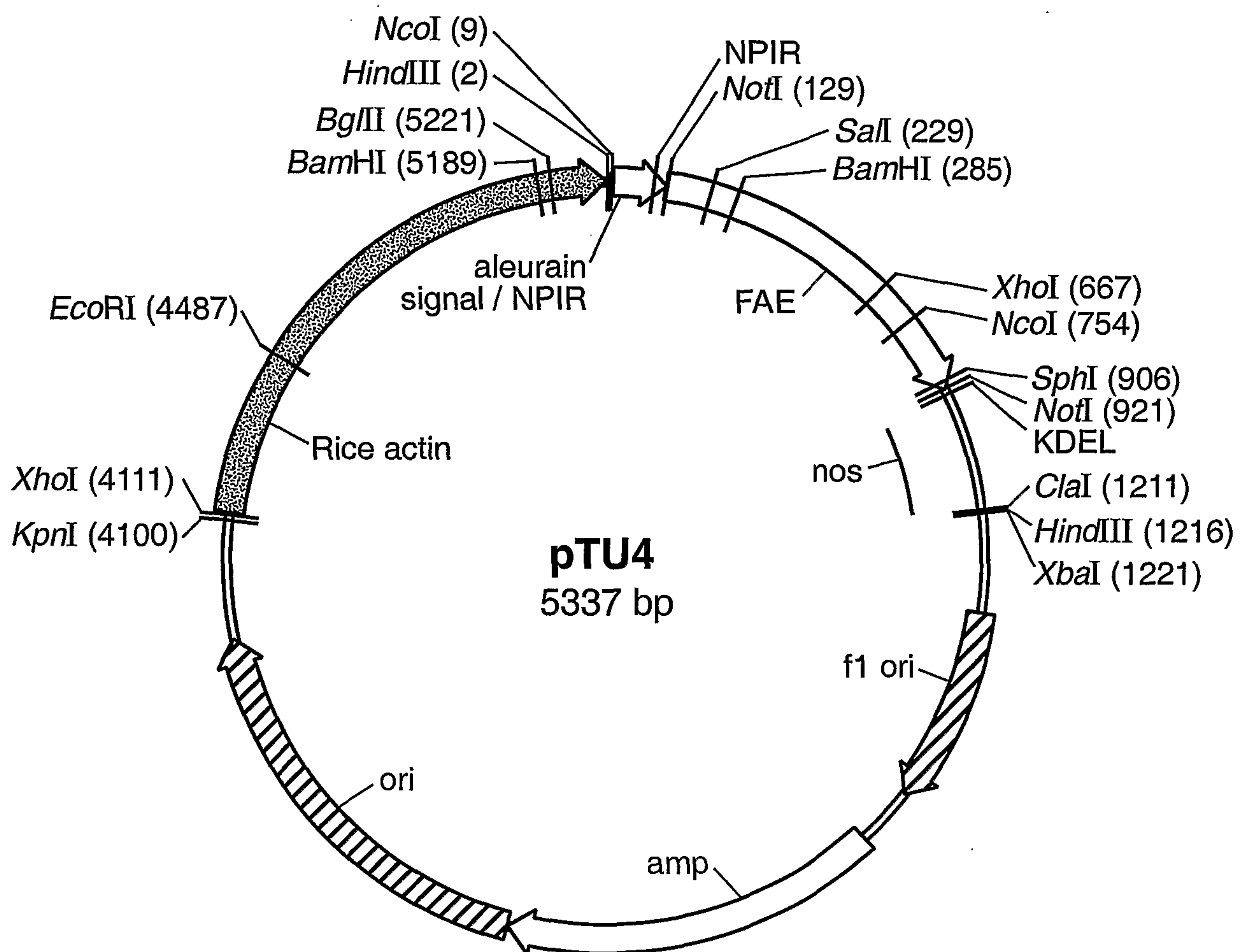
FIG._33C

| | | | | | | | |
|-----------|---------------|-------------|-------------|-------------|---------------|--------------|--------------|
| 2731 | AGCCGGTGAG | CGTGGGTCTC | GCGGTATCAT | TGCAGCACTG | GGCCAGATG | GTAAGCCCTC | CCGTATCGTA |
| 2801 | GTATCTACA | CGACGGGGAG | TCAGGCAACT | ATGGATGAAC | GAAATAGACA | GATCGCTGAG | ATAGGTGCCCT |
| 2871 | CAC TGATTAA | GCA TTGGTAA | CTGT CAGACC | AAGTTTACTC | ATATATACTT | TAGATTGATT | TAAAAACTTCA |
| 2941 | TTTTTTAAATT | AAAAGGATCT | AGGTGAAGAT | CCTTTT TGAT | AATCTCATGA | CCAAAAATCCC | TTAACGTGAG |
| 3011 | TTTTTCGTTCC | ACTGAGCGTC | AGACCCCGTA | GAAAAGATCA | AAGGATCTTC | TTGAGATCCT | TTTTTTTCTGC |
| 3081 | GC GTAATCTG | CTGCTTGCAA | ACAAAAAAAC | CACCGCTACC | AGCGGTGGTT | TGTTTGCCGG | ATCAAGAGCT |
| 3151 | ACCAACTCTT | TTTCCGAAGG | TAACTGGCTT | CAGCAGAGCG | CAGATACCAA | ATACTGTCCT | TCTAGTGTAG |
| 3221 | CCGTAGTTAG | GCCACCACCT | CAAGAACTCT | GTAGCACCGC | CTACATACTT | CGCTCTGCTA | ATCCTGTTAC |
| 3291 | CAGTGGCTGC | TGCCAGTGGC | GATAAGTCGT | GTCTTACCGG | GTTGGACTCA | AGACGATAGT | TACCGGATAA |
| 3361 | GGCGCAGCGG | TCGGGCTGAA | CGGGGGGTC | GTGCACACAG | CCCAGCTTGG | AGCGAAACGAC | CTACACCGAA |
| 3431 | CTGAGATACC | TACAGCGTGA | GCTATGAGAA | AGGCCACGCG | TTCCCGAAGG | GAGAAAGGCG | GACAGGTATC |
| 3501 | CGGTAAGCGG | CAGGTCGGA | ACAGGAGAGC | GCACGAGGGA | GCTTCCAGGG | GGAAACGCCCT | GGTATCTTTA |
| 3571 | TAGTCCCTGTC | GGGTTCGCC | ACCTCTGACT | TGAGCGTCCA | TTTTTTTG TGAT | GCTCGTCAGG | GGGGCGGAGC |
| 3641 | CTATGGAAAA | ACGCCAGCAA | CGCGGCCTTT | TTACGGGTTCC | TGGCCCTTTTG | CTGGCCCTTT | GCTCACATGT |
| 3711 | TC TTTCCCTGC | GTTATCCCCCT | GATCTGTGG | ATAACCGTAT | TACCGCCCTTT | GAGTGAGCTG | ATACCGCTCG |
| 3781 | CCGCAGCCGA | ACGACCGAGC | GCAGCGAGTC | AGTGAGCGAG | GAAAGCGGAAG | AGCGCCCAAT | ACGCAAACCG |
| 3851 | CCTCTCCCCG | CGCGTTGGCC | GATTCATTAA | TGCAGCTGGC | ACGACAGGTT | TCCCGACTGG | AAAGCGGGCA |
| 3921 | GTGAGCGCAA | CGCAATTAAAT | GTGAGTTAGC | TCACTCAATTA | GGCACCCCCAG | GCTTTACACT | TTATGCTTCC |
| 3991 | GGCTCGTATG | TTGTGTGGAA | TTGTGAGCGG | ATAACAAATT | CACACAGGAA | ACAGCTATGA | CCATGATTAC |
| KpnI XhoI | | | | | | | |
| 4061 | GCCAAGCGCG | CAATTAAACC | TCACTAAAGG | GAACAAAAGC | TGGGTACC GG | GCCCCCCTC | GAGGTCATTC |
| 4131 | ATATGCTTGA | GAAGAGAGTC | GGGATAGTCC | AAAATAAAAC | AAAGGTAAGA | TTACCTGGTC | AAAAGTGAAA |
| 4201 | ACATCAGTTA | AAAGGTGGTA | TAAGTAAAAAT | ATCGGTAATA | AAAGGTGGCC | CAAAGTGAAA | TTTACTCTTT |
| 4271 | TC TACTATTA | TAAAAAATTGA | GGATGTTTTG | TCGGTACTTT | GATACGTCAAT | TTTTTG TATGA | ATTGGTTTTT |
| 4341 | AAGTTTATTTC | GCGATTTGGA | AATGCATATC | TGTAATTTGAG | TCGGTTTT TA | AGTTCGTTGC | TTTTTG TAAAT |
| 4411 | ACAGAGGGAT | TTGTATAAGA | AAATATCTTTA | AAAAACCCAT | ATGCTAAATTT | GACATAATTT | TTGAGAAAAAA |
| EcoRI | | | | | | | |
| 4481 | TATATATTCA | GGCGAATTCC | ACAATGAACA | ATAATAAGAT | TAAAAATAGCT | TGCCCCCGTT | GCAGCGATGG |
| 4551 | GTATTTT TTTTC | TAGTAAAAATA | AAAGATAAAC | TTAGACTCAA | AACATTTTACA | AAAACAACCC | CTAAAGTCCCT |
| 4621 | AAAGCCCAA | GTGCTATGCA | CGATCCATAG | CAAGCCCAGC | CCAAACCCAAC | CCAACCCAAC | CCACCCAGT |
| 4691 | GCAGCCCAACT | GGCAAAATAGT | CTCCACCCCC | GGCACTATCA | CCGTGAGTTG | TCCGCACCCAC | CGCACGTCTC |
| 4761 | GCAGCCCAAAA | AAAAAAA AAG | AAAGAAAAAA | AAGAAAAAGA | AAAACAGCAG | GTGGGTCCGG | GTCTGTGGGG |
| 4831 | CCGGAAAAGC | GAGGAGGATC | GCGAGCAGCG | ACGAGGCCCG | GCCCTCCCTC | CGCTTCCAAA | GAAACGCCCC |

FIG._33D

| | | | | | | | |
|------|-------------|------------|------------|-------------|------------|------------|------------|
| 4901 | CCATCGCCAC | TATATACATA | CCCCCCCCTC | TCCTCCCATC | CCCCCAACCC | 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 | GGCGTCTCCG |
| | | BamHI | | | BglII | | |
| | | ~~~~~ | | | ~~~~~ | | |
| 5181 | GGCGTGAGTC | GGCCCGGATC | CTCGCGGGA | ATGGGGCTCT | CGGATGTAGA | TCTTCTTCT | TTCTTCTTT |
| 5251 | TGTGGTAGAA | TTTGAATCCC | TCAGCATTGT | TCAATCGGTAG | TTTTTCTTTT | CATGATTTGT | GACAAATGCA |
| 5321 | GCCTCGTGCG | GAGCTTTT | 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

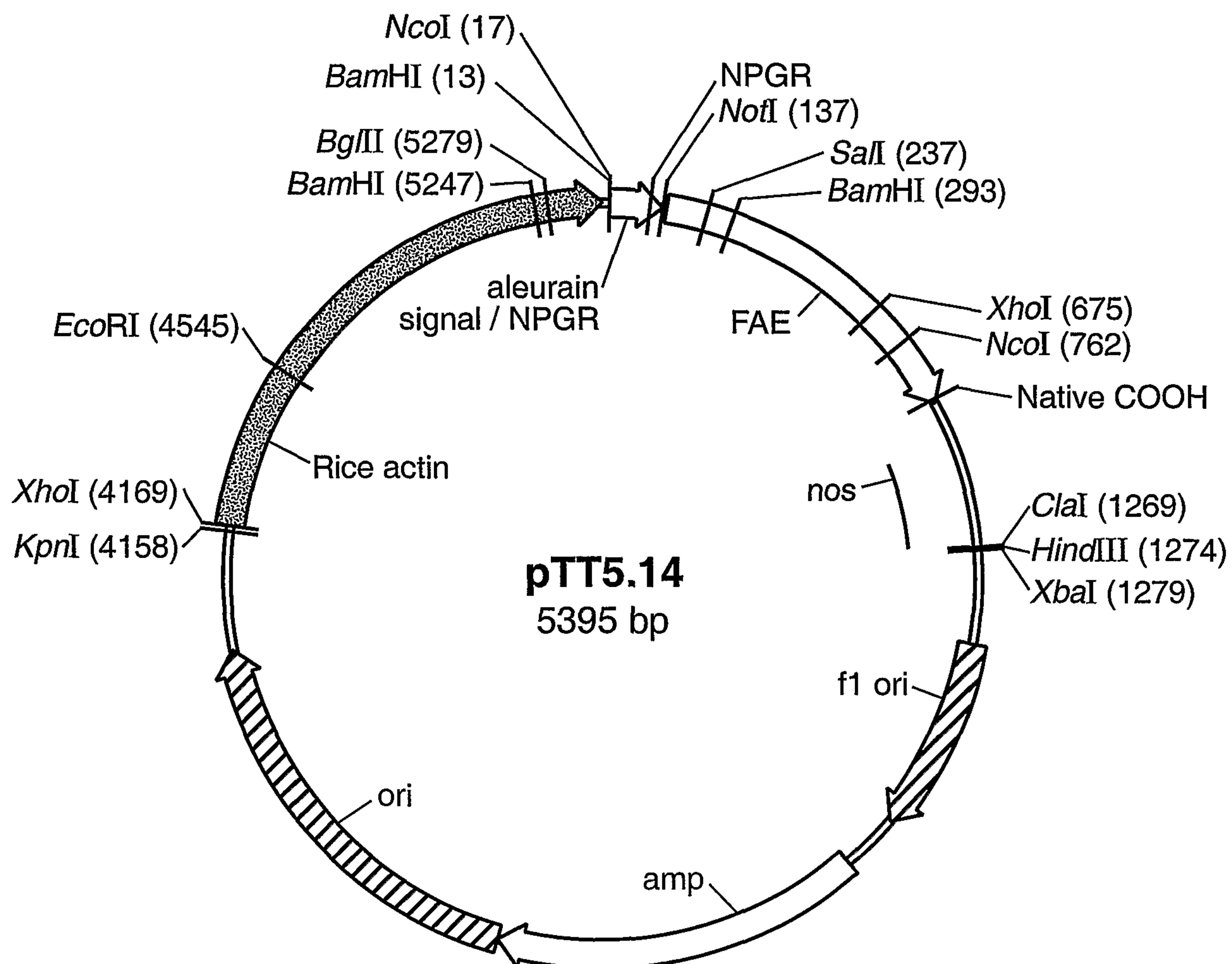
701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCGGTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
• W S V D P Y S A Q N T F V C T G D E V Q C C E  
771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCAC TGGG ATGAAGTGCA GTGCTGTGAG  
SphI  
~~~~~  
A Q G G Q G V N N A H T T Y F G M T S G A C T W
841 GCCCAGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC GCATGCACCT
NotI
~~~~~  
• P V A A A E P L K D E L \*  
911 GGCCGGTCGC GGCCGCGGAA CCACTGAAGG ATGAGCTGTA AAGAA GCAGA TCGTTCAAAC ATTGGCAAT  
981 AAGGTTTCTT AAGATTGAAT CCTGTTGCCG GTC TTGCGAT GATTATCATA TAATTTCTGT 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 AAACCTGGC GTTACCCAAC  
1331 TTAAATCGCCT TGCAGCACAT CCCCTTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCC  
1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCTGTAGCG CGCTCCTTTC GCGGCGGGT  
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 TAGTGACTC TTGTTCCAAA  
1751 CTGGAACAAC ACTCAACCCT ATCTCGGTCT ATTC TTTGA TTTATAAGG ATTTTGCCGA TTTCGGCCTA  
1821 TTGGTTAAAA AATGAGCTGA TTAAACAAA AATTTAACG AATTTAACA AAATAATAAC GCTTACAATT  
1891 TAGGTGGCAC TTTTCGGGGA AATGTGCGG GAACCCCTAT TTGTTTATTT TTCTAAATAT ATTCAAATAT  
1961 GTATCCGCTC ATGAGACAAAT AACCTGATA AATGCTTCAA TAATAATTGAA AAAGGAAGAG TATGAGTATT  
2031 CAACATTTCC GTGTCGCCCT TATTCCTTTT 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      | ACTATTAACT | GGCGAACTAC   | TTACTCTAGC  | TTCCCGGCAA  | CAATTAATAG   | ACTGGATGGA  | GGCGGATAAA |
| 2661      | GTTGCAGGAC | CAC TTCGCG   | CTCGGCCCTT  | CCGGCTGGCT  | GGTTTATTGC   | TGATAAAATCT | GGAGCCGGTG |
| 2731      | AGCGTGGGTC | TCGCGGTATC   | ATTGCAGCAC  | TGGGCCAGA   | TGGTAAGCCC   | TCCCCTATCG  | TAGTTATCTA |
| 2801      | CACGACGGGG | AGTCAGGCAA   | CTATGGATGA  | ACGAAATAGA  | CAGATCGCTG   | AGATAGGTGC  | CTCACTGATT |
| 2871      | AAGCATTGGT | AAC TGT CAGA | CCAA GTTTAC | TCATATATAC  | TTTAGATTGA   | TTTAAACCTT  | CATTTTAAAT |
| 2941      | TTAAAAGGAT | CTAGGTGAAG   | ATCCTTTT    | ATAATCTCAT  | GACCAAAATC   | CCTTAACGTG  | AGTTTTCGTT |
| 3011      | CCACTGAGCG | TCAGACCCCG   | TAGAAAAGAT  | CAAAGGATCT  | TCTTGAGATC   | CTTTTTCCT   | GCGCGTAATC |
| 3081      | TGCTGCTTGC | AAACA AAAA   | ACCACCGCTA  | CCAGCGGTGG  | TTTGTTTGCC   | GGATCAAGAG  | CTACCAACTC |
| 3151      | TTTTTCCGAA | GGTAAC TGGC  | TTCAGCAGAG  | CGCAGATACC  | AAATACTGTC   | CTTCTAGTGT  | AGCCGTAGTT |
| 3221      | AGGCCACCAC | TTCAAGAACT   | CTGTAGCAC   | GCCTACATAC  | CTCGCTCTGC   | TAA TCCTGTT | ACCAGTGGCT |
| 3291      | GCTGCCAGTG | GCGATAAGTC   | GTGTCTTACC  | GGTTGGACT   | CAAGACGATA   | GTTACC GGAT | AAGCGCAGC  |
| 3361      | GGTCGGGCTG | AACGGGGGT    | TCGTGCACAC  | AGCCAGCTT   | GGAGCGAACG   | ACCTACACCG  | AACTGAGATA |
| 3431      | CCTACAGCGT | GAGCTATGAG   | AAAGGCCAC   | GCTTCCGAA   | GGGAGAAAGG   | CGGACAGGTA  | TCCGGTAAGC |
| 3501      | GGCAGGGTCG | GAACAGGAGA   | GCGCACGAGG  | GAGCTTCCAG  | GGGAAACGC    | CTGGTATCTT  | TATAGTCCTG |
| 3571      | TCGGGTTTCG | CCACCTCTGA   | CTTGAGCGTC  | GATTTTGTG   | ATGCTCGTCA   | GGGGGGCGGA  | GCCTATGGAA |
| 3641      | AAACGCCAGC | AACGCGGCCT   | TTTTACGGTT  | CTTGGCCTT   | TGCTGGCCTT   | TTGCTCACAT  | GTTCTTTCCT |
| 3711      | GCGTTATCCC | CTGATTCTGT   | GGATAACCGT  | ATTACCGCT   | TTGAGTGAGC   | TGATACCGCT  | CGCCGCAGCC |
| 3781      | GAACGACCGA | GCGCAGCGAG   | TCAGTGAGCG  | AGGAAGCGGA  | AGAGCGCCCA   | ATACGCAAC   | CGCTCTCCC  |
| 3851      | CGCGCGTTGG | CCGATTCAAT   | AATGCAGCTG  | GCACGACAGG  | TTTCCC GACT  | GGAAAGCGGG  | CAGTGAGCGC |
| 3921      | AACGCAATTA | ATGTGAGTTA   | GCTCACTCAT  | TAGGCACCCC  | AGGCTTTACA   | CTTTATGCTT  | CCGGCTCGTA |
| 3991      | TGTTGTGTGG | AATTGTGAGC   | GGATAACAAT  | TTCACACACGG | AAACAGCTAT   | GACCATGATT  | ACGCCAAGCG |
| KpnI XhoI |            |              |             |             |              |             |            |
| 4061      | CGCAATTAA  | CCTCAC TAA   | GGGAACAAA   | GCTGGGTACC  | GGGCCCCCCC   | TCGAGGTCA   | TCATATGCTT |
| 4131      | GAGAAAGAG  | TCGGGATAGT   | CCAAAATAAA  | ACAAAGGTAA  | GATTACCTGG   | TCAAAAGTGA  | AAACATCAGT |
| 4201      | TAAAAGGTGG | TATAAGTAAA   | ATATCGGTAA  | TAAAAGGTGG  | CCCAAAGTGA   | AATTACTCT   | TTTCTACTAT |
| 4271      | TATAAAAATT | GAGGATGTTT   | TGTCGGTACT  | TTGATACGTC  | ATTTTGTGAT   | GAA TTGGTTT | TTAAGTTTAT |
| 4341      | TCGCGATTTG | GAAATGCATA   | TC TGTA TTG | AGTCGGTTT   | TAA GTTTCGTT | GCTTTTGTA   | ATACAGAGGG |
| 4411      | ATTTGTATA  | GAAATATCTT   | TAAAAAACCC  | ATATGCTAAT  | TTGACATAAT   | TTTTTGAGAAA | AAATATATAT |
| EcoRI     |            |              |             |             |              |             |            |
| 4481      | CAGGCGAATT | CCACAATGAA   | CAATAATAAG  | ATTAAAATAG  | CTTGCCCCCG   | TTGCAGCGAT  | GGGTATTTT  |
| 4551      | TCTAGTAAAA | TAAAAGATA    | ACTTAGACTC  | AAAACATTTA  | CAAAAACAC    | CCCTAAAGTC  | CTAAAGCCCA |

FIG.-34D

**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
TACAACCTCGA
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
ATGCCTTCCA
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 GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATAATGCGCA CACGACTTAT TTTGGGATGA CGAGCGGAGC  
. C T W \*  
911 CTGTACATGG TGATCAGTCA TTTCAGCCTC CCCGAGTGTA CCAGGAAAGA TGGATGTCCCT GGAGAGGGGG  
981 CCGCGTAACC ACTGAAGGAT GAGCTGTAAA GAAGCAGATC GTTCAAACAT TTGGCAATAA AGTTTCTTAA  
1051 GATTGAATCC TGTGCGCGGT CTTGCGATGA TTATCATATA ATTTCTGTTG AATTACGTTA AGCATGTAAT  
1121 AATTAAACATG TAAATGCATGA CGTTATTAT GAGATGGGT TTTATGATTA GAGTCCCGCA ATTATACATT  
1191 TAATACGCGA TAGAAAACAA AATATAGCGC GCAAAC TAGG ATAAATTATC GCGCGCGGTG TCATCTATGT

HindIII

Clal

XbaI

~~~~~

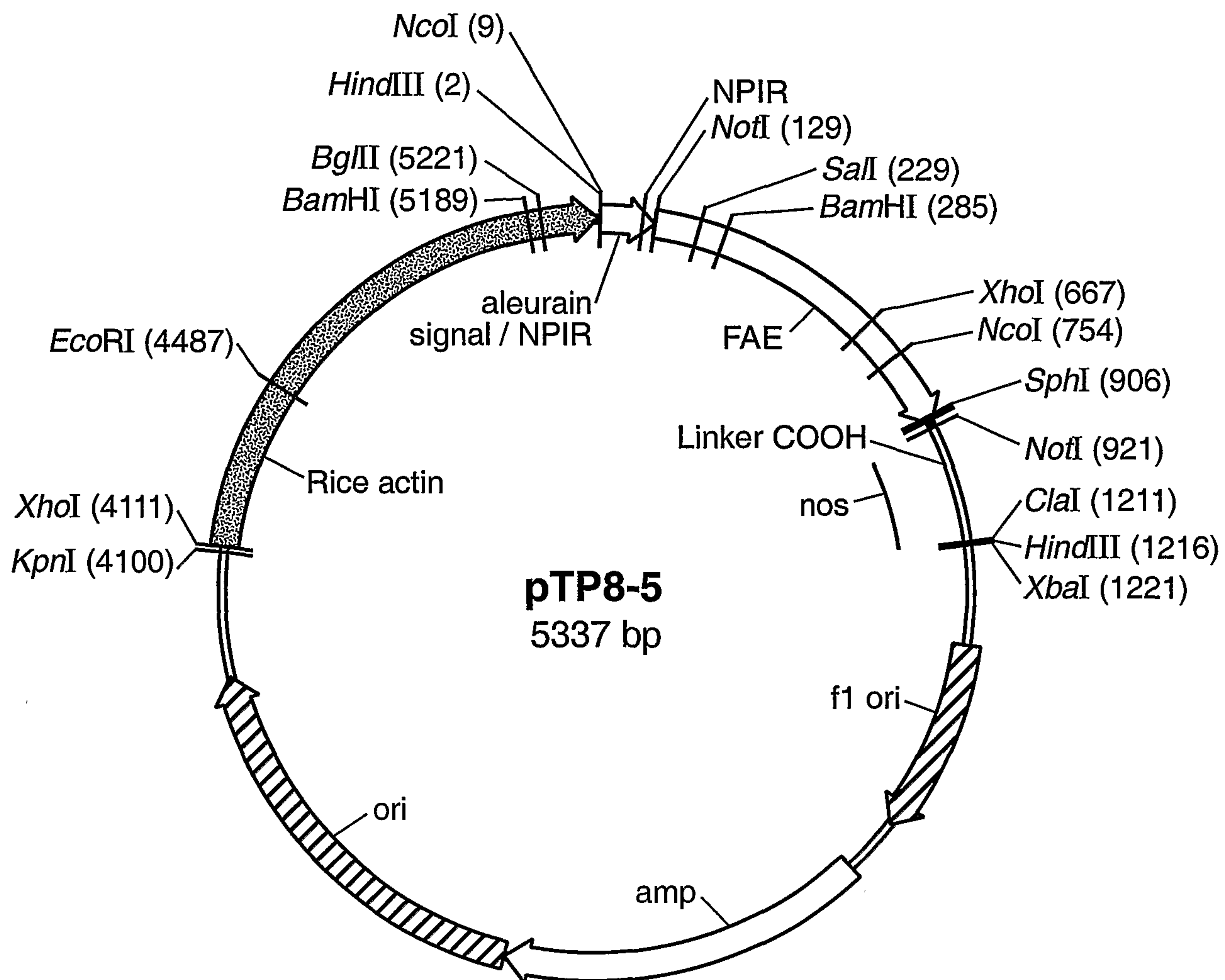
1261 TACTAGATCG ATAAGCTTCT AGAGCGGCCG GTGGAGCTCC AATTCGCCCT ATAGTGAGTC GTATTACGCG
1331 CGCTCACTGG CCGTCGTTT ACAACGTCGT GACTGGGAAA ACCCTGGCGT TACCCAACTT AATCGCCTTG
1401 CAGCACATCC CCTTTTCGCC AGCTGGCGTA ATAGCGAAGA GGCCCGCACG GCATTAAGCG CGCGGGGTG GTGGTTACG
1471 GCGCAGCCTG AATGGCGAAT GGGACGCGC CTGTAGCGCG CTCCTTTCGC TTTCTTCCCT TCTTTCCTCG
1541 CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG TAAATCGGG GCTCCCTTA GGGTTCGGAT TTAGTGCTTT
1611 CCACGTTTCG CCGCTTTCCC CGTCAAGCTC TAAATCGGG TCACGTAGTG GGCCATCGCC CTGATAGACG
1681 ACGGCACCTC GACCCCAAAA AACTTGATTA GGGTGATGGT TCACGTAGTG GGCCATCGCC CTGATAGACG
1751 GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTCTTTAATA GTGGACTCTT GTTCCAAACT GGAACAACAC
1821 TCAACCCCTAT CTCGGTCTAT TCTTTTGATT TATAAGGGAT TTTGCCGATT TCGGCCATT TCGTAAAAAA
1891 TGAGCTGATT TAACAAAAAT TTAACGCGAA TTTTAACAAA ATATTAAACG ATATTAA TTAACAATTG GTGGCATT
1961 TTCGGGGAAA TGTGCGCGGA ACCCTATT TTTTATT TTTTATACAT CTAAATATGT ATCCGCTCAT
2031 GAGACAATAA CCTGATAAA TGCTTCAATA ATATTGAAA AGGAAGAGTA TGAGTATTCA ACATTTCCTG
2101 GTCGCCCTTA TTCCCTTTT TCGGCAATT TGCCTTCC TGCTTCCCTG TTTTTCCTCA CCCAGAAACG CTGGTGAAAG
2171 TAAAAGATGC TGAAGATCAG TTGGGTGCAC GAGTGGGTTA CATCGAACTG GATCTCAACA GCGTAAGAT
2241 CCTTGAGAGT TTTCGCCCCG AAGAACGTTT TCCAATGATG AGCACTTTTA AAGTTCCTGCT ATGTGGCGCG
2311 GTATTATCCC GTATTGACGC CGGGCAAGAG CAACTCGGTC GCCGCATACA CTATTCTCAG AATGACTTGG
2381 TTGAGTACTC ACCAGTCACA GAAAAGCATC TTACGGATGG CATGACAGTA AGAGAAATTAT GCAGTGTGC
2451 CATAACCATG AGTGATAACA CTGCGGCCAA CTTACTTCTG ACAACGATCG GAGGACCGAA GGAGCTAACC
2521 GCTTTTTCG ACAACATGGG GGATCATGTA ACTCGCCTTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA
2591 TACCAAACGA CGAGCGTGAC ACCACGATGC 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 | AAGTTACTC | ATATATACTT | TAGATTGATT | TAAAACCTCA | TTTTTAAATT | AAAAGGATCT |
| 3011 | AGGTGAAGAT | CCTTTTGGAT | AATCTCATGA | CCAAAATCCC | TTAAACGTGAG | TTTTTCGTTCC | ACTGAGCGTC |
| 3081 | AGACCCCGTA | GAAAAGATCA | AAGGATCTTC | TTGAGATCCT | TTTTTTCTGC | GCGTAAATCTG | CTGCTTGCAA |
| 3151 | ACAAAAAAAC | CACCGCTACC | AGCGGTGGTT | TGTTTGCCGG | ATCAAGAGCT | ACCAACTCTT | TTTCCGAAGG |
| 3221 | TAACTGGCTT | CAGCAGAGCG | CAGATACCAA | ATACTGTCCT | TCTAGTGTAG | CCGTAGTTAG | GCCACCACCT |
| 3291 | CAAGAACTCT | GTAGCACCGC | CTACATACCT | CGTCTGCTA | ATCCTGTTAC | CAGTGGCTGC | TGCCAGTGGC |
| 3361 | GATAAGTCGT | GTCTTACCGG | GTTGGACTCA | AGACGATAGT | TACCGGATAA | GGCGCAGCGG | TCGGGCTGAA |
| 3431 | CGGGGGGTTT | GTGCACACAG | CCAGCTTGG | AGCGAACGAC | CTACACCGAA | CTGAGATACC | TACAGCGTGA |
| 3501 | GCTATGAGAA | AGGCCACGC | TTCCCGAAAG | GAGAAAGCGG | 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 | TCTTTCTCTGC | 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 | AAGTTATATC | GCGATTTGGA |
| 4411 | AATGCATATC | TGTATTTGAG | TCGGTTTTTA | AGTTCCGTTGC | 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 | GCAGCCAAA | AAAAAAAAG |
| 4831 | AAAGAAAAAA | AAGAAAAAGA | AAAACAGCAG | GTGGGTCCGG | GTCGTGGGG | CCGGAAAAAGC | GAGGAGGATC |

FIG. 35D

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 ACATTCCTGTC 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 TCTACAACATC GATACTAACT 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 TTGCGAT | GATTATCATA  | TAAATTTCTGT | TGAAATTACGT |
| 1051 | TAAGCATGTA  | ATAATTAAACA | TGTAATGCAAT | GACGTTATTT  | ATGAGATGGG  | TTTTTTATGAT | TAGAGTCCCG  |
| 1121 | CAATTATACA  | TTTAATACGC  | GATAGAAAAC  | AAAATATAGC  | GCGCAAAC TA | GGATAAA TTA | 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  | GCGCATT AAG | CGCGGCGGGT  |
| 1471 | GTGGTGGTTA  | CGCGCAGCGT  | GACCGCTACA  | CTTGCCAGCG  | CCCTAGCGCC  | CGCTCCCTTT  | GCTTTC TTCC |
| 1541 | CTTCC TTCT  | CGCCACG TTC | GCCGGCTTTC  | CCCGTCAAGC  | TCTAAATCGG  | GGGCTCCCTT  | TAGGGTTCCG  |
| 1611 | ATTTAGTGCT  | TTACGGCAC   | TCGACCCCAA  | AAACTTGAT   | TAGGGTGATG  | GTTCAACGTAG | TGGGCCATCG  |
| 1681 | CCCTGATAGA  | CGGTTTTCG   | CCCTTTGACG  | TTGGAGTCCA  | CGTTC TTAA  | TAGTGGACTC  | TTGTTCCAAA  |
| 1751 | CTGGAACAAC  | ACTCAACCTT  | ATCTCGGTCT  | ATTC TT TGA | TTTATAAGGG  | ATTTTGCCGA  | TTTCGGCCTA  |
| 1821 | TTGGTTAAAA  | AATGAGCTGA  | TTTAACA AAA | ATTTAACGCG  | AAATTTAACA  | AAATATT AAC | GCTTACAATT  |
| 1891 | TAGGTGGCAC  | TTTTCGGGGA  | AATGTGCGCG  | GAACCCCTAT  | TTGTTTATTT  | TTCTAAATAC  | ATTCAAATAT  |
| 1961 | GTATCCGCTC  | ATGAGACAAAT | AACCTGATA   | AATGCTTCAA  | TAATATTGAA  | 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  | TGAGCACTTT  | TAAAGTTCTG  |
| 2241 | CTATGTGGCG  | CGGTATTATC  | CCGTATTGAC  | GCCGGGCAAG  | AGCAACTCGG  | TCGCCGCATA  | CAC TATTCTC |
| 2311 | AGAAATGACTT | GGTTGAGTAC  | TCACCAGTCA  | CAGAAAAGCA  | TCTTACGGAT  | GGCATGACAG  | TAAGAGAA TT |

FIG.- 36C

|       |             |            |            |              |             |             |             |
|-------|-------------|------------|------------|--------------|-------------|-------------|-------------|
| 2381  | ATGCAGTGCT  | GCCATAACCA | TGAGTGATAA | CACTGCGGCC   | AACTTACTTC  | TGACAACGAT  | CGGAGGACCG  |
| 2451  | AAGGAGCTAA  | CCGCTTTT   | GCACAACATG | GGGATCATG    | TAAC        | TGATCGTTGG  | GAACCGGAGC  |
| 2521  | TGAATGAAGC  | CATACCAAC  | GACGAGCGTG | ACACACGAT    | GCC         | ATGGCAACAA  | CGTTGCGCAA  |
| 2591  | ACTATTAACT  | GGCGAACTAC | TTACTCTAGC | TTCCCGGCAA   | CAATTAA     | ACTGGATGGA  | GGCGGATAAA  |
| 2661  | GTTGCAGGAC  | CAC        | CTCGGCCCTT | CCGGCTGGCT   | GGTTTATTGC  | TGATAAACTCT | GGAGCCGGTG  |
| 2731  | AGCGTGGGTC  | TCGCGGTATC | ATTGCAGCAC | TGGGGCCAGA   | TGGTAAGCCC  | TCCCGTATCG  | TAGTTATCTA  |
| 2801  | CACGACGGGG  | AGTCAGGCAA | CTATGGATGA | ACGAAATAGA   | CAGATCGCTG  | AGATAAGGTGC | CTC         |
| 2871  | AAGCATTGGT  | AAC        | CTAGTTTAC  | TCATAATATAC  | TTTAGATTGA  | TTTAAAACTT  | CATTTT      |
| 2941  | TTAAAAGGAT  | CTAGGTGAAG | ATCCTTTT   | ATAATCTCAT   | GACCAAAATC  | CTTAACGTG   | AGTTTTCGTT  |
| 3011  | CCACTGAGCG  | TCAGACCCCG | TAGAAAAGAT | CAAAGGATCT   | TCTTGAGATC  | CTTTT       | GCGGTAATC   |
| 3081  | TGCTGCTTGC  | AAACAAAAA  | ACCACCGCTA | CCAGCGGTGG   | TTTGTTTGCC  | GGATCAAGAG  | CTACCAACTC  |
| 3151  | TTT         | GGTAAC     | TTCAGCAGAG | CGCAGATACC   | AAATAC      | CTTCTAGTGT  | AGCCGTAGTT  |
| 3221  | AGGCCACCAC  | TTCAAGAACT | CTGTAGCAC  | GCCTACATAC   | CTCGCTCTGC  | TAATCCTGTT  | ACCAGTGGCT  |
| 3291  | GCTGCCAGTG  | GC         | GTGTCTTACC | GGT          | CAAGACGATA  | GTTACCGGAT  | AAGGCGCAGC  |
| 3361  | GGTCGGGCTG  | AACGGGGGT  | TCGTGCACAC | AGCCAGCTT    | GGAGCGAACG  | ACCTACACCG  | AAC         |
| 3431  | CCTACAGCGT  | GAGCTATGAG | AAAGCGCCAC | GCTTCCCGAA   | GGAGAAAGG   | CGGACAGGTA  | TCCGGTAAGC  |
| 3501  | GGCAGGGTCG  | GAACAGGAGA | CGCACGAGG  | GAGCTTCCAG   | GGGAAACGC   | CTGGTATCTT  | TATAGTCCTG  |
| 3571  | TCGGGTTCG   | CCACCTCTGA | CTTGAGCGTC | GATTTT       | ATGCTCGTCA  | GGGGCGCGA   | GCCTATGGAA  |
| 3641  | AAACGCCAGC  | AACGGGCCCT | TTTTACGGTT | CCTGGCCCTT   | TGCTGGCCCTT | TTGCTCACAT  | GTTCTTTCCCT |
| 3711  | GCGTTATCCC  | CTGAT      | GGATAACCGT | ATTACCGCCT   | TTGAGTGAGC  | TGATACCGCT  | CGCCGCAGCC  |
| 3781  | GAACGACCGA  | GC         | TCAGTGAGCG | AGGAAGCGGA   | AGAGCGCCA   | ATACGCAAAC  | CGCCTCTCCC  |
| 3851  | CGCGCGTTGG  | CCGATT     | AATGCAGCTG | GCACGACAGG   | TTTCCCGACT  | GGAAGCGGG   | CAGTGAGCGC  |
| 3921  | AACGCAATTA  | ATGTGAGTTA | GCTCACTCAT | TAGGCACCCC   | AGGCTTTACA  | CTTTATGCTT  | CCGGCTCGTA  |
| 3991  | TGTTGTGTGG  | AATTGTGAGC | GGATAACAAT | TTCACACACAGG | AAACAGCTAT  | GACCATGATT  | ACGCCAAGCG  |
| ~~~~~ |             |            |            |              |             |             |             |
| 4061  | CGCAATTAA   | CCTCACTAAA | GGGAACAAA  | GCTGGGTACC   | GGGCCCCC    | TCGAGGTCAT  | TCATATGCTT  |
| 4131  | GAGAAAGAGAG | TCGGGATAGT | CCAAAATAAA | ACAAAGGTAA   | GATTACCTGG  | TCAAAAGTGA  | AAACATCAGT  |
| 4201  | TAAAAGGTGG  | TATAAGTAAA | ATATCGGTAA | TAAAAGGTGG   | CCCAAAGTGA  | AATTACTCT   | TTTCTACTAT  |
| 4271  | TATAAAAATT  | GAGGATGTTT | TGTCGGTACT | TTGATACGTC   | ATTTTGTAT   | GAATTGGTTT  | TTAAGTTTAT  |
| 4341  | TCGCGAATTG  | GAAATGCATA | TCTGTAATTG | AGTCGGTTT    | TAAGTTCGTT  | GC          | ATACAGAGGG  |
| 4411  | ATTGTATAA   | GAAATATCTT | TAAAAAACCC | ATATGCTAAT   | TTGACATAAT  | TTTTTGAGAAA | AATATATATT  |
| ~~~~~ |             |            |            |              |             |             |             |
| 4481  | CAGGCGAATT  | CCACAA     | CAATAATAAG | ATTAAAAATAG  | CTTGCCCCCG  | TTGCAGCGAT  | GGGTATTTT   |
| ~~~~~ |             |            |            |              |             |             |             |

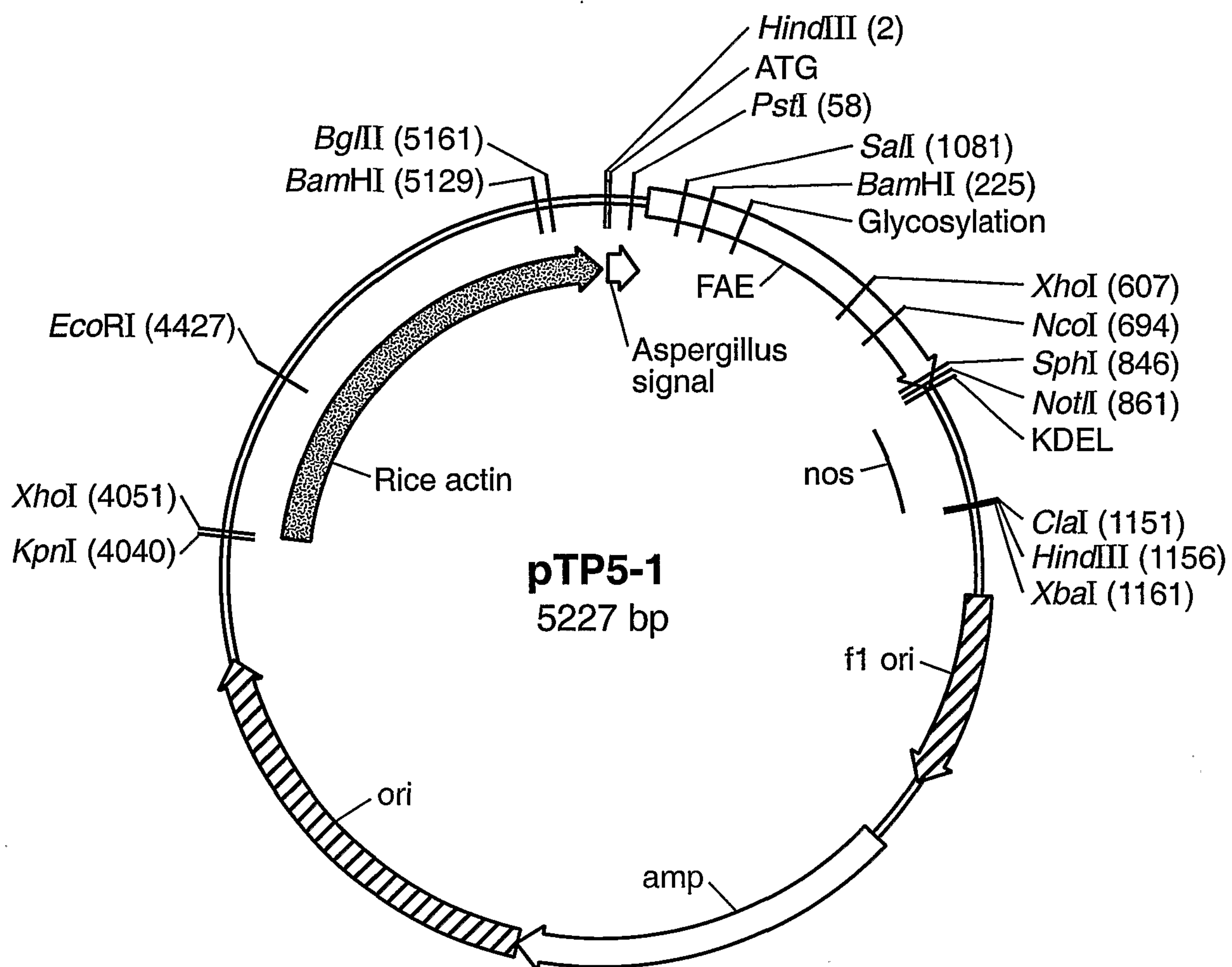
~~~~~  
KpnI
~~~~~  
~~~~~  
XhoI
~~~~~

ECORI

FIG.\_36D

|      |            |             |             |            |            |            |            |
|------|------------|-------------|-------------|------------|------------|------------|------------|
| 4551 | TCTAGTAAA  | TAAAGATAA   | ACTTAGACTC  | AAACATTTA  | CAAAACAAC  | CCCTAAAGTC | CTAAAGCCCA |
| 4621 | AAGTGCTATG | CACGATCCAT  | AGCAAGCCCA  | GCCCAACCCA | ACCCAACCCA | ACCCACCCCA | GTGCAGCCAA |
| 4691 | CTGGCAATA  | GTCCTCACCC  | CCGGCACTAT  | CACCGTGAGT | TGTCCGCACC | ACCGCACGTC | TCGCAGCCAA |
| 4761 | AAAAA      | AGAAAGAAA   | AAAGAAAAA   | GAAACACAGC | AGGTGGGTCC | GGTCGTGGG  | GGCCGGAAA  |
| 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 | GGGTGGGCGA |
| 5111 | GAGCGGCTTC | GTCGCCCAGA  | TCGGTGCGCG  | GGAGGGGCGG | GATCTCGCGG | CTGGCGTCTC | CGGGCGTGAG |
|      |            | BamHI       |             | BglII      |            |            |            |
|      | ~~~~~      |             |             | ~~~~~      |            |            |            |
| 5181 | TCGGCCCCGA | 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
TGAAGGCATC

CTCCGCCAAA
TCCGAAGACC

CACGTCCTCG
TCTACAGCCG

CAGTTGTGGT
TTTAGTCGAA

GACTGCAGGG
ATGGCCACTA

CACGCCCTTAG
TCTCCCAAGC

• H A L A
• M A T I S Q A •

PstI
~~~~~

• A Y A  
141

D L C N  
GACCTGTGCA

I P S  
ACATTCCGTC

T I I  
GACTATTATC

K G E K  
AAGGAGAGA

I Y N  
AAATTACAA

S Q T  
TTCTCAAAC

BamHI  
~~~~~

D I N G
211

W I L
GATGGATCCT

R D D
CCGCGACGAC

S S K E
AGCAGCAAAG

I I T
AAATAATCAC

V F R
CGTCTTCCGT

G T G S
GGCACTGGTA

• D T N
281

L Q L
D T N Y

T L T
P F D

T L P Q
T L P Q

C N G •
GTGATACGAA

TCTACAACCTC
GATACTAACT

ACACCCCTCAC
GCCTTTCGAC

ACCTACCAC
ACCTACCAC

AATGCAACGG
• C E V

H G G Y
Y I G

W V S
V Q D Q

V E S
L V K

TTGTGAAGTA
351

CACGGTGGAT
ATTATATTGG

ATGGGTCTCC
GTCCAGGACC

AAGTCGAGTC
GCTTGTCAAA

Q Q V S
421

Q Y P
D Y A

L T V T
G H X

L G A
S L A

A
CAGCAGGTTA

GCCAGTATCC
GGACTACGG

CTGACCCTGA
CCGGCCACKC

CCTCGGCGCC
TCCCTGGCGG

• L T A
491

A A Q L
S A T Y

D N I
R L Y

T F G E
P R S •

CACTCACTGC

CGCCAGCTG
TCTGCGACAT

ACGACAACAT
CCGCTGTAC

ACCTTCGGCG
AACC

CGCGCAG

XhoI
~~~~~

• G N Q  
561

A F A S  
Y M N

D A F  
Q A S

S S  
P D T

T Q Y  
CGGCAATCAG

GCCTTCGCGT  
CGTACATGAA

CGATGCCTTC  
CAAGCCTCGA

GCCCAGATAC  
GACGCAGTAT

• F R V  
631

T H A N  
D G I

P N L P  
P V E

Q G Y  
A H G

G  
TTCCGGGTCA

CTCATGCCAA  
CGACGGCATC

CCAAACCTGC  
CCCCGGTGGA

GCAGGGGTAC  
GCCCATGGCG

• V E Y  
701

W S V  
D P Y

S A Q N  
T F V

C T G D  
E V Q •

GTGTAGAGTA

CTGGAGCGTT  
GATCCTTACA

GCGCCAGAA  
CACATTTGTC

TGCAC

ATGAAGTGCA

• C C E  
771

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

NcoI  
~~~~~

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 | GGAAAGCGGG |
| 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 | ACAAAAGTAA | GATTACCTGG | TCAAAAGTGA |
| 4131 | AAACATCAGT | TAAAAGGTGG | TATAAGTAAA | ATATCGGTAA | TAAAAGGTGG | CCCAAAGTGA | AAATTACTCT |
| 4201 | TTTCTACTAT | TATAAAAATT | GAGGATGTTT | TGTCGGTACT | TTGATACGTC | ATTTTGTGAT | GAAATTGGTTT |
| 4271 | TTAAGTTTAT | TCGCGATTTG | GAAATGCATA | TCGTATTTG | AGTCGGTTTT | TAAAGTTCGTT | GCTTTTGTA |
| 4341 | ATACAGAGGG | ATTTGTATAA | GAAATATCTT | TAAAAAACCC | ATATGCTAAT | TTGACATAAT | TTTTTGAGAAA |
| | | | | 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 | CTGGCAATA | GTCCTCACCC | CCGGCACTAT | CACCGTGAGT | TGTCGCGACC | ACCGCACGTC |
| 4691 | TCGCAGCCAA | AAAAAATAAA | AGAAAGAAAA | AAAAGAAAAA | GAAAAACAGC | AGGTGGGTCC | GGGTGCTGGG |

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 | GGAGGGCGCG | GATCTCGCGG | CTGGCGTCTC |
| | | BamHI | | | | BglII | |
| | | ~~~~~ | | | | ~~~~~ | |
| 5111 | CGGGCGTGAG | TCGGCCCGGA | TCCTCGCGGG | GAATGGGGCT | CTCGGATGTA | GATCTTCTTT | CTTTCCTTCTT |
| 5181 | TTTGTTGGTAG | AATTGGAATC | CCTCAGCATT | GTTCATCGGT | AGTTTTTCTT | TTCATGATTT | GTGACAAATG |
| 5251 | CAGCCTCGTG | CGGAGCTTTT | TTGTAGC | | | | |

FIG._37E

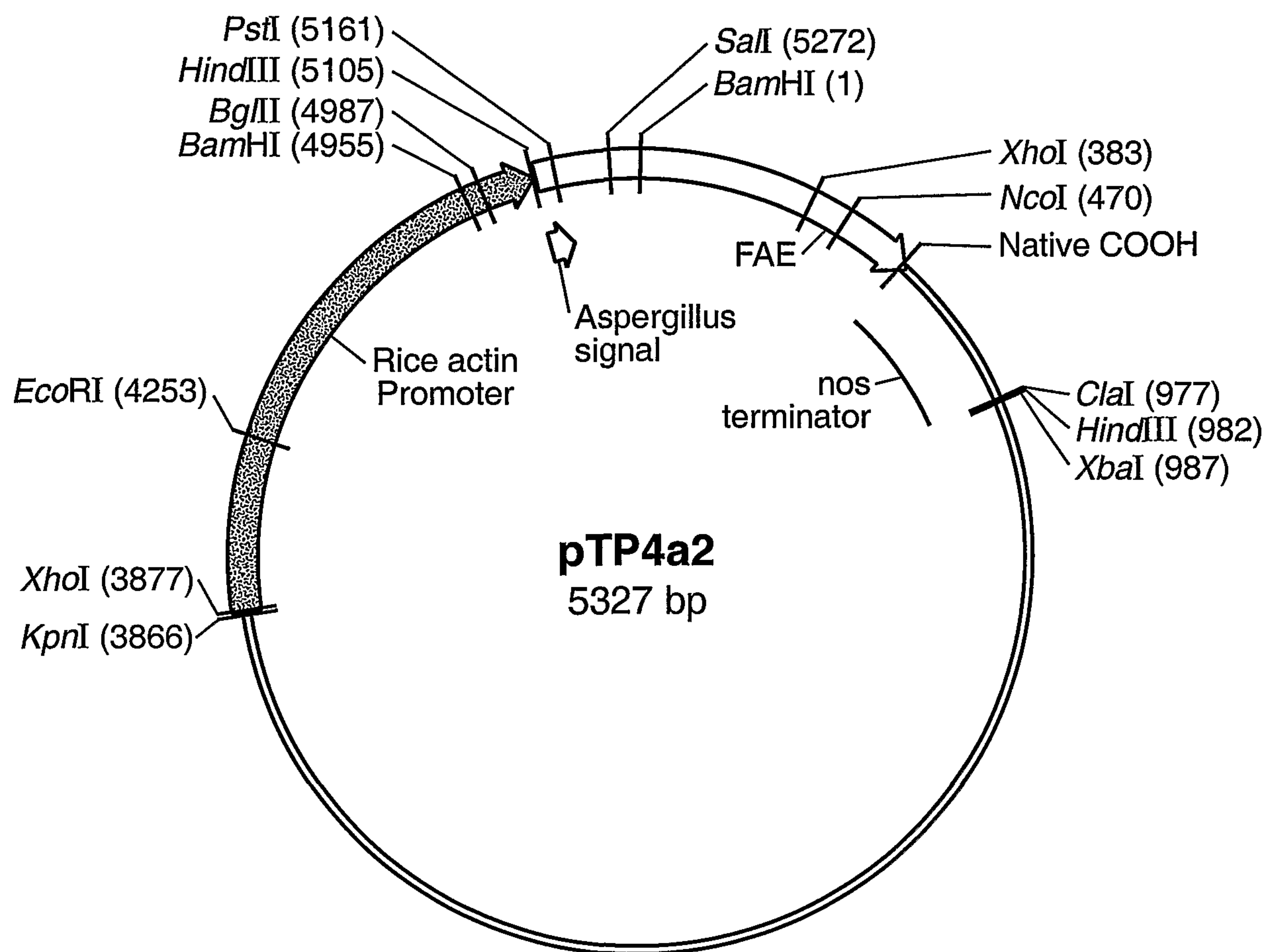
**FIG. 38A**

FIG. 38B

981 AAGCTTCTAG AGCGGCCGGT GGAGCTCCAA TTCGCCCTAT AGTGAGTCGT ATTACGCGCG CTCACTGGCC
1051 GTCGTTTAC AACGTCGTGA CTGGGAAAC CCTGGCGTTA CCAACTTAA CCGCTTGCA GCACATCCCC
1121 CTTTCGCCAG CTGGCGTAAT AGCGAGAGG CCGCACCGA TCGCCCTTCC TGCAGTTGC CAGCCTGAA
1191 TGGCGAATGG GACGCGCCCT TAGCGGCGC ATTAAGCGC GTGGGTGTGG GCGGTACGG TGGGTGACC
1261 GCTACACTTG CCAGCGCCCT AGCGCCGCT CTTTCGCTT CTTCCCTTC TCTCTCGCC ACGTTCGCC
1331 GCTTCCCCG TCAAGCTCTA AATCGGGGC TCCCTTTAGG GTTCCGATTT AGTGCTTTAC GGCACCTCGA
1401 CCCCAAAAA CTTGATTAGG GTGATGGTTC ACGTAGTGG CCAATCGCCCT CCATCGCCCT TTTTCGCCCT
1471 TTGACGTTGG AGTCCACGTT CTTTAATAGT GGACTCTTGT GGACTCTTGT TCCAAACTGG AACACACTC AACCTATCT
1541 CGGTCATTC TTTTGATTTA TAAGGGAATT TAAGGATTC TGCCGATTC GGCCTATTGG TTAAAAAATG AGCTGATTTA
1611 ACAAAAATTT AACGCGAATT TTAACAAAAT ATTAACGCTT ACAATTTAGG TGGCACTTTT CGGGGAAATG
1681 TCGCGGAAC CCTATTGT TTATTTTCT AAATACATTC AAATATGTAT AAATATGTAT CCGCTCATGA GACAATAACC
1751 CTGATAAATG CTTCAATAAT ATTGAAAAAG GAAGAGTATG AGTATTCAAC AGTATCCGT ATTTCCCTATT
1821 CCTTTTTTG CGGCATTTTG CCTTCCCTGT CTTGCTCAC TTTGCTCAC ATTCTCAGAA ATTTCTCCGT ATTATCCCT
1891 AAGATCAGTT GGTGCACGA GTGGTTACA TCGAACTGGA CACTTTTAA CACTTTTAA ATTTCTCAGAA TGAATTTAGG
1961 TCGCCCCGAA GAACGTTTTC CAATGATGAG CACTTTTAA CACTTTTAA ATTTCTCAGAA TGAATTTAGG
2031 ATTGACGCCG GGCAAGAGCA ACTCGGTCGC CGCATACACT TGACAGTAAG AGAATTATGC GGACCGAAG GGACTTGGT
2101 CAGTCACAGA AAAGCATCTT ACGGATGGCA TGACAGTAAG AGAATTATGC GGACCGAAG GGACTTGGT
2171 TGATAACACT GCGCCAAC TACTTCTGAC AACGATCGGA AACGATCGGA GGACCGAAG GGACTTGGT
2241 AACATGGGG ATCATGTAA TCGCCCTTGAT TAGCAATGG CAACAACGTT GCGCAAACTA TTAACCTGGC
2311 AGCGTGACAC CACGATGCC TTAGCAATGG CAACAACGTT GCGCAAACTA TTAACCTGGC
2381 TCTAGCTTCC CCGCAACAAT TAATAGACTG TAATAGACTG GATGAAGTTG GATAAAGTTG CAGGACCACT
2451 GCCCTTCCGG CTGGCTGGTT TATTGCTGAT TATTGCTGAT AAATCTGGAG CCGGTGAGCG TGGGTCTCG
2521 CAGCACTGGG GCCAGATGGT AAGCCCTCCC GTATCGTAGT TATCTACACG ACGGGGAGTC ACGGGGAGTC
2591 GGATGAACGA AATAGACAGA TCGCTGAGAT AGGTGCCCTCA CTGATTAGC CTGATTAGC ATTGGTAAC
2661 GTTTACTCAT ATATACTTTA GATTGATTTA AAACCTTCA TTTTAATTAA AAGGATCTAG GTGAAGATCC
2731 TTTTGTGATA TCTCATGACC AAAATCCCTT AACGTGAGTT TTCGTTCCAC TGAGCGTCAG ACCCGTAGA
2801 AAAGATCAAA GGATCTTCTT GAGATCCCTT TTTTCTGCGC GTAATCTGCT GTAACTAAC GCTTGCAAAC
2871 CCGCTACCAG CCGTGGTTTG TTTGCCGGAT CAAGAGCTAC CAACCTTTT TCCGAAGGTA ACTGGCTTCA
2941 GCAGAGCGCA GATACCAAAT ACTGTCCCTC TAGTGTAGG GTAGTTAGG CACCACTTCA AGAATCTGT
3011 AGCACCGCCT ACATACCTCG CTCTGCTAAT CCGGTGCTG GTGGCTGCTG CCAGTGGCGA TAAGTCGTGT
3081 CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAG CGCAGCGGT CAGCTGAAC GGGGTTCGT
3151 GCACACAGCC CAGCTTGGAG CGAACGACCT ACACCGA ACT GAGATACCTA CAGCTGAGC TATGAGAAAG
3221 CGCCACGCTT CCCGAAGGGA GAAAGCGGA CAGGTATCCG GTAAGCGGCA GGTTCGGAAC AGGAGAGCGC
3291 ACGAGGGAGC TTCCAGGGG AAACGCCTGG TATCTTTATA GTCCCTGTGG GTTTCGCCAC CTCGTACTTG
3361 AGCGTCGATT TTTGTGATGC TCCTCAGGG GCGGAGCCT ATGGAAAAAC GCCAGCAAC CGGCTTTT

FIG.-38C

| | | | | | | | |
|------|------------|------------|------------|-------------|-------------|-------------|-------------|
| 3431 | ACGGTTCCTG | GCCTTTTGCT | GGCCTTTTGC | TCACATGTTT | TTTCCTGCGT | TATCCCCCTGA | TTCTGTGGAT |
| 3501 | AACCGTATTA | CCGCCTTTGA | GTGAGCTGAT | ACCGCTCGCC | GCAGCCGAAC | GACCGAGCGC | AGCGAGTCAG |
| 3571 | TGAGCGAGGA | AGCGGAAGAG | CGCCCAATAC | GCAAACCGCC | TCCTCCCGCG | CGTTGGCCGA | TTCAATTAATG |
| 3641 | CAGCTGGCAC | GACAGGTTTC | CCGACTGGAA | AGCGGGCAGT | GAGCGCAACG | CAATTAATGT | GAGTTAGCTC |
| 3711 | ACTCATTAGG | CACCCAGGC | TTTACACTTT | ATGCTTCCGG | CTCGTATGTT | GTGTGGAATT | GTGAGCGGAT |
| 3781 | AACAAATTCA | CACAGGAAC | AGCTATGACC | ATGATTACGC | CAAGCGCGCA | ATTAACCCCTC | ACTAAAGGGA |
| | | KpnI | ~~~~~ | XhoI | ~~~~~ | | |
| 3851 | ACAAAAGCTG | GGTACCGGC | CCCCCTCGA | GGTCATTCT | ATGCTTGAGA | AGAGAGTCGG | GATAGTCCAA |
| 3921 | AATAAAACAA | AGGTAAGATT | ACCTGGTCAA | AAGTGAAAC | ATCAGTTAAA | AGGTGGTATA | AGTAAATAT |
| 3991 | CGGTAATAAA | AGGTGGCCCA | AAGTGAAATT | TACTCTTTTC | TACTATTATA | AAAATTGAGG | ATGTTTGTG |
| 4061 | GGTACTTTGA | TACGTCAATT | TTGTATGAAT | TGGTTTTAA | GTTTATTCGC | GATTGGAAA | TGCATATCTG |
| 4131 | TATTTGAGTC | GGTTTTTAAG | TTCGTTGCTT | TTGTAAATAC | AGAGGGAATT | GTATAAGAAA | TATCTTTTAA |
| | | | | | | ECORI | |
| | | | | | | ~~~~~ | |
| 4201 | AAACCCATAT | GCTAATTTGA | CATAATTTT | GAGAAAATA | TATAATTCAGG | CGAATTCAC | AATGAACAAT |
| 4271 | AATAAGATTA | AAATAGCTTG | CCCCCGTTG | AGCGATGGGT | ATTTTCTCTA | GTAAATAAAA | AGATAAACTT |
| 4341 | AGACTCAAAA | CATTTACAAA | AACAACCCCT | AAAGTCCTA | AGCCCAAGT | GCTATGCACG | ATCCATAGCA |
| 4411 | AGCCAGCC | AACCAACC | AACCAACC | ACCCAGTGC | AGCCAACTGG | CAAAATAGTCT | CCACCCCGG |
| 4481 | CACTATCAC | GTAGTTGTC | CGACCAACG | CACGTCTCG | AGCCAAAAA | AAAAAAGAA | AGAAAAAAA |
| 4551 | GAAAAAGAAA | AACAGCAGGT | GGGTCCGGGT | CGTGGGGCC | GGAAAAGCGA | GGAGGATCGC | GAGCAGCGAC |
| 4621 | GAGGCCCGGC | CCTCCCTCCG | CTTCCAAAGA | AACGCCCCC | ATGCCACATA | TATACATACC | CCCCCTCTC |
| 4691 | CTCCCATCCC | CCCAACCTA | CCACCAACAC | CACCACACC | TCCTCCCCC | TCGCTGCCGG | ACGACGAGCT |
| 4761 | CCTCCCCCCT | CCCCCTCCGC | CGCCGCCGGT | AACCACCCCG | CCCCCTCTCT | CTTCTCTTCT | CCGTTTTTT |
| 4831 | TTTCGTCTCG | GTCTCGATCT | TTGGCCTTGG | TAGTTTGGGT | GGCGGAGAGC | GGCTTCGTCT | CCCAGATCGG |
| | | | | | | BamHI | |
| | | | | | | ~~~~~ | |
| 4901 | TGCGCGGGAG | GGCGGGATC | TCGCGGCTGG | CGTCTCCGG | CGTGAGTCGG | CCCGGATCCT | CGCGGGGAAT |
| | | | | | BglII | ~~~~~ | |
| 4971 | GGGGCTCTCG | GATGTAGATC | TTCTTTCTTT | CTTCTTTTGG | TGGTAGAATT | TGAATCCCTC | AGCATTTGTC |
| | | | | | | HindIII | |
| | | | | | | ~~~~~ | |
| 5041 | ATCGGTAGTT | TTTCTTTTCA | TGATTTGTGA | CAAAATGCAGC | 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

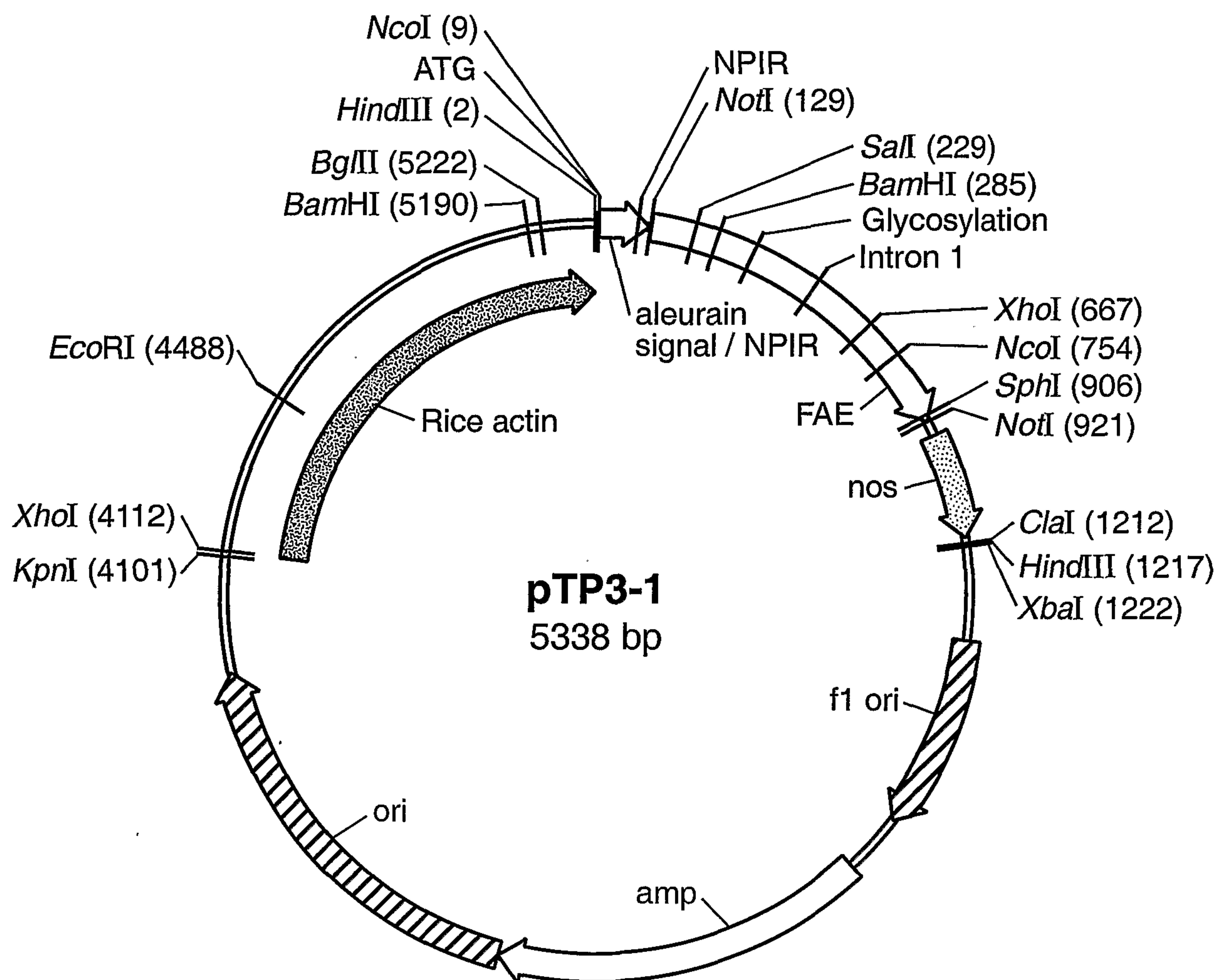
**FIG._39A**

FIG. 39B

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

GCCTTCGCGT CGTACATGAA CGATGCCCTC CAAGCCTCGA GCCAGATAC GACGCAGTAT TTCCGGGTCA

631

NcoI

~~~~~

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

CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA

701

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

CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCAC TGGG ATGAAGTGCA GTGCTGTGAG

771

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

GCCCAGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGG GCATGCACCT

841

NotI

~~~~~

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

GGCCGGTCGC GGCCGCGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAA CATTTGGCAA

911

TAAAGTTTCT TAAGATTGAA TCCTGTTGCC GGTCTTGCGA TGATTATCAT ATAATTCTG TTGAATTACG

981

TTAAGCATGT AATAATTAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTATGA TTAGAGTCCC

1051

GCAATTATAC ATTTAATACG CGATAGAAA CAAATATAG CGGCAAACT AGGATAAAT ATCGCGCGG

1121

HindIII

~~~~~

Clal XbaI

~~~~~

GTGTCACTA TGTACTAGA TCGATAAGCT TCTAGAGCG CCGGTGGAG TCCAATTGCG CCTATAGTGA

1191

GTCGTATTAC GCGCGCTCAC TGGCGTCTGT TTTACAACGT CGTGACTGGG AAAACCTGG CGTTACCCAA

1261

CTTAATCGCC TTGCAGCACA TCCCCCTTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC

1331

CTTCCCAACA GTTGCGCAGC CTGAATGGCG AATGGGACGC GCCCTGTAGC GGCGCATTA GCGCGCGGG

1401

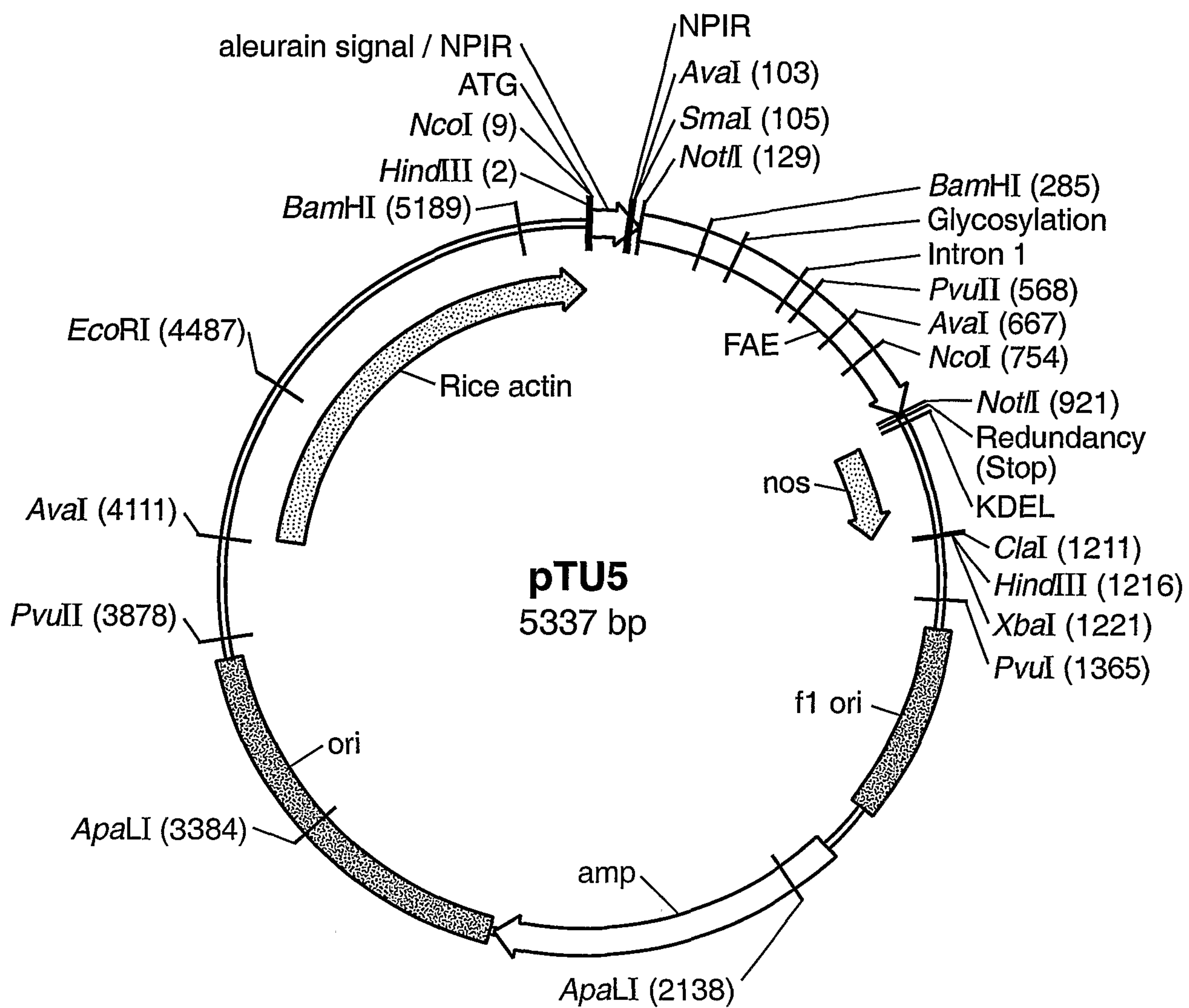
FIG.-39C

1471 TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCG CCGCTCCTTT CGCTTCTTC
1541 CTTTCCTTTC TCGCCACGTT CGCCGGCTTT CCCGTCAG GGTAAATCG GGGCTCCCT TTAGGGTTCC
1611 GATTAGTGC TTACGGCAC TTGACCCCA AAAACTTGA CTCAGCTGAT TTAGGGTAT GGTACGTA GTGGCCATC
1681 GCCCTGATAG ACGGTTTTC CACTCAACC TATCTCGGC AATTAAACA AATTTAAC GAATTTTAA AAAATATTAA CGCTTACAAT
1751 ACTGGAACAA CACTCAACC CACTCAACC TATCTCGGC TATCTTTTG TATTATAAGG GATTTGCCG ATTTCCGCC
1821 AATGGTTAAA AATGAGCTG AATGAGCTG AATTAACA AATTTAAC GAATTTTAA AAAATATTAA CGCTTACAAT
1891 TTAGGTGGCA CTTTTCGGG CATGAGACAA TAACCTGAT AAATGCTTCA ATAATATTGA AAAAGGAAGA GTATGAGTAT
1961 TGTATCCGCT CATGAGACAA TAACCTGAT AAATGCTTCA ATAATATTGA AAAAGGAAGA GTATGAGTAT
2031 TCAACATTTC CGTGTGCCC TTAATCCCTT TTTTGCGCA TTTTGCCCTTC CTGTTTTCG TCACCCAGAA
2101 ACGCTGGTGA AAGTAAAGA TGCTGAAGAT CAGTTGGTG CACGAGTGG TTACATCGAA CTGGATCTCA
2171 ACAGCGGTAA GATCCTTGAG AGTTTTCGCC CCGAAGAACG TTTTCCAATG ATGAGCACTT TTAAGTTCT
2241 GCTATGTGGC GCGGTATTAT CCCGTATTGA CGCCGGCAA CAGCAACTCG GTCGCCGAT GTCTATTTCT
2311 CAGAAATGACT TGGTTGAGTA CTCACAGTC ACAGAAAGC ATCTTACGGA TGGCATGACA GTAAGAGAA
2381 TATGCAGTGC TGCCATAACC ATGAGTGATA ACCTGCGC CAACTTACTT TGAACAACGA TCGGAGGACC
2451 GAAGGAGCTA ACCGCTTTT TGCACAACAT GGGGATCAT GTAACTCGCC TTGATCGTTG GGAACCGGAG
2521 CTGAATGAAG CCATACCAA CGACGAGCT GACACACGA TGCCGTAGC AATGGCAACA ACGTTGCGCA
2591 AACTATTAACT TGGCGAACTA CTTACTCTAG CTTCCCGCA ACAATTATA GACTGGATGG AGCGGATAA
2661 AGTTGCAGGA CCACTTCTGC GCTCGGCCCT TCCGGCTGGC TGGTTTATTG CTGATAAATC TGGAGCCGGT
2731 GAGCGTGGGT CTCGCGTAT CATGCAGCA CTGGGCGAG ATGGTAAGC ATGGTAAGC CTCCCGTATC GTAGTTATCT
2801 ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAAATAG ACAGATCGCT GAGATAGGTG CCTCAGTGAT
2871 TAAGCATTGG TAACGTGTCAG ACCAAGTTTA CTCATATATA CTTTAGATTG ATTTAAACT TCATTTTAA
2941 TTAAAGGA TCTAGGTGAA GATCCTTTT GATAATCTCA TGACCAAAAT CCGTTAACGT GAGTTTTCGT
3011 TCCACTGAGC GTCAGACCC GTAGAAAAGA TCAAGGATC TTCTTGAGAT TTCTTTTTC CCGCGTAAT TGCGCGTAAT
3081 CTGCTGCTTG CAAACAAAAG AACACCGCT ACCAGCGTG ACCAGCGTG GTTTGTTTG GGTAAACA GCTACCAACT
3151 CTTTTCCTGA AGGTAACCTG CTTAGCAGA CTTAGCATA CGCTACATA CTTGCTCTG CTAATCCTGT TACCAGTGGC
3221 TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCTACATA CTTGCTCTG CTAATCCTGT TACCAGTGGC
3291 TGCTGCCAGT GCGGATAAGT CGTGTCTTAC CCGGTTGGAC TCAAGACGAT AGTTACCGGA TAAGGCGCAG
3361 CCGTCGGGCT GAACGGGGG TTCGTGCACA CAGCCAGCT TGGAGCGAAC GACCTACACC GAACTGAGAT
3431 ACCTACAGCG TGAGCTATGA GAAAGCGCA CGCTTCCGA AGGAGAAAG GCGACAGGT ATCCGGTAAG
3501 CGGCAGGTC GGAACAGGAG AGGCACGAG GGAGCTTCCA GGGGAAACG CTGGGTATCT TTATAGTCCCT
3571 GTCGGGTTTC GCCACCTCTG ACTTGAGCGT CGATTTTGT GATGCTCGTC AGGGGGCGG AGCCTATGGA
3641 AAAACGCCAG CAACGCGCC TTTTACGGT TCCTGGCCT TTGCTCACA TTGCTTTC TGTCTTTC
3711 TGCGTTATCC CCTGATTCTG TGGATAACCG TATTACCGC TTTGAGTGAG CTGATACCGC TCGCCGAGC
3781 CGAACGACCG AGCGAGCGA GTCAGTGAG GAGGAAGCG GAGAGCGCC AATACGCAA CCGCTCTCC
3851 CCGCGCGTTG GCCGATTCT TAAATGAGT GGCACGACG GTTTCCCGAC TGGAAAGCGG GCAGTGAGCG

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 GATTTGTATA 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 CACCGCACGT CTCGCAGCCA
4761 AAAAAAAA AAGAAAGAAA AAAAGAAA AGAAACAG CAGGTGGGTC CAGGTCTGTG GGGCCGGAAA
4831 AGCGAGGAGG ATCGCGAGCA GCGACGAGG CCGGCCCTCC CTCGGCTTCC AAAGAAACGC CCCCATCGC
4901 CACTATATAC ATACCCCCC CTCTCTCC ATCCCCCAA CCTACCACC ACCACCAACA CCGGTAAACA CCGGCCCTC
4971 CCCCCTCGCT GCCGGACGAC GAGCTCCTCC CCCCTCCCC TCCGCCGCCG CCGGTAAACA CCGGCCCTC
5041 CTCCCTCTTC TTCTCTCCGT TTTTCTTCG TCTCGGTC TCATCTTGGC GATCTTGGC CTTGGTAGTT TGGGTGGCG
5111 AGAGCGGCTT CGTCGCCAG ATCGGTGCG ATCGGGGCG GATCTCGCG GGATCTCGCG CTTGGCTCT CCGGGCGTGA
BamHI
BgIII
~~~~~  
5181 GTCGGCCCCG ATCCTCGCG GGAATGGGC TCTCGGATGT AGATCTTCTT TCTTCTTCT TTTTGTGGTA  
5251 GAATTGGAAT CCTCAGCAT TGTTCATCG TAGTTTCTT TTTTCAATGAT 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<br>~~~~~         |             |            |            |            |
| 751  | GCCCATGGCG            | GTGTAGAGTA  | CTGGAGCGTT | GATCCTTACA | GCGCCAGAA  |
|      | CGGGTACCGC            | CACATCTCAT  | GACCTCGCAA | CTAGGAATGT | CGCGGGTCTT |
| 801  | CACATTTGTC            | TGCACTGGGG  | ATGAAGTGCA | GTGCTGTGAG | GCCCAGGGCG |
|      | GTGTAAACAG            | ACGTGACCCC  | TACTTCACGT | CACGACACTC | CGGGTCCCGC |
| 851  | GACAGGGTGT            | GAATAATGCG  | CACACGACTT | ATTTTGGGAT | GACGAGCGGC |
|      | CTGTCCCACA            | CTTATTACGC  | GTGTGCTGAA | TAAAACCCTA | CTGCTCGCCG |
|      | NotI<br>~~~~~         |             |            |            |            |
| 901  | GCATGCACCT            | GGCCGGTTCG  | 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 | TTTTTATGAT            | TAGAGTCCCG  | CAATTATACA | TTTAATACGC | GATAGAAAAC |
|      | AAAATACTA             | ATCTCAGGGC  | GTTAATATGT | AAATTATGCG | CTATCTTTTG |
| 1151 | AAAATATAGC            | GCGCAAAC TA | GGATAAATTA | TCGCGCGCGG | TGTCATCTAT |
|      | TTTTATATCG            | CGCGTTTGAT  | CCTATTTAAT | AGCGCGCGCC | ACAGTAGATA |
|      | XbaI<br>~~~~~         |             |            |            |            |
|      | ClaI HindIII<br>~~~~~ |             |            |            |            |
| 1201 | GTTACTAGAT            | CGATAAGCTT  | CTAGAGCGGC | CGGTGGAGCT | CCAATTCGCC |
|      | CAATGATCTA            | GCTATTTCGA  | GATCTCGCCG | GCCACCTCGA | GGTTAAGCGG |
| 1251 | CTATAGTGAG            | TCGTATTACG  | CGCGCTCACT | GGCCGTCGTT | TTACAACGTC |
|      | GATATCACTC            | AGCATAATGC  | GCGCGAGTGA | CCGGCAGCAA | AATGTTGCAG |
| 1301 | GTGACTGGGA            | AAACCCTGGC  | GTTACCCAAC | TTAATCGCCT | TGCAGCACAT |
|      | CACTGACCCT            | TTTGGGACCG  | CAATGGGTG  | AATTAGCGGA | ACGTCGTGTA |
|      | PvuII<br>~~~~~        |             |            |            |            |
| 1351 | CCCCCTTTTC            | CCAGCTGGCG  | TAATAGCGAA | GAGGCCCGCA | CCGATCGCCC |
|      | GGGGGAAAGC            | GGTCGACCGC  | ATTATCGCTT | CTCCGGGCGT | GGCTAGCGGG |
| 1401 | TTCCCAACAG            | TTGCGCAGCC  | TGAATGGCGA | ATGGGACGCG | CCCTGTAGCG |
|      | AAGGGTTGTC            | AACGCGTCGG  | ACTTACCGCT | TACCCTGCGC | GGGACATCGC |

**FIG. 40C**

SUBSTITUTE SHEET (RULE 26)

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

**FIG. 40D**

SUBSTITUTE SHEET (RULE 26)

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

**FIG. 40E**

SUBSTITUTE SHEET (RULE 26)

|      |             |            |            |            |             |
|------|-------------|------------|------------|------------|-------------|
| 3301 | GCGATAAGTC  | GTGTCTTACC | GGGTTGGACT | CAAGACGATA | GTTACCGGAT  |
|      | CGCTATTTCAG | 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 | AGGCCATTTCG |
| 3501 | GGCAGGGTCG  | GAACAGGAGA | GCGCACGAGG | GAGCTTCCAG | GGGGAAACGC  |
|      | CCGTCCCAGC  | CTTGTCTCT  | CGCGTGCTCC | CTCGAAGGTC | CCCCTTTGCG  |
| 3551 | CTGGTATCTT  | TATAGTCCTG | TCGGGTTCG  | CCACCTCTGA | CTTGAGCGTC  |
|      | GACCATAGAA  | ATATCAGGAC | AGCCCAAAGC | GGTGGAGACT | GAACTCGCAG  |
| 3601 | GATTTTGTG   | ATGCTCGTCA | GGGGGGCGGA | GCCTATGGAA | AAACGCCAGC  |
|      | CTAAAAACAC  | TACGAGCAGT | CCCCCGCCT  | CGGATACCTT | TTTGCGGTTCG |
| 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  | GTCACTCGCG | TTGCGTTAAT | TACACTCAAT | CGAGTGAGTA  |
| 3951 | TAGGCACCCC  | AGGCTTTACA | CTTTATGCTT | CCGGCTCGTA | TGTTGTGTGG  |
|      | ATCCGTGGGG  | TCCGAAATGT | GAAATACGAA | GGCCGAGCAT | ACAACACACC  |
| 4001 | AATTGTGAGC  | GGATAACAAT | TTCACACAGG | AAACAGCTAT | GACCATGATT  |
|      | TTAACACTCG  | CCTATTGTTA | AAGTGTGTCC | TTTGTGATA  | CTGGTACTAA  |
| 4051 | ACGCCAAGCG  | CGCAATTAAC | CCTCACTAAA | GGGAACAAAA | GCTGGGTACC  |
|      | TGCGGTTCGC  | GCGTTAATTG | GGAGTGATTT | CCCTTGTTTT | CGACCCATGG  |
|      |             |            | AvaI       |            |             |
|      |             |            | ~~~~~      |            |             |
| 4101 | GGGCCCCCCC  | TCGAGGTCAT | TCATATGCTT | GAGAAGAGAG | TCGGGATAGT  |
|      | CCCGGGGGGG  | AGCTCCAGTA | AGTATACGAA | CTCTTCTCTC | AGCCCTATCA  |

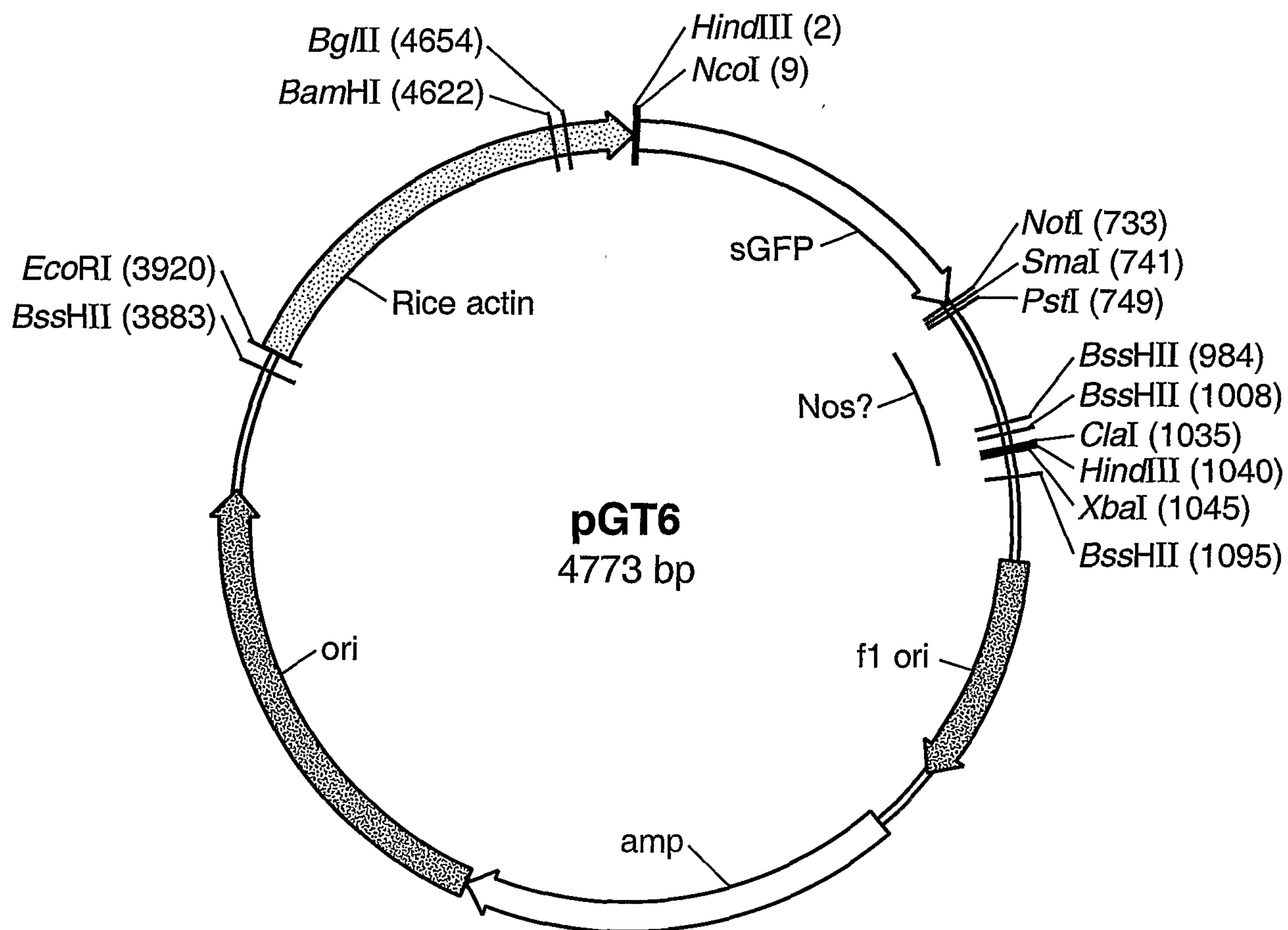
**FIG. 40F**

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  |
|      | AACATATGCAG | 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 GGCCGTTTGA CCGGCACGG ACCGGTGGG AGCACTGGTG GAAGTGGATG

211 GGCGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCACGACTT CTTCAAGTCC GCCATGCCCG
 CCGCACGTCA CGAAGTCGGC GATGGGGCTG GTGTACTTCG TCGTGCTGAA GAAATTTCAGG CGGTACGGGC

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 CCGGCTGTTT 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 GGTGAGGCGG

631 CTGAGCAAAG ACCCAACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT CGTGACCGCC GCCGGATCA
 GACTCGTTTC TGGGTTGCT CTTTCGCGCTA GTGTACCAGG ACGACCTCAA GCACTGGCGG CGGCCCTAGT

FIG._41B

| SmaI | | NotI | | PstI | |
|--------|--|--------|--|--------|--|
| ~~~~~ | | ~~~~~ | | ~~~~~ | |
| 701 | CTCACGGCAT GGACGAGCTG TACAAGTAA GCGCGCGCC GGGCTGCAGG GAAACCACTG AAGGATGAGC GAGTGCCGTA CCTGCTCGAC ATGTTCAATT CGCCGCGGG CCCGACGTCC CTTTGGTGAC TTCCTACTCG | | | | |
| 771 | TGTAAGAAG CAGATCGTTC AAACATTGG CAATAAGTT TCTTAAGATT GAATCCTGTT GCCGGTCTTG ACATTCTTC GTCTAGCAAG TTTGTAAACC GTTATTCTAA AGAATTCTAA CTTAGGACAA CGGCCAGAAC | | | | |
| 841 | CGATGATTAT CATATAATT CTGTTGAATT ACGTTAAGCA TGTAATAATT AACATGTAAT GCATGACGTT GCTACTAATA GTATATTAA GACAACTTAA TGCAATTCTG ACATTATTAA TTGTACATTA CGTACTGCAA | | | | |
| 911 | ATTTATGAGA TGGGTTTTA TGATTAGAGT CCCGCAATTA TACATTTAAT ACGCGATAGA AAACAAATA TAAATACTCT ACCCAAAAT ACTAATCTCA GGCCTTAAT ATGTAAATTA TCGCTATCT TTTGTTTAT | | | | |
| BssHII | | BssHII | | BssHII | |
| ~~~~~ | | ~~~~~ | | ~~~~~ | |
| 981 | TAGCGGCAA ACTAGGATAA ATTATCGCG GCGGTGTCAT CTATGTTACT AGATCGATAA GCTTCTAGAG ATCGCGCGTT TGATCCTATT TAATAGCGG CGCCACAGTA GATACAAATGA TCTAGCTATT CGAAGATCTC | | | | |
| BssHII | | BssHII | | BssHII | |
| ~~~~~ | | ~~~~~ | | ~~~~~ | |
| 1051 | CGGCCGGTGG AGCTCCAATT CGCCCTATAG TGAGTCGTAT TACGCGCGCT CACTGGCCGT CGTTTACAA GCCGGCCACC TCGAGGTTAA GCGGGATATC ACTCAGCATA ATGCGCGCGA GTGACCGGCA GCAAAATGTT | | | | |
| 1121 | CGTCGTGACT GGGAAAACCC TGGCGTTACC CAACTTAATC GCCTTGCAGC ACATCCCCCT TTCGCCAGCT GCAGCACTGA CCTTTTGGG ACCGCAATGG GTTGAAATAG CGGAACGTCG TGTAGGGGA AAGCGGTCGA | | | | |
| 1191 | GGCGTAATAG CGAAGAGGCC CGCACCGATC GCCCTTCCA ACAGTTGCGC AGCCTGAATG GCGAATGGGA CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAGGT TGTCAACGCG TCGGACTTAC CGCTTACCT | | | | |
| 1261 | CGCGCCCTGT AGCGGCGCAT TAAGCGCGG GGTGTGGTG GTTACGCGCA GCGTGACCGC TACACTTGCC GCGCGGACA TCGCCGCGTA ATTGCGCGC CACACACCAC CAATGCGCGT CGCACTGGCG ATGTGAACGG | | | | |

XbaI
~~~~~  
ClaI HindIII  
~~~~~

FIG. 41C

1331 AGCGCCCTAG CGCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCCGGC TTTCCCCGTC
TCGCGGGATC GCGGCGAGG AAGCGAAG AAGCGGGTG CAAAGGCCG AAGGGGCAG

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 CGCGAATTTT 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 GGAATAAGG GAAAAACGC

1891 GCATTTTGCC TTCCCTGTTT TGCTCACCCA GAAACGCTGG TGAAAGTAAA AGATGCTGAA GATCAGTTGG
CGTAAAACGG AAGGACAAA ACGAGTGGT CTTTGCAGC ACTTTCATT TCTACGACTT CTAGTCAACC

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

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

2101 CAAGAGCAAC TCGGTCGCC CATACACTAT TCTCAGAA'TG ACTTGGTTGA GTACTCACC'A GTCACAGAAA
GTTCTCGTTG AGCCAGCGC GTATGTGATA AGAGTCTTAC TGAACCAACT 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 ACCTCCGCC TATTCAACGT CCTGGTGAAG ACGCGAGCCG GGAAGGCCGA

2521 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTC TC GCGG TATCATTGCA GCAC TGGGGC
CCGACCAAAT AACGACTATT TAGACCTCGG CCACTCGCAC CCAGAGCGCC ATAGTAACGT CGTGACCCCG

2591 CAGATGGTAA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGAGTCAG GCAACTATGG ATGAACGAAA
GTC TACCA TT CGGAGGGCA TAGCATCAAT AGATGTGCTG CCC TCAGTC CGTTGATACC TACTTGCTTT

2661 TAGACAGATC GCTGAGATAG GTGCC TCACT GATTAAGCAT TGGTAACTGT CAGACCAAGT TTA CT CATAT
ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTGTA ACCATTGACA GTCTGGTTCA AATGAGTATA

2731 ATACTTTAGA TTGATTTAAA ACTTCATTTT TAATTTAAA GGATCTAGGT GAAGATCCTT TTTGATAATC
TATGAAATCT AACTAAATTT TGAAGTAAA ATTAATTTT CCTAGATCCA CTTCTAGGAA AAAC TATTAG

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

2871 ATCTTCTTGA GATCCTTTTT TTCTGCGCGT AATCTGCTGC TTGCAACAA AAAAACACC GCTACCAGCG
TAGAAGAACT CTAGGAAAAA AAGACGCGCA TTAGACGACG AACGTTTGT TTTT TGGTGG CGATGGTCGC

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 ACAATGGTCA CCGACGACGG TCACCGCTAT TCAGCACAGA ATGGCCCCAAC

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 TC CGGT 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 CAGCAACGCG GCCTTTTAC GGTTCCTGGC
ACACTACGAG CAGTCCCCC GCCTCGGATA CCTTTTTCG 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 AACCGGCTAA 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 | GATTACGCCA | AGCGCGCAAT | TAACCTCAC |
| | GTCCTTTGTC | GATACTGGTA | CTAATGCGGT | TCGCGCGTTA | ATTGGAGTG |
| | | EcoRI | | TAAAGTGTTA | |
| | | ~~~~~ | | TTCCTCCCTG | |
| 3921 | ATTCCACAAT | GAAACAATAAT | AAGATTAAAA | TAGCTTGCCC | CCGTTGCAGC |
| | TAAGGTGTTA | CTTGTTATTA | TTCCTAATT | ATCGAACGGG | GGCAACGTCG |
| 3991 | AAATAAAAGA | TAAACTTAGA | CTCAAAAACAT | TTACAAAAC | AACCCCTAAA |
| | TTTATTTTCT | ATTGAATCT | GAGTTTGTGTA | AATGTTTTTG | TTGGGGAATT |
| 4061 | ATGCACGATC | CATAGCAAGC | CCAGCCCAAC | CCAACCCACC | CCAGTGCAGC |
| | TACGTGCTAG | GTATCGTTTG | GGTCGGGTTG | GGTTGGGTG | GGTCACGTCG |
| 4131 | ATAGTCTCCA | CCCCCGGCAC | TATCACCGTG | AGTTGTCCGC | ACCACCGCAC |
| | TATCAGAGGT | GGGGCCCGTG | ATAGTGGCAC | TCAACAGCG | TGGTGGCGTG |
| 4201 | AAAAGAAAGA | AAAAAAAGAA | AAAGAAAAAC | AGCAGGTGGG | TCCGGGTCGT |
| | TTTTCCTTCT | TTTTTTTCTT | TTTCTTTT | TCGTCCACCC | AGGCCAGCA |
| 4271 | GGATCGCGAG | CAGCGACGAG | GCCCCGCCCT | CCCTCCGCTT | CCAAAGAAAC |
| | CCTAGCGCTC | GTCGCTGCTC | CGGGCCGGGA | GGAGGCGGAA | GGTTCTTTTG |
| 4341 | ACATACCCCC | CCCTCTCCTC | CCATCCCCC | AACCTACCA | CCACCACCAC |
| | TGTATGGGG | GGGAGAGGAG | GGTAGGGGG | TTGGGATGGT | GGTGGTGGTG |
| 4411 | CTGCCGGACG | ACGAGCTCCT | CCCCCCTCCC | CTCCGCCGC | CGCCGGTAAC |
| | GACGGCCTGC | TGCTCGAGGA | GGGGGAGGG | GGAGCGGCG | GCGCCATTG |
| 4481 | TCCTTCTCCG | TTTTTTT | CGTCTCGGTC | TCGATCTTTG | GCCTTGGTAG |
| | AGAAAGAGGC | AAAAAA | GCAGAGCCAG | AGCTAGAAAC | CGGAACCATC |
| 4551 | TTTCGTCGCC | AGATCGGTGC | GCGGGAGGG | CGGATCTCG | CGGCTGGCGT |
| | AAGCAGCGGG | TCTAGCCACG | CGCCCTCCC | GCCCTAGAGC | GCCGACCGCA |

FIG._41G

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

FIG. 41H

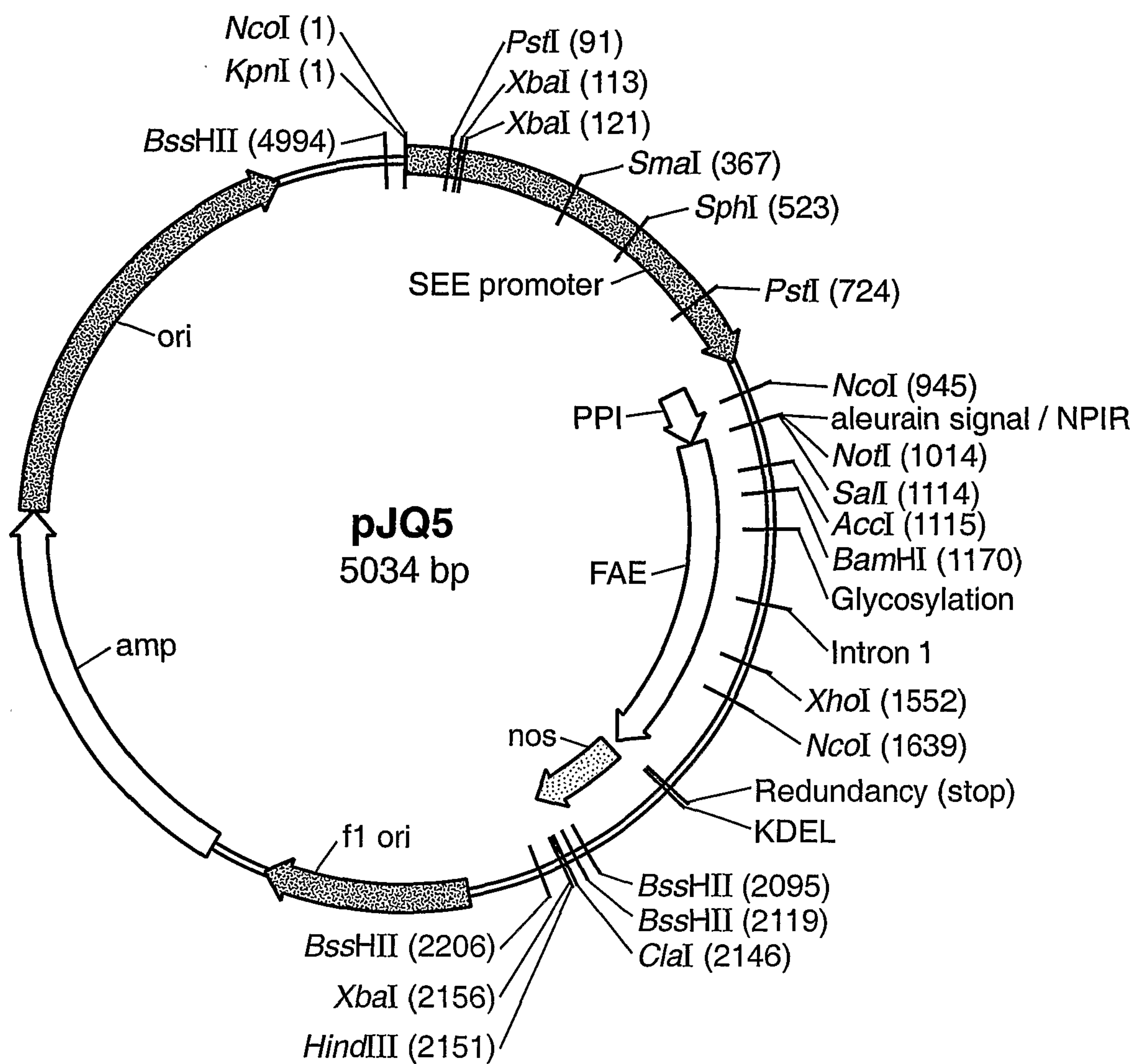
**FIG. 42A**

FIG. 42B

| | | | | | | | |
|------|--------------------------------|------------|------------|-------------|-------------|------------|------------|
| 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 | TTTCGTTACC | ACTTCGCACC | GCTACTCTAC | CCATATTTTG |
| 771 | CCCCGGCACC | GGGACGCGAG | CTCCCGCCTA | CCAGTACCAT | CTCGCCTCGC | TCCCCCTGCC | GGACGACCCA |
| | GGGGCCGTGG | CCCTGCGCTC | GAGGGCGGAT | GGTCATGGTA | GAGCGGAGCG | AGGGGGACGG | CCTGCTGGGT |
| 841 | GTAATAATACT | GTTGCCCACT | CGCCGGCGAG | ATGGMCGTGC | ACAAGGAGGT | SAACTTCGTS | GCCTACCTCC |
| | CATTTTATGA | CAACGGGTGA | GCGGCCGCTC | TACCKGCACG | TGTTCTCTCA | 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 | TACAATTCTC | AAACTGACAT | TAACGGATGG | ATCCTCCGCG | ACGACAGCAG |
| | AATAGTTCCC | TCTCTTTTAA | ATGTTAAGAG | TTTGACTGTA | ATTGCCCTACC | TAGGAGGCGC | TGCTGTCGTC |

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 TCAAACAGCA GGTTAGCCAG TATCCGGACT ACGCGCTGAC
AGAGGCAGGT CCTGGTTCAG CTCAGCGAAC AGTTTGTCGT CCAATCGGTC ATAGGCCCTGA TGC CGACTG

1401 CGTGACCGGC CACKCCCTCG GCGCCTCCCT GCGGGCACTC ACTGCCGCC AGCTGTCTGC GACATACGAC
GCACTGGCCG GTGMGGGAGC CCGGAGGGA CCGCCGTGAG TGACGGCGGG TCGACAGACG CTGTA TGCTG

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

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

1611 CCTGCCCCCG GTGGAGCAGG GGTAAGCCCA 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
TACGCGTGTG CTGAAATAAA CCCTACTGCT CGCCTCGGAC ATGTACCACT AGTCAGTAAA GTCGGAGGGG

1821 GAGTGTAACA GGAAAGATGG ATGTCTTGA GAGGGGGCCG CGTAACCACT GAAGGATGAG CTGTAAAGAA
CTCACATGGT CCTTTC TACC TACAGGACCT CTCCCCCGGC GCATTGGTGA CTTCCCTACTC GACATTTCTT

XhoI

~~~~~

NcoI

~~~~~

FIG.- 42D

1891

GCAGATCGTT CAAACATTTG GCAATAAAGT TTCTTAAGAT TGAATCCTGT TGCCGGTCTT GCGATGATTA CGTCTAGCAA GTTTGTAAAC CGTTATTCA AAGAAATCTA ACTTAGGACA ACGGCCAGAA CGCTACTAAT

1961

TCAATATAAT TCTGTTGAAT TACGTTAAGC ATGTAATAAT TAACATGTAA TGCATGACGT TATTATGAG AGTATATTAA AGACAACCTTA ATGCAATTCTG TACATTATTA ATTGTACATT ACGTACTGCA ATAAATATCTC

2031

ATGGGTTTTT ATGATTAGAG TCCCGCAATT ATACATTTAA TACGCGATAG AAAACAAAAT ATAGCGCGCA TACCCAAAAA TACTAATCTC AGGCGTTAA TATGTAAATT ATGCGCTATC TTTTGTTTTA TATCGCGCGT

2101

AAC TAGGATA AATATCGCG CGCGGTGCA TCTATGTTAC TAGATCGATA AGCTTCTAGA GCGGCCGGTG TTGATCCTAT TTAATAGCGC GCGCCACAGT AGATACAATG ATCTAGCTAT TCGAAGATCT CGCCGGCCAC

2171

GAGCTCCAAT TCGCCCTATA GTAGTCGTA TTACGCGCGC TCACTGGCCG TCGTTTACA ACGTCGTGAC CTCGAGGTTA AGCGGATAT CACTCAGCAT AATGCGCGCG AGTGACCGGC AGCAAAATGT TGCAGCACTG

2241

TGGGAAAACC CTGGCGTTAC CCAACTTAAT CGCCTTGCAG CACATCCCC TTTCCGCAGC TGGCGTAATA ACCCTTTTGG GACCGCAATG GGTTGAAATTA GCGGAACGTC GTGTAGGGG AAAGCGGTCG ACCGCATTAT

2311

GCGAAGAGGC CCGCACC GAT CGCCCTTCCC AACAGTTGCG CAGCCTGAAT GGCGAATGGG ACGCGCCCTG CGCTTCTCCG GCGGTGGCTA GCGGGAAGG TTGTCAACGC GTCGGACTTA CCGCTTACCC TGC CGGGAC

2381

TAGCGGCGCA TTAAGCGCGG CGGGTGTTGT GGTACGCGC AGCGTGACCG CTACACTTGC CAGCGCCCTA ATCGCCGCGT AATTCGCGCC GCCACACCA CCAATGCGCG TC GCACTGGC GATGTGAACG GTC CGGGGAT

2451

GCGCCCGCTC CTTTCGCTTT CTTCCCTTCC TTTCTCGCCA CGTTCGCCGG CTTTCCCCGT CAAGCTCTAA CGCGGGCGAG GAAAGCGAAA GAAGGAAGG AAAGAGCGGT GCAAGCGGCC GAAAGGGGCA GTTCGAGATT

BssHII

~~~~~

Clal HindIII

~~~~~

XbaI

~~~~~

BssHII

~~~~~

FIG. 42E

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

2591 TGATGGTTCA CGTAGTGGG CATGCCCTG 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 GCTGATTTA CAAAAATTA ACGCGAATTT
TTCCCTAAAA CGGCTAAAGC CGGATAACCA ATTTTTACT CGACTAAAT GTTTTTAAAT TGCGCTTAAA

2801 TAACAAAAATA TTAACGCTTA CAATTAGGT GGCACTTTTC GGGGAAATGT GCGCGGAACC CCTATTGTT
ATTGTTTAT AATTGCGAAT GTTAAATCCA CCGTGAAAG CCCCTTACA CGCGCCTTG GGATAAACAA

2871 TATTTTCTA AATACATTCA AATATGTATC CGCTCATGAG ACAATAACCC TGATAAATGC TTCAATAATA
ATAAAAAGAT TTAGTAAAGT TTATACATAG GCGAGTACTC TGTTATTGG ACTATTACG AAGTTATTAT

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

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

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

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

3221 CTCGGTCGCC GCATACACTA TTC TCAGAAAT 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 TGCAC TCATAA AGCAAGGTGA CTCGCAGTCT GGGGCATCTT TTCTAGTTTC CTAGAAGAAC

3991 AGATCCCTTTT TTTC TGCGCG TAATCTGCTG CTTGCAAACA AAAAACCCAC CGCTACCAGC GGTGGTTTGT
TCTAGGAAA AAAGACGCGC ATTAGACGAC GAACGTTTGT TTTT TTTGGTG GCGATGGTCG CCACCAAACA

4061 TTGCCGGATC AAGAGCTACC AACTCTTTT CCGAAGGTAA CTGGCTTCAG CAGAGCGCAG ATACCAAATA
AACGGCCTAG TTCTCGATGG TTGAGAAAAA GGCTTCCATT GACCGAAGTC GTCTCGCGTC TATGGTTTAT

4131 CTGTCCCTTCT AGTGTAGCCG TAGTTAGGCC ACCACTTCAA GAAC TCTGTA GCACCGCCTA CATACCTCGC
GACAGGAAGA TCACATCGGC ATCAATCCGG TGGTGAAAGTT CTTGAGACAT CGTGGCGGAT GTATGGAGCG

FIG. 42G

4201 TCTGCTAATC CTGTTACCAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGA^{CT}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^TTCCCCTC

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^TTTTGAG
CGGAAAAACGA GTGTACAAGA AAGGACGCAA TAGGGGACTA AGACACCTAT TGGCATAATG GCGGAAAC^TC

4691 TGAGCTGATA CCGCTCGCCG CAGCCGAACG ACCGAGCGCA GCGAGTCAGT GAGCGAGGAA GCGGAAGAGC
ACTCGACTAT GCGGAGCGGC GTCGGCTTGC TGGCTCGCGT CGCTCAGTCA CTCGCTCCTT CGC^TTTCTCG

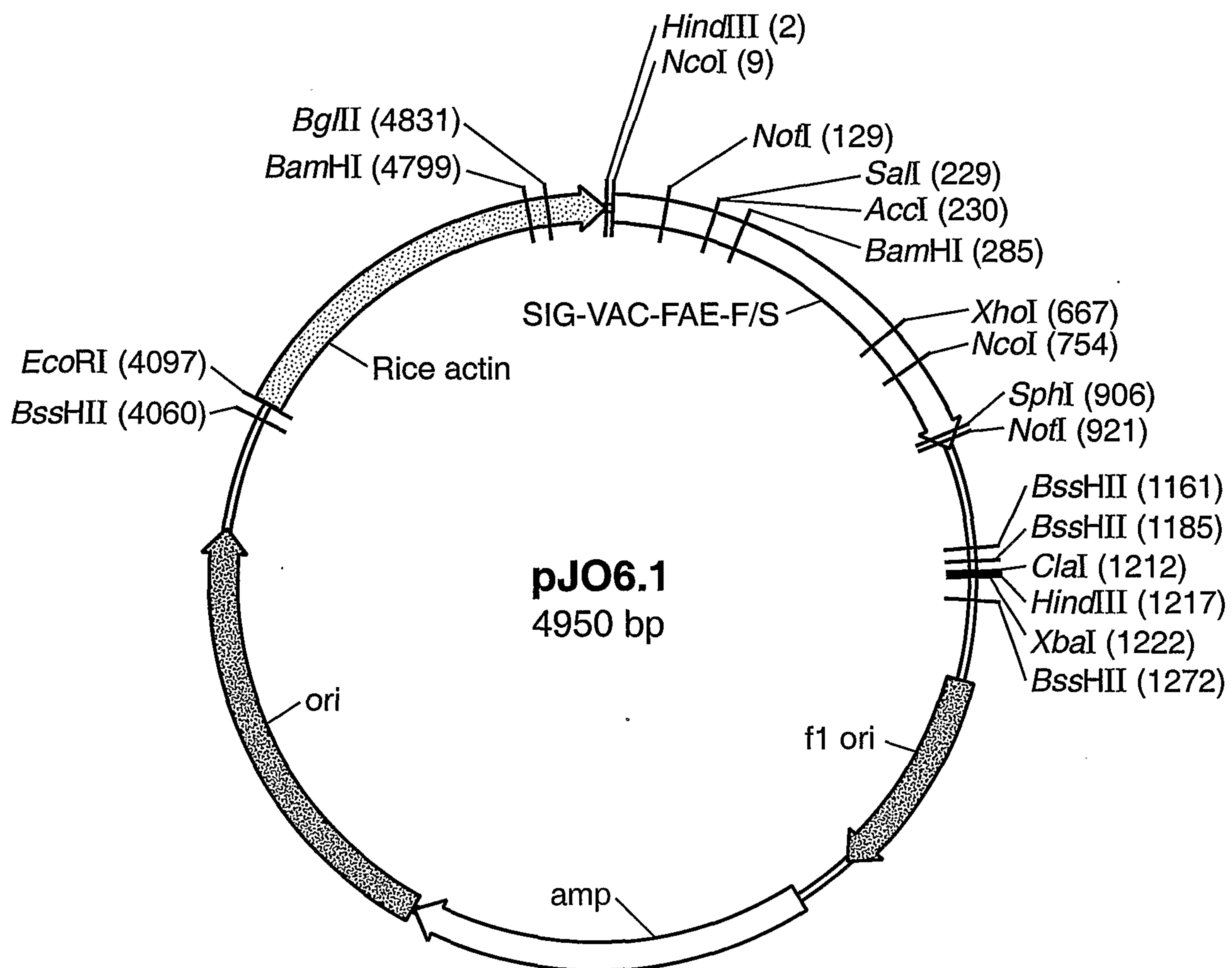
4761 GCCCAATACG CAAACCGCCT CTC^{CC}CGCGC GTTGGCCGAT TCATTAAATGC AGCTGGCACG ACAGGTTTCC
CGGGTTATGC GTTTGGCGGA GAGGGCGCG CAACCGGCTA AGTAAATTACG TCGACCGTGC TGTCCAAAGG

4831 CGACTGGAAA GCGGGCAGTG AGCGCAACGC AATTAATGTG AGTTAGCTCA CTCATTAGGC ACCCAGGCT
GCTGACCTTT CGCCCGTCAC TCGCGTTGCG TTAATTACAC TCAATCGAGT GAGTAAATCCG TGGG^TGTCCGA

4901 TTACACTTTA TGCTTCCGGC TCGTATGTTG TGTGGAATTG TGAGCGGATA ACAA^{TT}TCAC ACAGGAAACA
AATGTGAAAT ACGAAGGCCG AGCATACAAC ACACCTTAAC ACTCGCCTAT TGT^TAAAGTG TGTCC^TTTGT

FIG.- 42H

SUBSTITUTE SHEET (RULE 26)

**FIG. 43A**

| | |
|--------------|--|
| HindIII NcoI | |
| ~~~~~ ~~~~~~ | |
| 1 | AAGCTTACCA TGGCCACGC CCGGTCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTCGCCG
TTCGAATGGT ACCGGGTGCG GCGCAGGAG GAGGACCGCG AGCGGCACGA CCGGTGCCGG CGGCAGCGGC |
| NotI | |
| ~~~~~ | |
| 71 | TCGCCTCCTC CTCCTCCTTC GCCACTCCA ACCGATCCG GCCCGTCACC GACCGCGCGG CCGCCTCCAC
AGCGGAGGAG GAGGAGGAAG CCGCTGAGGT TGGGCTAGGC CGGGCAGTGG CTGGCGCGCC GCGGAGGTG |
| 141 | GCAGGGCATC TCCGAAGACC TCTACAGCCG TTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC
CGTCCCCTAG AGGCTTCTGG AGATGTCGGC AAATCAGCTT TACCGGTGAT AGAGGGTTCTG ACGGATGCCG |
| SalI | |
| ~~~~~ | |
| AccI | |
| ~~~~~ | |
| 211 | GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTTACAA TTCTCAAAC T GACATTAACG
CTGGACACGT TGTAAGGCAG CTGATAATAG TTCCCCTCTCT TTTAATATGTT AAGAGTTTGA CTGTAATTGC |
| BamHI | |
| ~~~~~ | |
| 281 | GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCAC TGGTA GTGATACGAA
CTACCTAGGA GCGCTGCTG TCGTCGTTTC TTATA TTAGTG GCAGAAGGCA CCGTGACCAT CACTATGCTT |
| 351 | TCTACAAC TC GATACTAACT ACACCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA
AGATGTTGAG CTATGATTGA TGTGGGAGTG CGGAAAGCTG TGGGATGGTG TTACGTTGCC AACACTTCAT |
| 421 | CACGGTGGAT ATTATATTGG ATGGGTCCTC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA
GTGCCACCCTA TAAATATAACC TACCCAGAGG CAGGTCCTGG TTCAGCTCAG CGAACAGTTT GTCGTCCAAT |
| 491 | GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGCGGCC TCCCTGGCGG CACTCACTGC
CGGTCATAGG CCTGATGCGC GACTGGCACT GGCCGGTGMG GGAGCCGCGG AGGGACCGCC GTGAGTGACG |
| 561 | CGCCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGGCAG CGGCAATCAG
GCGGGTCGAC AGACGCTGTA TGCTGTTGTA GCGGACATG TGGAAGCCGC TTGGCGCGTC GCCGTTAGTC |

FIG. 43B

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

FIG. 43C

1191

GTGTCAATCTA
CACAGTAGAT

TGTTACTAGA
ACAAATGATCT

TCGATAAGCT
AGCTATTTCGA

TCTAGAGCGG
AGATCTCGCC

CCGGTGGAGC
GGCCACCTCG

TCCAATTTCG
AGGTTAAGCG

CCTATAGTGA
GGATATCACT

1261

GTCGTATTAC
CAGCATAATG

GCGCGCTCAC
CGCGCGAGTG

TGGCCGTCGT
ACCGGCAGCA

TTTACAACGT
AAATGTTGCA

CGTGACTGGG
GCACTGACCC

AAAACCTTGG
TTTTGGGACC

CGTTACCCAA
GCAATGGGT

1331

CTTAATCGCC
GAAATAGCGG

TTGCAGCACA
AACGTCGTGT

TCCCCCTTTC
AGGGGAAAG

GCCAGCTGGC
CGGTCGACCG

GTAATAGCGA
CATTATCGCT

AGAGGCCCGC
TCTCCGGGCG

ACCGATCGCC
TGGCTAGCGG

1401

CTTCCCAACA
GAAGGGTTGT

GTTGCGCAGC
CAACGCGTGG

CTGAATGGCG
GACTTACCGC

AATGGGACGC
TTACCCCTGCG

GCCCTGTAGC
CGGGACATCG

GGCGCATTAA
CCGCGTAATT

GCGCGGCGGG
CGCGCCGCC

1471

TGTGGTGGTT
ACACCACCAA

ACGCGCAGCG
TGC GCGTCGC

TGACCGCTAC
ACTGGCGATG

ACTTGCCAGC
TGAACGGTCG

GCCCTAGCGC
CGGGATCGCG

CCGCTCCTTT
GGCGAGGAAA

CGCTTTCCTC
GCGAAAGAA

1541

CCTTCCCTTC
GGAAGGAAAG

TCGCCACGTT
AGCGGTGCAA

CGCCGGCTTT
GCGGCCGAAA

CTCTAAATCG
GGGCAGTTC

GGGGTCCCT
CCCCGAGGGA

TTAGGGTTC
AATCCCAAG

1611

GATTTAGTGC
CTAAATCACG

TTTACGGCAC
AAATGCCGTG

CTCGACCCCA
GAGCTGGGGT

AAAAACTTGA
TTTTTTGAACT

TTAGGGTGAT
AATCCCACTA

GGTTCACGTA
CCAAGTGCAT

GTGGGCCATC
CACCCGGTAG

1681

GCCCTGATAG
CGGGACTATC

ACGGTTTTTC
TGCCAAAAG

GCCCTTTGAC
CGGAAAAC TG

GTTGGAGTCC
CAACCTCAGG

ACGTTCTTTA
TGCAAGAAAT

ATAGTGGACT
TATCACCTGA

CTTGTTCCAA
GAACAAGGTT

1751

ACTGGAAACA
TGACCTTGTT

CAC TCAACCC
GTGAGTTGGG

TATCTCGGTC
ATAGAGCCAG

ATTATAAAGG
TAAATATTCC

GATTTTGCCG
CTAAAACGGC

ATTTCGGCCT
TAAAGCCGGA

1821

ATTGGTTAAA
TAACCAATTT

AAATGAGCTG
TTTACTCGAC

ATTTAACAAA
TAAATTGTTT

AATTTAACGC
TTAAAATTGCG

AAAATATTAA
TTTTTATAATT

CGCTTACAAT
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 TTTTCCTTCT CATACTCATA

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

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 AAAAGTTAC 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 AGCTCCTGG

2451 GAAGGAGCTA ACCGCTTTT TGCACAACAT GGGGATCAT GTAACTCGCC TTGATCGTTG GGAACCGGAG
CTTCTTCGAT 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 TAACTGTCAG ACCAAGTTTA CTCATATATA CTTTAGATTG ATTTAAAACT TCATTTTAA
ATTCTAACC ATTGACAGTC TGGTTCAAAT GAGTATATAT GAAATCTAAC TAAATTTTGA AGTAAAAAAT

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

3011 TCCACTGAGC GTCAGACCCC GTAGAAAAA GTAGAAAGATC TTCTTGAGAT CTTTTTTTTC TGC CGGTAAT
AGGTGACTCG CAGTCTGGG CATCTTTCT AGTTCCCTAG AAGAACTCTA GGAATAAAG ACGGCATTA

3081 CTGCTGCTTG CAAACAAAAA AACACCGCT ACCAGCGGTG GTTTGTTTG CGGATCAAGA GCTACCAACT
GACGACGAAC GTTTGTTTT TTGGTGCGA TGGTCGCCAC CAAACAAACG GCCTAGTTCT CGATGGTTGA

3151 CTTTTTCCGA AGGTAAC TGTAGCAC CTTCAGCAGA GCGCAGATAC CAAATACTGT CTTCTAGTG TAGCCGTAGT
GAAAAAGGCT TCCATTGACC GAAAGTCGTCT CGGCTCTATG 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 AGTTC TGCTA TCAATGGCCT ATTCCGCGTC

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

3431 ACCTACAGCG TGAGCTATGA GAAAGCGCCA CGCTTCCCGA AGGGAGAAAG GCGGACAGGT ATCCGGTAAG
TGGATGTCGC ACTCGATACT CTTTCGCGGT GCGAAGGGCT TCCCTCTTTC CGCCTGTCCA TAGGCCATT

3501 CGGCAGGGTC GGAACAGGAG AGGCACGAG GGAGCTTCCA GGGGAAACG CCTGGTATCT TTATAGTCTT
GCCGTCCCAG CCTGTCTC 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 AAAAA TGCCA AGGACCGGAA AACGACCGGA AACGAGTGT ACAAGAAAGG

3711 TCGGTTATCC CCTGATTCTG TGGATAACCG TATTACCGCC TTTGAGTGAG CTGATACCGC TCGCCGCAGC
ACGCAATAGG GGACTAAGAC ACCTATTGGC ATAA TGCGG AAAC TCACTC GACTATGGCG AGCGCGTCCG

3781 CGAACGACCG AGCGAGCGA GTCAGTGAGC GAGGAAGCGG AAGAGCGCCC AATACGCAA CCGCCTCTCC
GCTTGCTGGC TCGCGTCGCT CAGTCACTCG CTCCTTCGCC TTC TC CGGG 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

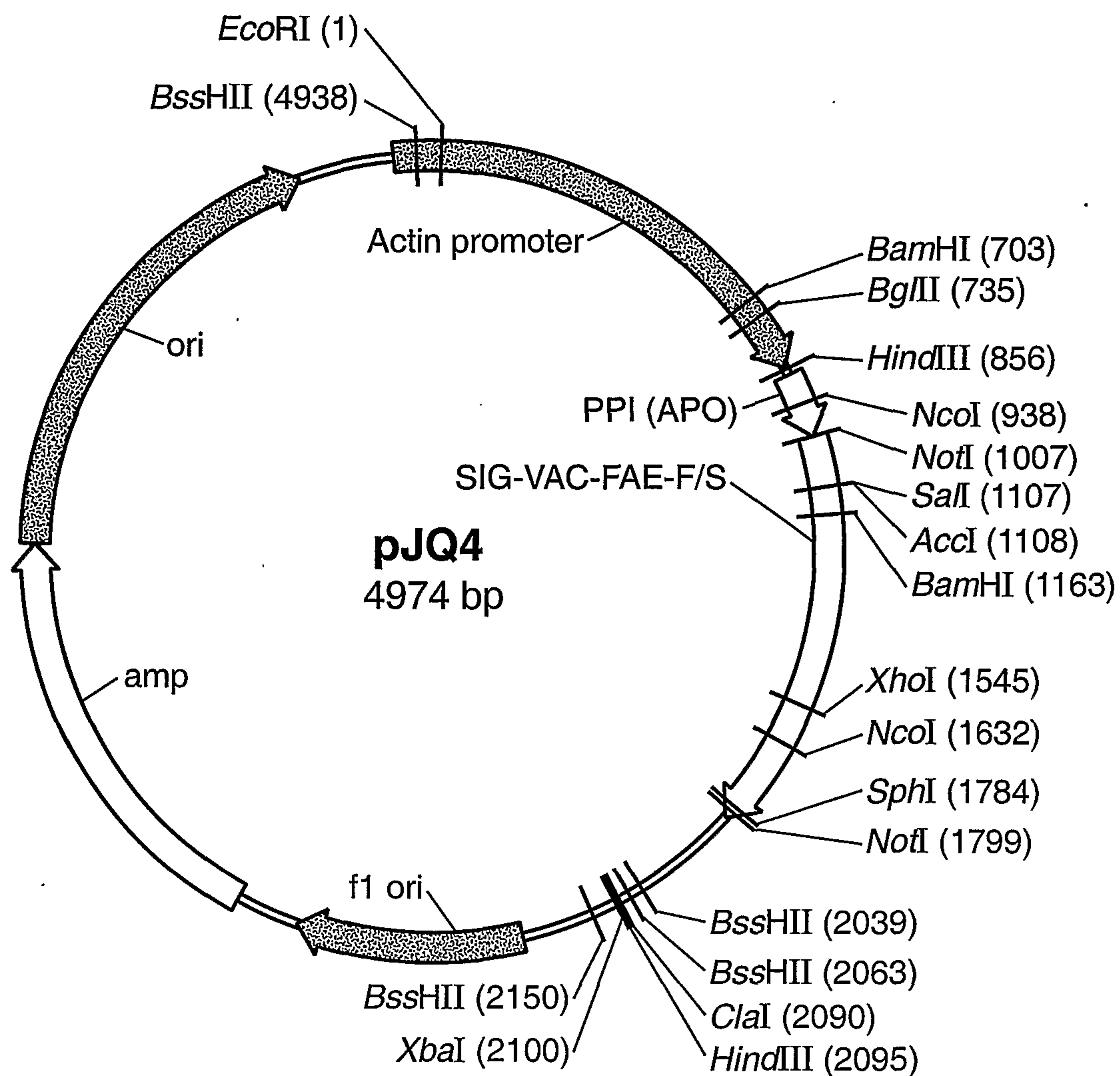
4131 CTTGCCCCCG TTGCAGCGAT GGGTATTTT TCTAGTAAA TAAAAGATAA ACTTAGACTC AAAACATTTA  
GAACGGGGC AACGTCGCTA CCCATAAAA AGATCATTTT ATTTTCTATT TGAATCTGAG TTTTGTAAT

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

FIG.\_ 43G

|       |             |            |            |            |            |             |             |
|-------|-------------|------------|------------|------------|------------|-------------|-------------|
| 4271  | ACCCAACCCA  | ACCCACCCCA | GTGCAGCCAA | CTGGCAAATA | GTCTCCACCC | CGGCACTAT   | CACCGTGAGT  |
|       | TGGGTGGGT   | TGGTGGGT   | CACGTCGGT  | GACCGTTTAT | CAGAGGTGG  | GGCCGTGATA  | GTGGCACTCA  |
| 4341  | TGTCCGCACC  | ACCGCACGTC | TCGCAGCCAA | AAAAAAAAAA | AGAAAGAAAA | AAAAAGAAAA  | GAAAAACAGC  |
|       | ACAGGCGTGG  | TGGCGTGCAG | AGCGTCGGT  | TTTTTTT    | TCCTTCTTT  | TTTTCTTTT   | CTTTTGTG    |
| 4411  | AGGTGGGTCC  | GGGTCGTGGG | GGCCGGAAAA | CGAGGAGGA  | TCGCGAGCAG | CGACGAGGCC  | CGGCCCTCCC  |
|       | TCCACCCAGG  | CCAGCACCC  | CCGGCTTTT  | CGCTCCTCT  | AGCGCTCGTC | GCTGCTCCGG  | GCCGGGAGGG  |
| 4481  | TCCGCTTCCA  | AAGAAACGCC | CCCATCGCC  | ACTATATACA | TACCCCCCC  | TCTCCTCCA   | TCCCCCAAC   |
|       | AGCGGAAGGT  | TTCCTTGCGG | GGGTAGCGG  | TGATATATGT | ATGGGGGGG  | AGAGGAGGT   | AGGGGGT     |
| 4551  | CCTACCACCA  | CCACCACCAC | CACCTCCTC  | CCCCTCGCTG | CCGGACGACG | AGCTCCTCCC  | CCCTCCCCCT  |
|       | GGAATGGTGGT | GGTGGTGGTG | GTGGAGGAGG | GGGAGCGAC  | GGCTGCTGC  | TCGAGGAGGG  | GGGAGGGGA   |
| 4621  | CCGCCGCCGC  | CGGTAACCAC | CCCGCCCTC  | TCTCTTTCT  | TTCCTCGTTT | TTTTTTTCGT  | CTCGGTCCTG  |
|       | GGCGCGGCG   | GCCATGGTG  | GGCGGGGAG  | AGGAGAAAG  | AAGAGGCAAA | AAAAAAGCA   | GAGCCAGAGC  |
| 4691  | ATCTTTGGCC  | TTGGTAGTTT | GGTGGGCGA  | GAGCGCTTC  | GTGCCCAGA  | TCGGTGCGCG  | GGAGGGCGG   |
|       | TAGAAACCGG  | AACCATCAAA | CCCACCCGT  | CTCGCCGAAG | CAGCGGTCT  | AGCCACGCGC  | CCTCCCCGCC  |
| BamHI |             |            |            |            |            |             |             |
|       | ~~~~~       |            |            |            |            |             | BglII       |
| 4761  | GATCTCGCGG  | CTGGCGTCTC | CGGGCGTGAG | TCGGCCCCGA | TCCTCGCGGG | GAATGGGGCT  | CTCGGATGTA  |
|       | CTAGAGCGCC  | GACCGCAGAG | GCCCGCACTC | AGCCGGGCT  | AGGAGCGCCC | CTTACCCCGA  | GAGCCTACAT  |
| BglII |             |            |            |            |            |             |             |
|       | ~~~~~       |            |            |            |            |             |             |
| 4831  | GATCTTCTTT  | CTTCTTCTT  | TTTGTGGTAG | AATTGAATC  | CCTCAGCAT  | GTTCAATCGGT | AGTTTTTCTT  |
|       | CTAGAAGAAA  | GAAAGAAAG  | AAACACCATC | TTAAACTTAG | GGAGTCGTAA | CAAGTAGCCA  | TCAAAAAAGAA |
| 4901  | TTCAATGATTT | GTGACAAATG | CAGCCTCGTG | CGGAGCTTTT | TTGTAGGTAG |             |             |
|       | AAGTACTAAA  | CACGTTTTAC | 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  | TAGTTTGT   | AAATGTTTTT  | GTTGGGATT   | TCAGGATTTC  | GGTTTCACG   |  |  |
| 141   | TATGCACGAT  | CCATAGCAAG | CCCAGCCCAA | CCAACCCAC   | 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  | CCCTCTCTCT | CCCATCCCC  | CAACCCCTACC | ACCACCACTC  | CCACCACTC   | CTCCCCCTC   |  |  |
|       | ATGTATGGG   | GGGAGAGGA  | GGTAGGGG   | GTTGGGATGG  | TGGTGGTGGT  | GGTGGTGGAG  | GAGGGGGAG   |  |  |
| 491   | GCTGCCGGAC  | GACGAGCTCC | TCCCCCTCTC | CCCTCCGGCCG | CCGCCGGTAA  | CCACCCCGCC  | CCTCTCCTCT  |  |  |
|       | CGACGGCCTG  | CTGCTCGAGG | AGGGGGGAGG | GGAGGCCGC   | GGCGGCCATT  | GGTGGGGCGG  | GGAGAGGAGA  |  |  |
| 561   | TTCTTTCTCC  | GTTTTTTTTT | TCGTCTCGGT | CTCGATCTTT  | GGCCTTGGTA  | GTTTGGGTGG  | GCAGAGCGGG  |  |  |
|       | AAGAAAGAGG  | CAAAAAAAA  | AGCAGAGCCA | GAGCTAGAAA  | CCGGAACCAT  | CAAACCCACC  | CGCTCTCGCC  |  |  |
| 631   | CTTCGTCCG   | CAGATCGGTG | CGCGGGAGGG | GGGGATCTC   | GGGCTGGCG   | TCTCCGGGCG  | TGAGTCGGCC  |  |  |
|       | GAAGCAGCGG  | GTCTAGCCAC | CGCCCTCTCC | 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<br>~~~~~               |            |             |             |             |            |             |
| 841  | TTTTTTGTAG                     | GTAGAAGCTT | ACMATGGMCG  | TGCACAAGGA  | GGTSAACTTC  | GTSGCCTACC | TCCTGATCGT  |
|      | AAAAAACATC                     | CATCTTCGAA | TGKTACCKGC  | ACGTGTTTCT  | CCASTTGAAG  | CASCGGATGG | AGGACTAGCA  |
|      | NcoI<br>~~~~~                  |            |             |             |             |            |             |
| 911  | SCTCGGCCCTC                    | CTCTTGCTCG | TSTCCGCCAT  | GGAGCAGTG   | GACGCCAAGG  | CCTGCACCK  | CGAGTGCGGC  |
|      | SGAGCCGGAG                     | GAGAACGAGC | ASAGCGGTA   | CCTCGTGCAC  | CTGCGGTTCC  | GGACGTGGGM | GCTCACGCCG  |
|      | NotI<br>~~~~~                  |            |             |             |             |            |             |
| 981  | AACCTCGGCT                     | TCGGCATCTG | CCCGCGGCC   | GCCTCCACGC  | AGGGCATCTC  | CGAAGACCTC | TACAGCCGTT  |
|      | TTGGAGCCGA                     | AGCCGTAGAC | GGCCGCCCG   | CGGAGGTGCG  | TCCCCTAGAG  | GCTTCTGGAG | ATGTCGGCAA  |
|      | Sall<br>~~~~~<br>AccI<br>~~~~~ |            |             |             |             |            |             |
| 1051 | TAGTCGAAAT                     | GGCCACTATC | TCCCAAGCTG  | CCTACGCCGA  | CCTGTGCAAC  | ATTCCGTCGA | CTATTATCAA  |
|      | ATCAGCTTTA                     | CCGGTGATAG | AGGGTTCGAC  | GGATGCCGGCT | GGACACGTTG  | TAAGGCAGCT | GATAATAGTT  |
|      | BamHI<br>~~~~~                 |            |             |             |             |            |             |
| 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 TTCATGT | 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 | CTTCGCAACT  | 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 TGTCGCCCTT ATTCCCCTTT  
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 GAATGCCCTAC CGTACTGTCA TTCCTCTAAT ACGTCACGAC GGTA TTGGTA CTCAC TATTG

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

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 AACGTCCCTGG 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 CTTAACGTGA GTTTTCGTTC CACTGAGCGT CAGACCCCGT AGAAAAGATC  
ATTAGAGTAC TGGTTTTAGG GAATTGCACT CAAAAGCAAG GTGACTCGCA GTCTGGGCA TCTTTTCTAG

3921 AAAGGATCTT CTTGAGATCC TTTTTTCTG CGCGTAATCT GCTGCTTGCA AACAAAAAA CCACCGCTAC  
TTTCCCTAGAA GAACTCTAGG AAAAAAGAC GCGCATTAGA CGACGAACGT TTGTTTTTTT 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 AGTTCTTGAG ACATCGTGGC

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

4201 GGTTTGGACTC AAGACGATAG TTACCGGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGTTT CGTGCACACA  
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  
TAAAACACACT ACGAGCAGTC CCCCGCCTC GGATACCTTT TTGCGGTCGT TGCGCCGGA AAATGCCAAG

4551 CTGGCCTTT GCTGGCCTTT TGCTCACATG TTCCTTCCTG 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 TTCCCAGCTG GAAAGCGGC AGTGAGCGCA ACGCAATTA TGTGAGTTAG CTCACTCAT  
GTGCTGTCCA AAGGGCTGAC CTTTCGCCCG TCACTCGCGT TGCGTTAATT ACACTCAATC GAGTGAGTAA

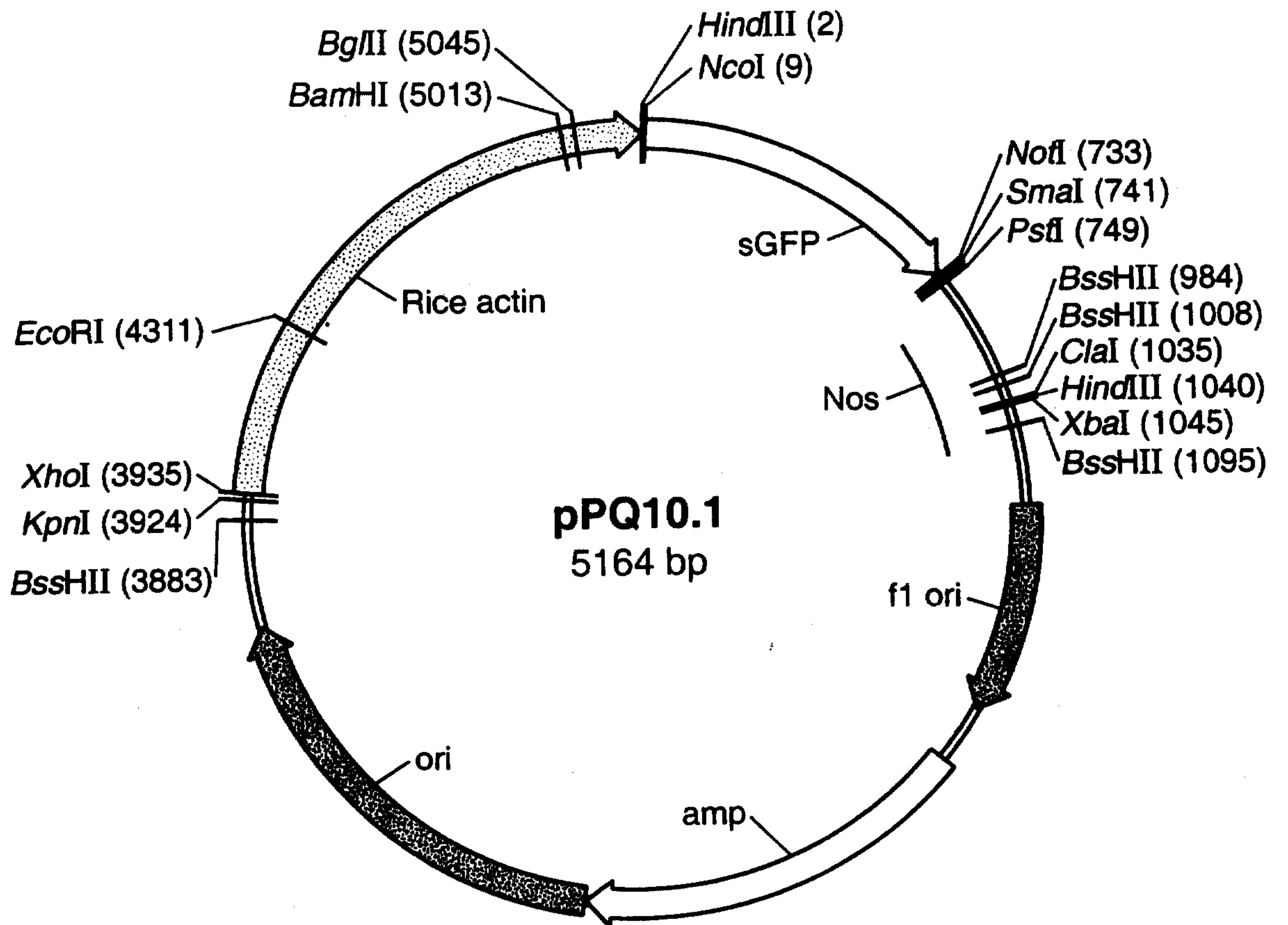
4831 AGGCACCCCA GGCATTACAC TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTGAGCG GATAACAAAT  
TCCGTGGGGT CCGAAATGTG AAATACGAAG GCCGAGCATA CAACACACCT TAACACTCGC CTATTGTTAA

BssHII  
~~~~~

4901 TCACACAGGA AACAGCTATG ACCATGATTA CGCCAAGCGC GCAATTAAAC CTCACTAAAG GGAACAAAAG
AGTGTGTCCCT TTGTCGATAC TGGTACTAAT GCGGTTGCGC CGTTAATTGG GAGTGATTTC CCTTGTTTC

EcoR
4971 CTGG
GACC

FIG.- 44H

**FIG. 45A**

HindIII NcoI
~~~~~ ~~~~~~  
1 AAGCTTACCA TGGTGAGCAA GGGCGAGGAG CTGTTACCG GGGTGGTGCC CATCCTGGTC 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 GGCGTGCACT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCAGCACTT CTTCAAGTCC GCCATGCCCG  
CCGCACGTCA CGAAGTCGGC GATGGGGCTG GTGTACTTCG TCGTGCTGAA GAAGTTCAGG CCGTACGGGC  
281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC AAGACCCGCG CCGAGGTGAA  
TTCCGATGCA GGTCTTCGCG TGGTAGAAGA AGTTCCTGCT GCCGTGATG TTCTGGGCGC GGCTCCACTT  
351 GTTCGAGGGC GACACCCCTGG TGAACCGCAT CGAGCTGAAG GGCATCGACT TCAAGGAGGA CCGCAACATC  
CAAGCTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA AGTTCCCTCCT GCCGTGTGTAG  
421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT GGCCGACAAG CAGAAGAACG  
GACCCCGTGT TCGACCTCAT GTTGATGTTG TCGGTGTTGC AGATATAGTA CCGGCTGTTC GTCTTCTTGC  
491 GCATCAAGGT GAACTTCAAG ATCCGCCACA ACATCGAGGA CGGCAGCGTG CAGCTCGCCG ACCACTACCA  
CGTAGTTCCA CTTGAAGTTC TAGGCGGTGT TGTAGCTCCT GCCGTGCGAC GTCGAGCGGC TGGTGATGGT  
561 GCAGAACACC CCATCGGCG ACGGCCCGGT 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 TTCTTACTCG | |
| 771 | TGTAAGAAG CAGATCGTTC AACATTTGG CAATAAAGTT TCTTAAGATT GAATCCTGTT GCCGGTCTTG
ACATTTCTTC GTCTAGCAAG TTTGTAAACC GTTATTTCAG AGAATTCTAA CTTAGGACAA CGGCCAGAAC | |
| 841 | CGATGATTAT CATATAATT CTGTTGAATT ACGTTAAGCA TGTAAATAAT AACATGTAAT GCATGACGTT
GCTACTAATA GTATATTAA GACAACTTAA 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
~~~~~ |
| 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 | GGCGTAATAG 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
GCGCTTAAAA 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 TTCCGTGTGCG CCTTATTCC CTTTTTGGC
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 GAAAATTCA AGACGATACA CCGCGCCATA 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 TGTCATTCTC TTAATACGTC ACGACGGTAT TGGTACTCAC TATTGTGACG

2241 GGCCAACTTA CTCTTGACAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT TTTTGCACAA CATGGGGGAT
CCGGTTGAAT GAAGACTGTT GCTAGCCTCC TGGCTTCCTC GATGGCGAA AAAACGTGTT GTACCCCTTA

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

2381 CGATGCCCTGT AGCAATGGCA ACAACGTTGC GCAAACCTATT 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 GCACTGGGGC
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 TAAATTAAAA 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

FIG. 45F

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

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

3921 TACCGGGCCC CCCCTCGAGG TCATTCAAT GCTTGAGAAG AGAGTCGGGA TAGTCCAAA TAAACAAAG
ATGGCCCCGG GGGGAGCTCC AGTAAAGTATA CGAACTCTTC TCTCAGCCCT ATCAGGTTT ATTTTGTTTC

3991 GTAAGATTAC CTGGTCAAAA GTGAAACAT CAGTTAAAG GTGGTATAAG TAAATATCG GTAATAAAG
CATCTAATG GACCAGTTT CACTTTTGTA GTCAAATTTT CACCATAATC ATTTATATAGC CATTAATTTTC

4061 GTGGCCCCAA GTGAAATTTA CTCTTTTCTA CTATTATAAA AATTGAGGAT GTTTTGTCGG TACTTTGATA
CACCGGGTTT CACTTTAAAT GAGAAAGAT GATAATATT TTAACCTCTA CAAACACAGCC ATGAAACTAT

4131 CGTCATTTT GTATGAATTG GTTTTAAAG TTATTCGCGA TTGGGAAATG CATATCTGTA TTTGAGTCGG
GCAGTAAAA CATACTTAAC CAAAATTCA AATAAGCGCT AAACCTTTAC GTATAGACAT AAACCTCAGCC

4201 TTTTAAAGTT CGTTGCTTT GTAAATACAG AGGATTGT ATAAGAAATA TCTTTAAAA ACCCATATGC
AAAAATTCAA GCAACGAAAA CATTTATGTC TCCCTAAACA TATCTTTAT AGAAATTTT TGGGTATACG

4271 TAATTGACA TAAATTTTGA GAAAAATA TATCAGCGG AATCCACAA TGAACAATA TAAGATTAA
ATTAAACTGT ATTAAAACT CTTTTATAT ATAAGTCCGC TTAAGGTGTT ACTTGTTAT ATTCTAATT

4341 ATAGCTTGCC CCCGTTGCAG CGATGGGTAT TTTTCTAGT AAAATAAAG ATAACTTAG ACTCAAAACA
TATCGAACGG GGGCAACGTC GCTACCCATA AAAAGATCA TTTTATTTT TATTGAATC TGAGTTTGT

4411 TTACAAAA CAACCCCTAA AGTCCTAAG CCCAAAGTC TATGCAGAT CCATAGCAAG CCCAGCCCA
AAATGTTTT GTTGGGATT TCAGGATTTC GGGTTTCAG ATACGTGCTA GGATCGTTC GGTCTGGGT

KpnI

BssHII

XhoI

EcoRI

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 AAAAAAAA
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 GGGGAGAGA 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
~~~~~

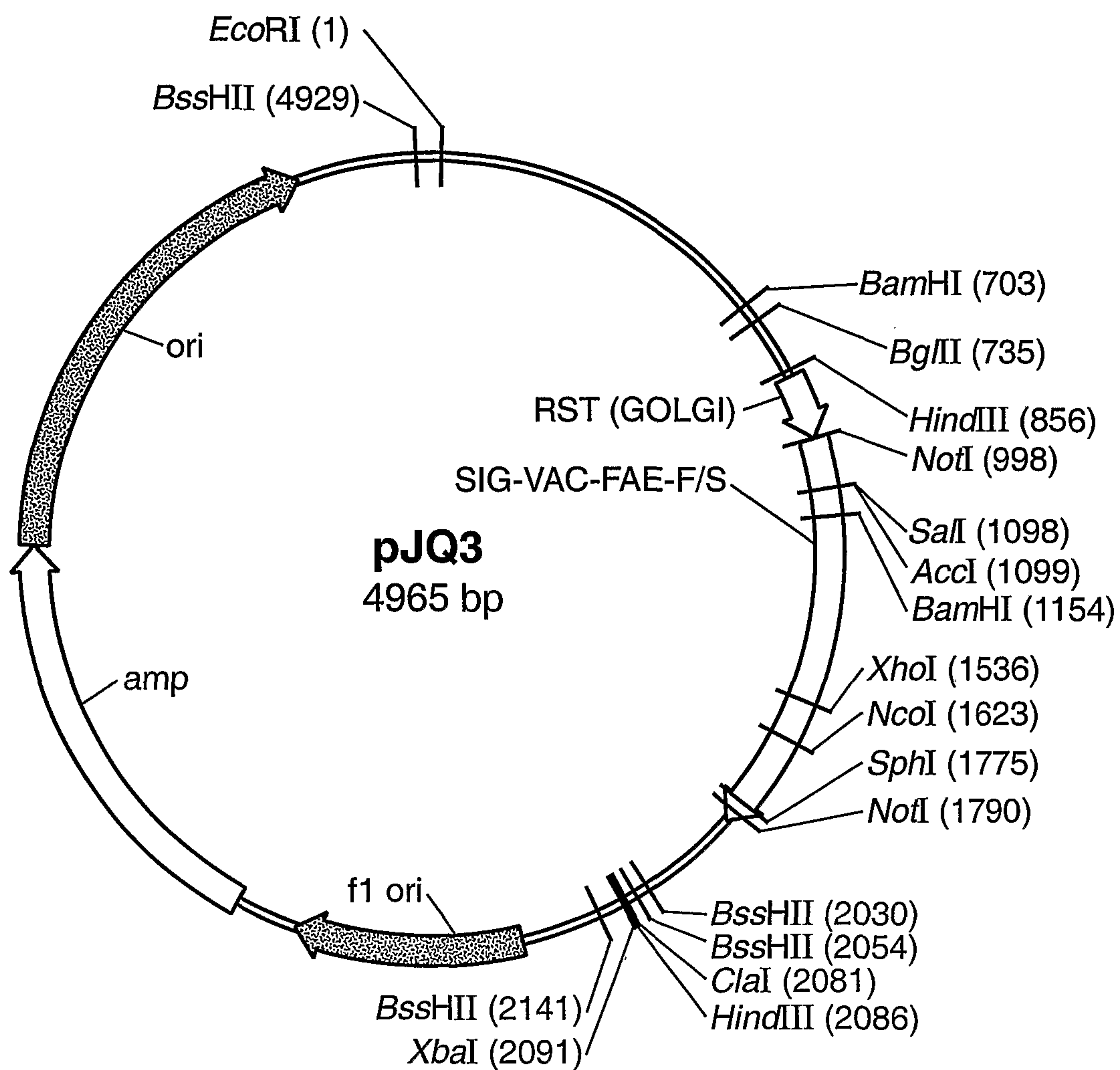
4971 GCGGGATCTC GCGGCTGGCG TCTCCGGGCG TGAGTCGGCC CGGGAATGG GGTCTCGGA  
CGCCCTAGAG CGCCGACCGC AGAGGCCCGC ACTCAGCCGG GCCTAGGAGC GCCCTTACC CCGAGAGCCT

BglII  
~~~~~

5041 TGTAGATCTT CTTTCTTTCT TCTTTTGTG GTAGAAATTG AATCCCTCAG CATGTTCAT CGGTAGTTT
ACATCTAGAA GAAAGAAAGA AGAAAAACAC CATCTTAAC TTAGGGAGTC GTAACAAGTA GCCATCAAA

5111 TCTTTTCATG ATTTGTGACA AATGCAGCCT CGTGGGAGC 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 CCCCTCTCTT CCATCCCCC CAACCTACC ACCACACCA CCACACCTC CTCCCCCCTC
ATGTATGGG GGGAGAGGA GGTAGGGG GTTGGGATGG TGGTGTGGT GGTGTGGAG 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 CAAAAAATAA AGCAGAGCCA GAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC |
| 631 | CTTCGTCCG CAGATCGGTG CGGGGAGGG GCGGATCTC GCGCTGGCG TCTCCGGCG TGAGTCGGCC
GAAGCAGCG GTCTAGCCAC GCGCCCTCCC CGCCGACCGC AGAGGCCCGC ACTCAGCCCG |
| BamHI | |
| ~~~~~ | |
| 701 | CGGATCCTCG CGGGGAATGG GGCTCTCGGA TGATGATCTT CTTTCTTTCT TCTTTTGTG GTAGAATTG
GCCTAGGAG GCCCCTTACC 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       |             |            |            |            |             |             |   |   |  |
|      | CGCAAACTA    | 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  | GGTCGACCGC  |   |   |  |
| 2241 | BssHII       |             |            |            |            |             |             |   |   |  |
|      | TAATAGCGAA   | GAGCCCGCA   | CCGATCGCCC | TTCCCAACAG | TTGCGCAGCC | TGAATGGCGA  | ATGGGACGCG  |   |   |  |
|      | ATTATCGCTT   | CTCCGGGCGT  | GGCTAGCGGG | AAGGTTGTC  | AACGCGTCGG | ACTTACCGCT  | TACCC TGCGC |   |   |  |
| 2311 | BssHII       |             |            |            |            |             |             |   |   |  |
|      | CCCTGTAGCG   | GCGCATTAA   | CGCGGCGGGT | GTGGTGGTTA | CGCGCAGCGT | GACCGCTACA  | CTTGCCAGCG  |   |   |  |
|      | GGGACATCGC   | CGCGTAATTC  | GCGCCGCCCA | CACCACCAAT | GCGCGTCGCA | CTGGCGATGT  | GAACGGTCGC  |   |   |  |
| 2381 | BssHII       |             |            |            |            |             |             |   |   |  |
|      | CCCTAGCGCC   | CGCTCCTTTC  | GCTTTCCTTC | CTTCCCTTCT | CGCCACGTTT | GCCGGCTTTC  | CCCGTCAAGC  |   |   |  |
|      | GGGATCGCGG   | GCGAGGAAAG  | CGAAGAAGG  | GAAGGAAAGA | GCGGTGCAAG | CGGCCGAAAG  | GGCAGTTTCG  |   |   |  |
| 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 TAAAACGGCT AAAGCCGGAT AACCAATT TT TACTCGACT AAATTGTTTT TAAATTGCGC

2731 AATTTTAACA AAATATTAC GCTTACAATT TAGGTGGCAC TTTTCGGGA AATGTGCGG GAACCCCTAT  
TTAAAAATTGT TTTATAATTG CGAATGTTAA ATCCACCGTG AAAAGCCCCT TTACACGCGC CTTGGGGATA

2801 TTGTTTATTT TTCTAAATAC ATTCAAAATAT 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 TTTGCCCTTCC TGTTTTGTGCT CACCAGAAA 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 TTTCCAAATGA 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  
AGAA TGCCTA 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 CTTGGCCCTCG ACTTACTTCG GTATGGTTTG CTGCTCGCAC TGTGGTGCTA

FIG. 46F

3431 GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAACT GGCGAACTAC TTACTCTAGC TTCCCGGCAA  
CGGACATCGT TACCGTTGTT GCAACGCGTT TGATAATTGA CCGCTTGATG AATGAGATCG AAGGGCCGTT

3501 CAATTAAATAG ACTGGATGGA GGCGGATAAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGCTGGCT  
GTTAATTATC TGACCTACCT CCGCCTATT CAACGTCCTG GTGAAGACGC GAGCCGGGA GGCAGACCGA

3571 GGTTTATTGC TGATAAATCT GGAGCCGGTG AGCGTGGTC TCGCGGTATC ATTGCAGCAC TGGGGCCAGA  
CCAAATAACG ACTATTTAGA CCTCGGCCAC TCGCACCCAG AGCGCCATAG TAACGTCGTG ACCCCGGTCT

3641 TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA  
ACCAATCGGG AGGCATAGC ATCAATAGAT GTGCTGCCCC TCAGTCCGTT GATACCTACT TGC'TTTATCT

3711 CAGATCGCTG AGATAGGTGC CTCAC'TGATT AAGCAT'TGGT AACTGTCAGA CCAAGTTTAC TCATATATAC  
GTCTAGCGAC TCTATCCACG GAGTGACTAA TTTCGTAACCA TTGACAGTCT GGTTCAAATG AGTATATATG

3781 TTTAGATTGA TTTAAAACTT CATTTTAAAT TTAAAAAGGAT CTAGGTGAAG ATCCTTTTIG ATAACTCAT  
AAATCTAACT AAATTTTGAA GTAAAAATTA AATTTTCCTA GATCCACTTC TAGGAAAAAC TATTAGAGTA

3851 GACCAAATC CCTAACGTG AGTTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT  
CTGGTTT'TAG GGAATTGCAC TCAAAAGCAA GTGACTCGC AGTCTGGGC ATCTTTTCTA GTTTCCTAGA

3921 TCTTGAGATC CTTT'TT'TCT GCGCGTAATC TGCTGCTTGC AAACAAAAA ACCACCGCTA CCAGCGTGG  
AGAACTCTAG GAAAAAAGA CGCGCATTAG ACGACGAACG TTTGT'TTTT TGGTGGCGAT GGTGCGCAC

3991 TTTGT'TTGCC GGATCAAGAG CTACCAACTC TTTT'TCCGAA GGTAAC'TGGC TTCAGCAGAG CGCAGATACC  
AAACAAACGG C'TAGTTCTC GATGGTTGAG AAAAAGGCTT CCATTGACCG AAGTCGTCTC GCGTCTATGG

4061 AAATACTGTC CTTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAACT CTGTAGCACC GCTACATAC  
TTTATGACAG GAAGATCACA TCGGCATCAA TCCGGTGGTG AAGTTC'TTGA GACATCGTGG CGGATGTATG

4131 CTCGCTCTGC TAA'TCCTGTT ACCAGTGGCT GTGCCAGTG GCGATAAGTC GTGTCTTACC GGGTTGGACT  
GAGCGAGACG ATTAGGACAA TGGTCACCGA CGACGGTCAC CGCTATT'CAG CACAGAAATGG CCCAACCTGA

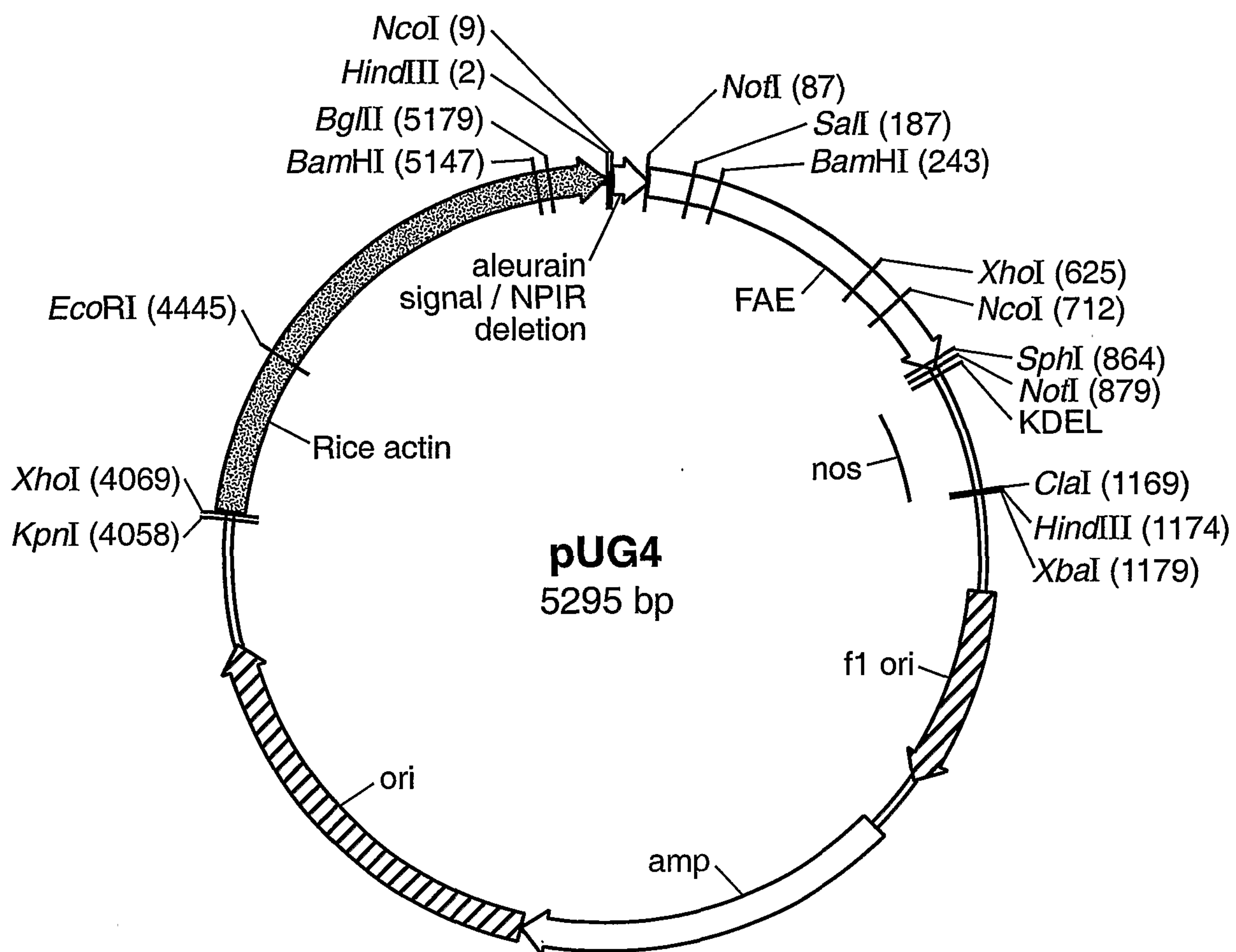
4201 CAAGACGATA GTTACCGGAT AAGGCGCAGC GGTGCGGCTG AACGGGGGT TCGTGCACAC AGCCAGCTT  
GTTCTGCTAT CAA'TGGCCTA TTCCGCGTCG CCAGCCCGAC TTGCCCCCCA AGCACGTGTG TCGGGTCGAA

FIG. - 46G

|        |              |             |             |             |             |              |              |
|--------|--------------|-------------|-------------|-------------|-------------|--------------|--------------|
| 4271   | GGAGCGAACG   | ACCTACACCG  | AAC TGAGATA | CCTACAGCGT  | GAGCTATGAG  | AAAGGCCAC    | GCTTCCCGAA   |
|        | CCTCGCTTGC   | TGGATGTGGC  | TTGACTCTAT  | GGA TGTCGCA | CTCGATACTC  | TTTCGCGGTG   | CGAAGGGCTT   |
| 4341   | GGGAGAAAGG   | CGGACAGGTA  | TCCGGTAAGC  | GGCAGGGTCG  | GAACAGGAGA  | GCGCACGAGG   | GAGCTTCCAG   |
|        | CCCTCTTTCC   | GCCTGTC CAT | AGGCCATT CG | CCGTCCCAGC  | CTTGTCCTCT  | CGCGTGCTCC   | CTCGAAGGTC   |
| 4411   | GGGGAAACGC   | CTGGTATCTT  | TATAGTCCTG  | TCGGGTTTCG  | CCACCCTCTGA | CTTGAGCGTC   | GATTTTGTG    |
|        | CCCC'TTTCG   | GACCATAGAA  | ATATCAGGAC  | AGCCCAAAGC  | GGTGGAGACT  | GAACTCGCAG   | CTAAAAAACAC  |
| 4481   | ATGCTCGTCA   | GGGGGGCGGA  | GCCTATGGAA  | AAACGCCAGC  | AACGCGGCCT  | TTTTACGGTT   | CCTGGCCTTT   |
|        | TACGAGCAGT   | CCCCCCGCCT  | CGGATACCTT  | TTTGCGGT CG | TTGCGCCGGA  | AAAATGCCAA   | GGACCGGAAA   |
| 4551   | TGCTGGCCTT   | TTGCTCACAT  | GTTCTTTCC T | GCGTTATCCC  | CTGATTCTGT  | GGATAAACC GT | ATTACCGCCT   |
|        | ACGACCGGAA   | AACGAGTGTA  | CAAGAAAGGA  | CGCAATAGGG  | GACTAAGACA  | CCTATTGGCA   | TAATGGCGGA   |
| 4621   | TTGAGTGAGC   | TGATACCGCT  | CGCCGCAGCC  | GAACGACCGA  | GCGCAGCGAG  | TCAGTGAGCG   | AGGAAGCGGA   |
|        | AAC TCAC TCG | ACTATGGCGA  | GCGGCGTCGG  | CTTGCTGGCT  | CGCGTCGCTC  | AGTCAC TCGC  | TCC'TTCGCC T |
| 4691   | AGAGCGCCCA   | ATACGCAAAC  | CGCCTCTCCC  | CGCGCGTTGG  | CCGATTCA TT | AATGCAGCTG   | GCACGACAGG   |
|        | TCTCGCGGGT   | TATGCGTTTG  | GCGGAGAGGG  | GCGCGCAA CC | GGCTAAGTAA  | TTACGT CGAC  | CGTGCTGTCC   |
| 4761   | TTTCCCGACT   | GGAAAGCGGG  | CAGTGAGCGC  | AACGCAATTA  | ATGTGAGTTA  | GCTCACTCAT   | TAGGCACCCC   |
|        | AAAGGGCTGA   | CCTTTCGCC C | GTCAC TCGCG | TTGCGTTAAT  | TACAC TCAAT | CGAGTGAGTA   | ATCCGTGGGG   |
| 4831   | AGGC'TTTACA  | CTTTATGCTT  | CCGGCTCGTA  | TGTTGTGTGG  | AATTGTGAGC  | GGATAACAAT   | TTCACACAGG   |
|        | TCCGAAATGT   | GAAATACGAA  | GGCCGAGCAT  | ACAACACACC  | TTAAACACTCG | CCTATTGTTA   | AAGTGTGTCC   |
| BssHII |              |             |             |             |             |              |              |
| ~~~~~  |              |             |             |             |             |              |              |
| 4901   | AAACAGCTAT   | GACCATGATT  | ACGCCAAGCG  | CGCAATTAA C | CCTCACTAAA  | GGGAACAAAA   | GCTGG        |
|        | TTTGTGCGATA  | CTGGTACTAA  | TGCGGTTTCG  | GCGTTAATTG  | GGAGTGATTT  | CCCTTGTTT T  | CGACC        |

EcoRI

FIG.\_ 46H

**FIG.\_47A**

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 CTCTGGCGC TCGCCGTGCT GGCCACGGC GCCGTGCGCG  
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
Sall
~~~~~  
A T I S Q A A Y A D L C N I P S T I I K G E K  
141 GGCACTATC 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 ATTTACAATT 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 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
~~~~~  
F G E P R S G N Q A F A S Y M N D A F Q A S S  
561 CTCGGCGGAA 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 | CCTTTTGTAT | 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 | GTCTTACC GG | GTTGGACTCA |
| 3291 | AGACGATAGT | TACCGGATAA | GGCGAGCGG | TCGGGCTGAA | CGGGGGGTTT | GTGCACACAG | CCAGCTTGG |
| 3361 | AGCGAACGAC | CTACACCGAA | CTGAGATACC | TACAGCGTGA | GCTATGAGAA | AGCGCCACGC | TTCCCGAAGG |
| 3431 | GAGAAAGGCG | GACAGGTATC | CGGTAAGCGG | CAGGGTCGGA | ACAGGAGAGC | GCACGAGGGA | GCTTCCAGGG |
| 3501 | GGAAACGCCT | GGTATCTTTA | TAGTCCCTGTC | GGGTTTCGCC | ACCTCTGACT | TGAGCGTCCA | TTTTTTGTGAT |
| 3571 | GCTCGTCAGG | GGGCGGAGC | CTATGGAAA | ACGCCAGCAA | CGCGGCCTTT | TTACGGTTC | TGGCCCTTTG |
| 3641 | CTGGCCCTTT | GCTCACATGT | TCTTTCCCTGC | GTTATCCCTT | GATTCTGTGG | ATAACCGTAT | TACCGCCCTT |
| 3711 | GAGTGAGCTG | ATACCGCTCG | CCGAGCCGA | ACGACCGAGC | GCAGCGAGTC | AGTGAGCGAG | GAAAGCGGAA |
| 3781 | AGCGCCCAAT | ACGCAAAACCG | CCTCTCCCGG | 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 | GAGGTCATTC | ATATGCTTGA | GAAGAGAGTC | GGGATAGTCC | AAAATAAAAC | AAAGGTAAGA |
| 4131 | TTACCTGGTC | AAAAGTGAAA | ACATCAGTTA | AAAGGTGGTA | TAAGTAAAT | ATCGGTAATA | AAAGGTGGCC |
| 4201 | CAAAGTGAAA | TTTACTCTTT | TCTACTATT | TAAAAATTGA | GATGTTT | TCGGTACTTT | GATACGTCA |
| 4271 | TTTTGTATGA | ATTGGTTTTT | AAGTTTATTC | GCATTTTGA | AATGCATATC | TGTATTTGAG | TCGGTTTTTA |
| 4341 | AGTTCGTTGC | TTTTGTAAAT | ACAGAGGGAT | TTGTATAAGA | ATAATCTTTA | 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 |

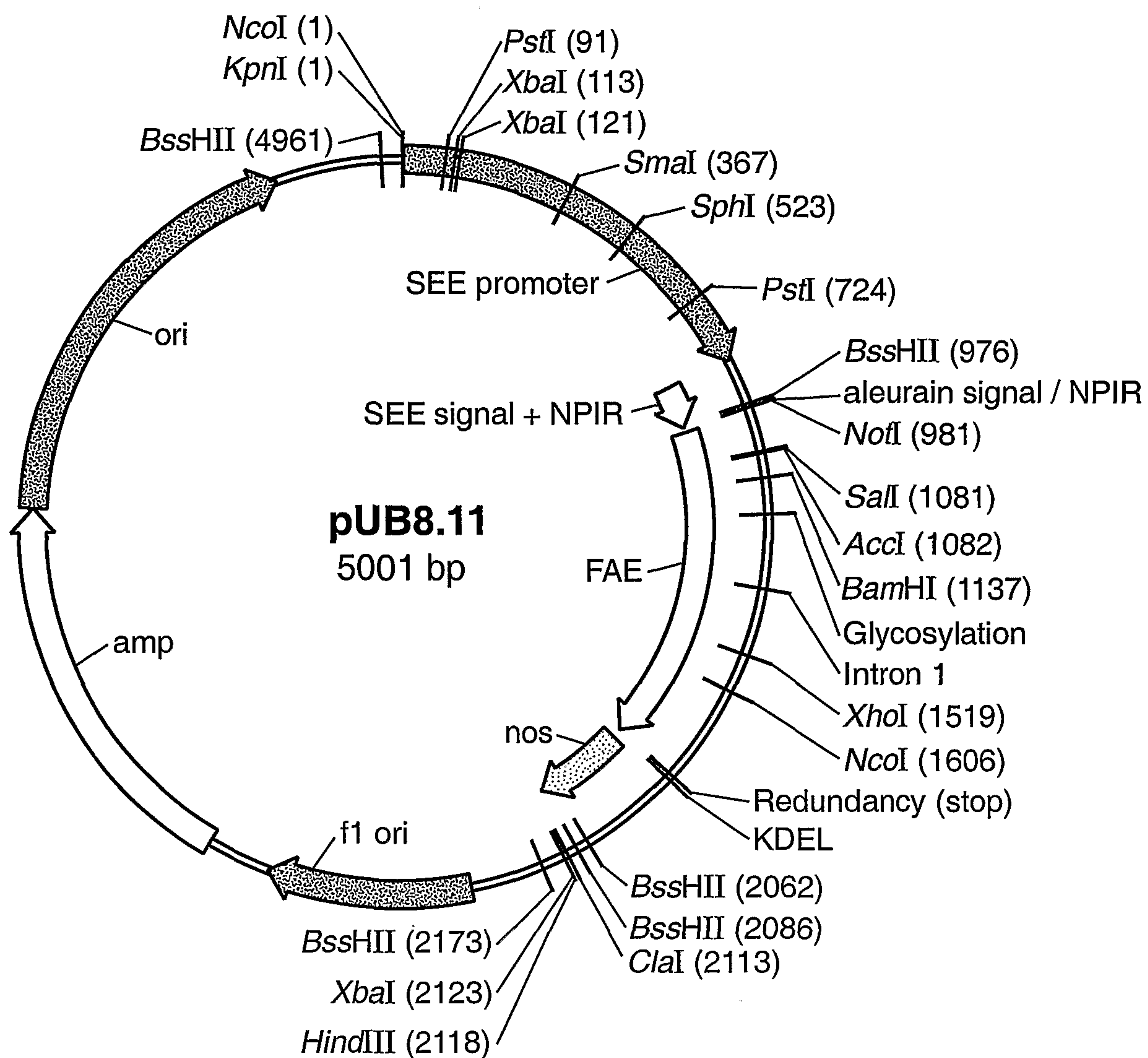
KpnI
~~

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  | CTCTTTCTTT | CTCCGTTT    | TTTTTCGTCT  | CGGTCTCGAT |
| 5041 | CTTTGGCCCT  | GGTAGTTTGG  | GTGGGCGAGA | GCGGCTTCGT | CGCCACAGATC | GGTGC GCGGG | AGGGGCGGGA |
|      |             |             |            | BamHI      |             |             | BglII      |
|      |             |             |            | ~~~~~      |             |             |            |
| 5111 | TCTCGCGGCT  | GGCGTCTCCG  | GGCGTGAGTC | GGCCCCGATC | CTCGCGGGGA  | ATGGGGCTCT  | CGGATGTAGA |
|      | BglII       |             |            |            |             |             |            |
|      | ~~~         |             |            |            |             |             |            |
| 5181 | TCTTCTTCT   | TTCTTCTTTT  | TGTGGTAGAA | TTTGAATCCC | TCAGCATTGT  | TCATCGGTAG  | TTTTTCTTTT |
| 5251 | CATGATTTGT  | GACAAATGCA  | GCCTCGTGCG | GAGCTTTTTT | GTAGC       |             |            |

FIG.- 47E

**FIG. 48A**

|       |                                                                                |       |       |
|-------|--------------------------------------------------------------------------------|-------|-------|
| NcoI  |                                                                                |       |       |
| ~~~~~ |                                                                                |       |       |
| KpnI  |                                                                                |       |       |
| ~     |                                                                                |       |       |
| 1     | CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG  |       |       |
|       | GTACCCGGGC CATATTAAATA CCTATAGAG TTCGTTTATT AGCTTTATAG TGGTAACCGA TGTTATAGAC   |       |       |
|       |                                                                                | PstI  | XbaI  |
|       | ~~~~~                                                                          | ~~~~~ | ~~~~~ |
| 71    | AGCTCCGAGT TCTGACTGCA GTC TGGA TGA CGCGTGTTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC |       |       |
|       | TCGAGGCTCA AGACTGACGT CAGACCTACT GCGCACAA CA TAGATCTTGA GATCTATCGT GTCGGTGTCG  |       |       |
| 141   | ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT   |       |       |
|       | TGGATGTCC T CACGCTGTGA ACACCTGACA TCATCACAAC CTCTGCCCTCG AGAAAGGATG GAGGACTGCA |       |       |
| 211   | TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC   |       |       |
|       | ACGGCGGCAA CAGGTAAGGT TGCCGTAGTG AGAGTTGGTT AGTGCGCGAG GGTGTGTTTA TAGCAGGGGG   |       |       |
| 281   | ATGTC TTGGC GGAGAGAGAG TACATACATG CTGTCGCGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC   |       |       |
|       | TACAGAACCG CCTCTCTCTC ATGTATGTAC GACAGCGCGG CAAAACAGA CTTAGAGCGA AGGTGACCGG    |       |       |
|       |                                                                                | SmaI  |       |
|       | ~~~~~                                                                          |       |       |
| 351   | AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG 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 AAATAAAAC    |       |       |
| 561   | TTGGAGTTGA ATGCTGATTT GTTG TGTA AA 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 GGTCAATGGTA GAGCGGAGCG AGGGGACGG CCTGCTGGGT

841 GTAAATAACT GTTGCCCACT CGCCGGCGAG ATGGCCACG GCCGCAATCCT CTTCTTGGCG CTCGCCGTCT
CATTTATGA CAACGGGTGA GCGCCGCTC TACCGGGTGC CGCGTAGGA GAAGAACCGC GAGCGGCAGA

BsSHII
~~~~~  
NotI  
~~

911 TGGCCACCGC CGGGTGGCC GCCGCATCNT TGGCGGACTC CAACCCGATC CGGCCCGTCA CCGAGCGCGC  
ACCGTGGCG GCGCACCCGG CCGCGTAGNA 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 | | |
| | GACTGTAATT 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 CCTACCCAGA 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

SUBSTITUTE SHEET (RULE 26)

FIG. 48E

2451 CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC GGGGGCTCCC TTTAGGGTTC CGATTTAGTG
GAGCGGTGCA AGCGGCCGAA AGGGGCAGTT CGAGATTTAG CCCCAGAGG AAATCCCAAG GCTAAATCAC

2521 CTTTACGGCA CCTCGACCCC AAAAAACTTG ATTAGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA
GAAATGCCGT GGAGCTGGGG TTTTTTTGAAC 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 AAATTTAAG CGAATTTTAA CAAAATATTA ACGCTTACAA TTTAGGTGGC
TTTTTACTCGA CTAAATTGTT TTTAAATTGC GCTTAAAAAT GTTTTATAAT TGCGAATGTT AAATCCACCG

2801 ACTTTTCGGG GAAATGTGG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT ACATTCAAAT ATGTATCCGC
TGAAAAGCCC CTTTACACGC GCCTTGGGA TAAACAATA AAAAGATTA TGTAAGTTTA TACATAGGCG

2871 TCATGAGACA ATAACCTGA 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 TTATAAGTTC TGCATGTGG
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 GATT TAAAC 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 GACGCGCAT 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 AGCACAGAAT GGCCCAACCT GAGTTC TGCT ATCAATGGCC 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 CCTCCCTTTG CGGACCATAG AAATATCAGG ACAGCCCAA

4481 CGCCACCTCT GACTTGAGCG TCGATTTTG TGATGCTCGT CAGGGGGCG GAGCCTATGG AAAACGCCA
GCGGTGGAGA CTGAACTCGC AGCTAAAAAC ACTACGAGCA GTCCCCCGC CTCGGATACC TTTTTCGGT

4551 GCAACGCGC CTTTTTACGG TTCCTGCCCT TTTGCTGGC TTTTGCTCAC ATGTTCTTTC CTGCGTTATC
CGTTGCGCG GAAAAATGCC AAGGACCGA AAACGACCG 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 CGGCAATTA  
CCTTAACACT CGCCTATTGT TAAAGTGTGT CCTTGTCTGA TACTGGTACT AATGCGGTTC GCGCGTTAAT

4971

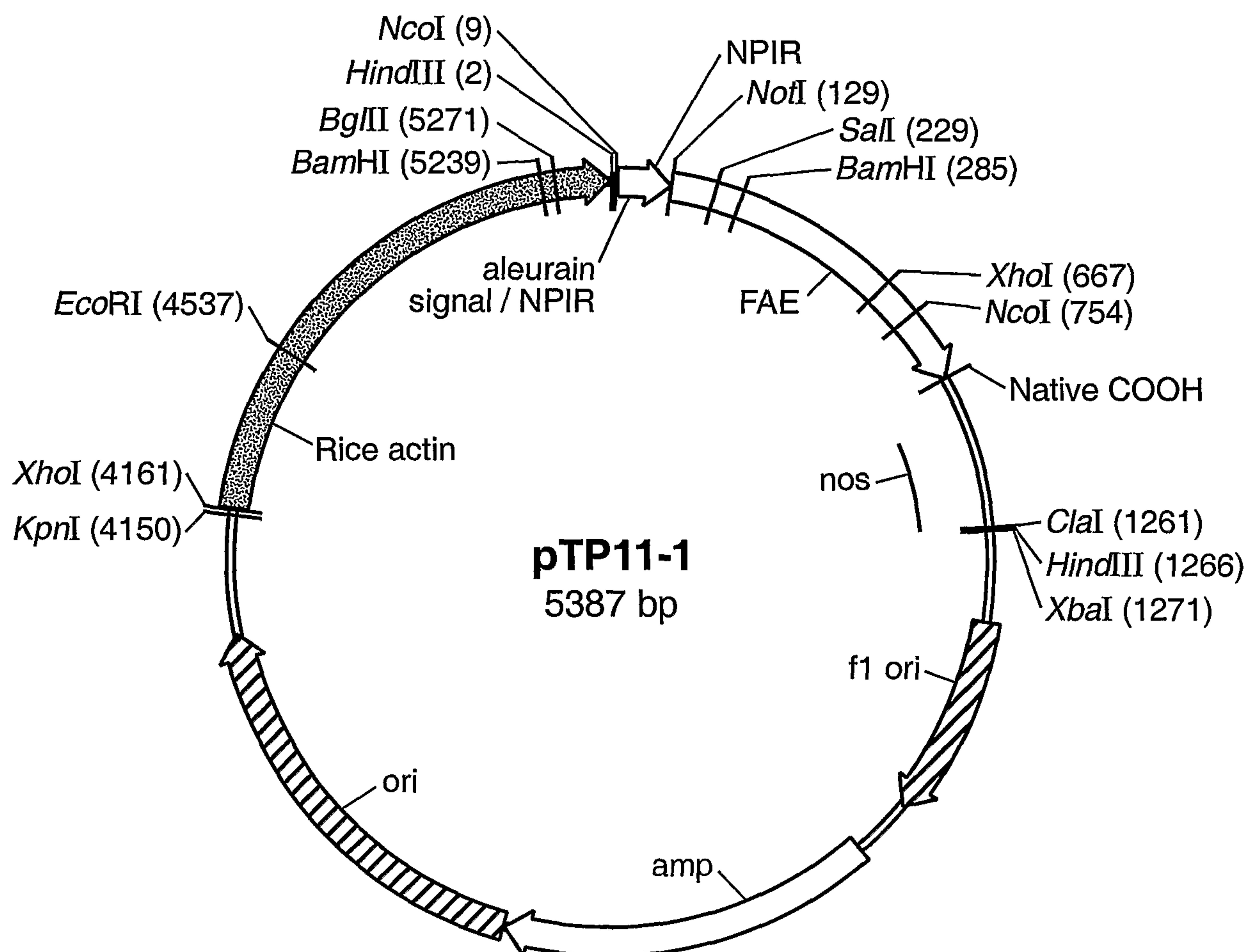
NcoI

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 GCCGTCGCCG
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 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 CGATGCCCTTC CAAGCCTCGA GCCAGATAC GACGCAGTAT TTCCGGGTCA

FIG. 49B

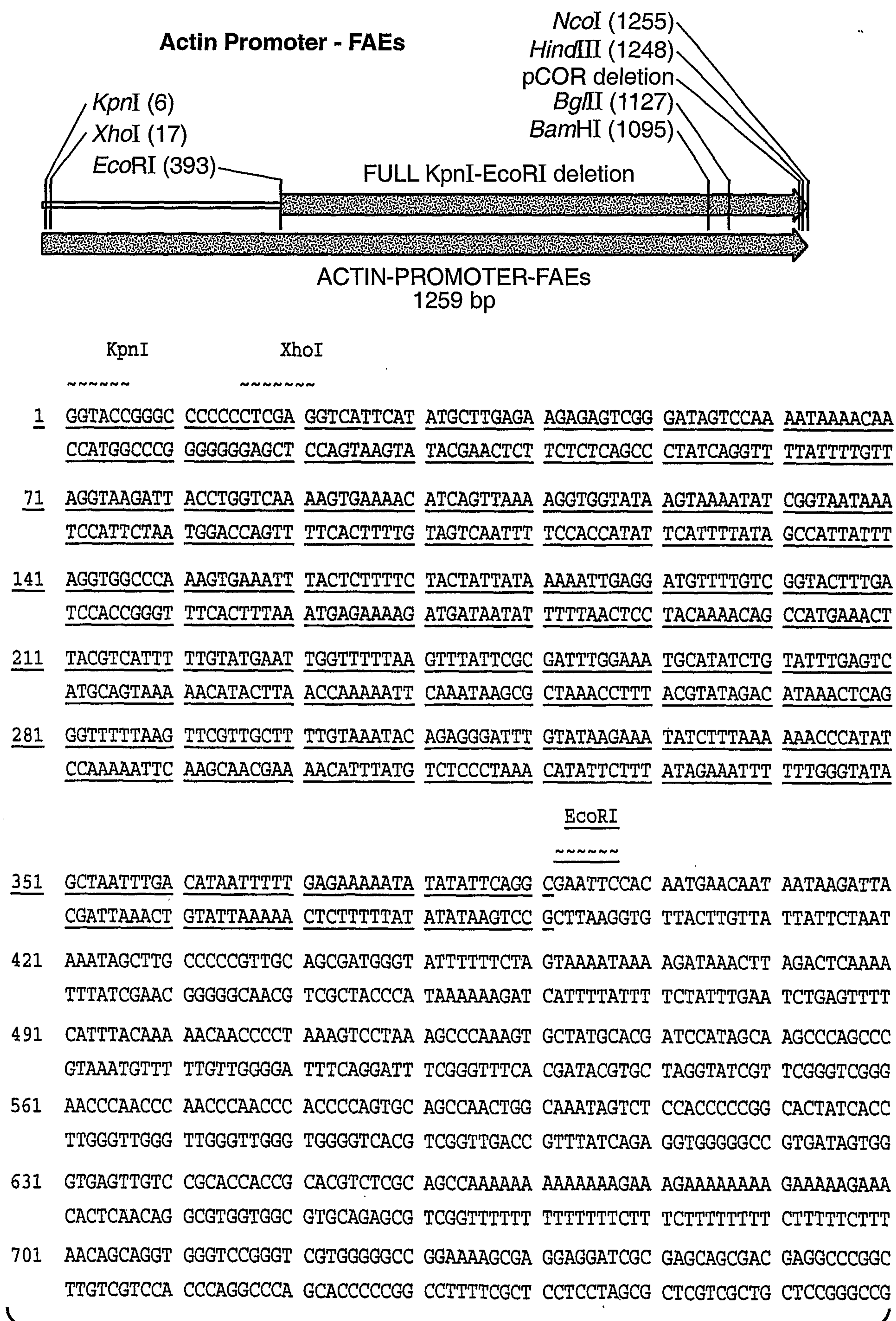
**FIG. 49C**

|      |             |              |             |             |             |             |             |
|------|-------------|--------------|-------------|-------------|-------------|-------------|-------------|
| 2451 | TGAGTGATAA  | CACTGCGGCC   | AACTTACTTC  | TGACAAACGAT | CGGAGGACCG  | AAGGAGCTAA  | CCGCTTTT    |
| 2521 | GCACAACATG  | GGGATCATG    | TAACTCGCCT  | TGATCGTTGG  | GAAACGGAGC  | TGAATGAAGC  | CATACCAAAC  |
| 2591 | GACGAGCGTG  | ACACCACGAT   | GCCTGTAGCA  | ATGGCAACAA  | CGTTGCGCAA  | ACTATTAACT  | GGCGAACTAC  |
| 2661 | TTACTCTAGC  | TTCCCGGCAA   | CAATTAAATAG | ACTGGATGGA  | GGCGGATAAA  | GTTGCAGGAC  | CACCTTCTGG  |
| 2731 | CTCGGCCCTT  | CCGGCTGGCT   | GGTTTATTGC  | TGATAAATCT  | GGAGCCGGTG  | AGCGTGGGTC  | TCGCGGTATC  |
| 2801 | ATTGCAGCAC  | TGGGGCCAGA   | TGGTAAGCCC  | TCCCGTATCG  | TAGTTATCTA  | CACGACGGGG  | AGTCAGGCAA  |
| 2871 | CTATGGATGA  | ACGAAATAGA   | CAGATCGCTG  | AGATAGGTGC  | CTCACTGATT  | AAGCATTTGGT | AACTGTCCAGA |
| 2941 | CCAAAGTTTAC | TCATATATAC   | TTTAGATTGA  | TTTAAAACTT  | CATTTTTAAT  | TTAAAAAGGAT | CTAGGTGAAG  |
| 3011 | ATCCTTTTIG  | ATAATCTCAT   | GACCAAAATC  | CCTTAAACGTG | AGTTTTCGTT  | CCACTGAGCG  | TCAGACCCCG  |
| 3081 | TAGAAAAAGAT | CAAAGGATCT   | TCTTGAGATC  | CTTTTTTTCT  | GGCGTAAATC  | TGCTGCTTGC  | AAACAATAAA  |
| 3151 | ACCACCGCTA  | CCAGCGGTGG   | TTTGTTTGCC  | GGATCAAGAG  | CTACCAACTC  | TTTTTTCGAA  | GGTAACTGGC  |
| 3221 | TTCAGCAGAG  | CGCAGATACC   | AAATACTGTC  | CTTCTAGTGT  | AGCCGTAGTT  | AGGCCACCAC  | TTCAAGAACT  |
| 3291 | CTGTAGCACC  | GCCTACATAC   | CTCGCTCTGC  | TAATCCTGTT  | ACCAGTGGCT  | GCTGCCAGTG  | GGATAAAGTC  |
| 3361 | GTGCTTTACC  | GGGTTGGACT   | CAAGACGATA  | GTTACCCGGAT | AAGCGCAGC   | GGTCGGGCTG  | AACGGGGGGT  |
| 3431 | TCGTGCACAC  | AGCCAGCTT    | GGAGCGAACG  | ACCTACACCG  | AAC TGAGATA | CCTACAGCGT  | GAGCTATGAG  |
| 3501 | AAAGCGCCAC  | GCTTCCCGAA   | GGGAGAAAGG  | CGGACAGGTA  | TCCGGTAAGC  | GGCAGGGTCG  | GAAACAGGAGA |
| 3571 | GCGCAGGAGG  | GAGCTTCCAG   | GGGAAACGC   | CTGGTATCTT  | TATAGTCCCTG | TCGGGTTTCG  | CCACCTCTGA  |
| 3641 | CTTGAGCGTC  | GATTTTGTG    | ATGCTCGTCA  | GGGGGCGGA   | GCCTATGGAA  | AAACGCCAGC  | AACGCGGCCT  |
| 3711 | TTTTACGGTT  | CCTGGCCCTT   | TGCTGGCCCTT | TTGCTCACAT  | GTTCTTTCCCT | GCGTTATCCC  | CTGATTCTGT  |
| 3781 | GGATAACCGT  | ATTACCGCCT   | TTGAGTGAGC  | TGATACCGCT  | CGCCGCAGCC  | GAAACGACCGA | GCGCAGCGAG  |
| 3851 | TCAGTGAGCG  | AGGAAGCGGA   | AGAGCGCCA   | ATACGCAAAC  | CGCTCTCTCC  | CGCGCGTTGG  | CCGATTTCATT |
| 3921 | AATGCAGCTG  | GCACGACAGG   | TTTCCCGACT  | GGAAAGCGGG  | CAGTGAGCGC  | AACGCAATTA  | ATGTGAGTTA  |
| 3991 | GCTCACTCAT  | TAGGCACCCC   | AGGCTTTACA  | CTTTATGCTT  | CCGGCTCGTA  | TGTTGTGTGG  | AATTGTGAGC  |
| 4061 | GGATAACAA   | TTCACACACAGG | AAACAGCTAT  | GACCATGATT  | ACGCCAAGCG  | CGCAATTAA   | CCTCACTAAA  |
|      |             | KpnI         |             | XhoI        |             |             |             |
| 4131 | GGGAACAAA   | GCTGGGTACC   | GGGCCCCCCC  | TCGAGGTCAT  | TCATATGCTT  | GAGAAAGAGAG | TCGGGATAGT  |
| 4201 | CCAAAATAAA  | ACAAAAGGTAA  | GATTACCTGG  | TCAAAAGTGA  | AAACATCAGT  | TAAAAAGGTGG | TATAAGTAAA  |
| 4271 | ATATCGGTAA  | TAAAAGGTGG   | CCCAAAGTGA  | AATTACTCT   | TTTCTACTAT  | TATAAAAATT  | GAGGATGTTT  |
| 4341 | TGTCGGTACT  | TTGATACGTC   | ATTTTGTGAT  | GAATTGGTTT  | TTAAGTTTAT  | TCGCCGATTTG | GAAATGCATA  |
| 4411 | TCTGTATTTG  | AGTCGGTTTTT  | TAAGTTCGTT  | GCTTTTGTA   | ATACAGAGGG  | ATTTGTATAA  | GAAATATCTT  |
|      |             |              |             |             | ECORI       |             |             |
| 4481 | TAAAAAACCC  | ATATGCTAAT   | TTGACATAAT  | TTTTGAGAAA  | AAATATATAT  | CAGGCGAATT  | CCACAATGAA  |
| 4551 | CAATAATAAG  | ATTAAAAATAG  | CTTGCCCCCG  | TTGCAGCGAT  | GGGTAATTTT  | TCTAGTAAAA  | TAAAAAGATAA |
| 4621 | ACTTAGACTC  | AAAACATTTA   | CAAAAACAAC  | CCCTAAAGTC  | CTAAAGCCCA  | AAGTGCTATG  | CACGATCCAT  |

FIG.- 49D

4691 AGCAAGCCCA GCCAAACCCA ACCCAACCA ACCCAACCA GTGAGCCAA CTGGCAATA GTCTCCACCC  
4761 CCGGCACTAT CACCGTGAGT TGTCGCACAC ACCGCACGTC TCGCAGCCAA AAAAAAAA AGAAAGAAAA  
4831 AAAAGAAAA GAAAAACAGC AGGTGGGTCC AGGTGGGTCC GCGCGGAAA ACTATATACA TACCCCCCC  
4901 CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCATCGCC CACTATACA TACCCCCCC  
4971 TCTCCTCCCA TCCCCCACA CCTACCACA CACCTCCTCC CACCTCCTG CCCCTCGCTG CCGACGACG  
5041 AGCTCCTCCC CCTCCTCCT CCTCCTCCT CCGTAACAC CCGCCCCCTC TCCTCTTTCT TTCTCCGTTT  
5111 TTTTTTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT GGTGGGCGA GAGGGCTTC GTCGCCAGA  
BamHI  
~~~~~  
5181 TCGGTGCGG GGAGGGGCGG GATCTGCGG CTGGCTCTC CCGGCGTGAG TCGGCCCGA TCCTCGCGG
BglII
~~~~~  
5251 GAATGGGGCT CTCGGATGTA GATCTTCTTT CTCTCTCTT TTTGTGGTAG AATTGAATC CCTCAGCAT  
5321 GTTCATCGGT AGTTTTCCTT TTATGATTT GTGACAAATG CAGCCTCGTG CGGAGCTTTT 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

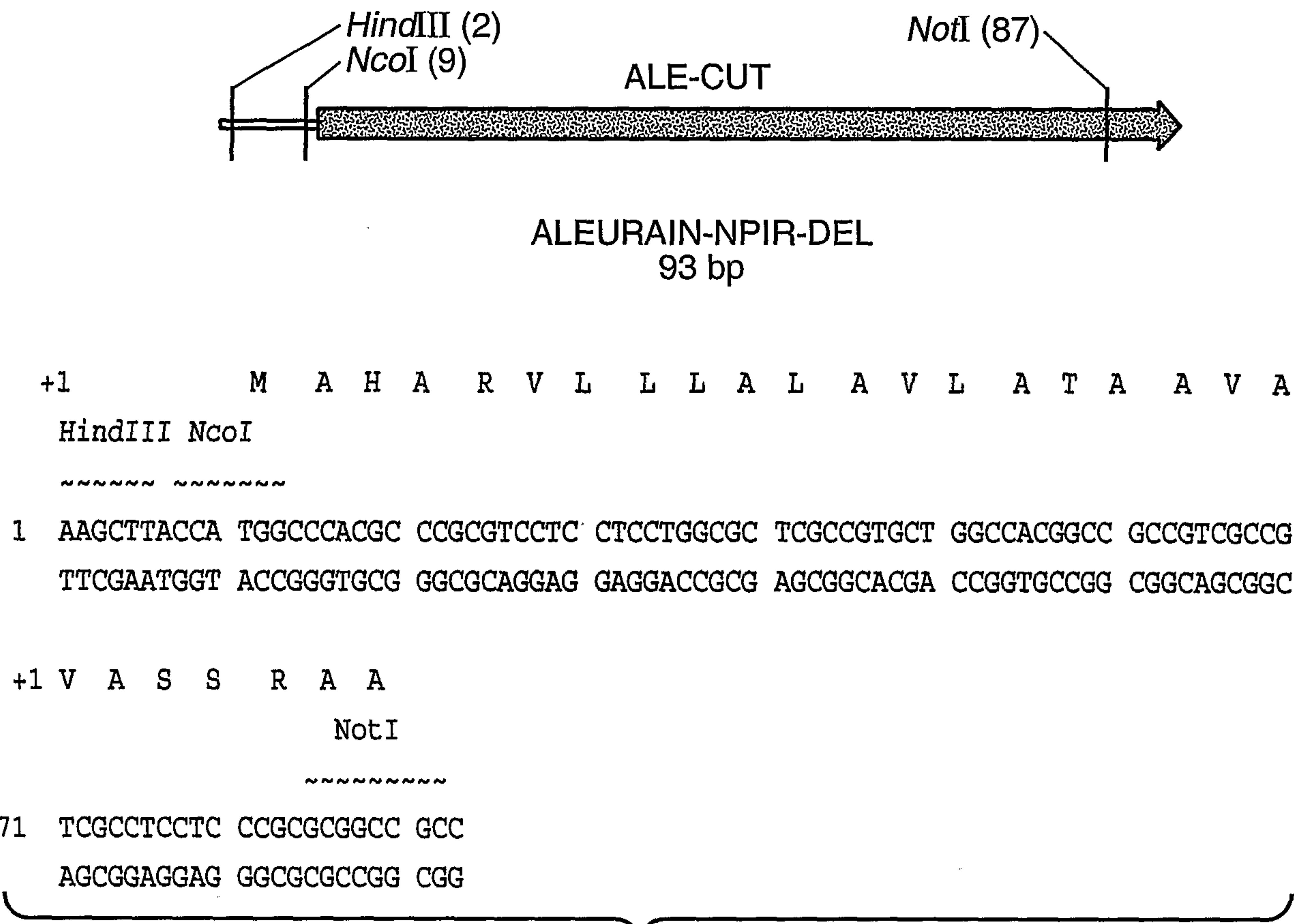
~~~~~

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



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