



(19)



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 325 908**

(51) Int. Cl.:
C12N 15/82 (2006.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(96) Número de solicitud europea: **01273842 .3**

(96) Fecha de presentación : **16.11.2001**

(97) Número de publicación de la solicitud: **1337652**

(97) Fecha de publicación de la solicitud: **27.08.2003**

(54) Título: **Manipulación del contenido de ácido fenólico y de la digestibilidad de paredes de células vegetales mediante la expresión dirigida de genes que codifican enzimas degradativas de la pared celular.**

(30) Prioridad: **17.11.2000 US 249608 P**

(73) Titular/es: **GENENCOR INTERNATIONAL, Inc.**
925 Page Mill Road
Palo Alto, California 94304, US

(45) Fecha de publicación de la mención BOPI:
24.09.2009

(72) Inventor/es: **Dunn-Coleman, Nigel;**
Langdon, Timothy y
Morris, Phillip

(45) Fecha de la publicación del folleto de la patente:
24.09.2009

(74) Agente: **Ponti Sales, Adelaida**

ES 2 325 908 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Manipulación del contenido de ácido fenólico y de la digestibilidad de paredes de células vegetales mediante la expresión dirigida de genes que codifican enzimas degradativas de la pared celular.

Campo de la invención

La presente invención se refiere a procedimiento para aumentar la disponibilidad de carbohidratos fermentables.

Antecedentes de la invención

La presente crisis en la agricultura del ganado ha provocado un resurgimiento en el interés por animales que se alimentan de hierba. Sin embargo, aunque se puede desear una dieta alta en forraje, actualmente no satisface las demandas de producción moderna de animales. Para que el animal haga un uso eficaz del forraje que consume, debe satisfacerse la demanda de energía de los microorganismos en el rumen y sincronizarse con la disponibilidad de las proteínas de las plantas. En otro caso, la falta de sincronía conducirá a (a) proteínas y otros nutrientes que se utilizan escasamente en el rumen, b) pérdida de nitrógeno, en orina y heces y, por lo tanto, el medio y (c) la necesidad de administrar cantidades en exceso de concentrados proteicos como suplementos de la dieta del rumiante.

La celulosa y la hemicelulosa en los tejidos de hierba y maíz podrían cumplir con los requisitos de energía del rumiante o proporcionar nuevas reservas de alimentación para la fermentación industrial a etanol. Este potencial no se realiza actualmente ya que las paredes celulares se lignifican y los polisacáridos de la pared celular se reticulan a nivel elevado con residuos fenólicos y lignina, dando lugar a velocidades bajas de digestión de la pared celular de las plantas en comparación con las velocidades de rotura de las proteínas en los rumiantes. Este es un problema particular para los forrajes más importantes en Europa, la hierba de centeno *Lolium perenne* y *L. multiflorum*, así como uno de los impedimentos principales para un uso más amplio de especies mejor adaptadas, tales como *Festuca arundinacea*, como cultivo de forraje. El aumento del índice de digestibilidad de la hierba ha sido por tanto un objetivo principal en el cultivo durante varias décadas, pero el progreso ha sido lento debido a las dificultades en la fijación de la variación natural en las variedades sintéticas derivadas de estas especies exogámicas (Hayward, *et al.*, TAG 70:48 (1985)).

Se sabe que la eliminación de ácidos fenólicos lábiles mediante tratamiento químico con álcali aumenta la biodegradabilidad y el valor nutricional de la alimentación de baja calidad, tales como paja de cereal, y se utiliza comercialmente para mejorar la alimentación. La reducción de la reticulación fenólica de los carbohidratos de las paredes celulares es por lo tanto una manera predecible de mejorar la velocidad de digestión y la digestibilidad de la hierba de centeno. Sin embargo, la modificación química puede tener otras desventajas. Por lo tanto, la modificación genética sería un método preferible para el cambio de la química de la pared celular de variedades altamente digeribles. Mucha gente de este campo persigue esta estrategia. Sin embargo, una alternativa es utilizar modificaciones genéticas para reducir los niveles de ácidos fenólicos en las paredes celulares disponibles para la reticulación mediante la rotura directa de enlaces éster que unen los ácidos fenólicos y ligninas a los polisacáridos de la pared celular o evitando la ferulación excesiva de los carbohidratos de las paredes celulares antes de su incorporación en la pared celular.

La presente invención satisface ésta y otras necesidades mediante la utilización de la expresión dirigida de las enzimas degradativas de la pared celular en plantas.

Descripción resumida de la invención

En la presente invención se dan a conocer procedimientos para aumentar la disponibilidad de carbohidratos fermentables. En un aspecto, se proporciona un procedimiento para aumentar la digestibilidad de una planta, comprendiendo el procedimiento introducir en la planta un cassette de expresión que comprende una secuencia de ADN que codifica esterasa de ácido ferúlico. La secuencia de ADN que codifica la esterasa del ácido ferúlico está unida operativamente a una secuencia promotora. El promotor puede ser constitutivo o inducible. El cassette de expresión comprende una secuencia de reconocimiento que dirige la expresión del polipéptido esterasa del ácido ferúlico a la vacuola.

En una realización preferida, la enzima que degrada la pared celular deriva de una fuente fúngica. En una realización más preferida, la esterasa de ácido ferúlico fúngica es una esterasa de ácido ferúlico de *Aspergillus*, preferiblemente *A. niger*.

En otro aspecto de la presente invención, se proporciona una planta transformada con el cassette de expresión. La planta se puede seleccionar del grupo que consiste en *Festuca*, *Lolium*, *Avena* y *Zea*. En una realización preferida, la planta es hierba de forraje. En otra realización, la planta es maíz.

Otros objetos, características y ventajas de la presente invención serán evidentes a partir de la siguiente descripción detallada.

Breve descripción de las figuras

La figura 1 ilustra un mapa de restricción de un fragmento de ADN que contiene el gen que codifica la esterasa de ácido ferúlico de 38 kd.

ES 2 325 908 T3

Las figuras 2A-E ilustran el ADN completo, con highlight para indicar la secuencia señal, el intrón y varios sitios de endonucleasas de restricción, y la secuencia de aminoácidos correspondiente al gen que codifica la esterasa de ácido ferúlico de 38 kd aislada de *Aspergillus niger*.

5 La figura 3 ilustra la secuencia de ADN del gen que codifica la esterasa de 38 kd.

La figura 4 ilustra la construcción de la esterasa de ácido ferúlico sin intrón aislado de *Aspergillus niger*.

10 La figura 5 ilustra que el solapamiento de los productos de PCR producidos con cebadores FAE-I5 FAE-I3 crea dos posibles marcas de lectura ininterrumpidos - la parte superior en la figura siguiente es funcional (la serina destacada es un sitio activo), la parte inferior está inactivada.

15 La figura 6 ilustra las posibles construcciones de vectores útiles en la presente invención. Son posibles varias combinaciones. Aunque se representa un gen FAE, se puede utilizar otra enzima que degrada la pared celular sola (es decir, en lugar de) o conjuntamente con el gen FAE. Amp = gen de resistencia a ampicilina

La figura 7 ilustra pCOR105

20 La figura 8 ilustra un vector ALE-TER genérico.

La figura 9 ilustra las secuencias de retención KDEL-COOH ER.

La figura 10 ilustra la estructura y secuencia de FAE-LINKER-FRAMESHIFT

25 La figura 11 ilustra los cassettes de transformación de plantas.

La figura 12 es una tabla de los vectores utilizados de la presente invención.

30 La figura 13 representa la secuencia señal vacuolar y en apoplasto de aleuraina de cebada.

La figura 14 ilustra la estructura y secuencia de sialil transferasa de rata.

La figura 15 ilustra la estructura y secuencia del motico del inhibidor II de proteasa de patata (PPI).

35 La figura 16 ilustra la expresión dirigida de gfp a compartimentos de células diferentes. También se muestran esquemas de los vectores utilizados.

40 La figura 17 ilustra la actividad FAE en hojas de *Festuca arundinacea* transgénica de diferentes edades bajo secuencias de reconocimiento ER y APO.

La figura 18 ilustra la actividad FAE en hojas de *Festuca arundinacea* transgénica de diferentes edades bajo secuencia de reconocimiento Vac.

45 La figura 19 ilustra la actividad FAE en hojas de *Lolium mutflorum* transgénica de diferentes edades.

La figura 20 ilustra la actividad FAE en hojas de *Lolium mutflorum* transgénica bajo secuencias de reconocimiento Vac, ER y APO.

50 La figura 21 ilustra los niveles de ácidos hidroxicinámicos monoméricos y diméricos esterificados en plantas de *Festuca arundinacea* que expresan FAE bajo la secuencia de reconocimiento Vac.

La figura 22 ilustra los niveles de ácidos hidroxicinámicos monoméricos y diméricos esterificados en plantas de *Festuca arundinacea* que expresan FAE bajo la secuencia de reconocimiento APO y ER.

55 La figura 23 ilustra la digestibilidad de materia seca *in vitro* de tejido de hoja de plantas *Festuca arundinacea* madura que expresan FAE bajo un promotor de actina.

60 La figura 24 ilustra la digestibilidad de materia seca *in vitro* de tejido de hoja de plantas *Lolium mutflorum* madura que expresan FAE bajo un promotor de actina.

La figura 25 ilustra la velocidad de fermentación y producción de gas acumulado en células de *Festuca arundinacea*.

65 La figura 26 ilustra la fermentación *in vitro* de paredes celulares de *Festuca arundinacea* de cultivos celulares que expresan FAE1 recombinante.

La figura 27 ilustra el tiempo para la velocidad de digestión máxima para células de *Festuca arundinacea*.

ES 2 325 908 T3

La figura 28 ilustra la producción de gas total en células de *Festuca arundinacea*.

La figura 29 ilustra la cinética de la actividad de FAE mediante la liberación de ácido ferúlico de las paredes celulares bajo la auto-digestión en *Festuca arundinacea* y estimulación por xilanasa.

La figura 30 ilustra la actividad de beta-glucoronidasa bajo el promotor de senescencia de *Lolium See1* en hojas de plantas transgénicas de *Lolium mutflorum*.

La figura 31 ilustra la liberación de HCAs monoméricas y diméricas en la auto-digestión de hojas de plantas que expresan FAE dirigida vacuolar.

La figura 32 es un esquema del vector pTP10-1. También se observa la secuencia de nucleótidos del vector de 5338 pb.

La figura 33 es un esquema del vector pUA4-4. También se observa la secuencia de nucleótidos del vector de 5345 pb.

La figura 34 es un esquema del vector pTU4. También se observa la secuencia de nucleótidos del vector de 5337 pb.

La figura 35 es un esquema del vector pTT5.14. También se observa la secuencia de nucleótidos del vector de 5395 pb.

La figura 36 es un esquema del vector Ptt8-5. También se observa la secuencia de nucleótidos del vector de 5337 pb.

La figura 37 es un esquema del vector pTP5-1. También se observa la secuencia de nucleótidos del vector de 5277 pb.

La figura 38 es un esquema del vector pTP4a2. También se observa la secuencia de nucleótidos del vector de 5327 pb.

La figura 39 es un esquema del vector pTP3-1. También se observa la secuencia de nucleótidos del vector de 5338 pb.

La figura 40 es un esquema del vector pTU5. También se observa la secuencia de nucleótidos del vector de 5337 pb.

La figura 41 es un esquema del vector pGT6. También se observa la secuencia de nucleótidos del vector de 4773 pb.

La figura 42 es un esquema del vector pJQ5. También se observa la secuencia de nucleótidos del vector de 5034 pb.

La figura 43 es un esquema del vector pJO6.1. También se observa la secuencia de nucleótidos del vector de 4950 pb.

La figura 44 es un esquema del vector pJQ4. También se observa la secuencia de nucleótidos del vector de 4974 pb.

La figura 45 es un esquema del vector pPQ10.1. También se observa la secuencia de nucleótidos del vector de 5164 pb.

La figura 46 es un esquema del vector pJQ3. También se observa la secuencia de nucleótidos del vector de 4965 pb.

La figura 47 es un esquema del vector pUG4. También se observa la secuencia de nucleótidos del vector de 5295 pb.

La figura 48 es un esquema del vector pUB8.11. También se observa la secuencia de nucleótidos del vector de 5001 pb.

La figura 49 es un esquema del vector pTP11-1. También se observa la secuencia de nucleótidos del vector de 5387 pb.

La figura 50 ilustra el promotor de actina y su correspondiente secuencia de nucleótidos.

La figura 51 ilustra la estructura de Aleuraína con NPIR eliminado. También se observan las secuencias de nucleótidos correspondientes.

La figura 52 ilustra la secuencia promotora de SEE1 (senescencia aumentada).

La figura 53 ilustra la secuencia promotora de SEE1 (senescencia aumentada) más la secuencia señal de aleuraína vacuolar/NPIR.

Descripción detallada de la invención

A continuación, se describirá la invención haciendo referencia únicamente a las siguientes definiciones y ejemplos.

A menos que se defina contrario, todos los términos científicos y técnicos utilizados en la presente invención tienen el mismo significado que se entiende habitualmente por un experto en la materia a la que pertenece la invención. Singleton *et al.*, DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2D ED, John Wiley and Sons, Nueva York (1994), y Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) proporcionan a un experto con un diccionario general de muchos los términos utilizados en la presente invención. Aunque en la práctica o ensayo se puede utilizar cualquier método y material similar o equivalente a los descritos en la presente invención, se describen los procedimientos y materiales preferidos. Los intervalos numéricos incluyen los números que definen el intervalo, A menos que se indique lo contrario, los ácidos nucleicos se escriben de izquierda a derecha en la orientación 5' a 3'; las secuencias de aminoácidos se escriben de izquierda a derecha en la orientación de amino a carboxi, respectivamente. Para las definiciones y términos de la técnica los profesionales se pueden dirigir a Sambrook *et al.*, 1989 y Ausubel FM *et al.*, 1993. Debe entenderse que esta invención no se limita a la metodología, protocolos y reactivos descritos concretos, ya que éstos pueden variar.

Los encabezamientos que se proporcionan en la presente invención no son limitaciones de los diversos aspectos o realizaciones de la invención que se pueden obtener por referencia a la memoria en su totalidad. Por consiguiente, los términos definidos inmediatamente a continuación están completamente definidos por referencia a la memoria en su totalidad.

Definiciones

Debe indicarse que, tal como se utiliza en esta memoria y en las reivindicaciones adjuntas, las formas singulares “un”, “una” y “el/la” incluyen los referentes plurales a menos que el contenido dictamine claramente lo contrario. De este modo, por ejemplo, la referencia a una composición que contiene “un compuesto” incluye una mezcla de dos o más compuestos. Debe indicarse que el término “o” se utiliza generalmente en el sentido que incluye “y/o”, a menos que el contenido dictamine claramente lo contrario.

“Variantes modificadas de manera conservativa” se aplica tanto a secuencias de aminoácidos como de polinucleótidos. Con respecto a polinucleótidos concretos, las variantes modificadas de manera conservativa se refieren a aquellos polinucleótidos que codifican secuencias de aminoácidos idénticas o esencialmente idénticas, o cuando el polinucleótido no codifica una secuencia de aminoácidos, a secuencias esencialmente idénticas. Debido a la degeneración del código genético, un mayor número de polinucleótidos funcionalmente idénticos codifican cualquier proteína determinada. Por ejemplo, los codones GCA, GCC, GCG y GCU codifican el aminoácido alanina. De este modo, en cada posición en la que una alanina está especificada por un codón, el codón se puede alterar a cualquiera de los codones correspondientes descritos sin alterar el polipéptido codificado. Dichas variaciones de ácido nucleico son “variaciones silenciosas”, que son una muestra de variaciones modificadas de manera conservativa. Cada polinucleótido de la presente invención que codifica un polipéptido también describe cada posible variación silenciosa del ácido nucleico. Un experto reconocerá que cada codón en un polinucleótido (excepto AUG, que es normalmente el único codón para metionina) se puede modificar para obtener una molécula funcionalmente idéntica. Por consiguiente, cada variación silenciosa de un polinucleótido que codifica un polipéptido está implícita en cada secuencia descrita. Para los objetivos de la expresión de la proteína, existen “codones subóptimos”. Estos son codones que no son preferidos por un género o especie particular. La alteración de estos “codones sub-óptimos” a “codones preferidos” es una mutación silenciosa en la que el aminoácido codificado por los codones es el mismo pero un codón es preferencialmente expresado por el género concreto, por ejemplo *Triticum spp.*

En cuanto a las secuencias de aminoácidos, un experto reconocerá que sustituciones, eliminaciones o adiciones individuales a una secuencia de polinucleótido, péptido, polipéptido o proteína que altera, añade o elimina un único aminoácido o un pequeño porcentaje de aminoácidos en la secuencia codificada es una “variante modificada de manera conservativa”, donde la alteración da lugar a la sustitución de un aminoácido con un aminoácido químicamente similar. Las tablas de sustitución conservativa que proporcionan aminoácidos funcionalmente similares son bien conocidos en la técnica.

Los siguientes seis grupos contienen cada uno aminoácidos que son sustituciones conservativas por otros:

- 1) Alanina (A), Serina (S), Treonina (T);
- 2) Ácido Aspártico (D), Ácido Glutámico (E);

3) Asparagina (N), Glutamina (Q);

4) Arginina (R), Lisina (K);

5) Isoleucina (I), Leucina (L), Metionina (M), Valina (V); y

6) Fenilalanina (F), Tirosina (Y), Triptófano (W).

(véase, por ejemplo, Creighton, Proteins (1984)).

“Ácido Piroglutámico” es la amida interna ciclada del ácido L-glutámico.

La frase “que controla el nivel de ácido fenólicos” se refiere a la manipulación de la expresión de ácidos fenólicos en plantas, particularmente paredes celulares de plantas. La manipulación puede ser positiva; por ejemplo, aumentando los niveles de ácidos fenólicos; o negativa, por ejemplo, disminuyendo el nivel de ácidos fenólicos; o neutra, por ejemplo, cambiando las cantidades relativas de ácidos fenólicos específicos en las paredes celulares pero manteniendo la cantidad total relativamente igual. La acción de manipulación puede ser durante el crecimiento de las plantas o después del crecimiento de las plantas, por ejemplo, después de que la planta haya sido cortada o arrancada del suelo o ingerida. Las “paredes celulares de plantas” se refieren a las paredes celulares de cualquier célula de la planta.

El término “derivado” significa que un polinucleótido o proteína se refiere a otro polinucleótido o proteína. Las relaciones pueden ser de homología, por ejemplo, nucleótidos y proteínas de ciertas especies son homólogos a polinucleótidos y proteínas similares de otras especies; analogía, por ejemplo, las proteínas realizan la misma función y, por tanto, se relacionan entre sí independientemente del organismo de origen. La relación pueden ser manipuladas, por ejemplo, una proteína (y un polinucleótido) se puede derivar de otra proteína mediante mutación; o manipulación química (peptidomiméticos). Además, se puede derivar una proteína o un polinucleótido de un organismo si, en el estado natural, la proteína o polinucleótido se encuentra en un organismo, pero se reproduce recombinantemente en otro.

El término “polinucleótido exógeno” se refiere a un polinucleótido que se introduce en la planta mediante algún medio diferente de un cruzamiento sexual o reproducción sexual. A continuación se describen ejemplos de medios mediante los cuales esto se puede realizar e incluyen transformación mediada por *Agrobacterium*, métodos biolísticos, electroporación, técnicas *in planta*, y similares. Dicha planta que contiene el polinucleótido exógeno se refiere en la presente invención como una planta transgénica de generación R1. Las plantas transgénicas que surgen del cruzamiento sexual o mediante autopolinización son la progenie de dicha planta.

El término “molécula de polinucleótido aislada” o “proteína aislada” se refiere a un polinucleótido o proteína que está esencialmente libre de otros componentes celulares con los que está asociada en el estado natural. Se encuentra preferiblemente en un estado homogéneo aunque puede estar en una solución pura o acuosa. La pureza y homogeneidad se determinan habitualmente utilizando técnicas de química analítica, tales como electroforesis en gel de poliacrilamida o cromatografía líquida de alto rendimiento. La proteína que es la especie predominante presente en una preparación se purifica sustancialmente. En particular, un gen *FAE1* aislado se separa de los marcos de lectura abiertos que flanquean el gen y codifica una proteína diferente de *FAE1*.

Un “polinucleótido que codifica *FAE1*” es una secuencia de ácido nucleico que comprende (o consiste en) una región codificante de un gen *FAE1* o que codifica un polipéptido de *FAE1*. Los polinucleótidos de *FAE1* también se pueden identificar por su capacidad para hibridarse bajo condiciones de baja fidelidad (ver a continuación) a sondas de ácidos nucleicos que tienen una secuencia de 8 a 300 bases, preferiblemente una secuencia de 80 a 100 bases en la secuencia descrita en WO 98/14594.

El término “ácido nucleico que codifica”, “secuencia de ácido nucleico que codifica” o “polinucleótido que codifica” se refiere a un polinucleótido que dirige la expresión de una proteína o péptidos específico. Los polinucleótidos incluyen tanto la secuencia de la hebra de ADN que se transcribe en el ARN como la secuencia de ARN que se traduce en la proteína. Los polinucleótidos incluyen tanto los polinucleótidos de longitud completa como las secuencias más cortas derivadas de las secuencias de longitud completa. Se entiende que un polinucleótido concreto incluye los codones degenerados de la secuencia nativa o las secuencias que se pueden introducir para proporcionar la preferencia de codón en una célula huésped específica. El polinucleótido incluye la hebra tanto de sentido como antisentido como hebras únicas individuales o en forma de cadenas dobles.

El término “unido operativamente” se refiere a la unión funcional entre un promotor y una segunda secuencia, en la que la secuencia promotora inicia la transcripción del ARN correspondiente a la segunda secuencia.

El término “plásmido” se refiere a una molécula de ADN de doble cadena circular que comprende la secuencia codificante de interés, elementos reguladores, un marcador de selección y opcionalmente un marcador de amplificación. Un plásmido puede transformar células procariotas o transfectar células eucariotas. Un “cassette de expresión” significa una parte de un plásmido (o el plásmido completo) que contiene los elementos reguladores deseados para la

transcripción, traducción y/o expresión y la región codificante de un polinucleótido. Un plásmido puede contener uno o más cassettes de expresión. Si se introducen múltiples cassettes de expresión en una planta.

Se pueden introducir simultáneamente o a tiempos diferentes. Si se desea la introducción simultánea, los cassettes de expresión pueden ser en un plásmido o más. Habitualmente, un cassette de expresión comprende un promotor, una cola Poly A+ y secuencias señal que dirigen el polipéptido expresado a una región específica de una célula o para secretarse, si se desea. Entre los ejemplos de secuencias señal que “dirigen la expresión” de esterasa de ácido ferúlico se incluyen secuencias localizadas en dirección 5’ de la secuencia codificante de FAE. El polinucleótido que codifica la secuencia señal se encuentra preferiblemente en los 100 nucleótidos “en dirección 5’” desde el codón de inicio (AUG). Más preferiblemente, el polinucleótido que codifica la secuencia señal se encuentra en los 50 nucleótidos en dirección 5’ desde el codón de inicio. Los orgánulos celulares reconocidos por las secuencias señales utilizadas en la presente invención son vacuolas. Además de las secuencias señales en dirección 5, el cassette de expresión de la presente invención puede incluir un polinucleótido que codifica una secuencia señal en el extremo 3’. Preferiblemente, el polinucleótido que codifica la secuencia señal se encuentra directamente en dirección 3’ desde la secuencia codificante, más preferiblemente menos de 100 pares de bases desde el codón de parada, más preferiblemente menos de 20 pares de bases desde el codón de parada.

El término “polinucleótido,” “polinucleótido” o “secuencia de ácido nucleico” se refiere a desorribonucleótidos o ribonucleótidos y polímeros de los mismos en forma de cadena individual o doble. A menos que se limite de manera específica, el término comprende polinucleótidos que contienen análogos conocidos de nucleótidos naturales que presentan propiedades de unión similares al polinucleótido de referencia y se metabolizan de una manera similar a los nucleótidos naturales. A menos que se indique lo contrario, un polinucleótido de FAE1 particular se la presente invención también comprende implícitamente variantes modificadas de manera conservativa (por ejemplo, sustituciones de codones degenerados) y secuencias complementarias y también la secuencia explícitamente indicada. Específicamente, las sustituciones de codones degenerados se pueden conseguir mediante la generación de secuencias en las que la tercera posición de uno o más codones seleccionados (o todos) se sustituye por residuos de base mixta y/o desoxiinosina (Batzner *et al.*, Nucleic Acid Res. 19:5081 (1991); Ohtsuka *et al.*, J. Biol. Chem. 260:2605-2608 (1985); y Cassol *et al.*, 1992; Rossolini *et al.*, Mol. Cell. Probes 8:91-98 (1994)). El término polinucleótido se utiliza indistintamente con gen, ADNc, y ARNm codificado por el gen.

El término “polipéptido,” “péptido” y “proteína” se utilizan indistintamente y se refieren a aminoácidos conectados por enlaces peptídicos. Los polipéptidos pueden ser proteínas completas o partes de las mismas. Por ejemplo, un polipéptido de FAE1 puede referirse a la proteína de FAE1 completa o fragmentos de la proteína de FAE1. Una “esterasa de ácido ferúlico con un sitio de glicosilación alterado” se refiere a una proteína de FAE en la que una mutación ha cambiado el patrón de glicosilación de la proteína. Las mutaciones que realizan dichos cambios son conocidos en la técnica e incluyen, pero sin limitación, sustituciones de aminoácidos, y mutaciones en las proteínas del Aparato de Golgi y retículo endoplasmático que realizan la glicosilación de las proteínas.

El término “promotor” se refiere a un polinucleótido que dirige la expresión de una secuencia codificante. Un promotor puede ser constitutivo, es decir, relativamente independiente de la etapa de diferenciación de la célula en la que está contenido o puede ser inducible, es decir, inducido por factores ambientales específicos, tales como la duración del día, la temperatura, etc., o un promotor puede ser específico de tejido, es decir, que dirige la expresión de la secuencia codificante en células de un cierto tipo de tejido. Un promotor de “senescencia” es un promotor inducible que provoca que se inicie la transcripción tras un cierto suceso referente a la edad del organismo. Un “promotor del choque térmico” es un promotor inducible que provoca que se inicie la transcripción tras un cambio en la temperatura. Un ejemplo de un promotor de proteína del choque térmico es el promotor Gmhsp de soja. Además de estos promotores inducibles, un experto entenderá que se pueden utilizar otros promotores inducibles. Por ejemplo, un promotor inducido por herida, como LAP, Patente de Estados Unidos No. 5.962.670.

El término “purificado” indica que un polinucleótido o proteína da lugar a esencialmente una banda en un gel electroforético. Particularmente, significa que el polinucleótido o proteína es por lo menos un 85% puro, más preferiblemente por lo menos un 95% puro, y lo más preferente por lo menos un 99.5% puro.

El término “se hibrida específicamente” se refiere a una sonda de ácidos nucleicos que hibrida, forma cadenas dobles o se une a una secuencia de ADN o ARN diana concreta cuando las secuencias diana están presentes en una preparación de ADN o ARN celular total. Secuencias de ácido nucleico “complementarias” o “diana” se refieren a aquellas secuencias de ácido nucleico que se hibridan selectivamente a una sonda de ácido nucleico. Las condiciones de hibridación más adecuadas dependen, por ejemplo, de la longitud de la sonda, la composición de bases, y el número de desapareamientos y su posición en la sonda, y frecuentemente se deben determinar de manera empírica. Para discusiones sobre el diseño de la sonda de ácidos nucleicos y las condiciones de hibridación, véase, por ejemplo, Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989) (“Sambrook”) o CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, F. Ausubel *et al.*, ed. Greene Publishing y Wiley-Interscience, New York (1987) (“Ausubel”).

El término “condiciones de fidelidad” en el contexto de los experimentos de hibridación de polinucleótidos, tales como hibridaciones Southern y Northern se refiere a entornos de unión y lavado dependientes de la secuencia. Una amplia guía para la hibridación de polinucleótidos se encuentra en Tijssen (1993) LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY-HYBRIDIZATION WITH NUCLEIC ACID PROBES parte

I capítulo 2 “overview of principles of hybridization and the strategy of nucleic acid probe assays”, Elsevier, New York. Generalmente, las condiciones de hibridación y lavado con fidelidad elevada se seleccionan para que sean aproximadamente 5°C inferior al punto de fusión térmico (T_m) para la secuencia específica a una fuerza iónica y pH definidos. La T_m es la temperatura (bajo una fuerza iónica y pH definidos) a la que el 50% de la secuencia
 5 diana se hibrida a un sonda de emparejamiento perfecto. Las condiciones de fidelidad muy elevada se seleccionan para que sean iguales a la T_m para una sonda particular. Un ejemplo de condiciones de hibridación de fidelidad para la hibridación de polinucleótidos complementarios que tienen más de 100 residuos complementarios en un filtro en una transferencia Southern o Northern en formalina al 50% con 1 mg de heparina entre 40 y 50°C, preferiblemente 42°C, con la hibridación llevada a cabo durante toda la noche. Un ejemplo de condiciones de lavado de fidelidad
 10 elevada es 0,15M NaCl de 70 a 80°C, siendo preferible 72°C durante aproximadamente 15 minutos. Un ejemplo de condiciones de lavado de fidelidad es un lavado 0,2x SSC a aproximadamente 60 a 70°C, preferiblemente 65°C durante 15 minutos (véase, Sambrook, *supra* para una descripción del tampón de SSC). A menudo, un lavado de fidelidad elevada está precedida por un lavado de fidelidad baja para eliminar la señal de sonda de fondo. Un ejemplo de lavado de fidelidad media para una doble cadena de, por ejemplo, más de 100 nucleótidos, es 1x SSC de 40 a
 15 50°C, preferiblemente 45°C durante 15 minutos. Un ejemplo de lavado de fidelidad baja para una cadena doble de, por ejemplo, más de 100 nucleótidos, es 4-6x SSC de 35 a 45°C, siendo 40°C preferible, durante 15 minutos. En general, una proporción de señal con respecto a ruido de 2x (o superior) que la observada para una sonda no relacionada en el ensayo de hibridación concreto indica la detección de una hibridación específica. Los polinucleótidos que no se hibridan entre sí bajo condiciones de fidelidad son aún sustancialmente idénticos si los polipéptidos que codifican son
 20 sustancialmente idénticos. Esto tiene lugar, por ejemplo, cuando una copia de un polinucleótido se crea utilizando la máxima degeneración del codón permitida por el código genético.

El término “planta transgénica” se refiere a una planta en la que los polinucleótidos exógenos se han introducido y su progenie. Habitualmente, las células de una planta se transforman con el polinucleótido exógeno y se regenera
 25 una planta transgénica a partir de las células transformadas. A continuación, la planta regenerada se desarrolla para producir una cepa de plantas transgénicas.

“Xilanasa” (EC 3.2.1.8) se refiere a una clase bien descrita de glicosil hidrolasas que hidrolizan xilano. Las aplicaciones comerciales de xilanasa incluyen la degradación y blanqueamiento de la pulpa de madera para fabricar papel. La
 30 xilanasa también se puede añadir a la alimentación para animales para mejorar la digestibilidad de la materia vegetal. Habitualmente, la xilanasa comercial deriva de los hongos. Una xilanasa preferida deriva de *Trichoderma*.

Realizaciones preferentes

35 Las paredes de las células vegetales contienen un rango de ácidos fenólicos unidos a éster lábiles a álcali.

En particular, las paredes celulares de la hierba se caracterizan por la presencia de grandes cantidades de ácidos ferúlico y p-coumárico esterificados (principalmente en sus configuraciones E), unidos a arabinoxilanos en el C5 de arabinosa. Se liberan como oligosacáridos ferulados (FAX y PAX) mediante el tratamiento con celulasa, pero *in vivo* proporcionan un sustrato para la reticulación catalizada por peroxidasa de los polisacáridos de la pared celular y lignina. Los niveles elevados de estos ácidos fenólicos y sus dímeros tienen una fuerte influencia en las propiedades mecánicas, digestibilidad y velocidades de digestión de la hierba por los rumiantes.

Las investigaciones previas han mostrado que el ácido ferúlico es el ácido p-hidroxicinámico predominante este-
 45 rificado con polisacárido de la hierba, pero hasta hace poco, el único deshiodrómero de ácido ferúlico que se había aislado era el ácido 5,5'-diferúlico. Recientemente, se han aislado dímeros de deshiodróferulato y mezclas de dímeros del tipo ciclobutano a partir de las paredes de células vegetales (Waldron, *et al.*, Phytochemical Analysis 7:305 (1996)). Tal como se puede observar en la figura 1, estas mezclas están presentes en grandes cantidades en células de hierba. También se han aislado a partir de paredes celulares dímeros ácido ferúlico unido a éter-coniferil alcohol (Jacquet, *et al.*, Polyphenol Comm. Bordeaux pp 451 (1996)), estableciendo por primera vez que los esters de ferulato se copolimerizan de manera oxidativa con precursores de lignina que pueden anclar ligninas a los polisacáridos de las paredes celulares. La producción de estos dímeros en células de hierba indica que la reticulación de deshiodrómeros de ácido
 50 ferúlico de los polisacáridos de las paredes celulares es mucho más amplia que lo que se pensaba anteriormente.

Se ha descrito un sistema de enzimas de endomembranas de perejil que cataliza la ferulación de aceptores de
 55 polisacáridos endógenos a partir de feruloil CoA, apuntando a ER/golgi como el sitio de esterificación de polisacáridos y el éster CoA como el cosustrato fisiológico (Meyer, *et al.*, FEBS Lett. 290:209 (1991)). Se ha observado evidencias adicionales de esto en polisacáridos extracelulares solubles en agua excretados en grandes cantidades en el medio por los cultivos de células de hierba. Este material está muy esterificado con ácido ferúlico y p-coumárico a niveles
 60 similares a las paredes celulares de las células cultivadas.

Se ha detectado actividad de feruloil esterasa en varias muestras fúngicas incluyendo, hongos anaeróbicos de la tripa, levaduras, actinomicetos y varias bacterias ruminales que degradan la fibra, lo que permite desesterificar los
 65 arabinoxilanos y pectinas.

Se han aislado dos esterasas de ácido ferúlico (FAE) de *Aspergillus niger*, diferenciadas en base al peso molecular y la especificidad de sustrato, y se ha observado que hidrolizan cuantitativamente el ácido ferúlico y liberan dímeros de deshiodróferulato de las paredes de las células vegetales. Además, se ha observado que las FAE actúan sinérgicamente

con xilanasa para liberar ácido ferúlico de las paredes de células vegetales a una mayor velocidad. Recientemente, se ha clonado un gen de la esterasa de ácido ferúlico (FAE) de *Aspergillus niger* (Michelson, *et. al.* Solicitud de Patente Europea No. 9510370.1). Los inventores han observado que la enzima recombinante libera ácido ferúlico y dímeros de diferulato de paredes celulares de la hierba de una manera dependiente de la concentración y que esta enzima es estable a 30°C pH 5,0 en presencia de sustrato y tiene una vida media de 61 h a 30°C en presencia de extractos vacuolares (pH 4,6) de células de hierba. Por tanto, este gen era un candidato para la expresión dirigida e inducible de FAE en hierbas (por ejemplo, *Lolium multiflorum*).

La presente invención proporciona procedimientos de cambio de la estructura de la pared celular de plantas transgénicas y, por lo tanto, las hace más digeribles. El procedimiento comprende introducir una secuencia codificante de esterasa de ácido ferúlico en las células de una planta. Unido operativamente a la secuencia codificante se encuentra un promotor que puede ser constitutivo o inducible y secuencias señales que sirven para dirigir la expresión de la secuencia codificante en los orgánulos deseados en la célula deseada de la planta. Las secuencias señales pueden ser N-terminal o C-terminal o ambas.

Opcionalmente, se introduce una segunda y/o tercera secuencia codificante en la planta. Se prefiere que una secuencia codificante de xilanasa fúngica se coexpresa con la secuencia codificante de FAE.

La presente invención también proporciona plantas transgénicas que contienen las secuencias codificantes de FAE1, donde la expresión de la secuencia codificante de FAE1 está dirigida a la vacuola que conduce a las hierbas más digeribles.

En general, la nomenclatura y los procedimientos de laboratorio en la tecnología de ADN recombinante descritos a continuación son bien conocidos y se utilizan habitualmente en la técnica. Las técnicas estándar se utilizan para la clonación, el aislamiento, amplificación y purificación de ADN y ARN. En general, las reacciones enzimáticas que implican ADN ligasa, ADN polimerasa, endonucleasas de restricción y similares se realizan según las especificaciones del fabricante. Los textos bases que describen los métodos generales de uso en la presente invención incluyen Sambrook, *et al.*, MOLECULAR CLONING, A LABORATORY MANUAL, 2ND ED. (1989); Kriegler, GENE TRANSFER AND EXPRESSION: A LABORATORY MANUAL (1990); y Ausubel *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (1994)).

A. Aislamiento de polinucleótidos

El aislamiento de los polinucleótidos, por ejemplo, FAE1 y xilanasa de la invención, puede llevarse a cabo mediante varias técnicas. Ver, por ejemplo, la solicitud de Estados Unidos 08/952,445 en trámite que describe el aislamiento de FAE a partir de *Aspergillus niger*, y la solicitud de Estados Unidos 09/658,772 en trámite que describe el aislamiento de una xilanasa a partir de *T. reesei*.

Por ejemplo, pueden utilizarse las sondas de oligonucleótidos basadas en las secuencias citadas en la presente invención para identificar el gen deseado en una biblioteca de ADNc o ADN genómico de una especie deseada. Para construir bibliotecas genómicas, se generan grandes segmentos de ADN genómico mediante fragmentación aleatoria, por ejemplo utilizando endonucleasas de restricción, y se ligan con un ADN de vector para formar concatémeros que pueden empaquetarse en el vector apropiado. Para preparar una biblioteca de ADNc a partir de un cultivo celular específico, por ejemplo, *Aspergillus niger*, se aísla ARNm del cultivo y se prepara a partir del ARNm una biblioteca de ADNc que contiene las transcripciones génicas del ARNm.

A continuación puede cribarse la biblioteca de ADNc o genómica utilizando una sonda basada en la secuencia de un polinucleótido conocido, tal como los polinucleótidos citados en la presente invención. Pueden utilizarse sondas para hibridarse con secuencias de ADN genómico o ADNc para aislar genes homólogos en especies vegetales iguales o diferentes. Además de las sondas derivadas de polinucleótidos conocidos, pueden utilizarse sondas degeneradas. Las técnicas de fabricación y utilización de sondas degeneradas son bien conocidas en el sector y pueden encontrarse en Sambrook y Ausubel.

De modo alternativo, los polinucleótidos de interés pueden amplificarse a partir de muestras de polinucleótidos utilizando técnicas de amplificación. Por ejemplo, puede utilizarse la tecnología de la reacción en cadena de la polimerasa (PCR) para amplificar las secuencias de los genes directamente a partir de ARNm, a partir de ADNc, a partir de bibliotecas genómicas o bibliotecas de ADNc. También pueden ser útiles la PCR y otros procedimientos de amplificación *in vitro*, por ejemplo, para clonar polinucleótidos que codifican para proteínas a expresar, para producir polinucleótidos para usarlos como sondas para detectar la presencia del ARNm deseado en muestras, para la secuenciación de polinucleótidos y para otros propósitos.

A partir de las comparaciones de las secuencias proporcionadas en la presente invención se generan cebadores y sondas apropiadas para identificar genes específicos de esterasa de ácido ferúlico, así como secuencias de xilanasa, de tejidos vegetales y fúngicos. Para una visión general de PCR ver PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, (Innis, M, Gelfand, D., Sninsky, J. y White, T., eds.), Academic Press, San Diego (1990). Los componentes de la reacción habitualmente son: Tris-HCl 10 mM, pH 8,3, cloruro potásico 50 mM, cloruro magnésico 1,5 mM, gelatina al 0,001%, dATP 200 μ M, dCTP 200 μ M, dGTP 200 μ M, dTTP 200 μ M, cebadores 0,4 μ M, y 100

unidades por mL de Taq polimerasa. Programa: 96°C durante 3 minutos, 30 ciclos de 96°C durante 45 segundos, 50°C durante 60 segundos, 72°C durante 60 segundos, seguidos de 72°C durante 5 minutos.

También pueden sintetizarse polinucleótidos mediante técnicas bien conocidas como las descritas en la literatura técnica. Ver, por ejemplo, Carruthers, *et al.*, Cold Spring Harbor Symp. Quant. Biol. 47:411-418 (1982), y Adams, *et al.*, J. Am. Chem. Soc. 105:661 (1983). Por tanto, pueden obtenerse fragmentos de ADN de doble cadena mediante la síntesis de la cadena complementaria y la hibridación de las cadenas juntas bajo condiciones apropiadas, o mediante la adición de la cadena complementaria utilizando ADN polimerasa con una secuencia de cebador apropiada.

En esta invención las fuentes adecuadas de la esterasa de ácido ferúlico utilizadas incluyen, pero sin limitación, *Neurospora crassa*, *Aspergillus spp.* y específicamente, *Aspergillus niger*. La xilanasa utilizada en esta invención puede derivarse de cualquier fuente adecuada incluyendo, pero sin limitación, *Trichoderma*, preferiblemente *Trichoderma reesei* y *Aspergillus spp.*

B. Preparación de Vectores Recombinantes

Para utilizar secuencias aisladas en las técnicas anteriores, se preparan vectores de ADN recombinante adecuados para la transformación de células vegetales. En la literatura técnica y científica se conocen bien y se describen las técnicas para transformar una amplia variedad de especies vegetales. Ver, por ejemplo, Weising, *et al.*, Ann. Rev. Genet. 22:421-477 (1988). Una secuencia de ADN que codifica el polipéptido deseado, por ejemplo una secuencia de ADNc que codifica la proteína FAE1 de longitud completa, preferiblemente se combinará con el inicio transcripcional y traduccional y las secuencias reguladoras de reconocimiento que dirigirán la transcripción de la secuencia del gen en los tejidos pretendidos de la planta transformada bajo las condiciones deseadas.

Los cassettes de expresión utilizados en la presente invención comprenden una secuencia de polinucleótidos que codifica una secuencia señal que reconoce la expresión de esterasa de ácido ferúlico a la vacuola.

Pueden identificarse promotores mediante el análisis de las secuencias 5' de un gen deseado. Pueden utilizarse secuencias características de las secuencias de promotor para identificar el promotor. Se han estudiado ampliamente las secuencias que controlan la expresión de genes eucariotas. Por ejemplo, los elementos de secuencias de promotor incluyen la secuencia de consenso caja TATA (TATAAT), la cual tiene normalmente de 20 a 30 pares de bases en dirección 5' del sitio de inicio de transcripción. En la mayoría de los casos, la caja TATA es necesaria para la iniciación de la transcripción precisa. En las plantas, más arriba en dirección 5' de la caja TATA, en las posiciones de -80 a -100, hay habitualmente un elemento promotor con una serie de adeninas que rodean el trinucleótido G (o T) N G. Messing, *et al.*, in GENETIC ENGINEERING IN PLANTS, pp. 221-227 (Kosage, Meredith and Hollaender, eds. (1983)).

Se conocen varios procedimientos por los expertos en la materia para identificar y caracterizar regiones promotoras en ADN genómico vegetal (ver, por ejemplo, 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); y Zhang, *et al.*, Plant Physiology 110:1069-1079 (1996)).

En la construcción de cassettes de expresión recombinantes de la invención, puede emplearse un fragmento de promotor vegetal que dirigirá la expresión del gen en todos los tejidos de una planta regenerada. Dichos promotores en la presente invención son referidos como promotores "constitutivos" y son activos bajo la mayoría de las condiciones ambientales y estados de desarrollo o diferenciación celular. Como ejemplos de promotores constitutivos se incluyen la región de iniciación de la transcripción 35S del virus del mosaico de la coliflor (CaMV), el promotor 1'- o 2'-derivado de T-ADN de *Agrobacterium tumefaciens*, los promotores de actina y ubiquitina y otras regiones de iniciación de transcripción de varios genes de plantas conocidos por los expertos en la materia. Un promotor constitutivo particularmente preferente es el promotor de actina del arroz (ver, McElroy, Plant Cell, 2:163 (1990)).

Alternativamente, el promotor de plantas puede dirigir la expresión del polinucleótido de la invención en un tejido específico (promotores específicos de tejido) o de lo contrario puede estar bajo un control ambiental más preciso (promotores inducibles). Como ejemplos de promotores específicos de tejido bajo control de desarrollo se incluyen promotores que inician la transcripción sólo en ciertos tejidos, tales como las hojas, las raíces o las semillas.

En un aspecto de la presente invención, la expresión de FAE tiene lugar tras haber cortado la planta, haberla arrancado del suelo o ingerida. De este modo, un promotor apropiado sería un promotor de senescencia. Por ejemplo, recientemente se ha demostrado que *BFN1* es una nucleasa expresada en hojas senescentes, Perez-Amador, *et al.*, Plant Physiol. 122:169 (2000). De un modo similar, SAG12, una cisteína proteasa también se encuentra en hojas senescentes (Noh & Amasino, Plant Mol. Biol. 41:181 (1999)). En una realización preferente, se utiliza el promotor del gen *gem* de *Festuca pratensis* para dirigir la expresión de FAE en hojas senescentes.

En otro aspecto, el FAE se expresaría tras la ingestión por un animal de forraje. Como promotores de ejemplo para este aspecto se incluiría el promotor Soybean Gmhs17.5 y el promotor de la leucina aminopeptidasa (LAP). El promotor GMhs17.5 es de un gen de una proteína de choque térmico e inicia la expresión si se incrementa la temperatura del ambiente. En el laboratorio, el choque térmico preferido es un incremento de 15°C durante 2 horas. No obstante, en condiciones que no son de laboratorio se darán incrementos adecuados de temperatura en silos y en el rumen de animales que han ingerido las plantas de la presente invención. El promotor LAP inicia la expresión del gen de FAE después

de dañar a la planta. Dicho daño tendrá lugar después de cortar la planta o después de la masticación por parte de un animal de forraje. Los promotores específicos de tejido que podrían utilizarse en esta invención incluyen promotores de genes que están expresados diferencialmente en las hojas de las hierbas. Un ejemplo de un promotor específico de hoja es el promotor *rbcS* de tomate (Proc. Natl Acad. Sci. USA 84:7104 (1987)). Este promotor normalmente regula un gen determinado por tener importancia en la fotosíntesis.

Para una expresión correcta del polipéptido, debe incluirse una región de poliadenilación en el extremo 3' de la región codificante. La región de poliadenilación puede derivarse del gen fúngico natural, de una serie de genes fúngicos o vegetales, o del T-ADN. Estas secuencias son bien conocidas y están disponibles fácilmente para los expertos en la materia.

Los vectores de expresión de la presente invención también contienen secuencias señal. Éstas son polinucleótidos hallados en los extremos 5' y/o 3' de la región codificante y sirven para dirigir la expresión del gen a los orgánulos celulares específicos. Estas secuencias señal pueden estar ambas en dirección 5' ó 3' de la región codificante. Algunos ejemplos preferentes de secuencias señal en dirección 5' incluyen la secuencia de aleuraina de cebada (Rogers, Proc. Nat'l Acad. Sci. USA 82:6512 (1985) que reconoce vacuolas.

El vector que comprende los cassettes de expresión (por ejemplo, promotores y/o regiones codificantes) de la invención comprenderá habitualmente un gen marcador que conferirá un fenotipo seleccionable en las células vegetales. Por ejemplo, el marcador puede codificar resistencia a biocida, particularmente resistencia a antibiótico, tal como resistencia a la higromicina, la canamicina, el G418, la bleomicina, o resistencia a herbicida, tal como la resistencia a cloroslufuron o Basta.

C. Producción de Plantas Transgénicas

Las construcciones de ADN de la presente invención se pueden introducir en el genoma de la planta huésped deseada mediante una variedad de técnicas convencionales. Por ejemplo, la construcción de ADN se puede introducir directamente el tejido vegetal utilizando métodos balísticos, tales como el bombardeo de partículas de ADN o las construcciones se pueden introducir directamente en el ADN genómico de la célula vegetal utilizando técnicas, tales como electroporación y microinyección de protoplastos de células vegetales. Alternativamente, las construcciones de ADN se pueden combinar con regiones flanqueantes de T-ADN adecuadas e introducir en un vector huésped *Agrobacterium tumefaciens* convencional. Las funciones de virulencia del *Agrobacterium tumefaciens* del huésped dirigirán la inserción de la construcción y el marcador adyacente en el ADN de la célula vegetal cuando la célula es infectada por la bacteria.

Véase, 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) para métodos de ejemplo para el cultivo y transformación de hierbas.

Las técnicas de microinyección son conocidas en el sector y están bien descritas en la literatura científica y de patentes. La introducción de construcciones de ADN utilizando la precipitación con polietilenglicol se describe en Paszkowski, *et al.*, Embo J. 3:2717-2722 (1984). Las técnicas de electroporación se describen en Fromm, *et al.*, Proc. Natl. Acad. Sci. USA 82:5824 (1985).

Las técnicas de transformación mediadas por *Agrobacterium tumefaciens*, incluyendo el desarme y la utilización de vectores binarios, están bien descritas en la literatura científica. Véase, por ejemplo, Horsch, *et al.*, Science 233:496-498 (1984), y Fraley, *et al.*, Proc. Natl. Acad. Sci. USA 80:4803 (1983). La Patente de Estados Unidos 5.591.616 describe técnicas de transformación mediadas por *Agrobacterium* en monocotiledóneas.

Las técnicas de transformación balísticas se describen en Klein, *et al.*, Nature 327:70-73 (1987). En una realización preferida, se utiliza una pistola de bombardeo de partículas (PIG) para transformar las células vegetales de la presente invención.

Las células vegetales transformadas que se derivan mediante cualquiera de las técnicas de transformación anteriores se pueden cultivar para regenerar una planta completa que posee el genotipo transformado y, de este modo, el fenotipo deseado, tal como una digestibilidad mejorada. Dichas técnicas de regeneración dependen de la manipulación de ciertas fitohormonas en un medio de crecimiento para cultivos de tejidos, habitualmente dependen de un marcador de biocida y/o herbicida que se han introducido juntos con las secuencias de nucleótidos deseadas. La regeneración de plantas de protoplastos cultivados se describe en Evans, *et al.*, PROTOPLASTS ISOLATION AND CULTURE, HANDBOOK OF PLANT CELL CULTURE, pp. 124-176, Macmillan Publishing Company, New York, 1983; y Binding, REGENERATION OF PLANTS, PLANT PROTOPLASTS, pp. 21-73, CRC Press, Boca Raton, 1985. La regeneración también se puede obtener de callos de plantas, explantes, órganos, o partes de los mismos. Dichas técnicas de regeneración se describen en general en Klee, *et al.*, Ann. Rev. of Plant Phys. 38:467-486 (1987).

Para determinar la presencia o el aumento de la actividad de FAE1, se puede utilizar un ensayo enzimático o un ensayo para medir los aumentos y disminuciones en las velocidades de fermentación. Estos ensayos están fácilmente disponibles en la literatura y los expertos en la materia los pueden encontrar fácilmente.

ES 2 325 908 T3

Un experto reconocerá que se pueden utilizar otros ensayos para detectar la presencia o ausencia de FAE1. Estos ensayos incluyen, pero sin limitación, inmunoensayos y ensayos de detección electroforética (con tinción o transferencia western).

Los polinucleótidos de la presente invención se pueden utilizar para conferir características deseadas esencialmente en cualquier planta. Sin embargo, la principal utilidad de la presente invención se encuentra en la mejor digestibilidad de las plantas de forraje. De este modo, se prevé que las plantas transgénicas de la presente invención incluirán, pero sin limitación, los siguientes géneros *Lolium*, *Festuca*, *Triticum*, *Avena*, y *Medicago*. Los genes FAE1 de la presente invención son particularmente útiles en la producción de plantas transgénicas en el género *Lolium*.

Un experto en la material reconocerá que después de que el cassette de expresión se incorpore de manera estable en plantas transgénicas y se confirme que sea operable, se puede introducir en otras plantas mediante cruzamiento sexual. Se puede utilizar cualquiera de un conjunto de técnicas de reproducción estándar dependiendo de la especie a cruzar.

Tal como se ha mencionado anteriormente, las plantas transgénicas de la presente invención se pueden utilizar como un cultivo de forraje para animales, tales como ganado, ovejas, cabras y caballos. Además, los procedimientos de la presente invención se pueden utilizar para transformar cualquier planta en la cual se desee la expresión de FAE. Por ejemplo, resulta ventajoso romper las paredes celulares durante la conversión de la biomasa o durante el procesamiento de las plantas para alimentación. La presente invención ayudaría a conseguir este objetivo de manera eficaz y económica.

Los procedimientos de la presente invención también se pueden utilizar para proporcionar enzimas adicionales para aumentar la disponibilidad de azúcares fermentables en plantas. Los carbohidratos de las plantas se pueden someter a una modificación adicional, de manera exógena o endógena, mediante la acción de otras enzimas. Entre dichas enzimas se incluyen, pero sin limitación, endoglucanasas, xilosidasas y/o biohidrolasas celulares. Estas enzimas se pueden disponer en un cassette de expresión proporcionado para la presente invención (es decir, endógeno) o aplicado a las paredes de células vegetales (es decir, exógeno) para aumentar la disponibilidad de monosacáridos y/o disacáridos.

Otras plantas diferentes a hierbas pueden ser útiles en la presente invención. Por ejemplo, se considera específicamente que el maíz es útil. La hierba *Festuca* es similar al maíz en la estructura de la pared celular y, por tanto, proporciona un buen modelo de la capacidad para aumentar los carbohidratos fermentables en el maíz. Otras plantas útiles que se consideran para su uso en la presente invención son *Festuca*, *Lolium*, *Zea*, *Avena*, *Sorghum*, *Millet* (cereales tropicales), *Miscanthus* (una hierba con potencial para su uso como cultivo para energía como biomasa), *Cenchrus*, *Dichanthium*, *Brachiaria* y *Paspalum* (hierbas de tipo tropical apomítica) y *Poa* (Kentucky bluegrass).

Las paredes celulares de hierbas de forraje forman un 30-80% de materia seca de forraje que representa un fuente principal de energía para los rumiantes, pero menos de un 50% de esta fracción es digerida por el animal. La conversión de la biomasa de poco valor en azúcares y etanol también es inferior a óptimo debido a la indisponibilidad de los carbohidratos de los alimentos, incluyendo, pero sin limitación, bagazo, paja de arroz, forraje de maíz y fibras de maíz.

El ácido ferúlico y otros ácidos hidroxycinámicos son ésteres unidos a redifuso arabinosilo en arabinoxilanos, y juegan un papel principal en la reticulación de xilanos con lignina, dando lugar a paredes celulares menos degradables. La esterasa del ácido ferúlico (FAE) puede liberar ácido ferúlico (FA) tanto monomérico como dimérico a partir de arabinoxilanos, haciendo la pared celular más susceptible de un ataque enzimático posterior. Las plantas transgénicas se produjeron expresando un gen de FAE siguiendo un bombardeo de microproyectiles de cultivos celulares. Las mediciones del nivel de actividad de FAE de diferentes vectores que dirigen FAE a la vacuola, ER y apoplasto bajo promotores constitutivos o inducibles (choque térmico) muestra que por lo menos para la expresión constitutiva de FAE dirigido a la vacuola, la actividad era la más elevada en hojas jóvenes y aumentaba a lo largo de las láminas de las hojas. También se demuestra que la expresión de FAE da lugar a la liberación de FA monomérico y dimérico de las paredes celulares en la muerte celular y esto se aumentó varias veces mediante la adición de xilanas. Se observa un efecto de la expresión de FAE en el contenido de ferulato monomérico y dimérico unido a éster de la pared celular en comparación con las plantas de control (no transformadas). En general, cuanto más bajos son los niveles de monómeros y, en particular, dímeros de ácidos hidroxycinámicos en las hojas, mayor es la digestibilidad y/o disponibilidad de carbohidratos complejos para la conversión.

La senescencia es la fase terminal en el desarrollo de las hojas y aparece sin crecimiento o morfogénesis. Por lo tanto, el metabolismo/la fisiología de esta etapa de la vida de las hojas se puede dirigir directamente para la alteración con el impacto perjudicial mínimo sobre un desarrollo inicial. La senescencia prosigue a la madurez de las hojas y esta asociada con la expresión de genes específicos. Estos genes y sus elementos de control se pueden aprovechar para manipular las características de desarrollo, adaptación, productividad y calidad en plantas de cosecha. Parece haber una buena conversación de la fisiología de la senescencia a lo largo del rango de especies de plantas superiores y, de este modo, estos promotores son útiles en la presente invención.

Los siguientes parámetros y ejemplos se ofrecen para permitir a un experto en la material entender más claramente y llevar a la práctica la presente invención.

ES 2 325 908 T3

En la siguiente descripción experimental. Se aplican las siguientes abreviaturas: eq (equivalentes); M (Molar); μ M (micromolar); N (Normal); mol (moles); mmol (milimoles); μ mol (micromoles); nmol (nanomoles); g (gramos); mg (miligramos); kg (kilogramos); μ g (microgramos); L (litros); ml (mililitros); μ l (microlitros); cm (centímetros); mm (milímetros); μ m (micrómetros); nm (nanómetros); °C. (grados Centígrados); h (horas); min (minutos); s (segundos); ms (milisegundos); Ci (Curies) mCi (miliCuries); μ Ci (microCuries); TLC (cromatografía de capa fina); Et (etilo), Me (metilo).

Ejemplo 1

Preparación de secuencias de ADN que codifican enzimas

Se utiliza un clon genómico para FAE 1 (véase las figuras 1-3) como el punto de partida para la preparación de una secuencia de ADN que codifica FAE1 sin intrón. La secuencia para el clon genómico se proporciona en las figuras 2 y 3. Los fragmentos separados para ambos exones de FAE se recuperaron por PCR a partir de un fragmento EcoRI de 5,5 kb del clon genómico en pLITMUS28, y se creó "ADNc" mediante el solapamiento de PCR. Véase la figura 4.

Se utilizaron dos cebadores 5'. FAE-S5 que amplifica el marco de lectura completo (incluyendo la señal de Aspergillus), y FAE-N5 que amplifica sólo la proteína madura (es decir, no tiene señal). Se optimizan un conjunto de codones (subrayados en las secuencias de cebadores siguientes=). El producto de solapamiento puede derivar de cebadores FAE-I5 (tipo salvaje) o FAE-I3 (Ser conservada cambiada a Ala), lo que permite la producción de proteína enzimáticamente inactiva para comprobar la toxicidad. Tal como se observa en la figura 5, el solapamiento de los productos de PCR realizado con FAE-I5 y FAE-I3 crea dos posibles marcos de lectura ininterrumpidos. Si el complemento a FAE-I5 sirve como plantilla cuando se recombina, entonces la proteína codificada mantiene el grupo serina y la esterasa es funcional (la serina destacada está en el sitio activo). Si el cebador FAE-I3 sirve como plantilla, la serina se sustituye con una alanina y la esterasa se inactiva (la alanina destacada en la secuencia de aminoácidos inferior se proporciona en la figura 5).

Cuando sea posible se ha optimizado el uso de los codones en los marcos de lectura construidos (elección del codón en base a las preferencias de cebada publicadas).

```
FAE-I5 (SEC ID NO:_)
GGCGCCGAGGGAGTGGCCGGTCACGGTCAGCGCGTAGTCC 40-mer
FAE-I3 (SEC ID NO:_)
CCGGCCACGCCCTCGGCGCCTCCCTGGCGGCACTC 35-mer
FAE-N5 (SEC ID NO:_)
CTAAAGCTTACCATGGCGGGCCGCTCCACGCAGGGCATCTCCGA 44-mer
FAE-S5 (SEC ID NO:_)
CTAAAGCTTAACATGAAGCAGTTCTCCGCCAA 32-mer
FAE-3 (SEC ID NO:_)
TCTAAGCTTGCGGCCGCGACCGGCCAGGTGCATGCGCCGCTCGTCATCCC 50-MER
```

Ejemplo 2

Preparación de vectores

Los vectores presentaban la estructura general mostrada en la figura 6.

A. Serie de vectores para la transformación de plantas

Los vectores de expresión iniciales se basaron en pCOR105 [promotor de la actina del arroz - McElroy *et al.* MGG 231:150-160 (1991)] (Figura 7). Los sitios Not y SstI de pCOR105 se destruyeron en primer lugar [cortados con NotI y SstI, seguido por inactivación por calor y tratamiento con T4 ADN polimerasa en presencia de dNTPs] utilizando métodos estándar tal como se describe en Maniatis *et al.* o siguiendo las instrucciones para enzimas del fabricante para simplificar posteriormente la manipulación del cassette Not y permitir el uso de un único sitio Sst (ver continuación).

ES 2 325 908 T3

Los terminadores nos de **pMA406** (Ainley & Key (1990) PMB 14:949-60) se amplificó por PCR utilizando los cebadores TER5 y TER3 para generar un fragmento con la siguiente secuencia (SEC ID NO: _):

(Pst1) (Not 1)

(AGACTGCAGACCATGGCGGCCGCGKAACCACTGAAGGATGAGCTGTAAAG
 AAGCAGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTT
 GCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGT
 AATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTA
 GAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGC
 AAAGTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGATA
 AGCTT CTA GATCT (donde K = G o T)

(HindIII) (XbaI)

Una redundancia en el cebador TER5 (GCGKAA) crea fragmentos que tienen o un codón de parada (TAA) o un codón de glutamato (GAA) en un marco de lectura. El codón de glutamato se encuentra en el marco con un motivo KDEL en dirección 3'.

El fragmento y vector pCOR105 modificado se cortaron con PstI y XbaI, según las instrucciones de los fabricantes, se purificaron en gel los fragmentos relevantes, se ligaron con T4 ADN ligasa y se transformaron en *E. coli*. A continuación, los clones resultantes se secuenciaron para establecer qué TER5 alternativas estaban presentes.

A continuación, los vectores de expresión FAE iniciales se construyeron a partir de estos vectores mediante la inserción de productos PCR FAE-S5/FAE-3 (T4 ADN polimerasa "pulida" en presencia de dNTPs, purificados y digeridos con NotI, clonados en un vector digerido con EcoRV y NotI) o productos PCR FAE-N5/FAE-3 (purificados y digeridos con NotI, clonados en un vector digerido con NotI y tratado con fosfatasa alcalina intestinal de ternera).

Los clones terminadores pCOR105-nos también se modificaron mediante la adición de productos PCR Ale-5/ALE-3 (que codifican péptidos señal de aleuraina de cebada de tipo salvaje y modificados, véase a continuación para más detalles). Los productos se "pulieron" con T4 ADN polimerasa en presencia de dNTPs, se purificaron y cortaron con NotI, a continuación se clonaron en vectores digeridos con EcoRV y NotI. La adición de las secuencias ALE crea una serie de vectores que pueden expresar un marco de lectura insertado en los sitios NotI o NcoI como una fusión a la señal de aleuraina de cebada, con o sin el motivo de reconocimiento vacuolar, y con o sin un motivo de retención en ER. Los sitios HindIII que flanquean el codón de inicio de la traducción y el terminador transcripcional permiten un sencillo movimiento de las unidades de transcripción entre los vectores de expresión que proporcionan secuencias promotoras diferentes. (Véase la figura 8 que representa el vector ALE-TER genérico).

Las secuencias de los vectores se confirmaron mediante secuenciación. Se encontraron dos artefactos. En primer lugar, se observó que el codón redundante en TER5 era AAA en un clon, el cual se utilizó posteriormente como la fuente de todas las fusiones de KDEL (es decir, la secuencia del péptido es KPLKDEL, en lugar de EPLKDEL tal como se diseñó). Véase la figura 9. En segundo lugar, se observa una base adicional en el sitio del codón redundante en un clon, creando un péptido terminal desplazado en el marco de lectura (ETTEG, Figura 10) que se utilizó como control en algunas construcciones.

La explotación de la disposición modular de los péptidos señales en la serie de vectores anteriores permitió que se crearan varias combinaciones de FAE y motivos de reconocimiento utilizando procedimientos estándar de biología molecular (es decir, digesto de restricción, purificación de fragmentos relevantes y unión según sea apropiado). Por ejemplo, el fragmento NotI que contenía el marco de lectura de FAE se insertó en el sitio NotI del clon desplazado en el marco de lectura descrito anteriormente para crear el vector **pTP3.1**. Se insertó COOH terminal de *Aspergillus* nativo en un clon FAE-S5/FAE-3 como un fragmento SphI (T4 ADN polimerasa pulida) - NcoI del gen genómico FAE (sustituyendo el fragmento NotI (T4 ADN polimerasa pulida) - NcoI, creando el vector **pTP4a2**, que a continuación codifica la FAE de *Aspergillus* completa no modificada. La sustitución del fragmento Sall/XbaI de pTP3.1 con el de pTP4a2 creó entonces **pTP11.1**, que no codifica la FAE con un COOH terminal de *Aspergillus* nativo, sino una señal N-terminal de aleuraina de cebada.

Brevemente, otros vectores fabricados en esta serie fueron: **pTP8.5**, el fragmento NotI de FAE insertado en el sitio NotI de una construcción de COOH terminal desplazado en el marco-ALE, N-terminal de aleuraina; **pTP5.1**, sustitución del COOH terminal de *Aspergillus* nativo por un péptido KDEL (intercambio de fragmentos NotI/XbaI), la señal N-terminal de *Aspergillus* permanece; **pTU4.4**, fragmento BamHI de pTP11.1 sustituye el fragmento BamHI

ES 2 325 908 T3

de pTP5.1, crea el marco de lectura de FAE fusionado a los extremos N y C terminales heterólogos (señal de aleuraina y KDEL).

Los vectores en los que el motivo de reconocimiento vacuolar de aleuraina NPIR se sustituyó por NPGR (hallado como inactivo en algunos ensayos con plantas) se crearon mediante la sustitución de un fragmento EcoRV/NotI por un producto PCR de ALE que había sido cortado con AccI (T4 ADN polimerasa pulida) y NotI (vectores **pTT5.5** y **pTT5.14**, extremo COOH de *Aspergillus*). El fragmento BamHI de pTT5.5 se utiliza para sustituir el de pTP5.1 para producir **pTU5**, creando un marco de lectura de FAE fusionado a los extremos N y C terminales heterólogos (modificación NPGR de la señal de aleuraina y KDEL). La señal de aleuraina también se modificó mediante mutagénesis de PCR para eliminar el motivo NPIR de reconocimiento vacuolar en su totalidad (dirigido por el cebador ALECUT, que contiene un sitio NotI para permitir el intercambio de fragmentos BgIII/NotI. La eliminación de NPIR se creó de esta manera en pTP11.1 (creando **pUA4.4**), y en pTP5.1 mediante el intercambio de fragmentos BamHI por pUA4.4 (creando **pUG4**).

Finalmente, la mutagénesis por PCR, utilizando el solapamiento de fragmentos generados por los cebadores GLY3 y GLYB, también se utilizó para alterar un sitio de glicosilación potencial (un codón de asparagina cambió a aspartato, tal como se lleva a cabo por ejemplo en Chen, H.M., C. Ford & P. J. Reilly (1994) Biochem J 301 275-281 Substitution of asparagine residues in *Aspergillus awamori* glucoamylase by site-directed mutagenesis to eliminate N-glycosylation and inactivation by deamidation; véase los datos de las secuencias para el cambio exacto, vector **pTP10.1**).

Cebadores de PCR

TER-5 (SEC ID NO: _)

AGACTGCAGACCATGGCGGCCGCG**KA**ACCACTGAAGGATGAGCTGTAAAG
AAGCAGATCGTTCAAACATTTG 72-MER (El codón de parada KDEL
está subrayado)

TER-NOT (SEC ID NO: _)

AAGACTGCAGACCATGGCGG 20-MER

TER-3 (SEC ID NO: _)

AGATCTAGAAGCTTATCGATCTAGTAACATAGATGACACC

ALECUT (SEC ID NO: _)

CTAGGCGGCCGCGCGGGAGGAGGCGACGGCGAC

GLYB (SEC ID NO: _)

GAGGGTGTATTTCGGTATCGAGTTGCAGGTTTCGTATC

GLY3 (SEC ID NO: _)

CTCGATACCCATTACACCCTCACGCCTTTCGA

B. Construcción de diferentes vectores promotores.

Se utilizaron diversos promotores para optimizar la expresión y para establecer la constitutividad, la inducción al choque térmico y el aumento de la senescencia.

i. Promotor de actina de arroz y primer intrón

Se construyeron los vectores iniciales (figuras 11 y 12) a partir de pCOR105 que se observó posteriormente que contenía una eliminación de 5 pb en relación con la secuencia publicada que destruye el sitio AccI (GTAGGTAGAC, bases eliminadas subrayadas) y puede afectar al corte y empalme en el sitio 3' adyacente. La secuencia de actina de arroz original en esta región (GTAGGTAG) se restauró por tanto utilizando el oligonucleótido **NCO-ACT** (CTCAC-CATGGTAAGCTTCTACC TACAAAAAAGCTCCGCA) mediante la sustitución del fragmento BgIII/HindIII por un producto PCR para producir el vector **pPQ10.1**.

Se encuentra presente un elemento repetitivo del arroz en la región de dirección 5' del promotor de actina utilizado en **pCOR105**; dado que esto podría tener efectos impredecibles sobre el vector de expresión, se extrajo de **pPQ10.1**

mediante la eliminación del fragmento KpnI/EcoRI (completado en el extremo con T4 polimerasa y unido tras el digesto, restaurando EcoRI, pero no KpnI), para producir el vector **pGT6**. El fragmento HindIII que contenía el marco de lectura de FAE y el terminador nos de **pTP3.1** (véase el Ejemplo 2A) se insertó a continuación en pGT6 para producir la construcción **pJO6.3**.

ii. Promotor del choque térmico de soja

Se obtuvo un promotor de choque térmico de soja de un HSP de 23 kD a partir de pMA406 (Ainley & Key (1990) PMB 14: 949-60). Se había observado previamente que cuando este promotor se fusionaba a β -glucuronidasa (Jefferson *et al* 1987 EMBO J 6:3901-3907) era inducible por un choque térmico a 10°C y mostraba una expresión estable durante 24-48 horas (datos no mostrados). Las fusiones con β -glucuronidasa son un marcador de fusión sensible y versátil en plantas superiores. La construcción de los vectores HS de cointegración se indica a continuación.

iii. Promotor de expresión del aumento de la senescencia (*See1*) de *Lolium multiflorum*

El promotor y la secuencia señal (incluyendo el motivo NPIR) del gen LSee1 se amplificó a partir de cv Tribune de *Lolium multiflorum* con los oligonucleótidos SEE-NCO y SEE-VAC, y se clonaron como una sustitución Asp718/NotI de la región promotora del vector **pTP11.1**. Tras la secuenciación para cribar los artefactos de PCR, se eligió uno de los tres clones idénticos (**pUB8.11**).

El promotor See1 del maíz se clonó previamente y el número de acceso EMBL es AX050343. Véase WO0070061.

La versión de See1 de *Lolium* también se clonó previamente (Qiang Li (2000) Studies on leaf senescence and its genetic manipulation in *Lolium mutiflorum* PhD Thesis University of Wales, Aberystwyth) y se ha observado que era inducible por senescencia cuando se utilizaba para dirigir tanto la β -glucuronidasa como el gen *ipt* de *Agrobacterium*.

Se construyó un derivado dirigido a apoplastos mediante la amplificación del motivo del Inhibidor de Proteasa de Patata (PPI) con los cebadores PPI-AP6 y SEE-ATG y la clonación del producto como un fragmento NgoMIV/NotI en pUB8.11 (digesto parcial de NgoMIV), para producir un vector **pJQ5.2**. Este vector presenta tanto el promotor inducido por senescencia como la secuencia dirigida de apoplasto con el gen a expresar insertado en la zona de dirección 3' de la secuencia del apoplasto.

Cebadores de PCR

SEE-VAC (SEC ID NO: _)

AACCATGGCGGCCGCGCTCGGTGACGGGCCGGAT

SEE-NCO (SEC ID NO: _)

TTCGGTACCATGGCCAGGTATAATTATGG

SEE-ATG (SEC ID NO: _)

CTGCGCCGGCGAGATGGMCGTGCACAAGGAG

C. Construcción de secuencias de reconocimiento

Con el fin de examinar si la localización de la enzima tendría efecto o no sobre el contenido de ácido fenólico de la pared celular, se utilizaron varias secuencias señal. Las secuencias de reconocimiento se añadieron al extremo N-terminal o al extremo c-terminal del gen de interés.

i. Secuencias señal N-terminal

Se utilizaron seis secuencias señal N-terminal

(a) El extremo de FAE de *Aspergillus* nativo más la señal de excreción [localización de apoplasto]

Es del clon original y tiene la secuencia de péptido:

MKQFSAKHVLAVVVTAGHALAAASTQGI.

ES 2 325 908 T3

(b) El extreme de Aspergillus maduro sin señal de excreción [localización citoplasmática]

La secuencia del péptido es MAAASTQGI (el motivo subrayado es común en todas las construcciones). El truncamiento de la secuencia señal en (a) anterior se llevó a cabo mediante PCR con el cebador mutagénico FAE-N5.

(c) La señal de aleuraina de cebada, incluyendo el motivo NPIR intacto [localización de la vacuola]

La secuencia señal vacuolar de aleuraina de cebada (véase la figura 13; Base de datos Swissprot de número de acceso P05167) derivaba completamente del solapamiento de los cebadores (ALE-5, ALE-3, ALE-CUT ALE-CAP-5 y ALE CAP-3). Tras la hibridación de cebadores a 37°C y la extensión con T4 ADN polimerasa en presencia de dNTPs según las instrucciones del fabricante, se llevó a cabo la PCR con los cebadores flanqueantes ALE-5 y ALE-3. El producto se “pulio” con T4 ADN polimerasa, se purificó, se digirió con NotI y se clonó en el vector pCOR105-terminador nos digerido con EcoRV/NotI (ver anteriormente). ALE-3 contiene redundancias, de manera que se pueden recuperar los clones que codifican los motivos NPIR o NPGR. Se produjeron dos versiones de la señal, son y sin el motivo de reconocimiento de la vacuola, para obtener posibles secuencias señales NPIR vacuolar y de apoplasto (NPGR).

Cebadores de PCR

ALE-5 (SEC ID NO: _)

GGAATTCGTAGACAAGCTTACMATGGCCACGCCGCGTCCT 41-MER

ALE-3 (SEC ID NO: _)

TATCCATGGCGGCCGCGCGGTGCGGTGACGGGCCGGMCGGGTTGGAGTC GGCGAA

55-MER

ALE-CUT (SEC ID NO: _)

CTAGGCGGCCGCGCGGGAGGAGGCGACGGCGAC 33-mer

ALECAP-5 (SEC ID NO: _)

GCGACGGCGACGGCGGCCGTGGCCAGCACGGCGAGCGCCAGGAGGAGG ACGCGG

54-MER

ALECAP-3 (SEC ID NO: _)

TCGCCGTCGCCTCCTCCTCCTTCGCCGACT 33-MER

(d) La señal de aleuraina de cebada mutada a un motivo NPGR [localización citoplasmática]

(e) El motivo de reconocimiento en golgi de la sialil transferasa de rata [localización en golgi]

Se fabricó un vector de reconocimiento del golgi, **pJQ3.2**, mediante la inserción de un marco de lectura que codifica el motivo de sialil transferasa de rata (RST) pertinente (Véase la figura 14. Se observa que el motivo RST actúa en plantas según 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) en el vector pPQ10.1, y la sustitución del fragmento promotor/señal EcoRI/NotI de **pJO6.3** por el fragmento de este vector. Brevemente, el motivo RST se construyó mediante la hibridación de oligonucleótidos RST-F1A, RST-F1 B, RST-F2A y RST-F2B, y la amplificación del producto con RST-5AD y RST-3A. Este producto se clonó y secuenció. Se observó que los clones presentaban una eliminación que se corrigió mediante PCR con RST-RPT, seguido de solapamiento PCR y la clonación de los productos.

Cebadores de PCR

RST-5AD (SEC ID NO: _)

ACTAAGCTTAAGGAGATATAACAATGATCCACACCAACCTCAA

RST-F1A (SEC ID NO: _)

ES 2 325 908 T3

TTCCATGATCCACACCAACCTCAAAAAGAAGTTCTCCCTCTTCAT

RST-F1B (SEC ID NO: _)

AGAGTGATCACGGCGAAGAGGAGGAAGACGAGGATGAAGAGGGAGAACTTCT TTT

RST-F2A (SEC ID NO: _)

TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGCCCTCACCTCCAA

GCCAAGGA

RST-F2B (SEC ID NO: _)

CATTTGGAACCTCCTTGGCTTGGAGGGTG

RST-3A (SEC ID NO: _)

AACCATGGCGGCCGCCATTTGGAACCTCCTTGGCT

RST-RPT (SEC ID NO: _)

TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGCCCTCACCTCC

AAGCCAAGGA

(f) Motivo [localización citoplasmática]

(g) El motivo en apoplasto del inhibidor de proteasa de patata II (PPI) [localización en apoplasto]

Se diseñó un marco de lectura de reconocimiento en apoplasto para codificar el motivo de inhibidor de proteasa de patata II (PPI) relevante (véase la figura 15) y se clonó en **pJO6.3**, para producir el vector **pJQ4.9**. Brevemente, el motivo PPI se construyó mediante la hibridación de los oligonucleótidos PPI-AP1, PPI-AP2, PPI-AP3, PPI-AP4, PPI-AP5 y PPI-AP6, y la clonación de este producto como un fragmento HindIII/NotI en el vector pPQ10.1; el fragmento de promotor/señal EcoRI/NotI de pJO6.3 se substituyó a continuación por el fragmento equivalente del vector pPQ10.1 modificado.

Cebadores de PCR

PPI-AP1 (SEC ID NO: _)

GGAATTCGTAGACAAGCTTACMATGGMCGTGCACAAGGAGGT

PPI-AP2 (SEC ID NO: _)

GATCAGGAGGTAGGCWACGAAGTTWACCTCCTTGTGC

PPI-AP3 (SEC ID NO: _)

CCTACCTCCTGATCGTSCCTCGGCCTCCTCTTGCTCGT

PPI-AP4 (SEC ID NO: _)

CCTTGGCGTCCACGTGCTCCATGGCGGAWACGAGCAAGAGGAG

PPI-AP5 (SEC ID NO: _)

GTGGACGCCAAGGCCTGCACCCCKCGAGTGCGGCAACCTC

PPI-AP6 (SEC ID NO: _)

GGAATTCGCGGCCCGCCGGGCAGATGCCGAAGCCGAGGTTGCCGCACT

ii. *Secuencias señal de los extremos C-terminal*

Se utilizaron cuatro secuencias señal C-terminales

(a) Extremo de *Aspergillus* nativo, [CTW] (vectores de vacuola y apoplasto). Derivaba directamente del clon genómico (véase el ejemplo 1) como un fragmento NcoI-SphI (extremo Sph completado con T4 polimerasa) que sustituye la región NcoI-NotI de un vector estándar de actina-FAE (extremo NotI completado con T4 ADN polimerasa).

ES 2 325 908 T3

(b) Enlazador del vector de expresión solo [CTW-PVAAA] (extreme C-terminal optimizado de planta para vectores de vacuola, golgi y apoplasto)

CTW es la secuencia peptídica del extreme COOH de FAE de *Aspergillus* y se proporciona aquí mediante el oligo FAE3. En este cebador el marco de lectura se extiende para proporcionar los aminoácidos adicionales PVAAA que son codificados parcialmente por el sitio NotI utilizado para clonar las señales en dirección 3'. Véase c) y d) a continuación. Algunos aminoácidos/motivos con COOH pueden afectar al reconocimiento de compartimento, se espera que las secuencias PVAAA sean neutras en este aspecto aunque el extremo de *Aspergillus* puede no serlo.

(c) Enlazador más KPLKDEL [primer K es el artefacto cebador que presente ser E] {vectores de retención en ER)

Estas secuencias son proporcionadas por el cebador TER5 introducido durante la PCR para generar el fragmento terminador nos, e identificado mediante secuenciación en un clon específico. El reconocimiento de KDEL se ha demostrado en plantas según 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) Enlazador más ETTEG [desplazamiento del marco de (c)] (pérdida de vectores de vacuola de retención en ER)

Estas secuencias son proporcionadas por el cebador TER5 introducido durante la PCR para generar el fragmento terminador nos, e identificado mediante la secuenciación en un clon específico (véase el Ejemplo 2A).

La señal de KDEL es para la retención en ER, mientras que otras proporcionan controles. Se utilizó un desplazamiento del marco en la región [A adicional] en las posteriores construcciones para destruir la señal de retención KDEL en ER.

El enlazador utilizado en las secuencias de reconocimiento C-terminal fue PVAAA.

D. Vectores de co-integración y co-transformación

Vectores de co-transformación

Se utilizó un gen de resistencia a higromicina dirigido por un promotor CaMV345S (pRob5) (35S-HYG-CMV en pUC18 (HYG modificado, derivado de pGL2) Bilang *et al* (1991) Gene 100:247-50) para experimentos de co-transformación con vectores **pTT3** y **pTP3.1**, **pJQ4.9**, **pJQ3.2**, **pJO6.3**, **pJQ5.2**, **pUB8.1**.

Vectores de Co-integración

1. Construcciones de promotores de actina - pTR2.22, pTR6.1, pTR8.1, pTR9.4, pTR7.1, pTT5.5 y 5.1

La región CAMV35S-hyg de pAJEB64TCA [un vector de expresión de planta construido por Andy Bettany en IGER que contiene CaMV-HYG de pTRA151 (Zheng *et al* 1991 Plant Physiol 97: 832-835) (CaMV35S-HYG-terminador tml como cassette clonable en pUC4) clonado en un sitio KpnI de pCOR105] se añadió como un fragmento HindIII en el sitio KpnI (extremo romo de T4 polimerasa) de pTP4a2, en orientación divergente a FAE para crear **pTR2.22**. El fragmento FAE/Nos HindIII de este vector se substituyó tal como se indica en los vectores de co-expresión. De pTP5.1 por **pTR6.1**, de pTP10.1 por **pTR8.1**, de pTP11.1 por **pTR9.4**, las secuencias señales de FAE en pTR2.22 se substituyeron como fragmentos HindIII/BglIII en **pTR7.1** (fragmento de pT09.1). Los productos PCR (ALE5/ALE-G) se digirieron con AccI y T4 polimerasa, se pulió, seguido de la digestión con NotI y la clonación en pTR2.22 digerido con EcoRV/NotI para producir los clones **pTT5.5** y **5.1**.

Cebador de PCR

ALE-G

TATCCATGGCGGCCGCGCGGTTCGGTGACGGGCCGCGCCGGTGGAGTC GGCGAA

2. Construcciones de promotores de actina -pUF1, pUA1 K3, pUH4, pUH5, pUH6, pUH7, pUH8, pUH9

El gen HygR de pAJEB64TCA, dirigido por el promotor CaMV, se clonó en primer lugar como un fragmento HindIII de extremo completo en el sitio XbaI de extremo completo de pTP3.1 para producir **pHOX3**. Para facilitar la clonación el sitio HindIII en dirección 3' se destruyó para crear **pUA1 K3** y la sustitución del fragmento HindIII con FAE/terminador Nos en este vector se llevó a cabo tal como se indica a continuación. De pTP5.1 por **pUF1**, de pTP11.1 por **pUH4**, de pTP8.5 por **UH5**, de pTT5 por **pUH6**, de pUA4.4 por **pUH7**, de pTU5 por **pUH8** y de pUG4 por **pUH9**.

3. Construcciones de promotores del choque térmico - pUH10, pUH12, pUC5.11

Se fabricó un vector de co-transformación en el que la FAE se expresa a partir del promotor de choque térmico de la soja mediante la modificación de pMA406 para eliminar el terminador nos (linealizado con BglII y purificado por gel, digerido con KpnI, pulido con T4 ADN polimerasa en presencia de dNTPs y recirculado), y a continuación la inserción del fragmento HindIII de la FAE de pTP11.1, creando pTP3.1, que codifica la señal de aleuraina completa y el extremo COOH de Aspergillus nativo.

Después de los ensayos de varias construcciones, se construyeron los vectores de co-integración con genes de FAE y HygR dispuestos en tándem.

El gen HygR de pAJEB-64-TCA, dirigido por el promotor CaMV, se clonó en primer lugar como un fragmento HindIII de extremo completo en el sitio XbaI de extremo completo de pTP3.1 para obtener pHOX3 y posteriormente escindir como un fragmento HindIII/SacI (digestión parcial con SacI, sitios pertinentes hallados en las secuencias pTP3.1 flanqueantes), el cual se clonó en los sitios HindIII/SacI de pMA406, en orientación tandem (vector pUH1a20). A continuación, las secuencias de FAE se clonaron en el sitio HindIII de pUH1a20 en dirección 3' con respecto del promotor de choque térmico (fragmento HindIII de pTU5 por pUH10, fragmento HindIII de pTT5 por pUH12). Se fabricó un derivado pTP3.1 mediante la clonación del cassette CaMV/HygR HindIII de pAJEB-64-TCA en orientación tándem en dirección 3' con respecto al gen de FAE en pTP3.1, la inactivación del sitio HindIII central mediante la digestión parcial y el relleno del extremo, y la escisión del cassette FAE/HygR combinado como un fragmento HindIII único, el cual se insertó en el sitio HindIII en pMA406 para producir pUC5.11.

Ejemplo 3

Transformación de células vegetales

Se bombardearon cultivos de suspensiones de *F. arundinacea* y *L. multiflorum* de ocho a 10 semanas de vida embriogénica con un único vector de ADN plasmídico de co-integración que contenía los genes de FAE y resistencia a hyg o con un vector de co-transformación que contenía la FAE y con el plásmido pROB5 que confería resistencia a higromicina (CAMV35S-hpt-nos) utilizando una pistola de bombardeo de partículas (PIG) (Finer *et al.* (1992) Development of the particle inflow gun for DNA delivery to plant cells Plant Cell Reports 11:323-328) y partículas de oro de 1,5-3,0 μm como en Dalton *et al.* (Dalton *et al.* (1999) Co-transformed diploid Lolium perenne (Perennial ryegrass), *Lolium multiflorum* (Italian ryegrass) and *Lolium temulentum* (Dandelion) plants produced by microprojectile bombardment. Plant Cell Reports. 18: 721-726) y 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). Los transformantes se seleccionan con higromicina (25 a 50 mg/l) durante un periodo de selección de 10-12 semanas a 25°C bajo una luz blanca fluorescente continua ($60 \mu\text{E m}^{-2} \text{s}^{-1}$) y las plantas se regeneran a través de la embriogénesis somática como en Dalton *et al.* 1999, *supra*. Las plantas regeneradas se cribaron según la actividad FAE en la transferencia a la tierra y la expresión de plantas desarrolladas hasta la madurez en una habitación de crecimiento de contención a 18°C bajo luces fluorescentes durante 16 horas ($350 \mu\text{E m}^{-2} \text{s}^{-1}$). Se volvió a analizar en las plantas maduras (6-8 semanas de vida) la actividad FAE y se recogió nuevo tejido para un análisis Southern, Northern y Western y para el análisis por autodigestión. El tejido restante se liofilizó y se pulverizó para el análisis de la estructura de la pared celular, determinaciones de la digestibilidad de la materia seca *in vitro* (IVDMD) y para las determinaciones con la producción de gas *in vitro* de las velocidades de digestión del tejido.

Ejemplo 4

Reconocimiento del producto de expresión

Para verificar que las secuencias de reconocimiento son eficaces en la liberación del gen, las secuencias de reconocimiento se unieron operativamente a una proteína verde fluorescente GFP. Las construcciones de vectores se muestran en la figura 16. Las células se transformaron mediante el bombardeo de partículas como en el ejemplo 3. La localización de la GFP se podía visualizar bajo un microscopio 1 día después del bombardeo (es decir, el disparo). Véase la figura 16.

Ejemplo 5

Actividad FAE1

Las plantas regeneradas a partir de células transformadas mostraron actividad FAE en todos los tejidos vegetales evaluados. Las células se transformaron como se ha indicado anteriormente bajo la dirección de las secuencias de reconocimiento en ER y APO. La actividad FAE en hojas de *Festuca arundinacea* transformadas de diferentes edades se evaluó en comparación con las plantas de control (no transformadas). Ver Figuras 17 y 18.

ES 2 325 908 T3

Se observaron resultados similares con hojas de *Lolium multiflorum* de diferentes edades transformadas como se ha indicado anteriormente bajo la dirección de la secuencia de reconocimiento vacuolar, ER y APO. Ver Figuras 19 y 20.

También puede inducirse la expresión de FAE bajo un promotor de choque térmico. (Datos no mostrados).

De este modo, hemos demostrado la expresión de FAE en hojas de *Festuca* y *Lolium* bajo promotores constitutivos y HS con el reconocimiento eficaz de FAE en vac, ER y apo.

Ensayo de FAE

Se determinó la actividad FAE en extractos solubles de hojas frescas (o congeladas a -70°C) o cultivos celulares frescos (0,5 g) con NaAc 0,1 M, tampón de extracción de pH 5,0. Se incubaron los extractos con EF 24 mM (4-hidroxi-3-metoxycinamato de etilo) o FAXX al 1% como sustrato, a 28°C durante 24 horas y se calculó la actividad FAE como la cantidad de ácido ferúlico liberado. También se determinó la actividad FAE mediante la medida de la liberación de ácido ferúlico monomérico y dimérico de muestras de hojas cultivo celular auto-digeridas. Se molieron hojas o cultivos celulares frescos, o congelados (0,5 g) en NaAc 0,1 M, tampón de extracción de pH 5,0 en presencia o ausencia de xilanasa (1000 U GC140/muestra) sin sustrato añadido y se incubaron a 28°C durante 72 horas. Después de la incubación, y la centrifugación, se cargaron los extractos solubles en una columna C18 μ Nova sep-pak (Waters) de fase inversa activada, se eluyó con MeOH al 100% y se analizó la muestra de MeOH mediante HPLC.

Ejemplo 6

Análisis Químico de Extractos de Paredes Celulares

Se extrajeron los componentes de enlace éster de hojas en polvo liofilizadas o cultivos celulares (50-100 mg) con NaOH (5 ml de 1 M) seguido de la incubación a 25°C durante 23 horas bajo N₂. Tras la centrifugación y la acidificación del extracto soluble con HCl concentrado, se cargaron los compuestos fenólicos extraídos en una columna C18 μ Nova seppak (Waters) de fase inversa activada y se eluyó con MeOH al 100% y la muestra de MeOH se analizó mediante HPLC.

El HPLC se llevó a cabo con metanol: ácido acético al 5% ya sea con un gradiente de MeOH al 35-65% en 15 minutos (ensayo de FAE) o con un gradiente de MeOH al 30-70% en 25 minutos (componentes monoméricos y diméricos de la pared celular) a 2 ml/min en una μ Nova Pak C18 8x10 RCM (Waters). Se detectaron y se cuantificaron los extractos con un detector de red de diodos (240-400 nm Waters 996PDA) monitorizado a 280 nm para aldehídos y 340 nm para ácidos hidroxycinámicos.

Se reducen los niveles de ácidos hidroxycinámicos monoméricos y diméricos en plantas *Festuca arundinacea* que expresan FAE bajo secuencias de reconocimiento en VAC, ER y APO en comparación con plantas de control (no transformadas). Pueden verse los resultados en las Figuras 21 y 22, respectivamente. De este modo, demostramos que esto no da lugar a una reducción de los compuestos fenólicos de la pared celular en plantas en desarrollo con reconocimiento en vac pero sí da lugar a una reducción de compuestos fenólicos con reconocimiento en ER y apo.

No se reducen significativamente los niveles de ácidos hidroxycinámicos monoméricos y diméricos esterificados en plantas *Festuca arundinacea* que expresan FAE cuando se reconoce FAE en VAC (Fig 21) que está tal y como se predijo para el reconocimiento vacuolar correcto, pero se reducen significativamente, tal y como se predijo, en algunas plantas cuando FAE estaba reconocido en ER y APO, en comparación con las plantas de control (no transformadas). Pueden verse los resultados en la Figura 22.

Ejemplo 7

Digestibilidad de materia seca in vitro. (IVDMD)

La digestibilidad de la materia seca *in vitro* (IVDMD) se estimó en 1,0 g de peso en seco de tejido de hoja o tejido de cultivo celular utilizando el procedimiento de pepsina/celulasa de Jones y 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).

Demostramos que la presencia de FAE en las plantas da lugar a una mayor digestibilidad de las hojas. Esto puede deberse a actividad de FAE interna que actúa en las paredes celulares normales con FAE localizada en las vacuolas y tanto a la actividad de FAE como a la menor reticulación de la pared celular con FAE de ER y reconocido en APO (como también se encuentra con cultivos celulares).

El punto final de la digestibilidad tal y como se determina mediante IVDMD fue mayor en tejido de hoja de algunas plantas transformadas de *Festuca* que expresan FAE, en comparación con las plantas de control (no transformadas). Los ejemplos se muestran allí donde el FAE dirigido a vacuola, ER o opoplasto bajo un promotor de actina ha sido efectiva en el incremento de IVDMD. Se obtuvieron resultados similares con hojas de *Lolium*, pero menos pronunciados.

En las Figuras 23 y 24 pueden verse los resultados.

Ejemplo 8

Mediciones de producción de gas *in vitro*

En cada experimento, se fermentaron 1,0 g de muestras de tejido o cultivo celular de hoja en polvo liofilizado en tres botellas de suero de 165 ml de capacidad de acuerdo con la técnica del transductor de presión de 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). Se tomaron muestras al azar de la digestión en el rumen a las 8.00 h antes de la alimentación matinal de carneros fistulados alimentados con heno de hierba, y transportados al laboratorio en un frasco de vacío precalentado (39°C). Se prepararon el inóculo microbiano y el medio de cultivo tal y como se describe en Theodorou *et al.* (1994). Cada botella de suero recibió 10 ml de inóculo microbiano, 85 ml de tampón y 4 ml de agente reductor.

Al final del periodo de incubación, (144 h) se filtró el contenido de cada botella de suero a través de embudos de cristal sinterizado pre-pesados y se liofilizó hasta peso constante. Se calculó la pérdida de materia seca como la diferencia entre el peso seco de la muestra antes y después de la incubación. Adicionalmente, se determinó la concentración de ácidos grasos volátiles (VFA) en la fracción líquida de los medios de cultivo al final del periodo de incubación de 144 h mediante cromatografía de gas. Para la cuantificación de VFA se utilizó un cromatógrafo Chrompack CP 9000 acoplado con un tomador de muestras automático (Chrompack 911) y un detector de ionización de llama, conectado a un PC Dell con software de integración A1-450.

Los datos de producción de gas se ajustaron al modelo de 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.) utilizando el paquete MLP (Ross, G.J.S. 1987. MLP, Maximum Likelihood Program Version 3.08. Oxford Numerical Algorithms Group). La ecuación está en la forma, $Y = A\{1 - e^{[-b(t-T) - c(\sqrt{t} - \sqrt{T})]}\}$ donde Y es la producción de gas acumulado (ml), A es la asíntota (es decir el conjunto de gas), T es el intervalo de tiempo, y b (h^{-1}) y c ($h^{-0.5}$) son constantes de la velocidad de descomposición. Se calculó una velocidad fraccional combinada (h^{-1}) de producción de gas (μ) como, $\mu = b + c/2\sqrt{t}$, donde t es el tiempo de incubación (h).

Puede observarse que en la *Festuca arundinacea* (indicada como BN en la Figura 25) los cultivos celulares tienen una mayor velocidad de digestión y producción de gas acumulado en presencia de FAE y que la adición de una xilanasa exógena mejora adicionalmente la disponibilidad de los carbohidratos fermentables. Se encuentran resultados similares en cultivos de expresión de FAE sin FAE añadido. Se incrementan adicionalmente las velocidades de fermentación en comparación con los controles mediante la adición de FAE o xilanasa exógenos ya que estos cultivos que expresan FAE presentaban una composición fenólica reducida de la pared celular frente a los controles. Figuras 26-28.

Ejemplo 9

Plantas transformadas con FAE y xilanasa

La adición de xilanasa exógena (GC140) incrementó enormemente la liberación mediada por FAE de compuestos fenólicos de hojas de *Festuca* y *Lolium* que expresan FAE de *A. niger*. Ver Figuras 29-31 que demuestran que se incrementa la liberación de compuestos fenólicos desde las paredes celulares de las hojas en todas las plantas que expresan FAE en la muerte celular y es estimulada por xilanasa independientemente del reconocimiento. Por tanto, se analiza la expresión de una xilanasa fúngica en células vegetales.

Se modifica el cassette de expresión de FAE para comprender un gen de la xilanasa fúngica (ya sea *T. reesei* o *A. niger*) para producir un cassette de expresión de FAE-xilanasa. Se utiliza el cassette de expresión de FAE-xilanasa para transformar células vegetales de un modo similar a los descritos en el Ejemplo 3. Las células transformadas se dejan crecer y se seleccionan en un medio apropiado. Las enzimas así expresadas incrementan la disponibilidad de carbohidratos fermentables hasta una extensión mayor que el cassette de expresión de FAE.

Referencias citadas en la descripción

Esta lista de referencias citadas por el solicitante está prevista únicamente para ayudar al lector y no forma parte del documento de patente europea. Aunque se ha puesto el máximo cuidado en su realización, no se pueden excluir errores u omisiones y la OEP declina cualquier responsabilidad en este respecto.

Documentos de patente citados en la descripción

- WO 9814594 A [0023]

- US 5962670 A [0029]
- EP 9510370 A [0039]
- 5 • US 08952445 B [0044]
- US 09658772 B [0044]
- US 5591616 A [0065]
- 10 • WO 0070061 A [0101]

Documentos que no son patentes citados en la descripción

- 15 • **HAYWARD** *et al.* TAG, 1985, vol. 70, 48 [0003]
- **SINGLETON** *et al.* DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY. *John Wiley and Sons*, 1994 [0012]
- 20 • **HALE; MARHAM.** THE HARPER COLLINS DICTIONARY OF BIOLOGY. *Harper Perennial*, 1991 [0012]
- **BATZER** *et al.* *Nucleic Acid Res.*, 1991, vol. 19, 5081 [0027]
- **OHTSUKA** *et al.* *J. Biol. Chem.*, 1985, vol. 260, 2605-2608 [0027]
- 25 • **ROSSOLINI** *et al.* *Mol. Cell. Probes*, 1994, vol. 8, 91-98 [0027]
- **SAMBROOK** *et al.* MOLECULAR CLONING: A LABORATORY MANUAL. *Cold Spring Harbor Laboratory*, 1989, vol. 1-3 [0031]
- 30 • CURRENT PROTOCOLS IN MOLECULAR BIOLOGY. *Greene Publishing and Wiley-Interscience*, 1987 [0031]
- overview of principles of hybridization and the strategy of nucleic acid probe assays. **TIJSSEN.** LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY-HYBRIDIZATION UIT NUCLEIC ACID PROBES. *Elsevier*, 1993 [0032]
- 35 • **WALDRON** *et al.* *Phytochemical Analysis*, 1996, vol. 7, 305 [0036]
- 40 • **JACQUET** *et al.* *Polyphenol Comm. Bordeaux*, 1996, 451 [0036]
- **MEYER** *et al.* *FEBS Lett.*, 1991, vol. 290, 209 [0037]
- **SAMBROOK** *et al.* MOLECULAR CLONING, A LABORATORY MANUAL. 1989 [0043]
- 45 • **KRIEGLER.** GENE TRANSFER AND EXPRESSION: A LABORATORY MANUAL. 1990 [0043]
- CURRENT PROTOCOLS IN MOLECULAR BIOLOGY. 1994 [0043]
- 50 • PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS. *Academic Press*, 1990 [0048]
- **CARRUTHERS** *et al.* *Cold Spring Harbor Symp. Quant. Biol.*, 1982, vol. 47, 411-418 [0049]
- **ADAMS** *et al.* *J. Am. Chem. Soc.*, 1983, vol. 105, 661 [0049]
- 55 • **WEISING** *et al.* *Ann. Rev. Genet.*, 1988, vol. 22, 421-477 [0051]
- N G. **MESSING** *et al.* GENETIC ENGINEERING IN PLANTS. 1983, 221-227 [0053]
- 60 • **JORDANO** *et al.* *Plant Cell*, 1989, vol. 1, 855-866 [0054]
- **BUSTOS** *et al.* *Plant Cell*, 1989, vol. 1, 839-854 [0054]
- **GREEN** *et al.* *EMBO J.*, 1988, vol. 7, 4035-4044 [0054]
- 65 • **MEIER** *et al.* *Plant Cell*, 1991, vol. 3, 309-316 [0054]
- **ZHANG** *et al.* *Plant Physiology*, 1996, vol. 110, 1069-1079 [0054]

- **MCELROY**. *Plant Cell*, 1990, vol. 2, 163 [0055]
- **PEREZ-AMADOR** *et al.* *Plant Physiol.*, 2000, vol. 122, 169 [0057]
- 5 • **NOH; AMASINO**. *Plant Mol. Biol.*, 1999, vol. 41, 181 [0057]
- *Proc. Nat'l Acad. Sci. USA*, 1987, vol. 84, 7104 [0058]
- 10 • **ROGERS**. *Proc. Nat'l Acad. Sci. USA*, 1985, vol. 82, 6512 [0060]
- **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, vol. 18 (9), 721-726 [0063]
- 15 • **PASZKOWSKI** *et al.* *Embo J.*, 1984, vol. 3, 2717-2722 [0064]
- **FROMM** *et al.* *Proc. Natl. Acad. Sci. USA*, 1985, vol. 82, 5824 [0064]
- **BORSCH** *et al.* *Science*, 1984, vol. 233, 496-498 [0065]
- 20 • **FRALEY** *et al.* *Proc. Nat'l. Acad. Sci. USA*, 1983, vol. 80, 4803 [0065]
- **KLEIN** *et al.* *Nature*, 1987, vol. 327, 70-73 [0066]
- 25 • **EVANS** *et al.* PROTOPLASTS ISOLATION AND CULTURE, HANDBOOK OF PLANT CELL CULTURE. *Macmillian Publishing Company*, 1983, 124-176 [0067]
- **BINDING**. REGENERATION OF PLANTS, PLANT PROTOPLASTS. *CRC Press*, 1985, 21-73 [0067]
- 30 • **KLEE** *et al.* *Ann. Rev. of Plant Phys.*, 1987, vol. 38, 467-486 [0067]
- **MCELROY** *et al.* *MGG*, 1991, vol. 231, 150-160 [0084]
- **AINLEY; KEY**. *PMB*, 1990, vol. 14, 949-60 [0085] [0099]
- 35 • **CHEN**, H.M.; **C. FORD**; **P. J. REILLY**. *Biochem J*, 1994, vol. 301, 275-281 [0094]
- **JEFFERSON** *et al.* *EMBO J*, 1987, vol. 6, 3901-3907
- 40 • **BOEVINK P; OPARKA K; CRUZ SS; MARTIN B; BETTERIDGE A; HAWES C.** *PLANT JOURNAL*, 1998, vol. 15, 441-447 [0109]
- **DENECKE** *et al.* *EMBO J*, 1992, vol. 11, 2345-2355 [0115]
- 45 • **BILANG** *et al.* *Gene*, 1991, vol. 100, 247-50 [0119]
- **ZHENG** *et al.* *Plant Physiol*, 1991, vol. 97, 832-835 [0120]
- **FINER** *et al.* Development of the particle inflow gun for DNA delivery to plant cells. *Plant Cell Reports*, 1992, vol. 11, 323-328 [0126]
- 50 • **DALTON** *et al.* Co-transformed diploid *Lolium perenne* (Perennial ryegrass), *Lolium multiflorum* (Italian ryegrass) and *Lolium temulentum* (Damel) plants produced by microprojectile bombardment. *Plant Cell Reports*, 1999, vol. 18, 721-726 [0126]
- 55 • **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, vol. 58, 149-154 [0126]
- **JONES; 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, vol. 26, 711-718 [0137]
- 60 • **THEODOROU** *et al.* A new gas production method using a pressure transducer to determine the fermentation kinetics of ruminant feeds. *Animal Feed Science and Technology*, 1994, vol. 48, 185-197 [0141]
- 65 • **FRANCE, J.; DHANOA, M.S.; THEODOROU, M.K; LISTER, S.J.; DAVIES. D.R.; ISAC, D.** A model to interpret gas accumulation profiles associated with *in vitro* degradation of ruminant feeds. *Journal of Theoretical Biology*, 1993, vol. 163, 99-111 [0143]

ES 2 325 908 T3

- **ROSS**, G.J.S. MLP, *Maximum Likelihood Program Version 3.08*. 1987 [0143]

5

10

15

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

1. Procedimiento para aumentar la digestibilidad de una planta, comprendiendo el método la introducción en la
5 planta de una cassette de expresión que comprende un promotor unido operativamente a un polinucleótido que codifica la esterasa de ácido ferúlico, en el que el cassette de expresión comprende además una secuencia de polinucleótidos que codifica una secuencia señal que reconoce la expresión del polipéptido de esterasa de ácido ferúlico en la vacuola.
2. Procedimiento según la reivindicación 1, en el que el polinucleótido que codifica la esterasa de ácido ferúlico
10 deriva de *Aspergillus niger*.
3. Procedimiento según la reivindicación 2, en el que el polinucleótido codifica FAE1 de *Aspergillus niger*.
4. Procedimiento según la reivindicación 3, en el que el polinucleótido codifica una esterasa de ácido ferúlico con
15 un sitio de glicosilación alterado.
5. Procedimiento según la reivindicación 3, en el que el polinucleótido codifica una esterasa de ácido ferúlico con una sustitución, de manera que se altera la glicosilación.
6. Procedimiento según la reivindicación 3, en el que el polinucleótido comprende además un polinucleótido que
20 codifica CTWPVAAA en el extremo 3'.
7. Procedimiento según la reivindicación 3, en el que los codones subóptimos se modifican a codones preferidos de *Triticum spp*.
- 25 8. Procedimiento según la reivindicación 1, en el que el promotor es un promotor inducible.
9. Procedimiento según la reivindicación 1, en el que el promotor es un promotor de senescencia.
10. Procedimiento según la reivindicación 1, en el que el promotor es un promotor de choque térmico.
- 30 11. Procedimiento según la reivindicación 1, en el que el promotor es un promotor constitutivo.
12. Procedimiento según la reivindicación 1, en el que la secuencia de polinucleótidos que codifica la secuencia
35 señal se encuentra en la dirección del extremo 5' del polinucleótido que codifica la esterasa de ácido ferúlico.
13. Procedimiento según la reivindicación 12, en el que el polinucleótido que codifica la secuencia señal deriva de un gen de reconocimiento vacuolar.
- 40 14. Procedimiento según la reivindicación 13, en el que el gen de reconocimiento es un gen de aleuraina de cebada.
15. Procedimiento según la reivindicación 13, en el que el polinucleótido que codifica la secuencia señal deriva de un gen de senescencia de reconocimiento vacuolar.
- 45 16. Procedimiento según la reivindicación 15, en el que el gen de senescencia es un gen de *Lolium Seel*.
17. Procedimiento según la reivindicación 1, en el que el polinucleótido que codifica la secuencia señal deriva de *Solanum tuberosum*.
- 50 18. Procedimiento según la reivindicación 1, en el que la secuencia de polinucleótido que codifica la secuencia señal se encuentra en la dirección del extremo 3' del polinucleótido de esterasa de ácido ferúlico.
19. Procedimiento según la reivindicación 1, que comprende además la introducción simultánea en la planta de un
55 segundo cassette de expresión que comprende un promotor unido operativamente a un polinucleótido que codifica un gen de xilanas.
20. Procedimiento según la reivindicación 19, en el que el segundo polinucleótido codifica una xilanasa fúngica.
21. Procedimiento según la reivindicación 20, en el que el polinucleótido que codifica la xilanasa es de *Trichoderma
60 reesei*.
22. Procedimiento según la reivindicación 19, en el que el primer y segundo cassette de expresión están presentes en plásmidos separados.
23. Procedimiento según la reivindicación 1, en el que la planta se selecciona del grupo que consiste en *Festuca*,
65 *Lolium*, *Zea* y *Avena*.
24. Procedimiento según la reivindicación 23, en el que la planta es *Festuca arundinacea*.

ES 2 325 908 T3

25. Procedimiento según la reivindicación 1, en el que la introducción del polinucleótido de esterasa de ácido ferúlico en la planta es mediante la transformación de cultivos celulares.

26. Procedimiento según la reivindicación 25, en el que los cultivos celulares se regeneran en plantas.

27. Procedimiento según la reivindicación 11, en el que el promotor es un promotor de actina.

28. Planta transgénica que comprende un cassette de expresión que comprende un promotor unido operativamente a un polinucleótido que codifica la esterasa de ácido ferúlico, en la que el cassette de expresión comprende además una secuencia de polinucleótidos que codifica una secuencia señal que reconoce la expresión del polipéptido de esterasa de ácido ferúlico en la vacuola.

29. Planta según la reivindicación 28, en la que el polinucleótido que codifica la esterasa de ácido ferúlico deriva de *Aspergillus niger*.

30. Planta según la reivindicación 29, en la que el polinucleótido codifica FAE1 de *Aspergillus niger*.

31. Planta según la reivindicación 30, en la que el polinucleótido codifica una esterasa de ácido ferúlico con un sitio de glicosilación alterado.

32. Planta según la reivindicación 30, en la que el polinucleótido codifica una esterasa de ácido ferúlico con una sustitución, de manera que se altera la glicosilación.

33. Planta según la reivindicación 30, en la que el polinucleótido comprende además un polinucleótido que codifica CTWPVAAA en el extremo 3'.

34. Planta según la reivindicación 28, en la que el promotor es un promotor inducible.

35. Planta según la reivindicación 28, en la que el promotor es un promotor de senescencia.

36. Planta según la reivindicación 28, en la que el promotor es un promotor de choque térmico.

37. Planta según la reivindicación 28, en la que el promotor es un promotor constitutivo.

38. Planta según la reivindicación 28, en la que la secuencia de polinucleótidos que codifica la secuencia señal se encuentra en la dirección del extremo 5' del polinucleótido que codifica la esterasa de ácido ferúlico.

39. Planta según la reivindicación 38, en la que el polinucleótido que codifica la secuencia señal deriva de un gen de reconocimiento vacuolar.

40. Planta según la reivindicación 39, en la que el gen de reconocimiento es un gen de aleuraina de cebada.

41. Planta según la reivindicación 39, en la que el polinucleótido que codifica la secuencia señal deriva de un gen de senescencia de reconocimiento vacuolar.

42. Planta según la reivindicación 41, en la que el gen de senescencia es un gen de *Lolium Seel*.

43. Planta según la reivindicación 28, en la que el polinucleótido que codifica la secuencia señal deriva de *Solanum tuberosum*.

44. Planta según la reivindicación 28, en la que la secuencia de polinucleótido que codifica la secuencia señal se encuentra en la dirección del extremo 3' del polinucleótido de esterasa de ácido ferúlico.

45. Planta según la reivindicación 28, que comprende además un segundo cassette de expresión que comprende un promotor unido operativamente a un polinucleótido que codifica un gen de xilanasa.

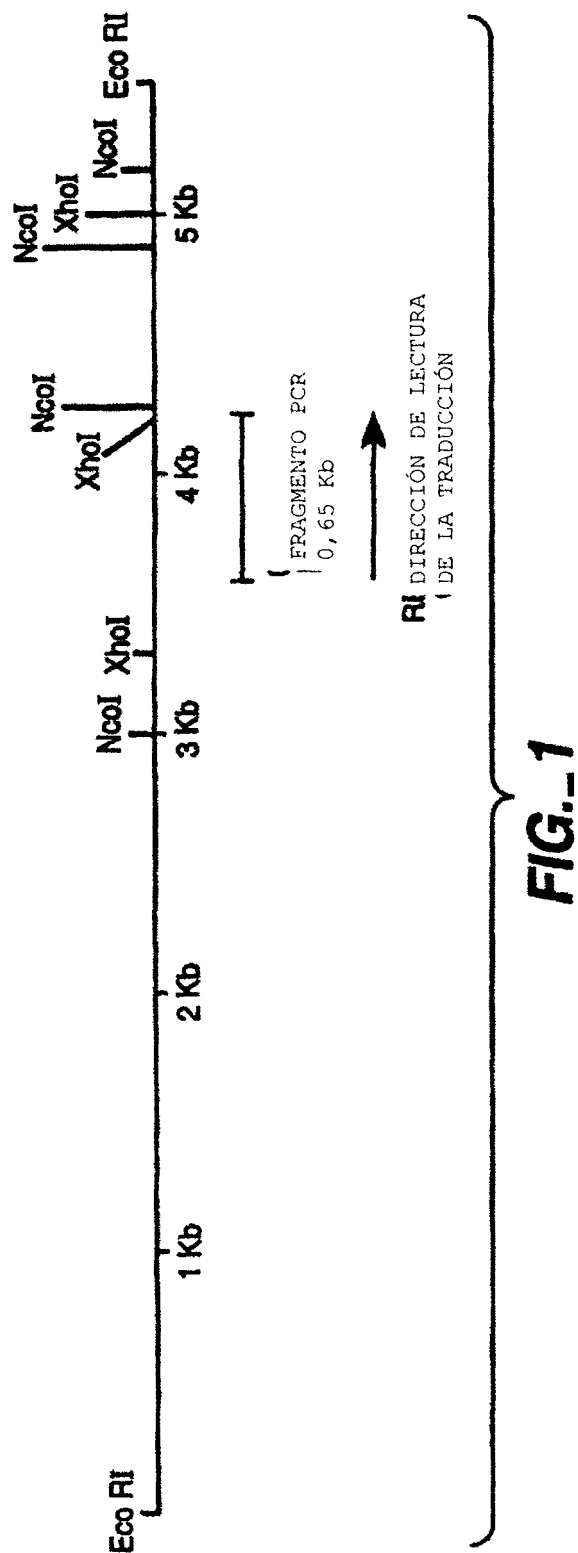
46. Planta según la reivindicación 45, en la que el segundo polinucleótido codifica una xilanasa fúngica.

47. Planta según la reivindicación 46, en la que el polinucleótido que codifica la xilanasa es de *Trichoderma reesei*.

48. Planta según la reivindicación 28, en la que la planta se selecciona del grupo que consiste en *Festuca*, *Lolium*, *Zea* y *Avena*.

49. Planta según la reivindicación 48, en la que la planta es *Festuca arundinacea*.

50. Planta según la reivindicación 37, en la que el promotor es un promotor de actina.



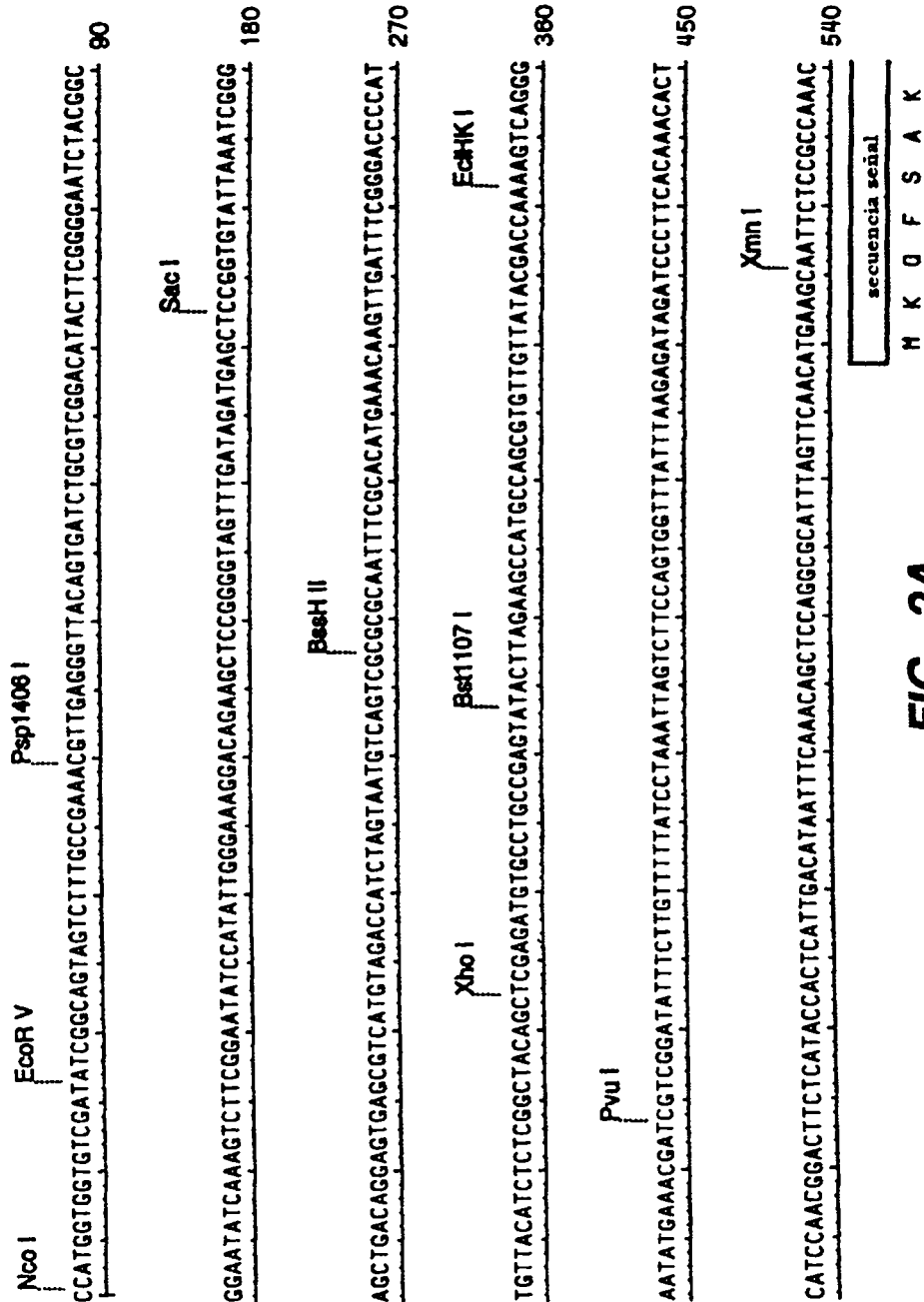


FIG.-2A

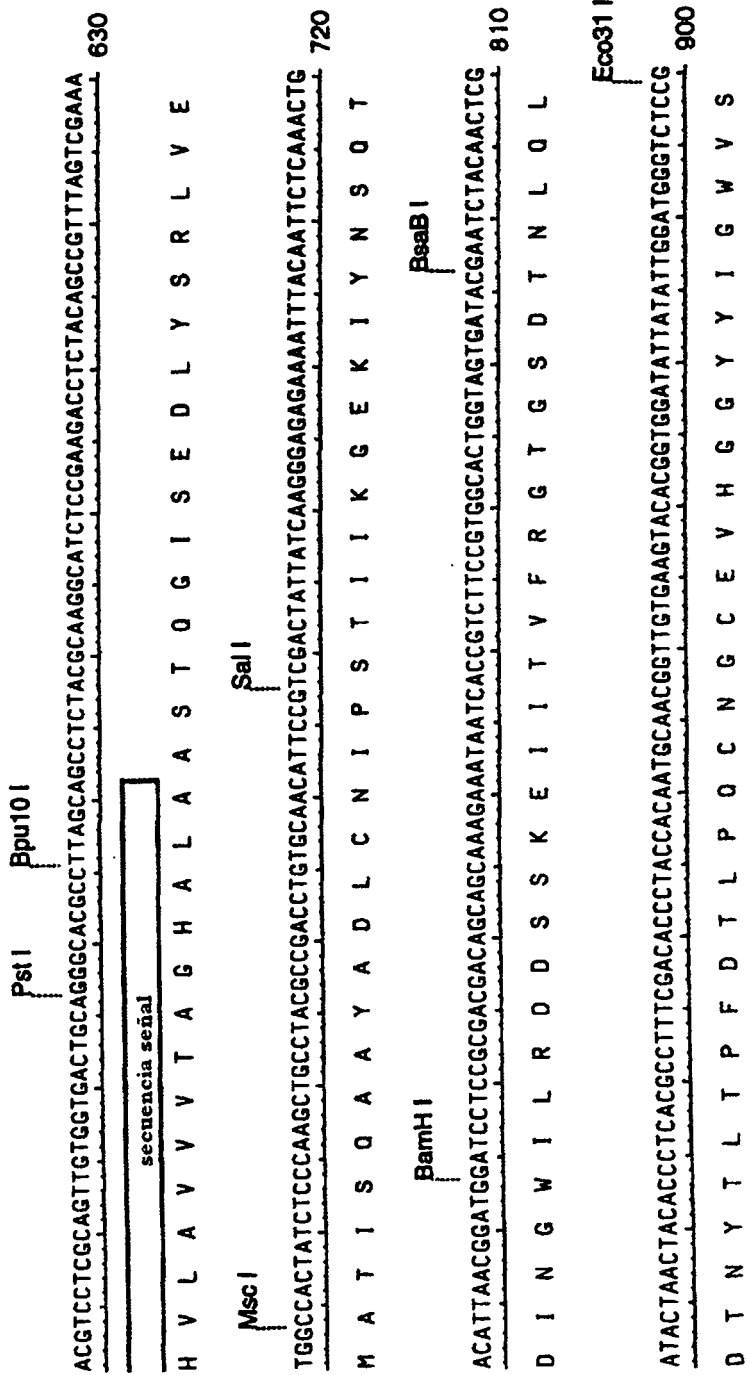


FIG. 2B

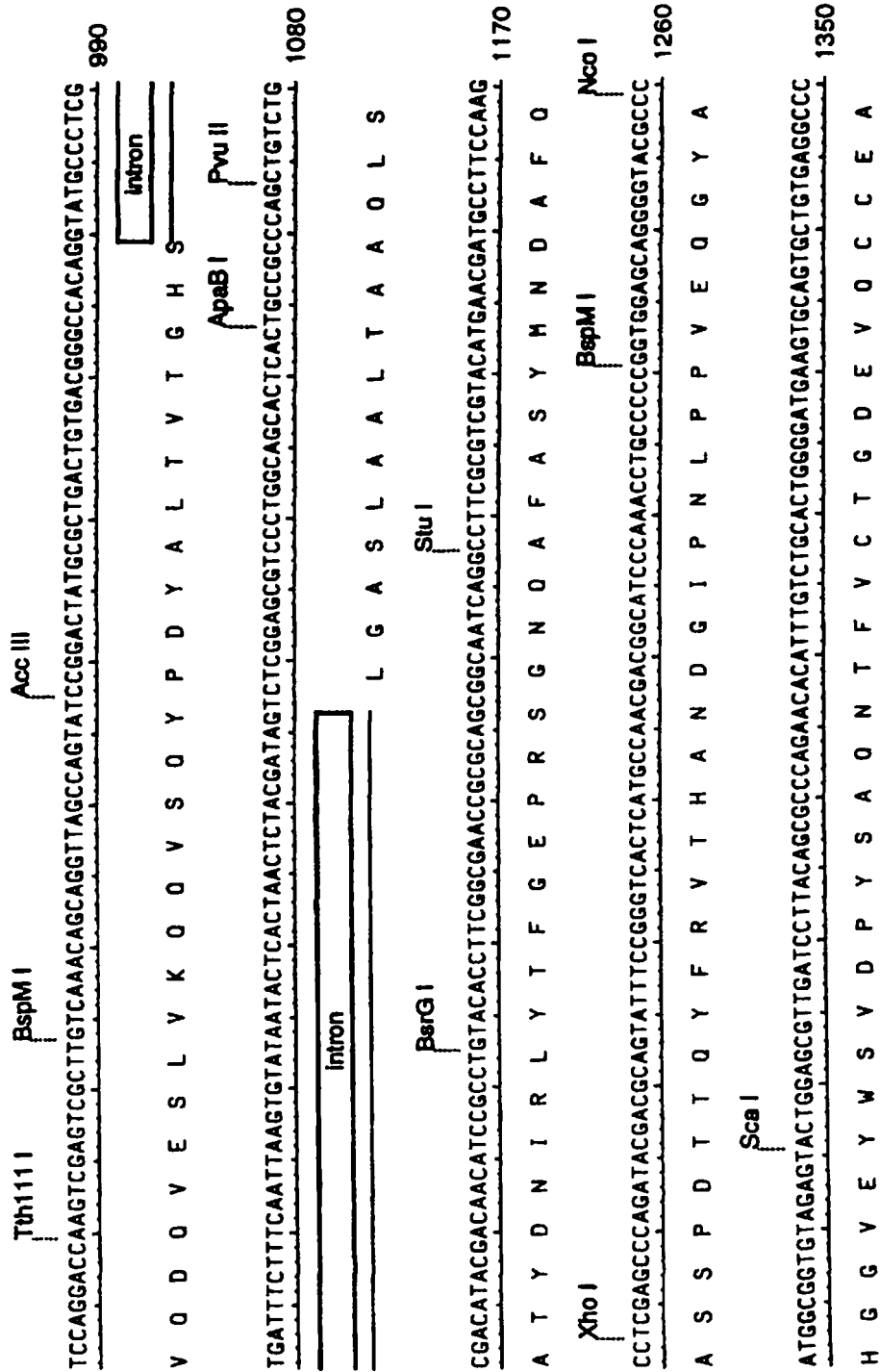


FIG..2C

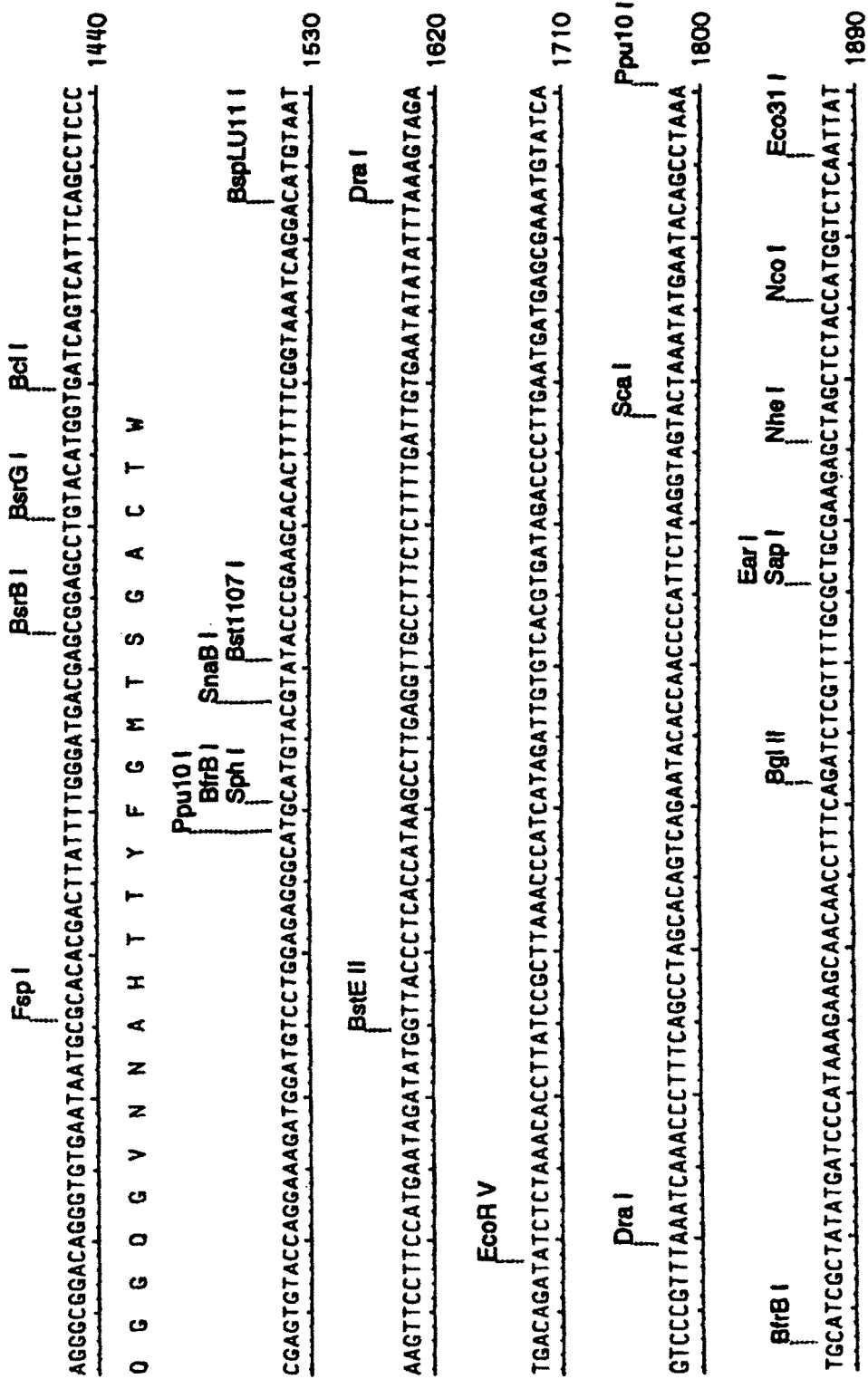


FIG._2D

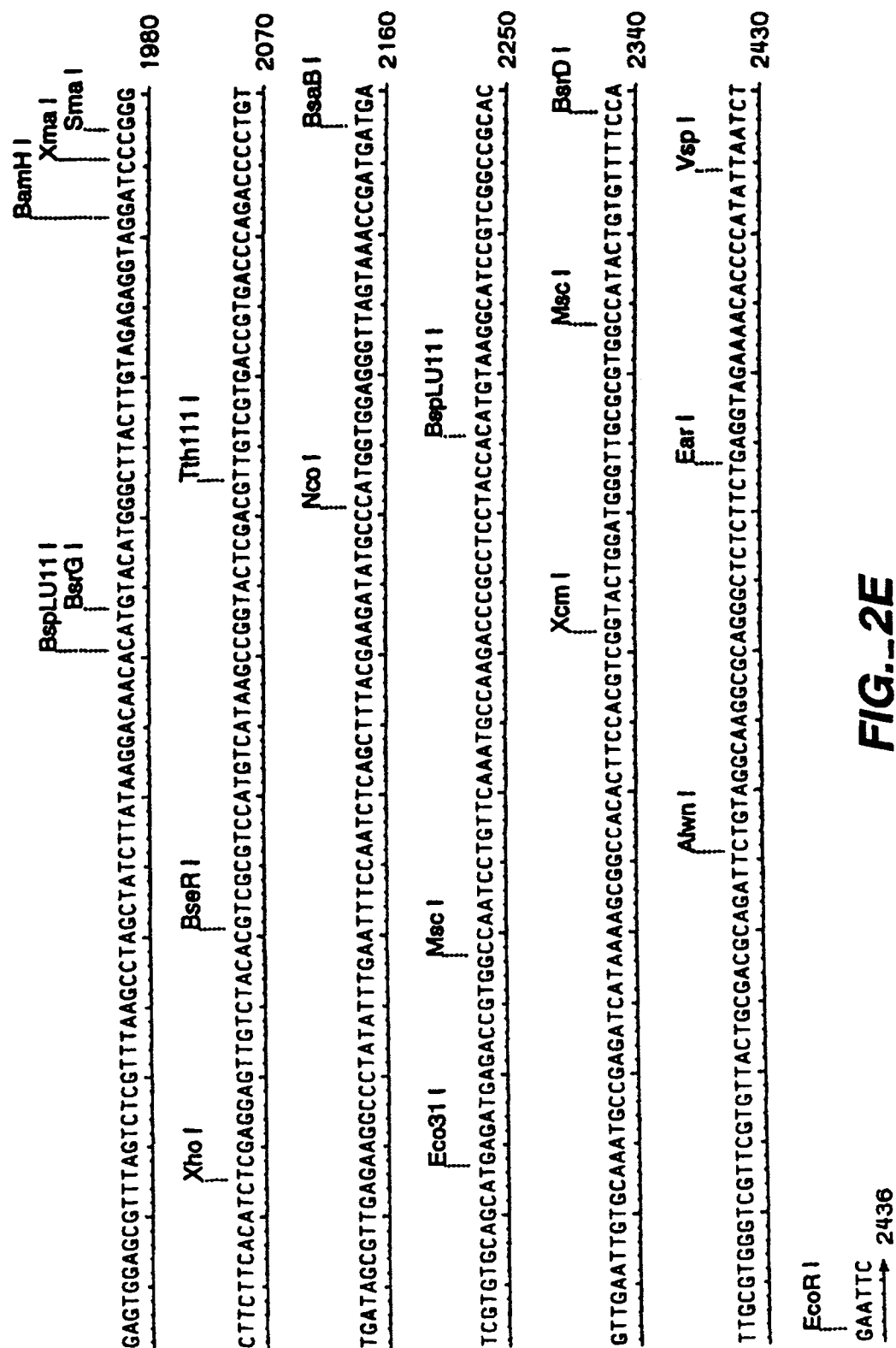


FIG. 2E

CCA TGG TGT CGA TAT CGG CAG TAG TCT TTG CCG AAA C G T T G A G G G T T A C A G T G A T C T G C G T C G G A C A T A C T T
 C G G G A A T C T A C G G C G A A T A T C A A A G T C T T C G G A A T A T C C A T A T T G G G A A A G G A C A G A A G C T C C G G G T A G T T T
 G A T A G A T G A G C T C C G G T G T A T T A A T C G G G A G C T G A C A G A G T A G C G T C A T G T A G A C C A T C T A G T A A T G T C A G T
 C G C G C A A T T T C G C A C A T G A A A C A A G T T G A T T T C G G A C C C C A T T G T T A C A T C T C T C G G C T A C A G C T C G A G A T G
 T G C C T G C C G A G T A C T T A G A A G C C A T G C C A G C G T G T T G T T A T A C G A C C A A A A G T C A G G G A A T A T G A A A C G A T C G
 T C G G A T A T T T C T T G T T T A T C C T A A A T T A G T C T C C A G T G T T A T T A A G A G A T A G A T C C C T T C A C A A A C A C T
 C A T C C A A C G G A C T T C T C A T A C C A C T C A T T G A C A T A A T T T C A A A C A G C T C C A G G C A C G C C T T A G C A G C C T C T A C G C A A G G C A T C T
 A A T T C T C G C C A A A C A C G T C C T C G C A G T T G T G T G A C T G C A G G C A C G C C T T A G C A G C C T A C G C C A G C C T G T G C A A C A T T C
 C G A A G A C C T C T A C A G C C G T T A G T C G A A A T G G C C A C T A T C T C C A A G C T G C C T A C G C C A G C C T G T G C A A C A T T C
 C G T C G A C T A T T A T C A A G G G A G A A A A T T T A C A A T T C T C A A A C T G A C A T T A A C G G A T G G A T C C T C C G C G A C G A C A
 G C A G C A A A G A A A T A A T C A C G T C T T C C G T G G C A C T G G T A G T A C G A A T C T A C A A C T C G A T A C T A A C T A C A C C C
 T C A C G C C T T T C G A C A C C C T A C C A C A T A G C A A C G G T T G T G A A G T A C A C G G T G G A T A T T A T A T T G G A T G G G T C T C C G
 T C C A G G A C C A A G T C G A G T C G C T T G T C A A A C A G C A G G T T A G C C A G T A T C C G G A C T A T G C G C T G A C T G T G A C G G G C C
 A C A G G T A T G C C C T C G T G A T T T C T T T C A A T T A A G T A T A A T A C T C A C T A A C T C C G C C T G T A C A C T T C G G A T A G T C T C G G A G C G T C C C
 T G G C A G C A C T C A C T G C C G C C A G C T G T C T G C G A C A T A C G A C A C A C T C C G C C T G T A C A C C T T C G G C G A A C C G C G C A
 G C G G C A A T C A G G C C T T C G C G T C G T A C A T G A A C G A T G C C T T C C A A G C C T G G A G C C C A G A T A C G A C C G A G A T A T T C C
 G G G T C A C T C A T G C C A A C G A C G G C A T C C C A A A C C T G C C C C G G T G G A G C A G G G T A C G C C C A T G G C C C A T G G C G T G T A G A G T
 A 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 C A G A A C A C A T T T T G T C T G C A C T G G G G A T G A A G T G C A G T G C T G T G A G G C C C
 A G G G C G G A C A G G G T G T G A A T A A T G C G C A C A C G A C T T A T T T T G G G A T G A C G A G C G G A G C C T G T A C A T G G T G A T C A G
 T C A T T T C A G C C T C C C G A G T G T A C C A G G A A A G A T G G A T G T C C T G G A G A G G G C A T G C A T G T A C G T A T A C C C G A A G C
 A C A C T T T T C G G T A A T C A G G A C A T G T A A T A A G T T C C T T C C A T G A A T A G A T A T G G T T A C C C T C A C C A T A A G C C T T
 G A G G T T G C C T T T C T C T T T T G A T T G T G A A T A T A T A T T A A A G T A G A T G A C A G A T A T C T C T A A A C A C C T T A T C C G C T
 T A A C C C A T C A T A G A T T G T G T C A C G T G A T A G A C C C C T T G A A T G A T G A C G A A A T G T A T C A G T C C C G T T T A A A T C A
 A A C C C T T T C A G C C T A G C A C A G T C A G A A T A C A C C A A C C C C A T T C T A A G G T A G T A C T A A A T A T G A A T A C A G C C T A A A
 T G C A T C G C T A T A T A T G A T C C C A T A A A A G A G C A A C A A C C T T T C A G A T C T C G T T T G C G C T G C G A A G A G C T A G C T C T A C
 C A T G G T C T C A A T A T A G A T G G A G C G T T A G T C T C G T T T A A G C C T A G C T A T C T T A T A A G G A C A A C A C A T G T A C A T G
 G G C T T A C T T G T A G A G A G G T A G G A T C C C G G G C T T C T C A C A T C T C G A G A G T T G T C T A C A C G T C G C G T C C A T G T C A
 T A A G C C G G T A C T C G A C G T T G T C G T A C C C G T G A C C C A G A C C C C T G T T G A T A G C G T T G A A A G G C C C T A T A T T T G A A
 T T T C C A A T C T C A G C T T A C G A A G A T A T G C C C A T G G T G A G G G T T A G T A A A C C G A T G A T G A T C G T G T G C A G C A T G A
 G A T G A G A C C G T G G C C A A T C C T G T T C A A A T G C C A A G A C C C G C C T C C T A C C A C A T G T A A G G C A T C C G T C G G C C G C A C
 G T T G A A T T G T G C A A A T G C C G A G A T C A T A A A A G C G G C C A C A C T T C C A C G T C G G T A C T G G A T G G G T T G C G C G T G G C C
 A T A C T G T G T T T T C C A T T G C G T G G G T C G T C G T G T T A C T G C G A C G C A G A T T C T G T A G G C A A G G C G C A G G G C T C T C T
 T C T G A G G T A G A A A A C A C C C C A T A T T A A T C T G A A T T C

FIG. 3

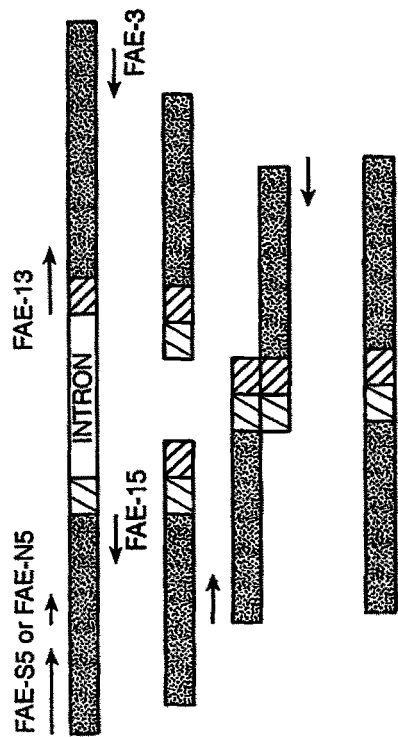


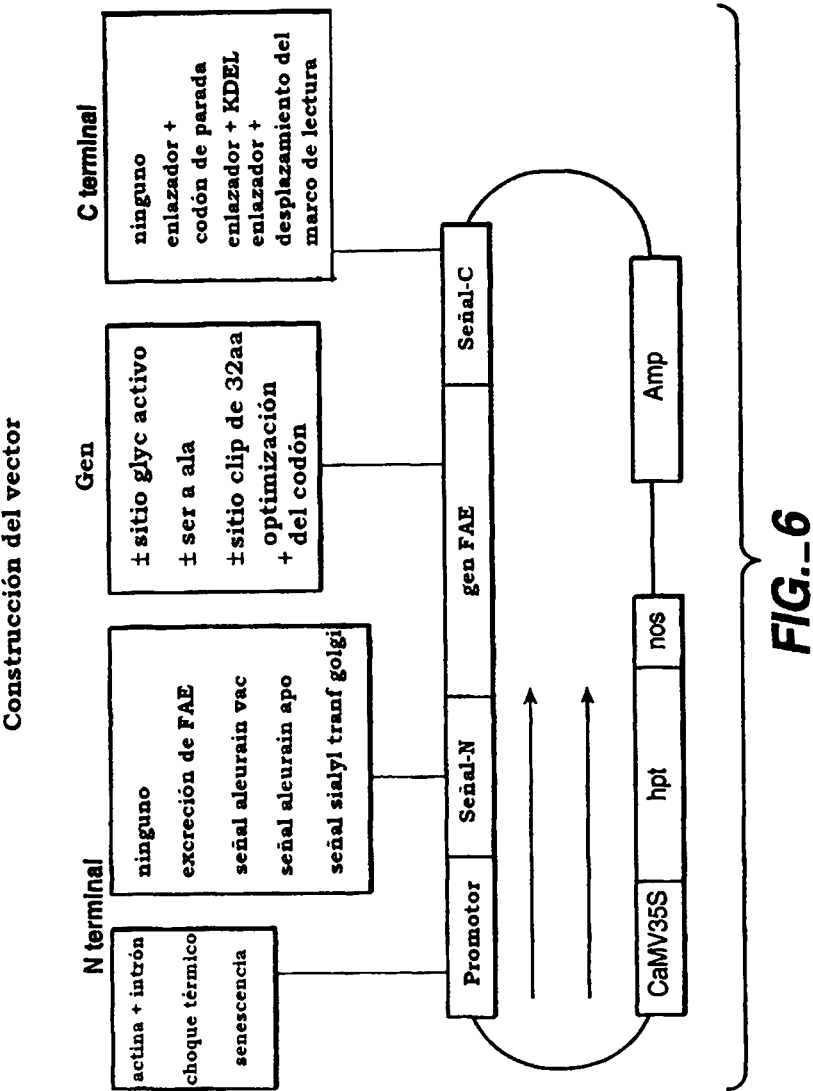
FIG. 4

FAE-13 CCGGCCACGCCCTCGGGCCCTCCCTGGCGGCACTC 35-mer
FAE-15 GGCGCCGAGGAGTGCCCGGTCA CGGTCA GCGCGTAGTCC 40-mer

posición del intrón en el original

Y A L T V T G H S L G A S L A A L
GGACTACGCGCTGACCGTGACCGGCCACCCCTCGGGCC
complemento, FAE-15
Y A L T V T G H A L G A S L A A L
FAE-13

FIG. 5



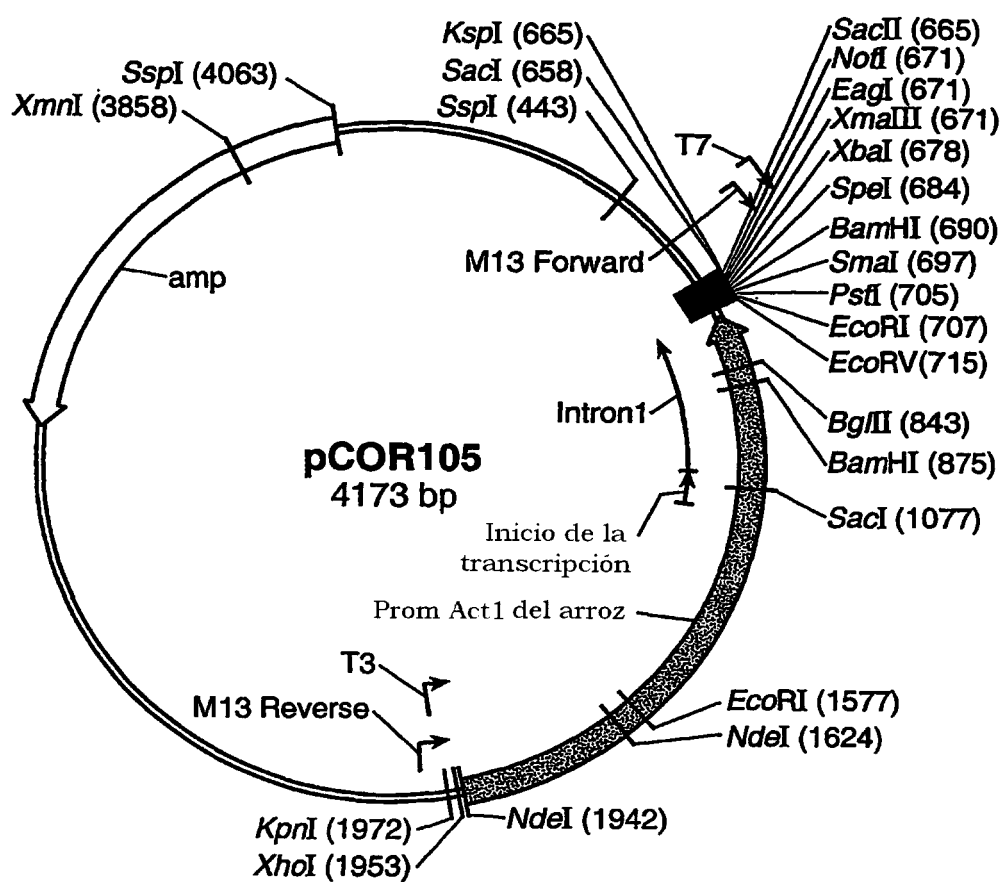


FIG._7

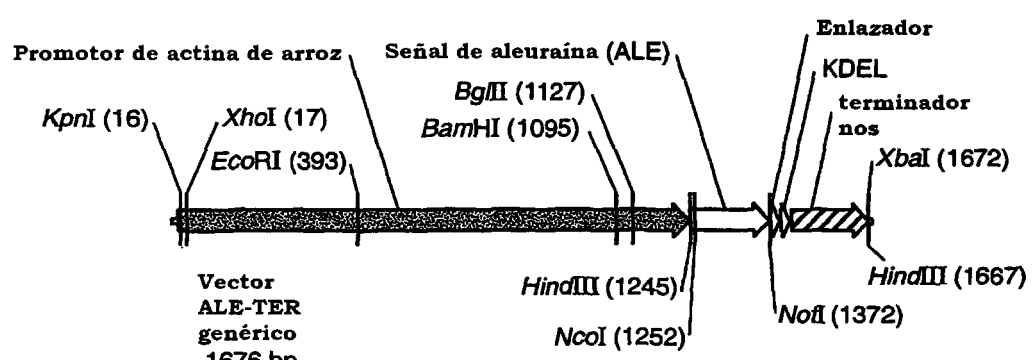


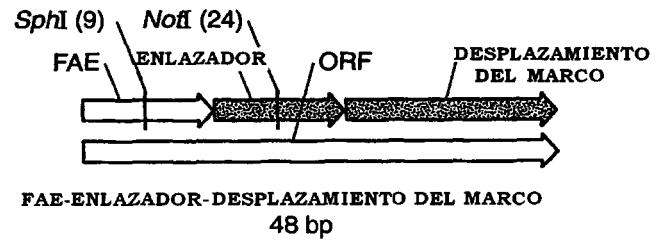
FIG._8

Secuencia de retención del ER KDEL-COOH

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

FIG.\_9

Estructura y secuencia de FAE-ENLAZADOR-DESPLAZAMIENTO DEL MARCO



+1 G A C T W P V A A A E T T E G  
SphI NotI  
~~~~~  
1 GGCGCATGCA CCTGGCCGGT CGCGGCCGCG GAAACCACTG AAGGATGA
CCGCGTACGT GGACCGGCCA GCGCCGGCGC CTTTGGTGAC TTCCTACT

FIG._10

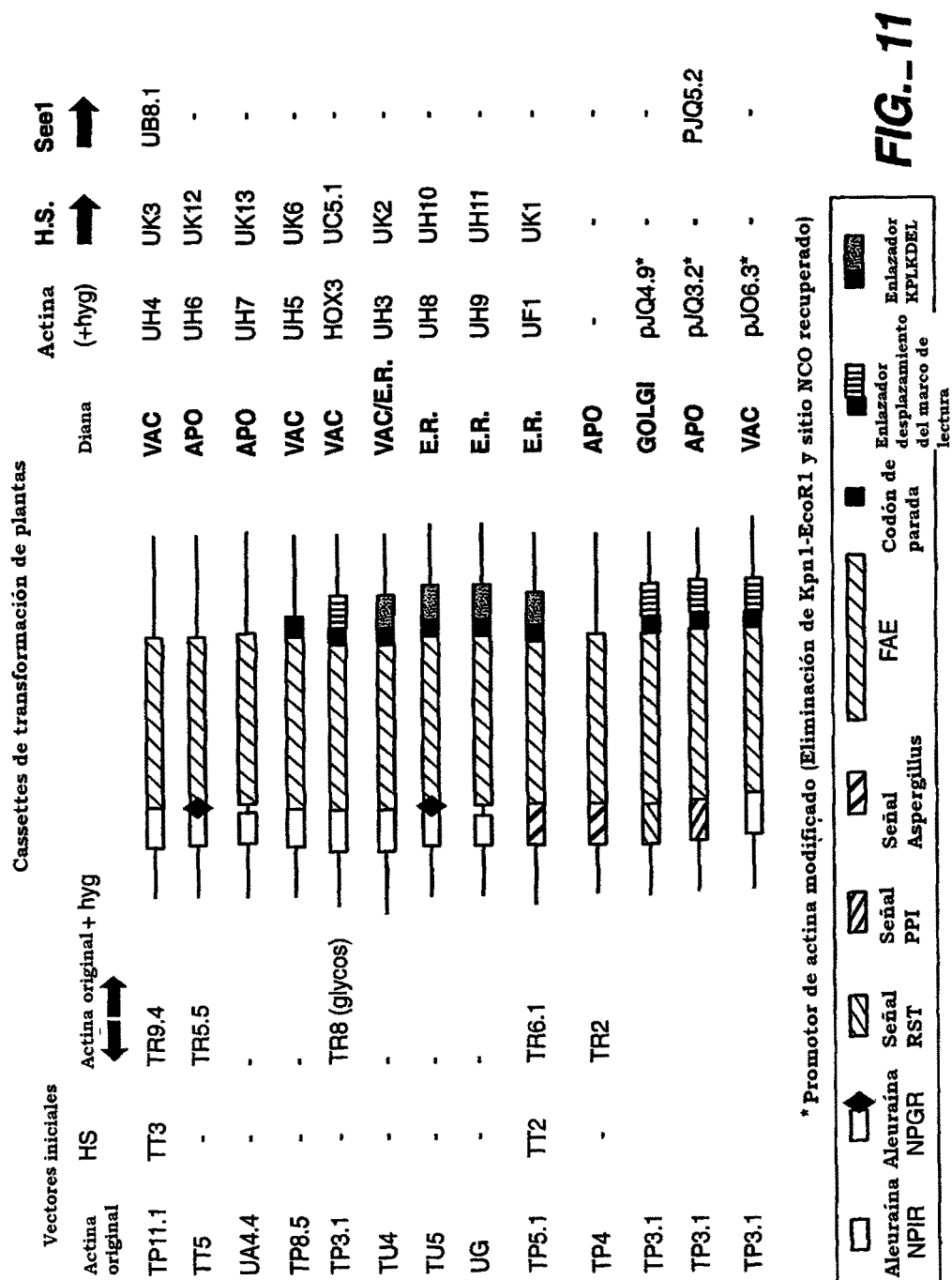


FIG. 12

Vectores**Promotor de actina original en pCOR105**

	Diana	Secuencias señal	Vectores
(i)	APO	-aleuraina-NPGR-FAE	pUH6, pTT5, TT5.5, pTT5.1
(ii)	ER	-aleuraina-delINPIR-FAE	pUH7, pUA4,4
		-aleuraina-NPGR-FAE-enlazador-KDEL	pTU5, pUH8
		-aleuraina-delINPIR-FAE-enlazador-KDEL	pUG4, pUH9
(iii)	VAC	-aleuraina-NPIR-FAE	pTP11.1, pTR9.4, pUH4, pUK3,
(iv)	ER/VAC	-aleuraina-NPIR-FAE-enlazador-KDEL	pTU4, pUH3,
(v)	VAC	-aleuraina-NPIR-FAE-enlazador-desplazamiento de marco	pUA1K3, pTP3.1, pUC5.11
(vi)	VAC	-aleuraina-NPIR-FAE-enlazador-parada	pTP8.5, pUH5
(vii)	ER	-señal Aspergillus-FAE-KDEL	pTP5.1, pTP6.1, pUF1,

Promotor de actina modificado (eliminación Kpn1-EcoR1 y restauración del sitio NCO)

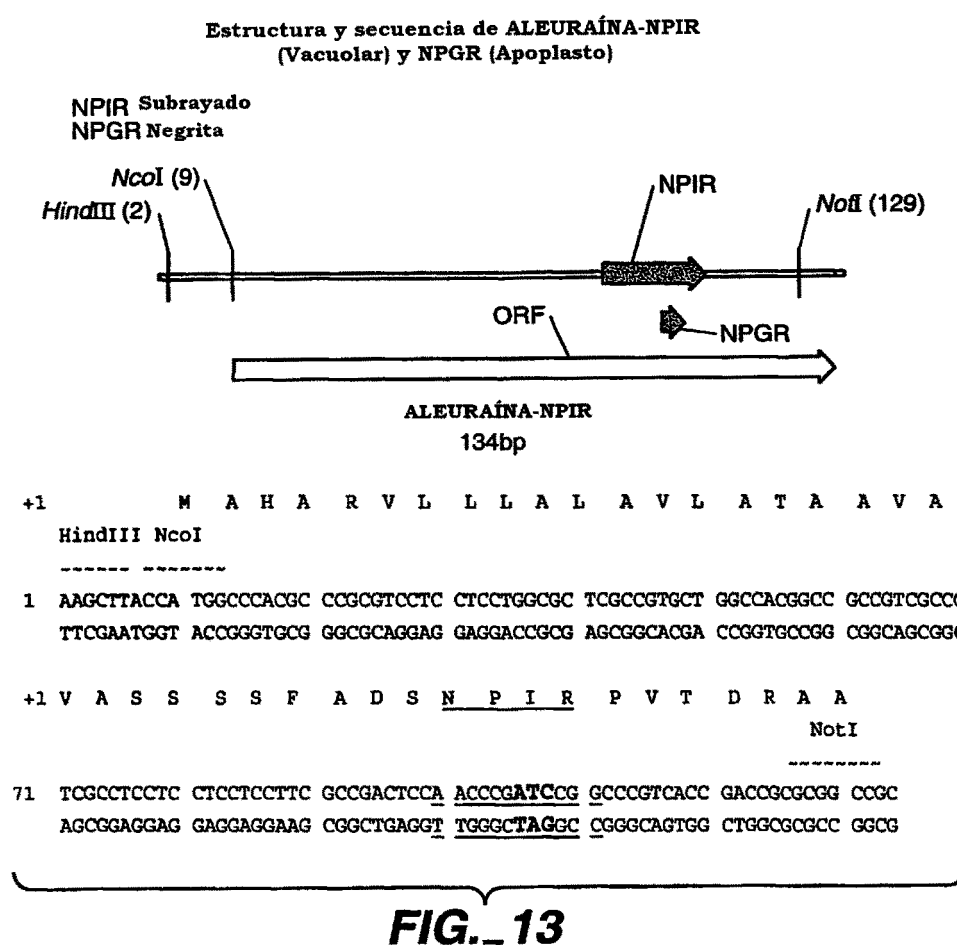
(i)	VAC	-aleuraina-NPIR-FAE-enlazador-desplazamiento de marco	pJ06.3
(ii)	GOLGI	-RST-FAE-enlazador-desplazamiento de marco	pJQ3.2
(iii)	APO	-PPI-FAE-enlazador-desplazamiento de marco	pJQ4.9

Promotor de choque térmico

(i)	APO	-aleuraina-NPGR-FAE	pUH12
		-aleuraina-delINPIR-FAE	pUH13
		-señal Aspergillus-FAE	pTP4a2, pTR2.22
(ii)	ER	-aleuraina-NPGR-FAE-enlazador-KDEL	pUH10
		-aleuraina-delINPIR-FAE-enlazador-KDEL	pUH11
(iii)	VAC	-aleuraina-NPIR-FAE	pUK3, pTT3
(iv)	ER/VAC	-aleuraina-NPIR-FAE-enlazador-KDEL	pUK2
(v)	VAC	-aleuraina-NPIR-FAE-enlazador-desplazamiento de marco	pUC5.11, pHOX3
(vi)	VAC	-aleuraina-NPIR-FAE-enlazador-parada	pUK6
(vii)	ER	-señal Aspergillus-FAE-KDEL	pUK1, pTT2

Promotor de senescencia

(i)	APO	See1-PPI-FAE-enlazador-desplazamiento de marco	pJQ5.2
(ii)	VAC	-See1-aleuraina-NPIR eliminado-FAE	pUB8.1



SIALIL TRASNFERASA DE RATA Secuencia señal de Golgi

```
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 AAGAAGTTCT CCGCTTTGAT CCGGTCTTC CTCCCTTCG
      . 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 CGGTGGAG AAGGCTCG ACTACGAGC CTCACCTC CAAGCCAGG AGTTCCAAAT

NotI
-----
      . A A
141 GCGGCGCG
```

FIG._14

INHIBIDOR DE PROTEASA II DE PATATA Secuencia señal de apoplasto

```
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 CAAGGAGGTG AACTTCGTG CTAACCTCT GATCGTBTCT
      GGCCTCTCT

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 TGCTGTSTC CGCCATGGAG CACGTGAGC CCAAGGCTG CACCCGAG TGGGCAACC
      TCGGCTCGG

NotI
-----
      . I C P A A A
141 CATCTGCCG GCGGCGCC
```

FIG._15

Expresión de reconocimiento de gfp en Compartimentos Celulares Diferentes

Vectores de reconocimiento del Promotor de Actina

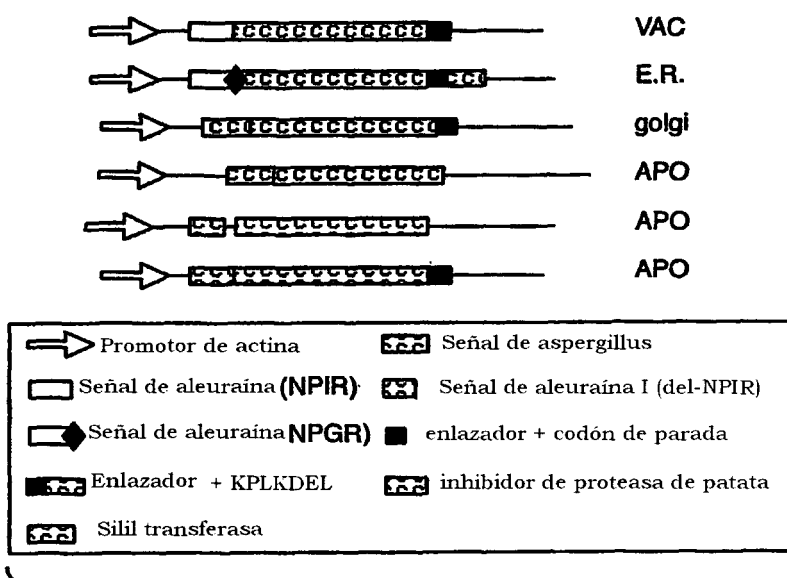


FIG._16A

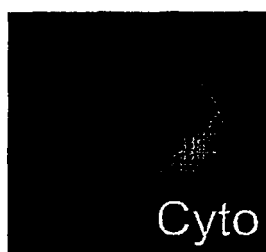


FIG._16B

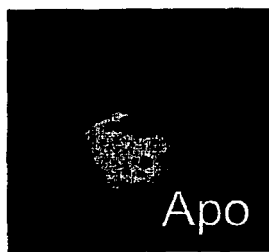


FIG._16C

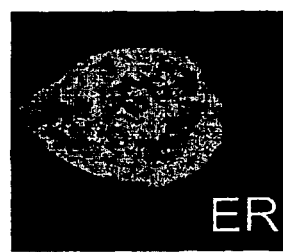


FIG._16D



FIG._16E

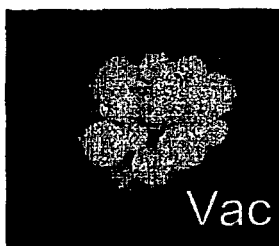


FIG._16F

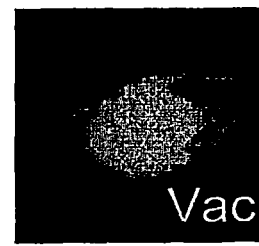


FIG._16G

Actividad de FAE en hojas de *Festuca arundinacea* transgénica de edades diferentes bajo las secuencias de reconocimiento en ER y APO

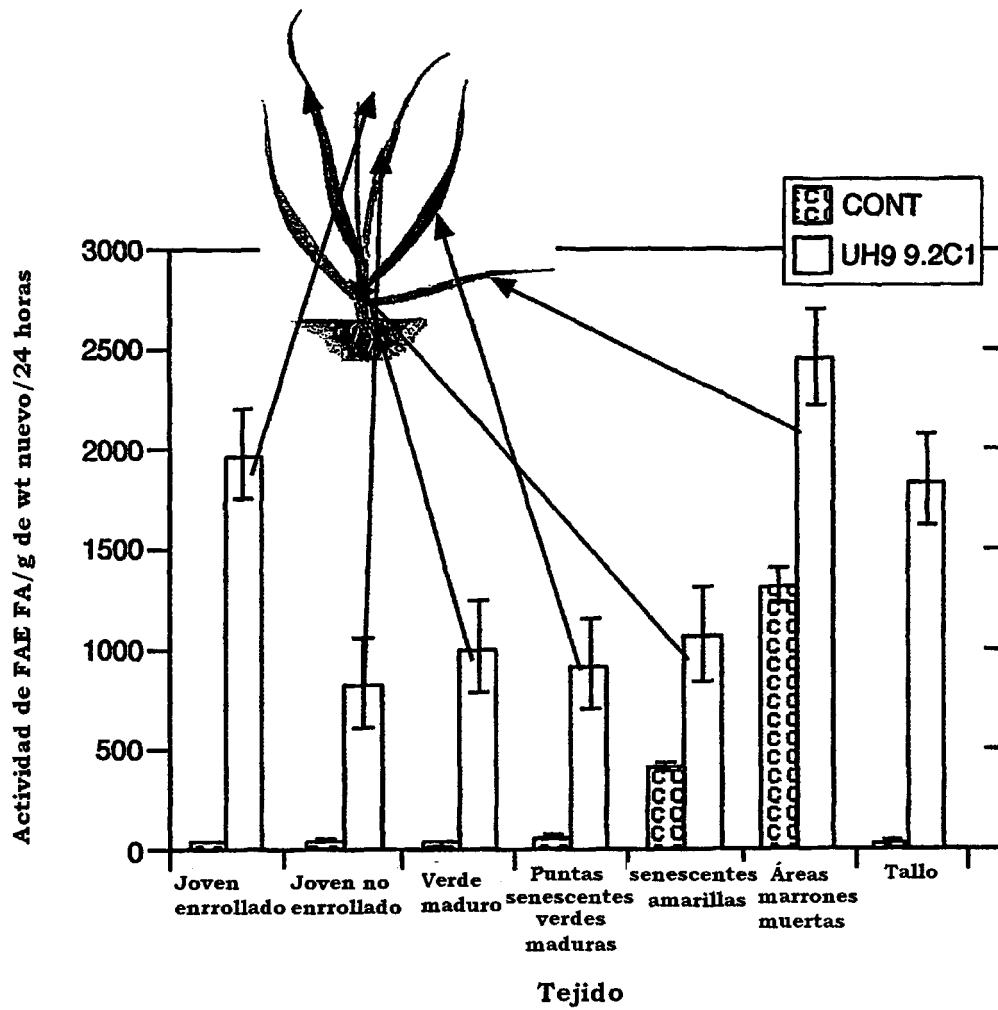


FIG. 17A

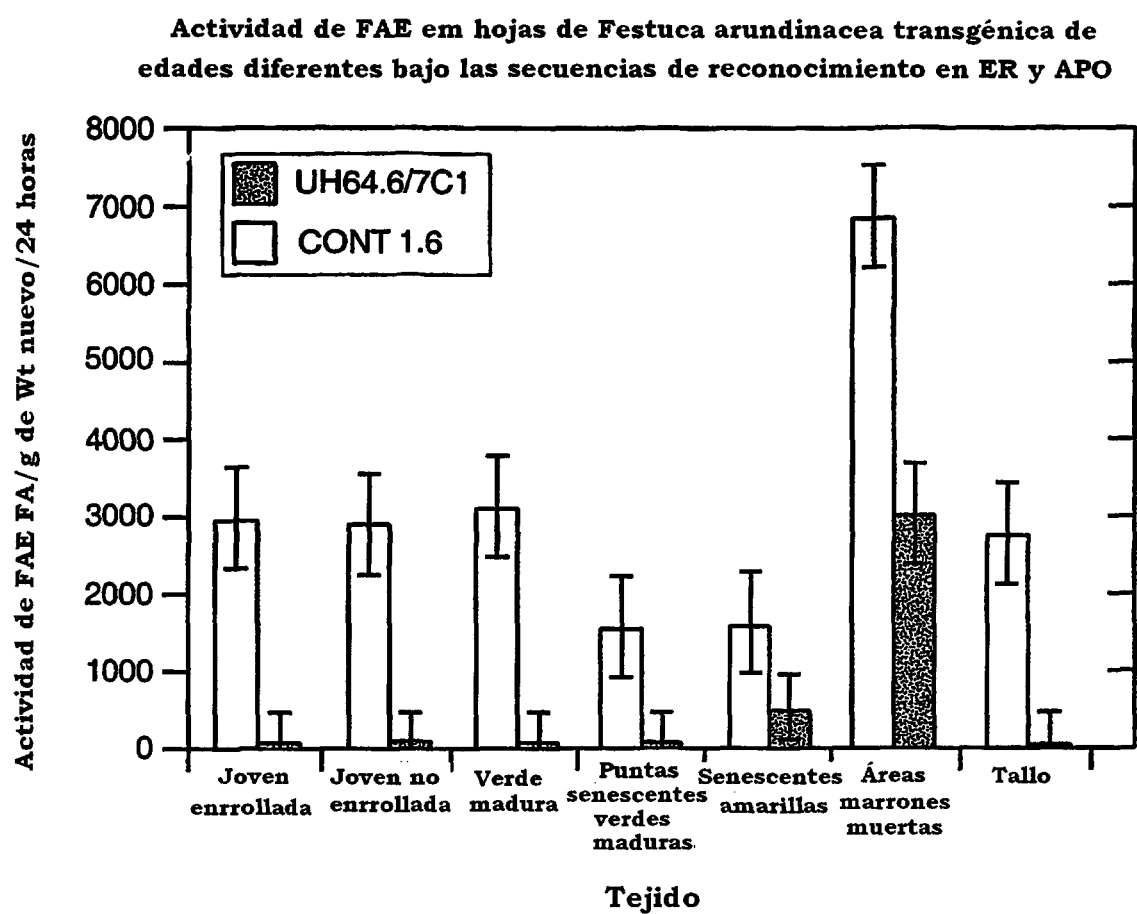


FIG._17B

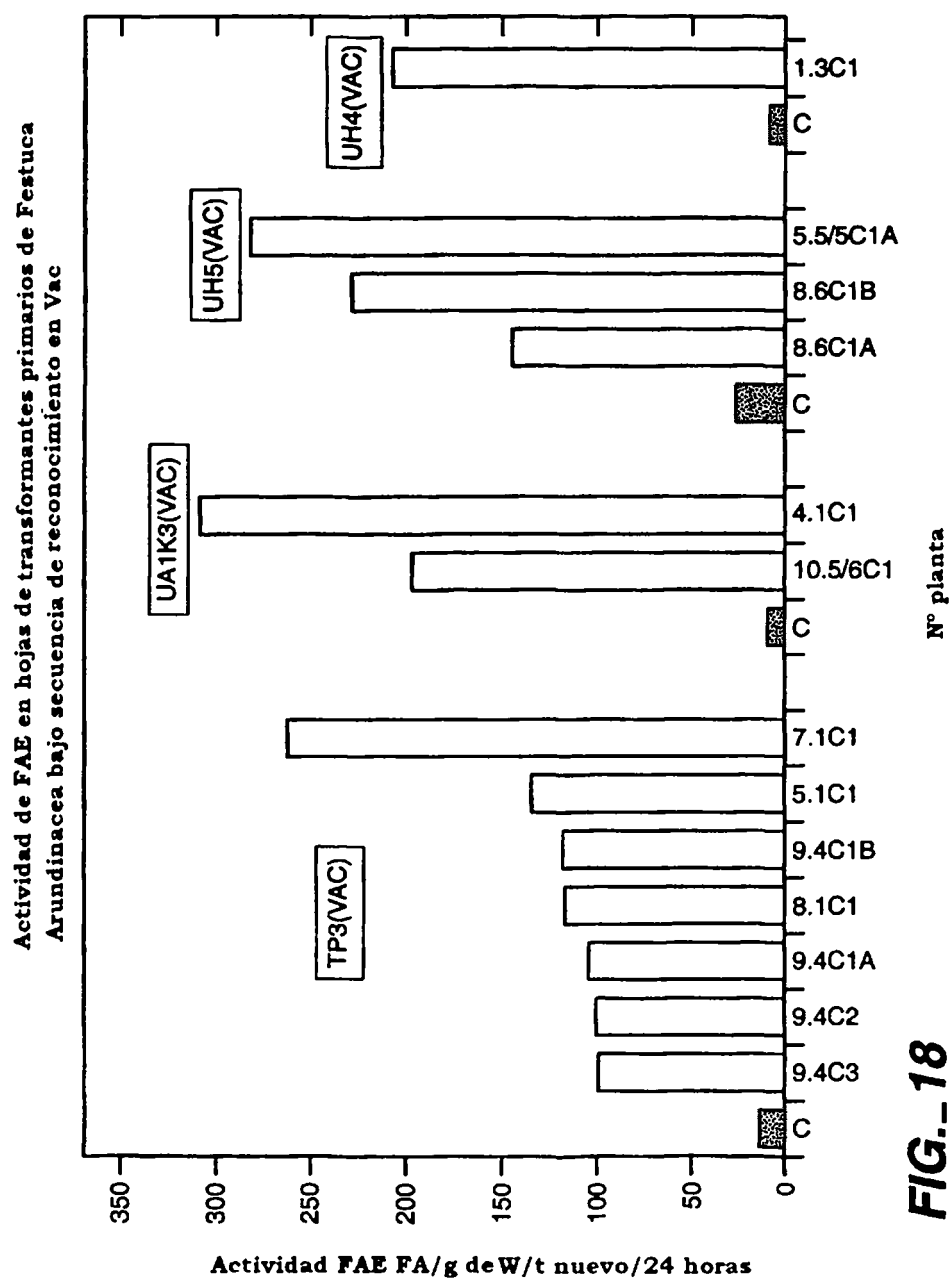


FIG._18

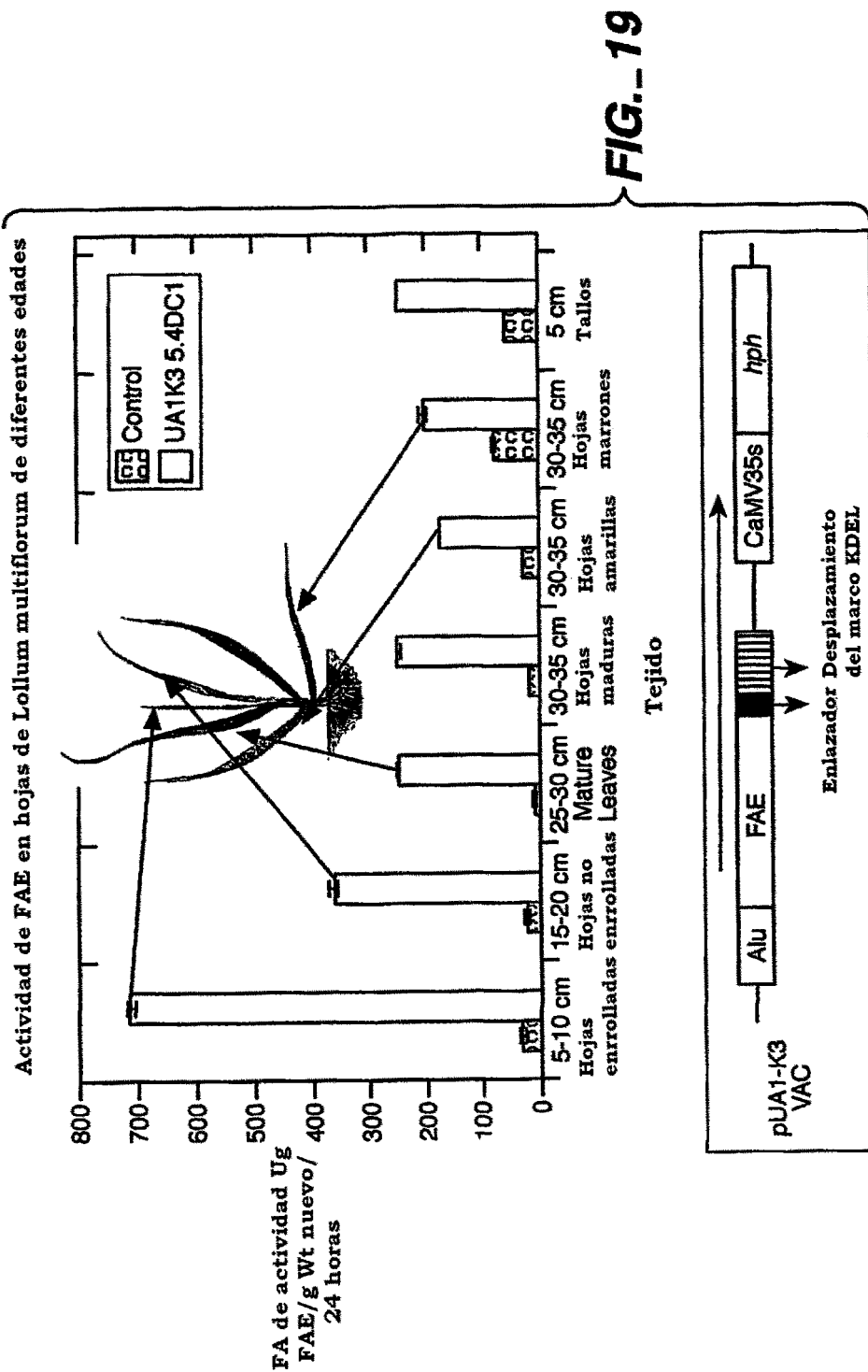


FIG.-19

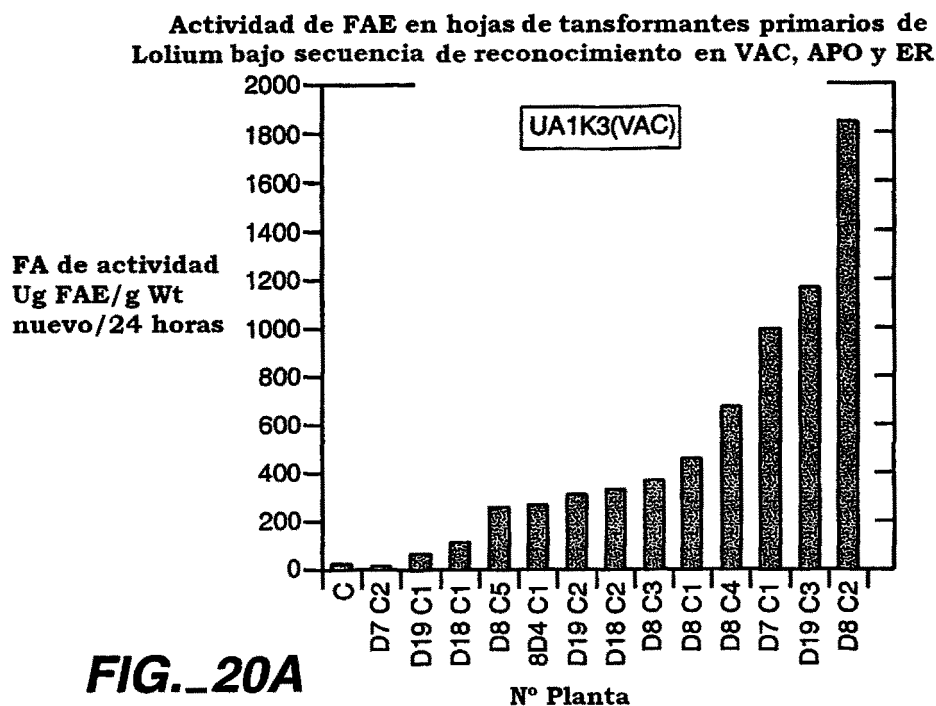


FIG._20A

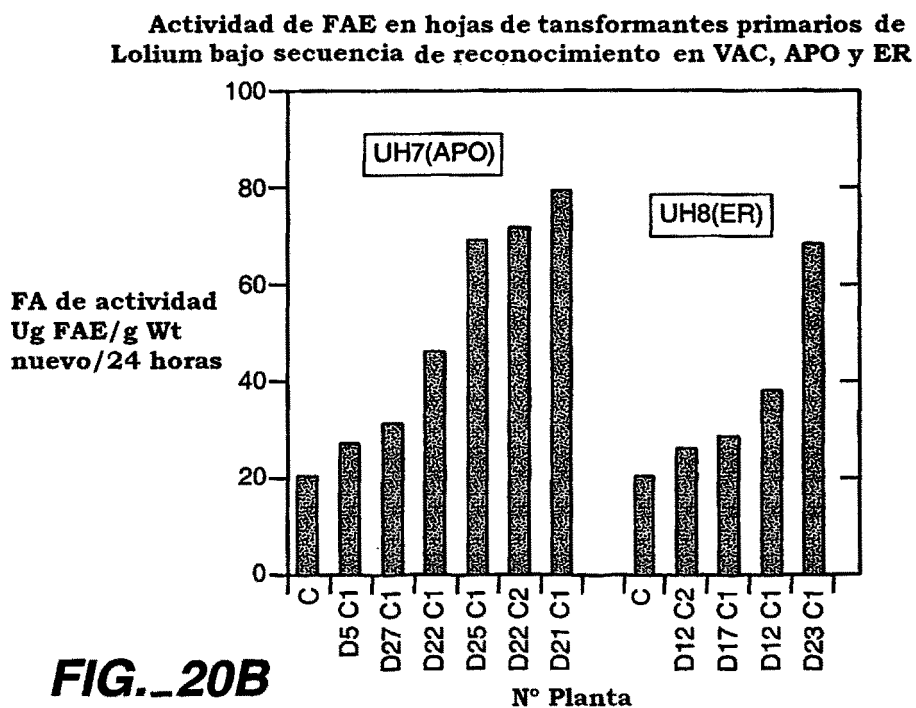


FIG._20B

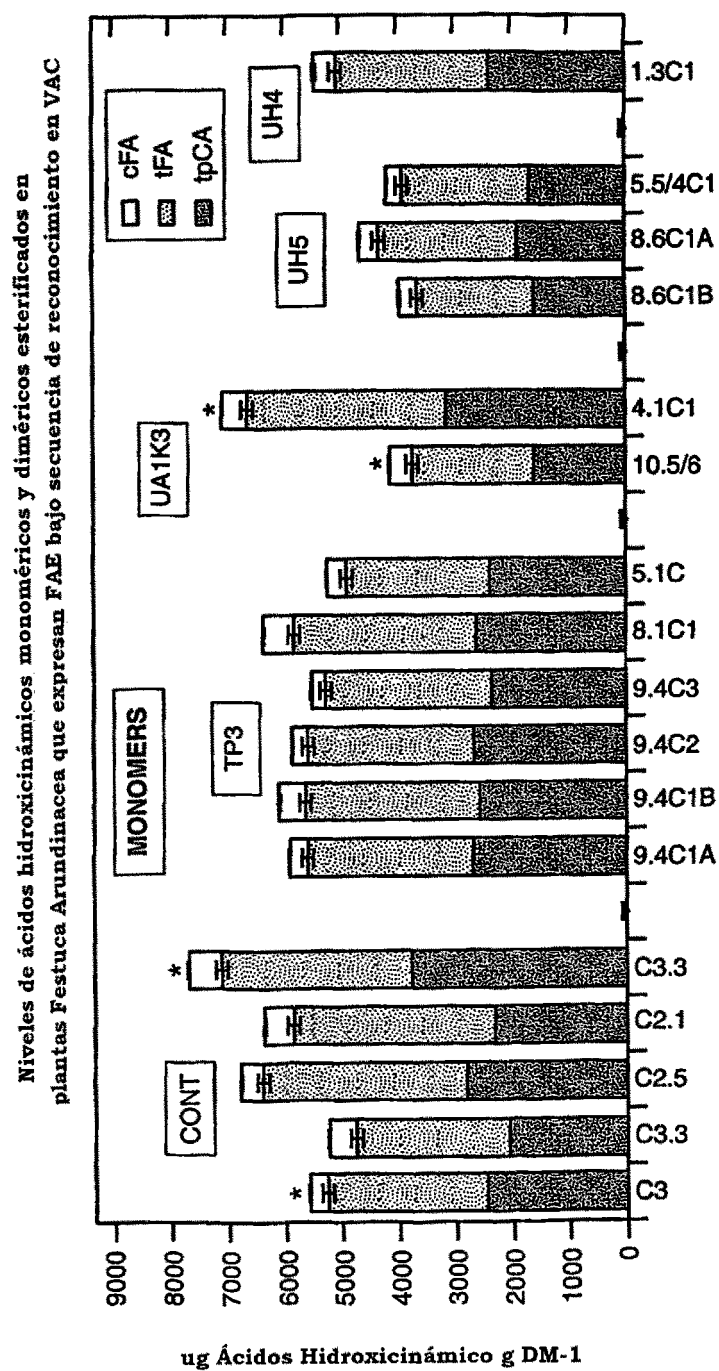
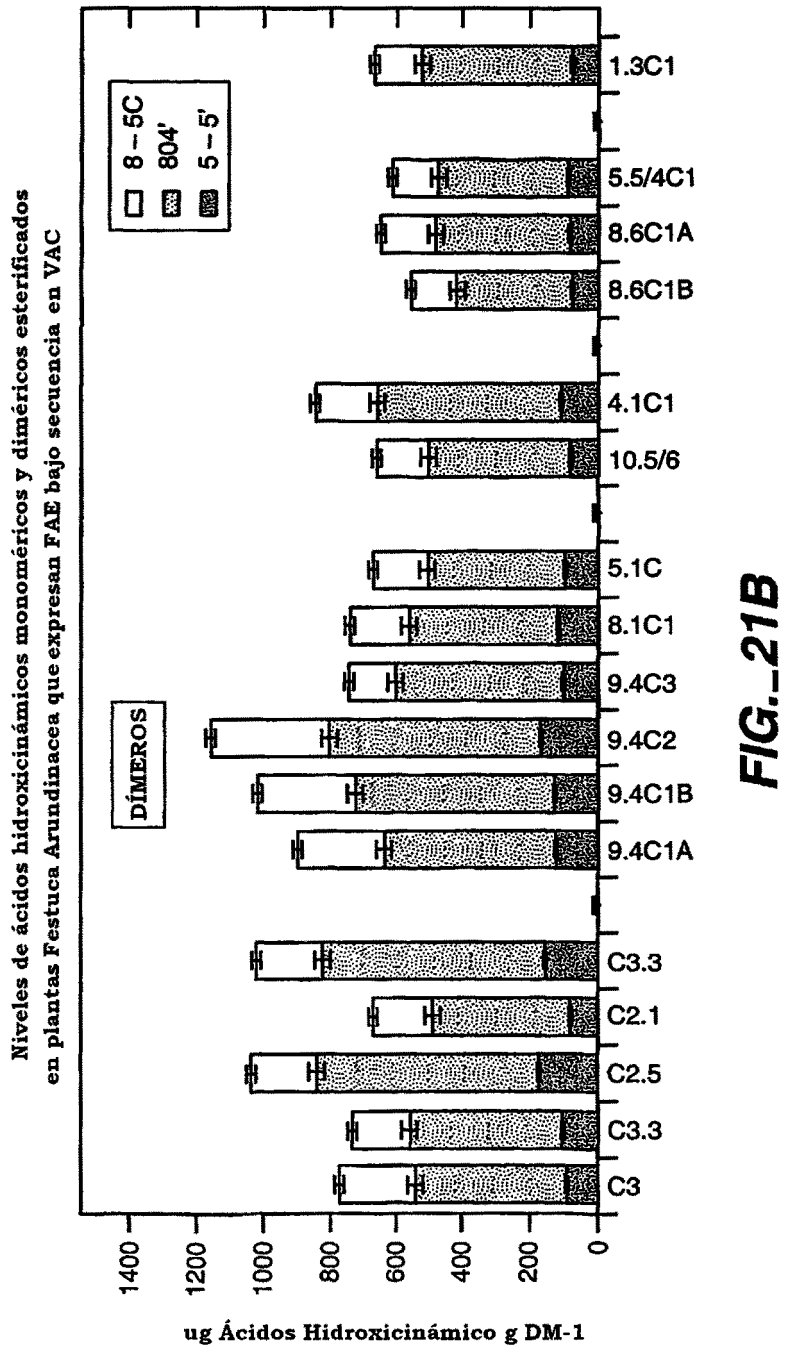
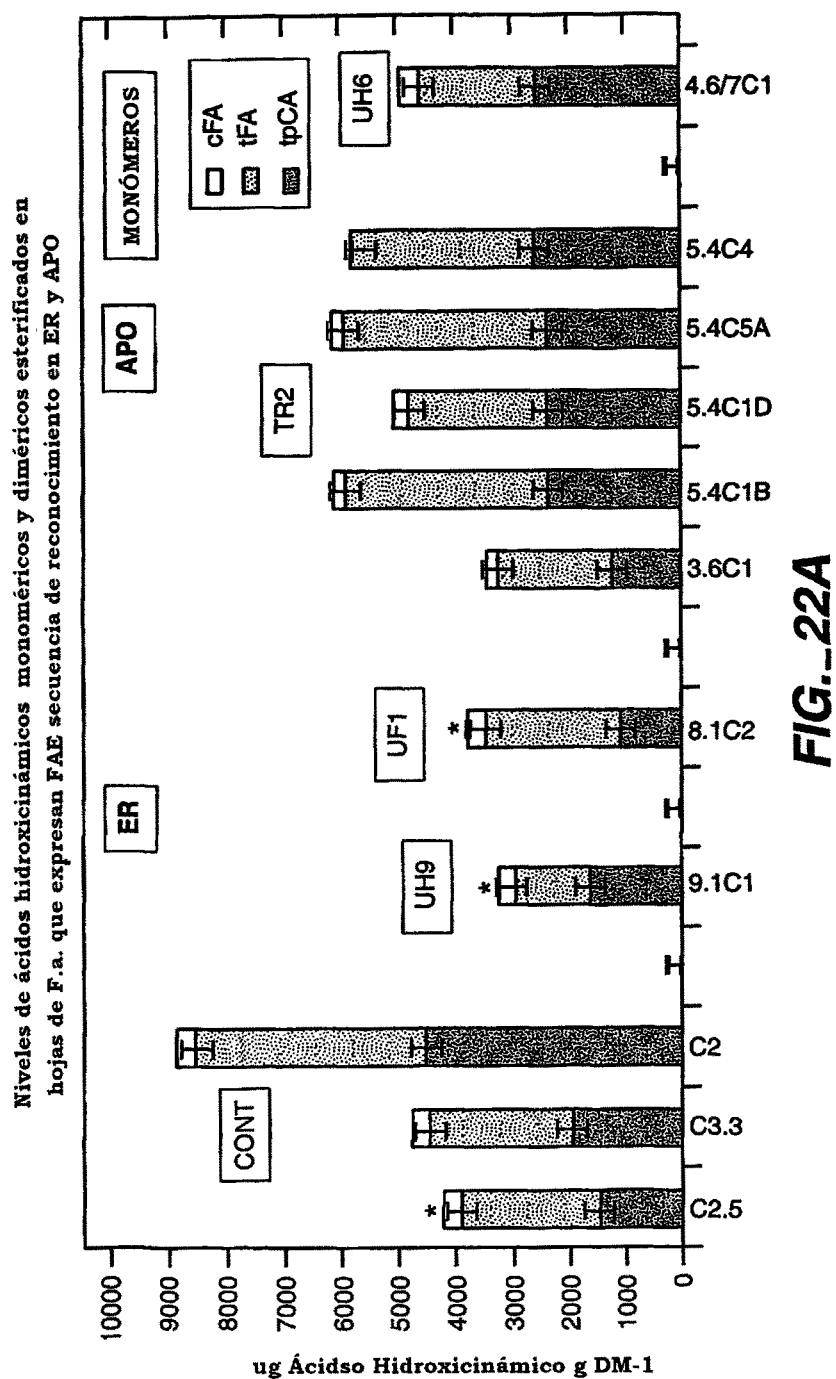
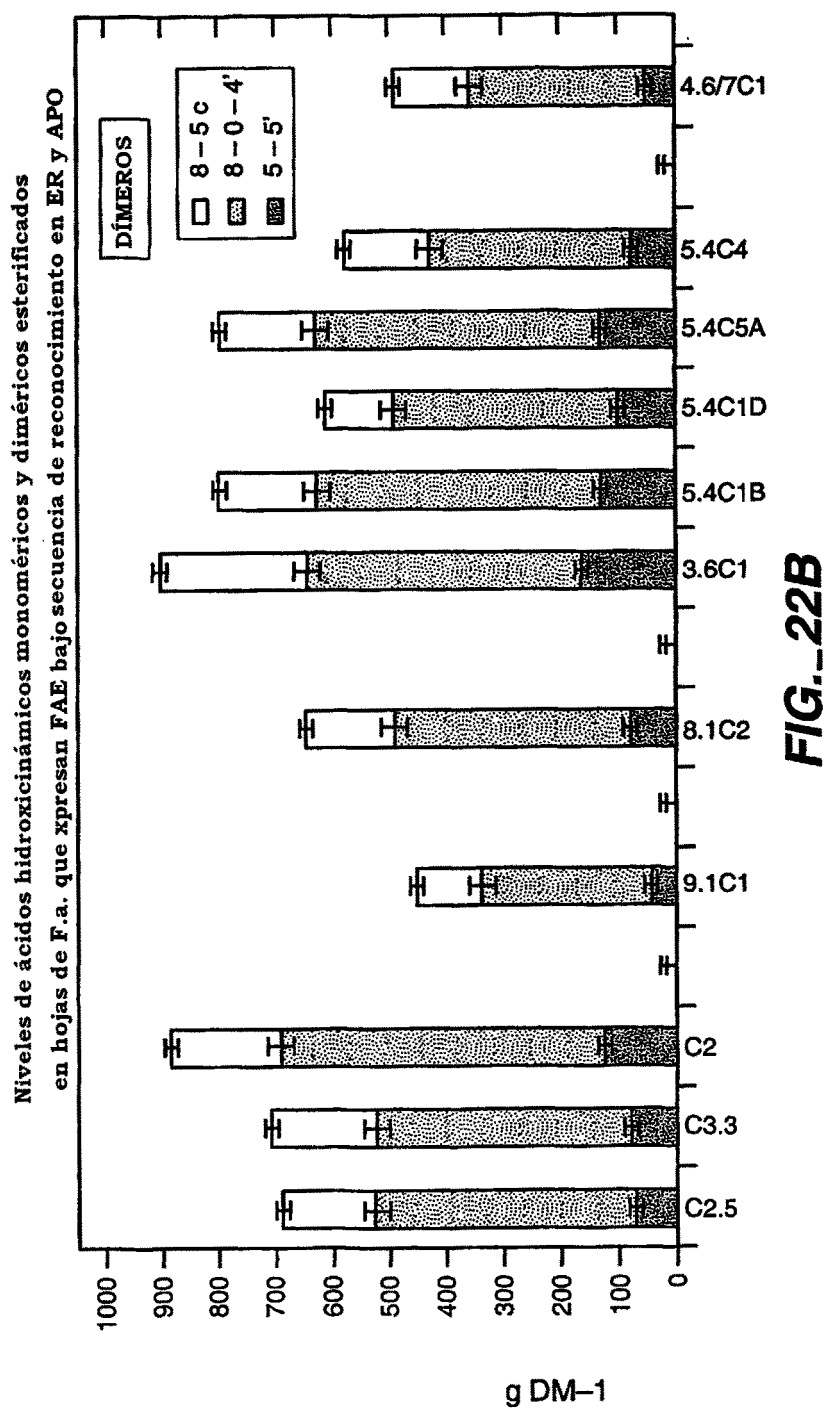


FIG. 21A







Digestibilidad de materia seca in vitro de tejido de hoja de plantas Festuca arundinacea maduras que expresan FAE bajo un promotor de actina

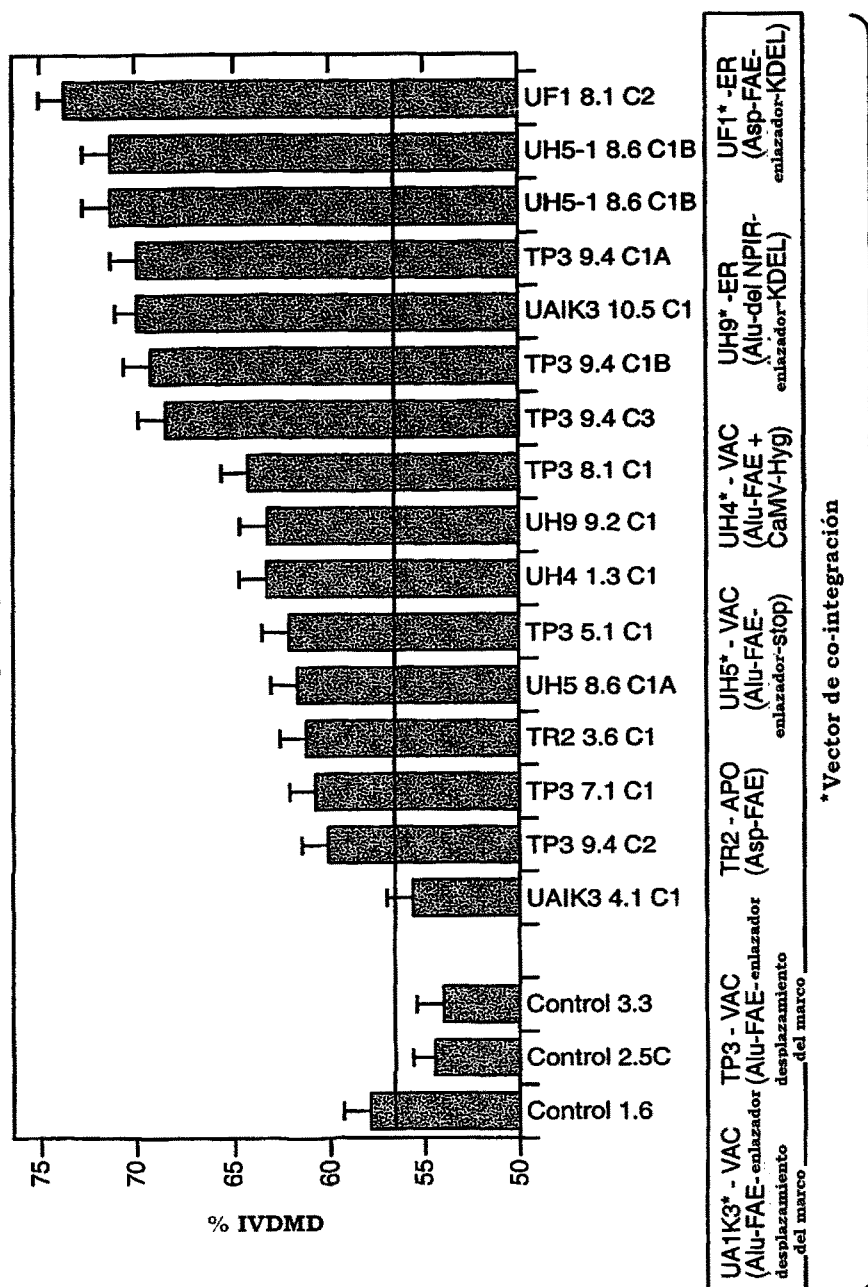


FIG._23

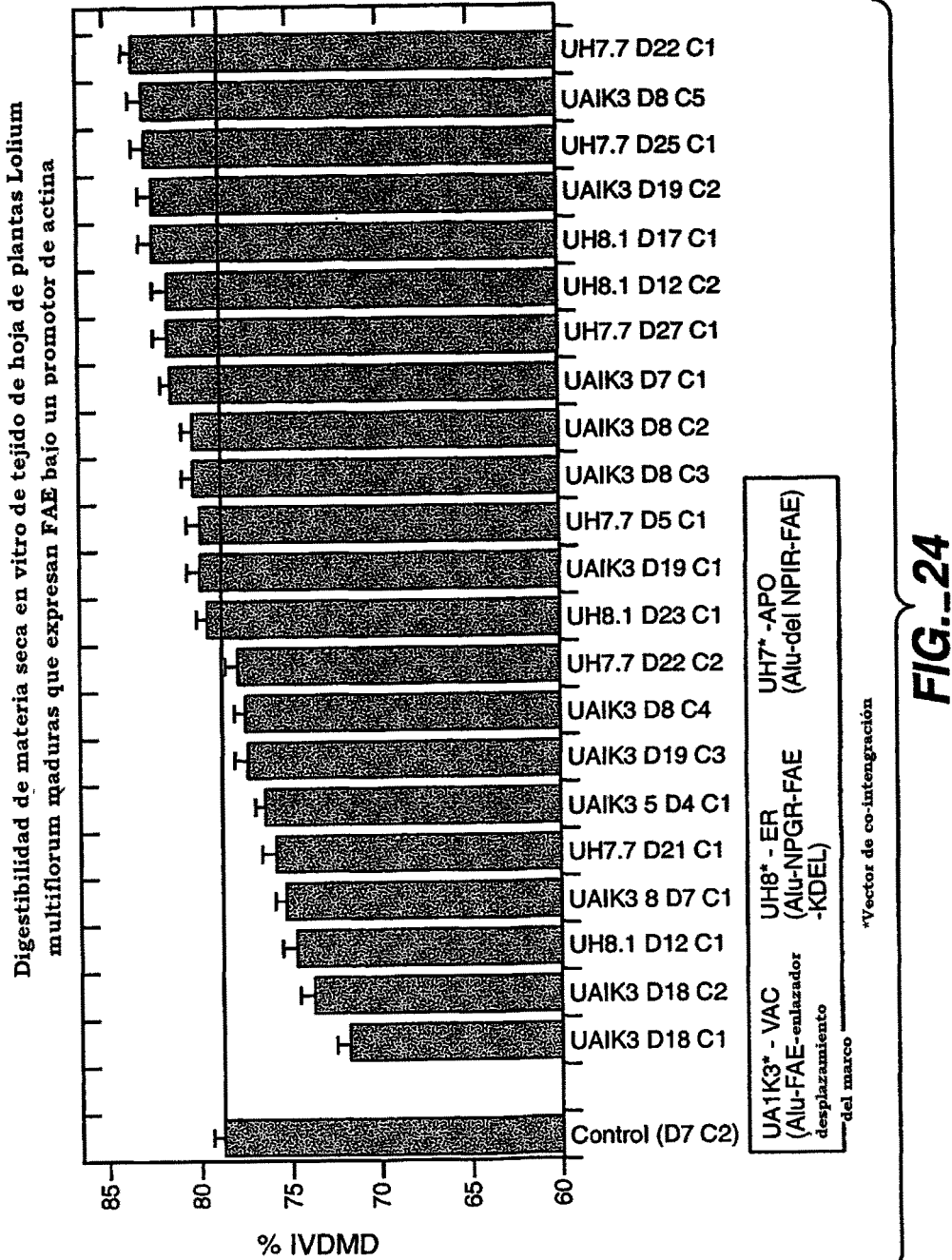


FIG..24

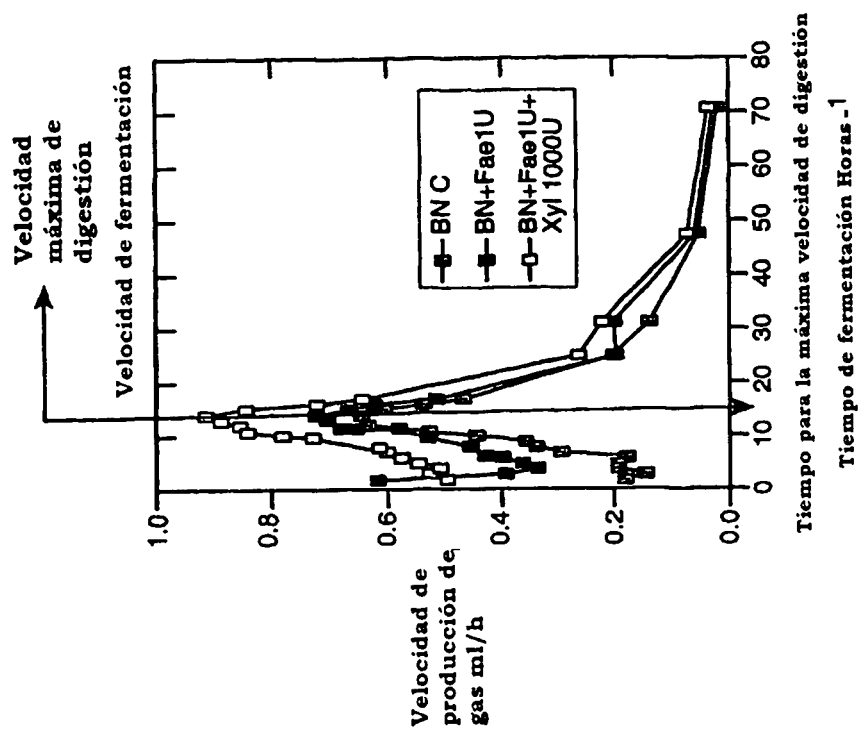


FIG. 25A

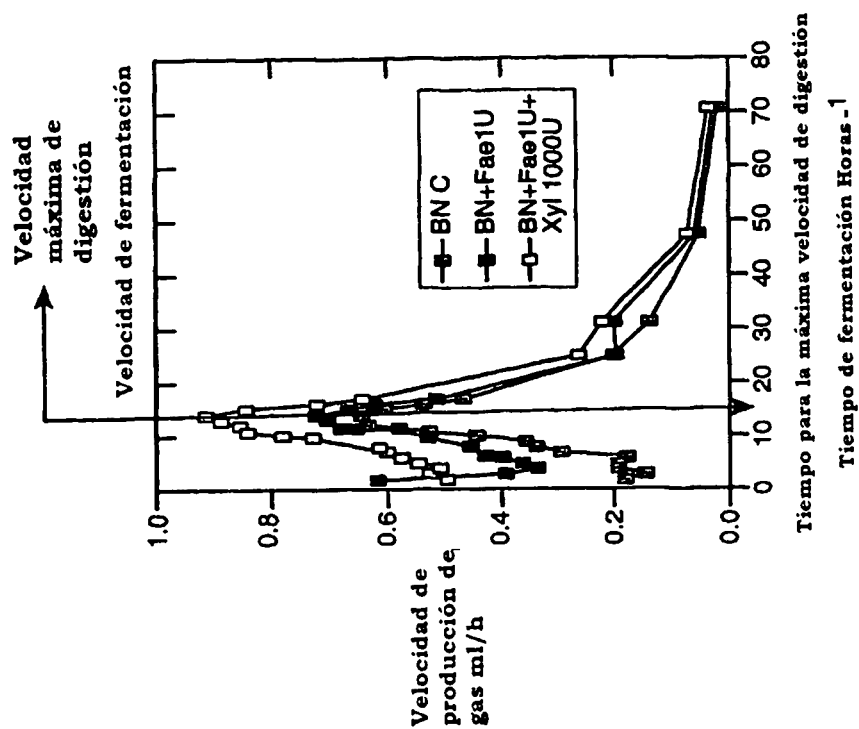
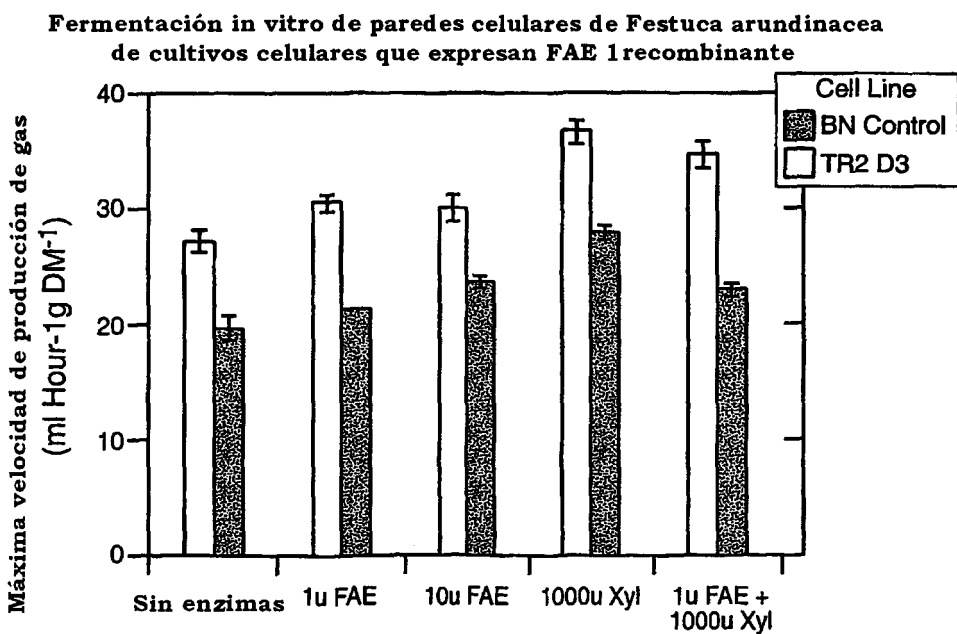
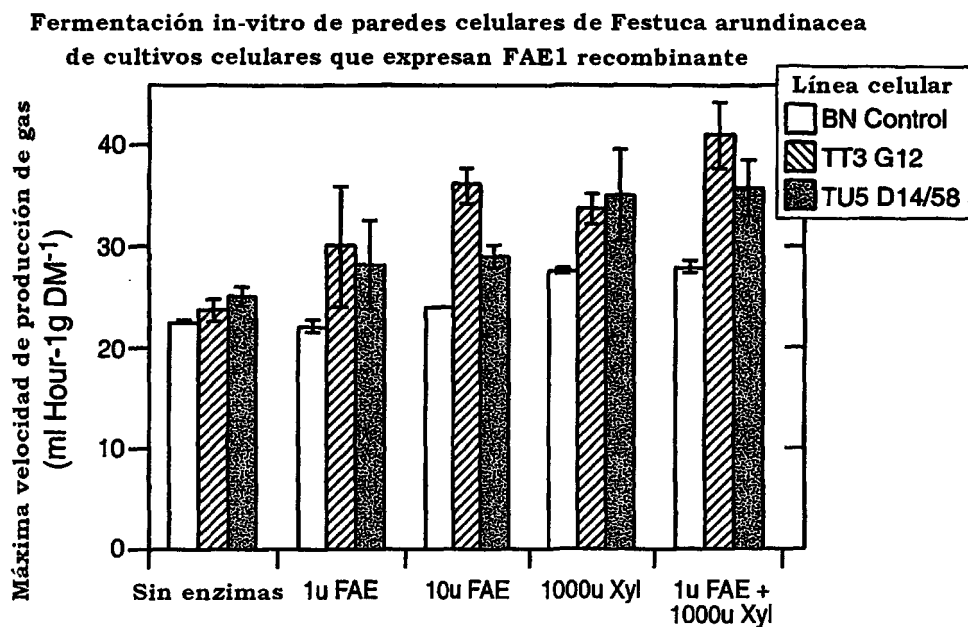


FIG. 25B



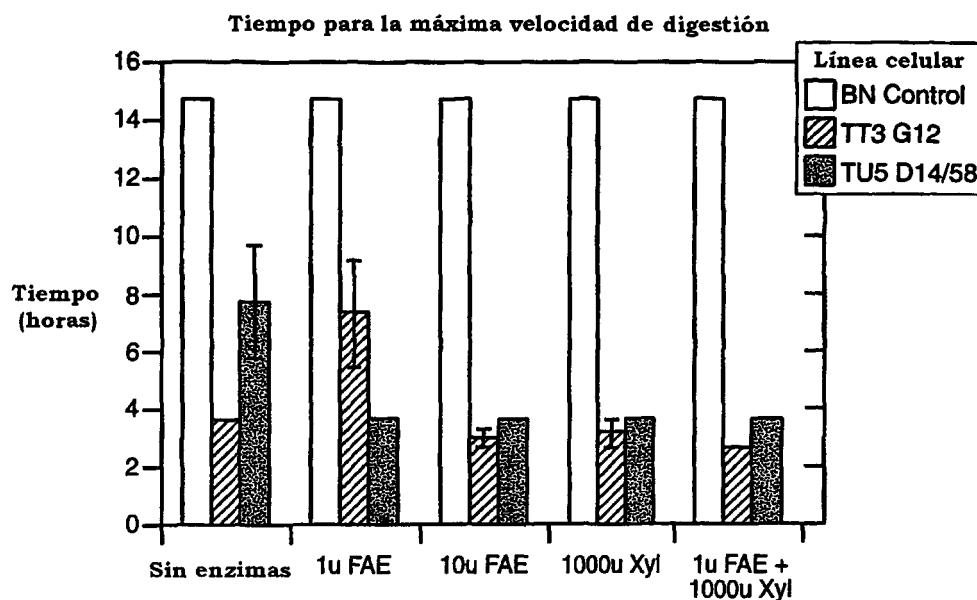


FIG._27A

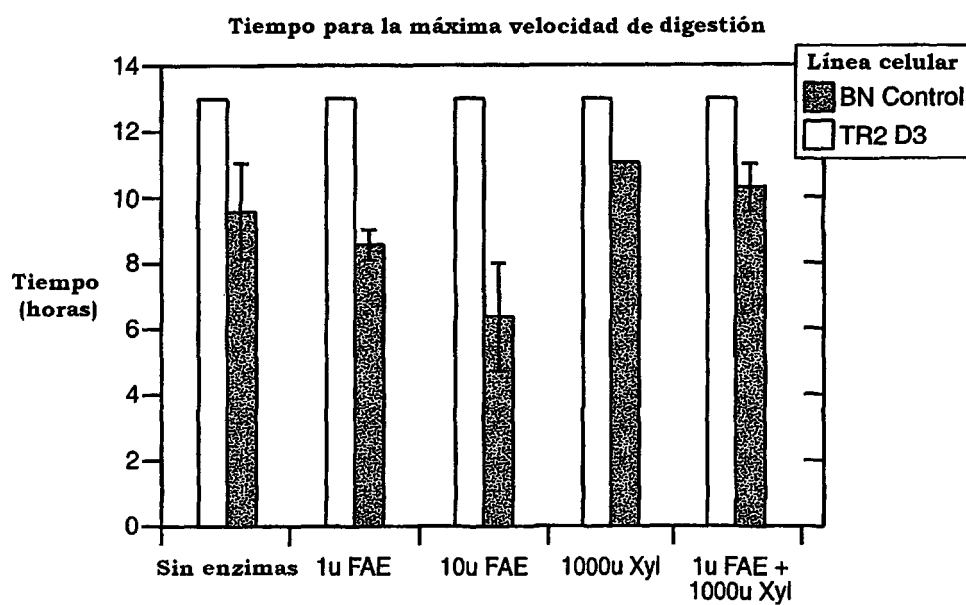


FIG._27B

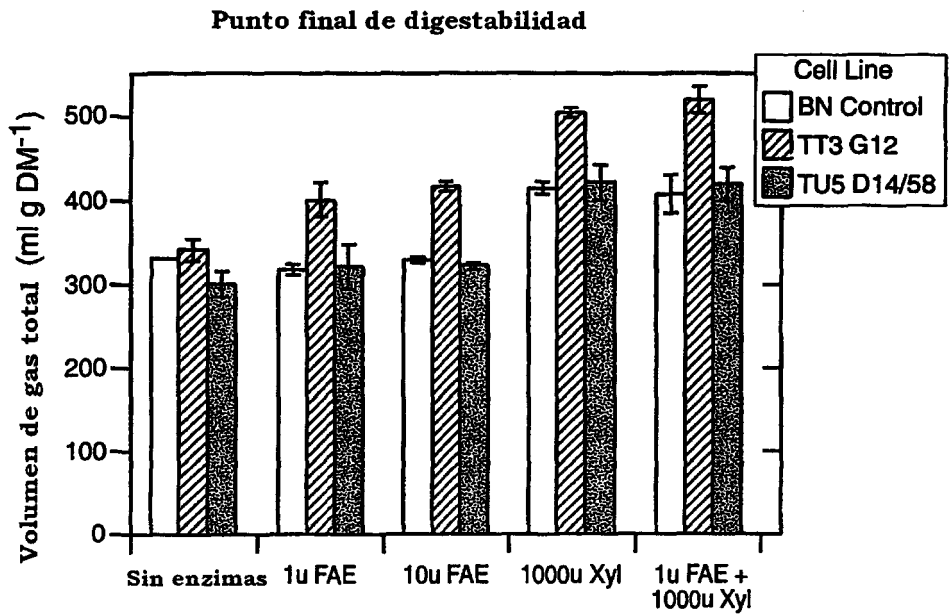


FIG._28A

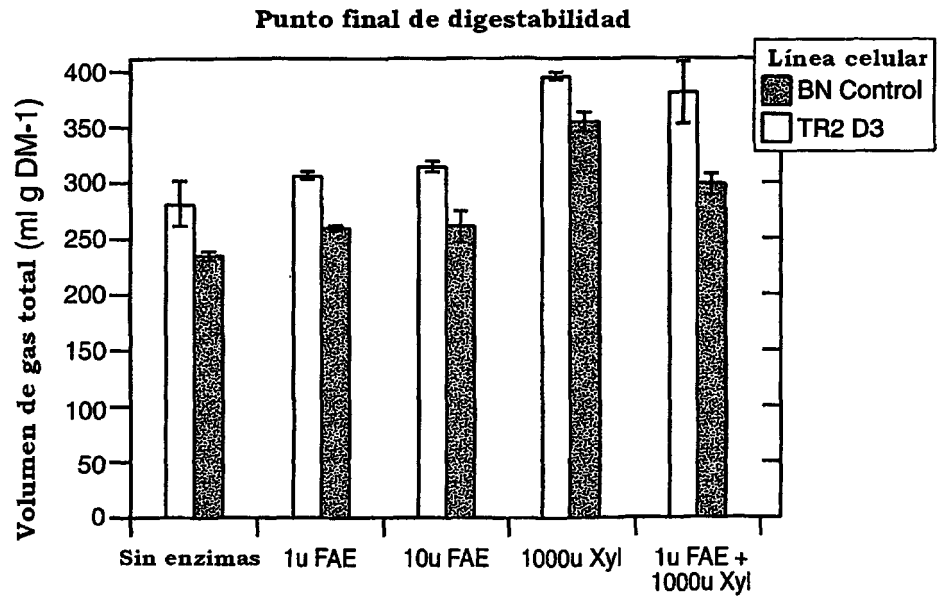
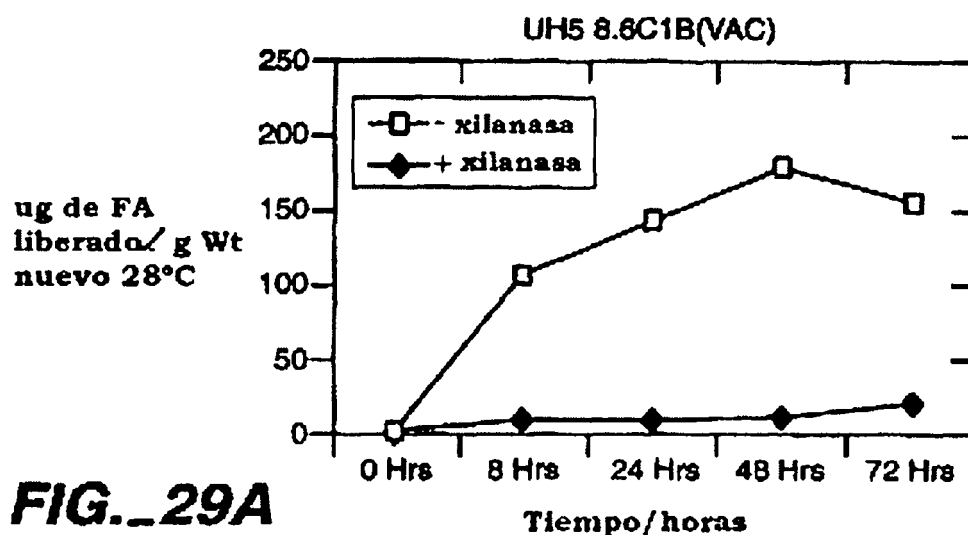
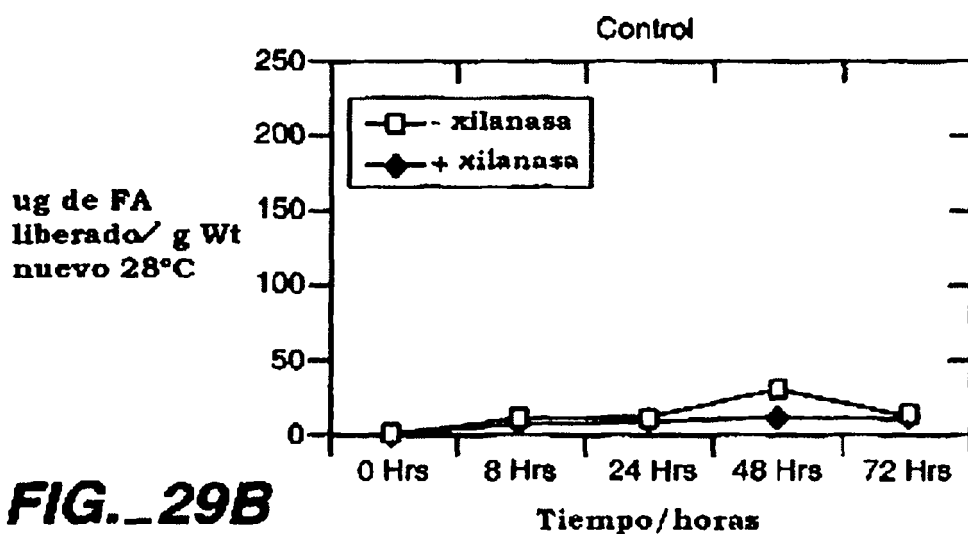


FIG._28B

Cinética de actividad de FAE mediante la liberación de ácido ferúlico de la pared celular bajo autodigestión en *Festuca arundinacea* y estimulación mediante xilanasa



Cinética de actividad de FAE mediante la liberación de ácido ferúlico de la pared celular bajo autodigestión en *Festuca arundinacea* y estimulación mediante xilanasa



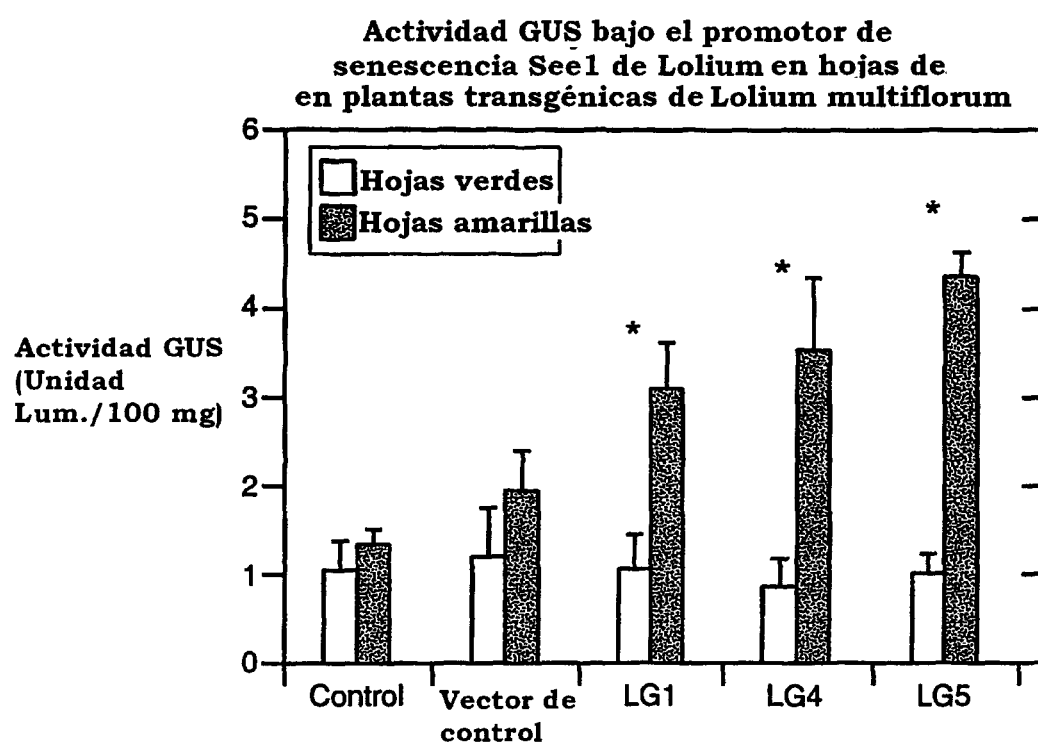


FIG._30

Liberación de monoméricos y diméricos en la autodigestión de hojas de plantas que expresan FAE reconocida en vacuolas

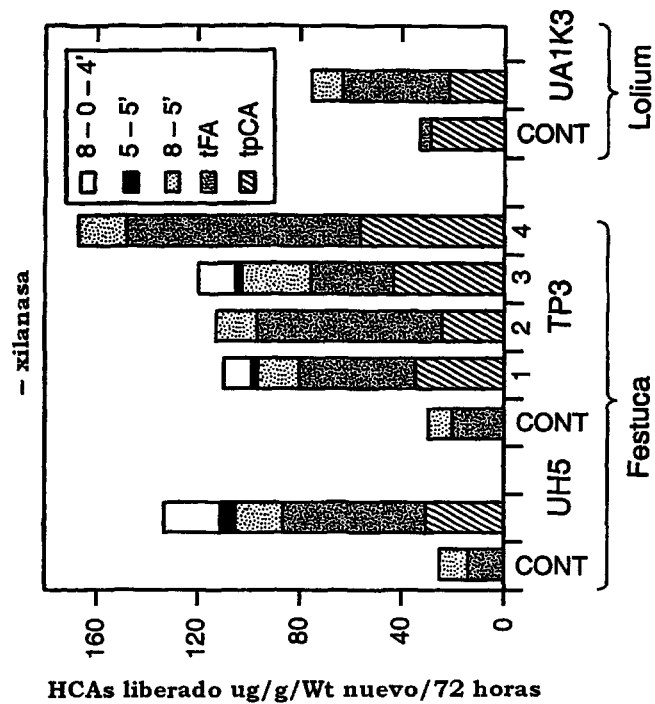


FIG. 31B

Liberación de monoméricos y diméricos en la autodigestión de hojas de plantas que expresan FAE reconocida en vacuolas

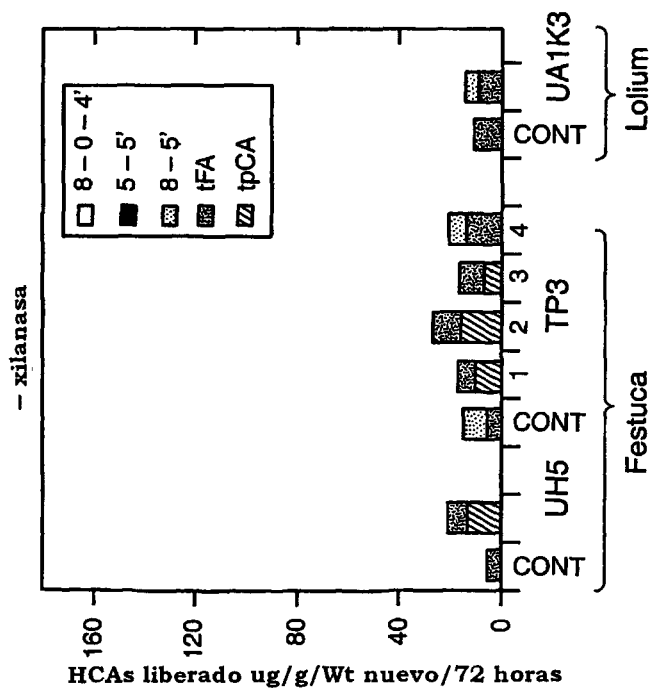


FIG. 31A

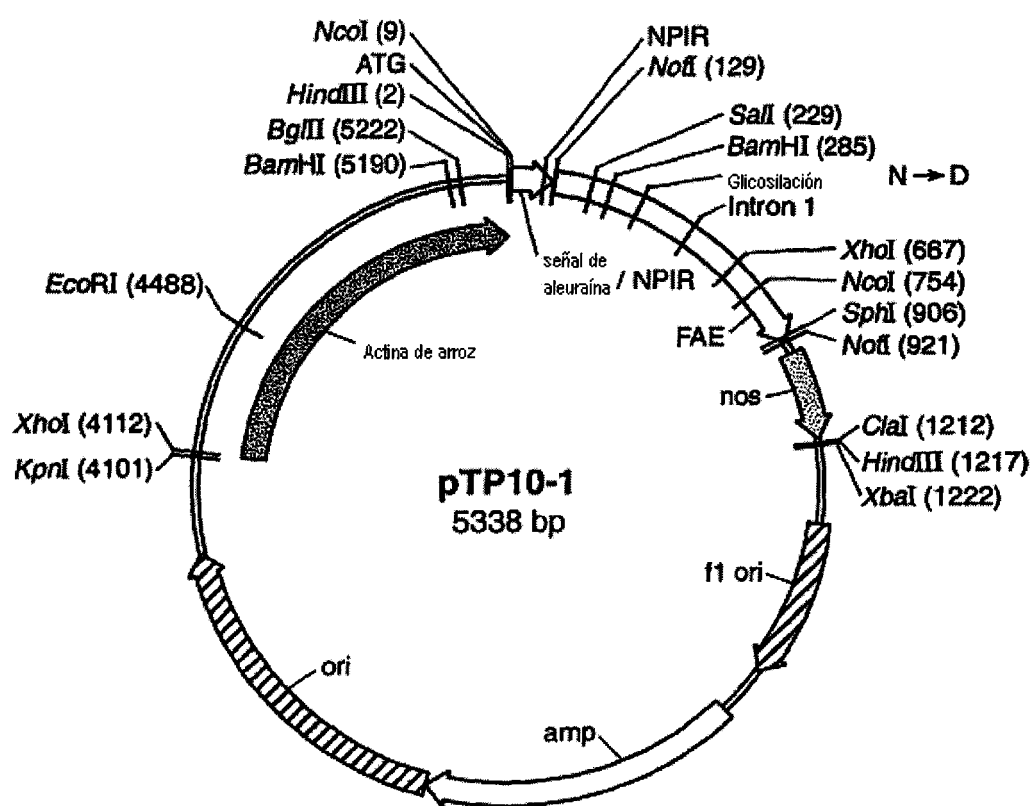


FIG._32A

```

NGOI
~~~~~
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 CTCCTGGCGC TCGCCGTGCT GGCACGGCC GCGTCGCCG

      . A S S S S F A D S N P I R P V T D R A A A S T .
71  TCGCCTCCTC CTCCTCCTC GCGACTCCA ACCGATCCG GCCGTCACC GACCGCGCG 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 TCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC

      NPIR
      ~~~~~
      NotI
      ~~~~~
      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 AAGGGAGAGA AATTATCAA TTCTCAACT GACATTACG

      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 GATGATCCT CCGCGACGAC AGCAGCAAG AATAATCAC CGTCTCCGT GGCATGGTA GTGATACGAA

      Glycosylation
      ~~~~~
      . L Q L D T D Y T L T P F D T L P Q C N G C E V
351 TCTACAACTC GATACTGACT ACACCTCACC GCCTTCGAC ACCCTACCAC ATGCAACGG 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 CACGGTGAT ATTATATTGG ATGGTCTCC GTCCAGACC AAGTCGATC GTTGTCAA CAGCAGGTA
      . 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 GACTACCG CTGACCGTGA CCGGCCACG CCTCGCGCC TCCCTGGCG 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 ACACCAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG

```

FIG._32B

```

XhoI
-----
631  A F A S Y M N D A F Q A S S P D T T Q Y F R V T
    GCCTTCGGT CGTACATGAA CGATGCCTTC CAAGCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA

NcoI
-----
701  . 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 CCAACCTGC CCGGCTGGA 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 CACATTGTG TGCACCTGGG ATGAAGTGCA GTGCTGTGAG

SphI
-----
841  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 GAATAATCG CACAGACTT ATTTGGGAT GACGAGCGGC GCATGCACCT

NotI
-----
911  . P V A A A E T T E G *
    GCGCGGTGCG GCGCGCGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAA CATTGGCAA
981  TAAAGTTTCT TAAGATTGAA TCCTGTGCG GGTCTGCGA TGATTATCAT ATATTTCTG TTGAATTACG
1051 TTAAGCATGT AATAATTNAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTATGA TTAGAGTCCC
1121 GCAATTATAC ATTAATATCG CGATAGAAA CAAATATAG CGCGCAAACT AGGTAAATT ATCGCGCGCG

HindIII
-----
1191 GTGTCATCTA TGTTACTAGA TCGATAAGCT TCTAGAGCG CCGGTGGAGC TCCAATTGCG CCTATAGTGA
1261 GTCGTATTAC GCGCGCTCAC TGGCGTCTGT TTTACAAAGT CGTGACTGGG AAAACCTGCG CGTTACCCAA
1331 CTTAATCGCC TTGCAGCACA TCCCGCTTTC GCGAGCTGGC GTANTAGCGA AGAGGCCCGC ACCGATCGCC
1401 CTTCCCAACA GTTGGCGCAGC CTGAATGGCG ATGGAGAGC GCCCTGTAGC GCGCATTTAA GCGCGCGCGG

ClaI
-----
XbaI
-----

```

FIG._32C

1471 TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCGC CCGCTCCTTT CGCTTCTCTC
 1541 CCTTCCCTTC TCGCCACGTT CGCCGGCTTT CCCCGTCAAG CTCTAAATCG GGGGCTCCCT TTAGGGTTCC
 1611 GATTTAGTGC TTTACGGCAC CTCGACCCCA AAAAATCTGA TTAGGGTGAT GGTTCACGTA GTTACGCTA GTGGGCCATC
 1681 GCCCTGATAG ACGGTTTCTC CACTCAACCC TATCTCGGTC TATCTTTTG GATTTGCGG ATTTGCGCTT TTTGTTCCAA
 1751 ACTGGAACAA CACTCAACCC TATCTCGGTC TATCTTTTG GATTTGCGG ATTTGCGCTT TTTGTTCCAA
 1821 ATTGGTTAAA AAATGAGCTG AATTAACAAA AATTAACGC GAATTTTAA CAAATATTA AAAATATTA CATTTACAAT
 1891 TTAGGTGGCA CTTTTCGGGG AAATGAGCTG AATTAACGC GAATTTTAA CAAATATTA AAAATATTA CATTTACAAT
 1961 TGTATCCGCT CATGAGACAA TAACCCCTGAT AAATGCTTCA AATATATTA AAAAGGAAGA GTATGAGTAT
 2031 TCAACATTTT CGTGTGCCCC TATTTCCCTT TTTTGGCGCA TTTTGGCTTC CTGTTTTCG TCACCCAGAA
 2101 ACGCTGGTGA AGTAAAGA TGTGAAGAT CAGTTGGGTG CACGAGTGGG TTACATCGAA CTGGATCTCA
 2171 ACAGCGGTAA GATCCTTGAG AGTTTTCGCC CCGAAGAAGG TTTTCCAATG ATGAGCACTT TTAAGTTCTT
 2241 GCTATGTGGC GCGGTATTAT CCCGTATTGA CGCCGGGCAA GAGCAACTCG GTCCCGCAT ACATATTCT
 2311 CAGAAATGACT TGGTTGAGTA CTCACCACTG ACAGAAAGC ATCTTACGGA TGGCATGACA GTAAGAGAAAT
 2381 TATGCAGTGC TGCCATAACC ATGAGTGATA ACACCTCGGC CAACTTACTT CTGACAAAGA TCGGAGGACC
 2451 GAAGGAGCTA ACCGCTTTT TGCACAACAT GGGGATCAT GTAACTCGCC TTGATCGTTG GGAACCGGAG
 2521 CTGAATGAAG CCATACCAA CGACGAGCGT GACACCACGA TGCCGTGAGC AATGGCAACA ACGTTGCGCA
 2591 AACTATTAACT TGGCGAACTA CTTACTCTAG CTTCCCGGCA ACAATTAATA GACTGGATGG AGCGGATAA
 2661 AGTTGCAGGA CCACCTCTGC GCTCGGCCCT TCCGGCTGGC TGCTTTATTG CTGATATAATC TGGAGCCGGT
 2731 GAGCGTGGGT CTCGCGGTAT CATTCGACGA CTGGGGCCAG ATGATAAGCC CTCCCGTATC GTAGTTATCT
 2801 ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAAATAG ACAGATCGCT GAGATAGGTG CCTCACCTGAT
 2871 TAAGCATGGG TAACTGTCTAG ACCAAGTTTA GATCCTTTT GATAATCTCA TGACCAAAAT CCTTAAACGT
 2941 TTTAAAGGA TCTAGGTGAA GATCCTTTT GATAATCTCA TGACCAAAAT CCTTAAACGT
 3011 TCCACTGAGC GTCAGACCCC GTAGAAAAGA TCAAGGATC TTTCTTGAGAT CTTTCTTTTC TGCGCGTAAT
 3081 CTGCTGCTTG CAAACAAAAA AACCAACCGT ACCAGCGGTG GTTTGTTTGC CGGATCAAGA GCTACCAACT
 3151 CTTTTCCTGA AGGTAACCTGG CTTACAGCAG GCGCAGATAC CAAATACTGT CCTTCTAGTG TAGCCGTAGT
 3221 TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCCTACATA CCAATACTGT CCTTCTAGTG TAGCCGTAGT
 3291 TGCTGCCAGT GCGGATAAGT CGTGTCTTAC CCGGTTGGAC TCAAGACGAT AGTTACCGGA TAAGGCGCAG
 3361 CGGTCGGGCT GAACGGGGG TTCTGTGCACA CAGCCAGCT TGGAGCGAAG GACCTACACC GAACGTGAGT
 3431 ACCTACAGCG TGAGCTATGA GAAAGCGCCA CGCTTCCCGA AGGGAGAAAG CGGACAGGT ATCCGGTAAG
 3501 CGGCAGGGTC GGAACAGGAG AGCGCACGAG GGAGCTTCCA GGGGGAACG GATGCTCGTC AGGGGGCGG
 3571 GTCGGGTTTC GCCACCTCTG ACTTGAAGCGT CGATTTTGT GATGCTCGTC AGGGGGCGG
 3641 AAACGCCAG CAACGCGGCC TTTTAAACCG TCCTGGCTTT TTTGCTCACA TGTCTTTTC
 3711 TGCCTTATCC CCTGATTTCTG TGGATAACCG TTTTACCGCC TTTGAGTGAG TGATATACCG
 3781 CGAACGACCG AGCGCAGCGA GTCAATGAGT GAGGAAGCGG AAGAGCGCCC AATACGAAA CCGCTCTCTC
 3851 CCGCGCGTTG GCCGATTCTA TAATGCAAGT GGCACGACAG GTTTCCCGAC TGGAAAGCGG GCAGTGAGCG

FIG._32D

```

3921 CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT
3991 ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC

                                KpnI
                                ~~~~~
4061 GCGCAATTAA CCTCACTAA AGGGAACAAA AGCTGGGTAC CGGGCCCCCC CTCGAGGTCA TTCATATGCT
4131 TGAGAAAGAGA GTCGGGATAG TCCAAAATAA AACAAAGGTA AGATTACCTG GTCAAAAGTG AAACATCAG
4201 TTAAAGGTG GTATNAGTAA AATATCGGTA ATAAAAGGTG GCCCAAAGTG AAATTACTC TTTTCTACTA
4271 TTATAAAAAT TGAGGATGTT TTGTCCGTAC TTGTATACGT CATTTTGTG TGAATTGGT TTTAAGTTA
4341 TTCGCGATTG GGAATGCAT ATCTGTATT GAGTCGGTTT TTAAGTTCGT TGCCTTTGTA AATACAGAGG
4411 GATTGTGATA AGAATATCT TTAATAAAC CATATGCTAA TTGACNTAA TTTTGAAGAA AATATATAT

                                EcoRI
                                ~~~~~
4481 TCAGGCGAAT TCCACAATGA ACAATAATA GATTAAATA GCTTGCCCCC GTTGCAGCGA TGGGTATTTT
4551 TTCTAGTAAA ATAAAAGATA AACTTAGACT CAAAACATTT ACATAAACAA CCCCTAAAAGT CCTAAAGCCC
4621 AAAGTGCTAT GCACGATCCA TAGCAAGCCC AGCCCAACCC AACCCACCCC AGTGCAGCCA
4691 ACTGGCAAT AGTCTCCACC CCGGCACCTA TCACCGTGAG TTGTCCGCAC CACCGCACGT CTCGCAGCCA
4761 AAAAAAAA AAGAAAGAAA AAAAAAGAAA AGAAAGAAA CAGGTGGGTC CGGTCTGTGG GGGCCGGAAA
4831 AGCGAGGAGG ATCGCGAGCA GCGACGAGGC CCGGCCCTCC CTCGCTTCC AAAGAAACGC CCCCCATCGC
4901 CACTATATAC ATACCCCCC CTCTCTCTCC ATCCCCCCTA CCTACCACC ACCACCACCA CCACCTCCTC
4971 CCCCCTCGCT GCCGACGAC GAGCTCCTCC CCCCCTCCCC TCCGCCGCCG CCGGTAAACCA CCCCGCCCTT
5041 CTCCTCTTTC TTCTCCGTT TTTTCTTTCG TCTCGGTCTC GATCTTTGGC CTTGGTAGTT TGGGTGGGCG
5111 AGAGCGGCTT CGTCGCCCAG ATCGGTGCGC GGAAGGGGCG GGAATCTCGC GCTGGCGTCT CCGGGCGTGA

                                BamHI
                                ~~~~~
5181 GTCGGCCCGG ATCCTCGCGG GGAATGGGGC TCTCGGATGT AGATCTTCTT TCCTTCTTCT TTTTGTGGA
5251 GAAATTGAAAT CCTCAGCAT TGTTCATCGG TAGTTTCTT TTTTCATGAT TGTGACAAAT GCAGCCTCGT
5321 GCGGAGCTTT TTTGTAGC

```

FIG..32E

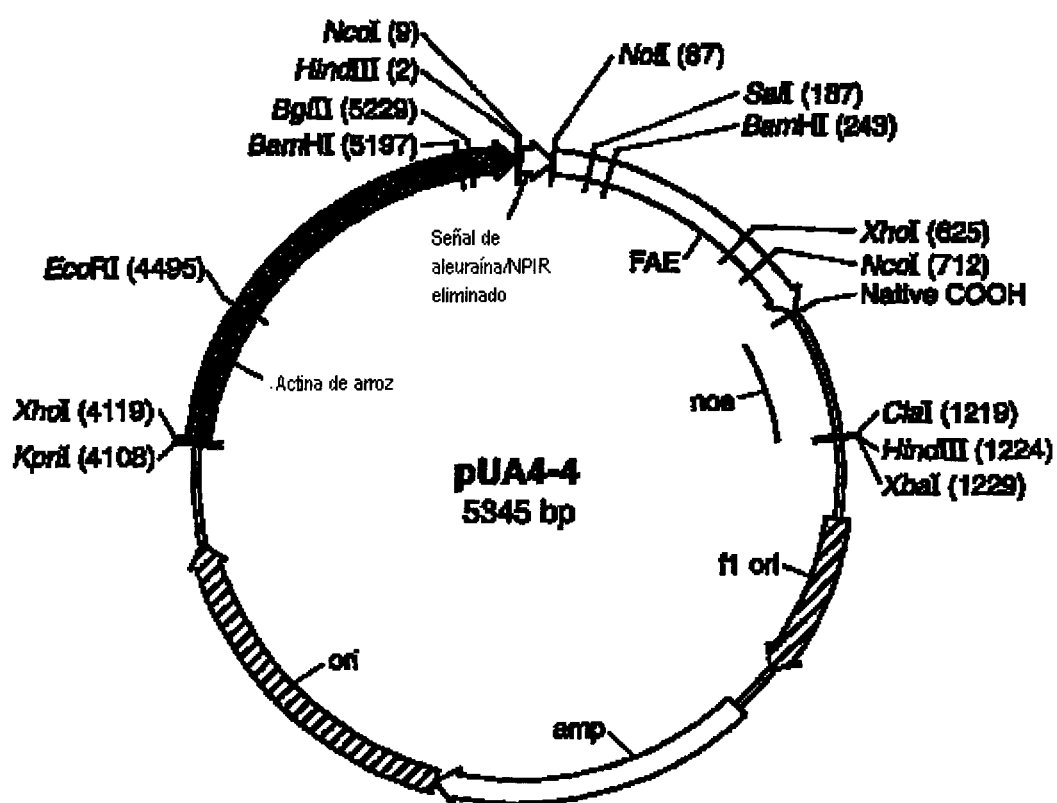


FIG._33A

```

      NcoI
      ~~~~~
HindIII
~~~~~
      M A H A R V L L L A L A V L A T A A V A V
1  AAGCTTACCA TGGCCCAAGC CCGGTCCTC CTCCTGGCGC TCGCCGTGCT GGCACGCGC 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 GCCTCCAGC 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 ATTCCGTGCA 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 ATTACAAATT CTCAAACTGA CATTAACGGA TGGATCCTCC GCGACGACAG CAGCAAGAA ATATCACC
      . F R G T G S D T W L Q L D T N Y T L T P F D T .
281 TCTTCGTTGG CACTGGTAGT GATACGAATC TACAACCTGA TACTAATAC ACCCTACGC 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 TGCNACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGTCTCCGT 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 TTGTCAACA GCAGGTIAGC CAGTATCCG ACTACGCGCT GACCGTACC GGCACACCC
      . 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 CTCACTGCC CTCAGCTGTC TCGGACATAC GACACATCC 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 CTTCCGGCGAA CCGCGCAGC GCAATCAGC CTTCCGCTCG TACATGACG ATGCTTCCA AGCCTCGAGC
      P D T T Q Y F R V T H A N D G I P N L P P V E Q
631 CCAGATACGA CGCAGTATT CCGGTCACT CATGCCACG 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

Clai Xbair

FIG. 33C

```

2731 AGCCGGTGAG CGTGGGCTCTC GCGGTATCAT TGCAGCACATG GGGCCAGATG GTAAAGCCCTC CCGTATCGTA
2801 GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG ATAGGTGCCT
2871 CACTGATTAA GCATTGGTAA CTGTGAGACC AGTTTACTC ATATATACTT TATATTGATT TAAACTTCA
2941 TTTTAAATT AAAGGATCT AAAGCCCGTA AGATGAGTCA GAAAGATCTC AGGATCTT TTTTCTGC
3011 TTTTCGTTCC ACTGAGCGTC ACATAAAGC AAGATGATCA CACCGTACC AGCGTGGTT TGTGTCGG
3081 GCATAATCTG CTGCTTGCAA AAAAAAAGC CACCGTACC CAGCATACCA CAGATACCAA ATACTGCTC TCTAGTGTAG
3151 ACCAACTCTT TTTCCGAAGG TAACTGGCTT CAAGAAGCTT GTAGCACCGC GTTACACTCA AGCGATAGT TACCGGATAA
3221 CCGTAGTTAG GCCACCACTT TGCCAGTGGC GATAAGTCTGT GTCTTACCGG GTTGGACTCA CCGGATAGT TACCGGATAA
3291 CAGTGGCTGC TCGGCTGAA CCGGGGCTTC GATTAAGTCTGT GTCTTACCGG GTTGGACTCA CCGGATAGT TACCGGATAA
3361 GCGCAGCGG TACAGCGTGA CAGGCTCGGA ACCTCTGACT TGAGCGTCCA TTAGCGTACC CCGGATAGT TACCGGATAA
3431 CTGAGATACC TACAGCGTGA CAGGCTCGGA ACCTCTGACT TGAGCGTCCA TTAGCGTACC CCGGATAGT TACCGGATAA
3501 CCGTAAGCGG CAGGCTCGGA ACCTCTGACT TGAGCGTCCA TTAGCGTACC CCGGATAGT TACCGGATAA
3571 TAGTCTGTC GGGTTTCGCC ACCTCTGACT TGAGCGTCCA TTAGCGTACC CCGGATAGT TACCGGATAA
3641 CTATGGAAAA ACGCCAGCAA CCGGCTGCTT GATTCGTTGG TTAGCGTACC CCGGATAGT TACCGGATAA
3711 TCTTCTCTGC GTTATCCCTT GATTCGTTGG TTAGCGTACC CCGGATAGT TACCGGATAGT TACCGGATAGT
3781 CCGCAGCGG ACGCCAGGAG CAGCGGAGT GCAGCGAGT AGTAGCGGAG TTAGCGTACC CCGGATAGT TACCGGATAGT
3851 CCTCTCCCC CCGCTTGGCC GATTCATTA TTAGCGTACC CCGGATAGT TACCGGATAGT TACCGGATAGT
3921 GTGAGCGCAA CCGCAATTAA GTGAGTTAGC TCACTCATTA GGCACCCAG GGCACCCAG GGCACCCAG GGCACCCAG
3991 GGCTCGTATG TTGTGTGGA TTGTGAGCGG ATAACAATTT CACACAGGAA CACACAGGAA CACACAGGAA CACACAGGAA

4061 GCCAAGCGCG CAATTAACCC TCACTAAGG GAACAAAAG TGGGTACCAG GCGCCCTCTC GAGGTCAATC
4131 ATATGCTTGA GAAGAGAGTC GGGATAGTCC AAAATAAAC AAAGGTAAAG AACCTTGGT TACCTGCTC
4201 ACATCAGTTA AAAGGTGGTA TAAATTAAT ATCGGTAAFA AAAGGTGGCC CAAAGTGAAA TTTACTCTTT
4271 TCTACTATTA TAAAAATTGA GGAATGTTTG TCGGTACTTT GATACGTGAT TTTTGTATGA ATTTGTTT
4341 AAGTTTATTC GCGATTGGA AATGCATATC TGTATTTGAG TCGGTTTTTA AGTTCTGTTG TTTTGTAAAT
4411 ACAGAGGGAT TTGTATAAGA AATATCTTTA AAAAACCCAT ATGCTAATTT GACNTAATTT TTGNGNAAAA

4481 TATATATTCA GCGCAATTCC AAAAAATTC AATATGAACA ATATAAGAT TAAATAGCT TGGCCCTCTC GCAGCGATGG
4551 GTATTTTTC TAGTAAATA TAAAGATAAC AAAGATAAAC AACATTTACA AACCAACCC CTAAAGTCTC
4621 AAAGCCCAAA GTGCTATGCA CGATCCATAG CAGCCGAGC CCAACCCAG CCAACCCAG CCAACCCAG
4691 GCAGCCAACT GGCATAATAG TCCACCCCTT CCGGTATGTA CCGGTATGTA CCGGTATGTA CCGGTATGTA
4761 GCAGCCAAAA AAAAAAAG AAGAAAAAG AAGAAAAAG AAGAAAAAG AAGAAAAAG AAGAAAAAG
4831 CCGNAAAAGC GAGGAGGATC GCGAGCAGCG ACGAGGCGCG GCGCTCTCTC GCGCTCTCTC GCGCTCTCTC

```

FIG._33D

```

4901 CCATCGCCAC TATATACATA CCCCCCCTC TCCTCCCATC CCCCACACC TACCACCACC ACCACCACCA
4971 CCTCCTCCCC CCTCGCTGCC GGACGACGAG CTCCTCCCCC CTCCTCCCTCC GCCGCCGCCG GTAACCAACCC
5041 CGCCCCCTCTC CTCCTTCTTT CTCCTGTTTT CTCCTGCTCT TTTTCTCGCT CGGTCTCGAT CTTTGGCCCT GGTAGTTTGG
5111 GTGGGCGAGA GCGGCTTCGT CGCCAGATC GGTGCGCGG GGTGCGCGG TCTCGCGCT GCGCTCTCCG
                               BamHI
                               ~~~~~
5181 GGCCTGAGTC GGCCCGGATC CTCGCGGGGA ATGGGCTCT CGGATGTAGA TCTTCTTCT TTCTTCTTTT
5251 TGTGGTAGAA TTTGAATCCC TCAGCATTGT TCATCGGTAG TTTTCTTTT CATGATTTGT GACAAATGCA
5321 GCCTCGTGCG GAGCTTTT GTAGC

```

FIG._33E

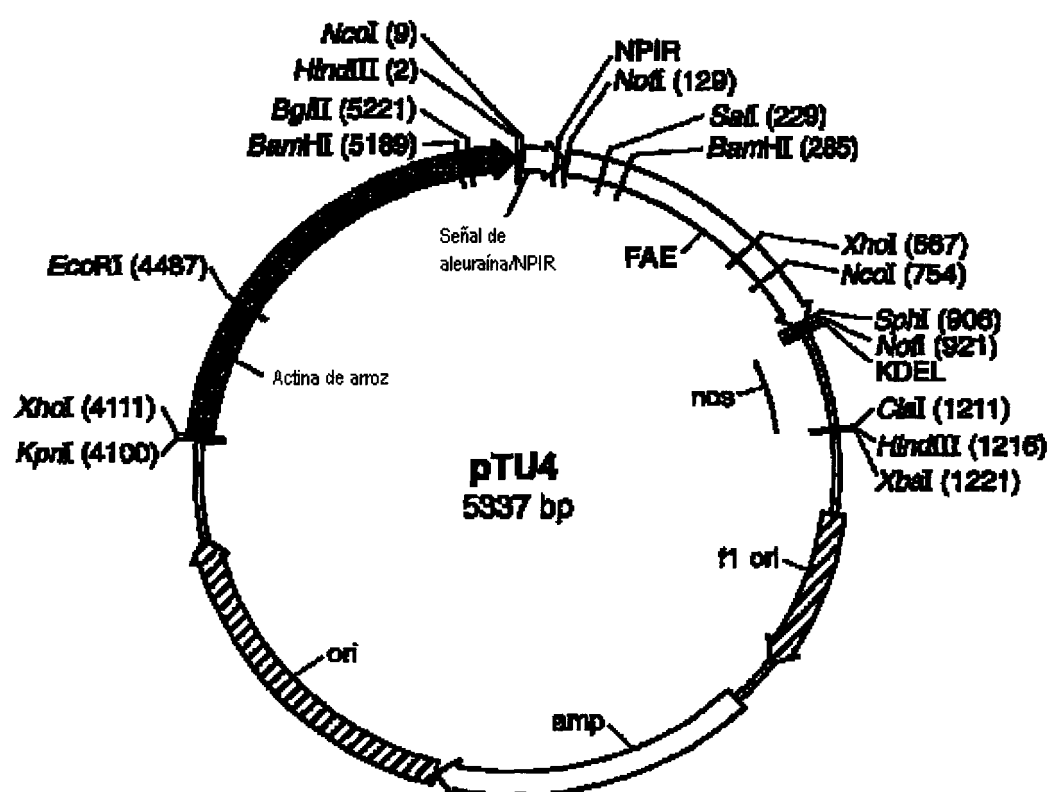


FIG. 34A

```

      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 CCGGTCTCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTGCGCG
      NotI
      ~~~~~
      . A S 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 GCCGTTCACC GACCGGCGCG 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 ACATTCCTC GACTATTATC AAGGGAGAGA AATTATACAA TTCTCAAACT GACATTAAAG
      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 AGCAGCAAG AATAATCAC CGTCTTCCGT GGCACGTGTA 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 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 GGACTACGG 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 CGACGGGATC CCAAACTGCG CCCCAGTGGG 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 GCGCCGAGAA CACATTTGTC TGCACGTGGG ATGAAATGCA 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 GCCCAGGGCG GACAGGGTGT GAATTAATCG CACACGACTT ATTTTGGGAT GACGAGCGCG GCATGCACCT
    NotI
      -----
      . P V A A A E P L K D E L *
911 GGCCGGTCCG GCGCGCGGAA CCAGTGAAGG ATGAGCTGTA AAGAAAGAGA TCGTTCAAAAC ATTGGGCAAT
981 AAAGTTTCTT AAGATTGAAT CCTGTGGCCG GTCTTCCGAT GATTAATCATA TAATTTCTGT TGAATTACGT
1051 TAAGCATGTA ATAATTAACTA TGTAAATGCAAT GACGTTATTT ATGAGATGGG TTTTATATGAT TAGAGTCCCG
1121 CAATTATACA TTTAATACGC GATAGAAAGC AAAATATAGC GCGCAAACTA GGTAAATTA TCGCGCGCGG
    HindIII
      -----
      ClaI      XbaI
      -----
1191 TGTCACTAT GTTACTAGAT CGATAGCTT CTAGAGCGCG CGGTGGAGCT CCAATTCGCC CTATAGTGAG
1261 TCGTATTACG CCGCTCACT GCGCTGCTT TTACAAACGC GTGACTGGGA AAACCTGGCG GTTACCCCAAC
1331 TTAATCGCCT TGCAGCACAT CCGCTTTGCG CCAGTGGCG TAATAGCGAA GAGGCCGCA CCGATCGCCC
1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCCTAGCGCG CCGTCCCTTC CGCGCGGGGT
1471 GTGGTGGTAA CGGCAAGCGT GACCGCTACA CTGCGCAAGC TCTAAATCGG GGGCTCCCTT TAGGGTTCCG
1541 CTTCCTTTCT CGCCAGGTTT GCGGCTTC CCGGCTTC AAACTTGT TAGGGTATG GTTACAGTAG TGGGCCATCG
1611 ATTTAGTGCT TTACGGCACC TCGACCCGAA AAACTTGT TAGGGTATG GTTACAGTAG TGGGCCATCG
1681 CCTGATAGA CGCTTTTTCG CCGTTTGGAG TTGGAGTCCA CGTTCTTAA CCGTTCCAAA TTGTTCCTAA
1751 CTGGAACAAC ACTCAACCCCT ATCTCGGCTT ATCTTTTGA TTTATAAGGG ATTTTGGCGA TTTTCGGCCTA
1821 TTGGTTAAAA AATGAGCTGA TTTAACAAA ATTTAACGCG AATTTTAACTA AAATATTAAC GTTACAAAT
1891 TAGGTGGCAC TTTTCGGGGA AATGTCGCG GAACTCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT
1961 GTATCCGCTC ATGAGACAAAT AACCTTGATA AATGCTTCAA TAATATTGAA AAGGAGAG TATGAGTATT
2031 CAACATTTCC GTGTCGCCCT TATTTCCCTT TTTGCGGCAAT TTTGCTTCTC TGTTTTTGT CACCCAGAAA
2101 CGCTGGTGAAG AGTAAAGAT GCTGAAGATC AATGTCGCG ACAGTGGTGC ACAGTGGTGC TGGATCTCAA
2171 CAGCGGTAAAG ATCTTTGAGA GTTTTCGCC CCAAGAACGT TTTCCAAATGA TGAACACTTT TAAAGTTCTG
2241 CTATGTGGCG CGGTATTATC CCGTATTGAC GCGGGGCAAG AGCAACTCGG TCGCCCGCAT CACTATTCTC
2311 AGAATGACTT GGTGAGTAC TCACCAATCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAAAT
2381 ATGCAGTGT GGCATAACCA TGAATGATA CACTGCGGCC AACTTACTTC TGACAAAGAT CGGAGGACCG

```

FIG.-34C

```

2451 AAGGAGCTAA CCGCTTTTTC GCACAACATG GGGGATCATG TAACTCGCCT TGAATCGTTGG GAACCGGAGC
2521 TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT ACCTGTAGCA ACTGCGCAA CGTTGCGCAA
2591 ACTATTAACT GCGCAACTAC TTACTCTAGC TTCCCGGCAA CAATTATAG CAATGATGGA ATGGCAACAA
2661 GTTGCAGGAC CACTCTGCG CACGCGCCTT CCGGCTGGCT GGTATTATGC GGTATTATGC GGAGCCCGTG
2731 AGCGTGGGTC TCGCGGTATC ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCGATATCG TAGTTATCTA
2801 CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAATAGA CAGATCGCTG AGATAGGTGC CTCACATGAT
2871 AAGCATTTGGT AACTGTGAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAAACCTT CATTTTAAAT
2941 TTAAGAAGAT CTAGGTGAAG ATCCTTTTTC ATAATCTCAT GACCAAAATC CCTTAACGTG AGTTTTCGTT
3011 CCATGAGCG CTAGACCCCG TAGAAAAGAT CAAAGGATCT TCTTGAGATC TTTTATTTCT GGCGGTAAATC
3081 TGCTGCTTGC GGTAACTGGC TTCAAGAACT ACCACCGTA CAGCGGTGG CAAAGGATCT TTTTATTTCT
3151 TTTTCCGAA GTTAACTGGC TTCAAGAACT ACCACCGTA CAGCGGTGG CAAAGGATCT TTTTATTTCT
3221 AGGCCACAC GGTAACTGGC TTCAAGAACT ACCACCGTA CAGCGGTGG CAAAGGATCT TTTTATTTCT
3291 GCTGCCAGTG GGTAACTGGC TTCAAGAACT ACCACCGTA CAGCGGTGG CAAAGGATCT TTTTATTTCT
3361 GGTCCGGCTG AACGGGGGT TCGTGACAC AGCCAGCTT GAGCGAAGC GAGCGAAGC ACCTACACCG
3431 CCTACAGCGT GAGCTATGAG AACGGGGGT TCGTGACAC AGCCAGCTT GAGCGAAGC GAGCGAAGC ACCTACACCG
3501 GGCAGGGTCG GACAGGAGA GACAGGAGA GACAGGAGA GACAGGAGA GACAGGAGA GACAGGAGA GACAGGAGA
3571 TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTGTG ATGCTCGTCA GGGGGCGGA GTTCTTTCTT
3641 AACCGCCAGC AACCGGGCT TTTTACGGT CTTGAGCGTC TGTGCGCTT TGTGCGCTT TGTGCGCTT
3711 GCGTTATCCC CTGATTTCTGT GGTATACCGT ATTACCGCT TTGAGTGAGC TGTGAGTGAGC TGTGAGTGAGC
3781 GACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAGCGGA AGGAGCGGA AGGAGCGGA AGGAGCGGA AGGAGCGGA
3851 CGCGGTTTGG CCGATTCAAT AATGCACTG GCACGACAGG TTTCCCGACT TTTCCCGACT TTTCCCGACT
3921 AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCCC TAGGCACCCC TAGGCACCCC TAGGCACCCC
3991 TGTGTGTGTG AATTGTGAGC GGATAACAAT TTCACACAGG AACAGCTAT GACCATGAT ACGCCAAGCG

                                KpnI
                                ~~~~~
4061 CGCAATTAA CCTCACTAAA GGGAAACAAA GCTGGGTACC GGGCCCCCCC TCGAGGTCAAT TCATATGCTT
4131 GAGAAAGAG TCGGGTAGT CCATAATAA ACAAGGTAA GATTACCTGG GATTACCTGG GATTACCTGG
4201 TAAAGGTGG TATAGTAA ATATCGGTAA TAAAGGTGG CCAAGGTGA AATTACTCT TTTCTACTAT
4271 TATAAATTT GAGGATGTT TGTGCGTACT TTGATACGTC ATTTTGTAT GAATTTGTTT TTAAGTTTAT
4341 TCGCGATTTG GAATGCTA TCTGTATTTG AGTCGGTTTT TAAGTTGTTA GCTTTTGTAA ATACAGAGGG
4411 ATTTGTATAA GAATATCTT TAAATAACCC ATATGCTAAT TTGACATAT TTTTGAGAAA AATATATATT

                                EcoRI
                                ~~~~~
4481 CAGGCGAATT CCACAATGAA CAATAATAG ATTAAATAG CTTGCCCCCG TTGCAGCGAT GGGTATTTTT
4551 TCTAGTAAA TAAAGTAA ACTTAGACT AAACATTTA CAAAACAC CCCTAAGTC CTTAAGCCCA

```

FIG.-34D

```

4621 AAGTGCTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCCACCCA GTGCAGCCAA
4691 CTGGCAATA GTCTCCACCC CCGGCACTAT CACCGTGAGT TGTCGACACC ACCGCACGTC TCGCAGCCAA
4761 AAAAAAAAAA AGAAGAAAA AAAAGAAAA GAAAACAGC AGTGGGTCC GGTCTGTGGG GGCCGGAAAA
4831 GCGAGGAGGA TCGCGAGCAG CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC
4901 ACTATATACA TACCCCCCCC TCTCCTCCCA TCCCCCAAC CTAACACCA CCACCAACAC CACCTCCTCC
4971 CCCCTCGCTG CCGGACGACG AGCTCCTCCC CCTCCCCCT CGCCGCCGC CGGTAACAC CCCGCCCTC
5041 TCCTCTTCT TTCTCCGTTT TTTTTCGT CTCGGTCTCG ATCTTGGCC TTGGTAGTTT GGGTGGGCGA
5111 GAGCGGCTTC GTCGCCCAGA TCGGTGCGCG GGAGGGGCGG GATCTCGCG CTGGCGTCTC CGGGCGTGAG
      BamHI                               BglII
      ~~~~~                               ~~~~~
5181 TCGGCCCCGA TCCTCGCGGG GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT TTTGTGGTAG
5251 AATTGAAATC CCTCAGCATT GTTCATCGGT AGTTTCTTCTT TTCATGATTT GTGACAAATG CAGCCTCGTG
5321 CGGAGCTTTT TTGTAGC

```

FIG._34E

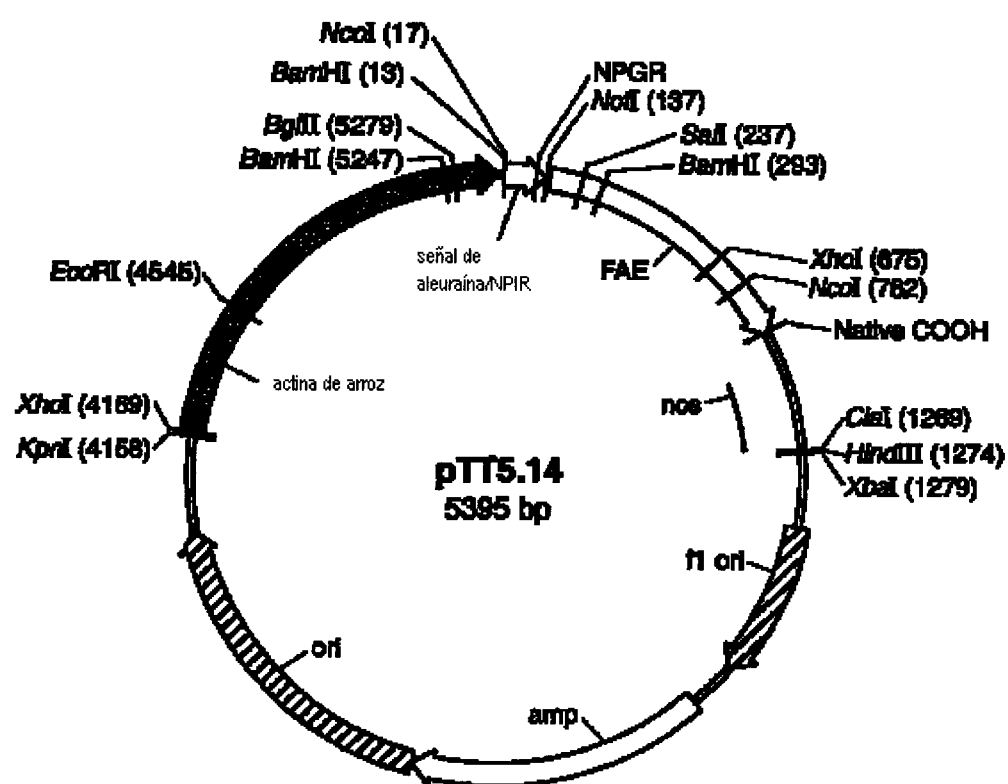


FIG..35A

```

NGOI
~~~~~
BamHI
~~~~~
      M A E A R V L L L A L A V L A T A A .
1 CCTGACGCCG AGGATCCATG GCCCAGCCC GCGTCCTCTT 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 GCCTCCTCCT CCTCCTTCGC CGACTCCAAC CCGGCCGGC 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 TAGTCGAAT 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 ATTCCGTGCA CTATTATCAA GGGAGAGAAA ATTACAAATT CTCNAACTGA
      BamHI
      ..
      . I N G W I L R D D S S K E I I T V P R G T G S
281 CATTAACGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA ATATCACCG 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 TACAACCTGA TACTAACTAC ACCCTCAGC CTTTCGACAC CCTACCAAC TGCNACGGTT
      . 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 GGGTCTCCGT CCAGGACCAA GTCGAGTCGC TTGTCAAAACA
      . 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 ACTACGGCGT GACCGTGACC GGCCACKCCC TCGGGCCCTC 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 CTCAC TGCCG CCGAGCTGTC TCGGACATAC GACAAATATC GCCTGTATAC CTTTGGCGAA 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 CTTGCGGTG TACATGAACG ATGCCTTCCA AGCCTCGAGC CCAGATACGA CGCAGTATTT
      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 AAACCTGCC CCAGTGGAGC AGGGGTAGC 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 TCCTTACAG GCCCAGAACA CATTTGTCTG CACTGGGAT GAGTGCAGT
 . C E A Q G G Q G V N A H T T Y F G M T S G A .
 841 GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATAATGCCA CACGACTTAT TTGGGATGA CGAGCGGAGC
 . C T W *
 911 CTGTACATGG TGATCAGTCA TTTCAGCCTC CCGAGTGTA CCAGGAAAGA TGGATGTCCT GGAGAGGGGG
 981 CCGCGTAACC ACTGAAGGAT GAGCTGTAA GAGCAGATC GTTCAAACAT TTGGCAATA AGTTTCTTAA
 1051 GATTGAATCC TGTTGCCGGT GTTGCATGA TTATCATATA ATTTCTGTTG AATTACGTTA AGCATGTAAT
 1121 AATTAAACAT TAATGCATGA CGTTATTAT GAGATGGGT TTTATGATTA GAGTCCGCA ATTATACATT
 1191 TAATACGCGA TGAANAACAA AATATAGCG CCAAACTAGG ATAAATTATC GCGCGCGGTG TCATCTATGT
 HindIII
 ~~~~~  
 ClaI ~~~~~ XbaI  
 ~~~~~  
 1261 TACTAGATCG ATAAGCTTCT AGAGCGGCGG GTGGAGCTCC AATTCGCCCT ATAGTAGTC GTATTACGCG
 1331 CGCTCACTGG CCGTCGTTT ACAACGTCGT GACTGGGAA ACCCTGGCGT TACCCAATT AATCGCCTTG
 1401 CAGCACATCC CCTTTTCGCC AGCTGGCGTA ATAGCGAAGA GGCCCGCAC GATCGCCCTT CCCAACAGTT
 1471 GCGCAGCCTG AATGGCGAAT GGGACGCGC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG
 1541 CGCAGCGTGA CCGCTTTCCC CGTCAAGCTC TAAATCGGG GCTCCCTTTA GGGTTCCGAT TTAGTGCCTT
 1611 CCACGTTCCG CGGCTTTCCC GAGTCCACG TTCTTTAATA GTGGACTCTT GTTCCAACT GGAACAACAC
 1681 ACGGCACCTC GACCCCAAA AACTTGATTA GGGTGTGGT TCACGTAGTG GGCCATCGCC CTGATAGACG
 1751 GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTTTAAACAA ATATTACGC TTACAAATTIA GGTGGCACTT
 1821 TCAACCTAT CTCGGTCTAT TCTTTTGATT TATAAGGGAT TTTGCCGAT TCGGCCATTT GGTAAAAA
 1891 TGAGCTGATT TAACAAAAAT TTAACGCGAA TTTTAAACAA ATATTACGC TTACAAATTIA GGTGGCACTT
 1961 TTCGGGGAAA TGTGCGCGGA ACCCTATTT GTTTATTTT CTAATAATAT TCAATATATG ATCCGCTCAT
 2031 GAGACAATA CCTGATAAA TGCTTCATA ATATTGAANA AGGAAGAGTA TGAGTATTC AATTTCCGT
 2101 GTCGCCCTTA TTCCCTTTT TCGGCGATTT TGCCCTTCCTG TTTTGTCTCA CCCAGAAAAG CTGGTGAAG
 2171 TAAAAGATGC TGAAGATCAG TTGGGTGCAC GAGTGGGTTA CATCGAACTG GATCTCAACA GCGGTAAGAT
 2241 CCTTGAGAGT TTTGCGCCCG AAGAAGCTT TCCAATGATG AGCACTTTTA AGTTCTGTCT ATGTGGCGCG
 2311 GTATTATCCC GTATTGACGC CGGGCAGAG CAACTCGGTC GCCGATACA CTATTTCTCAG AATGACTTGG
 2381 TTGAGTACTC ACCAGTCACA GAAAAGCATC TTACGGGATGG CATGACAGTA AGAGAAATTAT GCAGTGTGC
 2451 CATAACCATG AGTGATAACA CTGCGGCCAA CTTACTTCTG ACAACGATCG GAGGACCGAA GGAGCTAACC
 2521 GCTTTTTTGC ACAACATGGG GGATCATGTA ACTGCTCTTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA
 2591 TACCAAACGA CGAGCGTGAC ACCACGATGC CTGTAGCAAT GGCACAACG TTGCGCAAC TATTAACCTG
 2661 CGAACTACTT ACTCTAGCTT CCGGCAACA ATTAATAGAC TGGATGGAG CGGATAAAGT TGCAGGACCA

FIG. 35C

2731 CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG CGTGGGTCTC
 2801 GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA GTTATCTACA CGACGGGGAG
 2871 TCAGGCAACT ATGGATGAAC GAAATAGAAC GATCGCTGAG ATAGGTGCCT CACTGATTA GCAATTGGTAA
 2941 CTGTCAAGAC AAGTTTACTC AATATACTTT TAGATTGATT TAAACTTTCA TTTTAAATTT AAAGGATCT
 3011 AGGTGAAGAT CCTTTTGTAT AATCTCATGA CCAAAATCCC TTAACGTGAG TTTTTCGTTCC ACTGAGCGTC
 3081 AGACCCCGTA GAAAGATCA AAGGATCTTC TTGAGATCTC TTTTTCCTGC GCGTAAATCTG CTGCTTGCAA
 3151 ACAAAAAAC CACCGCTACC AGCGGTGGTT TGTGTGCGG ATCAAGAGCT ATCAACTCTT TTTCCGAAGG
 3221 TAACTGGCTT CAGCAGAGCG CAGATACCAA ATACTGTCTT TCTAGTGTAG CCGTAGTTAG GCCACCACTT
 3291 CAAGAATCTT GTAGCACCGC CTACATACCT CGCTCTGCTA ATCCTGTTAC CAGTGGCTGC TGCCAGTGGC
 3361 GATAAGTCGT GTCTTACC GGTTGGACTCA AGACGATAGT TACC GGATAA GCGGCAGCGG TCGGGCTGAA
 3431 CGGGGGGTTT GTGCACACAG CCCAGCTTGG AGCGAACGAC CTACACCGAA CTGAGATACC TACAGCGTGA
 3501 GCTATGAGAA AGCGCCACGC TTCCCGAAGG GAGAAAGCGG GACAGGTATC CCGTAAGCGG CAGGTCGGA
 3571 ACAGGAGAGC GCACGAGGA GCTTCCAGGG GGAACGCCCT GGTATCTTTA GGTATCTGTC GGGTTTCGCC
 3641 ACCTCTGACT TGAGCGTCTG TTTTGTGTAT GCTCGTCTAG GGGGCGGAGC CTATGGAAAA CGCCAGCAA
 3711 CGCGGCCCTT TTACGGTTCC TGGCCCTTTG CTGGCCCTTT GCTCACATGT TCTTCCCTGC GTTATCCCCT
 3781 GATCTGTGG ATAACCGTAT TACCGCCCTT GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCCGAGC
 3851 GCAGCGAGTC AGTGAGCGAG GAAGCGGAAG AGCGCCCAAT ACAGCAACCG CCTCTCCCCG CGCGTTGGCC
 3921 GATTCAATTA TGCAGCTGGC ACAGCAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAT
 3991 GTGAGTTAGC TCACTCAATTA GGCACCCCAAG GCTTTACACT TATGCTTCC GGTCTGTATG GGCTGTGGA
 4061 TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCAATGATTAC GCGAAGCGCG CAATTAACCC

 4131 TCACTAAAGG GAACAAAAGC TGGGTACCGG GCGCCCTCTC GAGGTCATTC ATATGCTTGA GAAGAGAGTC
 4201 GGGATAGTCC AAAATAAAC AAAGGTAAGA TTACCTGGTC AAAAGTGAAA ACATCAGTTA AAAGGTGGTA
 4271 TAAGTAAAT ATCGGTAATA AAAGTGGCC CAAAGTGAAA TTTACTCTTT TCTACTATTA TAAAAATTGA
 4341 GGATGTTTTG TCGGTACTTT GATACGTCAAT TTTTGTATGA ATTGGTTTTT AAGTTTATTC GCGATTGGA
 4411 AATGCATATC TGTATTGAG TCGGTTTTTA AGTTCGTTGC TTTTGTAAAT ACAGAGGGAT TTGTATAAGA

 4481 AATATCTTTA AAAAAACCAT ATGCTAATTT GACATAATTT TTGAGAAAAA TATATAATCA GCGAATTC
 4551 ACAATGAACA ATAATAAGAT TAAAAAGCT TGCCCTCGTT GCAGCGATGG GTATTTTTTC TAGTAAAAATA
 4621 AAAGATAAAC TTAGACTCAA AACATTTACA AAAACAACCC CTAAGTCTCT TAAAGCCAAA GTGCTATGCA
 4691 CGATCCATAG CAAGCCACGC CCAACCCCAAC CCACCCCACT GCAGCCCACT GCGCAATAGT
 4761 CTCCACCCCC GGCATATCA CCGTGAGTTG TCCGCACCAAC CGCACGCTCT GCGACCAAAA AAAAAAAG
 4831 AAAGAAAAAA AAAAAAGA AAAACAGCAG GTGGGTCCGG GTCGTGGGGG CCGGAAAAAGC GAGGAGGATC

KpnI
 XhoI
 EcoRI

FIG._35D

```

4901 GCGAGCAGCG ACGAGGCCCG GCCCTCCCTC CGCTTCCAAA GAAACGCCCC CCATCGCCAC TATATACATA
4971 CCCCCCCTC TCCTCCCATC CCCCACACC TACCACCACC ACCACCACCA CCTCCTCCCC CCTCGCTGCC
5041 GGACGACGAG CTCCTCCCCC CTCCTCCCTCC GCGCCGCCG GTAACCAACC CGCCCCCTCTC CTCCTTCTTT
5111 CTCCTGTTTT TTTTTCGTCT CCGTCTCGAT CTTTGGCCTT GGTAGTTGG GTGGCGGAGA GCGGCTTCGT
                                     BamHI
5181 CGCCCAGATC GGTGCGCGGG AGGGCGGGA TCTCGCGGT GCGTCTCCG GCGGTGAGTC GGCCCCGGATC
                                     BglII
                                     ~~~~~~
5251 CTCGCGGGGA ATGGGCTCT CGGATGTAGA TCTTCTTCT TCTTCTTTT TGTGGTAGAA TTTGAATCCC
5321 TCAGCATTGT TCATCGGTAG TTTTCTTTT CATGATTTGT GACAAATGCA GCCTCGTGCG GAGCTTTTTT
5391 GTAGC

```

FIG._35E

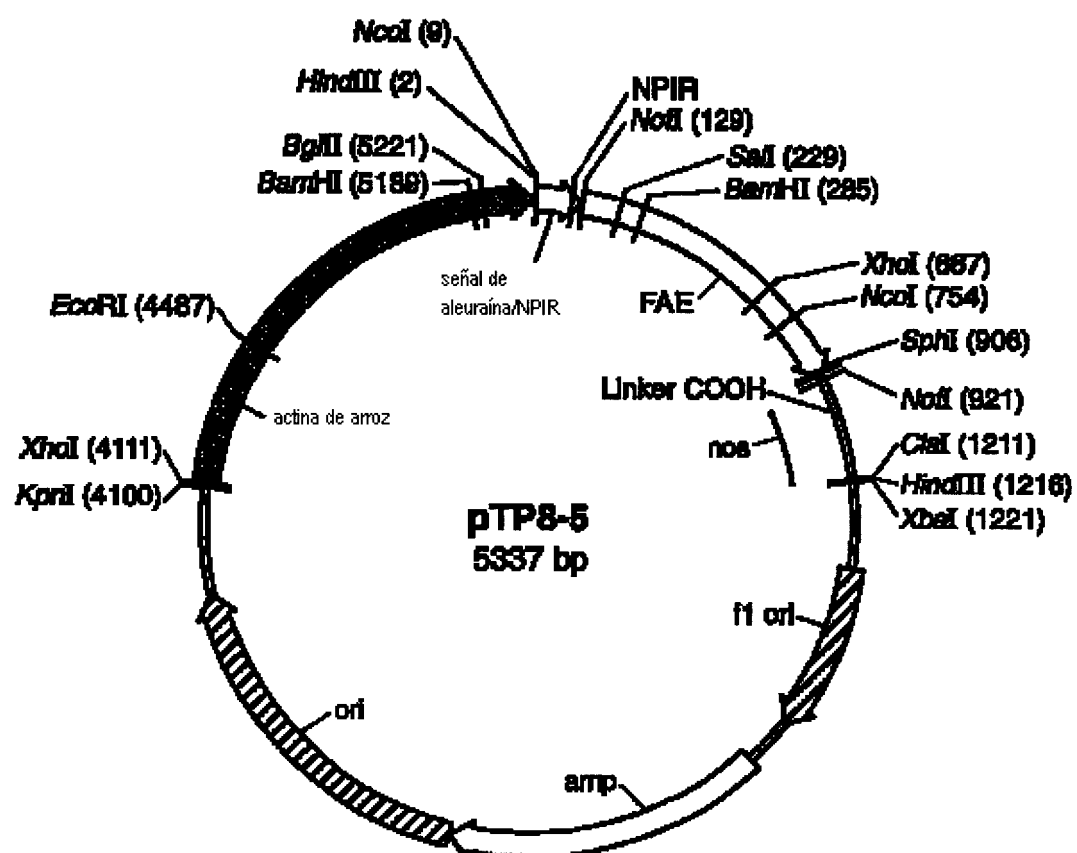


FIG._36A

```

                                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 TCCTCGTCT GGCACGGCC GCGTCGCCG
                                NotI
                                ~~~~~
      . A S S S F A D S N P I R P V T D R A A A S T .
71  TCGCTCCTC CTCCTCCTC GCCGACTCCA ACCGATCCG GCCGTCAAC 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 TCGGAGACC 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 TTCTCAACT GACATTACG
                                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 GATGATCCT CCGCGACGAC AGCAGCAAG AAATAATCAC CGTCTCCGT GGCCTGTA GTGATACGAA
      . L Q L D T N Y T L T P F D T L P Q C N G C E V
351 TCTACAATC GATACTACT ACACCTCAC GCCTTTCGAC ACCTACCAC AATGCACCG 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 CACGTGGAT ATTATATTG ATGGGCTCC GTCCAGACC AAGTCAGTC 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 CCGGCAKCK CTCGGCGCC TCCTGGCGC 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 CGCCGAGCTG TCTGCGACAT ACGACACAT CCGCTGTAT ACCTTCGGCG AACCGCGCAG CGGCATCAG
                                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 COTACATGAA CGATGCCTTC CAAGCTCGA GCCAGATAC GACGCAATAT TTCGCGTCA
                                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 CCCCCTGGGA 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 TGCACCTGGG ATGAAGTGCA GTGCTGTGAG
    Sphi
    ~~~~~
    A Q G G Q G V N N A H T T Y F G N T S G A C T W
841 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGC GCATGCACCT
    NotI
    ~~~~~
    . P V A A A *
911 GGCCGGTCCG GCGCGGTAA CCACTGAAGG ATGAGCTGTA AAGAAGCAGA TCGTTCAAC ATTTGGCAAT
981 AAAGTTTCTT AAGATTGAAT CCTGTGCGG GTCTTGCGAT GATTATCATA TAAATTCTGT TGAATTACGT
1051 TAAGCATGTA ATAATAACA TGTAAATGCA TACGTTATTT ATGAGATGGG TTTTATGAT TAGAGTCCCG
1121 CAATTATACA TTTAATACGC GATAGAAAAC AAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGG
    HindIII
    ~~~~~
    ClaI XbaI
    ~~~~~
1191 TGTCACTAT GTTACTAGAT CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTCGCC CTATAGTGAG
1261 TCGTATTACG CGCGCTCACT GCGCGTCTGT TTACAACGTC GTGACTGGGA AAACCTGGC GTTACCCCAAC
1331 TTAATCGCCT TGCAGCACAT CCCCCTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CGGATCGCCC
1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCTGTAGCG GCGCATTAAG CCGCGCGGGT
1471 GTGGTGGTTA CCGCAGCGT GACCGCTACA CTGCGCAGCG CCTAGCGCC CGCTCCTTTC GCTTCTTTC
1541 CTTCCCTTCT CGCCACGTTT CCGGCGTTTC CCCGTCAAGC TCTAAATCGG GGGCTCCCTT TAGGTTCCG
1611 ATTTAGTGCT TTACGGCACC TCGACCCCAA AAAACTTGAT TAGGGTGATG GTTCACGTAG TGGGCCATCG
1681 CCGTGATAGA CCGTTFTTTCG CCTTTGACG TTGGAGTCCA CGTCTTTAA TAGTGGACTC TTGTCCCAA
1751 CTGGAACAAC ACTCAACCTT ATCTGGTCT TTTTACAAA ATTTTACCGA ATTTTACCGA TTTTCGGCTA
1821 TTGGTTAAA AATGAGCTGA TTAAACAAA AATTAACCGG AATTTTAAAC AAATATTAAC GCTTACAAAT
1891 TAGGTGGCAC TTTTCGGGA AATGTGCGG GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT
1961 GTATCCGCTC ATGAGACAT AACCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT
2031 CAACATTTCC GTGTGCGCCT TATTCCTTTT TTGCGGCGAT TTGCTCTCC TGTTTTGTCT CACCCAGAAA
2101 CGCTGGTGAA AGTAAGAAT GCTGAAGATC AGTTGGGTGC ACAGTGGGT TACATCGAAC TGGATCTCAA
2171 CAGCGGTAAG ATCCTTGAGA GTTTTCGCC CGAAGAACGT TTTCCAATGA TAGACACTTT TAAAGTTCTG
2241 CTATGTGGCG CCGTATTATC CCGTATTGAC GCGGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC
2311 AGAATGACTT GGTGAGTAC TCACCAGTCA CAGAAAAACA TCTTACGGAT GGCATGACAG TAAGAGAAAT

```

FIG._36C

2381 ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG
 2451 AAGGAGCTAA CCGCTTTTTC GCACAACATG GGGGATCATG TAACCTCGCT TAATCTGTTG GAACCGGAGC
 2521 TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA AATGGCAACAA CGTTGCGCAA
 2591 ACTATTAACT GCGGAACCTAC CACTCTGCG ATTGCGATC CTGCGCCCTT CCGGCTGGCT GGTATTATTC TGAATAATCT
 2661 GTTGACGAGC CACTCTGCG ATTGCGATC CTGCGCCCTT CCGGCTGGCT GGTATTATTC TGAATAATCT TGAATAATCT
 2731 AGCGTGGGTC TCAGGATGATC AGTCAGGCA CTATGGATGA ATTGCGATC CTGCGCCCTT CCGGCTGGCT TGAATAATCT
 2801 CACGACGGGG AGTCAGGCA CTATGGATGA ATTGCGATC CTGCGCCCTT CCGGCTGGCT TGAATAATCT TGAATAATCT
 2871 AAGCATGGT AACTGTGAGA CCAAGTTTTC ATAATCTCAT GACCAAAATC CTTTAAACCTG AGTTTTCGTT
 2941 TTTAAAGGAT CTAGGTGAAG ATCCTTTTTC TAGAATAAGT CAAAGGATCT TCTTGAGATC CTTTAAACCTG AGTTTTCGTT
 3011 CCACTGAGCG TCAGACCCCG TAGAATAAGT CAAAGGATCT CAAAGGATCT TCTTGAGATC CTTTAAACCTG AGTTTTCGTT
 3081 TGCTGCTTGC AAACAATAA ACCACCGCTA CCAGCGGTGG TTTGTTTGGC GGATCAAGAG CTACCAACTC
 3151 TTTTTCGAA GGTAACTGGC TTCAAGAACT TTTCAGCAGG CCGAGATACC AAATACTGTC TTTGTTTGGC GGATCAAGAG
 3221 AGCCACACAC TTCAAGAACT TTTCAGCAGG CCGAGATACC AAATACTGTC TTTGTTTGGC GGATCAAGAG CTACCAACTC
 3291 GCTGCCAGTG GCGATAAGT AACGGGGGT TCGTGACAC ACCTTACACG AACTTACACG AACTTACACG AACTTACACG
 3361 GGTGCGGCTG AACGGGGGT TCGTGACAC ACCTTACACG AACTTACACG AACTTACACG AACTTACACG AACTTACACG
 3431 CCTACAGCGT GAGCTATGAG AAGCGGCCAC GCTTCCCGAA GGGAGAAAGG CGGACACGTA TCCGGTAAGC
 3501 GGCAGGGTCG GAACAGGAGA GCGCACGAGG GAGCTTCCAG GGGAGAAACG CTGGTATCTT TATAGTCTTG
 3571 TCGGGTTTCG CCACCTCTGA CTTCAGCGTC GATTTTGTG GATCTGTC AATCTGTC AATCTGTC AATCTGTC
 3641 AAACGCCAGC AACGGGGGT TTTTACGGTT CTTGCGCTTT TGGCTCACAT TTGCTCACAT TTGCTCACAT TTGCTCACAT
 3711 GCGTTATCCC CTGATTCTGT GGATAACCGT ATTACCGCTT TGGAGTGAGC TGATACCGCT CGCCGACGCG
 3781 GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCGCCA ATACGCAAC CGCTCTCCC
 3851 GCGCGGTTGG CCGATTCTAT AATGACGCTG GCACGACAGG TTTCCCGACT GGAAGCGGG CAGTGAGCGC
 3921 AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCCC AGGCTTTACA CTTTATGCTT CCGGCTCGTA
 3991 TGTGTGTGG AATTGTGAGC GGATAACAAT TTCACACAGG AAACAGCTAT GACCATGATT ACGCCAAAGC

 4061 CGCAATTAC CCTCACTAAA GGGAAACAAA GCTGGGTACC GGGCCCCCCC TCGAGGTGAT TCATATGCTT
 4131 GAGAGAGAG TCGGGATAGT CCATAATAA ACAAGGTAA GATTACCTGG GATTACCTGG TCAAAAGTGA AAACATCAGT
 4201 TAAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAGGTGG CCCAAAGTGA AATTACTCT TTTCTACTAT
 4271 TATAAAATTT GAGGATGTTT TGTGGGTACT TGTGGGTACT TGTGGGTACT TGTGGGTACT TGTGGGTACT TGTGGGTACT
 4341 TCGCGAATTT GAAATGCATA TCTGTATTGG AGTCGGTTT TAAGTTCTGTT GCTTTTGTAA ATACAGAGGG
 4411 ATTTGTATAA GAAATATCTT TAAAACACCC ATATGCTAAT TTGACATAAT TTTTGAGAAA AATATATATT

 ECORI
 4481 CAGCGCAATT CCACAATGAA CAATAATAAG ATTAAATAG CTTGCCCCCG TTGCAGCGAT GGGTATTTTT

FIG._36D

4551 TCTAGTAAA TAAAGATAA ACTTAGACTC AAAACATTTA CAAAACACAC CCCTAAAGTC CTAAGCCCCA
4621 AAGTGCTATG CACGATCCAT AGCAAGCCCCA GCCCAACCCA ACCCAACCCA ACCCAGCCCA GTGCAGCCAA
4691 CTGGCAATA GTCTCCACCC CCGGCACTAT CACCGTGAGT TGTCGGCACC ACCGCACGTC TCGCAGCCAA
4761 AAAAAAAAA AGAAGAAAA AAAAGAAAA CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC
4831 GCGAGGAGGA TCGCGAGCAG TACCCCTCCC TCTCTTCCA CCTACCAACA CCACCAACAC CACCTCCTCC
4901 ACTATATACA TACCCCTCCC AGCTCCTCCC CCTTCCCCCT CCGCCGCCGC CGGTAACAC CCGCCCCCTC
4971 CCCCTCGCTG CCGGACGACG TTCTCCGTTT TTTTTCGT TCTGGTCTCG ATCTTTGGCC TTGGTAGTTT GGGTGGGCGA
5041 TCCTCTTTCT TTCTCCGTTT TTTTTCGT TCTGGTCTCG ATCTTTGGCC TTGGTAGTTT GGGTGGGCGA
5111 GAGCGGCTTC GTCGCCAGA TCGGTGCGCG GGAGGGGCGG GATCTCGCG CTGGCGTCTC CGGGCGTQAG
BamHI BglII
~~~~~  
5181 TCGGCCCCGA TCCTCGCGGG GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT TTTGTGGTAG  
5251 AATTGAAATC CCTCAGCATT GTTCATCGGT AGTTTTTCTT TTCATGATTT GTGACAAATG CAGCCTCGTG  
5321 CGGAGCTTTT TTGTAGC

FIG.\_36E

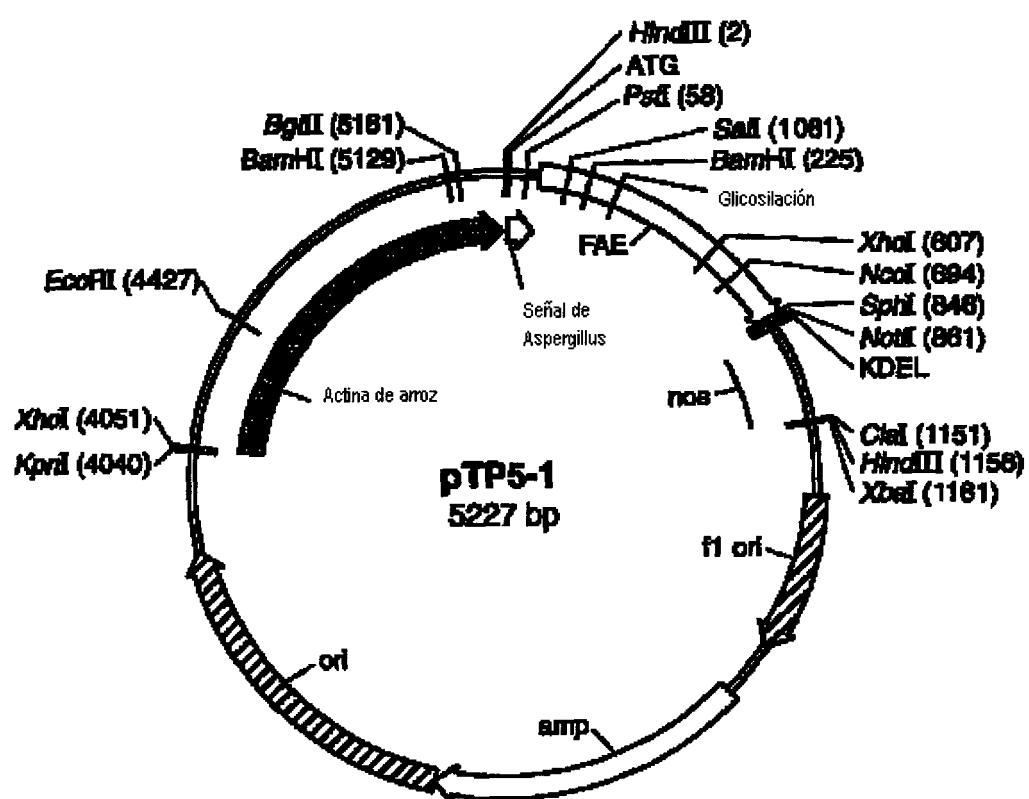


**FIG. 37A**

```

HindIII
~~~~~
1 AAGCTTAACA TGAAGCAGTT CTCCGCCAAA CACGTCTCTCG CAGTTGTGGT GACTGCAGGG CACGCCTTAG
 . A S T Q G I S E D L Y S R L V E M A T I S Q A .
71 CAGCCTCTAC GCAAGGCATC TCCGAAGACC TCTACAGCCG TTAGTCGAA ATGCCCACTA TCTCCCAAGC

 PstI
                                     ~~~~~
                                     M K Q F S A K H V L A V V V T A G H A L A
1  AAGCTTAACA TGAAGCAGTT CTCCGCCAAA CACGTCTCTCG CAGTTGTGGT GACTGCAGGG CACGCCTTAG
   . A S T Q G I S E D L Y S R L V E M A T I S Q A .
71  CAGCCTCTAC GCAAGGCATC TCCGAAGACC TCTACAGCCG TTAGTCGAA ATGCCCACTA TCTCCCAAGC

                                     SalI
                                     ~~~~~
 . A Y A D L C N I P S T I I K G E K I Y N S Q T
141 TGCCTACGCC GACCTGTGCA ACATTCCGTC GACTATATATC AAGGAGAGA AATTACAA TTCTCAAACT

 BamHI
                                     ~~~~~
   D I N G W I L R D D S S K E I I T V F R G T G S
211 GACATTAACG GATGGATCCT CCGCGACGAC AGCAGCAAG AATAATATC CGTCTCCGT GGCACGTGTA
   . D T N L Q L D T N Y T L T P F D T L P Q C N G .
281 GTGATACGAA TCTACAAC TC GATACTAAT ACACCTTCAC GCCTTTCGAC ACCCTACAC AATGCAACGG
   . C E V H G G Y Y I G W V S V Q D Q V E S L V K
351 TTGTGAAGTA CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAAA
   Q Q V S Q Y P D Y A L T V T G H X L G A S L A A
421 CAGCAGGTTA GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG
   . L T A A Q L S A T Y D N I R L Y T F G E P R S .
491 CACTCACTGC CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG

                                     XhoI
                                     ~~~~~
 . G N Q A F A S Y M N D A F Q A S S P D T T Q Y
561 CGGCAATCAG GCCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT

 NcoI
                                     ~~~~~
   F R V T H A N D G I P N L P P V E Q G Y A H G G
631 TTCCGGGTCA CTCATGCCAA CGACGGCATC CCAAACTGCG CCCCAGTGA GCAGGGGTAC GCCCATGGCG
   . V E Y W S V D P Y S A Q N T F V C T G D E V Q .
701 GTGTAGAGTA CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTTGTC TGCACCTGGG ATGAAGTCCA
   . C C E A Q G G Q G V N N A H T T Y F G M T S G
771 GTGCTGTGAG GCCCAGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGC

```

FIG.\_37B

```

SphI                                     NotI
-----
A C T W P V A A A E P L K D E L *
841 GCATGCACT GGCGGTGCG GGCGCGGAA CCACTGAAG ATGAGCTGA AGAAGCAGA TCGTTCAAC
911 ATTGGCAAT AAGTTTCTT AAGATTGAAT CCTGTGCGG GTCTTCCGAT GATATCATA TAATTTCTGT
981 TGAATTACGT TANGCATGA ATAAATTAACA TGTATGCAAT GACGTTATT ATGAGATGGG TTTTATGAT
1051 TAGAGTCCG CAATTATACA TTTAATACGC GATAGAAGC AATAATAGC GCGCAACTA GGATAAATTA
HindIII
-----
ClaI                                     XbaI
-----
1121 TCGCGCGCG TGTATCTAT GTTACTAGAT CGATAAGCTT CTAGAGCGG CGGTGAGCT CCAATTCGCC
1191 CTATAGTGA TCGTATACG CCGCTCACT GCGCGTCTT TTACAGCTC GTGACTGGG AAACCTGGC
1261 GTTACCCAA TTAATCGCT TCGAGCACAT CCCCCTTTG CAGCTGCG TAATAGCGA GAGGCCCGA
1331 CCGATCGCC TTCCCAACG TTGCGACGC TGAATGGCA ATGGCAGCG CCTGTAGCG CGCATTTAG
1401 CGCGGCGGT GTGTGTGTA CGCGCAGCT GACCGCTCA CCGGCTTTC CCTGTAGCG CGCTCTTTC
1471 GCTTCTTTC ATTTAGTCT TTACGGCAC TCGACCCCA AAGCTTGA TAGGTGATG GTTCACGTAG
1541 TAGGGTTCC ATTTAGTCT TTACGGCAC TCGACCCCA AAGCTTGA TAGGTGATG GTTCACGTAG
1611 TGGGCCATC CCTGTATGA CGGTTTTCG CCTTTGACG TTGAGTCCA CGTCTTTAA TAGTGGACTC
1681 TTGTTCCAA CTGGAACAC ACTCAACCT ATCTGGTCT ATCTTTGA TTATATAGG ATTTGCGGA
1751 TTTCGGCCTA TTGGTTAAA ATGAGCTGA TTTACAAA ATTTACCGG AATTTAACA AATATTAAAC
1821 GCTTACAAAT TAGTGGCAC TTTTCGGGA AATGTGCGG GAACCTCAT TTGTTTATT TTCTAAATAC
1891 ATTCAATAT GTATCGCTC ATGAGACAT AACCTGATA AATGCTCAA TAATATTGA AAGGAAGAG
1961 TATGAGTAT CAACATTTC GTGTGCGCT TATTCCTTT TTGCGCAT TTGCTTTC TGTTTTGT
2031 CACCCAGAA CGCTGTGAA AGTAAAGAT GCTGAAGAT AGTTGGTGC ACGATGGGT TACATCGAAC
2101 TGAATCTCA CAGCGGTAG ATCTTGAGA GTTTTCGCC CGAGGACGT TTTCCAATGA TGAGCACCTT
2171 TAAAGTTCT CTATGTGCG AGATGACTT GTTTGAGTAC TCACGAGTCA CAGAAAGCA TCTTACGGAT GGCATGACAG
2241 CACTATTCT TAAGAGATT ATGAGTCTT GGCATAACA TGAATGATA CACTCGGCC ACTTACTTC TGACACGAT
2311 CAGAGGACCG AAGGAGCTA CCGCTTTT TTGAGTATG GGGATCATG TAACTCGCT TGATCGTTG
2381 GAACCGGAGC TGAATGAGC CATACCAAT CACGAGCTG ACACCTCAT GCTGTAGCA ATGGCAACA
2451 GGTGCGGCA ACTATTACT GCGGAACAT TACTCTAGC TTCCCGCAA CAATTAATAG ATGGATGGA
2521 GCGGATATA GTTGCAGAC CACTTCTGCG CTCGGCCCT CCGCTGCTT GGTATTTC TGAATATCT
2591 GAGCCCGGT AGCGTGGGT TCGCGGTATC ATTCAGCAC TGGGCGCAGA TGGTAAGCCC TCCCGTATCG
2661

```

FIG..37C

```

2731 TAGTTATCTA CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC
2801 CTCACGTGAT AAGCATTTGGT AACTGTTCAGA CCAAGTTTAC CCATATATATAC TTATAGATTGA TTTAAAACCTT
2871 CATTTTAAAT TTAAAAGGAT CTAGGTGAAG ATCTTTTGTG ATAACTCTCAT GACTCAAAATC CTTTAAACGTG
2941 AGTTTTCGTT CCACGTAGCG TCAGACCCCG TCAGAACAGT TAGAAAGATC CAAGGATCT TCTTGAGATC CTTTTTTTCT
3011 GCGCGTAATC TGCTGCTTGC AAACAANAAG GGTAACTGGC TTCAAGAACT CTGTAGCAC CCCTACATAC GTTACCGGAT
3081 CTACCAACTC TTTTTCGAA TTTTTCGAA GGTAACTGGC TTCAAGAACT CTGTAGCAC CCCTACATAC GTTACCGGAT
3151 AGCGTAGTT AGGCCACAC GCTGCCAGTG GGTAGGCTG AACGGGGGT GGTGTCTTACC GGTGTGGACT CAAGACGATA GTTACCGGAT
3221 ACCAGTGGCT GCTGCCAGTG GGTAGGCTG AACGGGGGT GGTGTCTTACC GGTGTGGACT CAAGACGATA GTTACCGGAT
3291 AAGGCGCAGC GGTGCGGCTG CTTACAGCGT GAGCTATGAG AAAGCGCCAC GGTGTCTTACC GGTGTGGACT CAAGACGATA GTTACCGGAT
3361 AACTGAGATA CCTACAGCGT GAGCTATGAG AAAGCGCCAC GGTGTCTTACC GGTGTGGACT CAAGACGATA GTTACCGGAT
3431 TCCGTAAGC GGCAGGCTG GAGCTATGAG AAAGCGCCAC GGTGTCTTACC GGTGTGGACT CAAGACGATA GTTACCGGAT
3501 TATAGTCTTG TCGGCTTTCG CCACCTCTGA CTTGAGCGTC GATTTTGTG ATGCTCGTCA GGGGGGCGGA CTGGTATCTT
3571 GCCTATGGAA AAACGCCAGC AACGCGGCTT TTTTACGGTT CTTGAGCGTC GATTTTGTG ATGCTCGTCA GGGGGGCGGA CTGGTATCTT
3641 GTTCTTTTCT GCTTATCTCC CTGATTTCTG GATTAACCGT ATTACCGCTT TTGAGTGAAG TGTATACCGCT
3711 CCGCGCAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAAC
3781 CGCTCTTCCC CGCGCGTTGG CCGATTCTAT AATGACGCTG GCACGACAGG TTTCCCGACT TTTCCCGACT GGAAGCGGG
3851 CAGTGAGCGC AACGCAATTA ATGTGAGTGA GCTCACTCAT TAGGCACAGG TAGGCACAGG AATCAGCTAT GACCATGATT
3921 CCGGCTCGTA TGTGTGTGTG AATTGTGAGC GATTAACAAAT TTCACACAGG TTCACACAGG AATCAGCTAT GACCATGATT

                                     KpnI
                                     -----
3991 ACGCCAAAGC CGCAATTAA CCTCACTAAA GGGAAACAAA GCTGGGTACC GGGCCCCCCC TCGAGGTCAAT
4061 TCATATGCTT GAGAAGAGAG TCGGGATAGT CCATAATAAA ACAAGGTAA GATTACCTGG TCAAAAAGTGA
4131 AAACATCAGT TAAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAAGGTGG CCAAAAGTGA AATTACTCTT
4201 TTTCTACTAT TATAAAAATT GAGGATGFTT TGTGCGTACT TGTATACGTC ATTTTGTAT GAAATGGGTT
4271 TTAAGTTTAT TCGCGATTG GAAATGCATA TCTGTATTTG AGTCGGTTTT TAAATTCGTT TAAGTTCGTT GCTTTGTATA
4341 ATACAGAGGG ATTTGTATTA GAAATATCTT TAAAAAACCC ATATGCTAAT TTGACATAAT TTTTGAGAAA

                                     EcoRI
                                     -----
4411 AATATATATT CAGGCGAATT CCACAATGAA CAATAATAAG ATTAATAATAG CTTGCCCCCG TTGCAGCGAT
4481 GGGTATTTTT TCTAGTAAAA TAAAAGATAA ACTTAGACTC AAAACATTTA CAAAACAAAC CCTTAAAGTC
4551 CTAAGCCCCA AAGTGTATG CACGATCCAT AGCAAGCCCA GCCCAACCA ACCCAACCA ACCCAACCA
4621 GTGCAGCCAA CTGGCAATA GTCTCCACC CCGGCACTAT CACCGTAGT TGTCCGAC ACCGACGTC
4691 TCGCAGCCAA AAAAAAAGAA AGAAGAAAA AAAGAAAAA GAAAAACAGC AGGTGGGTCC GGTTCGTGGG

```

FIG.\_37D

```

4761  GGCCGGGAAA  GCGAGGAGGA  TCGCGAGCAG  CGACGAGGCC  CGGCCCTCCC  TCCGCTTCCA  AAGAAACGCC
4831  CCCCATCGCC  ACTATATACA  TACCCCTCCC  TCTCTCTCCA  TCCCCCAAC  CCTACCACCA  CCACCACCAC
4901  CACCTCCTCC  CCCCTCGCTG  CCGGACGACG  AGCTCCTCCC  CCTCCCCCT  CCGCCGCCGC  CGGTAACCCAC
4971  CCGCCCCCTC  TCCCTCTTCT  TTCTCCGTTT  TTTTCTTCGT  CTCGGTCTCG  ATCTTTGGCC  TTGGTAGTTT
5041  GGGTGGGCGA  GAGCGGCTTC  GTCGCCCAGA  TCGGTGCGCG  GGAGGGGCGG  GATCTCGCGG  CTGGCGTCTC

                               BamHI
                               ~~~~~
5111 CGGGCGTGAG TCGGCCCGGA TCCTCGCGGG GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT
5181 TTTGTGGTAG AATTGAATC CCTCAGCATT GTTCATCGGT AGTTTCTCTT TTCATGATTT GTGACAAATG
5251 CAGCCTCGTG CGGAGCTTTT TTGTAGC

```

**FIG.\_37E**

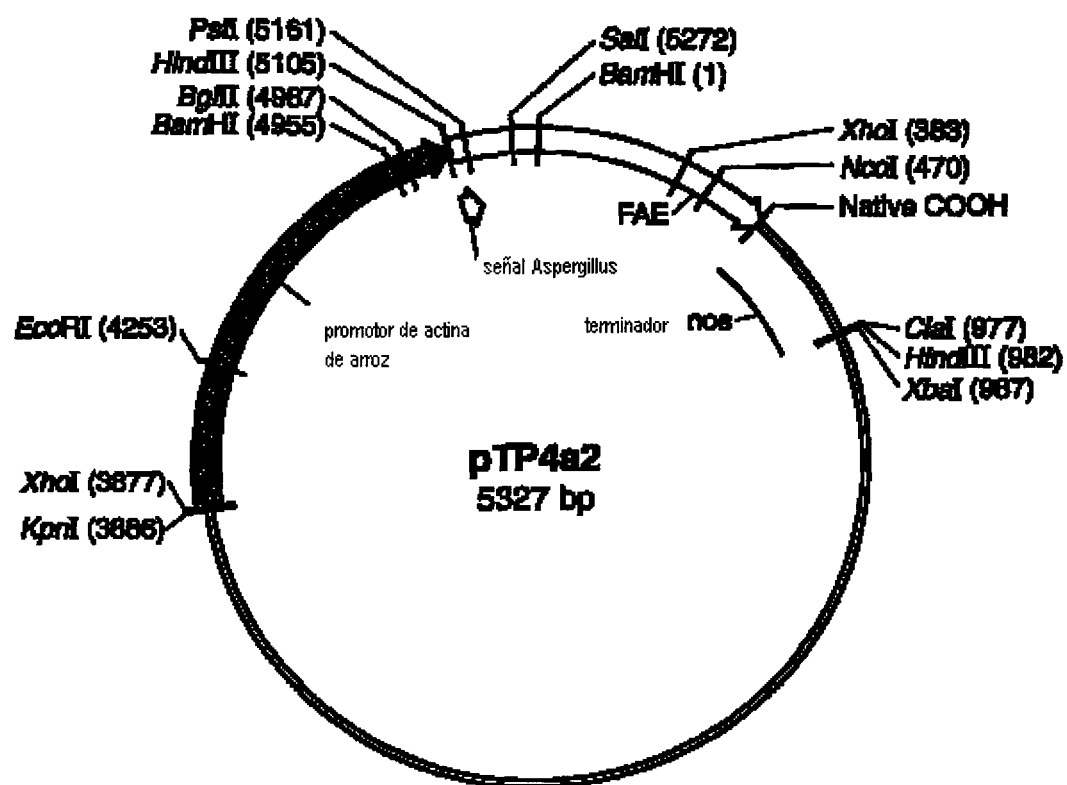


**FIG. 38A**

```

 BamHI

1 . I L R D D S S K E I I T V F R G T G S D T N L
 GATCCTCCGC GACGACAGCA GCAAAGAAAT AATCACCCTC TTCGTGGCA CTGCTAGTGA TACGAATCTA
 Q L D T N Y T L T P F D T L P Q C N G C E V H G
71 CAACTCGATA CTAATACAC CCTCAGCCT TTCGACACC TACCACAATG CAACGGTTGT GAAGTACACG
 . G Y Y I G W V S V Q D Q V E S L V K Q Q V S Q .
141 GTGGATATTA TATGGATGG GTCTCCGTCC AGGACCAAGT CGAGTCGCTT GTCAAACAGC AGGTTAGCCA
 . Y P D Y A L T V T G H X L G A S L A A L T A A
211 GTATCCGGAC TACGCGCTGA CCGTGACGG CCACKCCCTC GCGGCTCCC TGCGGCACT CACTGCCGCC
 Q L S A T Y D N I R L Y T F G E P R S G N Q A F
281 CAGCTGTCTG CGACATACGA CAACATCCGC CTGTACACCT TCGGCGAAC CCGCAGCGGC AATCAGGCCT
 XhoI

 . A S Y M N D A F Q A S S P D T T Q Y F R V T H .
351 TCGCGTCTGA CATGAACGAT GCCTTCCAG CCTCGAGCCC AGATACGACG CAGTATTTC GGTCACTCA
 NcoI

 . A N D G I P N L P P V E Q G Y A H G G V E Y W
421 TGCCAACGAC GGCATCCCA ACCTGCCCC GGTGAGCAG GGTACGCC ATGGCGTGT AGAGTACTGG
 S V D P Y S A Q N T F V C T G D E V Q C C E A Q
491 AGGTTGATC CTTACAGCG CCAGAACACA TTGTCTGCA CTGGGATGA AGTCAGTGC TGTGAGGCC
 . G G Q G V N N A H T T Y F G M T S G A C T W * .
561 AGGGCGGACA GGGTGTGAAT AATGGGCACA CCACTTATT TGGGATGAC AGCGAGCCT GTACATGGTG
 . *
631 ATCAGTCATT TCAGCCTCCC CGAGTGTACC AGGAAGATG GATGCTCTGG AGAGGGGCC GCGTAACCCAC
701 TGAAGGATGA GCTGTAAAGA AGCAGATCGT TCAACATTT GGCATAAAG TTTCCTAAGA TTGAATCCTG
771 TTGCCGGTCT TGCAGATGAT ATCATATAAT TTCTGTGAA TTACGTTAAG CATGTATAA TTACATGTA
841 ATGCATGACG TTATTATGA GATGGGTTT TATGATTAGA GTCCCGCAAT TATACATTTA ATACCGGATA
 ClaI

911 GAAACAAAA TATAGCGCGC AAAC TAGGAT AAATTATCGC GCAGGCTGTC ATCTATGTTA CTAGATCGAT
 HindIII

```

FIG.-38B

981 AAGCTTCTAG AGCGGCGGT GGAGCTCCAA TTGCGCCCTAT AGTGAGTCGT ATTACGGCG CTCACATGGCC  
 1051 GTCTGTTTAC AACGTCTGTA CTGGGAAAC CTGGCGTTA CCCAACTTAA TCGCCCTTGA GCACATCCCC  
 1121 CTTTCGCCAG CTGGCGTAAT AGCGAAGAG CCAGCAACGA TCGCCCTTCC CAACAGTTGC GCAGCCTGAA  
 1191 TGGCAATGG GACGCGCCT GTAGCGGCG ATTAAGCGG GCGGCTGCG GCGGCTGCG CAGCGTGACC  
 1261 GCTACACTTG CCTACGCCCT AGCGCGCCCT AGCGCGCCCT TCTTCTGCTT TCTTCTGCTT CAGCTGCGCG  
 1331 GCTTTCGCCG TCAGCTCTA AATCGGGGG GTCGCGATT GTTCGGAATT AGTGCCTTAC GGCACCTTGA  
 1401 CCCAAAAA CTTGATTAG AGTCCACGTT CTTTAAATAG GGTATGTTG CCATCGCCTT ATTCGCGCTT  
 1471 TTGACGTTGG TTTTGATTTA TAAAGGAATT TGCCGATTTG TCCAAACTGG AACAACTATC TTTTCGCTT  
 1541 CGGCTATTTT ACAGCAATTT AACGCAATTT TTAACAAAT ATTAACGCTT ACATTTTAG GGCCTATTTG  
 1611 TCGCGGGAAC CCTATTTGT TTAATTTTCT AATACATTC AATATGTAT CCGCTCATGA GACAATAACC  
 1751 CTGATAAATG CTTCAATAAT ATTGAAAAAG GAAGAGTATG AGTATTCAC ATTTCCGCTT CGCCCTTAT  
 1821 CCTTTTCTTG CGCATTTTG CGCATTTTG CTTTCTGCTT TTTGCTCAC TTTGCTCAC CAGAACGCTT  
 1891 AAGATCAGTT GGTGCAAGA GTGGTTTACA GTGGTTTACA TCGAATGGA TCTCAACAGC GGTAGAGTTT  
 1961 TCGCCCGGAA GACGTTTTC CAATGATGAG CACTTTTAAA GTTCTGCTAT GTGCGCGCTT ATTTCCCTT  
 2031 ATTGACGCCG GGCAGAGCA ACTCGGTGCG CGCATACACT ATTTCTCAGA TGACTTGTGT GAGTACTCAC  
 2101 CAGTCACAGA AAGCATCTT ACGGATGCGA TGACAGTAA AGAATATAG GAGCCGAGG AGCTAACCG  
 2171 TGATACACT GCGGCCACT TACTTCTGAC AACGATCGGA GAGCCGAGG GGTAGAGTTT TTTTTCGAC  
 2241 AACATGGGG ATCATGTAAC TCGCTTGTAT GTAGCAATGG CAACAACGTT CGGCAACTA TTAAGCTGCG  
 2311 AGCGTGACAC CACGATGCTT GTAGCAATGG CAACAACGTT CGGCAACTA TTAAGCTGCG AACTACTTAC  
 2381 TCTAGCTTCC CGGCACCAAT TAATAGACTG GATGAGGCG GATTAAGTTG CAGGACCACT TCTGCGCTCG  
 2451 GCCCTTCCGG CTGGCTGGTT TATGCTGAT AATCTGGAG CCGTGAGCG TATCTACAC AGCGGAGTC  
 2521 CAGCACTGGG GCCAGATGGT AAGCCCTCCC GTATCTGAT AGTGCCTCA CTGATTAAGC ATTGCTAAT  
 2591 GGATGAACGA AATAGACAGA TCGCTGAGT AGTGCCTCA CTGATTAAGC ATTGCTAAT GTGAGAGTCC  
 2661 GTTTACTCAT ATATACTTTA GATTGATTTA AACTTTCAT TTTAATTTAA AAGGATCTAG GTGAGAGTCC  
 2731 TTTTGTATA TCTCATGACC AATATCCCTT AACGTGAGT TTCGTTCCAC TGAGCGTCTAG ACCCGTAGA  
 2801 AAGATCAAA GGATCTTCTT GAGATCCCTT TTTTCTGCG CAAGAGCTAC CAACTCTTTT TCCGAAGTA  
 2871 CCGCTACCAG CCGTGGTTTG GATACCAAT ACTGTCTTC TAGGTAGGC CAGCACTTCA AGAATCTGT  
 2941 GCAGAGCGCA GATACCAAT ACTGTCTTC TAGGTAGGC CAGCACTTCA AGAATCTGT CAGTGGCTTCA  
 3011 AGCACCGCCT ACATACCTCG CTCTGCTAAT CCTGTTACCA GTGGTGTCT GAGTGGCGA TAAATCTGT  
 3081 CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATACAG CGCAGCGGT CGCAGCGGT GGGGTTCTGT  
 3151 GCACACAGCC CAGCTTGGAG CGAACGACCT ACACCGAAT GAGATACCTA CAGCGTGAGC TATGAGAAAG  
 3221 CGCCACGCTT CCGAAGGGA GAAGGCGGA CAGGTATCCG GTAAAGCGCA GGTGCGGAC AGGAGGCGC  
 3291 ACGAGGAGC TTCCAGGGGG AAACGCTGG TATCTTTATA GTCTGTCTG GTTTCGCCAC CTCGACTTG  
 3361 AGCGTCGATT TTTGTGATGC TCGTCAGGG GCGGAGCCT ATGGAATAAC GCCAGCAACG CGGCCCTTTT

**FIG. 38C**

```

3431 ACGGTTCCCTG GCCTTTGTG GGCCTTTTGC TCACATGTTT TATCCTCGCGT TATCCTCCCTGA TTCTGTGGAT
3501 AACCGTATTA CCGCCTTTGA GTGAGCTGAT ACCGCTCGCC GACGCCGAAC GACCGCGCGC AGCGAGTCAG
3571 TGAGCGAGGA AGCGGAAGAG CGCCCAATAC GCAACCCGCC TCTCCTCCCGC GGTGGCCGGA TTCATTAAATG
3641 CAGCTGGCAC GACAGGTTTC CCGACTGGAA AGCGGGCAGT GAGCGCAACG CAATTAATGT GAGTTAGCTC
3711 ACTCATTAGG CACCCAGGC TTACACATTT ATGCTTCCGG CTCGTATGTT GTGTGGAAAT GTGAGCGGAT
3781 AACCAATTCA CACAGGAAC AGCTATGACC ATGATTACGC CAAGCGCGCA ATTAAACCTC ACTAAAGGGA

 KpnI
                                   ~~~~~
3851 ACAAAGCTG GGTACCGGC CCCCCCTCGA GGTCAATTCAT ATGCTTGAGA AGAGAGTCGG GATAGTCCAA
3921 AATAAACAAGGTAGATT ACCTGGTCAA AAGTGAACAAC ATCAGTTAAA AGGTGGTATA AGTAAATAT
3991 CGGTAATAAA AGGTGGCCCA AAGTGAATTT TACTCTTTTC TACTATTATA AAAATTGAGG ATGTTTTGTC
4061 GGTACTTTGA TACGTCAATTT TTGTATGAAT TGGTTTTTAA GTTTATTCGC GATTTGGAAA TGCAATATCTG
4131 TATTGAGTC GGTTTTTAAG TTCGTTGCTT TTGTAATAC AGAGGGATTT GTATAAGAAA TATCTTTAAA

                                   EcoRI
                                   ~~~~~
4201 AAACCCATAT GCTAATTGTA CATAATTTTT GAGAAAAATA TATATTCAGG CGAATTCAC AATGAACAAT
4271 AATAAGATTA AATAGCTTG CCCCCTGTGC AGCGATGGGT ATTTTTTCTA GTAAAAATA AGTAAACTT
4341 AGACTCAAAA CATTTACAAA AACCAACCCCT AAAGTCTTAA AGCCCAAGT GCTATGCAG ATCCATAGCA
4411 AGCCAGCCC AACCCAACCC AACCCAACCC ACCCAGTGC AGCCAACTGG GCAATAGTCT CCACCCCGG
4481 CACTATCACG GTGAGTTGTC CGCACCAACC CACGTCTCG CAGCCAAAAA AAAAAAGAA AGAAAAAAA
4551 GAAAAAGAAA AACAGCAGGT GGGTCCGGGT CGTGGGGGCC GGAAGAAGCA GGAGGATCG GAGCAGCGAC
4621 GAGGCCGGC CCTCCTCCG CTTCACAAA CACGCCCCC ATCGCCACTA TATACATACC CCCCCCTCTC
4691 CTCCCATCCC CCAACCCCTA CCAACCCAC CACCAACACC TCCCTCCCCC TCGCTGCCGG ACGACGAGCT
4761 CCTCCCCCT CCCCCTCCG CGCGCCGGT AACCAACCCG CCCCCTCTCT CTTTCTTTCT CCGTTTTTTT
4831 TTTCTGTCG GTCTCGATCT TTGGCCTTGG TAGTTTGGGT GGGCGAGAGC GGCTTCGTCT CCCAGATCGG

 BamHI
                                   ~~~~~
4901 TCGCGGGGAG GGGCGGGATC TCGCGGCTGG CGTCTCCGG CGTGAGTCGG CCCGGATCCT CGCGGGGAAT

                                   BglII
                                   ~~~~~
4971 GGGGCTCTCG GATGTAGATC TTCTTTCTTT CTTCCTTTTG TGGTAGAAT TGAATCCCTC AGCATTTGTC

 HindIII
                                   ~~~~~
5041 ATCGGTAGTT TTTCTTTTCA TGATTGTGA CAATGCAGC CTCGTGCGGA GCTTTTTTGT AGCAAGCTTA

```

FIG.\_38D

M K Q F S A K H V L A V V V T A G H A L A A S .
5111 ACATGAAGCA GTTCTCCGCC AAACACGTCC TCGCAGTTGT GGTGACTGCA GGGCACGCCT 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 AGCTGCCTAC
SalI
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
PstI

FIG.\_38E

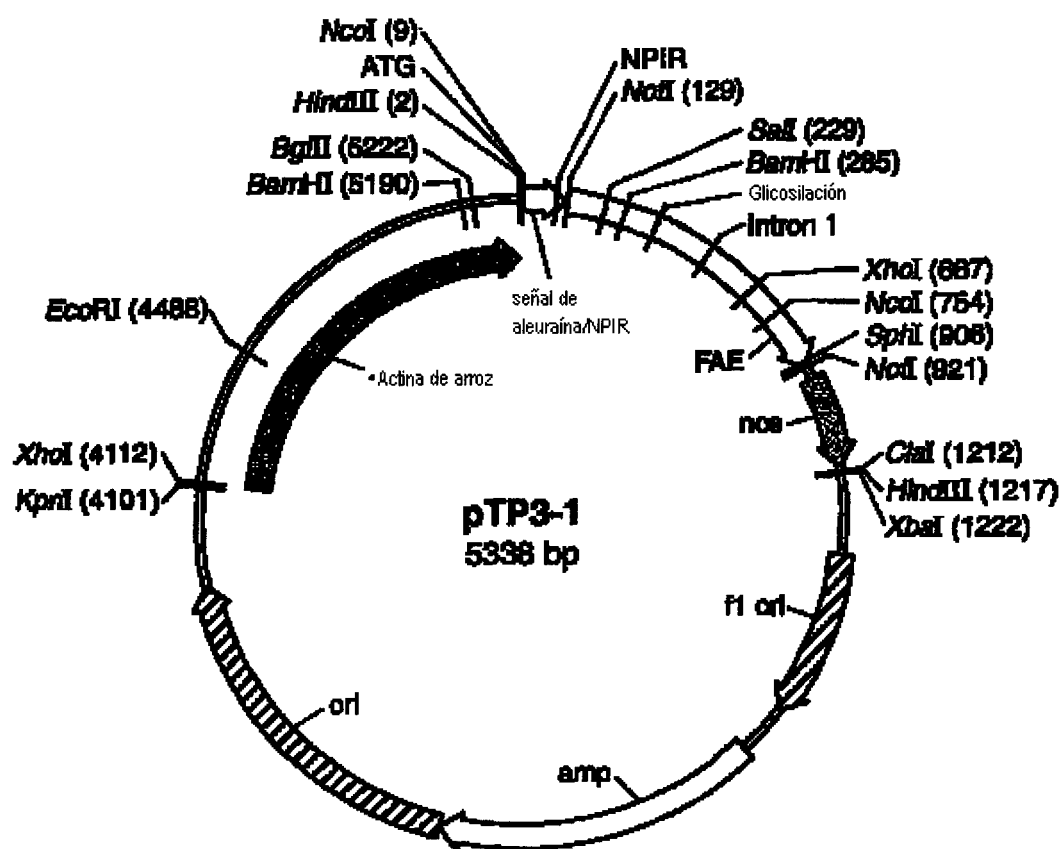


**FIG. 39A**

```

      McoI
      ~~~~~
HindIII
~~~~~
      M A H A R V L L L A L A V L A T A A V A V
1  AAGCTTACCA TGGCCACGCG CCGGTCTC CTCTGGCGC TGGCGTCT GGCACGGCC GCGTGGCG

      NPIR
      ~~~~~
 . A S S S F A D S N P I R P V T D R A A A S T .
71 TGGCTCTCTC CTCTCTCTC GCGACTCC ACGGATCC GCGGTACC 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 TCGNAGACC TCTACAGCG TTATGCGAA ATGGCAGTA 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 TTCTCAACT GACATTAAAG

      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 GATGATCCT CCGCGACGAC AGCAGCAAG AATAATCAC CGTCTCCGT GGCATGGTA GTGATACGAA

 Glycosylation
      ~~~~~
      . 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 GCCTTCGAC ACCCTACC ACCTACGCG 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 ATGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGTTA
      . Q Y P D Y A L T V T G H X L G A S L A L T A .
491 GCCAGTATCC GGACTACCG CTGACCGTGA CCGGCCAC CCGCGCGCC TCCCTGGCG 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 CCGCCAGCTG TCTGGGACAT ACGACACAT CCGCTGTAC ACCTCGCG AACCGCGCAG CCGCAATCAG

```

FIG.\_39B

```

                                XhoI
                                ~~~~~
631 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 CGATGCCCTTC CAGCCTCGA GCCCAGATAC GACGCAGTAT TTCGGGTCA

 NcoI
                                ~~~~~
701  . 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 CCAACCTGC CCGCGGTGA 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 TGCCTGGG ATGAGTGA GTGCTGTGAG

                                SphI
                                ~~~~~
841 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 GACAGGCTGT GAATAATCG CACACGACTT ATTTGGGAT GACGAGCGGC GCATGCACCT

 NotI
                                ~~~~~
                                KDEL
                                ~~~~~
911 . P V A A A E T T E G *
 GGCCGGTCG GCGCGCGAA ACCACTGAG GATGAGCTGT AAGAAGCAG ATCGTTCAAA CATTTGCCA
981 TAAAGTTTCT TAAGATTGAA TCCTGTGCC GGTCTTCCG TGAATTCAT ATAAATTCG TTGAATTCG
1051 TTAAGCATGT AATAATTAC ATGTAATGA TACGCTTAT TATGAGATG GTTTTATGA TTAGAGTCCC
1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAATATAG GCGCAAACT AGGATAAAT ATCGCGCGCG

 HindIII
                                ~~~~~
                                ClaI
                                ~~~~~
 XbaI
                                ~~~~~
1191 GTGTCATCTA TGTACTAGA TCGATAAGCT TCTAGAGCG CCGGTGGAGC TCCAATTCGC CCTATAGTGA
1261 GTCGTATTAC GCGCGCTCAC TGGCGTCTGT TTACAAACGT CGTACTGGG AAACCTTGG CGTTACCCAA
1331 CTTAATCGCC TTGCAGCACA TCCCTCTTC GCCAGCTGC GTATATGCA AGAGGCGCG ACCGATCGCC
1401 CTTCCCAACA GTTGGCGAGC CTGAATGGC AATGGGACG GCGCTGTAG GCGCATTA GCGCGCGCGG

```

FIG.-39C

1471 TGTGGTGGTT ACAGCGACGG TGACCGGTAC ACTTGGCAGC GCCCTAGCGC CGCTCTCTTT CGCTTCTTTC  
 1541 CCTTCTTTTC TCGCCACGTT CGCCGGCTTT CGCCGTCAAG CTCCTAAATCG GGGGCTCCCT TTAGGGTTCC  
 1611 GATTAGTGC TTACGGGCAC TCGACCCCA AAAACCTGA ATAGGTGAT GTTTCACGTA GTTGGCCATC  
 1681 GCCCTGATAG ACGGTTTTTC GCCCTTTGAC TATCTGGGC TATCTTTTG ATTATAGG GATTTGCGG ATTTGTTCCAA  
 1751 ACTGGACAA CACTCAACCC TATCTGGGC TATCTTTTG ATTATAGG GATTTGCGG ATTTGTTCCAA  
 1821 ATTGGTTAA AATGAGCTG ATTTAACAA AATTTAAGC GAATTTTAA AATATATTAA CGCTTACAA  
 1891 TTAGGTGGCA TTTTCTGGG AATGTGCGC AATTTAAGC GAATTTTAA TTTCTAAATA CATTCAATA  
 1961 TGTATCCGCT CATGAGCAA AATGCTTCA ATATATTGA AAAAGGAGA GTATGAGTAT  
 2031 TCAACATTC CGTGTGCCC TATTTCCCT TTTTGGGCA TTTTCCCTC CTGTCTTTC TCACCCGAA  
 2101 ACCTGCTGA AAGTAAAGA TGCTGAAGT CAGTTGGGT CACAGTGGG TTACATCGAA CTGGATCTCA  
 2171 ACAGCGGTAA GATCCTTGA GATTTTGGC CCGAAGACG TTTTCAATG ATGAGCACTT TTAAGTTCT  
 2241 GCTATGTGG GCGGTATTAT CCGTATTGA CCGGGGCA GAGCAACTCG CTGCGCGAT ACATATTCT  
 2311 CAGAAAGCT TGGTTAGTA CTCACCAATG ACAGAAAGC ATCTACGGA TGGCATGCA GTAAAGGAAT  
 2381 TATGAGTGC TGCCATACC ATGAGTATA ACCTCGGC CAATTTACTT CTGACACGA TCGGAGGACC  
 2451 GAAGAGCTA ACCGCTTTT TGCACAACAT GGGGGATCAT GTAACTCGCC TTGATCGTTG GGAACCGAG  
 2521 CTGAATGAAG CCATACCAA CGACGAGCT GACACACGA TCCCTGTAG AATGGCACTA AGCTTGGCA  
 2591 AACTATTAA TGCGCACTA CTACTCTAG CTTCGCGCA ACATTAATA GACTGGATG AGCGGATAA  
 2661 AGTTGAGGA CCATTTCTG CTGCGGCTT CCGGGTGGC TGGTTATTG CTGATAAATC TGGAGCCGCT  
 2731 GAGCGTGGT CTGCGGCTT CATTTGACA CTGGGCGCAG ATGGTAAGC CTCCCGTATC GTAGTTATCT  
 2801 ACACGACGG GAGTCAGCA ACTATGGAT AACGAATAG ACAGATCGCT GAGATAGGTG CCTACTGAT  
 2871 TAAGCATGG TAACTGTGAG ACCAAGTTA CTCATATA CTTAGATTG ATTTAAACT TCATTTTAA  
 2941 TTTAAAGGA TCTAGGTGA GATCTTTT GATATCTCA TGACCAAAAT CCCTTAACGT GAGTTTCTGT  
 3011 TCCACTGAG GTCAAGACCC GTAGAAAGA TCAAGGATC TCTTGAGAT CCTTTTTC TCGCGTAAAT  
 3081 CTGCTGCTG CAACAATAA AACCCGCT GTAGAGATG GTTGTGATG CGATCAAGA GCTACCAACT  
 3151 CTATTTCCGA AGGTAACTG CTTCAGCA GCGCAGATAC CAAATACTGT CTTCTAGTG TAGCCGTAGT  
 3221 TAGGCCACCA CTTCAGAAC TCTGTAGCA CCGCTACATA CTTGCTCTG CTAACTCTGT TACCAGTGGC  
 3291 TGTGCTCAGT GCGATTAAT GTGTCTTAC CCGGTGGAC TCAAGACGAT AGTTACCGA TAAGGCGCAG  
 3361 CGGTGCGCT GAACGGGGG TCTGTGACA CAGCCAGCT TGGAGCGAC GACCTACACC GAACTGAGAT  
 3431 ACCTACAGC TGAGCTATGA GAAAGCGCA CGCTTCCGA AGGAGAAAG GCGGACAGGT ATCCGTTAAG  
 3501 CGGAGGGTC GGAACAGGAG AGCGACGAG GAGCTTCCA GGGGAAACG CCTGGTATCT TTATAGTCT  
 3571 GTCGGGTTT GGCACCTCT ACTTGAGCT CGATTTTGT GATCTCTGC AGGGGGCGG AGCTATGGA  
 3641 AAAAGCCAG CAACCGGCC TTTTACGCT TCCTGGCTT TTGCTCACA TGTCTTTTC  
 3711 TCGGTTATCC CTTGATCTG TGGATACCG TATTAAGCC TTTGAGTGG CTGATACCG TCGCCGAGC  
 3781 CGAAGACCG AGCGAGCGA GTCAGTGAG GAGGAGCGG AAGAGCGCC AATACGCAA CGGCTCTCC  
 3851 CCGCGCGTTG GCGGATTCAT TAATGCACT GGCACGACAG GTTCTCCGNC TGGAAAGCGG CAGTGGAGG

FIG. 39D

```

3921 CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT
3991 ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC
                                     KpnI
4061 GCGCAATTAA CCTTCACTAA AGGGAACAAA AGCTGGGTAC CGGGCCCCC CTGAGGTCA TTCATATGCT
4131 TGAGAAGAGA GTCGGGATAG TCCAAATATA AACAAAGTGA AGATTACCTG GTCAAAAGTG AAACATCAG
4201 TTAAAGGTG GTATAAGTAA AATATCGGTA ATAAAGGTG GCCAAAGTG AAAATTACTC TTTTCTACTA
4271 TTATAAATAA TGAGGATGTT TTGTCGGTAC TTGATACGTT CATTTTGTG TGAATGTTT TTTAAGTTTA
4341 TTCGCGAATT GGAATGCAAT ATCTGTATTT GAGTCGGTTT TTAAGTTCGT TGCTTTGTA AATACAGAGG
4411 GATTGTGATA AGAATATATCT TTAATAAAC CATATGCTAA TTTGACATAA TTTTGGAGAA AAATATATAT
                                     EcoRI
4481 TCAGGCGAAT TCCACAATGA ACAATAATA GATTAAATA GCTTGCCCC GTTGCAGCGA TGGGTATTTT
4551 TTCTAGTAAA ATAAAGATA AACTTAGACT CAAACATTT ACMAAACAA CCGCTAAGT CCTAAGCCCC
4621 AAAGTGCTAT GCACGATCCA TAGCAAGCCC AGCCCAACCC AACCCAGCCC AGTGCAGCCA
4691 ACTGGCAAT AGTCTCCACC CCGGCACTA TCACCGTAG TTGTCCGAC CACCGACGT CTCGCAGCCA
4761 AAAAAAATAA AAGAAAGAA AAAAAAGAA AGAAACAG AGGTGGGTG CGGTCTGTG GGGCCGGAAA
4831 AGCGAGGAGG ATCGCGAGCA GCGACGAGGC CCGGCCCTCC CTCCGCTTC AAGAAACGC CCGCCATTCG
4901 CACTATATAC ATACGCCCCC CTCTCTCTCC ATCCGCCCAA CCCTACCAAC ACCAACCCA CCACCTCCTC
4971 CCCCCTCGCT GCCGAGAGAC GAGTCTCTCC CCCCCTCCCC TCCGCCGCGG CCGGTAAACA CCGCGCCCCC
5041 CTCCTCTTTC TTCTCTCGTT TTTTCTTTCG TCTCGGTCTG GATCTTTGGC CTGTGTAGTT TGGGTGGGCG
5111 AGAGCGGCTT CGTCGCCCG ATCGGTGCGC GGGAGGGGCG GGAATCGCG BglII
                                     BamHI
5181 GTCGGCCCCG ATCCTCGCGG GGAATGGGGC TCTCGGATGT AGATCTTCTT TCTTTCTTCT TTTTGTGGTA
5251 GAATTTGAAAT CCTCAGCAT TGTTCATCGG TAGTTTTTCT TTTCAATGAT TGTGACAAAT GCAGCCTCGT
5321 GCGGAGCTTT TTTGTAGC

```

FIG.\_39E

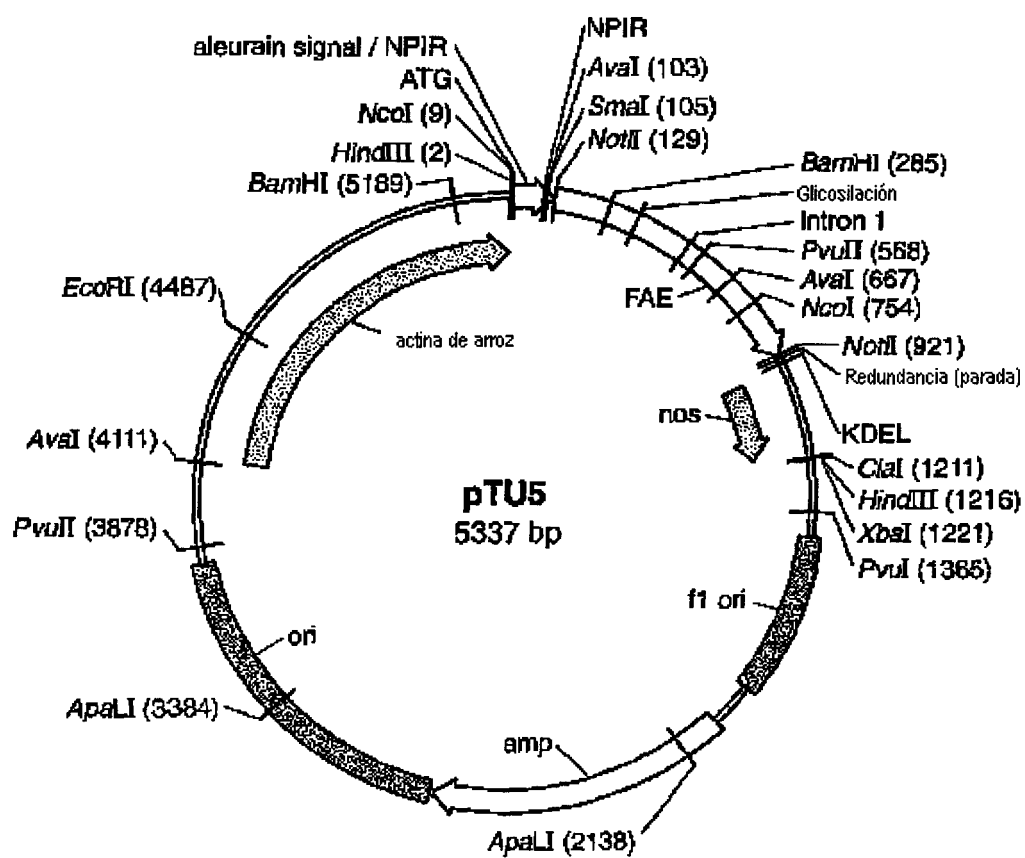


**FIG. 40A**

**HindIII NcoI**  
 ~~~~~

1 AAGCTTACCA TGGCCCAACG CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT
 TTCGAATGGT ACCGGGTGCG GCGCAGGAG GAGGACCGCG AGCGGCACGA

51 GGCACAGGCC GCGGTGCGCG TCGCCTCCTC CTCCTCCTTC GCGACTCCA
 CCGGTGCCGG CCGCAGCGCG AGCGGAGGAG GAGGAGGAAG CCGCTGAGGT

SmaI
 ~~~~~

**AvaI** **NotI**  
 ~~~~~

101 ACCCGGGCCG GCGCGTCACC GACCGCGCGG CCGCCTCCAC GCAGGGCATC
 TGGGCCCGGC CCGGCAGTGG CTGGCGCGCC GCGGAGGGTG CGTCCCGTAG

151 TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCAAAGC
 AGGCTTCTGG AGATGTCGCG AAATCAGCTT TACCGGTGAT AGAGGGTTCG

201 TGCCTACGCC GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA
 ACGGATGCGG CTGGACACGT TGTAAGGCAG CTGATAATAG TTCCCTCTCT

BamHI
 ~~~~~

251 AAATTTACAA TTCTCAAAC TACATTAAAG GATGGATCCT CCGCGACGAC  
 TTAAATGTT AAGAGTTTGA CTGTAATTGC CTACCTAGGA GCGCGTGTG

301 AGCAGCAAAG AAATAATCAC CGTCCTCCGT GGCCTGGTA GTGATACGAA  
 TCGTCGTTTC TTTATTAGTG GCAGAAGGCA CCGTGACCAT CACTATGCTT

351 TCTACAAC TC TACTAACT ACACCCAC GCCTTTCGAC ACCCTACCAC  
 AGATGTTGAG CTATGATTGA TGTGGGAGTG CCGAAAGCTG TGGGATGGTG

401 AATGCAACGG TTGTGAAGTA CACGGTGGAT ATTATATTGG ATGGGTCTCC  
 TTACGTTGCC AACACTTCAT GTGCCACCTA TAATATAACC TACCCAGAGG

451 GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA GCCAGTATCC  
 CAGGTCTTGG TTCAGCTCAG CGAACAGTTT GTCGTCCAAT CGGTCATAGG

501 GGACTACGCG CTGACCGTGA CCGGCCACAC CCTCGGCGCC TCCCTGGCGG  
 CCTGATGCGC GACTGGCACT GGCCGGTGAG GGAGCCGCGG AGGGACCGCC

**PvuII**  
 ~~~~~

551 CACTCACTGC CGCCCAAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC
 GTGAGTGACG GCGGGTCGAC AGACGCTGTA TGCTGTTGTA GCGGACATG

601 ACCTTCGGCG AACCGCGCAG CGGCAATCAG GCCTTCGCGT CGTACATGAA
 TGGAAAGCCGC TTGGCGCGTC GCGGTTAGTC CGGAAGCGCA GCATGTACTT

AvaI
 ~~~~~

651 CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA  
 GCTACGGAAG GTTCGGAGCT CCGGTCTATG CTGCGTCATA AAGGCCAGT

**FIG.\_40B**

```

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGG GCAGGGGTAC
    GAGTACGGTT GCTGCCGTAG GGTTCGGACG GGGGCCACCT CGTCCCCATG

      NcoI
      ~~~~~
751 GCCCATGGCG GTGTAGAGTA CTGGAGCCTT GATCCTTACA GCGCCAGAA
 CGGGTACCGC CACATCTCAT GACCTCGCAA CTAGGAATGT CCGGGGTCTT

801 CACATTTGTC TGCCTGGGG ATGAAGTGCA GTGCTGTGAG GCCCAGGGCG
 GTGTAAACAG ACGTGACCCC TACTTCACGT CACGACACTC CGGGTCCCCG

851 GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC
 CTGTCCCACA CTTATTACGC GTGTGCTGAA TAAAACCCTA CTGCTCGCCG

 NotI
      ~~~~~
901 GCATGCACCT GCGCGGTGCG GCGCGCGGAA CCACTGAAGG ATGAGCTGTA
    CGTACGTGGA CCGGCCAGCG CCGGCGCCTT GGTGACTTCC TACTCGACAT

951 AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCCT AAGATTGAAT
    TTCTTCGTCT AGCAAGTTTG TAAACCGTTA TTCAAAGAA TTCTAACTTA

1001 CCTGTTGCCG GTCTTGCGAT GATTATCATA TAATTTCTGT TGAATTACGT
    GGACAACGGC CAGAACGCTA CTAATAGTAT ATTAAAGACA ACTTAATGCA

1051 TAAGCATGTA ATAATTAACA TGTAATGCAT GACGTTATTT ATGAGATGGG
    ATTCGTACAT TATTAATTGT ACATTACGTA CTGCAATAAA TACTCTACCC

1101 TTTTATGAT TAGAGTCCCG CAATTATACA TTTAATACGC GATAGAAAAC
    AAAAATACTA ATCTCAGGGC GTTAATATGT AAATTATGCG CTATCTTTTG

1151 AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGG TGTCATCTAT
    TTTTATATCG CGCGTTTGAT CCTATTTAAT AGCGCGCGCC ACAGTAGATA

      XbaI
      ~~~~~
 ClaI HindIII
      ~~~~~
1201 GTTACTAGAT CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTCGCC
    CAATGATCTA GCTATTGAA GATCTCGCCG GCCACCTCGA GGTAAAGCGG

1251 CTATAGTGAG TCGTATTACG CCGGCTCACT GGCCGTCGTT TTACAACGTC
    GATATCACTC AGCATAATGC GCGCGAGTGA CCGGCAGCAA AATGTTGCAG

1301 GTGACTGGGA AAACCCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT
    CACTGACCCT TTTGGGACCG CAATGGGTTG AATTAGCGGA ACGTCGTGTA

      PvuII
      ~~~~~
1351 CCCCCTTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC
 GGGGGAAAGC GGTGACCGC ATTATCGCTT CTCCGGGCGT GGCTAGCGGG

1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCCTGTAGCG
 AAGGGTTGTC AACGCGTCGG ACTTACCGCT TACCCTGCGC GGGACATCGC

```

**FIG.\_40C**

|      |                           |                           |                           |                          |                          |
|------|---------------------------|---------------------------|---------------------------|--------------------------|--------------------------|
| 1451 | GCGCATTAAAG<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 | CTTCCTTTC<br>GAAGGAAAGA  |
| 1551 | CGCCACGTTT<br>GCGGTGCAAG  | GCCGGCTTTC<br>CGGCCGAAAG  | CCCCTCAAGC<br>GGGCAGTTCG  | TCTAAATCGG<br>AGATTTAGCC | GGGCTCCCTT<br>CCCAGGGAA  |
| 1601 | TAGGGTTCCG<br>ATCCCAAGGC  | ATTTAGTGCT<br>TAAATCACGA  | TTACGGCACC<br>AATGCCGTGG  | TCGACCCCAA<br>AGCTGGGGTT | AAAACTTGAT<br>TTTTGAACTA |
| 1651 | TAGGGTGATG<br>ATCCCACTAC  | GTTACAGTAG<br>CAAGTGCATC  | TGGGCCATCG<br>ACCCGGTAGC  | CCCTGATAGA<br>GGGACTATCT | CGGTTTTTCG<br>GCCAAAAAGC |
| 1701 | CCCTTTGACG<br>GGGAAACTGC  | TTGGAGTCCA<br>AACCTCAGGT  | CGTTCTTTAA<br>GCAAGAAATT  | TAGTGGACTC<br>ATCACCTGAG | TTGTTCCAAA<br>AACAAGGTTT |
| 1751 | CTGGAACAAC<br>GACCTTGTTG  | ACTCAACCTT<br>TGAGTTGGGA  | ATCTCGGTCT<br>TAGAGCCAGA  | ATTCTTTTGA<br>TAAGAAAAC  | TTTATAAGGG<br>AAATATTCCC |
| 1801 | ATTTTGCCGA<br>TAAACGGCT   | TTTCGGCCTA<br>AAAGCCGGAT  | TTGGTTAAAA<br>AACCAATTTT  | AATGAGCTGA<br>TTACTCGACT | TTTAACAAAA<br>AAATTGTTTT |
| 1851 | ATTTAACGCG<br>TAAATTGCGC  | AATTTTAACA<br>TTAAAAATTGT | AAATATTAAC<br>TTTATAATTG  | GCTTACAATT<br>CGAATGTTAA | TAGGTGGCAC<br>ATCCACCGTG |
| 1901 | TTTTCGGGGA<br>AAAAGCCCC   | AATGTGCGCG<br>TTACACGCGC  | GAACCCCTAT<br>CTTGGGGATA  | TTGTTTATTT<br>AACAAATAAA | TTCTAAATAC<br>AAGATTTATG |
| 1951 | ATTCAAATAT<br>TAAGTTTATA  | GTATCCGCTC<br>CATAGGCGAG  | ATGAGACAAT<br>TACTCTGTTA  | 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  | TTTGCCCTTC<br>AAACGGAAGG  | TGTTTTTGCT<br>ACAAAAACGA | CACCCAGAAA<br>GTGGGTCTTT |
|      |                           |                           |                           | ApaLI<br>~~~~~           |                          |
| 2101 | CGCTGGTGAA<br>GCGACCACCT  | 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 | GTTTTCGCCC<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  | GTTTGAGTAC<br>CCAACCTCATG | TCACCAGTCA<br>AGTGGTCAGT | CAGAAAAGCA<br>GTCTTTTCGT |

**FIG.\_40D**

|      |                          |                           |                           |                          |                          |
|------|--------------------------|---------------------------|---------------------------|--------------------------|--------------------------|
| 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>CGTGTGTGAC  | 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>TGATAAATGA |
| 2601 | GGCGAACTAC<br>CCGCTTGATG | TTACTCTAGC<br>AATGAGATCG  | TTCCCGGCAA<br>AAGGGCCGTT  | CAATTAATAG<br>GTTAATTATC | ACTGGATGGA<br>TGACCTACCT |
| 2651 | GGCGGATAAA<br>CCGCTATTTT | GTTGCAGGAC<br>CAACGTCCCTG | 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>ACCATTCCGG  | TCCCGTATCG<br>AGGGCATAGC | TAGTTATCTA<br>ATCAATAGAT |
| 2801 | CACGACGGGG<br>GTGCTGCCCC | AGTCAGGCAA<br>TCAGTCCGTT  | CTATGGATGA<br>GATACCTACT  | ACGAAATAGA<br>TGCTTTATCT | CAGATCGCTG<br>GTCTAGCGAC |
| 2851 | AGATAGGTGC<br>TCTATCCACG | CTCACTGATT<br>GAGTGACTAA  | AAGCATTGGT<br>TTCGTAACCA  | AACTGTCAGA<br>TTGACAGTCT | CCAAGTTTAC<br>GGTTCAAATG |
| 2901 | TCATATATAC<br>AGTATATATG | TTTAGATTGA<br>AAATCTAACT  | TTTAAACTT<br>AAATTTTGAA   | CATTTTAAAT<br>GTAAAAATTA | TTAAAAGGAT<br>AATTTTCCTA |
| 2951 | CTAGGTGAAG<br>GATCCACTTC | ATCCTTTTTG<br>TAGGAAAAAC  | ATAATCTCAT<br>TATTAGAGTA  | GACCAAAATC<br>CTGGTTTTAG | CCTTAACGTG<br>GGAATTGCAC |
| 3001 | AGTTTTTCGT<br>TCAAAAGCAA | CCACTGAGCG<br>GGTGACTCGC  | TCAGACCCCG<br>AGTCTGGGGC  | TAGAAAAGAT<br>ATCTTTTCTA | CAAAGGATCT<br>GTTTCCTAGA |
| 3051 | TCTTGAGATC<br>AGAACTCTAG | CTTTTTTTCT<br>GAAAAAAGA   | GCGCGTAATC<br>CGCGCATTAG  | TGCTGCTTGC<br>ACGACGAACG | AAACAAAAAA<br>TTTGTTTT   |
| 3101 | ACCACCGCTA<br>TGGTGCGGAT | CCAGCGGTGG<br>GGTCGCCACC  | TTGTTTGCC<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  | AGGCCACCCAC<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**

3301 GCGATAAGTC GTGTCTTACC GGGTTGGACT CAAGACGATA GTTACCGGAT  
CGCTATTTCAG CACAGAATGG CCCAACCTGA GTTCTGCTAT CAATGGCCTA

ApaLI  
\*\*\*\*\*

3351 AAGGCGCAGC GGTGCGGGCTG AACGGGGGGT TCGTGACACAC AGCCCAGCTT  
TCCGCGTCTG CCAGCCCGAC TTGCCCCCA AGCACGTGTG TCGGGTCGAA

3401 GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG  
CCTCGCTTGC TGGATGTGGC TTGACTCTAT GGATGTGCA CTCGATACTC

3451 AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CGGACAGGTA TCCGGTAAGC  
TTTCGCGGTG CGAAGGGCTT CCTCTTTCC GCCTGTCCAT AGGCCATTGC

3501 GGCAGGGTCTG GAACAGGAGA GCGCACGAGG GAGCTTCCAG GGGGAAACGC  
CCGTCCCAGC CTTGTCTCTT CCGGTGCTCC CTCGAAGGTC CCCCTTTGCG

3551 CTGGTATCTT TATAGTCCTG TCGGGTTTCG CCACCTCTGA CTTGAGCGTC  
GACCATAGAA ATATCAGGAC AGCCCAAAGC GGTGGAGACT GAACCTGCGAG

3601 GATTTTTGTG ATGCTCGTCA GGGGGGCGGA GCCTATGGAA AAACGCCAGC  
CTAAAAACAC TACGAGCAGT CCCCCGCT CGGATACCTT TTTGCGGTG

3651 AACGCGGCCT TTTACGGTT CCTGGCCTT TGCTGGCCTT TTGCTCACAT  
TTGCGCCGGA AAAATGCCAA GGACCGGAA ACGACCGGA AACGAGTGT

3701 GTTCTTTCTT GCGTTATCCC CTGATTCTGT GGATAACCGT ATTACCGCT  
CAAGAAAGGA CGCAATAGGG GACTAAGACA CCTATTGCA TAATGGCGGA

3751 TTGAGTGAGC TGATACCGCT CGCCGCGAGC GAACGACCGA GCGCAGCGAG  
AACTCACTCG ACTATGGCGA GCGGCGTCGG CTTGCTGGCT CCGGTGCTC

3801 TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAAC CGCCTCTCCC  
AGTCACTCGC TCCTTCGCC TCTCGCGGGT TATGCGTTTG GCGGAGAGGG

PvuII  
\*\*\*\*\*

3851 CGCGCGTTGG CCGATTCAAT AATGCAGCTG GCACGACAGG TTTCCCGACT  
GCGCGCAACC GGCTAAGTAA TTACGTCGAC CGTGCTGTCC AAAGGGCTGA

3901 GGAAAGCGGG CAGTGAGCGC AACGCAATTA ATGTGAGTTA GCTCACTCAT  
CCTTTCGCCC GTCACTCGCG TTGCGTTAAT TACACTCAAT CGAGTGAGTA

3951 TAGGCACCCC AGGCTTTACA CTTTATGCTT CCGGCTCGTA TGTGTGTGG  
ATCCGTGGGG TCCGAAATGT GAAATACGAA GGCCGAGCAT ACAACACACC

4001 AATTGTGAGC GGATAACAAT TTCACACAGG AAACAGCTAT GACCATGATT  
TTAACACTCG CCTATTGTTA AAGTGTGTCC TTTGTGATA CTGGTACTAA

4051 ACGCCAAGCG CGCAATTAACT CCTCACTAAA GGGAAACAAA GCTGGGTACC  
TGCGGTTTCG CGGTTAATTG GGAGTGATTT CCCTTGTTTT CGACCCATGG

AvaI  
\*\*\*\*\*

4101 GGGCCCCCCC TCGAGGTCAT TCATATGCTT GAGAAGAGAG TCGGGATAGT  
CCCGGGGGGG AGCTCCAGTA AGTATACGAA CTCTTCTCTC AGCCCTATCA

FIG.\_40F

```

4151 CCAAAATAAA ACAAGGTAA GATTACCTGG TCAAAAGTGA AAACATCAGT
GGTTTTATTT TGTTTCCATT CTAATGGACC AGTTTTCACT FTTGTAGTCA

4201 TAAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAAGGTGG CCCAAAGTGA
ATTTCCACC ATATTCATTT TATAGCCATT ATTTCCACC GGGTTTCACT

4251 AATTACTCT TTTCTACTAT TATAAAATT GAGGATGTT TGTGGTACT
TTAATGAGA AAAGATGATA ATATTTTAA CTCCTACAAA ACAGCCATGA

4301 TTGATACGTC ATTTTGTAT GAATTGGTTT TTAAGTTTAT TCGCGATTTG
AAGTATGCAG TAAAAACATA CTTAACCAA AATTCAAATA AGCGCTAAAC

4351 GAAATGCATA TCTGTATTTG AGTCGGTTTT TAAGTTCGTT GCTTTTGTAA
CTTTACGTAT AGACATAAAC TCAGCCAAA ATTCAAGCAA CGAAAACATG

4401 ATACAGAGGG ATTTGTATAA GAAATATCTT TAAAAACCC ATATGCTAAT
TATGTCTCCC TAAACATATT CTTATAGAA ATTTTGGG TATACGATTA

 EcoRI
                                     ~~~~~~
4451 TTGACATAAT TTTTGAGAAA AATATATATT CAGGCGAATT CCACAATGAA
AATGTATTA AAAACTCTTT TTATATATAA GTCCGCTTAA GGTGTTACTT

4501 CAATAATAAG ATTAABATAG CTTGCCCCCG TTGCAGCGAT GGGTATTTTT
GTTATTATTC TAATTTTATC GAACGGGGGC AACGTCGCTA CCCATAAAAA

4551 TCTAGTAAAA TAAAGATAA ACTTAGACTC AAAACATTTA CAAAAACAAC
AGATCATTTC ATTTTCTATT TGAATCTGAG TTTTGTAAAT GTTTTGTGTG

4601 CCCTAAAGTC CTAAAGCCCA AAGTGCTATG CACGATCCAT AGCAAGCCCA
GGGATTTTCA GATTTCTGGT TTCACGATAC GTGCTAGGTA TCGTTCGGGT

4651 GCCCAACCCA ACCCAACCCA ACCCACCCA GTGCAGCCAA CTGGCAATA
CGGTTGGGT TGGGTGGGT TGGGTGGGT CACGTCGGT GACCGTTTAT

4701 GTCTCCACCC CCGGCACAT CACCGTGAGT TGTCCGCACC ACCGCACGTC
CAGAGGTGGG GGCCGTGATA GTGGCACTCA ACAGGCGTGG TGGCGTGCAG

4751 TCGCAGCCAA AAAAAAAAAA AGAAAGAAAA AAAAGAAAAA GAAAAACAGC
AGCGTCGGTT TTTTTTTTTT TCTTCTTTT TTTTCTTTT CTTTTGTGCG

4801 AGGTGGGTCC GGTTCGTGGG GGCCGAAAA GCGAGGAGGA TCGCGAGCAG
TCCACCCAGG CCCAGCACCC CGGCCTTTT CGCTCCTCCT AGCGCTCGTC

4851 CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC
GCTGCTCCGG GCGGGAGGG AGGCGAAGGT TTCTTTGCGG GGGGTAGCGG

4901 ACTATATACA TACCCCCCCC TCTCCTCCCA TCCCCCAAC CCTACCACCA
TGATATATGT ATGGGGGGGG AGAGGAGGGT AGGGGGGTTG GGATGGTGGT

4951 CCACCACCAC CACCTCCTCC CCCCTCGCTG CCGGACGACG AGCTCCTCCC
GGTGGTGGTG GTGGAGGAGG GGGGAGCGAC GGCCTGCTGC TCGAGGAGGG

5001 CCCTCCCCCT CCGCCGCCGC CGGTAACCAC CCCGCCCTC TCCTCTTTCT
GGGAGGGGGA GCGCGCGCG GCGATTGGTG GGGCGGGGAG AGGAGAAAGA

```

**FIG. 40G**

```

5051  TTCTCCGTTT TTTTTTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT
      AAGAGGCAAA AAAAAAAGCA GAGCCAGAGC TAGAAACCGG AACCATCAAA

5101  GGGTGGGCGA GAGCGGCTTC GTCGCCCAGA TCGGTGCGCG GGAGGGGCGG
      CCCACCCGCT CTCGCCGAAG CAGCGGGTCT AGCCACGCGC CCTCCCCGCC

                                     BamHI
                                     ~~~~~~
5151 GATCTCGCGG CTGGCGTCTC CGGGCGTGAG TCGGCCCGGA TCCTCGCGGG
 CTAGAGCGCC GACCGCAGAG GCCCGCACTC AGCCGGGCCT AGGAGCGCCC

5201 GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT TTTGTGGTAG
 CTTACCCCGA GAGCCTACAT CTAGAAGAAA GAAAGAAGAA AACACCATC

5251 AATTTGAATC CCTCAGCATT GTTCATCGGT AGTTTTTCTT TTCATGATTT
 TTAAACTTAG GGAGTCGTAA CAAGTAGCCA TCAAAAAGAA AAGTACTAAA

5301 GTGACAAATG CAGCCTCGTG CGGAGCTTTT TTGTAGC
 CACTGTTTAC GTCGGAGCAC GCCTCGAAAA AACATCG

```

**FIG.\_40H**

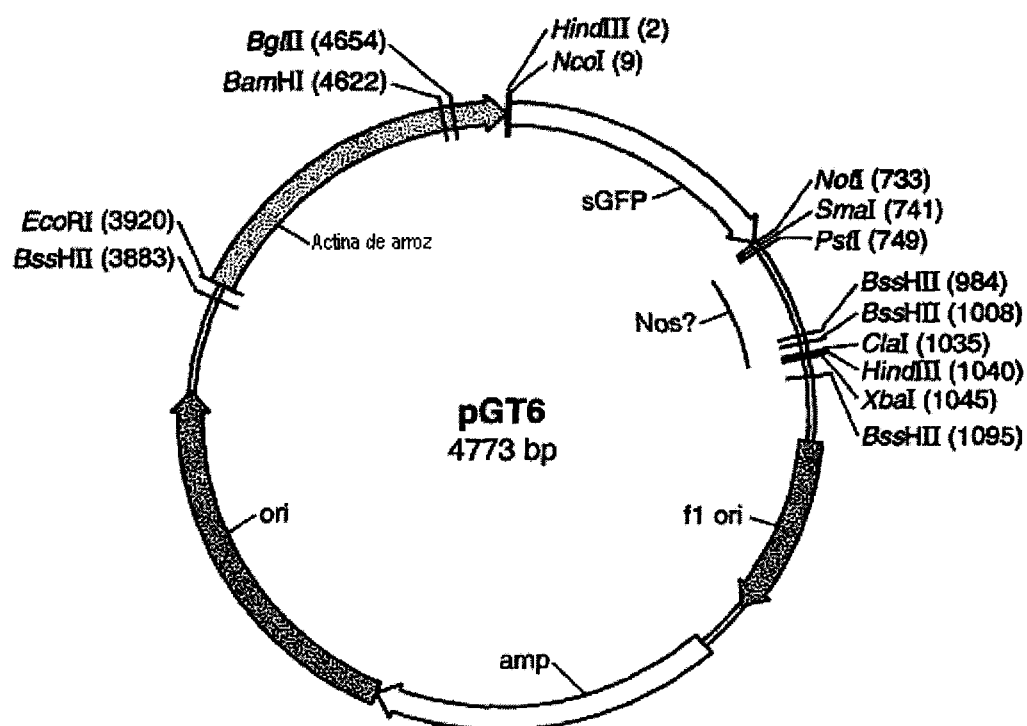


**FIG.\_41A**

HindIII NcoI  
 ~~~~~  
 1 AAGCTTACCA TGGTAGCAA GGGCGAGGAG CTGTTACCCG GGGTGGTGCC CATCTGGTC GAGCTGACG  
 TTCGAATGGT ACCACTCGTT CCGCTCCTC GACAAGTGGC CCCACCACGG GTAGGACCAG CTCGACCCTGC  
 71 GCGACGTGAA CGGCCACAAG TTCAGCGTGT CCGGCGAGGG CQAGGGCGAT GCCACCTACG GCAAGCTGAC  
 CGCTGCACCT GCCGGTGTTC AAGTCGCACA GGGCGCTCCC GCTCCCGCTA CCGTGGATGC CGTTCGACTG  
 141 CCTGAAGTTC ATCTGCACCA CCGGCAAGCT GCCCGTGCCC TGGCCCAACC TCGTGACCAC CTTACACCTAC  
 GGACTTCAAG TAGACGTGGT GGCCGTTCGA CGGGCACGGG ACCGGGTGGG AGCACTGGTG GAAGTGGATG  
 211 GCGGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCAGGACTT CTTCAAGTCC GCCATGCCCG  
 CCGCACGTCA CGAAGTCGGC GATGGGGCTG GTGTACTTCG TCGTGTCTGA GAA GTTCAGG CGGTACGGGC  
 281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC AAGACCCGCG CCGAGGTGAA  
 TTCCGATGCA GGTCTCTCGC TGGTAGAAGA AGTTCCTGCT GCCGTTGATG TTCTGGGCGC GGCTCCACTT  
 351 GTTCGAGGGC GACACCCCTGG TGAACCGCAT CGAGCTGAAG GGCATCGACT TCAAGGAGGA CCGCAACATC  
 CAAGCTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA AGTTCCCTCT GCCGTGTGAG  
 421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT GGCCGACAAAG CAGAAGAACG  
 GACCCCGTGT TCGACCTCAT GTTGATGTTG TCGGTGTTGC AGATATAGTA CCGGCTGTTT GTCTTCTTGC  
 491 GCATCAAGGT GAACCTCAAG ATCCGCCACA ACATCGAGGA CCGCAGCGTG CAGCTCGCCG ACCACTACCA  
 CGTAGTTCCA CTTGAAGTTC TAGGCGGTGT TGTAGCTCCT GCCGTCCGAC GTCGAGCGGC TGGTGATGGT  
 561 GCAGAACACC CCATCGGCG ACGGCCCCGT GCTGCTGCCC GACAACCACT ACCTGAGCAC CCAGTCCGCC  
 CGTCTTGTGG GGGTAGCCGC TGCCGGGGCA CGACGACGGG CTGTTGGTGA TGGACTCGTG GGTACGGCGG  
 631 CTGAGCAAG ACCCCAACGA GAAGCGCGAT CACATGGTCC TGGTGGAGTT CGTGACCGCC GCCGGGATCA  
 GACTCGTTTC TGGGGTTGCT CTTGCGGCTA GTGTACCAGG ACGACCTCAA GCACTGGCGG CGGCCCTAGT

**FIG.\_41B**

SmaI

NotI

PstI

701

CTCACGGCAT GGACGAGCTG TACAAGTAAA GCGGCCGCC CCGCTGCAGG GAAACCAC TG AAGGATGAGC GAGTGGCCGA CCTGCTCGAC ATGTTCAATT CGCCGGCGGG CCCGACGTCC CTTTGGTGAC TTCCTACTCG

771

TGTAAGAAG CAGATCGTTC AAACATTGG CAATAAAGTT TCTTAAGATT GAATCCTGTT GCCGGTCTTG ACATTTCTTC GTCTAGCAAG TTGTAAACC GTTATTTC AA AGAATTC TAA CTTAGGACAA CGGCCAGAAC

841

CGATGATTAT CATATAATTT CTGTTGAATT ACGTTAAGCA TGTATAAATT AACATGTAAT GCATGACGTT GCTACTAATA GTATATTAAA GACAACTTAA TGCAATTCGT ACATTATTAA TTGTACATTA CGTACTGCAA

911

ATTTATGAGA TGGGTTTTTA TGATTAGAGT CCCGCAATTA TACATTTAAT ACGCCATAGA AAACAAAATA TAAATACTCT ACCCAAAAT ACTAATCTCA GGGCGTTAAT ATGTAAATTA TCGCGTATCT TTTGTTTTAT

BssHII

BssHII

981

TAGCGGCGCA ACTAGGATAA ATTATCGCGC GCGGTGTCTAT CTATGTTACT AGATCGATAA GCTTCTAGAG ATCGCGCGTT TGATCCCTATT TAATAGCGCG CGCCACAGTA GATACAATGA TCTAGCTATT CGAAGATCTC

XbaI

Clal

HindIII

1051

CGGCCGGTGG AGCTCCAATT CGCCCTATAG TGAGTCGTAT TACGCGCGCT CACTGGCCGT CGTTTTACAA GCCGGCCACC TCGAGGTTAA GCGGGATATC ACTCAGCATA ATGCGCGCGA GTGACCGGCA GCMAATGTT

1121

CGTCGTGACT GGGAAAACCC TGGCGTTACC CAACTTAATC GCCTTGCAGC ACATCCCCCT TTCGCCAGCT GCAGCACTGA CCTTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCG TGTAGGGGGA AAGCGGTCTGA

1191

GGCGTAATAG CGAAGAGGCC CGCACCGATC GCCCTTCCCA ACAGTTGCGC AGCCTGAATG GCGAATGGGA CCGCATATAT GCTTCTCCGG GCGTGGCTAG CGGGAAGGTT TGTCAACGCG TCGGACTTAC CGCTTACCCCT

1261

CGCGCCCTGT AGCGGCGCAT TAAGCGCGGC GGGTGTGGTG GTTACGCGCA GCGTGACCGC TACACTTGCC GCCGGGACA TCGCCGCGTA ATTGCGCGCG CCCACACCAC CAATGCGCGT CGCACTGGCG ATGTGAACGG

FIG.\_41C

112

```

1331 AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCCGGC TTTCGCCGTC
TCGCGGGATC GCGGCGGAGG AAGCGAAG AAGCGAAGG AAGAGCGGTG CAAGCGGCGG AAAGGGGCGAG

1401 AAGCTCTAAA TCGGGGGCTC CCTTTAGGGT TCGGATTTAG TGCTTTACGG CACCTCGACC CCAAAAAACT
TTGAGATTT AGCCCCCGAG GGAATCCCA AGGCTAAATC ACGAATGCC GTGGAGCTGG GGTTTTTTGA

1471 TGATTAGGGT GATGGTTTAC GTAGTGGGCC ATCGCCCTGA TAGACGGTTT TTGCCCCCTT GACGTTGGAG
ACTAATCCCA CTACCAGTG CATCACCCGG TAGCGGACT ATCTGCCAA AAGCGGAAA CTGCAACCTC

1541 TCCACGTTCT TTAATAGTGG ACTCTGTTC CAACTGGAA CAACACTCAA CCTATCTCG GTCTATTCTT
AGGTGCAAGA AATTATCACC TGAGAACAG GTTGACCTT GTTGTGAGTT GGGATAGAGC CAGATAAGAA

1611 TTGATTTATA AGGGATTTTG CCGATTCGG CCTATTGGTT AAAAAATGAG CTGATTTAAC AAAAAATTAA
AACTAAATAT TCCCTAAAAC GGCTAAAGCC GGATAACCA TTTTCTACTC GACTAAATTG TTTTAAATT

1681 CGCGAATTT AACAAAATAT TAACGCTTAC AATTAGGTG GCCTTTTCG GGGAAATGTG CGCGGAACCC
GCGCTAAAA TTGTTTATA ATTGCGAATG TTAAATCCAC CGTGAAGAAG CCTTTACAC GCGCCTTGGG

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

1821 TCAATAATAT TGA AAAAGGA AGAGTATGAG TATTCACAT TTCCGTGTCG CCCTATTTC CTTTTTGGC
AGTTATTATA ACTTTTCTCT TCTCATACTC ATAGTTGTA AAGGCACAGC GGGAAATAGG GAAAAACGC

1891 GCATTTTGGC TTCTGTGTTT TGCTCACCCA GAACGCTGG TGAAGTAAA AGATGCTGAA GATCAGTTGG
CGTAAAACGG AAGGACAAAA ACGAGTGGGT CTTTGCAGC ACTTTCATT TCTACGACTT CTAGTCNACC

1961 GTGCACGAGT GGGTTACATC GAAC TGATC TCAACAGCG TAAGATCCTT GAGAGTTTC GCCCGAAGA
CACGTGCTCA CCCAATGTAG CTTGACCTAG AGTTGTGCGC ATTCTAGAA CTCTCAAAAG CCGGGCTTCT

2031 ACGTTTCCA ATGATGAGCA CTTTAAAGT TGTGCTATGT GCGCGGTAT TATCCGTAT TGACGCCGGG
TGCAAAAGGT TACTACTCGT GAAATTTCA AGACGATACA CCGCGCCATA ATAGGCATA ACTGCGGCC

2101 CAAGAGCAAC TCGGTGCGCG CATACACTAT TCTCAGATG ACTTGTGTA GTACTCACCA GTCACAGAAA
GTTCTCGTTG AGCCAGCGGC GTATGTGATA AGAGTCTTAC TGAACCACT CATGAGTGGT CAGTGTCTTT

```

FIG. 41D

2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGTGTCATTA ACCATGAGTG ATAACACTGC  
 TCGTAGAATG CCTACCCOTAC TGTCAATCTC TTAATACGTC ACGACGGTAT TGGTACTCAC TATTGTGACG  
 2241 GGCCAACTTA CTTCAGCAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT TTTTGCACAA CATGGGGAT  
 CCGGTTGAAT GAAGACTGTT GCTAGCCTCC TGGCTTCCTC GATTGGGAA AATACGTGTT GTACCCCTTA  
 2311 CATGTAATC GCCTTGATCG TTGGGAACCG GAGCTGAATG AAGCCATACC AATGACGAG CGTGACACCA  
 GTACATTGAG CGGAACCTAGC AACCTTGGC CTCGACTTAC TTCGGTATGG TTTGCTGCTC GCACTGTGGT  
 2381 CGATGCCCTGT AGCAATGGCA ACAACGTTGC GCAACTATT GCACTTACTC TAGCTTCCC  
 GCTACGGACA TCGTTACCGT TGTTCGAACG CGTTGATAA TTGACCGCTT GATGAATGAG ATCGAAGGGC  
 2451 GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTTC TGGCTCGGC CCTTCCGGCT  
 CGTTGTTAAT TATCTGACCT ACCTCCGCCT ATTTCAACGT CTGCTGTAAG ACGGAGCCG GGAAGGCCGA  
 2521 GGTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCG TATCATTGCA GCACTGGGC  
 CCGACCAAT AACGACTATT TAGACCTCG CCACCTCGAC CCAGAGCGCC ATAGTAACGT CGTGACCCCG  
 2591 CAGATGGTAA GCCCTCCGT ATCTAGTTA TCTACACGAC GGGGATCAG GCACTATAG ATGAACGAAA  
 GTCTACCAAT CGGGAGGCA TAGCATCAAT AGATGTGCTG CCCCTCAGTC CGTTGATACC TACTTGCTTT  
 2661 TAGACAGATC GCTGAGATAG GTGCTCACT GATTAAGCAT TGGTAACTGT CAGACCAAGT TTACTCATAT  
 ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTCGTA ACCATTGACA GTCTGGTTCA AATGAGTATA  
 2731 ATACTTTTGA TTGATTTAAA ACTTCATTTT TAATTTAAA GATCTAGGT GAAGATCCTT TTTGATTAATC  
 TATGAATATCT AACTAAATTT TGAAGTAAA ATTAAATTTT CCTAGATCCA CTCTAGGAA AACTATATTAG  
 2801 TCATGACCAA AATCCCTTAA CGTGAGTTT CGTTCCACTG AGCGTCAGAC CCGGTAGAAA AGATCAAAGG  
 AGTACTGGTT TTAGGGAATT GCACTCAAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTT TCTAGTTTCC  
 2871 ATCTTCTTGA GATCCTTTT TTCTGCGCGT AATCTGCTGC TTGCAACAA AAAAACCAAC GCTACCAAGC  
 TAGAAGAACT CTAGGAAAAA AAGACGCGCA TTAGACGAGG AACGTTGTT TTTTGGTGG CGATGGTCCG  
 2941 GTGGTTTGT TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA  
 CACCAACAA ACGGCCTAGT TCTCGATGGT TGAGAAAAAG GCTTCCATTG ACCGAAGTCC TCTCGCGTCT

**FIG. 41E**

3011 TACCATAATAC TGTCTTCTTA GTGTAGCCGT AGTTAGGCCA CCACTTCNAG AACCTCTTAG CACCGCTTAC  
 ATGGTTTATG ACAGGAAGAT CACATCGCA TCAATCCGGT GGTGAAGTTC TTGAGACATC GTGGCGGATG  
 3081 ATACCTCGCT CTGCTAATCC TGTTACCAAT GGTCTCTGCC AGTGGCGATA AGTCGTGTCT TACCGGGTTG  
 TATGGAGCGA GACGATTAGG ACAATGGTCA CCGAGCAGCG TCACCGCTAT TCAGCACAGA ATGCCCGAAC  
 3151 GACTCAAGAC GATAGTTACC GGATAAGCG CAGCGTCTGG GCTGAACGGG GGGTTCTGTC ACACAGCCCA  
 CTGAGTCTTG CTATCAATGG CCTATTTCGC GTGCCAGCC CCACTTGCCC CCCAAGCAGG TGTGTCCGGT  
 3221 GCTTGGAGCG AACGACCTAC ACCGAACTGA GATACCTACA GGTGAGCTA TGAGAAAGCG CCACGCTTCC  
 CGAACCTCGC TTGCTGGATG TGGCTTGACT CTATGCGAT GTGACTCGAT ACTCTTTCGC GGTGCGAAGG  
 3291 CGAAGGGAGA AAGCGGACA GGTATCCGT AAGCGGCGAG GTCGGAACAG GAGAGCGCAC GAGGAGCTT  
 GCTTCCCTCT TCCGCGCTGT CCATAGGCCA TTCGCGCTCC CAGCCTTGTG CTCTCGCGTG CTCCTCTCGAA  
 3361 CCAGGGGAA ACGCTGGTA TCTTTATAGT CCTGTGCGGT TTCGCCACT CTGACTTQAG CGTCGATTTT  
 GGTCCCCCTT TCGGACCAT AGAATATCA GGCAGGCCA AAGCGGTGA GACTGAACCTC GCAGCTAATA  
 3431 TGTGATGCTC GTCAGGGGG CGGAGCCTAT GGAAACGC CAGCAACGCG GCTTTTAC GGTTCCTGGC  
 ACACCTACGAG CAGTCCCCC GCTTCGGATA CTTTTGCG GTCTTTCGC CGGAAATG CCAAGGACCG  
 3501 CTTTTGCTGG CTTTTGCTC ACATGTTCTT TCTGCGTTA TCCCTGAT CTGTGGATAA CCGTATTACG  
 GAAACGACC GGAAACGAG TGTACAGAA AGGACGCAT AGGGGACTAA GACACCTATT GGCATTAATG  
 3571 GCTTTTGAAT GAGCTGATAC GCTCGCCGC AGCCGAAACGA CCGAGCGCAG CGAGTCTAGT AGCGAGGAG  
 CGAAACTCA CTCGACTATG GCGAGCGCG TCGGCTTGT GGTCTGCTAC TCGCTCTCTC  
 3641 CGGAAGAGCG CCCAATACGC AAACCGCTC TCCCGCGCG TTGGCGGAT CATTATGCA GCTGGCACGA  
 GCTTCTCGC GGGTTATGCG TTTGGCGGAG AGGGCGCGC AACCGCTAA GTAATTACGT CGACCTGCT  
 3711 CAGGTTTCCC GACTGGGAAG CGGGCAGTGA GCGCAACGCA ATTAAATGTA GTTAGCTCAC TCATTAGGCA  
 GTCCAAAGGG CTGACCTTTC GCGGTGCTCT TAATTACACT CAATCGAGTG AGTAATCCGT  
 3781 CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT GTGGAATTGT GAGCGGATAA CAATTTGACA  
 GGGGTCCGAA ATGTGAATA CGAAGGCCG GCATACACA CAGCTTACA CTCGCTTATT GTTAAAGTGT

FIG.\_41F

```

3851 CAGGAACAG CTAAGACCAT GATTACGCCA AGCGCCCAAT TAACCTCTAC TAAAGGGAAC AAAGCTGGA
 GTCCCTTGTC GATACTGGTA CTAATGCGGT TCGCGCGTTA ATTGGGATG ATTCCCTTG TTTTCGACCT
 BssHII
 5' 3'
 3' 5'
 EcoRI
 5' 3'
3921 ATTCACAAT GAAACAATAA AAGATTAAJA TAGTTGCCC CCGTTGCAGC GATGGGTATT TTTCTAGTA
 TAAGGTGTTA CTTGTTATTA TTCTAATTTT ATCGAACGGG GCGAACGTCT CTACCCATTA AAAGATCAT
3991 AATATAAGA TAACTTAGA CTCAAAACAT TTACAAAAC AACCCCTAAA GTCTCTAAGC CCAAGTGTCT
 TTTATTTTCT ATTTGAATCT GAGTTTGTGA AATGTTTTG TTGGGGAATT CAGGATTCG GGTTCACGA
4061 ATGCACGATC CATAGCAAGC CCAGCCCAAC CCACCCCAAC CCAGTGCAGC CACTGGCA
 TACGTGCTAG GTATCGTTCT GGTGCGGTTG GGTGCGGTTG GGTACACGTCT GTTGACCGTT
4131 ATAGTCTCCA CCCCCTGAC TATCACCTGT AGTTGTCTGC ACCACCTGAC GTCTCGCAGC CAAAAA
 TATCAGAGGT GGGGCGCGTG ATAGTGCAC TCACACAGCG TGGTGGCGTG CAGAGCGTCT GTTTTTTTT
4201 AAAAGAAAGA AAAAAGAA AAAGAAAGAA AGCAGGTGGG TCCGGGTCTGT GGGGGCCGGA AAAGCGAGGA
 TTTTCTTTCT TTTTCTTTCT TTTCTTTTG TCGTCCACCC AGCCCCAGCA CCCCCGCCCT TTTCTGCTCT
4271 GGATCGGAG CAGCGACGAG GCCCCGCCCT CCGTCCGCTT CCAAGAAAC GCCCCCCTAT GCCACTATAT
 CCTAGCGCTC GTCGCTGCTC CCGGCCGGGA GGGAGGCGAA GGTTCCTTG CCGGGGGTAG CGGTGATATA
4341 ACATACCCCC CCCTCTCTCT CCATCCCCC AACCTTACCA CCACCACTCC CACCCCTCTG TCCCCCTCTG
 TGTATGGGG GGGAGAGGAG GGTAGGGGG TTGGATGAT GTGTGTGAGG AGGGGGGAGC
4411 CTGCGGACG ACGAGCTCTT CCCCCCTCC CCGCGGCGC CCGCGGTAA CACCCCGCCC CTCCTCTCTT
 GACGGCTGCT TGCTCGAGGA GGGGGGAGGG GAGGGCGGG GCGGCCATTG GTGGGGCGGG GAGAGGAGAA
4481 TCTTTCTCCG TTTTCTTTT TGTCTCGGTC TCGATCTTTG GCCTTGGTAG TTTGGGTGG CAGAGCGGCG
 AGAAGAGGG AAAAAAAA GCAGAGCCAG AGCTAGAAAC CCGAACCATC AAACCCNCCC GCTCTGCGG
4551 TTCGTGCCC AGATCGGTGC GCGGGAGGG GGGATCTCG CCGGTGCGCT CTCGGGCGT GAGTCGGCCC
 AAGCAGCGGG TCTAGCCACG CCCCCCTCCC GCGCTAGAGC GCGAACCGCA GAGCCCCGCA CTCAGCCCGG

```

FIG.\_41G

	BamHI	BglII	
4621	GGATCCTCGC GGGGAATGGG GCTCTCGGAT GTAGATCTTC TTCTTTCTTT CTTTGTGTGG TAGAATTGGA		
	CCTAGGAGCG CCCCTTACCC CGAGAGCCTA CATCTAGAGG AAGAAGAA GAAAGACACC ATCTTAAACT		
4691	ATCCCTCAGC ATTGTTTCATC GGTAGTTTTT CTTTTCATGA TTGTGACAA ATGCAGCCTC GTGCGGAGCT		
	TAGGAGATCG TAACAAGTAG CCATCAAAA GAAAGTACT AAACACTGTT TACGTCGGGG CACGCCCTCGA		
4761	TTTTGTAGG TAG		
	AAAACATCC ATC		

**FIG.\_41H**

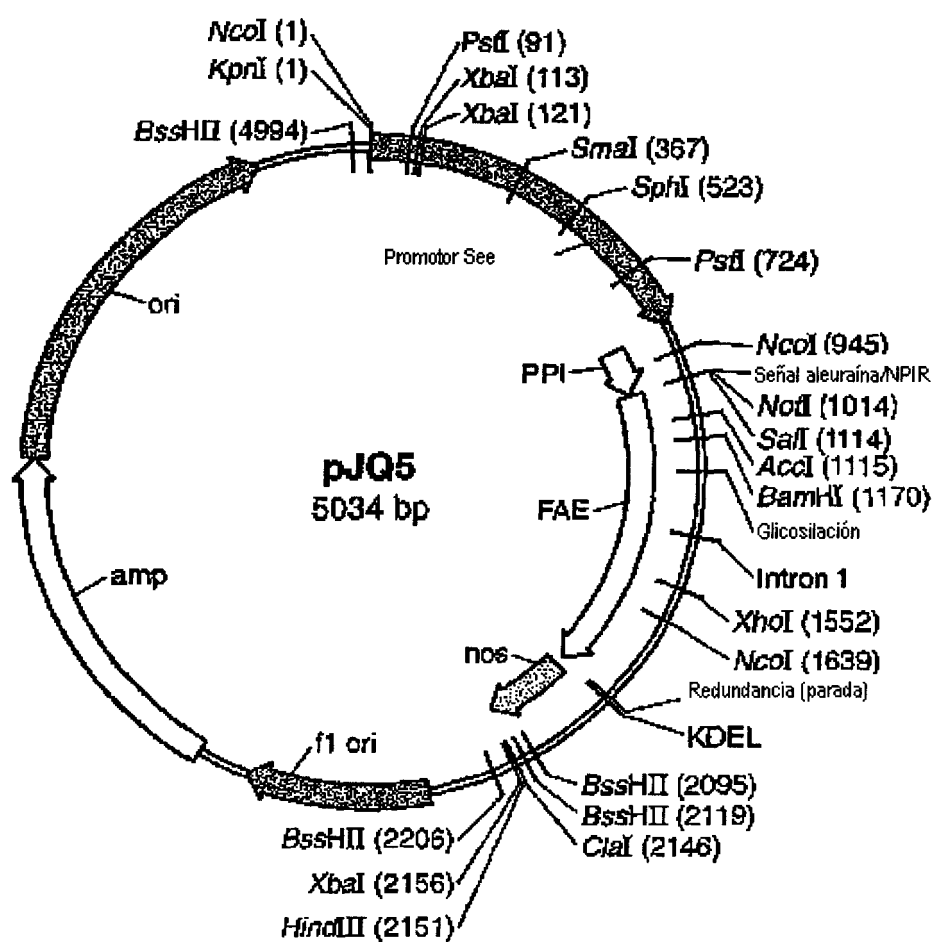


**FIG.\_42A**

```

NcoI

KpnI

1 CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAATATA TCGAATATC ACCATTGGCT ACAATATCTG
 GTACCCGGTC CATATTATAA CCTATAGAG TTCGTTTATT AGCTTATAG TGGTAACCGA TGTATATAGC

 PstI

71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGTG ATCTAGAACT CTAGATAGCA CAGCCACAGC
 TCGAGGCTCA AGACTGACGT CAGACCTACT GCGCACACAA TAGATCTTGA GATCTATCGT GTCGGTGTCTG

141 ACCTACAGGA GTCCGACACT TGTGGAAGTGT AGTAGTGTGG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
 TGGATGTCTT CACGCTGTGA ACACCTGACA TCATCACACAC CTCGCTCTG AGAAGGATG GAGGACTGCA

211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGGCTC CCAACAAAT ATCTGTCCTCC
 ACGCGGCCAA CAGGTAAGGT TGCCGTAGTG AGAGTTGGTT AGTCCGGAG GGTGTGTTTA TAGCAGGGGG

281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTGCGGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
 TACAGAACCG CCTCTCTCTC ATGTATGTAC GACAGCGCGG CAAAACAGA CTTAGAGCGA AGGTGACCGG

 SmaI

351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCTCTG TCACCCCTGG CGTCAATGGG
 TTAGTCGAGT CGAGGGCCCT CGAGTGAGTA AGTTCTAGGG TAGCAGCAGC AGTGGGGACC GCAGTACCTT

421 TGGAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
 ACCTTTCTT GGAGGCAACG AGCCTACTCA GTCGGTATAG GGGCTGTCT CATGACGTTT TATTGGGTTA

 SphI

491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTG CTTAATTGAC TTTATTTTGG
 AGCTAAGGG GGTATATCTT TTCAATCTGT ACGAAGGCC AAAACAAAC GAATTAACG AATATAAAC

561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGCTGG
 AACCTCAACT TAGGACTAAA CAACACATTT TACGGGTTGG TAGACTTATA GCTCTGCCTA TTATCCGACC

```

FIG.\_42B

```

631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
 GATTAATTAA ATATCGTTCT AAGACATCAC GTGTAGCGTT TATAGAAAGA CCCGTAATGT CGACCTCCGA

 PstI
                                ~~~~~
701  TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAG
    AGTAGTCGGA CTTTGTGAGA CGTCTCGGAC TTCGTTCAAC ACTTCGCACC GCTACTCTAC CCATATTTTG

771  CCCCCGCACC GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCCTGC TCCCTCTGCC GACCGACCCA
    GGGGCCGTGG CCTGCGCTC GAGGGCGGAT GGTCAATGTA GAGCGGAGCG AGGGGACGCG CCTGCTGGGT

841  GTAAATACT GTTCCCACT CGCCGCGGAG ATGCMCGTGC ACAAGGAGGT SAACCTCGTS GCCTACCTCC
    CATTTTATGA CAACGGGTGA GCGGCCGCTC TACCKGCACG TGTTCCTCCA STGAAGCAB CGGATGGAGG

                                NotI
                                ~~~~~
911 TGATCGTSTC CGCCTCCTC TTGCTCGTST CCGCCATGGA GCACGTGGAC GCCAAGGCTT GCACCCKCGA
 ACTAGCABGA GCGGAGGAG AACGAGCABA GCGGTACCT CGTGACCTG CGGTTCCGGA CGTGGGAGCT

 NotI
                                ~~~~~
981  GTGCGGCAAC CTCGGCTTCG GCATCTGCCC GCGCGCCGCC TCCACGCAGG GCATCTCCGA AGACCTCTAC
    CACGCCGTTG GAGCCGAAGC CGTAGACGGG CCGCCGCGCG AGGTGCGTCC CGTAGAGGCT TCTGGAGATG

                                SalI
                                ~~~~~
 AclI
                                ~~~~~
1051 AGCCGTTTAG TCGAATGGC CACTATCTCC CAAGTGCCT ACGCCGACCT GTGCAACATT CCGTCGACTA
    TCGGCAATC AGCTTTACCG GTGATAGAGG GTTCGACGGA TCGGGCTGGA CACGTTGTAA GGCAGCTGAT

                                BamHI
                                ~~~~~
1121 TTATCAAGG AGAGAAATT TACAATTCTC AAATGACAT TAACGGATGG ATCTCCCGG ACGACAGCAG
 AATAGTTCCC TCTCTTTTAA ATGTTAAGG TTTGACTGTA ATTGCCTACC TAGGAGGCC TGTGTCTGTC

```

FIG. 42C

```

1191 CAAAGAAATA ATCACCGTCT TCGTGGCAC TGGTAGTGAT ACGAATCTAC AACTCGATAC TAACTACACC
 GTTCTTTTAT TAGTGGCAGA AGGCACCGTG ACCATCACTA TGCTTAGATG TTGAGCTATG ATTGATGTGG

1261 CTCACGCGCTT TCGACACCGT ACCACAATGC AACGGTTGTG AAGTACACGG TGGATATTAT ATTGGATGGG
 GAGTGGGAA AGCTGTGGGA TGGTGTACG TTGCCAACAC TTCAATGTGCC ACCTATAATA TAACTACCCC

1331 TCTCGGTCCA GGACCAAGTC GAGTCGCTTG TCAAAACAGCA GGTTAGCCAG TATCCGGACT ACGCGCTGAC
 AGAGGCAGGT CCTGGTTTCAG CTCAGCGAAC AGTTTGTGCT CCAAATCGGT ATAGGCTGGA TCGCGCACTG

1401 CGTGACCGGC CACKCCCTCG GCGCCTCCCT GCGGGCACTC ACTGCCGCC AGCTGTCTGC GACATACGAC
 GCACGTGGCG GTGNGGGAGC CCGGAGGGGA CCGCCGTGAG TGACGGCGGG TCGACAGACG CTGTATGTCTG

1471 AACATCCGCC TGTACACCTT CGGCGAACCG CGCAGCGGCA ATCAGGCTTT CGCGTCGTAC ATGLACGATG
 TTGTAGGCGG ACATGTGGAA GCGCTTGGC GCGTCGCCGT TAGTCCGGAA GCGCAGCATG TACTTGCTAC

 XhoI
      ~~~~~
1541 CCTTCCAAGC CTCGAGCCCA GATACGACGC AGTATTTCG GGTCACTCAT GCCACGACG GCATCCCAAA
      GGAAGGTTG GAGCTCGGT CTATGCTCG TCATAAAGGC CCAATGAGTA CGGTTGCTGC CGTAGGGTTT

      NgoI
      ~~~~~
1611 CCTGCCCCCG GTGGAGCAGG GGTACGCCCA TGGCGGTGTA GAGTACTGGA GCGTTGATCC TTACAGCGCC
 GGACGGGGC CACCTCGTCC CCATGCGGT ACCGCCACAT CTCATGACCT CGCAACTAGG AATGTCGCGG

1681 CAGAACACAT TTGTCTGCAC TGGGGATGAA GTGCAGTGCT GTGAGGCCCA GGGCGGACAG GGTGTGAATA
 GTCATTGTGA AACAGACGTG ACCCTACTT CAGGTCACGA CACTCGGGT CCGCCTGTC CCACACTTAT

1751 ATGCCGACAC GACTTATTTT GGGATGACGA GCGGAGCCTG TACATGTTGA TCAGTCATTT CAGCCTCCCC
 TACGCGGTG CTGAATAAA CCCTACTGCT CGCCTCGGC ATGTACCACT AGTCAGTAAA GTCGGAGGGG

1821 GAGGTACCA GGAAGATGG ATGTCTTGA GAGGGGGCG CGTAACCACT GAAGGATGAG CTGTAAAGAA
 CTCACATGGT CCTTCTACC TACAGGACCT CTCCCCCGGC GCATTGGTGA CTTCTTACTC GACATTTCTT

```

**FIG.\_42D**

1891 GCAGATCGTT CAAACATTTG GCAATAAAGT TTCTTAAGAT TGAATCCTGT TGCCGGTCTT GCGATGATTA  
CGTCTAGCAA GTTTGTAAAC CGTTATTTCA AAGAATCTA ACTTAGGACA ACGGCCAGAA CGCTACTAAT

1961 TCATATAATT TCTGTTGAAT TACGTTAAGC ATGTAAATAT TAACATGTAA TGCATGACGT TATTATGAG  
AGTATATTAA AGACAACTTA ATGCAATTCC TACATATTAA ATTGTACATT ACCTACTGCA ATAAATACTC

2031 ATGGGTTTT ATGATTAGAG TCCGCAATT ATACATTAA TAGCGGATAG AAAACAAAAT ATAGCCGCA  
TACCCAAAA TACTAATCTC AGGGCGTTAA TAATGAAAT ATGCGCTATC TTTTGTTTA TATCGCGCGT

2101 AACTAGGATA AATTATCGC GCGGTGTCA TCTATGTTAC TAGATCGATA AGCTTCTAGA GCGGCCGGTG  
TTGATCCTAT TTAATAGCG CCGCCACAGT AGATACAATG ATCTAGCTAT TCGAAGATCT GCGCGGCCAC

2171 GAGCTCCAAT TCGCCCTATA GTGAGTCGT TTAGCGCGC TCACTGGCCG TCGTTTACA ACGTCGTGAC  
CTCGAGGTTA AGCGGATAT CACTCAGCAT AATCGCGCG AGTGACCGGC AGCAAAATGT TGCAGCACTG

2241 TGGGAAAACC CTGGCGTTAC CCAACTTAAT CGCTTGCAG CACATCCCC TTTGCCCAGC TGGCGTAATA  
ACCCTTTGG GACCGCAATG GGTGMAATTA GCGGAACGTC GTGTAGGGGG AAGCGGTCTG ACCGCATTAT

2311 GCGAAGAGGC CCGCACCGAT CGCCCTTCCC AACAGTTGCG CAGCCTGAAT GCGGAATGGG ACGCGCCCTG  
CGCTTCTCCG GCGGTGGCTA GCGGGMAGGG TTGTCAACGC GTCGGACTTA CCGCTTACCC TCGCGGGGAC

2381 TAGCGGCGCA TTAAGCGCG GGGGTGTGGT GGTACGCGC AGCGTGACCG CTACACTTGC CAGCGCCCTA  
ATCGCGCGT AATTCGCGC GCCCACCA CCAATGCGC TCGCACTGGC GATGTGAACG GTCGCGGGAT

2451 GCGCCCGCTC CTTTCGCTTT CTTCCCTTCC TTCTCGCCA CGTTCCGCCG CTTTCCCGGT CAAGCTCTAA  
CGCGGCGGAG GAAAGCGAAA GAAGGGMAGG AAGAGCGGT GCAAGCGGCC GAAGGGGCA GTTCGAGATT

FIG..42E

2521 ATCGGGGGCT CCTTTAGGG TTCCGATTTA GTGCTTTACG GCACCTCGAC CCCAAAAAAC TTGATTAGGG  
 TAGCCCCCGA GGGAAATCCC AAGGCTAAT CACGAAATGC CGTGGAGCTG GGGTTTTTTG AACTAATCCC  
 2591 TGATGGTTCA CGTAGTGGG CATCGCCCTG ATAGACGGTT TTTCGCCCTT TGACGTGGGA GTCCACGTT  
 ACTACCAAGT GCATCACCCG GTAGCGGAC TATCTGCCA AAGCGGGAA ACTGCAACCT CAGGTGCAAG  
 2661 TTTAATAGTG GACTCTTGTT CCAACTGGA ACAACACTCA ACCCTATCTC GGTCTATTCT TTTGATTAT  
 AAATTATCAC CTGAGAACAA GGTTTGACCT TGTGTGAGT TGGGATAGAG CCAGATAAGA AACTAATA  
 2731 AAGGGATTTT GCCGATTTCG GCCTATTGGT TAAAAAATGA GCTGATTTAA CAAAAATTTA ACGCGAATTT  
 TTCCCTAAAA CGGCTAAAGC CGGATAACCA ATTTTACT CGACTAAAT GTTTTAAAT TGCGCTTAAA  
 2801 TAAACAAATA TTAACGCTTA CAATTAGGT GGCACCTTTC GGGGAATGT CGCGGGAACC CCTATTGGTT  
 ATTGTTTTAT AATTGCGAAT GTTAAATCCA CCGTGAAAG CCCCTTTACA CGCGCTTGG GATAAACAA  
 2871 TATTTTCTA AATACATTCA AATATGTATC CGCTCATGAG ACAATAACCC TGATAAATGC TTCAATAATA  
 AAAAAAGAT TTATGTAAGT TTATACATAG GCGAGTACTC TGTATTGGG ACTATTACG AAGTTATTAT  
 2941 TTGAAAAAGG AAGAGTATGA GTATTCAACA TTTCGGTGTG GCCCTTATTC CCTTTTTCG GGCATTTTTC  
 AACTTTTTTC TTCTCATACT CATAAAGTTGT AAAGGCACAG CCGGAATAAG GGAATAAACG CCGTAAACG  
 3011 CTTCCCTGTTT TTGCTCACCC AGAAACGCTG GTGAAAGTAA AAGATGCTGA AGATCAGTTG GGTGCACGAG  
 GAAGGACAAA AACGAGTGGG TCTTTGCGAC CACTTTCATT TTCTACGACT TCTAGTCAAC CCACGTGCTC  
 3081 TGGGTTACAT CGAACTGGAT CTCAACAGCG GTAAGATCCT TGAGAGTTTT CGCCCCGAAG AACGTTTTTC  
 ACCCAATGTA GCTTGACCTA GAGTTGTGCG CATCTAGGA ACTCTCAAA GCGGGGCTTC TTGCAAAAGG  
 3151 AATGATGAGC ACTTTTAAAG TTCTGCTATG TGGCGCGGTA TTATCCCGTA TTGACGCCGG GCAAGAGCAA  
 TTACTACTCG TGAANAATTC AAGACGATAC ACCGCGCCAT AATAGGGCAT AACTGCGGCC CGTTCTCGTT  
 3221 CTCGGTCGCC GCATACACTA TTCTCAGAAAT GACTTGGTTG AGTACTCACC AGTCACAGAA AAGCATCTTA  
 GAGCCAGCGG CGTATGTGAT AAGAGTCTTA CTGAACCAAC TCATGAGTGG TCAGTGTCTT TTCGTAGAAT  
 3291 CGGATGGCAT GACAGTAAGA GAATTATGCA GTGCTGCCAT AACCATGAGT GATAACACTG CGGCCAACTT  
 GCCTACCGTA CTGTCAATTCT CTTAATACGT CACGACGGTA TTGGTACTCA CTATTGTGAC GCCGGTTGAA

FIG..42F

3361 ACTTCGTGACA ACGATCGGAG GACCGAAGGA GCTAACCGCT TTTTGCACA ACATGGGGGA TCATGTAACT  
 TGAAGACTGT TGCCTAGCCTC CTGGCTTCCT CGATTGGCGA AAAACGTGT TGTACCCCTT TGTACATTGA  
 3431 CGCCTTGATC GTTGGGAACC GGAAGTGAAT GAAGCCATAC CAAACGACGA GCGTGACACC ACGATGCCCTG  
 GCGGAACCTAG CAACCCCTGG CCTCGACTTA CTTCGGTATG GTTGTCTGCT CGCACTGTGG TGCTACGGAC  
 3501 TAGCAATGGC AACAACTGTG CGAAACTAT TAAGTGGGA ACTACTTACT CTAGCTTCCC GGCACCAATTT  
 ATCGTTACCG TTGTTGCAAC GCCTTTGATA ATTGACCGCT TGATGAATGA GATCGAAGGG CCGTTGTATA  
 3571 AATAAGCTGG ATGGAGGCGG ATAAAGTTGC AGGACCACCTT CTGGCTCGG CCTTCCGGC TGGCTGGTTT  
 TTATCTGACC TACCTCGCC TATTTCAACG TCCTGGTGA GACGCGAGCC GGGAGGCCG ACCGACCAAA  
 3641 ATTGCTGATA AATCTGGAGC CGGTGAGCGT GGTCTCGCG GTATCATTCG AGCACTGGGG CCAGATGGTA  
 TAACGACTAT TTAGACCTCG GCCACTCGCA CCCAGAGCGC CATAGTAACG TCGTGACCCC GGTCTACCAT  
 3711 AGCCCTCCCG TATCGTAGTT ATCTACACGA CGGGGAGTCA GGCAACTATG GATGAACGAA ATAGACAGAT  
 TCGGGAGGGC ATAGCATCAA TAGATGTGCT GCCCTCAGT CCGTTGATAC CTACTTGCTT TATCTGTCTA  
 3781 CGCTGAGATA GGTGCCCTCAC TGATTAAGCA TTGGTAACTG TCAGACCAAG TTTACTCATA TATACTTTAG  
 GCGACTCTAT CCACGGAGTG ACTAATTCGT AACCATTTGAC AGTCTGGTTC AAATGAGTAT ATATGAAATC  
 3851 ATTGATTTAA AACTTCATTT TTAATTTAAA AGGATCTAGG TGAAGATCCT TTTTGATAAT CTCATGACCA  
 TAACTAAATT TTGAAGTANA AATTAATTT TCCTAGATCC ACTTCTAGGA AAAACTATTA GAGTACTGGT  
 3921 AAATCCCTTA ACGTGAGTTT TCGTTCCACT GAGCGTCAGA CCCCCTAGAA AAGATCAAG GATCTTCTTG  
 TTTAGGGAAT TGCACTCANA AGCAAGGTGA CTCGCAGTCT GGGGCATCTT TTCTAGTTTC CTAGAGAAC  
 3991 AGATCCCTTT TTTCTGCGCG TAATCTGCTG CTTCGAAACA AAAAACCAC CGCTACCAGC GGTGGTTTGT  
 TCTAGGAAA AAGACGCGC ATTAGACGAC GAACGTTTGT TTTTGTGGTG GCGATGGTGG CCACCAAAAC  
 4061 TTGCCGGATC AAGAGCTACC AACTCTTTT CCGAAGGTAA CTGGCTTCAG CAGAGCGCAG ATACCAATA  
 AACGGCCTAG TTCTCGATGG TTGAGAAAAA GGCTTCCATT GACCGAAGTC GTCTCGCGTC TATGGTTTAT  
 4131 CTGTCTTCTT AGTGTAGCCG TAGTTAGGCC ACCACTTCAA GAATCTGTA GCACCGCCTA CATACCTGCG  
 GACAGGAGA TCACATCGGC ATCAATCCGG TGGTGAAGTT CTTGAGACAT CGTGGCGGAT GTATGGAGCG

**FIG..42G**

4201 TCTGCTAATC CTGTTACCAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGACTCAAGA  
 AGACGATTAG GACAATGGTC ACCGACGACG GTACCCGCTA TTCAGCACAG AATGGCCCA CCTGAGTTCT  
 4271 CGATAGTTAC CGGATAAGGC GCAGCGGTGC GGTGAACCG GGGTTGCTG CACACAGCCC AGCTTGGAGC  
 GCTATCAATG GCCTATTCCG CGTCGCCAGC CCGACTTGCC CCCCAGGCAC GTGTGTGCGG TCGAACCTCG  
 4341 GAACGACCTA CACCGAATG AGATACCTAC AGCTGAGCT ATGAGAAAGC GCCACGCTTC CCGAAGGGAG  
 CTTGCTGGAT GTGGCTTGAC TCTATGGATG TCGCACTCGA TACTCTTTCG CGGTGCGAAG GGCTTCCCTC  
 4411 AAGGCGGAC AGGTATCCGG TAAGCGGCGG GGTGCGAACA GGAGAGCGCA CGAGGGAGCT TCCAGGGGGA  
 TTTCCGCCTG TCCATAGGCC ATTGCGCCGTC CCAGCCTTGT CCTCTCGCGT GCTCCCTCGA AGGTCCCCCT  
 4481 AACGCCCTGT ATCTTTATAG TCCTGTGCGG TTTGCGCCAC TCTGACTTGA GCGTCGATTT TTGTGATGCT  
 TTGCGGACCA TAGAATATC AGGACAGCCC AAGCGGTGG AGACTGAACT CGCAGCTAAA AACACTACGA  
 4551 CGTCAGGGGG GCGGAGCCTA TGGAAAAACG CCAGCAACGC GGCCTTTTA GGGTTCCTGG CCTTTTGGTG  
 GCAGTCCCCC CGCTTCGGAT ACCTTTGTGC GGTGTTGCG CCGAAAAAT GCCAAGGACC GGAAACCGAC  
 4621 GCCTTTGTCT CACATGTTCT TTCCTGCGTT ATCCCTGAT TCTGTGGATA ACCGTATTAC CGCCTTTGAG  
 CGGAAAAACGA GTGTACAAGA AAGGACGCAA TAGGGACTA AGACACCTAT TGGCATAATG GCGGAACTC  
 4691 TGAGCTGATA CCGCTCGCCG CAGCCGAACG ACCGAGCGCA GCGAGTCAGT GAGCGAGGAA GCGGAAGAGC  
 ACTCGACTAT GCGGAGCGGC GTCGGCTTGC TGGCTCGCGT CGCTCAGTCA CTCGCTCCTT CGCCTTCTCG  
 4761 GCCCAATACG CAACCCGCTT CTCCCCGCGC GTTGGCCGAT TCATTAAATG AGCTGGCACG ACAGGTTTCC  
 CCGGTTATGC GTTTGGCGGA GAGGGGCGCG CAACCGCTA AGTAATTACG TCGACCGTGC TGTCCAAAGG  
 4831 CGACTGGAAA GCGGGCAGTG AGCGCAACGC AATTAATGTG AGTTAGCTCA CTCATTAGGC ACCCCAGGCT  
 GCTGACCTTT CGCCCGTCAC TCGCGTTGCG TTAATTACAC TCAATCGAGT GAGTAATCCG TGGGGTCCGA  
 4901 TTACACTTTA TGCTTCCGGC TCGTATGTTG TGTGGAATTG TGAGCGGATA ACAATTTCAC ACAGGAACA  
 AATGTGAAT ACGAAGGCCG AGCATACAA ACACCTAAC ACTCGCCTAT TGTTAAAGTG TGTCTTTGT

FIG.\_42H

4971

BssHII

~~~~~

NcoI

KpnI

~~~~~

GCTATGACCA

TGATTACGCC

AAGCGCGCAA

TTAACCCCTCA

CTAAAGGGAA

CAAAAGCTGG

GTAC

CGATACTGGT

ACTAATGCGG

TTCGCGCGTT

AATTGGGAGT

GATTTCCTT

GTTTTCGACC

CATG

FIG.\_42I

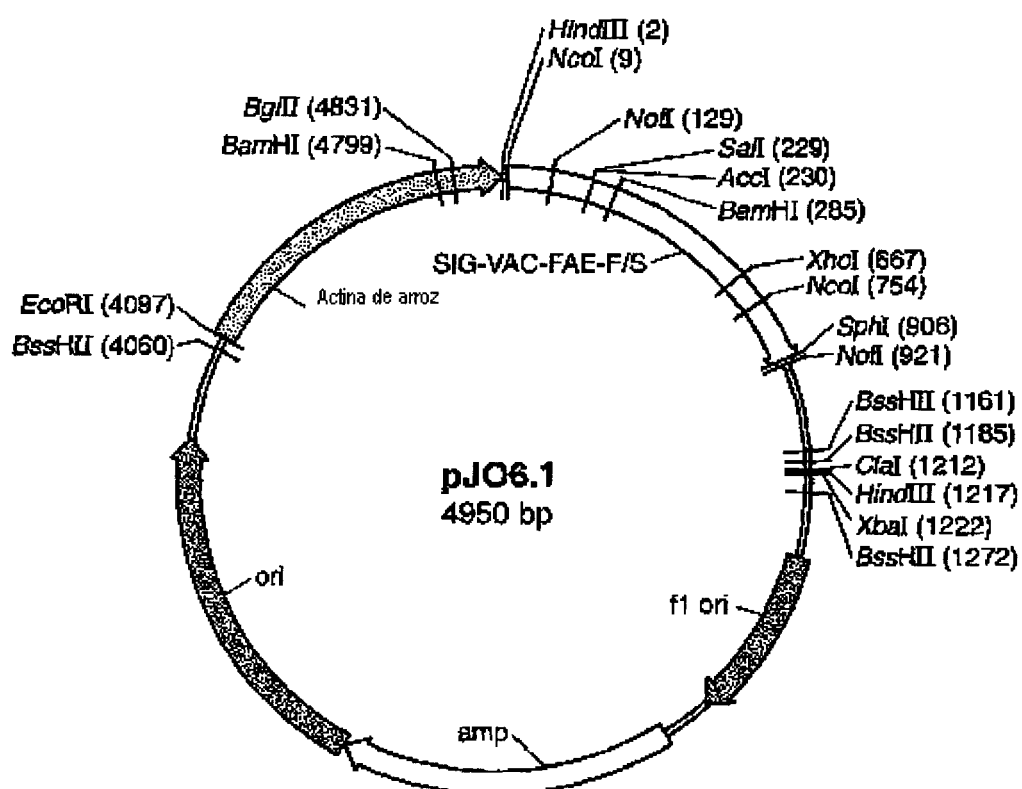


**FIG. 43A**

```

HindIII NcoI
~~~~~
1  AAGCTTACCA TGGCCCAAGC CCGGTCTC CTCCTGGGC TCCTCGTCTT GGCACGGCC GCGTCCCG GCGCAGCGGC
   TTCGAATGGT ACCGGGTGCG GCGCGCAGGAG GAGGACCGCG AGCGGCACGA CCGTCCCG GCGCAGCGGC

71  TCGCCTCCTC CTCCTCCTTC GCCGACTCCA ACCCGATCCG GCCCGTCACC GACCGCGCG GCGCCTCCAC
   AGCGGAGGAG GAGGAGGAG CCGCTGAGGT TGGCTAGGC CCGGCAGTGG CTGGCGCGC GCGCGAGGTG

141 GCAGGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC
   CGTCCCCTAG AGGCTTCTGG AGATGTCCGC AAATCAGCTT TACCGGTGAT AGAGGTTCG ACGGATGCCG

                               SalI
                               ~~~~~
 AccI
                               ~~~~~

211 GACCTGTGCA ACATTCGGTC GACTATTATC AAGGGAGAGA AAATTTACAA TTCTCAAACT GACATTAACG
   CTGGACACGT TGTAAGGCAG CTGATAATAG TTCCCTCTCT TTTAAATGTT AAGAGTTTGA CTGTAATTGC

                               BamHI
                               ~~~~~

281 GATGGATCCT CCGCGACGAC AGCAGCAAG AAATAATCAC CGTCTTCCGT GGCACITGTA GTGATACGAA
 CTACCTAGGA GCGGCTGCTG TCGTCGTCTT TTTATATAGT GCAGNAGGCA CCGTGNCCAT CACTATGCTT

351 TCTACAACTC GATACTAACT ACACCCCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAGTA
 AGATGTTGAG CTATGATGA TGTGGGAGTG CGGAAGCTG TGGGATGGTG TTACGTTGCC AACACTTCAT

421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA
 GTGCCACCTA TAATATAACC TACCCAGAGG CAGGTCTTGG TTCAGCTCAG CGAACAGTTT GTCGTCCAAT

491 GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACAC CCGTGGCGCC TCCCTGGCGG CACTCACTGC
 CCGTCATAGG CCTGATGCGC GACTGGCACT GCGCGGTGAG GGAGCCCGCG AGGACCCGCC GTGAGTGACG

561 CGCCCAAGCTG TCTGCGACAT ACGACAACAT CCGCTGTATC ACCTTCGGCG AACCGCGCAG CCGCAATCAG
 GCGGGTCGAC AGACGCTGTA TGCTGTGTA GCGGACATG TGGAGCCCGC TTGGCGCGTC GCGGTTAGTC

```

**FIG..43B**

**XhoI**  
 .....  
 631 GCCTTCGCGT CGTACATGAA CGATGCCCTTC CAAGCCTCGA GCCAGATAC GACGAGTAT TTCCGGGTCA  
 CGGAAGCGCA GCATGTACTT GCTACGGAAG GTTCGGAGCT CGGCTCTATG CTGGTCAATA AAGGCCCAGT  
  
 .....  
 701 CTCAATGCCAA CGACGGGATC CCAAACTGC CCCCAGTGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
 GAGTACGGTT GCTGCCGTAG GGTTCGACG GGGGCCACCT CGTCCCCTAG CGGTACCGC CACATCTCAT  
  
 .....  
 771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCACGGG ATGAAGTGA GTGCTGTGAG  
 GACCTCGCA CTAGGAATGT CCGGGTCTT GTGTAAACAG ACGTACCCC TACTTCAGCT CACGACACTC  
  
 .....  
 841 GCCCAGGGCG GACAGGGTGT GAATAATGC CACAGACTT ATTTGGGAT GACGAGCGC GCATGCACCT  
 CCGGTCCCGC CTGTCCCA CA CTATATACG GTGTCTGA TAAACCTA CTGCTCGCG CGTACGTGA  
  
 .....  
 911 GGCCGGTCGC GGCCGGGAA ACCACTGAAG GATGAGCTGT AAGAAGCAG ATCCTCAAA CATTGGCAA  
 CCGGCCAGCG CCGCGCCCTT TGGTACTTC CTACTCGACA TTCTTCGTC TAGCAAGTTT GTAAACCGTT  
  
 .....  
 981 TAAAGTTCT TAAGATTGAA TCCTGTTGCC GGTCTGCGA TGAATATCAT ATATTTCTG TTGAATTACG  
 ATTTCAAAGA ATTCTAACTT AGGACAACGG CCAGAACGCT ACTAATAGTA TATTAAGAC AACTTAATGC  
  
 .....  
 1051 TTAAGCATGT AATAATTAC ATGTAATGA TGACGTTATT TATGAGTGG GTTTTATGA TTAGAGTCCC  
 AATTCGTACA TTATTAATTG TACATTAGT ACTGCAATAA ATACTCTACC CAAAATACT AATCTCAGGG  
  
 .....  
 1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAATATAG CGCGCAACT AGGATAAAT ATCGCGCGCG  
 CGTTAATATG TAAATTATGC GCTATCTTT GTTTATATC GCGGTTGA TCCTATTTAA TAGCGCGCGC

**NotI**  
 .....  
 911 GGCCGGTCGC GGCCGGGAA ACCACTGAAG GATGAGCTGT AAGAAGCAG ATCCTCAAA CATTGGCAA  
 CCGGCCAGCG CCGCGCCCTT TGGTACTTC CTACTCGACA TTCTTCGTC TAGCAAGTTT GTAAACCGTT  
  
 .....  
 981 TAAAGTTCT TAAGATTGAA TCCTGTTGCC GGTCTGCGA TGAATATCAT ATATTTCTG TTGAATTACG  
 ATTTCAAAGA ATTCTAACTT AGGACAACGG CCAGAACGCT ACTAATAGTA TATTAAGAC AACTTAATGC  
  
 .....  
 1051 TTAAGCATGT AATAATTAC ATGTAATGA TGACGTTATT TATGAGTGG GTTTTATGA TTAGAGTCCC  
 AATTCGTACA TTATTAATTG TACATTAGT ACTGCAATAA ATACTCTACC CAAAATACT AATCTCAGGG  
  
 .....  
 1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAATATAG CGCGCAACT AGGATAAAT ATCGCGCGCG  
 CGTTAATATG TAAATTATGC GCTATCTTT GTTTATATC GCGGTTGA TCCTATTTAA TAGCGCGCGC

**SphI**  
 .....  
 841 GCCCAGGGCG GACAGGGTGT GAATAATGC CACAGACTT ATTTGGGAT GACGAGCGC GCATGCACCT  
 CCGGTCCCGC CTGTCCCA CA CTATATACG GTGTCTGA TAAACCTA CTGCTCGCG CGTACGTGA  
  
 .....  
 911 GGCCGGTCGC GGCCGGGAA ACCACTGAAG GATGAGCTGT AAGAAGCAG ATCCTCAAA CATTGGCAA  
 CCGGCCAGCG CCGCGCCCTT TGGTACTTC CTACTCGACA TTCTTCGTC TAGCAAGTTT GTAAACCGTT  
  
 .....  
 981 TAAAGTTCT TAAGATTGAA TCCTGTTGCC GGTCTGCGA TGAATATCAT ATATTTCTG TTGAATTACG  
 ATTTCAAAGA ATTCTAACTT AGGACAACGG CCAGAACGCT ACTAATAGTA TATTAAGAC AACTTAATGC  
  
 .....  
 1051 TTAAGCATGT AATAATTAC ATGTAATGA TGACGTTATT TATGAGTGG GTTTTATGA TTAGAGTCCC  
 AATTCGTACA TTATTAATTG TACATTAGT ACTGCAATAA ATACTCTACC CAAAATACT AATCTCAGGG  
  
 .....  
 1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAATATAG CGCGCAACT AGGATAAAT ATCGCGCGCG  
 CGTTAATATG TAAATTATGC GCTATCTTT GTTTATATC GCGGTTGA TCCTATTTAA TAGCGCGCGC

**BssHII**  
 .....  
 911 GGCCGGTCGC GGCCGGGAA ACCACTGAAG GATGAGCTGT AAGAAGCAG ATCCTCAAA CATTGGCAA  
 CCGGCCAGCG CCGCGCCCTT TGGTACTTC CTACTCGACA TTCTTCGTC TAGCAAGTTT GTAAACCGTT  
  
 .....  
 981 TAAAGTTCT TAAGATTGAA TCCTGTTGCC GGTCTGCGA TGAATATCAT ATATTTCTG TTGAATTACG  
 ATTTCAAAGA ATTCTAACTT AGGACAACGG CCAGAACGCT ACTAATAGTA TATTAAGAC AACTTAATGC  
  
 .....  
 1051 TTAAGCATGT AATAATTAC ATGTAATGA TGACGTTATT TATGAGTGG GTTTTATGA TTAGAGTCCC  
 AATTCGTACA TTATTAATTG TACATTAGT ACTGCAATAA ATACTCTACC CAAAATACT AATCTCAGGG  
  
 .....  
 1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAATATAG CGCGCAACT AGGATAAAT ATCGCGCGCG  
 CGTTAATATG TAAATTATGC GCTATCTTT GTTTATATC GCGGTTGA TCCTATTTAA TAGCGCGCGC

**FIG. 43C**

**XbaI**  
**Clal HindIII**  
**BssHII**

1191 GTGTCATCTA TGTTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC TCCTAATTCG CCTATAGTGA  
CACAGTAGAT ACAATGATCT AGCTATTGGA AGATCTCGCC GCCCACCCTG AGGTTAAGCG GGATATCACT

1261 GTCGTATTAC GCGCGCTCAC TGGCCGTCTGT TTTACAAACGT CGTGACTGGG AAATCCCTGG CGTTACCCAA  
CAGCATAATG CCGCGGAGTG ACCGGCAGCA AAATGTTGCA GCACTGACCC TTTTGGGACC GCAATGGGTT

1331 CTTAATCGCC TTGCAGCACA TCCCCCTTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC  
GAATTAGCGG AACGTCTGTGT AGGGGGAAG CGGTGACCG CATTATCGCT TCTCCGGCG TGGCTAGCGG

1401 CTTCCCAACA GTTGGCGCAG CTGAATGGCG AATGGGACGC GCCCTGTAGC GGCGCATTA GCGCGGCGGG  
GAAGGGTTGT CACGCGCTG GACTTACCGC TTACCCCTGG CCGGACATCG CCGCGTAATT CCGCGCGCCC

1471 TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTGGCCAGC GCCCTAGCG CGCTCCCTTT CGCTTCTTTC  
ACACCACCAA TCGCGCTCGC ACTGGCGATG TGAACGGTCT CCGGATCGCG GGCGAGGAAA GCGAAGAAG

1541 CCTTCCTTTC TCGCCACGTT CGCCGGCTTT CCGCGTCAAG CTCATAATCG GGGGCTCCCT TTAGGGTTCC  
GGAAGGAAG AGCGGTGCAA GCGGCCGAAA GGGGCACTTC GAGATTTAGC CCCCAGGGA AATCCCAAGG

1611 GATTTAGTGC TTTACGGCAC CTCGACCCCA AAAAATTGA TTAGGGTGA TTGTCACGTA GTGGGCCATC  
CTAAATCAG AATGCCCCTG GAGCTGGGGT TTTTGAACCT AATCCCACTA CCAAGTGCAT CACCCGGTAG

1681 GCCCTGATAG ACGGTTTTTC GCCCTTTGAC GTTGGAGTCC ACGTTCTTTA ATAGTGGACT CTTGTTCCAA  
CGGGACTATC TGCCAAAAG CGGGAAACTG CAACCTCAGG TCCAAGAAAT TATCACCTGA GAAACAGGTT

1751 ACTGGAACAA CACTCAACCC TATCTCGGTC TATCTTTTGG ATTTATAAGG GATTTTGGCG ATTTGGGCTT  
TGACCTTGTT GTGAGTTGGG ATAGAGCCAG ATAAGAAAAC TAAATATTCC CTAAACGGC TAAAGCCGA

1821 ATTGGTTAAA AATGAGCTG ATTTAACAAA AATTAAACG GAATTTAAC AAATATTAA CGCTTACAAAT  
TAACCAATTT TTTACTCGAC TAAATGTTT TAAATGCG CTTAAATTTG TTTTATAATT CGGAATGTTA

**FIG.-43D**

1891 TTAGGTGGCA CTTTTCGGGG AATGTGCGC GGAACCCCTA TTTGTTTATT TTCTAAATA CATTCAATA  
 AATCCACCGT GAAAGCCCC TTACACGCG CCTTGGGGAT AACAAATAA AAGATTAT GTAAGTTAT  
 1961 TGTATCCGCT CATGAGACAA TAACCCCTGAT AAATGCTTCA ATAATATTGA AAAAGGAAGA GTATAGTAT  
 ACATAGGCGA GTACTCTGTT ATTGGGACTA TTACGGAAGT TATTATAACT TTTCCTTCT CATACTCATA  
 2031 TCAACATTTC CGTGTGCCC TTATTCCCTT TTTTGGGCA TTTTGCCTTC CTGTTTTGCT TCACCCAGAA  
 AGTGTAAAG GCACAGCGG AATAAGGGA AAAACGCCGT AAAACGGAAG GACAAAAACG AGTGGGTCTT  
 2101 ACGCTGGTGA AAGTAAAGA TGCTGAAGT CAGTTGGGTG CACGAGTGGG TTACATCGAA CTGGATCTCA  
 TGCAGCCACT TTCATTTTCT ACGACTTCTA GTCAACCCAC GTGCTCACCC AATGTAGCTT GACCTAGAGT  
 2171 ACAGCGGTAA GATCCTTGAG AGTTTTCGCC CCGAAGAAGC TTTTCCAATG ATGAGCACTT TTAAGTTCT  
 TGTGCGCCATT CTAGGAATC TCAAAAGCGG GCTTCTTTCG AAAAGTTAC TACTCGTGA AATTTCAAGA  
 2241 GCTATGTGGC GCGGTATTAT CCGGTATTGA CGCCGGGCAA GAGCAACTCG GTCCGCCGTA TGTGATNAGA  
 CGATACACCG CGCCATAATA GGCATAACT CGGCCCGCTT CTCGTTGAGC CAGCGGCGTA TGTGATNAGA  
 2311 CAGAAAGACT TGGTTGAGTA CTCACCAATC GAGTGGTCAG TGTCTTTTCG TAGAATGCTT ACCGTACTGT CATTCTCTTA  
 GTCTTACTGA ACCAACTCAT GAGTGGTCAG TGTCTTTTCG TAGAATGCTT ACCGTACTGT CATTCTCTTA  
 2381 TATGCAGTGC TGCCATAACC ATGAGTGATA ACACGCGGC CAACCTTACTT CTGACACGA TCGGAGGACC  
 ATACGTCACG ACGGTATTGG TACTCACTAT TGTGACGCCG GTTGAATGA GACTGTTGCT AGCCTCCTGG  
 2451 GAAGGAGCTA ACCGCTTTT TGCACAACAT GGGGGATCAT GTAACTCGCC TTGATCGTTG GGAACCGGAG  
 CTTCTCGAT TGGCGAANA ACGTGTGTA CCCCCTAGTA CATTGAGCGG AACTAGCAAC CCTTGGCCTC  
 2521 CTGAATGAAG CCATACCAA CGACGAGCGT GACACCAGA TGCCCTGTAGC AATGGCAACA ACGTTGCGCA  
 GACTTACTTC GGTATGGTTT GCTGCTCGCA CTGTGTGCTT ACGACATCG TTACCGTTGT TGCNACGCGT  
 2591 AACTATTAACT TGGCGAATA CTTACTCTAG CTTCCCGGCA ACAATTATA GACTGGATGG AGCGGATAA  
 TTGATAATTG ACCGCTTGAT GAATGAGATC GAAGGCCGT TGTAAATTAT CTGACCTACC TCCGCCATTT  
 2661 AGTTGCAGGA CCACCTCTGC GCTCGGCCCT TCCGGCTGGC TGGTTATTG CTGATAAATC TGGAGCCGGT  
 TCAACGTCCT GGTGAAGACG CGAGCCGGA AGCCGACCG ACCAATAAC GACTATTTAG ACCTCGGCCA

**FIG. 43E**

2731 GAGCGTGGGT CTCGCGGTAT CATTCAGCA CTGGGGCCAG ATGGTAAGCC CTCCTGGTATC GTAGTTATCT  
 CTCGCACCCA GAGCGCCATA GTAACGTCTGT GACCCCGGTC TACCATTCCG GAGGGCATAG CATCAATAGA  
 2801 ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAAATAG ACAGATCGCT GAGATAGGTG CCTCACTGAT  
 TGTGCTGCC CTGAGTCCGT TGATACCTAC TTGCTTTATC TGCTATGCCA CTCATATCCAC GGAGTGACTA  
 2871 TAAGCATGG TAACTGTCTAG ACCAAGTTTA CTCATATATA CTTTAGATTG ATTTAAAACT TCATTTTAA  
 ATTCTGTAAC ATTGACAGTC TGCTTCAAAT GAGTATATAT GAAATCTAAC TAAATTTTGA AGTAAAAATT  
 2941 TTTAAAAAGGA TCTAGGTGAA GATCCTTTT GATAATCTCA TGACCAAAAT CCCTTAACGT GAGTTTTCGT  
 AAATTTTCTT AGATCCACTT CTAGGAAAA CTATTAGAGT ACTGGTTTIA GGAATTTGCA CTCAAAAAGCA  
 3011 TCCACTGAGC GTCAGACCCC GTAGAAAA GA TCAAAGGATC TTCTTGAGAT CCTTTTTC TCGCGTAAAT  
 AGGTGACTCG CAGTCTGGGG CATCTTTTCT AGTTTCCTAG AAGAACTCTA GGAATAAAAG ACGCCATTA  
 3081 CTGCTGCTTG CAAACAAAA AACCAACCGT ACCAGCGGTG GTTTGTTTGC CGGATCAAGA GCTACCAACT  
 GACGACGAAC GTTTGTTTTT TTGGTGGCGA TGGTGCGCCAC CAAACAAACG GCCTAGTTCT CGATGGTTGA  
 3151 CTTTTTCCGA AGGTAACTGG CTTCAGCAGA GCGCAGATAC CAAATACTGT CCTTCTAGTG TAGCCGTAGT  
 GAAAAAGGCT TCCATTGACC GAAGTCGTCT CGCGTCTATG GTTTATGACA GGAAGATCAC ATCGGCATCA  
 3221 TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCTACATA CCTCGCTCTG CTAATCCTGT TACCAGTGGC  
 ATCCGGTGGT GAAGTCTTG AGACATCTGT GCGGATGTAT GAGCGAGAC GATTAGGACA ATGGTCACCG  
 3291 TGCTGCCAGT GGCATAAGT CGTGTCTTAC CGGTTGGAC TCAAGACGAT AGTTACCGGA TAAGGCGCAG  
 ACGACGGTCA CCGCTATTCA GCACAGAA TG GCCCAACCTG AGTTCTGCTA TCAATGGCCT ATTCGCGTC  
 3361 CGGTCGGGT GAACGGGGG TTCGTGCACA CAGCCCAGCT TGGAGCGAAC GACCTACACC GAACGTGAGT  
 GCCAGCCCCA CTTGCCCCCC AAGCACGTGT GTCGGGTGCA ACCTCGCTTG CTGGATGTGG CTTGACTCTA  
 3431 ACCTACAGCG TGAGCTATGA GAAAGCGCCA CGCTTCCGA AGGAGAAAG CGGACAGGT ATCCGGTAAG  
 TGGATGTGCG ACTCGATACT CTTTCGCGGT CGGAAGGCT TCCCTCTTTC CGCCTGTCCA TAGGCCATTTC  
 3501 CGGCAGGTC GGAACAGGAG AGCGCACGAG GGAGCTTCCA GGGGAAACG CCTGGTATCT TTATAGTCTT  
 GCCGTCCCAG CCTTGTCTC TCGGGTGCTC CCTCGAAGGT CCCCCTTTC GGACCATAGA AATATCAGGA

**FIG.. 43F**

```

3571 GTCCGGTTTC GCCACCTCTG ACTTGAGCGT CGATTTTGT GATGCTCGTC AGGGGGCGG AGCCTATGGA
 CAGCCCAAG CGGTGGAGAC TGAACCTGCA GCTAANAACA CTACGAGCAG TCCCCCGCC TCGGATACCT

3641 AAAACGCCAG CAACGCGGCC TTTTACGGT TCCTGGCCTT TTGCTGGCCT TTTGCTCACA TGTTCCTTCC
 TTTTGGGGTC GTTGGCGCGG AAAAATGCCA AGGACCGGAA AACGACCGGA AAACGAGTGT ACAAGAAAGG

3711 TCGGTTATCC CCTGATTCG TGGATTAACG TATTACCGCC TTTGAGTGAG CTGATACCGC TCGCCGCGAGC
 ACGCAATAGG GGAATAAGAC ACCTATTGGC ATAATGGCGG AAATCACTC GACTATGGCG AGCGCGGTCC

3781 CGAACGACCG AGCGCAGCGA GTCAGTGAGC GAGGAAGCGG AAGAGCGCCC AATACGCAAA CCGCCTCTCC
 GCTTGTGGC TCGCGTCGCT CAGTCACTCG CTCCTTCGCC TTCTCGCGGG TTATGGGTTT GCGCGGAGGG

3851 CCGCGCGTTG GCCGATTCAT TAATGCAGCT GGCACGACAG GTTTCCTCGAC TGGAAAGCGG GCAGTGAGCG
 GCGCGCGAAC CGGCTAAGTA ATTACGTGCA CCGTGCTGTC CAAAGGGCTG ACCTTTCGCC CGTCACTCGC

3921 CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCGGCTCGT
 GTTGCCTTAA TTACACTCAA TCGAGTGAGT AATCCGTGGG GTCCGAAATG TGAATATCGA AGGCCGAGCA

 BssHII **

3991 ATGTTGTGTG GAAATTGTAG CCGATAACAA TTTACACACG GAAACAGCTA TGACCATGAT TACGCCAAGC
 TACAACACAC CTTAACACTC GCCTATTGTT AAAGTGTGTC CTTTGTGAT ACTGGTACTA ATGCGGTTCC

 BssHII EcoRI *****

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

4131 CTTGCCCCCG TTGCAGCGAT GGGTATTTT TCTAGTAATA TAAAGATAA ACTTAGACTC AAAACATTTA
 GAACGGGGGC AACGTGCTA CCCATAAAAA AGATCATTTT ATTTCTATT TGAATCTGAG TTTTGTAAAT

4201 CAAAAACAAC CCTTAAAGTC CTAAGCCCA AAGTGCTATG CAGCATCCAT AGCAAGCCCA GCCCAACCCA
 GTTTTGTGTT GGGATTTCAG GATTTCGGGT TTCACGATAC GTGCTAGGTA TCGTTCGGGT CGGGTTGGGT

```

**FIG.\_43G**

```

4271 ACCCAACCA ACCACCCCA GTGCAGCCAA CTGGCAATA GTCTCCACCC CCGGCACAT CACCOTGAGT
TGGGTGGGT TGGGTGGGT CACGTGGGT GACCGTTAT CAGAGTGGG GCGCGTATA GTGGCACTCA

4341 TGTCGCAAC ACCGCACGC TCGCAGCCAA AAAAAAAAAA AGAAGAAA AGAAGAAA GAAAAACAGC
ACAGGCGTGG TGCGGTGCAG AGGTCGGTT TTTTTTTTT TCTTCTTTT TTTTCTTTT CTTTTGTCTG

4411 AGGTGGGTCC GGTTCGTGG GCGCGAATA GCGAGGAGG TCGCAGCAG CGACGAGGCC CGGCCCTCCC
TCCACCCAGG CCGAGCACCC CCGGCTTTT CGCTCTCTT AGCGTCTGC GCTGCTCCG GCCGCGAGGG

4481 TCGCTTCCA AAGAACGCC CCCATCGCC ACTATATACA TACCCCCCC TCTCTCTCCA TCCCCCAAC
AGCGAAGGT TTTCTTTCGG GGGTAGCGG TGATATATG ATGGGGGGG AGAGGAGGT AGGGGGGTG

4551 CCTACCACCA CCACACCCAC CACTCTCTC CCCCTCGTG CCGGACGAC AGTCTCTCCC CCCTCCCCCT
GGATGGTGGT GGTGGTGGT GTGGAGGAG GGGAGCGAC GGCCTGCTG TCGAGGAGGG GGGAGGGGA

4621 CCGCCGCCG CGTAACCA CCGCCCTCT TCTCTTTCT TCTCTCGTT TTTTTCGT CTGGTCTCG
GGCGCGCGG GCCATTGGT GGGCGGGGAG AGGAGAAA AGAGGCAAA AAAAAAGCA GAGCCAGAGC

4691 ATCTTTGGC TTGGTAGTTT GGGTGGGCA GAGCGCTTC GTGCCCCAGA TCGGTGCGG GGAGGGCGG
TAGAAACCGG AACCATCAA CCCACCCGT CTCGCCGAG CAGCGGTCT AGCCACGCGC CTTCCCCGCC

4761 GATCTCGCG GTGGCTCTC CGGCGGTGAG TCGCCCCGA TCCTCGCGG GAATGGGCT CTCGGATGTA
CTAGAGCGC GACCGCAGG GCCGCACTC AGCCGGGCT AGGAGCGGCC CTTACCCCGA GAGCCTACAT

BgIII
~~~~~
4831 GATCTCTTT CTCTCTCTT TTTGTGTAG AATTGAATC CCTCAGCAT GTTCATCGGT AGTTTCTT
CTAGAAGAA GAAAGAAGAA AACACCATC TTAACCTAG GGTGCTGTA CAAGTAGCCA TCAAAAAGAA

4901 TTCATGATTT GTGACAAATG CAGCTCTGT CGGAGCTTT TTGTAGGTAG
AAGTACTAAA CACTGTTTAC GTCGGAGCAC GCCTCGAAA AACATCCATC

```

BamHI

BgIII

FIG.\_43H

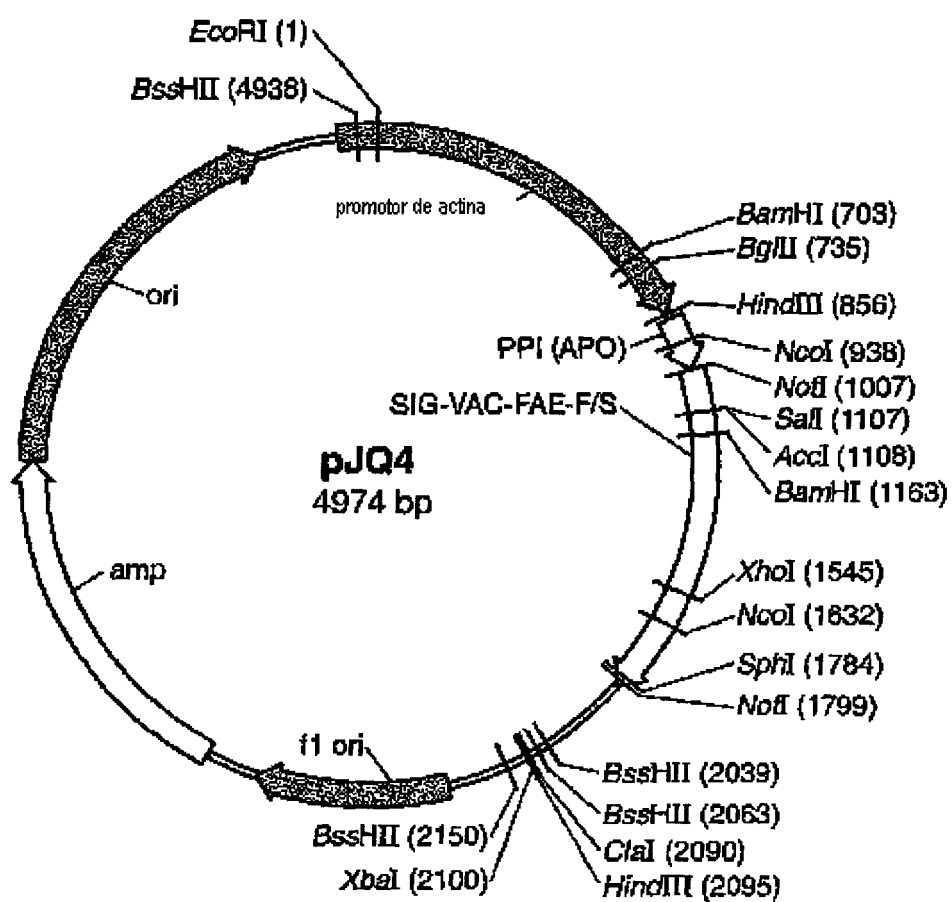


**FIG. 44A**

```

EcoRI
~~~~~
1 AATTCACAA TGAACAATAA TAAGATTAAA ATAGCTTGCC CCCGTTCCAG CGATGGGTAT TTTTCTAGT
 TTAAGGTGTT ACTGTTATT ATTCTAATTT TATCGAACGG GGGCAACGTC GCTACCCATA AAAAAATCA

71 AAAATAAAG ATAAACTTAG ACTCAAAACA TTTACAAAA CAACCCCTAA AGTCTTAAAG CCCAAGTGC
 TTTTATTTTC TATTGAATC TGAGTTTGT AAATGTTTT GTTGGGAT TCAGGATTC GGGTTTCACG

141 TATGCACGAT CCATAGCAAG CCCAGCCCA CCCAACCCAC CCCAGTGCAG GGTTCACGTC CCTACTGGCA
 ATACGTGCTA GGTATCGTTC GGGTCGGGT GGGTTGGGT GGGTACGTC GGTTCACCGT GGTTCACCGT

211 AATAGTCTCC ACCCCCGGCA CTATCACCGT GAGTTGTCCG CACCACCGCA CGTCTCGCAG CCAAAAAAA
 TTATCAGAGG TGGGGCCGT GATAGTGCA CTCACACGCG GTGTTGGCGT GCAGAGCGTC GGTTTTTTTT

281 AAAAAGAAAG AAAAAAAGA AAAAAGAAAA CAGCAGGTGG GTCCGGGTGG TGGGGGCCGG AAAAGCGAGG
 TTTTCTTTTC TTTTCTTTT TTTCTTTT GTCGTCCACC CAGGCCGAGC ACCCCCGGCC TTTTCGCTCC

351 AGGATCGCGA GCAGCGACGA GGCCCGGCC TCCTCCGCT TCCAAAGAAA CGCCCCCAT CGCCACTATA
 TCCTAGCGCT CGTCGCTGCT CCGGGCCGG AGGAGGCGA AGTTTCTTT GCGGGGGTA GCGTGTATAT

421 TACATACCCC CCCCTCTCCT CCCATCCCC CAACCTACC ACCACCAACA CCACCACTTC CTCCCCCTTC
 ATGTATGGGG GGGGAGAGGA GGGTAGGGG GTTGGGATGG TGGTGTGCT GGTGTGGAG GAGGGGGGAG

491 GCTGCCGGAC GACGAGCTCC TCCCCCTCC CCCTCCGCG CCGCCCGTAA CCACCCCGCC CCTCTCTCT
 CGACGGCCTG CTGCTCGAGG AGGGGGGAG GGGAGGCGGC GCGCGCCATT GGTGGGGCGG GGAGAGGAGA

561 TTCCTTCTCC GTTTTTTTT TCGTCTCGGT CTCGATCTTT GGCCTTGGTA GTTGGGTGG GCGAGAGCGG
 AAGAAAGAGG CAAAAAAGA AGCAGAGCCA GAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC

631 CTTCTGTCGC CAGATCGGTG CGCGGGAGGG GCGGATCTC GCGGTGGCG TCTCCGGGCG TGAATCGGCC
 GAAGCAGCGG GTCTAGCCAC GCGCCCTCCC CGCCCTAGAG CGCCGACCGC AGAGGCCCGC ACTCAGCCCG

BamHI
~~~~~
701  CGGATCCTCG  CCGGGAAATGG  GGCTCTCGGA  TGTAGATCTT  CTTCTTTCT  TCTTTTGTG  GTAGAATTGG
   GCCTAGGAGC  GCCCCTTACC  CCGAGAGCCT  ACATCTAGAA  GAAAGAAAGA  AGAAAAACAC  CATCTTAAC

```

**FIG.-44B**

```

771 AATCCCTCAG CATGTTCAT CGGTAGTTT TCTTTTCATG ATTTGTGACA AATGCAGCCT CGTCCGGAGC
   TTAGGGAGTC GTAAACAAGTA GCCATCAAAA AGAAAAGTAC TAAACACTGT TTACGTGGA GCACGCCCTCG

                               HindIII
                               ~~~~~
841 TTTTGTGTAG GTAGAACCTT ACNATGCMCG TGCACAAGGA GGTBAACTTC GTSGCCTACC TCCTGATCGT
 AAAAAACATC CATCTTCGAA TGTACCKGC ACGTGTTCCT CCATTTGAAG CASC GGATGG AGGACTAGCA

 NgoI
                               ~~~~~
911 SCTCGGCTC CTCTTGCTCG TSTCGCCAT GGAGCACGTG GACGCCAAGG CCTGCACCCK CGAGTCGGC
   SGAGCCGGAG GAGAACGAGC ABAGGCGGTA CCTCGTGCAC CTCGGTTC CACGCTGGG GCTCACGCCG

                               NotI
                               ~~~~~
981 AACCTCGGCT TCGGCATCTG CCCGGCGGCC GCCTCCACGC AGGCATCTC CGAAGACCTC TACAGCCGTT
 TTGGAGCCGA AGCCGTAGAC GGGCCGCCGG CGAGGTGCG TCCCGTAGAG GCTTCTGGAG ATGTCGGCAA

 SalI
                               ~~~~~
                               AccI
                               ~~~~~
1051 TAGTCGAAAT GGCACATATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC ATTCCGTGGA CTATTATCAA
 ATCAGCTTTA CCGGTGATAG AGGTTTCGAC GGATCCGGCT GGACACGTTG TAAGGCAGCT GATAATAGTT

 BamHI
                               ~~~~~
1121 GGGAGAGAAA ATTTACAATT CTCAAACTGA CATTAACGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA
   CCTCTCTTT TAAATGTTAA GAGTTTGACT GTAAATGCCCT ACCTAGGAGG CGCTGCTGTC GTCGTTTCTT

1191 ATAATCACCG TCTTCCGTGG CACTGGTAGT GATACGAATC TACAACTCGA TACTAACTAC ACCCTCACGC
   TATTAGTGGC AGAAGGCACC GTGACCATCA CTATGCTTAG ATGTTGAGCT ATGATTGATG TGGGAGTGGC

1261 CTTTCGACAC CCTACCACAA TGCACCGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT
   GAAAGCTGTG GGATGGTGTT ACGTTGCCAA CACTTCATGT GCCACCTATA ATATAACCTA CCCAGAGGCA

```

**FIG.\_44C**

1331 CCAGGACCMA GTCGAGTCGC TTGTCAAAACA GCAGGTTAGC CAGTATCCGG ACTACGCGCT GACCGTGACC  
GGTCCTGGTT CAGCTCAGCG AACAGTTTGT CGTCCAATCG GTCATAGGCC TGATGCGCGA CTGGCACTGG

1401 GGCCACKCCC TCGGCGCCTC CTTGGCGGCA CTCACCTGCC CCCAGCTGTC TCGGACATAC GACAAACATCC  
CCGGTGMGGG AGCCCGCGAG GGACCGCCGT GAGTGACGGC GGGTCGACAG ACGCTGTATG CTGTTGTAGG

1471 GCCTGTACAC CTTGCGGAA CCGCGCAGCG GCAATCAGGC CTTGCGGTTC TACATGAACG ATGCCTTCCA  
CGGACATGTG GAAGCCGCTT GCGCGGTTCG CGTTAGTCCG GAAGCGCAGC ATGTACTTGC TACGGAAAGT

XhoI  
~~~~~

1541 AGCCTCGAGC CCAGATACGA CGCAGTATTT CCGGGTCAC TATGCCAACG ACGGCATCCC AAACCTGCC
TCGGAGCTCG GGTCTATGCT GGTCTATGCT GGTCTATGCT GGTCTATGCT GGTCTATGCT GGTCTATGCT

NcoI
~~~~~

1611 CCGGTGGAGC AGGGGTACGC CCATGGCGGT GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCCGAACA  
GGCCACCTCG TCCCATGCG GGTACCGCCA CATCTCATGA CTTGCAACT AGGAATGTCT GGGTCTTGT

1681 CATTTGTCTG CACTGGGGAT GAAGTGCACT GCTGTGAGC CCAGGCGCGA CAGGGTGTGA ATATGCCCA  
GTAAACAGAC GTGACCCCTA CTTACAGTCA CGACACTCCG GGTCCGCTT GTCCACACT TATTACCGCT

SphI  
~~~~~

1751 CACGACTTAT TTTGGGATGA CGAGCGGCGC ATGCACCTGG CCGGTGCGGG CCGCGGAAC CACTGAAGGA
GTGCTGAATA AAACCTTACT GCTCGCGCG TACGTGGACC GGCCAGCGCC GCGCCCTTGG GTGACTTCTT

1821 TGAGCTGTAA AGAAGCAGAT CGTTCAAAACA TTTGGCAATA AGTTTCTTA AGATTGAATC CTGTTGCCGG
ACTCGACATT TCTTCGTCTA GCAAGTTTGT AAACCGTTAT TTCAAGAAT TCTAACTTAG GACAAACGCG

1891 TCTTGGGATG ATTATCATAT AATTCTGTT GAATTACGTT AAGCATGTAA TAATTAACAT GTATGCAATG
AGAACGCTAC TAATAGTATA TTAAGACAA CTTAATGCAA TTCGTACATT ATTAATTGTA CATTAACGTAC

1961 ACGTTATTTA TGAGATGGGT TTTTATGATT AGAGTCCCG AATTATACAT TTAATACGG ATAGAAACA
TGCAATAAAT ACTCTACCCA AAAATACTAA TCTCAGGCG TTAATATGTA AATTATGCG TATCTTTGT

FIG._44D

```

XbaI                                     ClaI HindIII
~~~~~
2031 AAATATAGCG CGCAAACTAG GATAAATTAT CGCGCGCGGT GTCATCTATG TTACTAGATC GATAAGCTTC
TTTATATCGC GCGTTTGATC CTATTTAATA CGCGCGCCCA CAGTAGATAC AATGATCTAG CTATTGGAAG

XbaI                                     BssHII
~~~~~
2101 TAGAGCGGCC GGTGGAGCTC CAAATCGCCC TATAGTGAGT CGATTTACGC GCGCTCACAG GCCGTGCTTT
ATCTCGCCCG CCACCTCGAG GTTAAGCGGG ATATCACTCA GCATAATGCG CCGGAGTGAC CGGCAGCAAA

2171 TACAACGTCT TGACTGGGAA AACCCCTGGC TTACCCAACT TAATCGCCTT GCAGCACATC CCCCTTTTCG
ATGTTGCAGC ACTGACCCCT TTGGGACCGC AATGGGTTGA ATTAGCGGAA CGTCGTGTAG GGGGAAAGCG

2241 CAGCTGGCGT AATAGCGAAG AGGCCCGCAC CGATCGCCCT TCCCAACAGT TCGCAGCCTT GAATGGCGAA
GTCGACCGCA TTATCGCTTC TCCGGCGGTG GCTAGCGGGA AGGTTGTCA ACGCTTCGGA CTTACCGCTT

2311 TGGGACGCGC CCTGTAGCGG CGCATTAAGC GCGCGGGGTG TGGTGGTTAC GCGCAGCTG ACCGCTACAC
ACCTGCGCGG GGACATCGCC GCGTAATTCT GCGCGCCAC ACCACCAATG CCGCTCGCAC TGGCGATGTG

2381 TTGCCAGCGC CCTAGCGGCC GCTCCTTTCT CTTTCTTCCC TTCTTTCTC GCCACGTTCT CCGGCTTTCC
AACGGTCTCG GATTCGCGGG CGAGGAAAGC GAAAGAAAGG AAGGAAAGAG CCGTCAAGC GCGCGAAAGG

2451 CCGTCAAGCT CTAATCGGG GGTCCCTTTT AGGTTCCGA TTTAGTGCTT TACGGCACCT CGACCCCAAA
GGCAGTTCTGA GATTTAGCCC CCGAGGAAA TCCCAAGGCT AATCACGA ATGCCGTGGA GCTGGGGTTT

2521 AAACTTGATT AGGTGATGG TTCACGTAGT GGGCCATCGC CCGATAGAC GGTTTTTCGC CTTTGCAGT
TTTGAACATA TCCCACATCC AAGTGCATCA CCCGGTAGCG GGACTATCTG CCAAAAAGCG GGAACCTGCA

2591 TGGAGTCCAC GTTCTTTTAA ATGGACTCT TGTTCCAAC TGGACACAA CTCACCCCTA TCTCGGTCTA
ACCTCAGGTG CAAGAAATTA TCACCTGAGA ACAAGGTTTG ACCTTGTGT GAGTTGGAT AGAGCCAGAT

2661 TTCCTTTTGT TTATAAGGGA TTTTGGCGAT TTCGGCCTAT TGGTTAATA ATGAGCTGAT TTAAACAAAA
AAGAAACTA AATATTCCCT AAAACGGATA AGCCGGATA ACCAATTTT TACTCGACTA AATTGTTTTT

```

FIG._44E

2731 TTTAACGGCA ATTTTAACAA AATATTAAAG CTTACAATTT AGGTGGCACT TTTCGGGAA ATGTGCGCGG
 AAATTGCGCT TAAATTTGTT TTATAATTGC GAATGTTAAA TCCACCGTGA AAAGCCCCCTT TACACGCGCC
 2801 AACCCCTATT TGTTTATTTT TCTAAATACA TTCAAATATG TATCCGCTCA TGAGACAATA ACCCTGATAA
 TTGGGGATAA ACAATATAAA AGATTATATG AAGTTTATAC ATAGCGAGT ACTCTGTAT TGGGACTATT
 2871 ATGCTTCAAT AATATTGAAA AAGGAAGAGT ATGAGTATTC AACATTTCCG TGTGCCCCCTT ATTCCCCTTTT
 TACGAAGTTA TTATAACTTT TTCTTCTCA TACTCATAG TGTAAAGGC ACAGCGGAA TAAGGGAAAA
 2941 TTGCGGCATT TTGCCCTTCCCT GTTTTGTCTC ACCCAGAAAC GCTGTTGAAA GTAAAAGATG CTGAAGATCA
 AACGCCGTAA AACGGAAAGGA CAAAAACGAG TGGGTCTTTG CGACCACTTT CATTTTCTAC GACTTCTAGT
 3011 GTTGGGTGCA CGAGTGGGT ACATCGAACT GGATCTCAAC AGCGGTAAAG TCCTTGAGAG TTTTCGCCCC
 CAACCCACGT GCTCACCCCA TGTAGCTTGA CCTAGAGTTG TCGCCATTCT AGGAATCTC AAAAGCGGGG
 3081 GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC CGTATTGACG
 CTTCTTGCAA AAGGTTACTA CTCGTGAAAA TTTCAGAGCG ATACACCGCG CCATAATAGG GCATAACTGC
 3151 CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC
 GGCCCGTTCT CGTTGAGCCA GCGCGTATG TGATAAGAGT CTTACTGAAAC CAACTCATGA GTGGTCAGTG
 3221 AGAAAGCAT CTTACGGATG GCATGACAGT AAGAGAAITA TGCAGTGTG TGCATAACCAT GAGTGATAAC
 TCTTTTCGTA GAATGCTTAC CGTACTGTCA TTCTCTTAAT ACGTCACGAC GGTATTGGTA CTCACTATTG
 3291 ACTGCGGCCA ACTTACTTCT GACAACGATC GGAGGACCGA AGGAGCTAAC CGCTTTTGTG CACAACATGG
 TGACGCCGGT TGAATGMAAG CTGTTGCTAG CCTCTGGCT TCCTCGATTG GCGAAAAAAC GTGTTGTACC
 3361 GGGATCATGT AACTCGCCCTT GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAAAG ACGAGCGTGA
 CCTAGTACA TTGAGCGGAA CTAGCAACCC TTGGCCTCGA CTTACTTCGG TATGTTTGC TGCTGCGACT
 3431 CACCACGATG CCTGTAGCAA TGGCAACAAC GTTTCGCMAA CTATTAACTG GCGAACTACT TACTCTAGCT
 GTGGTGTCTAC GGACNTCGTT ACCGTTGTTG CAACGCGTTT GATAATTGAC CGCTTGATGA ATGAGATCGA
 3501 TCCCGGCAAC AATTAAATGA CTGGATGGAG GCGGATAAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC
 AGGGCCGTTG TTAATTATCT GACCTACCTC CGCCTATTTT AACGTCTTGG TGAAGACGCG AGCCGGGAG

FIG. 44F

3571 CGGCTGGCTG GTTTATTGCT GATAAATCTG GAGCCGGTGA GCGTGGGTCT CGCGGTATCA TTGCAGCACT
 GCCGACCCGAC CAAATAACGA CTATTTAGAC CTCGGCCACT CGCACCCAGA CGCCCATAGT AACGTCGTGA
 3641 GGGGCCAGAT GGTAAAGCCCT CCCGTATCGT AGTTATCTAC ACGACGGGGA GTGAGGCAC TATGATGAA
 CCCCCTCTA CCATTCCGGA GGGCATAGCA TCAATAGATG TGCTGCCCTT CAGTCCCTTG ATACCTACTT
 3711 CGAATATGAC AGATCGCTGA GATAGGTGCC TCACTGATTA AGCATTGGTA ACTGTCAGAC CAAGTTTACT
 GCITTTATCTG TCTAGCGACT CTATCCACGG AGTGACTAAT TCGTAACCAT TGACAGTCTG GTTCAAAATGA
 3781 CATATATACT TTAGATTGAT TTAANAATTTC ATTTTAAAT TAAAGGATC TAGGTGAAGA TCCTTTTGA
 GTATATATGA AATCTAACTA AATTTTGAAG TAAATAATA ATTTTCTTAG ATCCACTTCT AGGAAAAACT
 3851 TAACTCTATG ACCAAAATCC CTTAACGTGA GTTTTCGTC CACTGAGCGT CAGACCCCGT AGAAAAGATC
 ATTAGAGTAC TGGTTTTAGG GAATTGCAC TAAAGCAAG GTGACTCGCA GTCTGGGGCA TCTTTTCTAG
 3921 AAAGGATCTT CTTGAGATCC TTTTTTTCTG CGCGTAATCT GCTGCTTGA AACMAAAJA CCACCGCTAC
 TTTCTTAGAA GAACCTTAGG AAAAAAGAC GCGCATTAGA CGACGAACGT TTGTTTTTTT GGTGGCGATG
 3991 CAGCGGTGGT TTGTTTCCG GATCAAGAGC TACCJAATCT TTTTCCGAAG GTAACTGGCT TCAGCAGAGC
 GTCGCCACCA AACAAACGGC CTAGTTCTCG ATGTTTGAAG AAAAGGCTTC CATTGACCGA AGTCGTCTCG
 4061 GCAGATACCA AATACTGTCC TTCTAGTGA GCGGTAGTTA GGCCACCAC TCAAGMACTC TGTAGCACCG
 CGTCTATGGT TTATGACAGG AAGATCACAT CGGCATCAAT CCGGTGGTGA AGTTCTTGAG ACATCGTGGC
 4131 CCTACATACC TCGCTCTGCT AATCCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTCTG TGTCTTACCG
 GGATGTATGG AGCGAGACGA TTAGGACAA TGGTACCAGC GACGGTCACC GCTATTTCAGC ACAGAAATGGC
 4201 GGTGGACTC AAGACGATAG TTACCGGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGGT TGCCCCCA
 CCAACCTGAG TTCTGCTATC AATGGCCTAT TCCGGCTGCG CAGCCCGACT TGCCCCCA GCACGTGTGT
 4271 GCCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA AAGCCCAAG
 CGGGTCGAAC CTCGCTTGCT GGATGTGGCT TGACTCTATG GATGTGCGAC TCGATACTCT TTCCCGGTGC
 4341 CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAGCG GCAGGGTCTG AACAGGAGAG CGACAGAGG
 GAAGGGCTTC CCTCTTTCCG CCTGTCCATA GGCCATTCTG CGTCCAGCC TTGTCTCTCT GCCTGCTCCC

FIG._44G

4411 AGCTTCCAGG GGAACGCC TGGTATCTTT ATAGTCTCTGT CGGGTTTCGC CACCTCTGAC TTGAGCGGTGG
TCGAAGGTCC CCTTTGCCG ACCATAGAAA TATCAGGACA GCCCAAGCG GTGGAGACTG AACTCGCAGC

4481 ATTTTGTGA TGCTCGTCAG GGGGCGGAG CCTATGAAA AACGCCAGCA ACGCGCCCTT TTTACGGTTC
TAAAAARCACT ACGAGCAGTC CCCCCGCCTC GGATACCTTT TTGCGGTCTGT TCGCGCGGAA AATGCCAAG

4551 CTGGCCTTTT GCTGGCCTTT TGCTCACATG TTCTTTCTCTG CGTTATCCCC TGAATCTGTG GATAACCGTA
GACCGGAAA CGACCGGAA ACGAGTGATC AAAAAAGGAC GCAATAGGGG ACTAAGACAC CTATTGGCAT

4621 TTACCGCCTT TGAGTGAGCT GATACCGCTC GCCGACGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA
AATGGCGGA ACTCACTCGA CTATGGCGAG CGCGGTCTGC TTGCTGGCTC GCGTCTCTCA GTCACTCGCT

4691 GGAAGCGGA GAGCGCCCA TACGCAAAAC GCCTCTCCCC GCGCGTTGGC CGATTCAATTA ATGCAGCTGG
CCTTCGCCCTT CTCGCGGGTT ATGCGTTTGG CGGAGAGGGG CGCGCAACCG GCTAAGTAAT TACGTGAGC

4761 CACGACAGGT TTCCCGACTG GAAAGCGGGC AGTGAGCGCA ACGCAATTA TGTGAGTTAG CTCACTCAAT
GTGCTGTCCA AAGGGCTGAC CTTTCGCCCG TCACTCGCGT TGCCTTAAT ACACCTCAATC GAGTGAGTAA

4831 AGGCACCCCA GGCTTTACAC TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTGAGCG GATAACAAAT
TCCGTGGGGT CCGAAATGTG AAATACGAAG GCCGAGCATA CAACACNCCT TAACNCTCGC CTATTGTTAA

BsSHII
~~~~~

4901 TCACACAGGA AACAGCTATG ACCATGATTA CGCCAAGCGC GCAATTAACC CTCACTAAAG GGAACAAAAG  
AGTGTGTCCT TTGTCGATAC TGGTACTAAT GCGGTTCTCGC GGTAAATTGG GAGTGATTC CCTTGTTTTTC

EcoR

4971 CTGG  
GACC

FIG..44H

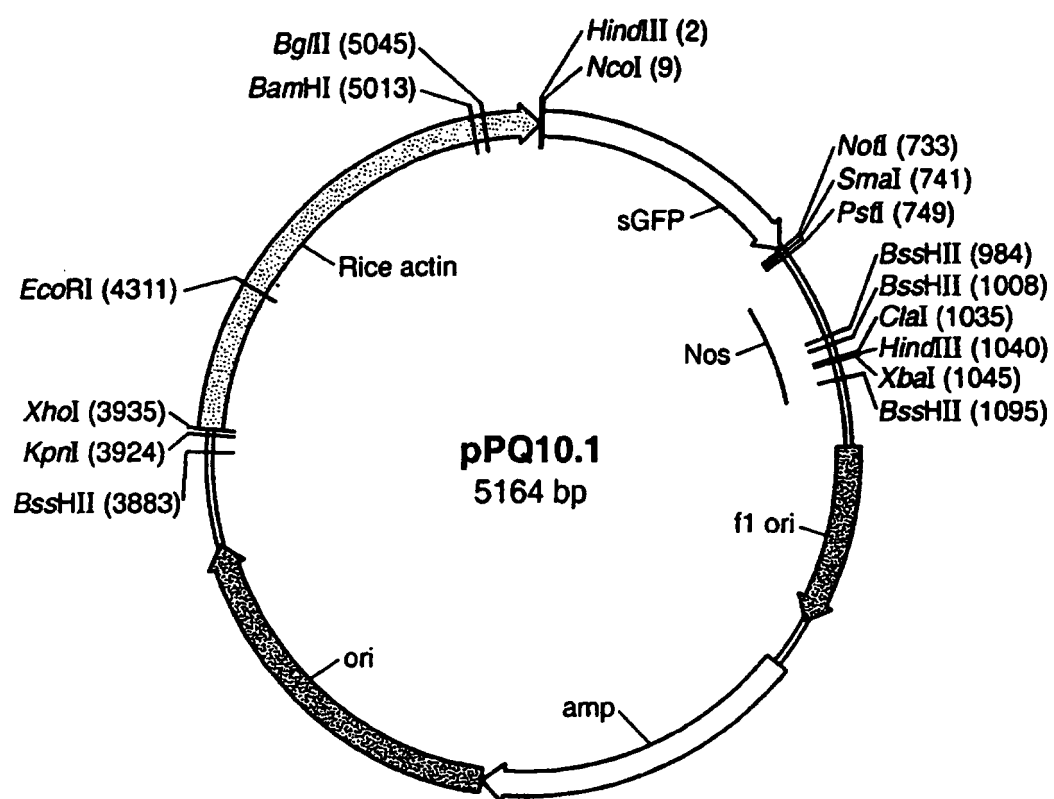


**FIG.\_45A**

HindIII NcoI  
 ~~~~~  
 1 AAGCTTACCA TGGTGAGCAA GGGCGAGGAG CTGTTACCCG GGGTGGTGCC CATCCTGGTC GAGCTGGACG
 TTCGAATGGT ACCACTCGTT CCCGCTCCTC GACAAGTGGC CCCACCACGG GTAGGACCAG CTCGACCCTGC
 71 GCGACGTGAA CGGCCACAAG TTCAGCGTGT CCGGCGAGGG CGAGGGCGAT GCCACCTACG GCAAGCTGAC
 CGCTGCACCT GCCGGTGTTC AAGTCGCACA GGCCGCTCCC GCTCCCGCTA CGGTGGATGC CGTTCGACTG
 141 CCTGAAGTTC ATCTGCACCA CCGGCAAGCT GCCGTGCCC TGGCCCCACC TCCTGACCAC CTTACACCTAC
 GGACTTCAAG TAGACGTGGT GGCCGTTCGA CCGGCACGG ACCGGGTGG AGCACTGGTG GAAGTGGATG
 211 GCGGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCAGCACTT CTTCAAGTCC GCCATGCCCG
 CCGCACGTCA CGAAGTCGGC GATGGGGCTG GTGTACTTCG TCCTGCTGAA GAAGTTCAGG CGGTACGGGC
 281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC AAGACCCGCG CCGAGGTGAA
 TTCCGATGCA GGTCCCTCGC TGGTAGAAGA AGTCCCTGCT GCCGTTGATG TTCTGGGCGC GGCTCCACTT
 351 GTTCGAGGCG GACACCCCTGG TGAACCCGAT CGAGCTGAAG GGCACTCGACT TCAAGGAGGA CGGCAACATC
 CAAGCTCCC CGTGTGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA AGTTCCTCCT GCCGTGTGAG
 421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT GGCCGACAAAG CAGAAGAAACG
 GACCCCGTGT TCGACCTCAT GTTGATGTTG TCGGTGTTGC AGATATAGTA CCGGCTGTTT GTCTTCTTGC
 491 GCATCAAGGT GAACTTCAAG ATCCGCCACA ACATCGAGGA CCGCAGCGTG CAGCTCGCCG ACCACTACCA
 CGTAGTTCCA CTTGAAGTTC TAGGCGGTGT TGTAGCTCCT GCCGTGCGAC GTCGAGCGGC TGGTGATGGT
 561 GCAGAACACC CCCATCGGCG ACBGCCCCGT GCTGCTGCCC GACAACCACT ACCTGAGCAC CCAGTCCGCC
 CGTCTTGTGG GGGTAGCCCG TGCCGGGGCA CGACGACGGG CTGTTGGTGA TGGACTCGTG GGTCAAGGCGG
 631 CTGAGCANAAG ACCCCACAAG GAAAGCGCGAT CACATGGTCC TGCTGGAGTT CGTGACCGCC GCCGGGATCA
 GACTCGTTTC TGGGGTTGCT CTTGCGGCTA GTGTACCAGG ACGACCTCAA GCACTGGCGG CGGCCCTAGT

FIG._45B

```

      SmaI
      -----
      NotI      PstI
      -----
701  CTCACGGCAT GGACGAGCTG TACAAGTAAA GCGGCCGCC GGGCTGCAGG GAAACCACTG AAGGATGAGC
    GAGTGGCCGT CCTGCTCGAC ATGTTCAATTT CGCCGGCGGG CCCGACGCTC CTTTGGTGAC TTCCTACTCG

771  TGTAAAGAA GAGATCGTTC AAACATTTGG CAATAAAGTT TCCTAAGATT GAATCCTGTT GCCGGTCTTG
    ACATTTCTC GTCTAGCAAG TTGTAAACC GTTATTTCAA AGAATTTCTAA CTTAGGACAA CGGCCAGAAC

841  CGATGATTAT CATATAAATTT CTGTTGAATT ACGTTAAGCA TGTAAATAAT AACATGTAAT GCATGACGTT
    GCTACTAATA GTATATTAAA GACAACCTTAA TGCAATTCGT ACATTATTAA TTGTACATTAA CGTACTGCAA

911  ATTTATGAGA TGGGTTTTTA TGATTAGAGT CCGGCAATTA TACATTTAAT ACGCGATAGA AAACAAAATA
    TAAATACTCT ACCCAAAAAT ACTAATCTCA GGGCGTTAAT ATGTAAATTA TCGGCTATCT TTGTTTTTAT

      XbaI
      -----
      BssHII      BssHII      ClaI HindIII
      -----
981  TAGCGCGCAA ACTAGGATAA ATTATCGCGC GCGGTGTCTAT CTATGTTACT AGATCGATAA GCTTCTAGAG
    ATCGCGCGTT TGATCCTATT TAATAGCGCG CGCCACAGTA GATACAATGA TCTAGCTATT CGAAGATCTC

      BssHII
      -----
1051 CGGCCGGTGG AGCTCCAAAT CGCCCTATAG TGAGTCGTAT TACGCGCGCT CACTGGCCGT CGTTTACAA
    GCCGGCCACC TCGAGGTTAA GCGGGATATC ACTCAGCATA ATGCGCGCGA GTGACCGGCA GCAAAATGTT

1121 CGTCGTGACT GGGAAAACCC TGGCGTTACC CMACTTAATC GCCTTGCAGC ACATCCCCCT TTCGCCAGCT
    GCAGCACTGA CCTTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCT GTTAGGGGA AAGCGGTCTGA

1191 GCGGTAATAG CGAAGAGGCC CGCACCGATC GCCCTTCCCA ACAGTTGCGC AGCCTGAATG GCGAATGGGA
    CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAGGGT TGTCAACGCG TCGGACTTAC CGCTTACCCT

```

FIG._45C

1261 CCGCCCTGT AGCGCGCAT TAAGCGGCG GGGTGTGGTG GTTACGCGCA GCGTACCGGC TACACTTGCC
 GCGCGGACA TCGCCGCGTA ATTGCGCGG CCCACACAC CAATGCGCGT CCGACTGGCG ATGTGAACGG
 1331 AGCGCCTAG CCGCCGCTCC TTTCGCTTC TTCCCTTCCT TTCTCGCCAC GTTCGCGGC TTTCGCCGTC
 TCGCGGATC GCGGCGAGG AAGCGGAAG AAGGGAAGGA AAGGCGGTG CAAGCGCGG AAGGGGCAG
 1401 AAGCTCTAA TCGGGGCTC CCTTAGGGT TCCGATTAG TGCTTTACGG CACCTCGACC CCAAAAA
 TTCGAGATT AGCCCCGAG GGAATCCCA AGCTAATC ACGAATGCC GTGGAGCTGG GGTTTTGA
 1471 TGATTAGGGT GATGGTTCAC GTAGTGGGC ATGCCCCGA TAGACGGTTT TTCGCCCTT GACGTTGGAG
 ACTAATCCCA CTACCAGTG CATACCCGG TAGCGGACT ATCTGCCAA AAGCGGAAA CTGCAACCTC
 1541 TCCACGTTCT TTAATAGTGG ACTCTGTTTC CAACCTGGA CAACACTCA CCTATCTCG GTCATTCCT
 AGGTGCAAGA AATTATCAC TGAGAACAG GTTGACCTT GTTGTAGTT GGGATAGAGC CAGATAAGAA
 1611 TTGATTATA AGGATTTTG CCGATTTGG CCTATTGGT AAAAAATGAG CTGATTTAAC AAAAAATTA
 AACTAATAT TCCCTAANAC GGCTAAGCC GDATAACCA TTTTTFACCT GACTAAATTG TTTTAAATTT
 1681 CCGGAATTT AACAAATAT TAACGCTTAC AATTAGGTG GCATTTTCG GGGAAATGTG CGCGGAACCC
 GCGCTTAAA TTGTTTATA ATTGCGMAT TTAATCCAC CGTGAAMAGC CCTTTTACAC GCGCCTTGGG
 1751 CTATTGTTT ATTTTCTAA ATACATTCA ATATGATCC GCTCATGAGA CAATAACCTT GATAAATGCT
 GATAACAAA TAAAGATT TATGTAAAT TATACATAGG CGAGTACTCT GTTATTGGGA CTATTTACGA
 1821 TCAATAATAT TGAANAAGGA AGAGTATGAG TATTCAACAT TTCCGTGTCG CCTTATTCC CTTTTTGGG
 AGTTATTATA ACTTTTCTT TCTCATCTC ATAAATTGA AAGCACAGC GGGAAATAGG GAAAAAACGC
 1891 GCATTTTGC TTCTGTTTT TGCTCACCCA GAACGCTGG TGAAGTAAA AGATGCTGAA GATCAGTTGG
 CGTAAACGG AAGACAAA ACGAGTGGGT CTTTGGGACC ACTTCATT TCTACGACTT CTAGTCNACC
 1961 GTGCACGAGT GGGTTACATC GAACGTGATC TCACACGCG TAAATCCTT GAGAGTTTTC GCCCCGAAAG
 CACGTGCTCA CCCAATGTAG CTTGACCTAG AGTTGTGCGC ATTCTAGGA CTCTCAAAAG CCGGCTTCT
 2031 ACGTTTCCA ATGATGAGCA CTTTAAAGT TCTGCTATGT GCGCGGTAT TATCCCCGTAT TGACGCCGGG
 TGCAAAAGGT TACTACTCGT GAAATTTCA AGACGATACA CCGCGCCATA ATAGGCATA ACTGCGGCC

FIG. 45D

2101 CAAGAGCAAC TCGGTGCGG CATACACTAT TCTCAGAATG ACTTGGTTGA GTACTCACCA GTACACAGAAA
 GTTCTCGTTG AGCCAGCGGC GTATGTGATA AGAGTCTTAC TGAACCAACT CATGAGTGGT CAGTGTCTTT

 2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA ACCATGAGTG ATAACACTGC
 TCGTAGAATG CCTACCGTAC TGTCATTCTC TTAATACTGC ACGACGGTAT TGGTACTCAC TATTGTGACG

 2241 GGCCAACTTA CTTCCTGACAA CGATCGGAGG ACCGAAGGAG CTAAACCGCTT TTTTGCACAA CATGGGGGAT
 CCGGTTGAAT GAAGACTGTT GCTAGCCTCC TGCTTCTC TCAGTGGGAA AAAACGTGTT GTACCCCTTA

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

 2381 CGATGCCTGT AGCAATGGCA ACAACGTTGC GCAAACTATT AACTGGCGAA CTACTTACTC TAGCTTCCCG
 GCTACGGACA TCGTTACCGT TGTTGCAACG CGTTTGATAA TTGACCGCTT GATGAATGAG ATCGAAGGBC

 2451 GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTTC TCGCTCGG CTTTCCGGCT
 CGTTGTTAAT TATCTGACCT ACCTCCGCT ATTTCAACGT CCTGGTGAG ACGGAGCCG GGAAGGCCGA

 2521 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCG TATCATTTGCA GCACGTGGGC
 CCGACCAAT AACGACTATT TAGACCTCGG CCACTCGCAC CCAGAGCGCC ATAGTAACGT CGTGACCCCG

 2591 CAGATGGTAA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGGAGTCAG GCAACTATGG ATGAACGAAA
 GTCTACCATT CGBGAGGGCA TAGCATCAAT AGATGTGCTG CCCCTCAGTC CGTTGATACC TACTTGCCTT

 2661 TAGACAGATC GCTGAGATAG GTGCCTCACT GATTAAAGCAT TGGTAACGT CAGACCAAGT TTACTCATAT
 ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTCGTA ACCATTGACA GTCTGGTTCA AATGAGTATA

 2731 ATACTTTAGA TTGATTTTAA ACTTCATTTT TAATTTTAAA GGATCTAGGT GAAGATCCTT TTTGATAATC
 TATGAATACT AACTAAATTT TGAAGTAAAA ATTAATTTTT CCTAGATCCA CTTCTAGGAA AAATATTAG

 2801 TCATGACCAA AATCCCTTAA CGTGAGTTT CATTCCACTG AGCTCAGAC CCCGTAGAAA AGATCAAGG
 AGTACTGGTT TTAGGGGAAT GCACCTCAAAA GCMAAGTGAC TCGCAGTCTG GGGCATCTTT TCTAGTTTCC

 2871 ATCTTCTTGA GATCCTTTT TTCTGGCGGT AATCTGCTGC TTGCMAACAA AAAAACCCACC GCTACCAAGC
 TAGAAGAACT CTAGGAAAAA AAGACCGCA TTAGACGACG AACGTTTGT TTTTGGTGG CGATGGTCCG

FIG._45E

2941 GTGGTTTGTT TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAAC TGGCTTCAGC AGAGGCGCAGA
 CACCAAAACA ACGGCCTAGT TCTCGATGGT TGAGAAAGA GCTTCCATTG ACCGAAGTCG TCTCGCGTCT

 3011 TACCAAAATAC TGTCTTTCTA GTGTAGCCGT AGTTAGGCCA CCACCTTCAAG AACTCTGTAG CACCGCCTTAC
 ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GTTGAAGTTC TTGAGACATC GTGGCGGATG

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

 3151 GACTCAAGAC GATAGTTACC GGATAAGCG CACCGGTCGG GCTGAACGGG GGGTTCTGTG ACACAGCCCA
 CTGAGTTCTG CTATCAATGG CCTATTCCGC GTCGCCAGCC CGACTTGCCC CCCAAGCACG TGTGTCGGGT

 3221 GCTTGGAGCG AACGACCTAC ACCGAACCTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC
 CGAACCTCGC TTGCTGGATG TGGCTTGACT CTATGQATGT CGCACTCGAT ACTCTTTCGC GTTGGGAAGG

 3291 CGAAGGGAGA AAGCGGACA GGTATCCGGT AAGCGGACGG GTCGGAACAG GAGAGCGCAC GAGGGAAGCTT
 GCTTCCCTCT TTCCGCCCTGT CCATAGGCCA TTGCGCCGTCC CAGCCTGTG CTTCTCGCGTG CTCCCTCGAA

 3361 CCAGGGGGA ACGCCTGGTA TCTTTATAGT CCTGTGCGGT TTGCGCACCT CTGACTTGA GGTGGAATTT
 GGTCCCCCTT TGGGAGCCAT AGAAATATCA GGACAGCCCA AAGCGGTGA GACTGAATC GCAGCTAATA

 3431 TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAAGCGG GCCTTTATAC GGTTCCTTGGC
 ACACTACGAG CAGTCCCCC GCCTCGGATA CCTTTTTCG GTCTGTGCGC CGGAJAJAATG CCAAGGACCG

 3501 CTTTGTGCTG CCTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCTCTGAT TCCCTGATTA CCGTATTACC
 GAAACGACC GGAAACGAG TGTACAAGAA AGGACGCAAT AGGGGACTAA GACACCTATT GGCATATATG

 3571 GCCTTTGAGT GAGCTGATAC CGTTCGCCGC AGCCGAACGA CCGAGCCCGAG CGAGTCAGTG AGCGAGGAAG
 CGGAACCTCA CTCGACTATG GCGAGCGGCG TCGGCTTGT GGTCTGCGTC GCTCAGTCAC TCGCTCCTTC

 3641 CGGAAGAGCG CCCAATACGC AAACCGCTC TCCCGCGCG TTGGCCGAT CATTAATGCA GCTGGCACGA
 GCCTTCTCGC GGGTTATGCG TTTGGCGGAG AGGGGCGCG AACCGGCTAA GTAAATACGT CGACCGTGT

 3711 CAGGTTTCCC GACTGGAAAG CGGGCAGTGA GCGCAACGCA ATTAATGTGA GTTAGCTCAC TCATTAGGCA
 GTCCAAAGGG CTGACCTTTC GCCCGTCNCT CGCGTTGCGT TAAATTACACT CAATCGAGTG AGTAATCCGT

FIG._45F

```

3781  CCCCAGGCTT TACACTTTAT GCTTCGGGCT COTATGTTGT GTGGAAATGT GAGCGGATAA CAATTTCCACA
      GGGGTCCGAA ATGTGAAATA CGAGGCCGGA GCATACAAACA CACCTTAACA CTCGCCATTT GTTAAAGTGT

      BssHII
3851  CAGGAAACAG CTATGACCAT GATTACGCCA AGCGCGCAAT TAACCTCTAC TAAAGGGAAAC AAAGCTGGG
      GTCCTTTGTC GATACTGGTA CTAAATCCGGT TCGCGCGTTA ATTGGGAGTG ATTTCCCTTG TTTTCGACCC

      KpnI
3921  TACCGGGGCC CCCCTCGAGG TCATTTCATAT GCTTGAGGAG AGAGTCGGGA TAGTCCAAAA TAAACAAAG
      ATGGCCCGGG GGGGAGCTCC AGTAAGTATA CGAACTCTTC TCTCAGCCCT ATCAGGTTTT ATTTGTTTC

      XhoI
3991  GTAAGATTAC CTGGTCAAAA GTGAACACAT CAGTTAAAG GTGGTATTAAG TAAATATCG GTAATAAAG
      CATTCTAATG GACCAGTTTT CACTTTTGTG GTCAATTTTC CACCATATTC ATTTTATAGC CATTAATTTTC

4061  GTGGCCCAAA GTGAATTTA CTCCTTTCTA CTATTATAAA AATTGAGGAT GTTTGTCTGG TACTTTGATA
      CACCGGGTTT CACTTTAAAT GAGAAAGAT GATAATATTT TTAACCTCTA CAAACAGGCC ATGAACCTAT

4131  CGTCATTTT GTATGAATTG GTTTTAAGT TTATTCGCGA TTTGGAAATG CATATCTGTA TTTGAGTCGG
      GCAGTAAAA CATACTTAAC CAATAATTCA AATAAGCGCT AACCTTTAC GTATAGACAT AAACCTCAGCC

4201  TTTTTAAGTT CGTTGCTTTT GTAATACAG AGGGAATTTGT ATAAGAAATA TCCTTAAAA ACCCATATGC
      AAAAATTCAG GCACGAAAA CATTATGTC TCCCTAAAA TATCTTTAT AGAAATTTTT TGGGTATACG

      EcoRI
4271  TAATTGACA TAATTTTGA GAAAAATATA TATTCAGGCG AATTCACAAA TGAACAAATA TAAGATTAAA
      ATTAAACTGT ATTAAAACT CTTTATATAT ATAAATCCGC TTAAGGTGTT ACTTGTTATT ATTCTAATTT

4341  ATAGCTTGCC CCCGTTGCAG CGATGGGTAT TTTTCTAGT AAATATAAG ATAAACTTAG ACTCAAAACA
      TATCGAACGG GGGCAACGTC GCTACCCATA AAAAAGATCA TTTATTTTC TATTTGAATC TGAATTTTGT

4411  TTTACAAAA CAACCCCTAA AGTCCTAAG CCCAAAGTGC TATGCACGAT CCATAGCAG CCCAGCCCCA
      AAATGTTTT GTTGGGATT TCAGGAATTC GGGTTTCAG ATACGTGCTA GGTATCGTTC GGTTCGGGTT

```

FIG. 45G

```

4481 CCCAACCCAA CCCAACCCAC CCCAGTGCAG CCAACTGGCA AATAGTCTCC ACCCCCGGCA CTATCACCGT
GGTTGGGTT GGGTTGGGTG GGGTCACGTC GGTTCACCGT TTATCAGAGG TGGGGCCCGT GATAGTGGCA

4551 GAGTTGTCCG CACCACCGCA CGTCTCGCAG CCAAAAAGAA AAAAAAGG AAAAAAGG AAAAAAGG
CTCAACAGGC GTGTGTGCGT GCAGAGCGTC GGTTTTATT TTTTCTTTC TTTTCTTTC TTTTCTTTC

4621 CAGCAGGTGG GTCCGGGTGG TGGGGGCCGG AAAAGCGAGG AGGATCGCGA GCAGCGACGA GGGCCGGCCC
GTCGTCCACC CAGGCCCAGC ACCCCCGGCC TTTTCGCTCC TCCTAGCGCT CGTCTGCTGCT CCGGGCCCGG

4691 TCCCTCCGCT TCCAAGAAA CGCCCCCCAT CGCCACTATA TACATACCCC CCCCCTCTCT CCCATCCCCC
AGGAGGCGA AGGTTCTTT GCGGGGGGTA GCGGTGATAT ATGTATGGGG GGGGAGAGGA GGTAGGGGG

4761 CAACCTACC ACCACCACCA CCACCACCTC CTCCCCCTC GTGCGCGGAC GACGAGCTCC TCCCCCTCC
GTTGGGATGG TGTGTGTGGT GGTGTGTGGG GAGGGGGGAG CGACGGCCTG CTGCTCGAGG AGGGGGGAGG

4831 CCTCCGCGG CGCCGGGTAA CCACCCCGCC CCTCTCTCT TCTTCTCTCC GTTTTCTTT TCGTCTCGGT
GGGAGGCGGC GCGGCCATT GGTGGGGCGG GGAGAGGAGA AAGAAAGAGG CAAAAAGGAG AGCAGAGCCA

4901 CTCGATCTTT GGCCTTGGTA GTTTGGGTGG GCGAGAGCGG CTTCTGTCGCC CAGATCGGTG CCGGGGAGGG
GAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC GAAGCAGCGG GTCTAGCCAC GCGCCCTCCC

                               BamHI
                               ~~~~~

4971 GCGGGATCTC GCGGCTGGCG TCTCCGGGCG TGAGTCGGCC CGGATCCTCG CGGGGATGG GGCTCTCGGA
CGCCCTAGAG CGCCGACCGC AGAGGCCCGC ACTCAGCCCG GCCTAGGAGC GCCCCCTACC CCGAGAGCCT

                               BglII
                               ~~~~~

5041 TGATAGATCTT CTTTCTTTCT TCTTTTGTG GTAGAAATTG AATCCCTCAG CATGTTCAT CGGTAGTTT
ACATCTAGAA GAAAGAAAGA AGAAAAACAC CATCTTAAAC TTAGGAGTC GTAACAAGTA GCCATCAAAA

5111 TCTTTTCATG ATTTGTGACA AATGAGCCT CGTGGGGAGC TTTTGTAG GTAG
AGAAAAAGTAC TAAACACTGT TTACGTGGA GCACGCCCTG AAAAAACATC CATC

```

FIG._45H

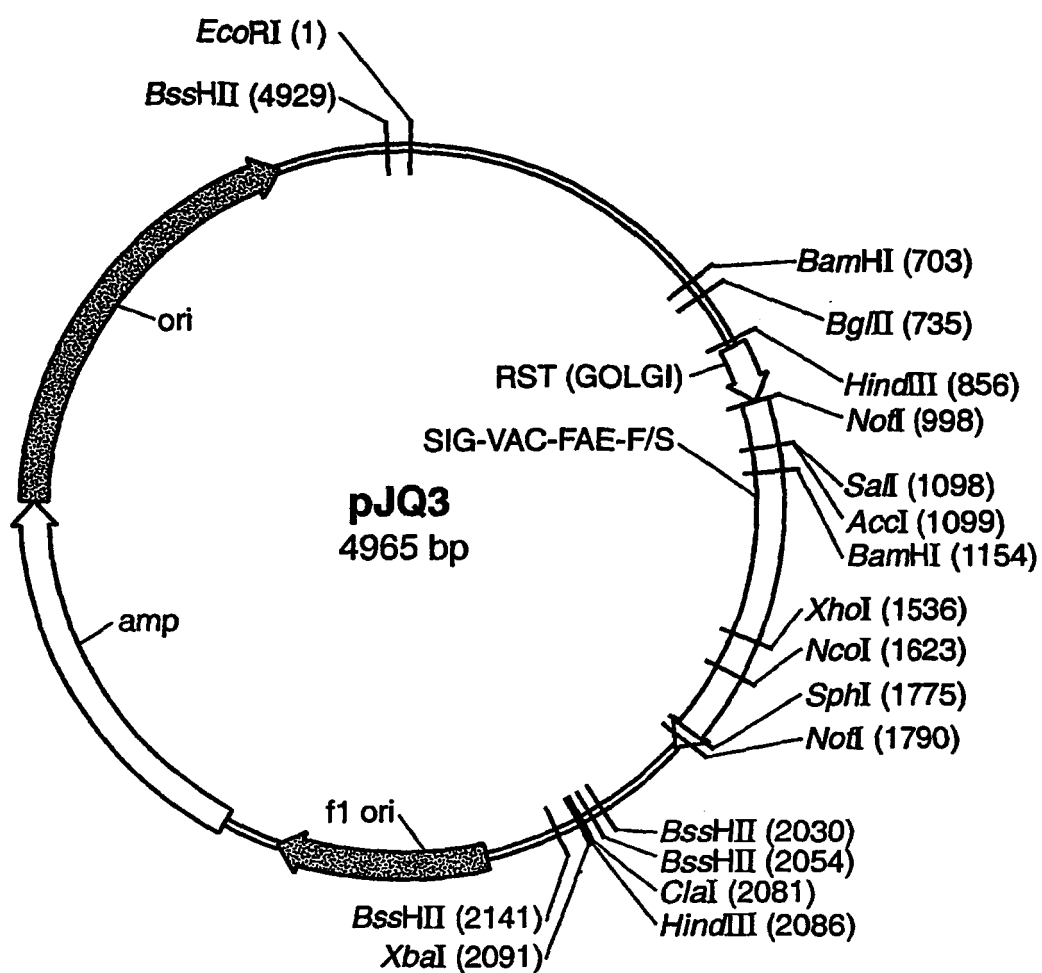


FIG. 46A

EcoRI

 1 AATTCACAA TGAACAATAA TAAGATTAAA ATAGCTTGCC CCGTTGTCAG CGATGGGTAT TTTTCTAGT
 TTAAGGTGTT ACTTGTTATT ATCTAAATT TATCGAACGG GGGCAGGTC GCTACCCATA AAAAGATCA
 71 AAAATAAAG ATAACTTAG ACTCAAAACA TTACAAAAA CAACCCCTAA AGTCCTAAG CCCAAGTGC
 TTTTATTTTC TATTGGAATC TGAGTTTGT AAATGTTTTT GTTGGGAAAT TCAGGATTTT GGGTTTCAAG
 141 TATGCACGAT CCATAGCAAG CCCAGCCCAA CCCAACCCAC CCCAGTGCAG CCAACTGGCA
 ATACGTGCTA GGTATCGTTC GGGTCGGGTT GGGTTGGGTT GGGTTCAGTC GGTTCACGTC GGTTCACCGT
 211 AATAGTCTCC ACCCCCGGCA CTATCACCGT GAGTTGTCCG CACCACGCA CGTCTCGCAG CCAAAAAA
 TTATCAGAGG TGGGGCCGT GATAGTGCA CTCAACAGGC GTGTGCGGT GCAGAGCGTC GGTTTTTT
 281 AAAAAGAAAG AAAAAAAGA AAAAAAATA CAGCAGGTGG GTCCGGTCC TGGGGGCCGG AAAAGCGAGG
 TTTTCTTTC TTTTTTTCT TTTTCTTTT GTCTGTCACC CAGGCCGAGC ACCCCCGGCC TTTTCGCTCC
 351 AGGATCGCGA GCAGCGACGA GGGCCGGCCC TCCCTCCGCT TCCAAAGAA CCCCCCCCAT CGCCACTATA
 TCCTAGCGCT CGTCGCTGCT CCGGCCGGG AGGAGGCGA AGTTTCTT GCGGGGGTA CGGTGATAT
 421 TACATACCCC CCCCTCTCTT CCCATCCCC CAACCTTACC ACCACACCA CCACCACTC CTCCCCCTC
 ATGTATGGG GGGGAGAGGA GGTAGGGGG GTTGGGATGG TGGTGTGGT GGTGTGGAG GAGGGGGAG
 491 GGTGCCGGAC GACGAGCTCC TCCCCCTCC CCCTCCGGCG CGCCCGTAA CCACCCCGCC CCTCTCTCT
 CGACGGCCTG CTGCTCGAGG AGGGGGAGG GGGAGCGGC GCGGCGCAT GTTGGGGCGG GAGAGGAGA
 561 TTCTTTCTCC GTTTTTTTT TCGTCTCGGT CTCGATCTTT GGCCTTGGTA GTTGGGTGG CGAGAGCGG
 AAGAAGAGG CAAAAAATA AGCAGAGCCA GAGCTAGAAA CCGGATCCAT CAACCCACCC CGCTCTCGCC
 631 CTTCTGTCGC CAGATCGGTG CGCGGGAGGG CGGGGATCTC GCGGCTGGCG TCTCCGGGCG TGAATCGGCC
 GAAGCAGCGG GTCTAGCCAC GCGCCCTCCC CCGCCTAGAG CCGCGACCGC AGAGGCCCGC ACTCAGCCCG

BamHI

 701 CGGATCCTCG CGGGGAATGG GGCTCTCGGA TGTAGTCTT CTTCCTTCT TCTTTTGTG GTAGAAATTG
 GCCTAGGAGC GCGCCCTACC CCGAGAGCCT ACATCTAGAA GAAAGAGA AGAAAAACAC CATCTTAAC

FIG. 46B

```

771 AATCCCTCAG CATTTGTCAT CGGTAGTTTT TCTTTTCATG ATTTGTGACA AATGCAGCCT CGTCCGGAGC
   TAGGGAGTC GTAAACAAGTA GCCATCAAAA AGAAAAGTAC TAAACACTGT TTACGTCCGA GCACGCCCTG

      HindIII
      -----
841 TTTTGTGTAG GTAGAGCTT ACCATGATCC ACACCAAGCT CAJAJAGAG TTCTCCCTCT TCATCCTCCT
   AAAAAACATC CATCTTCGAA TGTACTAGG TGTGTGTGGA GTTTTCTTTC AAGAGGGAGA AGTAGGAGCA

911 CTTCCTCCTC TTCGCCGTGA TCTGCGTGTG GAAGAAGGGC TCCGACTACG AGGCCCTCAC CCTCCAGGCC
   GAAGGAGGAG AAGCGCACT AGACGCACAC CTCTTCCCG AGGCTGATGC TCCGGGAGTG GGAGGTTCTG

      NotI
      -----
981 AAGGAGTTC AAATGCGGC CGCTCCACG CAGGGCATCT CCGAAGACCT CTACAGCCCT TTAGTCGAAA
   TTCTCAAGG TTTACCGCCG GCGAGGTGC GTCCCGTAGA GGCTTCTGGA GATGTCCGCA AATCAGCTTT

      SalI
      -----
      AcoI
      -----
1051 TGCCCACTAT CTCCCAAGCT GCTTACGCCG ACCTGTGCAA CATTCGCTCG ACTATTATCA AGGAGAGAAA
   ACCGTTGATA GAGGTTTGA CGGATGCGG TGGACACGTT GTAAGGCAGC TGATAATAGT TCCCTCTCTT

      BamHI
      -----
1121 AATTACAAT TCTCAAACTG ACATTAACGG ATGGATCCTC CGCGACGACA GCAGCAAGA AATAATCACC
   TTAATGTGA AGAGTTTGAC TGTAAATGCC TACCTAGGAG CGCTGCTGT CTCTGTTTCT TTATTAGTGG

1191 GTCTTCGGTG GCACTGGTAG TGAATACGAAT CTACAACCTG ATACTAACTA CAGCCTCACG CTTTTCGACA
   CAGAAGGCAC CGTGACCATC ACTATGCTTA GATGTTGAGC TATGATGAT GTGGAGTGC GGAAAGCTGT

1261 CCTTACCACA ATGCACGGT TGTGAAGTAC ACGGTGGATA TTATATTGGA TGGGTCTCCG TCCAGGACCA
   GGGATGGTGT TACGTTGCCA ACACTTCATG TGCCACCTAT AATATAACCT ACCCAGAGGC AGGTCTCTGT

```

FIG..46C

```

1331 AGTCGAGTCG CTTGTCJAAAC AGCAGGTTAG CCAGTATCCG GACTACCCGC TGACCTGAC CGGCCACGCC
    TCAGCTCAGC GACACAGTTG TCGTCCCATC GGTCAATGGC CTGATCGCG ACTGGCAC TG GCCGGTGGG
1401 CTCGGCGCCT CCCTGGCGGC ACTCACTGCC GCCAGCTGT CTGCGACATA CGACACATC CGCTGTACA
    GAGCCCGGA GGGACCGCG TGAOTGACCG CGGTCGACA GACGCTGTAT GCTGTTGTA GCGGACATGT
                                     XhoI
1471 CCTTCGGCGA ACCGCGCAGC GGCJAATCAG CCTTCGCGTC GTACATGAA C GATGCTTCC AAGCCTCGAG
    GGAAGCCGCT TGGCGCGTCG CCGTTAGTCC GGAAGCGCAG CATGTACTTG CTACGGAA G TTCCGAGCTC
1541 CCCAGATACG ACGCAGTATT TCGGGTCAC TCATGCCAAC GACGGCATCC CAACCTGCC CCGGGTGGAG
    GGGTCTATGC TCGTCAATAA AGGCCAGTG AGTACGGTTG CTGCCGTAGG GTTGGACCG GGGCCACCTC
                                     NcoI
1611 CAGGGGTACG CCCATGGCGG TGTAGATAC TGGAGCGTTG ATCCTTACAG CGCCAGAAC ACATTTGTCT
    GTCCCATGC GGGTACCGCC ACATCTCATG ACCTCGCAAC TAGGAATGTC GCGGTCCTTG TGTAAACAGA
1681 GCACTGGGGA TGAAGTGCAG TCGTGTGAG CCCAGGGCGG ACAGGGTGTG AATAATGCC ACACGACTTA
    CGTGACCCCT ACTTCACGTC ACGACACTCC GGGTCCCGCC TGTCCCAAC TTATTACCG TGTGCTGAAT
                                     SphI
1751 TTTTGGGATG ACGAGCGCGG CATGCACCTG GCCGTCGCG GCCCGGAAA CCACTGAA G ATGAGCTGTA
    AAAACCTAC TGTCTGCCG GTACGTGAC CGGCCAGCG CGCGCCCTT GTGACTTCC TACTCGACAT
                                     NotI
1821 AAGAAGCAGA TCGTTCAAA CATTGGCAAT AAGTTTCTT AAGATTGAAT CCTGTGCGG GTCTTGGCAT
    TTCCTCGTCT AGCAAGTTTG TAAACCTTA TTTCAAAGAA TTCTAACTTA GGACACGGC CAGAACGCTA
1891 GATTATCATA TAATTTCTGT TGAATTACGT TAAGCATGTA ATAATTACA TGTAAATGCAT GACGTTATTT
    CTAATAGTAT ATTAAGACA ACTTAATGCA ATTCGTACAT TATTAATGT ACATTACGTA CTGCATAAA

```

FIG._46D

```

1961 ATGAGATGGG TTTTATGAT TAGAGTCCG CAATTATACA TTTATATACG GATAGAAAC AAAATATAGC
      TACTCTACCC AAAATACTA ATCTCAGGC GTTAATATGT AATTATGCG CTATCTTTG TTTATATACG
      BssHII
      -----
      ClaI HindIII
      -----
2031 GCGCAAACTA GGTAAATTA TCGCGCGCG TGTCACTCAT GTTACTAGAT CGATAGCTT CTAGAGCGGC
      CGCGTTTGAT CCTATTAAAT AGCGCGCGC ACAGTAGATA CAATGATCTA GCTATTGGA GATCTCGCGC
      BssHII
      -----
      XbaI
      -----
2101 CGGTGAGCT CCAATTGCG CTATAGTAG TCGTATTACG CGCGCTCACT GCGCGTCTT TTACACGCTC
      GCCACCTCGA GGTAAAGCG GATATCACTC AGCATATGTC GCGCGAGTGA CCGGCAGCAA AATGTTGCAG
2171 GTGACTGGGA AAACCTTGG GTTACCCAA TTAATCGCT TGCAGCACAT CCCCCTTTG CCAGCTGGCG
      CACTGACCCT TTTGGGACC CAATGGGTTG AATTAGCGA ACGTCTGTGA GGGGGAAGC GGTGGAACGC
      BssHII
      -----
2241 TAATAGCGAA GAGGCCGCA CGGATCGCC TTCCCAACAG TTGCGAGCC TGAATGGCA ATGGGAGCG
      ATTATCGCTT CTCCGGGCGT GGTAGCGGG AAGGTTTGT AACGCTCGG ACTTACCGCT TACCTTGGC
2311 CCTGTAGCG GCGCATTAAG CGCGCGGGT GTGGTGGTTA CGCGCAGCGT GACCGCTACA CTTGCCAGCG
      GGGACATCGC CGCGTAATTC GCGCGGCCA CACCACCAAT GCGCGTCCA CTGGCGATGT GACCGGTGCG
      BssHII
      -----
2381 CCTAGCGCC CGCTCCTTC GCTTCTTTC CTTCTTTCT CGCCAGTTC GCGGCTTTC CCGTCAAGC
      GGGATCGCG GCGAGGAAAG CGAAGAAAG GAAGGAAAG CCGGTGCAAG CCGCCGAAAG GGGCACTTCG
2451 TCTAAATCG GGGTCCCTT TAGGTTCCG ATTTAGTGT TTACGGCAC TCGACCCCA AAAACTTGAT
      AGATTTAGC CCCGAGGGA ATCCCAAGC TAATCACA ATGCGGTGG AGCTGGGGT TTTTGAACTA
2521 TAGGGTGAT GTTCACGTAG TGGGCCATC CCCTGATGA CGGTTTTTC CCCTTTGAG TTGGAGTCCA
      ATCCCACTAC CAAGTGCATC ACCCGGTAG GGGACTATCT GCCAAGAAGC GGGAAACTGC AACCTCAGGT

```

FIG.-46E

```

2591 CGTCTTTTAA TAGTGACTC TTGTTCCAAA CTGGAACAAC ACTCAACCCCT ATCTCGGTCT ATTCCTTTTGA
    GCAAGAAATT ATCACCCTGAG AACAAAGGTTT GACCTTGTGTG TGAGTTGGGA TAGAGCCAGA TAAGAAAACCT

2661 TTTATAAGGG ATTTGCCGA TTTCCGCCCTA TTGGTTAAAA AATGAGCTGA TTTAACAAAA ATTTAACGCG
    AATATTTCCC TAAAAAGGCT AAGCCGGAT AACCAATTTT TTACTCGACT AAATGTGTTT TAAATTTGCGC

2731 AATTTAACA AAATATTAC GCTTACAATT TAGGTGGCAC TTTTCGGGGA AATGTGCGG GAACCCCTAT
    TTAAAAATGT TTTATAATTG CGAATGTAA ATCCACCCTG AAAAGCCCTT TTACACGCGC CTTGGGGATA

2801 TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC ATGAGACAAAT AACCTTGATA AATGCTTCAA
    AACAAATAAA AAGATTTATG TAAGTTTATA CATAGGCGAG TACTCTGTTA TTGGGACTAT TTACGAAGTT

2871 TAAATATTGA AAAGGAAGAG TATGAGTATT CAACATTTCC GTGTGCCCC TATTCCCTTT TTGCGGGCAT
    ATTATAACTT TTTCCCTTCT ATACTCATAA GTTGTAAGG CACAGCGGA ATAAGGAAA AACCGCCGT

2941 TTTGCCCTCC TGTTTTGTG CACCCAGAAA CGCTGGTGAA AGTAAAGAT GCTGAAGATC AGTTGGGTGC
    AAACGGAAGG ACAAAAACGA GTGGGTCTTT GCGACCACCT TCATTTTCTA CGACTTCTAG TCAACCCACG

3011 ACGAGTGGGT TACATCGAAC TGGATCTCAA CAGCGGTAAG ATCTTGAGA GTTTTCGCC CCAAGAACGT
    TGCTCACCCA ATGTAGCTG ACCTAGAGTT GTCGCCATTC TAGGAACCTT CAAAAGCGGG GCTTCTTGCA

3081 TTTCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTATTGAC GCCGGGCAAG
    AAAGGTTACT ACTCGTGAAA ATTTCAAGAC GATACACCGC GCCATAATAG GGCAATACTG CCGCCCGTTC

3151 AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTGAGTAC TCACCAGTCA CAGAAAAGCA
    TCGTTGAGCC AGCGCGTAT GTGATNAGAG TCTTACTGAA CCAACTCATG AGTGTGAGT GTCTTTTCGT

3221 TCTTACGGAT GGCATGACAG TAAGAGMAT ATGCAGTGT GCCATAACCA TGAGTGATA CACTGCGGCC
    AGAATGCCCTA CCGTACTGTC ATTCCTTAA TACGTACGA CGGTATTGGT ACTCACTATT GTGACGCGCG

3291 AACTTACTTC TGACAAACGAT CGGAGGACCG AAGGAGCTAA CCGCTTTTTT GCACAACTG GGGGATCATG
    TTGAATGAAG ACTGTTGCTA GCCTCCTGGC TTCTCTGATT GCGAAAAAA CGTGTGTGAC CCCCTAGTAC

3361 TAACCTCGCT TGATCGTTGG GAACCGGAGC TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT
    ATTGAGCGGA ACTAGCAACC CTTGGCCCTCG ACTTACTTCG GTATGGTTTG CTGCTCGCAC TGTTGTGCTA

```

FIG._46F

3431 GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAACT GGCGAACTAC TTACTCTAGC TTCCCGGCAA
 CGGACATCGT TACCGTTGTT GCAACGCCGT TGATAATTGA CCGCTTGATG AATGAGATCG AAGGGCCGTT
 3501 CAATTAATAG ACTGGATGGA GGCGGATAAA GTTCGAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT
 GTTAATTATC TGACCTACCT CCGCCTATTT CAACGTCCTG GTGAGACGCG GAGCCGGGAA GGCCGACCGA
 3571 GGTTTATTGC TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC ATTGCAGCAC TGGGGCCAGA
 CCAATAACG ACTATTAGA CCTCGGCCAC TCGCACCCAG NGCGCCATAG TAACGTCTGTG ACCCCGGTCT
 3641 TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGG AGTCAGGGCA CTATGGATGA ACGAAATAGA
 ACCATTCGGG AGGCATAGC ATCAATAGAT GTGCTGCCCC TCAGTCCGTT GATACCTACT TGCATTATCT
 3711 CAGATCGCTG AGATAGGTGC CTCACGTGAT AAGCATGGGT AACTGTGAGA CCAAGTTTAC TCATATATAC
 GTCTAGCGAC TCTATCCACG GAGTGACTAA TTCGTAAACA TTGACAGTCT GGTTCAAATG AGTATATATG
 3781 TTTAGATTGA TTTAAAACTT CATTTTAAAT TTTAAAGGAT CTAGGTGAAG ATCCTTTTGT ATAACTCTCAT
 AAATCTAACT AAATTTTGAA GTAAAAATTA AATTTTCTTA GATCCACTTC TAGGAAAAAC TATTAGAGTA
 3851 GACCAAATC CCTTAACGTG AGTTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT
 CTGGTTTAG GGAATTGCAC TCAAAAGCAA GGTGACTCGC AGTCTGGGGC ATCTTTTCTA GTTTCCTAGA
 3921 TCTTGAGATC CTTTTTTCCT GCGCGTAATC TGCTGCTTGC AAACAAAAAA ACCACCGCTA CCAGCGGTGG
 AGAACTCTAG GAAAAAAGA CCGCGCATTA ACGACGACG TTTGTTTTTT TGGTGGCGAT GGTGCGCCAC
 3991 TTTGTTTGGC GGATCAAGAG CTACCAACTC TTTTTCGGAA GGTAACCTGGC TTCAGCAGAG CGCAGATACC
 AAACAAACGG CCTAGTTCTC GATGGTTGAG AAAAAGGCTT CCATTGACCG AAGTCGTCTC GCGTCTATGG
 4061 AAATACTGTC CTTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAACT CTGTAGCACC GCTTACATAC
 TTTATGACAG GAAGATCACA TCGGCATCAA TCCGGTGGTG AAGTCTTGA GACATCGTGG CGGATGTATG
 4131 CTCGCTCTGC TAACTCTGTT ACCAGTGGCT GCTGCCAGTG GCGATAAGTC GTGCTTTACC GGGTTGGACT
 GAGCGAGACG ATTAGGACAA TGGTCAACCGA CGACGGTCAC CGCTATTTCAG CACAGAAATGG CCCAACTGA
 4201 CAAGACGATA GTTACCGGAT AAGCGGCAGC GGTGCGGCTG AACGGGGGGT TCGTGCACAC AGCCCCAGCTT
 GTTCTGCTAT CAATGGCCTA TTCCGCGTCTG CCAGCCCGAC TTGCCCCCCA AGCACGTGTG TCGGGTCGAA

FIG. 46G

```

4271 GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG AAAGCGCCAC GCTTCCCGAA
CCTCGCTTGC TGGATGTGGC TTGACTCTAT GGATGTGCA CTCGATCTC TTTCGGGTG CGAAGGCGTT

4341 GGGAGAAAGG CGGACAGGTA TCGGGTAAGC GGCAGGGTCG GAACAGGAGA GCGCAGGAG GAGCTTCCAG
CCCTCTTTCC GCCTGTCCAT AGGCCATTCC CCGTCCAGC CTTCCTCTCT CCGTGTCTCC CTCGAAGGTC

4411 GGGGAACGC CTGGTATCTT TATAGTCTG TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTGTG
CCCTTTTCCG GACCATAGAA ATATCAGGAC AGCCCAAGC GGTGAGACT GAACTCGCAG CTAAAAACAC

4481 ATGCTCGTCA GGGGGCGGA GCCTATGAA AAACGCCAGC AACGGGCTT TTTACGGTT CTGGGCTTT
TACGAGCAGT CCCCCGCCCT CGGATACCTT TTTGGGTCG TTGCGCCGA AAATGCCAA GGAACCGAAA

4551 TGCTGGCCTT TTGCTCACAT GTTCTTTCT GCGTTATCCC CTGATTTCTG GATTAACCGT ATTACGCGT
ACGACCGGAA AACGAGTGA CAAGAAAGGA CGCAATAGG GACTAAGACA CCTATTGCA TAATGGCGA

4621 TTGAGTGAGC TGAATACGCT CCGCGCAGCC GAACGACGGA GCGCAGCGG TCAGTGAGCG AGGAAGCGA
AACTCACTCG ACTATGGCGA GCGGCGTCGG CTTCCTGGCT CCGTCTGCTC AGTCACTCG TCCTTCGCTT

4691 AGAGCGCCCA ATACGCAAC CGCTCTCCC CCGCGGTTGG CCGATTCATT AATGCACTG GCAAGACAGG
TCTCGCGGT TATCGGTTG GCGGAGGG GCGCGCAACC GGCCTAAGTAA TTACGTGAC CGTGCTGTCC

4761 TTTCGGACT GGAAGCGGG CAGTGAGCGC AACGCAATTA ATGTAGTTA GCTCACTCAT TAGGCACCCC
AAAGGGCTGA CCTTTGCCC GTCACTCGG TTGCTTAAT TACACTCAAT CGATGAGTA ATCCGTGGG

4831 AGGCTTTACA CTTTAGCTT CCGGCTCGTA TGTGTGTGG AATTGTAGC GATTAACAT TTCACACAGG
TCCGAAATGT GAATACGAA GCGCGAGCAT ACAACACACC TTACACTCG CCTATTGTTA AAGTGTGTCC

BssHII
-----
EcoRI

4901 AAACAGCTAT GACCATGATT ACGCCAGCG CGCAATTAC CCTCACTAAA GGAACAAAA GCTGG
TTGTGCGATA CTGGTACTAA TCGGTTCCG CGGTTAATG GAGTGAATTT CCCTGTGTTT CGACC

```

FIG._46H

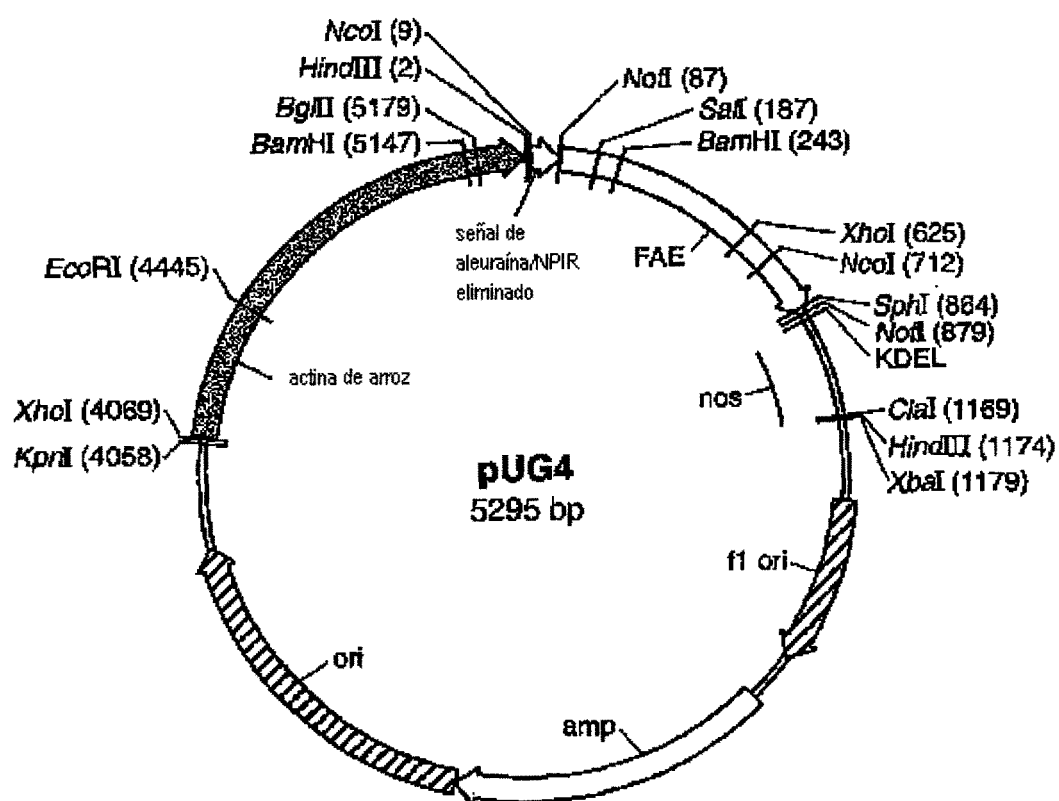


FIG._47A

```

      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 CCGCGTCCTC CTCTGGGCG TCGCCGTGCT GGCACAGCC 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

      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 CATTACGGA TGGATCCTCC GCGACGACAG CAGCAAGAA 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  TCCTCCGTGG CACTGGTAGT GATACGAATC TACAACCTGA TACTAATAC 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 GTGAGTACA 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 TTGTCAACA GCAGGTAGC 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 TCGGACATAC 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 GCAATCAGGC CTCGCGTCG TACATGAACG ATGCCCTTCCA 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 AAACCTGCCC CCGGTGGAGC

      NcoI
      ~~~~~

```

FIG. 47B

```

.   G   Y   A   H   G   G   V   E   Y   W   S   V   D   P   Y   S   A   Q   N   T   F   V   C   .
701 AGGGGTACGC CCATGGCGGT GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCCAGAACA CATTGTGCTG
.   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 GCTGTGAGGC CCAGGCGCGA CAGGGTGTGA ATAATCGCA CACGACTTAT

                               SphI
F   G   M   T   S   G   A   C   T   W   P   V   A   A   A   E   P   L   K   D   E   L   *
841 TTGGGATGA CGAGCGGCG ATGCACCTGG CCGTGCGG CCGCGAAC ACTGAGGAT GAGCTGTAA
911 GAAGCAGATC GTTCAAAACAT TTGGCAATAA AGTTCTTAA GATTGAATCC TGTGCGCGT CTTGCCGATGA
981 TTATCATATA ATTCTGTG AATTACGTTA AGCATGTAT AATTAACATG TAATGCATGA CGTTATTTAT
1051 GAGATGGGTT TTTATGATTA GAGTCCCGCA ATTATACATT TAATACGCGA TAGAAACAA AATATAGCGC

                               NotI
                               ClaI
                               XbaI
1121 GCAAACTAGG ATAAATTATC GCGCGCGGTG TCATCTATGT TACTAGATCG ATAACTTCT AGAGCGGCGG
1191 GTGGAGCTCC AATTCGCCCT ATAGTGAGTC GTATTAAGCG GTCTACTGG CCGTCGTTTT ACNACGTCGT
1261 GACTGGGAAA ACCCTGGCGT TACCCAACTT AATCGCCTTG CAGCACATCC CCCTTTCGCC AGCTGGCGTA
1331 ATAGCGAAGA GGCCTGCACC GATCGCCTT CCCAACAGTT GCGCAGCCTG AATGCGGAAT GGGACGCGCC
1401 CTGTAGCGGC GCATTAAGCG CCGCGGGTGT GTGTGTTACG CCGCTACACT TGCCAGCGCC
1471 CTAGCGCCCG CTCCTTTCGC TTCTTTCCTT TCTTGTCTTG CCAGGTTCCG CGGCTTCCC CGTCAAGCTC
1541 TAAATCGGGG GCTCCCTTTA GGGTTCCGAT TTAGTGTCTT ACGGCACCTC GACCCCAAAA AACTTGATTA
1611 GGGTGATGGT TCACGTAGTG GGCATCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT GAGTCCACG
1681 TTCTTTAATA GTGGACTCTT GTTCCAAACT GGAACAACAC TCACCTCTAT CTCGGTCTAT TCTTTGTATT
1751 TATAAGGAT TTTGCCGATT TCGGCTATT GGTAAAAA TGAGCTGATT TAACAAAAAT TTAACGCGAA
1821 TTTTAAACAA ATATTAACGC TTACAATTTA GGTGGCACCT TTCCGGGAAA TGTGCGCGGA ACCCTATT
1891 GTTTATTTT CTAATACAT TGAGTATTCA ACATTTCCGT GTGCGCCCTA TTCCCTTTT TCGGCAAT
1961 ATATTGAAA AGGAAGATG CCCAGAAAAC CTGGTGAAG TAAAAGATGC TGAAGATCAG TTGGGTGCAC
2031 TGCTTCTCTG TTTTGTCTCA CCAGAAACG CCGTGAAGAT CTTGAGAGT TTTGCGCCCG AAGAACGTTT
2101 GAGTGGGTTA CATCGAATG GATCTCAACA GCGTAAAGAT CTTGAGAGT TTTGCGCCCG AAGAACGTTT
2171 TCCAATGATG AGCACTTTTA AAGTCTGCT ATGTGCGCG GTATTATCCC GTATTGACGC CGGGCAAGAG
2241 CAACTCGGTC GCGCATACA CTATTTCTCAG AATGACTTGG TTGAGTACTC ACCAGTCACA GAAAAGCATC
2311 TTACGGATGG CATGACAGTA AGAGAATTAT GCAGTGCTGC CATTAACCATG AGTGATAACA CTGCGGCCAA
2381 CTTACTTCTG ACAACGATCG GAGGACGAA GAGCTAACCC GCTTTTTCG ACAACATGGG GGATCATGTA
2451 ACTCGCCTTG ATCGTTGGGA ACCGAGCTG AATGAAGCA TACCAAACGA CGAGCGTGAC ACCACGATGC
2521 CTGTAGCAAT GGCAACAACG TTGCGCAAACT TATTAACTGG CGAACTACTT ACTCTAGCTT CCCGGCAACA

```

FIG.-47C

2591 ATTAATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA CTTCGTGCGT CGCCCTTCC GCGTGGCTGG
 2661 TTTATTGCTG ATAAATCTGG AGCCGGTGAG CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG
 2731 GTAAGCCCTC CCGTATCGTA GTTATCTACA CGACGGGGAG TCAGGCAACT ATGATGAAAC GAATAGACA
 2801 GATCGCTGAG ATAGGTGCTT CACTGATTAA GCATTGGTAA CTTGTAGACC AAGTTTACTC ATATATACTT
 2871 TAGATTGATT TAAACTTCA TTTTAAATTT AAAAGGATCT AGGTGAAGAT CCTTTTGTAT AATCTCATGA
 2941 CCAAAATCCC TTAACGTGAG TTTTCGTTC CCGTAACTCC ACTGAGCGTC AGACCCCGTA GAAAAGATCA AAGGATCTTC
 3011 TTGAGATCCT TTTTTCCTGC GCGTAACTCC CCGTAACTCC CTGCTTGCAG ACACAAAAC AGCGGTGGTT
 3081 TGTTCGCCG ATCAAGAGCT ACCAACTCTT TTTCCGAAGG TAACTGGCTT CAGCAGAGCG CAGATACCAA
 3151 ATACTGTCTT TCTAGTGTAG CCGTAGTTAG GCCACCACTT CAAGAATCTT GTAGCACCGC CTACATACCT
 3221 CGCTCTGCTA ATCCTGTATC CAGTGGCTGC TGCCAGTGGC GATAAGTCGT GTCTTACCAG GTTGGACTCA
 3291 AGACGNTAGT TACCGGATAA GCGCAGCGG TCGGGGTGAA CCGGGGTTC GTGCACACAG CCCAGCTTGG
 3361 AGCGAACGAC CTACACCGAA CTGAGATACC TACAGCGTGA CAGGCTCGA ACAGGAGGC GCCTCCAGGG
 3431 GAGNAAGGCG GACAGGTATC CGGTAAAGCG TAGTCTCTGC GGGTTTCGCC ACCTCTGACT TGGCGTGGAT
 3501 GGAACGCCCT GGTATCTTTA GGTATCTTTA TAGTCTCTGC CTATGGAAAA CGCCAGCAA CGCGCTTTTG
 3571 GCTCGTCAGG GGGCGGAGC CTATGGAAAA TCTTCTCTGC GTTATCCCTT GATTCTGTGG ATAACCGTAT TACCGCTTTT
 3641 CTGGCCCTTT GCTCAGATGT CCGCAGCTCG CCGCAGCGAG ACAGCCGAGC GCAGCGAGTC AGTGAAGCGAG GAAGCGAAG
 3711 GAGTGAGCTG ATACCGCTCG CCGCAGCTCG CCGCAGCTCG CCGCAGCTCG GATTCATTAA TGCAGCTGGC ACGACAGGTT
 3781 AGCGCCCAAT ACGCAACCG CCGCAGCTCG CCGCAGCTCG CCGCAGCTCG GATTCATTAA TGCAGCTGGC ACGACAGGTT
 3851 TCCCGACTGG AAGCGGGCA GTGAGCGCA CGCAATTAAT GTGAGTTAGC TCACATCTTA GGCACCCAG
 3921 GCTTTACACT TTATGCTTCC GGTCTGTATG TTGTGTGGAA TTGTGAGCGG ATAACAAATTT CACACAGGAA
 KpnI

 3991 ACAGCTATGA CCATGATTAC GCCAAGCGCG CAATTAAACC TCACATAAAGG GAACAAAAGC TGGGTACCGG
 XhoI
 ~~~~~~  
 4061 GCGCCCTCTC GAGGTCAATC ATATGCTTGA GAAGAGAGTC GGGATAGTCC AAAATAAAAC AAGGTAAGA  
 4131 TTACCTGGTC AAAAGTGAAA ACATCAGTTA AAGGTGGTA TAAGTAAAT ATCGGTAATA AAGGTGGCC  
 4201 CAAAGTGAAA TTTACTCTTT TCTACTATTA TAAAAATGA GGATGTTTTG TCGGTACTTT GATACGTCA  
 4271 TTTTGTATGA ATTGGTTTTT AAGTTTATTC GCGATTGGA AATGCATATC TGTATTTGAG TCGGTTTTTA  
 4341 AGTTCGTTCG TTTTGTAAAT ACAGAGGGAT TTGTATAGA AATATCTTTA AAAACCCCAT ATGCTAAATTT  
 EcoRI  
 ~~~~~~  
 4411 GACATATATT TTGAGAAAAA TATATATTCA GCGCAATTC ACAATGAACA ATAATAAGAT TAAAAAGCT
 4481 TGCCCCGTT GCAGCGATGG GTATTTTCTC TAGTAAATA AAGATATAAC TTAGACTCAA AACATTTACA
 4551 AAACAAACC CTAAAGTCTT AAGCCCCAA GTGCTATGCA CGATCCATAG CAAGCCAGC CCAACCAAC
 4621 CCAACCCAAC CCACCCCAAGT GCAGCCCACT GGCATAATGT CTCCACCCCT GGCACATATCA CCGTGAAGTTG

FIG._47D

```

4691 TCCGCACCAC CGCACGTCTC GCAGCCAAA AAAA AAAAG AAGAAAAAGA AAAACAGCAG
4761 GTGGGTCCGG GTCGTGGGGG CCGGAAAAGC GAGGAGGATC GCGAGCAGCG ACGAGGCCCG GCCCTCCCTC
4831 CGCTTCCAAA GAAACGCCCC CCATCGCCAC TATATACATA CCCCCCCTC TCCCTCCCATC CCCCCAACCC
4901 TACCACCACC ACCACCACCA CCTCCTCCCC CCTCGCTGCC GGACGACGAG CTCCCTCCCC CTCCCCCTCC
4971 GCCGCCGCCG GTAACCAACC CCCCCCTCTC CTCCTTCTTT CTCCTTTTTC TTTTTCGTCT CCGTC'TCGAT
5041 CTTTGGCCCTT GGTAGTTTGG GTGGGCGAGA GCGGCTTTCGT CGCCACAGATC GGTGCGCGGG AGGGGCGGGA
                                     BamHI
                                     ~~~~~~
5111 TCTCGCGGCT GCGTCTCTCC GCGGTGAGTC GGCCCGGATC CTCGCGGGGA ATGGGCTCT CGGATGTAGA
                                     BglII
                                     ~~~~
5181 TCTTCTTCTT TCTTCTTCTT TGTGGTAGAA TTGGAATCCC TCAGCATTGT TCATCGGTAG TTTTTC'TTTT
5251 CATGATTTGT GACAAATGCA GCCTCGTGCG GAGCTTTTTC GTAGC

```

FIG. 47E

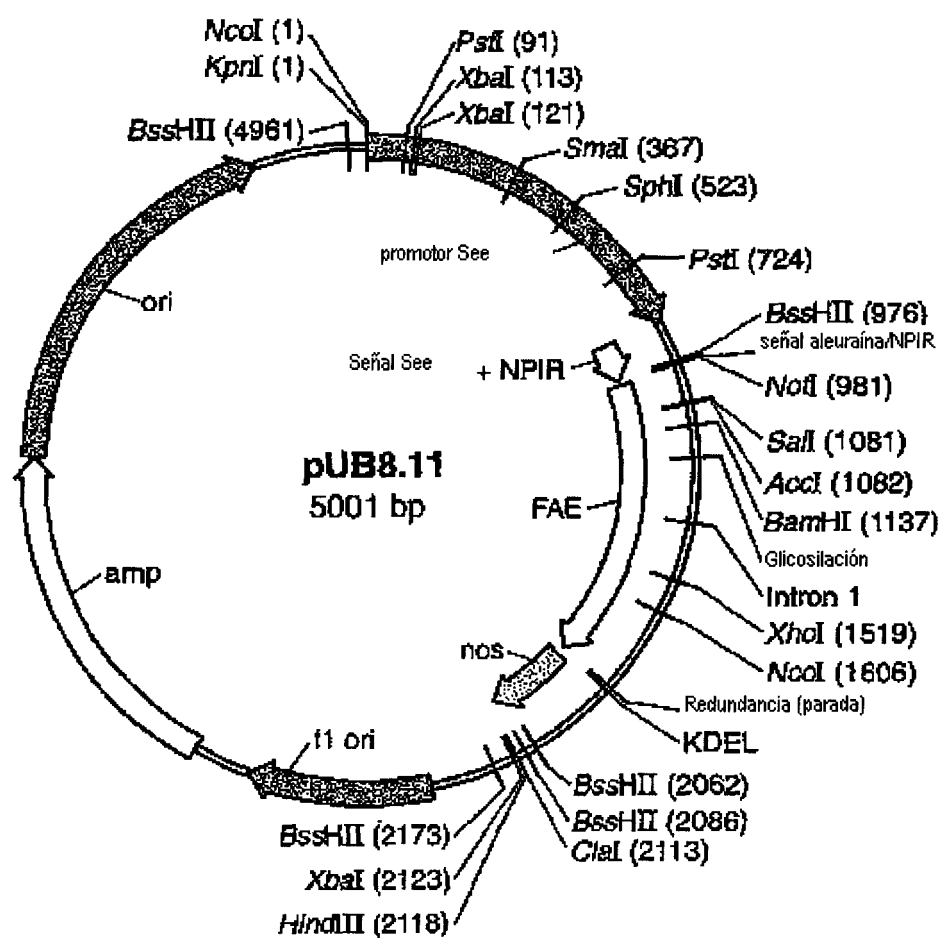


FIG._48A

```

NcoI
~~~~~
KpnI
~
1  CATGGGCCAG  GTATAAATTAT  GGGATATCTC  AAGCAAAATA  TCGAATATTC  ACCATTGGCT  ACAATATCTG
   GTACCCGGTC  CATATTAAATA  CCTATATAG  TTCGTTTATT  AGCTTTATAG  TGGTAACCGA  TGTATATAGC

           PstI
           ~~~~~
           XbaI
           ~~~~~
           XbaI
           ~~~~~
71  AGCTCCGAGT  TCTGACTGCA  GTCTGGATGA  CGCGTGTTGT  ATCTAGAACT  CTAGATAGCA  CAGCCACAGC
   TCGAGGCTCA  AGACTGACGT  CAGACCTACT  GCGCACAAAC  TAGATCTTGA  GATCTATCGT  GTCGGTGTGG

141  ACCTACAGGA  GTGCGACACT  TGTGGACTGT  AGTAGTGTGG  GAGACGGAGC  TCTTTCCTAC  CTCCTGACGT
   TGGATGTCC  CACGCTGTGA  ACACCTGACA  TCATCACAAC  CTCTGCCCTG  AGAAAGGATG  GAGGACTGCA

211  TGCCGCCGTT  GTCCATTCCA  ACGGCATCAC  TCTCAACCAA  TCACGGCGTC  CCAACAAAT  ATCGTCCCCC
   ACGCGGCCAA  CAGGTAAGGT  TGCCGTAGTG  AGAGTTGGTT  AGTCCCGAG  GGTGTGTTTA  TAGCAGGGGG

281  ATGTCCTTGG  GGAGAGAGAG  TACATACATG  CTGTGCGGCC  GTTTTGTGCT  GAAJCTCGCT  TCCACTGGCC
   TACAGAACCG  CCTCTCTCTC  ATGTATGTAC  GACAGCGCGG  CAAAAACAGA  CTTAGAGCGA  AGGTGACCGG

           SmaI
           ~~~~~
351  AATCAGCTCA  GCTCCCGGGA  GCTCACTCAT  TCAAGATCCC  ATCGTCTGTG  TCACCCCTGG  CGTCATGGGA
   TTAGTCGAGT  CGAGGCCCTT  CGAGTGAGTA  AGTCTAGGG  TAGCAGCAGC  AGTGGGGACC  GCAGTACCCCT

421  TGGAAAAGAA  CCTCCGTTGC  TCGGATGAGT  CAGCCATATC  CCCGAACAGA  GTACTGCAAG  ATAACCCAT
   ACCTTTCTTT  GGAGGCAACG  AGCCTACTCA  GTCGGGTATAG  GGGCTTGCT  CATGACGTTT  TATTGGGTTA

           SphI
           ~~~~~
491  TCAGATTCCC  CCAATAGAGA  AAGTATAGCA  TGCTTTCGGG  TTTTGTGTTG  CTTAATTGAC  TTTATTTTGG
   AGCTTAAGGG  GGTATCTCT  TTCATATCGT  ACGAAGGCC  AAAACAACC  GAATTAACTG  AAATAAAAC

561  TTGGAGTTGA  ATGCTGATTT  GTTGTGTAAA  ATGCCCAACC  ATCTGAATAT  CGAGACGGAT  AATAGGCTGG
   AACCTCAACT  TACGACTAAA  CAACACATTT  TACGGGTTGG  TAGACTTATA  GCTCTGCCCT  TTATCCGACC

```

FIG._48B

631 CTAATTAAAT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
GATTAAATTA ATATCGTTCT AAGACATCAC GTGTAGCGTT TATAGAAAGA CCCGTAATGT CGACCTCCGA

PstI
~~~~~

701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC  
AGTAGTCGGA CTTTGTGAGA CGTCTCGGAC TTCGTTTACC ACTTCGCACC GCTACTCTAC CCATATTTTG

771 CCCCAGCACC GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCCTCGC TCCCCCTGCC GGACGACCCA  
GGGGCCGTGG CCTTGGCTC GAGGGCGGAT GGTCATGGTA GAGCGGAGCG AGGGGACGG CCTGCTGGGT

841 GTAAATTAAT GTTGCCCACT CGCCGGCGAG ATGGCCACG GCCGCATCCT CTCTTTGGCG CTCGCCGTCT  
CATTTTATGA CAACGGGTGA GCGGCCGCTC TACCGGGTGC CGCGGTAGGA GAAGAAACCGC GAGCGGCAGA

BspHII  
~~~~~  
NotI
~~

911 TGGCCACCGC CGCGGTGGCC GCCGCATCNT TGGCGGACTC CAACCCGATC CGGCCCGTCA CCGAGCGCGC
ACCGGTGGCG GCGCCACCGG CGGCGTAGNA ACCGCCTGAG GTTGGGCTAG GCCGGGCAGT GGCTCGCGCG

NotI
~~~~~

981 GGCCGCCCTCC ACGCAGGGCA TCTCCGAAGA CCTCTACAGC CGTTTAGTGG AAATGGCCAC TATCTCCAA  
CCGGCGGAGG TCGTCCCGT AGAGGCTTCT GGAGATGTCT GCAAAATCAGC TTACCCGGTG ATAGAGGGTT

Sali  
~~~~~  
AccI
~~~~~

1051 GCTGCCCTACG CCGACCTGTG CAACATTCCG TCGACTATTA TCAAGGGAGA GAAAATTAC AATTCACAA  
CGACGGATGC GGCTGGACAC GTTGTAAAGC AGCTGATAT AGTTCCCTCT CTTTAAATG TTAAGAGTTT

FIG.-48C

```

BamHI
-----
1121 CTGACATTAA CGGATGGATC CTCCGCGACG ACAGCAGCAA AGAATAATC ACCGCTCTCC GTGGCACTGG
    GACTGTAATT GCCTACCTAG GAGGCGCTGC TGTGTCGTT TCTTTATTAG TGGCAGAAAG CACCGTGACC

1191 TAGTGATACG AATCTACAAC TCGATACTAA CTACACCCCTC ACGCCTTTCC ACACCCCTACC ACATGCAAC
    ATCACTATGC TTAGATGTTG AGCTATGATT GATGTGGGAG TCGGGAAGC TGTGGGATGG TGTACGTTG

1261 GGTGTGGAAG TACACGGGTGG ATATTATATT GGATGGGTCT CCGTCCAGGA CCAAGTCGAG TCGCTTGTCA
    CCAACACTTC ATGTGCCACC TATAATATAA CCTACCCAGA GGCAGGTCTT GGTTCAGCTC AGCGAACAGT

1331 AACAGCAGGT TAGCCAGTAT CCGGACTACG CGCTGACCGT GACCGGCCAC KCCCTCGGCG CCTCCCTGGC
    TTGTCTGTCCA ATCGGTTCATA GGCCTGATGC GCGACTGGCA CTGGCCGGTG MGGGAGCCGC GGAGGGACCG

1401 GGCAC'TCACT GCCGCCCAGC TGTCTGCGAC ATACGACAAAC ATCCGCC'TGT ACACCTTTCGG CGAACC'GCGC
    CCGTGAGTGA CGCGGGGTCTG ACAGACGCTG TATGCTGTGTG TAGGCGGACA TGTGGAAAGCC GCTTGGCGCG

XhoI
-----
1471 AGCGGCAATC AGGCTTTCGC GTCTGACATG AACGATGCCT TCCAAAGCCTC GAGCCCAAGT ACGACGCAGT
    TCGCCGTTAG TCCGGAAGCG CAGCATGTAC TTGCTACGGA AGGTTCCGAG CTCGGGTCTA TGTGCGTCA

NcoI
-----
1541 ATTTCCGGGT CACTCATGCC AACGACGGCA TCCCAAACCT GCCCCCGGTG GAGCAGGGGT ACGCCCATGG
    TAAAGGCCCA GTGAGTACGG TTGCTGCCGT AGGGTTTGA CGGGGGCCAC CTCGTCCCA TCGGGGTACC

1611 CGGTGTAGAG TACTGGAGCG TTGATCTTIA CAGCGCCCG CAGCATTTC TCTGCACTGG GGAATGAAAGT
    GCCACATCTC ATGACCTCGC AACTAGGAAT GTCGCGGGTC TTGTGTAAAC AGACGTGACC CCTACTTCAC

1681 CAGTGTGTG AGGCCCAGGG CGGACAGGGT GTGAATAATG CGCACACGAC TTATTTTGGG ATGACGAGCG
    GTCACGACAC TCCGGGTCCC GCCTGTCCCA CACTTATTAC GCGTGTGCTG AATAAAACCC TACTGCTTCG

1751 GAGCCTGTAC ATGGTGATCA GTCA'TTTCAG CCTCCCGGAG TGTACCAGGA AAGATGGATG TCTGGGAG
    CTCGGACATG TACCAC'TAGT CAGTAAAGTC GGAGGGGCTC ACATGGTCTT TCTTACCTAC AGGACCTCTC

```

**FIG.\_48D**

```

1821 GGGGCGCGT AACCACTGAA GGATGAGCTG TAAAGAAGCA GATCGTTCAA ACATTGGCA ATAAAGTTTC
    CCCCAGCGCA TTGGTGACTT CCTACTCGAC ATTCTTCGT CTAGCAAGTT TGTAAACCGT TATTCAAAAG

1891 TTAAGATTGA ATCCTGTTGC CGGTCTTGGG ATGATTATCA TATAATTTCT GTTGAATTAC GTTAAGCATG
    AATTCTAACT TAGGACAAAG GCCAGAACGC TACTAATAGT ATATTAAAGA CAACITTAATG CAATTCGTAC

1961 TAATAATTAA CATTGAATGC ATGACGTTAT TTATGAGATG GGTTTTATG ATTAGAGTCC CGCAATTATA
    ATTATTAAAT GTACATTACG TACTGCAATA AATACTCTAC CCAAAATATC TAATCTCAGG CGTTAATAT

                BssHII
                ~~~~~
2031 CATTTAATAC GCGATAGAAA ACAAAATATA GCGCGCAAC TAGGATAAAT TATCGCGCGC GGTGTCACT
 GTAAATTATG CGCTATCTTT TGTTTTATAT CGCGCGTTTG ATCCTATTTA ATAGCGCGCG CCACAGTAGA

 BssHII
                ~~~~~
                XbaI
                ~~~~~
 ClaI HindIII
                ~~~~~
2101 ATGTTACTAG ATCGATAAGC TTCTAGAGCG GCCGGTGGAG CTCCAATTCG CCTATAGTG AGTCGTATTTA
    TACAATGATC TAGCTATTTC AGATCTCGC CGGCCACCTC GAGGTTAAGC GGGATATCAC TCAGCATAAAT

                BssHII
                ~~~~~
2171 CGCGCGCTCA CTGGCCGTCTG TTTTACAACG TCGTGACTGG GAAAGCCCTG GCGTTACCCA ACTTAATCGC
 GCGCGCGAGT GACCGGCAGC AATAATGTTGC AGCACTGACC CTTTGGGAC CGCAATGGGT TGAATTAGCG

2241 CTTCGAGCAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC
 GACGTCGTG TAGGGGAAA GCGGTCGACC GCATTATCGC TTCTCGGGC GTGGCTAGCG GGAAGGTTG

2311 AGTTGCGCAG CCTGAATGGC GAATGGGACG CGCCCTGTAG CGGCGCATTA AGCGCGGCGG GTGTGGTGGT
 TCACGCGGTC GGACTTACCG CTTACCCCTG CGGGGACATC GCGCGTAAT TCGCGCGCGC CACACCACTA

2381 TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCTTT TCGCTTTCTT CCCTTCCTTT
 ATCGCGGTCG CACTGGCGAT GTGACGGTTC GCGGATCGC GGGCGAGGAA AGCGAAGAAA

```

**FIG. 48E**

2451 CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC GGGGGCTCCC TTTAGGGTTC CGATTTAGTG  
 GAGCGGTGCA AGCGGCCGAA AGGGGCAGTT CGAGATTAG CCCCCGAGGG AAATCCCAAG GCTAAATCAC  
 2521 CTTTACGGCA CCTCGACCCC AAAAALACTTG ATTAGGTGA TGGTTCAGT AGTGGGCCAT CGCCCTGATA  
 GAAATGCCGT GGAGCTGGGG TTTTTFGAAC TAAATCCACT ACCAAGTGCA TCACCCCGGT TCAGGACTAT  
 2591 GACGGTTTTT CGCCCTTTGA CGTTGGAGTC CAGGTCTTTT AATAGTGGAC TCTTGTTCCTA AACTGGAAACA  
 CTGCCAAAAA GCGGGAAACT GCACCTTCAG GTGCAAGAAA TTATCACCTG AGAACCAAGGT TTGACCTTGT  
 2661 ACACTCAAAC CTATCTCGGT CTATTCCTTT GATTATAGA GGAATTTGCC GATTTGGCC TATTTGTTAA  
 TGTGAGTTGG GATAGAGCCA GATAGAAAA CTAAATATTC CCTAAAAAGG CTAAAGCCGG ATAAACCAAT  
 2731 AAAATGAGCT GATTAAACAA AAATTTAACG CGAATTTTAA CAAAATATTA ACGCTTACAA TTTAGGTGGC  
 TTTTACTCGA CTAAATGTGT TTTAAATGTC GCTTAAAAAT GTTTTATAAT TCGGAATGTT AAATCCACCG  
 2801 ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT ACATTCAAAT ATGTATCCGC  
 TGAAGAGCCC CTTTACACGC GCCTTGGGA TAAACAAATA AAAAGATTTA TGTAAAGTTA TACATAGGCG  
 2871 TCATGAGACA ATAAACCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG AGTATGAGTA TTCATCATTT  
 AGTACTCTGT TATTGGGACT ATTTACGAAG TTATATATAAC TTTTCCCTTC TCATATCTAT AAGTTGTAAA  
 2941 CCGTGTGCGC CTTATTCCTT TTTTGGGGC ATTTGCTCTT CCTGTTTTTG CTACACCCAGA AACGCTGGTG  
 GGCACAGCGG GAATAAGGGA AAAAACGCCG TAAAAACGGA GGACAAAAAC GAGTGGGTCT TTGCGACCAC  
 3011 AAAGTAAAG ATGCTGAAGA TCAGTTGGGT GCACGAGTGG GTTACATCGA ACTGGATCTC AACAGCGGTA  
 TTTCAATTTT TACGACTTCT AGTCAAACCA CGTGTCTACC CAATGTAGCT TGACCTAGAG TTGTGCCCCAT  
 3081 AGATCCTTGA GAGTTTTCGC CCCGAAGAAC GTTTTCCAAAT GATGAGCACT TTATAAGTTC TGCTATGTGG  
 TCTAGGAACT CTCAAAAAGCG GGGCTTCTTG CAAAAGGTTA CTACTCGTGA AAATTTCAAG ACGATACACC  
 3151 CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC GGTGCGCCGA TACACTATTC TCAGAAATGAC  
 GCGCCATAAT AGGGCATAAC TCGGGCCCGT TCTGTTGAG TCGCGCCGT ATGTGATAAG AGTCTTTACTG  
 3221 TTGGTTGAGT ACTCACCAGT CACAGAAAAA CATCTTACGG ATGGCATGAC AGTAAAGAAA TTATGCAAGTG  
 AACCNACTCA TGAATGGTCA GTGTCTTTTC GTAGAAATGCC TACCGTACTG TCATTTCTTT AATACGTCTAC

**FIG.\_48F**

```

3291 CTGCCATTAAC CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAAAG ATCGGAGGAC CGAAGGAGCT
GACGGTATTG GTACTCACTA TTGTGACGCC GGTGGAATGA AGACTGTTCG TAGCCTCCTG GCTTCCTCGA

3361 AACCGCTTTT TTGCACAACA TGGGGGATCA TGTAACTCGC CTTGATCGTT GGGAAACCGGA GCTGAAATGAA
TTGGCGAATA AACGTGTTGT ACCCCCTAGT ACATTGAGCG GAACTAGCAA CCCCTTGGCCT CGACTTACTT

3431 GCCATACCAA ACGACGAGCG TGACACCACG ATGCCCTGAG CAATGGCAAC AACGTTGCGC AAATATATTAA
CGGTATGGTT TGCTGCTCGC ACTGTGTGTC TACGGACATC GTTACCCTTG TTGCAACGCG TTGTGATAATT

3501 CTGGCGAACT ACTTACTCTA GCTTCCCGGC AACAAATTAAT AGACTGGATG GAGGCGGATA AAGTTGCAGG
GACCGCTTGA TGAATGAGAT CGAAGGGCCG TTGTTAATTA TCTGACCTAC CTCCGCCCTAT TTCAAACGTCC

3571 ACCACTTCG CGCTCGGCC TTCCGGCTGG CTGGTTTATT GCTGATAAAT CTGGAGCCGG TGAGCGGTGG
TGGTGAAGAC GCGAGCCCGG AAGGCCGACC GACCAATATA CGACTATTTA GACCTCGGCC ACTCGCACCC

3641 TCTCGCGGTA TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT CGTAGTTATC TACACGACGG
AGAGCGCCAT AGTAACGTG TGACCCCGGT CTACCATTCG GGAGGGCATA GCATCAATAG ATGTGCTGCC

3711 GGAGTCAGGC AACTATGGAT GAACGAATA GACAGATCGC TGAGATAGGT GCCTCACTGA TTAAGCATTG
CCTCAGTCCG TTGNTACCTA CTTGCTTTAT CTGTCTAGCG ACTCTATCCA CGGAGTGACT AATTCGTAAAC

3781 GTAACGTGCA GACCAAGTTT ACTCATATAT ACTTTAGATT GATTTAAAC TTCAATTTTA ATTTAAAGG
CAITGACAGT CTGGTTCAAA TGAGTATATA TGAANTCTAA CTAAATTTTG AAGTANAAAT TAAATTTTCC

3851 ATCTAGGTGA AGATCCCTTT TGATATCTC ATGACCAAAA TCCCTTAAAC TGAGTTTTCG TTCCACTGAG
TAGATCCACT TCTAGGNAAA ACTATTAGNG TACTGTGTTT AGGGAAATGC ACTCAAAAGC AAGGTGACTC

3921 CGTCAGACCC CGTAGAAAAG ATCAAGGAT CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA TCTGTGCTT
GCAGTCTGGG GCATCTTTTC TAGTTTCCCTA GAAGAACTCT AGGMAAATAA GACCGGCATT AGACGACGAA

3991 GCAACAAATA AACCAACCGC TACCAGCGGT GGTGTGTTTG CCGGATCAAG AGCTACCAAC TCTTTTTCGG
CGTTTGTTTT TTTGGTGGCG ATGGTCGCCA CCAACAACAC GGCCTAGTTC TCGATGGTTG AGMAAAGGC

4061 AAGGTAACTG GCTTCAGCAG AGCGCAGATA CCAATACTG TCCTTCTAGT GTAGCCGTAG TTAGGCCACC
TTCCATTGAC CGAAGTCGTC TCGCGTCTAT GGTATTGAC AGGAAGATCA CATCGGCATC AATCCGGTGG

```

**FIG.\_48G**

4131 ACTTCAAGAA CTCGTGTAGCA CCGCTACAT ACCTGGTCTT GCTAATCCTG TTACCAAGTGG CTGCTGCCAG  
 TGAAGTTCTT GAGACATCGT GCGGATGTA TGGAGCGAGA CGATTAGGAC AATGGTCACC GACGACGGTC  
 4201 TGGCGATAAG TCGTGTCTTA CCGGTTTGA CTCAAGACGA TAGTTACCGG ATAAAGGCGCA GCGGTCGGGC  
 ACCGCTATTG AGCACAGAAAT GGCCCAACCT GAGTTCTGCT ATCAATGGCC TATTCGCGGT CGCCAGCCCG  
 4271 TGAACGGGGG GTTCGTGCAC ACAGCCACGCTTGGAGCGAA CGACCTACAC CGAAGTGA GA TACCTACAGC  
 ACTTGCCCCC CAAGCACGTA TGTGGGTG AACTCTGCTT GCTGATGTG GCTTGACTCT ATGGATGTG  
 4341 GTGAGCTATG AGAAAGCGCC ACGCTTCCG AAGGAGAA GGCGGACAGG TATCCGGTAA GCGGCAGGGT  
 CACTCGATAC TCTTTCGCGG TCGAAGGGC TTCCCTCTTT CCGCTGTCC ATAGGCCATT CGCCGTCCCA  
 4411 CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGAAAC GCCTGGTATC TTATAGTCC TGTCGGGTTT  
 GCCTTGTCCT CTCGCGTGCT CCTCGAAG CCTCCCTTTG CGGACCATAG AATATCAGG ACAGCCCAAA  
 4481 CGCCACCTCT GACTTGAGCG TCGATTTTG TGAATCTCT GAGGGGGCG GAGCCTATGG AAAAAAGCCA  
 GCGGTGGAGA CTGAACCTCG AGCTAAAAAC ACTACGACA GTCCCCCGC CTCGGATACC TTTTTCGGGT  
 4551 GCAACGCGGC CTTTTTACGG TTCCCTGGCT TTTGCTGACC TTTTGTCTTC ATGTTCTTTC CTGCGTTATC  
 CGTTGCGCCG GAAAAATGCC AAGGACCGA AAGGACCGG AAGCGAGTG AAAACGAGTG TACNAGAAAG GACGCAATAG  
 4621 CCCTGATTCT GTGGATAACC GTATTACCG CTTTGAGTGA GCTGATACC CTGCGCGCAG CCGAACGACC  
 GGGACTAAGA CACCTATTGG CATAATGGCG GAAACTCACT CGACTATGGC GAGCGGCGTC GGCITGCTGG  
 4691 GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGAGCGCC CAATACGCA ACCGCTCTC CCGCGCGGT  
 CTCGCGTCGC TCAGTCACTC GCTCCTTCGC CTTCTCGCG GTTATGCGT TGGCGGAGG GGGCGCGCAA  
 4761 GGCCGATTCA TTAATGCAGC TGGCAGACA GGTTCGCCA CTGAAAGCG GGCAGTGA GC AACGCAAT  
 CCGGCTAAGT AATTACGTG ACCGTGCTGT CCAAAAGGCT GACCTTTCG CCGTCACTCG CGTTGCGTTA  
 4831 TAATGTGAGT TAGCTCACTC ATTAGGCACC CCAGGCTTTA CACTTTATGC TTCCGGCTCG TATGTTGTGT  
 ATTACACTCA ATCGAGTGAG TAATCCGTGG GTCCGAAAT GTGAATACG AAGGCCGAGC ATACNACACA

FIG.\_48H

4901

B88HII

~~~~~

GGAATTGTGA GCGGATAACA ATTCACACA GGAACAGCT ATGACCATGA TTACGCCAAG CGCGCAATTA  
CCTTAACACT CGCCTATTGT TAAAGTGTGT CCTTGTCTGA TACTGGTACT AATGCGGTTC GCGCGTTAAT

4971

NcoI

KpnI

~~~~~

ACCTCACA AAGGGAACA AAGCTGGGTA C  
TGGGAGTGAT TTCCCTTGTT TTCGACCCAT G

FIG.\_48I

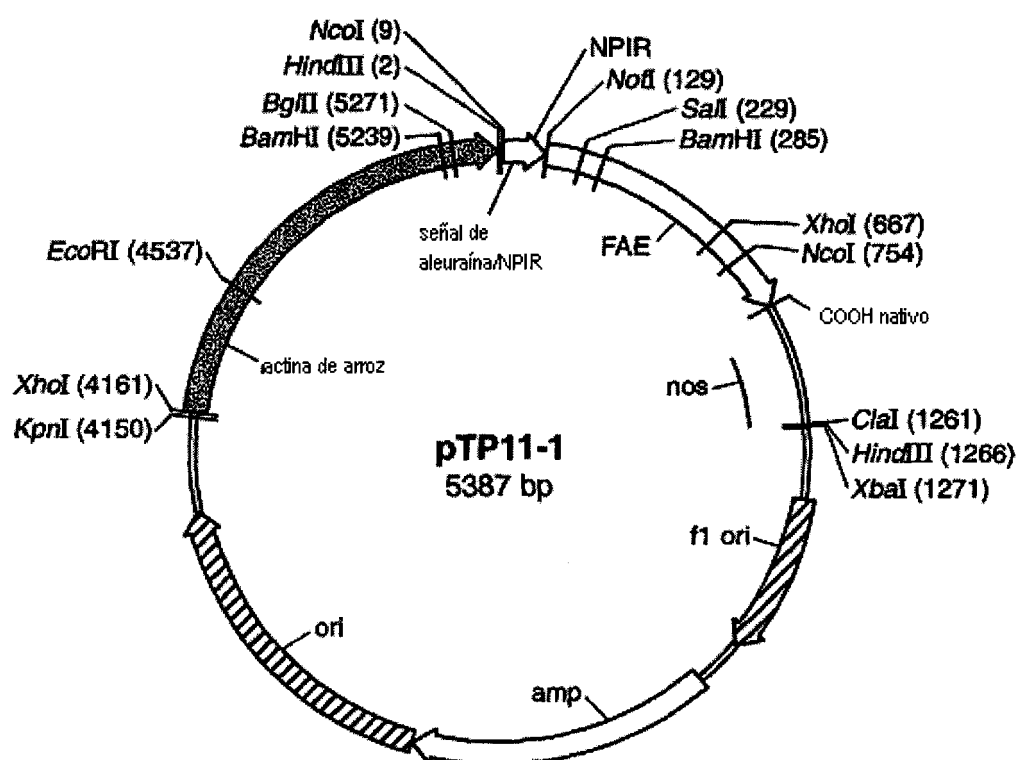


**FIG. 49A**

```

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 CCGGTCTCTC CTCCTGGCGC TCGCCGTGCT GGCACGGCC GCGTCTGCCG
 NotI
      ~~~~~
      . A S S S S F A D S N P I R P V T D R A A A S T .
71 TGGCTCTCTC CTCCTCTCTC GCGACTCCA ACCGATCCG GCGGTACAC GACGGCGCG GCGCTCCAC
      . 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 TCGAGAGACC TCTACAGCCG TTTAGTCGAA ATGCCACTA 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 AATTACAA TTCTCAACT GACATTAAAG
 BamHI
      ~~~~~
      . W I L R D D S S K E I I T V F R Q T G S D T N .
281 GATGGATCCT CCGGACGAC AGCAGCAAG AATAATCAC CGTCTCCGT GGCATGGTA GTGATACGAA
      . L Q L D T N Y T L T P F D T L P Q C N G C E V
351 TCTACAATC GATACTACT ACACCTCAC GCCTTCGAC ACCCTACCAC AATGCACCG 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 ATGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGTTA
      . Q Y P D Y A L T V T G H X L G A S L A L T A .
491 GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCAGTGC
      . 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 CCGCTGTAT ACCTTCGCG 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 CGATGCTTC CAAGCTCGA GCCAGATAC GACGCAGTAT TTCGGGGTCA

```

**FIG.\_49B**

```

 . H A N D G I P N L P P V E Q G Y A H G G V E Y .
701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCGGTGA GCAGGGGTAC GCCCATGCG 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 TGCACCTGGG ATGAAGTGA GTGCTGTGAG
 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 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGA GCCTGTACAT
 . *
911 GGTGATCAGT CATTTCAGCC TCCCGAGTG TACCAGGAA GATGGATGTC CTGGAGAGGG GGCCGCGTAA
981 CCACTGAAGG ATGAGCTGTA AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT AAGATTGAAT
1051 CCTGTGCGG GTCTTGCGAT GATTATCATA TAAATTCNGT TGAATACGT TAAGCATGTA ATAAATTAACA
1121 TGTAAATGCAT GACGTTATT ATGAGATGGG TTTTATGAT TAGAGTCCCG CAATTATACA TTTAATACGC
 ClaI
1191 GATAGAAAC AATATATAGC GCGCAACTA GGTAAATTA TCGCGCGCG TGTCACTAT GTTACTAGAT
 HindIII
 ClaI XbaI
    ~~~~~
1261 CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTCCGC CTATAGTGAG TCGTATTACG CGCGTCACT
1331 GCGCGTCTGT TTACAACTGC GTGACTGGGA AAGCCCTGGC GTTACCCAAC TTAAATCGCT TGCAGCACAT
1401 CCCCTTTTCG CCAGCTGGCG TAATAGCGAA AAGGCCGCGA CCGATCGCCC TTCCCAACAG TTGCGCAGCC
1471 TGAATGGCGA ATGGGACGCG CCCTGTAGCG GCGCAATTAAG CCGCGCGGT GTGTGTGTTA CGCGCAGCGT
1541 GACCGCTACA CTTGCCAGCG CCTAGCGCC CGCTCCTTTC GCTTCTTTC CTTCCTTCT CGCCACGTTT
1611 GCGGCTTTC CCCGTCAAGC TCTAAATCGG GGGTCCCTT TAGGGTTCCG ATTTAGTGT TTACGGCACC
1681 TCGACCCCAA AAAACTTGAT TAGGGTGATG GTTCACGTAG TGGGCCATCG CCTGATAGA CGGTTTTTCG
1751 CCTTTGACG TTGGAGTCCA TTATTAAGG ATTTGGCGA TTTCGGCTA TTGGTTAAA AATGAGCTGA
1821 ATCTCGGTCT ATTTTTCG AATTTAACG AATTTAACG GCTTACAAT TAGGTGGCAC TTTTCGGGA
1891 TTTAACAAA ATTTAACG GAAACCCCTAT TTGTTATTT TTCTAATAT ATTCAAATAT GTATCCGCTC ATGAGACAT
1961 AATGTGCGG GAAACCCCTAT TTGTTATTT TTCTAATAT ATTCAAATAT GTATCCGCTC ATGAGACAT
2031 AACCTGATA AATGCTTCAA TAATATTGA AAGGAAGAG TATGATATT CAACATTTCC GTGTCCGCTT
2101 TATTCCTTT TTTGCGGCAT TTTGCTTCC TGTTTTCGT CACCCAGAA CGCTGGTGA AGTAAAGAT
2171 GCTGAAGATC AGTTGGGTGC ACAGTGGGT TACATCGAC TGAATCTCA CAGCGTAA GATCCTTGAG
2241 GTTTTCGCC CGAAGAACGT TTTCCAATGA TGAGCACATT TAAAGTTCTG TATGTGGCG CGGTATTATC
2311 CCGTATTGAC GCCGGGCAAG AGCAACTCG TCGCCGCTA CACTATTCTC AGAATGACTT GGTGAGTAC
2381 TCACCAGTCA CAGAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAAAT ATGCAGTGT GCCATAACCA

```

FIG.\_49C

2451 TGAGTGATAA CACTGGGGCC AACTTACTTC TGACAACGAT CGGAGGACCG AAGGAGCTAA CCGCTTTTTT  
 2521 GCACAACATG GGGGATCATG TAACTCGCCT TGAATGATGG GAACCGGAGC TGAATGAGC CATACCAAAC  
 2591 GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAACT GCGGAACTAC  
 2661 TTAATCTAGC TTCCCGGCAA CAATTAAATAG ACTGGATGGA GCGCGATAAA GTTGCAGGAC CACTTCTGCG  
 2731 CTCGGCCCTT CCGCTGGCT GGTTAATGCT TGAATAAATCT GGAGCCGGTG AGCGTGGGTC TCGGCGTATC  
 2801 ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCGATATCG TAGTTATCTA CACGACGGGG AGTCAGGCCA  
 2871 CTATGGATGA ACGAAATAGA CAGATCGCTG AGTATAGTGC CTCACTGATT AAGCATTTGGT AACTGTCTAGA  
 2941 CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAAACTT CATTTTAAAT TTAAATGGAT CTAGGTGAG  
 3011 ATCTTTTTG ATAAATCTCAT GACCAAAATC CCTTAAGGTG AGTTTTCGTT CCACGTGAGC TCAGACCCCG  
 3081 TAGAAAAGAT CAAAGGATCT TCTTGAGATC CTTTTTCT GCGCGTAATC GCGGTGCTGC AAACAAAATA  
 3151 ACCACCGCTA CCAGCGGTGG TTTGTTTGCC GGATCAAGAG GTTACCAATC CTACCAATC TTTTCCGAA GGTAACTGGC  
 3221 TTCAGCAGAG CGCAGATACC AATACTGTC CTCTAGTGT TACTAGTGT AGCCGTAGTT AGGCCACAC TTAAGAACT  
 3291 CTGTAGCAC CCCTACATAC CTCGCTCTGC TAACTCTGTT ACCAGTGGCT AAGCGCAGC GGTGCGGCTG AACGGGGGT  
 3361 GTGTCTTACC GGTGTGGACT CAAGACGATA GTTACCGGAT GTTACCGGCT CCTACAGCGT GAGCTATGAG  
 3431 TCGTGACAC AGCCACGCTT GGAGCGAACG ACCTACACCG AACTGAGATA AACTGAGTGC GGTAAAGTC  
 3501 AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CCGACAGGTA TCCGATTAAG TCCGAGGTG GACACAGGGA  
 3571 GCGCACGAG GAGCTTCCAG GGGGAAACGC CTGATATCTT TATAGTCTTG TATAGTCTTG CCACCTCTGA  
 3641 CTTGAGCGTC GATTTTGTG ATGCTCGTCA GGGGGCGGA GCTATGGA GCTATGGA AACGCCAGC AACCGGCCCT  
 3711 TTTTACGGTT CCTGGCCTTT TGCTGGCCTT TGATCAGAT TATCTTCTT GCTTATCCC GGTATCTGT  
 3781 GGAATACCGT ATTACCGCT TTGAGTGAGC TGATACAAAC CCGCTCTCC CCGCGCAGG GCGACGCGAG  
 3851 TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAC ATACGCAAC CCGCTCTCC CCGCGCAGG CCGATTCTAT  
 3921 AATGCAGCTG GCACGACAGG TTTCCCGACT GGAAGCGGG GAGTGAGCGC CAGTGAGCGC AACGCAATTA ATGTGAGTTA  
 3991 GCTCACTCAT TAGGCACCCC AGGCTTTACA CTTATGCTT CCGCTCTGTA TGTGTGTGG AATTGTGAGC  
 4061 GGATAACAA TTCACACAGG AACAGCTAT GACCATGATT ACGCCAAGCG CGCAATTAA CCTCCTACAAA  
 4131 GGGAAACAAA GCTGGGTACC GGGCCCCCCC TCGAGGTCTT TCATATGCTT GAGNAGAGAG TCGGGATAGT  
 4201 CCAAAATAAA ACAAGGTAA GATTACCTGG TCATAAGTGA AACATCAGT TAAAGGTGG TATAAGTAAA  
 4271 ATATCGGTAA TAAAGGTGG CCCAAGTGA AATTACTCT TTTCTACTAT TATAAAAT TATAGATGTT  
 4341 TGTGCTACT TTGATACGTC ATTTTGTGAT GAATTGGTTT TTAAGTTTAT TCAGGATTTG GAAATGCAAT  
 4411 TCTGTATTG AGTCGGTTTT TAAGTTCTGT GCTTTTGTAA ATACAGAGGG ATTTGTATTA ECGRI  
 4481 TAAAAAACCC ATATGCTAAT TTGACATAAT TTTTGAGAAA AATATATATT CAGGCGAATT CCACAAATGA  
 4551 CAATAATAAG ATTAAATAG CTGCCCCCG TTGACGGAT TTGATATTTT GGTATTTTT TCTAGTAAA TAAAGATAA  
 4621 ACTTAGACTC AAAACATTTA CAAAAACAAC CCCTAAGTGC CTAAAGCCCC AAGTGTATG CACGATCCAT

FIG.\_49D

```

4691 AGCAAGCCCC GCCCAACCCA ACCAACCCTA ACCACCCCA GTGCAGCCAA CTGGCAATA GTCTCCACCC
4761 CCGGCACTAT CACCGTGAGT TGTCCGCACC ACCGCACGTC TCGCAGCCAA AAAAAAATA AGAAAGAAAA
4831 AAGAAGAAA GAAAAACAGC AGGTGGGTCC GGTTCGTGGG GGCCTGGAAA GCGAGGAGGA TCGCGAGCAG
4901 CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC ACTATATACA TACCCCCCCC
4971 TCTCCTCCCA TCCCCCAAC CTAACACCA CCACCAACC CACCTCCTCC CCGCTCGCTG CCGGACGACG
5041 AGCTCCTCCC CCTTCCCCCT CCGCCGCCGC CGTAACCA CCGGCCCTC TCCCTCTTCT TTCTCCGTTT
5111 TTTTTTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT GGTGGGCCA GAGCGGCTTC GTCGCCCCAGA
                                     BamHI
                                     ~~~~~
5181 TCGGTGCGCG GGAGGGCGG GATCTCGCG GATCGCTCTC CGGCGGTGAG TCGGCCCGGA TCCTCGCGGG
 BglII
                                     ~~~~~
5251 GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTCCTTCTT TTTGTGGTAG AATTGAATC CCTCAGCATT
5321 GTTCATCGGT AGTTTCTCTT TTCAATGATT GTGACAAATG CAGCCTCGTG CGGAGCTTTT TTGTAGC

```

**FIG.\_49E**



```

771 CCTCCCTCCG CTTCCTAAGA AACGCCCCCC ATCGCCACTA TATACATACC CCCCCCTCTC CTCCCATCCC
    GGAGGGAGGC GAAGGTTTCT TTGCGGGGGG TAGCGGTGAT ATATGTATGG GGGGGGAGAG GAGGGTAGGG

841 CCCAACCTTA CCACCACCAC CACCACCACC TCCTCCCCCC TCGCTGCCGG AGGACGAGCT CCTCCCCCCT
    GGGTTGGGAT GGTGGTGGTG GTGGTGGTGG AGGAGGGGGG AGCGACGGCC TGCTGCTCGA GGAGGGGGGA

911 CCCCCCTCCG CGCCGCGCGT AACCAACCCG CCCCTCTCCT CTTTCTTTCT CCGTTTTTTT TTTCGTCTCG
    GGGGGAGGCG GCGGCGGCCA TTGGTGGGGC GGGGAGAGGA GAAAGAAAGA GGCACAAAAA AAAGCAGAGC

981 GTCTCGATCT TTGGCCTTGG TAGTTTGGGT GGGCGAGAGC GGCTTCGTGG CCCAGATCGG TGCGCGGGAG
    CAGAGCTAGA AACCGGAACC ATCAAACCCA CCCGCTCTCG CCGAAGCAGC GGGTCTAGCC ACGCGCCCTC

                                BamHI
                                ~~~~~

1051 GGGCGGATC TCGCGGCTGG CGTCTCCGGG CGTGAGTCGG CCGGATCCT CGCGGGGAAT GGGGCTCTCG
 CCGGCCCTAG AGCGCGGACC GCAGAGGCCG GCACTCAGCC GGGCCTAGGA GCGCCCTTA CCCCAGAGC

 BglII
                                ~~~~~

1121 GATGTAGATC TTCTTTCTTT CTCTTTTTTG TGGTAGAATT TGAATCCCTC AGCATTGTTT ATCGGTAGTT
    CTACATCTAG AAGAAAGAAA GAAGAAAAC ACCATCTTAA ACTTAGGGAG TCGTAACAAG TAGCCATCAA

                                HindIII  NcoI
                                ~~~~~

1191 TTTCTTTTCA TGATTGTGA CAAATGCAGC CTGTGCGGA GCTTTTTTGT AGGTAGAAGC TTACCATGG
 AAAGAAAAGT ACTAAACACT GTTTACGTCG GAGCAGCCT CGAAAAACA TCCATCTTCG AATGGTACC

```

KpnI-EcoRI – eliminación subrayada y sitio NCO restaurado en negrita en los vectores pJQ4.9, pJQ3.2 y pJO6.3

**FIG.\_50B**

Estructura y secuencia (Apoplasto) de ALEURAINA-NPIR eliminado

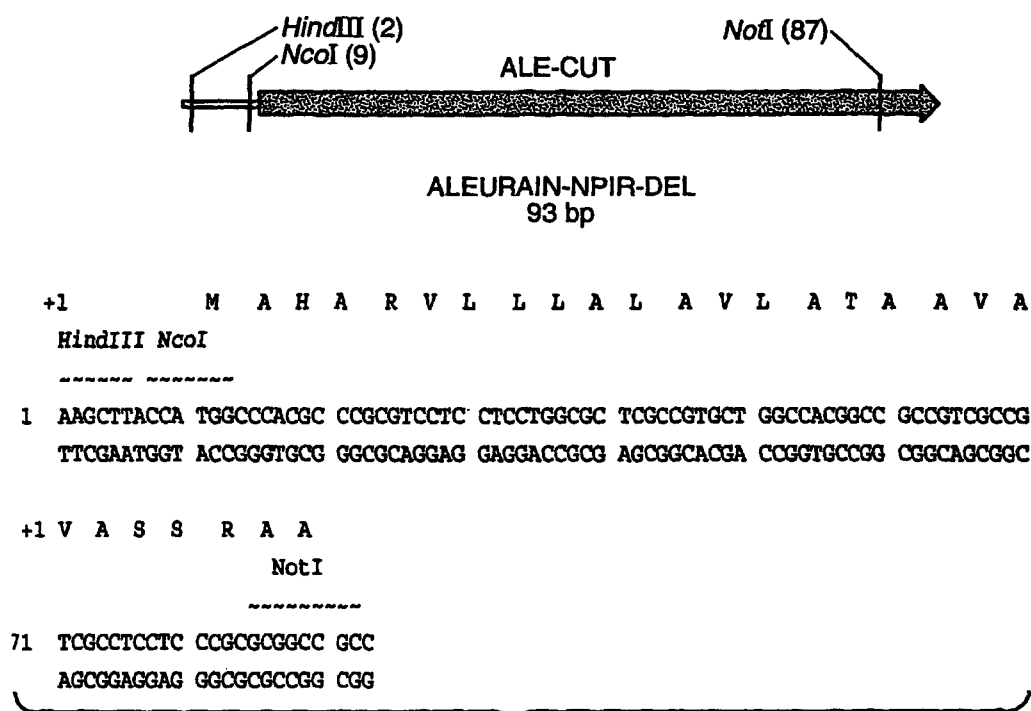


**FIG. 51**

SEE1 Secuencia de promotor SEE1 (aumento de senescencia)

```

1 CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
 PstI XbaI XbaI

71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTGCGGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
 SmaI

351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTGTGTG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
 SphI

491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTGG CTTAATTGAC TTTATTTTTG
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG
631 CTAATTAAAT TATAGCAAGA TTCTGTAGTG CACATGCGAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
 PstI

701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCGGCACC GGGACGCGAG CTCGCCCTA CCAATACCAT CTCGCTCGC TCCCCCTGCC GGACGACCCA
841 GTAAAATACT GTTGCCCACT CGCCGCGCAG ATG

```

## FIG.\_52

SEE1 Secuencia de promotor SEE1 (aumento de senescencia) más secuencia SEÑAL de aleurina/NPIR vacuolar

```

1 CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
 PstI XbaI XbaI

71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTGCGGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
 SmaI

351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTGTGTG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
 SphI

491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTGG CTTAATTGAC TTTATTTTTG
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG
631 CTAATTAAAT TATAGCAAGA TTCTGTAGTG CACATGCGAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
 PstI

701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCGGCACC GGGACGCGAG CTCGCCCTA CCAATACCAT CTCGCTCGC TCCCCCTGCC GGACGACCCA
 M A H G R I L F L A L A V L
841 GTAAAATACT GTTGCCCACT CGCCGCGCAG ATGGCCCAAG GCCGCATCCT CTTCTTGGCG CTCGCCGTCT
 BsaHII
 NotI

911 . A T A A V A A A S L A D S N P I R P V T E R A .
 NotI

981 . A A
 GCGCGCC

```

## FIG.\_53