



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 282 994**

51 Int. Cl.:

**C12N 15/64** (2006.01)

**C12N 15/67** (2006.01)

**C12N 15/85** (2006.01)

**C12N 9/72** (2006.01)

**C12N 5/10** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Número de solicitud europea: **95928192 .4**

86 Fecha de presentación : **28.07.1995**

87 Número de publicación de la solicitud: **0770136**

87 Fecha de publicación de la solicitud: **02.05.1997**

54

Título: **Procedimiento para la selección de células huésped de expresión elevada.**

30

Prioridad: **05.08.1994 US 286740**

45

Fecha de publicación de la mención BOPI:  
**16.10.2007**

45

Fecha de la publicación del folleto de la patente:  
**16.10.2007**

73

Titular/es: **GENENTECH, Inc.**  
**460 Point San Bruno Boulevard**  
**South San Francisco, California 94080-4990, US**

72

Inventor/es: **Crowley, Craig W.**

74

Agente: **Ponti Sales, Adelaida**

ES 2 282 994 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**

Procedimiento para la selección de células huésped de expresión elevada.

**5 Antecedentes de la invención****Campo de la invención**

10 La presente invención se refiere a un procedimiento de elección de células huésped de expresión elevada, a un procedimiento de producción de una proteína de interés a elevados rendimientos y a un procedimiento de producción de células eucarióticas que presentan copias múltiples de una secuencia codificante de una proteína de interés.

**Descripción de los antecedentes y técnica relacionada**

15 El descubrimiento de procedimientos para introducir ADN en células huésped vivas en una forma funcional ha proporcionado la clave para la comprensión de muchos procesos biológicos fundamentales, y ha hecho posible la producción de proteínas importantes y otras moléculas en cantidades comercialmente útiles.

20 A pesar del éxito general de estos procedimientos de transferencia génica, existen varios problemas frecuentes que podrían limitar la eficacia con la que puede introducirse un gen codificante de una proteína deseada en una célula huésped y expresarse en la misma. Un problema es llegar a conocer cuándo se ha transferido con éxito el gen hasta el interior de las células receptoras. Un segundo problema es distinguir entre aquellas células que contienen el gen y aquéllas que han sobrevivido a los procedimientos de transferencia pero que no contienen el gen. Un tercer problema es identificar y aislar aquéllas células que contienen el gen y que expresan niveles elevados de la proteína codificada por el gen.

En general, los procedimientos conocidos para introducir genes en células eucarióticas tienden a ser altamente ineficientes. De las células en un cultivo dado, sólo una proporción reducida incorpora y expresa ADN añadido exógenamente, y una proporción incluso menor mantiene establemente ese ADN.

30 La identificación de aquellas células que han incorporado un gen de producto codificante de una proteína deseada típicamente se consigue introduciendo en las mismas células otro gen, con frecuencia denominado gen seleccionable, que codifica un marcador seleccionable. Un marcador seleccionable es una proteína que resulta necesaria para el crecimiento o supervivencia de una célula huésped bajo las condiciones de cultivo particulares seleccionadas, tal como un enzima que proporciona resistencia a un antibiótico o a otro fármaco, o un enzima que compensa un defecto metabólico o catabólico en la célula huésped. Por ejemplo, entre los genes seleccionables utilizados con frecuencia en las células eucarióticas se incluyen los genes de la aminoglucósido fosfotransferasa (APH), la higromicina fosfotransferasa (hyg), dihidrofolato reductasa (DHFR), la timidina quinasa (tk), neomicina, puromicina, glutamina sintetasa y asparagina sintetasa.

40 El procedimiento de identificación de una célula huésped que ha incorporado un gen, basado en la expresión por la célula huésped de un segundo gen incorporado codificante de un marcador seleccionable se denomina cotransfección (o contranfección). En este procedimiento se introducen típicamente un gen codificante de un polipéptido deseado y un gen de selección en la célula huésped simultáneamente, aunque pueden introducirse secuencialmente. En el caso de la cotransfección simultánea, el gen codificante del polipéptido deseado y el gen seleccionable pueden encontrarse presentes en una sola molécula de ADN o en moléculas de ADN separadas antes de su introducción en las células huésped (Wigler *et al.*, Cell 16:777, 1979). Las células que han incorporado el gen codificante del polipéptido deseado seguidamente se identifican o se aíslan cultivando las células bajo condiciones que permiten preferentemente el crecimiento o la supervivencia de aquellas células que sintetizan el marcador seleccionable codificado por el gen seleccionable.

55 El nivel de expresión de un gen introducido en una célula huésped eucariótica depende de múltiples factores, incluyendo el número de copia del gen, la eficiencia de transcripción, el procesamiento, estabilidad y eficiencia de traducción del ARN mensajero (ARNm). Por consiguiente, un nivel elevado de expresión de un polipéptido deseado típicamente implicará optimizar uno o más de esos factores.

60 Por ejemplo, el nivel de producción de proteína puede incrementarse uniéndose covalentemente la secuencia codificante del gen a un promotor o intensificador "fuerte" que proporcionará niveles elevados de transcripción. Los promotores e intensificadores son secuencias de nucleótidos que interaccionan específicamente con proteínas en una célula huésped que se encuentran implicadas en la transcripción (Kriegler, Meth. Enzymol. 185:512, 1990; Maniatis *et al.*, Science 236:1237, 1987). Los promotores se encuentran situados corriente arriba de la secuencia codificante de un gen y facilitan la transcripción del gen por la ARN polimerasa. Entre los promotores eucarióticos que han sido identificados como promotores fuertes para la expresión de nivel elevado se encuentran el promotor temprano de SV40, el promotor tardío mayor de adenovirus, el promotor de la metalotioneína-I de ratón, la repetición terminal larga del virus del sarcoma de Rous y el promotor temprano inmediato del citomegalovirus humano (CMV).

Los intensificadores estimulan la transcripción desde un promotor ligado. Al contrario que los promotores, los intensificadores se encuentran activos cuando se sitúan corriente abajo del sitio de inicio de transcripción o a distancias

considerables del promotor, aunque en la práctica los intensificadores pueden solaparse física y funcionalmente con los promotores. Por ejemplo, todos los promotores fuertes indicados anteriormente también contienen intensificadores fuertes (Bendig, *Genetic Engineering* 7:91, Academic Press, 1988).

5 El nivel de producción de proteínas también puede incrementarse, incrementando el número de copia génica en la célula huésped. Un procedimiento para obtener un número de copia génica elevado es introducir directamente en la célula huésped múltiples copias del gen, por ejemplo utilizando un exceso molar elevado del gen de producto respecto al gen seleccionable durante la cotransfección (Kaufman, *Meth. Enzymol.* 185:537, 1990). Sin embargo, con este procedimiento, sólo una proporción reducida de las células cotransfectadas contendrá el gen de producto  
10 en un número de copia elevado. Además, debido a que no existe ningún procedimiento conveniente de aplicación general para distinguir estas células de las mayoría de las células que contienen menos copias del gen de producto, típicamente resultan necesarios procedimientos de cribado laboriosos y que requieren mucho tiempo para identificar los transfectantes deseados de elevado número de copia.

15 Otro procedimiento para obtener un número de copia elevado implica clonar el gen en un vector que sea capaz de replicarse autónomamente en la célula huésped. Entre los ejemplos de estos vectores se incluyen vectores de expresión de mamífero derivados de virus de Epstein-Barr o virus del papiloma bovino, y vectores plásmido de 2 micrómetros de levadura (Stephens y Hentschel, *Biochem. J.* 248:1, 1987; Yates *et al.*, *Nature* 313:812, 1985; Beggs, *Genetic Engineering* 2:175, Academic Press, 1981).

20 Todavía otro procedimiento para obtener un número de copia génica elevado implica la amplificación génica en la célula huésped. La amplificación génica se produce naturalmente en las células eucarióticas a una frecuencia relativamente reducida (Schimke, *J. Biol. Chem.* 263:5989, 1988). Sin embargo, la amplificación génica también puede inducirse, o por lo menos seleccionarse, mediante la exposición de las células huésped a una presión selectiva apropiada. Por ejemplo, en muchos casos resulta posible introducir un gen de producto junto con un gen amplificable en una célula huésped y posteriormente seleccionar para la amplificación del gen marcador mediante la exposición de las células cotransfectadas a concentraciones secuencialmente crecientes de un agente selectivo. Típicamente el gen de producto se coamplifica con el gen marcador bajo estas condiciones.

30 El gen amplificable más ampliamente utilizada para este fin es un gen DHFR, que codifica un enzima dihidrofolato reductasa. El agente de selección utilizado conjuntamente con un gen DHFR es el metotrexato (Mtx). Se cotransfecta una célula huésped con un gen de producto codificante de una proteína deseada y un gen DHFR, y los transfectantes se identifican cultivando en primer lugar las células en medio de cultivo que contiene Mtx. Una célula huésped adecuada cuando se utiliza un gen DHFR de tipo salvaje es la línea celular de ovario de hámster chino (CHO) deficiente en actividad DHFR, preparada y propagada tal como se describe en Urlaub y Chasin, *Proc. Nat. Acad. Sci. USA* 77:4216, 1980. A continuación, las células transfectadas se exponen a cantidades sucesivamente más elevadas de Mtx. Esto conduce a la síntesis de múltiples copias del gen DHFR, y concomitantemente, de múltiples copias del gen de producto (Schimke, *J. Biol. Chem.* 263:5989, 1988; Axel *et al.*, patente US nº 4.399.216; Axel *et al.*, patente US nº 4.634.665. Entre otras referencias dirigidas a la cotransfección de un gen conjuntamente con un marcador genético que permite la selección y posterior amplificación se incluyen Kaufmann, en *Genetic Engineering*, editor J. Setlow (Plenum Press, New York), vol. 9, 1987; Kaufman y Sharp, *J. Mol. Biol.* 159:601, 1982; Ringold *et al.*, *J. Mol. Appl. Genet.* 1:165-175, 1981; Kaufman *et al.*, *Mol. Cell Biol.* 5:1750-1759, 1985; Kaetzel y Nilson, *J. Biol. Chem.* 263:6244-6251, 1988; Hung *et al.*, *Proc. Natl. Acad. Sci. USA* 83:261-264, 1986; Kaufman *et al.*, *EMBO J.* 6:87-93, 1987; Johnston y Kucey, *Science* 242:1551-1554, 1988; Urlaub *et al.*, *Cell* 33:405-412, 1983.

45 Para extender el procedimiento de amplificación de DHFR a otros tipos celulares, puede utilizarse un gen DHFR mutante que codifica una proteína con sensibilidad reducida al metotrexato, conjuntamente con células huésped que contienen un número normal de un gen DHFR de tipo salvaje endógeno (Simonsen y Levinson, *Proc. Natl. Acad. Sci. USA* 80:2495, 1983; Wigler *et al.*, *Proc. Natl. Acad. Sci. USA* 77:3567-3570, 1980; Haber y Schimke, *Somatic Cell Genetics* 8:499-508, 1982).

55 Alternativamente, pueden cotransfectarse células huésped con el gen de producto, un gen DHFR, y un gen seleccionable dominante, tal como un gen *neo<sup>r</sup>* (Kim y Wold, *Cell* 42:129, 1985; Capon *et al.*, patente US nº 4.965.199. Los transfectantes se identifican en primer lugar cultivando las células en medio de cultivo que contiene neomicina (o el fármaco relacionado G418) y los transfectantes identificados de esta manera seguidamente se seleccionan para la amplificación del gen DHFR y del gen de producto mediante la exposición a cantidades sucesivamente crecientes de Mtx.

60 Tal como se apreciará del presente comentario, la selección de células huésped recombinantes que expresan niveles elevados de una proteína deseada es un procedimiento multietapa. En la primera etapa, se seleccionan los transfectantes iniciales que han incorporado el gen de producto y el gen seleccionable. En etapas posteriores, los transfectantes iniciales se someten a selección adicional para la expresión de nivel elevado del gen seleccionable y después a cribado aleatorio para la expresión de nivel elevado del gen de producto. Para identificar las células que expresan niveles elevados de la proteína deseada, típicamente debe cribarse un gran número de transfectantes. La mayoría de los transfectantes produce niveles no máximos de la proteína deseada. Además, la resistencia a Mtx de los transformantes de DHFR la proporciona, por lo menos parcialmente, grados diversos de amplificación génica (Schimke, *Cell* 37:705-713, 1984). Los problemas ocasionados por la coexpresión del gen no seleccionada han sido informados por Wold *et al.*, *Proc. Natl. Acad. Sci. USA* 76:5684-5688, 1979. La inestabilidad del ADN amplificado ha sido informada por

## ES 2 282 994 T3

Kaufman y Schimke, Mol. Cell Biol. 1:1069-1076, 1981; Haber y Schimke, Cell 26:355-362, 1981; y Fedespiel *et al.*, J. Biol. Chem. 259:9127-9140, 1984.

5 Se han descrito varios procedimientos para seleccionar directamente dichas células huésped recombinantes en una sola etapa. Una estrategia implica cotransfectar células huésped con un en de producto y un gen DHFR, y seleccionar aquellas células que expresan niveles elevados de DHFR mediante cultivo directo en medio que contiene una concentración elevada de Mtx. Muchas de las células expresadas de esta manera también expresan el gen de producto cotransfectado a niveles elevados (Page y Sydenham, Bio/Technology 9:64, 1991. Este procedimiento de selección en una sola etapa adolece de determinadas desventajas que limitan su utilidad. Las células de expresión elevada obtenidas mediante cultivo directo en medio que contiene un nivel elevado de un agente de selección pueden presentar características de crecimiento y estabilidad pobres, limitando de esta manera su utilidad para los procedimientos de producción de largo plazo (Page y Snyderman, Bio/Technology 9:64, 1991). La selección en una sola etapa para la resistencia de nivel elevado al Mtx puede producir células con un enzima DHFR resistente a Mtx alterado, o células que presentan propiedades de transporte de Mtx alteradas, y no células que contienen genes amplificados (Haber *et al.*, J. Biol. Chem. 256:9501, 1981; Assaraf y Schimke, Proc. Natl. Acad. Sci. USA 84:7154, 1987).

Otro procedimiento implica la utilización de vectores de expresión de ARNm policistrónico que contienen un gen de producto en el extremo 5' de la región transcrita y un gen seleccionable en el extremo 3'. Debido a que la traducción del gen seleccionable en el extremo 3' del ARN policistrónico es ineficiente, estos vectores manifiestan una traducción preferente del gen de producto y requieren niveles elevados de ARNm policistrónico para sobrevivir a la selección (Kaufman, Meth. Enzymol. 185:487, 1990; Kaufman, Meth. Enzymol. 185:53, 1990; Kaufman *et al.*, EMBO J. 6:187, 1987). Por consiguiente, las células que expresan niveles elevados del producto proteína deseado pueden obtenerse en una sola etapa mediante el cultivo de los transfectantes iniciales en medio que contiene un agente de selección de utilización apropiada con el gen seleccionable particular. Sin embargo, la utilidad de estos vectores es variable debido a la influencia impredecible del marco de lectura del producto corriente arriba sobre la traducción del marcador seleccionable y debido a que el marco de lectura corriente arriba en ocasiones resulta delecionado durante la amplificación con metotrexato (Kaufman *et al.*, J. Mol. Biol. 159:601-621, 1982; Levinson, Methods in Enzymology, San Diego: Academic Press, Inc., 1990). Los vectores posteriores incorporaron un sitio interno de inicio de traducción derivado de miembros de la familia de los picornavirus que se encuentra situado entre el gen de producto y el gen seleccionable (Pelletier *et al.*, Nature 334:320, 1988; Jang *et al.*, J. Virol. 63:1651, 1989).

Un tercer procedimiento para la selección en una sola etapa implica la utilización de un constructo de ADN con un gen seleccionable que contiene un intrón dentro del cual se encuentra situado un gen codificante de la proteína de interés (ver la patente US nº 5.043.270 y Abrams *et al.*, J. Biol. Chem. 264(24):14016-14021, 1989). En todavía otro procedimiento de selección en una sola etapa, se cotransfectan las células huésped con un gen seleccionable de intrón modificado y con un gen codificante de la proteína de interés (ver la patente WO nº 92/17566, publicada el 15 de octubre de 1992). El gen de intrón modificado se prepara insertando en la región transcrita de un gen seleccionable un intrón de longitud suficiente para que el intrón resulte correctamente procesado a partir del precursor de ARNm correspondiente a eficiencia reducida, de manera que la cantidad de marcador seleccionable producido a partir del gen seleccionable de intrón modificado sea sustancialmente inferior a la producida a partir del gen seleccionable de partida. Estos vectores ayudan a garantizar la integridad del constructo de ADN integrado, aunque el ligamiento transcripcional no se consigue debido a que el gen seleccionable y el gen de proteína se encuentran controlados por promotores separados.

45 Se han descrito otros vectores de expresión de mamífero que presentan unidades de transcripción única. Se han construido vectores retrovíricos (Cepko *et al.*, Cell 37:1053-1062, 1984) en los que se inserta un ADNc entre los sitios donador y aceptor de empalme del virus endógeno de leucemia murina de Moloney (M-MuLV) que se encuentran seguidos por un gen de resistencia a la neomicina. Este vector se ha utilizado para expresar una diversidad de productos génicos tras la infección retrovírica de varios tipos celulares.

50 Teniendo en cuenta las desventajas anteriormente indicadas, es un objetivo de la presente invención incrementar el nivel de homogeneidad con respecto a los niveles de expresión de los clones estables transfectados con un gen de producto de interés, mediante la expresión de un marcador seleccionable (DHFR) y la proteína de interés a partir de un solo promotor.

55 Es otro objetivo proporcionar un procedimiento para seleccionar células huésped recombinantes estables que expresan niveles elevados de un producto proteína deseado, procedimiento que resulte de realización rápida y conveniente, y que reduzca el número de células transfectadas que resulte necesario cribar. Además, es un objetivo que resulte posible generar con rapidez niveles elevados de polipéptidos de una y dos unidades a partir de clones o bancos de células huésped transfectantes.

60 Es un objetivo adicional proporcionar vectores de expresión con preferencia para sucesos de integración activa (es decir, presentan una tendencia incrementada a generar transformantes en los que el constructo de ADN se inserta en una región del genoma de la célula huésped que resulta en la expresión de nivel elevado del gen de producto) y que puedan acoger una diversidad de genes de producto sin necesidad de ser modificados.

## Sumario de la invención

Por consiguiente, la presente invención se refiere a un constructo de ADN (molécula de ADN) que comprende un sitio de inicio transcripcional y un sitio de terminación transcripcional, un gen seleccionable que es un gen amplificable y un gen de producto proporcionado en orientación 3' respecto al gen seleccionable, una región reguladora transcripcional que regula la transcripción de tanto el gen seleccionable y del gen de producto, el gen seleccionable situado dentro de un intrón definido por un sitio donador de empalme y un sitio aceptor de empalme. El sitio donador de empalme preferentemente comprende una secuencia donadora de empalme efectiva tal como se define en la presente invención, regulando de esta manera la expresión del gen de producto utilizando la región reguladora transcripcional.

En otra realización, la invención proporciona un procedimiento para producir un producto de interés que comprende cultivar una célula eucariótica que ha sido transfectada con el constructo de ADN indicado anteriormente, de manera que se expresa el gen de producto y se recupera el producto.

En una realización adicional, la invención proporciona un procedimiento para producir células eucarióticas que presentan múltiples copias del gen de producto, que comprende transfectar células eucarióticas con el constructo de ADN indicado anteriormente (en el que el gen seleccionable es un gen amplificable), cultivando las células en un medio selectivo que comprende un agente de amplificación durante un tiempo suficiente para que se produzca la amplificación, y seleccionar células que presentan múltiples copias del gen de producto. Preferentemente la transfección de las células se consigue utilizando la electroporación.

Tras la transfección de las células huésped, la mayoría de los transfectantes no llegan a mostrar el fenotipo seleccionable característico de la proteína codificada por el gen seleccionable, aunque inesperadamente una proporción reducida de los transfectantes sí manifiestan el fenotipo seleccionable, y entre esos transfectantes, la mayoría se observa que expresan niveles elevados del producto deseado codificado por el gen de producto. De esta manera, la invención proporciona un procedimiento mejorado para la selección de células huésped recombinantes que expresan niveles elevados de un producto deseado, procedimiento que resulta útil con una amplia diversidad de células huésped eucarióticas y evita los problemas inherentes a la tecnología existente de selección de células.

## Breve descripción de los dibujos

La Figura 1A-1D ilustra esquemáticamente diversos constructos de ADN comprendidos dentro del alcance de la invención. Las flechas grandes representan el gen seleccionable y el gen de producto, la V formada por las líneas discontinuas muestra la región del ARN precursor interna al sitio donador de empalme 5' (SD) y el sitio aceptor de empalme 3' (SA) que se extrae de los vectores que contienen un SD funcional. La región reguladora transcripcional, gen seleccionable, gen de producto y sitio de terminación transcripcional se ilustran en la Figura 1A. La Figura 1B ilustra los constructos de ADN del Ejemplo 1. Se ilustran las diversas secuencias donadoras de empalme, es decir, la secuencia donador de empalme ras de tipo salvaje (ras WT), la secuencia donadora de empalme de ras mutante (ras MUTANTE) y la secuencia donadora de empalme no funcional ( $\Delta$ GT). Las sondas utilizadas para el análisis de transferencia northern en el Ejemplo 1 se muestran en la Figura 1B. La Figura 1C ilustra los constructos de ADN del Ejemplo 2, y la Figura 1D ilustra el constructo de ADN del Ejemplo 3 utilizado para la expresión de  $V_H$  anti-IgE.

La Figura 2 ilustra esquemáticamente el constructo de ADN de control utilizado en el Ejemplo 1.

Las Figuras 3A-Q ilustran la secuencia de nucleótidos (SEC ID n° 1) del vector de expresión DHFR/intrón- (SD de ras WT)-tPA del Ejemplo 1.

La Figura 4 es un gráfico de columnas que muestra el número de colonias que se forman en medio selectivo tras la electroporación de ADNs miniprep linearizados por duplicado que se habían preparado a partir de tres vectores mostrados en la Figura 1B (es decir con secuencia donadora de empalme ras de tipo salvaje [ras WT], secuencia donadora de empalme ras mutante [ras MUTANTE] y secuencia donadora de empalme no funcional [ $\Delta$ GT]) y del vector de control que presenta DHFR bajo control del promotor de SV40 y tPA bajo el control del promotor de CMV (ver la Figura 2). Las células se seleccionaron en medio libre de nucleósidos y se contaron con un contador automático de colonias.

Las Figuras 5A-C son gráficos de columnas que ilustran la expresión de tPA a partir de pools y clones estables generados a partir de los vectores mostrados en la Figura 1B. En la Figura 5A, se mezclaron más de 100 clones de cada transfección de vector, se sembraron en placas de 24 pocillos, y se sometieron a ensayo mediante ELISA de tPA "a nivel de saturación". En la Figura 5B, se sometieron a ensayo mediante ELISA de tPA "a nivel de saturación" veinte clones seleccionados aleatoriamente derivados de cada uno de los vectores. En la Figura 5C, los pools indicados en la Figura 5A (excepto el pool  $\Delta$ GT) se expusieron a Mtx 200 nM para seleccionar para la amplificación de DHFR y después se agruparon y se sometieron a ensayo para la expresión de tPA.

Las Figuras 6A-P ilustran la secuencia de nucleótidos (SEC ID n° 2) del vector de expresión DHFR/intrón- (SD de ras WT)-TNFr-IgG del Ejemplo 2.

Las Figuras 7A-B son gráficos de columnas que ilustran la expresión de TNFr-IgG utilizando vectores dicistrónicos o de control (ver el Ejemplo 2). Se construyeron vectores que contenían TNFr-IgG (aunque aparte de ello idénticos

## ES 2 282 994 T3

a aquellos descritos para la expresión de tPA en el Ejemplo 1), se introdujeron en células dp12.CHO mediante electroporación, se agruparon y se sometieron a ensayo para la expresión de producto antes (Figura 7A) y después (Figura 7B) de someterse a amplificación en Mtx 200 nM.

5 La Figura 8 ilustra esquemáticamente el constructo de ADN utilizado para la expresión de la V<sub>L</sub> de anti-IgE en el Ejemplo 3.

Las Figuras 9A-O ilustran la secuencia de nucleótidos (SEC ID n° 3) del vector de expresión de V<sub>H</sub> de anti-IgE del Ejemplo 3.

10 Las Figuras 10A-Q ilustran la secuencia de nucleótidos (SEC ID n° 4) del vector de expresión de V<sub>L</sub> de anti-IgE del Ejemplo 3.

La Figura 11 es un gráfico de columnas que ilustra la expresión de anti-IgE en el Ejemplo 3.

15 Se construyeron vectores de expresión de cadena pesada (V<sub>H</sub>) y ligera (V<sub>L</sub>), se coelectroporaron en células CHO y los clones se seleccionaron y se sometieron a ensayo para la expresión de anticuerpos. Además, se establecieron pools y se evaluaron con respecto a la expresión antes y después de la selección en Mtx a 200 nM y 1 μM.

### 20 Descripción de las realizaciones preferentes

#### Definiciones

25 El "constructo de ADN" dado a conocer en la presente invención comprende una molécula de ADN de origen no natural que puede proporcionarse aislada o integrada en otra molécula de ADN, por ejemplo en un vector de expresión o como el cromosoma de una célula huésped eucariótica.

30 La expresión "gen seleccionable" tal como se utiliza en la presente memoria se refiere a un ADN que codifica un marcador seleccionable necesario para el crecimiento o supervivencia de una célula huésped bajo las condiciones de cultivo celular particulares seleccionadas. Por consiguiente, una célula huésped que se transforme con un gen seleccionable será capaz de crecer o de sobrevivir bajo determinadas condiciones de cultivo celular, en las que una célula huésped no transfectada no será capaz de crecer o de sobrevivir. Típicamente un gen seleccionable proporcionará resistencia a un fármaco o compensará un defecto metabólico o catabólico en la célula huésped. Se proporcionan ejemplos de genes seleccionables en la tabla siguiente (ver también Kaufman, Methods in Enzymology 185:537-566, 35 1990, para una revisión de los mismos).

TABLA 1

Genes seleccionables y sus agentes de selección	
Agente de selección	Gen seleccionable
Metotrexato	Dihidrofolato reductasa
45 Cadmio	Metalotioneína
PALA	CAD
50 Xyl-A-or adenosina y 2' - desoxicoformicina	Adenosina desaminasa
Adenina, azaserina y coformicina	Adenilato desaminasa
55 6-Azauridina, pirazofurán	UPM Sintetasa
Ácido micofenólico	IMP 5' -deshidrogenasa
60 Ácido micofenólico con xantina limitante	Xantina-guanina fosforibosiltransferasa
Hipoxantina, aminopterina y 65 timidina (HAT)	HGPRTasa mutante o timidina quinasa mutante
5-Fluorodesoxiuridina	Timidilato sintetasa

Múltiples fármacos, por ejemplo adriamicina, vincristina o colchicina	P-glucoproteína 170
Afidicolina	Ribonucleótido reductasa
Metionina sulfoximina	Glutamina sintetasa
$\beta$ -Aspartil hidroxamato o Albizina	Asparagina sintetasa
Canavanina	Arginosuccinato sintetasa
$\alpha$ -Difluorometilornitina	Ornitina descarboxilasa
Compactina	HMG-CoA reductasa
Tunicamicina	N-Acetilglucosaminil transferasa
Borrelidina	Treonil-ARNt sintetasa
Ouabaína	Na <sup>+</sup> K <sup>+</sup> -ATPasa

Según la presente invención, el gen seleccionable es un gen amplificable. Tal como se utiliza en la presente invención, la expresión "gen amplificable" se refiere a un gen que se amplifica (es decir, se generan copias adicionales del gen que sobreviven en forma intracromosómica o extracromosómica) bajo determinadas condiciones. El gen amplificable habitualmente codifica un enzima (es decir, un marcador amplificable) que resulta necesario para el crecimiento de las células eucarióticas bajo dichas condiciones. Por ejemplo, el gen puede codificar DHFR que se amplifica cuando una célula huésped transformada con el mismo se cultiva en Mtx. Según Kaufman, los genes seleccionables en la Tabla 1 proporcionada anteriormente también pueden considerarse genes amplificables. Un ejemplo de un gen seleccionable que generalmente no se considera que sea un gen amplificable es el gen de resistencia a la neomicina (Cepko *et al.*, *supra*).

Tal como se utiliza en la presente invención, la expresión "medio selectivo" se refiere a la solución nutritiva utilizada para cultivar células eucarióticas que presentan el gen seleccionable y por lo tanto incluye un "agente de selección". Los medios comercialmente disponibles, tales como F-10 de Ham (Sigma), medio mínimo esencial ([MEM], Sigma), RPMI-1640 (Sigma) y medio de Eagle modificado por Dulbecco ([DMEM], Sigma) son soluciones nutritivas ejemplares. Además, puede utilizarse cualquiera de los medios indicados en Ham y Wallace, *Meth. Enz.* 58:44, 1979; Barnes y Sato, *Anal. Biochem.* 102:255, 1980; patentes US n° 4.767.704, n° 4.657.866, n° 4.927.762 o n° 4.560.655; patentes WO n° 90/03430, n° 87/00195; patentes US n° Re. 30.985 o n° 5.122.469) como medios de cultivo. Cualquiera de estos medios puede suplementarse según resulte necesario con hormonas y/o otros factores de crecimiento (tales como insulina, transferrina o factor de crecimiento epidérmico), sales (tales como cloruro sódico, calcio, magnesio y fosfato), tampones (tales como HEPES), nucleósidos (tales como adenosina y timidina), antibióticos (tales como el fármaco Gentamicin<sup>TM</sup>), elementos traza (definidos como compuestos inorgánicos habitualmente presentes a concentraciones finales en el rango micromolar) y glucosa o una fuente de energía equivalente. También puede incluirse cualquier otro suplemento necesario a concentraciones apropiadas, que resultarían conocidas por los expertos en la materia. La solución nutritiva preferente comprende suero de feto bovino.

La expresión "agente de selección" se refiere a una sustancia que interfiere con el crecimiento o supervivencia de una célula huésped que es deficiente en un gen seleccionable particular. Se presentan ejemplos de agentes de selección en la Tabla 1 anteriormente proporcionada. El agente de selección preferentemente comprende un "agente de amplificación" que se define para los fines de la presente invención como un agente para amplificar copias del gen amplificable, tal como Mtx si el gen amplificable es DHFR. Ver la Tabla 1 para ejemplos de agentes de amplificación.

Tal como se utiliza en la presente invención, la expresión "sitio de inicio transcripcional" se refiere al ácido nucleico en el constructo de ADN correspondiente al primer ácido nucleico incorporado en el transcrito primario, es decir, el precursor ARNm, sitio generalmente proporcionada en el extremo 5' del constructo de ADN o contiguo al mismo.

La expresión "sitio de terminación transcripcional" se refiere a una secuencia de ADN, normalmente representada en el extremo 3' del constructo de ADN, que causa que la ARN polimerasa termine la transcripción.

Tal como se utiliza en la presente invención, la expresión “región reguladora transcripcional” se refiere a una región del constructo de ADN que regula la transcripción del gen seleccionable y el gen de producto. La región reguladora transcripcional normalmente se refiere a una secuencia de promotor (es decir, una región de ADN implicada en la unión de la ARN polimerasa para iniciar la transcripción) que puede ser constitutiva o inducible y, opcionalmente, un intensificador (es decir, un elemento de ADN que actúa en *cis*, habitualmente de entre aproximadamente 10 y 300 pb, que actúa sobre un promotor incrementando su transcripción).

Tal como se utiliza en la presente invención, “gen de producto” se refiere a ADN que codifica una proteína deseada o producto polipéptido. Puede utilizarse cualquier gen de producto que es capaz de expresión en una célula huésped, aunque los procedimientos de la invención resultan particularmente adecuados para obtener la expresión de nivel elevado de un gen de producto que no sea también un gen seleccionable o amplificable. Por consiguiente, la proteína o polipéptido codificada por un gen de producto típicamente será uno que no resulte necesario para el crecimiento o la supervivencia de una célula huésped bajo las condiciones de cultivo celular particulares seleccionadas. Por ejemplo, los genes de producto convenientemente codifican un péptido, o pueden codificar una secuencia polipeptídica de aminoácidos para la que la longitud de cadena resulte suficiente para producir niveles más elevados de estructura terciaria y/o cuaternaria.

Entre los ejemplos de polipéptidos o proteínas bacterianos se incluyen, por ejemplo, fosfatasa alcalina y  $\beta$ -lactamasa. Entre los ejemplos de polipéptidos o proteínas de mamífero se incluyen moléculas, tales como renina; una hormona de crecimiento, incluyendo la hormona de crecimiento humana, y la hormona de crecimiento bovina; el factor de liberación de hormona de crecimiento; la hormona paratiroidea; la hormona estimulante del tiroides; lipoproteínas; antitripsina alfa-1; cadena A de la insulina; cadena B de la insulina; proinsulina; hormona folículo-estimulante; calcitonina; hormona luteinizante; glucagón; factores de coagulación, tales como el factor VIII, factor IX, factor tisular y factor de von Willebrands; factores de anticoagulación, tales como la Proteína C; factor natriurético atrial; surfactante pulmonar; un activador del plasminógeno, tal como uroquinasa u orina humana o activador del plasminógeno de tipo tisular (t-PA); bombesina; trombina, factor de crecimiento hematopoyético; factor de necrosis tumoral alfa y beta; encefalinasa; RANTES (regulada tras su activación, normalmente expresado y secretado por células T); proteína inflamatoria de macrófagos humanos (MIP-1-alfa); una albúmina sérica, tal como la albúmina de suero humano; sustancia inhibidora mulleriana; cadena A de la relaxina; cadena B de la relaxina; prorelaxina; péptido asociado a gonadotropina de ratón; una proteína microbiana, tal como beta-lactamasa; ADNasa; inhibina; activina; factor de crecimiento endotelial vascular (VEGF); receptores de hormonas o factores de crecimiento; integrina; proteína A o D; factores reumatoides; un factor neurotrófico, tal como el factor neurotrófico derivado de hueso (BDNF), neurotrofina 3, 4, 5 ó 6 (NT-3, NT-4, NT-5 o NT-6) o un factor de crecimiento nervioso, tal como NGF- $\beta$ ; factor de crecimiento derivado de plaquetas (PDGF); factor de crecimiento fibroblástico, tal como aFGF y bFGF; factor de crecimiento epidérmico (EGF); factor de crecimiento transformante (TGF), tal como TGF-alfa y TGF-beta, incluyendo TGF- $\beta$ 1, TGF- $\beta$ 2, TGF- $\beta$ 3, TGF- $\beta$ 4 o TGF- $\beta$ 5; factor I y II de crecimiento similar a la insulina (IGF-I e IGF-II); des(1-3)-IGF-I (IGF-I cerebral), proteínas de unión a factor de crecimiento similar a la insulina; proteínas CD, tales como CD-3, CD-4, CD-8 y CD-19; eritropoyetina; factores osteoinductivos; inmunotoxinas; una proteína morfogénica ósea (BMP); un interferón, tal como interferón-alfa, beta y gamma; factores estimulantes de colonias (CSFs), por ejemplo M-CSF, GM-CSF y G-CSF; interleuquinas (ILs), por ejemplo IL-1 a IL-10; superóxido dismutasa; receptores de células T; proteínas membranales de superficie; factor acelerador de degradación; antígeno vírico, tal como, por ejemplo, una parte de la cubierta del SIDA; proteínas de transporte; receptores de alojamiento; adhesinas; proteínas reguladoras; anticuerpos; proteínas quiméricas, tales como inmunoadhesinas y fragmentos de cualquiera de los polipéptidos anteriormente indicados.

El gen de producto preferentemente no consiste de una secuencia antisentido para inhibir la expresión de un gen presente en el huésped. Las proteínas preferentes en la presente invención son proteínas terapéuticas, tales como TGF- $\beta$ , TGF- $\alpha$ , PDGF, EGF, FGF, IGF-I, ADNasa, activadores del plasminógeno, tales como t-PA, factores de coagulación, tales como factor tisular y factor VIII, hormonas, tales como la relaxina y la insulina, citoquinas, tales como IFN- $\gamma$ , proteínas quiméricas, tales como receptor de TNF-inmunoadhesina IgG (TNFr-IgG) o anticuerpos, tales como anti-IgE.

El término “intrón” tal como se utiliza en la presente invención se refiere a una secuencia de nucleótidos presente dentro de la región transcrita de un gen o dentro de un precursor de ARN mensajero, secuencia de nucleótidos que es capaz de resultar extraída, o procesada, a partir del precursor de ARN mensajero por una célula huésped previamente a la traducción. Los intrones adecuados para la utilización en la presente invención se preparan convenientemente mediante cualquiera de entre varios procedimientos que son bien conocidos en la técnica, tales como la purificación a partir de un ácido nucleico de origen natural o la síntesis *de novo*. Los intrones presentes en muchos genes eucarióticos de origen natural han sido identificados y caracterizados (Mount, Nucl. Acids Res. 10:459, 1982). Los intrones artificiales que comprenden sitios de empalme funcionales también han sido descritos (Winey *et al.*, Mol. Cell Biol. 9:329, 1989; Gattermann *et al.*, Mol. Cell Biol. 9:1526, 1989). Los intrones pueden obtenerse a partir de ácidos nucleicos de origen natural, por ejemplo mediante digestión de un ácido nucleico de origen natural con una endonucleasa de restricción adecuada, o mediante clonación por PCR utilizando cebadores complementarios a las secuencias en los extremos 5' y 3' del intrón. Alternativamente, pueden prepararse intrones de secuencia y longitud definidas de manera sintética utilizando diversos procedimientos de la química orgánica (Narang *et al.*, Meth. Enzymol. 68:90, 1979; Caruthers *et al.*, Meth. Enzymol. 154:287, 1985; Froehler *et al.*, Nucl. Acids Res. 14:5399, 1986).

Tal como se utiliza en la presente invención, la expresión “sitio donador de empalme” o “SD” se refiere a la secuencia de ADN inmediatamente circundante al límite exón-intrón en el extremo 5' del intrón, donde el “exón”

comprende el ácido nucleico 5' respecto al intrón. Muchos sitios donadores de empalme han sido caracterizados y Ohsima *et al.*, J. Mol. Biol. 195:247-259, 1987, proporciona una revisión de los mismos. La expresión "una secuencia donadora de empalme eficiente" se refiere a una secuencia de ácidos nucleicos codificante de un sitio donador de empalme en la que la eficiencia de procesamiento de los precursores de ARN mensajero con la secuencia donadora de empalme se encuentra comprendida entre aproximadamente 80% y 99%, y preferentemente entre 90% y 95% según determinación mediante PCR cuantitativa. Entre los ejemplos de secuencias donadoras de empalme eficientes se incluyen la secuencia donadora de empalme ras de tipo salvaje (WT) y la secuencia GAC:GTAAGT del Ejemplo 3. Pueden seleccionarse fácilmente otras secuencias donadoras de empalme eficientes utilizando las técnicas para medir la eficiencia de procesamiento dada a conocer en la presente invención.

Las expresiones "PCR" y "reacción en cadena de la polimerasa" tal como se utiliza en la presente invención se refieren al procedimiento de amplificación *in vitro* descrito en la patente US nº 4.683.195 (publicada el 28 de julio de 1987). En general, el procedimiento de PCR implica ciclos repetidos de síntesis de extensión de cebador, utilizando dos cebadores de ADN capaces de hibridarse preferentemente con un molde de ácidos nucleicos que comprende la secuencia de nucleótidos que debe amplificarse. El procedimiento de PCR puede utilizarse para clonar secuencias de ADN específicas a partir del ADN genómico total, ADNc transcrito a partir de ARN celular y ADNs víricos o plasmídicos (Wang y Mark, en: PCR Protocols, páginas 70-75, Academic Press, 1990; Scharf, en: PCR Protocols, páginas 84-98; Kawasaki y Wang, en: PCR Technology, páginas 89-97, Stockton Press, 1989). Puede utilizarse la reacción en cadena de transcripción inversa-polimerasa (PCR-RT) para analizar muestras de ARN que contienen mezclas de transcritos de ARNm procesados y no procesados. Los cebadores etiquetados fluorescentemente diseñados para abarcar el intrón se utilizan para amplificar dianas tanto procesadas como no procesadas. Los productos de amplificación resultantes seguidamente se separan mediante electroforesis en gel y se cuantifican midiendo la emisión fluorescente de la banda o bandas apropiadas. Se lleva a cabo una comparación para determinar la cantidad de transcritos procesados y no procesados presentes en la muestra de ARN.

Una secuencia donadora de empalme preferente es una "secuencia donadora de empalme de consenso". Las secuencias de nucleótidos circundantes a los sitios de empalme de intrón, secuencias que se encuentran altamente conservadas evolutivamente, se denominan "secuencias donadoras de empalme de consenso". En los ARNm de los eucariotas superiores, el sitio de empalme 5 se encuentra dentro de la secuencia de consenso AG:GUAAGU (en la que los dos puntos indican el sitio de corte y ligación). En los ARNm de levadura, el sitio de empalme 5' está limitado por la secuencia de consenso :GUAUGU (Padgett *et al.*, Ann. Rev. Biochem. 55:1119, 1986).

La expresión "sitio aceptor de empalme" o "SA" se refiere a la secuencia inmediatamente circundante al límite intrón-exón en el extremo 3' del intrón, en el que "exón" comprende el ácido nucleico 3' respecto al intrón. Muchos sitios aceptores de empalme han sido caracterizados, y Ohsima *et al.*, J. Mol. Biol. 195:247-259, 1987, proporcionan una revisión de los mismos. El sitio aceptor de empalme preferente es un sitio aceptor de empalme eficiente que se refiere a una secuencia de ácidos nucleicos codificante de un sitio aceptor de empalme en el que la eficiencia de procesamiento de los precursores de ARN mensajero con el sitio aceptor de empalme se encuentra comprendida entre aproximadamente 80% y 99%, y preferentemente entre 90% y 95%, según determinación mediante PCR cuantitativa. El sitio aceptor de empalme puede comprender una secuencia de consenso. En los ARNm de los eucariotas superiores, el sitio aceptor de empalme 3' se encuentra dentro de la secuencia de consenso (U/C)<sub>11</sub>NCAG:G. En los ARNm de levadura, el sitio aceptor de empalme 3' está limitado por la secuencia de consenso (C/U)AG: (Padgett *et al.*, *supra*).

Tal como se utiliza en la presente invención, "cultivar durante un tiempo suficiente para permitir que se produzca la amplificación" se refiere al acto de cultivar físicamente las células huésped eucarióticas que han sido transformadas con el constructo de ADN en medios de cultivo celular que contienen el agente de amplificación, hasta que el número de copia del gen amplificable (y preferentemente también el número de copia del gen de producto) en las células huésped se ha incrementado respecto a las células transformadas previamente a este cultivo.

El término "expresión" tal como se utiliza en la presente invención se refiere a la transcripción o a la traducción que se produce dentro de una célula huésped. El nivel de expresión de un gen de producto en una célula huésped puede determinarse a partir de la cantidad del ARNm correspondiente que se encuentra presente en la célula, o de la cantidad de la proteína codificada por el gen de producto que es producida por la célula. Por ejemplo, el ARNm transcrito a partir de un gen de producto deseablemente se cuantifica mediante hibridación northern (Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, páginas 7.3-7.57 (Cold Spring Harbor Laboratory Press, 1989)). La proteína codificada por un gen de producto puede cuantificarse mediante ensayo de la actividad biológica proteína o utilizando ensayos que son independientes de esta actividad, tales como transferencia western o radioinmunoensayo utilizando anticuerpos capaces de reaccionar con la proteína (Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, páginas 18.1 a 18.88 (Cold Spring Harbor Laboratory Press, 1989)).

### Modos para llevar a cabo la invención

Se proporcionan procedimientos y composiciones para incrementar la estabilidad y/o número de copia de una secuencia transcrita con el fin de permitir la producción de niveles elevados de una secuencia de ARN de interés. En general, los procedimientos de la presente invención implican transfectar una célula huésped eucariótica con un vector de expresión que comprende tanto un gen de producto codificante de un polipéptido deseado y un gen seleccionable (preferentemente un gen amplificable).

## ES 2 282 994 T3

Pueden obtenerse genes seleccionables y genes de producto a partir de ADN genómico, ADNc transcrito a partir de ARN celular, o mediante síntesis *in vitro*. Por ejemplo, se criban bibliotecas con sondas (tales como anticuerpos u oligonucleótidos de aproximadamente 20 a 80 bases) diseñadas para identificar el gen seleccionable o el gen de producto (o la proteína o proteínas codificadas en el mismo). El cribado del ADNc o la biblioteca genómica con la sonda seleccionable puede llevarse a cabo utilizando procedimientos estándar tal como se describe en los capítulos 10 a 12 de Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual* (New York: Cold Spring Harbor Laboratory Press, 1989). Un medio alternativo de aislar el gen seleccionable o el gen de producto es utilizar metodología de PCR tal como se describe en la sección 14 de Sambrook *et al.*, *supra*.

Un procedimiento preferente de poner en práctica la presente invención es utilizar secuencias oligonucleótidas cuidadosamente seleccionadas para cribar bibliotecas de ADNc de diversos tejidos que es conocido que contienen el gen seleccionable o el gen de producto. Las secuencias oligonucleótidas seleccionadas como sondas deben ser de suficiente longitud y suficientemente inequívocas para que se minimicen los falsos positivos.

El oligonucleótido generalmente se marca de manera que pueda detectarse tras la hibridación con ADN en la biblioteca que se está cribando. El procedimiento preferente de marcaje es la utilización de ATP marcado con <sup>32</sup>P con polinucleótido quinasa, como es bien conocido en la técnica, para marcar radioactivamente el oligonucleótido. Sin embargo, pueden utilizarse otros procedimientos para marcar el oligonucleótido, incluyendo, aunque sin limitarse a ellos, la biotilación o el marcaje enzimático.

Ocasionalmente, el ADN codificante del gen seleccionable y del gen de producto se encuentra precedido por ADN codificante de una secuencia de señal que presenta un sitio de corte específico en el extremo N-terminal de la proteína madura o polipéptido. En general, la secuencia de señal puede ser un componente del vector de expresión, o puede ser una parte del gen seleccionable o del gen de producto que se inserta en el vector de expresión. Si se utiliza una secuencia de señal heteróloga, preferentemente es una que resulte reconocida y procesada (es decir, cortada por una peptidasa de señal) por la célula huésped. Para la secreción en levadura, la secuencia de señal nativa puede sustituirse por, por ejemplo, el líder de invertasa de levadura, el líder del factor alfa (incluyendo los líderes de factor  $\alpha$  de *Saccharomyces* y *Kluyveromyces*, estando descritos estos últimos en la patente US n° 5.010.182, publicada el 23 de abril de 1991) o el líder de fosfatasa ácida, el líder de la glucoamilasa de *C. albicans* (patente EP n° 362.179, publicada el 4 de abril de 1990) o la señal descrita en la patente WO n° 90/13646, publicada el 15 de noviembre de 1990. En la expresión en células de mamífero, la secuencia de señal nativa de la proteína de interés es satisfactoria, aunque pueden resultar adecuadas otras secuencias de señal de mamífero, tales como secuencias de señal de polipéptidos secretados de la misma especie o de una especie relacionada, así como líderes secretorios víricos, por ejemplo la señal gD de Herpes simplex. El ADN para esta región precursora se encuentra ligado en el mismo marco de lectura que el gen seleccionable o que el gen de producto.

Tal como se muestra en la Figura 1ª, el gen seleccionable se proporciona en el extremo 5' del constructo de ADN y a este gen seleccionable le sigue el gen de producto. Por lo tanto, el mensaje de longitud completa (no procesado) contendrá DHFR como el primer marco de lectura abierto y por lo tanto generará proteína DHFR para permitir la selección de los transfectantes estables. El mensaje de longitud completa no se espera que genere cantidades apreciables de la proteína de interés debido a que el segundo AUG en el mensaje dicistrónico es un iniciador ineficiente de traducción en las células de mamífero (Kozak, *J. Cell Biol.* 115:887-903, 1991).

El gen seleccionable se encuentra situado dentro de un intrón. Los intrones son secuencias de nucleótidos no codificantes, normalmente presentes dentro de muchos genes eucarióticos, que se eliminan de los precursores ARNm nuevamente transcritos en un procedimiento multietapa denominado colectivamente empalme.

Se cree que un mecanismo único es responsable del empalme de los precursores ARNm en células de mamífero, vegetales y de levadura. En general, el procedimiento de empalme requiere que los extremos 5' y 3' del intrón se corten correctamente y los extremos resultantes del ARNm se unan con exactitud, de manera que se produce un ARNm maduro con el marco de lectura apropiado para la síntesis de proteína. El análisis de una diversidad de genes mutantes de origen natural y construidos sintéticamente ha demostrado que los cambios de nucleótidos en muchas de las posiciones dentro de las secuencias de consenso en los sitios de empalme 5' y 3' presenta el efecto de reducir o eliminar la síntesis de ARNm maduro (Sharp, *Science* 235:766, 1987; Padgett *et al.*, *Ann. Rev. Biochem.* 55:1119, 1986; Green, *Ann. Rev. Genet.* 20:671, 1986). Los estudios mutacionales también han demostrado que las estructuras secundarias de ARN que implican sitios de empalme pueden afectar a la eficiencia del empalme (Solnick, *Cell* 43:667, 1985; Konarska *et al.*, *Cell* 42:165, 1985).

La longitud del intrón también puede afectar a la eficiencia del empalme. Realización mutaciones por delección de diferentes tamaños dentro del intrón grande del gen de la beta-globina de conejo, Wieringa *et al.* determinaron que la longitud mínima del intrón necesaria para el empalme correcto es de aproximadamente 69 nucleótidos (*Cell* 37:915, 1984). Algunos estudios similares del intrón de la región E1A del adenovirus han demostrado que una longitud de intrón de aproximadamente 78 nucleótidos permite que se produzca un empalme correcto, aunque con una eficiencia reducida. El incremento de la longitud del intrón hasta 91 nucleótidos restaura la eficiencia normal del empalme, mientras que el truncado del intrón a 63 nucleótidos elimina el empalme correcto (Ulfendahl *et al.*, *Nucl. Acids Res.* 13:6299, 1985).

## ES 2 282 994 T3

Para que resulte útil en la invención, el intrón debe presentar una longitud que permita que el procesamiento del intrón a partir del ARNm resulte eficiente. La preparación de intrones de diferentes longitudes es una cuestión rutinaria, implicando procedimientos bien conocidos en la técnica, tal como síntesis de novo o la mutagénesis por delección *in vitro* de un intrón existente. Típicamente, el intrón presentará una longitud de por lo menos aproximadamente 150 nucleótidos, debido a que los intrones que son más cortos que esta longitud tienden a procesarse menos eficientemente. El límite superior para la longitud del intrón puede ser de hasta 30 kB o más. Sin embargo, como proposición general, el intrón presenta una longitud generalmente inferior a aproximadamente 10 kB.

El intrón se modifica para contener el gen seleccionable no presente normalmente dentro del intrón utilizando cualquiera de los diversos procedimientos conocidos para modificar un ácido nucleico *in vitro*. Típicamente, se introduce un gen seleccionable en un intrón en primer lugar cortando el intrón con una endonucleasa de restricción, y después uniendo covalentemente los fragmentos de restricción resultantes con el gen seleccionable en la orientación correcta para la expresión de la célula huésped, por ejemplo mediante ligación con un enzima ADN ligasa.

El constructo de ADN es dicistrónico, es decir, el gen seleccionable y el gen de producto se encuentran ambos bajo el control transcripcional de una sola región reguladora transcripcional. Tal como se ha indicado anteriormente, la región reguladora transcripcional comprende un promotor. Entre las secuencias promotoras adecuadas para la utilización con huéspedes levadura se incluyen los promotores para la 3-fosfoglicerato quinasa (Hitzeman *et al.*, J. Biol. Chem. 255:2073, 1980) u otros enzimas glucolíticos (Hess *et al.*, J. Adv. Enzyme Reg. 7:149, 1968, y Holland, Biochemistry 17:4900, 1978), tales como enolasa, gliceraldehído-3-fosfato deshidrogenasa, hexoquinasa, piruvato descarboxilasa, fosfofructoquinasa, glucosa-6-fosfato isomerasa, 3-fosfogliceratomutasa, piruvato quinasa, trifosfato isomerasa, fosfoglucosa isomerasa y glucoquinasa.

Otros promotores de levadura, que son promotores inducibles que presentan la ventaja adicional de que la transcripción se encuentra controlada por las condiciones de cultivo, son las regiones promotoras para la alcohol deshidrogenasa 2, isocitocromo C, fosfatasa ácida, enzimas degradativos asociados con el metabolismo del nitrógeno, metalotioneína, gliceraldehído-3-fosfato deshidrogenasa, y enzimas responsables de la utilización de la maltosa y de la galactosa. Los vectores y promotores adecuados para la utilización en la expresión de levadura se describen adicionalmente en Hitzeman *et al.*, patente EP n° 73.657A. Los intensificadores de levadura se utilizan ventajosamente con promotores de levadura.

Las secuencias de control de expresión son conocidas para los eucariotas. Virtualmente todos los genes eucarióticos presentan una región rica en AT situada aproximadamente 25 a 30 bases corriente arriba del sitio en el que se inicia la transcripción. Otra secuencia situada 70 a 80 bases corriente arriba del inicio de transcripción de muchos genes es una región CXCAAT en la que X puede ser cualquier nucleótido.

La transcripción de gen de producto a partir de vectores en células huésped de mamífero se encuentra controlada por promotores obtenidos de los genomas de virus, tales como el virus de polio, virus de la viruela aviar (patente UK n° 2.211.504 publicada el 5 de julio de 1989), adenovirus (tales como el adenovirus 2), virus del papiloma bovino, virus del sarcoma aviar, citomegalovirus, un retrovirus, virus de la hepatitis B y más preferentemente virus 40 del simio (SV40) a partir de promotores de mamífero heterólogos, por ejemplo el promotor de actina o un promotor de inmunoglobulina, a partir de promotores de choque térmico, y a partir del promotor asociado normalmente con el gen de producto, con la condición de que estos promotores sean compatibles con los sistemas de las células huésped.

Los promotores tempranos y tardíos del virus SV40 se obtienen convenientemente como fragmento de restricción de SV40 que también contiene el origen de replicación vírico de SV40 (Fiers *et al.*, Nature 273:113, 1978; Mulligan y Berg, Science 209:1422-1427, 1980; Pavlakis *et al.*, Proc. Natl. Acad. Sci. USA 78:7398-7402, 1981). El promotor temprano inmediato del citomegalovirus humano (CMV) se obtiene convenientemente en forma de fragmento E de restricción HindIII (Greenaway *et al.*, Gene 18:355-360, 1982). Se da a conocer un sistema para expresar ADN en huéspedes de mamífero utilizando el virus del papiloma bovino como vector en la patente US n° 4.419.446. Se describe una modificación de este sistema en la patente US n° 4.601.978. Ver también Gray *et al.*, Nature 295:503-508, 1982 sobre la expresión de ADNc codificante de interferón inmunológico en células de mono; Reyes *et al.*, Nature 297:598-601, 1982 sobre la expresión del ADNc del  $\beta$ -interferón humano en células de ratón bajo el control de un promotor de la timidina quinasa del virus del herpes simplex (Canaani y Berg, Proc. Natl. Acad. Sci. USA 79:5166-5170, 1982, sobre la expresión del gen del interferón  $\beta$ 1 humano en cultivo de células de ratón y de conejo, y Gorman *et al.*, Proc. Natl. Acad. Sci. USA 79:6777-6781, 1982, sobre la expresión de secuencias CAT bacterianas en células de riñón de mono CV-1, fibroblastos de embrión de pollo, células de ovario de hámster chino, células HeLa y células NIH-3T3 de ratón utilizando la repetición terminal larga del virus del sarcoma de Rous como promotor.

Preferentemente, la región reguladora transcripcional en eucariotas superiores comprende una secuencia intensificadora. Los intensificadores son relativamente independientes de la orientación y de la posición, al encontrarse en orientación 5' (Lainins *et al.*, Proc. Natl. Acad. Sci. USA 78:993, 1981) y 3' (Lusky *et al.*, Mol. Cell Biol. 3:1108, 1983) respecto a la unidad de transcripción, dentro de un intrón (Banerji *et al.*, Cell 33:729, 1983), así como dentro de la secuencia codificante misma (Osborne *et al.*, Mol. Cell Bio. 4:1293, 1984). En la actualidad se conocen muchas secuencias intensificadoras de genes de mamífero (globina, elastasa, albúmina,  $\alpha$ -fetoproteína e insulina). Típicamente, sin embargo, se utiliza un intensificador de un virus de célula eucariótica. Entre los ejemplos se incluyen el intensificador de SV40 en el lado tardío del origen de replicación (pb de 100 a 270), el intensificador del promotor temprano de citomegalovirus (CMV), el intensificador de polio en el lado tardío del origen de replicación y los intensificadores

## ES 2 282 994 T3

de adenovirus (ver también Yaniv, Nature 297:17-18, 1982) sobre los elementos intensificadores para la activación de promotores eucarióticos). El intensificador puede empalmarse en el vector en una posición 5' o 3' del gen de producto, pero preferentemente se sitúa en un sitio 5' respecto al promotor.

5 El constructo de ADN presenta un sitio de inicio transcripcional después de la región reguladora transcripcional y una región de terminación transcripcional después del gen de producto (ver la Figura 1A). Estas secuencias se proporcionan en el constructo de ADN utilizando técnicas que son bien conocidas en la técnica.

10 El constructo de ADN normalmente forma parte de un vector de expresión que puede presentar otros componentes, tales como un origen de replicación (es decir, una secuencia de ácidos nucleicos que permite que el vector se replique en una o más células huésped seleccionadas) y, si se desea, uno o más genes seleccionables adicionales. La construcción de vectores adecuados que contienen las secuencias codificantes y de control deseadas utiliza técnicas de ligación estándares. Los plásmidos aislados o fragmentos de ADN se cortan, se ajustan y se religan en la forma deseada para generar los plásmidos requeridos.

15 Generalmente, en los vectores de clonación, el origen de replicación es uno que permite que el vector se replique independientemente del ADN cromosómico del huésped, e incluye orígenes de replicación o secuencias de replicación autónoma. Estas secuencias son bien conocidas. El origen de replicación del plásmido  $2\mu$  es adecuado para la levadura y diversos orígenes víricos (SV40, poliovirus, adenovirus, VSV o BPV) resultan útiles para los vectores de clonación en las células de mamífero. Generalmente, el componente origen de replicación no resulta necesario para los vectores de expresión de mamífero (el origen de SV40 típicamente puede utilizarse sólo porque contiene el promotor temprano).

20 La mayoría de los vectores de expresión son vectores "lanzadera", es decir, son capaces de replicación en por lo menos una clase de organismos, pero pueden transfectarse en otro organismo para la expresión. Por ejemplo, se clona un vector en *E. coli* y después el mismo vector se transfecta en células de levadura o de mamífero para la expresión, aunque no es capaz de replicarse independientemente del cromosoma de la célula huésped.

25 Para el análisis de confirmación de las secuencias correctas en los plásmidos, los plásmidos de los transformantes se preparan, se analizan mediante restricción y/o se secuencian mediante el procedimiento de Messing *et al.*, Nucleic Acids Res. 9:309, 1981, o mediante el procedimiento de Maxam *et al.*, Methods in Enzymology 65:499, 1980.

30 El vector de expresión con el constructo de ADN preparado tal como se ha comentado anteriormente se transforma en una célula huésped eucariótica. Las células huésped adecuadas para la clonación o la expresión de los vectores de la presente invención son células de levadura o de eucariotas superiores.

35 Los microbios eucarióticos, tales como hongos filamentosos o levaduras, son huéspedes adecuados para vectores que contienen el gen de producto. *Saccharomyces cerevisiae*, o levadura común de panadería, es el utilizado más frecuentemente de entre los microorganismos huésped eucarióticos inferiores. Sin embargo, se encuentran disponibles comúnmente y resultan útiles en la presente invención varios otros géneros, especies y cepas, tales como *S. pombe* (Beach y Nurse, Nature 290:140, 1981), *Kluyveromyces lactis* (Louvencourt *et al.*, J. Bacteriol. 737, 1983), *Yarrowia* (patente EP nº 402.226), *Pichia pastoris* (patente EP nº 183.070), *Trichoderma reesia* (patente EP nº 244.234), *Neurospora crassa* (Case *et al.*, Proc. Natl. Acad. Sci. USA 76:5259-5263, 1979), y huéspedes *Aspergillus*, tales como *A. nidulans* (Ballance *et al.*, Biochem. Biophys. Res. Commun. 112:284-289, 1983, Tilburn *et al.*, Gene 26:205-221, 1983; Yelton *et al.*, Proc. Natl. Acad. Sci. USA 81:1470-1474, 1984, y *A. niger* (Kelly y Hynes, EMBO J. 4:475-479, 1985).

40 Las células huésped adecuadas para la expresión del gen de producto se derivan organismos multicelulares. Estas células huésped son capaces de actividades complejas de procesamiento y glucosilación. Como regla general, puede utilizarse cualquier cultivo celular eucariótico superior, procedente de cultivo de vertebrado o de invertebrado. Entre los ejemplos de células de invertebrado se incluyen las células vegetales y de insecto. Se han identificado numerosas cepas y variantes de baculovirus y células huésped de insecto permisivos correspondientes, tales como células huésped de *Spodoptera frugiperda* (oruga), *Aedes aegypti* (mosquito), *Aedes albopictus* (mosquito), *Drosophila melanogaster* (mosca de la fruta) y *Bombyx mori*. Ver, por ejemplo, Lucow *et al.*, Bio/Technology 6:47-55, 1988; Miller *et al.*, en: Genetic Engineering, Setlow, J.K. *et al.*, editores, vol. 8 (Plenum Publishing, 1986), páginas 277 a 279; y Maeda *et al.*, Nature 315:592-594, 1985. Se encuentran a disposición del público una diversidad de dichas cepas víricas, por ejemplo la variante L-1 de Autographa californica NPV y la cepa Bm-5 de *Bombyx mori* NPV, y estos virus pueden utilizarse como el virus según la presente invención, particularmente para la transfección de células de *Spodoptera frugiperda*.

45 Pueden utilizarse como huéspedes cultivos de células vegetales de algodón, maíz, patata, soja, petunia, tomate y tabaco. Típicamente, se transfectan células vegetales mediante incubación con determinadas cepas de la bacteria *Agrobacterium tumefaciens*, que han sido previamente manipuladas para contener el gen de producto. Durante la incubación del cultivo de células vegetales con *A. tumefaciens*, el gen de producto se transfiere al huésped célula vegetal de manera que se resulte transfectado y exprese, bajo las condiciones apropiadas, el gen de producto. Además, se encuentran disponibles secuencias reguladoras y de señal compatibles con las células vegetales, tal como el promotor de la nopalina sintasa y las secuencias de señal de poliadenilación (Depicker *et al.*, J. Mol. Appl. Gen. 1:561, 1982). Además, algunos segmentos de ADN aislados de la región corriente arriba del gen 780 de ADN-T son capaces de activar o de incrementar los niveles de transcripción de genes expresables en plantas en tejido vegetal que contiene ADN recombinante. Patente EP nº 321.196 publicada el 21 de junio de 1989.

Sin embargo, el mayor interés ha sido en las células de vertebrado, y la propagación de las células de vertebrado en cultivo (cultivo de tejido) se ha convertido en un procedimiento rutinario en años recientes (Tissue Culture, Academic Press, Kruse y Patterson, editores, 1973). Son ejemplos de líneas celulares de mamífero huésped útiles la línea CV1 de riñón de mono transformada por SV40 (COS-7, ATCC CRL 1651); la línea renal embrionaria humana (293 o células 293 subclonadas para el crecimiento en cultivo en suspensión, Graham *et al.*, J. Gen. Virol. 36:59, 1977); las células renales de hámster neonato (BHK, ATCC CCL 10); células de ovario de hámster chino/-DHFR (CHO, Urlaub y Chasin, Proc. Natl. Acad. Sci. USA 77:4216, 1980); células CHO.dp12 (patente EP n° 307.247 publicada el 15 de marzo de 1989); células de sertoli de ratón (TM4, Mather, Biol. Reprod. 23:243-251, 1980); células renales de mono (CV1 ATCC CCL 70); células renales de mono verde africano (VERO-76, ATCC CRL-1587); células de carcinoma cervical humano (HELA, ATCC CCL 2); células renales caninas (MDCK, ATCC CCL 34); células hepáticas de rata búfalo (BRL 3A, ATCC CRL 144); células pulmonares humanas (W138, ATCC CCL 75); células hepáticas humanas (HepG2, HB 8065); tumor mamario de ratón (MMT 060562, ATCC CCL51); células TRI (Mather *et al.*, Annals N.Y. Acad. Sci. 383:44-68, 1982); células MRC 5; células FS4; y una línea de hepatoma humano (HepG2).

Las células huésped se transforman con los vectores de expresión o de clonación anteriormente descritos de la presente invención y cultivarse en medios nutritivos convencionales modificados según resulte apropiado para inducir promotores, seleccionar transformantes, o amplificar los genes codificantes de las secuencias deseadas.

Se utiliza la infección con *Agrobacterium tumefaciens* para la transformación de determinadas células vegetales, tal como describen Shaw *et al.*, Gene 23:315, 1983 y la patente WO n° 89/05859, publicada el 29 de junio de 1989. Para las células de mamífero sin estas paredes celulares, puede utilizarse el procedimiento de precipitación con fosfato de calcio de Graham y van der Eb, Virology 52:456-457, 1978. Los aspectos generales de las transformaciones de sistemas de células huésped de mamífero han sido descritos por Axel en la patente US n° 4.399.216, publicada el 16 de agosto de 1983. Las transformaciones en levadura típicamente se llevan a cabo según el procedimiento de Van Solingen *et al.*, J. Bact. 130:946, 1977 y Hsiao *et al.*, Proc. Natl. Acad. Sci. (USA) 76:3829, 1979. Sin embargo, también pueden utilizarse otros procedimientos para introducir ADN en las células, tales como la inyección nuclear o mediante fusión de protoplasto.

En la realización preferente, el ADN se introduce en las células huésped utilizando la electroporación (ver Anderson, J. Tiss. Cult. Meth. 15:56-62, 1993, para una revisión de las técnicas de electroporación que resultan útiles para poner en práctica la invención reivindicada. Se ha descubierto que las técnicas de electroporación para introducir el constructo de ADN en las células huésped resultan preferibles a las técnicas de precipitación con fosfato de calcio en la medida en que estas últimas podrían provocar la rotura del ADN y la formación de concatámeros.

Las células huésped de mamífero utilizadas para expresar el gen de producto en la presente invención pueden cultivarse en una diversidad de medios, tal como se comenta en la sección de definiciones anterior. El medio contiene el agente de selección utilizado para seleccionar las células huésped transformadas que han incorporado el constructo de ADN (como elemento intracromosómico o extracromosómico). Para conseguir la selección de las células eucarióticas transformadas, las células huésped pueden cultivarse en placas de cultivo celular y colonias individuales que expresan el gen seleccionable (y de esta manera el gen de producto) pueden aislarse y cultivarse en medio de cultivo hasta el agotamiento de los nutrientes. A continuación, las células huésped se analizan para la transcripción y/o transformación tal como se comenta posteriormente. Las condiciones de cultivo, tales como la temperatura, pH y similares, son aquéllas utilizadas anteriormente con la célula huésped seleccionada para la expresión, y resultarán evidentes para el experto ordinario en la materia.

La amplificación y/o expresión génica puede medirse en una muestra directamente, por ejemplo mediante transferencia southern, transferencia northern convencionales para cuantificar la transcripción en ARNm (Thomas, Proc. Natl. Acad. Sci. USA 77:5201-5205, 1980), transferencia de puntos (análisis de ADN), o hibridación *in situ*, utilizando una sonda apropiadamente marcada, basada en las secuencias proporcionadas en la presente invención. Pueden utilizarse diversos marcajes, más frecuentemente isótopos radioactivos, particularmente <sup>32</sup>P. Sin embargo, también pueden utilizarse otras técnicas, tales como la utilización de nucleótidos modificados con biotina para la introducción en un polinucleótido. A continuación, la biotina sirve como el sitio para la unión a avidina o a anticuerpos, que pueden marcarse con una amplia diversidad de marcajes, tales como radionucleidos, fluorescentes, enzimas o similares. Alternativamente, pueden utilizarse anticuerpos que pueden reconocer dúplex específicos, incluyendo dúplex de ADN, dúplex de ARN, y dúplex híbridos ADN-ARN o dúplex de ADN-proteína. Los anticuerpos a su vez pueden marcarse y el ensayo llevarse a cabo con el dúplex unido a una superficie, de manera que al formarse el dúplex sobre la superficie, puede detectarse la presencia de anticuerpo unido al dúplex.

Alternativamente, la expresión génica puede medirse mediante procedimientos inmunológicos, tales como la tinción inmunohistoquímica de secciones de tejido y el ensayo de un cultivo celular o de líquidos corporales, para cuantificar directamente la expresión del gen de producto. Con las técnicas de tinción inmunohistoquímica, se prepara una muestra celular, típicamente mediante deshidratación y fijación, seguida de la reacción con anticuerpos marcados específicos para el producto génico acoplado, en el que los marcajes habitualmente son visualmente detectables, tales como marcajes enzimáticos, marcajes fluorescentes, marcajes luminiscentes, y similares. Una técnica de tinción particularmente sensible adecuada para la utilización en la presente invención se describe en Hsu *et al.*, Am. J. Clin. Path. 75:735-738, 1980.

## ES 2 282 994 T3

En la realización preferente, el ARNm se analiza mediante PCR cuantitativa (para determinar la eficiencia de empalme) y la expresión de proteínas se mide utilizando ELISA tal como se describe en el Ejemplo 1 en la presente memoria.

5 El producto de interés preferentemente se recupera del medio de cultivo como polipéptido secretado, aunque también puede recuperarse de los lisados de células huésped cuando se expresa directamente sin una señal secretoria. Cuando el gen de producto se expresa en una célula recombinante que no es de origen humano, el producto de interés se encuentra completamente libre de proteínas o de polipéptidos de origen humano. Sin embargo, resulta necesario purificar el producto de interés de las proteínas o polipéptidos celulares recombinantes para obtener preparaciones que  
10 son sustancialmente homogéneas respecto al producto de interés. Como primera etapa, el medio de cultivo o lisado se centrifuga para eliminar los residuos celulares particulados. Seguidamente, el producto de interés se purifica de proteínas y polipéptidos solubles contaminantes, por ejemplo mediante fraccionamiento en columnas de inmunoafinidad o de intercambio iónico; precipitación con etanol; HPLC en fase reversa; cromatografía en resina de sílice o en una resina de intercambio catiónico, tal como DEAE; cromatografía en fase reversa; SDS-PAGE; precipitación con sulfato amónico; electroforesis en gel utilizando, por ejemplo Sephadex G-75; cromatografía en columnas de plasminógeno para unir  
15 el producto de interés y columnas de proteína A-sefariosa para eliminar contaminantes tales como IgG.

Los ejemplos siguientes se proporcionan a título de ilustración únicamente y no pretenden limitar la invención en modo alguno

### 20 Ejemplo 1

#### *Producción de tPA utilizando los vectores de expresión dicistrónicos*

25 Se buscó incrementar el nivel de homogeneidad con respecto a los niveles de expresión de clones estables mediante la expresión de un marcador seleccionable (tal como DHFR) y la proteína de interés a partir de un solo promotor. Estos vectores desvían la mayor parte del transcrito a la expresión de producto, ligándola simultáneamente a una proporción fija a la expresión de DHFR a través del procesamiento diferencial.

30 Se construyeron vectores que se derivaron del vector pRK (Suva *et al.*, Science 237:893-896, 1987) que contiene un intrón entre el promotor temprano inmediato del citomegalovirus (CMV) y el ADNc que codifica el polipéptido de interés. El intrón de pRK es de 139 nucleótidos de longitud, presenta un sitio donador de empalme derivado del gen temprano inmediato de citomegalovirus (CMVIE) y un sitio aceptor de empalme procedente de un gen de región variable (V<sub>H</sub>) de cadena pesada de IgG (Eaton *et al.*, Biochem. 25:8343, 1986).

35 Los vectores de DHFR/intrón se construyeron insertando un linker EcoRV en el sitio BSTX1 presente en el intrón de pRK7. Se insertó un fragmento de 830 pares de bases que contenía un fragmento codificante de DHFR de ratón para obtener vectores de expresión de DHFR-intrón que diferían únicamente en la secuencia que comprende el sitio donador de empalme. Estas secuencias se alteraron solapando la mutagénesis por PCR para obtener secuencias con  
40 sitios donadores de empalme correspondientes en los exones 3 y 4 de genes Ras normales y mutantes. También se utilizó la PCR para destruir el sitio donador de empalme.

Se insertó un fragmento de ADNc de DHFR de ratón (Simonsen *et al.*, Proc. Natl. Acad. Sci. USA 80:2495-2499, 1983) en el intrón de este vector 59 nucleótidos más abajo del sitio donador de empalme. El sitio donador de empalme  
45 de este vector se alteró mediante mutagénesis para modificar la proporción entre mensaje procesado a no procesado en las células transfectadas. Anteriormente se había demostrado que un solo cambio de nucleótido (G a A) convirtió un sitio donador de empalme relativamente eficiente en el gen ras normal en un sitio de empalme eficiente (Cohen *et al.*, Nature 334:119-124, 1988). Este efecto ha sido demostrado en el contexto del gen ras y ha sido confirmado cuando estas secuencias se transfirieron en constructos de hormona de crecimiento humana (Cohen *et al.*, Cell 58:461-  
50 472, 1989). Además, se construyó un sitio de empalme 5' no funcional (GT CA) como control ( $\Delta$ GT). Se insertó un polilinker 35 nucleótidos más abajo del sitio de empalme 3' para aceptar el ADNc de interés. Un vector que contenía tPA (Pennica *et al.*, Nature 301:214-221, 1983) se linearizó más abajo del sitio de poliadenilación antes de introducirlo en células CHO (Potter *et al.*, Proc. Natl. Acad. Sci. USA 81:7161, 1984).

55 Se introdujeron plásmidos de ADN que contenían DHFR/intrón, tPA y (a) ras de tipo salvaje (ras WT), ver la Figura 3 (SEC ID n° 1), (b) ras mutante, o (c) sitio donador de empalme no funcional ( $\Delta$ GT) en células CHO DHFR-negativas mediante electroporación. Se linearizaron cada uno de los vectores de intrón más abajo del sitio de poliadenilación mediante tratamiento con endonucleasa de restricción. El vector de control se linearizó más abajo del segundo sitio de poliadenilación. Los ADNs se precipitaron con etanol tras la extracción con fenol/cloroformo y se resuspendieron  
60 en 20  $\mu$ l de Tris EDTA 1/10. A continuación, se incubaron 10  $\mu$ g de ADN con 10<sup>7</sup> células CHO.dp12 (patente EP n° 307.247, publicada el 15 de marzo de 1989) en 1 ml de PBS sobre hielo durante 10 minutos antes de la electroporación a 400 voltios y 330  $\mu$ f utilizando un electroporador celular BRL.

Las células se introdujeron nuevamente en hielo durante 10 minutos antes de sembrarse en placas en medio no selectivo. Tras 24 horas, las células se alimentaron con medio libre de nucleósidos para seleccionar los clones DHFR<sup>+</sup>  
65 estables, que se agruparon. Los clones DHFR<sup>+</sup> agrupados se lisaron y se prepararon los ARNm.

## ES 2 282 994 T3

Para preparar el ARNm, se extrajo el ARN de  $5 \times 10^7$  células que se cultivaron a partir de pools de más de 200 clones derivados de la transfección estable de los tres vectores, la construcción esencial de los cuales se muestra en la Figura 1B y a partir de células CHO no transfectadas. Se purificó el ARN sobre oligo-DT celulasa (Collaborative Biomedical Products). A continuación, se sometieron  $10 \mu\text{g}$  de ARNm a transferencia northern, que implicó correr el ARNm en un gel de 1,2% de agarosa y 6,6% de formaldehído, y transferirlo a un filtro de nilón (membrana Stratagene Duralon-UV), prehibridado, sondeado y lavado siguiendo las instrucciones del fabricante.

El filtró se sondeó secuencialmente utilizando sondas (mostradas en la Figura 1B) que podían detectar: (a) el mensaje de longitud completa, (b) el mensaje de tanto longitud completa como procesado, o (c) beta-actina. El sondeo con la sonda larga demostró que el vector que contenía el sitio donador de empalme eficiente (es decir ras WT) generaba predominantemente un ARNm del tamaño predicho para el producto procesado, mientras que los otros dos vectores dieron lugar principalmente a un ARNm de un tamaño que correspondía al del mensaje no procesado. La sonda de DHFR detectó sólo el mensaje de longitud completa y demostró que el vector derivado de sitio donador de empalme ras WT generaba muy poco mensaje de longitud completa con el que proporcionar un fenotipo DHFR-positivo.

La Figura 4 muestra el número de colonias DHFR-positivas tras electroporaciones por duplicado con los tres vectores de intrón indicados anteriormente y de un vector convencional que presenta un promotor de CMV que regula tPA y un promotor de SV40 que regula DHFR (ver la Figura 2). El incremento del número de colonias se produjo en paralelo al incremento del mensaje de longitud completa acumulado con la modificación de los sitios donadores de empalme. El vector convencional genera colonias eficientemente y no es significativamente diferente del constructo  $\Delta\text{GT}$ .

El nivel de expresión de tPA se determinó sembrando células en 1 ml de F12:DMEM (50:50 con 5% de FBS) en placas de 24 pocillos hasta casi confluencia. El crecimiento de las células continuó hasta agotar el medio. A continuación, se sometió a ensayo el medio mediante ELISA para la producción de tPA. Brevemente, se recubrieron los pocillos de una placa de microtitulación de ELISA con anticuerpo anti-tPA, se añadieron muestras de medio a los pocillos, seguido de lavado. A continuación, se cuantificó la unión del antígeno (tPA) utilizando anticuerpo anti-tPA marcado con peroxidasa de rábano picante (HRPO).

La Figura 5A ilustra los títulos de proteína tPA secretada tras agrupar los clones de cada grupo mostrado en la Figura 4. Aunque el número de colonias se incrementó con un debilitamiento de la función donadora de empalme, se observó lo contrario con respecto a la expresión de tPA. Los niveles de expresión eran consistentes con los productos ARN que se observaron; a medida que se procesa más mensaje dicistrónico, una cantidad creciente de mensaje contendrá tPA como el primer marco de lectura abierto, resultando en una expresión incrementada de tPA. Una mutación de GT a CA en el sitio donador de empalme resulta en una abundancia de colonias DHFR-positivas que expresan niveles indetectables de tPA, resultando posiblemente de la utilización ineficiente del segundo AUG. Se subraya que la Figura 5A también muestra que los niveles de expresión obtenidos de uno de los vectores dicistrónicos (con SD de ras WT) eran aproximadamente tres veces superiores que los obtenidos con el vector de control que contiene un promotor/intensificador de CMV que regula tPA, un promotor/intensificador de SV40 que controla DHFR, y señales de poliadenilación de SV40 que controlan la expresión de tPA y DHFR.

Además, se investigó la homogeneidad de la expresión en los pools. La Figura 5B muestra que la totalidad de los 20 clones generados por los vectores dicistrónicos de sitio donador de empalme ras WT expresaban niveles detectables de tPA, mientras que únicamente 4 de 20 clones generados por el vector de control expresaban tPA. Ninguno de los clones transfectados con el vector no procesado ( $\Delta\text{GT}$ ) expresó niveles de tPA detectables mediante ELISA. Este resultado es consistente con observaciones anteriores de que relativamente pocos clones generados por vectores convencionales sintetizan niveles útiles de proteína.

Se incrementó la expresión de tPA tras la amplificación con metotrexato de los pools. La Figura 5C muestra que se incrementó marcadamente (en 8,4 y en 7,7 veces) la expresión de dos de los pools derivados de vector dicistrónico (es decir, sitios SD de ras WT y de ras mutante), mientras que el pool generado por el vector convencional se incrementó sólo ligeramente (2,8 veces) al someter cada uno a Mtx 200 nM. Se obtuvo un incremento global de 9 veces utilizando el mejor vector dicistrónico (SD de ras WT) frente al vector convencional tras la amplificación. El crecimiento del pool amplificado de expresión más elevada en medio de producción rico en nutrientes rindió títulos de  $4,2 \mu\text{g/ml}$  de tPA.

Se demostró que la manipulación de la secuencia donadora de empalme altera la proporción entre mensaje procesado y mensaje de longitud completa y el número de colonias que se forman en medio selectivo. También se demostró que los vectores de expresión dicistrónicos generan clones que expresan niveles elevados de proteínas recombinantes. Inesperadamente, resultó posible aislar expresores elevados que presentaban el sitio donador de empalme eficiente de ras WT mediante selección para las células DHFR<sup>+</sup> a pesar de la eficiencia con la que el gen DHFR se procesa a partir de los precursores de ARN formados en estas células.

## Ejemplo 2

*Producción de TNFr-IgG utilizando vectores de expresión dicistrónicos*

5 Para demostrar la aplicabilidad general de este enfoque, se evaluó un segundo producto en el sistema de vector dicistrónico que contenía, como el ADN de interés, una inmunoadhesina (TNFr-IgG) capaz de unirse a factor de necrosis tumoral (TNF) (Ashkenazi *et al.*, Proc. Natl. Acad. Sci. USA 88:10535-10539, 1991). Los experimentos descritos en el Ejemplo 1 anteriormente se repitieron esencialmente excepto en que el gen de producto codificaba la inmunoadhesina TNFr-IgG. Se introdujeron plásmidos de ADN que contenían un ADNc de TNFr-IgG y (a) ras de WT, 10 ver la Figura 6 (SEC ID n° 2), (b) ras mutante, o (c) sitio donador de empalme no funcional ( $\Delta$ GT) se introdujeron en las células CHO.dp12 tal como se ha comentado en el Ejemplo 1. Ver la Figura 1C para una ilustración de los constructos de ADN.

15 Se descubrió que el número de colonias DHFR-positivas generadas por tres de estos vectores era similar al observado con los constructos de tPA. La expresión de TNFr-IgG también fue comparable a la observada con los constructos de tPA (Figura 7A). La amplificación de pools de dos de los constructos mostró un marcado incremento de la expresión de la inmunoadhesina (9,6 y 6,8 veces) (Figura 7B). Los mejores de estos pools amplificados expresaron 9,5  $\mu$ g/ml al cultivarlos en medio de producción rico en nutrientes.

20 De esta manera, nuevamente se demostró que los vectores de expresión dicistrónicos generan clones que expresan niveles elevados de proteínas recombinantes. Además, al contrario de lo esperado, se descubrió que el aislamiento de células huésped DHFR<sup>+</sup> de expresión de producto elevada resultaba posible utilizando un sitio donador de empalme eficiente (es decir, el sitio donador de empalme de ras WT).

## Ejemplo 3

*Producción de anticuerpo utilizando un vector de expresión dicistrónico*

25 Se evaluó la utilidad de este sistema para la expresión de anticuerpo mediante el ensayo de la producción de un anticuerpo dirigido contra IgE (Presta *et al.*, Journal of Immunology 151:2623-2632, 1993). Además, se sometió a ensayo la flexibilidad del sistema con respecto al inicio de la transcripción mediante la sustitución del promotor/intensificador de CMV presente en los vectores anteriores, con el promotor/intensificador derivado de la región temprana del virus SV40 (Griffin, B., Structure and Genomic Organization of SV40 and Polyoma Virus, en: J. Tooze [Editor] DNA Tumor Viruses, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). La cadena pesada del anticuerpo se 35 insertó más abajo de DHFR, tal como se ha descrito en los constructos anteriores de tPA y de TNFr-IgG. Además, se construyó una nueva secuencia de sitio donador de empalme (GAC:GTAAGT) en el vector que correspondía al sitio donador de empalme de consenso más estrechamente que los sitios donadores de empalme presentes en los vectores sometidos a ensayo en los Ejemplos 1 y 2. Se muestra el vector de expresión resultante en las Figuras 1D y 9.

40 Se descubrió que dicho vector producía menos colonias que los vectores anteriormente sometidos a ensayo, y produjo predominantemente un producto de ARN procesado. Se construyó un segundo vector para que la cadena ligera del anticuerpo estuviese bajo el control del promotor/intensificador de SV40 y poli-A, y el gen de resistencia a la higromicina B bajo el control del promotor/intensificador de CMV y poli-A de SV40. Estos vectores se linearizaron en sitios HpaI únicos más abajo de la señal poli-A, se mezclaron en una proporción de vector de cadena ligera a 45 vector de cadena pesada de 10:3 y se electroporaron en células CHO utilizando un protocolo optimizado (tal como se comenta en los Ejemplos 1 y 2).

50 La Figura 11 muestra los niveles de anticuerpo expresados por clones y pools tras la selección en higromicina B seguido de la selección para la expresión de DHFR. La totalidad de los 20 clones analizados expresó niveles elevados de anticuerpo al cultivarlos en medio rico en nutrientes y presentó diferencias entre ellos de sólo un factor de cuatro. Se generó un pool de clones productores de anticuerpo y se sometió a ensayo poco después de su establecimiento. Se cultivó el pool continuamente durante 6 semanas sin reducción significativa de la productividad, demostrando que su estabilidad era suficiente para generar cantidades de gramos de proteína en su cultivo a gran escala.

55 El pool se sometió a amplificación con metotrexato a 200 nM y a 1  $\mu$ M y alcanzó un incremento de más de 2 veces del título de anticuerpo. El pool resistente a Mtx 1  $\mu$ M alcanzó un título de 41 mg/l al cultivarlo bajo condiciones óptimas en cultivo en suspensión.

60 Se examinó la estructura del anticuerpo expresado. Las proteínas expresadas por el pool resistente a metotrexato 200 nM y por un clon de expresión bien caracterizado generado por vectores convencionales (Presta *et al.*, *supra*, 1993) se marcaron metabólicamente con S<sup>35</sup> cisteína y S<sup>35</sup> metionina. En particular, se marcaron metabólicamente placas celulares confluyentes de 35 mm con 50  $\mu$ Ci cada una de S-35-metionina y de S-35-cisteína (Amersham) en suero libre de cisteína y F12:DMEM libre de metionina. Tras una hora, se añadió medio de producción rico en nutrientes y se permitió que las proteínas marcadas se “marcasen” en el medio durante seis horas adicionales. Las 65 proteínas se corrieron en SDS/PAGE al 12% (NOVEX) no reducido o tras su reducción con B-mercaptoetanol. Los geles secos se expusieron a película durante 16 horas. También se marcaron células CHO de control.

## ES 2 282 994 T3

La mayoría de la proteína anticuerpo se segrega con un peso molecular de aproximadamente 155 kilodaltons, consistente con una molécula de anticuerpo apropiadamente ligada con disulfuro con 2 cadenas ligeras y 2 cadenas pesadas. Tras la reducción, el peso molecular se desplaza a 2 proteínas de abundancia aproximadamente igual de 22,5 y 55 kilodaltons. La proteína generada de este pool no puede distinguirse del anticuerpo producido por el clon de expresión bien caracterizado, sin incremento aparente de cadena pesada o ligera libre expresada por el pool.

### Conclusión

El sistema de expresión eficiente descrito en la presente invención utiliza vectores que consisten de elementos de promotor/intensificador seguidos de un intrón que contenía la secuencia codificante de marcador seleccionable, seguido del ADNc de interés y de una señal de poliadenilación.

Se sometieron a ensayo varias secuencias de sitio donador de empalme para su efecto sobre el número de colonias y la expresión del ADNc de interés. Se utilizó un sitio donador de empalme no funcional, sitios donadores de empalme que se encuentran en un intrón entre los exones 3 y 4 de las formas mutante (ras mutante) y normal (ras WT) del gen Ras de Harvey y otro sitio SD eficiente (ver el Ejemplo 3). Los vectores se diseñaron para dirigir la expresión de transcritos primarios dicistrónicos. Dentro de una célula transfectada algunos de los transcritos conservan la longitud completa, mientras que el resto son procesados extrayendo la secuencia codificante de la DHFR. Cuando el sitio donador de empalme se debilita o resulta destruido, se observa un incremento del número de colonias.

Los niveles de expresión mostraron el patrón inverso, generando los sitios donadores de empalme más eficientes los niveles más elevados de tPA, de inmunoadhesina TNF4 o de VH de anti-IgE.

La homogeneidad de la expresión de los clones generados por los vectores de DHFR-intrón sitio donador de empalme ras se comparó con clones generados a partir de un vector convencional con un promotor/intensificador y una señal de poliadenilación separadas para cada DHFR y tPA. El vector de intrón DHFR da lugar a colonias que son mucho más homogéneas con respecto a la expresión que aquéllas generadas por el vector convencional. Los clones no expresantes derivados del vector convencional pueden ser el resultado de roturas en el dominio de tPA o de TNFr-IgG del plásmido durante la integración en el genoma, o el resultado de la metilación de elementos promotores (Busslinger *et al.*, Cell 34:197-206, 1983; Watt *et al.*, Genes and Development 2:1136-1143, 1988) de control de la expresión de tPA o de TNFr-IgG. El silenciamiento de un promotor mediante metilación, o roturas en los vectores de DHFR-intrón muy probablemente los incapacitaría para proporcionar un fenotipo DHFR-positivo.

Se descubrió que los pools generados por los vectores DHFR-intrón podían amplificarse en metotrexato y que incrementan la expresión en un factor de 8,4 (tPA) o de 9,8 (TNFr-IgG). Los pools de los vectores convencionales se incrementaron sólo en 2,8 y en 3,0 veces para tPA y TNFr-IgG al amplificarlos de manera similar. Los pools amplificados resultaron en niveles de tPA 9 veces más elevados y niveles de TNFr-IgG 15 veces más elevados en comparación con los pools amplificados con el vector convencional.

Sin limitarse a ninguna teoría en particular, el incremento de expresión de los pools resistentes a metotrexato derivados de los vectores dicistrónicos probablemente se debe al ligamiento transcripcional de DHFR y producto; cuando las células se seleccionan para la expresión incrementada de DHFR, sobreexpresan consistentemente el producto. Los enfoques convencionales carecen de ligamiento de expresión de marcador seleccionable y ADNc y, por lo tanto, la amplificación con metotrexato frecuentemente da lugar a la sobreexpresión de DHFR sin incremento concomitante de la expresión de producto.

Se obtuvo un incremento adicional de 4 y 6,3 veces al transferir pools amplificados de tPA y de TNFr-IgG desde los medios utilizados para las selecciones y las amplificaciones a un medio de producción rico en nutrientes.

En el Ejemplo 3, el vector de expresión presentaba un sitio donador de empalme de correspondencia más estrecha con la secuencia donadora de empalme de consenso y presentaba inserción corriente arriba de la cadena pesada de un anticuerpo anti-IgE humanizado. Este vector se linearizó y se co-electroporó con un segundo vector linearizado que expresaba el gen de resistencia a la higromicina y la cadena ligera del anticuerpo cada uno bajo el control de su propio promotor/intensificador y señales de poli-A. Se utilizó un exceso de vector de expresión de cadena ligera sobre el vector de expresión dicistrónico de cadena pesada para inclinar la expresión en favor de la cadena ligera. Se generaron clones y un pool tras las selecciones con higromicina B y con DHFR. Se descubrió que los clones expresaban niveles elevados relativamente consistentes de anticuerpo, al igual que el pool. El pool de 1  $\mu$ M alcanzó un título de 41 mg/l al cultivarlo bajo condiciones óptimas en cultivo en suspensión.

Se evaluó el anticuerpo anti-IgE mediante marcaje metabólico seguido de SDS/PAGE bajo condiciones reductoras y no reductoras y se descubrió que no podía distinguirse de la proteína expresada por una línea celular clonal altamente caracterizada. Es de particular importancia el resultado de que no se observase ninguna cadena ligera libre en el pool respecto al clon.

Se ha desarrollado un sistema de expresión estable para las células CHO que produce niveles elevados de proteínas recombinantes rápidamente y con menos esfuerzo que el requerido por otros sistemas de expresión. El sistema de vector genera clones estables que expresan niveles consistentemente elevados, reduciendo de esta manera el número de clones que deben cribarse para obtenerse una línea clonal altamente productiva. Alternativamente, se han

## ES 2 282 994 T3

utilizado pools para generar convenientemente niveles moderados a elevados de proteína. Este enfoque puede resultar particularmente útil cuando debe expresarse y compararse varias proteínas relacionadas.

5 Sin limitarse a esta teoría, resulta posible que los vectores que presenten sitios donadores de empalme muy eficientes generen clones muy productivos debido a que muy poco transcrito queda sin procesar, de manera que sólo los sucesos de integración que conducen a la generación de niveles elevados de ARN producen suficiente proteína DHFR para dar lugar a colonias en medio selectivo. El nivel elemento de mensaje procesado de estos clones seguidamente se traduce en cantidades abundantes de la proteína de interés. Los pools de clones preparados concurrentemente mediante la introducción de vectores convencionales expresaron niveles inferiores de proteína y eran inestables con respecto a la expresión de largo plazo, y la expresión no pudo incrementarse apreciablemente al someter las células a amplificación con metotrexato.

15 El sistema desarrollado en la presente invención es versátil en que permite generar con rapidez niveles elevados de polipéptidos de una o múltiples subunidades a partir de clones o pools de transfectantes estables. Este sistema de expresión combina las ventajas de los sistemas de expresión transitorios (generación rápida y no laboriosa de cantidades investigativas de proteína), con el desarrollo simultáneo de líneas celulares de producción estables altamente productivas.

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

- 5 1. Constructo de ADN que comprende un sitio de inicio transcripcional, un sitio de terminación transcripcional, un gen seleccionable, un gen de producto proporcionado en orientación 3' respecto al gen seleccionable, una región reguladora transcripcional que regula la transcripción tanto del gen seleccionable como del gen de producto, estando situado el gen seleccionable dentro de un intrón que presenta un sitio donador de empalme en orientación 5' respecto al intrón, en el que el sitio de la secuencia donadora de empalme regula la expresión del gen de producto utilizando la región reguladora transcripcional y el gen seleccionable es un gen amplificable.
- 10 2. Constructo de ADN según la reivindicación 1, en el que el sitio donador de empalme comprende una secuencia donadora de empalme de consenso.
- 15 3. Constructo de ADN según la reivindicación 1, en el que el sitio donador de empalme comprende la secuencia GACGTAAGT.
4. Constructo de ADN según cualquiera de las reivindicaciones 1 a 3, en el que el gen amplificable es DHFR.
- 20 5. Constructo de ADN según la reivindicación 1, en el la región reguladora transcripcional comprende un promotor y un intensificador.
6. Vector que comprende el constructo de ADN según cualquiera de las reivindicaciones anteriores.
- 25 7. Vector según la reivindicación 6 que es capaz de replicación en un huésped eucariótico.
8. Célula huésped eucariótica que comprende el vector según la reivindicación 6 ó 7.
9. Célula huésped eucariótica que comprende el constructo de ADN según cualquiera de las reivindicaciones 1 a 5.
- 30 10. Célula huésped eucariótica que comprende el constructo de ADN según cualquiera de las reivindicaciones 1 a 5 integrado en un cromosoma de la célula huésped.
11. Célula huésped según la reivindicación 10 que es una célula de mamífero.
- 35 12. Procedimiento para producir un producto de interés, que comprende cultivar la célula huésped según cualquiera de las reivindicaciones 8 a 11 de manera que expresen el gen de producto, y recuperar el producto a partir del cultivo de la célula huésped.
- 40 13. Procedimiento según la reivindicación 13, que comprende recuperar el producto del medio de cultivo.
14. Procedimiento para producir un producto de interés que comprende cultivar la célula huésped según cualquiera de las reivindicaciones 8 a 11, de manera que expresen el gen de producto en un medio selectivo que comprende un agente de amplificación durante tiempo suficiente para permitir que se produzca la amplificación, y recuperar el producto.
- 45 15. Procedimiento para producir células eucarióticas que presentan múltiples copias de un gen de producto que comprende transformar células eucarióticas con el constructo de ADN según cualquiera de las reivindicaciones 1 a 5, cultivar las células en un medio selectivo que comprende un agente de amplificación durante un tiempo suficiente para que se produzca la amplificación, y seleccionar las células que presentan múltiples copias del gen de producto.
- 50 16. Procedimiento según la reivindicación 15, en el que el constructo de ADN se introduce en las células eucarióticas mediante electroporación.
- 55 17. Procedimiento según la reivindicación 15, que comprende además cultivar las células seleccionadas de manera que expresen el gen de producto, y recuperar de las células seleccionadas el producto de interés.
- 60
- 65

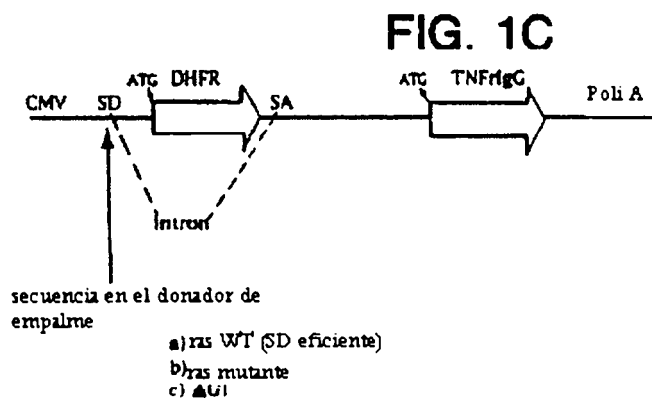
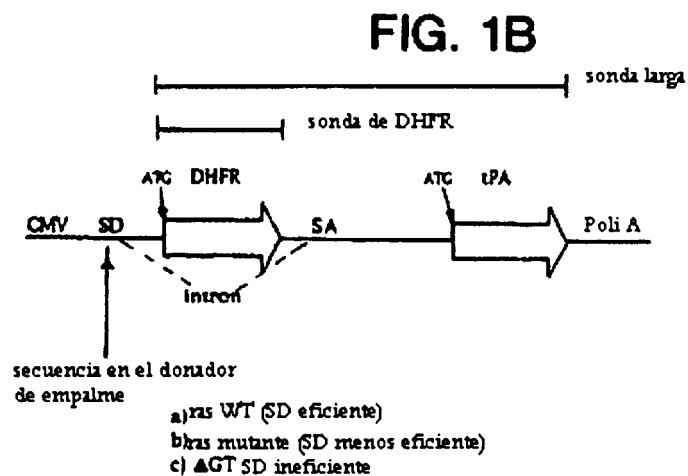
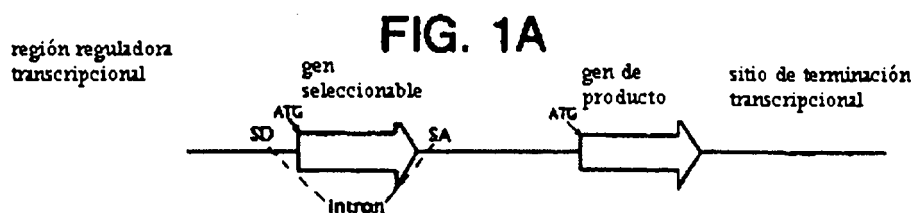


FIG. 1D

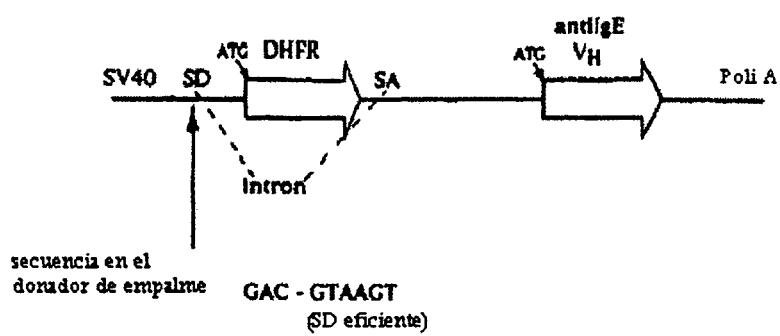


FIG. 2

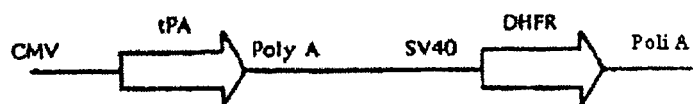


FIG. 3A

```
alul  
setI  
sacI  
hgiI/III  
hgiAI/ampl  
ecII/III  
bspI/II  
bsiEAI  
bmyI  
benII  
tqiI  
1. TTGAGCTCG CCOCAGATG ATTATTCAGT AGTATATTAAT AGTAATCAAT TAGOOGITCA ITAGITCATA GCCCATATAT GCCATATAT GCCATTCOCG CTATACATAC  
AACTCGAGC GGCCTGTAC TAATACCTGA TCATATATTA TCATTAATTA ATGOCOCAGT ATTCAGATAT CGCGTATATA CCTCAGGCGC CAATGTATTG  
rmaI tru9I  
mseI mseI  
speI -aseI/ascI/vspI  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000
```

FIG. 3B

```

          aaelt
          hmlt/eyt
          abalt/bsalt
          aattt
          plel
          acit hmlt
          bsalt
          rsl
          csp6t
          401 GGTTTGGCA GTACATCAAT GGGCTGGAG AGCGGTTTGA CTCAGGGGA TTTCAGATG TCACCGCAT TCCACCCAT TGAGTTCAT GGGGTTTTT TTTGGACGA
          CCMAACCT CATGTAGTTA CCCGACCTA TCGCCAACT GATGCCCT GATGCCCT AAAGTTTCCG AGCTGGGTA ACTCGAGTTA CCTCMAACA AACCCGTTGT
          elul
          actt
          actt
          hgtt
          hgtt/aspht
          ecilt6tt
          bsilt8t
          bsilt8at
          bmyt
          bmyt
          rsl
          csp6t
          mllt
          501 AAATCAAG GACTTCCAA ATGTCTAA CAKTTCCGC CCATGACGC AATGGGCG TAGCGGTA CGCTGGGAG TCTATATAG CAGACTGCT
          TTTAGTTCC CTGMAAGCTT TTACAGATT GTTGGGGCG GTTACTGGC GTTACTGGC TTACTGGC ATCCCGCAT GCGACCTCC AGATATATT CTCTGAGCA
          baeltt,pelt
          mctt
          eegt/xmalt/ecilt
          eact
          cirt
          frot8t
          actt
          thal
          frotlt/mvt
          eaptt
          scrtt
          sval bsalt
          eortt
          dtav
          bst8t hmlt/eyt
          apyt{dcm}
          sault{gsuf/bpm}
          mbot/ndelt{dam-}
          dnt{dam} hgtt {oh}
          dnt{dam-} abalt/bsalt
          601 TTAGTGACC GTGATGTC CTGGAGCC CATCCAGCT GTTTGACT CTATAGATA CAGCGGACC GATCCAGCT CCGGCGCC GAGCGGTCGA
          AATCACTGG CAGTCTAGC GACTCTGG GTAGTTGGG CAATCTGGA GTTATTTCT GTGGCCTGG CTAGTTGGA GCGCGGCC CTTCGALGT
  
```

FIG. 3C

```

tfii
aciI
thai b1efI
fnuDI1/mvnl
bstVI
bsh12362
701 TTGGAAAGCCG GATTCGCCCT GCGAAGAGTG CTTGAAAGTAC CCGCTATAGA GCGAATAGAG GATTTTATCC CCGCTGCCAT CATGCTTCSA CCGATTGACT
AACCTTGGCC CTAAAGGCA CCGTTCTCAC CCGTTTACT CCGATATCT CCGTATCTC CTAAATAGG GCGCAGCGTA GTACCAAGCT GGTAACTTGA
thai
fnuDI1/mvnl
bstVI
bsh12362
801 sfANI
pfIMI
bali
bmaAI
bmaI
bali
901 GCACTGTCC CCGTCCGAA AATATGGGA TTGCGAAGA TTGCGAAGTA CCGCAACTA CCGTCCCTC CCGTCCGTA CCGTCCGTA CCGTCCGTA CCGTCCGTA CCGTCCGTA
CGTCCGAGG GCGACGGGT TTATGCCCT AACCTTCTT AACCTTCTT AACCTTCTT AACCTTCTT AACCTTCTT AACCTTCTT AACCTTCTT AACCTTCTT AACCTTCTT
scfII
mvaI
ecobII
dcaV
bstNI
apyl(dcm+)
szAI
901 AACCTCTTCA GTGGAGGTA AACAGATCT GCTGATATG CCGTCCGAAA CCGTCCCTC CCGTCCGTA CCGTCCGTA CCGTCCGTA CCGTCCGTA CCGTCCGTA CCGTCCGTA
TTGGAGAGT CACCTTCAT TTGCTTTAGA CCACTATATC CCACTCTTTT GCGACAGAG GTAGAGACTC TTCTTAGTC GAAATTTCT GCTTATATA
almI
sstI
sacI
hplI1
hplI1/ampII
ocl136II
bsp1286
bsiMDAI
bmyI
banI
1001 ATAGTTTCA GTAGAGACT CAAGAGACA CCGAAGAGG CCGATTTCT TCCCAAGAT TTGAGATG CCGTCCGTA CCGTCCGTA CCGTCCGTA CCGTCCGTA CCGTCCGTA
TATCAAGAT CATCTTTGA GTTCTTTGT GGTCTCTCC CCGTAAAGA ACGTTTCA AACCTTAC CCGATTTCA ATACTTGT GBCCTTACC

```

FIG. 3D

```

                                     baeIII/psII
                                     haeI
accI nlaIII                               nlaIII          sauIAT
sacPI srfPI                               sbol/sbolII(dam-)
nvaI                                         qmI(dam-)
ecorII  ecorII                             sbol/sbolII(dam-)
dsav  tfII  dsav  pleI                       qmI(dam-)
bstXI  nlaIII  bstXI  ddeI                   qmII(dam-)
apyl(dam+)  binfI  apyl(dam+)  binfI  maeIII  aleI(dam-)
1101 CAAGTAAAGT AGACATGCTT TGGATAGTGG GAGCGGATTC TGTATACCG GAGCCAGCA ATCAACAGG CCACCTTAGA CTCCTTGTCA CAGGACACT
GTTCAFTCA TCTGTACCA ACCATATGCG CTCGCTCAG ACAATATGTC CTCGCTACT TACTGTTC GGTGAATCT GAGAAATCT GTTCTTAGTA
                                     mnlI
                                     binII/acyI  sacPI
                                     shaII/baeII  mvaI
sacPI                                         ecorII
nvaI  eccoII  sau96I
dsav  dsav  mvaII
bstXI  ball  esuI  mnlI
apyl(dam+)  mnlI  bstXI
hsaJI  hgaI  ddeI  apyl(dam+)
1201 GCAGGATTT GAAAGTGACA GTTTTCCC AGAATATGAT TTGGGAAAT ATAACTCTT CCCAGATATC CCAGGCTCC TCTGTAGG7 CCAGAGGAA
CTGCCTTAA CTTCACCTT GCAAAGGCG TCTTACTA AACCTTTA TATTGAGA GCGTCTAG GCTGGCAGG AGAACTTCA GCTCTCTT
                                     mnlI
                                     binII/acyI  sacPI
                                     shaII/baeII  mvaI
sacPI                                         ecorII
nvaI  eccoII  sau96I
dsav  dsav  mvaII
bstXI  ball  esuI  mnlI
apyl(dam+)  mnlI  bstXI
hsaJI  hgaI  ddeI  apyl(dam+)
1301 AAGGCATCA AGTATAGTT TGAAGTCTAC GAGAGGAAAG ACTAAGGAA AGATCTTTC CTCCTCTT AAAGCTATCC ATTTTATTA
TTTGGTAGT TCAATATCA ATTCTAGTTC CTCCTCTT TCACTTCTT TCACTAAG TCAAGGAC GAGGAGGA TTTCATAC TAAATATAT
                                     fnu4HI
                                     eciI
nlaIII                                         tbaI
styI  fnuDI/mni  krVI
ncoI  bstXI  mscI
dsai  styI  bstXI
bsaJI  bstXI  behI26I  aaeI/asmI/vspI
1401 GACCAAGGA CTTTGTGCG CTTAGACC CTTGTGCTTC CTTAGACC GCTTCACTT ATACTATC CTTATATTC ATACAGAG ATTAGTGA
CTGTACTCT GAAAGGACC GAATCTGCG GAATCTGCG CAATCTGCG CAATCTGCG TATATATTC GAATACAG TATGTATC TAAATCTCT

```



FIG. 3F

```

bpmI      sau96I      haeIII/palI      haeIII/palI
nlaIV     haeIII/palI      asuI   rraI      sruIII/mvnI
bglCI    asuI   rraI      ecoRII     dcaV      bstUI sau96I
bapI286  eco0109I/draII      aluNI  csp6I ddeI      hmlI/bsaHI
bmyI     aluNI  csp6I ddeI      hmlI/bsaHI      hmlI/bsaHI
1901 GGCGGACCT GCGAGGCGC CCTGACTTC TGAGATTTC TGTCCGATC GCGGAGGCA TTGCTGGGA AGTCTGTGA AATGATGAC AGGCGTAGT
    CCCCCTGGA CCGTTCCTC CGGCATGAG AGTCTMAGC ACAGCGTAC GCGGCTTCT AAGGACCT TACGACACT TATCTATAG TCCCGTCCA

scrFI     hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
mvaI     hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
ecoRII   hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
dcaV     hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
bstNI    hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
bsaJI    hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
sau96I   hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
avaII    hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
asul     hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
mnlI     hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
2001 GCTACAGGA CAGGCGGC ACCTACGCG GCGGCTGAG CAGGCGGC CAGGCGGC AGTCTGCGA CTGAGAGAG AGGCGTGG CCGAGAGCC
    CGATCTCT GCTTCCCTG TCGATCTCC CCGGCGGC CCGGCGGC CCGGCGGC TCGCTGCT GACTTCTG TCCCGGACC GCGTCTTGG

scrFI     hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
mvaI     hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
ecoRII   hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
dcaV     hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
bstNI    hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
bsaJI    hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
sau96I   hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
avaII    hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
asul     hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
mnlI     hmlI/bsaHI      hmlI/bsaHI      hmlI/bsaHI
2101 CTACAGGCG GCGGCGGC AGGCGTGG AGGCGTGG AGGCGTGG AGGCGTGG AGGCGTGG AGGCGTGG AGGCGTGG AGGCGTGG
    GACTTCTG TCCCGGACC CCGGCGGC CCGGCGGC CCGGCGGC TCGCTGCT GACTTCTG TCCCGGACC GCGTCTTGG

```





**FIG. 3I**

```

      wpi      scfi      pvu11      aluI      hgiJII      haeIII/palI
      hpaII    ptaI    fnu4HI    astI      hgiNI/aspHI
      scrFI    bepAI    bovi    sacI      ecII36FI
      nciI    sau96I    nsp8II    bsp1286
      dsav    avuII    fnu4HI    bsrI    bmlNKAI
      caulI    asuI    bbwI    mspi    bmyI
      drafiI    bslI    beeJI    aciI    bsgI    aluI    hpaII    hpaII    nlaIIII    hmaAI
      2 901  GCACATGATG  CCTTCCCGG  GCGGACCTG  AGCTCCCGA  CTGAGCGAG  TGTGACTG  GCGCTACG  CAGAGCTAG  GCTTGTCTC  CTTTCATTC
      CCTGACAC  GGAGGGGG  CCGCTGAG  TGAGCGCT  GACTGTCTC  AACTTGAA  GCGGATTC  GTTGACTC  CGAGAGAG  GAAGATAG

      Enu4HI    fnu4HI
      aciI    nsp8II
      fnd4II    aciI
      rslI    fonI    fnu4EII    trv9I    hphI    alaIII
      bsrBI    eco57I    mnlI    nlaIII    csp6I    aluNI    bowI    msaI    maeIII    nspHI    sauI/bpaI
      3 001  GGAGCGCTG  AAGGAGCTC  ATGTCAGACT  GTACCGATCC  ACCCGCTGA  CATGCAACA  TTTACTTAC  MGAAGCTCA  CGAGACTC  GCTGTGCTT
      CCTGCGAC  TTCTCCGAG  TACAGTCTA  CATGGTGGG  TGCGACTG  GTAGTGTGT  AATGAAATG  TCTGTGAGT  GCTGTGTTA  GCGACACA

      sau96I    nla2V    haeIII/palI    scrFI    mvaI
      asuI    nlaIV    scrFI    ecorII    dsav
      hgiJII    hgiJII    hgiJII    hgiJII    hgiJII    hgiJII    hgiJII    hgiJII    hgiJII    hgiJII    hgiJII    hgiJII    hgiJII    hgiJII
      3 101  GGAGCAGCT  GAGCGGGG  GCGCCAGCG  AACTTCAGC  AACTTCAGC  AACTTCAGC  AACTTCAGC  AACTTCAGC  AACTTCAGC  AACTTCAGC  AACTTCAGC  AACTTCAGC  AACTTCAGC  AACTTCAGC
      CCTGTCTAG  CTTCCCGG  CCGGATTC  TTGACTGTC  TGCGACTG  CCGCTGAG  CCGCTGAG  CCGCTGAG  CCGCTGAG  CCGCTGAG  CCGCTGAG  CCGCTGAG  CCGCTGAG  CCGCTGAG
  
```



FIG. 3K

```

styI
aciI
fnu4HI sau96I
bglI nlaIII
sfii ncoI hseIII/palI
hseIII/palI
caeI dsal asuI
cfrI bseJI fnu4HI aluI
fnu4HI bsvI
360I GATGCGCC ATGCGCCAC TGTATTATG CAGCTTATTA TGGTTAGAAA TAAGCAATA GCAATCACAA TTTCACAAAT AMBCAATTT TTTCACCTCA
CTACCGCCG TACCGCTTG AACAAATAC GTCAATATT ACCAATGTTT ATTGCTAT CTATGTTTTT AAAGTGTTA TTTCCTTAAA AAAGTACGT
seu3AI
aboI/ndaII(dam-)
dpmI(dam+)
dpmI(dam-)
pvuII/bspCI
mcrI
taqI(dam-) tcrUI
clal/bspI06(dam-)
sau3AI seeI styI fnu4HI heeI
aboI/ndaII(dam-) bbvI ncoI
dpmI(dam+) xmiI hinPI dsal haeIII/palI
dpmI(dam-) asuI/asnI/vspI bseJI
nlaIII alvi(dam-) asp700 hbaI/cfoI nlaIII molI
370I TTCTAGTTGT CCTTTCTCA AACTCATCA TGTATCTTAT CATCTCTGGA TCCATGCGA ATTAAATGCG CCGACCACTA TGGCTGTAAA TAACCTCTGA
AAGATCAACA CCAACAGGT TTGATGATG TCAATAGATA GTACAGACTT AGCTAGCCCT TAATTAGCC GCTCTGCTGT ACCGACTT ATTGAGACT
kmaI
seeI
380I AAGAGAACT TGGTTAGGTA CCTTCTGAGG CCGAAGAAC CAGCTGTGGA ATGTCTTCA GTTAGGTTGT GGAAATCTCC CAGCTCTCCC AGCAAGCAGA
TTCTCTTCA ACCAATCCAT CGAAGACTCC CCTTCTTCTG GTCCAGACTT TACAGACTT CAATCCACA CCTTCTCAGG GTCGAGAGCG TCTCTCTCT
tsal
csp6I
nlaIV
kpnI
hgICI
banI
asp718 mnlI
acc65I ddeI aciI
mnlI
380I AAGAGAACT TGGTTAGGTA CCTTCTGAGG CCGAAGAAC CAGCTGTGGA ATGTCTTCA GTTAGGTTGT GGAAATCTCC CAGCTCTCCC AGCAAGCAGA
TTCTCTTCA ACCAATCCAT CGAAGACTCC CCTTCTTCTG GTCCAGACTT TACAGACTT CAATCCACA CCTTCTCAGG GTCGAGAGCG TCTCTCTCT

```



FIG. 3M

4301 *ssu3AI* *mbol/ndeII(dam-)* *hincPI* *fnu4HI*  
*ssu96I* *hhaI/cfoI* *hincPI* *theI*  
*haeIII/palI* *ssuI* *dpnI(dam-)* *hincPI* *theI*  
*ssuI* *dpnII(dam-)* *hincPI* *theI*  
*mnlI* *acif* *pruI/bspCI* *hincPI* *theI*  
*ssuI/bsp32I* *acri* *hincPI* *theI*  
 4301 *CGAAGAGCC* *CCCAACGATC* *GCCCTGCCA* *ACAGTTCCT* *AGCTTGAATG* *GCATATGGC* *CGTATGGC* *TATTTCCTC* *TTACCCATCT* *GTCCCAATT*  
*CGTTCCTCG* *CGTGGCTAG* *CGGGAAGGT* *TGTCAGCCA* *TCCGACTTAC* *CCCTACGCC* *GCATACGCC* *ATTAAGAGG* *AATGCCTAGA* *CACGCCATA*  
  
 4401 *acii* *maeII* *hincPI* *theI* *fnu4HI* *hincPI*  
*TCACACCCA* *TACCTCAAG* *CAGCATAGT* *ACGCGCCCTG* *TACCGCCCA* *TTAAGCCCG* *CGGCTGTGCT* *GTATACGCC* *AGCTGACCG* *CTACACTTC*  
*AGTGTGCTT* *ATGCACCTT* *CTTGATATCA* *TCCCGCGAC* *ATCCGCTCT* *ATTTCGCTC* *GCACACCA* *GCATATGGC* *TCCACTGCG* *CATCTAGAC*  
  
 450: *trai* *hhaI/cfoI* *fnu4HI* *tru9I* *acii* *hincPI* *theI* *fnu4HI*  
*hhaI/cfoI* *hincPI* *theI* *fnu4HI* *hincPI* *theI* *fnu4HI*  
*hincPI* *hseII* *hhaI/cfoI* *hincPI* *theI* *fnu4HI* *hincPI*  
*hhaI/cfoI* *hincPI* *theI* *fnu4HI* *hincPI* *theI* *fnu4HI*  
*hhaI/cfoI* *hincPI* *theI* *fnu4HI* *hincPI* *theI* *fnu4HI*  
 450: *CAGCCCTA* *CGCCCGCTC* *CTTTCCTT* *CTTCCCTTC* *TTTCTGCTC* *CTTCCCTTC* *TTTCTGCTC* *CTTCCCTTC* *CTTCCCTTC* *CTTCCCTTC*  
*CTCCCGAT* *CTCCCGCTG* *GAAAGCCAA* *GAAAGCCAA* *GAAAGCCAA* *GAAAGCCAA* *GAAAGCCAA* *GAAAGCCAA* *GAAAGCCAA* *GAAAGCCAA*

4601 *TTCCGATTA* *GTCTTTACG* *GCACCTCAC* *CCCAAAAC* *TTGATTTGG* *TGATGTTCA* *CTTATGCGC* *CATCGCTCG* *ATAGCGCTT* *TTTCCGCTT*  
*AGGCTAAT* *CACAAATCC* *CGTGGAGCTG* *CGTTCCTTC* *ACTAATCC* *ACTACCAAT* *GCATACGCC* *GTATACGCC* *TATTCGCCA* *AAAGCCGCA*

4701 *ssuII* *pleI* *trui* *pleI*  
*didi* *hincPI* *maeII* *hincPI* *hincPI*  
 4701 *TGAGTTGCA* *GTCCAGCTC* *TTTAAATAG* *GACCTCTCT* *CAACATCA* *ACCTATCTC* *GCATATCT* *TTTAAATAG* *AGGCAATTT*  
*ACTGCACCT* *CAGTGCAG* *AAATATATC* *CTAAGACA* *CTTTCCTCT* *TGATATCT* *GCATATCT* *TTTAAATAG* *AGGCAATTT*

## FIG. 3N

```

4801 GCGCAATTCG GCGTATGCGT TAAAAATCA GCTGATTTAA CAAGAATTA ACCCAATTT TACKAATAA TTAACUTTTA CAATTTTAAQ GTCCACTTTC
CGCCTAAAGC CCGAANACA ATTTTTACT CCAATTAAT CTTTTAAK ATTGCTTAA ATGCTTTTAT ATTUCAAAT GTTAAATATC CACCTGAGAG
      kguII/aspHI
      bspI286
      balKRAI
      bmyI ddeI
      spali/snoI
      aI=4II/snoI
      haeIII/palI
      trsI
      msel
      haeIII/palI
      aluI
      sruI
      msel
      haeIII/palI
      aluI
      sruI
      msel
      haeIII/palI
      haeIII/palI
      msbII
      bpuAI
      lbaSI
      haeIII/palI
      sau96I
      aulI
      ecc01091/draII
      mliI
      haeIII/palI
      haeIII/scyI
      ahaII/bsaHI
      astII
      ddeI maeIII

4901 AGTAAATCT GCTGTATGC CCGTATGCA AGCCACTCC GCTATGCTA CCAATGCGT FCAAGCTGC GCGCGACAC CCGCGACAC
TCAUTTTAGA CCGACTAGC CCGATCAAT TGCTTGAGG CAGTACAT GTACTAGCC GACTAGCC CGCCTGCTG GCGCTGCTG
      hinPI
      hbaI/cfoI
      tbaI
      fnuDII/wmI
      bstOI
      bacOI
      nspIII bshI236I
      nspII
      hpaII
      dsav
      fnu4HI
      sfaHI
      nciI
      dsav foki
      cmaII acII
      TCCCCTGAC TCCCCTGAT CCGTTACGA CAGCTCTGA CAGCTCTCG GAGCTCTCG GAGCTCTCA ATGCTCTCA
      scrPI
      nciI
      nspI
      hpaII
      dsav
      esp3I
      fnu4HI
      bspI
      cmaII acII
      aluI
      balI
      caulI
      aluI
      nlaIII
      mliI
      hbaI
      hspI
      haeIII/palI
      msbII
      bpuAI
      lbaSI
      haeIII/palI
      sau96I
      aulI
      ecc01091/draII
      mliI
      haeIII/palI
      haeIII/scyI
      ahaII/bsaHI
      astII
      ddeI maeIII

5101 CCGCGAGGC AGRATTCTT AAGACAAAG GCGCTCTGA TACGCTAT TTTAAAGTT AANFCAITA TAAATAGT TTTCTGAGG TCAGTGGCA
CGCCTCTCC CCAACAGC AGCCCTGTA CCGGCACT ATCGGATTA AATATCCA ITTCAGTACT ATATTAACA AAGAACTGC AGTCCACGT
      hinPI
      hbaI/cfoI
      tbaI
      fnuDII/wmI
      bstOI
      bshI236I
      hpaII
      dsav
      fnu4HI
      sfaHI
      nciI
      dsav foki
      cmaII acII
      TCCCCTGAC TCCCCTGAT CCGTTACGA CAGCTCTGA CAGCTCTCG GAGCTCTCG GAGCTCTCA ATGCTCTCA
      scrPI
      nciI
      nspI
      hpaII
      dsav
      esp3I
      fnu4HI
      bspI
      cmaII acII
      aluI
      balI
      caulI
      aluI
      nlaIII
      mliI
      hbaI
      hspI
      haeIII/palI
      msbII
      bpuAI
      lbaSI
      haeIII/palI
      sau96I
      aulI
      ecc01091/draII
      mliI
      haeIII/palI
      haeIII/scyI
      ahaII/bsaHI
      astII
      ddeI maeIII

```

FIG. 30

```

nlaIV
aciI
chaI
fnuDII/mvni
bstOI
behI236I
hinPI
bhaI/cfoI

rcal
bspHI
berBI
bsaHI
aciI
nlaIII

5201 CTTTTCGGG AMATGTCG GAAAGCCCTA TTTTFTTART TTTCTAATA CATTCAMATA TGTATGCT CATAGCCGA TAAAGCTGAT AAATGCTTCA
GAAAGCCG TTTACACCG CTTGGGAT AAACAATAA AAGATTAT GTATATTAT ACATAGCGA GTACTCTT ATTGGACTA TTTAGGAT

5301 sspI          eakI/kp63I          hpiI/aspHI
ATAATATCA AAAAGGAGA GTATAGTAT TCACATTC CTTGCGCC TTATTCCTT TTTGCGCA TTTGCTTC GGTGTTTC TCACCGAAA
TATATPACT TTTTCTTCT CATACTATA AATTATAG GCACGCGG AATAGCGAA AAAGCCCT AAAGCGAAG CACAAAGC GTGGGTC

5401 hpiI          sfaI          mboI/dam-          hpiI/aspHI
ACCTGCTCA AAGTAAAGA TGTCTAGAT CAGTGGTG CACGATGG TTACATCAA CTGATCTCA ACAGCGTAA GATCTTGG AGTTTTCGC
TGGAGCACT TCAATTCT ACACCTCA CTCACCCAC GGTCTACC AATGTAGCT GACTAGCT TGTGCTT TGTGCTT CTAGACTE TCAAGCCG

5501 maeII          pspI406I          xmiI          asp700          mboI-          scrPI
CGGAGACG TTTTCAATC ATGAGCACT TTAAGTCT GCTATGCG CCGTATTAT CCGTATGA CCGCGCGAA GAGCACTCG GTCCCGCAT
GGCTTCTGC AAAGCTTAC TACTCTGAA AATTTCAGA CATTACACG CGCATATA GGCCTACT GCGCCCTT CTCTTGAGC CAGCGGTA

nlaIV          aciI          chaI          fnuDII/mvni          bstOI          behI236I          hinPI          hpiI/cfoI          raaI
aciI          nspi          hpaII          daav          hinI/acyI          hpaI          caulI          ahaII/bseHI          bsp6I          bafI
fnuHI          bviI          nlaIII          fobI          sfaMI          maeII          hpiI          sspI          ksp6I          bstI
fnuHI          bviI          nlaIII          fobI          sfaMI          maeII          hpiI          sspI          ksp6I          bstI
ACACTATCT CAGATGACT TGTTCAGTA CTCACGTC ACAGAAAG ATCTTAGCA TGCCTGCA GTAMGAAAT TATGCTGC TCCCTAAC
TGTATAGA GTCTTACT ACCACTAT GAGTGTG TGTCTTTC TAGAATGCT ACCACTCT CATTCTCA ATAGCTGAC AGGTATG

```



FIG. 3Q

```

trai          sau3AI
sau3AI hphi  nboI/ndeII(dam-)
nboI/ndeII(dam-)
dphi(dam+)  phiII(dam+)
dphiII(dam-) phiII(dam-)
          lru9I bctYI/rhoII alwI(dam-) alaIII      meeII
          meeI alwI(dam-) betYI/rhoII          rcaI      cru9I
          aheIII/draI meeI aheIII/draI meeI nboII(dam-) bsp8II meeI
6201  CTCATATA CTTCAGATTC ATTAAACT TCATTTTAA TTTAAAGCA TCTAAGTCA GATCTTTT GATPACTCA TACCCAAAT CCTTACCT
      GACTATAT GAATCTAC TAATTTCA AGTAAATTA AATTTCT AGCCACTT CTAGAAA CTATTAGAT ACTGTTTT GGGANTTCA

          sau3AI
          nboI/ndeII(dam-)
          phiII(dam-) sau3AI
          dphiII(dam-) nboI/ndeII(dam-) tbaY          sauBII/mnI
          bctYI/rhoII dphi(dam+)          betVI
          sau3AI alwI(dam-) dphiII(dam-)          behI236I
          nboI/ndeII(dam-) alwI(dam-)          hlnpY          fnu8II
          dphi(dam+) nboII(dam-)          betYI/rhoII          hhaI/cfoI          hbfI
6301  GATTTTCT TCCACTGAC CTCAGACCC CTCAGAAC TCAPANGTC TTTTNGAT SCITTTTC TCCCTATAT CTGCTCTTC CAACAAAA
      CTCAAAGCA AGCTACTG CAGTCTGCG CATCTTCT AGTTCTCTAG AAGACTTA GAAANAAG ACCCCACTA GACCCAGAC GTTCTTTT

          sau3AI
          nboI/ndeII(dam-)
          dphi(dam+)
          dphiII(dam-)
          alwI(dam-)
          mspI
          aciI          nsp8II          hpaII          aluI          bcrI          hinPI
6401  AACCAAGCT ACAGGCTG GTTCTTTC CCGATCAGA CTCAGACACT CTTTTCCA AGCTACTG CTCAGACA CCGCAAGAC CAATPACTT
      TCTGCGCA TCTCTGAC CAACAAAG CCTTACTT CAGTCTCA GAANAAGT TCCATTTCC GAAATGCT CCGCTCATC GTTTATGCA

          rnaI          bacIII/paII
          meeI          haeI
6501  CTTCTATG TACCTVAT TAGCCAGCA CTCAGACAC CTCCTGAC CCGTACTA CTCCTCTG CTAACTCT EMCATGTC TCTCTCAT
      GAAATCAC ATCCGATCA ATCCGATCT GAATTTCT AGACTCTG GCATATAT GAGCGAC GATTGACA ATGCTACC ACCAGCTCA

```



FIG. 3S

```

    thal
    (rudII/mvnl
    bstUI
    bah1261
    hinPI
    hhal/cfoI
    thal
    (rudII/mvnl
    bstUI haeIII/palI aluI
    bah1261 trc9I pvaII
    baII eelI cfiI aseI/asnI/vspI
    aciI cfrI hiniI meelI nspII
    7101 GTCAATGAC GAGCAGCG MAGACGCC ANTACAAA CCGCTCTCC
    CACGTCAGTC CTCCTTCCC TTTCGCGG TTTGCTTT GCGCGAGG
    GCGCGCAC CGCTTAGTA ATTAGTGA CCGCTCTC CAAGGCTG
    bstI
    acI
    7201 TCGAAGCG GCACTCAGC CAAGCANT ANTGCATT ACCTACTCA TTAGGCACC CAGGCTTAC ACTTATGCT TCGGCTCT ATTTGTGTG
    ACCTTTCCC CGTCACTGC GTTGCTTA TTACTACA TGAATGAT ATCCCTCG GTCCGAAATG TGAATAGA AGCCCGGCA TACAGCAC
    mspI hpaII
    hinPI haeIII
    hhal/cfoI
    aciI
    7301 GAATTCAG CGATACAA TTTCACAC GAAACGCTA TACCAAGAT TACCAATTA
    CTTACACTC CCTTATTGTT AAGTGTCT GTTGTGAT ACTGACTA ATGCTTAAT
    >length: 7360
  
```

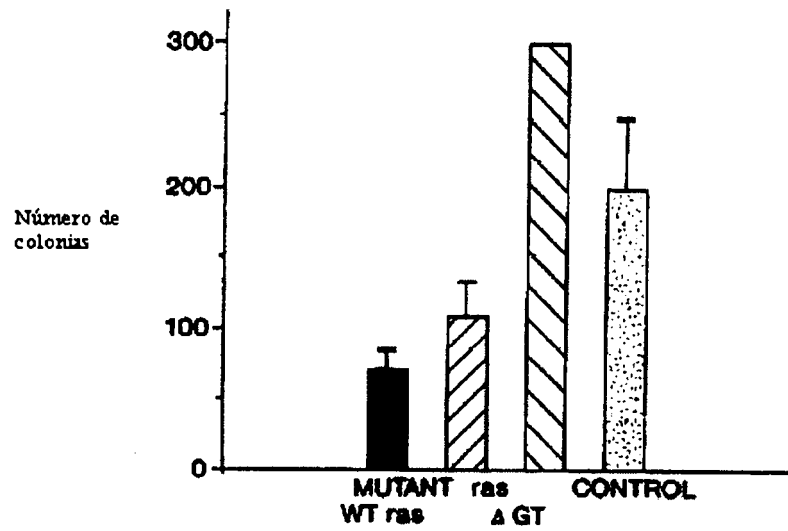


FIG. 4

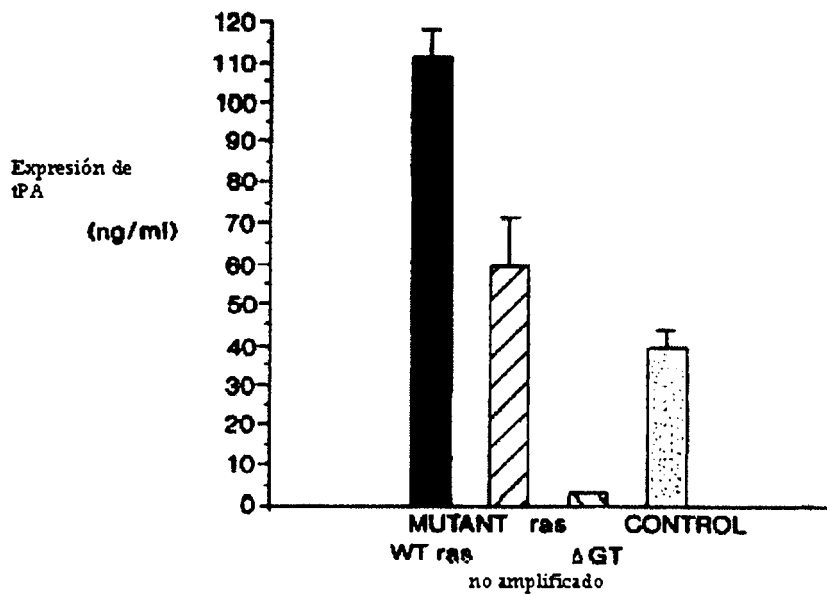


FIG. 5A

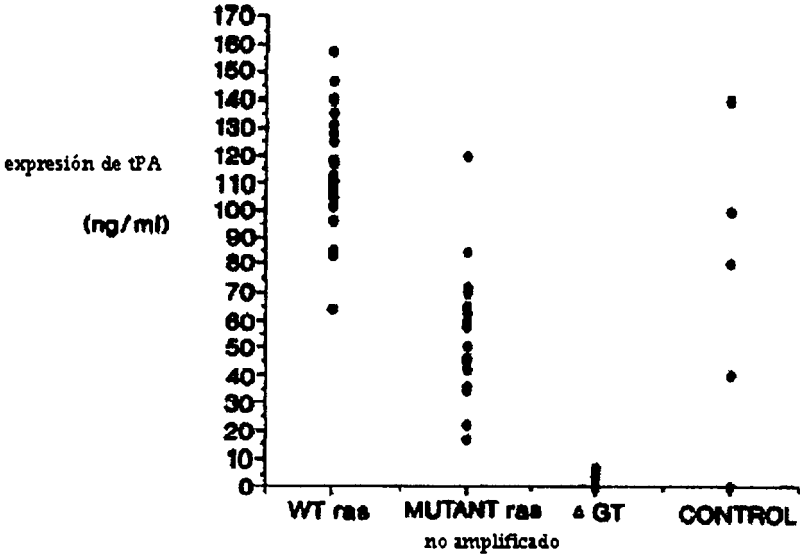


FIG. 5B

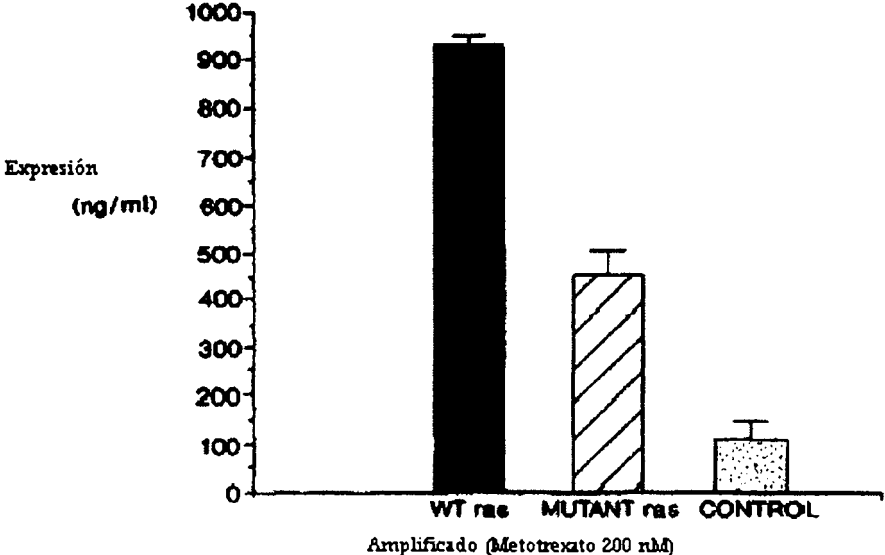


FIG. 5C

FIG. 6A

```

aluI
sacI
sacI
hgiIII
hgAI/aSPII
ecII/phiI
bspI236
bsIIRKAI
bmyI
banII
teqI
1 TTAGAGCTCG CCCGACATG ATTATTGACT AGTATTATAT ACTAATCAAT TACCGGTCA TTAGTCATA GCCCATATAT GCGATTGAT GCGATTGAC
AAGCTCGAGC GCGCTGTAC TAAATCACTA TCAATATTA TCAATATTA ATGCCCAT ATCACTATAT GCGTATATA CCTGAGGCG CAATGATTTG
scrFI
swaI
ecorII
dsvI
acII
bglI/bstNI
sau96I
heII/palI acII
asuI spyI(dcm+)
101 TTAGGTTAA TGGCGGCT GCGTACCG CCAGACCC CCGCCATG ACCTCAAAA TACGATAT TCCGATAT TCCGATATA ACCGCAATG GCGATTGCA
AATCCATTT ACCCGGGA CCGACTGCG GCTTCTGCG GCGGATAC TCAATATTA ACTGATATA AGGATATAT TCGGTTATC CCTGAAAGT
maeII
hinII/acyI
shaII/bstKI
satII
201 TTAGAGCTCA TCGGTGATG ATTTACGTA AACTGCGAC TTGCGATAC ATCAATGCTA TCAATGCA ACTACGCGC CTATTGAT CTATTGATG CAATGCGCT
ACTGCGCT ACCCACTCA TAAATCCAT TTGACGCTG ACCGCTATG TAGTTCACT AGTATAGCT TATTCGCGG GATTAATGCA GTTACTGCA
maeII
hinII/acyI
shaII/bstKI
satII
rsal
csp6I
bglI
mdeI
rsal
csp6I
maeII
hinII/acyI
shaII/bstKI
satII
301 AATGCGCG CCTGCAAT TCGCCATAC TCGCCATAT GCGACTTCC TACTTGGAG TACATGACG TATTAGTAT CCTATTATC CCTATTATC ATGCGATGC
TTTACCGGC GGCCTAAT ACGGTATG TACTGATATA CCTGAAAG ATTAACGCTC ATGATATG ATTAATGATA GCGATATG TACCCTACG
scrFI
swaI
ecorII
acII
bglI/dsvI
sau96I/bstNI
heII/palI
asuI spyI(dcm+)
baeI nlaIII
nlaIII
ncoI
dpsI hgiI acII
baeI sbaII
csp6I haeII
styI

```

FIG. 6B

```

rsai      rsaII      p1eI      p1eII      n1aIV      n1aIV
csp6I     csp6I     acII      acII      hpiCI     hpiCI
401 GCTTTTGGCA GTTACATCAAT GCGCTGCGAT AGCGTTTGA CTAAGGGGA TTCCAAZTC TCCACCGCAT TCACTCCAAAT GCGMTTGTGT TTTGGCAACA
CTAAACGCT CATGTAGTTA CCGCACTA TCGCCAACT GAGTCCCT AAGGTTTCAG AGGTGGGGTA ACTGCGTTTA CCTCAACA AACCGCTGT

                                     maeII     maeII     hgaI     hgaI     mnlI     mnlI
                                     acII      acII      csp6I     csp6I
                                     acII      acII      acII      acII
501 AATCAAGCC GACTTTGCA ATGTGCTTA CACTTCCG CCAITGAGC AATGCGCG TAGCGTCTA CCGTGGCAG TCTATATAG CAGAGCTGT
TTTAGTTGCC CTGAAGGTT TTACAGCTT GTTGAGGGG GGTACTCGG GTTACCGCC ATTCCGACAT CCGACCTCC AGATATATTC GTTCCAGCA

                                     maeII     acII      acII      hgaI     hgaI     mnlI     mnlI
                                     acII      acII      acII      acII      acII      acII
601 TTATCGAAC GTCAGATCG CTGAGAGCC CATTCCGCT GTTTGACTT CCAATAGCA CACCGAGC GATCCAGCT CCGCGCGCC GAGCGTCCA
AATCACTTGG CAGTCTAGCG GACTTCTCG GTAGCTGGA CAATCTTCT GTCCCTTGG CTAAGTCCG GCGCGCGCC GTTCCAGCT

```

FIG. 6C

```

t411
scii
thai hinfI
fnu011/mni
bst01
bsrl2161
701 TTGAAACCGG GATTCGCCGT GCGNAGTG CTGTANGTAC CCGCTATAGA GCGATAGAG GATTATAC CCGCTCCCA? CAGCTTGA GATTAGCT
AACCTTGGC CTAGCGCG?A CCGTCTCAC GACATATCT GCGATATCT GCGATATCT CTAATATAG GCGNCGGA GTACCAAGCT GGTAACTGA

fnu011/mni
bbvI
asp011
scii
taqI
mla111
701 TTGAAACCGG GATTCGCCGT GCGNAGTG CTGTANGTAC CCGCTATAGA GCGATAGAG GATTATAC CCGCTCCCA? CAGCTTGA GATTAGCT
AACCTTGGC CTAGCGCG?A CCGTCTCAC GACATATCT GCGATATCT GCGATATCT CTAATATAG GCGNCGGA GTACCAAGCT GGTAACTGA

thai
fnu011/mni
bst01
bsrl2161
mluI
berBI
ad1111
rsmI
scii
mmI
cepI1
mlI
dclI
asp700
acaI
801 GCACTCTGCG CGTGTCCCA AATATGGGA TTGCGAGAA GCGAGACTA CGTGGCCTC CACTCAGAA CCGCTCAGG TACTTCCAAA GAGTACGAC
CGTAGCGCG SCNAGGG?T TTATACCGT AACCTTCTT GCGTCTGAT GCGACGGGAG GCGATCTCTT GCGNAGTTG ATGAGGGTTT CTACTAGTG

scfI1
evai
ecolI1
dsav
t411
taqI
truI1
trvI1
mboI1
hinfI
alwI1
hphI
dclI
hinI1
abvI1
drrI
aseI1/asmI/vspI
mli
901 AACCTTGA CTGAGGGA MACAGACT? GCGATATG GGTAGGAAA CTTGCTTC CACTCTCAG AGAATCCAG CTTTAAAGA CAGATTAAT
TTGAGAGT CACTTGCAT TTGCTTGA CCGTATAC CCACTCTTT GCGAGAGAG GTAGAGCT TCTTACTG CAATTTCT CTCTTAATTA

alul
sstI
sacI
hgiI1
hglAI/espHI
ec11611
bsp1286
be100AI
bmyI
banI1
dclI
bclI
mli
1001 ATAGTTCTA GTAGAGACT CAAGAGCA CCAAGGCG CCACTTCTT TCGCAAGT TTGAGAGG CTTAGAGG CTTAGAGG CTTAGAGG CCGAATTCG
TATCAGAGT CACTCTGA GTTCTTGT GTCTCTCT GAGTANAGA ACGTTTCA AACCTTCA AACCTTCA AACCTTCA AACCTTCA AACCTTCA

```

FIG. 6D

```

        haeIII/palI
        haeI
1101   accI nlaIII          mli
        SCFPI          icrPI          nlaIII          sauIaI          sauIaI
        mvaI          ecorII          mvaI          sbal/ndelI(dam-)  sauIaI
        deaV          t f i I          deaV          pleI          dpaI(dam+)
        betHI          nlaIII          betHI          ddeI          dpaI(dam-)
        apyI(dam+) hlnI      apyI(dam+) hlnI      maeIII alwI(dam-)
        GAGCAGTTC TGCATACTG GAGCAGTTC GAGCAGTTC ATCAACCTG CCACCTTGA CTTCTTGTG CAGGACATAT
        CTTCTTTTA TCTGTACAA ACCATATGC CTCCTGCAG ACAATATGT CTTCTGTAC TTGTGTGTC GACAAACT GTTCTGTATG
        mli
        hlnI/acyI          acrPI          acryI
        abalI/bsalI      mvaI          ecorII          deaV
        scfPI          ecorII          ecorII
        mvaI          ecorII          sauIaI
        deaV          deaV          deaV
        betHI          bsalI          asuI          mli
        apyI(dam+)      bsalI          hyl          betHI
        mli              mli              hyl          hyl          ddeI          apyI(dam-)
        GACAGATTT GAAATGACA CTTTTTCC AAAAAATTG TTGCGAAT ATAACTCTT CCAAGATAT CCAAGCTTC TCTCTCATG CCAAGAGGA
        CTTCTTTAA CTTTACTGT GCAAAAGG TCTTTACTA AACCTTGA TATTTGGA GCTCTTATG GCTCCGAG AGAGTCTCA GCTCTCTT
        maeII
        apol              maeIII
        afiIII
        sfalI          accI          sbalI          mli
1201   AAAGCATCA AGTTAAGTT TGAAGCTAC GAGAGAGAG ACTACAGAC AACTGCTTC AGTTCCTG CTTCTTCTT AAAGCTATGC ATTTTATTA
        TTTCCATAT TCAATATTA ACTTCAGAT CTTCTCTTC TGATGTCT TCTACAGAG TTCAAGAGAC GAGGAGGA TTCTGATAG TAAATATAT
        emdBI
        acII
        tbaI
        kmulI/svmI      kruI
        betUI          styI          mli
        bsalI          bsalI          styI          styI          mli
1301   GACCATGCA CTTTCTGT CTTAGACC CTTTCTTTC GTTAGACC GTTAGACC GTTAGACC GTTAGACC GTTAGACC GTTAGACC
        CTTGTACCT GAAACACCC GAATCTGG GAACTGG GAACTGG CAACTGG CAACTGG CAACTGG CAACTGG CAACTGG CAACTGG
        nlaIII
        styI
        ncoI
        deaI
        bsalI
1401   GACCATGCA CTTTCTGT CTTAGACC CTTTCTTTC GTTAGACC GTTAGACC GTTAGACC GTTAGACC GTTAGACC GTTAGACC
        CTTGTACCT GAAACACCC GAATCTGG GAACTGG GAACTGG CAACTGG CAACTGG CAACTGG CAACTGG CAACTGG CAACTGG
        hpaI          maeIII
    
```



FIG. 6F

```
scrFI
nciI
mepI
hpaII
dswV
cauII
xmaI/ppaI
smaI
scrFI
nciI
daaV
cauII
bsiI
sau6I
baeII/palI
suI

scrFI
mvaI/bamI
ecorII
dswV

rraI
cepI
bpl401I
nlaIV
fru4II
scrFI
petI
bagI
nlaIV
fru4II
scrFI
nhiI
eco57I
mliI
dtaI
mliI

1801 CCTACTTGTA CAATCACTCT CTAGCCGCG CCCCTGCTTG CTTACTCTT CCAGCGGAA GTCCGCACTT CTTTCGTG NGTCGTCG
CGATGACAT CTTACTGCA GATCCGCGC GCGCGGATG GAGTGTCAG GAGTCCCTT CAGCCCTTA GAAAGCCAC TCGACACTG
CGATCCCGC CCGCTGCTG CCTAGCTC CTTACTCTT CCAGCGGAA GTCCGCACTT CTTTCGTG NGTCGTCG

fru4II
aluI
pvuII
nsp8II
dtaI
mliI bbvI

nboII
nciI
mepI
hpaII
daaV
cauII
sau6I
awaII
suI
dxaII
bbvI
mboII
berI
cfrI01
scrFI
nciI
mepI
hpaII
petI
bagI
fru4II
bbvI
mboII
berI
cfrI01
mepI
hpaII
scrFI
petI
bagI
fru4II
bbvI
mboII
berI
cfrI01
mepI
hpaII

1901 CCTCAGCTGC TCCAAATGCC GAAAGGAAAT GGCTAGCTG CAGATCTCTT CTTCCACTT GAGCGCGAG ACCCTGCTG CTTCGCGGAA CAACCACTAC
CGAGTCGKAG AGCTTACGG CTTCTCTTA CCAGTCCAC CTTAGTGA GAGCTCTA CTTGCTCTG TGGACACAC CAGCTCTT CTTGCTGATC
```

FIG. 6G

2001 CGCCATTATT CGATGTAANA CCTTTCGCG ACCTTCAATAT GCAGCCCTCTE CCTCANTGCG ACCTGCTACG TCTCTCTTCCA GCGAAMAMG AKACCCOCTOT  
 GCGGTATATA CCTCACTTTE GCGAAGGCTC ACCGAMTTAA CCTCGAGAC CGMGTATACG TCGCACTGCTG ACGCAGCTT CCTCTTTGTC TTCTGACCA  
 hgiAI/asphI bspI286 hgiAI/asphI bspI286 hgiAI/asphI  
 bspI286 bspI286 bspI286  
 bsiNRAI bsiNRAI bsiNRAI  
 sau96I svaII bmyI molI scrfI wvaI  
 svaII bmyI molI svaII bmyI molI scrfI wvaI  
 sauI spali/snoI bsciI dsaV scrfII scrfII  
 sauI spali/snoI bsciI dsaV scrfII scrfII  
 mli nli alw44I/snoI spyI(dcm+) draIII alw44I/snoI  
 mli mni bmyI mni mli nli alw44I/snoI spyI(dcm+) draIII alw44I/snoI  
 aJw4I fndIII berI hmi mni bmyI mni mli nli alw44I/snoI spyI(dcm+) draIII alw44I/snoI  
 fndIII hmi mni bmyI mni mli nli alw44I/snoI spyI(dcm+) draIII alw44I/snoI  
 bspHI nleIII ddeI bseAI mneIII scfI bnfI alw44I/snoI  
 bspHI nleIII ddeI bseAI mneIII scfI bnfI alw44I/snoI  
 3101 GCACCTGCCA TGCCAGTTC TTCTGACG AAGAGAGT TCCTCTCTG KCTACTCTA AGAAGACT GAGAGTACG AAGTGTGCT TACCCGAGAT  
 CCTGACCGT ACCTCCAAAG AAGATTCCT TTTCGTGAC AAGAGGACA TCAATGCAAT TCTTTCTGCA CCTCAGTCTC TTCACACAGG ATGGCCTCTA  
 hgiAI/asphI bspI286 hgiAI/asphI bspI286 hgiAI/asphI  
 bspI286 bspI286 bspI286  
 bsiNRAI bsiNRAI bsiNRAI  
 bmyI mneIII hphI nleIII  
 bmyI mneIII hphI nleIII  
 bmyI mneIII hphI nleIII  
 2201 TGACAAATGT AAGGCACTC AAGACTCAG CACACAGAC AAGAGATTC AGCTCAAAK CCTCACTTCTT GACACACTC ACACATGCTT ACGGTGCTG  
 ACTCTTACAA TTCCTGTGC CCTGTGCTC GTGTGCTCTG TTCTCTGAC TCGATGTTG GCGTGCACCA CTCTGTGAG TCTGTACGCG TCGCTACGCT  
 hgiAI/asphI bspI286 hgiAI/asphI bspI286 hgiAI/asphI  
 bspI286 bspI286 bspI286  
 bsiNRAI bsiNRAI bsiNRAI  
 bmyI mneIII hphI nleIII  
 bmyI mneIII hphI nleIII  
 bmyI mneIII hphI nleIII  
 2301 GAGCCCAAA CTTCGACAC AACTCCCGG TGCCAGCTT GCGCAAGCT CAATCTTGT GACACACTC CCGATGCTT ACGGTGCTG GAGCCTAAAT  
 CTCGGTCTTA GACACTCTC TTGAGGGGG ACGGTGCTG CCGTCTCTG GTTTAGACA CTGTGTGAG GCGTACAGG TCGCAGGGT CTCGGTCTTA  
 hgiAI/asphI bspI286 hgiAI/asphI bspI286 hgiAI/asphI  
 bspI286 bspI286 bspI286  
 bsiNRAI bsiNRAI bsiNRAI  
 bmyI mneIII hphI nleIII  
 bmyI mneIII hphI nleIII  
 bmyI mneIII hphI nleIII

FIG. 6H

```
          eam1105I
          sau96I
    scriI   bsp126   bsp11   mboII   mboII   styI
    nvaI    nlaIV   nlaIV   bpaAI   bpaAI   bpaAI
    ecorII  hgiCI   hgiCI   asuI    asuI    asuI
    dsaV    dsaI    bsaJI  spvI(dsw) spvI(dsw) spvI(dsw)
    bsaJI   bsaJI   bsaJI  bsaJI   bsaJI   bsaJI
    .maeIII mnlI  nlaIII bsaI  aIwHI  bsaII  mboII   mboII   styI
    CTATGACAC ACCTCGCGC TCGCCAGCGT GCCGAGACC TCAACTCTCG GAGAGCGGT CAGTCCTCT CTTCGCGCCA AAACCCAGGC ATACGCTTAT
    GAACACTCG TCGAGCGGC AGCGTCGCA CGCGTCGTCG ACTTGGGAC CTTCCTGGCA CTCAGAACCA GAAGCGCGT TTTGGTTCT TATCGAATA
    2401

    sau96I
    nlaIV
    nvaII
    sauI
    mspi
    hpaII
    scriI
    nciI
    dsaV
    cauII
    maeII
    pmlI
    eco72I
    mnlI  bsaAI
    ddeI  maeIII
    eco8II bbrFI
    bsu36I/mcII/suI
    maeII
    gctTGTGAC GTGAGCCAG GTCAGCGGT GCTGTGTGT GTCAGCGGT GTCAGCGGT GTCAGCGGT GTCAGCGGT GTCAGCGGT
    CTAAAGCGGC TGGGACTGC ATGTGAGCA CCACACTTC CACTCGTGC TTTGGGTGC TTTGGGTGC TTTGGGTGC TTTGGGTGC TTTGGGTGC
    2501

    acII
    theI
    fruDI1/mvnl
    bstOI
    bsh1236I
    sacII/satII
    nsp8II
    kspI
    deaI
    bsaJI
    aciI
    fru4HI mnlI
    maeII
    hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI
    mnlI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI
    bsaII bsaII bsaII bsaII bsaII bsaII bsaII bsaII bsaII bsaII
    scriFI mvaI bsrI
    ecoRII
    dsaV
    ecoMI bsrMI
    hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI
    mnlI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI hpaI
    rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI rsaI
    csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I csp6I
    2601 AATGCCAAGA CAAGGCCGC GGAGGAGCAG TTCAGACGCA GTTGGGTGT GTCAGCGGT CTACGCTGC TCGAGCGCA CTGCTGTGC CTCAGCGGT GCGAGCGGT
    TTACGTTCT GTTTCGGCG CCTCTGTCT AGTTTGTGT GAGGAGCAG CGATGCGAG AGTGTGTCT TACGACTTC CCTTCTCA
```

FIG. 6I

```

2701  ACAGTCCAA  GCTCTCCAAAC  AAGGCTCTCC  CAGCTCCCAT  CAGAGAACCC  ATCTCCAJAA  CCAAAGCMA  CCTCCGMAA  CCGCCGMAA  CCGCAGCTAT  ACGCTCTCC
      bsmA1      base1      mnl1      taq1      ava1      rna1      bell
      base1      csp41      bsp16071
TCTTCTCCTT  CCGAGCTTC  TTCTCCGAG  GTCCGCGTA  GCTCTTTTG  TAAAGTCTT  GCTTCTCTC  CAGAGCTCTT  GCTATCCACA  TATCCGAGC
      scrF1      ncl1      wpp1      hpa11      dsaV      csu11      rna1/pspA1
      sse1      scrF1      mva1      sco11      dsaV      batM1      spt1(dcm+)      gsp1      sc11      rsp011      fmdH1      fna4H1
      foki      bsl1      ava1      mnl1      sexA1      bspM1      spt1(dcm-)      bsl1      bsaJ1      bboV      bboV
2801  CCGATCCCG  GAGGAGATCA  CCAAGAACCA  GGTCACTCT  ACTCTCTCG  TCAAGCTT  CTACCCGAC  GACATCCGC  TCCAGTCCA  TCCATCTCCA  GAGCAGCCG
      foki      bsl1      ava1      mnl1      sexA1      bspM1      spt1(dcm-)      bsl1      bsaJ1      bboV      bboV
GCTTCTACT  CTTCTTACT  GCTTCTTGT  CAGTCCGAC  TCCAGCAGC  AATTCTCGA  GATCCGCTG  CATTACGAC  ACTCTACT  CTCTCCGCC
      nsp1      hpa11      mnl1      nla111      hnf1      nla1V      mbo11      scr1      alu1      bsaJ1      bspM1      bboV
2901  CAGCCGAGA  ACGACTACA  CAGCAGCCT  CCGATCTCG  ACTCCGCG  CTTCTTCTC  CTCTCAGCA  AGCTCAGCT  CCGCAGAGC  ACGTCCGAC  C
      gtcgctc  TCTTCTTGT  GTCTCCGGA  GCTTCTGCT  TCAAGCTT  CTACCCGAC  GACATCCGC  TCCAGTCCA  TCCATCTCCA  GAGCAGCCG
      scrF1      ncl1      wpp1      hpa11      dsaV      csu11      rna1/pspA1      sse1      scrF1      mva1      sco11      dsaV      batM1      spt1(dcm+)      gsp1      sc11      rsp011      fmdH1      fna4H1      bsl1      bsaJ1      bboV      bboV
3001  ACGGACAT  CTCTCAGC  TCCGATGC  ATGAGCTCT  GCAGACGC  TTCAGAGC  AAGCTCTC  CCTCTCCG  CCGTAAATG  TCCAGCAGC
      rna1/pspA1      asp701      nla111      sfam1      mnl1      ac11      ear1/bsp121      bsl1      cau11      scrF1      ncl1      wpp1      hpa11      dsaV      csu11      rna1/pspA1      sse1      scrF1      mva1      sco11      dsaV      batM1      spt1(dcm+)      gsp1      sc11      rsp011      fmdH1      fna4H1      bsl1      bsaJ1      bboV      bboV
TCTCTTGA  GAGAGTAC  ACGACTAC  TACTCCGAG  CTTCTCCG  CCGTAAATG  TCCAGCAGC  CCGTAAATG  TCCAGCAGC
      scrF1      ncl1      wpp1      hpa11      dsaV      csu11      rna1/pspA1      sse1      scrF1      mva1      sco11      dsaV      batM1      spt1(dcm+)      gsp1      sc11      rsp011      fmdH1      fna4H1      bsl1      bsaJ1      bboV      bboV

```









FIG. 6N

```

nlaIV
aciI
tbaI
fmdII/mvni
bcuI
bsh1236I
himpI
hhaI/cfoI
hnlI/acyI
ahaII/bshI
eatII
dclI/melI
4701 AATAATGCTT TCTTAGGCT CAGTGGCCAC TTTCGGCA MATGCGCC GAGCGACTT TGTATTAT TCTAATATC ATTCAANTAT GTATCGGCTC
TTATTACCA ACAATCTGCA GTCCACGGTC AAGAGCGCT TNCGCGCC CTGGGGCA ACAATATA AAGATTATG TAAGTTTATA CATAGCCDAG
rcaI
bepII
barBI
aciI nlaIII
fwdIII
4801 ATCGACCAAT AACCGTATA AATCTTCA AATCTTCA TATATGAA AAGCGAGAG TATGACTT CAGCTTCC GTGTGCGT TATTCGCTT TTTCGGCTAT
TACTCTGTA TTGGACTAT TTACGAGTT ATTATACT TTCTCTTC ATACTATAA GTTGTNAGG CACAGCGGA ATAGGGGAAA AAGCGCGTA
espI
bakI/kep12I
4901 TTGCGCTCC TGTCTTCT CACCAGAAA CCGTGTGAA AGTAAAGAT GCTGTGTGC AGTGTGTC AGAGTGGT TACATCGAC TCGATCTCAA
AAGCGAGG ACAANAACA GTGGCTTT CCGACACTT TCAITTTCTA CAGCTCTTAC TCGACTGCA TCGTGTGCA ATGTACTGCT ACTGAGTT
hphI
nboI/ndaiI(dam-)
dpoII(dam-)
bep1286
bsh1236I
hnlI/acyI
ahaII/bshI
eatII
nboII
bep1286
bsh1236I
hnlI/acyI
ahaII/bshI
eatII
5001 CAGGCTAAG ATCCCTGGA GTTTTGGCC CGAAGAGCT TTTCGATGA TCGGACTTT TAAATCTG CAGTGTGCG CAGTATATC CCGTATGAC
GTCCGCTTC TAGGACTCT CAAAGCGG GCTTCTTCA AAGGTACT ACTGTGAA ATTCTGAGC GATGACCGC GCGATATAG GCGTACTCT
scaI
nboI/ndaiI(dam-)
dpoII(dam-)
aciI
nspIII
scrFI
nclI
nspI
hpaII
dbsV
cruII
bcgI mcrI fwdIII
aciI
rseI
csp4I
scaI
hphI
nclI
sfaII
fokI
n.laIII
5101 GCGGGCAAG AGCAATCGG TCGCCGATA CACTATCTC AGAATGACT CATTATCTC TCGCCCTCA CAGAAACA TCTTNGCAT GCGATACAG
CGCCGCTTC TCGTTGAGCC AGCCCGGAT GTGTANGAG TCTTACTTGA CAGTCAATG AGTGGTCACT GCTTTTCT GCGATGCTA CCGACTATC

```

FIG. 60

```

sau96I          sau96I          sau96I          sau96I          sau96I          sau96I
avali          avali          avali          avali          avali          avali
sauJAI          sauJAI          sauJAI          sauJAI          sauJAI          sauJAI
mboI/ndeII(dam-) mboI/ndeII(dam-) mboI/ndeII(dam-) mboI/ndeII(dam-) mboI/ndeII(dam-) mboI/ndeII(dam-)
dpoII(dam-)    dpoII(dam-)    dpoII(dam-)    dpoII(dam-)    dpoII(dam-)    dpoII(dam-)
pruI/bapCI    pruI/bapCI    pruI/bapCI    pruI/bapCI    pruI/bapCI    pruI/bapCI
mcrI          mcrI          mcrI          mcrI          mcrI          mcrI
aluI          aluI          aluI          aluI          aluI          aluI
aciI          aciI          aciI          aciI          aciI          aciI
530I TAGAGAMIT ATGCACTGCT GCGATAGCCA TGAATGATAA CACTGCGCC NACTTACTTC TGACACAGCT CGAGAGCGG AMGAGACTAA CCGCTTTTIT
ATTCTCTTAA TAGCTACAGA CGATATTGCT ACTGACTATT GTGAGCGGG TTGATGAGAG ACTGTGCTGA GCTCTCTGCG TTCTCTGANT GCGGAAAAA

fnddHI          fnddHI          fnddHI          fnddHI          fnddHI          fnddHI
bbvI          bbvI          bbvI          bbvI          bbvI          bbvI
nlaIII          nlaIII          nlaIII          nlaIII          nlaIII          nlaIII
mceIII          mceIII          mceIII          mceIII          mceIII          mceIII
530I GCGACAGATG GGGATGATC TACTGCTCT TGAATGCTG GAGCGCGAG TGATGATGAG CATACAGAAC GAGCGAGCGG ACACAGAGAT GCGAGAGACA
CGTGTGTGAC CCGCTGATGAC ATTAGAGGGA ACTAGAGACC GTTGGCGTGG ACTTACTTCC GTATGTTTG CTTCTGCGAC TGTGTGCTGA CCGTGTGTOI

nspi          nspi          nspi          nspi          nspi          nspi
hpeII          hpeII          hpeII          hpeII          hpeII          hpeII
scriI          scriI          scriI          scriI          scriI          scriI
aluI          aluI          aluI          aluI          aluI          aluI
rwal          rwal          rwal          rwal          rwal          rwal
mceI          mceI          mceI          mceI          mceI          mceI
sauI          sauI          sauI          sauI          sauI          sauI
540I ATGCGACAGC CTTGCGGCAA ACTATTAGCT GCGAGACTAG TTACTTAGC TTCCCGGCAA CATTATAG ACATGATGTA GCGCGATAAA GTTGCGAGAC
TAGCTTGTGT GCAACCGCTT TGAATATGTA CCGCTTGATG ATGATGATG ATGGGCGCTT GTTAAATATC TGTACTACTT CCGCTTATT CAGCTGCTG

nspi          nspi          nspi          nspi          nspi          nspi
hpeII          hpeII          hpeII          hpeII          hpeII          hpeII
cfr-101        cfr-101        cfr-101        cfr-101        cfr-101        cfr-101
nlaIV          nlaIV          nlaIV          nlaIV          nlaIV          nlaIV
hpaII          hpaII          hpaII          hpaII          hpaII          hpaII
550I CACTTCTGCG CTGCGGCTT CCGCGTGGCT GGTATTATGC TGAATATGTA GAGCGCGTGG AGCTGCGTGC TCGCGTATC ATTGCGAGAC TGGCGCGAGA
GTGAGAGCGC GAGCGCGGAA GCGCGAGGTA CCGAATATAG ACATATTTAGA CTTGCGGCTAG TCGCGAGCGAG AGCGCGATAG TAACTGCTG ACCCGCGTCT

bglI          bglI          bglI          bglI          bglI          bglI
sau96I          sau96I          sau96I          sau96I          sau96I          sau96I
heeIII/paII    heeIII/paII    heeIII/paII    heeIII/paII    heeIII/paII    heeIII/paII
hinPI          hinPI          hinPI          hinPI          hinPI          hinPI
hhaI/cfoI      hhaI/cfoI      hhaI/cfoI      hhaI/cfoI      hhaI/cfoI      hhaI/cfoI
560I TAGTAAAGCC TCGCGATGCG TAGTTATCTA CAGAGCGGG AGTGAAGGAA CTATGATGTA AGCAATATGA CAGATGCTG AGATAGCTGC CTCAGCTGTT
ACCATTTGCG AGGCGATAGC ATCAATATAT GTGCTGCGCC CAGTGCCTT GATAAGCTACT TCGTTATCT GTTAGAGAC TCTATTCAGS GAGTACTATA

```

FIG. 6P

```

rmaI sauJA1
sauJA1 hphI mboI/ndelII(dam-)
mboI/ndelII(dam-)
dprII(dam+) dprII(dam+)
dprII(dam-) dprII(dam-)
aha:II/draI maeI
tru9I cru9I bstYI/xhoII bstYI/xhoII
maeI maeI alwI(dam-) mboII(dam-)
ahaII/draI
5701 AACGCTTGGT AACCTGTCA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAACTT CATTTTAT TTAANGAT CTAGGTTAG ATGCTTTTGG
TTCTAACCA TTGAGNGCTT GGTTCNATG AGTATATAG AAATCTACT AMTTTTOA CTAAATAA ATTTTCTA GATCCACTT TAGCAAAAAC
maeIII
oleIII maeII
rcalI tru9I
baphI maeI
oleIII maeII
5801 ATAAATTCAT GACCAAAATC CCTTAACTG AGTTTCTT CCACCTGCG TCAGACCCCG TAGAAAGAT CAAGAGACT TCTTGGAT CTTTTTTCT
TATTAGATA CTGTTTTC GBAATTCNC TCAAAACA GGTACTGCG AGCTGCGCG ACTTTTCTA GTTCTTGA AGAATCTTAT GAAAAAAGA
oleIII maeII hgaI
rcalI tru9I ddelI
baphI maeI
oleIII maeII haeIII/palI
5901 GCGCGTAACT TCTCTCTTCC AACAAAAA ACCACCTA CCAAGCCTG TTTTCTTCC GBAJTCAGAC CTCACACTT TTTTTCGAA GGTACTGCG
CGCGTTTTC ACCACGACG TTTTCTTTC TGTGCTTAT GGTCTCCACC AACAAAGCG CTATTCTTC GATGTTTAT AAAAGGCTT CCATTGACC
hinPI rmaI
hhaI/cfoI maeI
6001 TTCAGGAGG GCGGATACC AAATACTGTC CTCTTACTGT ACCCTGATT AGGCTACAC TTCACAGACT CTTCAGACC GCTTACTAT CTCGCTCTGC
AGTCTCTC GGTCTATG TTTATGAGAG GAAATACCA TCGGATCA TOGGTGGTG AGTTCTTGA GACATCTG GAGTCTAT GAGGGAGAG

```



FIG. 6R

```

haeIII/palI
haeI
scrFI
sr/I b/I
ecrII
dsuV
nleIII
nspI
haeIII/palI nspHE
spVI/dcm-I haeI
650) CTTGACCTT TACTGCTT TTGTCACAT GTTCTTTCY GGTATACC CTAATCTT GGTATACCT ATTACGCTT TTACGTGAC TFAFACGCT
G/PACCGANA AGACCGAA AACGAGTGA CAGAAAGA CCGATAGC GACTATGCA CTAATGCA TAAATGCGA AACTCACTG ACTATGCGA
cheI
fnuDII/mrnI
bstXI
behI236I
hinfI
hhaI/cfoI
thaI
fnuDII/mrnI
betUI haeIII/palI
BehI236I
660) CACCGAGCC GAGGACGA GCGACGAG TCAGTGACG AGAAAGCGA AGAGCGCA ATACCGAAC CCGCTTCC CCGCGTTCC CCGATTCY
C/CACGCTCC CTTCCTGCT ACTCACTCC TACTTCCCT TATCCCGCT TATCCCTTC GCGAGAGG CCGCGTACC CCGTACTAA
scrFI
mvaI
ecrIII
dsuV
nlaIV htaMI
hgICl sppI{dcm-}
670) AATCGACTG GCGACGACG TTTCGACT GAAAGCGG CMTGACGC AACCGATTA ATGTGACTTA CTTCACTCAT TAGGACCC AGCCITFACA
TTAGCTGAC COTGCTGCC AAGGCTGA CTTTCCGC GTACTGCG TACTGTTAT TACTACTAT GAGTGAATA ATCGTGGG TCGAAATGT
truSI
mseI
seel/seml/vsp I
680) CTTATGCTT CCGCGCTGA TTTTGTGCG AATGTGCG GATACAT TTACAGCG AACACTAT GACTGATY ACCAATTA
GAATACGA GCGCGAGCT TACACTCG CTAATGTA AAGTGTCC TTGTGATA CTGTACTAA TCGTAAAT
>length: 6889

```

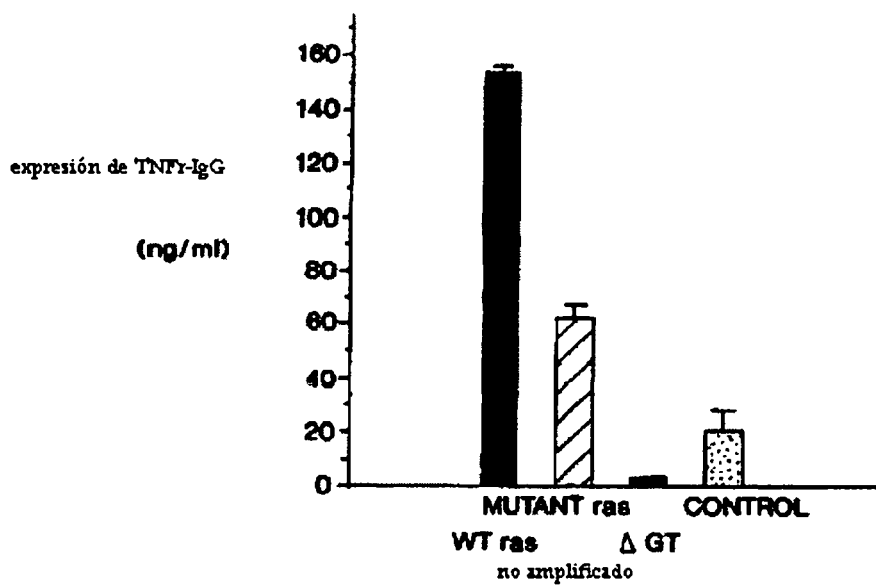


FIG. 7A

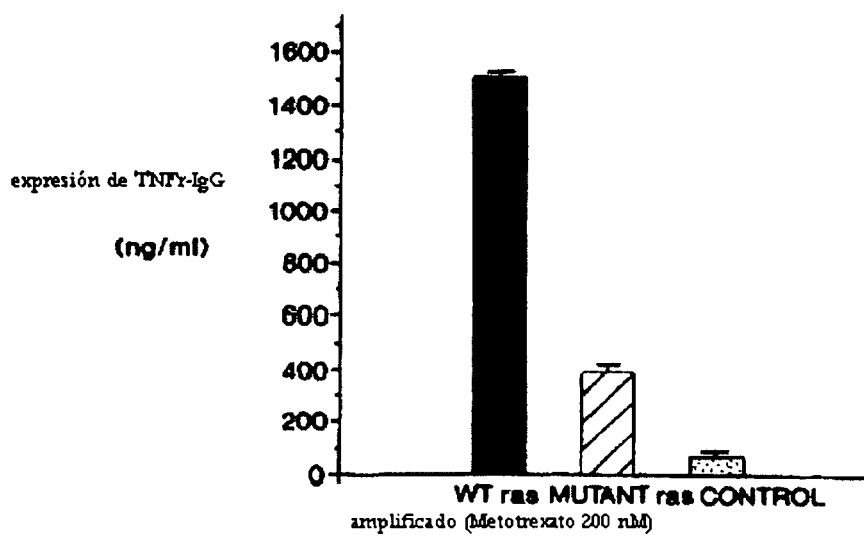


FIG. 7B

FIG. 8

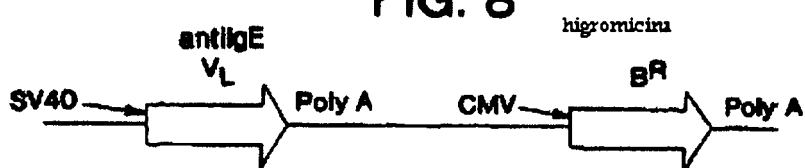
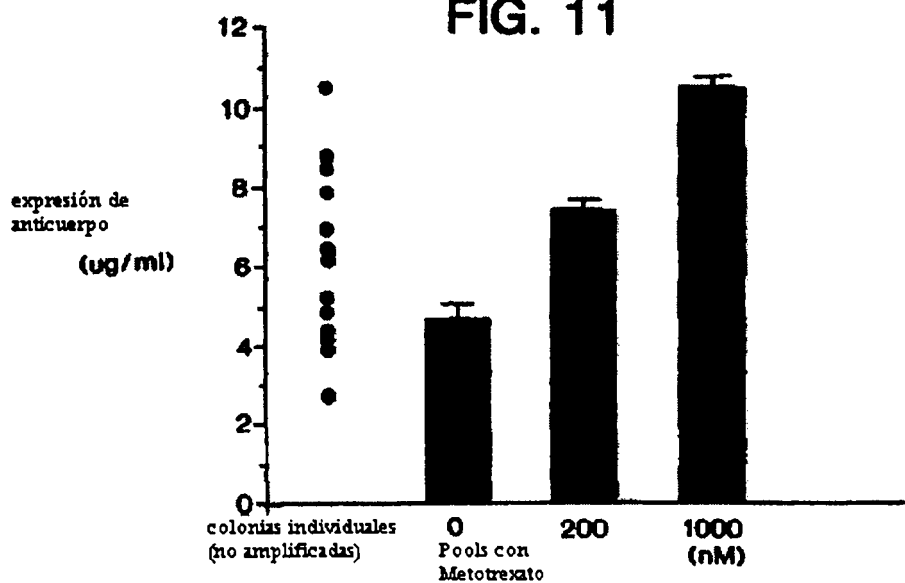


FIG. 11



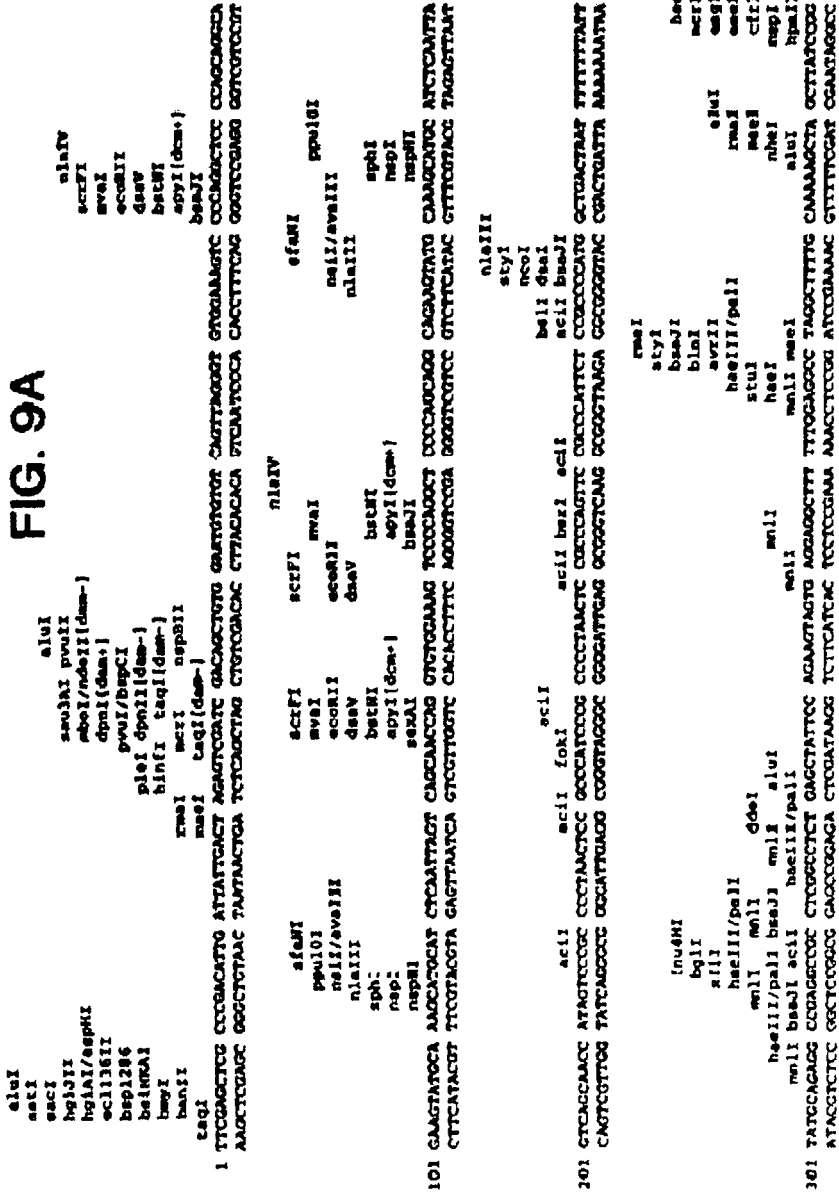






FIG. 9D

	maeIII	scfI	fokI	bsII	bsaJI	mnlI	laqI	ecorI
	hphI	scfI	fokI	bsII	bsaJI	bsaJI	claiI/bspI06	spol
1201	ATAGATTTA	GGTGACACTA	TAGATAACAT	CCACTTACC	TTCTCTCCA	CAAGTGTCCA	CTCCAGAGTC	CAACTGCACC
	TATGCTAAT	CCACTGTGAT	ATCTATTGTA	GGTGAAACGG	AAGAGAGGT	GTCCACAGGT	GAGGTGCG	GTTCAGGTGG
								AACCCAGATA
								GCTAACTTA
	nlaIII	styI	pflMI	neoI	dseI	bsaJI	fokI	maeI
	bsaJI	fokI	nlaIII	fokI	maeI	rmel	rmaI	rmaI
	CCACCATGG	ATGCTCATGT	ATCAGCTTT	TTCTAGTAG	AACTGCACCT	GTCTAGTACT	GGAGTACTT	CAAGAGTTCA
1301	GGTGATACC	TACCGTACA	TATGAGAAA	AAGATCATCG	TTAGCTTGA	CCTCAGTTGA	CTTCATAGT	GGACCACTTC
								AAGCCACAC
								CGGACCACTG

cau96I	scfI	maeI	ecorII	bsaI	bsaCI	fnu4HI
avaII	scfI	maeI	ecorII	bsaI	bsaCI	fnu4HI
asuI	scfI	maeI	ecorII	bsaI	bsaCI	fnu4HI
scrFI	scfI	maeI	ecorII	bsaI	bsaCI	fnu4HI
scfI	scfI	maeI	ecorII	bsaI	bsaCI	fnu4HI
scfI	scfI	maeI	ecorII	bsaI	bsaCI	fnu4HI
scfI	scfI	maeI	ecorII	bsaI	bsaCI	fnu4HI
scfI	scfI	maeI	ecorII	bsaI	bsaCI	fnu4HI

FIG. 9E

```

scrPI          scrPI          scrPI          scrPI          scrPI
ncIX           ncIX           ncIX           ncIX           ncIX
mspI           mspI           mspI           mspI           mspI
hpaII          hpaII          hpaII          hpaII          hpaII
dseV          dseV          dseV          dseV          dseV
cmaII         cmaII         cmaII         cmaII         cmaII
bclI           bclI           bclI           bclI           bclI
xmaI/pspAI    xmaI/pspAI   xmaI/pspAI   xmaI/pspAI   xmaI/pspAI
smuI          smuI          smuI          smuI          smuI
scrPI          scrPI          scrPI          scrPI          scrPI
ncIX           ncIX           ncIX           ncIX           ncIX
dseV          dseV          dseV          dseV          dseV
cmaII         cmaII         cmaII         cmaII         cmaII
sau96I        sau96I        sau96I        sau96I        sau96I
nlaIV         nlaIV         nlaIV         nlaIV         nlaIV
ecoRI         ecoRI         ecoRI         ecoRI         ecoRI
dseV          dseV          dseV          dseV          dseV
sau3AI        sau3AI        sau3AI        sau3AI        sau3AI
nboI/ndcII(dse-) nboI/ndcII(dse-) nboI/ndcII(dse-) nboI/ndcII(dse-) nboI/ndcII(dse-)
hpaII          hpaII          hpaII          hpaII          hpaII
mspI          mspI          mspI          mspI          mspI
hpaII          hpaII          hpaII          hpaII          hpaII
nroI          nroI          nroI          nroI          nroI
hpaIII        hpaIII        hpaIII        hpaIII        hpaIII
hpaII          hpaII          hpaII          hpaII          hpaII
hpaII          hpaII          hpaII          hpaII          hpaII
nlaIV         nlaIV         nlaIV         nlaIV         nlaIV
bclI/xhoII    bclI/xhoII    bclI/xhoII    bclI/xhoII    bclI/xhoII
bamHI         bamHI         bamHI         bamHI         bamHI
nlaI          nlaI          nlaI          nlaI          nlaI
nlaI(dse+)    nlaI(dse+)    nlaI(dse+)    nlaI(dse+)    nlaI(dse+)
hpaI          hpaI          hpaI          hpaI          hpaI
accIII        accIII        accIII        accIII        accIII
hpaI          hpaI          hpaI          hpaI          hpaI
bclI          bclI          bclI          bclI          bclI
aIuNI         aIuNI         aIuNI         aIuNI         aIuNI
1401 GCGGGGGC TCACTCCCTT TGTCTGTGC ACTTCTGCG TACTCCATCA CTTCTCGATA TACTCGAAC TCGATTCGTC AGGCCCCGG TTAGGGCCCTG
CGTCCCGC AGTAGGCA ACAGGACG TCAAGCG ATGAGGTAT GAGGCTTAT ATCGACTTA ACTTAGCGG TCCGGCGC ATTCCCGGAC
scrPI          scrPI          scrPI          scrPI          scrPI
ncIX           ncIX           ncIX           ncIX           ncIX
dseV          dseV          dseV          dseV          dseV
cmaII         cmaII         cmaII         cmaII         cmaII
sau3AI        sau3AI        sau3AI        sau3AI        sau3AI
nboI/ndcII(dse-) nboI/ndcII(dse-) nboI/ndcII(dse-) nboI/ndcII(dse-) nboI/ndcII(dse-)
hpaII          hpaII          hpaII          hpaII          hpaII
mspI          mspI          mspI          mspI          mspI
hpaII          hpaII          hpaII          hpaII          hpaII
nroI          nroI          nroI          nroI          nroI
hpaIII        hpaIII        hpaIII        hpaIII        hpaIII
hpaII          hpaII          hpaII          hpaII          hpaII
hpaII          hpaII          hpaII          hpaII          hpaII
nlaIV         nlaIV         nlaIV         nlaIV         nlaIV
bclI/xhoII    bclI/xhoII    bclI/xhoII    bclI/xhoII    bclI/xhoII
bamHI         bamHI         bamHI         bamHI         bamHI
nlaI          nlaI          nlaI          nlaI          nlaI
nlaI(dse+)    nlaI(dse+)    nlaI(dse+)    nlaI(dse+)    nlaI(dse+)
hpaI          hpaI          hpaI          hpaI          hpaI
accIII        accIII        accIII        accIII        accIII
hpaI          hpaI          hpaI          hpaI          hpaI
bclI          bclI          bclI          bclI          bclI
aIuNI         aIuNI         aIuNI         aIuNI         aIuNI
1501 GAATGGGTTG CATCGATTAC GTATCGCGA TCGACTTACT ATAGCCCTAG CGTCAGGCG CGTATCGACTA TACTCGCGA CGATTCGAA AACACCTTCT
CTTAGCCGAC GTTACTTATG CATAGCGCT ACTTAGTGA TTATGACTG GCAATTCGG GCAATGATAT ATTAGCGCT GCAAGCTT TTGTATAGA

```

FIG. 9F

```

scfI          scrfI
mvaI          mvaI
ecorII       ecorII
dsv          dsv
bclII        bclII
spyI{dcm+}  spyI{dcm+}
hinfI
nhaI/cfoI   nhaI/cfoI
nlaIV       nlaIV
sarI        sarI
kssI        kssI
hinfI/acyI hinfI/acyI
hgiCI       hgiCI
haeIII      haeIII
banI        banI
abaII/baeII abaII/baeII

nliI
xhoI
paeK7I
evaI
hpaI/aspHI hpaI/aspHI
bepI286    bepI286
baeI286A   bepI286A
bmyI taqI bbvI
mliI
dseI drdI
1601 ACTTCGAGAT GAAAGCCTG CCGCCVAGG ACACTGCGT CTATTATTTT GATVAGACA GCGACTATTT CCGCCCTGCG CACTGCGCG CTCGGCGTCA
TGCACCTTA CTTGTGGAC GCACGCTCC TGTACGCGA GATATACCA CGACTTCGT CCGATATAA GCGCGCGAC GTVAGCGCC ACACCTCCG

scfI          scrfI
mvaI          mvaI
ecorII       ecorII
dsv          dsv
bclII        bclII
spyI{dcm+}  spyI{dcm+}
bseAI maeIII
nlaIV       nlaIV
abaII/aspHI abaII/aspHI
bepI286    bepI286
baeIII/paiI baeIII/paiI
1701 ACCAGACTT GTCACCTCT CCGCGCGTC CACTGCGCT TCGCCCTGCG ACCCGCTCC ACCCGCTCC CCGCGCGAC ACCCGCTCA
TCCCTGGAC CATGCGAGA GAGCCCGAG CCGTCTCCG GGTACCGCA ACCCGCTCC TGGAGCGG TCTCTGCGA GAGCCCGCG TGGCGCGAC
```

FIG. 9G

```

scrFI          hinfI          hgiAI/aspHI
swaI           nsiI           bspI386
ecoRII        hinfI/acyI       bspI386
dseV          hgiCI           bsiWAI
batWI        haeII           bspI
bclI         hpaII           fnu4HI
spyI10m*)   cfrIbI        shalI/beaNI
fnu4HI       baMI tthI111/aspI  acII spALI/snoI dseV
bbvI         bclI egeI maeIII  ddeI hhaI/cfoI nsp8II alu48I/snoI swaII
1801 GGTCTGCTGG TCAGAGACTA CTTCCTCCGAAA CCGGTGACGG TGTCTGTGAA CTGACGACCT GTCAGCAGCG GGTGACGACT GTTCCTGACT scfI
CCGACCCAGC AGTTCCTGAT GAAAGGACTT GGCACACTGC ACAGCACTT GACTCTCCCG GACTCTCCCG GGCAGCTTGG GAAAGGCTGA CAGGATGTC

ddeI dleI    nlaIV    hgiCI
mliI hinfI   hpaII   fnu4HI
ecoRII      mliI bbvI    bclI
bsu36I/mstII/snaI ddeI bbvI bspI286   hpaII
bsu36I/mstII/snaI ddeI hpaI bspI      maeIII
1901 CCTCAGACT CTACTGCTC AGCAGGCTGG TGACTGTCC CTCTAGGCG CTCTAGGCG AGACTACAG CTCCAGCTG ATCCAGCAGC CCGCCAGC
CGACTCTGA GATGAGGAG TGTCTCCAGC ACTGACAGCG GAGATCTCG ACCCTCTGG TCTGATGTA GACTCTGAC TTTGCTTTCG GGTCTCTGTC

nlaIV    hgiCI
scrFI    swaI    hpaII
ecoRII   dseV
batWI    bclWI   hpaII
spyI10m*)  hpaII   hpaII
fnu4HI     bclI    bspI286
bbvI       hpaII   hpaII
2001 CAGGTGGAC AGAAGCTG ACCCCAAATC TTGTGACAA ACTCCACAT GGCACCTG CCGACCTG CCGACCTG GACTCTGCTC CCGACCTG AGTCTCTGTC
CTTCACCTE TTCTTCAC TCGGTTTG AACACTTCTT TATGTGTA GGTGTGTGA CCGGTGTGAC GGTGTGTGA CTTGAGAC CCGTGTGAG TCGAAGAG

```



**FIG. 9I**

```

scrFI      scrFI
ncII      nvaI      scrFI
mspI      ecorII      mvaI      scrFI
hpaII     dsav     ecorII
dsvV      bstXI     dsav
csmI      sxaI     bsrI
xmaI/pspAI  sxaI     spxI(dcm+)
bnaI      scrFI
scrFI      mvaI      scrFI
ncII      ecorII      mvaI      scrFI
dsav     dsav     ecorII
caulI     fckI     mboII
          rseI     bell     bsaJI     sscI
          -sp6I     bpi1497I  bsi     avai     szi/ksp612I
2401  cccgcagmac cagcagctga cagcagctga cagcagctga cagcagctga cagcagctga cagcagctga cagcagctga cagcagctga
          ccgcctcttc cttctctctc cttctctctc cttctctctc cttctctctc cttctctctc cttctctctc cttctctctc cttctctctc
          dsal     bell     bsaJI     sspI
          fnu4HI     fnu4HI     bbaVI     bpaII
2501  acatccctcc cagctgcgag acgaaatggc acccggagaa caactacagc acccagcttc cctctctgca cttccagctc tctctcttcc tctacagca
          ttatagcagca cttctctctc tctctctctt tctctctctt tctctctctt tctctctctt tctctctctt tctctctctt tctctctctt
          mspI
          fnu4HI     bpaII     nlaIII
          dsal     bsaJI     fnu4HI     mmi     mboII     nlaIII
          hpaII     bsp6I     bsvI     sep700     nlaIII
2601  cttcaccctg cagcagacca gttggacagca gggaaacctc ttctctctc cctctctgca tgacgctctg cagacacctc acagcagca cagcctctc
          ctagctggac cttctctctc ctagctctgt cccctctgag aacagtagca gccctctctc actccggagc gttctctgca tctctctctc cttccagctc
          bpaII     tagI
          dsal     ncII     saII     fnu4HI     sau96I
          mspI     hincII/hindII     rnaI     hincII/hindII     sfiI     ncoI     haeIII/paI
          hpaII     maeI     bccI     sfiI     haeIII/paI
          dsav     sau96I     pleI     psi     eaeI     dsal     esuI     alaI
          caulI     haeIII/paI     bspI     aluI     cfrI     bsaJI     fru4HI
          bsaI     asuI     hinfI     bsp6I     hindIII     bglI     nlaIII     bbvI     maeIII
2701  cttctctctg cttaaatgag cccagccccc ttatgtggac ttgcagagc ttgcagagc ttgcagagc ttgcagagc ttgcagagc ttgcagagc
          gacagagcc catttacta cctctctctc kttctctgct cagctctctc acccctctcc acccctctcc acccctctcc acccctctcc acccctctcc
          scrFI      taqI
          ncII     ssaI
          mspI     hincII/hindII     rnaI     hincII/hindII     sfiI     ncoI     haeIII/paI
          hpaII     maeI     bccI     sfiI     haeIII/paI
          dsav     sau96I     pleI     psi     eaeI     dsal     esuI     alaI
          caulI     haeIII/paI     bspI     aluI     cfrI     bsaJI     fru4HI
          bsaI     asuI     hinfI     bsp6I     hindIII     bglI     nlaIII     bbvI     maeIII
          gacagagcc catttacta cctctctctc kttctctgct cagctctctc acccctctcc acccctctcc acccctctcc acccctctcc acccctctcc

```

FIG. 9J

```

                teqi(dam-)
                ciaz/bsp106(dam-)
                seuzAI
                aboi/ndelII(dam-)
                dprII(dam+)
                dprII(dam-)
                alvII(dam-)
    2801
    sfaWI  apoI
    MAAACAATG CACCAAAAT TTCACAAATA AAGCATTTT TTCACCCCAI TCTNCTGTA GTTTCACAA ACYATCAAF STACTATAC ATCTCTGAT
    TTTCATTATC GTACTTTTA AAGCTTTTAT TCTGTAANA AATGACATA MATGACACAC CAACAGCTT TGNSTGTTA CEMANATAG TACAGACTTA
                rnaII
                bami
                nraI
                ssaI
                csp6I
                nleIV
                hspI
                hgICr
                bami
                asp71B
                waiI
                ddel acII
                pvuII
                nsp8II
    2901
    CMTGCGAAA TTAATTCGC GCACACCAAT GCGCTGAAT AACCTCGAA ACTCTGAA AKAZCACTT GCTTAGACAC CTCTAGCGC GAAAGACAC ACCTGCGAA
    CTAACCTT PATTAAGCCG CTTCTGCTTA CCGACTTTA TTGAGACTT TCTCTGAAA CEMATCTAG GAAACTCGC CTTCTCTG TCGAAGCTT
                ssaI
                csp6I
                nleIV
                hspI
                hgICr
                bami
                asp71B
                waiI
                ddel acII
                pvuII
                nsp8II
                nleIV
                scfFI
                mvaI
                ppu10I
                nraI
                nraI/svniI
                nleIV
                betMI
                sphI
                nspI sfamI
                bsaJI
                sfaMI
                ppu10I
                mvaI
                scfFI
                nraI
                dsaV
                betMI
                sphI
                nspI sfamI
                bsaJI
                sfaMI
    3001
    TGTGTGTCAT TTAGCTGTA CAAGTCCC ACACCTCCA GACAGCAGA STATAUAAO STACTATATC AATTATGTCG CACAGCGTA TCGAAGTTC
    ACAACAGATC ATATCCACAC CTTTCCCGG TCCGAGCGT CAGTCTTTC CATACTTTC GFACTUATAG TTAATGATC STTCTGTCAC ACCTTTCAG
                nleIV
                scfFI
                mvaI
                ppu10I
                nraI/svniI
                nleIV
                sphI
                nspI
                nspMI
                nraI(dam+)
                bsaJI
    3101
    CAGACTTCC CAGCGACAC AACTATCCA ACCTGATTC TCAATTATC ACGACAGAA GTTCCCCTC TACTCTGTC TACTCCCGCC CTTACTCTCC
    GATCCCGCGC GTTCCCTCC TTCAATGCTT TCTGTGTAG AATTATGTC TCTGTGAT CAGGCGCG ATTGAGCCG GTAGGCGCG GATTGAGCG
                teqi(dam-)
                ciaz/bsp106(dam-)
                seuzAI
                aboi/ndelII(dam-)
                dprII(dam+)
                dprII(dam-)
                alvII(dam-)

```





FIG. 9M

```

4101 ACAGCTGCTCTGATCCGCGATAGTTAAGCCCAACTTCCCTATCCCTAGCTGCTGCTCAATCCGCTCCGACACACCCGACACACACCCGCTGACACACCC
TCTTAAAGCA GACTACAGCGG TCCATATCC GTTAGCGCA TACCACTCA CTGACCTGAT ACCACACCGG GCTCTGTCG GCTTATCGCC GACTCTCCCG
      acii      fnu4HI      tru9I      maeII      hinpI      hnsII/mnhI
      fnu4HI      tru9I      maeII      hinpI      hnsII/mnhI
      sfahI      maeI      acII      hnsII/mnhI      hnsII/mnhI
      scrPI      hpaII      hpaII      hpaII      hpaII
      nciI      scrPI      mspI      mspI      mspI
      dsaV      sfahI      acII      hnsII/mnhI      hnsII/mnhI
      cauII      foki      aluI      maeIII      cruII      bbsI      hiaIII      mliI      hphI      hpaII
4201 CTGACCGCTTCTCTGCTCCCGCATCCGCTTACAGACAA GCTCTACCGG TCTCTGGAG CTGATATCTT CAGAGCTTTT CAGCTCTATC ACCGAAACG
GACTCCCGCA AAGACAGAG GCTCTAGCGG AATCTCTGTT CGACACTGCG NAGGCGCTC GACTACACA GTTCCGAAA GTCGACTAG TCGCTTTTCC
      nspi      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII
      scrPI      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII
      dsaV      sfahI      acII      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI
      cauII      foki      aluI      maeIII      cruII      bbsI      hiaIII      mliI      hphI      hpaII
4301 GCGACCGCTATTCTTGAG AGCAAGCCGCTCTCTATAC GCTATTTTT ATAGCTTAT GTCAATATA TAATGCTTC TTAGCTCA GCTGCGACTT
CGCTCCGCA TAGAGACTTCTCTCTCCG GAGCACTAG CGGATAAAA TATCCAAATA CACTACTATT ATTACCAAG ATCTCTCGT CCACCTCTAAA
      mliI      haeIII/paiI      nlaIII      rcaI      hpaII      hpaII
      sbolI      hpaII      spu96I      asuI      tru9I      rcaI      nlaIII      maeII      hialI/asyI
      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII
      mliI      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII      hpaII
4401 TTGCGCGAAA TGTGCGCGG ACCCTATTTT CTTTATTTT CTAAATGAT TGAATATAT ATCCCTCAT GAGCAATAA CCGGATATA TCGTTCAATA
AGCCCTTTT AGACCGCCTT TCGGATAAA CAATATAAA GATTTTATA AGTTTATCA TAGGAGATA CTCTGTATT GCGACTATT ACGAATATA
      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI
      sbolI      cauII/ksp632      hpaII      hpaII
4501 ATATTGAAA AGCAAGCTA TGAATATCA ACATTCCTT GTCGCGCTA TCCCTTTT TCGGATTT TCGCTCTCG TTCTTCAATA CCGCAAGCC
TATPACTTTT TCTCTCTCAT ACTATATAT TGTAAAGCGA CAGCGGAT AAGGATAAA ACGCTTAA ACCGAAACG AAGAGACTT GCGCTCTTCC
      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI      hnsII/mnhI
      sbolI      cauII/ksp632      hpaII      hpaII

```

FIG. 9N

```

hgiAI/aspNI
bepI286
    sauJAI
    aboI/ndaiI(dam-)
    dpmI(dam-) bmy?
    dpeI(dam-)
    eco57I
    stamI aboI(dam-)
    4601 CTGGTCAGG TAAAGATGC TGAAGTCAG TTTGATGCAC GAGTGGTTA CATGATGTC GATTCACAA GCGCTAAGG CCGTAAAGT TTTGCGCCCG
    GACCACTTC NTTTTCTAGC ACTTCTAGC ACCCAAGTC CTACCAANT CTAGCTTAG CAGAGTTCT CCGCATCTTA GCAACTCTTA AAGCGGGCC
    hgiAI/aspNI
    bepI286
    sauJAI
    aboI/ndaiI(dam-)
    dpmI(dam-)
    dpeI(dam-)
    eco57I
    stamI aboI(dam-)
    4701 AAGAGCGTT TCGAATGATG AGCACTTTA AGCTTCGCT ATGCGCGCG CTATTAATCC CTGATGATCC GCGCGATAG CCGCGAAGG CAACTGCGTC GCGCGATACA
    TCTTTCGAA AGCTTACTAC TCTTCAAAAT TTCAGAGCA TACAGCGCC CATATAGCG CACTACTCC GCGCGTCTC GTTAGCCAG CCGCTATAT
    maeII
    pepI406I
    xmiI
    4701 AAGAGCGTT TCGAATGATG AGCACTTTA AGCTTCGCT ATGCGCGCG CTATTAATCC CTGATGATCC GCGCGATAG CCGCGAAGG CAACTGCGTC GCGCGATACA
    TCTTTCGAA AGCTTACTAC TCTTCAAAAT TTCAGAGCA TACAGCGCC CATATAGCG CACTACTCC GCGCGTCTC GTTAGCCAG CCGCTATAT
    xsaI
    cap6I bsrI
    4801 CTATTTCTAG AATGACTTGC TTGCTACTC ACCGATGCA GAAAGCATC TTACGATGC CATGCACTA AGCAATTA GCAAGTATG GCACTGCTGC CATTAACGATG
    CAFAAGATC TTAGTGACC AACTCATGCG TCGTCTGTT CTTTCTGATG AATCGCTAG CTACTGCTC TCTTTTATA CTTCAGGAG CTATTTGATC
    haeIII/paiI
    eaeI
    cfiI
    fouIII
    aciI
    4901 AGTCATACA CTGCGGCAA CTACTTGC ACAGATGC GAGACGAAA GCGCTTTC GCGCTTTC ACAGATGC GCGCTTTC ACAGATGC GCGCTTTC ACAGATGC
    TCGATATGT GACCGCGTT GANTGAGAC TTTTCTTAGE CTGCTGCTT CTGCTGCTT CTGCTGCTT CTGCTGCTT CTGCTGCTT CTGCTGCTT CTGCTGCTT
    hgiAI/aspNI
    bepI286
    sauJAI
    aboI/ndaiI(dam-)
    dpmI(dam-)
    dpeI(dam-)
    eco57I
    stamI aboI(dam-)
    4601 CTGGTCAGG TAAAGATGC TGAAGTCAG TTTGATGCAC GAGTGGTTA CATGATGTC GATTCACAA GCGCTAAGG CCGTAAAGT TTTGCGCCCG
    GACCACTTC NTTTTCTAGC ACTTCTAGC ACCCAAGTC CTACCAANT CTAGCTTAG CAGAGTTCT CCGCATCTTA GCAACTCTTA AAGCGGGCC
    hgiAI/aspNI
    bepI286
    sauJAI
    aboI/ndaiI(dam-)
    dpmI(dam-)
    dpeI(dam-)
    eco57I
    stamI aboI(dam-)
    4701 AAGAGCGTT TCGAATGATG AGCACTTTA AGCTTCGCT ATGCGCGCG CTATTAATCC CTGATGATCC GCGCGATAG CCGCGAAGG CAACTGCGTC GCGCGATACA
    TCTTTCGAA AGCTTACTAC TCTTCAAAAT TTCAGAGCA TACAGCGCC CATATAGCG CACTACTCC GCGCGTCTC GTTAGCCAG CCGCTATAT
    xsaI
    cap6I bsrI
    4801 CTATTTCTAG AATGACTTGC TTGCTACTC ACCGATGCA GAAAGCATC TTACGATGC CATGCACTA AGCAATTA GCAAGTATG GCACTGCTGC CATTAACGATG
    CAFAAGATC TTAGTGACC AACTCATGCG TCGTCTGTT CTTTCTGATG AATCGCTAG CTACTGCTC TCTTTTATA CTTCAGGAG CTATTTGATC
    haeIII/paiI
    eaeI
    cfiI
    fouIII
    aciI
    4901 AGTCATACA CTGCGGCAA CTACTTGC ACAGATGC GAGACGAAA GCGCTTTC GCGCTTTC ACAGATGC GCGCTTTC ACAGATGC GCGCTTTC ACAGATGC
    TCGATATGT GACCGCGTT GANTGAGAC TTTTCTTAGE CTGCTGCTT CTGCTGCTT CTGCTGCTT CTGCTGCTT CTGCTGCTT CTGCTGCTT CTGCTGCTT

```

FIG. 90

```

mspI          hinfI
hpaII         msrI
bseMI        avIII/dspI  bsrI
              msrI  hhaI/cfoI  tru9I
              pspI406I  msrI
nlaIV  aluI
              msrIII  sfaMI  bbvI  fnu4HI  hinfI  hpaII  bsrI
              msrIII  sfaMI  bbvI  fnu4HI  hinfI  hpaII  bsrI
5001  ACCGTCAGT  ACCGTCAGT  ACCGTCAGT  ACCGTCAGT  ACCGTCAGT  ACCGTCAGT  ACCGTCAGT  ACCGTCAGT
TACGACCT  TACGACCT  TACGACCT  TACGACCT  TACGACCT  TACGACCT  TACGACCT  TACGACCT
              msrIII  sfaMI  bbvI  fnu4HI  hinfI  hpaII  bsrI
              msrIII  sfaMI  bbvI  fnu4HI  hinfI  hpaII  bsrI
5101  CGAATCTACT  ACTCTACT  CGGCGACA  ATTAATAGAC  TGAATGAG  GCGTAAAT  TCGAGACA  CTCTGCTC  GCGCTCTC  GCGCTCTC
CGTATGAA  TCGATCGA  GCGCTCTC  ACTACTCT  GCGTATTA  ACTCTCTC  GAAGCGAA  GCGCGAAG  GCGCGAAG
              tbaI
mspI          fndBII/wmI  haeIII/paII
hpaII         e-r10I  ssa96I
cfr10I        bctUI  ssa96I
nlaIV  hphI  bsaAI  aciI  fnu4HI  nlaIV
pauI/hpaI    bsaI  babI236I  bsvI  bsrI  ssaI  hinfI  sfaI  hpaII
              msrI
5201  TTATCTCT  ATAAATCT  ACCGTCAG  CGGCTCTC  GCGTATTA  TCGAGACA  GCGCTCTC  GCGCTCTC  GCGTATTA  GCGTATTA
AAATAGAC  TATTAGAC  TCGGAGTC  GCGCGAAG  GCGTATTA  TCGAGACA  GCGCTCTC  GCGCTCTC  GCGTATTA  GCGTATTA
              ddeI
p1eI         sau3AI  nlaIV
hinfI        mboI/mdeII(dam-)  mslI  tta9I
              dpmI(dam-)  hgiCI
eam1105I     dpmI(dam-)  benI  msrI  msrIII
              foki
5301  CAGCGAG  TCGCGACT  ATGATGAC  GAATAGAA  GATCGTAC  ATAGCTCT  CACTGATTA  GATGATTA  CTGCGACT  GCGTATTA
GCTGCTCT  ACTGCTTA  TACTACTCT  CTACTACT  CTAGGACT  TATCGAGA  CTACTAAT  CTACTAAT  GACTACTCT  TCGATGAG
              rmaI  sau3AI
              sau3AI  hphI  mboI/mdeII(dam-)
              dpmI(dam-)  dpmI(dam-)
              tru9I  dpmII(dam-)  dpmI(dam-)
              shaIII/dzeI  msrI  alwI(dam-)
              tru9I  bctVI/aboiI  bctVI/aboiI  nlaIII  msrIII
              msrI  msrI  alwI(dam-)  mboII(dam-)  rcaI  tru9I  msrI
5401  ATATATACT  TAGATGAT  TAATCTCA  TTTTATTT  AAAGATCT  AGTAAAT  CTTTTTAT  AGCTTAT  CCGAATCT  CCGAATCT
TATATAGA  ATCTACTTA  ATTTCAGT  AAAATATA  TTTTCTTA  TCCACTTA  GGAATACT  TTAGATCT  GCTTTTCT  ANTCCTCT

```





FIG. 9R

```

chaZ
fruDII/mvni
batO1
bsh12161
hinPI
hhaI/cfoI
thaX
fruDII/mvni
bctO1
bsh12361 hseIII/palI      tru9I   aloI
          eaeI   tflI aaeI/asnI/vspI   pvaII
          eaeI   ctrI   hinfI   mael   hspBII      bakI
          acII
mnlI   acII   mnlI   acII   mnlI   acII   mnlI   acII
6301  ACTGAGCGAG  GAGCGCGAG  AGCCGCCAT  AGCCAACCG  CCTCTCCCG  CCGCTTCCG  GATTCATTA  TCCAGCTGG  AGACAGGTT  TCCCGACTGG
TCACTGCCTC  CTTCGCCTC  TCCCGGCTA  TCCGCTTGC  GCGAGAGGC  GCGCAACCG  CTAACTAAT  AGCTGAGCG  TCCCTGCCA  AGCCCTAAC

          scrFI
          m-ai
          ecolII
          dsauV
          nlaIV betW1
          hgiCI apyI{dca}
          banI bsaJI      mspI
          hpaII
          nlaIV betW1
          hgiCI apyI{dca}
          banI bsaJI      mspI
          hpaII
          fru9I
          maeI
          eaeI/asnI/vspI
          xmiI
          asp700
          alwi
          nlaIII
          asp700
          alwi
          nlaIII
6501  TTGTAGCGG  ATACACAT  CACACAGGA  ACAGCTAGA  CCATGATTA  GAAATTA
          ANCWTCGCG  TATTGTTAA  GTGTGCTT  TGTGATACT  GGTACTAATG  CTAAAT

          acII
          bakBI
          maeI
          eaeI/asnI/vspI
          xmiI
          asp700
          alwi
          nlaIII
          asp700
          alwi
          nlaIII
          length: 6517
    
```



FIG. 10B

```

rsal          paeII          nlaIV
csp6I        binII/acyI    hgiCI
              ahaII/bamHI    haeII
              aeiII
401 GCTTTTGGCA GTACATCAAT GCGGTGGAT ACCGGTTTGA CTCAGGGGA TTTCCAGATC TCCACGGCA TCGGTCTCAT GCGAGTTGT TTTGGCAGCA
    CCANAACTT CATGTGTGTTA CCGCAACTA TCGCCAACT GAGTGGCTT AAGGTTTCAG AGTGGGGTA ACAGCGTTTA CCTCAAGCA AACCGTGGT

              paeII          nlaIV
              binII/acyI    hgiCI
              ahaII/bamHI    haeII
              aeiII
501 AATATCAAGG GACTTTCGA AATGCTTAA CAATCTGCC CGATTCAGC AATGGGGG TAGGGGTGA CCGTGGCAGG TCTATATAG CAGACTGCT
    TTTAGTTGC CTGAAGGTT TTACAGCATT GTTGGGGG GGTACTAGC GTTACGGC TTTACGGC ATCCGACAT GCGCCCTCC AGATATATC GTTCAGCA

rsal          paeII          nlaIV
csp6I        binII/acyI    hgiCI
              ahaII/bamHI    haeII
              aeiII
601 TTAGTCAAC GTCAGATCC CTGAGAGCC CATTCCAGCT GTTTGAGCT CCATAGAGA CACCGGACC CCGGGGGGG GAAAGCTGCA
    AATCACTTG CAGTCTAGG GACTCTGGG GTAGTGTCT GTGGCCCTGG CTAGTCTGGA GCGGGGGCC CTTCGCACT

```



FIG. 10D

```

1101 hlnpi hbel/efol mapi hpefi
    hbel hpefi
    acil mcoi
    hecili/pali bspphi
    mcri fndii/mvni bspphi
    eeg/zsaii/aci1 bsewfi
    eael betui tfii
    cfri bsh236i hlnpi
    s1ami fnd4hi bali acilii
    1101 GATCTATTG TTATCGCA CCTTCATCG CCCTCCCTC QMATICGDA MHTCTTTTMC ATTGCGGAT TCGCCBAG CCTCACTAT TCGCTCC
    CTAGCAATC AATATGCGT GAAGGTATG CCRGCGMG QETANGCT TCGCAGACTG TAACCGCTA ATCTCTTC QCATYCGAPA MHTTAGDQD
    stp1 fnd3ai acil
    fndii/mvni aboi/ndi1(dam-)
    bstci dhai dpo1(dam+) saulaj
    fnd4hi mcri hecili/pali dpo1(dam-) mboi/ndol1(dam-)
    bboi mapi bh1236i bsewfi s1ami fnd4hi dpo1(dam+)
    pcri hpefi msi e7e11i pvu1/bsp1 hecili/pali
    begi cfri01 ac5k hecl fok1 mci1 bwi cfri dpo1(dam-)
    nsp11 acii
    121 OCCUTMCA QGTUTAGE TITCAGACC TCGCTBAMC QMATICCC QCTTTTTC ACCCTCTC QMATICCC GATCCCTC GATCCCTC CTGCGGCA
    CGCMCTGT CCCAGACTC MACTTTTGT ACCTATG CTITAGCGM TCGM1236I TCGCGGAC QCTTCGTATC CTATCCGCTC
    sa196i evel1
    sa196i fcr11/cpf1
    hecili/pali acil -fii
    amu1 cpol hlnpi
    dcl acil
    bzr1
    1301 TCTFACCAQ ACGAGGGT TCGCGCTT CCGACCGCA QMATICCC QCATYCGAPA TATCTATG TACCGACTA ANATTATG QETANGCT MHTCTTMC
    AGATCCCTC TCTCGCTC ACCCGCTT CTTATCGCG CTATCGCGT TATCTATG TACCGACTA ANATTATG QETANGCT MHTCTTMC
    hlnpi hbel/efol
    hbel hpefi
    fndii/mvni
    betui
    ech1111/asp1
    drdi
    1401 TATCTG AACTGTAT QMATICCC CTGCTGCTT CCGCTCGCA QCATYCGAPA TITTCGCGCA QMATICCC QCATYCGAPA QCATYCGAC
    ATAGTACC TITCAGACTA CTTCTGTGT GAGTCAGCA QCATYCGAPA TCGACTA CTGCTTMC ANACCGCT CTTGCGCG CTGCGCG
    hlnpi
    acil
    hpc1
    bsi1
    epe1/epo1
    bali drcl1 msci1
    1401 GATCTATTG TTATCGCA CCTTCATCG CCCTCCCTC QMATICGDA MHTCTTTTMC ATTGCGGAT TCGCCBAG CCTCACTAT TCGCTCC
    CTAGCAATC AATATGCGT GAAGGTATG CCRGCGMG QETANGCT TCGCAGACTG TAACCGCTA ATCTCTTC QCATYCGAPA MHTTAGDQD
    stp1 mcri
    fnd4hi mcri hecili/pali dpo1(dam-) mboi/ndol1(dam-)
    bboi mapi bh1236i bsewfi s1ami fnd4hi dpo1(dam+)
    pcri hpefi msi e7e11i pvu1/bsp1 hecili/pali
    begi cfri01 ac5k hecl fok1 mci1 bwi cfri dpo1(dam-)
    nsp11 acii
    121 OCCUTMCA QGTUTAGE TITCAGACC TCGCTBAMC QMATICCC QCTTTTTC ACCCTCTC QMATICCC GATCCCTC GATCCCTC CTGCGGCA
    CGCMCTGT CCCAGACTC MACTTTTGT ACCTATG CTITAGCGM TCGM1236I TCGCGGAC QCTTCGTATC CTATCCGCTC
    sa196i evel1
    sa196i fcr11/cpf1
    hecili/pali acil -fii
    amu1 cpol hlnpi
    dcl acil
    bzr1
    1301 TCTFACCAQ ACGAGGGT TCGCGCTT CCGACCGCA QMATICCC QCATYCGAPA TATCTATG TACCGACTA ANATTATG QETANGCT MHTCTTMC
    AGATCCCTC TCTCGCTC ACCCGCTT CTTATCGCG CTATCGCGT TATCTATG TACCGACTA ANATTATG QETANGCT MHTCTTMC
    hlnpi hbel/efol
    hbel hpefi
    fndii/mvni
    betui
    ech1111/asp1
    drdi
    1401 TATCTG AACTGTAT QMATICCC CTGCTGCTT CCGCTCGCA QCATYCGAPA TITTCGCGCA QMATICCC QCATYCGAPA QCATYCGAC
    ATAGTACC TITCAGACTA CTTCTGTGT GAGTCAGCA QCATYCGAPA TCGACTA CTGCTTMC ANACCGCT CTTGCGCG CTGCGCG

```



**FIG. 10F**

<p>                     nlaIV                      aspi                      hpaII scfI                      bslI nciI                      xroI aspi                      bspHI hpaII                      bspEI(dam-)                      bsaWI dsav                      accIII(dam-)                      sauIAI caulI                      mboI/ndeII(dam-)                      dpmI(dam-)                      alwI(dam-)                 </p>	<p>                     heeII/paI                      scfI                      eagI/xmaIII/eclXI                      caci                      cfrI                      fru4HI                      acil                      chaI                      fru4II/mvni                      bscUI                      bebII16I seu96I                      hmgI avall                      hbaI/cfoI asuI                      acil                      reaI                      csp6I                      TOGGGAGTCC CCGGAGACTC TCGGGGCTAC ACAATATGCC CCGAGAGCC CCGCCCTCG CCGCTTCC TTTTATCCG GCGGAGAG CCGCTTCC                      TOGGTAGCA GCGTAGCCCT CCGCCCTGAC ACCCCGATG TTTTATCCG GCGCTTCC                 </p>	<p>                     scfI                      nciI                      mspI                      hpaII                      dsav                      xmaI/dspAI                      smaI                      scfI                      nciI                      dsav                      caulI                      bsaJI                      sval                      bsaJI                      neuIAI                      mboI/ndeII(dam-)                      dpmI(dam-)                      dpmII(dam-)                      alwI(dam-)                      nlaIV caulI                      batYI/zhoII                      bamHI bsaJI ecorI                      alwI(dam-) apoI                      ncri                      baII                      sfaBI                      mull                      bsaJI                      hmiI/acyI                      hgaI                      baII/bsaKI                      seu96I                      haeII/paI                      fru4II asuI                      bglI nlaIII                      sfiI styI                      eaeI rcoI                      cfrI daaI                      taqI haeIII/paII                      claI/bepID6 bsaJI                      GAGGATCC GAGGATCC GAGGATCC GAGGATCC GAGGATCC GAGGATCC GAGGATCC GAGGATCC GAGGATCC GAGGATCC                      ATCACCTTTC GCTCCGCGGT CCGTAGCAGG CCGCCCTTTC CTTATCTCAT CTAGCGTCC GCTCCCTAG GCGCTTAG TTTACTACCG GCGTACCG                 </p>
--	---	--

1801  
1901

### FIG. 10G

```

1001      aluI          #faNI      spoI      hmaI      hmaI
          fnu4HI      maeIII      #faNI      spoI      hmaI      hmaI
          bbvI        maeIII      #faNI      spoI      hmaI      hmaI
          CAACTGCTT AATTCAGCTT ATAATAGTA CATATAGAG AATAGCTCA CAATTTCAC AAATAGACA TTCTTTCAC TCCACTCTAG TTGTGCTTG
          GTTGACAAA TACGTGGAA TATACCAI GTTATTCG TTACTGAGT GTTATAGG TTTATTCGT AAATAAGTAC AGCTAGATC AACACAAAC
          sau3AI      mboI/ndeII[dm+]      spoI      hmaI      hmaI
          dpmI[dm+]      dpmII[dm+]      spoI      hmaI      hmaI
          pvuI/bspCI      mcrI      taeI[dm+]      tru9I      haeII/psaI
          clai/bspI06[dm+]      claI/bspI06[dm+]      haeI      haeI
          sau3AI      maeI      fnu4HI      styI      hmaI      hmaI
          mboI/ndeII[dm+]      mboI/ndeII[dm+]      bbyI      ncoI      hmaI      hmaI
          dpmI[dm+]      xmnI      hinPI      dsai      hmaI      hmaI
          dpmII[dm+]      asp700      aaeI/asmI/vspI      haeII      hmaI      hmaI
          nlaIII      alwI[dm+]      asp700      hbaI/cfoI      nlaIII      mliI
          101 TCCAACTCA TCAATGATC TTATCAATC TCGATGATC TCGATGATC TCGATGATC TCGATGATC TCGATGATC TCGATGATC TCGATGATC
          AGCTTAGTGT AGTACATAG AATAGTAG AGCTTAGT CCTTAATA AGCCGCTG TGTACCGA GTTATAGCA GACTTTCU TTAGACCA
          K88I      cep6I      nlaIV      hpaI01
          nlaIV      hpaI      ecorII      nsII/avaIII
          hpaI      hgiC1      hpaI      hpaI      hpaI      hpaI      hpaI      hpaI
          asp718      mliI      aluI      pvuII      hpaII      hpaI01
          acc65I      ddeI      acII      nsp8II      nsp8II      nsII/avaIII
          201 GATACCTTCT ENGCGGAAA GAACCACTG TCGATGATC TCGATGATC TCGATGATC TCGATGATC TCGATGATC TCGATGATC
          CCATGGAGA CTCGCCTT CTGTGCGAC ACCTTACAG CAGTCATCC GACACTTC AGGGTCGA GGGCTCCG GTCTTCATC GTTTCGTA
          scRFI      scrFI      nlaIV      hpaI01
          mvaI      mvaI      ecorII      nsII/avaIII      hpaI01
          dsav      dsav      hpaI      hpaI      hpaI      hpaI
          apyI[dm+]      apyI[dm+]      hpaI      hpaI      hpaI      hpaI
          aadI      hpaI      hpaI      hpaI      hpaI      hpaI
          301 ATCCANTTA GTGAGCACC AGGTGGAA AGTCCCAGG CTCGCCAG CTCGCCAG CTCGCCAG CTCGCCAG CTCGCCAG CTCGCCAG
          TAGAGTAAI EAGTCGTTG TCCACCTY TCGGGCTC GAGGGCTC GAGGGCTC GAGGGCTC GAGGGCTC GAGGGCTC GAGGGCTC
          scRFI      scrFI      nlaIV      hpaI01
          mvaI      mvaI      ecorII      nsII/avaIII      hpaI01
          dsav      dsav      hpaI      hpaI      hpaI      hpaI
          apyI[dm+]      apyI[dm+]      hpaI      hpaI      hpaI      hpaI
          aadI      hpaI      hpaI      hpaI      hpaI      hpaI
          acII      acII
    
```

FIG. 10H

```

    fnu4HI
    bglI
    sfiI
    haeIII/palI
    mnlI
    haeIII/palI
    mnlI bsaJI scII
    2401 GCCCTAACT CCCCTATCC CCCCGCTAAC TGCCGCGCAGT TCCCGCCCAT CTGCCCCCA TCCCTGCACTA ATTTTITTTA TTTATGCGAG GCGCGAGCC
    CCGGATTA GCGCGATTA GCGCGATTA GCGCGATTA ACCCGCTCA ACCGCGCTG ACCCGCTCA TAAAAAAT ATATACCTC CCGCGCCGG
    scfII
    acII
    mspI
    hpaII
    dnaV
    haeIII/palI
    mcrI
    engI/smaII/scIXI
    eaeI
    cfrI
    rpaI cauII
    haeIII/palI
    styI
    bsaJI
    bliI
    haeIII/palI
    stuI kaeI
    haeI meeI
    mnlI avrII
    mnlI
    2501 GCCTCGGCT CTGAGCTATT CCAGAGTAG TCGAGAGCT TTTTGTGAGT TCCAAAGCC TACCTATCC GCGCGGAGC GGTTCATTCG
    CGAGGCCGA GACTCGATA GCTTTCATC ACTCTCCGA AAATCTCC GATTCGAA ACTTTTTC ATCGATAGC CCGCGCTTG CCGCTTAC
    tfII
    bliII
    acII
    :hseI
    :nuDII/mvnl
    bclII
    bsh1236I
    2401 AAGCGGATT CCCCGCGCA AGAGTCAGT AAGTACGCG ATATCTGCA TATCGCGTG GCGGAGGCCA ACCAACTTTC CCGCGATTC ATTTTGTAT
    fnu4HI
    acII
    tbaI
    fnuDII/mvnl tra9I
    bclII
    bsh1236I aaeI/mnlI/wspI
    2401 AAGCGGATT CCCCGCGCA AGAGTCAGT AAGTACGCG ATATCTGCA TATCGCGTG GCGGAGGCCA ACCAACTTTC CCGCGATTC ATTTTGTAT
  
```

FIG. 10I

sau3AI  
nboI/ndelII(dam-)  
dpmI(dam+)  
dpmII(dam-)  
alwI(dam-)  
taqI(dam-)  
c1aI/bsp106(dam-)  
sau3AI  
nboI/ndelII(dam-)  
dpmI(dam+)  
dpmII(dam-)  
alwI(dam-)

fokI  
2701 ACCCTTTTGGG TGGATCTGAC ATCCACTTTT TCTTTTCTC CACAGGTGTC CACTCCCGCG TCCACTGCA CCTGGTTCG CGAGACTGCG  
TGGAAACTT AGCTAGGATG ACTGTGACTG TGGTGTGAAA AGAAGAGAG GTGTGGACAG GTGAGGATCC AGCTTGACTT GAGCCGAGCG OCTTCCACTCG

nlaIII  
styI  
pflMI  
ncoI  
dsaI

sfmI ecoRI  
fmu4HI taqI epoi bali fokI nlaIII fokI rnaI rnaI qsuI/bpmI aluI tth1111/aspI  
bbvI c1aI/bsp106 bsaJI bsaJI nlaIII fokI rnaI rnaI bsaI cap6I ecoRV nsp8II bsaI  
AACCCGAGCT AGCTAAGCTT AAGTGTGTACC CTACCGATAC ATAGTAGGAA AAGAGATCAFC GTTGAAGCTTG AACTCAATGTA AGCTTAGAG TCGACTGGGT

aluI  
sacI  
hga3II  
hgiAI/aspHI  
ec1136II  
bsp1286  
bbsHKA1  
bwyI  
banII

acII  
2801 GTCCCGGAGG TCCCTGTGCG CCTCTGTGCG CGATAGGAGTC ACCATCACTT CGCGTCCGCG TCGAGCCGTC GATTAGAGTG GTGATAGCTA CATGAACTGG  
CAGGGGCTCG AGGGACAGCC CGAGACACT OCTATGCGCG TGATAGTGA CGGACAGCTC AGCTCTCGAG CTATCTCTAC CACTATGGAT GTACTTGACC



FIG. 10K

```

seti
saci
hg12II
hg1AI/ampKI
ec1136II
hp1286
be192AI
hv1
baeII/palI
see96I alul
asm1 baqII
ec0109I/draII
hphI
meeII alwI ddaI
ec0109I/draII
GTCAAGGATC AGCGCTG
GGTGTGTGTG GCGGTGCTG TCCGAGACTC
3401 AGGACAGACAC CTACAGCTC AGCAACAGCC TCAAGCTGAG CAAGAGCAC TAGAGGAC ACAGAGCTA CCCTCCGA
TCTGTGTGTG GATGTGCGAG TGTGCTGG ACTGCACTC GTTTGTCTG ATGCTCTTG TTTTCTGAT CGGAGCCIT
GTGTTGCTGAC TCTTATTTC
3501 CTGGCCCTC ACAAGACCT TCAACAGGCG AGGTTTAA CTTCTGATG CCAGCGCC CCACTTCTT
TATTGAGCT TATTGAGCT ATATTAGCA ATATTAGCA ATATTAGCA ATATTAGCA ATATTAGCA ATATTAGCA
3601 CAATAGCATC ACAATTTCA CAATTAGAG ATTTTTC CTGCATCTA GTTTGTGTT TTGCAACTC
ATCAATCPAT CTATATGAT CTGATGCTG GTTATGTGAG TTTTAAAGT GTTATTTTC
TAAAAAAGT GAGTAAAGT GAGTAAAGT GAGTAAAGT GAGTAAAGT GAGTAAAGT
TCTGATGAC GAGTAAAGT GATTTTCTG ATTTTCTG CAGCTAAC CAGCTAAC CAGCTAAC CAGCTAAC CAGCTAAC CAGCTAAC

```

FIG. 10L

```

                                     baseII/pali
                                     baseI
trpI      stuNI styI
mseI      bwi ncoI
aseI/asnI/vspI dsal
xmiI      hicPI bseJI
asp700    hkaI/cfoI nlaIII mnlI mnlI
3701  CGCGANTTAA TTCCGCGCAG CACATGCC TGAATTAAC TCTGAAAGG GACTTGGT AGTACTTTC CTGAAAGAA AGAAGCCACT GTGAANTGG
      GCGCTTAAT AACCGGCTC GTGATACCG ACTTATGAG AGCTTCTTC CTGAAAGAA TCAAGGAAZ ACTCGGCTT TCTTGTGCA CACTTAAAC

                                     nlaIV
scrPI     sfaNI     scrPI     scrPI     nlaIV
svaI      ppvI01    svaI      svaI      mvaI
ecorIII   nsiI/aveIII   ecorIII   ecorIII
dsav      nlaIII     dsav      dsav      dsav
berNI     sphi      berNI     sphi      berNI
spyI(dcm+)  bseJI     spyI(dcm+)  bseJI     spyI(dcm+)  bseJI
bseJI     nspII     nspII     nspII     nspII
3801  TGTCAATTG GGTGTGGAA GTCCGAGC TCCCGAGC GCGAAGTAT GCGAAGCTG CACTTCAAT ATGCAAGCAAC CAGGTGGA AMGTCCCG
      AAGTCAATC GCACACTT CAGCGTCC AGCGGCTC GCTTTCATC CTTTCTATC GTAAATTA TCAATGCTG GTCCAGACT TTCAGGCTC

ppvI01
nslI/aveIII
nlaIII
sphi
nspI sfaNI
nspII
3901  GTCGCGAGC AGCGAAGT ATCGAAGCA TCCATCTCA TTACTGCA ACCATGAC CCGCTTAC TCCCGCTC CCGCTTAA CTCCGCGC
      CAGCGCTC TCGTCTTCA TACTTCTCT ACTTAAAT ATTCAAGT ATCAAGCT CCGCGATC AGCGGATC CCGCGATC GAGCGGCTC

                                     baseII/pali
                                     baseI
nlaIII
styI
ncoI
bsiZ dsal
scII bseJI
scII
4001  TTCGCTCAT TTCGCGCTT ATGCTCACT AATTTTTT ATTATGCG AGCGGAGC CCGCTTCC TCTAGCTAT TCCAGAGTA GTGAGAGC
      AAGCGGCTA AGAGCGGCG TACTTACTA TTAATAAAA TAATAGCT TCCGCTCC CCGAGCGG AGACTGATA AGCTTCTCT CACTTCTCC

```

FIG. 10M

styI                         scrPI  
 besvI                        mvAI  
                               ecoRII  
                               dsev  
                               bstRI  
                               spyl(dse-)                truBI  
                               besvI   maeIII              mseI  
                               malI   avtII  
                               aluI   hincII/hindII   bseI              maeIII   bseI  
                               aluI   hincII/hindII   bseI              maeIII   bseI  
 4101   TTTTTTRDAG   GCTTACGGCTT   TTCCAAAG   CTTATTACAG   CTTGCGACTC   GGCTGCGCTC   GGCTGGCTC   GCGAGCAGAA   ATGTTCCAGC   ACTGCAGCTT   TTGGCAACGC   ATGGCTTTCA  
                               AAAAAAAACT   CGATTCGAAA   AAGCTTTTTC   GACATTTGTC   GAGCGCTAAC   CGGCAACGAA   ATGTTCCAGC   ACTGCAGCTT   TTGGCAACGC   ATGGCTTTCA  
                               maeIII  
                               mboI/mdeII(dse-)  
                               spe6I                            qniI(dse)  
                               haeIII/palI  
                               mmlI   aciI                    qniII(dse-)  
                               mboII   asuI                    pviI/bspCI  
                               sarI/kpA32I                    mcrI  
                               aluI  
                               pvuII  
                               nap8II  
                               mboII   asuI                    pviI/bspCI  
                               sarI/kpA32I                    mcrI  
 4201   TAAITGCGCTT   GCGMCACATC   CGCCCTTCC   CAGCTTCCCT   AATGACGAGAG   AGGCGCCGC   CAGTATCGCT   TCCAGCAGCTT   TCCAGCAGCTT   TCCAGCAGCTT   TCCAGCAGCTT  
                               ATTAGCGGAA   CGTCCCTGAG   GCGGAAAGCG   GTCCAGCGCA   TTATCGCTTC   TCGCGGCTG   GCTATCGGCA   AGGCGCCGC   CAGTATCGCTT   TCCAGCAGCTT   TCCAGCAGCTT  
                               hinPI  
                               hbaI/cfoI  
                               nlaIV  
                               narI  
                               hesI  
                               hincII/acyI  
                               hpiCI  
                               hesII  
                               benI   sfarI                        aciI  
                               ahaeII/besvII  
 4301   TGCGGCTGCA   TCCGCTATT   TCTCGTTACG   CATCGTGGG   GTATTTCACA   CGCGATTCCT   GAAAGCAGCC   ATATGATCGC   CCCTGTAGCG   CCCTGTAGCG   GGCGATTACG   GGCGATTACG  
                               ACCGAGACT   ACCGCTTAAA   AAGAGTATCC   GTATGACAGC   CATATAGCTT   GGGTATCGCA   GTTTGTGGTG   TATGATGACG   GGCGATTACG  
                               fnu6HI  
                               hinPI  
                               hbaI/cfoI  
                               chaI  
                               fnu6II/mvtI  
                               bstCI  
                               beh1236I  
                               maeIII   bbvI   maeIII              mboII  
                               CGCGGCGCTT   GTGTGTGTTA   CAGCGACGCTT   GAGCGCTTACA   CTTCCAGCAG   CCGTGACGCC   CGGTGCTTCC   GTTTCTTCTC   CTTCTCTTCT   GCGGCGCTTC  
                               CGCGCGCGCA   CAGCCAGCAAT   GCGGTTCCCA   CTGCGCTGAT   CGAGGCTGGC   GGCGATTACG   GGCGATTACG   GGCGATTACG   GGCGATTACG   GGCGATTACG  
                               fnu6HI  
                               hinPI  
                               hbaI/cfoI  
                               chaI  
                               fnu6II/mvtI  
                               bstCI  
                               beh1236I  
                               maeIII   bbvI   maeIII              mboII  
                               CGCGGCGCTT   GTGTGTGTTA   CAGCGACGCTT   GAGCGCTTACA   CTTCCAGCAG   CCGTGACGCC   CGGTGCTTCC   GTTTCTTCTC   CTTCTCTTCT   GCGGCGCTTC  
                               CGCGCGCGCA   CAGCCAGCAAT   GCGGTTCCCA   CTGCGCTGAT   CGAGGCTGGC   GGCGATTACG   GGCGATTACG   GGCGATTACG   GGCGATTACG   GGCGATTACG

FIG. 10N

```

mspI      nlaIV      mslI
hpaII     hglIII    nlaIV
bspI286   bmyI      bglCI
nseI      bnlII     nlaIV      bnlI      teqI      NbsI
cfr101    cfr101    aIuI      GGGTGGCTT TAGGCTGCG ATTAGTCTC TTAGCCGAC TCGACCCGAA AAGACTGAT TCGGTGATG
CGCCGGAAG GCGTAGTTCG AGATTGAGC CCGCGGGA ATCCGAGCC TAAATCAGA AATCCCTCG ACTCGGTT TTTGACTA ACCCTACTC

aseII     haeIII/palI
dralI    sau96I
bnaI     sauI
4501 GTCACCTG TGGCCATCG CCTGATGCG GCGTTTTCG CCTTTCAG TGGATGCGA CCGTTTGA TACTGACTC TTTGTCGAAA CCGGACGAC
CAGTCGATC ACCCTGATC GCGACTATCT GCGAAGAGC GCGAAGTCC TAAAGCGCT AAGAGCGAT ACCGATTTT TACTGACTC AAGATGTTT TAAATGCGC
tbaI
fwaIII/mviI
bcoII
bahI236I
tru9I    kfu9I    tru9I    kfu9I    tru9I    kfu9I
4701 ACTCAACCT ACCTCGCT ATCTTTCA TTATAGCG ATTTGCGA TTTGCGCTA TTTGTTAAA AATGACTCA TTTAGAAA ATTTAGCGA
TCACTGCGA TCGAGCGCG TAGGAAACT AATATTCCT TAAAGCGCT AAGAGCGAT ACCGATTTT TACTGACTC AAGATGTTT TAAATGCGC
hgiI/aphI
bepI286
hai96AI
bwyI     ddeI
epaI/epI xmi
ale44I/soI cap6I
4801 AATTTACA AATATTAAC GTTACAAI TTAGCTGCGA CTTGCTGAC ATTCTCTCT GATGCGCAT AGTTAGCGA ACTGCGCTT CCGTACGCA
TTAAATGAT TTTATATTC GAAATTTTA AATGCACTC GAGTCTATG TTGAGCGA CTAAGCGTA TAAATGCTT TAAAGCGATA CCGGACTC

hlepI
hhaI/cfoI
tbaI
fwaIII/mviI
nlaIV    haeIII/palI
tru9I    kfu9I    nseI
4901 CTGCTGATG GCTGCGCTT GAGCAGCGCC AAGCAGCGCC GAGCAGCGCT GAGCAGCGCT GAGCAGCGCT GAGCAGCGCT GAGCAGCGCT
GAGCAGCGCT GAGCAGCGCT GAGCAGCGCT GAGCAGCGCT GAGCAGCGCT GAGCAGCGCT GAGCAGCGCT GAGCAGCGCT GAGCAGCGCT

```



FIG. 10P

```

5401  seuJAI      mboI/ndeII(dam-)  seuJAI      mboI/ndeII(dam-)
      qmII(dam+)    qmII(dam+)
      LaktI/afolI  qmII(dam-)      qmII(dam-)
      bari        nsp8II  alwi(dam-)    qmII(dam-)
      t8QI  alwi(dam-)  acII  baliI/ahelI  mboI/ndeII(dam-)
      AGCTTAGCT AGAGTGTGCG CGATTCTTAGG  TACTCTGAAA ACCCGGCTT CTTCGAAAGG GTTACTACTC GTAAATTT CAAAGCAGATA CACCGCCGCA
      acrfI
      ncfI
      mspI
      hpeII
      danV
      caufI
      hiniI/acyI
      hgeI
      abeli/bse8I  bcgI  mcrI  fmu4HI  acII
5501  ATTATTCGCT GATGACCGCG CGCGAGAGCGA ACTGCTGTGC CGCGATGCT ATTCTGAGAA TACTTTCGTT GACTACTGAC CAGTCTGCGA AAGCGATCTT
      TAATAGCGCA CTACTCGCG CGCTTCTCTCT TCGCGGCGCG GCTATATGTA TAAAGCTCTI ACTGAAAGCGA GACTATGATG GTGATGCTT TTTCTGTAGAA
      rpsI
      csp6I  bari
      scal  bpiI  me8II  sloni
      auu6I  auuI
      mboI/ndeII(dam-)
      qmII(dam-)
      qmII(dam-)
      pveI/bspCI
      mcrI  mliI
      seu6E
      auuII
5601  foki  nleII  fmu4HI  nleII  acII
      b6VI
      AGCGATGCGA TCACAGTACG AGAATATGCG AGTCCTGCGA TAAAGCTGCG TCGATGCT GCGCGGACT TACTTCTGAC AACCGTGA GAACCGAAGG
      TCGCTAGCT ACTGTCAITC TCTTATACG TCGAGCGGT ATTGTACTC ACTATATGTA CGCGCTTGA ATGAGACTG TTTCTGCTT CCTGCTTTC
      nleII
      seuJAI  me8II  mspI
      mboI/ndeII(dam-)  seuJAI  nleIV
      qmII(dam+)  mboI/ndeII(dam-)
      nleII(dam-)  qmII(dam-)  hpeII
      nleII(dam-)  nleII(dam-)  bse8I  aluI
5701  aluI  acII
      TTTTTCGAC AACATGCGCG ATGATGTACG TCGCTGTAT GCTTGTGAC CGAGCTTGA TGAAGCGATA TGAAGCGATA CGAAGCGCG ACCGTGAC
      TCGATTGCGG AAAAAGGCTG TTTTACCGCG TATTGCAITG ACCGAACTA GCAAGCTGCG GCTTCTGCTT ACTTCGCTT ACTTCGCTT TCGACTGCTG
  
```





FIG. 10S

```

6701 CCGCCAGCTT CCGAGCGGA GAAAGCGGA CAGGTATCG GTAGCGCA GCGTCGAC GCGTCGAC GCGTCGAC GCGTCGAC GCGTCGAC GCGTCGAC
    acil          acil          fnu4HI    hinfI  mslI  hhal/cfoI  aluI  apyI/dcm+  apyI/dcm+
    GCGTCGCA GCGCTTCCCT CTTTCCCT GTCCATPAGC CATTCGCT CCGGCTTG TCTCTCGG TCTCTCTG TCTCTCTG TCTCTCTG TCTCTCTG
    nspI  hpeII  bslI  bsaBI  fnu4HI    hinfI  mslI  hhal/cfoI  aluI  apyI/dcm+  apyI/dcm+
    mslI  ddi  hpaI  kqi  sfaHI          nlaIV  acil          fnu4III/mvnI  bshI236I
    TATCTTTTFA GTCTCTCG GTTTCCGAC CTTCTACTG AGCTCGAT TTTGATG TCTGCGGG GCGGAGCCT ATCGAAGAC GCGGAGAGG
    ATAGAAATAT CAGGACAGC CAAGCTGTA GAGACTGAC TCGACTTA AACACTG ACCAGTCCG CCGCTCGGA TACTTTTTT CCGTCTTCC
    haeIII/palI          haeIII/palI  nspI          tiiI  hinfI
    bslI          haeIII/palI  nspI          tiiI  hinfI
    haeIII/palI nlaIV haeI          haeIII/palI  haeI          acil
6901 CCGCTTTTT ACCGTTCTG GCGTTTCT GCGTTTCT TCGATGTTT TTTCTGCTT TATCTCTGAT TCTCTGAT AACCTATTA CCGCTTTTA
    CCGGAAAA TCCGAGGAC CCGAAGACG AGGTGACAG AAGAGGCA ATAGGAGCTA ATGAGAGCTA TTAGCATAT GCGGAGACT
    haeIII/palI
    scrPI
    mvaI
    ecoRII
    dsaV
    bstXI  bclI
    haeIII/palI nlaIV haeI          haeIII/palI  haeI          acil
    fnu4HI
    acil
    tbaI
    fnu4III/mvnI
    bstXI
    bshI236I

```



# ES 2 282 994 T3

## LISTA DE SECUENCIAS

### (1) INFORMACIÓN GENERAL:

- 5 (i) SOLICITANTE: GENENTECH, INC.
- (ii) TÍTULO DE LA INVENCIÓN: PROCEDIMIENTO PARA LA SELECCIÓN DE CÉLULAS HUÉSPED DE EXPRESIÓN ELEVADA
- 10 (iii) NÚMERO DE SECUENCIA: 4
- (iv) DIRECCIÓN POSTAL:
- (A) DESTINATARIO: Genentech, Inc.
  - 15 (B) CALLE: 460 Point San Bruno Blvd
  - (C) CIUDAD: South San Francisco
  - (D) ESTADO: California
  - (E) PAÍS: EUA

20 (F) ZIP: 94080
- (v) FORMA LEGIBLE POR ORDENADOR:
- 25 (A) TIPO DE MEDIO: 5,25 pulgadas, disquete de 360 Kb
  - (B) ORDENADOR: IBM PC Compatible
  - (C) SISTEMA OPERATIVO: PC-DOS/MS-DOS
  - (D) SOFTWARE: patin (Genentech)
- 30 (vi) DATOS ACTUALES DE LA SOLICITUD:
- (A) NÚMERO DE SOLICITUD:
  - (B) FECHA DE PRESENTACIÓN:

35 (C) CLASIFICACIÓN:
- (vii) DATOS ANTERIORES DE LA SOLICITUD:
- (A) NÚMERO DE SOLICITUD: 08/286740
  - 40 (B) FECHA DE PRESENTACIÓN: 05-AGOSTO-1994
- (viii) DATOS DEL CESIONARIO/AGENTE:
- (A) NOMBRE: Lee, Wendy M.
  - 45 (B) NÚMERO DE REGISTRO: 00.000
  - (C) NÚMERO DE REFERENCIA/EXPEDIENTE: 798PCT
- (ix) INFORMACIÓN DE TELECOMUNICACIONES:
- 50 (A) TELÉFONO: 415/225-1994
  - (B) TELEFAX: 415/952-9881
  - (C) TELEX: 910/371-7168

### (2) INFORMACIÓN DE LA SEC ID nº 1:

- 55 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 7360 bases
  - (B) TIPO: ácido nucleico

60 (C) CADENA: doble

  - (D) TOPOLOGÍA: lineal
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID nº 1:
- 65

ES 2 282 994 T3

TTCGAGCTCG CCCGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT 50  
TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC 100  
5 TTACGGTAAA TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCATTG 150  
ACGTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA 200  
10 TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCAC TTGGCAGTAC 250  
ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGGT 300  
15 AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAT GGGACTTTCC 350  
TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC 400  
20 GGTTTTGGCA GTACATCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA 450  
TTTCCAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA 500  
AAATCAACGG GACTTTCCAA AATGTGCTAA CAACTCCGCC CCATTGACGC 550  
25 AAATGGGCGG TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTCGT 600  
TTAGTGAACC GTCAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT 650  
30 CCATAGAAGA CACCGGGACC GATCCAGCCT CCGCGGCCCG GAACGGTGCA 700  
TTGGAACGCG GATTCCCOGT GCCAAGAGTG CTGTAAGTAC CGCCTATAGA 750  
35 GCGATAAGAG GATTTTATCC CCGCTGCCAT CATGGTTTGA CCATTGAACT 800  
GCATCGTCCG CGTGTCCCA AATATGGGGA TTGGCAAGAA CCGAGACCTA 850  
40 CCCTGCCCTC CGCTCAGGAA CCGTTCAAG TACTTCCAAA GAATGACCAC 900  
AACCTCTTCA GTGGAAGGTA AACAGAATCT GGTGATTATG GGTAGGAAAA 950  
45 CCTGGTTCTC CATTCTGAG AAGAATCGAC CTTTAAAGGA CAGAATTAAT 1000  
ATAGTTCTCA GTAGAGAACT CAAAGAACCA CCACGAGGAG CTCATTTTCT 1050  
50 TGCCAAAAGT TTGGATGATG CCTTAAGACT TATTGAACAA CCGGAATTGG 1100  
CAAGTAAAGT AGACATGGTT TGGATAGTCG GAGGCAGTTC TGTTTACCAG 1150  
55 GAAGCCATGA ATCAACCAGG CCACCTTAGA CTCTTTGTGA CAAGGATCAT 1200  
GCAGGAATTT GAAAGTGACA CGTTTTTCCC AGAATTGAT TTGGGGAAAT 1250  
ATAAACCTCT CCCAGAATAC CCAGGCGTCC TCTCTGAGGT CCAGGAGGAA 1300  
60 AAAGGCATCA AGTATAAGTT TGAAGTCTAC GAGAAGAAAG ACTAACAGGA 1350  
AGATGCTTTC AAGTTCTCTG CTCCCCTCCT AAGCTATGC ATTTTTATAA 1400  
65

ES 2 282 994 T3

GACCATGGGA CTTTGTGCTGG CTTTAGACCC CCTTGGCTTC GTTAGAACGC 1450  
5 GGCTACAATT AATACATAAC CTTATGTATC ATACACATAG ATTTAGGTGA 1500  
CACTATAGAA TAACATCCAC TTTGCCTTTC TCTCCACAGG TGTCACTCCA 1550  
10 GGTCAACTGC ACCTCGGTTT TAAGCTTGGG CTGCAGGTCG CCGTGAATTT 1600  
AAGGGACGCT GTGAAGCAAT CATGGATGCA ATGAAGAGAG GGCTCTGCTG 1650  
15 TGTGCTGCTG CTGTGTGGAG CAGTCTTCGT TTCGCCAGC CAGGAAATCC 1700  
ATGCCCGATT CAGAAGAGGA GCCAGATCTT ACCAAGTGAT CTGCAGAGAT 1750  
20 GAAAAACGC AGATGATATA CCAGCAACAT CAGTCATGGC TCGCCCTGT 1800  
GCTCAGAAGC AACCGGTGG AATATTGCTG GTGCAACAGT GGCAGGGCAC 1850  
25 AGTGCCACTC AGTGCTGTC AAAAGTTGCA GCGAGCCAAG GTGTTTCAAC 1900  
GGGGCACCT GCCAGCAGGC CCTGTACTTC TCAGATTTG TGTGCCAGTG 1950  
30 CCCCAGGA TTTGCTGGGA AGTGCTGTGA AATAGATACC AGGGCCACGT 2000  
GCTACGAGGA CCAGGGCATC AGCTACAGGG GCACGTGGAG CACAGCGGAG 2050  
35 AGTGGCGCCG AGTGCACCAA CTGGAACAGC AGCGCGTTGG CCCAGAAGCC 2100  
CTACAGCGGG CGGAGGCCAG ACGCCATCAG GCTGGGCTG GGGAAACCACA 2150  
40 ACTACTGCAG AAACCCAGAT CGAGACTCAA AGCCCTGGTG CTACGTCTTT 2200  
AAGGCGGGGA AGTACAGCTC AGAGTTCTGC AGCACCCCTG CCTGCTCTGA 2250  
45 GGGAAACAGT GACTGCTACT TTGGGAATGG GTCAGCCTAC CGTGGCACGC 2300  
ACAGCCTCAC CGAGTCGGGT GCCTCCTGCC TCCCGTGGAA TTCCATGATC 2350  
CTGATAGGCA AGGTTTACAC AGCACAGAAC CCCAGTGCCC AGGCACTGGG 2400  
55 CCTGGGCAAA CATAATTACT GCCCGAATCC TGATGGGGAT GCCAAGCCCT 2450  
GGTGCCACGT GCTGAAGAAC CGCAGGCTGA CGTGGGAGTA CTGTGATGTG 2500  
60 CCCTCCTGCT CCACCTGCGG CTTBAGACAG TACAGCCAGC CTCAGTTTCG 2550

65

ES 2 282 994 T3

CATCAAAGGA GGGCTCTTCG CCGACATCGC CTCCCACCCC TGGCAGGCTG 2600  
5 CCATCTTTGC CAAGCACAGG AGGTCGCCCG GAGAGCGGTT CCTGTCCGGG 2650  
GGCATACTCA TCAGCTCCTG CTGGATTCTC TCTGCCGCCC ACTGCTTCCA 2700  
10 GGAGAGGTTT CCGCCCCACC ACCTGACGGT GATCTTGGGC AGAACATACC 2750  
GGGTGGTCCC TGGCGAGGAG GAGCAGAAAT TTGAAGTCGA AAAATACATT 2800  
15 GTCCATAAGG AATTCGATGA TGACACTTAC GACAATGACA TTGGCTGTCT 2850  
GCAGCTGAAA TCGGATTCGT CCCGCTGTGC CCAGGAGAGC AGCGTGGTCC 2900  
20 GCACTGTGTG CCTTCCCCCG GCGGACCTGC AGCTGCCGGA CTGGACGGAG 2950  
TGTGAGCTCT CCGGCTACGG CAAGCATGAG GCCTTGTCTC CTTTCTATTG 3000  
25 GGAGCGGCTG AAGGAGGCTC ATGTCAGACT GTACCCATCC AGCCGCTGCA 3050  
CATCACAACA TTTACTTAAC AGAACAGTCA CCGACAACAT GCTGTGTGCT 3100  
30 GGAGACACTC GGAGCGGCGG GCCCCAGGCA AACTTGACCG ACGCCTGCCA 3150  
GGCGGATTCG GGAGGCCCCC TGGTGTGTCT GAACGATGGC CGCATGACTT 3200  
35 TGGTGGGCAT CATCAGCTGG GGCCTGGGCT GTGGACAGAA GGATGTCCCG 3250  
GGTGTGTACA CCAAGGTTAC CAACTACCTA GACTGGATTC GTGACAACAT 3300  
40 GCGACCGTGA CCAGGAACAC CCGACTCCTC AAAAGCAAAT GAGATCCCGC 3350  
CTCTTCTTCT TCAGAAGACA CTGCAAAGGC GCAGTGCTTC TCTACAGACT 3400  
45 TCTCCAGACC CACCACACCG CAGAAGCGGG ACGAGACCCT ACAGGAGAGG 3450  
GAAGAGTGCA TTTTCCAGA TACTTCCCAT TTTGGAAGTT TTCAGGACTT 3500  
50 GGTCTGATTT CAGGATACTC TGTCAGATGG GAAGACATGA ATGCACACTA 3550  
GCCTCTCCAG GAATGCCTCC TCCCTGGGCA GAAGTGGGGG GAATTCAATC 3600  
55 GATGGCCGCC ATGGCCCAAC TTGTTTATTG CAGCTTATAA TGGTTACAAA 3650  
60 TAAAGCAATA GCATCACAAA TTTCACAAAT AAAGCATTTT TTCACTGCA 3700

65

ES 2 282 994 T3

TTCTAGTTGT GGTTTGTCCA AACTCATCAA TGTATCTTAT CATGTCTGGA 3750  
5 TCGATCGGGA ATTAATTCCG CGCAGCACCA TGGCCTGAAA TAACCTCTGA 3800  
AAGAGGAACT TGGTTAGGTA CCTTCTGAGG CGGAAAGAAC CAGCTGTGGA 3850  
10 ATGTGTGTCA GTTAGGGTGT GGAAAGTCCC CAGGCTCCCC AGCAGGCAGA 3900  
AGTATGCAAA GCATGCATCT CAATTAGTCA GCRACCAGGT GTGGAAAGTC 3950  
15 CCCAGGCTCC CCAGCAGGCA GAAGTATGCA AAGCATGCAT CTCATTAGT 4000  
CAGCAACCAT AGTCCCGCCC CTAACTCCGC CCATCCCGCC CCTAACTCCG 4050  
20 CCCAGTCCG CCCATTCTCC GCCCATGGC TGACTAATTT TTTTATTTA 4100  
TGCAGAGGCC GAGGCCGCCT CGGCCTCTGA GCTATTCCAG AAGTAGTGAG 4150  
25 GAGGCTTTTT TGGAGGCCA GGCTTTTGCA AAAAGCTGTT AACAGCTTG 4200  
CACTGGCCGT CGTTTTACAA CGTCGTGACT GGGAAAACCC TGGCGTTACC 4250  
30 CAACTTAATC GCCTTGAGC ACATCCCCC TTCGCCAGCT GGCCTAATAG 4300  
CGAAGAGGCC CGCACCGATC GCCCTTCCA ACAGTTGCGT AGCCTGAATG 4350  
35 GCGAATGGCG CCTGATGCGG TATTTCTCC TTACGCATCT GTGCGGTATT 4400  
TCACACCGCA TACGTCAAAG CAACCATAGT ACGCGCCCTG TAGCGGCGCA 4450  
40 TTAAGCGCGG CGGGTGTGGT GGTACGCGC AGCGTGACCG CTACACTTGC 4500  
CAGCGCCCTA GCGCCCGCTC CTTTCGCTT CTTCCCTTC TTTCTCGCCA 4550  
45 CGTTCGCCGG CTTTCCCGT CAAGCTCTAA ATCGGGGGCT CCCTTAGGG 4600  
TTCCGATTTA GTGCTTTACG GCACCTCGAC CCCAAAAAAC TTGATTGGG 4650  
50 TGATGGTTCA CGTAGTGGC CATCGCCCTG ATAGACGGTT TTTGCCCCTT 4700  
55 TGACGTTGGA GTCCACGTTT TTTAATAGT GACTCTTGT CCAAAGTGA 4750  
ACAACACTCA ACCCTATCTC GGGCTAATCT TTTGATTAT AAGGGATTTT 4800  
60 GCCGATTTG GCCTATTGGT TAAAAATGA GCTGATTAA CAAAAATTA 4850

65

ES 2 282 994 T3

ACGCGAATTT TAACAAAATA TTAACGTTTA CAATTTTATG GTGCACTCTC 4900  
5 AGTACAATCT GCTCTGATGC CGCATAGTTA AGCCAACTCC GCTATCGCTA 4950  
CGTGACTIONG TCATGGCTGC GCCCCGACAC CCGCCAACAC CCGCTGACGC 5000  
10 GCCCTGACCG GCTTGTCTGC TCCCAGCATC CGCTTACAGA CAAGCTGTGA 5050  
CCGTCTCCGG GAGCTGCATG TGTCAGAGGT TTTACCCGTC ATCACCGAAA 5100  
15 CCGCGAGGC AGTATTCTTG AAGACGAAAG GGCCTCGTGA TACGCCTATT 5150  
TTTATAGGTT AATGTCATGA TAATAATGGT TTCTTAGACO TCAGGTGGCA 5200  
20 CTTTTCGGGG AAATGTGCGC GGAACCCCTA TTTGTTTATT TTTCTAAATA 5250  
CATTCAAATA TGTATCCGCT CATGAGACAA TAACCCTGAT AAATGCTTCA 5300  
25 ATAATATTGA AAAAGGAAGA GTATGAGTAT TCAACATTC CGTGTGCCCC 5350  
TTATTCCCTT TTTTGGCGCA TTTTGCCTTC CTGTTTTTGC TCACCCAGAA 5400  
30 ACCGTGGTGA AAGTAAAAGA TGCTGAAGAT CAGTTGGGTG CACGAGTGGG 5450  
TTACATCGAA CTGGATCTCA ACAGCGGTAA GATCCTTGAG AGTTTTCGCC 5500  
35 CCGAAGAACG TTTTCCAATG ATGAGCACTT TTAAGTTCT GCTATGTGGC 5550  
GCGGTATTAT CCCGTGATGA CGCCGGGCAA GAGCAACTCG GTCGCCGCAT 5600  
40 ACACTATTCT CAGAATGACT TGGTTGAGTA CTCACCAGTC ACAGAAAAGC 5650  
ATCTTACGGA TGGCATGACA GTAAGAGAAT TATGCAGTGC TGCCATAACC 5700  
45 ATGAGTGATA ACACTGCGGC CAACTTACTT CTGACAACGA TCGGAGGACC 5750  
GAAGGAGCTA ACCGCTTTTT TGCACAACAT GGGGGATCAT GTAACCTGCC 5800  
50 TTGATCGTTG GGAACCGGAG CTGAATGAAG CCATACCAA CGACGAGCGT 5850  
GACACCACGA TGCCAGCAGC AATGGCAAACA ACGTTGCGCA AACTATTAAC 5900  
55 TGGCGAACTA CTTACTCTAG CTTCCCGGCA ACAATTAATA GACTGGATGG 5950  
60 AGGCGGATAA AGTTGCAGGA CCACTTCTGC GCTCGGCCCT TCCGGCTGGC 6000

65

ES 2 282 994 T3

TGGTTTATTG CTGATAAATC TGGAGCCGGT GAGCGTGGGT CTCGCGGTAT 6050  
5 CATTGCAGCA CTGGGGCCAG ATGOTAAGCC CTCCTGATC GTAGTTATCT 6100  
ACACGACGGG GAGTCAGCA ACTATGGATG AACGAAATAG ACAGATCGCT 6150  
10 GAGATAGGTG CCTCACTGAT TAAGCATTGG TAACTGTCAG ACCAAGTTTA 6200  
CTCATATATA CTTTAGATTG ATTTAAACT TCATTTTTAA TTTAAAGGA 6250  
15 TCTAGGTGAA GATCCTTTTT GATAATCTCA TGACCAAAT CCCTTAACGT 6300  
GAGTTTCGT TCCACTGAGC GTCAGACCCC GTAGAAAAGA TCAAGGATC 6350  
20 TTCTGAGAT CCTTTTTTTC TGGCGTAAT CTGCTGCTTG CAAACAAAA 6400  
AACCACCGCT ACCAGCGGTG GTTGTTTGC CGGATCAAGA GCTACCAACT 6450  
25 CTTTTCCGA AGTAACTGG CTCAGCAGA GCGCAGATAC CAAACTGT 6500  
CCTCTAGTG TAGCCGTAGT TAGGCCACCA CTTCAAGAAC TCTGTAGCAC 6550  
30 CGCTACATA CCTGCTCTG CTAATCCTGT TACCAGTGGC TGCTGCCAGT 6600  
GGCGATAAGT CGTGTCTTAC CGGGTTGGAC TCAAGACGAT AGTTACCGGA 6650  
35 TAAGGCGCAG CGGTCGGGCT GAACGGGGG TTCGTGCACA CAGCCAGCT 6700  
TGGAGCGAAC GACCTACACC GAACTGAGAT ACCTACAGCG TGAGCATTGA 6750  
40 GAAAGCGCCA CGCTCCCGA AGGGAGAAAG GCGGACAGGT ATCCGGTAAG 6800  
CGCAGGGTC GGAACAGGAG AGCGCACGAG GGAGCTTCCA GGGGAAAACG 6850  
45 CCTGATATCT TTATAGTCTT GTCGGGTTT GCCACCTCTG ACTTGAGCGT 6900  
CGATTTTTGT GATGCTCGTC AGGGGGGCGG AGCCTATGGA AAAACGCCAG 6950  
50 CAACGCGGCC TTTTACGGT TCCTGCCCTT TTGCTGGCCT TTTGCTCACA 7000  
TGTTCTTCC TCGTTATCC CCTGATTCTG TGGATAACCG TATTACCGCC 7050  
55 TTTGAGTQAG CTGATACCGC TCGCCGAGC CGAACGACCG AGCGCAGCGA 7100  
60 GTCAGTGAGC GAGGAAGCGG AAGAGCGCCC AATACGCAA CCGCCTCTCC 7150

65

ES 2 282 994 T3

CCGCGCGTTG GCCGATTCAT TAATCCAGCT GGCACGACAG GTTCCCGAC 7200

5 TGGAAAGCGG GCAGTGAGCG CAACGCAATT AATGTGAGTT ACCTCACTCA 7250

TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT ATGTTGTGTG 7300

10 GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT 7350

TACGAATTAA 7360

15 (2) INFORMACIÓN DE LA SEC ID nº 2:

(i) CARACTERÍSTICAS DE LA SECUENCIA

(A) LONGITUD: 6889 pares de bases

20 (B) TIPO: ácido nucleico

(C) CADENA: doble

(D) TOPOLOGÍA: lineal

25 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID nº 2:

TTGAGCTCG CCCGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT 50

30 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTCCGC GTTACATAAC 100

TTACGGTAAA TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCATTG 150

35 ACGTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA 200

TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCAC TTGGCAGTAC 250

40 ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGOT 300

AAATGGCCCC CCTGGCATTG TGCCAGTAC ATGACCTTAT GGGACTTTCC 350

45 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC 400

GGTTTGGCA GTACATCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA 450

50 TTTCCAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA 500

AAATCAACGG GACTTTCCAA AATGTCGTAA CAACTCCGCC CCATTGACGC 550

55 AAATGGGCGG TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTCGT 600

TTAGTGAACC GTCAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT 650

65 CCATAGAAGA CACCGGGACC GATCCAGCCT CCGCGGCCGG GAACGGTGCA 700

ES 2 282 994 T3

TTGGAACGCG GATTCCTCGT GCCAAGAGTG CTGTAAGTAC CGCCTATAGA 750  
5 GCGATAAGAG GATTTTATCC CCGCTGCCAT CATGGTTCGA CCATTGAACT 800  
GCATCGTCGC CGTGTCCCAA AATATGGGGA TTGGCAAGAA CGGAGACCTA 850  
10 CCCTGCCCTC CGCTCAGGAA CGCGTTC AAG TACTTCCAAA GAATGACCAC 900  
AACCTCTTCA GTGGAAGGTA AACAGAATCT GGTGATTATG GGTAGGAAAA 950  
15 CCTGGTCTC CATTCTGAG AAGAATCGAC CTTTAAAGGA CAGAATTAAT 1000  
ATAGTTCTCA GTAGAGAACT CAAAGAACCA CCACGAGGAG CTCATTTTCT 1050  
20 TGCCAAAAGT TTGGATGATG CCTTAAGACT TATTGAACAA CCGGAATTGG 1100  
CAAGTAAAGT AGACATGGTT TGGATAGTCG GAGGCAGTTC TGTTTACCAG 1150  
25 GAAGCCATGA ATCAACCAGG CCACCTTAGA CTCTTTGTGA CAAGGATCAT 1200  
GCAGGAATTT GAAAGTGACA CGTTTTTCCC AGAAATTGAT TTGGGGAAAT 1250  
30 ATAAACCTCT CCCAGAATAC CCAGGCGTCC TCTCTGAGGT CCAGGAGGAA 1300  
AAAGGCATCA AGTATAAGTT TGAAGTCTAC GAGAAGAAAAG ACTAACAGGA 1350  
35 AGATGCTTTC AAGTTCTCTG CTCCCCTCCT AAAGCTATGC ATTTTTATAA 1400  
GACCATGGGA CTTTGTCTGG CTTTAGACCC CCTTGGCTTC GTTAGAACGC 1450  
40 GGCTACAATT AATACATAAC CTTATGTATC ATACACATAG ATTTAGGTGA 1500  
CACTATAGAA TAACATCCAC TTTGCCTTTC TCTCCACAGG TGCACTCCA 1550  
45 GGTCAACTGC ACCTCGGTTT TATCGATTGA ATTCCCCTGC CATAGCTGTC 1600  
TGGUATGGGC CTCTCCACCG TGCTGACCT GCTGCTGCCG CTGGTCTCC 1650  
50 TGGAGCTGTT GGTGGGAATA TACCCTCAG GGGTTATTGG ACTGGTCCCT 1700  
CACCTAGGGG ACAGGGAGAA GAGAGATAGT GTGTGTCCCC AAGGAAAATA 1750  
55 TATCCACCCT CAAAATAATT CGATTGCTG TACCAAGTGC CACAAAGGAA 1800  
60 CCTACTTGTA CAATGACTGT CCAGGCCCGG GGCAGGATAC GGACTGCAGG 1850

65

ES 2 282 994 T3

GAGTGTGAGA GCGGCTCCTT CACCGCTTCA GAAAACCACC TCAGACACTG 1900  
5 CCTCAGCTGC TCCAAATGCC GAAAGGAAAT GGGTCAGGTG GAGATCTCTT 1950  
CTTGCACAGT GGACCGGGAC ACCGTGTGTG GCTGCAGGAA GAACCCAGTAC 2000  
10 CGGCATTATT GGAGTGAAA CTTTTCCAG TGCTTCAATT GCAGCCTCTG 2050  
CCTCAATGGG ACCGTGCACC TCTCCTGCCA GGAGAAACAG AACACCGTGT 2100  
15 GCACCTGCCA TGCAGGTTTC TTTCTAAGAG AAAACGAGTG TGTCTCCTGT 2150  
AGTAACTGTA AGAAAAGCCT GGAGTGCACG AAGTTGTGCC TACCCAGAT 2200  
20 TGAGAATGTT AAGGGCACTG AGGACTCAGG CACCACAGAC AAGAGAGTTG 2250  
AGCTCAAAAC CCCACTTGGT GACACAAC TC ACACATGCCC ACGGTGCCCA 2300  
25 GAGCCCAAAT CTTGTGACAC ACCTCCCCCG TGCCACCGT GCCCAGAGCC 2350  
CAAATCTTGT GACACACCTC CCCCATGCCC ACGGTGCCCA GAGCCCAAAT 2400  
30 CTTGTGACAC ACCTCCCCCA TGCCACCGT GCCCAGCACC TGAACCTCTG 2450  
GGAGGACCGT CAGTCTTCCT CTTCCCCCA AAACCCAAGG ATACCCTTAT 2500  
35 GATTTCCCGG ACCCCTGAGG TCACGTCCGT GGTGGTGGAC GTGAGCCACG 2550  
AAGACCCCGA GGTCCAGTTC AAGTGGTACG TGGACGGCGT GGAGGTGCAT 2600  
40 AATGCCAAGA CAAAGCCCGG GGAGGAGCAG TTCAACAGCA CGTTCCGTGT 2650  
GGTCAGCGTC CTCACCGTCC TGCACCAGGA CTGGCTGAAC GGCAAGGAGT 2700  
45 ACAAGTGCAA GGTCTCCAAC AAAGCCCTCC CAGCCCCCAT CGAGAAAACC 2750  
ATCTCCAAA CCAAGGACA GCCCCGAGAA CCACAGGTGT ACACCCTGCC 2800  
50 CCCATCCCGG GAGGAGATGA CCAAGAACCA GTCAGCCTG ACCTGCCTGG 2850  
55 TCAAAGGCTT CTACCCACG GACATCGCCG TGGAGTGGGA GAGCAGCGGG 2900  
CAGCCGGAGA ACAACTACAA CACCAGCCT CCCATGCTGG ACTCCGACGG 2950  
60 CTCCTTCTC CTCTACAGCA AGCTCACCGT GGACAAGAGC AAGTGGCAGC 3000  
65

ES 2 282 994 T3

AGGGGAACAT CTTCTCATGC TCCGTGATGC ATGAGGCTCT GCACAACCCG 3050  
5 TTCACGCAGA AGAGCCTCTC CCTGTCTCCG GGTAAATGAG TGCQACGGCC 3100  
GGGGATCCTC TAGAGTCGAC CTGCAGAAGC TTGGCCGCCA TGGCCCAACT 3150  
10 TGTTTATGTC AGCTTATAAT GGTTACAAAT AAAGCAATAG CATCAQAAAT 3200  
TTCACAAATA AAGCATTITT TTCACTGCAT TCTAGTTGTG GTTTGTCCAA 3250  
15 ACTCATCAAT GTATCTTATC ATGTCTGGAT CGATCGGGAA TTAATTCGGC 3300  
GCAGCACCAT GGCCTGAAAT AACCTCTGAA AGAGGAACTT GGTTAGGTAC 3350  
20 CTTCTGAGGC GGAAGAACC AGCTGTGGAA TGTGTGTCAG TTAGGGTGTG 3400  
GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC 3450  
25 AATTAGTCAG CAACCAGGTG TGGAAAGTCC CCAGGCTCCC CAGCAGGCAG 3500  
AAGTATGCAA AGCATGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC 3550  
30 TAACTCCGCC CATCCCGCCC CTAATCCGC CCAGTCCGC CCATTCTCCG 3600  
CCCCATGGCT GACTAATTTT TTTTATTTAT GCAGAGGCCG AGGCCGCCTC 3650  
35 GGCCTCTGAG CTATTCCAGA AGTAGTGAGG AGGCTTTTTT GGAGGCCTAG 3700  
GCTTTTGCAA AAAGCTGTTA ACAGCTTGGC ACTGGCCGTC GTTTTACAAC 3750  
40 GTCGTGACTG GGAAAACCTT GGCCTTACCC AACTTAATCG CCTTGCAGCA 3800  
CATCCCCCT TCGCCAGCTG GCGTAATAGC GAAGAGGCCG GCACCGATCG 3850  
45 CCCTTCCCAA CAGTTGCGTA GCCTGAATGG CGAATGGCGC CTGATGCGGT 3900  
ATTTTCTCCT TACGCATCTG TGCGGTATTT CACACCGCAT ACGTCAAAGC 3950  
AACCATAGTA CGCGCCCTGT AGCGGCGCAT TAAGCGCGGC GGGTGTGGTG 4000  
55 GTTACGCGCA GCGTGACCGC TACACTTGCC AGCGCCCTAG CGCCCGCTCC 4050  
TTTCGTTTTT TTCCCTTCCT TTCTCGCCAC GTTCGCGGC TTTCCCGTC 4100  
60 AAGCTCTAAA TCGGGGGCTC CCTTTAGGGT TCCGATTTAG TGCTTTACGG 4150

65

ES 2 282 994 T3

CACCTOGACC CCAAAAAACT TGATTTGGGT GATGGTTCAC GTAGTGGGCC 4200  
5 ATCGCCCTGA TAGACGGTTT TTCGCCCTTT GACGTTGGAG TCCACGTTCT 4250  
TTAATAGTGG ACTCTTGTTT CAAACTGGAA CAACACTCAA CCCTATCTCG 4300  
10 GGCTATTCTT TTGATTTATA AGGGATTTTG CCGATTTCCG CCTATTGGTT 4350  
AAAAAATGAG CTGATTTAAC AAAAATTTAA CGCGAATTTT AACAAAATAT 4400  
15 TAACGTTTAC AATTTTATGG TGCACTCTCA GTACAATCTG CTCTGATGCC 4450  
GCATAGTTAA GCCAACTCCG CTATCGCTAC GTGACTGGGT CATGGCTCGG 4500  
20 CCCCACACC CGCCAACACC CGCTGACGCG CCCTGACGGG CTTGTCTGCT 4550  
CCCGGCATCC GCTTACAGAC AAGCTGTGAC CGTCTCCGGG AGCTGCATGT 4600  
25 GTCAGAGGTT TTCACCGTCA TCACCGAAC GCGCGAGGCA GTATTCTTGA 4650  
AGACGAAGG GCCTCTGAT ACGCCTATTT TTATAGTTA ATGTCATGAT 4700  
30 AATAATGGTT TCTTAGACGT CAGGTGGCAC TTTTCGGGGA AATGTGCGCG 4750  
GAACCCCTAT TTGTTTATTT TTCTAATAC ATTCAAATAT GTATCCGCTC 4800  
35 ATGAGACAAT AACCCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG 4850  
TATGAGTATT CAACATTTCC GTGTGCCCCT TATTCCTTT TTTGCGGAT 4900  
40 TTTGCCCTCC TGTTTTTGCT CACCCAGAAA CGCTGGTGAA AGTAAAAGAT 4950  
GCTGAAGATC AGTTGGGTGC ACGAGTGGT TACATCGAAC TGGATCTCAA 5000  
45 CAGCGTAAG ATCCTTGAGA GTTTTGCCC CGAAGAACGT TTTCCAATGA 5050  
TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTGATGAC 5100  
50 GCCGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT 5150  
GGTTGAGTAC TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG 5200  
55 TAAGAGAATT ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC 5250  
60 AACTTACTTC TGACAACGAT CGGAGGACCG AAGGAGCTAA CCGCTTTTTT 5300  
65

ES 2 282 994 T3

GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG GAACCGGAGC 5350  
5 TGAATGAAGC CATAACAAAC GACGAGCGTG ACACCACGAT GCCAGCAGCA 5400  
ATGGCAACAA CGTTGCGCAA ACTATTA ACT GCGGA ACTAC T TACTCTAGC 5450  
10 TTCCCGGCAA CAATTAATAG ACTGGATGGA GCGCGATAAA GTTGCAGGAC 5500  
CACTTCTGCG CTCGGCCCTT CCGGCTGGCT GGT T TATTGC TGATAAATCT 5550  
15 GGAGCCGGTG AGCGTGGGTC TCGCGTATC ATTGCAGCAC TGGGGCCAGA 5600  
TGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGG AGTCAGGCAA 5650  
20 CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC CTCACTGATT 5700  
AAGCATTGGT AACTGTCAGA CCAAGTTTAC TCATATATAC TTFAGATTGA 5750  
25 TTTAAACTT CATTTTTAAT TAAAAGGAT CTAGGTGAAG ATCCTTTTTG 5800  
ATAATCTCAT GACCAAATC CCTTAACGTG AGTTTTCGTT CCACTGAGCG 5850  
30 TCAGACCCCG TAGAAAAGAT CAAAGGATCT TCTTGAGATC CTTTTTTTCT 5900  
GCGCGTAATC TGCTGCTTGC AAACAAAAAA ACCACCGCTA CCAGCGGTGG 5950  
35 TTTGTTTGCC GGATCAAGAG CTACCAACTC TTTTCCGAA GGTA ACTGGC 6000  
TTCAGCAGAG CGCAGATACC AAATACTGTC CTTCTAGTGT AGCCGTAGTT 6050  
40 AGGCCACCAC TTCAAGAACT CTGTAGCACC GCCTACATAC CTCGCTCTGC 6100  
TAATCCTGTT ACCAGTGGCT GCTGCCAGTG GCGATAAGTC GTGTCTTACC 6150  
45 GGGTTGGACT CAAGACGATA GTTACCGGAT AAGCGCAGC GGTCCGGCTG 6200  
AACGGGGGGT TCGTGACAC ACCCAGCTT GGAGCGAACG ACCTACACCG 6250  
50 AACTGAGATA CCTACAGCGT GAGCATTGAG AAACCGCCAC GCTTCCCGAA 6300  
GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTCG GAACAGGAGA 6350  
55 GCGCACGAGG GAGCTTCCAG GGGGAAACGC CTGGTATCTT TATAGTCCTG 6400  
60 TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTTGTG ATGCTCGTCA 6450

65

ES 2 282 994 T3

GGGGGGCGGA GCCTATGGAA AAACGCCAGC AACGCGCCT TTTTACGGTT 6500  
5 CCTGGCCTTT TGCTGGCCTT TTGCTCACAT GTTCTTTCTT GCGTTATCCC 6550  
CTGATTCTGT GGATAACCGT ATTACCGCCT TTGAGTGAGC TGATACCGCT 6600  
10 CCGCGCAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA 6650  
AGAGCGCCCA ATACGCAAAC CGCCTCTCCC CGCGCGTTGG CCGATTCAIT 6700  
15 AATCCAGCTG GCACGACAGG TTTCCCGACT GGAAAGCGGG CAGTGAGCGC 6750  
AACGCAATTA ATGTGAGTTA CCTCACTCAT TAGGCACCCC AGGCTTTACA 6800  
20 CTTTATGCTT CCGGCTCGTA TGTGTGTGG AATTGTGAGC GGATAACAAT 6850  
TTCACACAGG AAACAGCTAT GACCATGATT ACGAATTAA 6889  
25

(2) INFORMACIÓN DE LA SEC ID nº 3:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 6557 pares de bases  
(B) TIPO: ácido nucleico  
(C) CADENA: doble  
(D) TOPOLOGÍA: lineal

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID nº 3:

30 TTCGAGCTCG CCCGACATTG ATTATTGACT AGAGTCGATC GACAGCTGTG 50  
GAATGTGTGT CAGTTAGGGT GTGGAAAGTC CCCAGGCTCC CCAGCAGGCA 100  
45 GAAGTATGCA AAGCATGCAT CTCAATTAGT CAGCAACCAG GTGTGGAAAG 150  
TCCCCAGGCT CCCGAGCAGG CAGAAGTATG CAAAGCATGC ATCTCAATTA 200  
50 GTCAGCAACC ATAGTCCCGC CCTAACTCC GCCCATCCCG CCCCTAACTC 250  
CGCCCAGTTC CGCCCATCTT CCGCCCCATG GCTGACTAAT TTTTTTTATT 300  
55 TATGCAGAGG CCGAGGCCGC CTCGGCCTCT GAGCTATTCC AGAAGTAGTG 350  
60 AGGAGGCTTT TTTGGAGGCC TAGGCTTTTG CAAAAGCTA GCTTATCCGG 400  
CCGGGAACGG TGCAATTGGAA CCGGATTCC CCGTCCCAAG AGTGACGTAA 450  
65 GTACCGCCTA TAGAGCGATA AGAGGATTTT ATCCCCGCTG CCATCATGGT 500

ES 2 282 994 T3

TCGACCATTG AACTGCATCG TCGCCGTOTC CCAAATATG GGGATTGGCA 550  
5 AGAACGGAGA CCTACCCTGG CCTCCGCTCA GGAACGAGTT CAAGTACTTC 600  
CAAAGAATGA CCACAACCTC TTCAGTGGAA GGTAAACAGA ATCTGGTGAT 650  
10 TATGGGTAGG AAAACCTGGT TCTCCATTCC TGAGAAGAAT CGACCTTTAA 700  
AGGACAGAAT TAATATAGTT CTCAGTAGAG AACTCAAAGA ACCACCACGA 750  
15 GGAGCTCATT TTCTTGCCAA AAGTTTGGAT GATGCCTTAA GACTTATTGA 800  
ACAACCGGAA TTGGCAAGTA AAGTAGACAT GGTTTGGATA GTCGGAGGCA 850  
20 GTTCTGTTTA CCAGGAAGCC ATGAATCAAC CAGGCCACCT TAGACTCTTT 900  
GTGACAAGGA TCATGCAGGA ATTTGAAAGT GACACGTTTT TCCCAGAAAT 950  
25 TGATTTGGGG AAATATAAAC CTCTCCAGA ATACCCAGGC GTCCTCTCTG 1000  
AGGTCCAGGA GGAAAAGGC ATCAAGTATA AGTTTGAAGT CTACGAGAAG 1050  
30 AAGACTAAC AGGAAGATGC TTCAAGTTC TCTGCTCCCC TCCTAAAGCT 1100  
ATGCATTTTT ATAAGACCAT GGGACTTTTG CTGGCTTTAG ATCCCCTTGG 1150  
35 CTTCTTTAGA ACGCAGCTAC AATTAATACA TAACCTTATG TATCATAAC 1200  
ATACGATTTA GGTGACACTA TAGATAACAT CCACTTTGCC TTTCTCTCCA 1250  
40 CAGGTGTCCA CTCCCAGGTC CAACTGCACC TCGGTTCTAT CGATTGAATT 1300  
CCACCATGGG ATGGTCATGT ATCATCCTTT TTCTAGTAGC AACTGCAACT 1350  
45 GGAGTACATT CAGAAGTTCA GCTGGTGGAG TCTGGCGGTG GCCTGGTGCA 1400  
GCCAGGGGGC TCACTCCGTT TGTCTGTGC AGTTTCTGGC TACTCCATCA 1450  
50 CCTCCGGATA TAGCTGGAAC TGGATCCGTC AGGCCCCGGG TAAGGGCCTG 1500  
GAATGGGTTG CATCGATTAC GTATGCCGGA TCGACTAACT ATAACCCTAG 1550  
55 CGTCAAGGGC CGTATCACTA TAAGTCGCGA CGATTCCAAA AACACATTCT 1600  
60 ACCTGCAGAT GAACAGCCTG CGTGCTGAGG ACACTGCCGT CTATTATTGT 1650

65

ES 2 282 994 T3

GCTCGAGGCA GCCACTATTT CGGCGCCTGG CACTTCGCGG TGTGGGGTCA 1700  
AGGAACCCTG GTCACCCTCT CCTCGGCTC CACCAAGGGC CCATCGGTCT 1750  
TCCCCCTGGC ACCCTCCTCC AAGAGCACCT CTGGGGGCAC AGCGGGCCCTG 1800  
GGCTGCCTGG TCAAGGACTA CTCCCCGAA CCGGTGACGG TGTGTTGAA 1850  
CTCAGGCGCC CTGACCAGCG GCGTGCACAC CTCCCCGGCT GTCTACAGT 1900  
CCTCAGGACT CTACTCCCTC AGCAGCGTGG TGA CTGTGCC CTCTAGCAGC 1950  
TTGGGCACCC AGACCTACAT CTGCAACGTG AATCACAAGC CCAGCAACAC 2000  
CAAGGTGGAC AAGAAAGTTG AGCCCAAATC TTGTGACAAA ACTCACACAT 2050  
GCCACCGTG CCCAGCACCT GAATCCTTG GGGGACCGTC AGTCTTCTCT 2100  
TTCCCCCAA AACCCAAGGA CACCCTCATG ATCTCCCGA CCGCTGAGGT 2150  
CACATGCGTG GTGGTGGACG TGAGCCACGA AGACCCTGAG GTCAAGTTCA 2200  
ACTGGTACGT GGACGGCGTG GAGGTGCATA ATGCCAAGAC AAGCCGCGG 2250  
GAGGAGCAGT ACAACAGCAC GTACCCTGTG GTCAGCGTCC TCACCCTCT 2300  
GCACCAGGAC TGGCTGAATG GCAAGGAGTA CAAGTGCAG GTCTCCAACA 2350  
AAGCCCTCCC AGCCCCATC GAGAAAACCA TCTCCAAGC CAAAGGGCAG 2400  
CCCCGAGAAC CACAGGTGTA CACCCTGCC CCATCCCGG AAGAGATGAC 2450  
CAAGAACCAG GTCAGCTGA CCTGCCTGT CAAAGGCTC TATCCAGCG 2500  
ACATCGCGT GGAGTGGAG AGCAATGGC AGCCGGAGAA CAACTACAAG 2550  
ACCACGCTC CCGTGTGGA CTCCGACGGC TCCTTCTTC TCTACAGCAA 2600  
GCTCACCGT GACAAGAGCA GGTGGCAGCA GGGGAACGTC TTCTCATGCT 2650  
CCGTGATGCA TGAGGCTCTG CACAACCACT ACACGCAGAA GAGCCTCTCC 2700  
CTGTCTCCGG GTAAATGAGT GCGACGGCCC TAGAGTGGAC CTGCAGAAGC 2750  
TTGGCCGCCA TGGCCCAACT TGTTTATTGC AGCTTATAAT GGTTACAAAT 2800

ES 2 282 994 T3

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

AAAAGCAATAG CATCACAAAT TTCACAAATA AAGCATTITT TTCACTGCAT 2850  
TCTAGTTGTG GTTTGTCCAA ACTCATCAAT GTATCTTATC ATGTCTGGAT 2900  
CGATCGGGAA TTAATTCGGC GCAGCACCAT GGCCTGAAAT AACCTCTGAA 2950  
AGAGGAACTT GGTTAGGTAC CTTCTGAGGC GGAAGAACC AGCTGTGGAA 3000  
TGTGTGTCAG TTAGGGTGTG GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA 3050  
GTATGCAAAG CATGCATCTC AATTAGTCAG CAACCAGGTG TGGAAAGTCC 3100  
CCAGGCTCCC CAGCAGGCAG AAGTATGCAA AGCATGCATC TCAATTAGTC 3150  
AGCAACCATA GTCCCGCCCC TAACTCCGCC CATCCCGCCC CTAACTCCGC 3200  
CCAGTCCGC CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTTAT 3250  
GCAGAGGCCG AGGCCGCCCTC GGCCTCTGAG CTATTCCAGA AGTAGTGAGG 3300  
AGGCTTTTTT GGAGGCCTAG GCTTTTGCAA AAAGCTGTTA CCTCGAGCGG 3350  
CCGCTTAATT AAGGCGGCC ATTTAAATCC TGCAGGTAAC AGCTTGGCAC 3400  
TGGCCGTCGT TITACAACGT CGTGACTGGG AAAACCCTGG CGTTACCCAA 3450  
CTTAATCGCC TTGCAGCACA TCCCCCCTC GCCAGCTGGC GTAATAGCGA 3500  
AGAGGCCCGC ACCGATCGCC CTTCCCAACA GTTGGGTAGC CTGAATGGCG 3550  
AATGGCGCCT GATGCGGTAT TTTCTCCTTA CGCATCTGTG CCGTATTTCA 3600  
CACCGCATA C GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGCGCATT 3650  
AGCGCGCGG GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTTGCCAG 3700  
CGCCCTAGCG CCGCTCCTT TCGCTTCTT CCCTTCCTTT CTCGCCACGT 3750  
TCGCCGGCTT TCCCCGTCAA GCTCTAAATC GGGGGCTCCC TTTAGGGTTC 3800  
CGATTTAGTG CTTTACGGCA CCTCGACCCC AAAAACTTG ATTTGGGTGA 3850  
TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA 3900  
CGTTGGAGTC CACGTTCTT AATAGTGGAC TCTTGTCCA AACTGGAACA 3950

ES 2 282 994 T3

ACACTCAACC CTATCTCGGG CTATTCTTTT GATTTATAAG GGATTTTGCC 4000  
GATTTGCGCC TATTGGTTAA AAAATGAGCT GATTTAACAA AAATTTAACG 4050  
CGAATTTTAA CAAAATATTA ACGTTTACAA TTTTATGGTG CACTCTCAGT 4100  
ACAATCTGCT CTGATGCCGC ATAGTTAAGC CAACTCCGCT ATCGCTACGT 4150  
GACTGGGTCA TGGCTGCGCC CCGACACCCG CCAACACCCG CTGACGCGCC 4200  
CTGACGGGCT TGTCTGCTCC CGGCATCCGC TTACAGACAA GCTGTGACCG 4250  
TCGCCGGGAG CTGCATGTGT CAGAGTTTTT CACCGTCATC ACCGAAACGC 4300  
GGGAGGCAGT ATTCTTGAAG ACGAAAGGGC CTCGTGATAC GCCTATTTTT 4350  
ATAGGTTAAT GTCATGATAA TAATGGTTTC TTAGACGTCA GGTGGCACTT 4400  
TTCGGGAAA TGTGCGCGGA ACCCCTATTT GTTTATTTTT CTAATACAT 4450  
TCAAATATGT ATCGGCTCAT GAGACAATAA CCCTGATAAA TGCTTCAATA 4500  
ATATTGAAA AGGAAGAGTA TGAGTATTCA ACATTTCCGT GTCGCCCTTA 4550  
TTCCCTTTTT TCGGCATTT TGCCTTCTG TTTTGTCTCA CCCAGAAACG 4600  
CTGGTGAAG TAAAAGATGC TGAAGATCAG TTGGGTGCAC GAGTGGGTTA 4650  
CATCGAACTG GATCTCAACA GCGGTAAGAT CCTTGAGAGT TTTCGCCCCG 4700  
AAGAACGTTT TCCAATGATG AGCACTTTTA AAGTTCTGCT ATGTGGCGCG 4750  
GTATTATCCC GTGATGACGC CGGGCAAGAG CAACTCGGTC GCCGCATACA 4800  
CTATTCTCAG AATGACTTGG TTGAGTACTC ACCAGTCACA GAAAAGCATC 4850  
TTACGGATGG CATGACAGTA AGAGAATTAT GCAGTGCTGC CATAACCATG 4900  
AGTGATAACA CTGCGGCCAA CTTACTTCTG ACAACGATCG GAGGACCGAA 4950  
GGAGCTAACC GCTTTTTTGC ACAACATGGG GGATCATGTA ACTCGCCTTG 5000  
ATCGTTGGGA ACCGGAGCTG AATGAAGCCA TACCAAACGA CGAGCGTGAC 5050  
ACCACGATGC CAGCAGCAAT GGCAACAACG TTGCGCAAAC TATTAECTGG 5100

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

ES 2 282 994 T3

CGAACTACTT ACTCTAGCTT CCCGCCAACA ATTAATAGAC TGGATGGAGG 5150  
5  
CGGATAAAGT TGCAGGACCA CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG 5200  
TTTATTGCTG ATAAATCTGG AGCCGGTGAG CGTGGGTCTC GCGGTATCAT 5250  
10  
TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA GTTATCTACA 5300  
CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG 5350  
15  
ATAGGTGCCT CACTGATTAA GCATTGGTAA CTGTCAGACC AAGTTTACTC 5400  
ATATATACTT TAGATTGATT TAAACTTCA TTTTAAATT AAAAGGATCT 5450  
20  
AGGTGAAGAT CCTTTTGAT AATCTCATGA CCAAATCCC TTAACGTGAG 5500  
TTTTCGTTCC ACTGAGCGTC AGACCCCGTA GAAAAGATCA AAGGATCTTC 5550  
25  
TTGAGATCCT TTTTTCTGC GCGTAATCTG CTGCTTGCAA ACAAAAAAAC 5600  
CACCGCTACC AGCGGTGTT TGTTCGCGG ATCAAGAGCT ACCAACTCTT 5650  
30  
TTCCGAAGG TAACTGGCTT CAGCAGAGCG CAGATACCAA ATACTGTCTT 5700  
TCTAGTGTAG CCGTAGTTAG GCCACCACTT CAAGAACTCT GTAGCACCGC 5750  
35  
CTACATACCT CGCTCTGCTA ATCCTGTTAC CAGTGGCTGC TGCCAGTGGC 5800  
GATAAGTCGT GTCTTACCGG GTTGGACTCA AGACGATAGT TACCGGATAA 5850  
40  
GGCGCAGCGG TCGGGCTGAA CGGGGGTTC GTGCACACAG CCCAGCTTGG 5900  
AGCGAACGAC CTACACCGAA CTGAGATACC TACAGCGTGA GCATTGAGAA 5950  
45  
AGCGCCACGC TTCCGAAGG GAGAAAGCG GACAGGTATC CGGTAAGCGG 6000  
CAGGGTCGGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG GGAAACGCCT 6050  
50  
GGTATCTTTA TAGTCTGTG GGGTTTCGCC ACCTCTGACT TGAGCGTCGA 6100  
TTTTGTGAT GCTCGTCAGG GGGGCGGAGC CTATGGAAAA ACGCCAGCAA 6150  
55  
CGCGGCCTTT TTACGGTTCC TGGCCTTTTG CTGGCCTTTT GCTCACATGT 6200  
60  
TCTTCTGTC GTTATCCCCT GATTCTGTGG ATAACCGTAT TACCGCCTTT 6250  
65

ES 2 282 994 T3

GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC GCAGCGAGTC 6300

5 AGTGAGCGAG GAAGCGGAAG AGCGCCCAAT ACGCAAACCG CCTCTCCCCG 6350

CGCGTTGGCC GATTCATTAA TCCAGCTGGC ACGACAGGTT TCCCGACTGG 6400

10 AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTACC TCACTCATTAA 6450

GGCACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA 6500

15 TTGTGAGCGG ATAACAATT CACACAGGAA ACAGCTATGA CCATGATTAC 6550

GAATTAA 6557

20

(2) INFORMACIÓN DE LA SEC ID nº 4:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

25

(A) LONGITUD: 7305 pares de bases

(B) TIPO: ácido nucleico

(C) CADENA: doble

30

(D) TOPOLOGÍA: lineal

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID nº 4:

35 TTCGAGCTCG CCCGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT 50

TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTCCGC GTTACATAAC 100

40 TTACGGTAAA TGGCCCGCCT GGCTGACCGC CCAACGACCC CGGCCATTG 150

ACGTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA 200

45

TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCAC TTGGCAGTAC 250

ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGGT 300

50

AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAT GGGACTTTCC 350

55

TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC 400

GGTTTTGGCA GTACATCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA 450

60

TTTCCAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA 500

AAATCAACGG GACTTTCCAA AATGTCGTAA CAACTCCGCC CCATTGACGC 550

65

AAATGGGCGG TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTCGT 600

ES 2 282 994 T3

TTAGTGAACC GTCAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT 650  
5 CCATAGAAGA CACCGGACCC GATCCAGCCT CCGCGGCCCG GAACGGTGCA 700  
TTGGAACGCG GATTCCCCGT GCCAAGAGTG ACGTAAGTAC CGCCTATAGA 750  
10 GTCTATAGGC CCACCCCTT GGCTTCGTTA GAACGCGGCT ACAATTAATA 800  
CATAACCTTA TGTATCATAC ACATACGATT TAGGTGACAC TATAGAATAA 850  
15 CATCCACTTT GCCTTCTCT CCACAGGTGT CCACTCCCAG GTCCAACTGC 900  
ACCTCGGTTT TAAGCTTATC GATATGAAA AGCCTGAACT CACCGCGACG 950  
20 TCTGTGAGA AGTTTCTGAT CGAAAAGTTC GACAGCGTCT CCGACCTGAT 1000  
GCAGCTCTCG GAGGCGAAG AATCTCGTGC TTTCAGCTTC GATGTAGGAG 1050  
25 GCGGTGGATA TGCTCTCGG GTAAATAGCT GCGCCGATGG TTTCTACAAA 1100  
GATCGTTATG TTTATCGGCA CTTTGCATCG GCCGCGCTCC CGATTCCGGA 1150  
30 AGTGCTTGAC ATTGGGGAAT TCAGCGAGAG CCTGACCTAT TGCATCTCCC 1200  
GCCGTGCACA GGGTGTACG TTGCAACACC TGCTGAAAC CGAACTGCCC 1250  
35 GCTGTCTGTC AGCCGGTTCG GGAGGCCATG GATGCGATCG CTGCGGCCGA 1300  
TCTTAGCCAG ACGAGCGGGT TCGGCCATT CGGACCGCAA GGAATCGGTC 1350  
40 AATACACTAC ATGGCGTGAT TTCATATGCG CGATTGCTGA TCCCATGTG 1400  
TATCACTGGC AACTGTGAT GGACGACACC GTCAGTCCGT CCGTCGCGCA 1450  
45 GGCTCTCGAT GAGCTGATGC TTTGGGCCGA GGAAGTCCCGC 1500  
ACCTCGTGCA CGCGGATTTT GGCTCCAACA ATGTCCTGAC GGACAATGGC 1550  
50 CGCATAACAG CGGTCATTGA CTGGAGCGAG GCGATGTTG GGGATTCCCA 1600  
ATACGAGGTC GCCAACATCT TCTTCTGGAG GCCGTGGTTG GCTTGTATGG 1650  
55 AGCAGCAGAC GACTTTCGAG CGGAGGCATC CGGAGCTTGC AGGATCGCCG 1700  
60 CCGCTCCGGG CGTATATGCT CCGCATTGGT CTTGACCAAC TCTATCAGAG 1750

65

ES 2 282 994 T3

CTTGGTTGAC GGCAATTTCC ATGATGCAGC TTGGCCGCAG GGTCCGATCCG 1800  
5 ACGCAATCGT CCGATCCGGA GCCGGGACTG TCGGGCGTAC ACAAATCGCC 1850  
CGCAGAAGCG CGGCCGTCTG GACCGATGGC TGTGTAGAAG TACTCGCCGA 1900  
10 TAGTGGAAAC CGACGCCCCA GCACTCGTCC GAGGGCAAAG GAATAGAGTA 1950  
GATGCCGACC GAAGGATCCC CGGGGAATTC AATCGATGGC CGCCATGGCC 2000  
15 CAACTTGTTT ATTGCAGCTT ATAATGGTTA CAAATAAAGC AATAGCATCA 2050  
CAAATTTTAC AAATAAAGCA TTTTTTTCAC TGCATTCTAG TTGTGGTTTTG 2100  
20 TCCAAACTCA TCAATGTATC TTATCATGTC TGGATCGATC GGGAAATTAAT 2150  
TCGGCCGAGC ACCATGGCCT GAAATAACCT CTGAAAGAGG AACTTGTTTA 2200  
25 GGTACCTTCT GAGGCCGAAA GAACCAGCTG TGGAATGTGT GTCAGTTAGG 2250  
GTGTGGAAAG TCCCCAGGCT CCCGAGCAGG CAGAAGTATG CAAAGCATGC 2300  
30 ATCTCAATTA GTCAGCAACC AGGTGTGGAA AATCCCCAGG CTCCCCAGCA 2350  
GGCAGAAGTA TGCAAAGCAT GCATCTCAAT TAGTCAGCAA CCATAGTCCC 2400  
35 GCCCCTAACT CCGCCCATCC CGCCCTAAC TCCGCCAGT TCCGCCATT 2450  
CTCCGCCCCA TGGCTGACTA ATTTTTTTTA TTTATGCAGA GGCCGAGGCC 2500  
40 GCCTCGGCCT CTGAGCTATT CCAGAAGTAG TGAGGAGGCT TTTTGGAGG 2550  
CCTAGGCTTT TGCAAAAAGC TAGCTTATCC GGCCGGGAAC GGTGCATTGG 2600  
45 AACGCGGATT CCCCCTGCCA AGAGTCAGGT AAGTACCGCC TATAGAGTCT 2650  
ATAGGCCCCAC CCCCTTGGCT TCGTTAGAAC GCGGCTACAA TTAATACATA 2700  
50 ACCTTTTGGG TCGATCCTAC TGACACTGAC ATCCACTTTT TCTTTTTCTC 2750  
CACAGGTGTC CACTCCCAGG TCCAAC TGCA CCTCGGTTCC CGAAGCTAGC 2800  
55 TTGGGCTGCA TCGATTGAAT TCCACCATGG GATGGTCATG TATCATCCTT 2850  
60 TTTCTAGTAG CAACTGCAAC TGGAGTACAT TCAGATATCC AGCTGACCCA 2900

65

ES 2 282 994 T3

GTCCCCGAGC TCCCTGTCCG CCTCTGTGGG CGATAGGGTC ACCATCACCT 2950  
5 GCCGTGCCAG TCAGAGCGTC GATTACGATG GTGATAGCTA CATGAACTGG 3000  
TATCAACAGA AACCAGGAAA AGCTCCGAAA CTA CTGATT TACGCGGCCTC 3050  
10 GTACCTGGAG TCTGGAGTCC CTCTCGCTT CTCTGGATCC GGTTCCTGGGA 3100  
CGATTTCAC TCTGACCATC AGCAGTCTGC AGCCCGAAGA CTTCGCAACT 3150  
15 TATTACTGTC AGCAAAGTCA CGAGGATCCG TACACATTTG GACAGGGTAC 3200  
CAAGGTGGAG ATCAAACGAA CTGTGGCTGC ACCATCTGTC TTCATCTTCC 3250  
20 CGCCATCTGA TGAGCAGTTG AAATCTGGAA CTGCCTCTGT TGTGTGCCCTG 3300  
CTGAATAACT TCTATCCCAG AGAGGCCAAA GTACAGTGGA AGGTGGATAA 3350  
25 CGCCCTCCAA TCGGGTAACT CCCAGGAGAG TGTCACAGAG CAGGACAGCA 3400  
AGGACAGCAC CTACAGCCTC AGCAGCACCC TGACGCTGAG CAAAGCAGAC 3450  
30 TAGGAGAAAC ACAAAAGTCTA CGCCTGCGAA GTCACCCATC AGGGCCCTGAG 3500  
CTCGCCCGTC ACAAAAGAGCT TCAACAGGGG AGAGTGTTAA GCTTCGATGG 3550  
35 CCGCCATGGC CCAACTTGTT TATTGCAGCT TATAATGGTT ACAATAAAG 3600  
CAATAGCATC ACAAAATTCA CAAATAAAGC ATTTTTTTCA CTGCATTCTA 3650  
40 GTTGTGGTTT GTCCAAACTC ATCAATGTAT CTTATCATGT CTGGATCGAT 3700  
CGGGAATTAA TTCGGCGCAG CACCATGGCC TGAATAACC TCTGAAAGAG 3750  
45 GAACTTGGTT AGGTACCTTC TGAGGCGGAA AGAACCAGCT GTGGAATGTG 3800  
TGTCAGTTAG GGTGTGGA AAA GTCCCCAGGC TCCCCAGCAG GCAGAAGTAT 3850  
50 GCAAAGCATG CATCTCAATT AGTCAGCAAC CAGGTGTGGA AAGTCCCCAG 3900  
GCTCCCCAGC AGGCAGAAGT ATGCAAAGCA TGCATCTCAA TTAGTCAGCA 3950  
55 ACCATAGTCC CGCCCTAAC TCCGCCCATC CCGCCCTAA CTCGCCCCAG 4000  
60 TTCCGCCCAT TCTCCGCCCC ATGGCTGACT AATTTTTTTT ATTTATGCAG 4050

65

ES 2 282 994 T3

AAGCCGAGGC CGCCTCGGCC TCTGAGCTAT TCCAGAAGTA GTGAGGAGGC 4100  
5 TTTTTGGAG GCCTAAGCTT TTGCAAAAAG CTGTTAACAG CTTGGCACTG 4150  
GCCGTCGTT TACAACGTCG TGACTGGGAA AACCCCTGGCG TTACCCAACT 4200  
10 TAATCGCCTT GCAGCACATC CCCCCTTCGC CAGCTGGCGT AATAGCGAAG 4250  
AAGCCCGCAC CGATCGCCCT TCCCAACAGT TCGTAGCCT GAATGGCGAA 4300  
15 TGGCGCCTGA TCGCGTATTT TCTCCTTACG CATCTGTGCG GTATTTCACA 4350  
CCGCATACCT CAAGCAACC ATAGTACCGG CCTGTAGCG GCCCATTAAAG 4400  
20 CCGCGCGGT GTGGTGGTTA CGCGCAGCGT GACCGCTACA CTTGCCAGCG 4450  
CCCTAGCGCC CGCTCCTTC GCTTCTTCC CTTCCTTCT CGCCACGTC 4500  
25 GCCGCTTC CCCGTCAGC TCTAAATCGG GGGCTCCCTT TAGGGTCCG 4550  
ATTTAGTCT TTACCGCACC TCGACCCCAA AAACTTGAT TTGGGTGATG 4600  
30 GTTCACGTAG TGGGCCATCG CCCTGATAGA CGGTTTTTCG CCCTTTGACG 4650  
TTGGAGTCCA CGTCTTTAA TAGTGGACTC TTGTTCCAAA CTGGAACAAC 4700  
35 ACTCAACCT ATCTCGGCT ATTCTTTGA TTTATAAGGG ATTTGCCGA 4750  
TTTCGGCCTA TTGGTTAAA AATGAGCTGA TTTAACAAA ATTTAACGCG 4800  
40 AATTTAACA AAATATTAAC GTTTACAATT TTATGGTGCA CTCTCAGTAC 4850  
AATCTGCTCT GATGCCGCAT AGTTAAGCCA ACTCCGCTAT CGCTACGTGA 4900  
CTGGGTCATG GCTGCCCCCC GACACCCGCC AACACCCGCT GACCGCCCT 4950  
50 GACGGGCTTG TCTGCTCCCG GCATCCGCTT ACAGACAAGC TGTGACCGTC 5000  
TCCGGGAGCT GCATGTGTCA GAGGTTTICA CCGTCATCAC CGAAACGCGC 5050  
55 GAGGCAGTAT TCTTGAAGAC GAAAGGGCCT CGTGATACGC CTATTTTTAT 5100  
AGGTTAATGT CATGATAATA ATGGTTTCTT AGACGTCAGG TGGCACTTTT 5150  
60 CCGGGAAATG TCGCGGAAC CCCTATTTGT TTATTTTCT AAATACATTC 5200

65

ES 2 282 994 T3

AAATATGTAT CCOCTCATGA GACAATAACC CTGATAAATG CTTCAATAAT 5250  
5 ATTGAAAAAG GAAGAGTATG AGTATTCAAC ATTTCCGTGT CGCCCTTATT 5300  
CCCTTTTTTG CGGCATTTTG CCTTCCTGTT TTTGCTCACC CAGAAACGCT 5350  
10 GGTGAAAGTA AAAGATGCTG AAGATCAGTT GGGTGCACGA GTGGGTTACA 5400  
TCGAACTGGA TCTCAACAGC GGTAAGATCC TTGAGAGTTT TCGCCCCGAA 5450  
15 GAACGTTTTT CAATGATGAG CACTTTTAAA GTTCTGCTAT GTGGCGCGGT 5500  
ATTATCCCGT GATGACGCCG GGCAAGAGCA ACTCGGTCCG CGCATACT 5550  
20 ATTCTCAGAA TGACTTGOTT GAGTACTCAC CAGTCACAGA AAAGCATCTT 5600  
ACGGATGGCA TGACAGTAAG AGAATTATGC AGTGCTGCCA TAACCATGAG 5650  
25 TGATAACACT GCGGCCAACT TACTTCTGAC AACGATCGGA GGACCGAAGG 5700  
AGCTAACCGC TTTTTGAC AACATGGGGG ATCATGTAAC TCGCCTTGT 5750  
30 CGTTGGGAAC CGGAGCTGAA TGAAGCCATA CCAAACGACG AGCGTGACAC 5800  
CACGATGCCA GCAGCAATGG CAACAACGTT GCGCAAATA TTAAGTGGCG 5850  
35 AACTACTTAC TCTAGCTTCC CGCAACAAT TAATAGACTG GATGGAGGCG 5900  
GATAAAGTTG CAGGACCACT TCTGCGCTCG GCCCTTCCGG CTGGCTGOTT 5950  
40 TATTGCTGAT AAATCTGGAG CCGGTGAGCG TGGOTCTCGC GGTATCATTG 6000  
CAGCACTGGG GCCAGATGGT AAGCCCTCCC GTATCGTAGT TATCTACACG 6050  
45 ACGGGGAGTC AGGCAACTAT GGATGAACGA AATAGACAGA TCGCTGAGAT 6100  
AGGTGCCTCA CTGATTAAGC ATTGGTAACT GTCAGACCAA GTTTACTCAT 6150  
50 ATATACTTTA GATTGATTTA AACTTCATT TTTAATTTAA AAGGATCTAG 6200  
GTGAAGATCC TTTTGTATAA TCTCATGACC AAAATCCCTT AACGTGAGTT 6250  
55 TTCGTTCCAC TGAGCGTCAG ACCCCGTAGA AAAGATCAA GGATCTTCTT 6300  
60 GAGATCCTTT TTTTCTGCGC GTAATCTGCT GCTTGCAAAC AAAAAACCA 6350

65

ES 2 282 994 T3

CCGCTACCAG CGGTGGTTTG TTTGCCGGAT CAAGAGCTAC CAACTCTTTT 6400  
5 TCCGAAGGTA ACTGGCTTCA GCAGAGCGCA GATACCAAAT ACTGTCCTTC 6450  
TAGTGTAGCC GTAGTTAGGC CACCACTTCA AGAACTCTGT AGCACCGCCT 6500  
10 ACATACCTCG CTCTGCTAAT CCTGTTACCA GTGGCTGCTG CCAAGTGGCGA 6550  
TAAGTCGTGT CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAGG 6600  
15 CGCAGCGGTC GGGCTGAACG GGGGGTTCGT GCACACAGCC CAGCTTGGAG 6650  
CGAACGACCT ACACCGAACT GAGATACCTA CAGCGTGAGC ATTGAGAAAG 6700  
20 CGCCACGCTT CCCGAAGGGA GAAAGGCGGA CAGGTATCCG GTAAGCGGCA 6750  
GGTTCGGAAC AGGAGAGCGC ACGAGGGAGC TTCCAGGGGG AAACGCCTGG 6800  
25 TATCTTTATA GTCCTGTCGG GTTTCGCCAC CTCTGACTTG AGCGTCGATT 6850  
TTTGTGATGC TCGTCAGGGG GGCAGGACCT ATGGAAAAC GCCAGCAACG 6900  
CGGCCTTTT ACGGTTCCTG GCCTTTTGCT GGCCTTTTGC TCACATGTTT 6950  
35 TTTCTGCGT TATCCCCTGA TTCTGTGGAT AACCGTATTA CCGCCTTGA 7000  
GTGAGCTGAT ACCGCTCGCC GCAGCCGAAC GACCGAGCGC AGCGAGTCAG 7050  
40 TGAGCGAGGA AGCGGAAGAG CGCCCAATAC GCAAACCGCC TCTCCCCGCG 7100  
CGTTGGCCGA TTCATTAATC CAGCTGGCAC GACAGGTTT CCGACTGGAA 7150  
45 AGCGGGCAGT GAGCGCAACG CAATTAATGT GAGTTACCTC ACTCATTAGG 7200  
CACCCCAGGC TTTACACTTT ATGCTTCCGG CTCGTATGTT GTGTGGAATT 7250  
GTGAGCGGAT AACAAATTCA CACAGGAAAC AGCTATGACC ATGATTACGA 7300  
55 ATTAA 7305