



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 268 711**

51 Int. Cl.:
C12N 15/53 (2006.01)
C12N 9/06 (2006.01)
C12N 15/82 (2006.01)
C12N 5/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Número de solicitud europea: **96934039 .7**
86 Fecha de presentación : **03.10.1996**
87 Número de publicación de la solicitud: **0859849**
87 Fecha de publicación de la solicitud: **26.08.1998**

54 Título: **Polipéptidos y polinucleótidos relacionados con las subunidades alfa y beta de las glutamato deshidrogenasas y procedimientos de utilización.**

30 Prioridad: **06.10.1995 US 541033**

45 Fecha de publicación de la mención BOPI:
16.03.2007

45 Fecha de la publicación del folleto de la patente:
16.03.2007

73 Titular/es: **The University of Florida
223 Grinter Hall
Gainesville, Florida 32611, US**

72 Inventor/es: **Schmidt, Robert, R. y
Miller, Philip**

74 Agente: **Curell Suñol, Marcelino**

ES 2 268 711 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

Polipéptidos y polinucleótidos relacionados con las subunidades alfa y beta de las glutamato deshidrogenasas y procedimientos de utilización.

5

Antecedentes de la invención

El nitrógeno inorgánico adquirido por las plantas se convierte finalmente en amonio antes de ser asimilado en el metabolismo del nitrógeno orgánico. Se ha postulado que un enzima implicado en el proceso de asimilación es la glutamato deshidrogenasa (GDH), un grupo de enzimas ubicuos presentes en casi todos los organismos desde los microbios a los vegetales superiores y animales [Srivastava, H.S., R.P. Singh, *Phytochem.* 26:597-610 (1987)]. La GDH cataliza la conversión reversible del α -cetoglutarato a glutamato mediante una aminación reductora que utiliza como cofactor al β -nicotinamida-adenina dinucleótido reducido (NADH) o el β -nicotinamida adenina dinucleótido fosfato reducido (NADPH). La función de las GDH vegetales en la asimilación del amonio en aminoácidos ha sido cuestionada desde el descubrimiento de la senda glutamina sintetasa/glutamato sintasa (GS/GOGAT) que se cree que es la senda preferida en la asimilación del amonio en los vegetales superiores [Mifflin, B.J., P.J. Lea *Phytochem.* 15:873-885 (1976)].

La principal objeción a que la GDH tenga una función importante en el metabolismo del nitrógeno de los vegetales es su baja afinidad por el amonio que necesitaría concentraciones de amonio intracelular elevadas para funcionar anabólicamente. Las evidencias iniciales indicaban que la GDH es un enzima catabólico que cataliza la desaminación del glutamato con solamente una función anabólica parcial en la síntesis del glutamato [Wallgrove, J.C., N.P. Hall, A.C. Kendall, *Plant Physiol.*, 83:155-158 (1987)]. La función fisiológica de grandes cantidades de GDH que se encuentran en diversos tejidos vegetales y orgánulos todavía está poco esclarecida y las condiciones bajo las que posiblemente la GDH podría tener una función significativa en el metabolismo del carbono y el nitrógeno no han sido determinadas.

La mayoría de la GDH vegetal caracterizada hasta la fecha se encuentra localizada en las mitocondrias; sin embargo, se ha caracterizado una especie de GDH que difiere en múltiples propiedades (p.ej., especificidad de cofactor, valores de K_m , localización en orgánulos y estabilidad térmica entre otras) a partir de los cloroplastos de un alga verde unicelular *Chlorella sorokiniana*. Se ha demostrado que las células de *C. sorokiniana* poseen una GDH constitutiva, mitocondrial, tetramérica específica del NAD (a partir de aquí designada "NAD-GDH") Meredith, M.J., R.M. Gro-nostajski, R.R. Schmidt, *Plant Physiol.*, 61:967-974 (1978)] y siete isozimas de GDH, específicas de NADP, homo- y heteroméricas, localizadas en los cloroplastos e inducibles por el amonio (a partir de aquí designadas como "NADP-GDH") (Prunkard, D.E. N.F. Bascomb, R.W. Robinson, R.R. Schmidt, *Plant Physiol.*, 81:349-355 (1986); Bascomb, M.F., R.R. Schmidt, *Plant. Physiol.*, 83:75-84 (1983)]. Se ha demostrado que las siete isozimas NADP-GDH cloroplásticas tienen diferentes movilidades electroforéticas durante PAGE-nativo, lo que puede ser consecuencia de la formación de homo y heterohexámeros compuestos por diferentes proporciones de α - y β -subunidades (53,5 y 52,3 kilodaltons, respectivamente).

Las células de *Chlorella* cultivadas en medio 1 a 2 mM en amonio acumulan solamente el]] α -homoheptámero (Bascomb y Schmidt, *supra*). La adición de concentraciones de amonio superiores (3,4 a 29 mM) a células cultivadas en nitrato produjo la acumulación tanto de la subunidad α como de la β en las holoenzimas NADP-GDH [Prunkard *et al.*, *supra*; Bascomb y Schmidt, *supra*; Bascomb, N.F., D.E. Prunkard, R.R. Schmidt, *Plant Physiol.*, 83:85-91 (1987)]. Prunkard *et al.*, (Prunkard D.E., N.F. Bascomb, N.F. W.T. Molin, R.R. Schmidt, *Plant Physiol.*, 81:413-422 [1986]) demostraron que la proporción de subunidades de la NADP-GDH y el patrón de holoenzimas están influidos tanto por la fuente de carbono y nitrógeno como por las condiciones de luz bajo las que se cultivan las células.

Los homoheptámeros α - y β -NADP-GDH purificados a partir de las células de *Chlorella* tienen valores K_m para el amonio notablemente diferentes; sin embargo, los valores de la K_m para sus otros sustratos son muy semejantes. El α -homoheptámero compuesto por seis α -unidades idénticas) que cataliza la biosíntesis de glutamato está regulada alostéricamente por el NADPH y posee una K_m para el amonio muy baja, comprendida entre 0,02 y 3,5 mM, dependiendo de la concentración de NADPH (Bascomb y Schmidt, *supra*). El valor de K_m para el amonio del α -homoheptámero es el valor K_m de amonio más bajo publicado hasta la fecha para GDH de cualquier vegetal. Por el contrario, el β -homoheptámero (forma catabólica) no es un enzima alostérico con un K_m para el amonio de aproximadamente 75 mM. De tales estudios que implican los enzimas purificados, se había postulado hasta la fecha que los heterohexámeros tienen grados variables de afinidad para el amonio, comprendidos entre los valores de K_m para los α - y β -homoheptámeros. Sorprendentemente, sin embargo, se ha descubierto que ciertos heterohexámeros pueden tener una proporción de actividad aminante:desaminante superior a la de cualquiera de los α - o β -homoheptámeros.

Aunque las α - y β -subunidades tienen velocidades de transformación *in vivo* diferentes (Bascomb *et al.*, *supra*) y los homoheptámeros correspondientes tienen valores de K_m para el amonio notablemente diferentes, las α - y β -subunidades provienen de proteínas precursoras de tamaños prácticamente idénticos (aproximadamente 58.000 Daltons) y se demostró que tienen mapas peptídicos muy semejantes (Prunkard *et al.*, *supra*; Bascomb and Schmidt, *supra*). Además, los anticuerpos policlonales preparados contra el β -homoheptámero son capaces de inmunoprecipitar todos los isozimas NADP-GDH (Yeung *et al.*, *Anal. Biochem.*, 10:216-228 (1981); Bascomb *et al.*, *supra*), pero no tienen reacción cruzada con el NAD-GDH mitocondrial.

Cock *et al.*, [*Plant Mol. Biol.*, 17:17-27 (1991)], describen una secuencia de nucleótidos que codifica la β -subunidad madura completa de la glutamato-deshidrogenasa (GDH) de *Chlorella Sorokiniana* y por lo menos el 98% de la α -subunidad madura. Dicho artículo también proporciona pruebas genómicas de clonaje y de transferencia Southern que indican que el genoma de *C. sorokiniana* solamente posee un gen estructural para NADP-GDH.

Los isozimas NADP-GDH cloroplásticos codificados en el núcleo de *C. sorokiniana* son las únicas secuencias de GDH localizadas en los cloroplastos aisladas y caracterizadas de vegetales.

Sumario de la invención

Se ha descubierto que las dos subunidades maduras de los isozimas GDH de *chlorella* se generan mediante el procesamiento específico de dos proteínas precursoras semejantes codificadas en dos ARNm formados mediante el ensamblaje alternativo de un pre-mARN derivado de un único gen nuclear. Además, se ha identificado el lugar de corte y la secuencia del péptido aminoterminal de las subunidades GDH funcionales maduras.

Más específicamente, la presente invención proporciona el aislamiento y la caracterización de dos ADNc de tamaño completo de ARNm aislados a partir del alga verde unicelular *Chlorella sorokiniana*. Los dos ADNc codifican las proteínas precursoras (α -precursor, 53,5 kD; β -precursor, 57,85 kD) que son procesadas para producir las α - y β -subunidades maduras (53,5 kD; 52,3 kD, respectivamente) que componen los isozimas de hexaméricos activo de NADP-GDH. La presente invención se refiere a un solo gen de NADP-GDH que se ensambla de modo alternativo para generar dos ARNm que codifican dos proteínas precursoras diferentes de los cloroplastos. Tales proteínas precursoras se pueden procesar a continuación para dar las subunidades α y β maduras de los isozimas de NADPH-GDH. También se describen fragmentos de utilidad o mutantes de las secuencias de nucleótidos y aminoácidos que retienen la actividad o utilidad dadas a conocer. Por ejemplo, ciertos fragmentos de las secuencias de aminoácidos proporcionadas en la presente memoria pueden ser de utilidad como péptidos de tránsito, que proporcionan a la proteína la capacidad de entrar y permanecer en ciertos compartimentos celulares. Las secuencias de nucleótidos que se describen en la presente memoria, y los fragmentos de tales secuencias de nucleótidos, pueden ser de utilidad, por ejemplo, como cebadores en procesos de amplificación o como sondas para hibridar a secuencias complementarias de interés. Las secuencias de aminoácidos y nucleótidos y los fragmentos de las mismas tal como se describen en la presente memoria también pueden ser de utilidad como marcadores del peso molecular o en la identificación y establecimiento del grado de relación de otras secuencias de nucleótidos, polipéptidos o isozimas relacionadas con NADP-GDH.

La presente invención proporciona además procedimientos con los que se puede alterar la asimilación de nitrógeno inorgánico en el metabolismo del nitrógeno orgánico de los vegetales superiores, mediante la expresión de la GDH de *C. sorokiniana* o las GDH aisladas de otros organismos. Las alteraciones en el metabolismo del nitrógeno pueden resultar en el incremento en la asimilación de nitrógeno que, tal como se conoce en la materia, puede afectar la composición del vegetal mediante un efecto inverso sobre el metabolismo del carbono, p.ej., la acumulación de hidratos de carbono. El objetivo de la presente invención también se refiere a construcciones de ADN destinados a la utilización en los procedimientos descritos. La presente invención incluye la identificación de las secuencias aminoterminales de α - y β -subunidades que se pueden ensamblar para formar isozimas NADP-GDH, p.ej., la NADP-GDH hexamérica nativa que se encuentra en los cloroplastos de *C. sorokiniana*. Dicha precisa información del peso molecular se puede utilizar para expresar NADP-GDH provista de las propiedades cinéticas especiales de los homohexameros α - y β -NADP-GDH cloroplásticos de *C. sorokiniana*. La presente invención proporciona asimismo células recombinantes u organismos, p.ej., cultivos o plantas transgénicas que, mediante la expresión de los genes de las secuencias de polinucleótidos descritas, con el fin de producir los correspondientes polipéptidos, pueden tener un rendimiento superior, mejores propiedades asimiladoras de nitrógeno que pueden ventajosamente incrementar su tolerancia a la toxicidad por el amonio, mejor tolerancia al estrés osmótico y mejor composición del cultivar o vegetal.

Breve descripción de los dibujos

La Figura 1 muestra un patrón de actividades de NADP-GDH en homogeneizados de células sincronizadas de *C. sorokiniana* cultivadas durante 240 minutos en medio con 29 mM amonio con luz continua. Se analizaron espectrofotométricamente alícuotas de homogeneizados aclarados, de células cosechadas a diferentes intervalos de tiempo, para determinar actividades de aminación (●), desaminación (○) de NADP-GDH.

La Figura 2 muestra patrones de acumulación de antígenos de NADP-GDH en células iluminadas cultivadas en medio con 29 mM amonio durante 240 minutos. A tiempo cero, se añadió amonio a las células hijas de *C. sorokiniana* y se cultivaron con iluminación. Las autorradiografías de las transferencias Western se analizaron mediante densitometría láser para determinar los niveles relativos de las α - (●) y β -subunidades (○) de NADP-GDH a lo largo de los 240 minutos del período de inducción.

Breve descripción de las secuencias

La SEC. ID n° 1, es el ADNc de la proteína precursora de la α -subunidad de una glutamato deshidrogenasa específica para NADP.

La SEC. ID n° 2, es la secuencia de aminoácidos deducida de la secuencia de polinucleótidos SEC. ID n° 1.

ES 2 268 711 T3

La SEC. ID nº 3 es el ADNc de la proteína precursora de la β -subunidad de una glutamato deshidrogenasa específica para NADP.

La SEC. ID nº 4 es la secuencia de aminoácidos deducida del polinucleótido de SEC. ID nº 3.

La SEC. ID nº 5 es la secuencia N-terminal de la α -subunidad de NADP-GDH.

La SEC. ID nº 6 es la secuencia N-terminal de la β -subunidad de NADP-GDH.

La SEC. ID nº 7 es la secuencia de ADNc en el clon referido como pBGDc 53.

La SEC. ID nº 8 es un cebador que se hibrida a las regiones conservadas de los pequeños ARNm de NADP-GDH.

La SEC. ID nº 9 es un polinucleótido poli(dT) utilizado como cebador adaptador según la presente invención.

La SEC. ID nº 10 es un polinucleótido utilizado como cebador según la presente invención.

La SEC. ID nº 11 es un polinucleótido utilizado como cebador según la presente invención.

La SEC. ID nº 12 es un polinucleótido utilizado como adaptador cebador según la presente invención.

La SEC. ID nº 13 es el polinucleótido inserto en el clon referido como pRGDc 60.

La SEC. ID nº 14 es el polinucleótido inserto en el clon designado como pRGDc 61.

La SEC. ID nº 15 es el polinucleótido utilizado como un cebador según la presente invención.

La SEC. ID nº 16 es el polinucleótido inserto en un clon designado como pGDc 63

La SEC. ID nº 17 es el polinucleótido inserto de un clon designado como pGDc 64.

La SEC. ID nº 18 es el polinucleótido resultante de la ligación de los fragmentos purificados correspondientes a los insertos en los clones designados pBGDc 53 y pGDc 63, según la presente invención.

La SEC. ID nº 19 es la secuencia de polinucleótidos resultante de la ligación de los insertos purificados de los clones designados pGDc 64 y pBGDc 53.

La SEC. ID nº 20 es un polinucleótido utilizado como cebador según la presente invención.

La SEC. ID nº 21 es un polinucleótido utilizado como cebador que se hibrida al extremo 3' de ADN molde según la presente invención.

La SEC. ID nº 22 es un polinucleótido utilizado como cebador según la presente invención.

La SEC. ID nº 23 es la secuencia de polinucleótidos (ADNc) de la α -subunidad de la NADP-GDH madura procesada.

La SEC. ID nº 24 es la secuencia de aminoácidos de la α -subunidad de NADP-GDH madura procesada.

La SEC. ID nº 25 es la secuencia de polinucleótidos (ADNc) de la β -subunidad de NADP-GDH madura procesada.

La SEC. ID nº 26 es la secuencia de aminoácidos de la β -subunidad de NADP-GDH madura procesada.

Descripción detallada de la invención

La presente invención proporciona a continuación secuencias de polinucleótidos no descritas, por ejemplo, ADNc de las proteínas precursoras de las α - y β -subunidades de la glutamato deshidrogenasa NAD específica inducible por el amonio localizada en los cloroplastos (a partir de aquí NADP-GDH) de *Chlorella sorokiniana*. La secuencia de nucleótidos para las proteínas precursoras de las α - y β -subunidades que forman la NADP-GDH se muestran en las SEC. ID nºs 1 y 3, respectivamente. Las secuencias de aminoácidos deducidas para las proteínas precursoras de las α - y β -subunidades del enzima NADP-GDH de *Chlorella sorokiniana* se muestran en las SEC. ID nºs 2 y 4, respectivamente.

Los huéspedes *E. coli* que comprenden los insertos de ADNc de la presente invención se depositaron en la American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland 20852 USA. A los cultivos se les asignaron los siguientes números de acceso en el repositorio:

ES 2 268 711 T3

Cultivo	Número de acceso	Fecha del depósito
<i>E. coli</i> DH5 α α -NADP-GDH SEC. ID n° 1 (+42 bp)	ATCC 69925	6 de Octubre 1995
<i>E. coli</i> DH5 α β -NADP-GDH SEC. ID n° 1 (- 42 bp)	ATCC 69926	6 de Octubre 1995

5 El análisis automático de la secuencia de aminoácidos identificó 20 y 10 aminoácidos aminoterminales de las α - y β -subunidades, respectivamente. La alineación de las secuencias de péptidos de las α - y β -subunidades revela que las dos subunidades son idénticas con excepción de una extensión de 11 aminoácidos presente en la α -subunidad más grande. Los anticuerpos monoclonales generados contra la α -subunidad demostraron reconocer la β -subunidad proporcionando pruebas adicionales de que las dos subunidades son casi idénticas. La identificación de los lugares
10 únicos de procesamiento de las α - y β -subunidades en las proteínas precursoras proporciona un mecanismo molecular para explicar las diferencias en las propiedades cinéticas de los isozimas homohexaméricos de la α - y β - NADP-GDH.

15 Los datos mencionados anteriormente proveen información aplicable a los vegetales genéticamente modificados con una GDH específica provista de propiedades cinéticas favorables que pueden influenciar tanto el metabolismo del nitrógeno como el del carbono. Debido al elevado contenido en guanina/citosina los ADNc son muy adecuados para la expresión heteróloga en los vegetales superiores. La introducción de una cualquiera o de las dos subunidades provistas de secuencias dirigidas a los cloroplastos o con secuencias dirigidas a otros orgánulos en sistemas vegetales heterólogos puede mejorar la asimilación de nitrógeno e influenciar el equilibrio carbono/nitrógeno.

20 Se ha descubierto que la localización a los cloroplastos está relacionada con, y puede depender de, el extremo N-terminal de la proteína α - o β -precursora. La escisión del extremo N-terminal de los precursores genera la proteína madura. En consecuencia, el péptido de tránsito a los cloroplastos comprende un péptido que forma, o es un fragmento activo de, el extremo N-terminal escindido de la proteína precursora. Los péptidos que tienen una secuencia semejante o equivalente, o que tienen estructura terciaria o conformación semejante a estos péptidos escindidos pueden funcionar
25 asimismo como péptidos de tránsito. El péptido de transición a los cloroplastos comprende el fragmento activo del péptido N-terminal escindido del α -precursor (un 40-mero) o del β -precursor (un 37-mero). Las secuencias de polinucleótidos que codifican los péptidos de tránsito a los cloroplastos se pueden utilizar por un experto en la materia para generar los péptidos de tránsito a los cloroplastos utilizados con los péptidos descritos en la presente memoria, u otros conocidos en la técnica.

30 La adición, eliminación o reemplazo del péptido de tránsito a los cloroplastos asociado con una proteína, p.ej., el enzima GDH, se puede utilizar para localizar la proteína según se necesite, mediante procedimientos bien conocidos en la técnica. Por ejemplo, se puede lograr la localización del enzima en un cloroplasto de una célula mediante la inserción de un péptido de tránsito a los cloroplastos en una secuencia de aminoácidos que carece de dicho péptido de tránsito. Se pueden añadir péptidos de tránsito específicos de la especie a los cloroplastos, o se pueden reemplazar los ya existentes para optimizar la inserción en los cloroplastos de una especie determinada. Además, la localización dentro del cloroplasto de una proteína que se expresa en el cloroplasto se puede lograr mediante la transformación directa del plasto con las secuencias de polinucleótidos que codifican una proteína expresada. De modo semejante, la
35 eliminación de un péptido de tránsito a los cloroplastos o la producción de una proteína recombinante que carece del péptido se puede utilizar para secuestrar la proteína en un compartimento celular distinto de los cloroplastos.

40 Las plantas transformadas que expresan el α -homohexamero pueden ser más tolerantes a la toxicidad del amonio, asimilar el amonio con mayor eficacia y responder más rápidamente al estrés osmótico que se encuentra en los suelos transitoriamente salinos mediante el aporte de glutamato, el precursor del osmoprotector prolina. La expresión de, por ejemplo, el β -homohexamero o los heterohexameros GDH se puede utilizar con el fin de alterar la velocidad de asimilación de nitrógeno, favoreciendo la acumulación de carbohidratos en los frutos y otros órganos de almacenamiento.

45 Inesperadamente, se descubrió que un hexámero que comprende por lo menos una α -subunidad y por lo menos una β -subunidad, es decir, un heterohexamero, puede tener una actividad ventajosa. Específicamente, la proporción de actividad aminante:desaminante (es decir, la capacidad biosintética para las síntesis de glutamato) de un isozima NADP-GDH cloroplástico se puede incrementar mediante la incorporación de tanto α - como β -subunidades a la proteína hexamérica en lugar de utilizar un homohexamero que comprenda solamente α - o solamente β -subunidades. En una forma de realización de la presente invención, puede ser ventajoso expresar simultáneamente ADNc que
50 codifican los dos tipos de subunidades en la misma planta en diferentes proporciones/niveles con el fin de obtener una proporción determinada de α -subunidades y β -subunidades en el heterohexamero. Por ejemplo, se ha descubierto que es preferible un heterohexamero de NADP-GDH que tiene por lo menos una de las subunidades en forma β para incrementar la proporción de actividad aminante:desaminante. Un heterohexamero más preferido tiene entre 2 y 5 β -subunidades. Tal velocidad diferencial en la expresión de los dos ADNc se puede lograr colocándolos bajo el control de promotores vegetales de diferente potencia o bajo el mismo promotor que se ha modificado para generar diferentes niveles de expresión. La utilización de este sistema de isozima NADP-GDH de alga en la biotecnología de los vegetales ofrece ventajas sobre los NADP-GDH de organismos tales como bacterias, que contienen una sola forma del enzima (es decir, no isozimas).

65 Se reconoce que los niveles de expresión de ciertas proteínas recombinantes en las plantas transgénicas se puede mejorar mediante el incremento de la expresión de transcritos de mRNA estabilizados; y que, por el contrario, la detección de tales transcritos de estabilizados se puede utilizar para medir la expresión de los productos de traducción (proteína). La baja expresión de ARN de proteínas en las plantas y, por consiguiente, de la baja expresión de proteí-

ES 2 268 711 T3

nas, se puede resolver mediante la utilización de un gen sintético mejorado que especifique la proteína deseada del organismo origen del gen.

Por consiguiente, en una forma de realización de la presente invención, las bacterias y plantas se pueden modificar genéticamente para lograr los niveles de expresión deseados para las nuevas proteínas con valor agrícola o comercial. Para proveer genes de expresión incrementada en los vegetales, la secuencia de ADN del gen se puede modificar para que comprenda los dones preferidos en los genes muy expresados de los vegetales, para lograr un contenido en los nucleótidos A + T sustancialmente como el encontrado en las plantas y preferentemente para formar asimismo una secuencia iniciadora de vegetal y para eliminar secuencias desestabilizadoras, de poliadenilación inapropiada, de degradación y de terminación del ARN y para evitar secuencias que constituyan estructuras secundarias en horquilla y sitios de ensamblaje ADN. Por ejemplo, en los genes sintéticos, los codones utilizados para especificar un aminoácido determinado se pueden seleccionar en relación con la distribución de frecuencias de utilización de codones empleada por los genes muy expresados en las plantas para especificar dicho aminoácido. Tal como aprecian los expertos en la materia, la distribución de frecuencias de utilización de codones utilizada en un gen sintético es un determinante del nivel de expresión.

Para los objetivos de la presente invención, “frecuencia de utilización de los codones preferidos” se refiere a la preferencia mostrada por una célula huésped específica en la utilización de codones de nucleótidos para especificar un aminoácido determinado. Para determinar la frecuencia de utilización de un codón particular en un gen, el número de ocurrencias de dicho codón en el gen se divide entre el número total de ocurrencias de todos los codones utilizados para especificar el mismo aminoácido en el gen. De modo semejante, la frecuencia de utilización de los codones preferidos mostrada por una célula huésped se puede calcular haciendo el promedio de la frecuencia de utilización de los codones preferidos en un gran número de los genes expresados por la célula huésped. Es preferible que dicho análisis se limite a genes de expresión elevada en la célula huésped.

Cuando se sintetiza un gen para mejorar la expresión en una célula huésped es deseable diseñar el gen de modo que su frecuencia de utilización de codones se aproxime a la frecuencia de utilización de los codones preferidos de la célula huésped.

El porcentaje de desviación de la frecuencia de utilización de codones preferidos de un gen sintético con respecto a la utilizada por una célula huésped se calcula primero mediante la determinación del porcentaje de desviación de la frecuencia de utilización de un solo codón con respecto a la de la célula huésped, seguido de la obtención del promedio de la desviación para todos los codones. Tal como se define en la presente memoria dicho cálculo incluye codones especiales (es decir, ATG y TGG). En términos generales la desviación media general de utilización de codones de un gen sintético relativa a la de la célula huésped se calcula utilizando la siguiente ecuación:

$$A = \sum_{n=1}^Z \frac{X_n - Y_n}{X_n} \times 100$$

en la que X_n = frecuencia de utilización del codón n en la célula huésped; Y_n = frecuencia de utilización del codón n en el gen sintético. En la que n representa un codón individual que especifica un aminoácido, el número total de codones es Z. La desviación general de la frecuencia de utilización de codones, A, para todos los aminoácidos preferentemente debe ser inferior a aproximadamente 25% y más preferentemente inferior a aproximadamente el 10%. Por consiguiente, se puede diseñar un gen de modo que su frecuencia de distribución de utilización de codones se desvíe, preferentemente no más del 25% de la de los genes vegetales de elevada expresión y, más preferentemente, no más de aproximadamente el 10%. Además, se considera el porcentaje de contenido G + C de la tercera base degenerada (las monocotiledóneas parecen favorecer G + C en esta posición, mientras que las dicotiledóneas no lo hacen). También se reconoce que el nucleótido XCG (donde X es A, T, C o G) es el codón menos preferido en las dicotiledóneas mientras que el codón XTA se evita tanto en las dicotiledóneas como en las monocotiledóneas. Los genes sintéticos de la presente invención preferentemente presentan asimismo índices de evitación de los dobletes CG y TA próximos a los del vegetal huésped elegido. Más preferentemente dichos índices no se desvían de los del huésped por más de entre aproximadamente el 10 y el 15%.

El ensamblaje del gen de NADP-GDH de la presente invención se puede realizar utilizando procedimientos estándar en la materia. Un gen estructural diseñado para una expresión incrementada en los vegetales de la forma de realización preferida se puede ensamblar enzimáticamente en un vector de ADN a partir de segmentos de oligonucleótidos de doble cadena sintetizados químicamente. A continuación se puede introducir el gen en la célula del vegetal huésped y expresar mediante procedimientos conocidos en la materia. Preferentemente, la proteína producida cuando se expresa el gen sintético en los vegetales es funcionalmente equivalente a una proteína nativa en que tiene una actividad aminante/desaminante comparable o mejorada. Según la presente invención, “funcionalmente equivalente” se refiere a identidad o casi identidad de funciones. El producto de un gen sintético que tiene por lo menos una propiedad relacionada con su actividad o función, que es igual o semejante a una proteína natural se considera funcionalmente equivalente a ésta. Se pueden realizar modificaciones en la secuencia de nucleótidos de la región codificante para alterar el contenido en A + T de la composición de bases del ADN de un gen sintético para que refleje el encontrado normalmente en los genes de las proteínas de expresión elevada nativas a la célula huésped. Preferentemente el contenido en A + T del gen sintético es sustancialmente igual al de dichos genes para las proteínas de expresión muy elevada.

En los genes que codifican proteínas vegetales de expresión elevada, el contenido en A + T es de aproximadamente el 55%. Es preferible que el gen sintético tenga un contenido en A + T próximo a este valor y no suficientemente elevado como para producir la desestabilización del ARN y, por consiguiente, niveles inferiores de expresión de proteína. Más preferentemente, el contenido en A + T no es superior a aproximadamente el 60% y más preferentemente es de aproximadamente el 55%. También, para la expresión final en las plantas, la secuencia de nucleótidos del gen puede ser modificada preferentemente para formar secuencias de inicio vegetales en el extremo 5' de la región codificante. Además, preferentemente se presta atención en asegurarse que se disponen lugares de restricción únicos en posiciones estratégicas para permitir el ensamblaje eficiente de los fragmentos de oligonucleótidos durante la construcción del gen sintético y para facilitar las modificaciones de nucleótidos subsiguientes. Como resultado de tales modificaciones en la región codificante del gen nativo, el gen sintético preferido se expresa en los vegetales a un nivel incrementado cuando se compara con el observado de los genes estructurales naturales.

Se conoce que la utilización relativa de codones sinónimos difiere entre las monocotiledóneas y las dicotiledóneas. En general, el factor más importante para discriminar entre los patrones de utilización de codones de las monocotiledóneas y dicotiledóneas es el porcentaje G + C de la tercera base degenerada. En las monocotiledóneas, 16 de 18 aminoácidos favorecen G + C en esta posición, mientras que las dicotiledóneas solamente favorecen G + C en 7 de 18 aminoácidos.

En la soja y el maíz, el patrón de utilización de codones del maíz se parece al de las monocotiledóneas en general, mientras que el patrón de utilización de codones de la soja es casi idéntico al patrón general de las dicotiledóneas.

Cuando se diseña un gen sintético que debe ser expresado en vegetales, es preferible eliminar las secuencias que interfieren con la eficacia de la expresión génica.

Se puede sintetizar un gen sintético con otros fines además del de ganar un incremento en el nivel de expresión. Por ejemplo, de acuerdo con la presente invención, una de las secuencias de nucleótidos que codifican la α -subunidad o la β -subunidad de la NADP-GDH se puede modificar de tal modo que los productos se expresen de modo diferencial, favoreciendo la expresión de una de las subunidades. Un resultado de tal expresión diferencial es un heterohexámero que comprende más de una de las subunidades que de la otra. La modificación puede comprender la sustitución de uno o más, pero no todos, segmentos de oligonucleótidos utilizados para construir el gen sintético por una región de la secuencia natural. Preferentemente, la expresión diferencial de la secuencia de nucleótidos que codifica las α - y β -subunidades de los polipéptidos de la NADP-GDH se puede utilizar para producir un heterohexámero que tenga por lo menos una β -subunidad, más preferentemente entre dos y cinco β -subunidades, y más preferentemente tres β -subunidades.

La molécula de ADN recombinante que comprende una secuencia de nucleótidos de la presente invención se puede introducir en el tejido de las plantas mediante procedimientos conocidos en la materia. El procedimiento utilizado para una especie vegetal determinada o para un tipo específico de tejido vegetal depende de las técnicas exitosas conocidas. A medida que se desarrollan nuevos procedimientos para la inserción estable de genes extraños en las células de las plantas y para la manipulación de las células modificadas, los expertos en la materia serán capaces de seleccionar de entre los procedimientos conocidos para lograr los resultados deseados. Los procedimientos para introducir ADN recombinante en los tejidos de los vegetales incluyen, pero sin limitarse a la adquisición directa de ADN [Paszkowski, J. *et al.*, EMBO J., 3:2717 (1984), electroporación (Fromm, M. *et al.*, Proc. Natl. Acad. USA, 82:5824 (1985), microinyección (Crossway, A. *et al.*, Mol. Gen. Genet., 202:179 (1986), o transferencia al tejido vegetal desde el *Agrobacterium tumefaciens* mediante T-ADN. No parece existir una limitación fundamental a la transformación de T-ADN a la clase de huésped natural de *Agrobacterium*. Se ha publicado la transformación con éxito mediante T-ADN de las monocotiledóneas [Hooykaas-Van Slogteren, G. *et al.*, Nature 311:763 (1984)], las gimnospermas [Dandekar, A. *et al.*, Biotechnology 5:587 (1987)] y las algas [Ausich, R., solicitud EPO 108.580]. En las siguientes referencias se describen los sistemas de vectores de T-ADN representativos: An, G. *et al.*, EMBO J., 4:277 (1985); Herrera-Estrella, L., *et al.*, Nature 303:209 (1983); Herrera-Estrella, L., *et al.*, EMBO J. 2:987 (1983); Herrera-Estrella, L., *et al.*, *Plant Genetic Engineering*, New York: Cambridge University Press, p 63 (1985). Una vez introducidas en el tejido vegetal, se puede ensayar la expresión del gen estructural mediante uno cualquiera de los procedimientos conocidos en la materia y se puede medir la expresión como mRNA transcrito o como proteína sintetizada. Se conocen procedimientos para el cultivo *in vitro* de tejidos vegetales y en múltiples casos, para la regeneración de plantas enteras. Los procedimientos conocidos para transferir el complejo de expresión introducido en cultivos de utilización comercial son conocidos por los expertos en la materia.

En una de las formas de realización preferidas la invención publicada en la presente memoria comprende la expresión en células vegetales de un gen de la NADP-GDH bajo el control de un promotor expresable en vegetales, es decir, insertando el gen en T-ADN bajo el control de un promotor expresable en vegetales e introduciendo el T-ADN que contiene el inserto en una célula vegetal utilizando procedimientos conocidos. Una vez que se obtienen células vegetales expresan el gen bajo el control de un promotor expresable en vegetales, se pueden generar tejidos vegetales y plantas enteras a partir de estas utilizando procedimientos y técnicas bien conocidas en la materia. Las plantas regeneradas se reproducen entonces mediante procedimientos convencionales y los genes introducidos se pueden transferir a otras variedades y cultivares mediante procedimientos convencionales de selección vegetal.

La introducción y expresión del gen de la NADP-GDH se puede utilizar para mejorar, por ejemplo, incrementar el rendimiento de un cultivo. Otras utilidades de la invención, que explotan las propiedades de los genes introducidos en las especies de vegetales serán evidentes para los expertos en la materia.

ES 2 268 711 T3

También existen diferencias en la elección de codones en los genes nucleares de los vegetales y en los cloroplastos. Los cloroplastos difieren de los vegetales superiores en que codifican solamente 30 especies de genes de ARNt. Debido a que los cloroplastos tienen restringidos sus genes de ARNt, la utilización de codones preferidos por las proteínas codificadas por los cloroplastos parece más extrema. Sin embargo, se ha publicado una correlación positiva entre el nivel de ARNt iso-aceptantes para un aminoácido determinado y la frecuencia con que se utiliza dicho codón en el genoma de los cloroplastos [Pfitzinger *et al.*, Nucl. Acids. Res., 15:1377-1386 (1987)]. En general, el perfil de codones de los cloroplastos se parece más de cerca al de los organismos unicelulares, con un sesgo fuerte hacia la utilización de A + T en la tercera base degenerada.

A continuación se dan ejemplos que ilustran los procedimientos, que incluyen el mejor modo de poner en práctica la presente invención. Estos ejemplos no se deben considerar limitativos. Todos los porcentajes son en peso y todas las proporciones de mezclas de disolventes son en volumen a menos que se indique de otro modo.

Ejemplos

Ejemplo 1

Cinética de las glutamato deshidrogenasas de los cloroplastos de C. sorokiniana

Las α - y β -isozimas de glutamato deshidrogenasa cloroplástica utilizadas en los experimentos siguientes se producen naturalmente por un organismo caracterizado como *Chlorella sorokiniana*.

Condiciones de cultivo de C. sorokiniana

Para la caracterización cinética tanto de la dirección aminorante como desaminante, los α - y β -holoenzimas se purificaron a partir de células que acumulaban solamente una forma de isozima homohexamérica GDH.

Las células de *C. sorokiniana* (UTEX-1230, University of Texas algal culture collection; 3B2NA, Robert R. Schmidt, University of Florida, Microbiology Cell Science Department) se cultivaron autotróficamente tal como han descrito anteriormente Prunkard *et al.*, en un medio basal modificado. El medio modificado contiene en concentraciones mM: CaCl₂, 0,34; K₂SO₄, 6,0; KH₂PO₄, 18,4; MgCl₂, 1,5; en concentración μ M: CoCl₂, 0,189; CuCl₂, 0,352; EDTA, 72; FeCl₃, 71,6; H₃BO₃, 38,8; MnCl₂, 10,1; NH₄VO₄, 0,20; (NH₄)₆MO₇O₂₄, 4,19; NiCl₂, 0,19; SnCl₂, 0,19; ZnCl₂, 0,734. El medio está suplementado con 1 mM NH₄Cl, 29 mM NH₄Cl o 29 mM KNO₃ como fuente de nitrógeno dependiendo de las condiciones experimentales. El medio que contiene NH₄Cl se ajustó a pH 7,4 y el medio que contiene KNO₃ se ajustó a pH 6,8 con KOH después de esterilizado en autoclave. Las células se suministraron con 2% (v/v) mezcla de CO₂-aire y una intensidad de luz suficiente para permitir la división celular en cuatro generaciones.

Purificación de las isozimas de NADP-GDH

Con el fin de purificar el α -isozima de glutamato deshidrogenasa, se cultivaron células de *C. sorokiniana* con luz continua en medio con 29 mM amonio en una cámara de plexiglás de 30 L tal como se describió con anterioridad [Baker, A.L., R.R. Schmidt, *Biochem. Biophys. Acta*, 74:75-83 (1963)]. Se cosecharon las células a 4,0 OD₆₄₀ mediante centrifugación a 30.000 rpm en una centrífuga Sharples y se lavaron dos veces en 10 mM Tris pH 8,5 a 4°C. Las células precipitadas (130 g) se almacenaron a -20°C en frascos de centrifuga de 250 ml hasta su utilización. La purificación de NADP-GDH se logró utilizando un procedimiento modificado de Yeung *et al.*, *supra*. Las modificaciones del proceso implicaron la sustitución de gel de Sephadex G-200 (Farmacia) por gel G-150 en la columna de filtración y la adición de NADP⁺ como estabilizador a una concentración final de 0,1 mM al tampón de filtración por gel y a todos los tampones de almacenamiento siguientes. Como modificación final, la etapa de resina de afinidad de NADP⁺ se omitió y se sustituyó por una etapa de PAGE preparativo no desnaturizante [Miller, P.W., W.D. Dunn, R.R. Schmidt, *BioRad US/EG Bulletin* (1987)].

La disolución de ensayo del enzima desaminante GDH está compuesta de 44 mM Tris, 20,4 mM glutamato y 1,02 mM NADP⁺, pH 8,8. La disolución de ensayo de aminación comprende 50 mM Tris, 25 mM α -cetoglutarato, 0,357 mM NADPH y 0,356 M (NH₄)₂SO₄, pH 7,4. Una unidad de actividad enzimática es la cantidad de NADP-GDH necesaria para reducir u oxidar 1,0 μ mol de NADP⁺ o NADPH por minuto a 38,5°C.

Las fracciones de la columna de Sephadex G-200 con actividad NADP-GDH se juntaron y concentraron mediante filtración Diaflow. El enzima soluble (68 mg) se protegió de la oxidación mediante la adición de DTT a una concentración final 10 mM y se dializó durante 30 minutos contra 28,8 mM Tris, 192 mM glicina, 2 mM DTT (pH 8,4). El dializado se aclaró mediante centrifugación a 20.000 g durante 10 minutos a 4°C y se combinó con 3 ml de 40% (peso/vol.) sacarosa y 1 ml de 0,02% azul de bromofenol.

Para PAGE no desnaturizante preparativo, se moldeó un gel de separación del 7% (peso/volumen, 28 acrilamida: 0,735 bis-acrilamida, pH 8,8) de 3 cm de altura, y un gel de apilamiento del 2% de acrilamida (peso/volumen, 1,6 acrilamida: 0,4 bis-acrilamida, pH 6,6) de 2 cm de altura en un tubo para geles de 28 mm ID Model 491 Prep Cell. Todas los almacenamientos se pretrataron con la resina de lecho mixto AG501-X8 para eliminar todo residuo de ácido acrílico contaminante con el fin de evitar la N-acilación *in vitro* de las proteínas durante la electroforesis. La muestra

ES 2 268 711 T3

de proteína se fraccionó por electroforesis a 15 mA de corriente constante durante 20 minutos y a continuación durante 3,5 horas a 30 mA de corriente constante. Se recogieron fracciones de seis mililitros y se ensayaron para determinar la actividad NADP-GDH desaminante y las fracciones que contenía GDH se juntaron. El enzima en las fracciones coleccionadas en 10 mM KPO₄ (pH 6,2), 0,1 mM NADP⁺ se concentró mediante ultrafiltración Diaflow a 1 mg/ml tal como se determinó por el procedimiento de Bradford, utilizando BSA como estándar. La preparación de enzima concentrada se almacenó a -20°C. La pureza de la preparación mediante tinte con plata para visualizar las proteínas fraccionadas mediante PAGE-SDS en geles del 10% (peso/volumen) Tris-Tricina [Schagger, H., G. von Jagow, *Anal. Biochem.*, 166:368-379 (1987)]

El β -isozima de NADP-GDH se purificó a partir de una mezcla de células cultivadas durante 240 minutos en medio 1 mM amonio (14 g), 90 minutos en medio 1 mM amonio (6 g) y durante 20, 40, 60 y 80 minutos en medio 29 mM amonio (1 g/punto de tiempo) según Bascomb y Schmidt, *supra*. El β -isozima NADP-GDH se purificó parcialmente utilizando un procedimiento modificado subescalado de Yeung *et al.*, *supra*. Las columnas de intercambio iónico de DEAE sephacel (pH 7,4 y pH 6) se subescalaron a 40 ml de volumen de lecho y se utilizó un gradiente lineal de 400 ml de KCl (0 a 0,4 M) para eluir las proteínas en fracciones de 3 ml. Las fracciones de la columna de intercambio iónico de DEAE pH 6 que contienen NADP-GDH se combinaron en dos grupos; correspondientes con las mitades delanteras y traseras del pico de actividad NADP-GDH. Las fracciones juntas por separado se dializaron contra 10 mM KPO₄ (pH 6,2), 2 mM DTT durante 16 horas y se purificaron por afinidad utilizando gel de afinidad NADP⁺ Type 3 (Pharmacia) tal como se describió anteriormente (Bascomb y Schmidt, *supra*). El NADP-GDH en las fracciones acumuladas se concentró mediante ultrafiltración Diaflow a 2 mg/ml de proteína, tal como se determinó mediante el procedimiento de Bradford [Bradford, M.M., *Anal. Biochem.*, 72:248-254 (1976)] y se almacenó a 4°C hasta su utilización. Una vez fraccionadas las proteínas mediante SDS-PAGE del 8% (peso/volumen) Tris-Tricina se determinó la pureza de la preparación mediante tinción con plata.

La Tabla 1 resume los valores K_m determinados para la reacción de aminación de tanto el isozima α - como el β -homohexamérico

TABLA 1

Isoforma GDH	Sustrato	Valor K _m (mM)
α -homohexamérico	NADPH	0,14
	NH ₄ ⁺	0,02-3,5
	α -cetoglutarato	0,35*
β -homohexamérico	NADPH	0,14
	NH ₄ ⁺	77
	α -cetoglutarato	12

*según Shatilov, V.R., W.L. Kretovich, *Mol. Cell Biochem.*, 15:201-212 (1977)

La Tabla 2 resume los valores de K_m determinados para la reacción de desaminación de tanto el isozima α - como el β -homohexamérico.

TABLA 2

Isoforma GDH	Sustrato	Valor de K _m (mM)
α -homohexamérico	NADP ⁺	0,04
	Glutamato	38,2
β -homohexamérico	NADP ⁺	0,04
	Glutamato	32,3

Actividad de los α - y β -heterohexamémeros

Las actividades aminantes y desaminantes de la mezcla de isozimas nativos NADP-GDH (heterohexamémeros compuestos de múltiples proporciones de α - y β -subunidades) se midieron también con niveles saturantes de sustratos durante los 240 minutos del período de inducción (Figura 1). Las actividades aminantes y desaminantes mostraron

retrasos de inducción iniciales de 20 y 40 minutos, respectivamente. La actividad aminante incrementó rápidamente durante los primeros 100 minutos, descendió abruptamente entre los minutos 100 y 140 e incrementó abruptamente otra vez entre 140 y 240 minutos. Por el contrario, la actividad desaminante incrementó de modo casi lineal durante la inducción después del retraso de inducción inicial.

5 Durante los 240 minutos del período de inducción en el medio con 29 mM amonio, también se examinaron los patrones de acumulación de las α - y β -subunidades de la NADP-GDH de *Chlorella sorokiniana* en isozimas mediante la utilización de un proceso de inmunodetección por transferencia Western a continuación de una electroforesis en geles de SDS poliacrilamida (ver la Figura 2). La β -subunidad de NADP-GDH se detectó a T_0 e incrementó durante 10 los primeros 40 minutos seguido de un descenso gradual a lo largo del resto del período de inducción. La α -subunidad se detectó en los primeros 20 minutos. Dicha subunidad se acumuló a poca velocidad durante los primeros 80 minutos, mostró un incremento marcado durante los 80 y 100 minutos y a continuación se acumuló de un modo lineal a una velocidad más lenta durante el resto del período de inducción. La transición desde que la β -subunidad es la especie prominente hasta que la α -subunidad es prominente tuvo lugar entre los 60 y 80 minutos.

15 El cociente de actividad aminante:desaminante y el cociente de subunidades α : β se calculó para determinar si los cambios en la proporción de subunidades en la mezcla de isozimas NADP-GDH se correlacionan con el cociente de actividad aminante:desaminante predicho durante el transcurso del período de inducción (Tabla 3). Sorprendentemente, el mayor cociente aminante:desaminante se observó a los 60 minutos cuando el cociente de las subunidades 20 demostró que la β -subunidad era el antígeno NADP-GDH predominante, mientras que la α -subunidad fue la forma predominante cuando el cociente de actividad aminante:desaminante era el más bajo. Este último resultado no fue predecible por adelantado.

25 Anteriormente al presente descubrimiento, los estudios cinéticos de los sustratos de los α - y β -homohexámeros purificados, el α -homohexámero, con su muy elevada afinidad por el amonio (relativa a la del β -homohexámero), se había asumido que era la forma de isozima con la actividad aminante superior (es decir, capacidad biosintética para la síntesis de glutamato). Los resultados sugirieron que las subunidades individuales actuaban independientemente con respecto a sus propiedades cinéticas en los homo y heterohexámeros.

30 La comparación del cociente de actividad aminante:desaminante con el cociente de subunidades α : β durante la inducción de 240 en medio 29 mM en amonio reveló una correlación inesperada entre los máximos en estas proporciones (Tabla 3).

TABLA 3

35 *Actividad aminante:desaminante de la NADP-GDH y proporción de α -subunidad: β -subunidad durante el período de inducción con amonio en *C. sorokiniana**

40	Tiempo (minutos)	Actividad Am:Deam	Subunidades α : β
	0	2,87	0,28
	20	2,96	0,58
45	40	3,81	0,49
	60	4,51	0,80
	80	3,49	1,57
	100	2,73	8,74
50	140	1,61	11,23
	240	1,13	34,79

55 El pico del cociente aminación:desaminación tuvo lugar a los 60 minutos, tiempo en el que la β -subunidad fue la prominente aunque no el antígeno exclusivo, mientras que la α -subunidad fue prominente cuando el cociente aminante:desaminante era inferior. Interesantemente, la actividad aminante fue superior cuando se encontraban presentes las dos subunidades, lo que sugiere que los heterohexámero(s), formados mediante la combinación de de las α - y β -subunidades pueden tener una actividad aminante superior que un homohexámero. Con base en que K_m del α -homohexámero purificado es muy inferior a la del β -homohexámero por el amonio, se había predicho anteriormente que el α -homohexámero tendría una actividad aminante superior que un heterohexámero cualquiera compuestos de las dos subunidades (Bascomb y Schmidt, 1987).

65

Ejemplo 2

*Secuenciación de polipéptidos y polinucleótidos*5 *Secuenciación aminoterminal de las subunidades maduras*

Una alícuota de una preparación de α -subunidad de NADP-GDH purificada (120 pmol) y una preparación parcialmente purificada de α -subunidad de NADP-GDH (80 pmol) y β -subunidad (50 pmol) se fraccionaron mediante SDS-PAGE Tris-Tricina al 8% (p/v) y se electrotransferieron a una membrana de PVDF (Immobilon-P^{8Q}, Millipore) tal como se describen Plough *et al.*, [Plough, M., A.L. Jensen, V. Barkholt, *Anal. Biochem.*, 181:33-39 (1989)]. Con el fin de evitar la acilación *in vitro* de los residuos aminotermiales de la proteína, todas las disoluciones de poliácridamida utilizadas en PAGE se trataron con la resina de lecho mixto AG501-X8 para eliminar el ácido acrílico contaminante. Se utilizó un secuenciador de fase gaseosa modelo 470A de Applied Biosystems Inc. para el análisis automatizado de la secuencia de aminoácidos por degradación Edman. Los derivados PTH-aa se identificaron mediante RP-HPLC. El análisis de la secuencia de la proteína electrotransferida lo proveyó el Interdisciplinary Center for Biotechnology Research Protein Chemistry Core Facility de la Universidad de Florida.

Se determinó la siguiente secuencia N-terminal para la α -subunidad: AVSLEEIQISAMDATTGDFTA (SEC. ID n° 5). Se determinó la siguiente secuencia N-terminal para la β -subunidad (DATTGDFTAL (SEC. ID n° 6). Estas secuencias son idénticas al ORF identificado en los dos ADNc de NADP-GDH e indican la posición de los lugares de escisión internos utilizados para eliminar la secuencia del péptido de dirección a los cloroplastos. Las secuencias de los péptidos de dirección a los cloroplastos (o péptidos de tránsito a los cloroplastos) pueden ser de utilidad para dirigir a los compartimentos celulares con estas u otras secuencias de aminoácidos. Los polinucleótidos que codifican los péptidos de tránsito a los cloroplastos se pueden utilizar con otras secuencias de polinucleótidos para codificar péptidos de tránsito a los cloroplastos.

Aislamiento y secuenciación de secuencias de ADNc

Se resuspendió 1 a 10 (peso/volumen) un precipitado de células de *C. sorokiniana* almacenado a -70°C en un tampón de disrupción de ARN; 0,1 M Tris (pH 8,5), 0,4 M LiCl., 10 mM EGTA, 5 mM EDTA, 100 unidades/ml heparina sódica (Sigma, 100 unidades/mg) y 1 mM ácido aurintricarboxílico (Sigma). La suspensión de células se centrifugó a 7000 g durante 5 minutos a 4°C y se descartó el sobrenadante. Se suspendió el precipitado de células 1 a 10 (peso/volumen) en tampón de disrupción de ARN y se rompió haciéndolo pasar a través de una prensa French a una presión de 20.000 p.s.i. Se recogió el homogeneizado celular en un tubo cónico desechable de 50 ml que contenía 0,05 veces el volumen de 20% (peso/volumen) SDS, 0,05 veces el volumen de 0,5 M EDTA (pH 8), 200 μ g/ml proteinasa K y se incubó a temperatura ambiente durante 15 minutos. Se añadió la mitad del volumen de fenol equilibrado con tampón TE (Tris 10 mM; EDTA 1mM, pH 8,0) y después de una incubación de 3 minutos se añadió medio volumen de cloroformo:alcohol isoamílico (24:1, v/v) y se mezcló durante 10 minutos en un agitador con acción de muñeca. El homogeneizado extraído se transfirió a un tubo Corex de 30 ml siliconizado y se centrifugó a 1000 g durante 10 minutos a 4°C. La fase acuosa superior se recogió y se extrajo repetidamente con un volumen igual de cloroformo:alcohol isoamílico (24:1, v/v), tal como se describió anteriormente, hasta que la interfase acuosa se hizo transparente. Después de la extracción final, la fase acuosa se combinó con un volumen igual de 2X tampón LiCl-Urea (4 M LiCl, 4 M urea, 2 mM EDTA, 1 mM ácido aurintricarboxílico; Sigma) y se precipitó el ARN en hielo durante 16 horas a 4°C. el precipitado de ARN se centrifugó a 4000 g durante 20 minutos a 4°C y el precipitado resultante se limpió una vez con tampón 1X LiCl-urea y se centrifugó otra vez para precipitar el ARN. El precipitado de ARN se disolvió en TE (pH 7,5) y se cuantificó una alícuota espectrofotométricamente a 260 nm. Después de la cuantificación, la fracción de mARN se aisló del ARN celular total utilizando un equipo de columna spin de oligo(dT). Se combinó el ARN poli(A)+ de cada preparación (50 μ g) y se utilizó para la producción comercial de una genoteca de ADNc a medida en λ Uni-ZAP XR de *C. sorokiniana* (Stratagene Cloning System, Palo Alto, CA).

La genoteca λ ZAP amplificada, que contiene 2×10^{10} pfu/ml, se plaqueó en veinte placas de petri de 150 mm a 50.000 pfu por placa, cribado de un total de 1×10^6 pfu. Las placas de fago se absorbieron para generar duplicados en membranas circulares de 132 mm Hybond-N y se trataron según el protocolo de transferencia de placas de Amersham (1985, Amersham Internacional plc, Arlington Heights, IL). Se prehibridaron las membranas en un recipiente común en 20 ml de 2X PIPES (0,8 M NaCl, 20 mM PIPES, pH 6,5), 50% (peso/volumen) formamida, 0,5% (peso/volumen) SDS, 100 μ g/ml ADN de esperma de salmón desnaturalizado y cizalleado a 40°C. Las membranas bloqueadas se hibridaron a 42°C en diez bolsas selladas térmicamente (cuatro membranas por bolsa) en tampón de prehibridación que contiene 1×10^6 cpm/membrana de la sonda de ADNc de 242 bp de NADP-GDH HCR marcada con ³²P en un agitador de laboratorio. Las membranas se limpiaron tres veces en 200 ml de 0,1 X SSC, 0,1% (peso/volumen) SDS durante 20 minutos por lavado a 50°C. Las membranas por duplicado se envolvieron en plástico y se expusieron a película Kodak X-Omat AR a -70°C durante 28 horas. Las placas de ADNc putativo de NADP-GDH, detectadas en las membranas por duplicado, se rasparon de la placa y se purificaron por plaqueado y selección secundaria y terciaria con la sonda de la región conservada de 242 bp. Las colonias de fagos de ADNc putativos de NADP-GDH, seleccionadas en el cribado inicial, se combinaron y cribaron una segunda vez con un fragmento de Eco RI/Bgl II de 130 bp de ADNc marcado con ³²P aislado del extremo 5' del clon de ADNc de NADP-GDH con el extremo 5' más completo. Se subclonaron en pBluescript KS+ (Stratagene) diez clones de NADP-GDH purificados por placa y se transformaron en *E. coli* DH5 α F⁺ (Bethesda Research Laboratories, BRL) mediante un protocolo de escisión *in vivo* provisto por Stratagene. Todos los aislamientos de plásmidos se realizaron tal como describen Kraft *et al.* (Kraft, R., J.Tardiff, K.S.

Krauter, L.A. Leinwand, *Biotechniques* 6:544-547 (1988). El análisis de las secuencias reveló que los 10 clones eran idénticos en el extremo 3' y diferían en diferentes grados en el nivel de truncamiento en sus extremos 5'. El clon más largo de ADNc con un extremo 3' completo designada pBGDc 53 (SEC. ID n° 7) no fue lo suficientemente larga para codificar cualquiera de las dos subunidades; por consiguiente, las secuencias 5' terminales se determinaron mediante PCR RACE.

Las secuencias de ADNc 5'-terminales de NADP-GDH se clonaron utilizando un proceso de PCR anclado para la amplificación rápida de los extremos de ADNc [Frohman, M.A., en D.H. Gelford, J.J. Sninsky, T.J. White, eds, *PCR Protocols*, Academic Press, San Diego, CA, pp28-38; Jain, R., R.H. Gorner, J.J. Murtagh, *Biotechniques* 12:58-59 (1992)]. Una mezcla de ARN poly(A)⁺, utilizado en la síntesis de la genoteca en λ ZAP se utilizó para clonar el extremo 5' del mARN de NADP-GDH. Se combinaron cien nanogramos de la mezcla de mARN con 10 ng de un cebador específico del gen (5'-CTCAAAGGCAAGGAACCTTCATG-3', SEC. ID n° 8), diseñada para que se hibride con la región conservada de los ARNm de NADP-GDH, se calentaron durante 5 minutos y se enfriaron en hielo. La síntesis de la primera cadena se realizó utilizando la transcriptasa reversa Superscript™ (BRL) según los protocolos provistos por el fabricante. La reacción terminada de la transcriptasa reversa se trató con una unidad de ribonucleasa H durante 20 minutos a 37°C, 5 minutos a 95°C y se extrajo una vez con cloroformo:alcohol isoamílico (24:1, vol/vol). El exceso de cebadores y dNTPs se eliminó mediante centrifugación a 2000 rpm a través de un tubo filterfuge Ultrafree-MC (exclusión a 30.000 MW, Millipore) y el retenido se concentró en 10 μ l en un Savant Speedvac. Los productos de la síntesis de la primera cadena se combinaron con 10 μ l de mezcla de cola (tailing) (1X tampón de cola (Promega Corp.), 0,4 mM dATP, 10 unidades de deoxitranferasa Terminal) y se incubaron a 37°C durante 10 minutos. La mezcla de reacción se calentó a 95°C durante 5 minutos, se diluyó a 0,5 ml con TE (pH 8) y se utilizó con un reservorio de ADNc. Una mezcla de 5 μ l del reservorio de ADNc, 5 μ l de tampón de polimerasa Vent™ 10X (New England Biolabs), 200 μ M cada uno de los dNTPs, 25 pmol de cebador específico del gen (SEC. ID n° 8), 5 pmol de cebador adaptador de poli(dT) (5'-GGGTCGACATTCTAGACAGAATTCGTGGATCC(T)₁₈-3'; SEC. ID n° 9), 0,2 unidades del amplificador Perfectmatch™ para la ADN polimerasa (Stratagene) y 1 unidad de Vent™ polimerasa (NEB) en 50 μ l se amplificaron según Jain *et al.*, *supra*. Los productos de PCR se purificaron del exceso de cebador mediante centrifugación a 2.000 rpm a través de una unidad Ultrafree-MC. El retenido se recogió y se sometió a dos ciclos más de amplificación utilizando un nuevo cebador específico anidado del gen en cada ciclo (5'-GGACGAGTACTGCACGC-3', SEC. ID n° 10; 5'-GATCTCGGTCAGCAGCTG-3', SEC. ID n° 11, respectivamente) y un cebador adaptador (5'-GGGTCGACATTCTAGACAGAA-3', SEC. ID n° 12). Las amplificaciones por PCR se realizaron en un termociclador Model 480 (Perkin-Elmer Cetus) y todos los oligonucleótidos a la medida fueron sintetizados en las instalaciones de síntesis de ADN de ICBR, Universidad de Florida. La mezcla estándar de reacción de PCR comprendió 10 μ l de tampón 10X de polimerasa Vent™, 100 μ M de cada uno de los dNTP, 0,4 unidades de Perfectmatch™, 50 pmol de cada uno de los cebadores, 1 unidad de ADN polimerasa Vent™ en 100 μ l de volumen de reacción. Los productos 5' de la RACE-PCR se purificaron mediante geles, se subclonaron en el lugar SmaI de pUC18 y se subclonaron en *E. coli* DH5 α para su posterior caracterización. La PCR-RACE identificó dos clones de ADNc 5', que se solapan con el clon pBGDc 53 anteriormente identificado, que difirieron en un inserto de 42 nt identificado en uno de los clones designado pRGDc 60 (SEC. ID n° 13) y que no se encuentra en el segundo ADNc designado pRGDc 61 (SEC. ID n° 14).

Mediante amplificación RT-PCR de mARN se construyeron dos clones de ADNc adicionales en los que falta el poliligador de PCR RACE, pero que contienen el extremo-5' completo correspondiente a pRGDc 60 y 61, utilizando las condiciones de reacción descritas anteriormente y el par de cebadores específicos del gen (5'-CTTTCTGCTCGCCCT CTC-3', SEC. ID n° 15 y SEC. ID n° 11, anterior). Los dos productos de PCR se clonaron en el lugar SmaI de pBluescript SK+ (Stratagene) y se transformaron en *E. coli* DH5 α para más caracterización. El clon de ADNc que contenía el inserto de 42 nt se designó pGDc 63 (SEC. ID n° 16) mientras que el ADNc al que faltaba el inserto se designó pGDc 64 (SEC. ID n° 17).

Se construyó un ADNc de NADP-GDH de tamaño completo mediante el tratamiento de pGDc 63 y 64 con las endonucleasas de restricción EcoR I/Apa LI y purificación mediante gel de los fragmentos resultantes (264 bp; 222 bp, respectivamente). Los fragmentos purificados mediante gel se ligaron al fragmento de restricción purificado Apa LI/XhoI de pBGDc 53 y los productos de ligación de tamaño completo (SEC. ID n° 18, SEC. ID n° 19) se purificaron en un gel de agarosa y se utilizaron en las subsiguientes reacciones de PCR.

*Expresión de los α - y β -homohexámeros en *E. coli**

Utilizando el producto purificado mediante gel (SEC. ID n° 18), se realizó mutagénesis por PCR para eliminar la señal de direccionamiento a los cloroplastos del ADNc de tamaño completo dando ADNc que codifican específicamente las α - y β -subunidades maduras. Se diseñaron dos conjuntos de pares de cebadores para sintetizar los genes de las subunidades α - y β -GDH.

Se diseñó el siguiente cebador con el fin de añadir una metionina al extremo amino de la subunidad α -NADP-GDH madura y procesada (alanina-41) para permitir la inicialización de la traducción y generar un lugar NdeI 5' con fines de subclonación: 5'-CATATGGCCGTCTCGCTGGAGGAG-3' (SEC. ID n° 20). El siguiente segundo cebador se diseñó con el fin de hibridarlo al extremo 3' del ADN molde en la posición 20 nt 3' del codón de terminación TAA endógeno: 5'-GTTGGATTGCCGGTGAGCC-3' (SEC. ID n° 21).

El siguiente cebador se diseñó con el fin de añadir metionina al extremo amino del la β -subunidad procesada y madura (aspartato-38) para permitir la iniciación de traducción y para generar un lugar NdeI 5' con fines de subclonación.

ES 2 268 711 T3

naje: 5'-CATATGGACGCCACCACCGGC-3' (SEC. ID n° 22). El segundo cebador 3' utilizado en la amplificación mediante PCR fue el cebador de extremo 3' (SEC. ID n° 21) descritos para la amplificación de la α -subunidad.

Las condiciones para los ciclos de PCR fueron las siguientes: 95°C, 50 segundos; 64°C, 1 minuto; 72°C, 1 minuto 35 segundos (30 ciclos). Los cebadores, dNTP, polimerasa Vent y las concentraciones de otros componentes de la reacción fueron tal como se describió anteriormente. Los productos de PCR del gen de la subunidad α - NADP-GDH de 1506 bp (SEC. ID n° 23) y del gen de la subunidad β - NADP-GDH de 1473 bp (SEC ID n° 25) se purificaron mediante geles y provistos del nucleótido adenina colgante en el extremo 3' mediante la incubación del fragmento purificado con 100 μ M dATP y polimerasa Taq durante 15 minutos a 72°C. Los productos de PCR modificados se clonaron en el vector de clonaje PCRII T/A (Invitrogen) y se transformaron en células competentes de *E. coli*. Los clones portadores de los insertos se seleccionaron mediante selección de colonias azules-blancas, se purificó el plásmido y se digirió con Nde I/BamHI para seleccionar la orientación adecuada en un segundo vector de clonaje. Los plásmidos seleccionados se restringieron mediante NdeI y BamHI (lugar BamHI provisto por el vector) y se clonaron en la dirección adecuada bajo el control del promotor de la polimerasa T7 inducible por IPTG del vector de expresión bacteriano pET 11a y pET 15b bacteriano de (Novagen) linealizado con Nde I/BamHI y se transformaron en DH5 α . Los transformantes se cribaron mediante análisis de restricción con Nde I/BamHI y se seleccionaron los clones provistos de los ADNc de las α - y β -subunidades (SEC. ID n° 23; SEC. ID n° 25), se purificó el plásmido y se transformó en *E. coli* BL21(DE3) con el fin de expresar proteína.

Las células de *E. coli* BL21(DE) transformadas con las construcciones pET 11a- α -ADNc y pET 11a- β -ADNc se indujeron con 100 mM IPTG durante 1 hora. Los extractos de proteína extraídos de las células inducidas se ensayaron enzimáticamente para determinar actividad NADP-GDH y las proteínas desnaturalizadas se fraccionaron mediante electroforesis en geles de SDS y se visualizaron mediante tinción con azul coomassie. Las proteínas expresadas por el ADNc de la α -subunidad madura (SEC. ID n° 23) y el ADNc de la β -subunidad (SEC. ID n° 25) tiene la secuencia de aminoácidos mostradas por las SEC. ID n° 24 (α -subunidad) y la SEC. ID n° 26 (β -subunidad). Las unidades de GDH recombinante se verificaron mediante la reactividad cruzada con anticuerpos de conejo antiNADP-GDH de *Chlorella*.

En condiciones no optimizadas para una máxima inducción, las células de *E. coli*, que poseen ADNc de α - y β -GDH y están inducidas por IPTG, mostraron un incremento en la actividad de NADP-GDH comprendido entre 60- y 7.000 veces relativo a los controles no inducidos, respectivamente. Las α - y β - NADP-GDH están actualmente siendo analizadas para verificar sus propiedades cinéticas y bioquímicas.

La sobre expresión y el ensamblaje de las GDH cloroplásticas de *C. sorokiniana* en enzimas activos proporciona la prueba de que las construcciones de ADN construidas mediante PCR se transcriben y traducen en proteínas auténticas. Las construcciones mencionadas anteriormente, a continuación se utilizaron en la expresión citosólica de las GDH de alga en vegetales transgénicos.

Transformación de vegetales

Un procedimiento para producir vegetales genéticamente transformados que expresan niveles amplificados de una GDH específica necesita la introducción de una molécula recombinante de ADN de doble cadena en el genoma nuclear de una célula vegetal. La molécula de ADN debe (1) contener un ADN estructural para el enzima GDH que se debe introducir en la célula vegetal; (2) disponer de un promotor que funciona en los vegetales para regular la producción de secuencias de ARN de modo constitutivo o tejido específico mediante el enzima ARN polimerasa; y (3) disponer de una región 3'-no traducida que funciona para inducir la terminación de la transcripción y la adición de nucleótidos poliadenilados al extremo 3' del ARN. La molécula de ARN primaria resultante a continuación se procesa en el núcleo, un proceso que implica la eliminación de las secuencias intrónicas y la adición de nucleótidos poliadenilados al extremo 3' del mRNA.

Los promotores de utilidad en la presente invención son los que pueden iniciar la transcripción de modo constitutivo o de modo tejido específico en la que la producción o catabolismo del glutamato son deseables. Un ejemplo de promotor constitutivo de utilidad es el promotor amplificado CaMV 35S que dirige la síntesis de ARN de un modo tejido específico. Los promotores que causan la producción de GDH específicamente en semillas, tallos, raíces, hojas o tipos celulares específicos en tales tejidos son de utilidad en la presente invención. Por ejemplo, el promotor específico de las semillas de la Phaseolina es uno de tales promotores tejido específicos. Por consiguiente, los promotores nativos específicos para maíz, trigo, centeno y arroz se pueden obtener y utilizar en la presente invención así como también como los promotores heterólogos de otros organismos que se ha demostrado que funcionan de modo constitutivo/tejido específico.

Intrones

Generalmente, la expresión óptima en las plantas monocotiledóneas se obtiene cuando se inserta una secuencia de intrón entre la secuencia de promotor y la secuencia del gen estructural. Un ejemplo de tal secuencia intrónica es el intrón HSP 70 descrito en el documento WO 93/19189.

Señal de poliadenilación

Las construcciones de ADN de la presente invención pueden poseer una región 3' no traducida que funciona en los vegetales para dirigir la adición de nucleótidos poliadenilados al extremo 3' del ARN. Un ejemplo de región 3' no traducida adecuada es la señal de poliadenilación del plásmido inductor de tumores de *Agrobacterium*, es decir, el gen de la nopalina sintetasa el gen (NOS).

Secuencia directora a los plastos

Las construcciones de ADN de la presente invención pueden opcionalmente contener secuencias directoras a los plastos. Las secuencias directoras a los plastos dirigen la importación de las proteínas a los plastos y se elimina durante la importación. La secuencia directora a los plastos puede ser, de manera no limitativa, el péptido nativo de dirección a cloroplastos (CTP) identificado en ADNc de tamaño completo de la NADP-GDH de *C. sorokiniana* que codifica las proteínas precursoras. Una fusión de una secuencia directora a plastos seleccionada y las secuencias de las subunidades α - y β -NADP-GDH maduras se puede realizar mediante procedimientos estándar y ser utilizada en la presente invención. Las subunidades de GDH que carecen de estas secuencias de dirección se encuentran típicamente en el citoplasma de la célula. Dicha localización citosólica del enzima puede ser de utilidad en la captura de de amonio o glutamato compartimentalizado en el citosol de la célula.

Fuentes del gen GDH

El gen GDH utilizado en las construcciones de ADN de la presente invención puede ser cualquier gen GDH. No queda limitado a los genes de GDH de *C. sorokiniana* descritos anteriormente, aunque éstos son preferidos. Por ejemplo, se puede utilizar un gen GDH de bacteria u hongo. Los ejemplos proporcionados utilizan los genes α - y β -GDH de *C. sorokiniana*, pero no se deben interpretar como limitativos del alcance de la presente invención en modo alguno. Los expertos en la materia reconocerán que se puede utilizar otra multiplicidad de genes así como también alteraciones a los genes y procedimientos descritos en la presente invención sin apartarse del espíritu y alcance de la presente invención. Por ejemplo, se puede implementar la mutagénesis y el cribado estándar mediante procedimientos conocidos en la materia para generar variantes mutantes que carecen de la regulación mediante el cofactor NADPH.

Expresión transitoria en protoplastos de maíz

Para ensayar la expresión de las subunidades de GDH de *C. sorokiniana* y su ensamblaje en enzimas activos en células de *Zea mays*, se construyeron vectores que contienen el promotor CaMV E35S, la secuencia codificante de la α -subunidad madura (pMON2194) o la β -subunidad (pMON21905), la región de poliadenilación 3' no traducida NOS y la de resistencia a la kanamicina para seleccionar en *E. coli*. Los genes de las α - y β -subunidades se aislaron como fragmentos Xba I-EcoR I a partir de pET 11a- α -ADNc y pET 11a- β -ADNc, respectivamente. Se ligaron los genes GDH en la región portadora de resistencia a kanamicina en el promotor Xba I-EcoR I E35S, NOS '3 de pMON22072 para dar pMON21904 y pMON21905. Las construcciones de ADN se electroporaron en protoplastos de maíz y trigo según el procedimiento de Sheen *et al.* (The plant cell Vol 3, 225-245).

Análisis de los protoplastos de maíz transformados

Los precipitados de muestras de protoplastos transformados con pMON21904 (α -subunidad), pMON21905 (β -subunidad) pMON21709 (ADN de control kanamicina negativo) y no-ADN se descongelaron en 0,2 ml de tampón de lisis de células GDH (Yeung *et al, supra*) en hielo. Las células en cada suspensión se homogeneizaron dos veces durante 30 segundos, se enfriaron en hielo y se aclararon a 14.000 rpm durante 10 minutos. Los extractos de las células se ensayaron en la dirección de desaminación a 38,5°C según Yeung *et al., supra*. El contenido total de los extractos celulares se determinó utilizando el ensayo de microproteínas de BioRad según el protocolo de los fabricantes. Las actividades se normalizaron contra el contenido total de proteína para las comparaciones entre las diferentes preparaciones. Una unidad de actividad de GDH se define como la cantidad de enzima necesaria para reducir 1 μ mol de NADP por minuto a 38,5°C.

Los protoplastos transformados con el vector control pMON21709 (n = 3) o los protoplastos no transformados (n = 3) no presentaron actividad detectable NADP-GDH detectable. Los protoplastos transformados con pMON21904 (n = 3) expresaron 3,31 Unidades-mg⁻¹ de proteína con actividad GDH, mientras que los protoplastos transformados por pMON21905 (n = 3) tuvieron 1,96 Unidades-mg⁻¹ proteína.

El elevado nivel de actividad observado en los protoplastos transformados con los genes de α - y β - NADP-GDH de *C. sorokiniana* expresados en el citoplasma proporcionan pruebas de que las subunidades GDH se expresan en sistemas vegetales heterólogos. Además, los niveles de expresión demuestran que las subunidades se ensamblan en enzimas activos. Generalmente, resultaría evidente para los expertos en la materia que las secuencias superfluas añadidas a las secuencias descritas, o los fragmentos de las secuencias de nucleótidos o aminoácidos descritos en la presente memoria, que resultan en secuencias de polinucleótidos o aminoácidos que funcionan de modo semejante o equivalente al de las secuencias descritas expresamente en la presente memoria, se deben asimismo considerar parte de la presente invención. Se pueden producir fácilmente y rutinariamente mediante procedimientos conocidos en la materia, por ejemplo, mediante la digestión controlada en el tiempo con la exonucleasa Bal31 del ADN de tamaño completo, seguido de expresión de los fragmentos resultantes y cribado de rutina de los productos de expresión tal

ES 2 268 711 T3

como se describe en los ejemplos anteriores. Además, resultaría fácilmente evidente para los expertos en la materia que las funciones, propiedades o utilidad de las secuencias descritas pueden ser negadas mediante la inserción de mutaciones en las secuencias mediante la utilización de procedimientos y técnicas estándar. Tales mutaciones que, mediante implicación, sirven para eliminar efectivamente la propiedad o función inherente a la secuencia tal como se describió, se incluyen en la presente memoria expresamente como parte de la presente invención. Por ejemplo, una clara distinción entre las α - y β -subunidades de *C. sorokiniana* es la secuencia de polipéptido de 11 aminoácidos en el extremo N-terminal de la α -subunidad, pero que está ausente en la β -subunidad. Dicha secuencia puede afectar la afinidad, especificidad y modulación de los compuestos de amonio por el enzima. Por consiguiente, sería evidente que la inserción (si se encuentra ausente) o eliminación (si se encuentra presente) de la secuencia apropiada, o su equivalente funcional, para producir una diferencia en ciertas características de otros genes de GDH, o sus productos, sería fácil de realizar por tales personas.

15

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

- 5 1. Polinucleótido que comprende una secuencia de nucleótidos que codifica una secuencia de aminoácidos seleccionada de entre SEC. ID nº 2; SEC. ID nº 4; SEC. ID nº 24 y SEC. ID nº 26 o un fragmento de la misma que presenta actividad glutamato deshidrogenasa (GDH).
- 10 2. Polinucleótido que comprende una secuencia de nucleótidos seleccionada de entre SEC. ID nº 1; SEC. ID nº 3; SEC. ID nº 18; SEC. ID nº 19; SEC. ID nº 23 y SEC. ID nº 25; o un fragmento de la misma que codifica un péptido que presenta actividad GDH.
- 15 3. Polinucleótido según la reivindicación 2, que comprende una secuencia de nucleótidos seleccionada de entre SEC. ID nº 1; SEC. ID nº 3; SEC. ID nº 23 y SEC. ID nº 25, o un fragmento de la misma que codifica un péptido que presenta actividad GDH.
- 20 4. Polinucleótido según la reivindicación 2, que forma un gen quimérico capaz de expresión en células vegetales, estando dicho polinucleótido unido funcionalmente a un promotor expresable en vegetales y a una secuencia de poliadenilación vegetal.
- 25 5. Polinucleótido según la reivindicación 2, que comprende por lo menos una secuencias de nucleótidos que codifica una subunidad alfa de NADP-GDH y por lo menos una secuencia de nucleótidos que codifica una subunidad beta de NADP-GDH.
- 30 6. Polinucleótido según la reivindicación 5, en la que dicho polinucleótido comprende además un promotor para dicho polinucleótido que codifica una subunidad alfa de NADP-GDH y una subunidad beta de NADP-GDH.
- 35 7. Polinucleótido según la reivindicación 6, que comprende diferentes promotores para la subunidad alfa y para la subunidad beta.
- 40 8. Polinucleótido según la reivindicación 6, en el que dicho promotor tiene una actividad diferencial para las subunidades alfa y beta.
- 45 9. Polinucleótido según la reivindicación 5, que comprende una secuencia de nucleótidos seleccionada de entre la SEC. ID nº 1 y la SEC. ID nº 23 y una secuencia de nucleótidos seleccionada de entre SEC. ID nº 3 y la SEC. ID nº 25.
- 50 10. Célula huésped transformada que comprende un polinucleótido según la reivindicación 1.
- 55 11. Célula huésped transformada según la reivindicación 10, en la que dicho polinucleótido comprende una secuencia de nucleótidos seleccionada de entre SEC. ID nº 1, SEC. ID nº 3; SEC. ID nº 23 y SEC. ID nº 25.
- 60 12. Procedimiento para modular el metabolismo del nitrógeno en una célula vegetal, que comprende transformar dicha célula vegetal para que comprenda un polinucleótido según la reivindicación 1, y mantener dicha célula o descendientes de la misma en condiciones en las que dicho polipéptido se expresa.
- 65 13. Procedimiento según la reivindicación 12, en el que dicho polinucleótido tiene una secuencia de nucleótidos que codifica por lo menos una subunidad alfa de NADP-GDH y por lo menos una secuencia de nucleótidos que codifica una subunidad beta de NADP-GDH.
14. Procedimiento según la reivindicación 13, en el que dicho polinucleótido codifica un polipéptido que comprende una secuencia de aminoácidos seleccionada de entre SEC. ID nº 2 y SEC. ID nº 24 y una secuencia de aminoácidos seleccionada de entre SEC. ID nº 4 y SEC. ID nº 26.
15. Procedimiento según cualquiera de las reivindicaciones 12 a 14, en el que dicho polinucleótido se encuentra unido funcionalmente a un promotor expresable en los vegetales y a una secuencia de poliadenilación vegetal.
16. Procedimiento según cualquiera de las reivindicaciones 12 a 15, en el que la asimilación de nitrógeno inorgánico en nitrógeno orgánico es incrementada.
17. Procedimiento según cualquiera de las reivindicaciones 12 a 16, en el que se incrementa el rendimiento de las cosechas, o la asimilación de amonio, o la tolerancia al estrés osmótico o la composición de la planta es alterada.
18. Polipéptido que comprende una secuencia de aminoácidos seleccionada de entre SEC. ID nº 2; SEC. ID nº 4; SEC. ID nº 5; SEC. ID nº 6; SEC. ID nº 24 y SEC. ID nº 26 o un fragmento de las mismas que presenta actividad GDH.

ES 2 268 711 T3

19. Polipéptido que comprende por lo menos una subunidad alfa de NADP-GDH que comprende una secuencia de aminoácidos seleccionada de entre SEC. ID n° 2 y SEC. ID n° 24, o un fragmento de las mismas que presenta actividad GDH, y por lo menos una subunidad beta de NADP-GDH.

5 20. Polipéptido según la reivindicación 19, en la que la subunidad beta comprende una secuencia de aminoácidos seleccionada de entre SEC. ID n° 4 y SEC. ID n° 26.

10

15

20

25

30

35

40

45

50

55

60

65

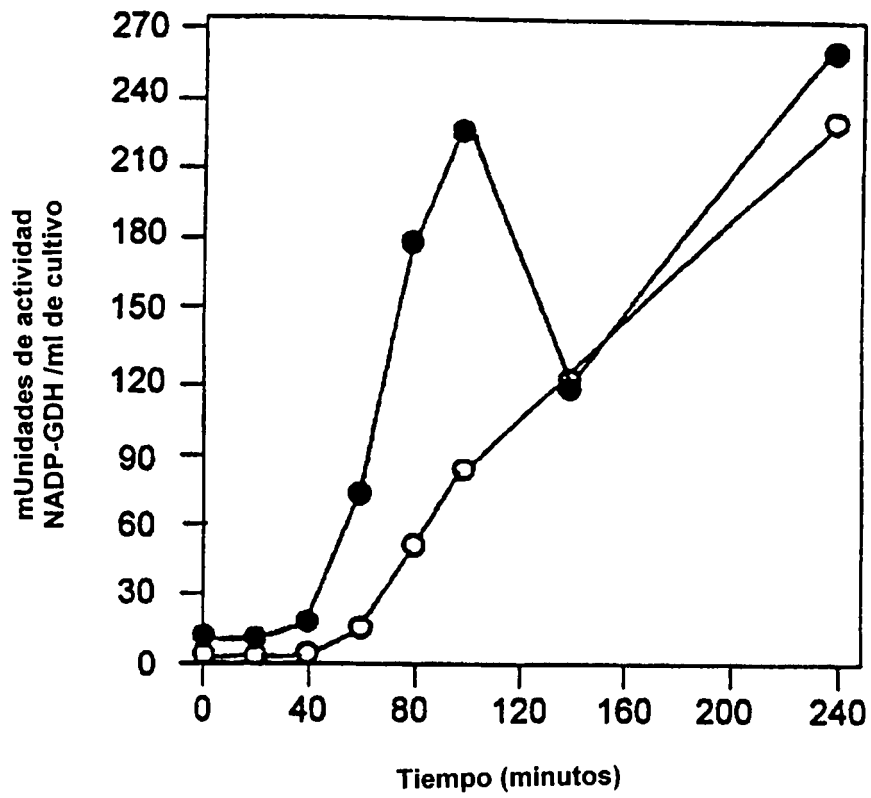


FIG. 1

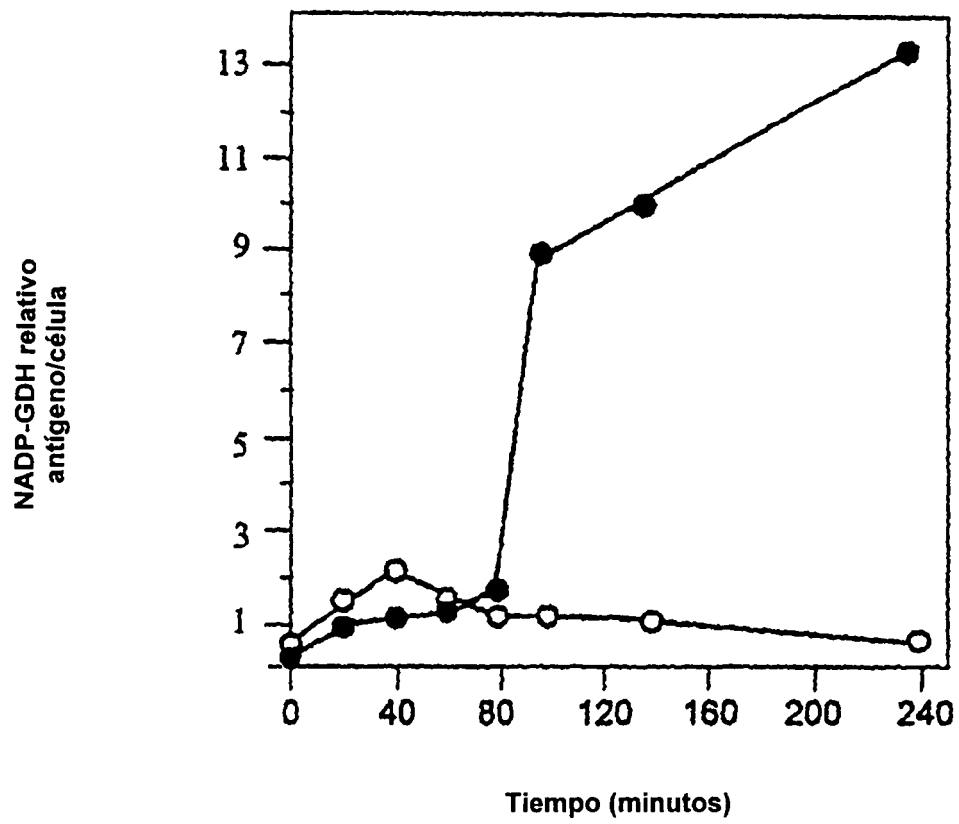


FIG. 2

ES 2 268 711 T3

LISTA DE SECUENCIAS

(1) Información General:

- 5 (i) Información del solicitante:
Nombre del solicitante(s): University of Florida
Dirección: 223 Grinter Hall
Ciudad: Gainesville
10 Estado/provincia: Florida
País: US
Código postal: 32611
Número de teléfono: (352) 392 8929 FAX: (352) 392 6600
- 15 (ii) Título de la invención: Polipéptidos y polinucleótidos relacionados con las α - y β -subunidades de las glutamato deshidrogenasas y procedimientos de utilización
- (iii) Número de las secuencias: 26
- 20 (iv) Dirección de correspondencia:
(A) Remitente: Saliwanchik & Saliwanchik
(B) Calle: 2421 N.W. 41st Street, Suite A-1
25 (C) Ciudad: Gainesville
(D) Estado: Florida
(E) País: USA
(F) Código postal: 32602
- 30 (v) Forma legible por ordenador:
(A) Tipo de medio: disquete
(B) Ordenador: IBM PC compatible
35 (C) Sistema operativo: PC-DOS/MS-DOS
(D) Software: PatentIn publicación # 1.0, versión # 1.25
- (vi) Datos de la solicitud actual:
40 (A) Numero de solicitud: US
(B) Fecha de solicitud:
(C) Clasificación:
- 45 (vii) Información del abogado/agente:
(A) Nombre: Whitelock, Ted W.
(B) Número de registro: 36.965
50 (C) Numero de referencia/certificado: UF155
- (viii) Información de telecomunicaciones:
(A) Teléfono: (904) 375 8100
55 (B) Telefax: (904) 372 5800

(2) Información de SEC. ID nº 1:

- (i) Características de la secuencia:
60 (A) Longitud: 2140 pares de bases
(B) Tipo: ácido nucleico
(C) Tipo de cadena: doble
65 (D) Topología: lineal
- (ii) Tipo de molécula: ADNc
- (ix) Característica:

ES 2 268 711 T3

(A) Nombre/código: CDS

(B) Localización: 33..1610

(xi) Descripción de la secuencia: SEC. ID nº 1:

5

**CTCCTTTCTG CTCGCCCTCT CTCGGTCCCG CC ATG CAG ACC GCC CTC GTC GCC
53**

10

Met Gln Thr Ala Leu Val Ala
1 5

**AAG CCT ATC GTG GCC GCC CCG CTG GCG GCA CGC CCG CGC TGC CTC GCG
101**

15

**Lys Pro Ile Val Ala Ala Pro Leu Ala Ala Arg Pro Arg Cys Leu Ala
10 15 20**

**CCG TGG CCG TGC GCG TGG GTC CGC TCC GCC AAG CGC GAT GTC CGC GCC
149**

20

**Pro Trp Pro Cys Ala Trp Val Arg Ser Ala Lys Arg Asp Val Arg Ala
25 30 35**

**AAG GCC GTC TCG CTG GAG GAG CAG ATC TCC GCG ATG GAC GCC ACC ACC
197**

25

**Lys Ala Val Ser Leu Glu Glu Gln Ile Ser Ala Met Asp Ala Thr Thr
40 45 50 55**

**GGC GAC TTC ACG GCG CTG CAG AAG GCG GTG AAG CAG ATG GCC ACC AAG
245**

30

**Gly Asp Phe Thr Ala Leu Gln Lys Ala Val Lys Gln Met Ala Thr Lys
60 65 70**

**GCG GGC ACT GAG GGC CTG GTG CAC GGC ATC AAG AAC CCC GAC GTG CGC
293**

35

**Ala Gly Thr Glu Gly Leu Val His Gly Ile Lys Asn Pro Asp Val Arg
75 80 85**

40

45

50

55

60

65

ES 2 268 711 T3

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

CAG CTG CTG ACC GAG ATC TTC ATG AAG GAC CCG GAG CAG CAG GAG TTC
 341
 Gln Leu Leu Thr Glu Ile Phe Met Lys Asp Pro Glu Gln Gln Glu Phe
 90 95 100

 ATG CAG GCG GTG CGC GAG GTG GCC GTC TCC CTG CAG CCC GTG TTC GAG
 389
 Met Gln Ala Val Arg Glu Val Ala Val Ser Leu Gln Pro Val Phe Glu
 105 110 115

 AAG CGC CCC GAG CTG CTG CCC ATC TTC AAG CAG ATC GTT GAG CCT GAG
 437
 Lys Arg Pro Glu Leu Leu Pro Ile Phe Lys Gln Ile Val Glu Pro Glu
 120 125 130 135

 CGC GTG ATC ACC TTC CGC GTG TCC TGG CTG GAC GAC GCC GGC AAC CTG
 485
 Arg Val Ile Thr Phe Arg Val Ser Trp Leu Asp Asp Ala Gly Asn Leu
 140 145 150

 CAG GTC AAC CGC GGC TTC CGC GTG CAG TAC TCG TCC GCC ATC GGC CCC
 533
 Gln Val Asn Arg Gly Phe Arg Val Gln Tyr Ser Ser Ala Ile Gly Pro
 155 160 165

 TAC AAG GGC GGC CTG CGC TTC CAC CCC TCC GTG AAC CTG TCC ATC ATG
 581
 Tyr Lys Gly Gly Leu Arg Phe His Pro Ser Val Asn Leu Ser Ile Met
 170 175 180

 AAG TTC CTT GCC TTT GAG CAG ATC TTC AAG AAC AGC CTG ACC ACC CTG
 629
 Lys Phe Leu Ala Phe Glu Gln Ile Phe Lys Asn Ser Leu Thr Thr Leu
 185 190 195

 CCC ATG GGC GGC GGC AAG GGC GGC TCC GAC TTC GAC CCC AAG GGC AAG
 677
 Pro Met Gly Gly Gly Lys Gly Gly Ser Asp Phe Asp Pro Lys Gly Lys
 200 205 210 215

 AGC GAC GCG GAG GTG ATG CGC TTC TGC CAG TCC TTC ATG ACC GAG CTG
 725
 Ser Asp Ala Glu Val Met Arg Phe Cys Gln Ser Phe Met Thr Glu Leu
 220 225 230

 CAG CGC CAC ATC AGC TAC GTG CAG GAC GTG CCC GCC GGC GAC ATC GGC
 773
 Gln Arg His Ile Ser Tyr Val Gln Asp Val Pro Ala Gly Asp Ile Gly
 235 240 245

 GTG GGC GCG CGC GAG ATT GGC TAC CTT TTC GGC CAG TAC AAG CGC ATC
 821
 Val Gly Ala Arg Glu Ile Gly Tyr Leu Phe Gly Gln Tyr Lys Arg Ile
 250 255 260

 ACC AAG AAC TAC ACC GGC GTG CTG ACC CCG AAG GGC CAG GAG TAT GGC
 869
 Thr Lys Asn Tyr Thr Gly Val Leu Thr Pro Lys Gly Gln Glu Tyr Gly
 265 270 275

 GGC TCC GAG ATC CGC CCC GAG GCC ACC GGC TAC GGC GCC GTG CTG TTT
 917
 Gly Ser Glu Ile Arg Pro Glu Ala Thr Gly Tyr Gly Ala Val Leu Phe
 280 285 290 295

ES 2 268 711 T3

5 GTG GAG AAC GTG CTG AAG GAC AAG GGC GAG AGC CTC AAG GGC AAG CGC
 965
 Val Glu Asn Val Leu Lys Asp Lys Gly Glu Ser Leu Lys Gly Lys Arg
 300 305 310

10 TGC CTG GTG TCT GGC GCG GGC AAC GTG GCC CAG TAC TGC GCG GAG CTG
 1013
 Cys Leu Val Ser Gly Ala Gly Asn Val Ala Gln Tyr Cys Ala Glu Leu
 315 320 325

15 CTG CTG GAG AAG GGC GCC ATC GTG CTG TCG CTG TCC GAC TCC CAG GGC
 1061
 Leu Leu Glu Lys Gly Ala Ile Val Leu Ser Leu Ser Asp Ser Gln Gly
 330 335 340

20 TAC GTG TAC GAG CCC AAC GGC TTC ACG CGC GAG CAG CTG CAG GCG GTG
 1109
 Tyr Val Tyr Glu Pro Asn Gly Phe Thr Arg Glu Gln Leu Gln Ala Val
 345 350 355

25 CAG GAC ATG AAG AAG AAG AAC AAC AGC GCC CGC ATC TCC GAG TAC AAG
 1157
 Gln Asp Met Lys Lys Lys Asn Asn Ser Ala Arg Ile Ser Glu Tyr Lys
 360 365 370 375

30 AGC GAC ACC GCC GTG TAT GTG GGC GAC CGC CGC AAG CCT TGG GAG CTG
 1205
 Ser Asp Thr Ala Val Tyr Val Gly Asp Arg Arg Lys Pro Trp Glu Leu
 380 385 390

35 GAC TGC CAG GTG GAC ATC GCC TTC CCC TGC GCC ACC CAG AAC GAG ATC
 1253
 Asp Cys Gln Val Asp Ile Ala Phe Pro Cys Ala Thr Gln Asn Glu Ile
 395 400 405

40 GAT GAG CAC GAC GCC GAG CTG CTG ATC AAG CAC GGC TGC CAG TAC GTG
 1301
 Asp Glu His Asp Ala Glu Leu Leu Ile Lys His Gly Cys Gln Tyr Val
 410 415 420

45 GTG GAG GGC GCC AAC ATG CCC TCC ACC AAC GAG GCC ATC CAC AAG TAC
 1349
 Val Glu Gly Ala Asn Met Pro Ser Thr Asn Glu Ala Ile His Lys Tyr
 425 430 435

50 AAC AAG GCC GGC ATC ATC TAC TGC CCC GGC AAG GCG GCC AAC GCC GGC
 1397
 Asn Lys Ala Gly Ile Ile Tyr Cys Pro Gly Lys Ala Ala Asn Ala Gly
 440 445 450 455

55 GGC GTG GCG GTC AGC GGC CTG GAG ATG ACC CAG AAC CGC ATG AGC CTG
 1445
 Gly Val Ala Val Ser Gly Leu Glu Met Thr Gln Asn Arg Met Ser Leu
 460 465 470

60 AAC TGG ACT CGC GAG GAG GTT CGC GAC AAG CTG GAG CGC ATC ATG AAG
 1493
 Asn Trp Thr Arg Glu Glu Val Arg Asp Lys Leu Glu Arg Ile Met Lys
 475 480 485

65 GAC ATC TAC GAC TCC GCC ATG GGG CCG TCC CGC AGA TAC AAT GTT GAC
 1541
 Asp Ile Tyr Asp Ser Ala Met Gly Pro Ser Arg Arg Tyr Asn Val Asp
 490 495 500

ES 2 268 711 T3

5
 10
 15
 20
 25
 30
 35

CTG GCT GCG GGC GCC AAC ATC GCG GGC TTC ACC AAG GTG GCT GAT GCC
 1589
 Leu Ala Ala Gly Ala Asn Ile Ala Gly Phe Thr Lys Val Ala Asp Ala
 505 510 515

GTC AAG GCC CAG GGC GCT GTT TAAGCTGCCC AGGCCCAAGC CACGGCTCAC
 1640
 Val Lys Ala Gln Gly Ala Val
 520 525

CGGCAATCCA ACCCAACCAA CTCAACGGCC AGGACCTTTT CGGAAGCGGC GCCTTTTTC
 1700

CAGCCAGGGC CCTCACCTGC CCTTTCATAA CCCTGCTATT GCCGCCGTGC CCCTGCAATT
 1760

CCACCCCAAG AAGAACTAGC GGCACCTGAC TGCATCAGGA CGGCTATTTT TTTCGCGACG
 1820

CGCGCTCACC CCGAGAGCCT CTCTCCCCCG AGCCCTAAGC GCTGACGTCG GCCCGACTTT
 1880

GCCTCGCACA TCGCTCGGTT TTGACCCCT CCAGTCTACC CACCCTGTTG TGAAGCCTAC
 1940

CAGCTCAATT GCCTTTTAGT GTATGTGCGC CCOCTCCTGC CCCCGAATTT TCCTGCCATG
 2000

AGACGTGCGG TTCCTAGCCT GGTGACCCCA AGTAGCAGTT AGTGTGCGTG CCTTGCCCTG
 2060

CGCTGCCCGG GATGCGATAC TGTGACCTGA GAGTGCTTGT GTAACACGA CGAGTCAAAA
 2120

AAAAAAAAA AAAAAAAAAA
 2140

(2) Información para SEC. ID nº 2

- 40 (i) Características de la secuencia:
 (A) Longitud: 526 amino ácidos
 (B) Tipo: aminoácido
 (D) Topología: lineal
- 45 (ii) Tipo de molécula: proteína
- (xi) Descripción de la secuencia: SEC. ID nº 2

50 Met Gln Thr Ala Leu Val Ala Lys Pro Ile Val Ala Ala Pro Leu Ala
 1 5 10 15

Ala Arg Pro Arg Cys Leu Ala Pro Trp Pro Cys Ala Trp Val Arg Ser
 20 25 30

55 Ala Lys Arg Asp Val Arg Ala Lys Ala Val Ser Leu Glu Glu Gln Ile
 35 40 45

60 Ser Ala Met Asp Ala Thr Thr Gly Asp Phe Thr Ala Leu Gln Lys Ala
 50 55 60

65

ES 2 268 711 T3

Val Lys Gln Met Ala Thr Lys Ala Gly Thr Glu Gly Leu Val His Gly
 65 70 75 80
 5 Ile Lys Asn Pro Asp Val Arg Gln Leu Leu Thr Glu Ile Phe Met Lys
 85 90 95
 Asp Pro Glu Gln Gln Glu Phe Met Gln Ala Val Arg Glu Val Ala Val
 100 105 110
 10 Ser Leu Gln Pro Val Phe Glu Lys Arg Pro Glu Leu Leu Pro Ile Phe
 115 120 125
 Lys Gln Ile Val Glu Pro Glu Arg Val Ile Thr Phe Arg Val Ser Trp
 130 135 140
 15 Leu Asp Asp Ala Gly Asn Leu Gln Val Asn Arg Gly Phe Arg Val Gln
 145 150 155 160
 Tyr Ser Ser Ala Ile Gly Pro Tyr Lys Gly Gly Leu Arg Phe His Pro
 165 170 175
 20 Ser Val Asn Leu Ser Ile Met Lys Phe Leu Ala Phe Glu Gln Ile Phe
 180 185 190
 Lys Asn Ser Leu Thr Thr Leu Pro Met Gly Gly Gly Lys Gly Gly Ser
 195 200 205
 Asp Phe Asp Pro Lys Gly Lys Ser Asp Ala Glu Val Met Arg Phe Cys
 210 215 220
 30 Gln Ser Phe Met Thr Glu Leu Gln Arg His Ile Ser Tyr Val Gln Asp
 225 230 235 240
 Val Pro Ala Gly Asp Ile Gly Val Gly Ala Arg Glu Ile Gly Tyr Leu
 245 250 255
 35 Phe Gly Gln Tyr Lys Arg Ile Thr Lys Asn Tyr Thr Gly Val Leu Thr
 260 265 270
 Pro Lys Gly Gln Glu Tyr Gly Gly Ser Glu Ile Arg Pro Glu Ala Thr
 275 280 285
 40 Gly Tyr Gly Ala Val Leu Phe Val Glu Asn Val Leu Lys Asp Lys Gly
 290 295 300
 Glu Ser Leu Lys Gly Lys Arg Cys Leu Val Ser Gly Ala Gly Asn Val
 305 310 315 320
 45 Ala Gln Tyr Cys Ala Glu Leu Leu Leu Glu Lys Gly Ala Ile Val Leu
 325 330 335
 Ser Leu Ser Asp Ser Gln Gly Tyr Val Tyr Glu Pro Asn Gly Phe Thr
 340 345 350
 50 Arg Glu Gln Leu Gln Ala Val Gln Asp Met Lys Lys Lys Asn Asn Ser
 355 360 365
 Ala Arg Ile Ser Glu Tyr Lys Ser Asp Thr Ala Val Tyr Val Gly Asp
 370 375 380
 55 Arg Arg Lys Pro Trp Glu Leu Asp Cys Gln Val Asp Ile Ala Phe Pro
 385 390 395 400
 Cys Ala Thr Gln Asn Glu Ile Asp Glu His Asp Ala Glu Leu Leu Ile
 405 410 415
 60 Lys His Gly Cys Gln Tyr Val Val Glu Gly Ala Asn Met Pro Ser Thr
 420 425 430

65

ES 2 268 711 T3

Asn Glu Ala Ile His Lys Tyr Asn Lys Ala Gly Ile Ile Tyr Cys Pro
435 440 445
Gly Lys Ala Ala Asn Ala Gly Gly Val Ala Val Ser Gly Leu Glu Met
450 455 460
Thr Gln Asn Arg Met Ser Leu Asn Trp Thr Arg Glu Glu Val Arg Asp
465 470 475 480
Lys Leu Glu Arg Ile Met Lys Asp Ile Tyr Asp Ser Ala Met Gly Pro
485 490 495
Ser Arg Arg Tyr Asn Val Asp Leu Ala Ala Gly Ala Asn Ile Ala Gly
500 505 510
Phe Thr Lys Val Ala Asp Ala Val Lys Ala Gln Gly Ala Val
515 520 525

20 (2) Información de SEC. ID nº 3:

(i) Características de la secuencia:

- (A) Longitud: 2099 pares de bases
- (B) Tipo: ácido nucleico
- (C) Tipo de cadena: doble
- (D) Topología: lineal

(ii) Tipo de molécula: ADNc

(ix) Característica:

- (A) Nombre/código: CDS
- (B) Localización: 33..1568

(xi) Descripción de la secuencia: SEC. ID nº 3:

CTCCTTTCTG CTCGCCCTCT CTCGGTCCCG CC ATG CAG ACC GCC CTC GTC GCC
53 Met Gln Thr Ala Leu Val Ala
1 5
AAG CCT ATC GTG GCC TGC GCG TGG GTC CGC TCC GCC AAG CGC GAT GTC
101 Lys Pro Ile Val Ala Cys Ala Trp Val Arg Ser Ala Lys Arg Asp Val
10 15 20

ES 2 268 711 T3

CGC GCC AAG GCC GTC TCG CTG GAG GAG CAG ATC TCC GCG ATG GAC GCC
 149
 Arg Ala Lys Ala Val Ser Leu Glu Glu Gln Ile Ser Ala Met Asp Ala
 25 30 35

5 ACC ACC GGC GAC TTC ACG GCG CTG CAG AAG GCG GTG AAG CAG ATG GCC
 197
 Thr Thr Gly Asp Phe Thr Ala Leu Gln Lys Ala Val Lys Gln Met Ala
 40 45 50 55

10 ACC AAG GCG GGC ACT GAG GGC CTG GTG CAC GGC ATC AAG AAC CCC GAC
 245
 Thr Lys Ala Gly Thr Glu Gly Leu Val His Gly Ile Lys Asn Pro Asp
 60 65 70

15 GTG CGC CAG CTG CTG ACC GAG ATC TTC ATG AAG GAC CCG GAG CAG CAG
 293
 Val Arg Gln Leu Leu Thr Glu Ile Phe Met Lys Asp Pro Glu Gln Gln
 75 80 85

20 GAG TTC ATG CAG GCG GTG CGC GAG GTG GCC GTC TCC CTG CAG CCC GTG
 341
 Glu Phe Met Gln Ala Val Arg Glu Val Ala Val Ser Leu Gln Pro Val
 90 95 100

25 TTC GAG AAG CGC CCC GAG CTG CTG CCC ATC TTC AAG CAG ATC GTT GAG
 389
 Phe Glu Lys Arg Pro Glu Leu Leu Pro Ile Phe Lys Gln Ile Val Glu
 105 110 115

30 CCT GAG CGC GTG ATC ACC TTC CGC GTG TCC TGG CTG GAC GAC GCC GGC
 437
 Pro Glu Arg Val Ile Thr Phe Arg Val Ser Trp Leu Asp Asp Ala Gly
 120 125 130 135

35 AAC CTG CAG GTC AAC CGC GGC TTC CGC GTG CAG TAC TCG TCC GCC ATC
 485
 Asn Leu Gln Val Asn Arg Gly Phe Arg Val Gln Tyr Ser Ser Ala Ile
 140 145 150

40 GGC CCC TAC AAG GGC GGC CTG CGC TTC CAC CCC TCC GTG AAC CTG TCC
 533
 Gly Pro Tyr Lys Gly Gly Leu Arg Phe His Pro Ser Val Asn Leu Ser
 155 160 165

45 ATC ATG AAG TTC CTT GCC TTT GAG CAG ATC TTC AAG AAC AGC CTG ACC
 581
 Ile Met Lys Phe Leu Ala Phe Glu Gln Ile Phe Lys Asn Ser Leu Thr
 170 175 180

50 ACC CTG CCC ATG GGC GGC GGC AAG GGC GGC TCC GAC TTC GAC CCC AAG
 629
 Thr Leu Pro Met Gly Gly Gly Lys Gly Gly Ser Asp Phe Asp Pro Lys
 185 190 195

55 GGC AAG AGC GAC GCG GAG GTG ATG CGC TTC TGC CAG TCC TTC ATG ACC
 677
 Gly Lys Ser Asp Ala Glu Val Met Arg Phe Cys Gln Ser Phe Met Thr
 200 205 210 215

60 GAG CTG CAG CGC CAC ATC AGC TAC GTG CAG GAC GTG CCC GCC GGC GAC
 725
 Glu Leu Gln Arg His Ile Ser Tyr Val Gln Asp Val Pro Ala Gly Asp
 220 225 230

65 ATC GGC GTG GGC GCG CGC GAG ATT .C TAC CTT TTC GGC CAG TAC AAG
 773

ES 2 268 711 T3

Ile Gly Val Gly Ala Arg Glu Ile Gly Tyr Leu Phe Gly Gln Tyr Lys
 235 240 245
 5 CGC ATC ACC AAG AAC TAC ACC GGC GTG CTG ACC CCG AAG GGC CAG GAG
 821
 Arg Ile Thr Lys Asn Tyr Thr Gly Val Leu Thr Pro Lys Gly Gln Glu
 250 255 260
 10 TAT GGC GGC TCC GAG ATC CGC CCC GAG GCC ACC GGC TAC GGC GCC GTG
 869
 Tyr Gly Gly Ser Glu Ile Arg Pro Glu Ala Thr Gly Tyr Gly Ala Val
 265 270 275
 15 CTG TTT GTG GAG AAC GTG CTG AAG GAC AAG GGC GAG AGC CTC AAG GGC
 917
 Leu Phe Val Glu Asn Val Leu Lys Asp Lys Gly Glu Ser Leu Lys Gly
 280 285 290 295
 20 AAG CGC TGC CTG GTG TCT GGC GCG GGC AAC GTG GCC CAG TAC TGC GCG
 965
 Lys Arg Cys Leu Val Ser Gly Ala Gly Asn Val Ala Gln Tyr Cys Ala
 300 305 310
 25 GAG CTG CTG CTG GAG AAG GGC GCC ATC GTG CTG TCG CTG TCC GAC TCC
 1013
 Glu Leu Leu Leu Glu Lys Gly Ala Ile Val Leu Ser Leu Ser Asp Ser
 315 320 325
 30 CAG GGC TAC GTG TAC GAG CCC AAC GGC TTC ACG CGC GAG CAG CTG CAG
 1061
 Gln Gly Tyr Val Tyr Glu Pro Asn Gly Phe Thr Arg Glu Gln Leu Gln
 330 335 340
 35 GCG GTG CAG GAC ATG AAG AAG AAG AAC AAC AGC GCC CGC ATC TCC GAG
 1109
 Ala Val Gln Asp Met Lys Lys Lys Asn Asn Ser Ala Arg Ile Ser Glu
 345 350 355
 40 TAC AAG AGC GAC ACC GCC GTG TAT GTG GGC GAC CGC CGC AAG CCT TGG
 1157
 Tyr Lys Ser Asp Thr Ala Val Tyr Val Gly Asp Arg Arg Lys Pro Trp
 360 365 370 375
 GAG CTG GAC TGC CAG GTG GAC ATC GCC TTC CCC TGC GCC ACC CAG AAC
 1205
 Glu Leu Asp Cys Gln Val Asp Ile Ala Phe Pro Cys Ala Thr Gln Asn
 380 385 390
 45 GAG ATC GAT GAG CAC GAC GCC GAG CTG CTG ATC AAG CAC GGC TGC CAG
 1253
 Glu Ile Asp Glu His Asp Ala Glu Leu Leu Ile Lys His Gly Cys Gln
 395 400 405
 50 TAC GTG GTG GAG GGC GCC AAC ATG CCC TCC ACC AAC GAG GCC ATC CAC
 1301
 Tyr Val Val Glu Gly Ala Asn Met Pro Ser Thr Asn Glu Ala Ile His
 410 415 420
 55 AAG TAC AAC AAG GCC GGC ATC ATC TAC TGC CCC GGC AAG GCG GCC AAC
 1349
 Lys Tyr Asn Lys Ala Gly Ile Ile Tyr Cys Pro Gly Lys Ala Ala Asn
 425 430 435
 60 GCC GGC GGC GTG GCG GTC AGC GGC CTG GAG ATG ACC CAG AAC CGC ATG
 1397
 Ala Gly Gly Val Ala Val Ser Gly Leu Glu Met Thr Gln Asn Arg Met
 440 445 450 455
 65

ES 2 268 711 T3

AGC CTG AAC TGG ACT CGC GAG GAG GTT CGC GAC AAG CTG GAG CGC ATC
1445
Ser Leu Asn Trp Thr Arg Glu Glu Val Arg Asp Lys Leu Glu Arg Ile
460 465 470

ATG AAG GAC ATC TAC GAC TCC GCC ATG GGG CCG TCC CGC AGA TAC AAT
1493
Met Lys Asp Ile Tyr Asp Ser Ala Met Gly Pro Ser Arg Arg Tyr Asn
475 480 485

GTT GAC CTG GCT GCG GGC GCC AAC ATC GCG GGC TTC ACC AAG GTG GCT
1541
Val Asp Leu Ala Ala Gly Ala Asn Ile Ala Gly Phe Thr Lys Val Ala
490 495 500

GAT GCC GTC AAG GCC CAG GGC GCT GTT TAAGCTGCCC AGGCCCAAGC
1588
Asp Ala Val Lys Ala Gln Gly Ala Val
505 510

CACGGCTCAC CGGCAATCCA ACCCAACCAA CTCACGGCC AGGACCTTTT CGGAAGCGGC
1648

GCCTTTTTC CAGCCAGGGC CCTCACCTGC CCTTTCATAA CCCTGCTATT GCCGCCGTGC
1708

CCCTGCAATT CCACCCCAAG AAGAACTAGC GGCACCTGAC TGCATCAGGA CGGCTATTTT
1768

TTTCGCGACG CGCGCTCACC CCGAGAGCCT CTCTCCCCCG AGCCCTAAGC GCTGACGTCC
1828

GCCCCACTTT GCCTCGCACA TCGCTCGGTT TTGACCCCT CCAGTCTACC CACCCTGTTG
1888

TGAAGCCTAC CAGCTCAATT GCCTTTTAGT GTATGTGCGC CCCCTCCTGC CCCC GAATTT
1948

TCCTGCCATG AGACGTGCGG TTCCTAGCCT GGTGACCCCA AGTAGCAGTT AGTGTGCGTG
2008

CCTTGCCCTG CGCTGCCCGG GATGCGATAC TGTGACCTGA GAGTGCTTGT GTAAACACGA
2068

CGAGTCAAAA AAAAAAAAAA AAAAAAAAAA A
2099

(2) Información de SEC. ID nº 4:

(i) Características de la secuencia:

(A) Longitud: 512 aminoácidos

(B) Tipo: aminoácido

(D) Topología: lineal

(ii) Tipo de molécula: proteína

(xi) Descripción de la secuencia: SEC. ID nº 4:

ES 2 268 711 T3

Met Gln Thr Ala Leu Val Ala Lys Pro Ile Val Ala Cys Ala Trp Val
 1 5 10 15
 Arg Ser Ala Lys Arg Asp Val Arg Ala Lys Ala Val Ser Leu Glu Glu
 20 25 30
 Gln Ile Ser Ala Met Asp Ala Thr Thr Gly Asp Phe Thr Ala Leu Gln
 35 40 45
 Lys Ala Val Lys Gln Met Ala Thr Lys Ala Gly Thr Glu Gly Leu Val
 50 55 60
 His Gly Ile Lys Asn Pro Asp Val Arg Gln Leu Leu Thr Glu Ile Phe
 65 70 75 80
 Met Lys Asp Pro Glu Gln Gln Glu Phe Met Gln Ala Val Arg Glu Val
 85 90 95
 Ala Val Ser Leu Gln Pro Val Phe Glu Lys Arg Pro Glu Leu Leu Pro
 100 105 110
 Ile Phe Lys Gln Ile Val Glu Pro Glu Arg Val Ile Thr Phe Arg Val
 115 120 125
 Ser Trp Leu Asp Asp Ala Gly Asn Leu Gln Val Asn Arg Gly Phe Arg
 130 135 140
 Val Gln Tyr Ser Ser Ala Ile Gly Pro Tyr Lys Gly Gly Leu Arg Phe
 145 150 155 160
 His Pro Ser Val Asn Leu Ser Ile Met Lys Phe Leu Ala Phe Glu Gln
 165 170 175
 Ile Phe Lys Asn Ser Leu Thr Thr Leu Pro Met Gly Gly Gly Lys Gly
 180 185 190
 Gly Ser Asp Phe Asp Pro Lys Gly Lys Ser Asp Ala Glu Val Met Arg
 195 200 205
 Phe Cys Gln Ser Phe Met Thr Glu Leu Gln Arg His Ile Ser Tyr Val
 210 215 220
 Gln Asp Val Pro Ala Gly Asp Ile Gly Val Gly Ala Arg Glu Ile Gly
 225 230 235 240
 Tyr Leu Phe Gly Gln Tyr Lys Arg Ile Thr Lys Asn Tyr Thr Gly Val
 245 250 255
 Leu Thr Pro Lys Gly Gln Glu Tyr Gly Gly Ser Glu Ile Arg Pro Glu
 260 265 270
 Ala Thr Gly Tyr Gly Ala Val Leu Phe Val Glu Asn Val Leu Lys Asp
 275 280 285
 Lys Gly Glu Ser Leu Lys Gly Lys Arg Cys Leu Val Ser Gly Ala Gly
 290 295 300
 Asn Val Ala Gln Tyr Cys Ala Glu Leu Leu Leu Glu Lys Gly Ala Ile
 305 310 315 320
 Val Leu Ser Leu Ser Asp Ser Gln Gly Tyr Val Tyr Glu Pro Asn Gly
 325 330 335
 Phe Thr Arg Glu Gln Leu Gln Ala Val Gln Asp Met Lys Lys Lys Asn
 340 345 350
 Asn Ser Ala Arg Ile Ser Glu Tyr Lys Ser Asp Thr Ala Val Tyr Val
 355 360 365
 Gly Asp Arg Arg Lys Pro Trp Glu Leu Asp Cys Gln Val Asp Ile Ala
 370 375 380
 Phe Pro Cys Ala Thr Gln Asn Glu Ile Asp Glu His Asp Ala Glu Leu
 385 390 395 400

ES 2 268 711 T3

```

Leu Ile Lys His Gly Cys Gln Tyr Val Val Glu Gly Ala Asn Met Pro
                    405                                410                    415

5 Ser Thr Asn Glu Ala Ile His Lys Tyr Asn Lys Ala Gly Ile Ile Tyr
                    420                                425                    430

Cys Pro Gly Lys Ala Ala Asn Ala Gly Gly Val Ala Val Ser Gly Leu
                    435                                440                    445

10 Glu Met Thr Gln Asn Arg Met Ser Leu Asn Trp Thr Arg Glu Glu Val
                    450                                455                    460

Arg Asp Lys Leu Glu Arg Ile Met Lys Asp Ile Tyr Asp Ser Ala Met
465                    470                                475                    480

15 Gly Pro Ser Arg Arg Tyr Asn Val Asp Leu Ala Ala Gly Ala Asn Ile
                    485                                490                    495

Ala Gly Phe Thr Lys Val Ala Asp Ala Val Lys Ala Gln Gly Ala Val
20                    500                                505                    510

```

(2) Información de SEC. ID nº 5:

- (i) Características de la secuencia:
 - (A) Longitud: 20 aminoácidos
 - (B) Tipo: aminoácido
 - (C) Tipo de cadena: única
 - (D) Topología: lineal
- (ii) Tipo de molécula: péptido
- (xi) Descripción de la secuencia: SEC. ID nº 5:


```

35 Ala Val Ser Leu Glu Glu Gln Ile Ser Ala Met Asp Ala Thr Thr Gly
   1          5          10          15
   Asp Phe Thr Ala
   20

```

(2) Información de SEC. ID nº 6:

- (i) Características de la secuencia:
 - (A) Longitud: 10 aminoácidos
 - (B) Tipo: aminoácido
 - (C) Tipo de cadena: única
 - (D) Topología: lineal
- (ii) Tipo de molécula: péptido
- (xi) Descripción de la secuencia: SEC. ID nº 6:


```

55 Asp Ala Thr Thr Gly Asp Phe Thr Ala Leu
   1          5          10

```

(2) Información de SEC. ID nº 7:

- (i) Características de la secuencia:
 - (A) Longitud: 1969 pares de bases
 - (B) Tipo: ácido nucleico
 - (C) Tipo de cadena: doble
 - (D) Topología: lineal
- (ii) Tipo de molécula: ADNc

ES 2 268 711 T3

(xi) Descripción de la secuencia: SEC. ID nº 7:

5 **CAGATCTCCG CGATGGACGC CACCACCGGC GACTTCACGG CGCTGCAGAA GGCGGTGAAG**
60

CAGATGGCCA CCAAGGCGGG CACTGAGGGC CTGGTGCACG GCATCAAGAA CCCCAGCGTG
120

10 **CGCCAGCTGC TGACCGAGAT CTTCATGAAG GACCCGGAGC AGCAGGAGTT CATGCAGGCG**
180

GTGCGCGAGG TGGCCGTCTC CCTGCAGCCC GTGTTCGAGA AGCGCCCGA GCTGCTGCCC
240

15 **ATCTTCAAGC AGATCGTTGA GCCTGAGCGC GTGATCACCT TCCGCGTGTG CTGGCTGGAC**
300

GACGCCGGCA ACCTGCAGGT CAACCGCGGC TTCCGCGTGC AGTACTCGTC CGCCATCGGC
360

20 **CCCTACAAGG GCGGCCTGCG CTCCACCCC TCCGTGAACC TGTCCATCAT GAAGTTCCTT**
420

GCCTTTGAGC AGATCTTCAA GAACAGCCTG ACCACCCTGC CCATGGGCGG CGGCAAGGGC
480

25 **GGCTCCGACT TCGACCCCAA GGGCAAGAGC GACGCGGAGG TGATGCGCTT CTGCCAGTCC**
540

TTCATGACCG AGCTGCAGCG CCACATCAGC TACGTGCAGG ACGTGCCCGC CGGCGACATC
600

30 **GGCGTGGGCG CGCGCGAGAT TGGCTACCTT TTCGGCCAGT ACAAGCGCAT CACCAAGAAC**
660

TACACCGGCG TGCTGACCCC GAAGGGCCAG GAGTATGGCG GCTCCGAGAT CCGCCCCGAG
720

35 **GCCACCGGCT ACGGCGCCGT GCTGTTTGTG GAGAACGTGC TGAAGGACAA GGGCGAGAGC**
780

40 **CTCAAGGGCA AGCGCTGCCT GGTGTCTGGC GCGGGCAACG TGGCCCAGTA CTGCGCGGAG**
840

CTGCTGCTGG AGAAGGGCGC CATCGTGCTG TCGCTGTCCG ACTCCAGGG CTACGTGTAC
900

45 **GAGCCCAACG GCTTCACGCG CGAGCAGCTG CAGGCGGTGC AGGACATGAA GAAGAAGAAC**
960

AACAGCGCCC GCATCTCCGA GTACAAGAGC GACACCGCCG TGTATGTGGG CGACCGCCGC
1020

50

55

60

65

ES 2 268 711 T3

AAGCCTTGGG AGCTGGACTG CCAGGTGGAC ATCGCCTTCC CCTGCGCCAC CCAGAACGAG
1080
 5 **ATCGATGAGC ACGACGCCGA GCTGCTGATC AAGCACGGCT GCCAGTACGT GGTGGAGGGC**
1140
GCCAACATGC CCTCCACCAA CGAGGCCATC CACAAGTACA ACAAGGCCGG CATCATCTAC
1200
 10 **TGCCCCGGCA AGGCGGCCAA CGCCGGCGGC GTGGCGGTCA GCGGCCTGGA GATGACCCAG**
1260
AACCGCATGA GCCTGAACTG GACTCGCGAG GAGGTTCCGG ACAAGCTGGA GCGCATCATG
1320
 15 **AAGGACATCT ACGACTCCGC CATGGGGCCG TCCCGCAGAT ACAATGTTGA CCTGGCTGCG**
1380
GGCGCCAACA TCGCGGGCTT CACCAAGGTG GCTGATGCCG TCAAGGCCCA GGGCGCTGTT
1440
 20 **TAAGCTGCCC AGGCCCAAGC CACGGCTCAC CGGCAATCCA ACCCAACCAA CTC AACGGCC**
1500
AGGACCTTTT CGGAAGCGGC GCCTTTTTTCC CAGCCAGGGC CCTCACCTGC CCTTTCATAA
1560
CCCTGCTATT GCGCGCGTGC CCCTGCAATT CCACCCCAAG AAGAACTAGC GGC ACTTGAC
1620
 30 **TGCATCAGGA CGGCTATTTT TTTCGCGACG CGCGCTCACC CCGAGAGCCT CTCTCCCCCG**
1680
AGCCCTAAGC GCTGACGTCC GCCCGACTTT GCCTCGCACA TCGCTCGGTT TTGACCCCT
1740
 35 **CCAGTCTACC CACCCTGTTG TGAAGCCTAC CAGCTCAATT GCCTTTTAGT GTATGTGCGC**
1800
CCCCTCCTGC CCCCGAATTT TCCTGCCATG AGACGTGCGG TTCCTAGCCT GGTGACCCCA
1860
 40 **AGTAGCAGTT AGTGTCGTG CCTTGCCCTG CGCTGCCCGG GATGCGATAC TGTGACCTGA**
1920
GAGTGCTTGT GTAACACGA CGAGTCAAAA AAAAAAAAAA AAAAAAAAAA
1969
 45

(2) Información de SEC. ID nº 8:

- 50 (i) Características de la secuencia:
- (A) Longitud: 22 pares de bases
 - (B) Tipo: ácido nucleico
 - (C) Tipo de cadena: doble
 - 55 (D) Topología: lineal

(ii) Tipo de molécula: ADNc

(xi) Descripción de la secuencia: SEC. ID nº 8:

60 CTCAAAGGCA AGGAACTTCA TG

22

(2) Información de SEC. ID nº 9:

- 65 (i) Características de la secuencia:
- (A) Longitud: 50 pares de bases
 - (B) Tipo: ácido nucleico

ES 2 268 711 T3

- (C) Tipo de cadena: doble
(D) Topología: lineal
- 5 (ii) Tipo de molécula: ADNc
(xi) Descripción de la secuencia: SEC. ID nº 9:
- GGGTCGACAT TCTAGACAGA ATTCGTGGAT CCTTTTTTTTTT TTTTTTTTTTTT 50
- 10 (2) Información de SEC. ID nº 10:
- (i) Características de la secuencia:
- 15 (A) Longitud: 17 pares de bases
(B) Tipo: ácido nucleico
(C) Tipo de cadena: doble
(D) Topología: lineal
- 20 (ii) Tipo de molécula: ADNc
(xi) Descripción de la secuencia: SEC. ID nº 10:
- GGACGAGTAC TGCACGC 17
- 25 (2) Información de SEC. ID nº 11:
- (i) Características de la secuencia:
- 30 (A) Longitud: 18 pares de bases
(B) Tipo: ácido nucleico
(C) Tipo de cadena: doble
(D) Topología: lineal
- 35 (ii) Tipo de molécula: ADNc
(xi) Descripción de la secuencia: SEC. ID nº 11:
- 40 GATCTCGGTC AGCAGCTG 18
- (2) Información de SEC. ID nº 12:
- (i) Características de la secuencia:
- 45 (A) Longitud: 21 pares de bases
(B) Tipo: ácido nucleico
(C) Tipo de cadena: doble
(D) Topología: lineal
- 50 (ii) Tipo de molécula: ADNc
(xi) Descripción de la secuencia: SEC. ID nº 12:
- 55 GGGTCGACAT TCTAGACAGA A 21
- (2) Información de SEC. ID nº 13:
- 60 (i) Características de la secuencia:
- (A) Longitud: 367 pares de bases
(B) Tipo: ácido nucleico
(C) Tipo de cadena: doble
(D) Topología: lineal
- 65 (ii) Tipo de molécula: ADNc

ES 2 268 711 T3

(xi) Descripción de la secuencia: SEC. ID nº 13:

5 **GGGTCGACAT TCTAGACAGA ATTCTGGAT CCTTTTTTTT TTTTTTTTTT TTTTTCTCC**
60

10 **TTTCTGCTCG CCCTCTCTCC GTCCCGCCAT GCAGACCGCC CTCGTGCGCA AGCCTATCGT**
120

15 **GGCCGCCCCG CTGGCGGCAC GCCCGCGCTG CCTCGCGCCG TGGCCGTGCG CGTGGGTCCG**
180

20 **CTCCGCCAAG CGCGATGTCC GCGCCAAGGC CGTCTCGCTG GAGGAGCAGA TCTCCGCGAT**
240

25 **GGACGCCACC ACCGGCGACT TCACGGCGCT GCAGAAGGCG GTGAAGCAGA TGGCCACCAA**
300

30 **GGCGGGCACT GAGGGCCTGG TGCACGGCAT CAAGAACCCC GACGTGCGCC AGCTGCTGAC**
360

35 **CGAGATC**
367

(2) Información de SEC. ID nº 14:

25 (i) Características de la secuencia:

- (A) Longitud: 325 pares de bases
- (B) Tipo: ácido nucleico
- 30 (C) Tipo de cadena: doble
- (D) Topología: lineal

(ii) Tipo de molécula: ADNc

35 (xi) Descripción de la secuencia: SEC. ID nº 14:

40 **GGGTCGACAT TCTAGACAGA ATTCTGGAT CCTTTTTTTT TTTTTTTTTT TTTTTCTCC**
60

45 **TTTCTGCTCG CCCTCTCTCC GTCCCGCCAT GCAGACCGCC CTCGTGCGCA AGCCTATCGT**
120

50 **GGCCTGCGCG TGGGTGCGCT CCGCCAAGCG CGATGTCCGC GCCAAGGCGG TCTCGCTGGA**
180

55 **GGACGAGATC TCCGCGATGG ACGCCACCAC CGGCGACTTC ACGGCGCTGC AGAAGGCGGT**
240

60 **GAAGCAGATG GCCACCAAGG CGGGCACTGA GGGCCTGGTG CACGGCATCA AGAACCCCGA**
300

65 **CGTGCGCCAG CTGCTGACCG AGATC**
325

(2) Información de SEC. ID nº 15:

(i) Características de la secuencia:

- (A) Longitud: 18 pares de bases
- 60 (B) Tipo: ácido nucleico
- (C) Tipo de cadena: doble
- (D) Topología: lineal

(ii) Tipo de molécula: ADNc

(xi) Descripción de la secuencia: SEC. ID nº 15:

CTTTCTGCTC GCCCTCTC

(2) Información de SEC. ID nº 16:

- 5 (i) Características de la secuencia:
- (A) Longitud: 308 pares de bases
 - (B) Tipo: ácido nucleico
 - (C) Tipo de cadena: doble
 - (D) Topología: lineal
- 10 (ii) Tipo de molécula: ADNc
- 15 (xi) Descripción de la secuencia: SEC. ID nº 16:

CTTTCTGCTC GCCCTCTCTC CGTCCCGCCA TGCAGACCGC CCTCGTCGCC AAGCCTATCG
60

TGGCCGCCCC GCTGGCGGCA CGCCCGCGCT GCCTCGCGCC GTGGCCGTGC GCGTGGGTCC
120

GCTCCGCCAA GCGCGATGTC CGCGCCAAGG CCGTCTCGCT GGAGGAGCAG ATCTCCGCGA
180

TGGACGCCAC CACCGGCGAC TTCACGGCGC TGCAGAAGGC GGTGAAGCAG ATGGCCACCA
240

AGGCGGGCAC TGAGGGCCTG GTGCACGGCA TCAAGAACCC CGACGTGCGC CAGCTGCTGA
300

CCGAGATC
308

(2) Información de SEC. ID nº 17:

- 35 (i) Características de la secuencia:
- (A) Longitud: 266 pares de bases
 - (B) Tipo: ácido nucleico
 - (C) Tipo de cadena: doble
 - (D) Topología: lineal
- 40 (ii) Tipo de molécula: ADNc
- 45 (xi) Descripción de la secuencia: SEC. ID nº 17:

CTTTCTGCTC GCCCTCTCTC CGTCCCGCCA TGCAGACCGC CCTCGTCGCC AAGCCTATCG
60

TGGCCTGCGC GTGGGTCCGC TCCGCCAAGC GCGATGTCCG CGCCAAGGCC GTCTCGCTGG
120

AGGAGCAGAT CTCCGCGATG GACGCCACCA CCGGCGACTT CACGGCGCTG CAGAAGGCGG
180

TGAAGCAGAT GGCCACCAAG GCGGGCACTG AGGGCCTGGT GCACGGCATC AAGAACCCCG
240

ACGTGCGCCA GCTGCTGACC GAGATC
266

(2) Información de SEC. ID nº 18:

- 65 (i) Características de la secuencia:
- (A) Longitud: 2137 pares de bases
 - (B) Tipo: ácido nucleico

ES 2 268 711 T3

(C) Tipo de cadena: doble

(D) Topología: lineal

(ii) Tipo de molécula: ADNc

(xi) Descripción de la secuencia: SEC. ID nº 18:

5

**CTTCTGCTC GCCCTCTCTC CGTCCCGCCA TGCAGACCGC CCTCGTCGCC AAGCCTATCG
60**

10

**TGGCCGCCCC GCTGGCGGCA CGCCCGCGCT GCCTCGCGCC GTGGCCGTGC GCGTGGGTCC
120**

15

**GCTCCGCCAA GCGCGATGTC CGCGCCAAGG CCGTCTCGCT GGAGGAGCAG ATCTCCGCGA
180**

**TGGACGCCAC CACCGGCGAC TTCACGGCGC TGCAGAAGGC GGTGAAGCAG ATGGCCACCA
240**

20

**AGGCGGGCAC TGAGGGCCTG GTGCACGGCA TCAAGAACCC CGACGTGCGC CAGCTGCTGA
300**

**CCGAGATCTT CATGAAGGAC CCGGAGCAGC AGGAGTTCAT GCAGGCGGTG CCGAGGTGG
360**

25

**CCGTCTCCCT GCAGCCCGTG TTCGAGAAGC GCCCCGAGCT GCTGCCCATC TTCAAGCAGA
420**

**TCGTTGAGCC TGAGCGCGTG ATCACCTTCC GCGTGTCTTG GCTGGACGAC GCCGGCAACC
480**

30

**TGCAGGTCAA CCGCGGCTTC CGCGTGCAGT ACTCGTCCGC CATCGGCCCC TACAAGGGCG
540**

**GCCTGCGCTT CCACCCCTCC GTGAACCTGT CCATCATGAA GTTCCTTGCC TTTGAGCAGA
600**

35

**TCTTCAAGAA CAGCCTGACC ACCCTGCCCA TGGGCGGCGG CAAGGGCGGC TCCGACTTCG
660**

**ACCCCAAGGG CAAGAGCGAC GCGGAGGTGA TGCCTTCTG CCAGTCCTTC ATGACCGAGC
720**

40

**TGCAGCGCCA CATCAGCTAC GTGCAGGACG TGCCCGCCGG CGACATCGGC GTGGGCGCGC
780**

**GCGAGATTGG CTACCTTTTC GGCCAGTACA AGGCATCAC CAAGAACTAC ACCGGCGTGC
840**

45

**TGACCCCGAA GGGCCAGGAG TATGGCGGCT CCGAGATCCG CCCCAGGCC ACCGGCTACG
900**

**GCGCCGTGCT GTTGTGGAG AACGTGCTGA AGGACAAGGG CGAGAGCCTC AAGGGCAAGC
960**

50

**GCTGCTGGT GTCTGGCGCG GGCAACGTGG CCCAGTACTG CGCGGAGCTG CTGCTGGAGA
1020**

**AGGGCGCCAT CGTGCTGTGG CTGTCCGACT CCCAGGGCTA CGTGACGAG CCCAACGGCT
1080**

55

**TCACGCGCGA GCAGCTGCAG GCGGTGCAGG ACATGAAGAA GAAGAACAAC AGCGCCCGCA
1140**

**TCTCCGAGTA CAAGAGCGAC ACCGCCGTGT ATGTGGGCGA CCGCCGCAAG CCTTGGGAGC
1200**

60

**TGGACTGCCA GGTGGACATC GCCTTCCCCT GCGCCACCCA GAACGAGATC GATGAGCACG
1260**

65

**ACGCCGAGCT GCTGATCAAG CACGGCTGCC AGTACGTGGT GGAGGGCGCC AACATGCCCT
1320**

ES 2 268 711 T3

5 **CCACCAACGA GGCCATCCAC AAGTACAACA AGGCCGGCAT CATCTACTGC CCCGGCAAGG**
1380

10 **CGGCCAACGC CGGCGGCGTG GCGGTCAGCG GCCTGGAGAT GACCCAGAAC CGCATGAGCC**
1440

15 **TGAACTGGAC TCGCGAGGAG GTTCGCGACA AGCTGGAGCG CATCATGAAG GACATCTACG**
1500

20 **ACTCCGCCAT GGGGCCGTCC CGCAGATACA ATGTTGACCT GGCTGCGGGC GCCAACATCG**
1560

25 **CGGGCTTCAC CAAGGTGGCT GATGCCGTCA AGGCCCAGGG CGCTGTTTAA GCTGCCCAGG**
1620

30 **CCCAAGCCAC GGCTCACCGG CAATCCAACC CAACCAACTC AACGGCCAGG ACCTTTTCGG**
1680

35 **AAGCGGCGCC TTTTCCCAG CCAGGGCCCT CACCTGCCCT TTCATAACCC TGCTATTGCC**
1740

40 **GCCGTGCCCC TGCAATTCCA CCCAAGAAG AACTAGCGGC ACTTGACTGC ATCAGGACGG**
1800

45 **CTATTTTTTT CGCGACGCGC GCTCACCCCG AGAGCCTCTC TCCCCGAGC CCTAAGCGCT**
1860

50 **GACGTCCGCC CGACTTTGCC TCGCACATCG CTCGGTTTTG ACCCCCTCCA GTCTACCCAC**
1920

55 **CCTGTTGTGA AGCCTACCAG CTCAATTGCC TTTTAGTGTA TGTGCGCCCC CTCCTGCCCC**
1980

60 **CGAATTTTCC TGCCATGAGA CGTGCGGTTT CTAGCCTGGT GACCCCAAGT AGCAGTTAGT**
2040

65 **GTGCGTGCCT TGCCCTGCGC TGCCCGGGAT GCGATACTGT GACCTGAGAG TGCTTGTGTA**
2100

AACACGACGA GTCAAAAAAAA AAAAAAAAAA AAAAAA
2137

(2) Información de SEC. ID nº 19:

(i) Características de la secuencia:

- (A) Longitud: 2096 pares de bases
- (B) Tipo: ácido nucleico
- (C) Tipo de cadena: doble
- (D) Topología: lineal

(ii) Tipo de molécula: ADNc

(xi) Descripción de la secuencia: SEC. ID nº 19:

ES 2 268 711 T3

CTTTCTGCTC GCCCTCTCTC CGTCCC GCCA TGCAGACCGC CCTCGTCGCC AAGCCTATCG
 60
 TGGCCTGCGC GTGGGTCCGC TCCGCCAAGC GCGATGTCCG CGCCAAGGCC GTCTCGCTGG
 5 120
 AGGAGCAGAT CTCCGCGATG GACGCCACCA CCGGCGACTT CACGGCGCTG CAGAAGGGCG
 180
 TGAAGCAGAT GGCCACCAAG GCGGGCACTG AGGGCCTGGT GCACGGGCATC AAGAACCCCG
 10 240
 ACGTGCGCCA GCTGCTGACC GAGATCTTCA TGAAGGACCC GGAGCAGCAG GAGTTCATGC
 300
 AGGCGGTGCG CGAGGTGGCC GTCTCCCTGC AGCCCGTGTT CGAGAAGCGC CCCGAGCTGC
 15 360
 TGCCCATCTT CAAGCAGATC GTTGAACCTG AGCGCGTGAT CACCTTCCGC GTGTCTGGC
 420
 TGGACGACGC CGGCAACCTG CAGGTCAACC GCGGCTCCG CGTGCAGTAC TCGTCCGCCA
 20 480
 TCGGCCCTTA CAAGGGCGGC CTGCGCTTCC ACCCCTCCGT GAACCTGTCC ATCATGAAGT
 540
 TCCTTGCTT TGAGCAGATC TTCAAGAACA GCCTGACCAC CCTGCCCATG GCGGGCGGCA
 25 600
 AGGGCGGCTC CGACTTCGAC CCCAAGGGCA AGAGCGACGC GGAGGTGATG CGCTTCTGCC
 660
 AGTCCTTCAT GACCGAGCTG CAGCGCCACA TCAGCTACGT GCAGGACGTG CCCGCCGGCG
 30 720
 ACATCGGCGT GGGCGCGCGC GAGATTGGCT ACCTTTTCGG CCAGTACAAG CGCATCACCA
 780
 AGAACTACAC CGGCGTGCTG ACCCCGAAGG GCCAGGAGTA TGGCGGCTCC GAGATCCGCC
 35 840
 CCGAGGCCAC CGGCTACGGC GCCGTGCTGT TTGTGGAGAA CGTGCTGAAG GACAAGGGCG
 900
 AGAGCCTCAA GGGCAAGCGC TGCCTGGTGT CTGGCGGGG CAACGTGGCC CAGTACTGGC
 40 960
 CGGAGCTGCT GCTGGAGAAG GCGCCATCG TGCTGTCCCT GTCCGACTCC CAGGGCTACG
 1020
 TGTACGAGCC CAACGGCTT ACGCGCGAGC AGCTGCAGGC GGTGCAGGAC ATGAAGAAGA
 45 1080
 AGAACAAACG CGCCCGCATC TCCGAGTACA AGAGCGACAC CGCCGTGTAT GTGGGCGACC
 1140
 GCCGCAAGCC TTGGGAGCTG GACTGCCAGG TGGACATCGC CTTCCCCTGC GCCACCCAGA
 1200
 ACQAGATCGA TGAGCACGAC GCCGAGCTGC TGATCAAGCA CGGCTGCCAG TACGTGGTGG
 50 1260
 AGGGCGCCAA CATGCCCTCC ACCAACGAGG CCATCCACAA GTACAACAAG GCCCGCATCA
 1320
 TCTACTGCCC CGGCAAGGGC GCCAACGCCG GCGGCGTGGC GGTCAAGCGC CTGGAGATGA
 55 1380
 CCCAGAACCG CATGAGCCTG AACTGGAATC GCGAGGAGGT TCGCGACAAG CTGGAGCGCA
 1440
 TCATGAAGGA CATCTACGAC TCCGCCATGG GGCCGTCCCG CAGATACAAT GTTGACCTGG
 60 1500
 CTGCGGGCGC CAACATCGCG GGCTTCACCA AGGTGGCTGA TGCCGTCAAG GCCCAGGGCG
 1560

65

ES 2 268 711 T3

CTGTTTAAGC TGCCCAGGCC CAAGCCACGG CTCACCGGCA ATCCAACCCA ACCAACTCAA
1620

5 CGGCCAGGAC CTTTTCGGAA GCGGCGCCTT TTTCCAGCC AGGGCCCTCA CCTGCCCTT
1680

CATAACCCTG CTATTGCCGC CGTGCCCTG CAATCCACC CCAAGAAGAA CTAGCGGCAC
1740

10 TTGACTGCAT CAGGACGGCT ATTTTTTTCG CGACGCGCGC TCACCCCGAG AGCCTCTCTC
1800

CCCCGAGCCC TAAGCGCTGA CGTCCGCCCG ACTTTGCCTC GCACATCGCT CGGTTTTGAC
1860

15 CCCCTCCAGT CTACCCACCC TGTGTGAAG CCTACCAGCT CAATTGCCTT TTAGTGTATG
1920

TGCGCCCCCT CCTGCCCCCG AATTTTCCTG CCATGAGACG TCGGGTTCCT AGCCTGGTGA
1980

20 CCCCAAGTAG CAGTTAGTGT GCGTGCCTTG CCCTGCGCTG CCCGGGATGC GATACTGTGA
2040

25 CCTGAGAGTG CTTGTGTAAG CACGACGAGT CAAAAAAAAA AAAAAAAAAA AAAAAA
2096

(2) Información de SEC. ID nº 20:

- 30 (i) Características de la secuencia:
- (A) Longitud: 25 pares de bases
 - (B) Tipo: ácido nucleico
 - (C) Tipo de cadena: doble
 - 35 (D) Topología: lineal
- (ii) Tipo de molécula: ADNc
- (xi) Descripción de la secuencia: SEC. ID nº 20:

CATATGGCCG TCTCGCTGGG AGGAG

25

(2) Información de SEC. ID nº 21:

- 45 (i) Características de la secuencia:
- (A) Longitud: 19 pares de bases
 - (B) Tipo: ácido nucleico
 - 50 (C) Tipo de cadena: doble
 - (D) Topología: lineal
- (ii) Tipo de molécula: ADNc
- (xi) Descripción de la secuencia: SEC. ID nº 21:

GTTGGATTGC CGGTGAGCC

19

(2) Información de SEC. ID nº 22:

- 60 (i) Características de la secuencia:
- (A) Longitud: 21 pares de bases
 - (B) Tipo: ácido nucleico
 - 65 (C) Tipo de cadena: doble
 - (D) Topología: lineal

ES 2 268 711 T3

(ii) Tipo de molécula: ADNc

(xi) Descripción de la secuencia: SEC. ID nº 22:

5 CATATGGACG CCACCACCGG C

21

(2) Información de SEC. ID nº 23:

(i) Características de la secuencia:

- 10 (A) Longitud: 1506 pares de bases
- (B) Tipo: ácido nucleico
- (C) Tipo de cadena: doble
- 15 (D) Topología: lineal

(ii) Tipo de molécula: ADNc

(ix) Característica:

- 20 (A) Nombre/código: CDS
- (B) Situación: 4..1464

(xi) Descripción de la secuencia: SEC. ID nº 23:

25	CAT 48	ATG	GCC	GTC	TCG	CTG	GAG	GAG	CAG	ATC	TCC	GCG	ATG	GAC	GCC	ACC
	Met	Ala	Val	Ser	Leu	Glu	Glu	Gln	Ile	Ser	Ala	Met	Asp	Ala	Thr	
			515					520					525			
30	ACC 96	GGC	GAC	TTC	ACG	GCG	CTG	CAG	AAG	GCG	GTG	AAG	CAG	ATG	GCC	ACC
	Thr	Gly	Asp	Phe	Thr	Ala	Leu	Gln	Lys	Ala	Val	Lys	Gln	Met	Ala	Thr
			530					535					540			
35	AAG 144	GCG	GGC	ACT	GAG	GGC	CTG	GTG	CAC	GGC	ATC	AAG	AAC	CCC	GAC	GTG
	Lys	Ala	Gly	Thr	Glu	Gly	Leu	Val	His	Gly	Ile	Lys	Asn	Pro	Asp	Val
			545			550						555				
40	CGC 192	CAG	CTG	CTG	ACC	GAG	ATC	TTC	ATG	AAG	GAC	CCG	GAG	CAG	CAG	GAG
	Arg	Gln	Leu	Leu	Thr	Glu	Ile	Phe	Met	Lys	Asp	Pro	Glu	Gln	Gln	Glu
	560					565					570				575	

45

50

55

60

65

ES 2 268 711 T3

5
 TTC ATG CAG GCG GTG CGC GAG GTG GCC GTC TCC CTG CAG CCC GTG TTC
 240
 Phe Met Gln Ala Val Arg Glu Val Ala Val Ser Leu Gln Pro Val Phe
 580 585 590

10
 GAG AAG CGC CCC GAG CTG CTG CCC ATC TTC AAG CAG ATC GTT GAG CCT
 288
 Glu Lys Arg Pro Glu Leu Leu Pro Ile Phe Lys Gln Ile Val Glu Pro
 595 600 605

15
 GAG CGC GTG ATC ACC TTC CGC GTG TCC TGG CTG GAC GAC GCC GGC AAC
 336
 Glu Arg Val Ile Thr Phe Arg Val Ser Trp Leu Asp Asp Ala Gly Asn
 610 615 620

20
 CTG CAG GTC AAC CGC GGC TTC CGC GTG CAG TAC TCG TCC GCC ATC GGC
 384
 Leu Gln Val Asn Arg Gly Phe Arg Val Gln Tyr Ser Ser Ala Ile Gly
 625 630 635

25
 CCC TAC AAG GGC GGC CTG CGC TTC CAC CCC TCC GTG AAC CTG TCC ATC
 432
 Pro Tyr Lys Gly Gly Leu Arg Phe His Pro Ser Val Asn Leu Ser Ile
 640 645 650 655

30
 ATG AAG TTC CTT GCC TTT GAG CAG ATC TTC AAG AAC AGC CTG ACC ACC
 480
 Met Lys Phe Leu Ala Phe Glu Gln Ile Phe Lys Asn Ser Leu Thr Thr
 660 665 670

35
 CTG CCC ATG GGC GGC GGC AAG GGC GGC TCC GAC TTC GAC CCC AAG GGC
 528
 Leu Pro Met Gly Gly Gly Lys Gly Gly Ser Asp Phe Asp Pro Lys Gly
 675 680

40
 AAG AGC GAC GCG GAG GTG ATG CGC TTC TGC CAG TCC TTC ATG ACC GAG
 576
 Lys Ser Asp Ala Glu Val Met Arg Phe Cys Gln Ser Phe Met Thr Glu
 690 695 700

45
 CTG CAG CGC CAC ATC AGC TAC GTG CAG GAC GTG CCC GCC GGC GAC ATC
 624
 Leu Gln Arg His Ile Ser Tyr Val Gln Asp Val Pro Ala Gly Asp Ile
 705 710 715

50
 GGC GTG GGC GCG CGC GAG ATT GGC TAC CTT TTC GGC CAG TAC AAG CGC
 672
 Gly Val Gly Ala Arg Glu Ile Gly Tyr Leu Phe Gly Gln Tyr Lys Arg
 720 725 730 735

55
 ATC ACC AAG AAC TAC ACC GGC GTG CTG ACC CCG AAG GGC CAG GAG TAT
 720
 Ile Thr Lys Asn Tyr Thr Gly Val Leu Thr Pro Lys Gly Gln Glu Tyr
 740 745 750

60
 GGC GGC TCC GAG ATC CGC CCC GAG GCC ACC GGC TAC GGC GCC GTG CTG
 768
 Gly Gly Ser Glu Ile Arg Pro Glu Ala Thr Gly Tyr Gly Ala Val Leu
 755 760 765

65
 TTT GTG GAG AAC GTG CTG AAG GAC AAG GGC GAG AGC CTC AAG GGC AAG
 816
 Phe Val Glu Asn Val Leu Lys Asp Lys Gly Glu Ser Leu Lys Gly Lys
 770 775 780

CGC TGC CTG GTG TCT GGC GCG GGC AAC GTG GCC CAG TAC TGC GCG GAG
 864
 Arg Cys Leu Val Ser Gly Ala Gly Asn Val Ala Gln Tyr Cys Ala Glu
 785 790 795

CTG CTG CTG GAG AAG GGC GCC ATC GTG CTG TCG CTG TCC GAC TCC CAG
 912
 Leu Leu Leu Glu Lys Gly Ala Ile Val Leu Ser Leu Ser Asp Ser Gln
 800 805 810 815

ES 2 268 711 T3

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

GGC TAC GTG TAC GAG CCC AAC GGC TTC ACG CGC GAG CAG CTG CAG GCG
 960
 Gly Tyr Val Tyr Glu Pro Asn Gly Phe Thr Arg Glu Gln Leu Gln Ala
 820 825 830

GTG CAG GAC ATG AAG AAG AAG AAC AAC AGC GCC CGC ATC TCC GAG TAC
 1008
 Val Gln Asp Met Lys Lys Lys Asn Asn Ser Ala Arg Ile Ser Glu Tyr
 835 840 845

AAG AGC GAC ACC GCC GTG TAT GTG GGC GAC CGC CGC AAG CCT TGG GAG
 1056

Lys Ser Asp Thr Ala Val Tyr Val Gly Asp Arg Arg Lys Pro Trp Glu
 850 855 860

CTG GAC TGC CAG GTG GAC ATC GCC TTC CCC TGC GCC ACC CAG AAC GAG
 1104
 Leu Asp Cys Gln Val Asp Ile Ala Phe Pro Cys Ala Thr Gln Asn Glu
 865 870 875

ATC GAT GAG CAC GAC GCC GAG CTG CTG ATC AAG CAC GGC TGC CAG TAC
 1152
 Ile Asp Glu His Asp Ala Glu Leu Leu Ile Lys His Gly Cys Gln Tyr
 880 885 890 895

GTG GTG GAG GGC GCC AAC ATG CCC TCC ACC AAC GAG GCC ATC CAC AAG
 1200
 Val Val Glu Gly Ala Asn Met Pro Ser Thr Asn Glu Ala Ile His Lys
 900 905 910

TAC AAC AAG GCC GGC ATC ATC TAC TGC CCC GGC AAG GCG GCC AAC GCC
 1248
 Tyr Asn Lys Ala Gly Ile Ile Tyr Cys Pro Gly Lys Ala Ala Asn Ala
 915 920 925

GGC GGC GTG GCG GTC AGC GGC CTG GAG ATG ACC CAG AAC CGC ATG AGC
 1296
 Gly Gly Val Ala Val Ser Gly Leu Glu Met Thr Gln Asn Arg Met Ser
 930 935 940

CTG AAC TGG ACT CGC GAG GAG GTT CGC GAC AAG CTG GAG CGC ATC ATG
 1344
 Leu Asn Trp Thr Arg Glu Glu Val Arg Asp Lys Leu Glu Arg Ile Met
 945 950 955

AAG GAC ATC TAC GAC TCC GCC ATG GGG CCG TCC CGC AGA TAC AAT GTT
 1392
 Lys Asp Ile Tyr Asp Ser Ala Met Gly Pro Ser Arg Arg Tyr Asn Val
 960 965 970 975

GAC CTG GCT GCG GGC GCC AAC ATC GCG GGC TTC ACC AAG GTG GCT GAT
 1440
 Asp Leu Ala Ala Gly Ala Asn Ile Ala Gly Phe Thr Lys Val Ala Asp
 980 985 990

GCC GTC AAG GCC CAG GGC GCT GTT TAAGCTGCCC AGGCCCAAGC CACGGCTCAC
 1494
 Ala Val Lys Ala Gln Gly Ala Val
 995

CGGCAATCCA AC
 1506

(2) Información de SEC. ID nº 24:

(i) Características de la secuencia:

(A) Longitud: 487 aminoácido

(B) Tipo: aminoácido

(D) Topología: lineal

ES 2 268 711 T3

(ii) Tipo de molécula: proteína

(xi) Descripción de la secuencia: SEC. ID nº 24:

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

ES 2 268 711 T3

5 **Leu Leu Glu Lys Gly Ala Ile Val Leu Ser Leu Ser Asp Ser Gln Gly**
 290 **295** **300**

10 **Tyr Val Tyr Glu Pro Asn Gly Phe Thr Arg Glu Gln Leu Gln Ala Val**
 305 **310** **315** **320**

15 **Gln Asp Met Lys Lys Lys Asn Asn Ser Ala Arg Ile Ser Glu Tyr Lys**
 325 **330** **335**

20 **Ser Asp Thr Ala Val Tyr Val Gly Asp Arg Arg Lys Pro Trp Glu Leu**
 340 **345** **350**

25 **Asp Cys Gln Val Asp Ile Ala Phe Pro Cys Ala Thr Gln Asn Glu Ile**
 355 **360** **365**

30 **Asp Glu His Asp Ala Glu Leu Leu Ile Lys His Gly Cys Gln Tyr Val**
 370 **375** **380**

35 **Val Glu Gly Ala Asn Met Pro Ser Thr Asn Glu Ala Ile His Lys Tyr**
 385 **390** **395** **400**

40 **Asn Lys Ala Gly Ile Ile Tyr Cys Pro Gly Lys Ala Ala Asn Ala Gly**
 405 **410** **415**

45 **Gly Val Ala Val Ser Gly Leu Glu Met Thr Gln Asn Arg Met Ser Leu**
 420 **425** **430**

50 **Asn Trp Thr Arg Glu Glu Val Arg Asp Lys Leu Glu Arg Ile Met Lys**
 435 **440** **445**

55 **Asp Ile Tyr Asp Ser Ala Met Gly Pro Ser Arg Arg Tyr Asn Val Asp**
 450 **455** **460**

60 **Leu Ala Ala Gly Ala Asn Ile Ala Gly Phe Thr Lys Val Ala Asp Ala**
 465 **470** **475** **480**

65 **Val Lys Ala Gln Gly Ala Val**
 485

(2) Información de SEC. ID nº 25:

- 40 (i) Características de la secuencia:
- (A) Longitud: 1473 pares de bases
 - (B) Tipo: ácido nucleico
 - 45 (C) Tipo de cadena: doble
 - (D) Topología: lineal
- (ii) Tipo de molécula: ADNc
- 50 (ix) Característica:
- (A) Nombre/código: CDS
 - (B) Localización: 4..1431
- 55 (xi) Descripción de la secuencia: SEC. ID nº 25:

60

65

ES 2 268 711 T3

CAT ATG GAC GCC ACC ACC GGC GAC TTC ACG GCG CTG CAG AAG GCG GTG
 48 Met Asp Ala Thr Thr Gly Asp Phe Thr Ala Leu Gln Lys Ala Val
 490 495 500

AAG CAG ATG GCC ACC AAG GCG GGC ACT GAG GGC CTG GTG CAC GGC ATC
 96 Lys Gln Met Ala Thr Lys Ala Gly Thr Glu Gly Leu Val His Gly Ile
 505 510 515

AAG AAC CCC GAC GTG CGC CAG CTG CTG ACC GAG ATC TTC ATG AAG GAC
 144 Lys Asn Pro Asp Val Arg Gln Leu Leu Thr Glu Ile Phe Met Lys Asp
 520 525 530

CCG GAG CAG CAG GAG TTC ATG CAG GCG GTG CGC GAG GTG GCC GTC TCC
 192 Pro Glu Gln Gln Glu Phe Met Gln Ala Val Arg Glu Val Ala Val Ser
 535 540 545 550

CTG CAG CCC GTG TTC GAG AAG CGC CCC GAG CTG CTG CCC ATC TTC AAG
 240 Leu Gln Pro Val Phe Glu Lys Arg Pro Glu Leu Leu Pro Ile Phe Lys
 555 560 565

CAG ATC GTT GAG CCT GAG CGC GTG ATC ACC TTC CGC GTG TCC TGG CTG
 288 Gln Ile Val Glu Pro Glu Arg Val Ile Thr Phe Arg Val Ser Trp Leu
 570 575 580

GAC GAC GCC GGC AAC CTG CAG GTC AAC CGC GGC TTC CGC GTG CAG TAC
 336 Asp Asp Ala Gly Asn Leu Gln Val Asn Arg Gly Phe Arg Val Gln Tyr
 585 590 595

TCG TCC GCC ATC GGC CCC TAC AAG GGC GGC CTG CGC TTC CAC CCC TCC
 384 Ser Ser Ala Ile Gly Pro Tyr Lys Gly Gly Leu Arg Phe His Pro Ser
 600 605 610

GTG AAC CTG TCC ATC ATG AAG TTC CTT GCC TTT GAG CAG ATC TTC AAG
 432 Val Asn Leu Ser Ile Met Lys Phe Leu Ala Phe Glu Gln Ile Phe Lys
 615 620 625 630

AAC AGC CTG ACC ACC CTG CCC ATG GGC GGC GGC AAG GGC GGC TCC GAC
 480 Asn Ser Leu Thr Thr Leu Pro Met Gly Gly Gly Lys Gly Gly Ser Asp
 635 640 645

TTC GAC CCC AAG GGC AAG AGC GAC GCG GAG GTG ATG CGC TTC TGC CAG
 528 Phe Asp Pro Lys Gly Lys Ser Asp Ala Glu Val Met Arg Phe Cys Gln
 650 655 660

TCC TTC ATG ACC GAG CTG CAG CGC CAC ATC AGC TAC GTG CAG GAC GTG
 576 Ser Phe Met Thr Glu Leu Gln Arg His Ile Ser Tyr Val Gln Asp Val
 665 670 675

CCC GCC GGC GAC ATC GGC GTG GGC GCG CGC GAG ATT GGC TAC CTT TTC
 624 Pro Ala Gly Asp Ile Gly Val Gly Ala Arg Glu Ile Gly Tyr Leu Phe
 680 685 690

ES 2 268 711 T3

5 GGC CAG TAC AAG CGC ATC ACC AAG AAC TAC ACC GGC GTG CTG ACC CCG
 672
 Gly Gln Tyr Lys Arg Ile Thr Lys Asn Tyr Thr Gly Val Leu Thr Pro
 695 700 705 710

10 AAG GGC CAG GAG TAT GGC GGC TCC GAG ATC CGC CCC GAG GCC ACC GGC
 720
 Lys Gly Gln Glu Tyr Gly Gly Ser Glu Ile Arg Pro Glu Ala Thr Gly
 715 720 725

15 TAC GGC GCC GTG CTG TTT GTG GAG AAC GTG CTG AAG GAC AAG GGC GAG
 768
 Tyr Gly Ala Val Leu Phe Val Glu Asn Val Leu Lys Asp Lys Gly Glu
 730 735 740

20 AGC CTC AAG GGC AAG CGC TGC CTG GTG TCT GGC GCG GGC AAC GTG GCC
 816
 Ser Leu Lys Gly Lys Arg Cys Leu Val Ser Gly Ala Gly Asn Val Ala
 745 750 755

25 CAG TAC TGC GCG GAG CTG CTG CTG GAG AAG GGC GCC ATC GTG CTG TCG
 864
 Gln Tyr Cys Ala Glu Leu Leu Leu Glu Lys Gly Ala Ile Val Leu Ser
 760 765 770

30 CTG TCC GAC TCC CAG GGC TAC GTG TAC GAG CCC AAC GGC TTC ACG CGC
 912
 Leu Ser Asp Ser Gln Gly Tyr Val Tyr Glu Pro Asn Gly Phe Thr Arg
 775 780 785 790

35 GAG CAG CTG CAG GCG GTG CAG GAC ATG AAG AAG AAG AAC AAC AGC GCC
 960
 Glu Gln Leu Gln Ala Val Gln Asp Met Lys Lys Lys Asn Asn Ser Ala
 795 800 805

40 CGC ATC TCC GAG TAC AAG AGC GAC ACC GCC GTG TAT GTG GGC GAC CGC
 1008
 Arg Ile Ser Glu Tyr Lys Ser Asp Thr Ala Val Tyr Val Gly Asp Arg
 810 815 820

45 CGC AAG CCT TGG GAG CTG GAC TGC CAG GTG GAC ATC GCC TTC CCC TGC
 1056
 Arg Lys Pro Trp Glu Leu Asp Cys Gln Val Asp Ile Ala Phe Pro Cys
 825 830 835

50 GCC ACC CAG AAC GAG ATC GAT GAG CAC GAC GCC GAG CTG CTG ATC AAG
 1104
 Ala Thr Gln Asn Glu Ile Asp Glu His Asp Ala Glu Leu Leu Ile Lys
 840 845 850

55 CAC GGC TGC CAG TAC GTG GTG GAG GGC GCC AAC ATG CCC TCC ACC AAC
 1152
 His Gly Cys Gln Tyr Val Val Glu Gly Ala Asn Met Pro Ser Thr Asn
 855 860 865 870

60 GAG GCC ATC CAC AAG TAC AAC AAG GCC GGC ATC ATC TAC TGC CCC GGC
 1200
 Glu Ala Ile His Lys Tyr Asn Lys Ala Gly Ile Ile Tyr Cys Pro Gly
 875 880 885

65 AAG GCG GCC AAC GCC GGC GGC GTG GCG GTC AGC GGC CTG GAG ATG ACC
 1248
 Lys Ala Ala Asn Ala Gly Gly Val Ala Val Ser Gly Leu Glu Met Thr
 890 895 900

ES 2 268 711 T3

CAG AAC CGC ATG AGC CTG AAC TGG ACT CGC GAG GAG GTT CGC GAC AAG
1296
Gln Asn Arg Met Ser Leu Asn Trp Thr Arg Glu Glu Val Arg Asp Lys
905 910 915

CTG GAG CGC ATC ATG AAG GAC ATC TAC GAC TCC GCC ATG GGG CCG TCC
1344
Leu Glu Arg Ile Met Lys Asp Ile Tyr Asp Ser Ala Met Gly Pro Ser
920 925 930

CGC AGA TAC AAT GTT GAC CTG GCT GCG GGC GCC AAC ATC GCG GGC TTC
1392
Arg Arg Tyr Asn Val Asp Leu Ala Ala Gly Ala Asn Ile Ala Gly Phe
935 940 945 950

ACC AAG GTG GCT GAT GCC GTC AAG GCC CAG GGC GCT GTT TAAGCTGCCC
1441
Thr Lys Val Ala Asp Ala Val Lys Ala Gln Gly Ala Val
955 960

AGGCCCAAGC CACGGCTCAC CGGCAATCCA AC
1473

(2) Información de SEC. ID nº 26:

(i) Características de la secuencia:

(A) Longitud: 476 aminoácidos

(B) Tipo: aminoácido

(D) Topología: lineal

(ii) Tipo de molécula: proteína

(xi) Descripción de la secuencia: SEC. ID nº 26:

Met Asp Ala Thr Thr Gly Asp Phe Thr Ala Leu Gln Lys Ala Val Lys
1 5 10 15

Gln Met Ala Thr Lys Ala Gly Thr Glu Gly Leu Val His Gly Ile Lys
20 25 30

Asn Pro Asp Val Arg Gln Leu Leu Thr Glu Ile Phe Met Lys Asp Pro
35 40 45

Glu Gln Gln Glu Phe Met Gln Ala Val Arg Glu Val Ala Val Ser Leu
50 55 60

Gln Pro Val Phe Glu Lys Arg Pro Glu Leu Leu Pro Ile Phe Lys Gln
65 70 75 80

Ile Val Glu Pro Glu Arg Val Ile Thr Phe Arg Val Ser Trp Leu Asp
85 90 95

Asp Ala Gly Asn Leu Gln Val Asn Arg Gly Phe Arg Val Gln Tyr Ser
100 105 110

Ser Ala Ile Gly Pro Tyr Lys Gly Gly Leu Arg Phe His Pro Ser Val
115 120 125

Asn Leu Ser Ile Met Lys Phe Leu Ala Phe Glu Gln Ile Phe Lys Asn
130 135 140

ES 2 268 711 T3

Ser Leu Thr Thr Leu Pro Met Gly Gly Gly Lys Gly Gly Ser Asp Phe
 145 150 155 160
 5 Asp Pro Lys Gly Lys Ser Asp Ala Glu Val Met Arg Phe Cys Gln Ser
 165 170 175
 Phe Met Thr Glu Leu Gln Arg His Ile Ser Tyr Val Gln Asp Val Pro
 180 185 190
 10 Ala Gly Asp Ile Gly Val Gly Ala Arg Glu Ile Gly Tyr Leu Phe Gly
 195 200 205
 Gln Tyr Lys Arg Ile Thr Lys Asn Tyr Thr Gly Val Leu Thr Pro Lys
 210 215 220
 15 Gly Gln Glu Tyr Gly Gly Ser Glu Ile Arg Pro Glu Ala Thr Gly Tyr
 225 230 235 240
 Gly Ala Val Leu Phe Val Glu Asn Val Leu Lys Asp Lys Gly Glu Ser
 245 250 255
 20 Leu Lys Gly Lys Arg Cys Leu Val Ser Gly Ala Gly Asn Val Ala Gln
 260 265 270
 Tyr Cys Ala Glu Leu Leu Leu Glu Lys Gly Ala Ile Val Leu Ser Leu
 275 280 285
 25 Ser Asp Ser Gln Gly Tyr Val Tyr Glu Pro Asn Gly Phe Thr Arg Glu
 290 295 300
 30 Gln Leu Gln Ala Val Gln Asp Met Lys Lys Lys Asn Asn Ser Ala Arg
 305 310 315 320
 Ile Ser Glu Tyr Lys Ser Asp Thr Ala Val Tyr Val Gly Asp Arg Arg
 325 330 335
 35 Lys Pro Trp Glu Leu Asp Cys Gln Val Asp Ile Ala Phe Pro Cys Ala
 340 345 350
 Thr Gln Asn Glu Ile Asp Glu His Asp Ala Glu Leu Leu Ile Lys His
 355 360 365
 40 Gly Cys Gln Tyr Val Val Glu Gly Ala Asn Met Pro Ser Thr Asn Glu
 370 375 380
 Ala Ile His Lys Tyr Asn Lys Ala Gly Ile Ile Tyr Cys Pro Gly Lys
 385 390 395 400
 45 Ala Ala Asn Ala Gly Gly Val Ala Val Ser Gly Leu Glu Met Thr Gln
 405 410 415
 Asn Arg Met Ser Leu Asn Trp Thr Arg Glu Glu Val Arg Asp Lys Leu
 420 425 430
 50 Glu Arg Ile Met Lys Asp Ile Tyr Asp Ser Ala Met Gly Pro Ser Arg
 435 440 445
 55 Arg Ty r Asn Val Asp Leu Ala Ala Gly Ala Asn Ile Ala Gly Phe Thr
 450 455 460
 Lys Val Ala Asp Ala Val Lys Ala Gln Gly Ala Val
 465 470 475
 60
 65