

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 647 828**

51 Int. Cl.:

**C12N 9/88** (2006.01)

**C12P 5/00** (2006.01)

**C12N 15/60** (2006.01)

**C12N 5/10** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **11.03.2014 PCT/US2014/023759**

87 Fecha y número de publicación internacional: **25.09.2014 WO14150599**

96 Fecha de presentación y número de la solicitud europea: **11.03.2014 E 14718824 (7)**

97 Fecha y número de publicación de la concesión europea: **16.08.2017 EP 2970934**

54 Título: **Polipéptidos de valenceno sintasa, moléculas de ácido nucleico que los codifican y usos de los mismos**

30 Prioridad:

**14.03.2013 US 201361852462 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**26.12.2017**

73 Titular/es:

**EVOLVA, INC. (100.0%)  
2277 Thunderstick Drive, Suite 300  
Lexington, KY 60505, US**

72 Inventor/es:

**SARAN, DAYAL y  
PARK, GRACE, EUNYOUNG**

74 Agente/Representante:

**PONS ARIÑO, Ángel**

ES 2 647 828 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 de valenceno sintasa, moléculas de ácido nucleico que los codifican y usos de los mismos

## 5 CAMPO DE LA INVENCION

Se proporcionan polipéptidos de valenceno sintasa, moléculas de ácido nucleico que codifican valenceno sintasas, células hospedadoras que contienen los ácidos nucleicos y procedimientos para producir productos cuya producción está catalizada por los polipéptidos. Se proporcionan también procedimientos para producir valenceno y nootkatona.

10

## ANTECEDENTES

Valenceno y nootkatona son sesquiterpenos que aparecen en aceites esenciales tales como aceites cítricos, incluyendo naranja y uva. El valenceno se produce por ciclación del precursor terpénico de pirofosfato acíclico difosfato de farnesilo (FPP) y la oxidación del valenceno da como resultado la formación de nootkatona. Valenceno y nootkatona se usan ambos en la industria de los perfumes y aromas.

El valenceno se extrajo originalmente de la piel de naranja de Valencia y la nootkatona se extrae de pieles de uva o se produce mediante la oxidación de valenceno extraído. Aunque se han intentado enfoques químicos para generar valenceno y nootkatona, las estructuras altamente complejas de estos compuestos han vuelto inalcanzables procesos sintéticos económicamente viables para su preparación en grandes cantidades. El valenceno se ha producido en células hospedadoras catalizadas por ácido nucleico que codifica una valenceno sintasa cítrica (véase p.ej. la patente de EE.UU. nº 7.442.785). Se buscan mejoras en la producción y medios alternativos de producción y la producción de valenceno y otros terpenos.

25

Por tanto, entre los objetos de la presente memoria, está la provisión de polipéptidos de valenceno sintasa y procedimientos para la producción de productos terpénicos cuya producción está catalizada por los polipéptidos. El documento US2012/246767 divulga polipéptidos de valenceno sintasa modificados y procedimientos de uso de polipéptidos de valenceno sintasa modificados. Divulga también procedimientos para producir terpeno sintasas modificadas.

30

El número de secuencia GU987104.1, que puede recuperarse de la base de datos en línea EMBLE, se refiere a ARNm de alfa-copaeno sintasa (Cop) de *Eleutherococcus trifoliatu*s, cds completa. Divulga una molécula de ARNm lineal con una longitud de secuencia de 1948.

## 35 RESUMEN

Se proporcionan valenceno sintasas de *Eryngium* (EgVS). Se proporcionan polipéptidos de valenceno sintasa de *Eryngium* aislados que tienen una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con el polipéptido de valenceno sintasa cuya secuencia se expone en la SEQ ID NO:1 o con un fragmento catalíticamente activo del mismo. La sintasa de *Eryngium glaciale* ejemplificada tiene la secuencia de aminoácidos expuesta en la SEQ ID NO:1. Las valenceno sintasas proporcionadas en la presente memoria catalizan la producción de valenceno y otros productos a partir de un precursor terpénico de pirofosfato acíclico tal como, pero sin limitación, difosfato de farnesilo. La producción puede efectuarse *in vitro* poniendo en contacto la sintasa aislada con un precursor terpénico de pirofosfato acíclico, o *in vivo* en una célula hospedadora adecuada que codifica la sintasa. Las células hospedadoras adecuadas producen o se modifican para producir un precursor de pirofosfato acíclico. Si las células son de origen humano, son células aisladas o cultivadas o un cultivo celular. Las células, tales como células de levadura, particularmente células que expresan un precursor terpénico de pirofosfato acíclico, que expresan EgVS producen muy altas cantidades de valenceno en comparación con las mismas células que expresan la valenceno sintasa cítrica (*Citrus sinensis*) (CVS) y variantes de la misma que se optimizan y/o modifican para expresar altos niveles de valenceno.

45

Se proporciona además un polipéptido de valenceno sintasa que es un fragmento catalíticamente activo del péptido de valenceno sintasa aislado de la presente invención, donde el fragmento catalíticamente activo cataliza la formación de valenceno a partir de un precursor terpénico de pirofosfato acíclico.

Se proporcionan también moléculas de ácido nucleico que codifican cualquiera de las valenceno sintasas proporcionadas en la presente memoria. Las moléculas de ácido nucleico incluyen moléculas de ácido nucleico aisladas y también ADNc. Por ejemplo, las moléculas de ácido nucleico proporcionadas en la presente memoria son aquellas que codifican la secuencia de aminoácidos expuesta en la SEQ ID NO:1, y también cualquiera que codifique una secuencia de aminoácidos que tenga al menos

55

un 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad de secuencia con un polipéptido de

60

valenceno sintasa cuya secuencia se expone en la SEQ ID NO: 1 o un fragmento catalíticamente activo del mismo, donde el polipéptido codificado cataliza la producción de valenceno a partir de un precursor terpénico de pirofosfato acíclico. Por ejemplo, la molécula de ácido nucleico aislada se selecciona de entre moléculas de ácido nucleico que contienen: (a) la secuencia de ácidos nucleicos expuesta en la SEQ ID NO:2; (b) una secuencia de ácidos nucleicos que tiene al menos un 85 % de identidad de secuencia con la secuencia de ácidos nucleicos expuesta en la SEQ ID NO:2 y que codifica un polipéptido de valenceno sintasa que tiene al menos un 90 % de identidad de secuencia con el polipéptido de SEQ ID NO:1 y (c) degeneraciones de (a) y (b). Se proporcionan también polipéptidos aislados y fragmentos catalíticamente activos codificados por las moléculas de ácido nucleico proporcionadas en la presente memoria. Los polipéptidos codificados pueden producirse mediante expresión en cualquier célula hospedadora, incluyendo células eucarióticas y procarióticas, tales como células de mamífero, células de levadura y células bacterianas. En ejemplos donde la célula hospedadora es una célula de mamífero que es una célula humana, la célula es una célula aislada o una célula en cultivo, tal como en un cultivo celular.

Se proporcionan también vectores que contienen las moléculas de ácido nucleico proporcionadas en la presente memoria. Los vectores incluyen vectores procarióticos y eucarióticos, incluyendo vectores víricos y vectores de levadura, tales como vectores de *Saccharomyces*.

Se proporcionan también células que contienen las moléculas de ácido nucleico que codifican una EgVS. Las células incluyen células procarióticas y células eucarióticas incluyendo, pero sin limitación, bacterias, células de levadura, insecto, planta y mamífero. En ejemplos donde la célula hospedadora es una célula de mamífero que es una célula humana, la célula es una célula aislada o una célula cultivada. En algunos ejemplos, el ácido nucleico es heterólogo de la célula. Las células de levadura proporcionadas en la presente memoria incluyen, pero sin limitación, células del género *Saccharomyces* y células del género *Pichia*. Las células bacterianas proporcionadas en la presente memoria incluyen células de *Escherichia coli*. Las células vegetales proporcionadas en la presente memoria incluyen protoplastos. Las células pueden producir el precursor acíclico, tal como difosfato de farnesilo, de forma nativa o pueden modificarse para producirlo o para producir más de una célula no modificada. Entre las células proporcionadas en la presente memoria están aquellas que codifican un polipéptido de valenceno sintasa donde: la célula hospedadora produce un precursor terpénico de pirofosfato acíclico; el polipéptido de valenceno sintasa es heterólogo del hospedador y la valenceno sintasa cataliza la producción de valenceno a partir del precursor terpénico de pirofosfato acíclico, tal como difosfato de farnesilo. Se proporcionan también plantas transgénicas que codifican los polipéptidos de valenceno sintasa proporcionados en la presente memoria. Pueden incluir ácido nucleico que codifica el polipéptido de valenceno sintasa o un vector adecuado que lo codifique. Las plantas transgénicas ejemplares incluyen plantas de tabaco.

Se proporcionan en la presente memoria procedimientos para producir un polipéptido de valenceno sintasa. Los procedimientos incluyen las etapas de introducir una molécula de ácido nucleico que codifica los polipéptidos de valenceno sintasa, o fragmentos catalíticamente activos de los mismos, proporcionados en la presente memoria, o un vector que codifica el polipéptido de valenceno sintasa o fragmentos catalíticamente activos del mismo, en una célula, por ejemplo mediante transfección o transformación; cultivar la célula *in vitro* o *in vivo* en condiciones adecuadas para la expresión del polipéptido de valenceno sintasa o fragmento catalíticamente activo del mismo y, opcionalmente, aislar el polipéptido de valenceno sintasa. Las células para uso en los procedimientos incluyen células hospedadoras eucarióticas y procarióticas incluyendo, pero sin limitación, bacterias, células de levadura, insecto, planta y mamífero. Las células de levadura para uso en los procedimientos incluyen, pero sin limitación, células del género *Saccharomyces* y células del género *Pichia*.

Se proporcionan también procedimientos para producir valenceno. Los procedimientos implican poner en contacto un precursor terpénico de pirofosfato acíclico con un polipéptido de valenceno sintasa o fragmento catalíticamente activo del mismo, proporcionado en la presente memoria, *in vitro* o *in vivo*, en condiciones adecuadas para la formación de valenceno a partir del precursor terpénico de pirofosfato acíclico, tal como difosfato de farnesilo. En algunos ejemplos, el valenceno se aísla o purifica. En los procedimientos proporcionados, la etapa de puesta en contacto del precursor terpénico de pirofosfato acíclico con el polipéptido de valenceno sintasa o fragmento catalíticamente activo del mismo puede efectuarse *in vitro* o *in vivo* en una célula hospedadora adecuada proporcionada en la presente memoria. Cuando se efectúa *in vivo*, pueden introducirse los ácidos nucleicos o vectores en las células como se describe en la presente memoria, de tal modo que el polipéptido de valenceno sintasa codificado o fragmento catalítico del mismo sea heterólogo de las células, y se cultivan las células en condiciones que se se exprese la valenceno sintasa codificada. Las células para uso en dichos procedimientos pueden modificarse para producir el precursor acíclico o para producir más del precursor acíclico que las células no modificadas. Las células y procedimientos ejemplares para seleccionar tales células incluyen aquellos descritos en las patentes de EE.UU. expedidas nº 8.609.371(pub. de EE.UU. nº 2010-015155) y 8.481.286(pub. de EE.UU. nº 2010-0151519). El valenceno puede convertirse en nootkatona, tal como mediante oxidación, que puede practicarse mediante procedimientos conocidos, incluyendo procedimientos biosintéticos y químicos.

Por tanto, se proporcionan también en la presente memoria procedimientos de producción de nootkatona. Tales procedimientos implican producir valenceno usando las etapas expuestas en cualquiera de los procedimientos proporcionados en la presente memoria para producir valenceno; aislar el producto valenceno; oxidar el valenceno para producir nootkatona y aislar la nootkatona. El valenceno puede oxidarse química o biosintéticamente.

En cualquiera de los procedimientos proporcionados, puede usarse cualquier procedimiento conocido por el especialista en la materia para aislar valenceno y/o nootkatona, así como cualquier otro producto de la reacción o reacciones, incluyendo aristoloqueno, que es el compuesto del pico 2 de la Figura 3A. Tales procedimientos de aislamiento incluyen, pero sin limitación, extracción con un disolvente orgánico y/o cromatografía en columna.

#### BREVE DESCRIPCIÓN DE LAS FIGURAS

La **FIGURA 1** es el esquema de reacción para producción de valenceno y nootkatona. Las valenceno sintasas son terpeno ciclasas o sintasas vegetales de clase 1 que convierten difosfato de farnesilo (FPP) en el sesquiterpeno valenceno. El valenceno puede oxidarse formando nootkatona.

**FIGURAS 2A-B.** La **Figura 2A** representa la cromatografía de gases del extracto de aceite vegetal de *Eryngium glaciale*. Se detecta valenceno (pico 1) a los 10,18 minutos. La **Figura 2B** representa el espectro de masas del pico a 10,18 minutos que corresponde al valenceno.

**FIGURAS 3A-B.** La **Figura 3A** representa el cromatograma de gases de una cepa de levadura ejemplar designada cepa ALX7-95 transformada con un gen que codifica la valenceno sintasa de *Eryngium glaciale*. Se observa valenceno (pico 1) a los 12,46 minutos. Se observa un compuesto adicional (designado pico 2) a los 12,37 minutos. La **Figura 3B** representa el espectro de masas del pico a 12,46, que corresponde a valenceno.

Las **FIGURAS 4A-G** representan alineamientos ejemplares de valenceno sintasa de *Eryngium glaciale* (EGVS) con otras valenceno sintasas. "\*" significa que los residuos o nucleótidos en esa columna son idénticos en todas las secuencias del alineamiento, ":" significa que se han observado sustituciones conservadas y "." significa que se observan sustituciones semiconservadas. Por ejemplo, la **Figura 4A** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con valenceno sintasa de *Citrus sinensis* (CVS) expuesta en la SEQ ID NO:14. La **Figura 4B** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con valenceno sintasa de *Citrus x paradisi* (CVS) expuesta en la SEQ ID NO:15. La **Figura 4C** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con valenceno sintasa de *Vitis vinifera* (VVS) expuesta en la SEQ ID NO:16. La **Figura 4D** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con valenceno sintasa de *Chamaecyparis nootkatensis* (CNVS) expuesta en la SEQ ID NO:17. La **Figura 4E** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con una valenceno sintasa CVS modificada designada V277 (descrita en la publicación de nº de serie de EE.UU. nº 2012-0246767 en tramitación junto con la presente), cuya secuencia se expone en la SEQ ID NO:31. La **Figura 4F** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con valenceno sintasa de *Chamaecyparis nootkatensis* (CNVS) expuesta en la SEQ ID NO:29. La **Figura 4G** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con valenceno sintasa de *Perilla frutescens* (PFVS) expuesta en la SEQ ID NO:38.

Las **FIGURAS 5A-J** representan alineamientos ejemplares de valenceno sintasa de *Eryngium glaciale* (EGVS) con otras terpeno sintasas. "\*" significa que los residuos o nucleótidos en esa columna son idénticos en todas las secuencias del alineamiento, ":" significa que se han observado sustituciones conservadas y "." significa que se observan sustituciones semiconservadas. Por ejemplo, la **Figura 5A** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con alfa-copaeno sintasa de *Eleutherococcus trifolius* (ETACS) expuesta en la SEQ ID NO:18. La **Figura 5B** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con germacreno-D sintasa de *Actinidia deliciosa* (ADGDS) expuesta en la SEQ ID NO:19. La **Figura 5C** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con (-)-germacreno-D sintasa de *Vitis vinifera* (VDGS) expuesta en la SEQ ID NO:20. La **Figura 5D** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con sesquiterpeno sintasa de *Santalum murrayanum* (SMSS) expuesta en la SEQ ID NO:21. La **Figura 5E** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con (+)-delta-cadineno sintasa de *Ricinus communis* (RCDCS) expuesta en la SEQ ID NO:22. La **Figura 5F** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con delta-cadineno sintasa de *Citrus x paradisi* (CDCS) expuesta en la SEQ ID NO:23. La **Figura 5G** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con 5-epiaristoloqueno sintasa de *Nicotiana tabacum* (TEAS) expuesta en la SEQ ID NO:24. La **Figura 5H** representa el alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con premnaspirodieno sintasa de *Hyoscyamus muticus* (HPS) expuesta en la SEQ ID NO:25. La **Figura 5I** representa un alineamiento de valenceno sintasa de *Eryngium glaciale* expuesta en la SEQ ID NO:1 con germacreno-D sintasa de *Citrus hystrix* expuesta en la SEQ ID NO:45. La **Figura 5J** representa un alineamiento de V277 expuesta en la SEQ ID NO:31 y valenceno sintasa cítrica expuesta en la SEQ ID NO:14.

La **Figura 6** representa los parámetros cinéticos en estado estacionario de la valenceno sintasa de *E. glaciale* (EgVS), cuya secuencia se expone en la SEQ ID NO:1 (**izquierda**) y la valenceno sintasa V277 (marcada CVS), cuya secuencia se expone en la SEQ ID NO:31 (**derecha**); FPP= difosfato de farnesilo.

La **Figura 7** demuestra que el valenceno producido en levadura catalizado por EVS es significativamente mayor que el producido en levadura catalizado por CVS.

## DESCRIPCIÓN DETALLADA

### Esquema

10

#### A. Definiciones

#### B. Visión general

##### 1. Estructura del valenceno y usos

##### 2. Nootkatona

15

#### 3. Valenceno sintasas

##### a. Estructura

##### b. Actividades

##### 4. Ensayos para detectar la actividad enzimática de polipéptidos de valenceno sintasa

20

#### C. Moléculas de ácido nucleico que codifican valenceno sintasa de *Eryngium glaciale* y polipéptidos codificados

##### 1. Aislamiento de ácido nucleico que codifica valenceno sintasas

##### a. Generación de un ácido nucleico modificado

##### 2. Vectores y células para la expresión de polipéptidos de valenceno sintasa

25

#### 3. Sistemas de expresión

##### a. Células procarióticas

##### b. Células de levadura

##### c. Plantas y células vegetales

##### d. Insectos y células de insecto

30

##### e. Células de mamífero

##### 4. Purificación

#### D. Polipéptidos de valenceno sintasa

##### 1. Polipéptidos de valenceno sintasa de *Eryngium glaciale*

35

#### 2. Modificaciones de polipéptidos de valenceno sintasa de *Eryngium glaciale*

##### a. Polipéptidos truncados

##### b. Polipéptidos con actividades o propiedades alteradas

##### c. Intercambios de dominio

##### d. Variantes adicionales

40

##### e. Proteínas de fusión

#### E. Procedimientos para producir terpenos y procedimientos para detectar tales productos y la actividad de polipéptidos de valenceno sintasa

##### 1. Producción de terpenos catalizada por valenceno sintasa de *Eryngium glaciale*

45

##### a. Células ejemplares

##### b. Cultivo de células

##### c. Aislamiento y valoración de productos

##### 2. Producción de nootkatona

50

#### F. EJEMPLOS

#### A. DEFINICIONES

55

60

A menos que se definan de otro modo, todos los términos técnicos y científicos usados en la presente memoria tienen el mismo significado que se entiende comúnmente por un especialista en la materia a la que pertenece la invención.

En el caso de que haya una pluralidad de definiciones para los términos de la presente memoria, prevalecen aquellos en esta sección. Cuando se hace referencia a una URL u otro de tales identificadores o direcciones, se entiende que tales identificadores pueden cambiar y la información particular en internet puede ir y venir, pero puede encontrarse información equivalente buscando en internet. La referencia a la misma evidencia la disponibilidad y

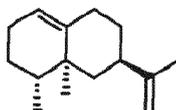
difusión pública de tal información.

Como se usa en la presente memoria, un precursor terpénico de pirofosfato acíclico es cualquier compuesto de pirofosfato acíclico que sea un precursor para la producción de al menos un terpeno incluyendo, pero sin limitación, 5 pirofosfato de farnesilo (FPP), pirofosfato de geranilo (GPP) y pirofosfato de geranilgeranilo (GGPP). Los precursores terpénicos de pirofosfato acíclico son por tanto sustratos de terpeno sintasas.

Como se usa en la presente memoria, un terpeno es un hidrocarburo insaturado basado en la unidad isopreno ( $C_5H_8$ ), y que tiene una fórmula general  $C_{5x}H_{8x}$ , tal como  $C_{10}H_{16}$ . La referencia a un terpeno incluye terpenos 10 acíclicos, monocíclicos y policíclicos. Los terpenos incluyen, pero sin limitación, monoterpenos que contienen 10 átomos de carbono; sesquiterpenos que contienen 15 átomos de carbono; diterpenos que contienen 20 átomos de carbono y triterpenos que contienen 30 átomos de carbono. La referencia a un terpeno incluye también estereoisómeros del terpeno.

15 Como se usa en la presente memoria, una terpeno sintasa es un polipéptido capaz de catalizar la formación de uno o más terpenos a partir de un precursor terpénico de pirofosfato acíclico, por ejemplo FPP, GPP o GGPP.

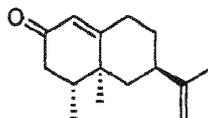
Como se usa en la presente memoria, el valenceno es un sesquiterpeno que tiene la siguiente estructura:



20

La referencia a valenceno incluye la referencia a cualquier isómero del mismo incluyendo, pero sin limitación, (+)-valenceno.

25 Como se usa en la presente memoria, la nootkatona es un sesquiterpenoide que tiene la siguiente estructura:



La referencia a nootkatona incluye la referencia a cualquier isómero de la misma.

30

Como se usa en la presente memoria, una "valenceno sintasa" o "polipéptido de valenceno sintasa" es un polipéptido capaz de catalizar la formación de valenceno a partir de un precursor terpénico de pirofosfato acíclico, tal como difosfato de farnesilo (FPP). Se incluyen entre los polipéptidos de valenceno sintasa divulgados en la presente memoria cualquiera que tenga más de o más de aproximadamente o un 63 %, 65 %, 70 %, 75 %, 76 %, 77 %, 78 %, 35 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia con la valenceno sintasa expuesta en la SEQ ID NO:1 cuando se alinean en toda su longitud o un fragmento catalíticamente activo del mismo. Estos polipéptidos catalizan la producción de valencia como único producto o uno entre una mezcla de productos formados a partir de la reacción de un precursor terpénico de pirofosfato acíclico con una valenceno sintasa. Típicamente, la valencia es el producto 40 más prevalente o está entre varios productos prevalentes. Por ejemplo, la cantidad de valenceno producida a partir de la reacción de una valenceno sintasa con un precursor terpénico de pirofosfato acíclico es típicamente de al menos o al menos aproximadamente un 1 %, 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 % o más de la cantidad total de terpeno producida en la reacción. En algunos aspectos, el valenceno es el producto terpénico predominante (*es decir*, presente en mayores cantidades que cualquier otro terpeno individual producido a partir de 45 la reacción de un precursor terpénico de pirofosfato acíclico con una valenceno sintasa).

La referencia a una valenceno sintasa incluye cualquier polipéptido que catalice la producción de valenceno incluyendo, pero sin limitación, un polipéptido producido recombinantemente, un polipéptido producido 50 sintéticamente y un polipéptido de valenceno sintasa extraído o aislado de células y material vegetal del que se ha aislado valenceno incluyendo, pero sin limitación, las especies de cardo como se proporcionan en la presente memoria. Otros polipéptidos de valenceno sintasa ejemplares incluyen aquellos aislados de frutas cítricas, flores de parra (*p.ej. Vitis vinifera* L. cv. Gewürztraminer y *Vitis vinifera* L. cv. Cabernet Sauvignon (véanse Lucker y col., (2004) Photochemistry 65(19): 2649-2659 y Martin y col., (2009) Proc. Natl. Acad. Sci, USA 106: 7245-7250) y Perilla (shiso verde). Las valenceno sintasas cítricas (CVS) incluyen, pero sin limitación, valenceno sintasa de *Citrus* 55 *sinensis* (naranja dulce) (SEQ ID NO: 14 y 34) y *Citrus x paradisi* (uva) (SEQ ID NO: 15, 26 y 27). Otros polipéptidos

- de valenceno sintasa ejemplares incluyen valenceno sintasa aislada de flores de parra, incluyendo *Vitis vinifera* L. cv. Gewürztraminer y *Vitis vinifera* L. cv. Cabernet Sauvignon (SEQ ID NO:16 y 28), valenceno sintasas aisladas de *Chamaecyparis nootkatensis pendula* (SEQ ID NO:17 y 29) y valenceno sintasa de *Perilla frutescens* (SEQ ID NO:38). La referencia a valenceno sintasa incluye valenceno sintasa de cualquier género o especie, e incluye
- 5 variantes alélicas o de especie, variantes codificadas por variantes de empalme y otras variantes de la misma, incluyendo polipéptidos que tienen al menos o al menos aproximadamente un 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 % o más identidad de secuencia con la valenceno sintasa expuesta en la SEQ ID NO: 1. Las valenceno sintasas incluyen también fragmentos de las mismas que retienen actividad valenceno sintasa.
- 10
- Como se usa en la presente memoria, los fragmentos catalíticamente activos de los polipéptidos de valenceno sintasa son polipéptidos de sintasa truncados que retienen la capacidad de catalizar la formación de valenceno a partir de un precursor terpénico de pirofosfato acíclico. Un especialista en la materia puede identificar fácilmente un fragmento de una sintasa que retenga actividad catalítica ensayándolo en cualquier ensayo adecuado, tal como
- 15 cualquiera descrito en la presente memoria, y detectando la formación de valenceno a partir del precursor.
- Como se usa en la presente memoria, "actividad valenceno sintasa" (a la que se hace referencia también en la presente memoria como actividad catalítica) hace referencia a la capacidad de catalizar la formación de valenceno a partir de un precursor terpénico de pirofosfato acíclico, tal como difosfato de farnesilo (FPP). Los procedimientos
- 20 para valorar la formación de valenceno a partir de la reacción de una sintasa con un precursor terpénico de pirofosfato acíclico, tal como FPP, son bien conocidos en la materia y se describen en la presente memoria. Por ejemplo, la sintasa puede expresarse en una célula hospedadora, tal como una célula de levadura, que produce también FPP. La producción de valenceno puede valorarse y cuantificarse entonces usando, por ejemplo, cromatografía de gases-espectrometría de masas (GC-MS) (véanse los ejemplos siguientes). Se considera que una
- 25 sintasa exhibe actividad valenceno sintasa o la capacidad de catalizar la formación de valenceno a partir de un precursor terpénico de pirofosfato acíclico tal como FPP si la cantidad de valenceno producida a partir de la reacción es de al menos o de al menos aproximadamente un 1 %, 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 % o más que la cantidad total de terpenos producidos en la reacción.
- 30 Como se usa en la presente memoria, "actividad catalítica aumentada" con referencia a la actividad de una valenceno sintasa significa que aumenta la capacidad de catalizar la formación de valenceno a partir de un precursor terpénico de pirofosfato acíclico, tal como difosfato de farnesilo (FPP), dando así como resultado una formación aumentada de valenceno. Con fines de la presente memoria, una valenceno sintasa exhibe actividad catalítica
- 35 aumentada si la cantidad de valenceno producida a partir de FPP por la valenceno sintasa modificada es de 10 a 500 %, de 10 a 250 %, de 50 a 250 %, de 100 a 500 % o de 100 a 250 % mayor que la cantidad de valenceno producida a partir de FPP por la valenceno sintasa expuesta en la SEQ ID NO: 1, tal como un 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 110 %, 120 %, 130 %, 140 %, 150 %, 160 %, 170 %, 180 %, 190 %, 200 %, 250 %, 300 %, 350 %, 400 %, 500 % o más mayor que la cantidad de valenceno producida a partir de FPP por la
- 40 valenceno sintasa expuesta en la SEQ ID NO:1. Por ejemplo, una valenceno sintasa exhibe actividad catalítica aumentada si la cantidad de valenceno producida a partir de FPP por la valenceno sintasa modificada es de al menos o de aproximadamente al menos un 110 %, 115 %, 120 %, 125 %, 130 %, 135 %, 140 %, 145 %, 150 %, 160 %, 170 %, 180 %, 200 %, 250 %, 300 %, 350 %, 400 %, 500 %, 1500 %, 2000 %, 3000 %, 4000 % o 5000 % de la cantidad de valenceno producida a partir de FPP por la valenceno sintasa de tipo silvestre expuesta en la SEQ ID NO:1 en las mismas condiciones.
- 45
- Como se usa en la presente memoria, "silvestre" o "nativa" con referencia a la valenceno sintasa hace referencia a un polipéptido de valenceno sintasa codificado por un gen de valenceno sintasa nativo o de origen natural, incluyendo variantes alélicas, que está presente en un organismo, incluyendo una planta, en la naturaleza. La referencia a valenceno sintasa de tipo silvestre sin referencia a una especie pretende englobar cualquier especie de
- 50 una valenceno sintasa de tipo silvestre. La secuencia aminoacídica de valenceno sintasas ejemplares se expone en la SEQ ID NO:1 (aislada de *Eryngium glaciale*), SEQ ID NO: 14 (aislada de *Citrus sinensis* cv. Valencia, *Citrus sinensis* cv. Cara Cara y *Citrus x paradisi*), SEQ ID NO:34 (aislada de *Citrus sinensis* cv. Valencia), SEQ ID NO:27 (aislada de *Citrus x paradisi*), SEQ ID NO:26 (aislada de *Citrus x paradisi*), SEQ ID NO:16 y 28 (aislada de *Vitis vinifera*), SEQ ID NO:29 (aislada de *Chamaecyparis nootkatensis pendula*) y SEQ ID NO:38 (aislada de *Perilla frutescens*).
- 55
- Como se usa en la presente memoria, variantes de especie hace referencia a variantes en polipéptidos entre diferentes especies.
- 60 Como se usa en la presente memoria, variantes alélicas hace referencia a variaciones en las proteínas codificadas entre los miembros de la misma especie.

Como se usa en la presente memoria, una variante de empalme hace referencia a una variante producida mediante un procesamiento diferencial de un transcrito primario de ADN genómico que da como resultado más de un tipo de ARNm.

5

Como se usa en la presente memoria, "polipéptido de valenceno sintasa modificado" hace referencia a un polipéptido de valenceno sintasa que tiene una o más diferencias aminoacídicas en comparación con un polipéptido de valenceno sintasa no modificado o de tipo silvestre. La una o más diferencias aminoacídicas pueden ser mutaciones aminoacídicas tales como uno o más reemplazos (sustituciones), inserciones o deleciones aminoacídicos, o pueden ser inserciones o deleciones de dominios completos, y cualquier combinación de los mismos. Típicamente, un polipéptido de valenceno sintasa modificado tiene una o más modificaciones en la secuencia primaria en comparación con un polipéptido de valenceno sintasa no modificado o de tipo silvestre. Por ejemplo, un polipéptido de valenceno sintasa modificado proporcionado en la presente memoria puede tener al menos 1, 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 15 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135 o más diferencias aminoacídicas en comparación con un polipéptido de valenceno sintasa no modificado. Se contempla cualquier modificación a condición de que el polipéptido resultante exhiba al menos una actividad valenceno sintasa asociada a un polipéptido de valenceno sintasa de tipo silvestre tal como, por ejemplo, actividad catalítica la capacidad de unirse a FPP y/o la capacidad de catalizar la formación de valenceno a partir de FPP.

Como se usa en la presente memoria, la referencia a un polipéptido de valenceno sintasa modificado productor de valenceno a partir de FPP en una cantidad que es mayor que la cantidad de valenceno producida a partir de FPP por una valenceno sintasa de referencia, tal como una valenceno sintasa de tipo silvestre, indica que la valenceno 25 sintasa modificada produce al menos o aproximadamente un 10 % más de valenceno a partir de FPP de lo que produce la valenceno sintasa de referencia. Por ejemplo, tal polipéptido de valenceno sintasa modificada puede producir al menos o al menos aproximadamente un 10 %, 11 %, 12 %, 13 %, 14 %, 15 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 110 %, 120 %, 130 %, 140 %, 150 %, 160 %, 170 %, 180 %, 190 %, 200 %, 250 %, 300 %, 350 %, 400 %, 500 %, 600 %, 700 %, 800 %, 900 %, 1000 %, 2000 %, 5000 % o más de valenceno a 30 partir de FPP en comparación con la cantidad de valenceno producida a partir de FPP por una valenceno sintasa de referencia. La cantidad de valenceno producida a partir de FPP por una valenceno sintasa puede valorarse mediante cualquier procedimiento conocido en la materia. Cuando se compara la cantidad de valenceno producida a partir de FPP por dos valenceno sintasas, tales como una valenceno sintasa modificada y una valenceno sintasa de referencia, tal como como una valenceno sintasa de tipo silvestre, se entiende que el ensayo se practica en las 35 mismas condiciones para cada sintasa. En un ejemplo, la cantidad de valenceno producida a partir de FPP por dos valenceno sintasas, tales como una valenceno sintasa modificada y una valenceno sintasa de referencia, se valora expresando la valenceno sintasa modificada y la valenceno sintasa de referencia separadamente en una célula de levadura de la misma cepa (donde la expresión es a partir del mismo vector de expresión) que produce también FPP, y cultivando las células en las mismas condiciones de tal modo que se produzca valenceno. La cantidad de 40 valenceno producida en el cultivo celular que expresa la valenceno sintasa modificada se compara con la cantidad de valenceno producida en el cultivo celular que expresa la valenceno sintasa de referencia, usando procedimientos de cuantificación bien conocidos en la materia, tales como GC-MS.

Como se usa en la presente memoria, residuos correspondientes hace referencia a los residuos que aparecen en 45 loci alineados. Los polipéptidos relacionados o variantes se alinean mediante cualquier procedimiento conocido por los especialistas en la materia. Tales procedimientos maximizan típicamente los apareamientos, e incluyen procedimientos tales como usar alineamientos manuales y usar los numerosos programas de alineamientos disponibles (por ejemplo, BLASTP) y otros conocidos por los especialistas en la materia. Alineando las secuencias de polipéptidos, un especialista en la materia puede identificar los residuos correspondientes, usando residuos 50 aminoacídicos conservados e idénticos como guías. Las posiciones correspondientes pueden estar también basadas en alineamientos estructurales, por ejemplo usando alineamientos simulados informáticamente de la estructura proteica. Por ejemplo, como se muestra en la Figura 4A, el Asp315 de valenceno sintasa de *E. glaciale* expuesta en la SEQ ID NO: 1 corresponde al Asp301 de valenceno sintasa de *C. sinensis* expuesta en la SEQ ID NO: 14.

Como se usa en la presente memoria, dominio o región (típicamente una secuencia de 3 o más, generalmente 5 o 7 55 o más aminoácidos) hace referencia a una porción de una molécula, tal como una proteína o los ácidos nucleicos codificantes, que es estructural y/o funcionalmente distinta de otras porciones de la molécula y es identificable. Una proteína puede tener uno, o más de uno, dominios distintos. Por ejemplo, un dominio puede identificarse, definirse o distinguirse por homología de la secuencia del mismo con miembros de la familia relacionados, tales como otras 60 terpeno sintasas. Un dominio puede ser una secuencia lineal de aminoácidos o una secuencia no lineal de aminoácidos. Muchos polipéptidos contienen una pluralidad de dominios. Tales dominios son conocidos, y pueden

identificarse, por los especialistas en la materia. Para ejemplificación en la presente memoria, se proporcionan definiciones, pero se entiende que está dentro de las competencias en la materia reconocer dominios particulares por el nombre. Si es necesario, puede emplearse un software apropiado para identificar dominios. Por ejemplo, como se discute anteriormente, pueden identificarse los dominios correspondientes en diferentes terpeno sintasas por alineamientos de secuencia, tal como usando herramientas y algoritmos bien conocidos en la materia (por ejemplo, BLASTP).

Como se usa en la presente memoria, un dominio funcional hace referencia a aquellas porciones de un polipéptido que se reconocen gracias a una actividad funcional, tal como actividad catalítica. Un dominio funcional puede distinguirse por su función, tal como por actividad catalítica, o una capacidad de interactuar con una biomolécula, tal como unión a sustrato o unión a metal. En algunos ejemplos, un dominio puede exhibir independientemente una función o propiedad biológica, de tal modo que el dominio independientemente, o fusionado con otra molécula, pueda practicar una actividad tal como, por ejemplo, actividad catalítica o unión a sustrato.

Como se usa en la presente memoria, un dominio estructural hace referencia a aquellas porciones de una cadena polipeptídica que pueden formar una estructura plegada independientemente en una proteína compuesta por uno o más motivos estructurales.

Como se usa en la presente memoria, "heterólogo" con respecto a una secuencia aminoacídica o de ácido nucleico hace referencia a porciones de una secuencia que no están presentes en el polipéptido nativo ni codificadas por el polinucleótido nativo. Por ejemplo, una porción de aminoácidos de un polipéptido, tal como un dominio o región o porción del mismo, de una valenceno sintasa es heteróloga del mismo si tales aminoácidos no están presentes en una valenceno sintasa nativa o de tipo silvestre (*p.ej.* como se expone en la SEQ ID NO:1) ni codificados por el polinucleótido que codifica una valenceno sintasa nativa o de tipo silvestre. Se hace referencia a los polipéptidos que contienen tales aminoácidos heterólogos o polinucleótidos que codifican los mismos como "polipéptidos quiméricos" o "polinucleótidos quiméricos", respectivamente.

Como se usa en la presente memoria, la frase "se mejora una propiedad de la terpeno sintasa modificada en comparación con la primera terpeno sintasa" hace referencia a un cambio deseable en una propiedad de una terpeno sintasa modificada en comparación con una terpeno sintasa que no contiene la modificación o modificaciones. Típicamente, la propiedad o propiedades se mejoran de tal modo que la cantidad de un terpeno deseado producido a partir de la reacción de un sustrato con la terpeno sintasa modificada aumente en comparación con la cantidad del terpeno deseado producida a partir de la reacción de un sustrato con una terpeno sintasa que no está modificada así. Las propiedades ejemplares que pueden mejorarse en una terpeno sintasa modificada incluyen, por ejemplo, producción de terpeno, actividad catalítica, distribución de producto, especificidad de sustrato, regioselectividad y estereoselectividad. Una o más de las propiedades pueden valorarse usando procedimientos bien conocidos en la materia para determinar si la propiedad había mejorado (*es decir* se había alterado para ser más deseable para la producción de un terpeno o terpenos deseados).

Como se usa en la presente memoria, la producción de terpeno (a la que se hace referencia también como rendimiento de terpeno) hace referencia a la cantidad (en peso o peso/volumen) de terpeno producida a partir de la reacción de un precursor terpénico de pirofosfato acíclico con una terpeno sintasa. La referencia a producción de terpeno total hace referencia a la cantidad total de todos los terpenos producidos a partir de la reacción, mientras que la referencia a la producción de terpeno específica hace referencia a la cantidad de un terpeno específico (*p.ej.* valenceno) producido a partir de la reacción.

Como se usa en la presente memoria, una producción de terpeno mejorada hace referencia a un aumento en la cantidad total de terpeno (*es decir*, producción de terpeno total mejorada) o a un aumento en la cantidad específica de terpeno (*es decir*, producción de terpeno específico mejorada) producida a partir de la reacción de un precursor terpénico de pirofosfato acíclico con una terpeno sintasa modificada en comparación con la cantidad producida a partir de la reacción del mismo precursor terpénico de pirofosfato acíclico con una terpeno sintasa que no está modificada así. La cantidad de terpeno (total o específico) producida a partir de la reacción de un precursor terpénico de pirofosfato acíclico con una terpeno sintasa modificada puede aumentarse en al menos o aproximadamente al menos un 5 %, 10 %, 15 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % o más en comparación con la cantidad de terpeno producida a partir de la reacción del mismo precursor terpénico de pirofosfato acíclico en las mismas condiciones con una terpeno sintasa que no está modificada así.

Como se usa en la presente memoria, especificidad de sustrato hace referencia a la preferencia de una valenceno sintasa por un sustrato diana frente a otro, tal como un precursor terpénico de pirofosfato acíclico (*p.ej.*, pirofosfato de farnesilo (FPP), pirofosfato de geranilo (GPP) o pirofosfato de geranilgeranilo (GGPP)) frente a otro. La especificidad de sustrato puede valorarse usando procedimientos bien conocidos en la materia, tales como aquellos

que calculan  $k_{cat}/K_m$ . Por ejemplo, la especificidad de sustrato puede valorarse comparando la  $K_{cat}/K_m$  relativa, que es una medida de la eficiencia catalítica, de la enzima frente a diversos sustratos (*p.ej.* GPP, FPP, GGPP).

5 Como se usa en la presente memoria, especificidad alterada hace referencia a un cambio en la especificidad de sustrato de un polipéptido de terpeno sintasa modificado (tal como un polipéptido de valenceno sintasa modificado) en comparación con una terpeno sintasa que no está modificada así (tal como, por ejemplo, una valenceno sintasa de tipo silvestre). La especificidad (*p.ej.*,  $k_{cat}/K_m$ ) de un polipéptido de terpeno sintasa modificado por un sustrato, tal como FPP, GPP o GGPP, puede alterarse en al menos o aproximadamente al menos un 10 %, 15 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % o más en comparación con la especificidad de una valenceno sintasa de partida por el mismo sustrato.

15 Como se usa en la presente memoria, especificidad de sustrato mejorada hace referencia a un cambio o alteración en la especificidad de sustrato hacia una especificidad más deseada. Por ejemplo, especificidad de sustrato mejorada puede incluir un aumento de la especificidad de sustrato de un polipéptido de terpeno sintasa modificado por un sustrato deseado, tal como FPP, GPP o GGPP. La especificidad (*p.ej.*,  $k_{cat}/K_m$ ) de un polipéptido de terpeno sintasa modificado por un sustrato, tal como FPP, GPP o GGPP, puede aumentarse en al menos o aproximadamente al menos un 10 %, 15 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % o más en comparación con la especificidad de una terpeno sintasa que no está modificada así.

20 Como se usa en la presente memoria, "distribución de producto" hace referencia a las cantidades relativas de diferentes terpenos producidos a partir de la reacción entre un precursor terpénico de pirofosfato acíclico, tal como FPP, y una terpeno sintasa, incluyendo los polipéptidos de valenceno sintasa modificados proporcionados en la presente memoria. La cantidad de terpeno producida puede representarse como un porcentaje de los productos totales producidos por la terpeno sintasa. Por ejemplo, la distribución de producto resultante de la reacción de FPP con una valenceno sintasa puede ser de 90 % (peso/volumen) de valenceno y 10 % (peso/volumen) de  $\beta$ -elemeno. Los procedimientos para valorar el tipo y cantidad de un terpeno en una solución son conocidos en la materia y se describen en la presente memoria e incluyen, por ejemplo, cromatografía de gases-espectrometría de masas (GC-MS) (véanse los ejemplos siguientes).

30 Como se usa en la presente memoria, una distribución de producto alterada hace referencia a un cambio en la cantidad relativa de terpenos individuales producidos a partir de la reacción entre un precursor terpénico de pirofosfato acíclico, tal como FPP, y una terpeno sintasa tal como valenceno sintasa. Típicamente, el cambio se valora determinando la cantidad relativa de terpenos individuales producidos a partir del precursor terpénico de pirofosfato acíclico usando una primera sintasa (*p.ej.* sintasa de tipo silvestre) y comparándola entonces con la cantidad relativa de terpenos individuales producidos usando una segunda sintasa (*p.ej.* una sintasa modificada). Se considera que aparece una distribución de producto alterada si la cantidad relativa de uno cualquiera o más de los terpenos aumenta o disminuye en al menos o aproximadamente en al menos un 0,5 %, 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 50 %, 60 %, 70 %, 80 % o más.

40 Como se usa en la presente memoria, una distribución de producto mejorada hace referencia a un cambio en la distribución de producto hacia uno que sea más deseable, *es decir* que contenga cantidades relativas de terpenos más deseables. Por ejemplo, una distribución de producto mejorada puede contener una cantidad aumentada de un terpeno deseado y una cantidad disminuida de un terpeno que no es tan deseado. La cantidad de terpeno deseado en una distribución de producción mejorada puede aumentar en al menos o aproximadamente en al menos un 0,5 %, 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 50 %, 60 %, 70 %, 80 % o más. La cantidad de un terpeno que no se desea en una distribución de producción mejorada puede disminuir en al menos o aproximadamente en al menos un 0,5 %, 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 50 %, 60 %, 70 %, 80 % o más.

50 Como se usa en la presente memoria, los ácidos nucleicos o moléculas de ácido nucleico incluyen ADN, ARN y análogos de los mismos, incluyendo ácidos peptidonucleicos (APN) y mezclas de los mismos. Los ácidos nucleicos pueden ser monocatenarios o bicatenarios. Cuando se hace referencia a sondas o cebadores, que están opcionalmente marcados tales como con un marcador detectable tal como un marcador fluorescente o radiomarcador, se contemplan moléculas monocatenarias. Tales moléculas son típicamente de una longitud tal que su diana es estadísticamente única o de bajo número de copias (típicamente menos de 5, generalmente menos de 3) para sondear o cebar una colección. Generalmente, una sonda o cebador contiene al menos 14, 16 o 30 nucleótidos contiguos de secuencia complementaria a o idéntica a un gen de interés. Las sondas y cebadores pueden ser de 10, 20, 30, 50, 100 o más ácidos nucleicos de largo.

60 Como se usa en la presente memoria, el término "polinucleótido" significa un polímero monocatenario o bicatenario de bases desoxirribonucleotídicas o ribonucleotídicas leídas del extremo 5' al 3'. Los polinucleótidos incluyen ARN y

ADN y pueden aislarse de fuentes naturales, sintetizarse *in vitro* o prepararse a partir de una combinación de moléculas naturales y sintéticas. La longitud de una molécula polinucleotídica se da en la presente memoria en términos de nucleótidos (abreviado "nt") o pares de bases (abreviado "pb"). El término nucleótido se usa para moléculas monocatenarias y bicatenarias donde permita el contexto. Cuando se aplica el término a moléculas bicatenarias, se usa para denotar la longitud global y se entenderá que es equivalente al término pares de bases. Se reconocerá por los especialistas en la materia que las dos hebras de un polinucleótido bicatenario pueden diferir ligeramente en longitud, y que los extremos de las mismas pueden estar escalonados; por tanto pueden no estar apareados todos los nucleótidos en una molécula polinucleotídica bicatenaria. Tales extremos no apareados no superarán, en general, los 20 nucleótidos de longitud.

10 Como se usa en la presente memoria, un ácido nucleico heterólogo es un ácido nucleico que no se produce normalmente *in vivo* por la célula en que se expresa o que se produce por la célula pero está en un locus diferente o se expresa diferentemente, o que media o codifica mediadores que alteran la expresión del ácido nucleico endógeno, tal como ADN, afectando a la transcripción, traducción u otros procesos bioquímicos regulables. El ácido nucleico heterólogo es generalmente no endógeno de la célula en que se introduce, sino que se ha obtenido de otra célula o preparado sintéticamente. El ácido nucleico heterólogo puede ser endógeno, pero es un ácido nucleico que se expresa a partir de un locus diferente o alterado en su expresión. Generalmente, aunque no necesariamente, tal ácido nucleico codifica ARN y proteínas que no se producen normalmente por la célula o del mismo modo en la célula en que se expresa. Puede hacerse referencia también al ácido nucleico heterólogo, tal como ADN, como ácido nucleico extraño, tal como ADN. Por tanto, el ácido nucleico heterólogo o ácido nucleico extraño incluye una molécula de ácido nucleico no presente en la orientación o posición exacta en que se encuentra en un genoma la contrapartida de molécula de ácido nucleico, tal como ADN. Puede hacerse referencia también a una molécula de ácido nucleico de otro organismo o especie (*es decir*, exógena).

25 Cualquier ácido nucleico, tal como ADN, que un especialista en la materia reconocería o consideraría como heterólogo o extraño para la célula en que se expresa el ácido nucleico, está englobado en la presente memoria por ácido nucleico heterólogo; el ácido nucleico heterólogo incluye ácido nucleico añadido exógenamente que se expresa también endógenamente. Los ejemplos de ácido nucleico heterólogo incluyen, pero sin limitación, ácido nucleico que codifica proteínas marcadoras trazables, tales como una proteína que confiere resistencia a fármacos, un ácido nucleico que codifica sustancias terapéuticamente efectivas tales como agentes anticancerosos, enzimas y hormonas, y un ácido nucleico, tal como ADN, que codifica otros tipos de proteínas tales como anticuerpos. Los anticuerpos que se codifican por ácido nucleico heterólogo pueden secretarse o expresarse sobre la superficie de la célula en que se ha introducido el ácido nucleico heterólogo.

35 Como se usa en la presente memoria, un péptido hace referencia a un polipéptido que es de 2 a 40 aminoácidos de longitud.

Como se usa en la presente memoria, los aminoácidos que aparecen en las diversas secuencias de aminoácidos proporcionadas en la presente memoria se identifican de acuerdo con sus abreviaturas conocidas de tres letras o una letra (Tabla 1). Los nucleótidos que aparecen en diversos fragmentos de ácido nucleico se designan con las denominaciones de una letra estándares usadas rutinariamente en la materia.

Como se usa en la presente memoria, un "aminoácido" es un compuesto orgánico que contiene un grupo amino y un grupo ácido carboxílico. Un polipéptido contiene dos o mas aminoácidos. Con fines de la presente memoria, los aminoácidos incluyen los 20 aminoácidos de origen natural, aminoácidos no naturales y análogos aminoacídicos (*es decir*, aminoácidos donde el carbono  $\alpha$  tiene una cadena lateral).

Manteniendo la nomenclatura polipeptídica estándar descrita en J. Biol. Chem., 243: 3557-3559 (1968), y adoptada en el título 37 del C.F.R. apartados 1.821-1.822, se muestran las abreviaturas para los residuos aminoacídicos en la Tabla 1:

**Tabla 1- Tabla de correspondencia**

SÍMBOLO		
Una letra	Tres letras	AMINOÁCIDO
Y	Tyr	Tirosina
G	Gly	Glicina
F	Phe	Fenilalanina
M	Met	Metionina
A	Ala	Alanina
S	Ser	Serina
I	Ile	Isoleucina
L	Leu	Leucina

T	Thr	Treonina
V	Val	Valina
P	Pro	Prolina
K	Lys	Lisina
H	His	Histidina
Q	Gln	Glutamina
E	Glu	Ácido glutámico
Z	Glx	Glu y/o Gln
W	Trp	Triptófano
R	Arg	Arginina
D	Asp	Ácido aspártico
N	Asn	Asparagina
B	Asx	Asn y/o Asp
C	Cys	Cisteína
X	Xaa	Desconocido u otro

Todas las secuencias de residuos aminoacídicos representadas en la presente memoria por fórmulas tienen una orientación de izquierda a derecha en la dirección convencional de extremo amino a extremo carboxilo. Además, la frase "residuo aminoacídico" se define ampliamente para incluir los aminoácidos listados en la Tabla de correspondencia (Tabla 1) y aminoácidos modificados e inhabituales, tales como aquellos a los que se hace referencia en el título 37 del C.F.R. apartados 1.821-1.822.

Además, un guion en el inicio o el final de una secuencia de residuo aminoacídico indica un enlace peptídico con una secuencia adicional de uno o más residuos aminoacídicos, o un grupo aminoterminal tal como NH<sub>2</sub> o un grupo carboxiterminal tal como COOH.

Como se usa en la presente memoria, "aminoácidos de origen natural" hace referencia a los 20 L-aminoácidos que aparecen en polipéptidos.

Como se usa en la presente memoria, "aminoácido no natural" hace referencia a un compuesto orgánico que contiene un grupo amino y un grupo ácido carboxílico que no es uno de los aminoácidos de origen natural listados en la Tabla 1. Los aminoácidos de origen no natural incluyen por tanto, por ejemplo, aminoácidos o análogos de aminoácidos distintos de los 20 aminoácidos de origen natural e incluyen, pero sin limitación, los D-estereoisómeros de aminoácidos. Los aminoácidos no naturales ejemplares son conocidos por los especialistas en la materia y pueden incluirse en cualquiera de los polipéptidos de valenceno sintasa modificados proporcionados en la presente memoria.

Como se usa en la presente memoria, modificación es con referencia a una modificación de una secuencia de aminoácidos de un polipéptido o una secuencia de nucleótidos en una molécula de ácido nucleico, e incluye delecciones, inserciones y reemplazos de aminoácidos y nucleótidos, respectivamente. Con fines de la presente memoria, los reemplazos aminoacídicos (o sustituciones), delecciones y/o inserciones pueden realizarse en cualquiera de las valenceno sintasas proporcionadas en la presente memoria. Las modificaciones pueden realizarse realizando reemplazos aminoacídicos conservativos y también sustituciones aminoacídicas no conservativas. Por ejemplo, pueden realizarse reemplazos aminoacídicos que alteran deseable o ventajosamente las propiedades de la valenceno sintasa. Por ejemplo, pueden realizarse reemplazos aminoacídicos en la valenceno sintasa de tal modo que la valenceno sintasa modificada resultante pueda producir más valenceno a partir de FPP en comparación con una valenceno sintasa no modificada.

Los reemplazos o sustituciones aminoacídicos contemplados incluyen sustituciones conservativas incluyendo, pero sin limitación, aquellas expuestas en la Tabla 2. Las sustituciones conservativas adecuadas de aminoácidos son conocidas por los especialistas en la materia y pueden realizarse en general sin alterar la conformación o actividad del polipéptido. Los especialistas en esta materia reconocen que, en general, las sustituciones aminoacídicas individuales en regiones no esenciales de un polipéptido no alteran sustancialmente la actividad biológica (véase, *p.ej.*, Watson y col. *Molecular Biology of the Gene*, 4ª edición, 1987, The Benjamin/Cummings Pub. co., pág. 224). Las sustituciones aminoacídicas conservativas se realizan, por ejemplo, de acuerdo con las expuestas en la Tabla 2 como sigue:

TABLA 2

Residuo original	Sustitución conservativa
Ala (A)	Gly; Ser; Abu
Arg (R)	Lys; Orn

Asn (N)	Gln; His
Cys (C)	Ser
Gln (Q)	Asn
Glu (E)	Asp
Gly (G)	Ala; Pro
His (H)	Asn; Gln
Ile (I)	Leu; Val
Leu (L)	Ile; Val
Lys (K)	Arg; Gln; Glu
Met (M)	Leu; Tyr; Ile
Ornitina	Lys; Arg
Phe (F)	Met; Leu; Tyr
Ser (S)	Thr
Thr (T)	Ser
Trp (W)	Tyr
Tyr (Y)	Trp; Phe
Val (V)	Ile; Leu; Met

Son también permisibles otras sustituciones conservativas y pueden determinarse empíricamente o de acuerdo con sustituciones conservativas conocidas. Los efectos de tales sustituciones pueden calcularse usando matrices de puntuación de sustitución tales como PAM120, PAM-200 y PAM-250 como se discute en Altschul (J. Mol. Biol. 219: 5 555-565 (1991)).

Como se usa en la presente memoria, "secuencia primaria" hace referencia a la secuencia de residuos aminoácidos en un polipéptido.

- 10 Como se usa en la presente memoria, "similitud" entre dos proteínas o ácidos nucleicos hace referencia al parentesco entre la secuencia de aminoácidos de las proteínas o las secuencias nucleotídicas de los ácidos nucleicos. La similitud puede estar basada en el grado de identidad y/u homología de secuencias de residuos y los residuos contenidos en las mismas. Los procedimientos para valorar el grado de similitud entre proteínas o ácidos nucleicos son conocidos por los especialistas en la materia. Por ejemplo, en un procedimiento para valorar la similitud de secuencia, se alinean dos secuencias aminoácidas o nucleotídicas de manera que procure el máximo nivel de identidad entre las secuencias. "Identidad" hace referencia a la extensión en que las secuencias aminoácidas o nucleotídicas son invariables. El alineamiento de secuencias aminoácidas, y en cierta medida de secuencias nucleotídicas, puede tener en cuenta también diferencias conservativas y/o sustituciones frecuentes en aminoácidos (o nucleótidos). Las diferencias conservativas son aquellas que conservan las propiedades físicoquímicas de los residuos implicados. Los alineamientos pueden ser globales (alineamiento de las secuencias comparadas por toda la longitud de las secuencias e incluyendo todos los residuos) o locales (el alineamiento de una porción de las secuencias que incluye solo la región o regiones más similares).

- 25 Como se usa en la presente memoria, los términos "homología" e "identidad" se usan para describir el parentesco entre polipéptidos (o moléculas de ácido nucleico codificantes). Identidad hace referencia a secuencias idénticas; homología puede incluir cambios aminoácidos conservativos. En general, para identificar las posiciones correspondientes, se alinean las secuencias aminoácidas de modo que se obtenga el máximo orden de apareamiento (véanse, *p.ej.* Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, Nueva York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, Nueva York, 1993; 30 Computer Analysis of Sequence Data, Parte I, Griffin, A.M. y Griffin, H.G., eds., Humana Press, Nueva Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. y Devereux, J., eds., M Stockton Press, Nueva York, 1991; Carrillo y col. (1988) SIAM J Applied Math 48: 1073).

- 35 Como se usa en la presente memoria, "identidad de secuencia" hace referencia al número de aminoácidos o bases nucleotídicas idénticos o similares en una comparación entre un polipéptido o polinucleótido de prueba y de referencia. La identidad de secuencia puede determinarse mediante alineamiento de secuencia de las secuencias de ácido nucleico o proteína para identificar regiones de similitud o identidad. Con fines de la presente memoria, se determina generalmente la identidad de secuencia por alineamiento para identificar residuos idénticos. El 40 alineamiento puede ser local o global, pero con fines de la presente memoria es generalmente un alineamiento global donde se compara toda la longitud de cada secuencia. Pueden identificarse los apareamientos, desapareamientos y huecos entre secuencias comparadas. Los huecos son aminoácidos o nucleótidos nulos insertados entre los residuos de secuencias alineadas de modo que se alineen caracteres idénticos o similares. Generalmente, puede haber huecos internos y terminales. La identidad de secuencia puede determinarse teniendo

en cuenta los huecos como el número de residuos idénticos/longitud de la secuencia más corta x 100. Cuando se usan penalizaciones de hueco, puede determinarse la identidad de secuencia sin penalización para huecos terminales (*p.ej.*, los huecos terminales no se penalizan). Como alternativa, puede determinarse la identidad de secuencia sin tener en cuenta los huecos como el número de posiciones idénticas/longitud de la secuencia alineada total x 100.

Como se usa en la presente memoria, un "alineamiento global" es un alineamiento que alinea dos secuencias desde el inicio al final, alineando cada letra en cada secuencia solo una vez. Se produce un alineamiento independientemente de si hay similitud o identidad o no entre las secuencias. Por ejemplo, una identidad de 10 secuencias de 50 % en "alineamiento global" significa que, en un alineamiento de toda la secuencia de dos secuencias comparadas, cada una de 100 nucleótidos de longitud, el 50 % de los residuos son los mismos. Se entiende que puede usarse también el alineamiento global en la determinación de la identidad de secuencia incluso cuando la longitud de las secuencias alineadas no sea la misma. Se tendrán en cuenta las diferencias en los extremos terminales de las secuencias en la determinación de la identidad de secuencia, a menos que se seleccione "sin penalización para huecos terminales". Generalmente, se usa un alineamiento global en secuencias que comparten una similitud significativa por la mayoría de su longitud. Los algoritmos ejemplares para practicar un alineamiento global incluyen el algoritmo de Needleman-Wunsch (Needleman y col. *J. Mol. Biol.* 48: 443 (1970). Los programas ejemplares para practicar alineamientos globales están públicamente disponibles e incluyen la Herramienta de Alineamiento de Secuencia Global disponible en el sitio web del Centro Nacional para Información Biotecnológica (NCBI) ([ncbi.nlm.nih.gov/](http://ncbi.nlm.nih.gov/)), y el programa está disponible en [deepc2.psi.iastate.edu/aat/align/align.html](http://deepc2.psi.iastate.edu/aat/align/align.html).

Como se usa en la presente memoria, un "alineamiento local" es un alineamiento que alinea dos secuencias, pero solo alinea aquellas porciones de las secuencias que comparten similitud o identidad. Por ello, un alineamiento local determina si están presentes subsegmentos de una secuencia en otra secuencia. Si no hay similitud, no se devolverá alineamiento. Los algoritmos de alineamiento local incluyen BLAST o el algoritmo de Smith-Waterman (*Adv. Math.* 2: 482 (1981)). Por ejemplo, un 50 % de identidad de secuencia basada en "alineamiento local" significa que, en un alineamiento de la secuencia completa de dos secuencias comparadas de cualquier longitud, una región de similitud o identidad de 100 nucleótidos de longitud tiene un 50 % de los residuos que son iguales en la región de similitud o identidad.

Con fines de la presente memoria, puede determinarse la identidad de secuencia mediante programas de algoritmo de alineamiento estándares usados con penalizaciones por hueco por defecto establecidas por cada suministrador. Los parámetros por defecto para el programa GAP pueden incluir: (1) una matriz de comparación unaria (que contiene un valor de 1 para identidades y de 0 para no identidades) y la matriz de comparación ponderada de Gribskov y col. *Nucl. Acids Res.* 14: 6745 (1986), como se describe por Schwartz y Dayhoff, eds., *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, pág. 353-358 (1979); (2) una penalización de 3,0 por cada hueco y una penalización adicional de 0,10 por cada símbolo en cada hueco y (3) sin penalización por huecos terminales. Puede determinarse si dos moléculas de ácido nucleico cualesquiera tienen secuencias nucleotídicas, o dos polipéptidos cualesquiera tienen secuencias aminoacídicas, que son al menos un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % "idénticas" u otras variaciones similares que enumeran una identidad porcentual, usando algoritmos informáticos conocidos basados en alineamiento local o global (véase *p.ej.* [wikipedia.org/wiki/Sequence\\_alignment\\_software](http://wikipedia.org/wiki/Sequence_alignment_software), que proporciona enlaces a docenas de bases de datos y programas conocidos y públicamente disponibles). Generalmente, con fines de la presente memoria, se determina la identidad de secuencia usando algoritmos informáticos basados en alineamiento global, tale como la herramienta de alineamiento de secuencia global de Needleman-Wunsch, disponible en NCBI/BLAST ([blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&Page\\_TYPE=BlastHome](http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&Page_TYPE=BlastHome)); LAlign (implementación por William Pearson del algoritmo de Huang y Miller (*Adv. Math.* (1991) 12: 337-357)); y el programa de Xiaoqui Huang disponible en [deepc2.psi.iastate.edu/aat/align/align.html](http://deepc2.psi.iastate.edu/aat/align/align.html). Generalmente, cuando se comparan las secuencias nucleotídicas de la presente memoria, se usa un alineamiento con penalización por huecos terminales. El alineamiento local puede usarse también cuando las secuencias que se están comparando son sustancialmente de la misma longitud.

Por lo tanto, como se usa en la presente memoria, el término "identidad" representa una comparación o alineamiento entre un polipéptido o polinucleótido de prueba y de referencia. En un ejemplo no limitante, "al menos un 90 % idéntico a" hace referencia a identidades porcentuales de 90 a 100 % respecto al polipéptido o polinucleótido de referencia. La identidad a un nivel del 90 % o más es indicativa del hecho de que, suponiendo con fines de ejemplificación que se compara una longitud de polipéptido o polipéptido de prueba y de referencia de 100 aminoácidos o nucleótidos, no más del 10 % (*es decir*, 10 de 100) de los aminoácidos o nucleótidos en el polipéptido o polinucleótido de prueba difiere de los de los polipéptidos de referencia. Pueden hacerse comparaciones similares entre un polinucleótido de prueba y de referencia. Pueden representarse tales diferencias como mutaciones puntuales distribuidas aleatoriamente por toda la longitud de una secuencia aminoacídica o pueden agruparse en

una o más localizaciones de longitud variable de hasta el máximo permisible, p.ej., diferencia aminoacídica de 10/100 (aproximadamente 90 % de identidad). Las diferencias pueden ser debidas también a deleciones o truncamientos de residuos aminoacídicos. Las diferencias se definen como sustituciones, inserciones o deleciones de ácido nucleico o aminoácido. Dependiendo de la longitud de las secuencias comparadas, al nivel de homologías o identidades por encima de aproximadamente un 85-90 %, el resultado puede ser independiente de los programas y parámetros de hueco fijados; tales altos niveles de identidad pueden valorarse fácilmente, a menudo sin confiar en software.

Como se usa en la presente memoria, se entiende también que los términos "sustancialmente idéntico" o "similar" varían con el contexto como se entiende por los especialistas en la materia relevante, pero que los especialistas pueden valorarlos.

Como se usa en la presente memoria, una secuencia alineada hace referencia al uso de homología (similitud y/o identidad) para alinear las posiciones correspondientes en una secuencia de nucleótidos o aminoácidos. Típicamente, se alinean dos o más secuencias que están relacionadas con un 50 % o más de identidad. Un conjunto alineado de secuencias hace referencia a 2 o más secuencias que están alineadas en las posiciones correspondientes y que pueden incluir alinear secuencias derivadas de ARN, tales como EST y otros ADNc, alineadas con la secuencia de ADN genómica.

Como se usa en la presente memoria, el polipéptido o proteína aislado o purificado o porción biológicamente activa del mismo está sustancialmente exento de material celular u otras proteínas contaminantes de la célula de tejido de la que deriva la proteína, o sustancialmente exento de precursores químicos u otros productos químicos cuando se sintetiza químicamente. Puede determinarse que las preparaciones son sustancialmente exentas si parecen exentas de impurezas fácilmente detectables como se determina por procedimientos estándares de análisis, tales como cromatografía en capa fina (TLC), electroforesis en gel y cromatografía líquida de alta resolución (HPLC), usados por los especialistas en la materia para valorar tal pureza, o suficientemente puras de tal modo que una purificación adicional no alteraría detectablemente las propiedades físicas y químicas, tales como actividades proteolíticas y biológicas, de la sustancia. Los procedimientos para la purificación de los compuestos para producir compuestos sustancialmente químicamente puros son conocidos por los especialistas en la materia. Sin embargo, un compuesto sustancialmente químicamente puro puede ser una mezcla de estereoisómeros. En tales aspectos, una purificación adicional podría aumentar la actividad específica del compuesto.

El término sustancialmente exento de material celular incluye preparaciones de valenceno sintasa y productos de terpeno en que la valenceno sintasa o producto de terpeno se separa de los componentes celulares de las células de las que se aísla o produce. En una realización, el término sustancialmente exento de material celular incluye preparaciones de valenceno sintasa o productos de terpeno que tienen menos de aproximadamente un 30 %, 20 %, 10 %, 5 % o menos (en peso seco) de proteínas o productos no de valenceno sintasa o terpeno, incluyendo medio de cultivo celular.

Como se usa en la presente memoria, la producción por procedimientos recombinantes usando procedimientos de ADN recombinante hace referencia al uso de los procedimientos bien conocidos de biología molecular para expresar proteínas codificadas por ADN clonado.

Como se usa en la presente memoria, vector (o plásmido) hace referencia a elementos de ADN discretos que se usan para introducir ácido nucleico heterólogo en células para la expresión o replicación del mismo. Los vectores permanecen típicamente episómicos, pero pueden diseñarse para efectuar la integración de un gen o porción del mismo en un cromosoma del genoma. Se contemplan también vectores que son cromosomas artificiales, tales como cromosomas artificiales bacterianos, cromosomas artificiales de levadura y cromosomas artificiales de mamífero. La selección y uso de tales vehículos son bien conocidos para los especialistas en la materia.

Como se usa en la presente memoria, expresión hace referencia al proceso mediante el cual se transcribe el ácido nucleico a ARNm y se traduce a péptidos, polipéptidos o proteínas. Si el ácido nucleico deriva de ADN genómico, la expresión puede incluir, si se selecciona una célula u organismo hospedador eucariótico apropiado, procesamiento tal como empalme del ARNm.

Como se usa en la presente memoria, un vector de expresión incluye vectores capaces de expresar ADN que está conectado operativamente con secuencias reguladoras, tales como regiones promotoras, que son capaces de efectuar la expresión de tales fragmentos de ADN. Tales segmentos adicionales pueden incluir secuencias promotoras y terminadoras, y opcionalmente pueden incluir uno o más orígenes de replicación, uno o más marcadores seleccionables, un potenciador, una señal de poliadenilación y similares. Los vectores de expresión derivan generalmente de ADN de plásmido o vírico, o pueden contener elementos de ambos. Por tanto, un vector de

expresión hace referencia a un constructo de ADN o ARN recombinante, tal como un plásmido, un fago, virus recombinante u otro vector que, tras la introducción en una célula hospedadora apropiada, da como resultado la expresión del ADN clonado. Los vectores de expresión apropiados son bien conocidos por los especialistas en la materia e incluyen aquellos que son replicables en células eucarióticas y/o procarióticas y aquellos que permanecen episómicos o aquellos que se integran en el genoma de la célula hospedadora.

Como se usa en la presente memoria, vector incluye también "vectores de virus" o "vectores víricos". Los vectores víricos son virus genomanipulados que están conectados operativamente con genes exógenos para transferir (como vehículos o lanzaderas) los genes exógenos a células.

Como se usa en la presente memoria, un adenovirus hace referencia a cualquiera de un grupo de virus que contienen ADN que causa conjuntivitis e infecciones del tracto respiratorio superior en seres humanos.

Como se usa en la presente memoria, ADN desnudo hace referencia a ADN exento de histona que puede usarse para vacunas y terapia génica. El ADN desnudo es el material genético que se pasa de célula a célula durante un proceso de transferencia génica llamado transformación o transfección. En la transformación o transfección, el ADN purificado o desnudo que se capta por la célula receptora dará a la célula receptora una nueva característica o fenotipo.

Como se usa en la presente memoria, conectado operativamente cuando se hace referencia a segmentos de ADN significa que los segmentos se organizan de modo que funcionen en sintonía con sus fines pretendidos, *p.ej.*, la transcripción se inicia en el promotor y prosigue a lo largo del segmento codificante hasta el terminador.

Como se usa en la presente memoria, una "proteína quimérica" o "proteína de fusión" hace referencia a un polipéptido conectado operativamente con un polipéptido diferente. Una proteína quimérica o de fusión proporcionada en la presente memoria puede incluir uno o más polipéptidos de valenceno sintasa, o una porción de los mismos, y uno o más polipéptidos distintos para una cualquiera o más de señales de control transcripcional/traduccionales, secuencias señal, marcaje para localización, marcaje para purificación, parte de un dominio de una inmunoglobulina G y/o un agente de orientación. Un polipéptido de valenceno sintasa quimérico incluye también aquellos que tienen sus dominios o regiones endógenos del polipéptido intercambiados con otro polipéptido. Estas proteínas quiméricas o de fusión incluyen aquellas producidas por medios recombinantes como proteínas de fusión, aquellas producidas por medios químicos tales como por acoplamiento químico mediante, por ejemplo, acoplamiento con grupos sulfhidrilo, y aquellas producidas por cualquier otro procedimiento mediante el que se conecta al menos un polipéptido (*es decir*, valenceno sintasa) o una porción del mismo directa o indirectamente mediante un conector o conectores con otro polipéptido.

Como se usa en la presente memoria, la enumeración de que un polipéptido "consiste esencialmente en" una secuencia enumerada de aminoácidos significa que solo está presente la secuencia enumerada, o un fragmento de la misma, del polipéptido de longitud completa. El polipéptido incluirá opcional y generalmente aminoácidos adicionales de otra fuente o puede insertarse en otro polipéptido.

Como se usa en la presente memoria, las formas singulares "un", "una" y "el/la" incluyen los referentes plurales a menos que el contexto dicte claramente otra cosa. Por tanto, por ejemplo, la referencia a un polipéptido que comprende "un reemplazo aminoacídico" incluye polipéptidos con uno o una pluralidad de reemplazos aminoacídicos.

Como se usa en la presente memoria, los intervalos y cantidades pueden expresarse como "aproximadamente" un valor o intervalo particular. Aproximadamente incluye también la cantidad exacta. Por ello, "aproximadamente 5 %" significa "aproximadamente 5%" y también "5 %".

Como se usa en la presente memoria, "opcional" u "opcionalmente" significa que el evento o circunstancia descrito a continuación ocurre o no, y que la descripción incluye aspectos en que dicho evento o circunstancia ocurre y aspectos en que no. Por ejemplo, una etapa opcional de aislamiento de valenceno significa que el valenceno se aísla o no se aísla.

Como se usa en la presente memoria, las abreviaturas para cualquier grupo protector, aminoácidos y otros compuestos están, a menos que se indique otra cosa, de acuerdo con su uso común, las abreviaturas reconocidas o la Comisión sobre Nomenclatura Bioquímica de la IUPAC-IUB (véase, (1972) *Biochem.* 11: 1726).

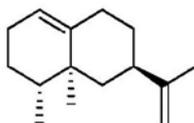
## 60 B. VISIÓN GENERAL

Se proporcionan en la presente memoria valenceno sintasas de *Eryngium glaciale*, y variantes y fragmentos catalíticamente activos de las mismas, que catalizan la producción de terpenos a partir de un precursor terpénico de pirofosfato acíclico. Los terpenos incluyen valenceno y opcionalmente otros sesquiterpenos. Tales valenceno sintasas catalizan la producción biosintética de valenceno a partir de un precursor de pirofosfato acíclico, tal como pirofosfato de farnesilo. Se proporcionan también en la presente memoria procedimientos para producir valenceno y otros sesquiterpenos a partir de tal precursor tales como, pero sin limitación, pirofosfato de farnesilo. Se proporcionan también en la presente memoria procedimientos para elaborar nootkatona a partir del valenceno resultante y a partir del precursor acíclico, tal como pirofosfato de farnesilo. Las valenceno sintasas de *Eryngium glaciale* (EgVS) proporcionadas proporcionan la producción de productos terpénicos valiosos, incluyendo valenceno, en cantidades comercialmente útiles y de manera económica y energéticamente eficiente. En particular, las EgVS catalizan la producción de muy altos niveles de valenceno en comparación con otras valenceno sintasas, incluyendo CVS.

### 1. Estructura del valenceno y usos

15

El valenceno (1,2,3,5,6,7,8,8a-octahidro-7-isopropenil-1,8a-dimetilnaftaleno (1) es un sesquiterpeno encontrado en aceites cítricos, tales como de naranja y uva. Hasta la fecha, se ha identificado valenceno en diversas plantas incluyendo frutas cítricas (*Citrus* sp.), flores de parra (*Vitis vinifera*), apio (*Apium graveolens*), mango (*Mangifera indica*), aceitunas (*Olea europea*) y coral. El valenceno se usa como aroma/fragancia de naranja en perfumes, bebidas y gomas de mascar, y se usa como material de partida para la producción de nootkatona.



Valenceno (1)

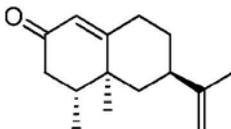
El valenceno se genera en plantas por la terpeno sintasa valenceno sintasa, que cataliza la reacción del precursor terpénico de pirofosfato acíclico pirofosfato de farnesilo (FPP) a valenceno (véase la Figura 1).

25

### 2. Nootkatona

La nootkatona (4,4a,5,6,7,8-hexahidro-6-isopropenil-4,4-a-dimetil-2(3H)-naftalenona (2) es un sesquiterpenoide encontrado en aceite de uva que proporciona el aroma a uva dominante. Típicamente, la nootkatona se aísla por extracción de la uva. La nootkatona se ha identificado también en diversas otras plantas incluyendo, por ejemplo, falso ciprés de Nootka (*Cupressus nootkatensis*), vetiver (*Chrysopogon zizanioides*) y *Alpinia oxyphylla* Miquel.

30



Nootkatona (2)

35

La nootkatona es un producto oxidado del valenceno. El valenceno puede experimentar hidroxilación regioselectiva formando 2-hidroxisvalenceno, que se oxida adicionalmente formando nootkatona (como se describe con más detalle en la sección E siguiente). La nootkatona es apreciada por su sabor y aroma a uva y se usa ampliamente en las industrias de perfumería y aromas. Además, se ha mostrado que la nootkatona es un repelente e insecticida efectivo contra garrapatas y mosquitos.

40

### 3. Valenceno sintasas

Las valenceno sintasas son terpeno ciclasas vegetales de clase 1, o terpeno sintasas, isoprenoide sintasas o terpenoide ciclasas, que convierten el difosfato de farnesilo en el sesquiterpeno valenceno. Hasta la fecha, las valenceno sintasas se han aislado de frutos cítricos, flores de parra y Perilla (shiso verde). La valenceno sintasa cítrica (CVS) se ha identificado en flavedo (piel externa) de *Citrus sinensis* (naranja dulce; naranja de Valencia) (SEQ

45

ID NO:14 y 34) y *Citrus x paradisi*(uva) (SEQ ID NO:15, 26 y 27) (véanse Chappell (2004) Trends Plant Sci., 9: 266; Sharon-Asa y col., (2003) The Plant Journal 36: 664-674; documento AF411120 y patentes de EE.UU. nº 7.273.735; 7.442.785; 7.790.426 y las solicitudes PCT internacionales nº WO2005021705 y WO2003025193). Se ha descrito una variante de valenceno sintasa que contiene los reemplazos aminoácidos A5171/1518V (Eyal, E. Master's Thesis, Department of Plant Sciences, Weizmann Institute of Science, Rehovot, Israel; enero de 2001; expuesta en la SEQ ID NO:37). Se han descrito una variedad de valenceno sintasas cítricas modificadas en la solicitud de patente de EE.UU. nº 2012-0246767. También se han identificado y aislado valenceno sintasas de flores de parra, incluyendo *Vitis vinifera* L. cv. Gewürztraminer y *Vitis vinifera* L. cv. Cabernet Sauvignon (véanse Lucker y col., (2004) Photochemistry 65(19): 2649-2659 y Martin y col., (2009) Proc. Natl. Acad. Sci, USA 106: 7245-7250) (SEQ ID NO:16 y 28). También se han aislado valenceno sintasas de *Chamaecyparis nootkatensis pendula* (véanse p.ej. las solicitudes PCT internacionales nº WO2011074954 y WO2012177129; SEQ ID NO: 17 y 29). La sintasa EgVS proporcionada en la presente memoria se muestra en la presente memoria que cataliza la producción de altos niveles de valenceno en comparación con otras valenceno sintasas, incluyendo otras cuyas secuencias se han modificado para producir niveles aumentados de valenceno.

15

#### a. Estructura

Las terpeno ciclasas vegetales de clase 1 incluyen un grupo diverso de terpeno sintasas monoméricas que comparten una arquitectura de hélice alfa común denominada pliegue de terpenoide ciclasa de clase 1 (véanse, p.ej., Christianson, D.W., (2008) Curr Opin Chem Biol 12(2): 141-150 y Bohlmann y col., (1998) Proc. Natl. Acad. Sci. USA 95: 4126-4133). Aunque existe una similitud de secuencia global relativamente baja, las terpeno ciclasas vegetales de clase 1 tienen estructuras homólogas y algunos motivos y/o residuos altamente conservados. En su sitio catalítico, cada terpeno ciclasa proporciona un molde que se une al sustrato isoprenoide flexible con una orientación y conformación tales que, tras ciclación, se forma un enlace carbono-carbono intramolecular. Por tanto, la estructura de cada sitio catalítico de la enzima dicta los monoterpenos, diterpenos y sesquiterpenos cíclicos resultantes.

Las estructuras de cristal de rayos X de 5-epiaristologueno sintasa de tabaco y pentaleneno sintasa revelaron que las terpeno ciclasas vegetales de clase 1 están compuestas por hélices alfa interconectadas por bucles y giros conectores cortos (véanse, p.ej., Starks y col., (1997) Science 277: 1815-1820 y Lesburg y col., (1997) Science 277: 1820-1824). Estas enzimas contienen dos dominios estructurales distintos, un dominio N-terminal, cuya estructura se asemeja a los núcleos catalíticos de las glicosilo hidrolasas, pero cuya función permanece desconocida en gran medida, y un dominio catalítico C-terminal. El dominio catalítico contiene dos motivos de unión a metal conservados, es decir regiones ricas en aspartato, que son responsables de la actividad catalítica de la enzima. El sitio catalítico contiene una gran cavidad central formada mayoritariamente por hélices alfa antiparalelas con las dos regiones ricas en aspartato localizadas en paredes opuestas. Las regiones ricas en aspartato median la unión de difosfatos sustrato mediante puentes con iones  $Mg^{2+}$ . La unión posterior del sustrato induce cambios conformacionales tales que la región N-terminal forma una caperuza sobre el núcleo catalítico que cierra el sitio activo al disolvente, estabilizando así los intermedios carbocatiónicos reactivos.

40

Las hélices alfa conservadas C, D, F, G y H constituyen el sitio catalítico o activo de las terpeno sintasas vegetales de clase 1. El sitio activo es un bolsillo hidrófobo revestido por residuos aromáticos para acomodar la cadena olefínica del sustrato. Los residuos aromáticos estabilizan los intermedios carbocatiónicos mediante interacciones catión- $\pi$ . Una región rica en aspartato 1 está localizada en la hélice D y se caracteriza por la secuencia conservada DDxxD (SEQ ID NO:35), que funciona también uniendo  $Mg^{2+}$  (véase, p.ej., Starks y col., (1997), Science 277: 1815-1820). Una segunda región de unión a metal conservada está localizada en la hélice H y se caracteriza por la secuencia conservada [N/D]xxx[S/T]xxxE (SEQ ID NO:36), a la que se hace referencia también como "motivo NSE/DTE". Estos dos motivos de unión a metal conservados coordinan la unión de tres iones  $Mg^{2+}$  al difosfato de isoprenoide.

50

La valenceno sintasa de *Eryngium glaciale* proporcionada en la presente memoria contiene un dominio N-terminal (aa 1-280 de SEQ ID NO:1) y un dominio catalítico C-terminal (aa 281-565 de SEQ ID NO:1). Dentro del dominio catalítico C-terminal está el sitio de unión a metal conservado que contiene las regiones ricas en aspartato 1 y 2. Basándose en el alineamiento con valenceno sintasas conocidas, tales como valenceno sintasa cítrica, la región rica en aspartato 1 que contiene el motivo DDxxD conservado corresponde a los aminoácidos D315, D316, T317, Y318 y D319 de SEQ ID NO:1. Asp315 y Asp319 se unen a restos difosfato de FPP mediante coordinación con  $Mg^{2+}$ . La región rica en aspartato 2, que contiene el motivo NSE/DTE, corresponde a los aminoácidos D461, D462, I463, G464, G465, H466, E467, F468 y E469 de SEQ ID NO:1. Esta región se une a un ión  $Mg^{2+}$  adicional mediante los aminoácidos Asp461, Gly465 y Glu469.

60

#### b. Actividades

La valenceno sintasa cataliza la formación de valenceno a partir de precursores de pirofosfato acíclicos, tales como el intermedio pirofosfato ubicuo difosfato de farnesilo (FPP), que se produce como parte de la ruta biosintética de isoprenoide dependiente de mevalonato en hongos y animales y la ruta biosintética de isoprenoide no dependiente de mevalonato en bacterias y plantas superiores. Los productos terpénicos adicionales que pueden producirse por valenceno sintasa a partir de precursores terpénicos de pirofosfato acíclico tales como FPP incluyen, pero sin limitación, germacreno A, beta-elemeno (el beta-elemeno se forma por descomposición espontánea de germacreno A),  $\beta$ -selineno,  $\tau$ -selineno, 7-*epi*- $\alpha$ -selineno y un compuesto adicional, aristoloqueno (véase, *p.ej.*, el pico 2 en la Figura 3).

10

En general, las terpeno ciclasas vegetales de clase 1, tales como valenceno sintasa, son ciclasas dependientes de metal que convierten difosfatos de isoprenoide todo-trans lineales, tales como difosfato de geranilo, difosfato de farnesilo y difosfato de geranilgeranilo, en monoterpenos, diterpenos y sesquiterpenos cíclicos. Las reacciones de ciclación proceden mediante alquilación electrófila en que se forman nuevos enlaces simples carbono-carbono mediante reacción de un carbocatión alílico deficiente en electrones altamente reactivo y un doble enlace carbono-carbono rico en electrones.

Las terpeno sintasas contienen iones metálicos divalentes, típicamente iones  $Mg^{2+}$  o a veces  $Mn^{2+}$ , en el centro activo de la enzima que son requeridos para catálisis enzimática. Más específicamente, son requeridos para la apertura de pirofosfato. Generalmente, las enzimas contienen dos motivos de unión a metal conservados que revisten el sitio catalítico, incluyendo el motivo DDxxD rico en aspartato (SEQ ID NO:35) que coordina la unión de dos iones  $Mg^{2+}$  y el motivo NSE/DTE (SEQ ID NO:36) que coordina un tercer ión  $Mg^{2+}$  (véanse Starks y col., (1997), *Science* 277: 1815-1820 y Lesburg y col. (1997), *Science* 277: 1820-1824). Las regiones ricas en aspartato del sitio activo catalítico median la unión de difosfatos de prenilo a través de puentes de iones  $Mg^{2+}$ . La unión de  $(Mg^{2+})_3$ -PP<sub>i</sub> induce cambios conformacionales tales que la región N-terminal forma una caperuza sobre el núcleo catalítico y por lo tanto estabiliza el sitio activo en una conformación cerrada que está exenta de disolvente bruto. La pérdida de pirofosfato (PP<sub>i</sub>) del sustrato unido a enzima da como resultado un carbocatión alílico altamente reactivo que ataca electrófilamente un doble enlace intramolecular más abajo de la cadena de terpeno efectuando el cierre de anillo. El anión PP<sub>i</sub> acepta enlaces de hidrógeno de residuos básicos conservados cuando se unen en conformación de sintasa cerrada y un bolsillo hidrófobo revestido por residuos aromáticos envuelve la cadena lateral de prenilo y probablemente hace de molde de la reacción de ciclación al reforzar conformaciones de sustrato particulares y estabilizar carbocationes mediante interacciones de apilamiento  $\pi$  (Noel y col., (2010) *ACS Chemical Biology* 5(4): 377-392).

#### 35 4. Ensayos para detectar la actividad enzimática de polipéptidos de valenceno sintasa

Un especialista en la materia está familiarizado con procedimientos y ensayos para detectar la actividad enzimática de polipéptidos de valenceno sintasa. Los polipéptidos de valenceno sintasa pueden sobreexpresarse y purificarse como se describe en la Sección C siguiente. Típicamente, se determina la actividad de una valenceno sintasa mediante incubación de una valenceno sintasa con un precursor terpénico de pirofosfato acíclico, tal como pirofosfato de farnesilo (FPP), e identificando, midiendo y/o cuantificando el valenceno y otros productos de reacción.

Por ejemplo, la actividad valenceno sintasa puede determinarse *in vitro* mediante incubación de una valenceno sintasa con un precursor terpénico de pirofosfato acíclico, tal como FPP, e identificando los productos de reacción. Los productos de reacción, incluyendo relaciones de los productos, pueden determinarse mediante cualquier procedimiento conocido por el especialista en la materia, incluyendo cromatografía de gases-espectrometría de masas (GC-MS), GC-FID, cromatografía líquida-espectrometría de masas (LC-MS), comparación con patrones conocidos y resonancia magnética nuclear de protón y carbono (RMN).

Como alternativa, la actividad valenceno sintasa puede determinarse *in vivo* mediante expresión de una valenceno sintasa en una cepa de levadura que produce un precursor terpénico de pirofosfato acíclico, tal como FPP, con lo que la expresión de la valenceno sintasa da como resultado la producción de valenceno y compuestos/productos de reacción o subproductos adicionales. El valenceno y compuestos adicionales pueden purificarse del medio de cultivo celular, por ejemplo, por extracción con un disolvente orgánico, con lo que el valenceno y otros productos se reparten en la fase acuosa como se describe en el Ejemplo 4, y los productos de reacción pueden identificarse y cuantificarse como se describe anteriormente. Las células de levadura ejemplares para expresión de valenceno sintasas y los procedimientos para la generación de células de levadura modificadas que producen un precursor terpénico de pirofosfato acíclico se describen con más detalle en la Sección E siguiente.

60 La cinética de la producción de valenceno puede determinarse mediante procedimientos estándares, tales como por ensayos de sintasa en que se emplean sustratos isoprenoides radiactivos, tales como  $^3H$  FPP o  $^{14}C$  FPP, con

concentraciones variables de sintasa. Se extraen los productos de la reacción en una fase orgánica y se mide la radiactividad usando un contador de centelleo líquido. Se determinan las constantes cinéticas a partir de ajustes directos de la ecuación de Michaelis-Menton a los datos.

### 5 C. MOLÉCULAS DE ÁCIDO NUCLEICO QUE CODIFICAN VALENCENO SINTASA DE *ERYNGIUM GLACIALE* (EGVS) Y POLIPÉPTIDOS CODIFICADOS

Se proporcionan en la presente memoria moléculas de ácido nucleico que codifican un polipéptido de valenceno sintasa, incluyendo moléculas de ADNc. Se proporcionan también en la presente memoria polipéptidos de valenceno sintasa, y fragmentos catalíticamente activos de los mismos, codificados por las moléculas de ácido nucleico proporcionadas en la presente memoria. Las valenceno sintasas codificadas por las moléculas de ácido nucleico proporcionadas en la presente memoria catalizan la formación de valenceno a partir de cualquier precursor terpenico de pirofosfato acíclico adecuado, incluyendo pirofosfato de farnesilo (FPP), pirofosfato de geranilo (GPP) y pirofosfato de geranilgeranilo (GGPP). Es de interés en la presente memoria la producción de valenceno a partir de FPP. En algunos ejemplos, las moléculas de ácido nucleico proporcionadas en la presente memoria que codifican los polipéptidos de valenceno sintasa son aquellas que son iguales a las que se aíslan del cardo *Eryngium glaciale*. En otros ejemplos, las moléculas de ácido nucleico y polipéptidos de valenceno sintasa codificados proporcionados en la presente memoria son variantes de aquellas aisladas a partir del cardo *Eryngium glaciale*.

Por ejemplo, se proporciona en la presente memoria una molécula de ácido nucleico que tiene una secuencia de nucleótidos expuesta en la SEQ ID NO: 2, y degeneraciones de la misma, que codifica un polipéptido de valenceno sintasa que tiene una secuencia de aminoácidos expuesta en la SEQ ID NO: 1. Se proporcionan también en la presente memoria moléculas de ácido nucleico que tienen al menos un 85 % de identidad de secuencia con una secuencia de nucleótidos expuesta en la SEQ ID NO:2 que codifica un polipéptido de valenceno sintasa. Por ejemplo, las moléculas de ácido nucleico proporcionadas en la presente memoria pueden exhibir al menos o al menos aproximadamente un 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % o más identidad de secuencia con una secuencia de nucleótidos expuesta en la SEQ ID N:2, a condición de que los polipéptidos de valenceno sintasa codificados exhiban actividad valenceno sintasa (es decir, la capacidad de catalizar la formación de valenceno). Puede determinarse la identidad porcentual por un especialista en la materia usando programas de alineamiento estándares. Se proporcionan también en la presente memoria secuencias degeneradas de la secuencia nucleotídica expuesta en la SEQ ID NO: 2, que codifica una valenceno sintasa que tiene una secuencia de aminoácidos expuesta en la SEQ ID NO: 1. En algunos ejemplos, las moléculas de ácido nucleico que codifican los polipéptidos de valenceno sintasa se aíslan del cardo *Eryngium glaciale*. En otros ejemplos, las moléculas de ácido nucleico y polipéptidos de valenceno sintasa codificados son variantes de aquellas aisladas a partir del cardo *Eryngium glaciale*.

Se proporcionan también en la presente memoria moléculas de ácido nucleico que codifican un polipéptido de valenceno sintasa modificado proporcionado en la presente memoria. Las modificaciones pueden realizarse en cualquier región de un gen que codifica la valenceno sintasa a condición de que el polipéptido de valenceno sintasa modificado codificado resultante retenga al menos actividad valenceno sintasa (es decir, la capacidad de catalizar la formación de valenceno a partir de un precursor terpenico de pirofosfato acíclico, típicamente FPP). Las modificaciones pueden incluir optimización de codón de los ácidos nucleicos y/o cambios que dan como resultado una modificación aminoacídica única en el polipéptido de valenceno sintasa codificado, tal como reemplazos (sustituciones), inserciones o deleciones aminoacídicas únicas, o múltiples modificaciones aminoacídicas, tales como múltiples reemplazos, inserciones o deleciones aminoacídicas, incluyendo intercambios de regiones o dominios del polipéptido.

La expresión de las moléculas de ácido nucleico proporcionadas en la presente memoria en un hospedador adecuado, por ejemplo una célula bacteriana o de levadura, da como resultado la expresión de valenceno sintasa. Pueden usarse tales células para producir la valenceno sintasa y/o para practicar reacciones *in vivo* para producir valenceno. Por ejemplo, puede generarse valenceno en una célula de levadura a partir de FPP, particularmente una célula de levadura que sobreproduce el precursor terpenico de pirofosfato acíclico FPP.

En ejemplos particulares, las moléculas de ácido nucleico proporcionadas en la presente memoria pueden estar optimizadas de codón, por ejemplo, para aumentar los niveles de expresión de la secuencia codificada. El uso de codón particular depende del organismo hospedador en que se expresa el polipéptido modificado. Un especialista en la materia está familiarizado con los codones óptimos para expresión en bacterias o levadura incluyendo, por ejemplo, *E. coli* o *Saccharomyces cerevisiae*. Por ejemplo, la información del uso de codón está disponible en la Base de Datos de Uso de codón disponible en [kazusa.or.jp/codon](http://kazusa.or.jp/codon) (véase Richmond (2000) Genome Biology, 1: 241 para una descripción de la base de datos). Véanse también Forsburg (2004) Yeast, 10: 1045-1047; Brown y col. (1991) Nucleic Acids Research, 19: 4298; Sharp y col. (1988) Nucleic Acids Research, 12: 8207-8211; Sharp y col.

(1991) Yeast, 657-678. En algunos ejemplos, las moléculas de ácido nucleico proporcionadas en la presente memoria que codifican un polipéptido de valenceno sintasa están optimizadas de codón para expresión en bacterias o levadura. En ejemplos particulares, las moléculas de ácido nucleico proporcionadas en la presente memoria que codifican una valenceno sintasa están optimizadas de codón para expresión basada en el uso de codón en

5 *Saccharomyces cerevisiae*.

Se proporcionan también en la presente memoria polipéptidos de valenceno sintasa codificados por cualquiera de las moléculas de ácido nucleico proporcionadas en la presente memoria. Los polipéptidos de valenceno sintasa y fragmentos activos de los mismos codificados por las moléculas de ácido nucleico proporcionadas en la presente

10 memoria pueden obtenerse mediante procedimientos bien conocidos en la materia para la generación y expresión de proteína recombinante. Pueden usarse tales polipéptidos de valenceno sintasa para producir valenceno a partir de un precursor terpénico de pirofosfato acíclico, tal como FPP, en la célula hospedadora en la que se expresa la sintasa, o *in vitro* después de la purificación de la sintasa. Puede usarse cualquier procedimiento conocido por los

15 especialistas en la materia para la identificación de ácidos nucleicos que codifican genes deseados para obtener el ácido nucleico que codifica una terpeno sintasa, tal como una valenceno sintasa. Por ejemplo, puede obtenerse un ácido nucleico que codifica polipéptidos de valenceno sintasa usando procedimientos bien conocidos a partir de un

20 fuente vegetal, tal como cardo. Los polipéptidos de valenceno modificados pueden genomanipularse entonces usando cualquier procedimiento conocido en la materia para introducir mutaciones en polipéptidos de valenceno sintasa no modificados o de tipo silvestre, incluyendo cualquier procedimiento descrito en la presente memoria tal como mutagénesis aleatoria del ácido nucleico codificante por PCR propensa a errores, mutagénesis dirigida a sitio, PCR de superposición u otros procedimientos recombinantes. Los ácidos nucleicos que codifican los polipéptidos pueden introducirse entonces en una célula hospedadora para expresarse heterológamente.

### 1. Aislamiento de ácido nucleico que codifica valenceno sintasas

25 Los ácidos nucleicos que codifican valenceno sintasas pueden clonarse o aislarse usando cualquier procedimiento disponible conocido en la materia para clonar y aislar moléculas de ácido nucleico. Tales procedimientos incluyen amplificación por PCR de ácidos nucleicos y cribado de colecciones, incluyendo cribado de hibridación de ácidos nucleicos. En algunos ejemplos, pueden usarse procedimientos para la amplificación de ácidos nucleicos para aislar

30 moléculas de ácido nucleico que codifican un polipéptido de valenceno sintasa, incluyendo por ejemplo procedimientos de reacción en cadena de la polimerasa (PCR). Puede usarse un material que contiene ácido nucleico como material de partida del que puede aislarse una molécula de ácido nucleico que codifica valenceno sintasa. Pueden usarse colecciones de ácido nucleico como fuente de material de partida. Los cebadores se diseñan para amplificar una molécula que codifica valenceno sintasa. Las moléculas proporcionadas en la presente memoria

35 pueden usarse como moldes para cebadores o hibridación o comparación.

Para aislar el ácido nucleico proporcionado aquí, se prepararon preparaciones de ADN y ARNm de cardo (*Eryngium* sp.), en este aspecto *Eryngium glaciale*, y se cribaron con cebadores basados en otras valenceno sintasas que se emplearon para cribado. Por ejemplo, se diseñaron cebadores basados en secuencias de ácido nucleico conocidas

40 que codifican un terpeno sintasa, tal como una valenceno sintasa, germacreno D sintasa y vetispiradieno (premnaspirodieno) sintasa, incluyendo aquellas cuyas secuencias se exponen en las SEQ ID NO: 39-44. Las moléculas de ácido nucleico generadas por amplificación pueden secuenciarse, e identificarse aquellas que codifican un polipéptido de valenceno sintasa.

45 Pueden unirse secuencias nucleotídicas adicionales a una molécula de ácido nucleico que codifica valenceno sintasa, incluyendo secuencias conectoras que contienen sitios de endonucleasa de restricción con el fin de clonar el gen sintético en un vector, por ejemplo un vector de expresión de proteína o un vector diseñado para la amplificación de las secuencias de ADN que codifican la proteína central. Además, pueden conectarse operativamente secuencias nucleotídicas adicionales que especifican elementos de ADN funcionales con una molécula de ácido nucleico que

50 codifica valenceno sintasa. Es más, puede incluirse también ácido nucleico que codifica otros restos o dominios de modo que la sintasa resultante sea una proteína de fusión. Por ejemplo, ácidos nucleicos que codifican otras enzimas, tales como FPP sintasa o marcajes de purificación de proteína, tales como marcajes His o Flag.

#### a. Preparación de un ácido nucleico modificado

55 Puede prepararse o generarse un ácido nucleico que codifica una valenceno sintasa modificada (descrita con más detalle en la Sección D siguiente) usando cualquier procedimiento conocido en la materia por efectuar modificación, particularmente inserciones, deleciones y reemplazos aminoacídicos. Los procedimientos para la modificación de moléculas de ácido nucleico incluyen mutagénesis racional y/o aleatoria de moléculas de ácido nucleico codificantes

60 (usando *p.ej.* PCR propensa a errores, mutagénesis de saturación dirigida a sitio aleatoria, barajado de ADN o mutagénesis dirigida a sitio racional tal como, por ejemplo, kits de mutagénesis (*p.ej.*, QuikChange disponible en

Stratagene) o procedimientos de síntesis en fase sólida). Además, pueden utilizarse técnicas de ADN recombinante rutinarias para generar ácidos nucleicos que codifican polipéptidos que contienen aminoácidos heterólogos. Por ejemplo, pueden generarse ácidos nucleicos que codifican polipéptidos quiméricos o polipéptidos que contienen una secuencia aminoacídica heteróloga usando un procedimiento de PCR en dos etapas y/o usando enzimas de restricción y metodologías de clonación para subclonación rutinaria de los componentes polipeptídicos quiméricos deseados.

Una vez generadas, las moléculas de ácido nucleico pueden expresarse en células para generar polipéptidos de valenceno sintasa modificados usando cualquier procedimiento conocido en la materia. Los polipéptidos de valenceno sintasa modificados pueden valorarse entonces cribando una propiedad o actividad deseada, por ejemplo la capacidad de producir un valenceno a partir de un sustrato adecuado, p.ej. FPP, como se describe en la Sección E. En ejemplos particulares, se generan las valenceno sintasas modificadas con las propiedades deseadas por mutación y se criba una propiedad de acuerdo con los ejemplos ejemplificados en la presente memoria. Típicamente, los polipéptidos de valenceno sintasa modificados producen valenceno a partir de FPP.

15

## 2. Vectores y células para la expresión de polipéptidos de valenceno sintasa

Para la expresión recombinante de uno o más de los polipéptidos de valenceno sintasa codificados por los ácidos nucleicos proporcionados en la presente memoria, puede insertarse el ácido nucleico que contiene toda o una porción de la secuencia nucleotídica que codifica la sintasa en un vector de expresión apropiado, *es decir* un vector que contiene los elementos necesarios para la transcripción y traducción de la secuencia codificante de proteína insertada. Dependiendo del sistema de expresión usado, pueden suministrarse también las señales transcripcionales y traduccionales necesarias para el promotor nativo por un gen de valenceno sintasa y/o sus regiones flanqueantes. Por tanto, se proporcionan también en la presente memoria vectores que contienen un ácido nucleico que codifica cualquier polipéptido de valenceno sintasa proporcionado en la presente memoria. Los vectores ejemplares incluyen pALX.63-71 expuesto en la SEQ ID NO:33 y descrito en el Ejemplo 4 siguiente.

Se proporcionan también células, incluyendo células procarióticas y eucarióticas, que contienen los vectores. Tales células incluyen células bacterianas, células de levadura, células fúngicas, Archea, células vegetales, células de insecto y células de animal. En ejemplos particulares, las células son levadura, tales como *Saccharomyces cerevisiae*, que expresan un precursor terpenico de pirofosfato acíclico tal como FPP. Las células se usan para producir una valenceno sintasa, haciendo crecer las células anteriormente descritas en condiciones mediante las que se expresa la valenceno sintasa codificada por la célula. En algunos aspectos, se purifica la sintasa expresada. En otros aspectos, la valenceno sintasa expresada convierte FPP en uno o más sesquiterpenos (*p.ej.* valenceno) en la célula hospedadora.

Puede usarse cualquier procedimiento conocido por los especialistas en la materia para la inserción de fragmentos de ADN en un vector para construir vectores de expresión que contienen un gen quimérico que contiene señales transcripcionales/traduccionales y secuencias codificantes de proteína apropiadas. Estos procedimientos pueden incluir técnicas de ADN recombinante *in vitro* y sintéticas y recombinantes *in vivo* (recombinación genética). La expresión de secuencias de ácido nucleico que codifican un polipéptido de valenceno sintasa, o dominios, derivados, fragmentos u homólogos del mismo, puede regularse mediante una segunda secuencia de ácido nucleico de modo que los genes o fragmentos de los mismos se expresen en un hospedador transformado con la molécula o moléculas de ADN recombinante. Por ejemplo, la expresión de las proteínas puede controlarse por cualquier promotor/potenciador conocido en la materia. En una realización específica, el promotor no es nativo de los genes que codifican una proteína valenceno sintasa. Los promotores que pueden usarse incluyen, pero sin limitación, promotores procarióticos, de levadura, de mamífero y vegetales. El tipo de promotor depende del sistema de expresión usado, descrito con más detalle a continuación

En una realización específica, se usa un vector que contiene un promotor conectado operativamente con ácidos nucleicos que codifican un polipéptido de valenceno sintasa, o un dominio, fragmento, derivado u homólogo del mismo, uno o mas orígenes de replicación y, opcionalmente, uno o más marcadores seleccionables (*p.ej.*, un gen de resistencia a antibiótico). Se describen vectores y sistemas de expresión de polipéptidos de valenceno sintasa.

## 55 3. Sistemas de expresión

Los polipéptidos de valenceno sintasa (modificados y no modificados) pueden producirse mediante cualquier procedimiento conocido en la materia para la producción de proteína, incluyendo procedimientos *in vitro* e *in vivo* tales como, por ejemplo, la introducción de moléculas de ácido nucleico que codifican la valenceno sintasa en una célula hospedadora o planta hospedadora para producción o expresión *in vivo* a partir de moléculas de ácido nucleico que codifican la valenceno sintasa *in vitro*. Los polipéptidos de valenceno sintasa pueden expresarse en

cualquier organismo adecuado para producir las cantidades y formas requeridas de un polipéptido de sintasa. Los hospedadores de expresión incluyen organismos procarióticos y eucarióticos tales como *E. coli*, levadura, plantas, células de insecto, células de mamífero incluyendo líneas celulares humanas y animales transgénicos. Los hospedadores de expresión pueden diferir en sus niveles de producción de proteína así como en los tipos de modificaciones postraduccionales que están presentes en las proteínas expresadas. La elección del hospedador de expresión puede realizarse basándose en estos y otros factores, tales como consideraciones regulatorias y de seguridad, costes de producción y la necesidad y procedimientos de purificación.

La expresión en células hospedadoras eucarióticas puede incluir la expresión en células de levadura tales como aquellas del género *Saccharomyces* (*p.ej.*, *Saccharomyces cerevisiae*) y el género *Pichia* (*p.ej.*, *Pichia pastoris*), células de insecto tales como células de *Drosophila* y células de *lepidópteros*, plantas y células vegetales tales como cítricos, tabaco, maíz, arroz, algas y *lemna*. Las células eucarióticas para expresión incluyen también líneas de células de mamífero tales como células de ovario de hámster chino (CHO) o células de riñón de hámster recién nacido (BHK). Los hospedadores de expresión eucarióticos incluyen también la producción en animales transgénicos, por ejemplo, incluyendo la producción en suero, leche y huevos.

Están disponibles muchos vectores de expresión y son conocidos por los especialistas en la materia para la expresión de una terpeno sintasa, tal como valenceno sintasa. Es un vector de expresión ejemplar pALX31-180.2, expuesto en la SEQ ID NO:32 y descrito en otro lugar de la presente memoria. La elección del vector de expresión está influida por la elección del sistema de expresión del hospedador. Tal selección está dentro del nivel de competencias del especialista en la materia. En general, los vectores de expresión pueden incluir promotores transcripcionales y opcionalmente potenciadores, señales traduccionales y señales de terminación transcripcional y traduccional. Los vectores de expresión que se usan para transformación estable tienen típicamente un marcador seleccionable que permite la selección y mantenimiento de las células transformadas. En algunos casos, puede usarse un origen de replicación para amplificar el número de copias de los vectores en las células.

Los procedimientos de producción de polipéptidos de terpeno sintasa, incluyendo polipéptidos de valenceno sintasa, pueden incluir la coexpresión de un precursor terpénico de pirofosfato acíclico, tal como FPP, en la célula hospedadora. En algunos aspectos, la célula hospedadora expresa naturalmente FPP. Tal célula puede modificarse para expresar cantidades mayores de FPP (véanse *p.ej.* las pat. de EE.UU. nº. 6.531.303, 6.689.593, 7.838.279 y 7.842.497). En otros aspectos, una célula hospedadora que no produce naturalmente FPP se modifica genéticamente para producir FPP.

#### a. Células procarióticas

Las células procarióticas, especialmente *E. coli*, proporcionan un sistema para producir grandes cantidades de polipéptidos de valenceno sintasa proporcionados en la presente memoria. La transformación de *E. coli* es una técnica sencilla y rápida bien conocida por los especialistas en la materia. Los vectores de expresión ejemplares para transformación de células de *E. coli* incluyen, por ejemplo, los vectores de expresión pGEM, los vectores de expresión pQE y los vectores de expresión pET (véase la patente de EE.UU. nº 4.952.496, disponible en Novagen, Madison, WI; véase también la bibliografía publicada por Novagen que describe el sistema). Tales plásmidos incluyen pET 11a, que contiene el promotor T7lac, el terminador T7, el operador lac de *E. coli* inducible y el gen represor lac; pET 12a-c, que contiene el promotor T7, el terminador T7 y la señal de secreción ompT de *E. coli* y pET 15b y pET 19b (Novagen, Madison, WI), que contienen una secuencia líder de His-Tag™ para uso en la purificación con una columna de Ni y un sitio de escisión por trombina que permite la escisión después de la purificación en la columna, la región promotora T7-lac y el terminador T7.

Los vectores de expresión de *E. coli* pueden contener promotores inducibles que son útiles para inducir altos niveles de expresión de proteína y para expresar proteínas que exhiben cierta toxicidad por las células hospedadoras. Los promotores procarióticos ejemplares incluyen, por ejemplo, el promotor de  $\beta$ -lactamasa (Jay y col., (1981) Proc. Natl. Acad. Sci. USA 78: 5543) y el promotor *tac* (DeBoer y col., (1983) Proc. Natl. Acad. Sci. USA 80: 21-25); véase también "Useful Proteins from Recombinant Bacteria": en Scientific American 242: 74-94 (1980)). Los ejemplos de promotores inducibles incluyen el promotor lac, el promotor trp, el promotor tac híbrido, los promotores de ARN T7 y SP6 y el promotor  $\lambda P_L$  regulado por temperatura.

Los polipéptidos de valenceno sintasa proporcionados en la presente memoria pueden expresarse en el entorno citoplasmático de *E. coli*. El citoplasma es un entorno reductor y para algunas moléculas esto puede dar como resultado la formación de cuerpos de inclusión insolubles. Pueden usarse para resolubilizar las proteínas agentes reductores tales como ditiotretol y  $\beta$ -mercaptoetanol y desnaturalizantes (*p.ej.*, tales como HCl de guanidina y urea). Es un enfoque alternativo la expresión de valenceno sintasas en el espacio periplásmico de bacterias, que proporciona un entorno oxidante y de tipo chaperonina y disulfuro isomerasas que conducen a la producción de

proteína soluble. Típicamente, se fusiona una secuencia líder con la proteína para expresar que dirige la proteína al periplasma. Se retira entonces la líder por peptidasas señal dentro del periplasma. Los ejemplos de secuencias líder de orientación periplásmica incluyen la líder pelB del gen de pectato liasa y la líder derivada del gen de fosfatasa alcalina. En algunos casos, la expresión periplásmica permite el escape de la proteína expresada al medio de cultivo. La secreción de proteínas permite una purificación rápida y sencilla a partir del sobrenadante de cultivo. Las proteínas que no se secretan pueden obtenerse a partir del periplasma por lisis osmótica. De forma similar a la expresión citoplasmática, en algunos casos las proteínas pueden volverse insolubles y pueden usarse desnaturizantes y agentes reductores para facilitar la resolubilización y repliegado. La temperatura de inducción y el crecimiento pueden influir también en los niveles de expresión y solubilidad. Típicamente, se usan temperaturas de entre 25 y 37 °C. Pueden usarse también mutaciones para aumentar la solubilidad de las proteínas expresadas. Típicamente, las bacterias producen proteínas aglicosiladas.

#### b. Células de levadura

Pueden usarse sistemas de levadura tales como, pero sin limitación, aquellos del género *Saccharomyces* (p.ej. *Saccharomyces cerevisiae*), *Schizosaccharomyces pombe*, *Yarrowia lipolytica*, *Kluyveromyces lactis*, y *Pichia pastoris* para expresar los polipéptidos de valenceno sintasa proporcionados en la presente memoria. Pueden usarse también los sistemas de expresión de levadura para producir terpenos, p.ej. valenceno, cuyas reacciones se catalizan por las sintasas (p.ej. valenceno sintasas). La levadura puede transformarse con vectores de replicación episómicos o por integración cromosómica estable por recombinación homóloga. En algunos ejemplos, se usan promotores inducibles para regular la expresión génica. Las secuencias promotoras ejemplares para expresión de polipéptidos de valenceno sintasa en levadura incluyen, entre otras, promotores de metalotionina, 3-fosfoglicerato cinasa (Hitzeman y col. (1980) J. Biol. Chem. 255: 2073) u otras enzimas glicolíticas (Hess y col. (1968) J. Adv. Enzyme Reg. 7: 149 y Holland y col. (1978) Biochem. 17: 4900) tales como enolasa, gliceraldehído fosfato deshidrogenasa, hexocinasa, piruvato descarboxilasa, fosfofructocinasa, glucosa fosfato isomerasa, 3-fosfoglicerato mutasa, piruvato cinasa, trifosfato isomerasa, fosfoglucosa isomerasa y glucocinasa.

Se describen adicionalmente otros vectores y promotores adecuados para uso en la expresión de levadura en Hitzeman, EPA-73.657 o en Fleer y col. (1991) Gene, 107: 285-195 y van den Berg y col. (1990) Bio/Technology, 8: 135-139. Otra alternativa incluye, pero sin limitación, el promotor ADH2 reprimible por glucosa descrito por Russell y col. (J. Biol. Chem. 258: 2674, 1982) y Beier y col. (Nature 300: 724, 1982) o un promotor ADH1 modificado. Los vectores lanzadera replicables en levadura y *E. coli* pueden construirse, por ejemplo, insertando secuencias de ADN de pBR322 para selección y replicación en *E. coli* (gen y origen de replicación Amp<sup>r</sup>) en los vectores de levadura anteriormente descritos.

Los vectores de expresión de levadura pueden incluir un marcador seleccionable tal como LEU2, TRP1, HIS3 y URA3 para selección y mantenimiento del ADN transformado. Los vectores ejemplares incluyen pALX31-108.2, descrito en otro lugar de la presente memoria, que contiene un marcador URA3. Las proteínas expresadas en levadura son a menudo solubles y la coexpresión con chaperoninas, tales como Bip y proteína disulfuro isomerasa, puede mejorar los niveles de expresión y la solubilidad. Adicionalmente, las proteínas expresadas en levadura pueden dirigirse para secreción usando fusiones de péptido señal de secreción tales como la señal de secreción de factor alfa de tipo apareamiento de levadura de *Saccharomyces cerevisiae* y fusiones con proteínas de superficie de célula de levadura tales como el receptor de adhesión por apareamiento Aga2p de glucoamilasa de *Arxula adeninivorans*. Puede genomanipularse un sitio de escisión de proteasa (p.ej., la proteasa Kex-2) para retirar las secuencias fusionadas de los polipéptidos a medida que salen de la ruta de secreción.

La levadura expresa naturalmente las proteínas requeridas, incluyendo FPP sintasa (gen ERG20 que puede producir FPP) para la ruta biosintética de isoprenoide dependiente de mevalonato. Por tanto, la expresión de terpeno sintasas, incluyendo polipéptidos de valenceno sintasa proporcionados en la presente memoria, en células de levadura puede dar como resultado la producción de sesquiterpenos tales como valenceno a partir de FPP. Las células de levadura ejemplares para la expresión de polipéptidos de valenceno sintasa incluyen levadura modificada para expresar niveles aumentados de FPP. Por ejemplo, las células de levadura pueden modificarse para producir menos escualeno sintasa o escualeno sintasa menos activa (p.ej. mutantes *erg9*; véanse p.ej. las patentes de EE.UU. nº 6.531.303 y 6.689.593 y las publicaciones de patente de EE.UU. nº. 2010-0151519 y 2010-0151555). Esto da como resultado la acumulación de FPP en la célula hospedadora a mayores niveles en comparación con las células de levadura de tipo silvestre, lo que a su vez puede dar como resultado rendimientos aumentados de sesquiterpenos (p.ej. valenceno).

En otro ejemplo, las células de levadura pueden modificarse para producir más FPP sintasa mediante la introducción de un gen de FPP sintasa, tal como FPPS de *A. annua* (véase, p.ej. Brodelius y col. (2002) Eur. J. Biochem. 269: 3570-3579), FPP de *N. crassa* y FPS1 y FPS2 de *A. thaliana*. En algunos ejemplos, puede eliminarse el gen de FPP

nativo en tal levadura. Otras modificaciones que posibilitan una producción aumentada de FPP en levadura incluyen, por ejemplo pero sin limitación, modificaciones que aumentan la producción de acetil CoA, inactivan genes que codifican enzimas que usan FPP y GPP como sustrato y sobreexpresan HMG-CoA reductasas, como se describe en la patente de EE.UU. nº 7.842.497. Las células de levadura modificadas ejemplares incluyen, pero sin limitación, las cepas de *Saccharomyces cerevisiae* modificadas CAL15-1 (*ura3, leu2, his3, trp1, Δerg9::HIS3, HMG2cat/TRP1::rDNA, dpp1, sue*), ALX7-95 (*ura3, his3, trp1, Δerg9::HIS3, HMG2cat/TRP1::rDNA, dpp1 sue*), ALX11-30 (*ura3, trp1, erg9<sup>def</sup>25, HMG2cat/TRP1::rDNA, dpp1, sue*), que son conocidas y se describen en una o más de las patentes de EE.UU. nº 6.531.303, 6.689.593, 7.838.279, 7.842.497 y solicitudes de patente de EE.UU. publicadas nº 2004-0249219 y 2011-0189717.

10

### c. Plantas y células vegetales

Pueden usarse células vegetales transgénicas y plantas para la expresión de polipéptidos de valenceno sintasa proporcionados en la presente memoria. Los constructos de expresión se transfieren típicamente a plantas usando transferencia de ADN directa tal como bombardeo de microproyectiles y transferencia mediada por PEG a protoplastos, y con transformación mediada por *Agrobacterium*. Los vectores de expresión pueden incluir secuencias promotoras y potenciadoras, elementos de terminación transcripcional y elementos de control traduccional. Los vectores de expresión y técnicas de transformación se dividen habitualmente entre hospedadores dicotiledóneos tales como *Arabidopsis* y tabaco y hospedadores monocotiledóneos, tales como maíz y arroz. Los ejemplos de promotores vegetales usados para expresión incluyen el promotor de virus del mosaico de la coliflor, el promotor de nopalina sintasa, el promotor de bisfosfato de ribosa carboxilasa y los promotores ubiquitina y UBQ3. Los marcadores seleccionables, tales como higromicina, fosfomanosa isomerasa y neomicina fosfotransferasa se usan a menudo para facilitar la selección y mantenimiento de células transformadas. Las células vegetales transformadas pueden mantenerse en cultivo como células, agregados (tejido calloso) o regenerarse hasta plantas enteras. Las plantas transgénicas pueden incluir también algas genomanipuladas para producir proteínas (véase, por ejemplo, Mayfield y col. (2003) Proc Natl Acad Sci USA 100: 438-442). Las plantas transformadas incluyen, por ejemplo, plantas seleccionadas de entre los géneros *Nicotiana*, *Solanum*, *Sorghum*, *Arabidopsis*, *Medicago* (alfalfa), *Gossypium* (algodón) y *Brassica* (colza). En algunos ejemplos, la planta pertenece a la especie *Nicotiana tabacum*, y se transforma con vectores que sobreexpresan valenceno sintasa y difosfato de farnesilo sintasa, tal como se describe en la publicación de patente de EE.UU. nº 2009-0123984 y la patente de EE.UU. nº 7.906.710.

30

### d. Insectos y células de insecto

Los insectos y células de insecto, particularmente un sistema de expresión de baculovirus, pueden usarse para expresar polipéptidos de valenceno sintasa proporcionados en la presente memoria (véase, por ejemplo, Muneta y col. (2003) J. Vet. Med. Sci. 65(2): 219-223). Las células de insecto y larvas de insecto, incluyendo la expresión en hemolinfa, expresan altos niveles de proteína y son capaces de la mayoría de las modificaciones postraduccionales usadas por eucariotas superiores. Los baculovirus tienen un intervalo hospedador restringido, lo que mejora la seguridad y reduce los problemas regulatorios de la expresión eucariótica. Típicamente, los vectores de expresión usan un promotor tal como el promotor de poliedrina de baculovirus para un alto nivel de expresión. Los sistemas de baculovirus usados comúnmente incluyen baculovirus tales como virus de la poliedrosis nuclear de *Autographa californica* (AcNPV), y virus de la poliedrosis nuclear de *Bombyx mori* (BmNPV) y una línea celular de insecto tal como Sf9, derivada de *Spodoptera frugiperda*, *Pseudaletia unipuncta* (A7S) y *Danaus plexippus* (DpN1). Para un alto nivel de expresión, se fusiona la secuencia nucleotídica de la molécula para expresar inmediatamente en dirección 3' del codón de inicio de poliedrina del virus. Las señales de secreción de mamífero se procesan correctamente en células de insecto y pueden usarse para secretar la proteína expresada en el medio de cultivo. Además, las líneas celulares *Pseudaletia unipuncta* (A7S) y *Danaus plexippus* (DpN1) producen proteínas con patrones de glicosilación similares a los sistemas de células de mamífero.

40

45

Es un sistema de expresión alternativo en células de insecto el uso de células transformadas establemente. Pueden usarse para expresión líneas celulares tales como las células Schnieder 2 (S2) y Kc (*Drosophila melanogaster*) y células C7 (*Aedes albopictus*) El promotor de metalotioneína de *Drosophila* puede usarse para inducir altos niveles de expresión en presencia de inducción por metales pesados con cadmio o cobre. Los vectores de expresión se mantienen típicamente mediante el uso de marcadores seleccionables tales como neomicina e higromicina.

55

### e. Células de mamífero

Pueden usarse sistemas de expresión de mamífero para expresar polipéptidos de valenceno sintasa proporcionados en la presente memoria y pueden usarse también para producir valenceno y otros terpenos cuya formación se cataliza por las valenceno sintasas. Pueden transferirse los constructos de expresión a células de mamífero mediante procedimientos estándares incluyendo, pero sin limitación, infección vírica tal como por adenovirus o por

60

transferencia de ADN directa tal como por liposomas, fosfato de calcio, DEAE-dextrano y por medios físicos tales como electroporación y microinyección. Los vectores de expresión para células de mamífero incluyen típicamente un sitio de caperuza de ARNm, una secuencia TATA, una secuencia de inicio traduccional (secuencia de consenso Kozak) y elementos de poliadenilación. Tales vectores incluyen a menudo promotores transcripcionales-potenciadores para un alto nivel de expresión, por ejemplo el promotor-potenciador de SV40, el promotor de citomegalovirus humano (CMV) o la repetición terminal larga del promotor del virus de sarcoma de Rous (RSV). Estos promotores-potenciadores están activos en muchos tipos celulares. Pueden usarse también para expresión promotores de tipo tejido y célula y regiones potenciadoras. Las regiones promotoras/potenciadoras ejemplares incluyen, pero sin limitación, aquella de genes tales como elastasa I, insulina, inmunoglobulina, virus de tumor mamario de ratón, albúmina, alfa-fetoproteína, alfa-1-antitripsina, beta-globina, proteína básica de mielina, cadena ligera 2 de miosina y control génico de la hormona liberadora de gonadotropina. Pueden usarse marcadores seleccionables para seleccionar y mantener células con el constructo de expresión. Los ejemplos de genes marcadores seleccionables incluyen, pero sin limitación, higromicina B fosfotransferasa, adenosina desaminasa, xantina-guanina fosforibosil transferasa, aminoglicósido fosfotransferasa, dihidrofolato reductasa y timidina cinasa.

15 La fusión con moléculas de señalización de superficie celular tales como TCR- $\xi$  y Fc $\epsilon$ RI- $\gamma$  pueden dirigir la expresión de las proteínas en estado activo sobre la superficie celular.

Están disponibles muchas líneas celulares para expresión de mamífero, incluyendo células de ratón, rata, ser humano, mono y pollo y hámster. Las líneas celulares ejemplares incluyen, pero sin limitación, BHK (es decir células BHK-21, 293-F, CHO, CHO Express (CHOX; Excellgene), Balb/3T3, HeLa, MT2, NS0 de ratón (no secretoras) y otras líneas celulares de mieloma, líneas celulares de hibridoma y heterohibridoma, linfocitos, fibroblastos, células Sp2/0, COS, NIH3T3, HEK293, 293S, 293T, 2B8 y HKB. Las líneas celulares están también disponibles adaptadas a medios exentos de suero, lo que facilita la purificación de proteínas secretadas de los medios de cultivo celular. Es uno de tales ejemplos la línea celular EBNA-1 exenta de suero (Pham y col. (2003) Biotechnol. Bioeng. 84: 332-42).

25

#### 4. Purificación

Los procedimientos de purificación de polipéptidos de valenceno sintasa a partir de células hospedadoras dependen de las células hospedadoras y los sistemas de expresión elegidos. Para moléculas secretadas, las proteínas se purifican generalmente a partir de los medios de cultivo después de retirar las células. Para expresión intracelular, pueden lisarse las células y purificarse las proteínas a partir del extracto. Cuando se usan para expresión organismos transgénicos tales como plantas y animales transgénicos, pueden usarse tejidos u órganos como material de partida para realizar un extracto de células lisadas. Adicionalmente, la producción animal transgénica puede incluir la producción de polipéptidos en leche o huevos, que pueden recogerse, y si es necesario pueden extraerse las proteínas y purificarse además usando procedimientos estándares en la materia.

Las valenceno sintasas pueden purificarse usando técnicas de purificación de proteína estándares conocidas en la materia incluyendo, pero sin limitación, PAGE-SDS, cromatografía de fraccionamiento por tamaño y exclusión por tamaño, precipitación con sulfato de amonio, cromatografía de quelato y cromatografía de intercambio iónico. Los constructos de expresión pueden genomanipularse también para añadir un marcaje de afinidad tal como un epítopo myc, fusión GST o His<sub>6</sub> y purificarse por afinidad con anticuerpo de myc, resina de glutation y resina de Ni, respectivamente, a una proteína. La pureza puede valorarse mediante cualquier procedimiento conocido en la materia incluyendo electroforesis en gel y tinción y técnicas espectrofotométricas.

#### 45 D. Polipéptidos de valenceno sintasa

Se proporcionan en la presente memoria polipéptidos de valenceno sintasa. En algunos ejemplos, los polipéptidos de valenceno sintasa están codificados por ácidos nucleicos aislados a partir del cardo *Eryngium glaciale*. En otros ejemplos, los polipéptidos de valenceno sintasa son variantes o fragmentos catalíticamente activos de aquellos codificados por ácidos nucleicos aislados del cardo *Eryngium glaciale*. Los polipéptidos de valenceno sintasa proporcionados en la presente memoria catalizan la formación de valenceno a partir de un precursor adecuado, tal como un precursor terpénico de pirofosfato acíclico, p.ej. pirofosfato de farnesilo (FPP).

Se proporcionan también en la presente memoria polipéptidos de valenceno sintasa modificados. Las modificaciones contempladas en la presente memoria incluyen, por ejemplo, reemplazos o sustituciones, adiciones o deleciones, truncamientos aminoacídicos o combinaciones de los mismos. Las modificaciones pueden realizarse en cualquier región de una valenceno sintasa a condición de que el polipéptido de valenceno sintasa modificado codificado resultante retenga al menos actividad valenceno sintasa (es decir, la capacidad de catalizar la formación de valenceno a partir de un precursor terpénico de pirofosfato acíclico, típicamente FPP). Los polipéptidos de valenceno sintasa proporcionados en la presente memoria pueden contener otras modificaciones, por ejemplo modificaciones fuera de la secuencia primaria del polipéptido incluyendo, por ejemplo, modificaciones postraduccionales.

60

Se proporcionan también en la presente memoria fragmentos catalíticamente activos de polipéptidos de valenceno sintasa. Por ejemplo, se proporcionan en la presente memoria fragmentos activos del polipéptido de valenceno sintasa que tienen la secuencia de aminoácidos expuesta en la SEQ ID NO:1. Los fragmentos activos retienen la capacidad de catalizar la formación de valenceno a partir de un precursor terpenico de pirofosfato acíclico, tal como pirofosfato de farnesilo (FPP). En ejemplos particulares, los polipéptidos de valenceno sintasa están truncados en el extremo N o C como se describe con más detalle a continuación. En algunos ejemplos, los fragmentos activos de polipéptidos de valenceno sintasa están modificados como se describe en la presente memoria. Tales fragmentos retienen una o más propiedades de un polipéptido de valenceno sintasa de longitud completa. Típicamente, los fragmentos activos exhiben actividad valenceno sintasa (es decir, catalizan la formación de valenceno).

Los polipéptidos de valenceno sintasa proporcionados en la presente memoria pueden generarse mediante cualquier procedimiento conocido por un especialista en la materia. En algunos ejemplos, los polipéptidos de valenceno sintasa proporcionados en la presente memoria se producen sintéticamente, tal como usando síntesis peptídica en fase sólida o fase de solución. Típicamente, los polipéptidos de valenceno sintasa proporcionados en la presente memoria se expresan en una célula hospedadora a partir de un ácido nucleico que codifica el polipéptido de valenceno sintasa, como se describe en la sección C anterior. Es una célula hospedadora ejemplar para la expresión de un polipéptido de valenceno sintasa una célula de levadura, p.ej., una célula de *S. cerevisiae*.

Los polipéptidos de valenceno sintasa proporcionados en la presente memoria pueden usarse para catalizar la producción de valenceno. Típicamente, los polipéptidos de valenceno sintasa proporcionados en la presente memoria catalizan la formación de valenceno a partir de FPP. Las reacciones pueden practicarse *in vivo*, tal como en una célula hospedadora en que se ha introducido un ácido nucleico que codifica el polipéptido de valenceno sintasa (como se describe en la Sección C anterior). Al menos uno de los polipéptidos, p.ej. el polipéptido de valenceno sintasa, será heterólogo del hospedador. Las reacciones pueden practicarse también *in vitro* poniendo en contacto el polipéptido de valenceno sintasa con el sustrato apropiado, p.ej. difosfato de farnesilo, en condiciones apropiadas para la generación de valenceno.

### 1. Polipéptidos de valenceno sintasa de *Eryngium glaciale*

Se proporciona en la presente memoria un polipéptido de valenceno sintasa que tiene la secuencia de aminoácidos expuesta en la SEQ ID NO:1. Se divulgan también en la presente memoria polipéptidos de valenceno sintasa que exhiben al menos un 50 % de identidad de secuencia aminoacídica con el polipéptido de valenceno sintasa expuesto en la SEQ ID NO:1. Por ejemplo, los polipéptidos de valenceno sintasa divulgados en la presente memoria pueden exhibir al menos o aproximadamente al menos un 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad de secuencia aminoacídica con el polipéptido de valenceno sintasa expuesto en la SEQ ID NO:1, a condición de que el polipéptido de valenceno sintasa resultante retenga al menos actividad valenceno sintasa (es decir, la capacidad de catalizar la formación de valenceno a partir de un precursor terpenico de pirofosfato acíclico, típicamente FPP). Puede determinarse la identidad porcentual por un especialista en la materia usando programas de alineamiento estándares.

De acuerdo con la presente invención, el polipéptido de valenceno sintasa tiene al menos un 90 % de identidad de secuencia con el polipéptido expuesto en la SEQ ID NO:1.

### 2. Modificaciones de polipéptidos de valenceno sintasa de *Eryngium glaciale*

Se proporcionan en la presente memoria polipéptidos de valenceno sintasa de *Eryngium glaciale* modificados. Se proporcionan también en la presente memoria moléculas de ácido nucleico que codifican cualquiera de los polipéptidos de valenceno sintasa modificados proporcionados en la presente memoria. Las modificaciones pueden realizarse en cualquier región de una valenceno sintasa a condición de que el polipéptido de valenceno sintasa modificado resultante retenga al menos actividad valenceno sintasa (es decir, la capacidad de catalizar la formación de valenceno a partir de un precursor terpenico de pirofosfato acíclico, típicamente FPP).

Las modificaciones pueden ser una modificación aminoacídica única, tal como un reemplazo (sustitución), inserción o delección único, o modificaciones aminoacídicas múltiples tales como reemplazos, inserciones o delecciones aminoacídicas múltiples. En algunos ejemplos, se intercambian dominios o regiones enteros o parciales, tales como cualquier dominio o región descrito en la presente memoria a continuación, con los dominios o regiones o porciones de los mismos correspondientes de otra terpeno sintasa. Son modificaciones ejemplares reemplazos aminoacídicos, incluyendo reemplazos aminoacídicos únicos o múltiples. Típicamente, la modificación es un reemplazo aminoacídico que puede ser una sustitución conservativa, tal como se expone en la Tabla 2, o una sustitución no conservativa. Por ejemplo, los polipéptidos de valenceno sintasa modificados proporcionados en la presente

memoria pueden contener al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 90, 95, 100, 105, 110, 115, 120 o más posiciones modificadas en comparación con el polipéptido de valenceno sintasa que no contiene la modificación.

Las modificaciones descritas en la presente memoria pueden estar en cualquier polipéptido de valenceno sintasa. Típicamente, las modificaciones se realizan en un polipéptido de valenceno sintasa proporcionado en la presente memoria. Por ejemplo, las modificaciones descritas en la presente memoria pueden estar en un polipéptido de valenceno sintasa como se expone en la SEQ ID NO:1 o cualquier variante del mismo, incluyendo cualquiera que tenga al menos un 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad de secuencia con el polipéptido de valenceno sintasa expuesto en la SEQ ID NO:1.

En particular, los polipéptidos de valenceno sintasa modificados proporcionados en la presente memoria contienen reemplazos o sustituciones, adiciones o deleciones, truncamientos aminoacídicos o combinaciones de los mismos con referencia al polipéptido de valenceno sintasa expuesto en la SEQ ID NO:1. Está dentro de nivel de un especialista en la materia realizar tales modificaciones en los polipéptidos de valenceno sintasa, tal como cualquiera expuesto en la SEQ ID NO:1 o cualquier variante del mismo. Los procedimientos ejemplares para generar polipéptidos de valenceno sintasa modificados se proporcionan en la Sección C.1.a anterior. Está dentro del nivel de un especialista en la materia generar una valenceno sintasa que contenga una cualquiera o más de las mutaciones descritas, y ensayar en cada una la actividad valenceno sintasa como se describe en la presente memoria.

También, en algunos ejemplos, se proporcionan en la presente memoria fragmentos activos modificados de polipéptidos de valenceno sintasa que contienen cualquiera de las modificaciones proporcionadas en la presente memoria. Tales fragmentos retienen una o más propiedades de una valenceno sintasa. Típicamente, los fragmentos activos modificados exhiben actividad valenceno sintasa (es decir, la capacidad de catalizar la formación de valenceno a partir de un precursor terpenico de pirofosfato acíclico, p.ej. FPP).

Las modificaciones en un polipéptido de valenceno sintasa pueden realizarse también en un polipéptido de valenceno sintasa que contiene también otras modificaciones, incluyendo modificaciones de la secuencia primaria y modificaciones no en la secuencia primaria del polipéptido. Por ejemplo, la modificación descrita en la presente memoria puede ser en un polipéptido de valenceno sintasa que es un polipéptido de fusión o polipéptido quimérico, incluyendo híbridos de diferentes polipéptidos de valenceno sintasa o diferentes polipéptidos de terpeno sintasa (p.ej. que contienen uno o más dominios o regiones de otra terpeno sintasa) y también polipéptidos de valenceno sintasa sintéticos preparados recombinantemente o sintetizados o construidos mediante otros procedimientos conocidos en la materia basándose en la secuencia de polipéptidos conocidos.

Para retener la actividad valenceno sintasa, las modificaciones no se realizan típicamente en aquellas posiciones que sean necesarias para la actividad valenceno sintasa, es decir, en el sitio activo de motivo DDxxD (SEQ ID NO:35) o motivo NSE/DTE (SEQ ID NO:36). Por ejemplo, generalmente no se realizan modificaciones en una posición correspondiente a la posición D315, D316 o D319 o en una posición correspondiente a la posición D461, G465 o E469, con referencia a una secuencia de aminoácidos expuesta en la SEQ ID NO:1.

Los polipéptidos de valenceno sintasa modificados pueden contener una o más sustituciones aminoacídicas, en cualquier combinación, con o sin modificaciones adicionales. Generalmente, pueden combinarse múltiples modificaciones proporcionadas en la presente memoria por un especialista en la materia, siempre que el polipéptido modificado retenga la capacidad de catalizar la formación de valenceno y/u otros terpenos a partir de cualquier precursor terpenico de pirofosfato acíclico adecuado incluyendo, pero sin limitación, FPP, GPP y GGPP. Típicamente, los polipéptidos de valenceno modificados catalizan la formación de valenceno a partir de FPP. En algunos ejemplos, el polipéptido de valenceno sintasa modificado resultante exhibe una producción de valenceno similar o aumentada a partir de FPP en comparación con el polipéptido de valenceno sintasa no modificado. En algunos aspectos, el polipéptido de valenceno sintasa modificado exhibe una producción de valenceno disminuida a partir de FPP en comparación con el polipéptido de valenceno sintasa no modificado.

Se proporcionan también en la presente memoria moléculas de ácido nucleico que codifican cualquiera de los polipéptidos de valenceno sintasa modificados proporcionados en la presente memoria. En ejemplos particulares, las moléculas de ácido nucleico pueden estar optimizadas de codón, por ejemplo, para aumentar los niveles de expresión de la secuencia codificada. En un ejemplo, las secuencias de ácido nucleico que codifican polipéptidos de valenceno sintasa modificados proporcionados en la presente memoria están optimizados de codón basándose en el uso de codón en *S. cerevisiae*.

Los polipéptidos modificados y moléculas de ácido nucleico codificantes proporcionados en la presente memoria pueden producirse por técnicas de ADN recombinante estándares conocidas por el especialista en la materia. Puede emplearse cualquier procedimiento conocido en la materia para efectuar la mutación de uno cualquiera o más aminoácidos en una proteína diana. Se proporcionan procedimientos ejemplares en la Sección C anterior. En algunos ejemplos, los polipéptidos de valenceno sintasa modificados se producen sintéticamente, tal como usando síntesis peptídica en fase sólida o fase de solución.

#### a. Polipéptidos truncados

- 10 Se proporcionan también en la presente memoria polipéptidos de valenceno sintasa truncados. Los polipéptidos de valenceno sintasa truncados pueden truncarse en el extremo N o el extremo C, siempre que los polipéptidos de valenceno sintasa truncados retengan la actividad catalítica de una valenceno sintasa. Típicamente, los polipéptidos de valenceno sintasa truncados exhiben actividad valenceno sintasa (*es decir*, la capacidad de catalizar la formación de valenceno a partir de un precursor terpénico de pirofosfato acíclico, tal como FPP). En algunos ejemplos, los
- 15 polipéptidos de valenceno sintasa proporcionados en la presente memoria están truncados en el extremo N. En otros ejemplos, los polipéptidos de valenceno sintasa proporcionados en la presente memoria están truncados en el extremo C. En aun otros ejemplos, los polipéptidos de valenceno sintasa proporcionados en la presente memoria están truncados en el extremo N y el extremo C.
- 20 En algunos ejemplos, los polipéptidos de valenceno sintasa están truncados en el extremo N, el extremo C o ambos extremos de un polipéptido de valenceno sintasa proporcionado en la presente memoria, tal como el truncamiento de la secuencia de aminoácidos expuesta en la SEQ ID NO:1. En otros ejemplos, están truncados cualquiera de las valenceno sintasas modificadas proporcionadas en la presente memoria. En algunos ejemplos, los polipéptidos de valenceno sintasa están truncados en su extremo N. Por ejemplo, cualquier polipéptido de valenceno sintasa
- 25 proporcionado en la presente memoria puede estar truncado en o aproximadamente o al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75 o más residuos aminoacídicos en el extremo N, a condición de que el polipéptido de valenceno sintasa retenga la actividad valenceno sintasa.
- 30 En otros ejemplos, cualquier polipéptido de valenceno sintasa proporcionado en la presente memoria puede estar truncado en o aproximadamente o al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75 o más residuos aminoacídicos en
- 35 el extremo C, a condición de que el polipéptido de valenceno sintasa retenga actividad valenceno sintasa.

#### b. Polipéptidos con actividades o propiedades alteradas

- Los polipéptidos de valenceno sintasa proporcionados en la presente memoria pueden exhibir también cambios en actividades y/o propiedades. La valenceno sintasa modificada puede exhibir, por ejemplo, una actividad catalítica aumentada, una unión a sustrato aumentada (*p.ej.* FPP), una estabilidad aumentada, una expresión aumentada en una célula hospedadora y/o una distribución de producto alterada (*es decir*, cantidades relativas y/o tipos de terpenos alterados) en comparación con un polipéptido de valenceno sintasa de tipo silvestre. Tales actividades y propiedades alteradas pueden dar como resultado una producción de valenceno aumentada a partir de pirofosfato
- 45 de farnesilo. Típicamente, la distribución de producto de terpenos producidos por una valenceno sintasa de tipo silvestre incluye valenceno, así como una serie de otros productos de terpeno (*p. ej.*, subproductos de terpeno o productos derivados de los mismos) incluyendo, por ejemplo,  $\beta$ -selineno,  $\tau$ -selineno, eremofilona, 7-*epi*- $\alpha$ -selineno, germacreno A,  $\beta$ -elemeno y aristoloqueno (pico 2 en la Figura 3A).
- 50 En algunos ejemplos, los polipéptidos de valenceno sintasa modificados pueden catalizar la formación terpenos distintos de valenceno a partir de cualquier sustrato adecuado tal como, por ejemplo, FPP, GPP o GGPP. Por ejemplo, las valenceno sintasas modificadas pueden producir uno o más monoterpenos, sesquiterpenos o diterpenos distintos de valenceno. Típicamente, los polipéptidos de valenceno sintasa modificados producen más valenceno que cualquier otro terpeno. Esto puede dar como resultado una producción aumentada de nootkatona. Las
- 55 modificaciones que dan como resultado una producción aumentada de valenceno a partir de FPP pueden identificarse usando los ensayos descritos en la presente memoria y bien conocidos en la materia, permitiendo por tanto la identificación de polipéptidos de valenceno sintasa modificados con capacidad mejorada de producir valenceno a partir de FPP.

#### 60 c. Intercambios de dominio

Se proporcionan en la presente memoria polipéptidos de valenceno sintasa modificados que son polipéptidos quiméricos que contienen un intercambio (delección e inserción) por delección de residuos aminoácidos de uno o más dominios o regiones del mismo o porciones del mismo e inserción de una secuencia heteróloga de aminoácidos. En algunos ejemplos, la secuencia heteróloga es una secuencia aleatorizada de aminoácidos. En otros ejemplos, la secuencia heteróloga es una secuencia contigua de aminoácidos del dominio o región o porción del mismo correspondiente de otro polipéptido de terpeno sintasa. La secuencia heteróloga que se reemplaza o inserta incluye generalmente al menos 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40 o más aminoácidos. En ejemplos en que la secuencia heteróloga es de un dominio correspondiente o porción del mismo de otra terpeno sintasa, la secuencia heteróloga incluye generalmente al menos un 50 %, 60 %, 70 %, 80 %, 90 %, 95 % o más aminoácidos contiguos del dominio o región o porción correspondiente. En tal ejemplo, los residuos adyacentes al dominio o región o porción del mismo correspondiente heterólogo pueden incluirse también en un polipéptido de valenceno sintasa modificado proporcionado en la presente memoria.

En un ejemplo de mutantes de intercambio proporcionados en la presente memoria, se reemplaza al menos un dominio o región o porción del mismo de un polipéptido de valenceno sintasa por una secuencia contigua de aminoácidos del dominio o región o porciones del mismo correspondiente de otro polipéptido de terpeno sintasa. En algunos ejemplos, se reemplazan 2, 3, 4, 5, 6, 7, 8, 9, 10 o más dominios o regiones o porciones de los mismos por una secuencia contigua de aminoácidos del dominio o región o porciones del mismo correspondiente de otro polipéptido de terpeno sintasa.

Puede reemplazarse cualquier dominio o región o porción del mismo de un polipéptido de valenceno sintasa por una secuencia heteróloga de aminoácidos, tal como una secuencia heteróloga del dominio o región correspondiente de otro terpeno. Un dominio o región puede ser un dominio estructural o un dominio funcional. Un especialista en la materia está familiarizado con los dominios o regiones de las terpeno sintasas. Los dominios funcionales incluyen, por ejemplo, el dominio N-funcional y el dominio catalítico C-terminal o una porción del mismo. Un dominio estructural puede incluir todo o una porción de un bucle, bucle desestructurado o dominio de hélice alfa.

Un especialista en la materia está familiarizado con las diversas terpeno sintasas y puede identificar los dominios o regiones correspondientes o porciones de aminoácidos del mismo. Por ejemplo, se describen dominios y regiones ejemplares de valenceno sintasa cítrica en la solicitud de patente de EE.UU. n° 2012-0246767. Las terpeno sintasas ejemplares incluyen, por ejemplo, sesquiterpeno sintasas. En ejemplos particulares de la presente memoria, los mutantes de intercambio de dominio polipeptídico de valenceno sintasa modificado proporcionados en la presente memoria contienen secuencias heterólogas de un dominio o región correspondiente o porción del mismo de un polipéptido de terpeno sintasa que es una valenceno sintasa de *Citrus* sp. (SEQ ID NO: 14-15, 26-27, 34 o 37), valenceno sintasa de *V. vinifera* (SEQ ID NO:16 o 28), valenceno sintasa de *P. frutescens* (SEQ ID NO:38) o valenceno sintasa de *C. nootkatensis* (SEQ ID NO: 17 o 29).

Típicamente, la valenceno sintasa modificada resultante exhibe actividad valenceno sintasa y la capacidad de producir valenceno a partir de FPP. Por ejemplo, los polipéptidos de valenceno sintasa modificados exhiben de 50 % a 5000 %, tal como de 50 % a 120 %, de 100 % a 500 % o de 110 % a 250 % de producción de valenceno a partir de FPP en comparación con el polipéptido de valenceno sintasa que no contiene la modificación (p.ej., el reemplazo aminoácido o intercambio de residuos aminoácidos de un dominio o región) y/o en comparación con el polipéptido de valenceno sintasa de tipo silvestre expuesto en la SEQ ID NO:1. Típicamente, los polipéptidos de valenceno sintasa exhiben una producción de valenceno aumentada a partir de FPP en comparación con el polipéptido de valenceno sintasa que no contiene la modificación, tal como en comparación con la valenceno sintasa de tipo silvestre expuesta en la SEQ ID NO:1.

Por ejemplo, los polipéptidos de valenceno sintasa modificados pueden producir valenceno a partir de FPP en una cantidad que es al menos o aproximadamente un 101 %, 102 %, 103 %, 104 %, 105 %, 106 %, 107 %, 108 %, 109 %, 110 %, 115 %, 120 %, 125 %, 130 %, 135 %, 140 %, 145 %, 150 %, 160 %, 170 %, 180 %, 200 %, 250 %, 300 %, 350 %, 400 %, 500 %, 1500 %, 2000 %, 3000 %, 4000 % o 5000 % de la cantidad de valenceno producida a partir de FPP por valenceno sintasa de tipo silvestre que no contiene la modificación en las mismas condiciones. Por ejemplo, la producción de valenceno aumenta al menos 1,2 veces, 1,5 veces, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces, 7 veces, 8 veces, 9 veces, 10 veces, 11 veces, 12 veces, 13 veces, 14 veces, 15 veces, 16 veces, 17 veces, 18 veces, 19 veces, 20 veces o más.

En otros ejemplos, los polipéptidos de valenceno sintasa modificados exhiben una producción de valenceno disminuida a partir de FPP en comparación con el polipéptido de valenceno sintasa que no contiene la modificación, tal como en comparación con la valenceno sintasa de tipo silvestre expuesta en la SEQ ID NO:1. Por ejemplo, los polipéptidos de valenceno sintasa modificados pueden producir valenceno a partir de FPP en una cantidad que es

- como máximo o de aproximadamente un 99 %, 98 %, 97 %, 96 %, 95 %, 94 %, 93 %, 92 %, 91 %, 90 %, 85 %, 80 %, 75 %, 70 %, 65 %, 60 %, 55 %, 50 %, 45 %, 40 %, 35 %, 30 %, 25 %, 20 %, 15 %, 10 % o menos de la cantidad de valenceno producida a partir de FPP por valenceno sintasa de tipo silvestre que no contiene la modificación en las mismas condiciones. Por ejemplo, la producción de valenceno sintasa disminuye de tal modo que la producción de FPP por una valenceno sintasa de tipo silvestre que no contiene la modificación en las mismas condiciones es de al menos 1,2 veces, 1,5 veces, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces, 7 veces, 8 veces, 9 veces, 10 veces, 11 veces, 12 veces, 13 veces 14 veces 15 veces, 16 veces, 17 veces, 18 veces, 19 veces, 20 veces o más mayor que la cantidad de valenceno producida a partir de FPP por el polipéptido de valenceno sintasa modificado.
- 5  
10 Puede usarse cualquier procedimiento conocido en la materia para generar polipéptidos quiméricos para reemplazar todo o una porción contigua de un dominio de una primera terpeno sintasa por todo o una porción contigua del dominio correspondiente de una segunda sintasa (véanse las patentes de EE.UU. n° 5.824.774, 6.072.045, 7.186.891 y 8.106.260 y la publicación de patente de EE.UU. n° 2011-0081703). También pueden emplearse procedimientos de barajado génico para generar polipéptidos quiméricos y/o polipéptidos con intercambios de dominio o región.

- Por ejemplo, pueden intercambiarse los dominios o regiones correspondientes de dos terpeno sintasas cualesquiera usando cualquier procedimiento recombinante adecuado conocido en la materia, o por síntesis *in vitro*. Es un ejemplo de un procedimiento recombinante un procedimiento de PCR de superposición de dos etapas tal como se describe en la presente memoria. En tales procedimientos, pueden emplearse cebadores que introducen mutaciones en una pluralidad de posiciones de codón en los ácidos nucleicos que codifican el dominio diana o porción del mismo en la primera terpeno sintasa, donde las mutaciones forman conjuntamente la región heteróloga (es decir, la región correspondiente de la segunda terpeno sintasa). Como alternativa, por ejemplo, pueden usarse aminoácidos aleatorizados para reemplazar dominios o regiones específicos. Se entiende que los errores de cebador, errores de PCR y/u otros errores en los procedimientos de clonación o recombinantes pueden dar como resultado errores tales que la región o dominio intercambiado o reemplazado resultante no exhiba una secuencia aminoacídica que sea idéntica a la región correspondiente de la segunda terpeno sintasa.
- 20  
25

- En un procedimiento basado en PCR ejemplar, la PCR de primera etapa usa (i) un cebador en dirección 3' que se reasocia en dirección 3' de la región que se está reemplazando con un cebador mutagénico que incluye aproximadamente 15 nucleótidos (o un número efectivo para efectuar la reasociación, tal como 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 25 nucleótidos o más) de secuencia homóloga en cada lado del dominio o región para intercambiar o aleatorizar flanqueante de la región para importar al gen diana y (ii) un cebador en dirección 5' que se reasocia en dirección 5' de la región que se está reemplazando junto con un cebador mutagénico de la hebra contraria que incluye también aproximadamente 15 nucleótidos (o un número efectivo para efectuar la reasociación, tal como 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 25 nucleótidos o más) de secuencia homóloga en cada lado del dominio o región para intercambiar o aleatorizar flanqueante de la región para importar al gen diana. Si se practica un reemplazo en que se reemplaza un dominio o región de un primer gen de terpeno sintasa por el dominio o región correspondiente de una segunda terpeno sintasa, los nucleótidos de los cebadores mutagénicos entre las regiones flanqueantes de la primera terpeno sintasa contienen codones de la región correspondiente de la segunda terpeno sintasa. En aspectos donde los aminoácidos en un dominio o región se vayan a aleatorizar, los nucleótidos de los cebadores mutagénicos entre las regiones flanqueantes de la primera terpeno sintasa contienen nucleótidos aleatorios. Se practica entonces una PCR de superposición para unir los dos fragmentos, usando los oligos en dirección 5' y dirección 3'. El producto de PCR resultante puede clonarse entonces en cualquier vector adecuado para la expresión de la terpeno sintasa modificada.
- 30  
35  
40  
45

Además, cualquiera de los polipéptidos de valenceno sintasa modificados que contienen mutaciones de intercambio en la presente memoria puede contener uno o más reemplazos aminoacídicos adicionales.

#### 50 d. Variantes adicionales

- Los polipéptidos de valenceno sintasa proporcionados en la presente memoria pueden modificarse mediante cualquier procedimiento conocido por un especialista en la materia para generar variantes proteicas incluyendo, pero sin limitación, barajado de ADN o génico, PCR propensa a errores, PCR de superposición u otros procedimientos recombinantes. En un ejemplo, pueden modificarse por barajado génico moléculas de ácido nucleico que codifican cualquier polipéptido de valenceno sintasa o polipéptido de valenceno sintasa variante proporcionado en la presente memoria. El barajado génico implica uno o más ciclos de fragmentación aleatoria y reconstrucción de al menos dos secuencias nucleotídicas, seguido de cribado para seleccionar las secuencias nucleotídicas que codifican polipéptidos con las propiedades deseadas. La recombinación puede *practicarse in vitro* (véanse Stemmer y col. (1994) Proc Natl Acad Sci USA 91: 10747-10751; Stemmer y col. (1994) Nature 370: 389-391; Cramieri y col. (1998) Nature 391: 288-291; patentes de EE.UU. n° 5.605.793, 5.811.238, 5.830.721, 5.834.252 y 5.837.458) o *in*
- 55  
60

vivo (véase la publicación de patente internacional . nº WO199707205). Las moléculas de ácido nucleico que codifican los polipéptidos pueden introducirse entonces en una célula hospedadora para expresar heterológicamente y ensayarse su actividad valenceno sintasa mediante cualquier procedimiento descrito en la sección E siguiente.

#### 5 e. Proteínas de fusión

Los polipéptidos de valenceno sintasa pueden utilizarse también o expresarse como fusiones proteicas. Por ejemplo, puede generarse una fusión para añadir una funcionalidad adicional a un polipéptido. Los ejemplos de proteínas de fusión incluyen, pero sin limitación, fusiones de una secuencia señal, un marcaje tal como para localización, *p.ej.* un marcaje his<sub>6</sub> o un marcaje myc, o un marcaje para purificación, por ejemplo una fusión de GST, fusión de GFP o fusión de CBP, y una secuencia para dirigir la secreción de proteína y/o la asociación a membrana. En otros ejemplos, puede fusionarse una sesquiterpeno sintasa, tal como un polipéptido de valenceno sintasa proporcionado en la presente memoria, con una FPP sintasa (véase, *p.ej.*, Brodelius y col. (2002) Eur. J. Biochem. 269: 3570-3579).

Se proporcionan también proteínas de fusión que contienen un polipéptido de valenceno sintasa y uno o más de otros polipéptidos. La conexión de un polipéptido de valenceno sintasa con otro polipéptido puede efectuarse directa o indirectamente a través de un conector. En un ejemplo, la conexión puede ser por conexión química, tal como a través de agentes heterobifuncionales o conexiones de tiol u otras de tales conexiones. La fusión puede efectuarse también por medios recombinantes. La fusión de una terpeno sintasa, tal como un polipéptido de valenceno sintasa, con otro polipéptido puede ser con el extremo N o C del polipéptido de valenceno sintasa.

Puede producirse una proteína de fusión mediante técnicas recombinantes estándares. Por ejemplo, los fragmentos de ADN que codifican diferentes secuencias polipeptídicas pueden ligarse conjuntamente en fase de acuerdo con técnicas convencionales, *p.ej.* empleando extremos romos o escalonados para ligamiento, digestión con enzima de restricción para proporcionar extremos apropiados, relleno de extremos cohesivos según sea apropiado, tratamiento con fosfatasa alcalina para evitar uniones indeseables y ligamiento enzimático. En otra realización, puede sintetizarse el gen de fusión mediante técnicas convencionales incluyendo sintetizadores de ADN automatizados. Como alternativa, puede llevarse a cabo la amplificación por PCR de fragmentos génicos usando cebadores de anclaje que dan lugar a salientes complementarios entre dos fragmentos génicos consecutivos que pueden reasociarse y reamplificarse posteriormente, generando una secuencia génica quimérica (véase, *p.ej.*, Ausubel y col (eds.) Current Protocols in Molecular Biology, John Wiley & Sons, 1992). Además, están comercialmente disponibles muchos vectores de expresión que ya codifican un resto de fusión (*p.ej.*, un polipéptido de GST). Por ejemplo, puede clonarse un ácido nucleico que codifica un polipéptido de valenceno sintasa en un vector de expresión tal que el resto de fusión esté conectado en fase con la proteína valenceno sintasa. La valenceno sintasa y el resto de fusión pueden estar conectados directamente, sin un conector, o como alternativa conectados indirectamente en fase con un conector.

#### 40 E. PROCEDIMIENTOS PARA PRODUCIR TERPENOS Y PROCEDIMIENTOS PARA DETECTAR TALES PRODUCTOS Y LA ACTIVIDAD DE POLIPÉPTIDOS DE VALENCENO SINTASA

Las valenceno sintasas proporcionadas en la presente memoria pueden usarse para, y valorarse por, su capacidad de producir terpenos tales como monoterpenos, sesquiterpenos y diterpenos, a partir de cualquier precursor terpénico de pirofosfato acíclico adecuado incluyendo, pero sin limitación, difosfato de geranilo (GPP), difosfato de farnesilo (FPP) o difosfato de geranilgeranilo (GGPP). Típicamente, los polipéptidos de valenceno sintasa proporcionados en la presente memoria catalizan la formación del sesquiterpeno valenceno a partir de FPP. Los polipéptidos de valenceno sintasa proporcionados en la presente memoria catalizan también la formación de un compuesto adicional, aristoloqueno (designado pico 2; véase la Figura 3A) a partir de FPP.

Puede usarse cualquier procedimiento conocido por un especialista en la materia para producir terpenos, incluyendo valenceno, con las valenceno sintasas proporcionadas en la presente memoria. Puede valorarse la capacidad de las valenceno sintasas proporcionadas en la presente memoria de catalizar la formación de valenceno u otros terpenos a partir de FPP usando estos procedimientos. En algunos ejemplos, se compara la cantidad de terpeno, tal como valenceno, producida a partir de FPP usando las valenceno sintasas proporcionadas en la presente memoria con la cantidad de terpeno, tal como valenceno, producida a partir del mismo sustrato usando una valenceno sintasa diferente, tal como una valenceno sintasa de una especie diferente, por ejemplo una valenceno sintasa cítrica, o cualquier otra valenceno sintasa conocida por un especialista en la materia.

Pueden valorarse también otras actividades y propiedades de las valenceno sintasas, tales como los polipéptidos de valenceno sintasa proporcionados en la presente memoria, usando procedimientos y ensayos bien conocidos en la materia. Además de valorar la actividad de las valenceno sintasas y su capacidad de catalizar la formación de

terpenos, puede valorarse la cinética de la reacción, regioquímica o estereoquímica modificada, utilización de sustrato alterada y/o distribución de producto alterada (es decir, cantidad alterada de los diferentes terpenos producidos a partir de FPP u otro sustrato) en comparación con las valenceno sintasas usando procedimientos bien conocidos en la materia. Por ejemplo, puede valorarse el tipo y la cantidad de los diversos terpenos producidos a partir de FPP, GPP o GGPP por los polipéptidos de valenceno sintasa proporcionados en la presente memoria por procedimientos de cromatografía de gases (p.ej., GC-MS), tales como aquellos descritos a continuación y en el Ejemplo 4. En algunos ejemplos, el perfil de los terpenos producidos por los polipéptidos de valenceno sintasa a partir de FPP incluye, pero sin limitación, valenceno,  $\beta$ -elemeno y aristoloqueno (pico 2). Se producen también por las células hospedadoras los alcoholes terpénicos acíclicos nerolidol y farnesol.

Se proporcionan a continuación procedimientos para la producción de terpenos, incluyendo valenceno y nootkatona, a partir de FPP usando las valenceno sintasas proporcionadas en la presente memoria.

### 1. Producción de terpenos catalizada por valenceno sintasa de *Eryngium glaciale*

Los polipéptidos de valenceno sintasa modificados pueden usarse para catalizar la formación de valenceno y otros terpenos a partir de un precursor terpénico de pirofosfato acíclico, tal como FPP. En algunos ejemplos, las valenceno sintasas proporcionadas en la presente memoria se expresan en células que producen o sobreproducen FPP, de tal modo que se produce valenceno mediante la ruta descrita anteriormente. En otros ejemplos, las valenceno sintasas proporcionadas en la presente memoria se expresan y purifican a partir de cualquier célula hospedadora adecuada, tal como se describe en la Sección C. Se combinan entonces las sintasas purificadas *in vitro* con FPP para producir valenceno.

En algunos ejemplos, la valenceno sintasa proporcionada en la presente memoria se sobreexpresa y purifica como se describe en la Sección C anterior. Se incuba entonces la valenceno sintasa con el sustrato difosfato de farnesilo y se produce valenceno. El pH de la solución que contiene FPP y valenceno sintasa puede afectar a la cantidad de valenceno producida (véase, p.ej. la publicación de patente de EE.UU. n.º 2010-0216186). Se añade un disolvente orgánico para repartir el valenceno en la fase orgánica para análisis. Se determinan entonces la producción de valenceno y la cuantificación de la cantidad de producto usando cualquier procedimiento proporcionado en la presente memoria, tal como cromatografía de gases (p.ej. GC-MS) usando un patrón interno. Como alternativa, se expresa la valenceno sintasa en células hospedadoras que producen también FPP, dando como resultado la producción de valenceno. El valenceno puede entonces extraerse del medio de cultivo celular con un disolvente orgánico y aislarse y purificarse posteriormente mediante cualquier procedimiento conocido, tal como cromatografía en columna o HPLC, y valorarse la cantidad y la pureza del valenceno recuperado. En algunos ejemplos, se convierte el valenceno por oxidación en nootkatona antes o después de la purificación.

#### a. Células ejemplares

El valenceno puede producirse expresando un polipéptido de valenceno sintasa proporcionado en la presente memoria en una línea celular que produce FPP como parte de la ruta biosintética de isoprenoide dependiente de mevalonato (p.ej. hongos, incluyendo células de levadura y células animales) o la ruta biosintética de isoprenoide independiente de mevalonato (p.ej. bacterias y plantas superiores). En ejemplos particulares, se produce valenceno expresando un polipéptido de valenceno sintasa proporcionado en la presente memoria en una línea celular que se ha modificado para sobreproducir FPP. Son ejemplos de tales células las células de levadura modificadas. Por ejemplo, las células de levadura que se han modificado para producir menos escualeno sintasa o escualeno sintasa menos activa (p.ej. mutantes *erg9*; véanse p.ej. las patentes de EE.UU. n.º 6,531,303 y 6.689.593 y las publicaciones de patente de EE.UU. n.º 2010-0151519 y 2010-0151555) son útiles en los procedimientos proporcionados en la presente memoria para producir valenceno. Una actividad escualeno sintasa reducida da como resultado la acumulación de FPP en la célula hospedadora a mayores niveles en comparación con células de levadura de tipo silvestre, lo que a su vez puede dar como resultado rendimientos aumentados de la producción de valenceno. Las células de levadura modificadas ejemplares incluyen, pero sin limitación, las cepas de *Saccharomyces cerevisiae* modificadas CALI5-1 (*ura3, leu2, his3, trp1,  $\Delta$ erg9::HIS3, HMG2cat/TRP1::rDNA, dpp1*), ALX7-95 (*ura3, his3, trp1,  $\Delta$ erg9::HIS3, HMG2cat/TRP1::rDNA, dpp1, sue*), ALX11-30 (*ura3, trp1, erg9<sup>def</sup>25, HMG2cat/TRP1::rDNA, dpp1, sue*) y aquellas descritas en las patentes de EE.UU. n.º 6.531.303 y 6.689.593 y las publicaciones de patente de EE.UU. publicadas n.º 2004-0249219, 2010-0151519 y 2010-0151555.

La cepa de *Saccharomyces cerevisiae* CALI5-1 es un derivado de SW23B#74 (descrita en las patentes de EE.UU. n.º 6.531.303 y 6.689.593 y Takahashi y col. (2007) (Biotechnol Bioeng. 97(1): 170-181), que derivaba de la cepa de tipo silvestre ATCC 28383 (MATa). Se generó CALI5-1 que tiene una actividad DPP1 fosfatasa disminuida (véase p.ej. la solicitud publicada de EE.UU. n.º 2004-0249219). La cepa de *Saccharomyces cerevisiae* CALI5-1 contiene, entre otras mutaciones, una mutación *erg9* (el alelo  $\Delta$ erg9::HIS3) así como una mutación que apoya la potenciación

de la captación aeróbica de esterol (*sue*). Contiene también aproximadamente 8 copias del gen *HMG2* truncado. La forma activada de HMG2 se activa por el promotor GPD y por lo tanto ya no está bajo una estrecha regulación, permitiendo un aumento del flujo de carbono hasta FPP. Contiene también una delección en el gen que codifica la enzima pirofosfato de diacilglicerol (DGPP) fosfatasa (*dpp*), que limita la desfororilación de FPP.

5

ALX7-95 y ALX11-30.1 son derivados de CALI5-1. ALX7-95 derivaba de CALI5-1 por corrección de la deficiencia  $\Delta leu2$  de CALI5-1 con un gen LEU2 funcional, de modo que no se requiere suplementar leucina a los medios (véase p.ej. el documento US2010/0151519). Se construyó ALX11-30 a partir de CALI5-1 en varias etapas, descritas en el Ejemplo 4 siguiente.

10

#### **b. Cultivo de células**

En procedimientos ejemplares, se expresa una valenceno sintasa proporcionada en la presente memoria en una línea de células hospedadoras que se ha modificado para sobreproducir difosfato de farnesilo, con lo que, tras la expresión de la valenceno sintasa, se convierte difosfato de farnesilo en valenceno. Se cultiva la célula hospedadora usando cualquier procedimiento adecuado bien conocido en la materia. En algunos ejemplos, tales como para cribado de alto rendimiento de células que expresan diversas valenceno sintasas, se cultivan las células que expresan las valenceno sintasas en pocillos individuales de una placa de 96 pocillos (véase, p.ej. el Ejemplo 4 siguiente). En otros ejemplos en que la célula hospedadora es levadura, se cultiva la célula que expresa los polipéptidos de valenceno sintasa y FPP usando procedimientos de fermentación tales como los descritos en los Ejemplos siguientes (véase, p.ej., el Ejemplo 5).

Pueden utilizarse una variedad de metodologías de fermentación para la producción de valenceno a partir de células de levadura que expresan los polipéptidos de valenceno sintasa proporcionados en la presente memoria. Por ejemplo, puede efectuarse la producción a gran escala por fermentación por lotes o continua. Es una fermentación por lotes clásica un sistema cerrado donde se fija la composición del medio al inicio de la fermentación y no se somete a alteraciones artificiales durante la fermentación. Por tanto, al inicio de la fermentación, se inocula el medio con el microorganismo o microorganismos deseados y se permite que ocurra la fermentación sin adición posterior de nutrientes. Típicamente, la concentración de la fuente de carbono en una fermentación por lotes está limitada y se controlan factores tales como pH y concentración de oxígeno. En sistemas por lotes, las composiciones de metabolito y biomasa del sistema cambian constantemente hasta el momento en que se detiene la fermentación. En los cultivos por lotes, las células se modulan típicamente mediante una fase de latencia estática hasta una fase de alto crecimiento logarítmico y finalmente una fase estacionaria donde la tasa de crecimiento disminuye o se detiene. Si no se tratan, las células en la fase estacionaria eventualmente morirán.

35

Es una variación del sistema por lotes estándar el sistema por lotes alimentado, que es similar a un sistema por lotes estándar con la excepción de que se añaden nutrientes a medida que progresa la fermentación. Los sistemas por lotes alimentados son útiles cuando la represión del catabolito tiende a inhibir el metabolismo de las células y donde sea deseable tener cantidades limitadas de sustrato en el medio. También la capacidad de alimentar nutrientes dará a menudo como resultado mayores densidades celulares en los procesos por lotes alimentados en comparación con procesos de fermentación por lotes. Generalmente, se miden y controlan factores tales como pH, oxígeno disuelto, concentraciones de nutriente y la presión parcial de gases de desecho tales como CO<sub>2</sub> en las fermentaciones por lotes alimentadas.

La producción de valenceno puede lograrse también con fermentación continua. La fermentación continua es un sistema abierto donde se añade continuamente un medio de fermentación definido a un biorreactor y se retira una cantidad igual de medio acondicionado simultáneamente para procesamiento. El sistema mantiene generalmente los cultivos a una densidad alta constante donde las células están principalmente en su fase de crecimiento logarítmico. La fermentación continua permite la modulación de cualquier número de factores que afecten al crecimiento celular o a la concentración de producto final. Por ejemplo, un procedimiento mantendrá un nutriente limitante tal como la fuente de carbono o el nivel de nitrógeno a una tasa fija y permitirá moderar todos los demás parámetros. En otros sistemas, pueden alterarse continuamente una serie de factores que afectan al crecimiento mientras que la concentración celular, medida por la turbidez del medio, permanece constante. Los sistemas continuos aspiran a mantener condiciones de crecimiento en estado estacionario, y por tanto la pérdida celular debida a la retirada de medio debe equilibrarse frente a la tasa de crecimiento celular en la fermentación. Los procedimientos de modulación de nutrientes y factores de crecimiento para procesos de fermentación continua, así como las técnicas para maximizar la tasa de formación de producto, son bien conocidos en la materia. Después del cultivo celular, puede recolectarse entonces el cultivo celular para obtener el valenceno producido.

En un procedimiento ejemplar, se hacen crecer las células hospedadoras que expresan los polipéptidos de valenceno sintasa (p.ej. *Saccharomyces cerevisiae* cepa CALI51, ALX7-95 o ALX11-30) en tanque de fermentación

de 3 l a 28 °C, pH 4,5, durante aproximadamente 132 horas, manteniendo la glucosa a entre 0 y 1 g/l. Después de la fermentación, se añade sulfato de sodio a una concentración final de 10-15 %. Se añade también aceite de soja y se agita, y se recupera el aceite que contiene el valenceno (y otros terpenos) por centrifugación.

### 5 c. Aislamiento y valoración de productos

- El valenceno producido usando los procedimientos anteriores con los polipéptidos de valenceno sintasa proporcionados en la presente memoria puede aislarse y valorarse mediante cualquier procedimiento conocido en la materia. En un ejemplo, se extrae el medio de cultivo celular con un disolvente orgánico para repartir el valenceno, y cualquier otro terpeno producido, en la fase orgánica. La producción de valenceno puede valorarse y/o aislarse el valenceno de otros productos usando cualquier procedimiento conocido en la materia tal como, por ejemplo, cromatografía de gases. Por ejemplo, la fase orgánica puede analizarse por cromatografía de gases usando cedreno y hexadecano como patrones internos. Este procedimiento se ejemplifica en los Ejemplos 3 y 4 siguientes.
- 15 La cantidad de valenceno producida puede determinarse mediante cualquier técnica cromatográfica estándar conocida útil para separar y analizar compuestos orgánicos. Por ejemplo, la producción de valenceno puede ensayarse mediante cualquier técnica cromatográfica conocida útil para la detección y cuantificación de hidrocarburos, tales como valenceno y otros terpenos incluyendo, pero sin limitación, cromatografía de gases acoplada con espectrometría de masas (GC-MS), cromatografía de gases usando un detector de ionización de llama (GC-FID), GC-MS capilar, cromatografía líquida de alta resolución (HPLC) y cromatografía en columna. Típicamente, se llevan a cabo estas técnicas en presencia de patrones internos conocidos, por ejemplo cedreno o hexadecano, que se usan para cuantificar la cantidad de terpeno producida.

- Por ejemplo, los terpenos, incluyendo sesquiterpenos tales como valenceno, pueden identificarse por comparación de los tiempos de retención y los espectros de masas con los de patrones auténticos en cromatografía de gases con detección de espectrometría de masas. Los patrones típicos incluyen, pero sin limitación, cedreno y hexadecano. En otros ejemplos, puede conseguirse la cuantificación por cromatografía de gases con detección por ionización de llama basada en curvas de calibración con cantidades conocidas de patrones auténticos y normalización al área de pico de un patrón interno. Estas técnicas cromatográficas permiten la identificación de cualquier terpeno presente en la fase orgánica incluyendo, por ejemplo, otros terpenos producidos por la valenceno sintasa modificada incluyendo, por ejemplo,  $\beta$ -elemeno y aristoloqueno (pico 2) y otros compuestos producidos por la células, incluyendo nerolidol y farnesol (véase, p.ej. el Ejemplo 4).

- En algunos ejemplos, puede determinarse la cinética de la producción de valenceno por ensayos de sintasa en que se utilizan sustratos isoprenoides radiactivos, tales como  $^3\text{H}$  FPP o  $^{14}\text{C}$  FPP, con concentraciones variable de sintasa. Los productos se extraen en una fase orgánica y se mide la radiactividad usando un contador de centello líquido. Se determinan las constantes cinéticas a partir de ajustes directos de la ecuación de Michaelis-Menton a los datos.

### 40 2. Producción de nootkatona

- La nootkatona, que es al aroma dominante de la uva, es un producto oxidado del valenceno. Las valenceno sintasas proporcionadas en la presente memoria catalizan la producción de valenceno, que puede oxidarse entonces hasta nootkatona. El valenceno puede experimentar hidroxilación regioselectiva formando 2-hidroxivalenceno, que se oxida adicionalmente formando nootkatona. La oxidación de valenceno puede llevarse a cabo mediante medios químicos o biosintéticos (véanse, p.ej. la patente de EE.UU. nº 5.847.226, Eur. Pat. nº EP1083233; Girhard y col., (2009) Microb. Cell. Fact. 8: 36; Fraatz y col., (2009) Appl Microbiol Biotechnol. 83(1): 35-41; Furusawa y col. (2005) Chem Pharm. Bull. 53: 1513-1514; Salvador y col., (2002) Green Chemistry, 4, 352-356). La oxidación bioquímica puede efectuarse por una lacasa, hidroxilasa u otra enzima oxidativa. En algunos ejemplos, se convierte valenceno en nootkatona usando trióxido de cromo o un catalizador de cromo (III) inmovilizado en fosfonato de sílice (como se describe, por ejemplo en la publicación de patente de EE.UU. . nº 2012-0246767). La formación de nootkatona puede confirmarse y/o cuantificarse mediante cualquiera de las técnicas cromatográficas descritas en la presente memoria.

### F. EJEMPLOS

- 55 Los siguientes ejemplos se incluyen solo con fines ilustrativos y no se pretende que limiten el alcance de la invención. Los ejemplos que caen fuera del alcance de las reivindicaciones se proporcionan solo con fines informativos.

### 60 Ejemplo 1

### Extracción e identificación de valenceno de *Eryngium glaciale*

Para confirmar que *Eryngium glaciale* produce valenceno, se ensayó en extractos de hojas de *Eryngium glaciale* la presencia de valenceno. Se adquirieron las hojas y ramas de *Eryngium glaciale* en el vivero alpino Wrightman (Ontario, Canadá). Se diseccionaron varias hojas y se sumergieron en acetato de etilo durante varios días. Se analizó posteriormente el aceite vegetal en acetato de etilo por cromatografía de gases-espectrometría de masas (GC-MS). Se practicó el análisis de GC-MS con el PerkinElmer AutoSystem XL equipado con el espectrómetro de masas TurboMass™. Se muestra el cromatógrafo de gases del extracto de aceite vegetal en la Figura 2A. Se detectó valenceno aproximadamente a los 10,18 minutos (pico 1) y era el pico principal. Se muestra en la Figura 2B el espectro de masas del pico a los 10,18 minutos (valenceno).

### Ejemplo 2

#### Identificación del ácido nucleico que codifica valenceno sintasa de *Eryngium glaciale*

Se identificó valenceno sintasa de *Eryngium glaciale* por secuenciación del transcriptoma total de *Eryngium glaciale* usando pirosecuenciación 454 de alto rendimiento (Roche Diagnostics, Branford, CT) y buscando homología con terpeno sintasas conocidas.

#### 20 Purificación de ARN

Se aisló ARN de un gramo de material vegetal, se ultracongeló en nitrógeno líquido y se molió posteriormente hasta un polvo usando mortero y almirez. Se descongeló la muestra en polvo en 10 ml de reactivo TRIzol® (Invitrogen Corp., Carlsbad, CA) para extraer ARN del tejido, siguiendo las instrucciones del fabricante con modificaciones menores. Se centrifugaron la mezcla de muestra en polvo y TRIzol® a 13.000 rpm durante 10 minutos a 4 °C para retirar los desechos celulares. Se transfirió el sobrenadante a tubos nuevos, se añadió cloroformo:alcohol isoamílico (24:1) y se agitó con vórtex la muestra. Después de centrifugación adicional, se añadieron dos volúmenes de isopropanol y se incubó la mezcla durante 10 minutos a temperatura ambiente. Se precipitó ARN por centrifugación a 13.000 rpm durante 10 minutos a 4 °C. Se lavó el ARN precipitado una vez con etanol al 75 % y se resuspendió en 20 µl de agua exenta de ARNasa. Se trató además el ARN con ADNasa I (New England Biolabs, Ipswich, MA) para retirar el ADN y se purificó y concentró posteriormente usando el kit de limpieza de ARN Qiagen MinElute (Qiagen, Hilden, Alemania).

#### Secuenciación del transcriptoma total

Se secuenció el transcriptoma total de *Eryngium glaciale* usando pirosecuenciación 454. Se remitió el ARN total extraído al Centro de Tecnología Genética Avanzada (AGTC) de la Universidad de Kentucky (UK) para análisis de secuenciación de transcriptoma. El conjunto de un total de 5512221 lecturas generó 15717 cóntigos, con un tamaño medio de cóntigo que era de 847 pares de bases. Se buscó en la base de datos resultante con BLAST frente a las secuencias de varias terpeno sintasas conocidas. Estas sintasas incluían: Valenceno sintasa de *Perilla frutescens* var. *frutescens* (SEQ ID NO:39; n° de acceso a GenBank AY917195.1); valenceno sintasa de *Citrus sinensis* (SEQ ID NO:40; n° de acceso a GenBank AF441124.1); vetispiradieno (o premnaspirodieno) de *Hyoscyamus muticus* (SEQ ID NO:41; n° de acceso a GenBank U20188.1); germacreno D sintasa de *Citrus hystrix* (ácido nucleico expuesto en la SEQ ID NO:42; n° de acceso a GenBank HQ652871.1; proteína expuesta en la SEQ ID NO:45) y valenceno sintasa de *Chamaecyparis nootkatensis* (SEQ ID NO:44). La homología entre estas oscilaba de aproximadamente 35 a 50 % de identidad al nivel nucleotídico. Se buscó en la base de datos BLAST generada usando las cinco secuencias. Se identificaron cinco secuencias parciales diferentes que codifican terpeno sintasas candidatas a partir de la base de datos generada para búsquedas BLAST. Después de la amplificación (descrita en el Ejemplo 3 para el gen que codifica la sintasa activa), se produjeron moléculas de ADN que codifican cada uno de los cinco posibles genes. Un gen, descrito en el Ejemplo 3, codificaba una sintasa activa que cataliza la producción de valenceno.

### Ejemplo 3

#### 55 Aislamiento de un gen que codifica la valenceno sintasa de *Eryngium glaciale*

Basándose en la secuencia de cóntigo seleccionada (SEQ ID NO:4), se diseñaron cebadores para la amplificación de los fragmentos de RACE 3' y 5' de tal modo que hubiera aproximadamente una superposición de 150 pb entre los fragmentos de RACE 5' y 3'. Se practicaron las síntesis de ADN complementario (ADNc) y PCR como se describe en los protocolos del kit (kit de amplificación de ADNc por RACE SMARTer™ (Clontech, Mountain View, CA, EE.UU.)) con algunas modificaciones de las condiciones de PCR. Se clonaron los fragmentos de PCR por RACE 5' y 3' en el

vector TOPO-TA (Invitrogen, San Diego) y se secuenció cada producto de amplificación.

Se usaron las siguientes condiciones de PCR para amplificar los fragmentos de RACE 3' y 5': desnaturalización inicial a 94 °C durante 2 minutos, 10 ciclos de 94 °C durante 30 segundos, reasociación a 61 °C durante 45 segundos, extensión a 72 °C durante 1 minuto y 20 segundos, seguido de 25 ciclos a 94 °C durante 30 segundos, reasociación a 55 °C durante 40 segundos, extensión a 72 °C durante 1 minuto y 20 segundos. Los cebadores usados para amplificar el gen que codifica la valenceno sintasa activa se exponen en la Tabla 3 siguiente y se incluyen como cebador directo 63-1-2-Fwd2 (SEQ ID NO:5) y cebador inverso 63-1-2-Rev2 (SEQ ID NO:6). Se usaron los cebadores anidado directo 63-1-2-NestFwd2 (SEQ ID NO:7) y anidado inverso 63-1-2-NestRev2 (SEQ ID NO: 8) para aumentar la especificidad de los productos de amplificación en la PCR.

<b>Tabla 3 Cebadores para RACE-PCR</b>		
<b>Cebador</b>	<b>Secuencia</b>	<b>SEQ ID NO</b>
63-1-2-Fwd2	GCTAGCTCATGTTGATACATTTTCTGCAGTCG	5
63-1-2-Rev2	GCTGTCAGTAACAACCTCTCGTCTTGAGC	6
63-1-2-NestFwd2	CGGTTAAAATCTAGTTTTGCAAACCTCCAAAAGC	7
63-1-2-NestRev2	CCCACGTGAAGTACCCTCTTGAGGACG	8

La secuencia nucleotídica completa del gen, designado EGVS y confirmado posteriormente que codifica una valenceno sintasa (véase el Ejemplo 3 siguiente), se expone en la SEQ ID NO:2 y la secuencia aminoacídica codificada se expone en la SEQ ID NO: 1.

#### Ejemplo 4

#### 20 Clonación de EGVS en el vector de expresión pAlx31-108.2

Se clonó el gen EGVS completo aislado en el vector de expresión pAlx31-108.2 (SEQ ID NO:32). Se amplificó el gen EGVS usando dos conjuntos de cebadores (véase la Tabla 4 siguiente) para retirar el sitio de restricción KpnI en la posición 964 de la secuencia codificante que incluye el cebador directo 63-52EG2FwdPart1 (SEQ ID NO:9) y el cebador inverso 63-52-EG2RevPart1 (SEQ ID NO:10); el cebador directo 63-52-EG2FwdPart2 (SEQ ID NO:11) y el cebador inverso 63-52-EG2RevPart2 (SEQ ID NO:12).

<b>Tabla 4 Cebadores</b>		
<b>Cebador</b>	<b>Secuencia</b>	<b>SEQ ID NO</b>
63-52-EG2FwdPart1	GCTGAATTTCGAGCTCGGTACCATTAAAAAAAATGTCTCT TAATGTACTTAGTACGTCAGG	9
63-52-EG2RevPart1	GAAGTTCTTTGAATGTACCATACACATCATACG	10
63-52-EG2FwdPart2	GATGTGTATGGTACATTCAAAGAACTTCTACTGTTCACT G	11
63-52-EG2RevPart2	TACGCGCACAAAAGCAGAGATTCTAGATTACAAAGGAAT AGGATCCACGAGCAGTG	12

Se retiró un sitio de restricción KpnI que empieza en la posición 964 de la secuencia codificante, ya que los sitios KpnI y XbaI se usan frecuentemente para clonar genes en el vector de expresión pAlx31-108.2. Se alteró el sitio KpnI sin afectar a la secuencia aminoacídica de EGVS, reemplazando la C en posición 969 del gen EGVS de tipo silvestre por una A, cambiando así el codón ACC (que codifica treonina) por un ACA (que codifica treonina) (secuencia nucleotídica expuesta en la SEQ ID NO:3). Se construyó el vector de expresión de EGVS final, designado pAlx63-71 (SEQ ID NO:33) usando la mezcla maestra Gibson Assembly™ (New England Biolabs) con los productos de PCR amplificados usando los cebadores 63-52-EG2FwdPart1, 63-52-EG2RevPart1, 63-52-EG2FwdPart2 y 63-52-EG2RevPart2 (véase la Tabla 4 anterior).

#### Ejemplo 5

#### 40 Expresión de EGVS en levadura y cribado de alto rendimiento de células que expresan valenceno

Se transformó el vector de expresión pAlx63-71 que codifica EGVS en las cepas de *Saccharomyces cerevisiae* modificadas ALX7-95 (ura3, his3, trp1,  $\Delta$ erg9::HIS3, HMG2cat/TRP1::rDNA, dpp1, sue) y ALX11-30 (ura3, trp1, erg9def25, HMG2cat/TRP1::rDNA, dpp1, sue) para expresión de valenceno sintasa y producción de valenceno usando un ensayo de cribado de alto rendimiento.

#### A. Generación de las cepas de *Saccharomyces cerevisiae* modificadas ALX7-95 y ALX7-95

Como se señala anteriormente, se derivaron ALX7-95 y ALX11-30 de la cepa de *Saccharomyces cerevisiae* CALI5-1, que es una cepa bien conocida y difundida (véase, p.ej., la solicitud de patente de EE.UU. nº 2012-0246767). La cepa CALI5-1 es un derivado de la cepa designada SW23B#74 (descrita, por ejemplo, en las patentes de EE.UU. nº 6.531.303 y 6.689.593; véase también Takahashi y col. (2007) *Biotechnol Bioeng.* 97(1): 170-181), que deriva de la cepa de tipo silvestre MATa, depositada con el número de acceso ATCC 28383. Se generó CALI5-1 que tiene una actividad DPP1 fosfatasa disminuida (véase, p.ej., la publicación de patente de EE.UU. nº 2004-0249219). La cepa CALI5-1 de *Saccharomyces cerevisiae* contiene, entre otras mutaciones, una mutación erg9 (el alelo  $\Delta$ erg9::HIS3) así como una mutación que apoya la potenciación de la captación aeróbica de esterol (sue). Contiene también aproximadamente 8 copias del gen HMG2 truncado. La forma activada de HMG2 se activa por el promotor GPD y por lo tanto ya no está bajo una estrecha regulación, permitiendo un aumento del flujo de carbono hasta FPP. Contiene también una delección en el gen que codifica la enzima pirofosfato de diacilglicerol (DGPP) fosfatasa (dpp1), que limita la desfororilación de FPP.

Se generó ALX7-95 a partir de CALI5-1 corrigiendo la deficiencia  $\Delta$ leu2 de CALI5-1 con un gen LEU2 funcional de modo que no se requiera el suplemento de leucina (véase, p.ej. el documento US2010/0151519). Se construyó ALX11-30 a partir de CALI5-1 en varias etapas a partir de una cepa designada ALX7-175.1 que se describe en la publicación de patente de EE.UU. nº 2010-0151519.

Brevemente, se produjo ALX7-175.1 como sigue. Se obtuvo HPS de ALX7-95 transformando un plásmido que contiene la prenaspirodieno sintasa (HPS) de *Hyoscyamus muticus* en la cepa ALX7-95. Se obtuvo el plásmido YEp-HPS clonando el gen de HPS en Yep-GW-URA, dando YEp-HPS-ura (YEp-HPS). Se practicó entonces una reacción de PCR propensa a errores del gen ERG9 y se transformó el ADN resultante en ALX7-95 que alberga YEpHPS. Se sembraron los transformantes en medio YP que carece de ergosterol y se cribó la producción de prenaspirodieno. Se guardaron aquellos que producían altos niveles de prenaspirodieno. Se transformó una cepa, ALX7-168.25 [ura3, trp1, his3, erg9<sup>del</sup>25, HMG2cat/TRP1::rDNA, dpp1, sue, YEpHPS] con un fragmento de PCR del gen HIS3 completo para crear un gen HIS3 funcional. Se aislaron los transformantes que eran capaces de crecer en ausencia de histidina en el medio. A partir de esta transformación, se aisló ALX7-175.1 [ura3, trp1, erg9def25, HMG2cat/TRP1::rDNA, dpp1, sue YEpHPS]. Finalmente, se retiró el plásmido YEpHPS haciendo crecer ALX7-175.1 varias generaciones en YPD (extracto de levadura 10 g/l, peptona 20 g/l, glucosa 20 g/l) y sembrando células en placas de YPD. Se identificaron las colonias que eran incapaces de crecer en medio SD sin uracilo (0,67 % de base nitrogenada de levadura Bacto sin aminoácidos, 2 % de glucosa, 0,14 % de medio defectivo sintético de levadura sin uracilo). Se designó esta cepa ALX11-30.

#### B. Cribado de alto rendimiento para la producción de valenceno

Se analizó en ocho colonias de cada transformación del vector pAlx63-71 en ALX7-95 y ALX11-30 la producción de sesquiterpeno en placas de microvaloración de pocillos profundos. Se cribó en los transformantes la producción de valenceno mediante análisis de microcultivo usando placas de 96 pocillos profundos. Se inocularon transformantes individuales en pocillos individuales de placas de microvaloración de 96 pocillos rellenos con 200  $\mu$ l de medio definido sintético (SD: 0,67 % de base nitrogenada de levadura Bacto sin aminoácidos, 2 % de glucosa y 0,14 % de suplemento de medio defectivo sintético de levadura sin histidina, leucina, triptófano o uracilo) para las cepas ALX11-30 y medios definidos sintéticos con ergosterol (SDE: 0,67 % de base nitrogenada de levadura Bacto sin aminoácidos, 2 % de glucosa, 0,14 % de suplemento de medio defectivo sintético de levadura sin histidina, leucina, triptófano o uracilo y ergosterol 40 mg/l) para las cepas ALX795 deficientes en erg9. Se incubaron las placas durante 2 a 3 días a 28 °C. Después del crecimiento hasta saturación, se usaron 10  $\mu$ l del cultivo saturado de cada pocillo para inocular una placa de 96 pocillos profundos que contiene 300  $\mu$ l de medio adecuado para el crecimiento y la producción de valenceno. Se selló la placa con cinta Airpore (Qiagen) y se incubó con agitación durante 3 días.

Se extrajeron los productos biosintéticos en primer lugar mediante la adición de 250  $\mu$ l de acetona que contiene cedreno en cada pocillo, sellando con sellador de placa de caucho y agitando con vórtex, seguido de la adición de 500  $\mu$ l de n-hexano que contiene hexadecano y agitación con vórtex adicional. Después de la separación de fases, se selló la placa con película de aluminio y se colocó en una bandeja de muestras de un automuestreador de cromatografía de gases, que retiraba un microlitro de la fase orgánica por cada pocillo para análisis de

sesquiterpenos. Se adicionaron a cada uno de la acetona y el hexano usados para extracción los patrones internos cedreno y hexadecano, respectivamente, para ayudar a la cuantificación de las muestras. Se analizaron las muestras extraídas por cromatografía de gases y se calculó la cantidad de valenceno a partir del área bajo el pico que representa valenceno.

5

### C. Resultados

La Tabla 5 siguiente proporciona los perfiles de distribución de producto y cantidad de valenceno producidos por valenceno sintasa de *Eryngium glaciale*, como se determina por cromatografía de gases, para los tres mayores  
10 productores de valenceno a partir de ALX7-95 y los cuatro mayor productores de valenceno a partir de ALX11-30, y sus denominaciones de nombre de cepa. Como se indica en la Tabla 5, se observaron diversos compuestos adicionales además de valenceno, incluyendo nerolidol, farnesol y el compuesto aristoloqueno, en el pico 2, que es aristoloqueno. El resultado demuestra que esta valenceno sintasa cataliza la producción de un perfil de producto  
15 diferente de la valenceno sintasa cítrica. Cedreno y hexadecano representan patrones internos usados para cuantificar la cantidad de valenceno producida. Se analizaron además las 7 cepas en el ensayo de matraz agitado descrito en el Ejemplo 4.

**Tabla 5 Distribución de producto para el ensayo de microcultivo**

Cepa de levadura	Área de pico						Valenceno mg/ml	Nombre de cepa
	Cedreno	Aristologueno	Valenceno	Nerolidol	Hexadecano	Farnesol		
7-95	15794	23576	75056	6718	15085	7475	120,10	Alx-63-70.1
	14916	18809	59731	6057	14930	7057	101,20	Alx-63-70.2
	14443	19245	61326	5368	13706	6378	107,31	Alx-63-70.3
11-30	14151	26590	80470		15281		143,71	Alx-63-70.4
	14491	33936	107516		16152		187,51	Alx-63-70.5
	14428	33512	105465		15969		184,73	Alx-63-70.6
	16093	19851	60353		17711		94,78	Alx-63-70.7

**Ejemplo 6****5 Ensayo de matraz agitado de células que expresan valenceno**

Se iniciaron cultivos de semilla en matraces de 250 ml inoculando 15 ml de medio SD o SDE (para cepas ALX11-30 y ALX7-95, respectivamente) con colonias de crecimiento reciente de las 7 cepas mostradas en la Tabla 5 anterior).

Se hicieron crecer los cultivos durante 24 horas, y se usaron 2,5 ml de cada cultivo para inocular 50 ml de medio de fermentación (2 % de sulfato de amonio, 2 % de fosfato de potasio, 0,1 % de NaCl, 0,6 % de MgSO<sub>4</sub>•7H<sub>2</sub>O, 0,4 % de extracto de levadura, 1 ml de solución mineral [0,028 % de FeSO<sub>4</sub>•7H<sub>2</sub>O, 0,029 % de ZnSO<sub>4</sub>•7H<sub>2</sub>O, 0,008 % de CuSO<sub>4</sub>•5H<sub>2</sub>O, 0,024 % de Na<sub>2</sub>MoO<sub>4</sub>•2H<sub>2</sub>O, 0,024 % de CoCl<sub>2</sub>•6H<sub>2</sub>O, 0,017 % de MnSO<sub>4</sub>•H<sub>2</sub>O, HCl 1 ml], 0,5 ml de glucosa al 50 %, 1,5 ml de solución de vitamina [0,001 % de biotina, 0,012 % de pantotenato de Ca, 0,06 % de inositol, 0,012 % de HCl de piridoxina, 0,012 % de HCl de tiamina] y 0,5 ml de CaCl<sub>2</sub> al 10 %) en un matraz con deflectores de 250 ml. Se hicieron crecer los cultivos a 28 °C. Después de 16 horas de incubación, se alimentaron a los cultivos 3,6 ml de glucosa al 50 % y 0,667 ml de extracto de levadura al 12,5 %. Se alimentaron los cultivos cada 24 h después de la alimentación inicial. Se ajustó el pH de los cultivos a 4,5 cada 24 horas con la adición de NaOH al 30 %. Después de aproximadamente 88 horas de incubación, se añadieron 0,1 ml de IGEPAL CA-630 y se incubó el cultivo con agitación para dispersar completamente el aceite vegetal. Después de 30 minutos, se tomó una muestra de cultivo de 2 ml para análisis. Se extrajo la muestra con 2 ml de solución de acetona/cedreno y se extrajo entonces con 4 ml de solución de hexano/hexadecano. Se analizó una alícuota de la fase orgánica por cromatografía de gases y se cuantificó la cantidad de valenceno producida calculando el área bajo el pico que representa valenceno. Se compararon los perfiles de distribución de producto y la producción de valenceno para cada una de las 2 cepas que contienen EGVS con una valenceno sintasa cítrica (CVS) de control y una valenceno sintasa V277 (descrita en la publicación de EE.UU. en tramitación junto con la presente de n° de serie 2012-0246767) que tiene una secuencia de nucleótidos expuesta en la SEQ ID NO:30 y una secuencia aminoacídica expuesta en la SEQ ID NO:31. V277 tiene aproximadamente un 83 % de identidad de secuencia con la CVS de tipo silvestre. Las células que expresan V277 producen más valenceno que las células que expresan CVS de tipo silvestre (secuencia aminoacídica expuesta en la SEQ ID NO:14).

La Tabla 6a siguiente proporciona perfiles de distribución de producto que incluyen la cantidad de valenceno producida a partir del ensayo de matraz agitado, como se determina por cromatografía de gases, de CVS 277 en comparación con dos cepas que producen EGVS en la cepa ALX11-30 (como se describe en el Ejemplo 3). Como se muestra en la Tabla 6, las cepas de valenceno sintasa de *Eryngium glaciale* producían más valenceno por ml en comparación con el mutante CVS V277, que a su vez produce más valenceno que la CVS de tipo silvestre.

Cedreno y hexadecano representan patrones internos usados para cuantificar la cantidad de valenceno producida. La Figura 3A muestra un cromatograma de gases ejemplar de una cepa ALX7-95 ejemplar, Alx-64-703, teniendo el valenceno un tiempo de retención de 12,46 min, y la Figura 3B muestra el espectro de masas del valenceno (pico a 12,46 min). La Figura 7 representa los resultados gráficamente.

<b>Tabla 6 Distribución de producto para el ensayo de matraz agitado</b>					
Cepa de levadura	Área de pico			Valenceno mg/ml	Nombre de cepa
	Cedreno	Valenceno	Hexadecano		
ALX11-30	13743	2565260	15877	356,1	YC33-11
	9770	378154	18747	714,2	Alx-63-70.6
	11179	472779	19251	780,3	Alx-63-70.7

La Figura 7 representa los datos de la Tabla 6b y muestra una comparación entre la producción de valenceno en células que expresan la valenceno sintasa cítrica con células que expresan la sintasa de *Eryngium glaciale*. Las células que expresan EG sintasa producen más de dos veces de cantidad de valenceno en comparación con la CVS sintasa.

## Ejemplo 7

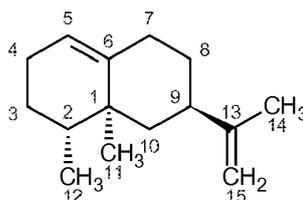
### Aislamiento de valenceno

5

Se filtraron a vacío los caldos de dos fermentaciones de ALX63-70.7 (2 l cada una) a través de un embudo de filtración grueso. Se transfirió la torta de filtrado (aceite de soja, células y desechos celulares) a un matraz de 4 l y se añadió acetona (2 l cada vez) a la torta de filtración y el filtrado (aceite de soja y células), respectivamente, y se agitaron las soluciones durante 1 h. Se filtraron las mezclas heterogéneas resultantes y se retiró el disolvente a presión reducida. Se sometió entonces el extracto orgánico bruto a cromatografía en gel de sílice usando hexanos como único disolvente. Después de la retirada del disolvente a presión reducida, se disolvió el material oleoso amarillo en acetona y se precipitaron los fosfolípidos a -20 °C. Se filtraron los fosfolípidos y se retiró la acetona en el filtrado a presión reducida, procurando un aceite transparente (25 g). Se purificó adicionalmente entonces el aceite transparente mediante destilación de película descendente (0,1 torr, 125 °C), procurando una fracción de valenceno enriquecida (3,13 g). Se purificó entonces el valenceno a partir de la fracción enriquecida por HPLC en fase inversa (C18) usando acetonitrilo isocrático como disolvente y un detector del índice de refracción.

Las resonancias magnéticas (RMN) de protón ( $^1\text{H}$ ) y carbono 13 ( $^{13}\text{C}$ ) de la muestra purificada mostraron que la muestra purificada tenía idénticos desplazamientos químicos que un patrón auténtico de valenceno (véanse los datos de RMN siguientes). Además, los experimentos de 1D-NOESY donde se irradió el grupo metilo cuaternario (C11) mostraron correlaciones con la metina en C9 y el grupo metilo en C12, sugiriendo por tanto que los grupos metilo de C11 y C12 y la metina de C9 estaban todos en la misma cara de la molécula. Se muestra a continuación la estructura del valenceno con la numeración de carbono asignada:

25



Finalmente, la cromatografía de gases (GC-FID), cromatografía de gases quiral y cromatografía de gases-espectrometría de masas de la muestra purificada mostraron que el presunto pico de valenceno tenía idénticos tiempos de retención y espectros de masas que un patrón auténtico de valenceno. Estos datos demuestran conjuntamente que la estructura del pico aislado es (+)-valenceno. RMN  $^1\text{H}$  (BENCENO- $d_6$ )  $\delta$ : 5,36 (dt,  $J = 4,8, 2,6$  Hz, 1H), 4,83 (s, 1H), 4,80 (s, 1H), 2,23-2,31 (m, 1H), 2,21 (s, 1H), 2,06 (ddd,  $J = 13,9, 4,1, 2,6$  Hz, 2H), 1,90-1,96 (m, 2H), 1,69 (d,  $J = 0,9$  Hz, 1H), 1,66 (s, 3H), 1,31-1,46 (m, 2H), 1,26 (d,  $J = 4,4$  Hz, 2H), 1,05 (t,  $J = 12,7$  Hz, 1H), 0,91 (s, 3H), 0,83 (d,  $J = 6,6$  Hz, 3H); RMN  $^{13}\text{C}$  (BENCENO- $d_6$ )  $\delta$ : 150,8, 143,3, 120,9, 109,3, 45,7, 41,6, 41,7, 38,5, 33,8, 33,5, 27,9, 26,7, 21,3, 18,9, 16,2. GC-MS ( $\text{EI}^+$ , 70 eV); (%): 55(44), 67(38), 79(82), 91(80), 93(77), 105(75), 107(73), 119(66), 133(56), 147(33), 161(100), 175(17), 189(34), 204(29).

## Ejemplo 8

### Comparación de los parámetros cinéticos de valenceno sintasa de *E. glaciale* (EgVS) y valenceno sintasa V277

En este ejemplo, se evaluó la actividad de valenceno sintasa de *E. glaciale* (EgVS; SEQ ID NO:1) y de la valenceno sintasa cítrica denominada V277 (SEQ ID NO:31) por análisis cinético del estado estacionario. Para comparar las actividades de estas enzimas *in vitro*, se marcaron con histidina las valenceno sintasas en el extremo C, se expresaron en *E. coli* se purificaron por cromatografía de quelato de níquel y se evaluó la actividad por análisis cinético en estado estacionario. Las condiciones de reacción eran las siguientes: EgVS 50 nM o valenceno sintasa V277 250 nM, Bis-Tris-propano 50 mM (pH 7,5),  $\text{MgCl}_2$  20 mM, KCl 50 mM. T= 30 °C. Detección: GC/MS-SIM.

Como se muestra en la Figura 6 y la Tabla 7 siguiente, EgVS exhibe una actividad total y específica drásticamente mayor cuando se compara directamente con la valenceno sintasa cítrica designada V277. Además de una actividad aumentada, la EgVS se recambiaba también significativamente más veces que la valenceno sintasa V277.

Tabla 7 Parámetros cinéticos en estado estacionario de CVS y EgVS

Enzima (purificada por marcaje de His)	$K_{m, FPP}$ ( $\mu\text{M}$ )	$k_{cat}$ ( $\text{min}^{-1}$ )	$V_{m\acute{a}x}$ ( $\mu\text{mol (g de prote\acute{i}na)^{-1} \text{min}^{-1}$ )
Valenceno sintasa V277	7,8	1,5	23,4
EgVS	21,4 $\pm$ 10,5	25,2 $\pm$ 3,8	376 $\pm$ 57

La EgVS sintasa proporcionada en la presente memoria, por tanto, no solo produce un perfil de terpeno diferente en comparación con otras sintasas tales como CVS, sino que también exhibe mayor actividad, incluso cuando se compara con una CVS sintasa que se ha modificado y optimizado para exhibir una actividad aumentada.

Puesto que las modificaciones resultarán evidentes para los especialistas en la materia, se pretende que esta invención esté limitada solo por el alcance de las reivindicaciones adjuntas.

## 10 LISTADO DE SECUENCIAS

- <110> Allylix, Inc.  
Saran, Dayal  
Park, Grace Eunyong
- 15 <120> POLIPÉPTIDOS DE VALENCENO SINTASA,  
MOLÉCULAS DE ÁCIDO NUCLEICO QUE LOS CODIFICAN Y USOS DE LOS MISMOS
- <130> 33322-237PC
- <140> Todavía no asignado
- <141> Con la presente
- 20 <150> 61/852.462
- <151> 14/03/2013
- <160> 45
- <170> FastSEQ para Windows versión 4.0
- <210> 1
- 25 <211> 565
- <212> PRT
- <213> Eryngium glaciale
- <220>
- <223> Proteína valenceno sintasa
- 30 <400> 1

ES 2 647 828 T3

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



ES 2 647 828 T3

ggagcgagat attacatatic cattedacgaa aaatttgatt cacacaataa attgcttttg 720  
 gaggttgcaa aactagatt taaccgactg cagaaaatgt atcaacatga gctagcccac 780  
 cttacaaggt ggtggaaga tttagatttt acaacaacac ttccatttgc aagagataga 840  
 attggtgagg gttacttttg gatcttagga atgtactttg agccagaacg taaggatgtc 900  
 aggaattct tgaacagagt atttgcactt attacagtag ttgatgacac gtatgatgtg 960  
 tatggtacct tcaaagaact tctactgttc actgatgcaa ttgaaagatg ggaactagt 1020  
 gatttggatc agctaccggg atatatgaga attatttacc aagctctcat ggatgtttat 1080  
 aatcaaatgg aggaaaagtt gtcaatgaaa gctgattgtc caacataccg tcttgagttt 1140  
 gcaatagaaa cagttaaagc catgttcaga tcatacctcg aagaagctag atggtccaaa 1200  
 gaacattata tcccatcgat ggaagagtat atgaccgtgg cactggtatc ggttggctac 1260  
 aaaaccatat taactaattc ctttgttggg atgggggata ttgcaacacg ggaagttttt 1320  
 gattgggtgt tcaatagtcc attgattatt agagcttccg acttaattgc cagattggga 1380  
 gatgatattg gaggccatga ggaggagcag aagaaaggag acgcagccac tgctatcgag 1440  
 tgttacataa aagagaatca tgaacaaaag catgaagctt atgatgaatt tcagaaacaa 1500  
 atgataatg cttggaagga tttgaataag gaagctctac gtccatttcc tgttccaatg 1560  
 actttcatca caagagtgt tcattttacg cgcgccatcc atgttattta tgccgacttt 1620  
 agtgatggtt acacacgttc agacaaggcg atcagaggtt acataacttc actgctcgtg 1680  
 gatcctattc ctttghtaa 1698

<210> 3

<211> 1698

5 <212> ADN

<213> Eryngium glaciale

<220>

<223> ADN c969a de valenceno sintasa

<400> 3

10

atgtctctta atgtacttag tacgtcaggt tcagctccaa caaccaaatic atctgagatt 60  
 actcgtagggt ccgctaatta tcatacctagt ttatggggag acaagttcct cgaatattcg 120  
 agcccagatc acctgaaaaa tgattcattc acagaaaaga aacatgaaca actcaaagaa 180  
 gagggtgaaga agatgctagt agaaaagggt caaaagcctc aacaacagct gaatctgatc 240  
 aacgaaatac aacgactagg tttatcatac ctttttgaac ccgaaattga ggctgcattg 300  
 caggaaatca gtgttaccta tgatgaattt tgttgttagta cagacgctga tgaccttcac 360  
 aatgtgtgctc tctctttccg aatacttaga gaacatggac ataatgtatc ttctgatgtg 420  
 tttcagaaat tcattgtag caatgggaag ttgaaagact acttggttaa tgatgctaga 480  
 ggactgttaa gcttgtacga agcaacacat tttcgggttc ataatgatga taaacttgaa 540  
 gatttgctgt cagtaacaac ctctcgtcct gagcatctca aatcccacgt gaagtaccct 600  
 cttgaggacg aatcagtag agcacttaag catcccctcc ataaagaact aatcgacta 660  
 ggagcgagat attacatatic cattedacgaa aaatttgatt cacacaataa attgcttttg 720  
 gaggttgcaa aactagatt taaccgactg cagaaaatgt atcaacatga gctagcccac 780  
 cttacaaggt ggtggaaga tttagatttt acaacaacac ttccatttgc aagagataga 840  
 attggtgagg gttacttttg gatcttagga atgtactttg agccagaacg taaggatgtc 900  
 aggaattct tgaacagagt atttgcactt attacagtag ttgatgacac gtatgatgtg 960  
 tatggtacat tcaaagaact tctactgttc actgatgcaa ttgaaagatg ggaactagt 1020  
 gatttggatc agctaccggg atatatgaga attatttacc aagctctcat ggatgtttat 1080  
 aatcaaatgg aggaaaagtt gtcaatgaaa gctgattgtc caacataccg tcttgagttt 1140  
 gcaatagaaa cagttaaagc catgttcaga tcatacctcg aagaagctag atggtccaaa 1200  
 gaacattata tcccatcgat ggaagagtat atgaccgtgg cactggtatc ggttggctac 1260  
 aaaaccatat taactaattc ctttgttggg atgggggata ttgcaacacg ggaagttttt 1320  
 gattgggtgt tcaatagtcc attgattatt agagcttccg acttaattgc cagattggga 1380  
 gatgatattg gaggccatga ggaggagcag aagaaaggag acgcagccac tgctatcgag 1440  
 tgttacataa aagagaatca tgaacaaaag catgaagctt atgatgaatt tcagaaacaa 1500  
 atgataatg cttggaagga tttgaataag gaagctctac gtccatttcc tgttccaatg 1560  
 actttcatca caagagtgt tcattttacg cgcgccatcc atgttattta tgccgacttt 1620  
 agtgatggtt acacacgttc agacaaggcg atcagaggtt acataacttc actgctcgtg 1680  
 gatcctattc ctttghtaa 1698

<210> 4

<211> 847

15 <212> ADN

<213> Eryngium glaciale

<220>

<223> Secuencia de contiguo que contiene EGVs parcial

<400> 4

20

ES 2 647 828 T3

```

tgtaataagt gcaataactc tgttcaagaa ttccttgaca tccttacgtt ctggctcaaa 60
gtacattcct aagatccaaa agtaaccctc aacaattcta tctcttgcaa atggaagttt 120
gtttgtaaaa tctaaatctt tccaccacct tgtaagggtg gctagctcat gttgatacat 180
tttctgcagt cggttaaaat ctagttttgc aaactccaaa agcaatttat tgtgtgaatc 240
aaatttttcg taaatggata tgtaatatct cgctcctagt cgatttagtt ctttatggag 300
gggatgctta agtgctctac tgatttcgtc ctcaagaggg tacttcacgt gggatttgag 360
atgctcaaga cgagagggtg ttactgacag caactottca agtttatcat cattatgaac 420
ccgaaaatgt gttgcttcgt acaagcttaa cagtcctcta gcatcattaa ccaagtagtc 480
tttcaacttc ccattgctat ccatgaattt ctgaaacaca tcagaagata cattatgtcc 540
atgttctcta agtattcgga aagagagagc aacattgtga aggtcatcag cgtctgtact 600
acaacaaaat tcatcatagg taacactgat ttcctgcaat gcagcctcaa tttcgggttc 660
aaaaaggtat gataaaccta gtcgttgat ttcggtgatc agattcagct gttgttgagg 720
cttttgaacc gtttctacta gcactctctt cacctcttct ttgagttgtt catgtttctt 780
ttctgtgaat gaatcatttt tcaggtgatc tgggctcgaa tattcgagga acttgtctcc 840
ccataaa 847

```

- <210> 5
- <211> 32
- 5 <212> ADN
- <213> Secuencia artificial
- <220>
- <223> Cebador de RACE 63-1-2-Fwd2
- <400> 5
- 10 gctagctcat gttgatacat tttctgcagt cg 32
- <210> 6
- <211> 29
- <212> ADN
- <213> Secuencia artificial
- 15 <220>
- <223> Cebador de RACE 63-1-2-Rev2
- <400> 6
- gctgtcagta acaactctc gtcttgagc 29
- <210> 7
- 20 <211> 33
- <212> ADN
- <213> Secuencia artificial
- <220>
- <223> Cebador de RACE 63-1-2-NestFwd2
- <400> 7
- 25 cggttaaaat ctagttttgc aaactccaaa agc 33
- <210> 8
- <211> 27
- <212> ADN
- 30 <213> Secuencia artificial
- <220>
- <223> Cebador de RACE 63-1-2-NestRev2
- <400> 8
- cccacgtgaa gtaccctctt gaggacg 27
- 35 <210> 9
- <211> 60
- <212> ADN
- <213> Secuencia artificial
- <220>
- 40 <223> 63-52-Cebador de ensamblaje de Gibson EG2FwdPart1
- <400> 9
- gctgaattcg agctcggtag cattaaaaaa aatgtctctt aatgtactta gtacgtcagg 60
- <210> 10
- <211> 33
- 45 <212> ADN
- <213> Secuencia artificial
- <220>
- <223> Cebador de ensamblaje de Gibson 63-52-EG2RevPart1
- <400> 10

gaagttcttt gaatgtacca tacacatcat acg 33  
 <210> 11  
 <211> 40  
 <212> ADN  
 5 <213> Secuencia artificial  
 <220>  
 <223> Cebador de ensamblaje de Gibson 63-52-EG2FwdPart2  
 <400> 11  
 gatgtgatg gtacattcaa agaactcta ctgttcactg 40  
 10 <210> 12  
 <211> 56  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 15 <223> 63-52-EG2RevPart2  
 <400> 12  
 tacgcgcaca aaagcagaga ttctagatta caaaggaata ggatccacga gcagtg 56  
 <210> 13  
 <211> 1647  
 20 <212> ADN  
 <213> Citrus sinensis  
 <220>  
 <223> ADN de valenceno sintasa cítrica  
 <300>  
 25 <308> GenBank GQ988384.1  
 <309> 20/10/2009  
 <400> 13

atgtcgtctg gagaacatt tcgtcctact gcagatttcc atcctagttt atggagaaac 60  
 catttcctca aaggtgcttc tgatttcaag acagttgatc atactgcaac tcaagaacga 120  
 cagcaggcac tgaagaaga ggtaaggaga atgataacag atgctgaaga taagcctggt 180  
 cagaagttac gcttgattga tgaagtacaa cgctggggg tggcttatca ctttgagaaa 240

30 gaaatagaag atgcaatata aaaattatgt ccaatctata ttgacagtaa tagagctgat 300  
 ctccacaccg tttcccttca ttttcgattg cttaggcagc aaggaatcaa gatttcatgt 360  
 gatgtgtttg agaagttcaa agatgatgag ggtagattca agtcatcggt gataaacgat 420  
 gttcaaggga tgtaagttt gtacgaggca gcatacatgg cagttcgcgg agaacatata 480  
 ttgatgaag ccattgcttt cactaccact cacctgaagt cattggtagc tcaggatcat 540  
 gtaaccctta agcttgcgga acagataaat catgctttat accgtcctct tcgtaaaacc 600  
 ctaccaagat tagagcgag gtattttatg tccatgatca attcaacaag tgatcattta 660  
 tacaataaaa ctctgtgaa ttttgcaag ttagatttta acatattgct agagctgcac 720  
 aaggaggaac tcaatgaatt aacaagtgg tggaaagatt tagacttcac tacaaaacta 780  
 ccttatgcaa gagacagatt agtggagtta tattttggg atttaggac atacttcgag 840  
 cctcaatatg ctttgggag aaagataatg acccaattaa attacatatt atccatcata 900  
 gatgatactt atgatcgta tggtagactt gaagaactca gcctctttac tgaagcagtt 960  
 caaagatgga atattgagc cgtagatatg cttccagaat acatgaaatt gatttacagg 1020  
 acactcttag atgcttttaa tgaattgag gaagatatgg ccaagcaagg aagatcacac 1080  
 tgcgtacggt atgcaaaaga ggagaatcaa aaagtaattg gagcactc tgttcaagcc 1140  
 aaatggttca gtgaaggtta cgtccaaca attgaggagt atatgcctat tgcactaaca 1200  
 agttgtgctt acacattcgt cataacaaat tccttccttg gcatgggtga ttttgcaact 1260  
 aaagaggttt ttgaatggat ctccaataac cctaaggttg taaaagcagc atcagttatc 1320  
 tgcagactca tggatgacat gcaaggtcat gagtttgagc agaagagagg acatggtgag 1380  
 tcagctattg aatgttacac gaagcagcat ggtgtctcta aggaagaggc aattaaaatg 1440  
 tttgaagaag aagttgcaa tgcatggaaa gatattaacg aggagttgat gatgaagcca 1500  
 accgtcgttg cccgaccact gctcgggacg attccttaac ttgctcgtgc aattgatttt 1560  
 atttacaaag aggacgacgg ctatacgcat tcttacctaa ttaaagatca aattgcttct 1620  
 tgcttaggag accacgttcc attttga 1647

<210> 14  
 <211> 548  
 35 <212> PRT  
 <213> Citrus sinensis  
 <220>  
 <223> Proteína valenceno sintasa cítrica

ES 2 647 828 T3

<400> 14

```

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

```

```

225                230                235                240
Lys Glu Glu Leu Asn Glu Leu Thr Lys Trp Trp Lys Asp Leu Asp Phe
                245                250                255
Thr Thr Lys Leu Pro Tyr Ala Arg Asp Arg Leu Val Glu Leu Tyr Phe
                260                265                270
Trp Asp Leu Gly Thr Tyr Phe Glu Pro Gln Tyr Ala Phe Gly Arg Lys
                275                280                285
Ile Met Thr Gln Leu Asn Tyr Phe Ile Leu Ser Ile Ile Asp Thr Tyr
                290                295                300
Asp Ala Tyr Gly Thr Leu Glu Glu Leu Ser Leu Phe Thr Glu Ala Val
305                310                315                320
Gln Arg Trp Asn Ile Glu Ala Val Asp Met Leu Pro Glu Tyr Met Lys
                325                330                335
Leu Ile Tyr Arg Thr Leu Leu Asp Ala Phe Asn Glu Ile Glu Glu Asp
                340                345                350
Met Ala Lys Gln Gly Arg Ser His Cys Val Arg Tyr Ala Lys Glu Glu
355                360                365
Asn Gln Lys Val Ile Gly Ala Tyr Ser Val Gln Ala Lys Trp Phe Ser
370                375                380
Glu Gly Tyr Val Pro Thr Ile Glu Glu Tyr Met Pro Ile Ala Leu Thr
385                390                395                400
Ser Cys Ala Tyr Thr Phe Val Ile Thr Asn Ser Phe Leu Gly Met Gly
                405                410                415
Asp Phe Ala Thr Lys Glu Val Phe Glu Trp Ile Ser Asn Asn Pro Lys
                420                425                430
Val Val Lys Ala Ala Ser Val Ile Cys Arg Leu Met Asp Asp Met Gln
435                440                445
Gly His Glu Phe Glu Gln Lys Arg Gly His Val Ala Ser Ala Ile Glu
450                455                460
Cys Tyr Thr Lys Gln His Gly Val Ser Lys Glu Glu Ala Ile Lys Met
465                470                475                480
Phe Glu Glu Glu Val Ala Asn Ala Trp Lys Asp Ile Asn Glu Glu Leu
                485                490                495
Met Met Lys Pro Thr Val Val Ala Arg Pro Leu Leu Gly Thr Ile Leu
500                505                510
Asn Leu Ala Arg Ala Ile Asp Phe Ile Tyr Lys Glu Asp Asp Gly Tyr
515                520                525
Thr His Ser Tyr Leu Ile Lys Asp Gln Ile Ala Ser Val Leu Gly Asp
530                535                540
His Val Pro Phe
545

```

<210> 15

<211> 548

5 <212> PRT

<213> Citrus x paradisi

<220>

<223> Proteína valenceno sintasa cítrica AAM00426

<300>

10 <308> Genbank AAM00426

<309> 02/04/2002

<400> 15

```

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

```

15

ES 2 647 828 T3

```

50          55          60
Leu Ile Asp Glu Val Gln Arg Leu Gly Val Ala Tyr His Phe Glu Lys
65          70          75          80
Glu Ile Glu Asp Ala Ile Leu Lys Leu Cys Pro Ile Tyr Ile Asp Ser
85          90          95
Asn Arg Ala Asp Leu His Thr Val Ser Leu His Phe Arg Leu Leu Arg
100         105         110
Gln Gln Gly Ile Lys Ile Ser Cys Asp Val Phe Glu Lys Phe Lys Asp
115         120         125
Asp Glu Gly Arg Phe Lys Ser Ser Leu Ile Asn Asp Val Gln Gly Met
130         135         140
Leu Ser Leu Tyr Glu Ala Ala Tyr Met Ala Val Arg Gly Glu His Ile
145         150         155         160
Leu Asp Glu Ala Ile Ala Phe Thr Thr Thr His Leu Lys Ser Leu Val
165         170         175
Ala Gln Asp His Val Thr Pro Lys Leu Ala Glu Gln Ile Asn His Ala
180         185         190
Leu Tyr Arg Pro Leu Arg Lys Thr Leu Pro Arg Leu Glu Ala Arg Tyr
195         200         205
Phe Met Ser Met Ile Asn Ser Thr Ser Asp His Leu Tyr Asn Lys Thr
210         215         220
Leu Leu Asn Phe Ala Lys Leu Asp Phe Asn Ile Leu Leu Glu Pro His
225         230         235         240
Lys Glu Glu Leu Asn Glu Leu Thr Lys Trp Trp Lys Asp Leu Asp Phe
245         250         255
Thr Thr Lys Leu Pro Tyr Ala Arg Asp Arg Leu Val Glu Leu Tyr Phe
260         265         270
Trp Asp Leu Gly Thr Tyr Phe Glu Pro Gln Tyr Ala Phe Gly Arg Lys
275         280         285
Ile Met Thr Gln Leu Asn Tyr Ile Leu Ser Ile Ile Asp Asp Thr Tyr
290         295         300
Asp Ala Tyr Gly Thr Leu Glu Glu Leu Ser Leu Phe Thr Glu Ala Val
305         310         315         320
Gln Arg Trp Asn Ile Glu Ala Val Asp Met Leu Pro Glu Tyr Met Lys
325         330         335
Leu Ile Tyr Arg Thr Leu Leu Asp Ala Phe Asn Glu Ile Glu Glu Asp
340         345         350
Met Ala Lys Gln Gly Arg Ser His Cys Val Arg Tyr Ala Lys Glu Glu
355         360         365
Asn Gln Lys Val Ile Gly Ala Tyr Ser Val Gln Ala Lys Trp Phe Ser
370         375         380
Glu Gly Tyr Val Pro Thr Ile Glu Glu Tyr Met Pro Ile Ala Leu Thr
385         390         395         400
Ser Cys Ala Tyr Thr Phe Val Ile Thr Asn Ser Phe Leu Gly Met Gly
405         410         415
Asp Phe Ala Thr Lys Glu Val Phe Glu Trp Ile Ser Asn Asn Pro Lys
420         425         430
Val Val Lys Ala Ala Ser Val Ile Cys Arg Leu Met Asp Asp Met Gln
435         440         445
Gly His Glu Phe Glu Gln Lys Arg Gly His Val Ala Ser Ala Ile Glu
450         455         460
Cys Tyr Thr Lys Gln His Gly Val Ser Lys Glu Glu Ala Ile Lys Met
465         470         475         480
Phe Glu Glu Glu Val Ala Asn Ala Trp Lys Asp Ile Asp Glu Glu Leu
485         490         495
Met Met Lys Pro Thr Val Val Ala Arg Pro Leu Leu Gly Thr Ile Leu
500         505         510
Asn Leu Ala Arg Ala Ile Asp Phe Ile Tyr Lys Glu Asp Asp Gly Tyr
515         520         525
Thr His Ser Tyr Leu Ile Lys Asp Gln Ile Ala Ser Val Leu Gly Asp
530         535         540
His Val Pro Phe
545

```

<210> 16

<211> 556

5 <212> PRT

<213> Vitis vinifera

<220>

<223> Proteína valenceno sintasa

<300>  
 <308> Uniprot C1JXK7  
 <309> 21/09/2011  
 <400> 16

5

```

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

ES 2 647 828 T3

Tyr Phe Ala Glu Ala Lys Trp Leu His Glu Glu His Val Pro Ala Phe  
 385 390 395 400  
 Glu Glu Tyr Met Arg Val Ala Leu Ala Ser Ser Gly Tyr Cys Leu Leu  
 405 410 415  
 Ala Thr Thr Ser Phe Val Gly Met Gly Glu Ile Ala Thr Lys Glu Ala  
 420 425 430  
 Phe Asp Trp Val Thr Ser Asp Pro Lys Ile Met Ser Ser Ser Asn Phe  
 435 440 445  
 Ile Thr Arg Leu Met Asp Asp Ile Lys Ser His Lys Phe Glu Gln Lys  
 450 455 460  
 Arg Gly His Val Ala Ser Ala Val Glu Cys Tyr Met Lys Gln Tyr Gly  
 465 470 475 480  
 Val Ser Glu Glu Gln Val Tyr Ser Glu Phe Gln Lys Gln Ile Glu Asn  
 485 490 495  
 Ala Trp Leu Asp Ile Asn Gln Glu Cys Leu Lys Pro Thr Ala Val Ser  
 500 505 510  
 Met Pro Leu Leu Ala Arg Leu Leu Asn Leu Thr Arg Thr Met Asp Val  
 515 520 525  
 Ile Tyr Lys Glu Gln Asp Ser Tyr Thr His Val Gly Lys Val Met Arg  
 530 535 540  
 Asp Asn Ile Ala Ser Val Phe Ile Asn Ala Val Ile  
 545 550 555

<210> 17

<211> 573

5 <212> PRT

<213> Chamaecyparis nootkatensis

<220>

<223> Proteína valenceno sintasa

<400> 17

10

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

Lys Ile Tyr Met Lys Lys Ile Leu Glu Leu Ala Ile Leu Asp Phe Asn  
 245 250 255  
 Ile Leu Gln Ser Gln His Gln His Glu Met Lys Leu Ile Ser Thr Trp  
 260 265 270  
 Trp Lys Asn Ser Ser Ala Ile Gln Leu Asp Phe Phe Arg His Arg His  
 275 280 285  
 Ile Glu Ser Tyr Phe Trp Trp Ala Ser Pro Leu Phe Glu Pro Glu Phe  
 290 295 300  
 Ser Thr Cys Arg Ile Asn Cys Thr Lys Leu Ser Thr Lys Met Phe Leu  
 305 310 315 320  
 Leu Asp Asp Ile Tyr Asp Thr Tyr Gly Thr Val Glu Glu Leu Lys Pro  
 325 330 335  
 Phe Thr Thr Thr Leu Thr Arg Trp Asp Val Ser Thr Val Asp Asn His  
 340 345 350  
 Pro Asp Tyr Met Lys Ile Ala Phe Asn Phe Ser Tyr Glu Ile Tyr Lys  
 355 360 365  
 Glu Ile Ala Ser Glu Ala Glu Arg Lys His Gly Pro Phe Val Tyr Lys  
 370 375 380  
 Tyr Leu Gln Ser Cys Trp Lys Ser Tyr Ile Glu Ala Tyr Met Gln Glu  
 385 390 395 400  
 Ala Glu Trp Ile Ala Ser Asn His Ile Pro Gly Phe Asp Glu Tyr Leu  
 405 410 415  
 Met Asn Gly Val Lys Ser Ser Gly Met Arg Ile Leu Met Ile His Ala  
 420 425 430  
 Leu Ile Leu Met Asp Thr Pro Leu Ser Asp Glu Ile Leu Glu Gln Leu  
 435 440 445  
 Asp Ile Pro Ser Ser Lys Ser Gln Ala Leu Leu Ser Leu Ile Thr Arg  
 450 455 460  
 Leu Val Asp Asp Val Lys Asp Phe Glu Asp Glu Gln Ala His Gly Glu  
 465 470 475 480  
 Met Ala Ser Ser Ile Glu Cys Tyr Met Lys Asp Asn His Gly Ser Thr  
 485 490 495  
 Arg Glu Asp Ala Leu Asn Tyr Leu Lys Ile Arg Ile Glu Ser Cys Val  
 500 505 510  
 Gln Glu Leu Asn Lys Glu Leu Leu Glu Pro Ser Asn Met His Gly Ser  
 515 520 525  
 Phe Arg Asn Leu Tyr Leu Asn Val Gly Met Arg Val Ile Phe Phe Met  
 530 535 540  
 Leu Asn Asp Gly Asp Leu Phe Thr His Ser Asn Arg Lys Glu Ile Gln  
 545 550 555 560  
 Asp Ala Ile Thr Lys Phe Phe Val Glu Pro Ile Ile Pro  
 565 570

<210> 18

<211> 556

5 <212> PRT

<213> Eleutherococcus trifolius

<220>

<223> Proteína alfa-copaeno sintasa

<300>

10 <308> GenBank ADK94034.1

<309> 01/04/2012

<400> 18

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

15

ES 2 647 828 T3

Asn Thr Lys Glu His Phe Arg Gln Leu Lys Glu Glu Val Lys Lys Met  
50 55 60  
Leu Val Asp Ala Gly Pro Asn Gln Gln Leu Asn Leu Ile Asp Asp Ile  
65 70 75 80  
Gln Arg Leu Gly Val Ala Tyr Gln Phe Glu Ala Glu Ile Asp Ala Ala  
85 90 95  
Leu Gln Arg Met Asn Val Ile Phe Gln Gly Asn Asp Asp Asp Leu His  
100 105 110  
Thr Ile Ser Leu Arg Phe Arg Leu Leu Arg Gln His Gly Tyr Asn Val  
115 120 125  
Ser Ser Asp Val Phe Arg Lys Phe Met Asp Asn Asn Gly Lys Phe Lys  
130 135 140  
Glu Cys Leu Ile Ser Asp Leu Arg Gly Val Leu Ser Leu Tyr Glu Ala  
145 150 155 160  
Thr His Phe Arg Val His Gly Glu Asp Ile Leu Glu Asp Ala Leu Glu  
165 170 175  
Phe Thr Thr Ser His Leu Glu Arg Leu Lys Ser His Leu Lys Asn Pro  
180 185 190  
Leu Ala Ala Gln Val Ile Arg Ala Leu Lys Cys Pro Ile His Lys Gly  
195 200 205  
Leu Asn Arg Leu Glu Ala Lys His Tyr Ile Ser Ile Tyr Gln Gln Glu  
210 215 220  
Asp Asp Ser His Asn Lys Val Leu Leu Asn Phe Ala Lys Leu Asp Phe  
225 230 235 240  
Asn Leu Leu Gln Lys Met His Gln Gly Glu Leu Ser His Ile Thr Arg  
245 250 255  
Trp Trp Lys Glu Leu Asn Phe Ala Lys Lys Leu Pro Phe Ala Arg Asp  
260 265 270  
Arg Val Val Glu Cys Tyr Phe Trp Ile Leu Gly Val Tyr Phe Glu Pro  
275 280 285  
Gln Tyr Leu Ile Ala Arg Arg Phe Leu Thr Lys Ile Ile Ala Met Ala  
290 295 300  
Ser Val Ala Asp Asp Ile Tyr Asp Val Tyr Gly Thr Leu Glu Glu Leu  
305 310 315 320  
Val Ile Leu Thr Asp Ala Ile Glu Arg Trp Asp Met Gly Ala Leu Asp  
325 330 335  
Gln Ile Pro Glu Cys Met Arg Val Tyr His Arg Ala Leu Leu Asp Val  
340 345 350  
Tyr Thr Glu Met Glu Glu Glu Met Ala Lys Thr Gly Arg Pro Ser Tyr  
355 360 365  
Arg Val His Tyr Ala Lys Glu Ala Tyr Lys Glu Leu Val Arg Gln Tyr  
370 375 380  
Leu Ala Glu Ala Lys Trp Phe Gln Glu Asp Tyr Asp Pro Thr Leu Glu  
385 390 395 400  
Glu Tyr Leu Pro Val Ala Leu Ile Ser Gly Gly Tyr Lys Met Leu Ala  
405 410 415  
Thr His Ser Phe Val Gly Met Gly Asp Leu Ala Thr Lys Glu Ala Phe  
420 425 430  
Asp Trp Val Ser Asn Asn Pro Leu Ile Val Lys Ala Ser Ser Val Ile  
435 440 445  
Cys Arg Leu Ser Asp Asp Met Val Gly His Glu Val Glu His Glu Arg  
450 455 460  
Gly Asp Val Ala Ser Ala Val Glu Cys Tyr Met Lys Gln Tyr Gly Val  
465 470 475 480  
Thr Lys Gln Glu Val Tyr Ile Glu Phe Gln Lys Gln Ile Ser Asn Ala  
485 490 495  
Trp Lys Asp Met Asn Gln Glu Cys Leu His Pro Thr Thr Val Thr Met  
500 505 510  
Pro Leu Leu Thr Val Ile Phe Asn Met Thr Arg Val Ile Asn Leu Leu  
515 520 525  
Tyr Asp Glu Glu Asp Gly Tyr Thr Asn Ser Asn Thr Arg Thr Lys Asp  
530 535 540  
Phe Ile Thr Ser Val Leu Ile Asp Pro Val Gln Ile

545

550

555

5 <210> 19  
<211> 565  
<212> PRT  
<213> Actinidia deliciosa

ES 2 647 828 T3

<220>  
 <223> Proteína germacreno D sintasa  
 <300>  
 <308> GenBank AAX16121.1  
 5 <309> 01/11/2005  
 <310> US7312323  
 <311> 25/12/2007  
 <400> 19

```

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

10

ES 2 647 828 T3

```

          355                360                365
Met Ala Lys Gln Gly Arg Ser Tyr Cys Val Asp Tyr Ala Lys Ser Ser
  370                375                380
Met Lys Ile Leu Val Arg Ala Tyr Phe Glu Glu Ala Lys Trp Phe His
  385                390                395
Gln Gly Tyr Val Pro Thr Met Glu Glu Tyr Met Gln Val Ala Leu Val
          405                410                415
Thr Ala Gly Tyr Lys Met Leu Ala Thr Ser Ser Phe Val Gly Met Gly
          420                425                430
Asp Leu Ala Thr Lys Glu Ala Phe Asp Trp Val Ser Asn Asp Pro Leu
          435                440                445
Ile Val Gln Ala Ala Ser Val Ile Gly Arg Leu Lys Asp Asp Ile Val
          450                455                460
Gly His Lys Phe Glu Gln Lys Arg Gly His Val Ala Ser Ala Val Glu
  465                470                475
Cys Tyr Ser Lys Gln His Gly Thr Thr Glu Glu Glu Ala Ile Ile Glu
          485                490                495
Leu Asp Lys Gln Val Thr His Ser Trp Lys Asp Ile Asn Ala Glu Cys
          500                505                510
Leu Cys Pro Ile Lys Val Pro Met Pro Leu Leu Ala Arg Val Leu Asn
          515                520                525
Leu Ala Arg Val Leu Tyr Val Ile Tyr Gln Asp Glu Asp Gly Tyr Thr
          530                535                540
His Pro Gly Thr Lys Val Glu Asn Phe Val Thr Ser Val Leu Ile Asp
  545                550                555
Ser Met Pro Ile Asn
          565

```

<210> 20

<211> 557

5 <212> PRT

<213> Vitis vinifera

<220>

<223> Proteína (-)-germacreno D sintasa

<300>

10 <308> NCBI XP\_003634696.1

<309> 07/12/2011

<400> 20

```

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

```

15

				165					170					175	
Ala	Leu	Ala	Phe	Thr	Thr	Thr	His	Leu	Gln	Ser	Ala	Thr	Lys	His	Ser
			180					185					190		
Ser	Asn	Pro	Leu	Ala	Glu	Gln	Val	Val	His	Ala	Leu	Lys	Gln	Pro	Ile
		195					200					205			
Arg	Lys	Gly	Leu	Pro	Arg	Leu	Glu	Ala	Arg	His	Tyr	Phe	Ser	Val	Tyr
	210					215					220				
Gln	Ala	Asp	Asp	Ser	His	Asn	Lys	Ala	Leu	Leu	Lys	Leu	Ala	Lys	Leu
225					230					235					240
Asp	Phe	Asn	Leu	Leu	Gln	Lys	Leu	His	Gln	Lys	Glu	Leu	Ser	Asp	Ile
			245						250					255	
Ser	Ala	Trp	Trp	Lys	Asp	Leu	Asp	Phe	Ala	His	Lys	Leu	Pro	Phe	Ala
		260						265					270		
Arg	Asp	Arg	Val	Val	Glu	Cys	Tyr	Phe	Trp	Ile	Leu	Gly	Val	Tyr	Phe
	275						280					285			
Glu	Pro	Gln	Phe	Phe	Phe	Ala	Arg	Arg	Ile	Leu	Thr	Lys	Val	Ile	Ala
	290					295					300				
Met	Thr	Ser	Ile	Ile	Asp	Asp	Ile	Tyr	Asp	Val	Tyr	Gly	Thr	Leu	Glu
305					310					315					320
Glu	Leu	Glu	Leu	Phe	Thr	Glu	Ala	Val	Glu	Arg	Trp	Asp	Ile	Ser	Ala
				325					330					335	
Ile	Asp	Gln	Leu	Pro	Glu	Tyr	Met	Arg	Val	Cys	Tyr	Gln	Ala	Leu	Leu
			340					345					350		
Tyr	Val	Tyr	Ser	Glu	Ile	Glu	Glu	Glu	Met	Ala	Lys	Glu	Gly	Arg	Ser
	355					360						365			
Tyr	Arg	Leu	Tyr	Tyr	Ala	Lys	Glu	Ala	Met	Lys	Asn	Gln	Val	Arg	Ala
	370					375					380				
Tyr	Tyr	Glu	Glu	Ala	Lys	Trp	Leu	Gln	Val	Gln	Gln	Ile	Pro	Thr	Met
385					390					395					400
Glu	Glu	Tyr	Met	Pro	Val	Ala	Leu	Val	Thr	Ser	Ala	Tyr	Ser	Met	Leu
				405					410					415	
Ala	Thr	Thr	Ser	Phe	Val	Gly	Met	Gly	Asp	Ala	Val	Thr	Lys	Glu	Ser
			420					425					430		
Phe	Asp	Trp	Ile	Phe	Ser	Lys	Pro	Lys	Ile	Val	Arg	Ala	Ser	Ala	Ile
	435					440						445			
Val	Cys	Arg	Leu	Met	Asp	Asp	Met	Val	Ser	His	Lys	Phe	Glu	Gln	Lys
	450				455						460				
Arg	Gly	His	Val	Ala	Ser	Ala	Val	Glu	Cys	Tyr	Met	Lys	Gln	His	Gly
465					470				475						480
Ala	Ser	Glu	Gln	Glu	Thr	His	Asn	Glu	Phe	His	Lys	Gln	Val	Arg	Asp
				485					490					495	
Ala	Trp	Lys	Asp	Ile	Asn	Glu	Glu	Cys	Leu	Ile	Pro	Thr	Ala	Val	Pro
			500					505					510		
Met	Pro	Ile	Leu	Met	Arg	Val	Leu	Asn	Leu	Ala	Arg	Val	Ile	Asp	Val
	515					520						525			
Ile	Tyr	Lys	Asn	Glu	Asp	Gly	Tyr	Thr	His	Ser	Gly	Thr	Val	Leu	Lys
	530					535					540				
Asp	Phe	Val	Thr	Ser	Met	Leu	Ile	Asp	Pro	Val	Pro	Ile			
545					550					555					

- <210> 21
- <211> 562
- 5 <212> PRT
- <213> Santalum murrayanum
- <220>
- <223> Proteína sesquiterpeno sintasa
- <300>
- 10 <308> Uniprot F6M8H7.1
- <309> 03/10/2012
- <400> 21

ES 2 647 828 T3

Met Glu Asn Gln Lys Met Pro Ile Ser Ser Val Pro Asn Leu Lys Asp  
1 5 10 15  
Leu Asn Met Ile Ser Arg Pro Ile Ala Asn Phe Pro Pro Ser Ile Trp  
20 25 30  
Gly Asp Arg Phe Ile Asn Tyr Thr Cys Glu Asp Glu Asn Asp Gln Thr  
35 40 45  
Gln Lys Glu Arg Gln Val Glu Glu Leu Lys Glu Gln Val Arg Arg Glu  
50 55 60  
Leu Ala Ala Thr Val Asp Lys Pro Leu Gln Gln Leu Asn Ile Ile Asp  
65 70 75 80  
Ala Thr Gln Arg Leu Gly Ile Ala Tyr Leu Phe Glu Asn Glu Ile Glu  
85 90 95  
Glu Ser Leu Lys His Ile Tyr Leu His Thr Tyr Val Glu Asn Asn Cys  
100 105 110  
Phe Glu Gly Ser Asp Asp Leu Tyr Ser Val Ala Leu Trp Phe Arg Leu  
115 120 125  
Leu Arg Gln Asn Gly Tyr Lys Val Ser Cys Asp Val Phe Asn Lys Phe  
130 135 140  
Arg Asp Asn Glu Gly Asn Phe Lys Asn Asn Leu Met Glu Asp Ala Lys  
145 150 155 160  
Gly Leu Leu Glu Leu Tyr Glu Ala Thr His Val Ser Ile His Gly Glu  
165 170 175  
Glu Met Leu Asp Asp Ala Leu Glu Phe Thr Lys Thr Arg Leu Glu Ser  
180 185 190  
Val Val Ser His Leu Asn Tyr Pro Leu Ala Glu Gln Val Arg His Ala  
195 200 205  
Leu Tyr Gln Pro Leu His Arg Gly Leu Pro Arg Leu Glu Ala Val Tyr  
210 215 220  
Phe Phe Arg Ile Tyr Glu Ala His Ala Ser His Asn Lys Ala Leu Leu  
225 230 235 240  
Lys Leu Ala Lys Leu Asp Phe Asn Leu Leu Gln Ser Phe His Lys Lys  
245 250 255  
Glu Leu Ser Asp Ile Ala Arg Trp Trp Lys Ser Leu Asp Phe Ala Ala  
260 265 270  
Lys Phe Pro Phe Ala Arg Asp Arg Leu Val Glu Gly Tyr Phe Trp Val  
275 280 285  
Leu Gly Val Tyr Phe Glu Pro Gln Tyr Ser Leu Ala Arg Lys Ile Ile  
290 295 300  
Ile Lys Val Phe Thr Met Ile Ser Thr Ile Asp Asp Ile Tyr Asp Ala  
305 310 315 320  
Tyr Gly Thr Leu Asp Glu Leu Lys Leu Phe Thr Lys Ala Met Gln Arg  
325 330 335  
Trp Asp Val Gly Ser Leu Asp Gln Leu Pro Glu Tyr Met Lys Pro Cys  
340 345 350  
Tyr Lys Ser Ile Leu Asp Val Tyr Asn Glu Ile Glu Glu Glu Met Ala  
355 360 365  
Asn Gln Gly Ser Leu Phe Arg Met His Tyr Ala Lys Glu Val Met Lys  
370 375 380  
Thr Ile Val Glu Gly Tyr Met Asp Glu Ala Lys Trp Cys His Glu Lys  
385 390 395 400  
Tyr Val Pro Thr Phe Gln Glu Tyr Met Ser Val Ala Leu Val Thr Ser  
405 410 415  
Gly Tyr Thr Phe Leu Thr Thr Ile Ser Tyr Leu Gly Met Gly Glu Ile  
420 425 430  
Ala Ser Lys Glu Ala Phe Asp Trp Leu Phe Ser His Pro Pro Val Ile  
435 440 445  
Glu Ala Ser Glu Ser Val Gly Arg Leu Met Asp Asp Met Arg Ser His  
450 455 460  
Lys Phe Glu Gln Glu Arg Gly His Val Ala Ser Gly Ile Glu Cys Tyr  
465 470 475 480  
Met Lys Gln Tyr Gly Val Thr Glu Glu Glu Ala His Asp Glu Phe Arg  
485 490 495

ES 2 647 828 T3

Lys Arg Leu Val Lys Ala Trp Lys Asp Ile Asn Glu Glu Cys Leu Arg  
 500 505 510  
 Pro Tyr Arg Val Pro Lys Pro Leu Leu Thr Arg Ile Leu Asn Leu Thr  
 515 520 525  
 Arg Val Ile Asp Val Ile Tyr Lys Asn Glu Asp Gly Tyr Thr His Val  
 530 535 540  
 Lys Lys Ala Met Lys Asp Asn Ile Ala Ser Leu Leu Ile Asp Pro Val  
 545 550 555 560  
 Ile Val

<210> 22

<211> 550

5 <212> PRT

<213> Ricinus communis

<220>

<223> Proteína (+)-delta-cadineno sintasa

<300>

10 <308> GenBank EEF38721.1

<309> 12/02/2009

<400> 22

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

15

Asp Asp Ile Tyr Asp Ala Tyr Gly Thr Leu Glu Glu Leu Val Val Phe  
 305 310 315 320  
 Thr Asp Ala Ile Asn Arg Trp Asp Ile Asp Cys Ile Asp Gln Leu Pro  
 325 330 335  
 Glu Tyr Met Lys Val Cys Tyr Lys Ala Leu Leu Asn Val Tyr Glu Glu  
 340 345 350  
 Ile Glu Arg Ala Leu Ser Glu Gln Gly Arg Ser Tyr Arg Leu His Tyr  
 355 360 365  
 Ala Lys Glu Ala Met Lys Lys Leu Val Gln Ala Tyr Leu Val Glu Ala  
 370 375 380  
 Asn Trp Met Asn Lys Asn Tyr Val Pro Thr Met Asp Glu Tyr Met Ser  
 385 390 395 400  
 Ile Ala Leu Val Ser Cys Ala Tyr Pro Leu Leu Thr Val Thr Ser Phe  
 405 410 415  
 Val Gly Met Gly Asp Ile Ala Thr Lys Glu Val Phe Asp Trp Ala Ser  
 420 425 430  
 Asn Asp Pro Lys Ile Val Arg Val Ala Ser Ile Ile Cys Arg Leu Met  
 435 440 445  
 Asp Asp Ile Val Ser His Glu Phe Glu Gln Lys Arg Gly His Ile Ala  
 450 455 460  
 Ser Ser Val Glu Cys Tyr Met Lys Gln Asn Gly Val Ser Glu Glu Ala  
 465 470 475 480  
 Thr Arg Asp Glu Phe Asn Lys Gln Ile Val Asp Ala Trp Lys Asp Ile  
 485 490 495  
 Asn Glu Glu His Leu Gln Pro Asn Tyr Val Pro Met Pro Phe Arg Thr  
 500 505 510  
 Arg Val Val Asn Ser Ala Arg Ile Met Asp Tyr Leu Tyr Lys Asp Asp  
 515 520 525  
 Asp Glu Tyr Thr His Val Gly Glu Leu Met Lys Gly Ser Val Ala Ala  
 530 535 540  
 Leu Leu Ile Asp Pro Ala  
 545 550

<210> 23

<211> 562

5 <212> PRT

<213> Citrus x paradisi

<220>

<223> Proteína delta-cadineno sintasa

<300>

10 <310> US7273735

<311> 25/09/2007

<400> 23

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

15

ES 2 647 828 T3

Arg Gln Asn Gly Tyr Lys Ile Ser Ala Asp Val Phe Lys Lys Phe Lys  
 130 135 140  
 Asp Thr Asp Gly Asn Phe Lys Thr Ser Leu Ala Lys Asp Val Arg Gly  
 145 150 155 160  
 Met Leu Ser Leu Tyr Glu Ala Thr His Leu Gly Val His Glu Glu Asp  
 165 170 175  
 Ile Leu Asp Glu Ala Leu Ala Phe Thr Ser His Leu Glu Ser Ile  
 180 185 190  
 Ala Thr His Gln Ile Arg Ser Pro Leu Val Glu Gln Val Lys His Ala  
 195 200 205  
 Leu Val Gln Pro Ile His Arg Gly Phe Gln Arg Leu Glu Ala Arg Gln  
 210 215 220  
 Tyr Ile Pro Ile Tyr Gln Glu Glu Ser Pro His Asn Glu Ala Leu Leu  
 225 230 235 240  
 Thr Phe Ala Lys Leu Asp Phe Asn Lys Leu Gln Lys Pro His Gln Lys  
 245 250 255  
 Glu Leu Gly Asp Ile Ser Arg Trp Trp Lys Glu Leu Asp Phe Ala His  
 260 265 270  
 Lys Leu Pro Phe Ile Arg Asp Arg Val Ala Glu Cys Tyr Phe Trp Ile  
 275 280 285  
 Leu Gly Val Tyr Phe Glu Pro Gln Tyr Ser Phe Ala Arg Arg Ile Leu  
 290 295 300  
 Thr Lys Val Ile Ser Met Thr Ser Val Ile Asp Asp Ile Tyr Asp Val  
 305 310 315 320  
 Tyr Gly Lys Ile Glu Glu Leu Glu Leu Phe Thr Ser Ala Ile Glu Arg  
 325 330 335  
 Trp Asp Ile Ser Ala Ile Asp Gln Leu Pro Glu Tyr Met Lys Leu Cys  
 340 345 350  
 Tyr Arg Ala Leu Leu Asp Val Phe Ser Glu Ala Glu Lys Asp Leu Ala  
 355 360 365  
 Pro Gln Gly Lys Ser Tyr Arg Leu Tyr Tyr Ala Lys Glu Ala Met Lys  
 370 375 380  
 Asn Met Val Lys Asn Tyr Phe Tyr Glu Ala Lys Trp Cys Leu Gln Asn  
 385 390 395 400  
 Tyr Val Pro Thr Val Asp Glu Tyr Met Thr Val Ala Leu Val Thr Ser  
 405 410 415  
 Gly Ser Pro Met Leu Ser Thr Thr Ser Phe Val Gly Met Gly Asp Ile  
 420 425 430  
 Val Thr Lys Glu Ser Phe Glu Trp Leu Phe Ser Asn Pro Arg Phe Ile  
 435 440 445  
 Arg Ala Ser Ser Ile Val Cys Arg Leu Met Asp Asp Ile Val Ser His  
 450 455 460  
 Lys Phe Glu Gln Ser Arg Gly His Val Ala Ser Ser Val Glu Cys Tyr  
 465 470 475 480  
 Met Lys Gln His Gly Ala Thr Glu Glu Glu Ala Cys Asn Glu Phe Arg  
 485 490 495  
 Lys Gln Val Ser Asn Ala Trp Lys Asp Ile Asn Glu Asp Cys Leu Arg  
 500 505 510  
 Pro Thr Val Val Pro Met Pro Leu Leu Met Arg Ile Leu Asn Leu Thr  
 515 520 525  
 Arg Val Ile Asp Val Ile Tyr Lys Tyr Glu Asp Gly Tyr Thr His Ser  
 530 535 540  
 Ala Val Val Leu Lys Asp Phe Val Ala Ser Leu Phe Ile Asn Pro Val  
 545 550 555 560  
 Pro Ile

<210> 24

<211> 548

5 <212> PRT

<213> Nicotiana tabacum

<220>

<223> Proteína 5-epiaristolóqueno sintasa de tabaco

<300>

10 <308> GenBank L04680.1

<309> 27/06/1994

<400> 24

ES 2 647 828 T3

Met Ala Ser Ala Ala Val Ala Asn Tyr Glu Glu Glu Ile Val Arg Pro  
1 5 10  
Val Ala Asp Phe Ser Pro Ser Leu Trp Gly Asp Gln Phe Leu Ser Phe  
20 25 30  
Ser Ile Lys Asn Gln Val Ala Glu Lys Tyr Ala Lys Glu Ile Glu Ala  
35 40 45  
Leu Lys Glu Gln Thr Arg Asn Met Leu Leu Ala Thr Gly Met Lys Leu  
50 55 60  
Ala Asp Thr Leu Asn Leu Ile Asp Thr Ile Glu Arg Leu Gly Ile Ser  
65 70 75 80  
Tyr His Phe Glu Lys Glu Ile Asp Asp Ile Leu Asp Gln Ile Tyr Asn  
85 90 95  
Gln Asn Ser Asn Cys Asn Asp Leu Cys Thr Ser Ala Leu Gln Phe Arg  
100 105 110  
Leu Leu Arg Gln His Gly Phe Asn Ile Ser Pro Glu Ile Phe Ser Lys  
115 120 125  
Phe Gln Asp Glu Asn Gly Lys Phe Lys Glu Ser Leu Ala Ser Asp Val  
130 135 140  
Leu Gly Leu Leu Asn Leu Tyr Glu Ala Ser His Val Arg Thr His Ala  
145 150 155 160  
Asp Asp Ile Leu Glu Asp Ala Leu Ala Phe Ser Thr Ile His Leu Glu  
165 170 175  
Ser Ala Ala Pro His Leu Lys Ser Pro Leu Arg Glu Gln Val Thr His  
180 185 190  
Ala Leu Glu Gln Cys Leu His Lys Gly Val Pro Arg Val Glu Thr Arg  
195 200 205  
Phe Phe Ile Ser Ser Ile Tyr Asp Lys Glu Gln Ser Lys Asn Asn Val  
210 215 220  
Leu Leu Arg Phe Ala Lys Leu Asp Phe Asn Leu Leu Gln Met Leu His  
225 230 235 240  
Lys Gln Glu Leu Ala Gln Val Ser Arg Trp Trp Lys Asp Leu Asp Phe  
245 250 255  
Val Thr Thr Leu Pro Tyr Ala Arg Asp Arg Val Val Glu Cys Tyr Phe  
260 265 270  
Trp Ala Leu Gly Val Tyr Phe Glu Pro Gln Tyr Ser Gln Ala Arg Val  
275 280 285  
Met Leu Val Lys Thr Ile Ser Met Ile Ser Ile Val Asp Thr Phe  
290 295 300  
Asp Ala Tyr Gly Thr Val Lys Glu Leu Glu Ala Tyr Thr Asp Ala Ile  
305 310 315 320  
Gln Arg Trp Asp Ile Asn Glu Ile Asp Arg Leu Pro Asp Tyr Met Lys  
325 330 335  
Ile Ser Tyr Lys Ala Ile Leu Asp Leu Tyr Lys Asp Tyr Glu Lys Glu  
340 345 350  
Leu Ser Ser Ala Gly Arg Ser His Ile Val Cys His Ala Ile Glu Arg  
355 360 365  
Met Lys Glu Val Val Arg Asn Tyr Asn Val Glu Ser Thr Trp Phe Ile  
370 375 380  
Glu Gly Tyr Thr Pro Pro Val Ser Glu Tyr Leu Ser Asn Ala Leu Ala  
385 390 395 400  
Thr Thr Thr Tyr Tyr Tyr Leu Ala Thr Thr Ser Tyr Leu Gly Met Lys  
405 410 415  
Ser Ala Thr Glu Gln Asp Phe Glu Trp Leu Ser Lys Asn Pro Lys Ile  
420 425 430  
Leu Glu Ala Ser Val Ile Ile Cys Arg Val Ile Asp Asp Thr Ala Thr

ES 2 647 828 T3

```

          435                440                445
Tyr Glu Val Glu Lys Ser Arg Gly Gln Ile Ala Thr Gly Ile Glu Cys
   450                455                460
Cys Met Arg Asp Tyr Gly Ile Ser Thr Lys Glu Ala Met Ala Lys Phe
465                470                475                480
Gln Asn Met Ala Glu Thr Ala Trp Lys Asp Ile Asn Glu Gly Leu Leu
          485                490
Arg Pro Thr Pro Val Ser Thr Glu Phe Leu Thr Pro Ile Leu Asn Leu
          500                505                510
Ala Arg Ile Val Glu Val Thr Tyr Ile His Asn Leu Asp Gly Tyr Thr
          515                520                525
His Pro Glu Lys Val Leu Lys Pro His Ile Ile Asn Leu Leu Val Asp
          530                535                540
Ser Ile Lys Ile
545

```

<210> 25

<211> 520

5 <212> PRT

<213> Hyoscyamus muticus

<220>

<223> Proteína prenaspirodieno sintasa

<300>

10 <308> GenBank: AAA86337.1

<309> 30/01/1996

<400> 25

```

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

```

				260						265					270
Asp	Asp	Thr	Phe	Asp	Ala	Tyr	Gly	Ile	Val	Lys	Glu	Leu	Glu	Val	Tyr
				275				280					285		
Thr	Asp	Ala	Ile	Gln	Arg	Trp	Asp	Ile	Ser	Gln	Ile	Asp	Arg	Leu	Pro
				290				295				300			
Glu	Tyr	Met	Lys	Ile	Ser	Tyr	Lys	Ala	Leu	Leu	Asp	Leu	Tyr	Asp	Asp
				305				310				315			320
Tyr	Glu	Lys	Glu	Leu	Ser	Lys	Asp	Gly	Arg	Ser	Asp	Val	Val	His	Tyr
				325				330							335
Ala	Lys	Glu	Arg	Met	Lys	Glu	Ile	Val	Gly	Asn	Tyr	Phe	Ile	Glu	Gly
				340				345					350		
Lys	Trp	Phe	Ile	Glu	Gly	Tyr	Met	Pro	Ser	Val	Ser	Glu	Tyr	Leu	Ser
				355				360					365		
Asn	Ala	Leu	Ala	Thr	Ser	Thr	Tyr	Tyr	Leu	Leu	Thr	Thr	Thr	Ser	Tyr
				370				375					380		
Leu	Gly	Met	Lys	Ser	Ala	Thr	Lys	Glu	His	Phe	Glu	Trp	Leu	Ala	Thr
				385				390				395			400
Asn	Pro	Lys	Ile	Leu	Glu	Ala	Asn	Ala	Thr	Leu	Cys	Arg	Val	Val	Asp
				405				410							415
Asp	Ile	Ala	Thr	Tyr	Glu	Val	Glu	Lys	Gly	Arg	Gly	Gln	Ile	Ala	Thr
				420				425					430		
Gly	Ile	Glu	Cys	Tyr	Met	Arg	Asp	Tyr	Gly	Val	Ser	Thr	Glu	Val	Ala
				435				440					445		
Met	Glu	Lys	Phe	Gln	Glu	Met	Ala	Asp	Ile	Ala	Trp	Lys	Asp	Val	Asn
				450				455				460			
Glu	Glu	Ile	Leu	Arg	Pro	Thr	Pro	Val	Ser	Ser	Glu	Ile	Leu	Thr	Arg
				465				470				475			480
Ile	Leu	Asn	Leu	Ala	Arg	Ile	Ile	Asp	Val	Thr	Tyr	Lys	His	Asn	Gln
				485				490						495	
Asp	Gly	Tyr	Thr	His	Pro	Glu	Lys	Val	Leu	Lys	Pro	His	Ile	Ile	Ala
				500				505						510	
Leu	Val	Val	Asp	Ser	Ile	Asp	Ile								
				515			520								

<210> 26

<211> 548

5 <212> PRT

<213> Citrus x paradisi

<220>

<223> Proteína valenceno sintasa

<300>

10 <310> US7273735

<311> 25/09/2007

<400> 26

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

ES 2 647 828 T3

		115					120					125					
Asp	Glu	Gly	Arg	Phe	Lys	Ser	Ser	Leu	Ile	Asn	Asp	Val	Gln	Gly	Met		
	130					135					140						
Leu	Ser	Leu	Tyr	Glu	Ala	Ala	Tyr	Met	Ala	Val	Arg	Gly	Glu	His	Ile		
145					150					155					160		
Leu	Asp	Glu	Ala	Ile	Ala	Phe	Thr	Thr	Thr	His	Leu	Lys	Ser	Leu	Val		
				165					170					175			
Ala	Gln	Asp	His	Val	Thr	Pro	Lys	Leu	Ala	Glu	Gln	Ile	Asn	His	Ala		
			180					185					190				
Leu	Tyr	Arg	Pro	Leu	Arg	Lys	Thr	Leu	Pro	Arg	Leu	Glu	Ala	Arg	Tyr		
		195					200					205					
Phe	Met	Ser	Met	Ile	Asn	Ser	Thr	Ser	Asp	His	Leu	Tyr	Asn	Lys	Thr		
210						215					220						
Leu	Leu	Asn	Phe	Ala	Lys	Leu	Asp	Phe	Asn	Ile	Leu	Leu	Glu	Leu	His		
225					230				235						240		
Lys	Glu	Glu	Leu	Asn	Glu	Leu	Thr	Lys	Trp	Trp	Lys	Asp	Leu	Asp	Phe		
				245					250					255			
Thr	Thr	Lys	Leu	Pro	Tyr	Ala	Arg	Asp	Arg	Leu	Val	Glu	Leu	Tyr	Phe		
			260					265					270				
Trp	Asp	Leu	Gly	Thr	Tyr	Phe	Glu	Pro	Gln	Tyr	Ala	Phe	Gly	Arg	Lys		
		275					280					285					
Ile	Met	Thr	Gln	Leu	Asn	Tyr	Ile	Leu	Ser	Ile	Ile	Asp	Asp	Thr	Tyr		
290					295						300						
Asp	Ala	Tyr	Gly	Thr	Leu	Glu	Glu	Leu	Ser	Leu	Phe	Thr	Glu	Ala	Val		
305					310					315					320		
Gln	Arg	Trp	Asn	Ile	Glu	Ala	Val	Asp	Met	Leu	Pro	Glu	Tyr	Met	Lys		
			325						330					335			
Leu	Ile	Tyr	Arg	Thr	Leu	Leu	Asp	Ala	Phe	Asn	Glu	Ile	Glu	Glu	Asp		
			340					345					350				
Met	Ala	Lys	Gln	Gly	Arg	Ser	His	Cys	Val	Arg	Tyr	Ala	Lys	Glu	Glu		
		355					360					365					
Asn	Gln	Lys	Val	Ile	Gly	Ala	Tyr	Ser	Val	Gln	Ala	Lys	Trp	Phe	Ser		
370					375					380							
Glu	Gly	Tyr	Val	Pro	Thr	Ile	Glu	Glu	Tyr	Met	Pro	Ile	Ala	Leu	Thr		
385					390					395					400		
Ser	Cys	Ala	Tyr	Thr	Phe	Val	Ile	Thr	Asn	Ser	Phe	Leu	Gly	Met	Gly		
			405					410						415			
Asp	Phe	Ala	Thr	Lys	Glu	Val	Phe	Glu	Trp	Ile	Ser	Asn	Asn	Pro	Lys		
			420					425					430				
Val	Val	Lys	Ala	Ala	Ser	Val	Ile	Cys	Arg	Leu	Met	Asp	Asp	Met	Gln		
		435					440					445					
Gly	His	Glu	Phe	Glu	Gln	Lys	Arg	Gly	His	Val	Ala	Ser	Ala	Ile	Glu		
450					455					460							
Cys	Tyr	Thr	Lys	Gln	His	Gly	Val	Ser	Lys	Glu	Glu	Ala	Ile	Lys	Met		
465					470					475					480		
Phe	Glu	Glu	Glu	Val	Ala	Asn	Ala	Trp	Lys	Asp	Ile	Asn	Glu	Glu	Leu		
			485					490					495				
Met	Met	Lys	Pro	Thr	Val	Val	Ala	Arg	Pro	Leu	Leu	Gly	Thr	Ile	Leu		
			500					505					510				
Asn	Leu	Ala	Arg	Ala	Ile	Asp	Phe	Ile	Tyr	Lys	Glu	Asp	Asp	Gly	Tyr		
		515					520					525					
Thr	His	Ser	Tyr	Leu	Ile	Lys	Asp	Gln	Ile	Ala	Ser	Val	Leu	Gly	Asp		
			530			535						540					
His	Val	Pro	Phe														
545																	

- <210> 27
- <211> 548
- 5 <212> PRT
- <213> Citrus x paradisi
- <220>
- <223> Proteína valenceno sintasa
- <300>
- 10 <310> US7442785
- <311> 28/10/2008
- <400> 27

ES 2 647 828 T3

Met Ser Ser Gly Glu Thr Phe Arg Pro Thr Ala Asp Phe His Pro Ser  
1 5 10 15  
Leu Trp Arg Asn His Phe Leu Lys Gly Ala Ser Asp Phe Lys Thr Val  
20 25 30  
Asp His Thr Ala Thr Gln Glu Arg His Glu Ala Leu Lys Glu Glu Val  
35 40 45  
Arg Arg Met Ile Thr Asp Ala Glu Asp Lys Pro Val Gln Lys Leu Arg  
50 55 60  
Leu Ile Asp Glu Val Gln Arg Leu Gly Val Ala Tyr His Phe Glu Lys  
65 70 75 80  
Glu Ile Glu Asp Ala Ile Leu Lys Leu Cys Pro Ile Tyr Ile Asp Ser  
85 90 95  
Asn Arg Ala Asp Leu His Thr Val Ser Leu His Phe Arg Leu Leu Arg  
100 105 110  
Gln Gln Gly Ile Lys Ile Ser Cys Asp Val Phe Glu Lys Phe Lys Asp  
115 120 125  
Asp Glu Gly Arg Phe Lys Ser Ser Leu Ile Asn Asp Val Gln Gly Met  
130 135 140  
Leu Ser Leu Tyr Glu Ala Ala Tyr Met Ala Val Arg Gly Glu His Ile  
145 150 155 160  
Leu Asp Glu Ala Ile Ala Phe Thr Thr Thr His Leu Lys Ser Leu Val  
165 170 175  
Ala Gln Asp His Val Thr Pro Lys Leu Ala Glu Gln Ile Asn His Ala  
180 185 190  
Leu Tyr Arg Pro Leu Arg Lys Thr Leu Pro Arg Leu Glu Ala Arg Tyr  
195 200 205  
Phe Met Ser Met Ile Asn Ser Thr Ser Asp His Leu Tyr Asn Lys Thr  
210 215 220  
Leu Leu Asn Phe Ala Lys Leu Asp Phe Asn Ile Leu Leu Glu Pro His  
225 230 235 240  
Lys Glu Glu Leu Asn Glu Leu Thr Lys Trp Trp Lys Asp Leu Asp Phe  
245 250 255  
Thr Thr Lys Leu Pro Tyr Ala Arg Asp Arg Leu Val Glu Leu Tyr Phe  
260 265 270  
Trp Asp Leu Gly Thr Tyr Phe Glu Pro Gln Tyr Ala Phe Gly Arg Lys  
275 280 285  
Ile Met Thr Gln Leu Asn Tyr Ile Leu Ser Ile Ile Asp Asp Thr Tyr  
290 295 300  
Asp Ala Tyr Gly Thr Leu Glu Glu Leu Ser Leu Phe Thr Glu Ala Val  
305 310 315 320  
Gln Arg Trp Asn Ile Glu Ala Val Asp Met Leu Pro Glu Tyr Met Lys  
325 330 335  
Leu Ile Tyr Arg Thr Leu Leu Asp Ala Phe Asn Glu Ile Glu Glu Asp  
340 345 350  
Met Ala Lys Gln Gly Arg Ser His Cys Val Arg Tyr Ala Lys Glu Glu  
355 360 365  
Asn Gln Lys Val Ile Gly Ala Tyr Ser Val Gln Ala Lys Trp Phe Ser  
370 375 380  
Glu Gly Tyr Val Pro Thr Ile Glu Glu Tyr Met Pro Ile Ala Leu Thr  
385 390 395 400  
Ser Cys Ala Tyr Thr Phe Val Ile Thr Asn Ser Phe Leu Gly Met Gly  
405 410 415  
Asp Phe Ala Thr Lys Glu Val Phe Glu Trp Ile Ser Asn Asn Pro Lys  
420 425 430  
Val Val Lys Ala Ala Ser Val Ile Cys Arg Leu Met Asp Asp Met Gln  
435 440 445

Gly His Glu Phe Glu Gln Lys Arg Gly His Val Ala Ser Ala Ile Glu  
 450 455 460  
 Cys Tyr Thr Lys Gln His Gly Val Ser Lys Glu Ala Ile Lys Met  
 465 470 475 480  
 Phe Glu Glu Glu Val Ala Asn Ala Trp Lys Asp Ile Asn Glu Glu Leu  
 485 490 495  
 Met Met Lys Pro Thr Val Val Ala Arg Pro Leu Leu Gly Thr Ile Leu  
 500 505 510  
 Asn Leu Ala Arg Ala Ile Asp Phe Ile Tyr Lys Glu Asp Asp Gly Tyr  
 515 520 525  
 Thr His Ser Tyr Leu Ile Lys Asp Gln Ile Ala Ser Val Leu Gly Asp  
 530 535 540  
 His Val Pro Phe  
 545

<210> 28

<211> 556

5 <212> PRT

<213> Vitis vinifera

<220>

<223> Proteína valenceno sintasa

<300>

10 <308> Uniprot Q6Q3H2

<309> 21/09/2011

<400> 28

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

15

ES 2 647 828 T3

Arg Asp Arg Leu Val Glu Gly Tyr Phe Trp Met His Gly Val Tyr Phe  
 275 280 285  
 Glu Pro Gln Tyr Leu Arg Gly Arg Arg Ile Leu Thr Lys Val Ile Ala  
 290 295 300  
 Met Thr Ser Ile Leu Asp Asp Ile His Asp Ala Tyr Gly Thr Pro Glu  
 305 310 315 320  
 Glu Leu Lys Leu Phe Ile Glu Ala Ile Glu Arg Trp Asp Ile Asn Ser  
 325 330 335  
 Ile Asn Gln Leu Pro Glu Tyr Met Lys Leu Cys Tyr Val Ala Leu Leu  
 340 345 350  
 Asp Val Tyr Lys Glu Ile Glu Glu Met Glu Lys Glu Gly Asn Gln  
 355 360 365  
 Tyr Arg Val His Tyr Ala Lys Glu Val Met Lys Asn Gln Val Arg Ala  
 370 375 380  
 Tyr Phe Ala Glu Ala Lys Trp Leu His Glu Glu His Val Pro Ala Phe  
 385 390 395 400  
 Glu Glu Tyr Met Arg Val Ala Leu Ala Ser Ser Gly Tyr Cys Leu Leu  
 405 410 415  
 Ala Thr Thr Ser Phe Val Gly Met Gly Glu Ile Ala Thr Lys Glu Ala  
 420 425 430  
 Phe Asp Trp Val Thr Ser Asp Pro Lys Ile Met Ser Ser Asn Phe  
 435 440 445  
 Ile Thr Arg Leu Met Asp Asp Ile Lys Ser His Lys Phe Glu Gln Lys  
 450 455 460  
 Arg Gly His Val Thr Ser Ala Val Glu Cys Tyr Met Lys Gln Tyr Gly  
 465 470 475 480  
 Val Ser Glu Glu Gln Val Tyr Ser Glu Phe Gln Lys Gln Ile Glu Asn  
 485 490 495  
 Ala Trp Leu Asp Ile Asn Gln Glu Cys Leu Lys Pro Thr Ala Val Ser  
 500 505 510  
 Met Pro Leu Leu Ala Arg Leu Leu Asn Phe Thr Arg Thr Met Asp Val  
 515 520 525  
 Ile Tyr Lys Glu Gln Asp Ser Tyr Thr His Val Gly Lys Val Met Arg  
 530 535 540  
 Asp Asn Ile Ala Ser Val Phe Ile Asn Ala Val Ile  
 545 550 555

<210> 29

<211> 589

5 <212> PRT

<213> Chamaecyparis nootkatensis

<220>

<223> Proteína valenceno sintasa

<400> 29

10

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

ES 2 647 828 T3

Leu Arg Leu Asn Arg Tyr Val Ala Ser Ser Asp Val Phe Lys Lys Phe  
 130 135 140  
 Lys Gly Glu Glu Gly Gln Phe Ser Gly Phe Glu Ser Ser Asp Gln Asp  
 145 150 155 160  
 Ala Lys Leu Glu Met Met Leu Asn Leu Tyr Lys Ala Ser Glu Leu Asp  
 165 170 175  
 Phe Pro Asp Glu Asp Ile Leu Lys Glu Ala Arg Ala Phe Ala Ser Met  
 180 185 190  
 Tyr Leu Lys His Val Ile Lys Glu Tyr Gly Asp Ile Gln Glu Ser Lys  
 195 200 205  
 Asn Pro Leu Leu Met Glu Ile Glu Tyr Thr Phe Lys Tyr Pro Trp Arg  
 210 215 220  
 Cys Arg Leu Pro Arg Leu Glu Ala Trp Asn Phe Ile His Ile Met Arg  
 225 230 235 240  
 Gln Gln Asp Cys Asn Ile Ser Leu Ala Asn Asn Leu Tyr Lys Ile Pro  
 245 250 255  
 Lys Ile Tyr Met Lys Lys Ile Leu Glu Leu Ala Ile Leu Asp Phe Asn  
 260 265 270  
 Ile Leu Gln Ser Gln His Gln His Glu Met Lys Leu Ile Ser Thr Trp  
 275 280 285  
 Trp Lys Asn Ser Ser Ala Ile Gln Leu Asp Phe Phe Arg His Arg His  
 290 295 300  
 Ile Glu Ser Tyr Phe Trp Trp Ala Ser Pro Leu Phe Glu Pro Glu Phe  
 305 310 315 320  
 Ser Thr Cys Arg Ile Asn Cys Thr Lys Leu Ser Thr Lys Met Phe Leu  
 325 330 335  
 Leu Asp Asp Ile Tyr Asp Thr Tyr Gly Thr Val Glu Glu Leu Lys Pro  
 340 345 350  
 Phe Thr Thr Thr Leu Thr Arg Trp Asp Val Ser Thr Val Asp Asn His  
 355 360 365  
 Pro Asp Tyr Met Lys Ile Ala Phe Asn Phe Ser Tyr Glu Ile Tyr Lys  
 370 375 380  
 Glu Ile Ala Ser Glu Ala Glu Arg Lys His Gly Pro Phe Val Tyr Lys  
 385 390 395 400  
 Tyr Leu Gln Ser Cys Trp Lys Ser Tyr Ile Glu Ala Tyr Met Gln Glu  
 405 410 415  
 Ala Glu Trp Ile Ala Ser Asn His Ile Pro Gly Phe Asp Glu Tyr Leu  
 420 425 430  
 Met Asn Gly Val Lys Ser Ser Gly Met Arg Ile Leu Met Ile His Ala  
 435 440 445  
 Leu Ile Leu Met Asp Thr Pro Leu Ser Asp Glu Ile Leu Glu Gln Leu  
 450 455 460  
 Asp Ile Pro Ser Ser Lys Ser Gln Ala Leu Leu Ser Leu Ile Thr Arg  
 465 470 475 480  
 Leu Val Asp Asp Val Lys Asp Phe Glu Asp Glu Gln Ala His Gly Glu  
 485 490 495  
 Met Ala Ser Ser Ile Glu Cys Tyr Met Lys Asp Asn His Gly Ser Thr  
 500 505 510  
 Arg Glu Asp Ala Leu Asn Tyr Leu Lys Ile Arg Ile Glu Ser Cys Val  
 515 520 525  
 Gln Glu Leu Asn Lys Glu Leu Leu Glu Pro Ser Asn Met His Gly Ser  
 530 535 540  
 Phe Arg Asn Leu Tyr Leu Asn Val Gly Met Arg Val Ile Phe Phe Met  
 545 550 555 560  
 Leu Asn Asp Gly Asp Leu Phe Thr His Ser Asn Arg Lys Glu Ile Gln  
 565 570 575  
 Asp Ala Ile Thr Lys Phe Phe Val Glu Pro Ile Ile Pro  
 580 585

<210> 30

<211> 1674

5 <212> ADN

<213> Secuencia artificial

<220>

<223> ADN de valenceno sintasa cítrica V277

<300>

10 <310> US20120246767

<311> 27/09/2012  
<400> 30

```

atgtctactc aagtctcagc atcttctcta gccagattc cccaacccaa aaatcgtcct 60
gtggc aaact ttcaccccaa ctttgggggt gaccaattta tcacctacac tcctgaagac 120
aaggttactc gtgcctgcaa agaggagcag attgaagctt tgaaggaaga agttagaaga 180
atgatttttag caaccggaag gaaaccaatt caaaaattga gattgattga tgaagttcaa 240
agattgggtg ttgcttacca ttttgaaaaa gaaattgaag atatgttggga tcacatttac 300
agagctgatc cttattttga ggctcatgaa tacaatgatt tgcatactgt ttctttgcat 360
ttcagattgt tgagacaaca aggtattaag atttcttctg atgttttoga acaatttaag 420
gatgatgaag gtagattcaa aagttctttg attaatgatg ttcaaggcat gttgtctttg 480
tatgaagctg cttatatggc tgttagaggt gaacatattt tggatgaagc tattgcattt 540
actactactc atttgcaatc tgcagctcca catttgaagt cacctttggc tgaacaaatt 600
aaccatgctt tgtatagacc attgagaaaa actttgccaa gattggaagc aagatacatt 660
atgtcagctc accaagatga agctttccat aacaagactt tgttaaattt cgctaagttg 720
gatttcaata ttttgttggg tttgcataaa gaagaattga acgaattgac taaatgggtg 780
caagatttgg attttactac taaattgccca tatgctagag atagattgggt tgaactgtac 840
ttttgggatt tggggactta ttttgaatca caatacgtt ttggtagaaa aatcatgact 900
aaattgaact acattttgtc cattattgat gatacctacg atgcttacgg tactttggaa 960
gaatgcacga tgttcagtga agctgttctc cgttgggaaca ttgaagctgt tgatatgttg 1020
ccagattata tgcgaattat ctacagaact ttgttggata cattcaacga aatagaagag 1080
gatatggcta aacaacggag atctcattgt gtaagatacg ctaaagaaga aattcaaaaag 1140
gttattgggtg cttattacgt tcaagctaag tggttttctg aaggttatgt ccctactatt 1200
gaagaataca tgccaattgc tttgacttct tgcgcttaca gatttgttat taccaattct 1260
tttttgggta tgggtgattt cgctacaaag gaagtattcg aatggatttc tggtaatcca 1320
aaagttgtta aatctgcttc tgttatttgt agattgatgg acgatatgca aggcacacgaa 1380
tttgaacaaa aaagaggcca cgttgcactc gcaattgaat gctatactaa acaacatggg 1440
gtttccaagg aagaggctat caagatgttc gaggaagatg ttgctaacgc ttggaaggat 1500
atcaatgaag aattaatgat gaaaccacca gttgttgcta gaccattggt aggtactatt 1560
ttgaatttgg ctagagctat cgattttatc tataaagaag atgacgggta cactcattct 1620
tatttgatta aggaacaaat cgcactctgt ttgggtgatc atgttccatt ttaa 1674

```

5

<210> 31  
<211> 557  
<212> PRT  
<213> Secuencia artificial

10 <220>

<223> Proteína valenceno sintasa cítrica V277

<300>

<310> US20120246767

<311> 27/09/2012

15 <400> 31

```

Met Ser Thr Gln Val Ser Ala Ser Ser Leu Ala Gln Ile Pro Gln Pro
 1          5          10          15
Lys Asn Arg Pro Val Ala Asn Phe His Pro Asn Ile Trp Gly Asp Gln
 20          25          30
Phe Ile Thr Tyr Thr Pro Glu Asp Lys Val Thr Arg Ala Cys Lys Glu
 35          40          45
Glu Gln Ile Glu Ala Leu Lys Glu Glu Val Arg Arg Met Ile Leu Ala
 50          55          60
Thr Gly Arg Lys Pro Ile Gln Lys Leu Arg Leu Ile Asp Glu Val Gln
 65          70          75          80
Arg Leu Gly Val Ala Tyr His Phe Glu Lys Glu Ile Glu Asp Met Leu

```

ES 2 647 828 T3

					85					90					95				
Asp	His	Ile	Tyr	Arg	Ala	Asp	Pro	Tyr	Phe	Glu	Ala	His	Glu	Tyr	Asn				
			100					105					110						
Asp	Leu	His	Thr	Val	Ser	Leu	His	Phe	Arg	Leu	Leu	Arg	Gln	Gln	Gly				
		115						120					125						
Ile	Lys	Ile	Ser	Cys	Asp	Val	Phe	Glu	Gln	Phe	Lys	Asp	Asp	Glu	Gly				
		130						135					140						
Arg	Phe	Lys	Ser	Ser	Leu	Ile	Asn	Asp	Val	Gln	Gly	Met	Leu	Ser	Leu				
145					150							155			160				
Tyr	Glu	Ala	Ala	Tyr	Met	Ala	Val	Arg	Gly	Glu	His	Ile	Leu	Asp	Glu				
					165									175					
Ala	Ile	Ala	Phe	Thr	Thr	Thr	His	Leu	Gln	Ser	Ala	Ala	Pro	His	Leu				
			180					185					190						
Lys	Ser	Pro	Leu	Ala	Glu	Gln	Ile	Asn	His	Ala	Leu	Tyr	Arg	Pro	Leu				
		195						200					205						
Arg	Lys	Thr	Leu	Pro	Arg	Leu	Glu	Ala	Arg	Tyr	Ile	Met	Ser	Val	Tyr				
		210						215					220						
Gln	Asp	Glu	Ala	Phe	His	Asn	Lys	Thr	Leu	Leu	Asn	Phe	Ala	Lys	Leu				
225					230						235			240					
Asp	Phe	Asn	Ile	Leu	Leu	Asp	Leu	His	Lys	Glu	Glu	Leu	Asn	Glu	Leu				
					245					250				255					
Thr	Lys	Trp	Trp	Gln	Asp	Leu	Asp	Phe	Thr	Thr	Lys	Leu	Pro	Tyr	Ala				
		260						265					270						
Arg	Asp	Arg	Leu	Val	Glu	Leu	Tyr	Phe	Trp	Asp	Leu	Gly	Thr	Tyr	Phe				
		275						280					285						
Glu	Ser	Gln	Tyr	Ala	Phe	Gly	Arg	Lys	Ile	Met	Thr	Lys	Leu	Asn	Tyr				
		290						295				300							
Ile	Leu	Ser	Ile	Ile	Asp	Asp	Thr	Tyr	Asp	Ala	Tyr	Gly	Thr	Leu	Glu				
305					310					315					320				
Glu	Cys	Thr	Met	Phe	Ser	Glu	Ala	Val	Ala	Arg	Trp	Asn	Ile	Glu	Ala				
					325					330				335					
Val	Asp	Met	Leu	Pro	Asp	Tyr	Met	Arg	Ile	Ile	Tyr	Arg	Thr	Leu	Leu				
			340					345					350						
Asp	Thr	Phe	Asn	Glu	Ile	Glu	Glu	Asp	Met	Ala	Lys	Gln	Arg	Arg	Ser				
		355						360					365						
His	Cys	Val	Arg	Tyr	Ala	Lys	Glu	Glu	Ile	Gln	Lys	Val	Ile	Gly	Ala				
		370						375				380							
Tyr	Tyr	Val	Gln	Ala	Lys	Trp	Phe	Ser	Glu	Gly	Tyr	Val	Pro	Thr	Ile				
385					390						395			400					
Glu	Glu	Tyr	Met	Pro	Ile	Ala	Leu	Thr	Ser	Cys	Ala	Tyr	Arg	Phe	Val				
					405						410			415					
Ile	Thr	Asn	Ser	Phe	Leu	Gly	Met	Gly	Asp	Phe	Ala	Thr	Lys	Glu	Val				
		420						425					430						
Phe	Glu	Trp	Ile	Ser	Gly	Asn	Pro	Lys	Val	Val	Lys	Ser	Ala	Ser	Val				
		435						440					445						
Ile	Cys	Arg	Leu	Met	Asp	Asp	Met	Gln	Gly	His	Glu	Phe	Glu	Gln	Lys				
		450						455					460						
Arg	Gly	His	Val	Ala	Ser	Ala	Ile	Glu	Cys	Tyr	Thr	Lys	Gln	His	Gly				
465					470					475					480				
Val	Ser	Lys	Glu	Glu	Ala	Ile	Lys	Met	Phe	Glu	Glu	Asp	Val	Ala	Asn				
					485					490				495					
Ala	Trp	Lys	Asp	Ile	Asn	Glu	Glu	Leu	Met	Met	Lys	Pro	Pro	Val	Val				
								505					510						
Ala	Arg	Pro	Leu	Leu	Gly	Thr	Ile	Leu	Asn	Leu	Ala	Arg	Ala	Ile	Asp				
		515						520					525						
Phe	Ile	Tyr	Lys	Glu	Asp	Asp	Gly	Tyr	Thr	His	Ser	Tyr	Leu	Ile	Lys				
		530						535					540						
Glu	Gln	Ile	Ala	Ser	Val	Leu	Gly	Asp	His	Val	Pro	Phe							
545					550					555									

- <210> 32
- <211> 7927
- 5 <212> ADN
- <213> Secuencia artificial
- <220>
- <223> ADN del vector de expresión pALx31-108.2
- <400> 32

cgcggccgcg tggaaatatt cggatatacct tttggtgttt ccgggtgtac aatatggact 60  
 tcctcttttc tggcaaccaa accatacat cgggattcct ataatacctt cgttggctctc 120  
 cctaacatgt aggtggcggg ggggagatat acaatagaac agataccaga caagacataa 180  
 tgggctaaac aagactacac caattacact gcctcattga tgggtgtaca taacgaacta 240  
 atactgtagc cctagacttg atagccatca tcatatcgaa gtttactac cttttttcca 300  
 tttgccatct attgaagtaa taataggcgc atgcaacttc ttttctttt ttttcttttc 360  
 tctctccccc gttgtgtct caccatatac gcaatgacaa aaaaatgatg gaagacacta 420  
 aaggaaaaaa ttaacgacaa agacagcacc aacagatgac gttgttccag agctgatgag 480  
 gggatctctg aagcacacga aactttttcc ttocctcatt cacgcacact actctctaact 540  
 gagcaacggg atacggcctt ccttccagtt acttgaattt gaaataaaaa aaagtttgct 600  
 gtcttgctat caagtataaa tagacctgca attattaatc ttttgtttcc tcgctcattgt 660  
 tctcgttccc tttcttctt gtttcttttt ctgcacaata tttcaagcta tacciaagcat 720  
 acaatcaact ccaagctgaa ttcgagctcg gtacccatca acaagtttgt acaaaaaagc 780  
 aggctatggc ccagctata gtgatgagta actacgaaga ggaggagatt gttcgccctg 840  
 ttgcagattt ttctccaagt ctttgggggtg atcgtttcca ttcatctca gttgacaatc 900  
 aggttgccga aaagtatgct caagagattg aaactttgaa ggaacaaaca agtactatgt 960  
 tgtctgctgc ttgtggaaca acattgactg agaaattgaa tttgatagat attattgagc 1020  
 gccttggaaat agcttatcat ttcgagaaac aaatagaaga tatgttggat cacatttaca 1080  
 gagctgatcc ttattttgag gctcatgaat acaatgattt aaacacttca tccgttcaat 1140  
 ttcgactact cagacaacat ggttacaacg tctctccaaa tatatttagc agattccaag 1200  
 atgcaaatgg caaattcaag gactctctta gaagcgacat caggggcta ctgaactta 1260  
 acgaagcttc acatgtaagg actcataaag aggatatttt ggaagaagca cttgtttttt 1320  
 ctggttggta tcttgaatct gcagctccac atttgaagtc acctctgagt aagcaagtga 1380  
 cacatgccct cgaacaatct ctccataaga gcattccaag agtcgagata cggtacttca 1440  
 tctccatcta cgaagaggag gaatttaaga atgatttgtt gcttcgattt gctaaattgg 1500  
 attacaactt acttcagatg ttgcacaagc atgaactcag tgaagtatca aggtgggtga 1560  
 aagatttggg tttcgtgaca acacttccat atgctagggg cagagcagtt gactgctact 1620  
 tttggacgat ggggggtgat gctgaacctc aatactccca ggctcgtgtc atgcttgcta 1680  
 agactatagc aatgatttcc atagtagatg acacattcga tgcttatggc atcgtaaaag 1740  
 aacttgaagt ctacacagat gccatacaga ggtgggatat tagtcaaatt gatcgactcc 1800  
 cggtaatatg gaaatcagt tataaggctc ttttggatct ctatgacgat tatgaaaagg 1860  
 agttgtcaaa ggatggtaga tccgatgttg tccactatgc aaaagaaaga atgaaggaga 1920  
 ttgtgagaaa ctattttatt gaagcaaaat ggtttattga gggatatatg ccatctgttt 1980  
 ccgagtacct tagcaatgca ctgactacta gcacatatta cttgctaact acgacatcct 2040  
 acttgggaat gaaatcagca accaaggaac attttgaatg gttggctacg aaccctaaa 2100  
 ttctggaagc taatgctaca ttatgccgag ttgttggatga catagccacg tatgaggttg 2160  
 agaagggtag ggggtcaaat gcaacaggaa ttgagtggtta tatgagggat tacgggtgat 2220  
 ccacagaagt agcaatggag aaattccaag aaatggctga catagcatgg aaggatgtaa 2280  
 atgaagaaat tcttcgacca acactgtct ctccagaaat tcttactcgt attctcaacc 2340  
 tcgctcgaat tatagatgct acttacaagc ataatacaaga tggatacact catcctgaa 2400  
 aagtactaaa acctcacatc atcgcttgg tgggtgattc tattgatatt tgaaccgac 2460  
 tttcttgtac aaagtggttg atggctcttc tagaatctct gcttttgtgc gcgatgttt 2520  
 atgtatgtac ctctctctct atttctatct ttaaaccacc ctctcaataa aataaaaaata 2580  
 ataaagtatt ttttaaggaaa agacgtgttt aagcactgac tttatctact ttttgtactg 2640  
 tttcattgat ataatgtgtt ttgtctctcc cttttctacg aaaatttcaa aaattgacca 2700  
 aaaaaaggaa tatataatag aaaaactatt atatttatat atcatagtgt tgataaaaaa 2760  
 tgtttatcca ttggaccgtg tatcagggcc cggatcctct aggcttggca ctggccgtcg 2820  
 ttttacaacg tcgtgactgg gaaaaccctg gcgttaccga acttaatcgc cttgcagcac 2880  
 atccccctt cgccagctgg cgtaatagcc aagaggcccg caccgatcgc ccttcccaac 2940  
 agttgcgcag cctgaatggc gaatggcgcc tgatgcggtg ttttctcctt acgcactctg 3000  
 ggggtatttc acaccgata gggtaataac tgatataatt aaattgaagc tctaatttgt 3060  
 gagtttagta tacatgcatt tacttataat acagtttttt agttttgctg gccgcactct 3120  
 ctcaaatatg cttcccagcc tgcttttctg taacgttcac cctctacctt agcatccctt 3180  
 ccctttgcaa atagtcctct tccaacaata ataatgtcag atcctgtaga gaccacatca 3240  
 tcacgggttc tatactgttg acccaatgcg tctcccttgt catctaaacc cacaccgggt 3300  
 gtcataatca accaatcgta accttcatct cttccacca tgtctctttg agcaataaag 3360

ES 2 647 828 T3

ccgataacaa aatccttgtc gctccttcgca atgtcaacag tacccttagt atattctcca 3420  
 gtagataggg agcccttgca tgacaattct gtaaacatca aaaggcctct aggttccttt 3480  
 gttacttctt ctgccgcctg cttcaaaccg ctaacaatac ctggggccac cacaccgtgt 3540  
 gcattogtaa tgtctgccca ttctgtctatt ctgtatacac cgcgagagta ctgcaatttg 3600  
 actgtattac caatgtcagc aaatctctgt tcttcggaaga gtaaaaaatt gtacttgccg 3660  
 gataatgcct ttagcggctt aactgtgccc tccatggaaa aatcagtcac gatatccaca 3720  
 tgtgttttta gtaaacaaat tttgggacct aatgcttcaa ctaactccag taattccttg 3780  
 gtggtacgaa catccaatga agcacacaag tttgtttgct tttcgtgcat gatattaaat 3840  
 agcttggcag caacaggact aggatgagta gcagcacggt ccttatatgt agctttcgac 3900  
 atgatttatc ttcgcttcgg tttttgttct gtgcagttgg gttagaataa ctgggcaatt 3960  
 tcatgtttct tcaacactac atatgcgtat atataccaat ctaagtctgt gctccttctc 4020  
 tctgttctcc ttctgttcgg agattaccga atcaaaaaaa tttcaaggaa accgaaatca 4080  
 aaaaaagaa taaaaaaaaa atgatgaatt gaaaagcact tgttaccat cattgaattt 4140  
 tgaacatccg aacctgggag tttccctga aacagatagt atatttgaac ctgtataata 4200  
 atatatagct tagcgtttta cggaaagaaa tgtatgtatt tgggttcctg gaaagaaat 4260  
 tgcattctatt gcataggtaa tcttgacagt cgcattcccc gtctattttc tgcgtttcca 4320  
 tcttgacatt caatagcata tctttgttaa cgaagcact gtgcttcatt ttgtaaaaaa 4380  
 aaaatgcaac gcgagagcgc taatttttca acaaaagaat ctgagctgca tttttacaga 4440  
 acagaaatcc aaccgaaagc cgctatttta ccaacgaaaga atctgtgctt catttttga 4500  
 aaacaaaaat gcaacgcgag agcgtcaatt tttcaaaaaa agaactctgag ctgcattttt 4560  
 acagaacaga aatgcaacgc gagagcgtca ttttaccac aaagaactca tacttctttt 4620  
 ttgttctaca aaaaatgcac cggagagcgc tatttttcta acaaagcact ttagattact 4680  
 ttttttctcc tttgtgcgct ctataatgca gtctcttgat aactttttgc actgtagtct 4740  
 cgttaagggt agaagaaggc tactttggtg tctattttct cttccataaa aaaagcctga 4800  
 ctccacttcc cgcgtttact gattactagc gaagctcgg gtgcattttt tcaagataaa 4860  
 ggcattcccc attatattct ataccgatgt ggattgcgca tactttgtga acagaaagt 4920  
 atagcgttga tgattcttca ttggtcagaa aattatgaac ggtttctctc atttgtctc 4980  
 tataactac gtataggaaa tgtttacatt ttogtattgt tttcgattca ctctatgaat 5040  
 agttcttact acaatttttt tgtctaaaga gtaatactag agataaacat aaaaaatgta 5100  
 gaggtcagat tttagatgcaa gttcaaggag cgaagggtg atgggtaggt tatatagga 5160  
 tatattgacag agatatatag caaagagata cttttgagca atggtttggt aagcggatt 5220  
 cgcaatatt tagtagctcg ttacagctcg gtgctttttt ggttttttga aagtgcgtct 5280  
 tcagagcgtc tttggttttc aaaagcgtc tgaagttcct atactttcta gctagagaat 5340  
 aggaacttcc gaataggaac ttcaaagcgt tccgaaaaac gagcgttcc gaaaatgcaa 5400  
 cgcgagctcc gcacatacag ctccactgttc acgtcgcacc tatatctgct gtttctgt 5460  
 atatatatat acatgagaag aacggcatag tgcgtgttta tgcttaaatg cgttatggtg 5520  
 cactctcagt acaatctgct ctgatgccgc atagttaaag cagccccgac acccgcaca 5580  
 acccgtgac gcgccctgac gggcttgtct gctcccggca tccgcttaca gacaagctgt 5640  
 gaccgtctcc gggagctgca tgtgtcagag gttttcaccg tcatcaccga aacgcgctg 5700  
 acgaaagggc ctctgtgatac gcctattttt ataggttaat gtcatgataa taatggtttc 5760  
 tttagcgtca ggtgacatt ttccgggaaa tgtgcccgga acccctattt gtttattttt 5820  
 ctaaatatct caaatatgt atccgctcat gagacaataa cctgataaa tgcctcaata 5880  
 atattgaaaa aggaagagta tgagatttca acatttccgt gtgccctta tccctttt 5940  
 tgcggcaatt tgccttctc tttttgctca cccagaaaag ctggtgaaag taaaagatgc 6000  
 tgaagatcag ttgggtgcac gagtgggtta catcgaactg gatctcaaca cgggtaagat 6060  
 ccttgagagt tttcgcctcc aagaacgctt tccaatgat agcactttta aagttctgct 6120  
 atgtggcgcg gtattatccc gtattgacgc cgggcaagag caactcgtc gccgataca 6180  
 ctattctcag aatgacttgg ttgagtactc accagtcaca gaaaagcact ttacggatgg 6240  
 catgacagta agagaattat gcagtgctgc cataaccatg agtgataaca ctgcccga 6300  
 ctacttctg acaacgatcg gaggaccgaa ggagctaacc gcttttttgc acaacatggg 6360  
 ggatcatgta actcgccttg atcgttggga accggagctg aatgaagcca taccaaaacg 6420  
 cgagcgtgac accacgatgc ctgtagcaat ggcaacaacg ttgcccgaac tattaactgg 6480  
 cgaactactt actctagctt ccccgcaaca attaatagac tggatggagg cggataaagt 6540  
 tgcaggacca cttctgocct cggcccttcc ggtggctgg tttattgctg ataaactctg 6600  
 agccgggtgag cgtgggtctc gcggtatcat tgcagcactg gggccagatg gtaagccctc 6660  
 ccgtatcgta gttatctaca cgacggggag tcaggcaact atggatgaa gaaatagaca 6720  
 gatcgctgag ataggtgcct cactgattaa gcattggtaa ctgtcagacc aagtttactc 6780  
 atataactt tagattgatt taaaacttca tttttaattt aaaaggatct aggtgaagat 6840  
 cctttttgat aatctcatga ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc 6900  
 agaccocgta gaaaagatca aaggatcttc ttgagatcct tttttctg cgtaatctg 6960  
 ctgcttgcac acaaaaaaac caccgctacc agcgggtggt tgtttgccc atcaagact 7020  
 accaactctt tttccgaag taactggctt cagcagagcg cagataccaa atactgttct 7080  
 tctaagttag ccgtagttag gccaccactt caagaactct gtagcaccgc ctacatacca 7140

ES 2 647 828 T3

```

cgctctgcta atcctgttac cagtggctgc tgccagtggc gataagtcgt gtcttaccgg 7200
gttgactca agacgatagt taccggataa ggcgacggc tccgggtgaa cgggggggtc 7260
gtgcacacag cccagcttgg agcgaacgac ctacaccgaa ctgagatacc tacagcgtga 7320
gctatgagaa agcgcacgac ttcccgaagg gagaaaggcg gacaggtatc cggtaagcgg 7380
cagggtcggg acaggagagc gcacgagggg gcttccaggg ggaaacgcct ggtatcttta 7440
tagtcctgtc gggtttcgcc acctctgact tgagcgtcga tttttgtgat gctcgtcagg 7500
ggggcggagc ctatggaaaa acgccagcaa cgcggccttt ttacggttcc tggccttttg 7560
ctggcctttt gctcacatgt tcttctctgc gttatcccct gattctgtgg ataaccgtat 7620
taccgccttt gagttagctg ataccgctcg ccgcagccga acgaccgagc gcagcagatc 7680
agtgagcggg gaagcgggag agcgcaccaat acgcaaaccg cctctccccg cgcgttggcc 7740
gattcattaa tgcagctggc acgacaggtt tcccgactgg aaagcgggca gtgacgcaac 7800
gcaattaatg tgagttagct cactcattag gcaccccagg ctttacactt tatgcttccg 7860
gctcgtatgt tgtgtggaat tgtgagcggg taacaatttc acacaggaaa cagctatgac 7920
atgatta
7972

```

<210> 33

<211> 7902

5 <212> ADN

<213> Secuencia artificial

<220>

<223> ADN del constructo de expresión pALX63-71

<400> 33

10

```

cgcggccgcg tggaaatatt cggatatacct tttgttgttt ccgggtgtac aatatggact 60
tcctcttttc tggcaaccaa acccatacat cgggattcct ataatacctt cgttggcttc 120
cctaacatgt aggtggcggg ggggagatat acaatagaac agataccaga caagacataa 180
tgggctaaac aagactacac caattacact gcctcattga tgggtgtaca taacgaacta 240
atactgtagc cctagacttg atagccatca tcatatcgaa gtttactac cctttttcca 300
tttgccatct attgaagtaa taataggcgc atgcaacttc ttttctttt ttttctttc 360
tctctccccg gttgtgtct caccatattc gcaatgacaa aaaaatgatg gaagacacta 420
aaggaaaaaa ttaacgacaa agacagcacc aacagatgtc gttgttccag agctgatgag 480
gggtatctcg aagcacacga aactttttcc ttcttctatt cacgcacact actctctaact 540
gagcaacggt atacgcctt ccttccagtt acttgaattt gaaataaaaa aaagtttgct 600
gtcttgctat caagtataaa tagacctgca attattaatc ttttgtttc tcgctcattgt 660
tctcgttccc tttcttccct gtttcttttt ctgcacaata tttcaagcta taccgaagct 720
acaatcaact ccaagctgaa ttcgagctcg gtaccattaa aaaaaatgtc tcttaatgta 780
cttagtacgt caggttcagc tccaacaacc aaatcatctg agattactcg taggtccgct 840
aattatcatc ctagtattat gggagacaag ttctcgaat attcagagcc agatcacctg 900
aaaaatgatt cattcacaga aaagaacat gaacaactca aagaagaggt gaagaagatg 960
ctagttagaaa cggttcaaaa gcctcaacaa cagctgaatc tgatcaacga aatacaacga 1020
ctaggtttat catacctttt tgaacccgaa attgaggctg cattgcagga aatcagtggt 1080
acctatgatg aatttgttg tagtacagac gctgatgacc ttcacaatgt tgctctctct 1140
ttccgaatac ttagagaaca tggacataat gtatcttctg atgtgtttca gaaattcatg 1200
gatagcaatg ggaagtgtga agactacttg gttaatgatg ctgagaggact gttaaagctg 1260
tacgaagcaa cacatttctg ggttcataat gatgataaac ttgaagagtt gctgtcagta 1320
acaacctctc gtcttgagca tctcaaatcc cacgtgaagt accctcttga ggacgaaatc 1380
agtagagcac ttaagcatcc cctccataaa gaactaaatc gactaggagc gagatattac 1440
atatccattt acgaaaaatt tgattcacac aataaattgc ttttgaggtt tgcaaaaacta 1500
gattttaacc gactgcagaa aatgtatcaa catgagctag cccaccttac aaggtgggtg 1560
aaagatttag attttcaaaa caaacttcca tttgcaagag atagaattgt tgagggttac 1620
ttttggatct taggaatgta ctttgagcca gaacgtaagg atgtcagggg attcttgaac 1680
agagtatttg cacttattac agtagttgat gacacgtatg atgtgtatgg tacattcaaa 1740
gaacttctac tgttctactga tgcaattgaa agatggggaa ctagtattt ggatcagcta 1800
ccgggatata tgagaattat ttatcaagct ctcatggatg tttataatca aatggaggaa 1860
aagttgtcaa tgaaagctga ttgtccaaca taccgtcttg agtttgcaat agaaaacagtt 1920
aaagccatgt tcagatcata cctcgaagaa gctagatggg ccaaagaaca ttatatcca 1980
tcgatggaag agtatatgac cgtggcactg gtatcggttg gctacaaaac catattaact 2040
aattcctttg ttggaatggg ggatattgca acacgggaag tttttgagtg ggtgttcaat 2100
agtccattga ttattagagc ttccgactta attgccagat tgggagatga tattggaggc 2160
catgaggagg agcagaagaa aggagacgca gccactgcta tcgagtgtta cataaaagag 2220
aatcatgtaa caaagcatga agcttatgat gaatttcaga acaaaattga taatgcttgg 2280
aaggatttga ataaggaagc tctacgtcca tttcctgttc caatgacttt catcacaaga 2340

```

ES 2 647 828 T3

gttgttcatt ttacgcgcgc catacatggt atttatgccg actttagtga tggttacaca 2400  
 cgttcagaca aggcgatcag aggttacata acttcactgc tcgtggatcc tattcctttg 2460  
 taatctagaa tctctgcttt tgtgcgcgta tgtttatgta tgtacctctc tctctatttc 2520  
 ttttttaaa ccacctctc aataaaataa aaataataaa gtatttttaa ggaaaagacg 2580  
 tgtttaagca ctgactttat ctactttttg tacgttttca ttgatataat gtgtttttgc 2640  
 tctccctttt ctacgaaaat ttcaaaaatt gaccaaaaaa aggaatata ataccgaaaa 2700  
 ctattatatt tatatatcat agtgttgata aaaaatgttt atccattgga cegtgtatca 2760  
 gggcccggt cctctaggct tggcactggc cgtcgtttta caacgtcgtg actgggaaaa 2820  
 ccctggcggt acccaactta atcgccttgc agcacatccc cctttcgcca gctggcgtaa 2880  
 tagcgaagag gcccgcaccg atcgccttcc ccaacagtgg cgcagcctga atggcgaatg 2940  
 ggcctgatg cggatatttc tccttacgca tctgtgcggt atttcacacc gcatagggta 3000  
 ataactgata taattaaatt gaagctctaa tttgtgagtt tagtatacat gcatttactt 3060  
 ataatacagt tttttagttt tgctggccgc atcttctcaa atatgcttcc cagcctgctt 3120  
 ttctgtaacg ttcaccctct accttagcat cccttccctt tgcaaatagt cctcttcaa 3180  
 caataatatt gtcagatcct gtagagacca catcatccac ggttctatc tgttgacca 3240  
 atgcgtctcc cttgtcatct aaaccacac cgggtgtcat aatcaaccaa tcgtaacctt 3300  
 catctcttcc acccatgtct ctttgagcaa taaagccgat aacaaaatct ttgtcgcctc 3360  
 tcgcaatgtc aacagtacc ttagtatatt ctccagtaga tagggagccc ttgcatgaca 3420  
 attcctcaat catcaaaagg cctctaggtt cctttgttac ttcttctgcc cctgacctca 3480  
 aaccgtaac aatacctggg cccaccacac cgtgtgcatt cgtaatgtct gccattctg 3540  
 ctattctgta tacaccgca gagtactgca atttgactgt attaccaatg tcagcaaat 3600  
 ttctgtcttc gaagagtaaa aaattgtact tggcggataa tgcctttagc ggttaactg 3660  
 tgcctccat ggaaaaatca gtcaagatat ccacatgtgt ttttagtaaa caaattttg 3720  
 gacctaatgc ttcaactaac tccagtaatt ccttgggtgg acgaacatcc aatgaagcac 3780  
 acaagtttgt ttgcttttcc tgcatgatataaaatagctt ggcagcaaca gactagat 3840  
 gagtagcagc acgttcotta tatgtagctt tcgacatgat ttatcttctg ttccggtttt 3900  
 gttctgtgca gttgggttaa gaatactggg caatttcatg ttcttcaac actacatg 3960  
 cgtatatata ccaatctaag tctgtgctcc ttcttctggt cttcttctg ttccggagatt 4020  
 accgaatcaa aaaaatttca aggaaaccga aatcaaaaaa aagaataaaa aaaaaatgat 4080  
 gaattgaaaa gcacttgta cccatcattg aattttgaa atccgaacct gggagtttc 4140  
 cctgaacag atagtatatt tgaacctgta taataatata tagtctagcg ctttacgaa 4200  
 gacaatgtat gtatttgggt tcctggagaa actattgcat ctattgcata ggtaatctt 4260  
 cacgtcgcac ccccggttca ttttctgcgt ttccatcttg cacttcaata gcatatctt 4320  
 gttaacgaag catctgtgct tcattttgta aaacaaaaat gcaacgcgag agcgctaatt 4380  
 tttcaacaa agaactctgag ctgcattttt acagaacaga aatgcaacgc gaaagcgtta 4440  
 ttttaaccaac gaagaatctg tgcttcattt ttgtaaaaca aaaatgcaac gcgagagcgc 4500  
 taatttttca aacaagaat ctgagctgca tttttacaga acagaaatgc aacgcgagag 4560  
 cgctatttta ccaacaaga atctatact cttttttgt ctacaaaaat gcatcccag 4620  
 agcgtatatt ttctaacaaa gcactttaga ttactttttt tctccttctg gcgctctata 4680  
 atgcagctctc ttgataactt tttgcaactg aggtccgtta aggttagaag aaggctactt 4740  
 tgggtgctat tttctcttcc ataaaaaag cctgactcca cttcccgcgt ttaactgata 4800  
 ctagcgaagc tgcgggtgca ttttttcaag ataaaggcat ccccgattat attctatacc 4860  
 gatgtggatt tgtgaacaga aagtgatagc gttgatgatt cttcattgat 4920  
 cagaaaatta tgaacggttt cttctatatt gtctctatat actacgtata ggaatgttt 4980  
 acattttctg attgttttcc atctactcta tgaatagttc ttactacaat ttttttct 5040  
 aaagagtaat actagagata aacataaaaa atgtagaggt cgagttaga tgcaagttca 5100  
 aggagcgaag ggtggatggg taggttatat agggatatag cacagagata tatagcaag 5160  
 agatactttt gagcaatggt tgtggaagcg gtattcgcaa tatttttagta gctcgttaca 5220  
 gtcgggtgcg tttttgggtt tttgaaagtg cgtcttcaga gcgcttttg ttttcaaaag 5280  
 cgctctgaag ttcctatact ttctagctag agaataggaa cttcgggata ggaacttcaa 5340  
 agcgtttccg aaaaacgagc cttccgaaaa tgcaacgcga gctgcgcaca tacagctcac 5400  
 tgttcacgtc gcacctatat ctgctgtgtg cctgtatata tatatacatg agaagaacgg 5460  
 catagtgcgt gtttatgctt aatgcgttca tgggtgactc tcagtacaat ctgctctgat 5520  
 gccgcatagt taagccagcc cggacaccgc ccaacaccgc ctgacgcgcc ctgacgggct 5580  
 tgtctgctcc cggcatccgc ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt 5640  
 cagaggtttt caccgtcatc accgaaacgc gcgagacgaa aggcctcgt gatagccta 5700  
 tttttatagg ttaatgtcat gataataatg gtttcttaga cgtcaggtgg cacttttcgg 5760  
 ggaaatgtgc gcggaacccc tattttgtta tttttctaaa tacattcaa tatgtatccg 5820  
 ctcatgagac aataaccctg ataaatgctt caataatatt gaaaaaggaa gagtatgagt 5880  
 attcaacatt tccgtgtcgc cottattccc ttttttgagg cattttgcct tctgttttt 5940  
 gctcaccag aaacgctggt gaaagtaaaa gatgctgaag atcagttggg tgcacgagtg 6000  
 ggttacatcg aactggatct caacagcggg aagatccttg agagttttgc ccccgagaa 6060  
 cgttttccaa tgatgagcac ttttaaagtt ctgctatgtg gcgcggtatt atcccgtatt 6120

ES 2 647 828 T3

```

gacgccgggc aagagcaact cggtcgccgc atacactatt ctcagaatga cttgggtgag 6180
tactcaccag tcacagaaaa gcactcttac gatggcatga cagtaagaga attatgcagt 6240
gctgccataa ccatgagtga taacactgcg gccaaactac ttctgacaac gatcgggagga 6300
ccgaaggagc taaccgcttt tttgcacaac atgggggatc atgtaactcg ccttgatcgt 6360
tggaaccgag agctgaatga agccatacca aacgacgagc gtgacaccac gatgcctgta 6420
gcaatggcaa caacgttgcg caaactatta actggcgaac tacttactct agcttccccg 6480
caacaattaa tagactggat ggaggcggat aaagttgcag gaccacttct gcgctcggcc 6540
cttccggctg gctggtttat tgctgataaa tctggagccg gtgagcgtgg gtctcgcggt 6600
atcattgcag cactggggcc agatggtaag ccctcccgta tcgtagttat ctacacgacg 6660
gggagtcagg caactatgga tgaacgaaat agacagatcg ctgagatagg tgctcactg 6720
attaagcatt ggtaactgtc agaccaagtt tactcatata tactttagat tgatttaaaa 6780
cttcattttt aattttaaag gatctaggtg aagatccttt ttgataatct catgaccaa 6840
atccctaac gtgagttttc gttccactga cgtcagacc ccgtagaaaa gatcaaagga 6900
tcttcttgag atcctttttt tctgcgcgta atctgctgct tgcaaaaaaaa aaaaccaccg 6960
ctaccagcgg tggtttgttt gccggatcaa gagctaccaa ctctttttcc gaaggtaact 7020
ggcttcagca gagcgcagat accaaatact gttcttctag tgtagccgta gttaggccac 7080
cacttcaaga actctgtagc accgcctaca taccacgctc tgctaactct gtaccagtg 7140
gctgctgcc a gtggcgataa gtcgtgtctt accgggttgg actcaagacg atagttaccg 7200
gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca cacagcccag cttggagcga 7260
acgacctaca ccgaactgag atacctacag cgtgagctat gagaagcgc cacgcttccc 7320
gaagggagaa aggcggacag gtatccggta agcggcaggg tcggaacagg agagcgcacg 7380
aggagcttc cagggggaaa cgcctggtat ctttatagtc ctgtcgggtt tcgccacctc 7440
tgacttgagc gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc 7500
agcaacgcgg cctttttacg gttcctggcc ttttgctggc cttttgctca catgttcttt 7560
cctgcgttat cccctgattc tgtggataac cgtattaccg cctttgagtg agctgatacc 7620
gctcgcgcga gccgaacgac cgagcgcagc gagtcagtga gcgaggaagc ggaagagcgc 7680
ccaatacgc aaccgcctct ccccgcgctg tggccgattc attaatgcag ctggcagcag 7740
aggtttcccg actggaaagc gggcagtgac gcaacgcaat taatgtgagt tagctcactc 7800
attaggcacc ccaggcttta cactttatgc ttccggctcg tatgttgtgt ggaattgtga 7860
cgggataaca atttcacaca ggaacagct atgacatgat ta 7920

```

<210> 34

<211> 548

5 <212> PRT

<213> Citrus sinensis

<220>

<223> Proteína valenceno sintasa cítrica

<300>

10 <308> Genbank AAQ04608

<309> 12/01/2004

<400> 34

```

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

```

ES 2 647 828 T3

```

145                               150                               155                               160
Leu Asp Glu Ala Ile Ala Phe Thr Thr Thr His Leu Lys Ser Leu Val
                               165                               170                               175
Ala Gln Asp His Val Thr Pro Lys Leu Ala Glu Gln Ile Asn His Ala
                               180                               185                               190
Leu Tyr Arg Pro Leu Arg Lys Thr Leu Pro Arg Leu Glu Ala Arg Tyr
                               195                               200                               205
Phe Met Ser Met Ile Asn Ser Thr Ser Asp His Leu Cys Asn Lys Thr
                               210                               215                               220
Leu Leu Asn Phe Ala Lys Leu Asp Phe Asn Ile Leu Leu Glu Leu His
225                               230                               235                               240
Lys Glu Glu Leu Asn Glu Leu Thr Lys Trp Trp Lys Asp Leu Asp Phe
                               245                               250                               255
Thr Thr Lys Leu Pro Tyr Ala Arg Asp Arg Leu Val Glu Leu Tyr Phe
                               260                               265                               270
Trp Asp Leu Gly Thr Tyr Phe Glu Pro Gln Tyr Ala Phe Gly Arg Lys
                               275                               280                               285
Ile Met Thr Gln Leu Asn Tyr Ile Leu Ser Ile Ile Asp Asp Thr Tyr
                               290                               295                               300
Asp Ala Tyr Gly Thr Leu Glu Glu Leu Ser Leu Phe Thr Glu Ala Val
305                               310                               315                               320
Gln Arg Trp Asn Ile Glu Ala Val Asp Met Leu Pro Glu Tyr Met Lys
                               325                               330                               335
Leu Ile Tyr Arg Thr Leu Leu Asp Ala Phe Asn Glu Ile Glu Glu Asp
                               340                               345                               350
Met Ala Lys Gln Gly Arg Ser His Cys Val Arg Tyr Ala Lys Glu Glu
                               355                               360                               365
Asn Gln Lys Val Ile Gly Ala Tyr Ser Val Gln Ala Lys Trp Phe Ser
                               370                               375                               380
Glu Gly Tyr Val Pro Thr Ile Glu Glu Tyr Met Pro Ile Ala Leu Thr
385                               390                               395                               400
Ser Cys Ala Tyr Thr Phe Val Ile Thr Asn Ser Phe Leu Gly Met Gly
                               405                               410                               415
Asp Phe Ala Thr Lys Glu Val Phe Glu Trp Ile Ser Asn Asn Pro Lys
                               420                               425                               430
Val Val Lys Ala Ala Ser Val Ile Cys Arg Leu Met Asp Asp Met Gln
                               435                               440                               445
Gly His Glu Phe Glu Gln Lys Arg Gly His Val Ala Ser Ala Ile Glu
                               450                               455                               460
Cys Tyr Thr Lys Gln His Gly Val Ser Lys Glu Glu Ala Ile Lys Met
465                               470                               475                               480
Phe Glu Glu Glu Val Ala Asn Ala Trp Lys Asp Ile Asn Glu Glu Leu
                               485                               490                               495
Met Met Lys Pro Thr Val Val Ala Arg Pro Leu Leu Gly Thr Ile Leu
                               500                               505                               510
Asn Leu Ala Arg Ala Ile Asp Phe Ile Tyr Lys Glu Asp Asp Gly Tyr
                               515                               520                               525
Thr His Ser Tyr Leu Ile Lys Asp Gln Ile Ala Ser Val Leu Gly Asp
                               530                               535                               540
His Val Pro Phe
545

```

<210> 35

<211> 5

5 <212> PRT

<213> Secuencia artificial

<220>

<223> Motivo de la región 1 rico en aspartato

<220>

10 <221> VARIANTE

<222> 3, 4

<223> Xaa= cualquier aminoácido

<400> 35

Asp Asp Xaa Xaa Asp

1

5

15

- <210> 36
- <211> 9
- <212> PRT
- <213> Secuencia artificial
- 5 <220>
- <223> Motivo NSD/DTE
- <220>
- <221> VARIANTE
- <222> 1
- 10 <223> Xaa= Asn o Asp
- <220>
- <221> VARIANTE
- <222> 2, 3, 4, 6, 7, 8
- <223> Xaa= cualquier aminoácido
- 15 <220>
- <221> VARIANTE
- <222> 5
- <223> Xaa= Ser o Thr
- <400> 36

20

Xaa	Glu								
1				5					

- <210> 37
- <211> 548
- 25 <212> PRT
- <213> Citrus sinensis
- <220>
- <223> Valenceno sintasa A517I, I518V
- <300>
- 30 <308> GenBank AAQ04608.1
- <309> 12/01/2004
- <400> 37

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

35

ES 2 647 828 T3

Gln Gln Gly Ile Lys Ile Ser Cys Asp Val Phe Glu Lys Phe Lys Asp  
 115 120 125  
 Asp Glu Gly Arg Phe Lys Ser Ser Leu Ile Asn Asp Val Gln Gly Met  
 130 135 140  
 Leu Ser Leu Tyr Glu Ala Ala Tyr Met Ala Val Arg Gly Glu His Ile  
 145 150 155 160  
 Leu Asp Glu Ala Ile Ala Phe Thr Thr Thr His Leu Lys Ser Leu Val  
 165 170 175  
 Ala Gln Asp His Val Thr Pro Lys Leu Ala Glu Gln Ile Asn His Ala  
 180 185 190  
 Leu Tyr Arg Pro Leu Arg Lys Thr Leu Pro Arg Leu Glu Ala Arg Tyr  
 195 200 205  
 Phe Met Ser Met Ile Asn Ser Thr Ser Asp His Leu Cys Asn Lys Thr  
 210 215 220  
 Leu Leu Asn Phe Ala Lys Leu Asp Phe Asn Ile Leu Leu Glu Leu His  
 225 230 235 240  
 Lys Glu Glu Leu Asn Glu Leu Thr Lys Trp Trp Lys Asp Leu Asp Phe  
 245 250 255  
 Thr Thr Lys Leu Pro Tyr Ala Arg Asp Arg Leu Val Glu Leu Tyr Phe  
 260 265 270  
 Trp Asp Leu Gly Thr Tyr Phe Glu Pro Gln Tyr Ala Phe Gly Arg Lys  
 275 280 285  
 Ile Met Thr Gln Leu Asn Tyr Ile Leu Ser Ile Ile Asp Asp Thr Tyr  
 290 295 300  
 Asp Ala Tyr Gly Thr Leu Glu Glu Leu Ser Leu Phe Thr Glu Ala Val  
 305 310 315 320  
 Gln Arg Trp Asn Ile Glu Ala Val Asp Met Leu Pro Glu Tyr Met Lys  
 325 330 335  
 Leu Ile Tyr Arg Thr Leu Leu Asp Ala Phe Asn Glu Ile Glu Glu Asp  
 340 345 350  
 Met Ala Lys Gln Gly Arg Ser His Cys Val Arg Tyr Ala Lys Glu Glu  
 355 360 365  
 Asn Gln Lys Val Ile Gly Ala Tyr Ser Val Gln Ala Lys Trp Phe Ser  
 370 375 380  
 Glu Gly Tyr Val Pro Thr Ile Glu Glu Tyr Met Pro Ile Ala Leu Thr  
 385 390 395 400  
 Ser Cys Ala Tyr Thr Phe Val Ile Thr Asn Ser Phe Leu Gly Met Gly  
 405 410 415  
 Asp Phe Ala Thr Lys Glu Val Phe Glu Trp Ile Ser Asn Asn Pro Lys  
 420 425 430  
 Val Val Lys Ala Ala Ser Val Ile Cys Arg Leu Met Asp Asp Met Gln  
 435 440 445  
 Gly His Glu Phe Glu Gln Lys Arg Gly His Val Ala Ser Ala Ile Glu  
 450 455 460  
 Cys Tyr Thr Lys Gln His Gly Val Ser Lys Glu Glu Ala Ile Lys Met  
 465 470 475 480  
 Phe Glu Glu Glu Val Ala Asn Ala Trp Lys Asp Ile Asn Glu Glu Leu  
 485 490 495  
 Met Met Lys Pro Thr Val Val Ala Arg Pro Leu Leu Gly Thr Ile Leu  
 500 505 510  
 Asn Leu Ala Arg Ile Val Asp Phe Ile Tyr Lys Glu Asp Asp Gly Tyr  
 515 520 525  
 Thr His Ser Tyr Leu Ile Lys Asp Gln Ile Ala Ser Val Leu Gly Asp  
 530 535 540  
 His Val Pro Phe  
 545

<210> 38

<211> 550

5 <212> PRT

<213> *Perilla frutescens* var. *frutescens*

<220>

<223> Proteína valenceno sintasa

<300>

10 <308> GenBank AAX16077.1

<309> 01/03/2005

&lt;400&gt; 38

Met Ala Ser Glu Gln Ala Gln Asn His Arg Pro Val Ala Asp Phe Ser  
 1 5 10 15  
 Pro Ser Leu Trp Gly Asp Gln Phe Val Lys Tyr Asp Ser Cys Pro Gln  
 20 25 30  
 Val Gln Lys Lys Tyr Ser Asn Thr Val Asp Val Leu Lys Lys Glu Val  
 35 40 45  
 Lys Gly Met Ile Thr Ala Pro Gly Thr Lys Met Val Asp Thr Met Glu  
 50 55 60  
 Leu Ile Asp Thr Ile Glu Arg Leu Gly Val Ser Phe His Phe Gln Asp  
 65 70 75 80  
 Glu Ile Glu Gln Lys Leu Gln Gln Phe Phe Asp Leu Lys Thr Asp Tyr  
 85 90 95  
 Cys Asn Asp Gly Asp Asp Ala Tyr Asp Leu Tyr Thr Val Ala Leu His  
 100 105 110  
 Phe Arg Leu Phe Arg Gln His Gly Tyr Arg Ile Ser Cys Asp Ile Phe  
 115 120 125  
 Gly Arg Trp Ile Asp Gly Asn Gly Lys Phe Lys Glu Gly Leu Lys Ser  
 130 135 140  
 Asp Gly Lys Ser Leu Leu Ser Leu Tyr Glu Ala Ser Tyr Leu Arg Thr  
 145 150 155 160  
 Arg Gly Glu Thr Ile Leu Asp Glu Ala Leu Asp Phe Ala Ala Ala Ser  
 165 170 175  
 Leu Lys Ser Ile Ala Pro His Leu Gln Ser Pro Leu Gly Lys Gln Val  
 180 185 190  
 Val His Ala Leu Val Gln Pro Leu His Phe Gly Asn Pro Arg Ile Glu  
 195 200 205  
 Ala Arg Asn Phe Ile Ser Ile Tyr Glu Glu Tyr Glu Gly Met Asn Glu  
 210 215 220  
 Ala Leu Leu Arg Leu Ala Lys Leu Asp Tyr Asn Leu Leu Gln Met Leu  
 225 230 235 240  
 His Lys Glu Glu Leu His Gln Val Ser Arg Trp Trp Lys Asp Leu Asp  
 245 250 255  
 Leu Ile Thr Lys Leu Pro Tyr Ala Arg Asp Arg Val Val Glu Cys Phe  
 260 265 270  
 Phe Trp Ala Val Gly Val Tyr His Glu Pro Gln Tyr Ser Arg Ala Arg  
 275 280 285  
 Val Met Leu Thr Lys Thr Ile Val Met Thr Ser Ile Ile Asp Asp Thr  
 290 295 300  
 Tyr Asp Ala Tyr Gly Thr Ile Glu Glu Leu Asp Ile Phe Thr Glu Ala  
 305 310 315 320  
 Ile Glu Arg Trp Asn Val Glu Glu Thr Lys Arg Leu Pro Glu Tyr Met  
 325 330 335  
 Lys Pro Leu Tyr Lys Ala Leu Leu Glu Leu Tyr Lys Gln Phe Glu Gln  
 340 345 350  
 Glu Leu Glu Lys Glu Gly Arg Ser Tyr Val Ala Tyr Tyr Ala Ile Glu  
 355 360 365  
 Ser Leu Lys Glu Leu Val Arg Ser Tyr Arg Ile Glu Ala Lys Trp Phe  
 370 375 380  
 Ile Gln Gly Tyr Leu Pro Tyr Glu Glu Tyr Leu Lys Asn Ala Leu  
 385 390 395 400  
 Ile Thr Cys Thr Tyr Cys Tyr His Thr Thr Thr Ser Leu Leu Gly Val  
 405 410 415  
 Glu Ser Ala Ile Lys Glu Asn Phe Glu Trp Leu Ser Asn Lys Pro Lys  
 420 425 430  
 Met Leu Val Ala Gly Leu Leu Ile Cys Arg Leu Ile Asp Asp Ile Ala

ES 2 647 828 T3

		435					440					445			
Thr	Tyr	Ile	Leu	Glu	Arg	Gly	Arg	Gly	Gln	Val	Ala	Thr	Gly	Ile	Glu
	450					455					460				
Ser	Tyr	Met	Lys	Asp	Asn	Gly	Ala	Thr	Gln	Glu	Glu	Pro	Ile	Ala	Lys
465					470					475					480
Ile	Phe	Glu	Ile	Ala	Thr	Asp	Ala	Trp	Lys	Asp	Ile	Asn	Asp	Glu	Cys
				485					490					495	
Leu	Arg	Pro	Ser	Leu	Tyr	Asn	Ser	Arg	Asp	Val	Leu	Met	Arg	Ile	Phe
			500					505					510		
Asn	Leu	Glu	Arg	Ile	Ile	Asp	Val	Thr	Tyr	Lys	Gly	Asn	Gln	Asp	Arg
		515				520						525			
Tyr	Thr	Gln	Pro	Glu	Lys	Val	Leu	Lys	Pro	His	Ile	Ile	Val	Phe	Phe
	530					535						540			
Phe	Asp	Pro	Ile	Leu	Ile										
545					550										

<210> 39

<211> 1653

5 <212> ADN

<213> *Perilla frutescens* var. *frutescens*

<220>

<223> ADN de valenceno sintasa

<300>

10 <308> GenBank AY917195.1

<309> 01/03/2005

<400> 39

```

atggcttcag aacaagcaca aaatcatcgc cccgtcgccg atttctccc cagcttgtgg 60
ggcgatcagt tcgtcaaata tgattcttgt ccacaggttc agaagaagta ttcgaacacg 120
gttgatgttt tgaagaagga agttaagggg atgataactg ctcctggaac caaatgggtc 180
gacacgatgg agctgattga cacgatcgaa cgtctagcgc tgctgtttca ctttcaagat 240
gagattgaac aaaaattgca gcagtttttt gatctcaaaa cagattattg caacgatggg 300
gatgacgcct atgatttgta cactgttgct cttcatttcc gattattcag gcaacatggc 360
taccgtatat cttgtgacat ttttggtaga tggatcgatg ggaatgggaa attcaaggag 420
ggactgaaga gtgatgggaa gaggttgcta agtctgtacg aggcgtcgta tctgagaaca 480
cgtggcgaaa ccatactcga cgaggcgtc gactttgctg cggctagtct gaagtcgata 540
gcccacacc tccaatcacc ccttgggaaa caagttgtgc acgccctagt gcagcctttg 600
cactttggca atccaagaat cgaagcgcgt aatttcatct ccatctatga ggaatatgaa 660
ggcatgaatg aagctctctt gaggttagct aaattagact ataactctatt gcaaagtcta 720
cataaggagg aacttcatca agtctcgagg tgggtgaaag atttggatct gatcacgaaa 780
cttccatagc caagagatag agtgggtggag tgtttctttt gggcagtcgg agtataccat 840
gagccacaat atttctgtgc tcgtgtaatg cttactaaaa ccatcgttat gacctctata 900
atagatgata cttatgatgc ttatgggtacc attgaagaac ttgatatttt cactgaagca 960
atagagaggt ggaatgttga agagactaaa aggtccccg agtacetgaa accattgtat 1020
aaagctcttc tggaactcta caagcagttt gaacaagaac tagaaaagga aggaagatcc 1080
tacgtggcat actatgccat cgaatctctt aaggaattgg tgagaagcta tcgcattgag 1140
gcaaagtggg ttatacaagg atatttacca ccttacgaag aatacctaaa gaatgcactg 1200
atcacctgca cttactgtta ccacacaacg acgtcgttgt tgggggttga atcagccatc 1260
aagggaaaact tcgaatggct aagcaacaaa cctaaaatgc ttgtagctgg cctoctaata 1320
tgtcgactca ttgatgacat agctacttat atcctggaga ggggtagggg tcaggttgct 1380
actggcatcg agtcctacat gaaggataat ggcgcaacac aagaagaacc catagctaaa 1440
attttcgaaa tagctacaga tgcatggaag gatataaatg acgaatgctt gagacctagc 1500
ctttacaact cgagggatgt tttgatgcga atatttaacc ttgaacgtat aatagacgct 1560
acttacaag gcaaccaaga tcgatacact caaccagaaa aggttttgaa gcctcacatc 1620
attgtcttct tcttcgatcc cattctcatt taa 1653

```

15

<210> 40

<211> 1647

<212> ADN

<213> *Citrus sinensis*

20 <220>

<223> ADN de valenceno sintasa cítrica

<300>

<308> GenBank AF441124.1

<309> 12/01/2004

<400> 40

```

atgctgctcg gagaaacatt tcgctcctact gcagatttcc atcctagttt atggagaaac 60
catttcctca aagggtcctc tgatttcaag acagttgatc atactgcaac tcaagaacga 120
cacgaggcac tgaaagaaga ggtaaggaga atgataacag atgctgaaga taagcctgtt 180
cagaagttac gcttgattga tgaagtacaa cgcctggggg tggcttatca ctttgagaaa 240
gaaataggag atgcaatata aaaattatgt ccaatctata ttgacagtaa tagagctgat 300
ctccacaccg tttcccttca ttttcggttg cttaggcagc aaggaatcaa gatttcatgt 360
gatgtgtttg agaagttcaa agatgatgag ggtagattca agtcatcgtt gataaacgat 420
gttcaaggga tgttaagttt gtacgaggca gcatacatgg cagttcggcg agaacatata 480
ttagatgaag ccattgcttt cactaccact cacctgaagt cattggtagc tcaggatcat 540
gtaaccctta agcttgccgga acagataaat catgctttat accgtcctct tcgtaaaacc 600
ctaccaagat tagaggcgag gtattttatg tccatgatca attcaacaag tgatcattta 660
tgcaataaaa ctctgctgaa ttttgcaaag tttagattta acatattgct agagctgcac 720
aaggaggaac tcaatgaatt aacaaagtgg tggaaagatt tagacttcac tacaaaacta 780
ccttatgcaa gagacagatt agtggagtta tatttttggg atttagggac atacttcgag 840
cctcaatatg catttgggag aaagataatg acccaattaa attacatatt atccatcata 900
gatgatactt atgatgcgta tggtagcact gaagaactca gcctctttac tgaagcagtt 960
caaagatgga atattgaggc cgtagatatg cttccagaat acatgaaatt gatttacagg 1020
acactcttag atgcttttaa tgaattgag gaagatatgg ccaagcaagg aagatcacac 1080
tgcgtacggt atgcaaaaaga ggagaatcaa aaagtaattg gagcatactc tgttcaagcc 1140
aaatggttca gtgaaggtta cgttccaaca attgaggagt atatgcctat tgcactaaca 1200
agttgtgctt acacattcgt cataacaaat tccttccttg gcatgggtga ttttgcaact 1260
aaagaggttt ttgaatggat ctccaataac cctaaggttg taaaagcagc atcagttatc 1320
tgcagactca tggatgacat gcaaggtcat gagtttgagc agaagagagg acatggttgcg 1380
tcagctattg aatgttacac gaagcagcat ggtgtctcta aggaagaggc aattaaaatg 1440
tttgaagaag aagttgcaaa tgcattgaaa gatattaacg aggagttgat gatgaagcca 1500
accgtcgttg cccgaccact gctcgggacg attcctaatac ttgctcgtgc aattgatttt 1560
atttacaaag aggacgacgg ctatagcgtat tcttacctaa ttaaagatca aattgcttct 1620
gtgctaggag accacgttcc attttga 1647

```

5 <210> 41

<211> 1750

<212> ADN

<213> Hyoscyamus muticus

<220>

10 <223> ADN de premnaspirodieno sintasa

<300>

<308> GenBank U20188.1

<309> 31/01/1996

<400> 41

15

```

gttgacaatc aggttgccgga aaagtatgct caagagattg aaactttgaa ggaacaaaca 60
agtactatgt tgtctgctgc ttgtggaaca acattgactg agaaattgaa tttgatagat 120
attattgagc gccttggaat agcttatcat ttcgagaaac aaatagaaga tatgttggat 180
cacatttaca gagctgatcc ttattttgag gctcatgaat acaatgattt aaacacttca 240
tccgttcaat ttcgactact cagacaacat ggttacaacg tctctccaaa tatatttagc 300
agattccaag atgcaaatgg caaattcaag gagtctctta gaagcgacat caggggccta 360
ctgaacttat acgaagcttc acatgtaagg actcataaag aggatatttt ggaagaagca 420
cttgtttttt ctggttggtca tcttgaatct gcagctccac atttgaagtc acctctgagt 480
aagcaagtga cacatgccct cgaacaatct ctccataaga gcattccaag agtcgagata 540
cggacttca tctccatcta cgaagaggag gaatttaaga atgatttgtt gcttcgattt 600
gctaaattgg attacaactt acttcagatg ttgcacaagc atgaactcag tgaagtatca 660
aggtggtgga aagatttggga tttcgtgaca acacttccat atgctaggga cagagcagtt 720

```

ES 2 647 828 T3

```

gagtgtact tttggacgat gggggtgat gctgaacctc aatactocca ggctcgtgtc 780
atgcttgcta agactatagc aatgatttcc atagtagatg acacattcga tgcttatggc 840
atcgtaaaag aacttgaagt ctacacagat gccatacaga ggtgggatat tagtcaaatt 900
gatcgactcc cggaatatat gaaaatcagt tataaggctc ttttggatct ctatgacgat 960
tatgaaaagg agttgtcaaa ggatggtaga tccogatgtt tccactatgc aaaagaaaga 1020
atgaaggaga ttgtgggaaa ctattttatt gaaggaaaat ggtttattga gggatatatg 1080
ccatctgttt ccgagtacct tagcaatgca cttagtacta gcacatatta cttgctaact 1140
acgacatcct acttgggaat gaaatcagca accaaggaac attttgaatg gtgggctacg 1200
aaccctaaaa ttctggaagc taatgctaca ttatgccgag ttgttgatga catagccacg 1260
tatgaggttg agaagggtag ggggtcaaatt gcaacaggaa ttgagtgtta tatgagggat 1320
tacgggtgat ccacagaagt agcaatggag aaattccaag aaatggctga catagcatgg 1380
aaggatgtaa atgaagaaat tcttcgacca acacctgtct cttcagaaat tcttactcgt 1440
atctcaacc ttgctcgaat tatagatgct acttacaagc ataatcaaga tggatgact 1500
catcctgaaa aagtactaaa acctcacatc atcgccttgg tgggtgattc tattgatatt 1560
tgaatcacca attgttgtgt acacctggga gcacttggtt cccaccccct caaataagtt 1620
tttgacagac acttgatgga tggatctctt gttgcttagt atacatggtt taatcgtgca 1680
gtgaagttac gtccttaatt cttttgatg attatttaca tttgaaatat aataattctg 1740
ctttttaact                                     1750

```

- <210> 42
- <211> 1839
- 5 <212> ADN
- <213> Citrus hystrix
- <220>
- <223> ADN de germacreno D sintasa
- <300>
- 10 <308> GenBank HQ652871.1
- <309> 28/06/2011
- <400> 42

```

atgtctgttg aaggttctgc aaattttcat ccaagcattt ggggtgatca tttccttcaa 60
tatactcgtg acttccagga aactggtgat cgaagtgtaa agcatctaga gctgaagaaa 120
gaaattagaa gaatgctaaa agctgtaaac aagacttcac atacactcga attgatagat 180
gcaattcagc ggtagtaggt gtcttaccat tttgaaagtg agattgatga aatcttggga 240
aagatgcaca agacttatcg agactgtgat ctttgtgata atgaaaatga tgagctttat 300
tatatctctc ttcagtttcc attgttcaga caaaatggct atagaatttc cgctgatgtt 360
ttcaatacgt tcaagggcag cgatgggaaa ttttaaggcat ctcttgcaaa agatgttcga 420
ggaatgttaa gcttgatga agctacacat ctgagggttc atggagaaaa tatacttgat 480
gagggcgttg ctttcaccac tagtcacctt gagtcagtag caaaacaagt ctgttctcca 540
ctagttgaac aagtcaagca cgccttagtt cagcctatcc acaagggtt agaaaggctt 600
gaggcaagac actacattcc tatctatcaa ggagaatctt cccacaatga agctctgta 660
acctttgcaa agttaggttt taatggattg caaaagctt accagaagga actcgggtgat 720
atltcaaggt ggtggaagaa attagacttt ggcataaagc tacccttctg aagagataga 780
attgcagagg tctacttttg gccagtagga gttcatttcg agccccaata ttcgtttact 840
agaaaactat ttacgaaagt gatttatatg acatctatca ttgatgacat ctatgatgtg 900
tatggcaaaa ttgaagaact tgatcttttt acttcagcta ttgaaagggtg ggatatcaat 960
gccatagatc aacttcctga gtatatgaaa ctgtgtatta gggcgcttat caatgtttac 1020
agtgaagtag agaaagattt ggtctcccag ggggaagtat accgactcca ttatgcgaaa 1080
gaagcaatga agaatcaagt taagcattac ttctttgaag ctaaagtgtg tcatcagaat 1140
tatgttccga cgggtgatga gtacatgacg gttgcattaa tttagctctgg ccacccaaat 1200
ttgtcaacca tatcttttgt tggcctggga gacattgtaa ctaaagaatc ttttgaatgg 1260
ttattcagca atcctagatc gattagggct tcttgtgcag ttggcagact aatgaatgac 1320
atggtgtcac acaagtttga acaaaagcaga gggcacgttg cctcaagcgt tgagtgttac 1380
atcaaccaat atggagcaac agaagaggaa gcatacagtg agttccggaa acaagtttca 1440
aatgcattgga aggatataaa tgaggaaatgc ctgctgctcc ctgttctgct agtgccactt 1500
cttatgcgaa ttctcaatct tacacgagct gcagatgtca tttacaagta taaagatggc 1560
tacacttact ccgaagagct gaaagatttt attgtttctc tgcttattaa tcctgtgccc 1620
atatgagcat gatattaagc ttatgtttca acaccctgct ggaaaataat tagtccattg 1680
ccgtttcagc ttggatgctt tttaaaaata aattgaaggg tgggtgatccc cgttgtaatc 1740
agagtgcact caaatagaca agtggttaata aagttttata gctctattta attaaattat 1800
ccagcagttt gcacgattaa aaaaaaaaa aaaaaaaaa 1839

```

- 15 <210> 43
- <211> 1722
- <212> ADN

ES 2 647 828 T3

<213> Chamaecyparis nootkatensis

<220>

<223> ADN de valenceno sintasa

<300>

5 <310> WO 2011/074954

<311> 15/12/2010

<400> 43

```

atgcccgatga aggacgcctc tcgtcggact ggaaatcatc atcctaactt gtggactgat 60
gatttcatac agtccctcaa ttctccatat tcggattcct cataccataa acatagggaa 120
atactaattg atgagattcg tgatagtttt tctaattggag aaggcgatga gttcgggtga 180
cttgaaaata tttggtttgg tgatgttgta caacgtttgg gaatagatcg acattttcaa 240
gaggaaatca aaactgcact tgattatata tacaagttct ggaatcatga tagtattttt 300
ggcgatctca acatgggtggc tctaggattt cggatactac gactgaatag atatgtcgcct 360
tcttcagatg tttttaaaaa gttcaaagggt gaagaaggac aattctctgg ttttgaatct 420
acggatcaag atgcaaaaatt agaaatgatg tttaaatttat ataaagcttc agaattagat 480
tttctctgat aagatatcct aaaagaagca agagcgtttg cttctatgta cctgaaacat 540
gttatcaaag aatatggtga catacaagaa tcaaaaaatc cacttctaag ggagatagag 600
tacactttta aatatccttg gagatgtagg cttccaagggt tggaggcttg gaactttatt 660
catataatga gacaacaaga ttgcaatata tcacttgcca ataaccttta taaaattcca 720
aaaaatata tgaaaaagat attggaacta gcaatactgg acttcaatat tttgcagtca 780
caacatcaac atgaaatgaa ataatatcc acatgggtgga aaaatcaag tgcaattcaa 840
ttggatttct ttcggcatcg tcacatagaa agttattttt ggtgggctag tccattattt 900
gaacctgagt tcagtagatg tagaattaat tgtaccaaat tatctacaaa aatgttcctc 960
cttgacgata tttatgacac atatgggact gttgaggaat tgaaaccatt cacaacaaca 1020
ttaacaagat gggatgtttc cacagttgat aatcatccag actacatgaa aattgctttc 1080
aatttttcat atgagatata taaggaaatt gcaagtgaag ccgaaagaaa gcatgggtccc 1140
tttgtttaca aataccttca atcttgctgg aagagttata tcgaggctta tatgcaagaa 1200
gcagaatgga tagcttctaa tcatatacca ggttttgatg aatacttgat gaatggagta 1260
aaaagtacg gcatgcgaat tctaatagata catgcactaa tactaatgga tactccttta 1320
tctgatgaaa ttttgagca acttgatata ccatcatcca agtcgcaagc tcttctatca 1380
ttaattactc gactagtgga tgatgtcaaa gactttgagg atgaacaagc tcatggggag 1440
atggcatcaa gtatagagt ctacatgaaa gacaaccatg gttctacaag ggaagatgct 1500
ttgaattatc tcaaaattcg tatagagagt tgtgtgcaag agttaaataa ggagcttctc 1560
gagccttcaa atatgcatgg atctttttaga aacctatata tcaatgttgg catgcgagta 1620
atatttttta tgctcaatga tggatgatctc tttacacact ccaatagaaa agagatacaa 1680
gatgcaataa caaaattttt tgtggaacca atcattccat ag 1722

```

10

<210> 44

<211> 1770

<212> ADN

<213> Chamaecyparis nootkatensis

15 <220>

<223> ADN de valenceno sintasa

<300>

<310> WO 2011/074954

<311> 15/12/2010

20 <400> 44

```

atggctgaaa tgtttaatgg aaattccagc aatgatggaa gttcttgcac gcccgatgag 60
gacgcccttc gtcggactgg aaatcatcat cctaacttgt ggactgatga tttcatacag 120
tccctcaatt ctccatattc ggattcttca taccataaac atagggaaat actaattgat 180
gagattcgtg atatgttttc taatggagaa ggcgatgagt tcgggtgact tgaaaatatt 240
tggtttggtt atgttgatca acgtttggga aatagatcgac attttcaaga ggaaatcaaa 300
actgcaactg attatatcta caagttctgg aatcatgata gtatttttgg cgatctcaac 360

```

ES 2 647 828 T3

```

atggtggctc taggatttgc gatactacga ctgaatagat atgtcgcttc ttcagatggt 420
tttaaaaagt tcaaagggtga agaaggacaa ttctctgggt ttgaatctag cgatcaagat 480
gcaaaattag aaatgatggt aaatttatat aaagcttcag aattagattt tcctgatgaa 540
gatatcttaa aagaagcaag agcgtttgct tctatgtacc tgaaacatgt tatcaaagaa 600
tatggtgaca tacaagaatc aaaaaatcca cttctaattg agatagagta cactttttaa 660
tatccttgga gatgtaggct tccaagggtg gaggcttggg actttattca tataatgaga 720
caacaagatt gcaatatatc acttgccaat aacctttata aaattccaaa aatatatag 780
aaaagatat tggaactagc aatactggac ttcaatattt tgcagtcaca acatcaacat 840
gaaatgaaat taatatccac atggtggaaa aattcaagt caattcaatt ggatttcttt 900
cggcatcgtc acatagaaag ttatttttgg tgggctagtc cattatttga acctgagttc 960
agtacatgta gaattaattg taccaaatta tctacaaaaa tgctcctcct tgacgatatt 1020
tatgacacat atgggactgt tgaggaattg aaaccattca caacaacatt aacaagatgg 1080
gatgtttcca cagttgataa tcatccagac tacatgaaaa ttgctttcaa tttttcatat 1140
gagatatata aggaaattgc aagtgaagcc gaaagaaagc atggtccctt tgtttacaaa 1200
taccttcaat cttgctggaa gagttatcct gaggcttata tgcaagaagc agaatggata 1260
gcttctaadc atataccagg ttttgatgaa tacttgatga atggagtaaa aagtagcggc 1320
atgccaattc taatgatata tgcaactaata ctaatggata ctcctttatc tgatgaaatt 1380
ttggagcaac ttgatatccc atcatccaag tcgcaagctc ttctatcatt aattactcga 1440
ctagtggatg atgtcaaaga ctttgaggat gaacaagctc atggggagat ggcatcaagt 1500
atagagtgct acatgaaaga caaccatggt tctacaaggg aagatgcttt gaattatctc 1560
aaaattcgtg tagagagttg tgtgcaagag ttaaataagg agcttctcga gccttcaaat 1620
atgcatggat cttttagaaa cctatatctc aatggtggca tgcgagtaat attttttatg 1680
ctcaatgatg gtgatctctt tacacactcc aatagaaaag agatacaaga tgcaataaca 1740
aaattttttg tggaaccaat cattccatag 1770

```

- <210> 45
- <211> 541
- 5 <212> PRT
- <213> Citrus hystrix
- <220>
- <223> Proteína germacreno D sintasa
- <300>
- 10 <308> GenBank ADX01384.1
- <309> 28/06/2011
- <400> 45

```

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

```

ES 2 647 828 T3

		195						200						205			
Tyr	Gln	Gly	Glu	Ser	Ser	His	Asn	Glu	Ala	Leu	Leu	Thr	Phe	Ala	Lys		
	210						215						220				
Leu	Gly	Phe	Asn	Gly	Leu	Gln	Lys	Leu	His	Gln	Lys	Glu	Leu	Gly	Asp		
225					230					235				240			
Ile	Ser	Arg	Trp	Trp	Lys	Glu	Leu	Asp	Phe	Ala	His	Lys	Leu	Pro	Phe		
				245					250					255			
Val	Arg	Asp	Arg	Ile	Ala	Glu	Val	Tyr	Phe	Trp	Ala	Val	Gly	Val	His		
			260					265					270				
Phe	Glu	Pro	Gln	Tyr	Ser	Phe	Thr	Arg	Lys	Leu	Phe	Thr	Lys	Val	Ile		
		275					280						285				
Tyr	Met	Thr	Ser	Ile	Ile	Asp	Asp	Ile	Tyr	Asp	Val	Tyr	Gly	Lys	Ile		
	290					295					300						
Glu	Glu	Leu	Asp	Leu	Phe	Thr	Ser	Ala	Ile	Glu	Arg	Trp	Asp	Ile	Asn		
305					310					315				320			
Ala	Ile	Asp	Gln	Leu	Pro	Glu	Tyr	Met	Lys	Leu	Cys	Ile	Arg	Ala	Leu		
				325					330					335			
Ile	Asn	Val	Tyr	Ser	Glu	Val	Glu	Lys	Asp	Leu	Val	Ser	Gln	Gly	Lys		
			340					345					350				
Leu	Tyr	Arg	Leu	His	Tyr	Ala	Lys	Glu	Ala	Met	Lys	Asn	Gln	Val	Lys		
		355					360						365				
His	Tyr	Phe	Phe	Glu	Ala	Lys	Trp	Cys	His	Gln	Asn	Tyr	Val	Pro	Thr		
	370					375					380						
Val	Asp	Glu	Tyr	Met	Thr	Val	Ala	Leu	Ile	Ser	Ser	Gly	His	Pro	Asn		
385					390					395				400			
Leu	Ser	Thr	Ile	Ser	Phe	Val	Gly	Leu	Gly	Asp	Ile	Val	Thr	Lys	Glu		
				405					410					415			
Ser	Phe	Glu	Trp	Leu	Phe	Ser	Asn	Pro	Arg	Ser	Ile	Arg	Ala	Ser	Cys		
				420				425					430				
Ala	Val	Gly	Arg	Leu	Met	Asn	Asp	Met	Val	Ser	His	Lys	Phe	Glu	Gln		
		435					440						445				
Ser	Arg	Gly	His	Val	Ala	Ser	Ser	Val	Glu	Cys	Tyr	Ile	Asn	Gln	Tyr		
	450					455						460					
Gly	Ala	Thr	Glu	Glu	Glu	Ala	Tyr	Ser	Glu	Phe	Arg	Lys	Gln	Val	Ser		
465					470					475				480			
Asn	Ala	Trp	Lys	Asp	Ile	Asn	Glu	Glu	Cys	Leu	Arg	Pro	Thr	Val	Val		
				485					490					495			
Pro	Val	Pro	Leu	Leu	Met	Arg	Ile	Leu	Asn	Leu	Thr	Arg	Ala	Ala	Asp		
				500				505					510				
Val	Ile	Tyr	Lys	Tyr	Lys	Asp	Gly	Tyr	Thr	Tyr	Ser	Glu	Glu	Leu	Lys		
		515					520						525				
Asp	Phe	Ile	Val	Ser	Leu	Leu	Ile	Asn	Pro	Val	Pro	Ile					
	530					535						540					

## REIVINDICACIONES

1. Un polipéptido de valenceno sintasa de *Eryngium* aislado que comprende una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con el polipéptido de valenceno sintasa cuya secuencia se expone en la SEQ ID NO: 1 o un fragmento catalíticamente activo de la misma, donde:  
5 el polipéptido de sintasa cataliza la formación de valenceno a partir de un precursor terpénico de pirofosfato acíclico; y
- las células que expresan el polipéptido de valenceno sintasa de SEQ ID NO:1 catalizan mayores cantidades de valenceno en comparación con células que expresan la valenceno sintasa cítrica (*Citrus sinensis*) (CVS) cuya secuencia se expone en la SEQ ID NO:14.  
10
2. El polipéptido de valenceno sintasa aislado de la reivindicación 1, que comprende o consiste en la secuencia de aminoácidos expuesta en la SEQ ID NO: 1 o un fragmento catalíticamente activo de la misma.  
15
3. El polipéptido de valenceno sintasa aislado de la reivindicación 1 o la reivindicación 2, donde el precursor terpénico de pirofosfato acíclico es difosfato de farnesilo.
4. Un polipéptido de valenceno sintasa que es un fragmento catalíticamente activo del péptido de valenceno sintasa aislado de cualquiera de la reivindicación 1 o reivindicación 2, donde el fragmento catalíticamente activo cataliza la formación de valenceno a partir de un precursor terpénico de pirofosfato acíclico.  
20
5. Una molécula de ácido nucleico aislada que codifica un polipéptido de valenceno sintasa de *Eryngium glaciale* o fragmento catalíticamente activo de cualquiera de las reivindicaciones 1-4.  
25
6. Una molécula de ácido nucleico que codifica un polipéptido de valenceno sintasa de *Eryngium glaciale* o fragmento catalíticamente activo de cualquiera de las reivindicaciones 1-4, donde la molécula de ácido nucleico es ADNc.
7. La molécula de ácido nucleico aislada de la reivindicación 5 o la reivindicación 6, donde:  
30 el polipéptido de valenceno sintasa codificado comprende una secuencia de aminoácidos expuesta en la SEQ ID NO:1, o el polipéptido de valenceno sintasa codificado comprende una secuencia de aminoácidos que tiene al menos un 95 % de identidad de secuencia con el polipéptido de valenceno sintasa cuya secuencia se expone en la SEQ ID NO: 1 o con un fragmento catalíticamente activo de la misma.  
35
8. La molécula de ácido nucleico aislada de la reivindicación 5, donde el polipéptido de valenceno sintasa codificado exhibe al menos un 95 % de identidad de secuencia aminoacídica con la secuencia de aminoácidos expuesta en la SEQ ID NO: 1 o un fragmento catalíticamente activo de la misma.  
40
9. La molécula de ácido nucleico aislada de cualquiera de las reivindicaciones 5-7, seleccionada de entre moléculas de ácido nucleico que comprenden:  
45 (a) la secuencia de ácidos nucleicos expuesta en la SEQ ID NO: 2, (b) una secuencia de ácidos nucleicos que tiene al menos un 85 % identidad de secuencia con una secuencia de ácidos nucleicos expuesta en la SEQ ID NO:2 y que codifica un polipéptido de valenceno sintasa que tiene al menos un 90 % de identidad de secuencia con el polipéptido de SEQ ID NO:1; y (c) degeneraciones de (a) y (b).  
50
10. La molécula de ácido nucleico aislada de cualquiera de las reivindicaciones 5-9, que comprende o consiste en la secuencia de nucleótidos expuesta en la SEQ ID NO: 2.
11. La molécula de ácido nucleico aislada de cualquiera de las reivindicaciones 5-10, donde el polipéptido de valenceno sintasa codificado comprende la secuencia de aminoácidos expuesta en la SEQ ID NO: 1.  
55
12. Un vector que comprende la molécula de ácido nucleico de cualquiera de las reivindicaciones 5-11.
13. El vector de la reivindicación 12, donde el vector es un vector procariótico, un vector vírico o un vector eucariótico, preferiblemente donde el vector es un vector de levadura.  
60

14. Una célula que comprende la molécula de ácido nucleico de cualquiera de las reivindicaciones 5 a 11, donde la molécula de ácido nucleico es heteróloga de la célula y, si la célula es una célula humana, es una célula aislada o está en un cultivo celular; o una célula hospedadora que:
- 5 a) codifica un polipéptido de valenceno sintasa o fragmento catalíticamente activo del mismo de cualquiera de las reivindicaciones 1 a 4, donde la célula hospedadora produce un precursor terpénico de pirofosfato acíclico; el polipéptido de valenceno sintasa es heterólogo del hospedador; el polipéptido de valenceno sintasa cataliza la producción de valenceno a partir del precursor y, si la célula hospedadora es una célula humana, es una célula aislada o una célula cultivada;
- 10 b) comprende el vector de la reivindicación 12 o la reivindicación 13, donde si la célula es una célula humana, es una célula aislada o está en un cultivo celular; o
- c) comprende un ácido nucleico que codifica el polipéptido de valenceno de la reivindicación 1, donde el polipéptido de valenceno sintasa codificado tiene al menos un 95 % de identidad de secuencia con el polipéptido de valenceno sintasa cuya secuencia se expone en la SEQ ID NO: 1 o con un fragmento catalíticamente activo del mismo;
- 15 el polipéptido de valenceno sintasa o fragmento catalíticamente activo del mismo cataliza la formación de valenceno a partir de un precursor terpénico de pirofosfato acíclico; el ácido nucleico es heterólogo de la célula hospedadora y si la célula es una célula humana, es una célula aislada o está en un cultivo celular.
- 20 15. La célula de la reivindicación 14, que es una célula procariótica o una célula eucariótica.
16. La célula de la reivindicación 14 o la reivindicación 15, que se selecciona de entre una bacteria, preferiblemente *Escherichia coli*, células de levadura, insecto, planta y mamífero.
- 25 17. La célula de cualquiera de las reivindicaciones 14 a 16, que es una célula de levadura y que es una célula del género *Saccharomyces*, preferiblemente *Saccharomyces cerevisiae* o un célula del género *Pichia*.
18. La célula de cualquiera de las reivindicaciones 14 a 17, que produce difosfato de farnesilo nativamente o está modificada para producir más difosfato de farnesilo en comparación con una célula no modificada.
- 30 19. La célula hospedadora de la reivindicación 14 que es una célula hospedadora de levadura, preferiblemente una célula de *Saccharomyces cerevisiae*.
20. La célula hospedadora de la reivindicación 14 o la reivindicación 19, donde el precursor es difosfato de
- 35 farnesilo.
21. Una planta transgénica que comprende el vector de cualquiera de las reivindicaciones 12-13 o una molécula de ácido nucleico de cualquiera de las reivindicaciones 5-11, donde la molécula de ácido nucleico es heteróloga de la planta.
- 40 22. La planta transgénica de la reivindicación 21 que es una planta de tabaco.
23. Un procedimiento para producir un polipéptido de valenceno sintasa, que comprende:
- 45 a) cultivar una célula de cualquiera de las reivindicaciones 14-20 en condiciones adecuadas para la expresión del polipéptido de valenceno sintasa, produciendo el polipéptido de valenceno sintasa o un fragmento catalíticamente activo del mismo;
- b) introducir una molécula de ácido nucleico de cualquiera de las reivindicaciones 5-11 o un vector de la reivindicación 12 o la reivindicación 13 en una célula;
- 50 cultivar la célula *in vitro* o *in vivo* en condiciones adecuadas para la expresión del polipéptido de valenceno sintasa, con lo que se expresa el polipéptido de valenceno sintasa codificado o un fragmento catalíticamente activo del mismo.
24. El procedimiento de la reivindicación 23, que comprende además aislar el polipéptido de valenceno
- 55 sintasa y/o fragmento catalíticamente activo del mismo.
25. Un procedimiento para producir valenceno que comprende:
- a) poner en contacto un precursor terpénico de pirofosfato acíclico con el polipéptido de valenceno sintasa o
- 60 fragmento catalíticamente activo de cualquiera de las reivindicaciones 1-4, o el polipéptido o fragmento codificado por la molécula de ácido nucleico de cualquiera de las reivindicaciones 5-11, en condiciones adecuadas para la

- formación de valenceno a partir del precursor terpénico de pirofosfato acíclico, donde la etapa de puesta en contacto del precursor terpénico de pirofosfato acíclico con el polipéptido de valenceno sintasa se efectúa *in vitro* o *in vivo*, con la condición de que, si se pone en contacto *in vivo* en una célula, el polipéptido de valenceno sintasa codificado es heterólogo de la célula hospedadora, y si la célula es una célula humana, es una célula aislada o está en un cultivo celular; o
- 5 b) cultivar una célula que contiene la molécula de ácido nucleico de cualquiera de las reivindicaciones 5-11 o un vector de cualquiera de las reivindicaciones 12 o 13 para producir valenceno, donde:
- la molécula de ácido nucleico es heteróloga de la célula;
- 10 la célula produce un precursor terpénico de pirofosfato acíclico; el polipéptido de valenceno sintasa codificado por la molécula de ácido nucleico o vector se expresa y el polipéptido de valenceno sintasa cataliza la formación de valenceno a partir de el precursor terpénico de pirofosfato acíclico para producir valenceno.
- 15 26. El procedimiento de la reivindicación 25, donde el precursor terpénico de pirofosfato acíclico es difosfato de farnesilo.
27. El procedimiento de la reivindicación 25 o 26, donde la célula es una célula procariótica o una célula eucariótica que se selecciona de entre una bacteria, célula de levadura, insecto, planta o mamífero.
- 20 28. El procedimiento de cualquiera de las reivindicaciones 25-27, donde la célula es una célula de levadura que es una célula del género *Saccharomyces*, preferiblemente *Saccharomyces cerevisiae*, o una célula del género *Pichia*.
- 25 29. El procedimiento de cualquiera de las reivindicaciones 25-28, donde la célula produce difosfato de farnesilo o se modifica para producir más difosfato de farnesilo en comparación con una célula no modificada.
30. El procedimiento de cualquiera de las reivindicaciones 25-29, que comprende además aislar el valenceno.
- 30 31. Un procedimiento para producir nootkatona, que comprende:
- (i)
- 35 (a) poner en contacto difosfato de farnesilo con el polipéptido de valenceno sintasa de cualquiera de las reivindicaciones 1-4 y/o el polipéptido codificado por la molécula de ácido nucleico de cualquiera de las reivindicaciones 5-11, con lo que se produce valenceno;
- (b) aislar el valenceno producido en la etapa (a);
- (c) oxidar el valenceno para producir nootkatona; y
- 40 (d) aislar la nootkatona; o
- (ii)
- (a) cultivar una célula que comprende una molécula de ácido nucleico de cualquiera de las reivindicaciones 5-11, o un vector de las reivindicaciones 12-13, en condiciones resultantes en la producción de valenceno;
- 45 (b) aislar el valenceno producido en la etapa (a);
- (c) oxidar el valenceno para producir nootkatona; y
- (d) aislar la nootkatona, donde la célula produce un precursor terpénico de pirofosfato acíclico;
- 50 el polipéptido de valenceno sintasa codificado por la molécula de ácido nucleico o vector se expresa y el polipéptido de valenceno sintasa cataliza la formación de valenceno a partir del precursor terpénico de pirofosfato acíclico.
32. El procedimiento de la reivindicación 30 o 31, donde se aísla el valenceno mediante extracción con un disolvente orgánico y/o cromatografía en columna.
- 55 33. El procedimiento de cualquiera de las reivindicaciones 24-32, donde la oxidación se practica biosintética o químicamente.
34. El procedimiento de cualquiera de las reivindicaciones 24-30 o 32, que comprende además oxidar el valenceno para producir nootkatona y aislar la nootkatona.
- 60

35. El procedimiento de cualquiera de las reivindicaciones 31-34, donde se aísla la nootkatona mediante extracción con un disolvente orgánico y/o cromatografía en columna.

36. El procedimiento de cualquiera de las reivindicaciones 31 (ii), 32 o 33, que comprende además aislar 5 aristoloqueno, que es el compuesto del pico nº 2 en la Figura 3A.

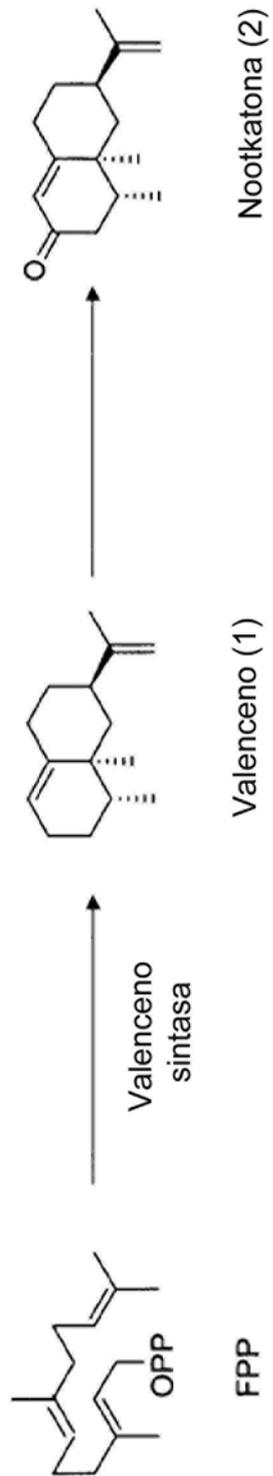


FIGURA 1

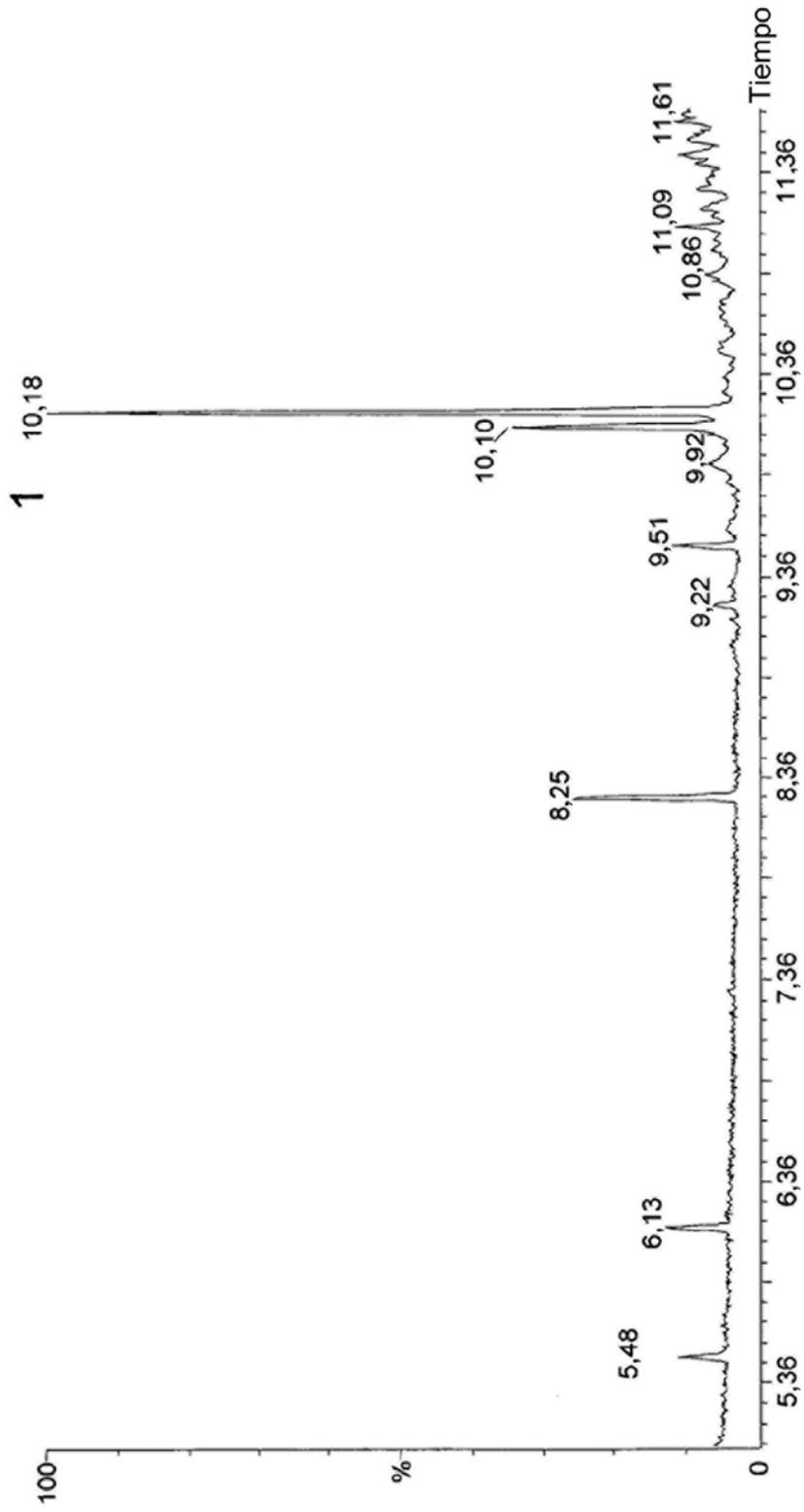


FIGURA 2A

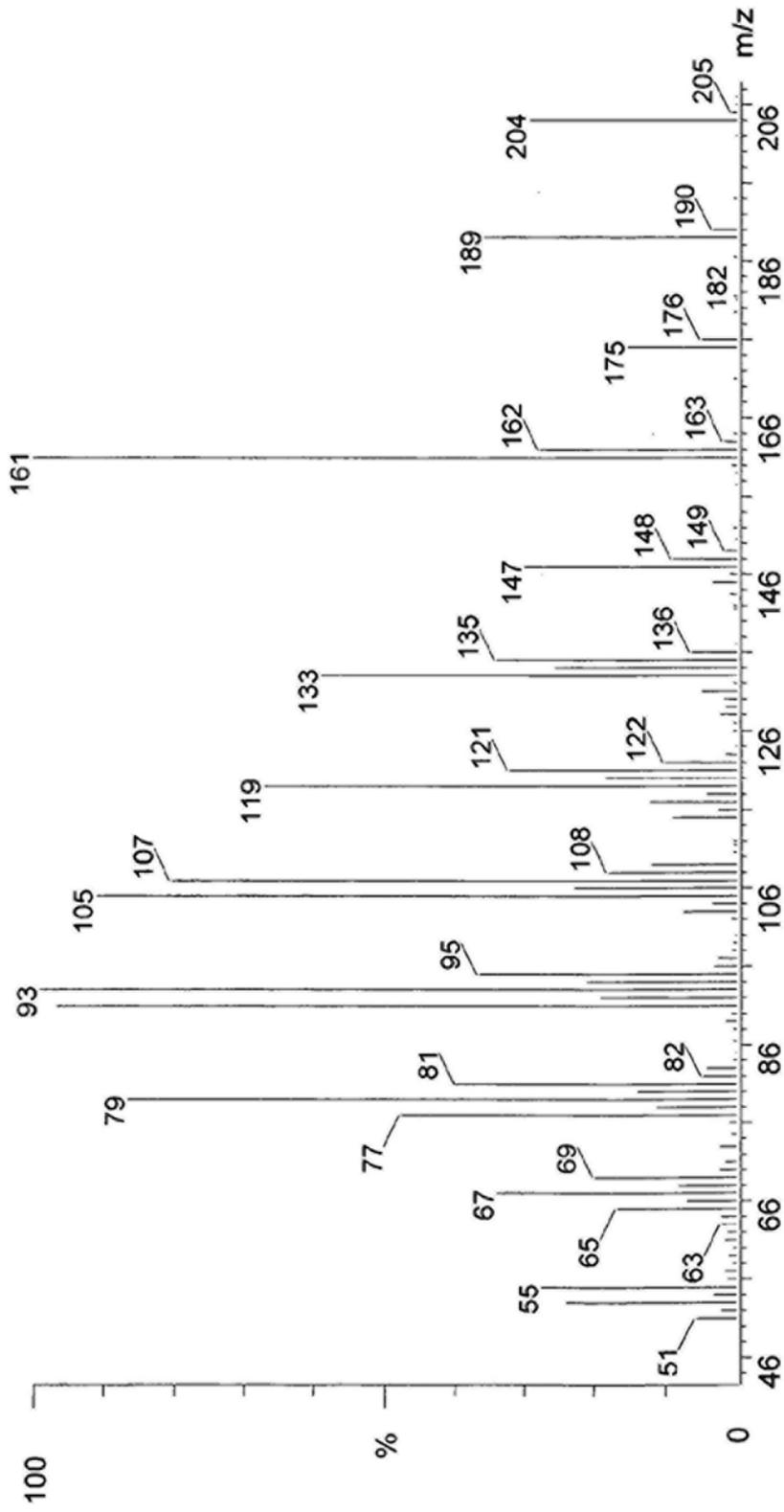


FIGURA 2B

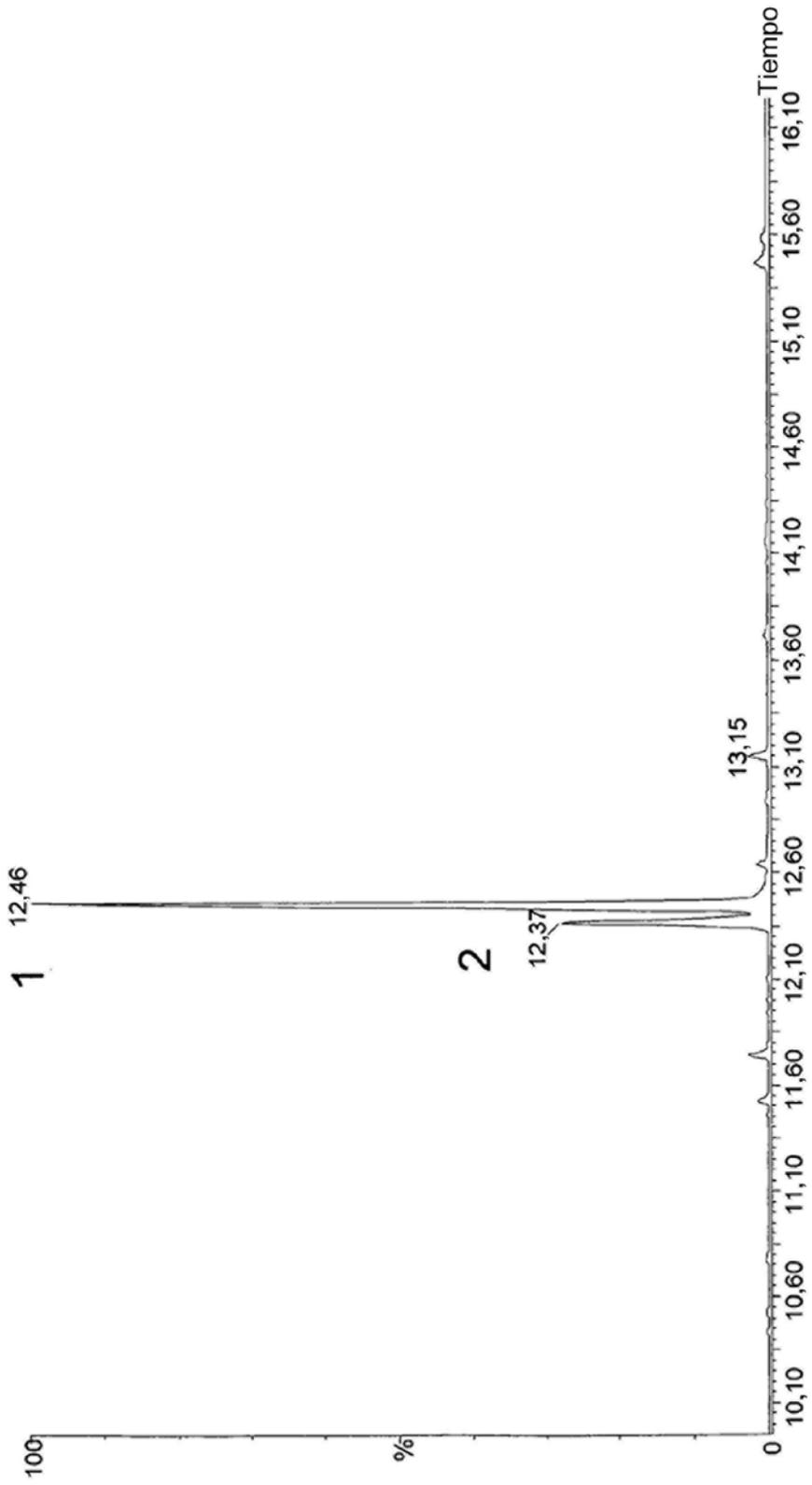


FIGURA 3A

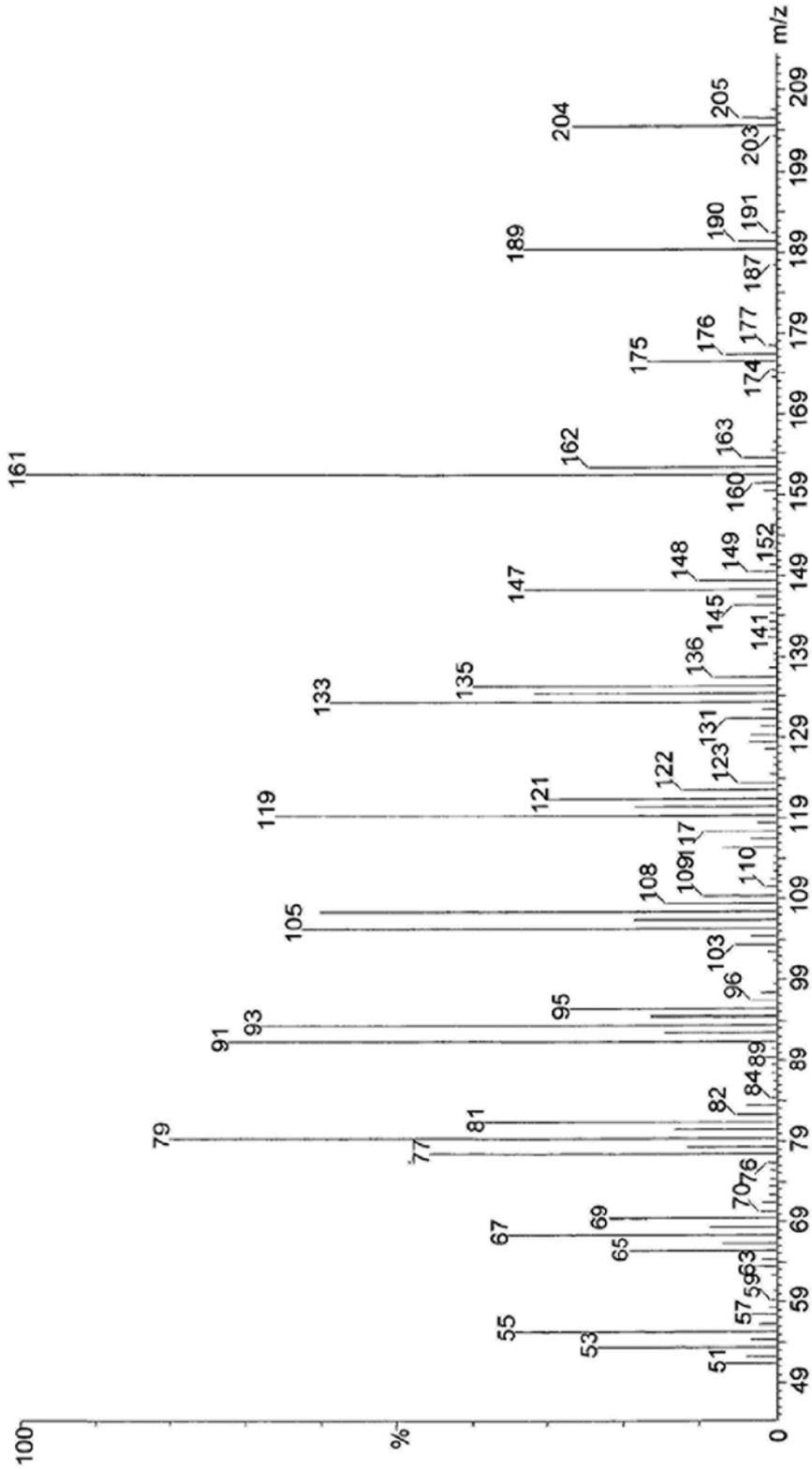


FIGURA 3B

```

EGVS (1)  MSLNVLSTSGSAPTTKSSEITRRSANYHPSLWGDKFLEYSSPDHLKNDSFTEKKHEQLKE 60
CVS (14)  -----MSSGETFRPTADFHPSLWRNHFLKGASDFKTV DHTATQERHEALKE 46
           .*. * :*:***** :*: * : : : *::** **

EGVS (1)  EVKMLVETVQKPQQQLNLINEIQRLGLSYLFEPEIEAALQEISVTYDEFCSTDADDLH 120
CVS (14)  EVRRMITDAEDKPVQKLRLIDEVQRLGVAYHFEKEIEDAIQKLCPIYID---SNRADLH 102
           **::*: : : * * :*:*****: * ** ** *::: . * : : : **

EGVS (1)  NVALSFRILREHGHNVSSDVQKFMDSNGKLDYLVNDARGLLSLYEATHFRVHNDCKLE 180
CVS (14)  TVSLHFRLLRQQGIKISCDVFEKFKDDEGRFKSSLINDVQGMLSLYEAAYMAVRGEHILD 162
           .*: * **::*: * :*:*****: * .::*: * :*:*****: : : *::: * :

EGVS (1)  ELLSVTTSRLEHLKS--HVKYPLEDEISRALKHPLHKLNLRGARYYISYIEKFDH--N 236
CVS (14)  EAIAFTTTHLKS LVAQDHVTPKLAEQINHALYRPLRKTLPRLRLEARYFMSMINSTSDHLYN 222
           * :*:*****: * : * . * :*:*****: * * * * * :*:*****: : . . * *

EGVS (1)  KLLLEFAKLDNFRLQKMYQHELALHTRWVKDLDFTNKLPFARDRIVEGYFWILGMYFEPE 296
CVS (14)  KTLNFAKLDNFNILLELHKEELNELTKWVKDLDFTTKLPYARDRLVELYFWDLGTYPFEPQ 282
           * **::***** * : : : * . * :*:*****: * * * * * :*:*****: * * * * * :

EGVS (1)  RKDVREFLNRFALITVVDITYDVYGTFKELLLFTDAIERWGTSDLDQLPGYMRIYQAL 356
CVS (14)  YAFGRKIMTQLNYILSIIDITYDAYGTLEELSLPTEAVQRWNIEAVDMLPEYMKLIYRTL 342
           * : : : : : : : * * * * * : * * * * * : * * * * * . . : * * * * * : * * * * *

EGVS (1)  MDVYNQMEEKLSMKADCPTYRLEFAIETVKAMFRSYLEEARWSKEHYIPSMEEYMTVALV 416
CVS (14)  LDAFNEIEEDMAKQG--RSHCVRYAKEENQKVGAYSVQAKWFSEGYVPTIEEYMPIALT 400
           :*::*:*****: : . : : : * * : : : * : * * . * :*:*****: * * .

EGVS (1)  SVGYKTILTNSFVGMGDIATREVFVWFNSPLIIRASDLIARLGGDGGHEEEQKKGDA 476
CVS (14)  SCAYTFVITNSFLGMGDFATKEVFEWISNNPKVVKAASVICRLMDDMQGHEFEQKRGHVA 460
           * . * . :*:*****: * * * * * : * * * * * : * * * * * : * * * * * : * * * * *

EGVS (1)  TAIECYIKENHVTKHEAYDEFQKQIDNAWKDLNKEALR-PPFVPMFITRVVHFTRAIHV 535
CVS (14)  SAIECYTKQHGVSKEEAIKMFEEEVANAWKDINEELMMKPTVVARPLLGTILNLARAIDF 520
           :***** * : * : * . * . * : : : * * * * * : * * : * * . : : : : * * * .

EGVS (1)  IYADFSGDYTRSDKAIRGYITSLLDPIPL 565
CVS (14)  IYKE-DDGYTHS-YLIKDQIASVLGDHVPF 548
           ** : . * * * * * : * . * : * * * * : * :
    
```

FIGURA 4A

EGVS (1) MSLNVLSTSGSAPTTSSEITRRSANYHPSLWGDKFLEYSSPDHLKNDSFTEKKHEQLKE 60  
 CVS (15) -----MSSGETFRPTADFPSPSLWRNHFLKGASDFKTVDHDTATQERHEALKE 46  
 .\*. \* :\*:\*\*\*\*\* :\*: \* : : : \* :\*:\*\* \*\*

EGVS (1) EVKMLVETVQKPQQQLNLINEIQRLGLSYLFEPEIEAALQEISVTYDEFCCSTDADDLH 120  
 CVS (15) EVRRMITDAEDKPVQKRLRLIDEVQRLGVAYHFEKEIEDAILKLCPIYID---SNRADLH 102  
 \* :\*:\*: : \* \* :\*:\*\*\*\*\* :\* \*\* \*\* \* : : . \* : : : \*\*

EGVS (1) NVALSFRILREHGHNVSSDVFQKFMDSNGKLDYLVNDARGLLSLYEATHFRVHNDKLE 180  
 CVS (15) TVSLHFRLRQQGIKISCDVFEKFKDDEGRFKSSLINDVQGMLSLYEAAYMAVRGEHILD 162  
 .\*: \* :\*:\*: : \* :\*:\*\*\*\*\* \* :\*:\*. \* :\*:\*\*\*\*\* :\* : : . \*

EGVS (1) ELLSVTTSRLEHLKS--HVKYPLEDEISRALKHPLHKELNRLGARYYISYIEKFDH--N 236  
 CVS (15) EAIAFTHLKSLSVAQDHVTPKLAEQINHALYRPLRKTLPRLARYFMSMINSTSDHLYN 222  
 \* : : \* :\*: \* : \* \* . \* :\*:\*\* :\*: \* \* \* \* \* :\*: : . . \* \*

EGVS (1) KLLLEFAKLDNRLQKMYQHELALHTRWVKDLDFTNKLPPFARDRIVEGYFWILGMYFEPE 296  
 CVS (15) KTLNFAKLDNILLEPHKEELNELTKWVKDLDFTTKLPYARDRLVELYFWDLGTYFEPQ 282  
 \* \* :\*:\*\*\*\*\* \* : : : \* . \* :\*:\*\*\*\*\* :\*:\*\*\*\*\* :\* \*\* \*\* \* \* \* \* :

EGVS (1) RKDVREFLNRFALITVVDDTYDVYGFKELLFTDAIERWGTSDLDQLPGYMRIIYQAL 356  
 CVS (15) YAFGRKIMTQLNYILSIIDDTYDAYGTLEELSPLTEAVQRWNIEAVDMLPEYMKLIYRTL 342  
 \* : : : : : : : \* \* \* \* \* :\*:\*\* \* \* \* : \* \* \* \* : \* \* \* \* : \* \* \* \* :

EGVS (1) MDVYNQMEEKLSMKADCPTYRLEFAIETVKAMFRSYLEEARWSKEHYIPSMEEYMTVALV 416  
 CVS (15) LDAFNEIBEDMAKQG--RSHCVRYAKEENQKVIGAYSVQAKWFSEGYVPTIEEYMPIALT 400  
 :\* :\*:\*\* : : : : : \* \* : : : \* :\*: \* \* \* :\*:\*\*\*\*\* :\* .

EGVS (1) SVGYKTILTNSFVGMGIATREVFVWFNSPLIIRASDLIARLGDDIGGHEEQKKGDA 476  
 CVS (15) SCAYTFVITNSFLGMGDFATKEVFEWISNPKVKAASVICRLMDDMQGHEFEQKRGHVA 460  
 \* . \* . :\*:\*\*\*\*\* :\*:\*\*\*\*\* :\* \* : : : : \* \* \* \* \* \* \* \* \* \* \* .

EGVS (1) TAIECYIKENHVTKHEAYDEFQKQIDNAWKDLNKEALR-PPVPMTFITRVVHFTRAIHV 535  
 CVS (15) SAIECYTKQHGVSKEEAIKMFEEVANAWKIDEBLMMKPTVVARPLLGTILNLARAIDF 520  
 :\*\*\*\*\* \* : \* : \* \* . \* : : : \* \* \* \* : \* \* \* . : : : \* \* \* .

EGVS (1) IYADFSGYTRSDKAIRGYITSLLDPIPL 565  
 CVS (15) IYKE-DDGYTHS-YLIKQIASVLGDHVPF 548  
 \*\* : . \* \* \* \* \* \* : . \* : \* \* \* \* \* \* :

FIGURA 4B

EGVS (1) MSLNVLSTSGSAPTTSSEITRRSANYHPSLWGDKFLYSSPDHLKNDSPTEKKHEQLKE 60  
VVS (16) MSTQVSASS-LAQIPQPK--NRPVANFHPNIWGDQFITYTPEDKVTR-ACKEEQIEDLKK 56  
\*\* : \* : : \* \* . . . . \* \*\* : \* \* : \* \* : \* \* : \* \* : \* \* : \* \* : \* \* : \* \* :

EGVS (1) EVK-KMLVETVQKPPQQLNLINEIQRLGLSYLFEPEIEAALQEISVTYDEFCCSTDADDL 119  
VVS (16) EVKRKLTAAAVANPSQLLNFIQVAVQRLGVAYHFEQEIEEALQHICNSFHD--CNDMDGDL 114  
\*\*\* \* : . : \* : \* . \* \*\* : \* : \* \* \* \* : \* \* \* \* \* \* . . . : \* . . \*\*

EGVS (1) HNVALSFRILREHGHNVSSDVQKFMDSNGKLDYLVNDARGLLSLEYEATHFRVHNDKDL 179  
VVS (16) YNIALGFRLLRQQGYTISCDIFNKFTDERGRFKEALISDVRGMLGLYEAAHLRVHGEDIL 174  
: \* : \* \* . \* \* : \* \* : \* \* : \* \* : \* \* : \* \* : \* \* : \* \* : \* \* : \* \* :

EGVS (1) ELLSVTTSRLEHLKSHVKYPLEDEISRALKHPLHKELNRLGARYYISYIEKFDShnkLL 239  
VVS (16) AKALAFTHLTKAMVESLGYHLAEQVAHALNRPKGLERLEARWYISVYQDEAFHDKTL 234  
: \* : \* \* : \* \* : . : \* \* : : : \* \* : \* \* : \* \* : \* \* : \* \* : \* \* : \* \* :

EGVS (1) LEFAKLDFNRLQKMYQHELALHTRWKKDLDFTNKLPFARDRIVEGYFWILGMYFEPERKD 299  
VVS (16) LELAKLDFNLVQSLHKEELSNLARWVKELDFATKLPFARDRLVEGYFWMHGVPYFEPQYLR 294  
\*\* : \* \* \* \* \* : \* . . . . \* \* : \* : \* \* \* \* : \* \* . \* \* \* \* \* : \* \* \* \* \* : \* \* \* \* \* :

EGVS (1) VREFLNRFALITVDDTYDVYGTFKELLLFTDAIERWGTSDLDQLPGYMRRIYQALMDV 359  
VVS (16) GRRIILTKVIAMTSLDDIHDAYGTPEELKLFIEAIERWDINSINQLPEYMKLCYVALLDV 354  
\* : \* : \* : \* : : : \* \* : \* . \* \* \* : \* \* \* \* : \* \* \* \* . . . : \* \* \* \* : \* \* \* \* \* :

EGVS (1) YNQMEEKLSMKADCPTYRLEFAIETVKAMFRSYLBEARWSKEHYIPSMEEYMTVALVSVG 419  
VVS (16) YKEIEEEMEKEGNQ--YRVHYAKEVMKNQVRAYFAEAKWLHEEHVPAFEEYMRVALASSG 412  
\* : : \* \* : . . : \* \* : : \* \* \* \* . \* : \* : \* \* \* \* : \* : : \* : \* \* \* \* \* \* \* \* :

EGVS (1) YKTILTNSFVGMGDIATREVFVWFNSPLIIRASDLIARLGDDIGGHEEQKKGDAATAI 479  
VVS (16) YCLLATTSFVGMGEIATKEAFDWTSDPKIMSSSNFITRLMDDIKSHKFEQKRGHVASAV 472  
\* : \* . \* \* \* \* : \* \* : \* . \* \* \* \* . \* \* \* : \* : \* \* \* \* \* \* \* \* : \* \* : \* \* . \* \* : \* :

EGVS (1) BCYIKENHVTKHEAYDEFQKQIDNAWKDLNKEALRPFVPMFTITRVVHFTRAIHVIYAD 539  
VVS (16) BCYMKQYGVSEEQVYSEFQKQIENAWLDINQELKPTAVSMPLLARLLNLRTRTMDVIYKE 532  
\* \* \* : \* : \* : . . \* . \* \* \* \* : \* \* : \* . \* \* \* . \* . \* . . : \* : \* : \* \* \* \* \* \* \* \* :

EGVS (1) FSDGYTRSDKAIRGYITSLLDVPIPL 565  
VVS (16) Q-DSYTHVGKVMRDNIASVFINAVI- 556  
\* . \* \* : . \* : \* . \* : \* : \* : \* : :

FIGURA 4C

```

EGVS (1)  MSLNVLSTSGSAPTTKSSEITRRSANYHPSLWGDKFLEYSSPDHLKNDSFTEKKHEQLKE 60
CNVS (17) -----MPVKDALRRRTGNHHPNLWTDDFIQSLNSPYS--DSSYHKHREILID 44
          .:  **::*:*** **.*:::  .. :  ** .*::* * :

EGVS (1)  EVKKMLVETVQKPOQLN---LINEIQRGLSYLFEPEIEAALQEISVTYDEFCCSTDAD 117
CNVS (17)  EIRDMFSNGEGDEFGVLENIWFVDVVQRLGIDRHFQEEIKTALDYI---YKFWNHDSIFG 101
          *::.*: :  .  *:  ::: :*****.  *:  ***::*: *  *. :  .:  .

EGVS (1)  DLHNVALSFRILREHGHNVSSDVFQKFMDNSGKLDYLVNDARG---LLSLYEATHFRV 173
CNVS (17)  DLNMVALGFRILRLNRYVASSDVFKKFKGEEGQFSGFESSDQDAKLEMLNLYKASELDF 161
          ***:  ***.***** :  :  .*****:**  ..*::::..:  .*  .  :*.*::*:..:  .

EGVS (1)  HNDDKLEELLSVTTSRLEHLKSHV-----KYPLEDEISRALKHPLHKELNRLGARYYIS 227
CNVS (17)  PDEDILKEARAFASMYLKHVIKEYGDIQESKNPLLMEIEYTFKYPWRCRLPRLEAWNFIH 221
          ::*  **  ::::  **::  ..  *  **  **  .:***:  :  .*  **  *  :*

EGVS (1)  IYEKFD-----HNKLLLEFAKLDNFRLQKMYQHELAHLTRWWKDLDFTNK 273
CNVS (17)  IMRQQDCNISLANNLYKIPKIYMKILELAILDFNILQSQHQHEMKLISTWWKNS-SAIQ 280
          *  .:  *.  :  *  :***:*  ****  **  .:***:  :  ***:  :  :

EGVS (1)  LPFARDRIVEGYFWILGMYFEPERKDVREFLNRVVALITVDDTYDVYGTFKELLFTDA 333
CNVS (17)  LDFFRHRHIESYFWWASPLFEPEFSTCRINCTKLSTKMFLLDDIYDTYGTVEELKPFPTT 340
          *  *  *.  *  :*.***  .  ****  .  *  .:  :  :  :**  **.*.***.***  **  :

EGVS (1)  IERWGTSDDLQPLGYMRIIYQALMDVYNQMBEELSMKADCPTYRLEFAIETVKAMFRSYL 393
CNVS (17)  LTRWDVSTVDNHPDYMKIAFNFSYEIYKEIASEAERKHGP--FVYKYLQSCWKSIEAYM 398
          :  **..*  *:  *.***:*  :  :  :*::::  .:  .  *  .  /  :  .  *  :  :*:

EGVS (1)  EEARWSKEHYIPSMEEYMTVALVSVGYKILTNSFVGMGDIATREVFWEVFN-SPLIIRA 452
CNVS (17)  QEAEWIASNHIPGFDEYLMNGVKSSGMRILMIHALILMDTPLSDEILEQLDIPSSKSQAL 458
          :**.*  .:***.***:  .:  *  *  :  :  :*:  .  :  *::*  :  *.

EGVS (1)  SDLIARLGDDIGGHEEBQKKGDAATAIECYIKENH-VTKHEAYDEFQKCIDNAWKDLNKE 511
CNVS (17)  LSLITRLVDDVKDFEQAHEMMASSIECYMKDNHGSTREDALNLYKIRIESCVQELNKE 518
          .**:*  **  :  .*:*  *:  *:  *****:*  **  *::*  :  :  :*..  :****

EGVS (1)  ALRPFVPMTFITRVVHFTRAIHVYADFSGYTRSD-KAIRGYITSLLVDFIPL 565
CNVS (17)  LLEPSNMHGSRNLYLNVGMRVIFFMLNDGDLFTHSNRKEIQDAITKFFVEPIIP 573
          *.  *  :  :*  .  :.  :  :  :  .  *  :*:  *  *.  **.*:***
    
```

FIGURA 4D

EGVS (1) MSLNVLSTSGSAPTTKSSEITRRSANYHPSLWGDKFLEYSSPDHLKNDSFTEKKHEQLKE 60  
V277 (31) MSTQVSASS-LAQIQPK--NRPVANFHPNIWGDQFITYTPEDKVTR-ACKEEQIEALKE 56  
\*\* : \* : \* \* . . . . \* \*\* : \* \* . : \* \* \* : \* : \* : \* . : \* : \* \* \* \*

EGVS (1) EVKKMLVETVQKPPQQNLNLINEIQRLGLSYLFEPEIEAALQEISVTYDEFCCSTDADDLH 120  
V277 (31) EVRRMILATGRKPIQKRLRLIDEVQRLGVAYHFEKEIEDMLDHIYR-ADPYFEAHEYNDLH 115  
\* \* : \* : \* \* : \* \* \* \* : \* \* \* \* : \* \* \* \* \* : \* \* \* \* \* : \* \* \* \* \* : \* \* \* \* \*

EGVS (1) NVALSFRILREHGHNVSSDVFQKFMDSNGKLDYLVNDARGLLSLYEATHFRVHNDKLE 180  
V277 (31) TVSLHFRLLRQQGIKISCDVFEQFKDDEGRFKSSLINDVQGMLSLYEAAYMAVRGEHILD 175  
. \* : \* \* \* : \* \* : \* \* \* : \* \* . : \* : \* . \* : \* \* . : \* \* \* \* : \* : \* . \* :

EGVS (1) ELLSVTTSRLEHLKSHVKYPLEDEISRALKHPLHKELNRLGARYYISYIEKFDShnKLLL 240  
V277 (31) EAIAFTTTHLQSAAPHLKSPLAEQINHALYRPLRKTLPRLARYIMSVMYQDEAFHNKTL 235  
\* : : \* \* : \* : \* : \* \* \* \* : \* : \*

EGVS (1) EFAKLDNRLQKMYQHELALTRWVKDLDFTNKLPFARDRIVEGYFWILGMYFEPERKDV 300  
V277 (31) NFAKLDNILLDLHKEELNELTKWQDLDFTTKLPYARDRLVELYFWDLGTIFYESQYAFG 295  
: \*

EGVS (1) REFLNRVFALITVDDTYDVYGTFKELLFTDAIERWGTSDDLQPLGYMRIIYQALMDVY 360  
V277 (31) RKIMTKLNYILSIIDTTYDAYGTLEECTMFSEAVARWNIEAVDMLPDYMRIIYRTLDDTF 355  
\* : : : : : : : \*

EGVS (1) NQMEEKLSMKADCPTYRLEFAIETVKAMFRSYLEEARWSKEHYIPSMEEYMTVALVSVGY 420  
V277 (31) NEIEEDMAQR--RSHCVRYAKEEIQKVGAYYVQAKWFSEGYVPTIEEYMPIALTSCAY 413  
\* : \* \* \* : : : : \* \* \* : : : \*

EGVS (1) KTILTNSFVGMGDIATREVFVFN SPLIIRASDLIARLGGDIGGHEEQKKGDAATAIE 480  
V277 (31) RFVITNSPLGMGDFATKEVFEWISGNPKVVKASVICRLMDDMQGHEFEQKRGHVASAIE 473  
: : \*

EGVS (1) CYIKENHVTKHEAYDEFQKQIDNAWKDLNKEALRPPFPMT-FITRVVHFTRAIHVIYAD 539  
V277 (31) CYTKQHGVSKEEAIKMFESDVANAWKDINEBLMMKPPVVARPLLGITILNLARAIIDFIYKE 533  
\* \* \* : \* \* \* \* \* \* : : : \*

EGVS (1) FSDGYTRSDKAIRGYITSLLDVPIPL 565  
V277 (31) -DDGYTHS-YLIKEQIASVLGDHVPF 557  
. \*

FIGURA 4E



EGVS (1) MSLNVLSTSGSAPTTKSSEITRRSANYHPSLWGDKFLEYSSPDHLKNDSTFEKKHEQLKE 60  
 PFVS (38) -----MASEQAQNHRPVADFPSPSLWGDQFVKYDSCPQVQKK--YSNTVDVLKK 46  
 :.:.: \* \*.: \*\*\*\*\*:.\*. :.:. .: : \*\*:

EGVS (1) EVKMLVETVQKPQQQLNLINEIQRLGLSYLFEPEIEAALQEISVTYDEFCCS-TDADDL 119  
 PFVS (38) EVKGMITAPGTMVDTMELIDTIERLGVSFHFQDEIEQKLQFPDLKTDYCNDDDAYDL 106  
 \*\*\* \*.: . \* : :.\*\*\*: \*;\*\*\*\*:\*: \* : \*\*\* \*\*.: :.\* . \*\* \*\*

EGVS (1) HNVALSFRILREHGHNSSDVPQKFMDSNGKLDYLVNDARGLLSLYEATHFRVHNDK 179  
 PFVS (38) YTVLHFRLFRQHGYRISCDIFGRWIDGNGKPKKGLKSDGKSLLSLYEASYLRTGETIL 166  
 :.\*\*\* \*\*.:\*:\*:\*:\*:\*:\* :.:\*.\*\*\*:\*: \* .\*. .\*\*\*\*\*:.\*.: \*

EGVS (1) EELLSVTTSRLEHLKSHVKYPLEDEISRALKHPLHKELNRLGARYYISIEYKFDShnkll 239  
 PFVS (38) DEALDFAAASLKSIAPHLQSPLGKQVVHALVQPLHFGNPRIEARNFISIEYEEYEGMNEAL 226  
 :\* \*.:.: \* : .\*: \*\* .: : \* : \*\*\* \* : \*\* :\*\*\*\*\*:.. \* : \*

EGVS (1) LEFAKLDFNRLQKMYQHELHLTRWKKDLDFTNKLPFARDRIVEGYFWILGMYFEPERKD 299  
 PFVS (38) LRLAKLDYNLLQMLHKEELHQVSRWKKDLDLITKLPYARDRVVECFWAVGVYHEPQYSR 286  
 \*.:\*\*\*\*\*:\* \*\* :.:.\* :.:\*\*\*\*\*: .\*\*\*;\*\*\*\*:\* \*\* :\*\* :\*:\*.\*.\*: .

EGVS (1) VREFLNRFALITVVDDTYDVGTFKELLLFTDAIERWGTSDLDQLPGYMRIIYQALMDV 359  
 PFVS (38) ARVMLTKTIVMTSIIDDTYDAYGTIEELDIFTEAIERWNVEETKRLPEYMKPLYKALLEL 346  
 .\* :\*.:.: :.:\*\*\*\*\*.\*\*\*:\* \*\* :\*\*\*\*\*. .: : \*\* \*\* :\*:\*:\*:.

EGVS (1) YNQMEEKLSMKADCPTYRLEFAIETVKAMFRSYLEEARWSKEHYIPSMEEYMTVALVSVG 419  
 PFVS (38) YKQFEQELEKEG--RSYVAYYAIESLKLVRSYRIEAKWFIQGYLPPYEEYLKNALITCT 404  
 \*\*:\*:\*:\* . . \* :\*\*\*\*\*:\* :.\*\*\* \*\*:\* :\*:\* .\*\*\*. \*\*:

EGVS (1) YKTILTNSFVGMGIATREVFVFNPLIIRASDLIARLGDIGGHEEQKKGDAATAI 479  
 PFVS (38) YCYHTTSSLGV-ESAIKENFEWLSNPKMLVAGLLICRLIDDIATYILERGRQVATGI 463  
 \* \*.:\*:\*: \* : \* :\* \*\*:\*.\* :. \* .\*\*.\* \*\* . : \* :\*:\*.\*.\*

EGVS (1) ECIKENHVTKHEAYDEFQKQIDNAWKDLNKEALRP-FPVPMTFITRVVHFTRAIHVIYA 538  
 PFVS (38) ESYMKDNATQEEPIAKIFEIATDAWKDINDECLRPSLYNSRDVLMRIFNLERIIDVTYK 523  
 \*.\*:\*:\* .\*:.\*. : : : \*\*\*\*\*:\*.\*.\* : . .: \*.: : \* \*.\* \*

EGVS (1) DFSDGYTRSDKAIRGYITSLLDPIPL 565  
 PFVS (38) GNQDRYTQPEKVLKPHIIVFFDPIPLI 550  
 . .\* \*\*.:\*.: : \* :.\*\*\* :

FIGURA 4G







EGVS (1) MSLNVLSTSGSAPTTKSSEITRRSANYHPSLWGDKFLEYSSPDHLKNDSTFEKKHEQLKE 60  
SMSS (21) MENQKMPISSVPLKDLNMI SRPIANFPPIWGD RFINYT CEDEN-DQTQKERQVEELKE 59  
\* . : : . \* . . . . \* : \* \* : \* : \* : \* : \* : \* . \* . : : . \* : : \* : \* \* \*

EGVS (1) EVKMLVETVQK PQQLNLINEIQRLGLSYLFEPEIEAALQEISV-TYDEFCCSTDADDL 119  
SMSS (21) QVRRELAATVDKPLQQLNIIDATQRLGIAYLPENEIEESLKH IYLHTYVENNCFEGSDDL 119  
: \* : : \* . \* : \* \* \* \* \* : \* : \* \* \* : \* \* \* \* \* : \* : \* \* \* \* \* . : \* \* \* \* \*

EGVS (1) HNVALSFRILREHGHNVSSDV FQKFMD SNGKLDYLVNDARGLLSLYEATHFRVHND DKL 179  
SMSS (21) YSVALWFRLLRQNGYK VSCDVFNKFRDNEGNFKNNLMEDAKGLELYEATHVSIHGE EML 179  
: . \* \* \* \* \* \* : \* : \* : \* : \* : \* \* \* \* \* \* \* \* . : \* : \* : \* : \* \* \* \* \* . : \* : : \* \*

EGVS (1) EELLSVTTSRLEHLKSHVKYPLEDEISRALKHPLHKELNRLGARYYIS IYEFDSHNKLL 239  
SMSS (21) DDALEFTKTRLESVVSHLNYPLAEQVRHALYQPLHRGLPRLEAVYFFRIYEASHNKAL 239  
: : \* . . \* . : \* \* \* : \* : \* : \* \* \* : : : \* \* : \* \* \* : \* \* \* \* \* : \* \* \* \* \* \* \* \*

EGVS (1) LEFAKLDFNRLQKMYQHELALHTRWVKDLDFTNKLPFARDRIVEGYFWILGMYFEPERKD 299  
SMSS (21) LKLAKLDFNLLQSFHKKELSDIARWWSLDFAAKFPFARDRLVEGYFWVLGVYFEPQYSL 299  
\* : \* \* \* \* \* \* \* \* : \* : \* : \* : \* : \* .

EGVS (1) VREFLN RVFALITVVD DTYDVYGT FKE LLLFTDAIERWGTSDDLQ LPGYMRIIYQALMDV 359  
SMSS (21) ARKIIIKVFTMISTIDDIYDAYGTLDELKLF TKAMQRWDVGS L DQLPEYMKPCYKSILDV 359  
. \* : : : \* : \* : \* : \* : \*

EGVS (1) YNQMEEKLSMKADCPT YRLEFAIETVKAMFRSYLEEARWSKEHYIPSMEEYMTVALVSVG 419  
SMSS (21) YNEIEEEMANQG--SLFRMHYAKEVMKTIVEGYMDEAKWCHEKYVPTFQ EYMSVALVTSG 417  
\* \* : \* \* \* : : . . : \* : \*

EGVS (1) YKTILTNSFVGMGDIATREVF EWVFN SPLIIRASDLIARLGDDIGGHEEEQKKGDAATAI 479  
SMSS (21) YTFLTTISYLGMEIASKEAFDWLFSHPPIEASESVGRLMDDMRSHKFEQERGHVASGI 477  
\* . : \* \* \* : \*

EGVS (1) ECIYIKENHVTKHEAYDEFQKQIDNAWKDLNKEALRPPVPMTFITRVVHFTRAIHVIYAD 539  
SMSS (21) ECYMKQYGVTEEEAHDEFKRLVKAWKDINEECLRPYRVPKPLLTRILNLTRVIDVIYKN 537  
\* \* \* : \* : \* \* : \* \* \* \* \* \* : \*

EGVS (1) FSDGYTRSDKAIRGYITSLLVDP IPL 565  
SMSS (21) -EDGYTHVKKAMKDNIASLLIDP VIV 562  
. \* \* \* \* : \* \* : . \* \* \* \* \* \* \* \* : :

FIGURA 5D



EGVS (1) MSLNVLSTSGSAPTTKSSEITRRSANYHPSLWGDKFLEYSSPDHLKNDSTFEKKHEQLKE 60  
 CDCS (23) MSLEVSASPAKVIQNAGKDSTRRSANYHPSIWGDHFLQYTC-DTQETDDGSNVKHLELKK 59  
 \*\*\*:\* :..... . . . : \*\*\*\*\*:\*\*\*:\*\*\*:\*. \* :.\*. : : \*\* :\*\*:

EGVS (1) EVKMLVETVQKPQQQLNLINEIQRLGLSYLFEPETEAALQEIS-VTYDEFCCSTDADDL 119  
 CDCS (23) EIRRMLKADN-KPSRTLQLIDAIQRLGVSYHFESIDEILGKMHKASQSDLCDNENDEL 118  
 \*:::\*\* \*\*.: \*:\*\*: \*\*\*\*\*:\*\* \*\*.\*: \* : : . : \* . \*.: \*\*:

EGVS (1) HNVALSFRILREHGHNVSSDVFQKFMDSNGKLDYLVNDARGLLSLYEATHFRVHNDKDL 179  
 CDCS (23) YYISLHFRLRQNGYKISADVFKFKDTDGNFKTSLAKDVRGMLSLYEATHLGVHEEDIL 178  
 : : : \* \*\*:\*:\*\*:\*\*\*:\*\*\*:\*\*\*:\*\*\* \* : : \* : \* \* . : \* . \* . \* . \* . \* . \* . \* . \* . \* . \*

EGVS (1) EELLSVTTSRLEHLKSH-VKYPLEDEISRALKHPLHKELNRLGARYYISIEYKFDShnkl 238  
 CDCS (23) DEALAFTTSHLESIATHQIRSPVQVQKHALVQPIHRGFQRLEARQYIPIYQEEsPNEA 238  
 : \* \* . : \* . \*

EGVS (1) LLEFAKLDFNRLQKMYQHELALHTRWVKDLDFTNKLPFARDRIVEGYFWILGMYFEPERK 298  
 CDCS (23) LLTFAKLDFNKLQKPHQKELGDISRWWKELDFAHKLPFIRDRVAECYFWILGVYFEPQYS 298  
 \*\* \*\*\*\*\*:\* \*\* :\*:\*..:\*\*\*:\*:\*\*:\*\*\* \*\*:\* . \* \*\*\*\*\*:\*\*\*: .

EGVS (1) DVREFLNRVFALITVVDITYDVYGTfKELLFTDAIERWGTSDDLQPLGYMRIIYQALMD 358  
 CDCS (23) FARRILTKVISMTSVIDDIYDVYKIEEELFTSAIERWDISAIDLPEYMKLCYRALLD 358  
 .\*.:\*.\*:\*\*\*: :\*:\* \*\* \*\* . : \* \* \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \*

EGVS (1) VYNQMBEKLsmKADCPTyrLEFAIETVKAMFRSYLEEARWSKEHYIPSMEEYMTVALVSV 418  
 CDCS (23) VFSEAEKDLAPQGKS--YRLYYAKEAMKNMVKNYFYEAkwclQNYVPTVDEYMTVALVTS 416  
 \*.: : \* . : \* . : \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \*

EGVS (1) GYKTILTNSFVGMGDIATREVFewVFNsPLIIRASDLIARLGDDIGGHEEEQKKGDAATA 478  
 CDCS (23) GSPMLSTTSFVGMGDIVTKESFEWLFsNPRFIRASSIVCRLMDDIVSHKFEQSRGHVASS 476  
 \* : \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \*

EGVS (1) IECYIKENHVTKHEAYDEFQKQIDNAWKDLNKEALRPPVPMTFITRVVHFTRAIHVIYA 538  
 CDCS (23) VECYMKQHGATEEEBACNEFRKQVSNawKDINEDCLRPTVVPMLLMRILNLTRVIDVIYK 536  
 :\*\*\*:\*: . : \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \*

EGVS (1) DFSDGYTRSDKAIRGYITSLLDPIPL 565  
 CDCS (23) -YEDGYTHSAVVLKDFVASLFINPVPI 562  
 : . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \* . \*

FIGURA 5F

ES 2 647 828 T3

EGVS (1) MSLNVLSTSGSAPTTKSSEITRRSANYHPSLWGDKFLEYSSPDHLKNDSSFTEKKHEQLKE 60  
 TEAS (24) -----MASAAVANYEEEEIVRPVADFSPSLWGDQFLSFSIKNQVAEK--YAKEIEALKE 51  
 :.:\* :. ..\*\*.\* \*.: \*\*\*\*\*:\*.:\* :.: :. \*: \* \*\*\*

EGVS (1) EVKMLVETVQKPOQQLNLINEIQRLGLSYLFEPEIEAALQEISVTYDFCCSTDADDLH 120  
 TEAS (24) QTRNMLLATGMKLADTLNLIIDTIERLGISYHFEKEIDDILDQIYNQ-----NSNCNDLC 105  
 :.:\*\*:\* \* \* : \*\*\*\*\*: \*:\*\*:\* \* \* \*\*; \*.:\* :.:\*\*

EGVS (1) NVALSFRILREHGHNVSSDVFQKFMDSNGKLDYLVNDARGLLSLYEATHFRVHNDKLE 180  
 TEAS (24) TSALQFRLLRQHGFNISPEIFSKFQDENGKFKESLASDVLGLLNLYEASHVRTHADDILE 165  
 . \*\*.\*:\*\*:\*.\*.\*.:\*\*.\* \*.\*:\*\*: \*..\* \*\*.\*:\*\*:\*.\*.\* \*\* \*\*

EGVS (1) ELLSVTTSRLEHLKSHVKYPLEDEISRALKHPLHKELNRLGARYIIS- IYEKFDShnKLL 239  
 TEAS (24) DALAFSTIHLESAAPHLKSPREQVTHALEQCLHKGVPVTRFFISSIYDKEQSKNNVL 225  
 : \*.:\*\*:\* \*\* .\*: \* \*\*.:\*\*:\*:\*\*: \*\* \* : \* :\*:\*\* \*\*:\* \*:\*\*:\*\*

EGVS (1) LEFAKLDFNRLQKMYQHELALTRWVKDLDFTNKLPFARDRIVEGYFWILGMYFEPERKD 299  
 TEAS (24) LRFAKLDFNLLQMLHKQELAQVSRWVKDLDFVTTLPYARDRVVECYFWALGVYFEPQYSQ 285  
 \*.\*:\*\*:\* \*\* :.:\*\*:\*:\*\*:\*:\*\*:\*..\*\*:\*:\*\*:\* \*\* \*\* \*\*:\*:\*\*: .:

EGVS (1) VREFLNRFALITVDDTYDVYGTFKELLLFTDAIERWGTSDDLQPLGYMRIIYQALMDV 359  
 TEAS (24) ARVMLVKTISMISIVDDTFDAYGTVKELEYTDAIQRDINEIDRLPDYMKISYKAILDL 345  
 .\* :\* :.:\*\*:\*:\*\*:\*:\*\*:\* \*\* \* :\*\*:\*:\*\*. :.:\*\*:\* \*\*:\* \*:\*\*:\*\*:

EGVS (1) YNQMEEKLSMKADCPTYRLEFAIETVKAMFRSYLEEARWSKEHYIPSMEEYMTVALVSVG 419  
 TEAS (24) YKDYEKELSSAG--RSHIVCHAIERMKEVVRNYNVESTWFIEGYTPPVSEYLSNALATTT 403  
 \*.: \*:\*\*:\* . : : .\*\* \* : \* .\*. \* \* : \* \* \* \*..\*\*:\* \*\*..

EGVS (1) YKTILTNSFVGMGDIATREVFVFN SPLIIRASDLIARLGDDIGGHEEQKKGDAATAI 479  
 TEAS (24) YYYLATTSYLGK-SATEQDFEWLSKNPKILEASVIIICRVIDDTATYEVEKSRGQIATGI 462  
 \* : \*.\*:\*\*:\* . \*\*.: \*\*:\* :.\* \*:\*\*:\* \*\*:\* \*\* . : \* \*:\*\*:\* \*\*.\*

EGVS (1) ECIKENHVTKHEAYDEFQKQIDNAWKDLNKEALRFPVPMTFITRVVHFTRAIHVIYAD 539  
 TEAS (24) ECCMRDYGISTKEAMAKFQNAETAWKDINEGLLRPTVPSTEFLLTPILNLRIVEVTYIH 522  
 \*\* :.: :.:\*\*:\* \*\*:\* :.\*:\*\*:\*:\*\*: \*\* \*\* . \*:\* :.:\*\*:\* .\*. \* .

EGVS (1) FSDGYTRSDKAIRGYITSLLDPIPL 565  
 TEAS (24) NLDGYTHPEKVLKPHIINLLVDSIKI 548  
 \*\*\*\*\*:.\*.: : \* .\*\*\*\*\*.\* :

FIGURA 5G

EGVS (1) MSLNVLSTSGSAPTTKSSEITRRSANYHPSLWGDKFLEYSSPDHLKNDSFTEKKHEQLKE 60  
HPS (25) -----VDNQVAEKYA-----QEIETLKE 18  
                                  .: : \* \*\*\*

EGVS (1) EVKKMLVETVQKPO-QQLNLINEIQRLGLSYLFEPFIEAALQEISVTYDEFCCSTDADDL 119  
HPS (25) QTSTMLSAACGTTLTKELNLDIIRLGIAYHFEKQIEDMLDHIYR-ADPYFEAHEYNDL 77  
                  : ...\*\* : .. :\*\*\*\*\*: \* :\*\*\*: \* \*\* :\*\* \* :.\* \* : : : \*\*

EGVS (1) HNVALSFRILREHGHNVSSDVFQKFMDNGLKLDYLVNDARGLLSLYEATHFRVHNDKDL 179  
HPS (25) NTSSVQFRLLRQHGYNVSPNIFSRFQDANGKFKESLRSDIRGLLNLYEASHVRTHKEDIL 137  
                  : . :. :\*\* :\*\* :\*\* :\*\* :. : \* \* :\*\* :\* : \* . \* \*\* \* , \*\* \* : \* . \* : \* \*

EGVS (1) EELLSVTTSRLEHLKSHVKYPLEDEISRALKHPLHKELNRLGARYYISIIYEFDSHNKLL 239  
HPS (25) EEALVFSVGHLESAAPHLKSPLSKQVTHALEQSLHKSIPRVEIRYFISIIYEEEEFKNDLL 197  
                  \*\* \* : . . : \*\* . \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \*

EGVS (1) LEFAKLDNFRLQKMYQHELAHLTRWVKDLDFTNKLPFARDRIVEGYFWILGMYFEPERKD 299  
HPS (25) LRFKLDYNLLQMLHKHELSEVSRWVKDLDFVTTLPYARDRAVECYFWTMGVYAEPQYSQ 257  
                  \* . \*\* \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \*

EGVS (1) VREFLNRFALITVVDDTYDVYGTFKELLFTDAIERWGTSDLDQLPGYMRIIYQALMDV 359  
HPS (25) ARVMLAKTIAMISIVDDTFDAYGIVKELEVYTDAIQRWDISQIDRLPEYMKISYKALLDL 317  
                  . \* : \* : . : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \*

EGVS (1) YNQMEEKLSMKADCPTYRLEFAIETVKAMFRSYLEEARWSKEHYIPSMEEYMTVALVSVG 419  
HPS (25) YDDYEKELSK--DGRSDVVHYAKERMKEIVGNFYFIEGKWFIEGYMPSVSEYLSNALATST 375  
                  \* : : \* : \* \* \* : : : \* \* \* : . : \* : \* : \* \* \* : \* : \* : \* : \* : \* : \*

EGVS (1) YKTILTNSFVGMGDIATREVFVFN SPLIIRASDLIARLGDDIGGHEEQKKGDAATAI 479  
HPS (25) YLLTTTSYLGM-KSATKEHFELATNPKILEANATLCRVVDDIATYEVEKGRGQIATGI 434  
                  \* : \* . : \* : \* \* \* : . \* \* : . \* . : . : \* : \* \* . : \* \* : : \* : \* \* \*

EGVS (1) ECIKENHVTKHEAYDEFQKQIDNAWKDLNKEALRPFVPMFITRVVHFTRAIHVIYAD 539  
HPS (25) ECYMRDYGVSTEVAMEKQEMADIAWKDVNEEILRPTFVSSEILTRILNLARIIDVYTKH 494  
                  \*\*\* : : : \* : . . \* : \* : \* : \* \* \* : \* \* \* \* \* . : \* : \* : \* : \* \* \* \*

EGVS (1) FSDGYTRSDKAIRGYITSLLDVPIPL 565  
HPS (25) NQDGYTHPEKVLKPHIIALVDSIDI 520  
                  . \* \* \* : . : \* : \* : \* \* \* : \*

FIGURA 5H



```

V277 (31) MSTQVSASSLAQIPQPKNRPVANFHPNIWGDQFITYTPEDK-VTRACKEEQIEALKEEVR 59
CVS (14) MSSGE-----TFRPTADFHPSLWRNHFLKGASDFKTVDHTATQERHEALKEEVR 49
**: . **.*;***.* :*:.. :.: * * :.:.*: *****

V277 (31) RMILATGRKPIQKRLRIDEVQRLGVAYHFEKEIEDMLDHIYRADPYFEAHEYNDLHTVSL 119
CVS (14) RMITDAEDKPVQKRLRIDEVQRLGVAYHFEKEIEDAIQKLC--PIYIDSNRADLHTVSL 106
*** : **;***** :.: : * : : *****

V277 (31) HFRLLRQQGIKISCDVFEQFKDDEGRFKSSLINDVQGMLSLYEAAVMVRGEHILDEAIA 179
CVS (14) HFRLLRQQGIKISCDVFEKFKDDEGRFKSSLINDVQGMLSLYEAAVMVRGEHILDEAIA 166
*****;*****;*****;*****;*****;*****;*****;*****;*****;*****

V277 (31) FTTHLQS--AAPHLKSPLAEQINHALYRPLRKTLPRLRLEARYIMSVY--QDEAFHNKTL 235
CVS (14) FTTHLKSLVAQDHVTPKLAEQINHALYRPLRKTLPRLRLEARYFMSMINSTSDHLYNKTL 226
*****;* * *;.. *****;**; .: :*****

V277 (31) NFAKLDNFILDLHKEELNELTKWQDLDFTTKLPYARDRLVELYFWDLGTYPESQYAFG 295
CVS (14) NFAKLDNFILLELHKEELNELTKWQDLDFTTKLPYARDRLVELYFWDLGTYPFEPQYAFG 286
*****;*****;*****;*****;*****;*****;*****;*****;*****;*****

V277 (31) RKIMTKLNYILSIIDDTYDAYGTLEECTMFSEAVARWNIEAVDMLPDYMRIIYRTLDDTF 355
CVS (14) RKIMTQLNYILSIIDDTYDAYGTLEELSLFTEAVQRWNIEAVDMLPEYMKLIYRTLDDAF 346
*****;*****;***** :*:** *****;*:*****;*:*****;*:*****

V277 (31) NEIEEDMAKQRRSHCVRYAKEEIQKVI GAYYVQAKWFSEGYVPTIEEYMPIALTSCAYRF 415
CVS (14) NEIEEDMAKQGRSHCVRYAKEENQKVI GAYSVQAKWFSEGYVPTIEEYMPIALTSCAYTF 406
***** ***** ***** ***** ***** ***** ***** ***** *****

V277 (31) VITNSFLGMGDFATKEVFEWISGNPKVVKASVICRLMDDMQGHEFEQKRGHVASAIECY 475
CVS (14) VITNSFLGMGDFATKEVFEWISNPKVVKAAVICRLMDDMQGHEFEQKRGHVASAIECY 466
*****;*****;*****;*****;*****;*****;*****;*****;*****;*****

V277 (31) TKQHGSKEEAIKMFEEVDANAWKDINEELMMKPPVVARPLLGTILNLARAIDFIYKEDD 535
CVS (14) TKQHGSKEEAIKMFEEVDANAWKDINEELMMKPTVVARPLLGTILNLARAIDFIYKEDD 526
*****;*****;*****;*****;*****;*****;*****;*****;*****;*****

V277 (31) GYTHSYLIKEQIASVLGDHVPF 557
CVS (14) GYTHSYLIKQIASVLGDHVPF 548
*****;*****

```

FIGURA 5J

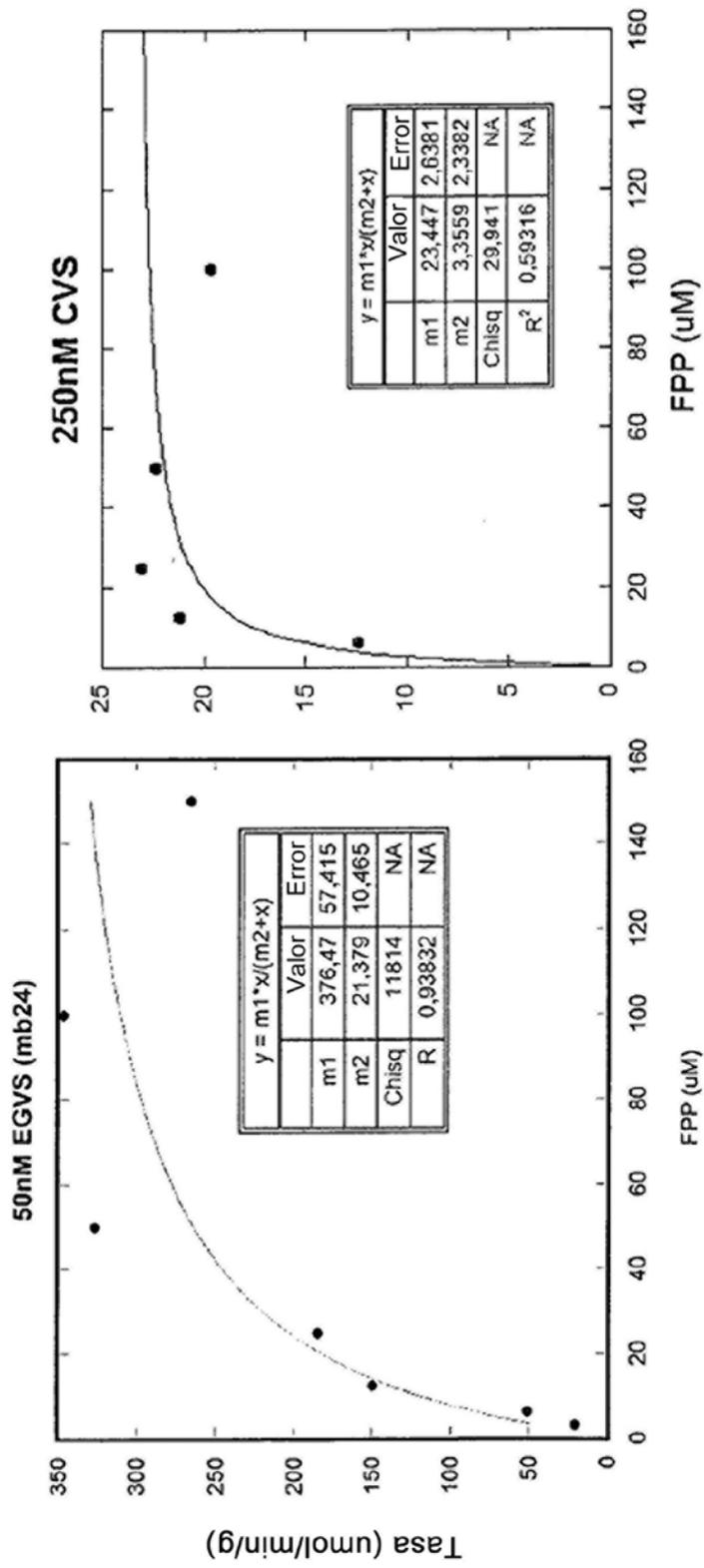


FIGURA 6

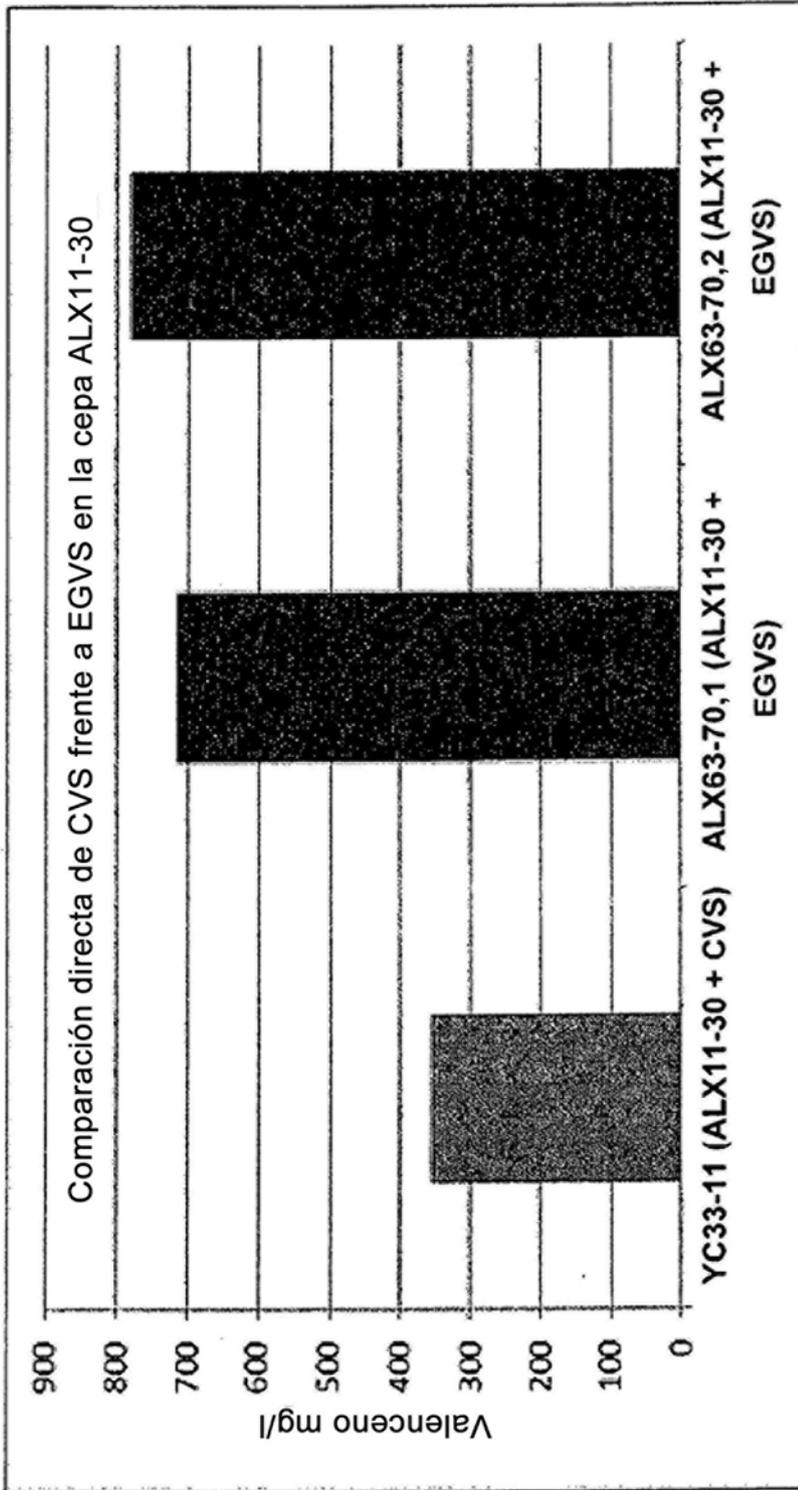


FIGURA 7