(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 556 221**

(51) Int. CI.:

***C12N 9/16*** (2006.01)
***C12P 7/02*** (2006.01)
***C12P 7/18*** (2006.01)
***C12N 15/52*** (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **26.01.2009 E 09706183 (2)**

(97) Fecha y número de publicación de la concesión europea: **14.10.2015 EP 2238256**

(54) Título: **Procedimiento de producción de esclareol**

(30) Prioridad:

**29.01.2008 EP 08101075**
**17.03.2008 EP 08102661**
**20.03.2008 EP 08102811**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
**14.01.2016**

(73) Titular/es:

**FIRMENICH S.A. (100.0%)**
**1, ROUTE DES JEUNES P.O. BOX 239**
**1211 GENEVA 8, CH**

(72) Inventor/es:

**SCHALK, MICHEL**

(74) Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

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

## DESCRIPCIÓN

Procedimiento de producción de esclareol

**Campo técnico**

La presente invención proporciona un procedimiento para producir esclareol, comprendiendo dicho procedimiento poner en contacto un polipéptido particular que tiene actividad de esclareol sintasa con difosfato de labdenodiol (LPP). En particular, dicho procedimiento se puede realizar *in vitro* o *in vivo* para producir esclareol, un compuesto muy útil en los campos de la perfumería y la condimentación. La presente invención también proporciona la secuencia de aminoácidos del polipéptido usado en el procedimiento. También son parte de la presente invención un ácido nucleico derivado de *Salvia sclarea* y que codifica el polipéptido de la invención, un vector de expresión que contiene dicho ácido nucleico, así como un organismo no humano o una célula transformados para albergar el mismo ácido nucleico.

**Técnica anterior**

El esclareol es un miembro de la familia de los terpenoides o terpenos, que comprende un gran número de productos naturales. Los terpenos se encuentran en la mayoría de los organismos (microorganismos, animales y plantas). Estos compuestos están compuestos de cinco unidades de carbono denominadas unidades de isopreno y se clasifican por el número de estas unidades presentes en su estructura. De ese modo, los monoterpenos, sesquiterpenos y diterpenos son terpenos que contienen 10, 15 y 20 átomos de carbono, respectivamente. Los diterpenos, por ejemplo, se encuentran ampliamente en el reino vegetal y se han descrito más de 2500 estructuras de diterpeno (Connolly y Hill, Dictionary of terpenoids, 1991, Chapman & Hall, Londres). Las moléculas de terpeno han sido de interés durante miles de años debido a sus propiedades de sabor y fragancia y sus efectos cosméticos, medicinales y antimicrobianos. Los extractos de plantas obtenidos mediante diferentes medios tales como destilación con vapor o extracción con disolventes se usan como fuente de terpenos. Las moléculas de terpeno se usan a menudo como tales, pero en algunos casos se usan reacciones químicas para transformar los terpenos en otras moléculas de alto valor.

La producción biosintética de terpenos implica enzimas denominadas terpeno sintasa. Estas enzimas convierten un precursor en uno o más productos de terpeno. En la mayoría de los casos, el precursor es un precursor de terpeno acíclico y, en particular, la mayoría de las diterpeno sintasas catalizan la ciclación del precursor acíclico pirofosfato de geranilgeranilo. No obstante, en algunos casos especiales, las terpeno sintasas catalizan la transformación de una molécula cíclica en uno o más productos de terpeno.

En la naturaleza se producen dos tipos de mecanismos de ciclación y están relacionados con dos tipos de diterpeno sintasas que se puede clasificar en la clase I y la clase II de diterpeno sintasas (Wendt y Schulz, 1998, Structure. 6(2):127-33). Para algunos diterpenos, el mecanismo de ciclación es similar al de los monoterpenos y sesquiterpenos ya que se inicia mediante la ionización de la función éster de difosfato de GGPP, seguido de la reacción del carbocatión el resultante con un doble enlace interno. Las diterpeno sintasas que catalizan este tipo de ciclación son diterpeno sintasas de clase I. El segundo modo de ciclación en la biosíntesis de los terpenos, catalizado por las diterpeno sintasas de clase II, se inicia mediante la protonación del doble enlace terminal de GGPP y conduce, después de transposición interna y eliminación de protón, a un compuesto intermedio de difosfato de diterpeno cíclico.

Se han clonado los genes y los ADNc que codifican las diterpeno sintasas de cada una de las dos clases y se han caracterizado las enzimas recombinantes. La disponibilidad de los genes que codifican los diferentes tipos de diterpeno sintasas proporciona información de las estructuras primarias de las enzimas. Algunos motivos de aminoácidos se conservan en las diterpeno sintasas y se relacionan con la ciclación dependiente de protonación o ionización. Se encuentra un motivo DDxxD en varias diterpeno sintasas de clase I. Dicho motivo esta implicado probablemente en la unión y la ionización del resto de difosfato. En las sintasas de clase II, se encuentra un motivo DxDD conservado, en el que el segundo resto de aspartato está implicado como donador de protones.

El esclareol es una molécula de diterpeno de origen natural usada ampliamente como material de partida para la síntesis de moléculas de fragancia con notas de ámbar gris. Estas síntesis se desarrollaron para proporcionar una alternativa al ámbar gris, una sustancia cérea secretada por el intestino del cachalote. El ámbar gris es muy apreciado por su olor agradable y se ha usado históricamente como ingrediente de perfume. Debido a su alto precio y a la demanda creciente del ámbar gris, y particularmente debido a la protección de las especies de ballenas, se han desarrollado síntesis químicas de constituyentes y moléculas de ámbar gris con carácter de ámbar gris. Entre estas moléculas, el sustituto más altamente apreciado para el ámbar gris es Ambrox® (marca registrada de Firmenich SA, Suiza). El material de partida usado con mayor frecuencia para la síntesis de Ambrox® es el diterpeno-diol esclareol.

Generalmente, el precio y la disponibilidad de los extractos naturales de plantas dependen de la abundancia, el rendimiento de aceite y el origen geográfico de las plantas. Además, la disponibilidad y la cantidad de extractos naturales depende en gran medida del clima y otras condiciones locales que conducen a una variabilidad anual, que algunos años hace muy difícil o incluso imposible el uso de tales ingredientes en la perfumería de alta calidad. Por lo

tanto, sería una ventaja proporcionar una fuente de esclareol, que se encuentre menos sometida a fluctuaciones de disponibilidad y calidad. La síntesis química parecería ser una opción evidente a la preparación del esclareol. Sin embargo, dada su estructura altamente compleja, un procedimiento sintético económico para la preparación de esclareol aún es difícil. Por lo tanto, una ruta bioquímica que conduzca a la síntesis del esclareol sería de gran interés.

La biosíntesis de terpenos en plantas y otros organismos se ha estudiado ampliamente y no se detalla adicionalmente en el presente documento, pero se hace referencia a Dewick, Nat. Prod. Rep., 2002, 19, 181-222, que revisa el estado de la técnica de las rutas biosintéticas de terpenos.

En la actualidad se han identificado varias diterpeno sintasas. En particular, el documento de Patente US 7.238.514 desde la cierto número de diterpeno sintasas, los ácidos nucleicos que las codifican, así como organismos unicelulares transformados para expresar cada una de estas sintasas junto con una GGPP sintasa, produciendo de ese modo diterpenos *in vivo*. No obstante, no se desvela específicamente en esa patente ningún procedimiento para la producción biosintética de esclareol que use un polipéptido que tenga una actividad de esclareol sintasa como se proporciona en el presente documento. Las secuencias de aminoácidos y nucleótidos desveladas en él son muy diferentes de las secuencias de la presente invención. Entre las diterpeno sintasas descritas en ese documento, lo más cercano a los polipéptidos de la presente invención son un ARNm de *Cucumis sativus* para una ent-kaureno sintasa designada mediante la SEC ID Nº: 389 en el documento de Patente US 7.238.514 y un ARNm de *Cucurbita maxima* para una ent-kaureno sintasa B designada por la SEC ID Nº: 395 en el documento de Patente US 7.238.514 y por el número de registro AAB39482.1. Estos polipéptidos y el de la invención solo comparten un 32 % de identidad. Además, no se hace ninguna sugerencia en este documento de la técnica anterior de que las diterpeno sintasas descritas sean útiles para la producción de esclareol.

También se han descubierto en las bases de datos de secuencias terpeno sintasas que tienen cierto porcentaje de identidad de secuencia con las secuencias de la presente invención. No obstante, el porcentaje de identidad entre las diterpeno sintasas conocidas y los polipéptidos de la invención es muy bajo. Las sintasas más próximas a las de la invención son una terpenoide ciclasa de función indefinida (número de registro NCBI AAS98912) que tiene un 36 % de identidad con el polipéptido de la invención, una ent-kaureno sintasa de *Cucumis sativus* (número de registro BAB 19275) que tiene un 32 % de identidad con el polipéptido de la invención, una ent-cassadieno sintasa de *Oryza sativa* (número de registro ABH10734 y publicada en Xu, Wilderman, Morrone, Xu, Roy, Margis-Pinheiro, Upadhyaya, Coates y Peters, Functional characterization of the rice kaureno synthase-like gene family, Phytochemistry, 68(3), 2007, 312-326) que tiene un 32 % de identidad con el polipéptido de la invención y una ent-kaureno sintasa de *Oryza sativa* (número de registro AAQ72559 y publicada en Margis-Pinheiro, Zhou, Zhu, Dennis y Upadhyaya, Isolation and characterization of a DS-tagged rice (Oryza sativa L.) GA-responsive dwarf mutant defective in an early step of the gibberellins biosynthesis pathway, Plant Cell Rep., 23(12), 2005, 819-833) que tiene un 32 % de identidad con el polipéptido de la invención. La capacidad potencial de cualquiera de estas secuencias para catalizar la producción de esclareol nunca se menciona en la técnica anterior.

Además de la diferencia entre las propias secuencias, también se ha de indicar que la estructura y las propiedades de ent-kaureno y ent-cassadieno son muy diferentes de las del esclareol. En particular, ent-kaureno es un diterpeno tricíclico que no contiene ningún grupo funcional alcohol, a diferencia del esclareol, que es un diol bicíclico. Además, ent-kaureno, que es un precursor de una hormona de planta reguladora del crecimiento, no tiene ningún uso en el campo de la perfumería ni de la condimentación, mientras que el esclareol tiene un gran interés en estos campos técnicos, como se ha explicado anteriormente.

Un documento de la técnica anterior se refiere específicamente a una esclareol sintasa (Banthorpe, Brown y Morris, Partial purification of farnesyl pyrophosphate: Drimenol cyclase and geranylgeranyl pyrophosphate: Sclareol cyclase, using cell culture as a source of material, Phytochemistry 31, 1992, 3391-3395). En esta referencia, se identifica una proteína parcialmente purificada de *Nicotiana glutinosa* como una esclareol sintasa, pero no se da ninguna indicación con respecto a la secuencia de aminoácidos de esa proteína, la secuencia de nucleótidos del ácido nucleico que la codifica o el uso de esa proteína en un procedimiento para la biosíntesis de esclareol *in vitro* o *in vivo*.

A pesar de los estudios exhaustivos de la ciclación de terpenos, el aislamiento y la caracterización de las enzimas es aún difícil, particularmente en plantas, debido a su baja abundancia, sus patrones de expresión a menudo transitorios, y la complejidad de la purificación de los mismos a partir de las mezclas de resinas y compuestos fenólicos en los tejidos donde se expresan.

Es un objetivo de la presente invención proporcionar procedimientos para fabricar esclareol de forma económica, como se ha indicado anteriormente. Por lo tanto, la presente invención tiene el objetivo de producir diterpenos mientras se producen pocos desperdicios, se usa un procedimiento más eficaz de energía y recursos y mientras se reduce la dependencia de combustibles fósiles. Es un objetivo adicional proporcionar enzimas capaces de sintetizar esclareol, que sean útiles como ingredientes de perfumería y/o aromas.

**Abreviaturas usadas**

| | |
|---|---|
| pb | pares de bases |
| kb | kilo bases |
| BSA | albúmina en suero bovino |
| ADN | ácido desoxirribonucleico |
| ADNc | ADN complementario |
| dT | desoxi timina |
| dNTP | desoxi nucleótido trifosfato |
| DTT | ditiotreitol |
| GC | cromatografía de gases |
| GGPP | pirofosfato de geranilgeranilo |
| IPTG | isopropil-D-tiogalacto-piranósido |
| LB | caldo de lisogenia |
| LPP | difosfato del labdenodiol |
| MOPSO | ácido 3-(N-morfolino)-2-hidroxipropanosulfónico |
| MS | espectrómetro de masas |
| ORF | marco abierto de lectura |
| PCR | reacción en cadena de la polimerasa |
| RMCE | intercambio de casete mediado por recombinasa |
| RT-PCR | reacción en cadena de la polimerasa con transcripción inversa |
| 3'-/5'-RACE | amplificación rápida 3' y 5' de los extremos de ADNc |
| ARN | ácido ribonucleico |
| ARNm | ácido ribonucleico mensajero |
| nt | nucleótido |
| RNasa | ribonucleasa |
| RuBisCO | ribulosa-1,5-bisfosfato carboxilasa |
| SDS-PAGE | electroforesis en gel de SDS-poliacrilamida |
| SsLPPs | difosfato de labdenodiol sintasa de *Salvia sclarea* |
| UTR | región no traducida |

**Descripción de la invención**

La presente invención proporciona un procedimiento para producir biosintéticamente esclareol de forma económica, fiable y reproducible.

Por lo tanto, un objeto de la presente invención es un procedimiento para producir esclareol que comprende

a) poner en contacto difosfato de labdenodiol (LPP) con al menos un polipéptido que tiene una actividad de esclareol sintasa y que comprende una secuencia de aminoácidos al menos un 70 % idéntica con SEC ID Nº: 1; y
b) opcionalmente, aislar el esclareol producido en la etapa a).

El procedimiento se puede realizar *in vitro* así como *in vivo*, como se explicará con mayor detalle posteriormente.

El esclareol y el LPP se definen mediante sus fórmulas que se representan en la Figura 1.

La invención también proporciona un polipéptido aislado que tiene una actividad de esclareol sintasa y que comprende una secuencia de aminoácidos al menos un 70 % idéntica a la secuencia de aminoácidos completa de SEC ID Nº: 1.

La invención también proporciona un ácido nucleico aislado que codifica un polipéptido de la invención.

Como una "esclareol sintasa" o como un "polipéptido que tiene una actividad de esclareol sintasa", los inventores pretenden indicar un polipéptido capaz de catalizar la síntesis de esclareol partiendo de (LPP). La capacidad de un polipéptido para catalizar la síntesis de esclareol se puede confirmar mediante el rendimiento del ensayo enzimático que se detalla en los Ejemplos.

De acuerdo con la presente invención, los polipéptidos también pretenden incluir polipéptidos truncados con la condición de que mantengan su actividad de esclareol sintasa como se define en el presente documento y que comparten al menos el porcentaje definido de identidad con la secuencia de aminoácidos completa de SEC ID Nº: 1.

De acuerdo con una realización preferente, el procedimiento para producir esclareol comprende poner en contacto LPP con un polipéptido que tiene una actividad de esclareol sintasa y que comprende una secuencia de aminoácidos al menos un 75 %, preferentemente al menos un 80 %, preferentemente al menos un 85 %, preferentemente al menos un 90 %, más preferentemente al menos un 95 % e incluso más preferentemente al menos un 98 % idéntica a SEC ID Nº: 1. De acuerdo con una realización más preferente, dicho polipéptido comprende la secuencia de aminoácidos de SEC ID Nº: 1. En una realización incluso más preferente, dicho polipéptido consiste en SEC ID Nº:

1.

También se desvela en el presente documento una esclareol sintasa que es un polipéptido truncado que comprende una secuencia de aminoácidos al menos un 50 % idéntica a SEC ID Nº: 102. Preferentemente, el polipéptido comprende una secuencia de aminoácidos al menos un 55 %, preferentemente al menos un 60 %, preferentemente al menos un 65 %, preferentemente al menos un 70 %, preferentemente al menos un 75 %, preferentemente al menos un 80 %, preferentemente al menos un 85 %, preferentemente al menos un 90 %, más preferentemente al menos un 95 % e incluso más preferentemente al menos un 98 % idéntica a SEC ID Nº: 96. También se desvela en el presente documento un polipéptido que comprende la secuencia de aminoácidos SEC ID Nº: 96. También se desvela en el presente documento un polipéptido que consiste en SEC ID Nº: 96.

El porcentaje de identidad entre dos secuencias peptídicas o nucleotídicas es función del número de restos de aminoácido o ácido nucleico que son idénticos en las dos secuencias cuando se ha generado una alineación de estas dos secuencias. Los residuos idénticos se definen como los residuos que son iguales en las dos secuencias en una posición dada de la alineación. El porcentaje de identidad de secuencia, como se usa en el presente documento, se calcula a partir de la alineación óptima tomando el número de residuos idénticos entre las dos secuencias dividido por el número total de residuos en la secuencia más corta y multiplicando por 100. La alineación óptima es la alineación en la que el porcentaje de identidad es el más alto posible. Se pueden introducir separaciones en una o ambas secuencias en una o más posiciones de la alineación para obtener la alineación óptima. En ese caso, estas separaciones se tienen en cuenta como restos no idénticos para el cálculo del porcentaje de la secuencia de identidad.

La alineación con fines de determinar el porcentaje de identidad de secuencia de aminoácidos o ácidos nucleicos se puede conseguir mediante diversas formas usando programas de ordenador y, por ejemplo, programas de ordenador disponibles públicamente disponibles en la *World Wide Web*. Preferentemente, se puede usar el programa BLAST (Tatiana y col., FEMS Microbiol Lett., 1999, 174:247-250, 1999) ajustado en sus parámetros por defecto, disponible en el Centro Nacional para Información Biotecnológica (NCBI) en http://www.ncbi.nlm.nih.gov/BLAST/b12seq/wblast2.cgi, para obtener una alineación óptima de secuencias peptídicas o nucleotídicas y para calcular el porcentaje de identidad de secuencia.

El polipéptido que se va a poner en contacto con LPP *in vitro* se puede obtener por extracción de cualquier organismo que lo exprese, usando tecnologías convencionales de extracción de proteínas o enzimas. Si el organismo hospedador es un organismo unicelular o una célula que libera el polipéptido de la invención en el medio de cultivo, el polipéptido se puede recoger simplemente del medio de cultivo, por ejemplo por centrifugación, seguido opcionalmente por etapas de lavado y resuspensión en soluciones tampón adecuadas. Si el organismo o célula acumula el polipéptido dentro de sus células, el polipéptido se puede obtener por ruptura o lisis de las células y extracción adicional del polipéptido del lisato celular.

Los polipéptidos, en forma aislada o junto con otras proteínas, por ejemplo, en un extracto de proteína sin procesar obtenido de células o microorganismos cultivados, se pueden suspender a continuación en una solución tampón a un pH óptimo. Si fuera adecuado, se pueden añadir sales, DTT, BSA y otras clases de cofactores enzimáticos, con el fin de optimizar la actividad enzimática. Más adelante se describen condiciones apropiadas con mayor detalle en los Ejemplos.

A continuación se puede añadir LPP a la suspensión o solución, que a continuación se incuba a una temperatura óptima, por ejemplo entre 15 y 40 ºC, preferentemente entre 25 y 35 ºC, más preferentemente a 30 ºC. Después de la incubación, el esclareol producido se puede aislar de la solución incubada mediante procedimientos de aislamiento convencionales, tales como extracción con disolventes y destilación, opcionalmente después de la retirada de polipéptidos de la solución.

Se puede obtener el LPP poniendo en contacto GGPP con una LPP sintasa aislada. Los Ejemplos 1 a 3 posteriores muestran un procedimiento para aislar un ADNc que codifica LPP sintasa de *Salvia sclarea*, un procedimiento para la expresión heteróloga de dicho ADNc en *E. coli*, un procedimiento para la purificación de la LPP sintasa producida de ese modo y un procedimiento para la producción *in vitro* de LPP que usa la LPP sintasa aislada.

De acuerdo con otra realización preferente, el procedimiento para producir esclareol se realiza *in vivo*. En este caso, la etapa a) del procedimiento descrito anteriormente comprende cultivar un organismo o célula no humano capaz de producir LPP y transformado para expresar un polipéptido que tiene una actividad de esclareol sintasa y que comprende una secuencia de aminoácidos al menos un 70 % idéntica a SEC ID Nº: 1 en condiciones que conduzcan a la producción de esclareol.

De acuerdo con una realización más preferente, el procedimiento comprende además, antes de la etapa a), transformar un organismo o célula no humano capaz de producir LPP con al menos un ácido nucleico que codifica un polipéptido que tiene una actividad de esclareol sintasa y que comprende una secuencia de aminoácidos al menos un 70 % idéntica a SEC ID Nº: 1, de modo que dicho organismo exprese dicho polipéptido.

De acuerdo con una realización preferente, el ácido nucleico usado para transformar el organismo o célula hospedador comprende una secuencia de nucleótidos al menos un 75 %, preferentemente al menos un 80 %,

preferentemente al menos un 85 %, preferentemente al menos un 90 %, más preferentemente al menos un 95 % e incluso más preferentemente al menos un 98 % idéntica a SEC ID Nº: 2 o el complemento de la misma. De acuerdo con otra realización preferente, el ácido nucleico comprende la secuencia de nucleótidos SEC ID Nº: 2 o el complemento de la misma. De acuerdo con una realización más preferente, el ácido nucleico consiste en SEC ID Nº: 2 o el complemento de la misma.

En el presente documento también se desvela un ácido nucleico truncado que comprende una secuencia de nucleótidos al menos un 50 %, preferentemente al menos un 55 %, preferentemente al menos un 60 %, preferentemente al menos un 70 %, preferentemente al menos un 75 %, preferentemente al menos un 80 %, preferentemente al menos un 85 %, preferentemente al menos un 90 %, más preferentemente al menos un 95 % e incluso más preferentemente al menos un 98 % idéntica a SEC ID Nº: 93 o el complemento de la misma. En el presente documento también se desvela un ácido nucleico que comprende la secuencia de nucleótidos SEC ID Nº: 93 o el complemento de la misma. En el presente documento también se desvela un ácido nucleico que consiste en SEC ID Nº: 93 o el complemento de la misma.

Estas realizaciones de la invención son particularmente ventajosas dado que es posible realizar el procedimiento *in vivo* sin aislar previamente el polipéptido. La reacción se produce directamente dentro del organismo o célula transformado para expresar dicho polipéptido.

Se pretende que el organismo o célula "exprese" un polipéptido, con la condición de que el organismo o célula se transforme para albergar el ácido nucleico que codifica dicho polipéptido, el ácido nucleico se transcriba a ARNm y el polipéptido se encuentre en el organismo o célula hospedador. El término "expresar" incluye "expresar heterólogamente" y "sobreexpresar", refiriéndose el último a los niveles de ARNm, polipéptido y/o actividad enzimática sobre y por encima de lo que se mide en un organismo o célula no transformado. Una descripción más detallada de los procedimientos adecuados para transformar un organismo o célula no humano se describirá posteriormente en la parte de la memoria descriptiva que se dedica a tales organismos o células no humanos transformados como objetos específicos de la presente invención y en los Ejemplos.

Se pretende que el organismo o célula particular sea "capaz de producir LPP" cuando produce LPP de forma natural o cuando no produce LPP de forma natural pero produce GGPP (o se transforma de ese modo) y se transforme para expresar una LPP sintasa, antes de la transformación con un ácido nucleico que codifica una esclareol sintasa o junto con dicho ácido nucleico. Los organismos o células transformados para producir una mayor cantidad de LPP que la del organismo o célula de origen natural también se incluyen en los "organismos o células capaces de producir LPP". De acuerdo con una realización preferente, el organismo acumula LPP de forma natural o se transforma para acumular LPP.

Los procedimientos para transformar organismos de modo que expresen una LPP sintasa, pueden ser cualquier procedimiento conocido en la técnica para transformar un organismo hospedador. Tales procedimientos se exponen con mayor detalle posteriormente y se da un ejemplo específico de la expresión de una LPP sintasa en *E. coli* en el Ejemplo 2. También se conocen en la técnica procedimientos para transformar un organismo para producir GGPP. Tales procedimientos se pueden encontrar, por ejemplo, en Huang, Roessner, Croteau y Scott, Engineering Escherichia coli for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol, Bioorg Med Chem., 9(9), 2001, 2237-2242.

Para realizar la invención *in vivo*, el organismo hospedador se cultiva en condiciones que conduzcan a la producción de esclareol. Por lo tanto, si el hospedador es una planta transgénica, se proporcionan condiciones de crecimiento óptimo, tales como condiciones de luz, agua y nutrientes óptimas, por ejemplo. Si el hospedador es un organismo unicelular, las condiciones que conduzcan a la producción de esclareol pueden comprender la adición de factores adecuados al medio de cultivo del hospedador. Además, se puede seleccionar el medio de cultivo de modo que maximice la síntesis de esclareol. Se describen condiciones de cultivo óptimas de una forma más detallada en los siguientes Ejemplos.

Algunos organismos no humanos adecuados para realizar el procedimiento de la invención *in vivo* pueden ser cualquier organismo multicelular o unicelular no humano. En una realización preferente, el organismo no humano usado para realizar la invención *in vivo* es una planta, un procariota o un hongo.

Se puede usar cualquier planta, procariota u hongo para realizar el procedimiento de la invención *in vivo*. Algunas plantas particularmente útiles son las que producen de forma natural grandes cantidades de terpenos. En una realización más preferente, la planta se selecciona entre la familia de *Solanaceae*, *Poaceae*, *Brassicaceae*, *Fabaceae*, *Malvaceae*, *Asteraceae* o *Lamiaceae*. Por ejemplo, la planta se selecciona entre los géneros *Nicotiana*, *Solanum*, *Sorghum*, *Arabidopsis*, *Brassica* (colza), *Medicago* (alfalfa), *Gossypium* (algodón), *Artemisia*, *Salvia* y *Mentha*. Preferentemente, la planta pertenece a la especie de *Nicotiana tabacum*.

En una realización más preferente, el organismo no humano es un microorganismo. De acuerdo con una realización incluso más preferente, dicho microorganismo es una bacteria o un hongo, y preferentemente dicho hongo es levadura. Más preferentemente, dicha bacteria es *E. coli* y dicha levadura es *Saccharomyces cerevisiae*.

La mayoría de estos organismos no producen LPP de forma natural. Para que sean adecuados para realizar el procedimiento de la invención, estos organismos se tienen que transformar para producir dicho precursor. Se pueden transformar de ese modo antes de la modificación con el ácido nucleico que codifica el polipéptido que tiene una actividad de esclareol sintasa o simultáneamente, como se ha explicado anteriormente.

También se pueden usar células eucariotas superiores aisladas, en lugar de organismos completos, como hospedadores para realizar el procedimiento de la invención *in vivo*. Algunas células eucariotas adecuadas pueden ser cualquier célula no humana, pero son preferentemente células de plantas.

De acuerdo con otra realización preferente, el polipéptido o el ácido nucleico usado en el procedimiento de cualquiera de las realizaciones anteriores deriva de *Salvia sclarea*.

Una herramienta importante para realizar el procedimiento de la invención es el propio polipéptido. Por lo tanto, es otro objeto de la presente invención un polipéptido aislado que tiene una actividad de esclareol sintasa y que comprende una secuencia de aminoácidos al menos un 70 % idéntica a la secuencia de aminoácidos completa de SEC ID Nº: 1.

De acuerdo con una realización preferente, la esclareol sintasa comprende una secuencia de aminoácidos al menos un 75 %, preferentemente al menos un 80 %, preferentemente al menos un 85 %, preferentemente al menos un 90 %, más preferentemente al menos un 95 % e incluso más preferentemente al menos un 98 % idéntica a SEC ID Nº: 1. De acuerdo con otra realización preferente, el polipéptido comprende la secuencia de aminoácidos SEC ID Nº: 1. De acuerdo con una realización más preferente, el polipéptido consiste en SEC ID Nº: 1.

De acuerdo con otra realización preferente de la invención, el polipéptido deriva de *Salvia sclarea*.

Como se usa en el presente documento, las expresiones "esclareol sintasa" o "polipéptido que tiene una actividad de esclareol sintasa" se refieren a un género de polipéptidos o fragmentos peptídicos que incluyen las secuencias de aminoácidos identificadas en el presente documento, así como polipéptidos truncados o variantes, con la condición de que mantengan su actividad de esclareol sintasa como se ha definido anteriormente y que compartan al menos el porcentaje definido de identidad con el correspondiente fragmento de SEC ID Nº: 1.

En el presente documento también se desvela una esclareol sintasa que comprende una secuencia de aminoácidos al menos un 50 % idéntica a SEC ID Nº: 96. Preferentemente, la esclareol sintasa comprende una secuencia de aminoácidos al menos un 55 %, preferentemente al menos un 60 %, preferentemente al menos un 65 %, preferentemente al menos un 70 %, preferentemente al menos un 75 %, preferentemente al menos un 80 %, preferentemente al menos un 85 %, preferentemente al menos un 90 %, más preferentemente al menos un 95 % e incluso más preferentemente al menos un 98 % idéntica a SEC ID Nº: 96. En el presente documento también se desvela un polipéptido que comprende la secuencia de aminoácidos SEC ID Nº: 96. En el presente documento también se desvela un polipéptido que consiste en SEC ID Nº: 96.

Algunos ejemplos de polipéptidos variantes son proteínas de origen natural que resultan de sucesos alternados de corte y empalme de ARNm o de la escisión proteolítica de los polipéptidos descritos en el presente documento. Algunas variaciones atribuibles a la proteolisis incluyen, por ejemplo, diferencias en los extremos N- o C-terminal después de la expresión en diferentes tipos de células hospedadoras, debido a la retirada proteolítica de uno o más aminoácidos terminales de los polipéptidos de la invención. Los polipéptidos codificados por un ácido nucleico obtenido por mutación de un ácido nucleico de la invención, como se describe posteriormente, también se incluyen en la invención.

El ácido nucleico que codifica el polipéptido que tiene una actividad de esclareol sintasa, como se ha definido anteriormente, es una herramienta necesaria para modificar organismos o células no humanos destinados a usarse cuando el procedimiento se realiza *in vivo*. Por lo tanto, es otro objeto de la invención un ácido nucleico que codifica un polipéptido como se define en cualquiera de las realizaciones anteriores.

De acuerdo con una realización preferente, el ácido nucleico comprende una secuencia de nucleótidos al menos un 75 %, preferentemente al menos un 80 %, más preferentemente al menos un 85 %, más preferentemente al menos un 90 %, más preferentemente al menos un 95 % e incluso más preferentemente al menos un 98 % idéntica a SEC ID Nº: 2 o el complemento de la misma. De acuerdo con una realización más preferente, el ácido nucleico comprende una secuencia de nucleótidos idéntica a SEC ID Nº: 2 o el complemento de la misma. De acuerdo con una realización incluso más preferente, el ácido nucleico consiste en SEC ID Nº: 2 o el complemento de la misma.

De acuerdo con otra realización preferente de la invención, el ácido nucleico deriva de *Salvia sclarea*.

Se puede definir que el ácido nucleico de la invención incluye polímeros de desoxirribonucleótido o ribonucleótido en forma mono o bicatenaria (ADN y/o ARN). Se debería entender que la expresión " secuencia de nucleótidos" comprende una molécula de polinucleótido o una molécula de oligonucleótido en forma de un fragmento separado o como un componente de un ácido nucleico mayor. Los ácidos nucleicos de la invención también incluyen ciertas secuencias de nucleótidos aisladas que incluyen las que están básicamente exentas de material endógeno contaminante. El ácido nucleico de la invención puede estar truncado, con la condición de que codifique un

polipéptido incluido por la presente invención, como se ha descrito anteriormente. Algunos ácidos nucleicos truncados particularmente útiles son los ácidos nucleicos al menos un 70 % idénticos a SEC ID Nº: 93 o el complemento de la misma.

Los ácidos nucleicos obtenidos por mutaciones de SEC ID Nº: 2 o el complemento de la misma también están incluidos en la invención, con la condición de que compartan al menos el porcentaje de identidad definido con el correspondiente fragmento de SEC ID Nº: 2 y que codifiquen polipéptidos que tengan una actividad de esclareol sintasa, como se ha definido anteriormente. Las mutaciones pueden ser cualquier clase de mutación de estos ácidos nucleicos, tales como mutaciones puntuales, mutaciones por supresión, mutaciones por inserción y/o mutaciones de desplazamiento de marco. Se pueden preparar ácidos nucleicos variantes con el fin de adaptar su secuencia de nucleótidos a un sistema de expresión específico. Por ejemplo, se conoce que los sistemas de expresión bacterianos expresan de forma más eficaz los polipéptidos si los aminoácidos se codifican por un codón preferente. Debido a la degeneración del código genético, en el que más de un codón puede codificar el mismo aminoácido, múltiples secuencias de ADN pueden codificar el mismo polipéptido, estando incluidas en la invención todas estas secuencias de ADN.

En el presente documento también se desvela un ácido nucleico que es un ácido nucleico truncado que comprende una secuencia de nucleótidos al menos un 50 %, preferentemente al menos un 55 %, preferentemente al menos un 60 %, preferentemente al menos un 70 %, preferentemente al menos un 75 %, preferentemente al menos un 80 %, preferentemente al menos un 85 %, preferentemente al menos un 90 %, más preferentemente al menos un 95 % e incluso más preferentemente al menos un 98 % idéntica a SEC ID Nº: 93 o el complemento de la misma. En el presente documento también se desvela un ácido nucleico que comprende la secuencia de nucleótidos SEC ID Nº: 93 o el complemento de la misma. En el presente documento también se desvela un ácido nucleico que consiste en SEC ID Nº: 93 o el complemento de la misma.

Hablando en términos generales, el ácido nucleico de la invención se puede aislar usando un enfoque de secuenciación masivo en paralelo, que se desarrolla ampliamente en los Ejemplos 5 y 6. La primera etapa de este procedimiento es la secuenciación global de la librería de ADNc. En primer lugar, la librería de ADNc se fragmenta por nebulización. A continuación, los fragmentos se amplifican mediante PCR y se realiza la reacción de secuenciación. Se obtienen secuencias cortas de 35 pases denominadas "lecturas". Las "lecturas" se reensamblan en secuencias contiguas ("*contigs*") usando un software con ajustes mínimos definidos de longitud y porcentaje de solapamiento. A continuación, se realiza una búsqueda en las "lecturas" y "*contigs*" para identidad de secuencia con enzimas conocidas del mismo tipo. En base a estas homologías, las "lecturas" y "*contigs*" se seleccionan y se usan para sintetizar cebadores con el fin de realizar la amplificación por PCR de la esclareol sintasa de longitud completa.

Otra herramienta importante para transformar organismos o células hospedadores adecuados para realizar el procedimiento de la invención *in vivo* es un vector de expresión que comprende al menos un ácido nucleico de acuerdo con cualquier realización de la invención. Por lo tanto, tal vector es también un objeto de la presente invención.

Un "vector de expresión", como se usa en el presente documento, incluye cualquier vector recombinante lineal o circular que incluye, pero no se limita a, vectores virales, bacteriófagos y plásmidos. El experto en la materia es capaz de seleccionar un vector adecuado de acuerdo con el sistema de expresión. En una realización, los vectores de expresión incluyen los ácidos nucleicos de la invención unidos operativamente al menos a una secuencia reguladora, que controla la transcripción, traducción, iniciación y terminación, tal como un promotor, operador o potenciador transcripcional, o un sitio de unión ribosomal de ARNm y, opcionalmente, que incluye al menos un marcador de selección. Las secuencias de nucleótidos se "unen operativamente" cuando la secuencia reguladora se relaciona funcionalmente con el ácido nucleico de la invención.

Los vectores de expresión de la presente invención se pueden usar en los procedimientos para preparar un organismo y/o célula hospedador genéticamente transformado, en organismos y/o células hospedadores que albergan los ácidos nucleicos de la invención y en los procedimientos para producir o fabricar polipéptidos que tienen una actividad de esclareol sintasa, como se desvela adicionalmente más adelante.

Los organismos y células no humanos recombinantes transformados para albergar al menos un ácido nucleico de la invención, de modo que expresen heterólogamente o sobreexpresen al menos un polipéptido de la invención, también son herramientas muy útiles para realizar el procedimiento de la invención. Por lo tanto, tales organismos y células no humanos son otro objeto de la presente invención.

Algunos organismos no humanos de la invención pueden ser cualquier organismo multicelular o unicelular no humano. En una realización preferente, el organismo no humano de la invención es una planta, un procariota o un hongo. Dicho organismo puede ser cualquier planta, procariota u hongo. Algunas plantas particularmente útiles son las que producen de forma natural grandes cantidades de terpenos. En una realización más preferente, la planta se selecciona entre la familia de *Solanaceae*, *Poaceae*, *Brassicaceae*, *Fabaceae*, *Malvaceae*, *Asteraceae* o *Lamiaceae*. Por ejemplo, la planta se selecciona entre los géneros *Nicotiana*, *Solanum*, *Sorghum*, *Arabidopsis*, *Brassica* (colza), *Medicago* (alfalfa), *Gossypium* (algodón), *Artemisia*, *Salvia* y *Mentha*. Preferentemente, la planta pertenece a la especie de *Nicotiana tabacum*.

En una realización más preferente, el organismo no humano es un microorganismo. De acuerdo con una realización incluso más preferente, dicho microorganismo es una bacteria o levadura y lo más preferentemente, dicha bacteria es *E. coli* y dicha levadura es *Saccharomyces cerevisiae*.

También se pueden transformar células eucariotas superiores aisladas, en lugar de organismos completos. Aquí, por células eucariotas superiores, se pretende indicar cualquier célula eucariota no humana excepto células de levadura. Algunas células eucariotas superiores preferentes son células de plantas o células fúngicas.

El término "transformado" se refiere al hecho de que el hospedador se somete a ingeniería genética para comprender una, dos o más copias de cualquiera de los ácidos nucleicos de la invención. Preferentemente, el término "transformado" se refiere a hospedadores que expresan heterólogamente los polipéptidos de la invención, así como sobreexpresan los mismos. Por lo tanto, en una realización, la presente invención proporciona un organismo transformado, en el que el polipéptido de la invención se expresa en una cantidad mayor que en el mismo organismo no transformado de ese modo.

Existen varios procedimientos conocidos en la técnica para la creación de organismos o células hospedadores transgénicos tales como plantas, hongos, procariotas, o cultivos celulares de organismos eucariotas superiores. Se describen vectores de clonación y expresión apropiados para su uso con hospedadores bacterianos, fúngicos, de levaduras, de plantas y de mamíferos, por ejemplo, en Pouwels y col., Cloning Vectors: A Laboratory Manual, 1985, Elsevier, Nueva York, y en Sambrook y col., Molecular Cloning: A Laboratory Manual, 2ª edición, 1989, Cold Spring Harbor Laboratory Press. Los vectores de clonación y expresión para plantas superiores y/o, en particular, células de plantas están disponibles para los expertos en la materia. Véase, por ejemplo, Schardl y col., Gene 61: 1-11,1987.

Los procedimientos para transformar organismos o células hospedadores para albergar ácidos nucleicos transgénicos, tales como los de la presente invención, son familiares para los expertos en la materia. Para la creación de plantas transgénicas, por ejemplo, los procedimientos actuales incluyen: electroporación de protoplastos de plantas, transformación mediada por liposomas, transformación mediada por *Agrobacterium*, transformación mediada por polietilenglicol, bombardeo de partículas, microinyección de células de plantas, y transformación usando virus.

En una realización, el ADN transformado se integra en un cromosoma de un organismo y/o célula hospedador no humano de modo que dé como resultado un sistema recombinante estable. Se puede usar cualquier procedimiento de integración cromosómica conocido en la técnica en la práctica de la invención, que incluye, pero no se limita a, intercambio de casete mediado por recombinasa (RMCE), inserción cromosómica específica de sitio viral, adenovirus, e inyección pronuclear.

Con el fin de realizar el procedimiento para producir esclareol in vitro, como se ha expuesto anteriormente en el presente documento, es muy ventajoso proporcionar un procedimiento para fabricar al menos un polipéptido que tenga una actividad de esclareol sintasa. Por lo tanto, la invención proporciona un procedimiento para producir al menos un polipéptido que tiene una actividad de esclareol sintasa que comprende

a) cultivar un organismo o célula no humano transformado con el vector de expresión de la invención, de modo que albergue un ácido nucleico de acuerdo con la invención y exprese o sobreexprese un polipéptido codificado por dicho ácido nucleico y que tenga una actividad de esclareol sintasa;
b) aislar el polipéptido que tiene una actividad de esclareol sintasa a partir del organismo o célula no humano cultivado en la etapa a).

De acuerdo con una realización preferente, dicho procedimiento comprende además, además de la etapa a), transformar un organismo o célula hospedador no humano con al menos un vector de expresión de la invención, de modo que albergue al menos un ácido nucleico de acuerdo con la invención y exprese o sobreexprese menos un polipéptido codificado por dicho ácido nucleico.

La transformación y el cultivo del organismo o célula no humano se puede realizar como se ha descrito anteriormente para el procedimiento para la fabricación del esclareol *in vivo*. La etapa b) se puede realizar usando cualquier técnica bien conocida en la técnica para aislar un polipéptido particular de un organismo o célula.

Una "variante de polipéptido", como se denomina en el presente documento, significa un polipéptido que tiene una actividad de esclareol sintasa y es básicamente homólogo a un polipéptido nativo, pero que tiene una secuencia de aminoácidos diferente de la codificada por cualquiera de las secuencias de ácidos nucleicos de la invención debido a una o más supresiones, inserciones o sustituciones.

Las variantes pueden comprender secuencias sustituidas conservativamente, que significa que un resto de aminoácido dado se reemplaza por un resto que tiene características fisicoquímicas similares. Algunos ejemplos de sustituciones conservativas incluyen la sustitución de un resto alifático por otro, tal como Ile, Val, Leu, o Ala por otro, o sustituciones de un resto polar por otro, tal como entre Lys y Arg; Glu y Asp; o Gln y Asn. Véase Zubay, Biochemistry, Addison-Wesley Pub. Co., (1983). Los efectos de tales sustituciones se pueden calcular usando matrices de calificación de sustitución tales como PAM-120, PAM-200, y PAM-250 como se discute en Altschul, (J. Mol. Biol. 219:555-65, 1991). Se conocen bien otras de tales sustituciones conservativas, por ejemplo sustituciones

de regiones enteras que tienen características de hidrofobicidad similares.

También se incluyen en la invención variantes de péptidos de origen natural. Algunos ejemplos de tales variantes son proteínas que resultan de sucesos de corte y empalme de ARNm alternantes o de la escisión proteolítica de los polipéptidos descritos en el presente documento. Las variaciones atribuibles a la proteolisis incluyen, por ejemplo, diferencias en los extremos N-terminal o C-terminal tras la expresión de diferentes tipos de células hospedadoras, debido a la retirada proteolítica de uno o más aminoácidos terminales de los polipéptidos codificados por las secuencias de la invención.

Se pueden usar variantes de los polipéptidos de la invención para obtener una mejora o una reducción deseada de la actividad enzimática, una modificación de la regioquímica o estereoquímica modificada, o una alteración de la utilización de sustrato o la distribución del producto. Además, las variantes se pueden preparar para que tengan al menos una propiedad modificada, por ejemplo un aumento de afinidad por el sustrato, una mejora de la especificidad para la producción de uno o más compuestos deseados; una diferente distribución de producto, una diferente actividad enzimática, un aumento de la velocidad de la reacción de la enzima, una mayor actividad o estabilidad en un entorno específico (pH, temperatura, disolvente, etc.), o una mejora del nivel de expresión en un sistema de expresión deseado. Una variante o un mutante dirigido a sitio se puede preparar mediante cualquier procedimiento conocido en la técnica. Como se ha indicado anteriormente, la invención proporciona polipéptidos recombinantes y no recombinantes, aislados y purificados, tales como de *Salvia sclarea*. Las variantes y los derivados de polipéptidos nativos se pueden obtener por aislamiento de variantes de origen natural, o la secuencia de nucleótidos de las variantes, de líneas o especies de plantas diferentes o iguales, o mediante mutaciones de programación artificial de secuencias de nucleótidos que codifican las esclareol sintasas nativas. Las alteraciones de la secuencia de aminoácidos nativa se pueden conseguir mediante cualquiera de cierto número de procedimientos convencionales.

Las variantes de polipéptidos que resultan de una fusión de secuencias de péptidos adicionales en los extremos amino y carboxilo terminales de los polipéptidos de la invención se pueden usar para mejorar la expresión de los polipéptidos, y pueden ser útiles en la purificación de la proteína o la mejora de la actividad enzimática del polipéptido en un entorno o sistema de expresión deseado. Tales secuencias de péptidos adicionales pueden ser péptidos de señal, por ejemplo. Por lo tanto, la presente invención incluye variantes de los polipéptidos de la invención, tales como las obtenidas por fusión con otros oligo o polipéptidos y/o las que se unen a péptidos de señal.

Por lo tanto, en una realización, la presente invención proporciona un procedimiento para preparar un polipéptido variante que tiene una actividad de esclareol sintasa y que comprende las etapas de:

(a) seleccionar un ácido nucleico de acuerdo con cualquiera de las realizaciones expuestas anteriormente;
(b) modificar el ácido nucleico seleccionado para obtener al menos un ácido nucleico mutante;
(c) transformar células u organismos unicelulares hospedadores con la secuencia de ácido nucleico mutante para expresar un polipéptido codificado por la secuencia de ácido nucleico mutante;
(d) analizar sistemáticamente el polipéptido para al menos una propiedad modificada; y,
(e) opcionalmente, si el polipéptido no tiene ninguna actividad de esclareol sintasa variante deseada, repetir las etapas de procedimiento (a) a (d) hasta que se obtenga un polipéptido con una actividad de esclareol sintasa variante deseada;
(f) opcionalmente, si se identificó un polipéptido que tiene una actividad de esclareol sintasa variante deseada en la etapa d), aislar el correspondiente ácido nucleico mutante obtenido en la etapa (c).

En la etapa (b), se pueden crear un gran número de secuencias de ácido nucleico mutantes, por ejemplo mediante mutagénesis aleatoria, mutagénesis específica de sitio, o barajado de ADN. Los procedimientos detallados del barajado génico se encuentran en Stemmer, DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. Proc Natl Acad Sci U S A., 1994, 91(22): 10747-1075. En resumen, el barajado de ADN se refiere a un procedimiento de recombinación aleatoria de secuencias conocidas *in vitro*, que implica al menos dos ácidos nucleicos seleccionados para recombinación. Por ejemplo, se pueden introducir mutaciones en sitios particulares mediante síntesis de oligonucleótidos que contienen una secuencia mutante, flanqueados por sitios de restricción que permiten la ligadura a fragmentos de la secuencia nativa. Después de la ligadura, la secuencia reconstruida resultante codifica un análogo que tiene la inserción, sustitución, o supresión de aminoácido deseada. Alternativamente, se pueden emplear procedimientos de mutagénesis específicos de sitio dirigidos por oligonucleótidos para proporcionar un gen alterado en el que se pueden alterar codones predeterminados por sustitución, supresión o inserción.

Por lo tanto, SEC ID Nº: 2 o 93 se pueden combinar con cualquier otro aminoácido que codifique diterpeno sintasa, por ejemplo aislado de un organismo distinto de *Salvia sclarea*. De ese modo, se pueden obtener y separar ácidos nucleicos mutantes, que se pueden usar para transformar células hospedadoras de acuerdo con procedimientos convencionales, por ejemplo tales como los desvelados en los presentes Ejemplos.

En la etapa (d), el polipéptido obtenido en la etapa (c) se analiza sistemáticamente para al menos una propiedad modificada, por ejemplo una actividad enzimática modificada deseada. Algunos ejemplos de actividades enzimáticas deseadas, para las que se puede analizar sistemáticamente un polipéptido expresado, incluyen aumento o reducción de la actividad enzimática, según se mide mediante el valor de $K_M$ o $V_{max}$, modificación de la regioquímica o

estereoquímica y alteración del uso de sustrato o la distribución de producto. El análisis sistemático de la actividad enzimática se puede realizar de acuerdo con procedimientos familiares para los expertos en la materia y los desvelados en los presentes Ejemplos.

La etapa (e) proporciona la repetición de las etapas de procedimiento (a)-(d), que se pueden realizar preferentemente en paralelo. Por lo tanto, mediante la creación de un número considerable de ácidos nucleicos mutantes, se pueden transformar numerosas células hospedadoras con diferentes ácidos nucleicos mutantes al mismo tiempo, permitiendo el análisis sistemático posterior de un elevado número de polipéptidos. De ese modo, se pueden aumentar las oportunidades de obtener un polipéptido variante deseado a discreción del experto en la materia.

**Descripción de las figuras**

Figura 1: estructuras de los diversos compuestos citados en la memoria descriptiva.
Figura 2: biosíntesis putativa de esclareol a partir de LPP, que está catalizada por SsTps1132 (SEC ID Nº: 1).
Figura 3: alineación de secuencias de aminoácidos de fragmentos de tipo diterpeno sintasa de clase I con la secuencia del estemodeno sintasa de *Oriza sativa* (número de registro AAZ76733).
Figura 4: alineación de la secuencia de aminoácidos deducida de SsTps1132 (SEC ID Nº: 1) y SsTps1137 (SEC ID Nº: 86) con secuencias de aminoácidos de diterpeno sintasas seleccionadas de la base de datos.
Figura 5: alineación de secuencias de aminoácidos de SsTps1132 (SEC ID Nº: 1) y 1132-2-5 (SEC ID Nº: 96) que se expresaron heterólogamente en *E coli*.
Figura 6: análisis por GC de los productos obtenidos después de la incubación de 1132 proteínas recombinantes diferentes con LPP. Los extractos de proteína sin procesar de *E. coli* que expresan las proteínas recombinantes SsTps1132 y 1132-2-5 (SEC ID Nº: 1 y 96) se incubaron con LPP en un volumen final de 1 ml de MOPSO 50 mM a pH 7 complementado con $MgCl_2$ 15 mM y DTT 1 mM.
Figura 7: análisis por GC-MS de los productos generados a partir de LPP mediante la proteína recombinante 1132-2-5. (A) Cromatograma iónico total de los productos obtenidos a partir de la incubación de LPP con un extracto de proteínas sin procesar de *E. coli* transformado con pET101-1132-2-5 (SEC ID Nº: 93). (B) Espectro de masas del pico para un tiempo de retención de 14,3. (C) Espectro de masas de un patrón de esclareol auténtico.
Figura 8: análisis por GC de los productos obtenidos después de incubación conjunta de las proteínas recombinantes SsTps1132 y 1132-2-5 (SEC ID Nº: 1 y 96) con la proteína recombinante SsLPPs3 (SEC ID Nº: 24) en presencia de GGPP.

**Realizaciones específicas de la invención o Ejemplos**

La invención se describirá a continuación con mayor detalle por medio de los siguientes Ejemplos.

**Ejemplo 1**

Aislamiento de los ADNc que codifican LPP sintasa de *Salvia sclarea* mediante un enfoque de PCR

***A. Material de planta y extracción de ADN***

Se recogieron las yemas florales (1,5 a 2 cm de longitud, 1-2 días de edad) que desarrolla la *Salvia sclarea* en los campos de Bassins (Suiza) y se congelaron directamente en nitrógeno líquido.

Se extrajo ARN total usando el Reactivo de ARN de Plantas Concert™ de Invitrogen (Carlsbad, CA) y el ARNm se purificó mediante cromatografía de afinidad oligodT-celulosa usando el Kit de aislamiento de ARNm FastTrack® 2.0 (Invitrogen, Carlsbad, CA) de acuerdo con las instrucciones del fabricante. Se construyó una librería de ADNc usando el Kit de Amplificación de ADNc Marathon™ (Clontech, Mountain View, CA).

***B. Reacciones en cadena de la polimerasa para la amplificación de los ADNc de diterpeno sintasas***

Se realizó la PCR usando el cebador directo PCR DT3F (5'-GAYRTNGAYGAYACNGCNATGG-3' (SEC ID Nº: 3)) y el cebador inverso DT4R (5'-GTYTTNCCNAKCCANACRTCRYYT-3' (SEC ID Nº: 4)). La mezcla de PCR contenía 0,4 µM de cada cebador, 300 µM de cada dNTP, 5 µl de tampón de ADN polimerasa 10X HotStartTaq® (Qiagen), 2 µl de ADNc diluido 100 veces, y 0,5 µl de ADN polimerasa HotStartTaq® en un volumen final de 50 µl. Las condiciones de ciclación fueron: 35 ciclos de 45 s a 94 ºC, 45 s a 50 ºC y 2 min a 72 ºC; y 10 min a 72 ºC. Los tamaños de los productos de PCR se evaluaron en un gel de agarosa al 1 %. Las bandas correspondientes al tamaño esperado se extrajeron del gel, se purificaron usando el Kit de Extracción de Gel QIAquick® (Qiagen) y se clonaron en el vector pCR®2.1-TOPO usando el Kit de clonación TOPO TA (Invitrogen, Carlsbad, CA). Los fragmentos de ADNc insertados se sometieron a continuación a secuenciación de ADN y la secuencia se comparó frente a la base de datos de proteínas no redundante de GenBank (NCBI) usando el algoritmo BLASTX (Altschul y col., J. Mol. Biol. 215, 403-410, 1990). Se obtuvo una secuencia de 354 pb denominada FN23 (SEC ID Nº: 5). Este fragmento de ADN poseía el tamaño esperado y mostró homología de secuencia con diterpeno sintasas.

***C. Aislamiento de ADNc de longitud completa mediante amplificación rápida de los extremos del ADNc (RACE)***

Se diseñaron los oligonucleótidos específicos para la secuencia FN23 (SEC ID Nº: 5): FN23-F1 (3'-GCACGGATAC-GACGTCGATCCAAATGTAC-5' (SEC ID Nº: 6)), FN23-F2 (3'-GGGCTGCTCAACTAAGATTTCCAGGAG-5' (SEC ID Nº: 7)) y FN23-F3 (5'-GGGTGATATCCGACCACTTATTTGATGAG-5' (SEC ID Nº: 8)). Estos cebadores se usaron en RT-PCR en combinación con cebadores de oligodT prolongados con una secuencia adaptadora (5'-AATTCGGTACCCGGGATCC$(T)_{17}$-3') (SEC ID Nº: 9). La composición de la mezcla de reacción de RT-PCR fue la siguiente: 10 µl de tampón 5X Qiagen OneStep RT-PCR, 400 µM de cada dNTP, 400 nM de cada cebador, 2 µl de Mezcla Enzimática Qiagen OneStep RT-PCR, 1 µl de Inhibidor de Ribonucleasa RNasin® (Promega Co., Madisson, WI) y 1250 ng de ARN total en un volumen final de 50 µl. Las condiciones del ciclado térmico fueron: 30 min a 50 ºC (transcripción inversa); 15 min a 95 ºC (activación de ADN polimerasa); 35 ciclos de 45 s a 94 ºC, 45 s a 50 ºC y 90 s a 72 ºC; y 10 min a 72 ºC. Se realizó una segunda ronda de PCR usando los productos de RT-PCR como plantilla con el cebador adapterP (5'-AATTCGGTACCCGGGATCC-3' (SEC ID Nº: 10)) en combinación con cebadores específicos de FN23 iguales o anidados. Este enfoque de PCR proporcionó un fragmento de ADNc de 1271 pb (FN30 (SEC ID Nº: 11)) que tenía una superposición perfecta de 192 pb con el fragmento FN23 (SEC ID Nº: 5) y que contenía el extremo 3' incluyendo el codón de terminación y la secuencia no codificante 3' del ADNc correspondiente.

Para la amplificación del extremo 5' del ADNc, se diseñaron oligonucleótidos antisentido específicos para FN23: FN23-R1 (5'-CATGGCATCTTCAACCCCAGCTTTATCTCATC-3' (SEC ID Nº: 12)), FN23-R2 (5'-GTGGTCGGATATCACCCATCTT-TCTTGAAGTCG-3' (SEC ID Nº: 13)), FN23-R3 (5'-CATTGGAGATGCAGACTCGACCGATTGACC-3' (SEC ID Nº: 14)). Estos cebadores se usaron para 5'RACE usando la librería de ADNc de *S. sclarea* siguiendo el protocolo del Kit de Amplificación de ADNc Marathon™ (Clontech, Mountain View, CA). Las condiciones de ciclación térmica fueron las que siguen a continuación: 1 min a 94 ºC, 5 ciclos de 30 s a 94 ºC y 4 min a 72 ºC, 5 ciclos de 30 s a 94 ºC y 4 min a 70 ºC, 20 ciclos de 30 s a 94 ºC y 4 min a 68 ºC. Esta 5'RACE proporcionó un fragmento de ADNc de 1449 pb (FN40 (SEC ID Nº: 15) que tenía una superposición perfecta de 227 pb con FN23 (SEC ID Nº: 5). La comparación con secuencias de diterpeno sintasas conocidas reveló que el fragmento FN40 (SEC ID Nº: 15) contenía el codón de iniciación de traducción y una región no codificante de 87 pb. El ensamblaje de los tres fragmentos de ADNc (FN23, FN30 y FN40 (SEC ID Nº: 5, 11 y 15) proporcionó una secuencia de ADNc de longitud completa (SaTpsl) de 2655 pb (SEC ID Nº: 16) con un marco abierto de lectura de 2355 pb que codifica una proteína de 785 restos (SEC ID Nº: 17).

**Ejemplo 2**

Expresión heteróloga de la LPP sintasa de *S. sclarea* en *E. coli*

El pETDuet-1 (Novagen, Madison, WI), diseñado para la expresión con el control de un promotor T7, se usó para la expresión en células de *E. coli*. Para construir el plásmido de expresión, se amplificó el marco abierto de lectura de SaTps1 (SEC ID Nº: 16) mediante PCR a partir de la librería de ADNc con los cebadores directo e inverso SaTps-Nde (3'- TACTGACAT-ATGACTTCTGTAAATTTGAGCAGAGCACC-5' (SEC ID Nº: 18)) y SaTps-Kpn (3'-TTGGTACCTCATACAACCG-GTCGAAAGAGTACTTTG-5' (SEC ID Nº: 19)) diseñados para introducir un sitio NdeI inmediatamente antes del codón de inicio y un sitio KpnI después del codón de terminación. Dado que el marco abierto de lectura contiene un sitio NdeI en la posición 1614 del marco abierto de lectura, esta amplificación se realizó en dos etapas por PCR de extensión superpuesta (Horton y col., Gene 77, 61-68, 1989), usando los cebadores SaTps-Nde (SEC ID Nº: 18) y SaTps-Kpn (SEC ID Nº: 19) en combinación con los cebadores Satps-mut1f (5'- GTTGGAGTGGATCCACATGCAGGAATGGTAC-3' (SEC ID Nº: 20)) y Satps-mut1r (3'- GTACCAT-TCCTGCATCTGGATCCACTCCAAC-5' (SEC ID Nº: 21)), diseñados para retirar el sitio NdeI sin alterar la secuencia de aminoácidos. El ADNc resultante se ligó en primer lugar en el plásmido PCR2.1-Topo usando el Kit de Clonación TOPO TA (Invitrogen, Carlsbad, CA) y las secuencias de los insertos se verificaron antes de la subclonación como fragmento NdeI-KpnI en el vector pETDuet-1.

El análisis de la secuencia de varios clones obtenidos mediante amplificación a partir de la librería de ADNc con los cebadores específicos SaTps1 mostró cierta variabilidad en varias posiciones de la secuencia de ADNc. Se identificaron siete posiciones, en las que se pueden encontrar dos aminoácidos diferentes. Se descubrió que una posición era la inserción de un resto de serina producido en algunos de los clones. Estas posiciones se enumeran en la siguiente tabla.

| **Posiciones** (con respecto a la secuencia de aminoácidos) | **Aminoácido** |
|---|---|
| 34 | Ile o Thr |
| 40 | Phe o Leu |
| 174 | Gln o His |
| 222 | Gly o Asp |
| 538 | Gln o His |

(continuación)

| Posiciones (con respecto a la secuencia de aminoácidos) | Aminoácido |
|---|---|
| 560 | Arg o Leu |
| 596 | Asn o Lys |
| 612 | Inserción de una Ser |

Estas variaciones parecieron producirse de una forma aleatoria en once clones diferentes secuenciados, lo que sugiere que están presentes al menos dos isoformas muy estrechamente relacionadas de una diterpeno sintasas en el genoma de *S. sclarea* y que el enfoque de amplificación de PCR condujo al barajado de las secuencias. Se seleccionaron dos clones, SsLPPs3 (SEC ID Nº: 22) y SsLPPs9 (SEC ID Nº: 23) representativos de la variabilidad de secuencias, para los experimentos de expresión heteróloga y de caracterización enzimática.

Los plásmidos pETDuet-SsLPPs3 y pETDuet-SsLPPs9 se transfirieron a células de *E. coli* B121(DE3) (Novagen, Madison, WI). Se usaron colonias individuales de células transformadas para inocular 5 ml de medio LB. Después de 5 a 6 horas de incubación a 37 ºC, los cultivos se transfirieron a una incubadora a 20 ºC y se dejaron 1 hora para equilibrado. La expresión de la proteína se indujo a continuación mediante la adición de IPTG 1 mM y el cultivo se incubó durante una noche a 20 ºC. Al día siguiente, las células se recogieron por centrifugación, se suspendieron en 0,1 volúmenes de MOPSO 50 mM a pH 7, glicerol al 10 % y se lisaron por sonicación. Los extractos se aclararon por centrifugación (30 min a 20.000 g), y los sobrenadantes que contenían las proteínas solubles se usaron para experimentos adicionales.

Los extractos de proteína sin procesar de las células transformadas pETDuet-SsLPPs3 y pETDuet-SsLPPs9 se analizaron mediante SDS-PAGE y se compararon con los extractos de proteína obtenidos a partir de las células transformadas con el plásmido vacío pETDuet. Se detectaron claramente las proteínas recombinantes SsLPPs3 y SsLPPs9 (SEC ID Nº: 24 y 25) y el peso molecular aparente se estimó en 90 kDa, un valor en concordancia con el peso molecular calculado de 83 kDa.

**Ejemplo 3**

Purificación de la LPP sintasa de *Salvia sclarea* y actividades enzimáticas

Los plásmidos PCR2.1-Topo que contienen el ADNc de SsLPPs3 y SsLPPs9 (SEC ID Nº: 22 y 23) (Ejemplo 2) se digirieron con NdeI y SacI y los insertos se ligaron en el plásmido pET28a(+) (Novagen). Los plásmidos de expresión resultantes (pET28-SsLPPs3 y pET28-SsLPPs9) contienen los ADNc con una modificación en el extremo en la posición 5' diseñada para expresar las proteínas con una marca de hexa-histidina N-terminal (marca de His). La purificación se realizó en condiciones nativas usando el Sistema de Purificación ProBond™ (Invitrogen) siguiendo el protocolo del fabricante excepto en que, para la elución, el imidazol se sustituyó con L-histidina para minimizar la inhibición de la enzima. Usando este enfoque, las enzimas recombinantes con "marca de His" en SsLPPs3 y SsLPPs9 (SEC ID Nº: 97 y 98) se pudieron purificar hasta una homogeneidad aparente.

Las enzimas purificadas por afinidad se incubaron 12 horas a 30 ºC con GGPP 200 µM y DTT 1 mM en MOPSO a pH 7, glicerol al 10 %, DTT 1 mM. No se observó producto de diterpeno mediante la extracción del medio de incubación con pentano y análisis del extracto con GC o GC-MS. El tratamiento del mismo extracto con fosfatasa alcalina (Sigma, 6 unidades/ml), seguido por extracción con pentano y análisis de GC, mostró la formación de labdenodiol y demostró la formación enzimática de difosfato de labdenodiol (LPP) como único producto de GGPP mediante la diterpeno sintasa recombinante.

El análisis de GC se realizó en un sistema 6890 Series GC de Agilent equipado con un detector de ionización a la llama usando un diámetro interno de 0,25 mm con columna de capilaridad SPB-1 de 30 m (Supelco, Bellefonte, PA). El gas vehículo era He a un flujo constante de 1,5 ml/min. La temperatura inicial del horno era de 100 ºC (mantenimiento de 1 min) seguido por un gradiente de 10 ºC/min a 300 ºC. El análisis de GC-MS se realizó en las mismas condiciones y los espectros se registraron en un detector de masas 5975 de Agilent.

**Ejemplo 4**

Enfoque de PCR para la clonación de homología de diterpeno sintasas de la clase I (esclareol sintasa) de *S. sclarea.*

La clonación y caracterización de SsLPPs3 (SEC ID Nº: 24) y SsLPPs9 (SEC ID Nº: 25), en los Ejemplos 1 a 3, sugieren que la biosíntesis del esclareol en *S. sclarea* implica a dos proteínas, la SsLPPs y una diterpeno sintasa de la clase I, la esclareol sintasa, que cataliza la conversión de LPP en esclareol.

Un enfoque de PCR se usó en un primer intento para el aislamiento de secuencias de ADNc de diterpeno sintasas de la clase I. Se diseñaron oligonucleótidos basados en secuencias conservadas en diterpeno sintasas de plantas y especialmente en diterpeno sintasas que catalizan el ciclado de ésteres de difosfato $C_{20}$ a través de un mecanismo

de ionización. Las secuencias con números de registro BAB19275, AAB39482, AAD30231, AAD34295, CAE05201, BAB12441, AAT49066, CAE05199, AAU05906, BAD17672, AAQ72565, AAL09965, AAK83563, AAS47691, AAS47690 y AAR13860, se seleccionaron a partir de las bases de datos de secuencias públicas (http://www.ncbi.nlm.nih.gov). Todas estas secuencias de proteínas corresponden a diterpeno sintasas de la clase I y contienen el motivo DDxxD (en el que x representa cualquier aminoácido) característico del mecanismo de ciclado dependiente de ionización en terpeno sintasas. A partir de la alineación de estas secuencias, en primer lugar se seleccionaron dos motivos conservados en la región N-terminal y se usaron para el diseño de oligonucleótidos sentido: YDT(A/S)WVA y (D/N)GSWG. En la secuencia de aminoácidos de la SsLPPs (SEC ID Nº: 24 y 25, Ejemplos 1 a 3) estos dos motivos también estaban conservados, aunque con algunas diferencias para el primer motivo (YDTAVIA). Por lo tanto, la secuencia de SsLPPs también se tuvo en cuenta para el diseño de los oligonucleótidos sentido. A partir del primer motivo, se diseñaron tres oligonucleótidos para cubrir todas las variaciones de las secuencias: DiTpsTB_F1, 5'-TATGATACNGCNGTNATDGC-3' (SEC ID Nº: 26); DiTpsTB_F2, 5'-TAT-GACACGGCAGTGATCGC-3' (SEC ID Nº: 27); DiTpsTB_F3, 3'-TATGACACGGCAKKGRTNGC-5' (SEC ID Nº: 28). A partir del segundo motivo, se diseñaron dos oligonucleótidos: DiTpsTB_F4, 5'-CAACTGGCTGATGGNTCNTGGGG-3' (SEC ID Nº: 29); DiTpsTB_F5, 5'-CAACTGGC'TGATGGCTCATGGGG-3' (SEC ID Nº: 30). El motivo DDxxD, localizado en la región C-terminal de las proteínas e implicado en la unión del resto de difosfato en el sitio activo, se usó para diseñar dos oligonucleótidos antisentido: DiTpsTB_R1, 5'-GATCCTCCAACRTCRWARARRTCRTC-3' (SEC ID Nº: 31), DiTpsTB_R2, 5'-GATCCTCCACGTCGWAGAAGTCGTC-3' (SEC ID Nº: 32).

Estos cebadores se usaron para amplificación de PCR a partir de una biblioteca de ADNc de *Salvia sclarea* (preparada como se ha descrito en el Ejemplo 1). Las PCR se realizaron usando la Mezcla de Polimerasa 2 Advantage® (Clontech). Cada mezcla de PCR contenía, en un volumen total de 50 µl, 5 µl de tampón de PCR 2 Advantage®, 200 µM de dNTPs, 200 nM de cada cebador de oligonucleótido, 5 µl de ADNc diluido 200 veces, 1 µl de Mezcla de Polimerasa 2 Advantage®. Para las amplificaciones se usaron las siguientes condiciones: 3 minutos de desnaturalización a 94 ºC; 15 ciclos de 1 minuto de desnaturalización a 94 ºC, 1 min de hibridación a 65 ºC para el primer ciclo y menos un grado para cada ciclo siguiente, y 2 minutos de prolongación a 72 ºC; 20 ciclos de 1 minuto de desnaturalización a 94 ºC, 1 min de hibridación a 58 ºC y 2 minutos de prolongación a 72 ºC; y por último 10 minutos de prolongación a 72 ºC. Se realizaron diferentes PCR con las posibles combinaciones de oligonucleótidos sentido y antisentido. Los amplicones se identificaron sistemáticamente para los tamaños esperados y para homología de secuencias con diterpeno sintasas. Desafortunadamente, usando este enfoque de PCR, no se pudo obtener ninguna secuencia relacionada con diterpeno.

**Ejemplo 5**

Secuenciación en paralelo de forma masiva de una biblioteca de ADNc de flor de *S. sclarea*.

Dado que el enfoque clásico de clonación basado en homología no tuvo éxito en la clonación de la diterpeno sintasa de la clase I de *S. sclarea*, los inventores se comprometieron a usar un enfoque basado en la secuenciación global de la biblioteca de ADNc. Los inventores usaron la tecnología de secuenciación en paralelo masiva de fragmentos pequeños de ADN desarrollada por Illumina (San Diego, California) para obtener información de secuencias de todos los transcritos (transcriptoma) presentes en las flores de *Salvia sclarea*. Esta técnica de secuenciación usa una química de secuenciación basada en terminador reversible y los aparatos de Estación de Agrupamiento y Secuenciador de Genoma desarrollados por Solexa e Illumina (www.illumina.com).

La tecnología y el equipo se ajustaron en Fasteris SA (Ginebra, Suiza) y la preparación de las muestras de ADN y la secuenciación se realizaron en Fasteris SA. Una alícuota (1 µg) de la biblioteca de ADNc generada a partir de flores en desarrollo de *S. sclarea* usando el Kit de Amplificación de ADNc Marathon™ (Clontech, Mountain View, CA) (Ejemplo 1), se trató usando el Kit de Preparación de Muestra Genómica (Illumina). En resumen, el ADN se fragmenta mediante nebulización, los extremos se preparan para generar extremos romos, los adaptadores se ligan a los extremos de los fragmentos de ADN y los fragmentos de ADN modificados con adaptador se amplifican con PCR. Después de controlar la calidad de la biblioteca mediante electroforesis en gel, la generación de los agrupamientos de ADN en la célula de flujo y la reacción de secuenciación se realiza en los equipos de Estación de Agrupamiento y Secuenciador de Genoma. Usando esta tecnología, se obtuvieron 1,9 millones de secuencias cortas (lecturas) de 35 bases.

El software Edena (Dr David Hernandez, Genomic Research Laboratory, University of Geneva Hospitals, Ginebra, Suiza) se usó para reensamblar secuencias contiguas. Las cinco últimas bases se retiraron primero de cada lectura debido a las posibles incorporaciones erróneas debido a la menor fidelidad en los últimos ciclos del procedimiento de secuenciación. Los parámetros del software se ajustaron de modo que se permitiera una longitud mínima de 15 bases para los solapamientos con identidad estricta (100 %). Se retuvieron las contigs (secuencias contiguas) con una longitud de al menos 50 bases. En estas condiciones, se pudieron reconstituir 2054 contigs de 50 a 1330 bases de longitud.

Para evaluar la calidad del ensamblaje, se buscaron las contigs para identidad de secuencias con la secuencia de ADN de SsLPPs, las diterpeno sintasas de la clase II aisladas primero a partir de la biblioteca de ADNc de *S. sclarea* (SsLPPs3 (SEC ID Nº: 22), Ejemplo 2). Esta búsqueda se realizó usando el procedimiento BLASTn (Altschul y col.,

J. Mol. Biol. 215, 403-410, 1990). De forma sorprendente, solamente se encontraron 3 contigs con longitudes de 81, 73 y 52 bases y solamente se habían usado cuarenta lecturas con el software Eland para generar estos contigs. La alineación con la secuencia de referencia de SsLPPs3 mostraba que los 3 contigs (SEC ID Nº: 33 a 35 cubrían solamente un 8,7 % de la secuencia de longitud completa aunque con una identidad de un 99 %).

En las bases de datos públicas se ha informado una información de secuencias limitadas para *Salvia sclarea.* La única secuencia genética disponible en la base de datos de NCBI (http://www.ncbi.nlm.nih.gov) era la secuencia de la subunidad grande de la ribulosa-1,5-bisfosfato carboxilasa (RuBisCO) de *Salvia sclarea* (Nº de registro en NCBI Z37450). La búsqueda de las contigs para la identidad de ADN con esta secuencia de ADN de la RuBisCO de *S. sclarea* (Búsqueda con BLASTn) proporcionó dos contigs de 870 y 547 bases respectivamente (SEC ID Nº: 36 y 37). La alineación de las dos contigs con la secuencia de RuBisCO presentaba una cobertura de un 98 %: solamente 27 bases (entre las posiciones 858 y 884) de 1420 bases no estaban presentes en las contigs. Además esta cobertura casi completa, la identidad entre la secuencia de referencia y las contigs era de un 99,5 %, lo que representa una diferencia solamente de 7 nucleótidos.

A continuación, se buscaron todas las lecturas (datos no ensamblados) para identidad de secuencia con la secuencia de SsLPPs3 (SEC ID Nº: 22). El software Eland (Illumina) se usó para realizar esta búsqueda permitiendo un máximo de 2 faltas de coincidencia con la secuencia de referencia. Se recuperó un total de 616 lecturas. La alineación de los fragmentos seleccionados con la secuencia de referencia revelaba que la secuencia de SsLPPs3 (SEC ID Nº: 22) se cubría en la longitud completa con una cobertura ligeramente más elevada (más lecturas) hacia el extremo en la posición 3'. La misma manipulación con la secuencia de la RuBisCO mostraba que se obtenían 1650 lecturas para esta secuencia. La cobertura de la secuencia de referencia con la estructura era mucho más elevada para la RuBisCo que para SsLPPs3 (SEC ID Nº: 22). Para SsLPPs3 (SEC ID Nº: 22), se encontraron varias regiones pequeñas sin cobertura y regiones con ambigüedad de secuencia entre lecturas. Esta cobertura incompleta evita el reensamblaje completo y es ciertamente la razón para generar solamente unos pocos contigs muy pequeños.

El número de lecturas obtenido para un ADNc dado es proporcional a la abundancia de este ADNc. Por lo tanto, las abundancias relativas se pueden calcular dividiendo el número de lecturas obtenidas para los ADNc dados entre sus longitudes totales. Al realizar este cálculo para la RuBisCO y SsLPPs3 (SEC ID Nº: 22) se proporcionaban valores de 1160 y 260 lecturas/kb respectivamente, lo que refleja una abundancia superior a 4,5 del ADNc de RuBisCO con respecto a los ADNc de SsLPPs. La RuBisCO es una enzima implicada en el metabolismo primario de plantas y que cataliza la fijación de carbono en el ciclo de Calvin. La abundancia relativamente más elevada de la RuBisCO refleja una representación elevada de genes implicados en metabolismo primarios en comparación con genes implicados en el metabolismo secundario tales como la síntesis de diterpeno. El análisis de búsqueda con BLAST con las contigs mostraba que otras enzimas de los ciclos de Calvin (por ejemplo, fosfoglicerato quinasa, gliceraldehído 3-fosfato deshidrogenasa, triosafosfato isomerasa) y metabolismo primario también se representaban de forma abundante en la biblioteca de ADNc usada en el presente documento. Por lo tanto, la codificación del ADNc para las enzimas implicadas en el metabolismo secundario y en particular en la biosíntesis de diterpeno estaba en una abundancia demasiado baja para obtener una cobertura suficiente y un reensamblaje completo.

**Ejemplo 6**

Extracción de secuencias del tipo diterpeno sintasas de la clase I a partir de los datos de secuenciación.

El algoritmo Blast (Altschul y col., J. Mol. Biol. 215, 403-410, 1990) se usó para buscar homología de las secuencias de aminoácidos deducidas con secuencias de diterpeno sintasas de la clase I.

EN primer lugar se realizó una búsqueda de Blastx frente a una base de datos de proteínas con los 2054 contigs. Esta búsqueda proporcionó solamente un contig (contig1610, SEC ID Nº: 38) que presentaba homología de secuencia con diterpeno sintasas de la clase I. La secuencia de aminoácidos deducida a partir de este contig contenía el motivo DDxxD característico de ciclación de difosfatos de prenilo iniciado mediante ionización.

A continuación se buscó una fracción de los datos sin procesar, que representa aproximadamente 3 x $10^5$ lecturas para homología con diterpeno sintasas de la clase I. Las lecturas se buscaron usando el algoritmo tBlastn con cinco secuencias de aminoácidos seleccionadas de diterpeno sintasa de la clase I (números de registro en NCBI AAC39443, BAB19275, BAB12441, AAD34295, AAS98912). Esta búsqueda seleccionó 462 lecturas, que a continuación se profesaron usando el programa CAP (Huang, Genomics 14 (1), 18-25, 1992) para identificar secuencias de solapamiento. Una pequeña porción de las lecturas se pudo ensamblar en contigs cortos de longitud máxima de 111 bases. Estos contigs, así como las lecturas aisladas restantes se usaron para búsqueda con Blastx frente a una base de datos de proteínas para confirmar su identidad con diterpeno sintasas de la clase I. Por último, se retuvieron 5 fragmentos de ADN (SEC ID Nº: 39 a 43).

Las secuencias de aminoácidos se dedujeron a partir de los fragmentos seleccionados (SEC ID Nº: 44 ha 48) y se alinearon con secuencias de diterpeno sintasas de referencia, permitiendo su colocación relativa. La Figura 3 muestra una alineación de estas secuencias con una secuencia de diterpeno sintasa de longitud completa, la estemodeno sintasa de *Oriza sativa* (Morrone y col., 2006; Nº de registro en NCBI AAZ76733) tomará como

referencia.

**Ejemplo 7**

Amplificación de PCR de los ADNc de diterpeno sintasas de la clase I de longitud completa.

Un conjunto de oligonucleótidos directos e inversos se dedujo de las secuencias de ADN relacionadas con las diterpeno sintasas seleccionadas a partir de la secuenciación de la biblioteca de ADNc de *S. sclarea* (Ejemplo 6). Estos cebadores se usaron en combinación con cebadores adaptadores de ADNc en amplificaciones de PCR de tipo 3'/5'RACE. Las amplificaciones se realizaron usando la biblioteca de ADNc de *S. sclarea*, preparada como se ha descrito en el Ejemplo 1, siguiendo el protocolo del Kit de amplificación de ADNc Marathon™ (Clontech, Mountain View, CA). Las condiciones de ciclado térmico fueron las que siguen a continuación: 1 min a 94 ºC, 5 ciclos de 30 segundos a 94 ºC y 4 min a 72 ºC, 5 ciclos de 30 segundos a 94 ºC y 4 min a 70 ºC, 20 ciclos de 30 segundos a 94 ºC y 4 min a 68 ºC.

Usando el cebador Cont250-Fwd (SEC ID Nº: 49) se obtuvo una secuencia de ADN de 547 pb (1130Cont250, SEC ID Nº: 81). El análisis de la secuencia reveló que correspondía al extremo en la posición 5' de un ADNc de diterpeno sintasa y contenía 348 pb de la región de codificación. Con el cebador Cont147_fw1 (SEC ID Nº: 51) y Cont147_fw2 (SEC ID Nº: 52) los inventores obtuvieron una secuencia de 1473 pb (1132Cont147, SEC ID Nº: 82) que contenía el extremo en la posición 3' y 1293 pb de la región de codificación de un ADNc de diterpeno sintasa. El cebador Cont224_fw (SEC ID Nº: 57) proporcionó un fragmento de ADN de 207 pb (1137Cont224, SEC ID Nº: 83) que codificaba los 43 aminoácidos C-terminales de una diterpeno sintasa con una secuencia distinta de la de 1132Cont147 (SEC ID Nº: 82). Los cebadores Cont147_rev1 (SEC ID Nº: 53) y Cont147_rev2 (SEC ID Nº: 54) permitían la amplificación de un fragmento de ADN de 464 pb (1134Cont147, SEC ID Nº: 84). El aminoácido deducido mostraba homología con diterpeno sintasas pero la alineación con otras secuencias de diterpeno sintasas sugería que aún faltaban de 200 a 300 codones para alcanzar el extremo en la posición 5'. Todas las secuencias obtenidas con estas series de amplificación diferían de forma significativa de las secuencias de la SsLPPs aisladas previamente (SEC ID Nº: 22 y 23). La PCR con los otros cebadores deducidos de las secuencias de ADN relacionadas con diterpeno sintasas (cebadores cont224-rev (SEC ID Nº: 58), cont250-rev (SEC ID Nº: 50), cont33-fw1 (SEC ID Nº: 55) y cont33-rev (SEC ID Nº: 56)) no proporcionaba secuencias relacionadas con diterpeno sintasa.

A partir de la única secuencia que contenía una región de inicio de la traducción evidente de una diterpeno sintasa (1130count250, SEC ID Nº: 81), se dedujeron oligonucleótidos sentido a partir de la región sin traducir en la posición 5' (UTR) (1130-fw1 (SEC ID Nº: 59) y 1130-fw2 (SEC ID Nº: 60) y a partir del extremo en la posición 5' del marco de lectura abierto (ORF) (1130-fw3, SEC ID Nº: 61). A partir de las dos secuencias que contenían la región de codón de parada de dos diterpeno sintasas distintas (1132Cont147 (SEC ID Nº: 82) y 1137Cont224 (SEC ID Nº: 83)), se dedujeron cebadores de sentido inverso de cualquiera del UTR en la posición 3' (1132-rev1 (SEC ID Nº: 65) y 1137-rev1 (SEC ID Nº: 62)) o del extremo en la posición 3' del marco de lectura abierto (1132-rev2 (SEC ID Nº: 64) y 1137-rev2 (SEC ID Nº: 63)). La PCR se realizó con diferentes combinaciones de estos cebadores directos e inversos. La combinación de cebadores deducida de la secuencia de 1130Cont250 (SEC ID Nº: 81) con los cebadores deducidos de la secuencia de 1137Cont224 (SEC ID Nº: 83) produjo un fragmento de 2388 pb (SEC ID Nº: 85) que codificaba una proteína de 795 aminoácidos (SsTps1137, SEC ID Nº: 86)). La comparación con secuencias publicadas presentaba homología es con diterpeno sintasas de la clase I y en particular con *ent*-kaureno sintasas B. La homología más elevada se produjo con una proteína sin caracterizar de *Vitis vinifera* (Nº de registro en NCBI CAO64942, identidad de un 59 %), una ent-kaureno sintasa de *Cucumis sativus* (Nº de registro en NCBI BAB19275, identidad de un 54 %) y una ent-kaureno sintasa de *Lactuca sativa* (Nº de registro en NCBI BAB12441, identidad de un 54 %). La secuencia de aminoácidos de SsTps1137 (SEC ID Nº: 86) contenía un motivo DDFFD habitual en terpeno sintasas dependientes de ionización (clase I) y no contenía el motivo característico de la clase II.

La combinación de los mismos cebadores directos con los cebadores inversos deducidos del 1132Cont147 (SEC ID Nº: 82) no permitía la amplificación de ningún fragmento, lo que confirma que estas dos secuencias no surgen del mismo ADNc. A continuación, se usó un enfoque de 5'RACE para identificar el extremo en la posición 5' del ORF correspondiente a la secuencia de 1132Cont147 (SEC ID Nº: 82). Usando los cebadores 1132_race1 (SEC ID Nº: 67) y 1132_race2 (SEC ID Nº: 68), se obtuvo una secuencia de 536 pb (1132RACE, SEC ID Nº: 87) que tenía un solapamiento de 41 bases con el fragmento de 1132Cont147 (SEC ID Nº: 82). Este producto de RACE era idéntico al de la secuencia de 1134Cont147 obtenida previamente (SEC ID Nº: 84) y no se observaba extensión en el extremo en la posición 5'. Como se ha observado previamente, esta secuencia tenía homología con diterpeno sintasas, pero parecía más corta en al menos 200 codones que todas las otras secuencias de diterpeno sintasas publicadas. Se realizaron experimentos de 5'RACE, para intentar prolongar la secuencia hacia el extremo en la posición 5' de la secuencia de 1132Cont147 (SEC ID Nº: 82) y para identificar el codón de inicio de la traducción verdadero. Se diseñaron varios conjuntos de oligonucleótidos (1132_race3 a 1132_race9, SEC ID Nº: 69 a 75) pero no se obtuvo información de secuencia adicional. Esto condujo a los inventores a suponer que uno del codón de ATG en la secuencia de 1134Cont147 (SEC ID Nº: 84) era realmente el codón de inicio del gen correspondiente de la diterpeno sintasa. La secuencia de nucleótidos de esta supuesta diterpeno sintasa (denominada SsTps1132, SEC ID Nº: 2) se reconstituyó a partir de las secuencias de 1132Cont147 (SEC ID Nº: 82) y 1132RACE (SEC ID Nº: 87). Tomando el primer ATG, el ORF de 1728 pb de SsTps1132 (SEC ID Nº: 2) codificaba una proteína de 575 aminoácidos (SEC ID Nº: 1). Esta proteína contenía el motivo dependiente de ionización (DDFFD) y compartía

homología, aunque relativamente baja, con diterpeno sintasas publicadas; siendo la secuencia más cercana la de una terpeno sintasa de *Nicotiana tabacum* (Nº de registro en NCBI AAS98912), con una identidad de un 37 %.

De forma sorprendente, la identidad entre las proteínas SsTps1137 (SEC ID Nº: 86) y SsTps1132 (SEC ID Nº: 1) era solamente de un 30 % y estas secuencias compartían solamente de un 21 a un 23 % de identidad con las SsLPPs de la clase II aisladas primero a partir de *S. sclarea* (SEC ID Nº: 24 y 25, Ejemplos 1-3). Una alineación de estas dos proteínas con secuencias seleccionadas de diterpeno sintasas se presenta en la figura 4. La alineación muestra que SsTps1132 (SEC ID Nº: 1) se trunca en el extremo N-terminal con 150 a 240 aminoácidos en comparación con las otras diterpeno sintasas. El procedimiento de ChloroP (Emanuelsson y col., Protein Science 8, 978-984, 1999; http://www.cbs.dtu.dk/services/ChloroP/) se usó para predecir la presencia de un péptido de tránsito de cloroplasto en cada secuencia de proteínas. Para SsTps1137 (SEC ID Nº: 86) y SsTps1132 (SEC ID Nº: 1), se predijeron péptidos de tránsito de cloroplasto de 22 y 51 aminoácidos respectivamente, lo que argumenta una localización en el cloroplasto de ambas proteínas.

La búsqueda de todas las lecturas de secuencias idénticas a las secuencias de ADN de SsTps1137 (SEC ID Nº: 85) y SsTps1132 (SEC ID Nº: 2), proporciono solamente 24 lecturas para SsTps1137 y 425 lecturas para SsTps1132. Esta diferencia en el número de lecturas generado a partir de cada transcrito refleja una diferencia significativa en los niveles de expresión. Basándose en el número relativo de lecturas obtenidas para cada transcrito, se puede calcular que el nivel de expresión de SsTps1132 (220 lecturas/Kb) era similar al nivel de expresión de SsLPPs (260 lecturas/Kb) y que SsTps1137 se expresaba a un nivel mucho menor (10 lecturas/Kb). Con la suposición de que las enzimas que catalizan etapas en la misma ruta metabólica se expresan generalmente a un nivel similar, se puede especular que SsTps1132 (SEC ID Nº: 1) en lugar de SsTps1137 (SEC ID Nº: 86) está implicado en la misma ruta metabólica que las SsLPPs.

Los contigs generados con el software Edena (Ejemplo 5) que buscaron para secuencias de ADN idénticas a las secuencias de estas dos nuevas y supuestas diterpeno sintasas de la clase I. país de inicial para SsTps1137 (SEC ID Nº: 85) no se encontró ningún contig de acuerdo con el supuesto nivel de baja expresión de esta enzima. Para SsTps1132 (SEC ID Nº: 2), se encontraron 4 contigs. El contig1610 (SEC ID Nº: 38) identificado previamente y tres contigs adicionales (de longitud de 53 a 96 pb) (SEC ID Nº: 88 a 90) no se identificaron previamente como fragmento de una diterpeno sintasa. La búsqueda de Blastx con estas tres secuencias no presentaba homología con secuencias de proteínas conocidas. El fallo para encontrar homología para estos contigs se debe a las cortas longitudes de estos fragmentos y a la baja homología de SsTps1132 (SEC ID Nº: 1) con las diterpeno sintasas presentes en las bases de datos.

La observación de una supresión N-terminal de SsTps1132 (SEC ID Nº: 1) en comparación con las otras diterpeno sintasas también explica después por qué no tuvo éxito el enfoque de PCR usado en primer lugar. De hecho, los cebadores directos se diseñaron a partir de regiones conservadas presentes en los primeros 150 aminoácidos de diterpeno sintasas, una región ausente en SsTps1132 (SEC ID Nº: 1). La secuencia de SsTps1137 (SEC ID Nº: 86) contiene los motivos conservados usados para diseñar los cebadores y las secuencias de ADN correspondientes son complementarias con las secuencias del cebador. Supuestamente, la amplificación de SsTps1137 (SEC ID Nº: 85) no tuvo éxito en el enfoque de PCR debido a la baja abundancia de este transcrito.

**Ejemplo 8**

Expresión heteróloga de las diterpeno sintasas de la clase I de *S. Sclarea* en *E. coli.*

Para asignar una actividad enzimática a SsTps1137 (SEC ID Nº: 86) y SsTps1132 (SEC ID Nº: 1), las proteínas recombinantes se expresaron en *E. coli*. Los ADNc de longitud completa se insertaron en el vector pet101/D-TOPO usando el Kit de Expresión Champion pET101 Directional TOPO.

Para cada enzima, se prepararon dos constructos: uno para expresar la proteína de longitud completa y uno para expresar una proteína truncada basándose en la predicción de péptido de tránsito de cloroplasto. Los marcos de lectura abiertos SsTps1137 (SEC ID Nº: 85) y SsTps1132 (SEC ID Nº: 2) de longitud completa se amplificaron a partir de la biblioteca de ADNc usando los pares de cebadores 1137-start (SEC ID Nº: 78) con 1137-stop (SEC ID Nº: 80) y 1132-start1 (SEC ID Nº: 76) con 1132-stop (SEC ID Nº: 66) respectivamente. Los cebadores 1137_start2 (SEC ID Nº: 79) y 1137_stop (SEC ID Nº: 80) se usaron para amplificar una versión truncada de 72 pb de SsTps1137 diseñada para expresar la proteína con 24 aminoácidos suprimidos en el extremo N-terminal. De la misma manera, los cebadores 1132_start2 (SEC ID Nº: 77) y 1132-parada (SEC ID Nº: 66) se usaron para preparar una versión truncada de SsTps1132 diseñada para expresar la proteína con una supresión N-terminal de 50 aminoácidos. Todas las amplificaciones de ADNc para la expresión de los constructos de expresión se realizaron usando *Pfu* ADN polimerasa (Promega), en un volumen final de 50 µl que contenía 5 µl de *Pfu* ADN polimerasa, tampón 10X, cada dNTP a 200 µM, cada cebador directo e inverso a 0,4 µM, 2,9 unidades de *Pfu* ADN polimerasa y 5 µl de ADNc diluido 100 veces (preparado como se describe en el presente documento en el Ejemplo 1 usando el Kit de Amplificación de ADNc Marathon™ (Clontech)). Las condiciones de ciclado térmico fueron las que siguen a continuación: 1,5 min a 95 ºC; 30 ciclos de 45 segundos a 95 ºC, 30 segundos a 58 ºC y 5 min a 72 ºC; y 10 min a 72 ºC.

Después de la ligadura en el vector pET101, se seleccionaron varios clones para cada constructo y se secuenciaron para asegurar que no se había introducido ninguna mutación durante la amplificación de PCR. Para SsTps1137, se seleccionaron los dos constructos 1137-B12 (SEC ID Nº: 91) y 1137-2-B12 (SEC ID Nº: 92) que contenían el ADNc de SsTps1137 respectivamente con y sin el péptido señal (que corresponde a secuencias de polipéptidos en las SEC ID Nº: 94 y 95). Para SsTps1132, se seleccionaron dos constructos: uno con la secuencia completa de SsTps1132 (SEC ID Nº: 2) y un constructo sin péptido señal (1132-2-5, SEC ID Nº: 93). La alineación de las dos secuencias de aminoácidos (SEC ID Nº: 1 y 96) deducida de estos constructos se muestra en la figura 5.

Los plásmidos pET101-1137-B12, pET101-1137-2-B12, pET101-SsTps1132, y pET101-1132-2-5 se transfirieron en células B121(DE3) de *E. Coli* (Novagene, Madison, WI). Para inocular 5 ml de medio LB se usaron colonias individuales de células transformadas. Después de 5 a 6 horas de incubación a 37 ºC, los cultivos se transfirieron a una incubadora a 20 ºC y se dejaron 1 hora para equilibrio. A continuación, la expresión de la proteína se indujo mediante la adición de IPTG 1 mM y el cultivo se incubó durante la noche a 20 ºC. Al día siguiente, las células se recogieron para centrifugación, se volvieron a suspender en 0,1 volúmenes de MOPSO 50 mM a pH 7, glicerol al 10 % y se lisaron mediante sonicación. Los extractos se eliminaron mediante centrifugación (30 min a 20.000 g), y los sobrenadantes que contenían las proteínas solubles se usaron para experimentos adicionales. Los extractos de proteína sin procesar se analizaron con SDS-PAGE y se compararon con extractos de proteína obtenidos de células transformadas con el plásmido pET101 vacío.

**Ejemplo 9**

Actividad enzimática de las diterpeno sintasas de la clase I de *S. Sclarea* recombinante en *E coli.*

Los extractos de proteína de *E. coli* sin procesar que contienen las proteínas recombinantes y se preparan como se ha descrito en el Ejemplo 8 se usaron para la caracterización de las actividades enzimáticas. Los ensayos enzimáticos se realizaron como se ha descrito en el Ejemplo 3. Todos los ensayos se realizaron en MOPSO 50 mM a pH 7, glicerol al 10 %, DTT 1 mM.

En primer lugar se evaluaron las actividades enzimáticas usando como sustrato cualquiera de GGPP o LPP, el producto de SsLPPs (SEC ID Nº: 22) y el supuesto compuesto intermedio en la biosíntesis de esclareol (Ejemplos 1 a 3). El GGPP se sintetizó como se describe en Keller y Thompson (J. Chromatogr 645 (1), 1993, 161-167) y el LPP se preparó esquemáticamente como se ha descrito en el Ejemplo 3. Los ensayos se realizaron en presencia de 10 a 100 µM de sustrato, $MgCl_2$ 15 mM y de 0,1 a 0,5 mg de proteínas sin procesar en un volumen total de 1 ml. Los tubos se incubaron de 4 a 12 horas a 30 ºC y se extrajeron dos veces con un volumen de pentano. Después de concentración con un flujo de nitrógeno, los extractos se analizaron con GC y GC-MS (usando las condiciones que se han descrito en el Ejemplo 3) y se compararon con los extractos del ensayo con proteínas de control (obtenidas de células transformadas con el plásmido vacío). Con el GGPP como sustrato, no se observó actividad con ninguna de las proteínas recombinantes (los datos no se muestran). Con el LPP como sustrato, no se observó actividad con los extractos de proteínas que contenían proteínas recombinantes de SsTps1137 pero con SsTps1132, la actividad se observó tanto con SsTps1132 como con 1132-2-5 (SEC ID Nº: 1 y 96) (Figura 6). Las enzimas también eran activas en ausencia de $MgCl_2$ y se observaron los mismos perfiles de producto con una actividad global aproximadamente igual. La identidad del producto se confirmó mediante concordancia de los tiempos de retención (Figura 6) y con el emparejamiento del espectro de masas con el espectro de un patrón auténtico (Figura 7). En todos los ensayos, se observó un solo pico de esclareol sin traza de producto adicional.

A continuación se realizaron ensayos con coincubación de las diterpeno sintasas de la clase II (SsLPPs3, SEC ID Nº: 24; Ejemplos 1-3) y las diterpeno sintasas de la clase I (serie 1132, SEC ID Nº: 1 y 96). Los ensayos se realizaron en MOPSO 50 mM a pH 7, glicerol al 10 %, DTT 1 mM, GGPP 50 µM, con $MgCl_2$ 1 mM y en presencia de 50 µl de los extractos de proteína sin procesar de *E. coli* que expresan los diferentes constructos. Por lo tanto, los ensayos en presencia de 50 µl de extractos de proteína sin procesar que contenían el SsLPPs3 (SEC ID Nº: 24) enzima recombinante y 50 µl de extractos que contenían SsTps1132 (SEC ID Nº: 1) o 1132-2-5 (SEC ID Nº: 96) Se evaluaron para la producción de productos de diterpeno. La Figura 8 muestra los perfiles de GC de extractos de tales incubaciones en presencia de $MgCl_2$. El esclareol se producía con ambos constructos de 1132 (SEC ID Nº: 1 y 96) (Figura 8), un resultado coherente con el ensayo que se ha descrito anteriormente con LPP como sustrato. No se observó ninguna diferencia significativa cuando se omitía el $MgCl_2$ de las incubaciones (los datos no se muestran).

En conclusión, la SsTps1132 (SEC ID Nº: 2) codifica la esclareol sintasa (SEC ID Nº: 1) y cataliza la conversión de LPP en esclareol.

Listado de secuencias

<110> Firmenich SA

<120> Procedimiento para producir esclareol

<130> 7260

<160> 98

<170> PatentIn versión 3,5

<210> 1
<211> 575
<212> PRT
<213> *Salvia sclarea*

<400> 1

```
Met Ser Leu Ala Phe Asn Val Gly Val Thr Pro Phe Ser Gly Gln Arg
1               5                  10                  15

Val Gly Ser Arg Lys Glu Lys Phe Pro Val Gln Gly Phe Pro Val Thr
            20                  25                  30

Thr Pro Asn Arg Ser Arg Leu Ile Val Asn Cys Ser Leu Thr Thr Ile
        35                  40                  45

Asp Phe Met Ala Lys Met Lys Glu Asn Phe Lys Arg Glu Asp Asp Lys
    50                  55                  60

Phe Pro Thr Thr Thr Thr Leu Arg Ser Glu Asp Ile Pro Ser Asn Leu
65                  70                  75                  80

Cys Ile Ile Asp Thr Leu Gln Arg Leu Gly Val Asp Gln Phe Phe Gln
                85                  90                  95

Tyr Glu Ile Asn Thr Ile Leu Asp Asn Thr Phe Arg Leu Trp Gln Glu
            100                 105                 110

Lys His Lys Val Ile Tyr Gly Asn Val Thr Thr His Ala Met Ala Phe
        115                 120                 125

Arg Leu Leu Arg Val Lys Gly Tyr Glu Val Ser Ser Glu Glu Leu Ala
    130                 135                 140

Pro Tyr Gly Asn Gln Glu Ala Val Ser Gln Gln Thr Asn Asp Leu Pro
145                 150                 155                 160

Met Ile Ile Glu Leu Tyr Arg Ala Ala Asn Glu Arg Ile Tyr Glu Glu
                165                 170                 175
```

```
Glu Arg Ser Leu Glu Lys Ile Leu Ala Trp Thr Thr Ile Phe Leu Asn
            180                 185                 190

Lys Gln Val Gln Asp Asn Ser Ile Pro Asp Lys Lys Leu His Lys Leu
        195                 200                 205

Val Glu Phe Tyr Leu Arg Asn Tyr Lys Gly Ile Thr Ile Arg Leu Gly
    210                 215                 220

Ala Arg Arg Asn Leu Glu Leu Tyr Asp Met Thr Tyr Tyr Gln Ala Leu
225                 230                 235                 240

Lys Ser Thr Asn Arg Phe Ser Asn Leu Cys Asn Glu Asp Phe Leu Val
                245                 250                 255

Phe Ala Lys Gln Asp Phe Asp Ile His Glu Ala Gln Asn Gln Lys Gly
            260                 265                 270

Leu Gln Gln Leu Gln Arg Trp Tyr Ala Asp Cys Arg Leu Asp Thr Leu
        275                 280                 285

Asn Phe Gly Arg Asp Val Val Ile Ile Ala Asn Tyr Leu Ala Ser Leu
    290                 295                 300

Ile Ile Gly Asp His Ala Phe Asp Tyr Val Arg Leu Ala Phe Ala Lys
305                 310                 315                 320

Thr Ser Val Leu Val Thr Ile Met Asp Asp Phe Phe Asp Cys His Gly
                325                 330                 335

Ser Ser Gln Glu Cys Asp Lys Ile Ile Glu Leu Val Lys Glu Trp Lys
            340                 345                 350

Glu Asn Pro Asp Ala Glu Tyr Gly Ser Glu Glu Leu Glu Ile Leu Phe
        355                 360                 365

Met Ala Leu Tyr Asn Thr Val Asn Glu Leu Ala Glu Arg Ala Arg Val
    370                 375                 380

Glu Gln Gly Arg Ser Val Lys Glu Phe Leu Val Lys Leu Trp Val Glu
385                 390                 395                 400

Ile Leu Ser Ala Phe Lys Ile Glu Leu Asp Thr Trp Ser Asn Gly Thr
                405                 410                 415
```

Gln Gln Ser Phe Asp Glu Tyr Ile Ser Ser Ser Trp Leu Ser Asn Gly
420 425 430

Ser Arg Leu Thr Gly Leu Leu Thr Met Gln Phe Val Gly Val Lys Leu
435 440 445

Ser Asp Glu Met Leu Met Ser Glu Glu Cys Thr Asp Leu Ala Arg His
450 455 460

Val Cys Met Val Gly Arg Leu Leu Asn Asp Val Cys Ser Ser Glu Arg
465 470 475 480

Glu Arg Glu Glu Asn Ile Ala Gly Lys Ser Tyr Ser Ile Leu Leu Ala
485 490 495

Thr Glu Lys Asp Gly Arg Lys Val Ser Glu Asp Glu Ala Ile Ala Glu
500 505 510

Ile Asn Glu Met Val Glu Tyr His Trp Arg Lys Val Leu Gln Ile Val
515 520 525

Tyr Lys Lys Glu Ser Ile Leu Pro Arg Arg Cys Lys Asp Val Phe Leu
530 535 540

Glu Met Ala Lys Gly Thr Phe Tyr Ala Tyr Gly Ile Asn Asp Glu Leu
545 550 555 560

Thr Ser Pro Gln Gln Ser Lys Glu Asp Met Lys Ser Phe Val Phe
565 570 575

<210> 2
<211> 1728
<212> ADN
<213> *Salvia sclarea*

<400> 2

atgtcgctcg ccttcaacgt cggagttacg cctttctccg gccaaagagt tgggagcagg 60

aaagaaaaat ttccagtcca aggatttcct gtgaccaccc ccaataggtc acgtctcatc 120

gttaactgca gccttactac aatagatttc atggcgaaaa tgaaagagaa tttcaagagg 180

gaagacgata aatttccaac gacaacgact cttcgatccg aagatatacc ctctaatttg 240

tgtataatcg acacccttca aaggttgggg gtcgatcaat tcttccaata tgaaatcaac 300

actattctag ataacacatt caggttgtgg caagaaaaac acaaagttat atatggcaat 360

gttactactc atgcaatggc atttaggctt ttgcgagtga aaggatacga agtttcatca 420

gaggagttgg ctccatatgg taaccaagag gctgttagcc agcaaacaaa tgacctgccg 480

atgattattg agctttatag agcagcaaat gagagaatat atgaagaaga gaggagtctt 540

gaaaaaattc ttgcttggac taccatcttt ctcaataagc aagtgcaaga taactcaatt 600

cccgacaaaa aactgcacaa actggtggaa ttctacttga ggaattacaa aggcataacc 660

ataagattgg gagctagacg aaacctcgag ctatatgaca tgacctacta tcaagctctg 720

aaatctacaa acaggttctc taatttatgc aacgaagatt ttctagtttt cgcaaagcaa 780

gatttcgata tacatgaagc ccagaaccag aaaggacttc aacaactgca aaggtggtat 840

gcagattgta ggttggacac cttaaacttt ggaagagatg tagttattat tgctaattat 900

ttggcttcat taattattgg tgatcatgcg tttgactatg ttcgtctcgc atttgccaaa 960

acatctgtgc ttgtaacaat tatggatgat tttttcgact gtcatggctc tagtcaagag 1020

tgtgacaaga tcattgaatt agtaaaagaa tggaaggaga atccggatgc agagtacgga 1080

tctgaggagc ttgagatcct ttttatggcg ttgtacaata cagtaaatga gttggcggag 1140

agggctcgtg ttgaacaggg gcgtagtgtc aaagagtttc tagtcaaact gtgggttgaa 1200

atactctcag ctttcaagat agaattagat acatggagca atggcacgca gcaaagcttc 1260

gatgaataca tttcttcgtc gtggttgtcg aacggttccc ggctgacagg tctcctgacg 1320

atgcaattcg tcggagtaaa attgtccgat gaaatgctta tgagtgaaga gtgcactgat 1380

ttggctaggc atgtctgtat ggtcggccgg ctgctcaacg acgtgtgcag ttctgagagg 1440

gagcgcgagg aaaatattgc aggaaaaagt tatagcattc tactagcaac tgagaaagat 1500

ggaagaaaag ttagtgaaga tgaagccatt gcagagatca atgaaatggt tgaatatcac 1560

tggagaaaag tgttgcagat tgtgtataaa aaagaaagca ttttgccaag aagatgcaaa 1620

gatgtatttt tggagatggc taagggtacg ttttatgctt atgggatcaa cgatgaattg 1680

acttctcctc agcaatccaa ggaagatatg aaatcctttg tcttttga 1728

<210> 3
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador directo

<220>
<221> misc_feature
<222> (6) .. (6)
<223> n es a, c, g, o t

<220>
<221> misc_feature
<222> (15) .. (15)
<223> n es a, c, g, o t

<220>
<221> misc_feature
<222> (18) .. (18)
<223> n es a, c, g, o t

<400> 3

gayrtngayg ayacngcnat gg 22

<210> 4
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador inverso

<220>
<221> misc_feature
<222> (6) .. (6)
<223> n es a, c, g, o t

<220>
<221> misc_feature
<222> (9) .. (9)
<223> n es a, c, g, o t

<220>
<221> misc_feature
<222> (15) .. (15)
<223> n es a, c, g, o t

<400> 4

gtyttnccna kccanacrtc ryyt 24

<210> 5
<211> 354
<212> ADN
<213> *Salvia sclarea*

<400> 5

gcgttaggct tctcaaaatg cacggatacg acgtcgatcc aaatgtacta aaacatttca 60

agcaacaaga tggtaaattt tcctgctaca ttggtcaatc ggtcgagtct gcatctccaa 120

tgtacaatct ttatagggct gctcaactaa gatttccagg agaagaagtt cttgaagaag 180

ccactaaatt tgcctttaac ttcttgcaag aaatgctagt caaagatcga cttcaagaaa 240

gatgggtgat atccgaccac ttatttgatg agataaagct ggggttgaag atgccatggt 300

acgccactct accccgagtc gaggctgcat attatctaga ccattatgct ggtt 354

<210> 6
<211> 29
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador diseñado específicamente para FN23

<400> 6

gcacggatac gacgtcgatc caaatgtac 29

<210> 7
<211> 27
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador diseñado específicamente para FN23

<400> 7

gggctgctca actaagattt ccaggag 27

<210> 8
<211> 29
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador diseñado específicamente para FN23

<400> 8

gggtgatatc cgaccactta tttgatgag 29

<210> 9
<211> 36
<212> ADN
<213> Secuencia artificial

<220>
<223> Secuencia adaptadora para prolongar los cebadores oligodT

<400> 9

aattcggtac ccgggatcct tttttttttt tttttt 36

<210> 10
<211> 19
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 10

aattcggtac ccgggatcc 19

<210> 11
<211> 1271
<212> ADN
<213> *Salvia sclarea*

<400> 11

aagaagttct tgaagaagcc actaaatttg cctttaactt cttgcaagaa atgctagtca 60

aagatcgact tcaagaaaga tgggtgatat ccgaccactt atttgatgag ataaagctgg 120

ggttgaagat gccatggtac gccactctac cccgagtcga ggctgcatat tatctagacc 180

attatgctgg ttctggtgat gtatggattg gcaagagttt ctacaggatg ccagaaatca 240

gcaatgatac atacaaggag cttgcgatat tggatttcaa cagatgccaa acacaacatc 300

agttggagtg gatccatatg caggaatggt acgacagatg cagccttagc gaattcggga 360

taagcaaaag agagttgctt cgctcttact ttctggccgc agcaaccata ttcgaaccgg 420

agagaactca agagaggctt ctgtgggcca aaaccagaat tctttctaag atgatcactt 480

catttgtcaa cattagtgga acaacactat ctttggacta caatttcaat ggcctcgatg 540

aaataattag tagtgccaat gaagatcaag gactggctgg gactctgctg gcaaccttcc 600

atcaacttct agacggattc gatatataca ctctccatca actcaaacat gtttggagcc 660

aatggttcat gaaagtgcag caaggagagg gaagcggcgg ggaagacgcg gtgctcctag 720

cgaacacgct caacatctgc gccggcctca acgaagacgt gttgtccaac aatgaataca 780

cggctctgtc caccctcaca aataaaatct gcaatcgcct cgcccaaatt caagacaata 840

agattctcca agttgtggat gggagcataa aggataagga gctagaacag gatatgcagg 900

cgttggtgaa gttagtgctt caagaaaatg gcggcgccgt agacagaaac atcagacaca 960

cgtttttgtc ggtttccaag actttctact acgatgccta ccacgacgat gagacgaccg 1020

atcttcatat cttcaaagta ctctttcgac cggttgtatg aaaaatattt taagctcgtc 1080

tgcagtccac gtagataatt attttaaaat aaaggataaa ttaacgagaa acgacgccat 1140

tttaaaataa tatgttaaga atggacccta aataagagcg tcgaaacatg cattgggata 1200

taatttatta attgttacac catttcggaa taaaatgatg ttatttcttt ttcatatgta 1260

aaaaaaaaaa a 1271

<210> 12
<211> 32
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador diseñado específicamente para FN23

<400> 12

catggcatct tcaaccccag ctttatctca tc 32

<210> 13
<211> 33
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador diseñado específicamente para FN23

<400> 13

gtggtcggat atcacccatc tttcttgaag tcg 33

<210> 14
<211> 30
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador diseñado específicamente para FN23

<400> 14

cattggagat gcagactcga ccgattgacc 30

<210> 15
<211> 1449
<212> ADN
<213> *Salvia sclarea*

<400> 15

caaaattctc ttccattttt aagataatag taatattcta attttccctc caaaaactcg 60

tgggaaattg aaaaatagaa aataaagatg acttctgtaa atttgagcag agcaccagca 120

gcgattatcc ggcgcaggct gcagctacag ccggaatttc atgccgagtg ttcatggctg 180

aaaagcagca gcaaacacgc gcccttgacc ttgagttgcc aaatccgtcc taagcaactc 240

tcccaaatag ctgaattgag agtaacaagc ctggatgcgt cgcaagcgag tgaaaaagac 300

atttcccttg ttcaaactcc gcataaggtt gaggttaatg aaaagatcga ggagtcaatc 360

gagtacgtcc aaaatctgtt gatgacgtcg ggcgacgggc gaataagcgt gtcaccctat 420

gacacggcag tgatcgccct gatcaaggac ttgaaagggc gcgacgcccc gcagtttccg 480

tcatgtctcg agtggatcgc gcaccaccaa ctggctgatg gctcatgggg cgacgaattc 540

ttctgtattt atgatcggat tctaaataca ttggcatgtg tcgtagcctt gaaatcatgg 600

aaccttcact ctgatattat tgaaaaagga gtgacgtaca tcaaggagaa tgtgcataaa 660

cttaaaggtg caaatgttga gcacaggaca gcggggttcg aacttgtggt tcctactttt 720

atgcaaatgg ccacagattt gggcatccaa gatctgccct atgatcatcc cctcatcaag 780

gagattgctg acacaaaaca acaaagattg aaagagatac ccaaggattt ggtttaccaa 840

atgccaacga atttactgta cagtttagaa gggttaggag atttggagtg ggaaaggcta 900

ctgaaactgc agtcgggcaa tggctccttc ctcacttcgc cgtcgtccac cgccgccgtc 960

ttgatgcata ccaaagatga aaaatgtttg aaatacatcg aaaacgccct caagaattgc 1020

gacggaggag caccacatac ttatccagtc gatatcttct caagactttg ggcaatcgat 1080

aggctacaac gcctaggaat ttctcgtttc ttccagcacg agatcaagta tttcttagat 1140

cacatcgaaa gcgtttggga ggagaccgga gttttcagtg gaagatatac gaaatttagc 1200

gatattgatg acacgtccat gggcgttagg cttctcaaaa tgcacggata cgacgtcgat 1260

ccaaatgtac taaaacattt caagcaacaa gatggtaaat tttcctgcta cattggtcaa 1320

tcggtcgagt ctgcatctcc aatgtacaat ctttataggg ctgctcaact aagatttcca 1380

ggagaagaag ttcttgaaga agccactaaa tttgccttta acttcttgca agaaatgcta 1440

gtcaaagat 1449

<210> 16
<211> 2655
<212> ADN
<213> *Salvia sclarea*

<400> 16

caaaattctc ttccattttt aagataatag taatattcta attttccctc caaaaactcg 60

tgggaaattg aaaaatagaa aataaagatg acttctgtaa atttgagcag agcaccagca 120

gcgattatcc ggcgcaggct gcagctacag ccggaatttc atgccgagtg ttcatggctg 180

aaaagcagca gcaaacacgc gcccttgacc ttgagttgcc aaatccgtcc taagcaactc 240

tcccaaatag ctgaattgag agtaacaagc ctggatgcgt cgcaagcgag tgaaaaagac 300

atttcccttg ttcaaactcc gcataaggtt gaggttaatg aaaagatcga ggagtcaatc 360

gagtacgtcc aaaatctgtt gatgacgtcg ggcgacgggc gaataagcgt gtcaccctat 420

gacacggcag tgatcgccct gatcaaggac ttgaaagggc gcgacgcccc gcagtttccg 480

tcatgtctcg agtggatcgc gcaccaccaa ctggctgatg gctcatgggg cgacgaattc 540

ttctgtattt atgatcggat tctaaataca ttggcatgtg tcgtagcctt gaaatcatgg 600

aaccttcact ctgatattat tgaaaaagga gtgacgtaca tcaaggagaa tgtgcataaa 660

cttaaaggtg caaatgttga gcacaggaca gcggggttcg aacttgtggt tcctactttt 720

atgcaaatgg ccacagattt gggcatccaa gatctgccct atgatcatcc cctcatcaag 780

gagattgctg acacaaaaca acaaagattg aaagagatac ccaaggattt ggtttaccaa 840

atgccaacga atttactgta cagtttagaa gggttaggag atttggagtg ggaaaggcta 900

ctgaaactgc agtcgggcaa tggctccttc ctcacttcgc cgtcgtccac cgccgccgtc 960

ttgatgcata ccaaagatga aaaatgtttg aaatacatcg aaaacgccct caagaattgc 1020

gacggaggag caccacatac ttatccagtc gatatcttct caagactttg ggcaatcgat 1080

aggctacaac gcctaggaat ttctcgtttc ttccagcacg agatcaagta tttcttagat 1140

cacatcgaaa gcgtttggga ggagaccgga gttttcagtg gaagatatac gaaatttagc 1200

gatattgatg acacgtccat gggcgttagg cttctcaaaa tgcacggata cgacgtcgat 1260

ccaaatgtac taaaacattt caagcaacaa gatggtaaat tttcctgcta cattggtcaa 1320

tcggtcgagt ctgcatctcc aatgtacaat ctttataggg ctgctcaact aagatttcca 1380

ggagaagaag ttcttgaaga agccactaaa tttgccttta acttcttgca agaaatgcta 1440

gtcaaagatc gacttcaaga aagatgggtg atatccgacc acttatttga tgagataaag 1500

ctggggttga agatgccatg gtacgccact ctaccccgag tcgaggctgc atattatcta 1560

gaccattatg ctggttctgg tgatgtatgg attggcaaga gtttctacag gatgccagaa 1620

atcagcaatg atacatacaa ggagcttgcg atattggatt tcaacagatg ccaaacacaa 1680

catcagttgg agtggatcca tatgcaggaa tggtacgaca gatgcagcct tagcgaattc 1740

gggataagca aaagagagtt gcttcgctct tactttctgg ccgcagcaac catattcgaa 1800

ccggagagaa ctcaagagag gcttctgtgg gccaaaacca gaattctttc taagatgatc 1860

acttcatttg tcaacattag tggaacaaca ctatctttgg actacaattt caatggcctc 1920

gatgaaataa ttagtagtgc caatgaagat caaggactgg ctgggactct gctggcaacc 1980

ttccatcaac ttctagacgg attcgatata tacactctcc atcaactcaa acatgtttgg 2040

agccaatggt tcatgaaagt gcagcaagga gagggaagcg gcggggaaga cgcggtgctc 2100

ctagcgaaca cgctcaacat ctgcgccggc ctcaacgaag acgtgttgtc caacaatgaa 2160

tacacggctc tgtccaccct cacaaataaa atctgcaatc gcctcgccca aattcaagac 2220

aataagattc tccaagttgt ggatgggagc ataaaggata aggagctaga acaggatatg 2280

caggcgttgg tgaagttagt gcttcaagaa aatggcggcg ccgtagacag aaacatcaga 2340

cacacgtttt tgtcggtttc caagactttc tactacgatg cctaccacga cgatgagacg 2400

accgatcttc atatcttcaa agtactcttt cgaccggttg tatgaaaaat attttaagct 2460

cgtctgcagt ccacgtagat aattatttta aaataaagga taaattaacg agaaacgacg 2520

ccattttaaa ataatatgtt aagaatggac cctaaataag agcgtcgaaa catgcattgg 2580

gatataattt attaattgtt acaccatttc ggaataaaat gatgttattt ctttttcata 2640

tgtaaaaaaa aaaaa 2655

<210> 17
<211> 785
<212> PRT
<213> *Salvia sclarea*

<400> 17

```
Met Thr Ser Val Asn Leu Ser Arg Ala Pro Ala Ala Ile Ile Arg Arg
1               5                  10                  15

Arg Leu Gln Leu Gln Pro Glu Phe His Ala Glu Cys Ser Trp Leu Lys
            20                  25                  30

Ser Ser Ser Lys His Ala Pro Leu Thr Leu Ser Cys Gln Ile Arg Pro
        35                  40                  45

Lys Gln Leu Ser Gln Ile Ala Glu Leu Arg Val Thr Ser Leu Asp Ala
    50                  55                  60

Ser Gln Ala Ser Glu Lys Asp Ile Ser Leu Val Gln Thr Pro His Lys
65                  70                  75                  80

Val Glu Val Asn Glu Lys Ile Glu Glu Ser Ile Glu Tyr Val Gln Asn
                85                  90                  95

Leu Leu Met Thr Ser Gly Asp Gly Arg Ile Ser Val Ser Pro Tyr Asp
            100                 105                 110

Thr Ala Val Ile Ala Leu Ile Lys Asp Leu Lys Gly Arg Asp Ala Pro
        115                 120                 125

Gln Phe Pro Ser Cys Leu Glu Trp Ile Ala His His Gln Leu Ala Asp
    130                 135                 140

Gly Ser Trp Gly Asp Glu Phe Phe Cys Ile Tyr Asp Arg Ile Leu Asn
145                 150                 155                 160

Thr Leu Ala Cys Val Val Ala Leu Lys Ser Trp Asn Leu His Ser Asp
                165                 170                 175

Ile Ile Glu Lys Gly Val Thr Tyr Ile Lys Glu Asn Val His Lys Leu
            180                 185                 190

Lys Gly Ala Asn Val Glu His Arg Thr Ala Gly Phe Glu Leu Val Val
        195                 200                 205

Pro Thr Phe Met Gln Met Ala Thr Asp Leu Gly Ile Gln Asp Leu Pro
    210                 215                 220

Tyr Asp His Pro Leu Ile Lys Glu Ile Ala Asp Thr Lys Gln Gln Arg
225                 230                 235                 240
```

```
Leu Lys Glu Ile Pro Lys Asp Leu Val Tyr Gln Met Pro Thr Asn Leu
                245                 250                 255

Leu Tyr Ser Leu Glu Gly Leu Gly Asp Leu Glu Trp Glu Arg Leu Leu
            260                 265                 270

Lys Leu Gln Ser Gly Asn Gly Ser Phe Leu Thr Ser Pro Ser Ser Thr
        275                 280                 285

Ala Ala Val Leu Met His Thr Lys Asp Glu Lys Cys Leu Lys Tyr Ile
    290                 295                 300

Glu Asn Ala Leu Lys Asn Cys Asp Gly Gly Ala Pro His Thr Tyr Pro
305                 310                 315                 320

Val Asp Ile Phe Ser Arg Leu Trp Ala Ile Asp Arg Leu Gln Arg Leu
                325                 330                 335

Gly Ile Ser Arg Phe Phe Gln His Glu Ile Lys Tyr Phe Leu Asp His
            340                 345                 350

Ile Glu Ser Val Trp Glu Glu Thr Gly Val Phe Ser Gly Arg Tyr Thr
        355                 360                 365

Lys Phe Ser Asp Ile Asp Asp Thr Ser Met Gly Val Arg Leu Leu Lys
    370                 375                 380

Met His Gly Tyr Asp Val Asp Pro Asn Val Leu Lys His Phe Lys Gln
385                 390                 395                 400

Gln Asp Gly Lys Phe Ser Cys Tyr Ile Gly Gln Ser Val Glu Ser Ala
                405                 410                 415

Ser Pro Met Tyr Asn Leu Tyr Arg Ala Ala Gln Leu Arg Phe Pro Gly
            420                 425                 430

Glu Glu Val Leu Glu Glu Ala Thr Lys Phe Ala Phe Asn Phe Leu Gln
        435                 440                 445

Glu Met Leu Val Lys Asp Arg Leu Gln Glu Arg Trp Val Ile Ser Asp
    450                 455                 460

His Leu Phe Asp Glu Ile Lys Leu Gly Leu Lys Met Pro Trp Tyr Ala
465                 470                 475                 480

Thr Leu Pro Arg Val Glu Ala Ala Tyr Tyr Leu Asp His Tyr Ala Gly
```

485 490 495

Ser Gly Asp Val Trp Ile Gly Lys Ser Phe Tyr Arg Met Pro Glu Ile
500 505 510

Ser Asn Asp Thr Tyr Lys Glu Leu Ala Ile Leu Asp Phe Asn Arg Cys
515 520 525

Gln Thr Gln His Gln Leu Glu Trp Ile His Met Gln Glu Trp Tyr Asp
530 535 540

Arg Cys Ser Leu Ser Glu Phe Gly Ile Ser Lys Arg Glu Leu Leu Arg
545 550 555 560

Ser Tyr Phe Leu Ala Ala Ala Thr Ile Phe Glu Pro Glu Arg Thr Gln
565 570 575

Glu Arg Leu Leu Trp Ala Lys Thr Arg Ile Leu Ser Lys Met Ile Thr
580 585 590

Ser Phe Val Asn Ile Ser Gly Thr Thr Leu Ser Leu Asp Tyr Asn Phe
595 600 605

Asn Gly Leu Asp Glu Ile Ile Ser Ser Ala Asn Glu Asp Gln Gly Leu
610 615 620

Ala Gly Thr Leu Leu Ala Thr Phe His Gln Leu Leu Asp Gly Phe Asp
625 630 635 640

Ile Tyr Thr Leu His Gln Leu Lys His Val Trp Ser Gln Trp Phe Met
645 650 655

Lys Val Gln Gln Gly Glu Gly Ser Gly Gly Glu Asp Ala Val Leu Leu
660 665 670

Ala Asn Thr Leu Asn Ile Cys Ala Gly Leu Asn Glu Asp Val Leu Ser
675 680 685

Asn Asn Glu Tyr Thr Ala Leu Ser Thr Leu Thr Asn Lys Ile Cys Asn
690 695 700

Arg Leu Ala Gln Ile Gln Asp Asn Lys Ile Leu Gln Val Val Asp Gly
705 710 715 720

Ser Ile Lys Asp Lys Glu Leu Glu Gln Asp Met Gln Ala Leu Val Lys
725 730 735

Leu Val Leu Gln Glu Asn Gly Gly Ala Val Asp Arg Asn Ile Arg His
740 745 750

Thr Phe Leu Ser Val Ser Lys Thr Phe Tyr Tyr Asp Ala Tyr His Asp
755 760 765

Asp Glu Thr Thr Asp Leu His Ile Phe Lys Val Leu Phe Arg Pro Val
770 775 780

Val
785

<210> 18
<211> 38
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 18

tactgacata tgacttctgt aaatttgagc agagcacc 38

<210> 19
<211> 36
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 19

ttggtacctc atacaaccgg tcgaaagagt actttg 36

<210> 20
<211> 31
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 20

gttggagtgg atccacatgc aggaatggta c 31

<210> 21
<211> 31
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 21

gtaccattcc tgcatctgga tccactccaa c 31

<210> 22
<211> 2358
<212> ADN
<213> *Salvia sclarea*

<400> 22

atgacttctg taaatttgag cagagcacca gcagcgatta cccggcgcag gctgcagcta 60

cagccggaat ttcatgccga gtgttcatgg ctgaaaagca gcagcaaaca cgcgcccttg 120

accttgagtt gccaaatccg tcctaagcaa ctctcccaaa tagctgaatt gagagtaaca 180

agcctggatg cgtcgcaagc gagtgaaaaa gacatttccc ttgttcaaac tccgcataag 240

gttgaggtta atgaaaagat cgaggagtca atcgagtacg tccaaaatct gttgatgacg 300

tcgggcgacg ggcgaataag cgtgtcaccc tatgacacgg cagtgatcgc cctgatcaag 360

gacttgaaag ggcgcgacgc cccgcagttt ccgtcatgtc tcgagtggat cgcgcaccac 420

caactggctg atggctcatg gggcgacgaa ttcttctgta tttatgatcg gattctaaat 480

acattggcat gtgtcgtagc cttgaaatca tggaaccttc actctgatat tattgaaaaa 540

ggagtgacgt acatcaagga gaatgtgcat aaacttaaag gtgcaaatgt tgagcacagg 600

acagcggggt tcgaacttgt ggttcctact tttatgcaaa tggccacaga tttgggcatc 660

caagatctgc cctatgatca tcccctcatc aaggagattg ctgacacaaa acaacaaaga 720

ttgaaagaga tacccaagga tttggtttac caaatgccaa cgaatttact gtacagttta 780

gaagggttag gagatttgga gtgggaaagg ctactgaaac tgcagtcggg caatggctcc 840

ttcctcactt cgccgtcgtc caccgccgcc gtcttgatgc ataccaaaga tgaaaaatgt 900

ttgaaataca tcgaaaacgc cctcaagaat tgcgacggag gagcaccaca tacttatcca 960

gtcgatatct tctcaagact ttgggcaatc gataggctac aacgcctagg aatttctcgt 1020

ttcttccagc acgagatcaa gtatttctta gatcacatcg aaagcgtttg ggaggagacc 1080

ggagttttca gtggaagata tacgaaattt agcgatattg atgacacgtc catgggcgtt 1140

aggcttctca aaatgcacgg atacgacgtc gatccaaatg tactaaaaca tttcaagcaa 1200

caagatggta aattttcctg ctacattggt caatcggtcg agtctgcatc tccaatgtac 1260

aatctttata gggctgctca actaagattt ccaggagaag aagttcttga agaagccact 1320

aaatttgcct ttaacttctt gcaagaaatg ctagtcaaag atcgacttca agaaagatgg 1380

gtgatatccg accacttatt tgatgagata aagctggggt tgaagatgcc atggtacgcc 1440

actctacccc gagtcgaggc tgcatattat ctagaccatt atgctggttc tggtgatgta 1500

tggattggca agagtttcta caggatgcca gaaatcagca atgatacata caaggagctt 1560

gcgatattgg atttcaacag atgccaaaca caacatcagt tggagtggat ccacatgcag 1620

gaatggtacg acagatgcag ccttagcgaa ttcgggataa gcaaaagaga gttgcttcgc 1680

tcttactttc tggccgcagc aaccatattc gaaccggaga gaactcaaga gaggcttctg 1740

tgggccaaaa ccagaattct ttctaagatg atcacttcat ttgtcaacat tagtggaaca 1800

acactatctt tggactacaa tttcaatggc ctcgatgaaa taattagtag tgccaatgaa 1860

gatcaaggac tggctgggac tctgctggca accttccatc aacttctaga cggattcgat 1920

atatacactc tccatcaact caaacatgtt tggagccaat ggttcatgaa agtgcagcaa 1980

ggagagggaa gcggcgggga agacgcggtg ctcctagcga acacgctcaa catctgcgcc 2040

ggcctcaacg aagacgtgtt gtccaacaat gaatacacgg ctctgtccac cctcacaaat 2100

aaaatctgca atcgcctcgc ccaaattcaa gacaataaga ttctccaagt tgtggatggg 2160

agcataaagg ataaggagct agaacaggat atgcaggcgt tggtgaagtt agtgcttcaa 2220

gaaaatggcg gcgccgtaga cagaaacatc agacacacgt ttttgtcggt ttccaagact 2280

ttctactacg atgcctacca cgacgatgag acgaccgatc ttcatatctt caaagtactc 2340

tttcgaccgg ttgtatga 2358

<210> 23
<211> 2355
<212> ADN
<213> *Salvia sclarea*

<400> 23

atgacttctg taaatttgag cagagcacca gcagcgatta tccggcgcag gctgcagcta 60

cagccggaat ttcatgccga gtgttcatgg ctgaaaagca gcagcaaaca cgcgcccttc 120

accttgagtt gccaaatccg tcctaagcaa ctctcccaaa tagctgaatt gagagtaaca 180

agcctggatg cgtcgcaagc gagtgaaaaa gacatttccc ttgttcaaac tccgcataag 240

gttgaggtta atgaaaagat cgaggagtca atcgagtacg tccaaaatct gttgatgacg 300

tcgggcgacg ggcgaataag cgtgtcaccc tatgacacgg cagtgatcgc cctgatcaag 360

gacttgaaag ggcgcgacgc cccgcagttt ccgtcatgtc tcgagtggat cgcgcaccac 420

caactggctg atggctcatg gggcgacgaa ttcttctgta tttatgatcg gattctaaat 480

acattggcat gtgtcgtagc cttgaaatca tggaaccttc aatctgatat tattgaaaaa 540

ggtgtgacgt acatcaagga gaatgtgcat aaacttaaag gtgcaaatgt tgagcacagg 600

acagcggggt tcgaacttgt ggttcctact tttatgcaaa tggccacaga tttgggcatc 660

caaggtctgc cctatgatca tcccctcatc aaggagattg ctgacacaaa acaacaaaga 720

ttgaaagaga tacccaagga tttggtttac caaatgccaa cgaatttact gtacagttta 780

gaagggttag gagatttgga gtgggaaagg ttactgaaac tgcagtcggg caatggctcc 840

ttcctcactt cgccgtcgtc caccgccgcc gtcttgatgc ataccaaaga tgaaaaatgt 900

ttgaaataca tcgaaaacgc cctcaagaat tgcgacggag gagcaccaca tacttatcca 960

gtcgatatct tctcaagact ttgggcaatc gataggctac aacgcctagg aatttctcgt 1020

ttcttccagc acgagatcaa gtatttctta gatcacatcg aaagcgtttg ggaggagacc 1080

ggagttttca gtggaagata tacgaaattt agcgatattg atgacacgtc catgggcgtt 1140

aggcttctca aaatgcacgg atacgacgtc gatccaaatg tactaaaaca tttcaagcaa 1200

caagatggta aattttcctg ctacattggt caatcggtcg agtctgcatc tccaatgtac 1260

aatctttata gggctgctca actaagattt ccaggagaag aagttcttga agaagccact 1320

aaatttgcct ttaacttctt gcaagaaatg ctagtcaaag atcgacttca agaaagatgg 1380

gtgatatccg accacttatt tgatgagata aagctggggt tgaagatgcc atggtacgcc 1440

actctacccc gagtcgaggc tgcatattat ctagaccatt atgctggttc tggtgatgta 1500

tggattggca agagtttcta caggatgcca gaaatcagca atgatacata caaggagctt 1560

gcgatattgg atttcaacag atgccaaaca caacatcagt tggagtggat ccagatgcag 1620

gaatggtacg acagatgcag ccttagcgaa ttcgggataa gcaaaagaga gttgcttcgc 1680

tcttactttc tggccgcagc aaccatattc gaaccggaga gaactcaaga gaggcttctg 1740

tgggccaaaa ccagaattct ttctaagatg atcacttcat ttgtcaacat tagtggaaca 1800

acactatctt tggactacaa tttcaatggc ctcgatgaaa taattagtgc caatgaagat 1860

caaggactgg ctgggactct gctggcaacc ttccatcaac ttctagacgg attcgatata 1920

tacactctcc atcaactcaa acatgtttgg agccaatggt tcatgaaagt gcagcaagga 1980

gagggaagcg gcggggaaga cgcggtgctc ctagcgaaca cgctcaacat ctgcgccggc 2040

ctcaacgaag acgtgttgtc caacaatgaa tacacggctc tgtccaccct cacaaataaa 2100

atctgcaatc gcctcgccca aattcaagac aataagattc tccaagttgt ggatgggagc 2160

ataaaggata aggagctaga acaggatatg caggcgttgg tgaagttagt gcttcaagaa 2220

aatggcggcg ccgtagacag aaacatcaga cacacgtttt tgtcggtttc caagactttc 2280

tactacgatg cctaccacga cgatgagacg accgatcttc atatcttcaa agtactcttt 2340

cgaccggttg tatga 2355

<210> 24
<211> 785
<212> PRT
<213> *Salvia sclarea*
<400> 24

```
Met Thr Ser Val Asn Leu Ser Arg Ala Pro Ala Ala Ile Thr Arg Arg
1               5                  10                  15

Arg Leu Gln Leu Gln Pro Glu Phe His Ala Glu Cys Ser Trp Leu Lys
            20                  25                  30

Ser Ser Ser Lys His Ala Pro Leu Thr Leu Ser Cys Gln Ile Arg Pro
        35                  40                  45

Lys Gln Leu Ser Gln Ile Ala Glu Leu Arg Val Thr Ser Leu Asp Ala
    50                  55                  60

Ser Gln Ala Ser Glu Lys Asp Ile Ser Leu Val Gln Thr Pro His Lys
65                  70                  75                  80

Val Glu Val Asn Glu Lys Ile Glu Glu Ser Ile Glu Tyr Val Gln Asn
                85                  90                  95

Leu Leu Met Thr Ser Gly Asp Gly Arg Ile Ser Val Ser Pro Tyr Asp
            100                 105                 110

Thr Ala Val Ile Ala Leu Ile Lys Asp Leu Lys Gly Arg Asp Ala Pro
        115                 120                 125

Gln Phe Pro Ser Cys Leu Glu Trp Ile Ala His His Gln Leu Ala Asp
    130                 135                 140

Gly Ser Trp Gly Asp Glu Phe Phe Cys Ile Tyr Asp Arg Ile Leu Asn
145                 150                 155                 160

Thr Leu Ala Cys Val Val Ala Leu Lys Ser Trp Asn Leu His Ser Asp
                165                 170                 175

Ile Ile Glu Lys Gly Val Thr Tyr Ile Lys Glu Asn Val His Lys Leu
            180                 185                 190

Lys Gly Ala Asn Val Glu His Arg Thr Ala Gly Phe Glu Leu Val Val
        195                 200                 205

Pro Thr Phe Met Gln Met Ala Thr Asp Leu Gly Ile Gln Asp Leu Pro
    210                 215                 220

Tyr Asp His Pro Leu Ile Lys Glu Ile Ala Asp Thr Lys Gln Gln Arg
225                 230                 235                 240
```

```
Leu Lys Glu Ile Pro Lys Asp Leu Val Tyr Gln Met Pro Thr Asn Leu
                245                 250                 255

Leu Tyr Ser Leu Glu Gly Leu Gly Asp Leu Glu Trp Glu Arg Leu Leu
            260                 265                 270

Lys Leu Gln Ser Gly Asn Gly Ser Phe Leu Thr Ser Pro Ser Ser Thr
        275                 280                 285

Ala Ala Val Leu Met His Thr Lys Asp Glu Lys Cys Leu Lys Tyr Ile
    290                 295                 300

Glu Asn Ala Leu Lys Asn Cys Asp Gly Gly Ala Pro His Thr Tyr Pro
305                 310                 315                 320

Val Asp Ile Phe Ser Arg Leu Trp Ala Ile Asp Arg Leu Gln Arg Leu
                325                 330                 335

Gly Ile Ser Arg Phe Phe Gln His Glu Ile Lys Tyr Phe Leu Asp His
            340                 345                 350

Ile Glu Ser Val Trp Glu Glu Thr Gly Val Phe Ser Gly Arg Tyr Thr
        355                 360                 365

Lys Phe Ser Asp Ile Asp Asp Thr Ser Met Gly Val Arg Leu Leu Lys
    370                 375                 380

Met His Gly Tyr Asp Val Asp Pro Asn Val Leu Lys His Phe Lys Gln
385                 390                 395                 400

Gln Asp Gly Lys Phe Ser Cys Tyr Ile Gly Gln Ser Val Glu Ser Ala
                405                 410                 415

Ser Pro Met Tyr Asn Leu Tyr Arg Ala Ala Gln Leu Arg Phe Pro Gly
            420                 425                 430

Glu Glu Val Leu Glu Glu Ala Thr Lys Phe Ala Phe Asn Phe Leu Gln
        435                 440                 445

Glu Met Leu Val Lys Asp Arg Leu Gln Glu Arg Trp Val Ile Ser Asp
    450                 455                 460

His Leu Phe Asp Glu Ile Lys Leu Gly Leu Lys Met Pro Trp Tyr Ala
465                 470                 475                 480
```

```
Thr Leu Pro Arg Val Glu Ala Ala Tyr Tyr Leu Asp His Tyr Ala Gly
                485                 490                 495

Ser Gly Asp Val Trp Ile Gly Lys Ser Phe Tyr Arg Met Pro Glu Ile
            500                 505                 510

Ser Asn Asp Thr Tyr Lys Glu Leu Ala Ile Leu Asp Phe Asn Arg Cys
        515                 520                 525

Gln Thr Gln His Gln Leu Glu Trp Ile His Met Gln Glu Trp Tyr Asp
    530                 535                 540

Arg Cys Ser Leu Ser Glu Phe Gly Ile Ser Lys Arg Glu Leu Leu Arg
545                 550                 555                 560

Ser Tyr Phe Leu Ala Ala Ala Thr Ile Phe Glu Pro Glu Arg Thr Gln
                565                 570                 575

Glu Arg Leu Leu Trp Ala Lys Thr Arg Ile Leu Ser Lys Met Ile Thr
            580                 585                 590

Ser Phe Val Asn Ile Ser Gly Thr Thr Leu Ser Leu Asp Tyr Asn Phe
        595                 600                 605

Asn Gly Leu Asp Glu Ile Ile Ser Ser Ala Asn Glu Asp Gln Gly Leu
    610                 615                 620

Ala Gly Thr Leu Leu Ala Thr Phe His Gln Leu Leu Asp Gly Phe Asp
625                 630                 635                 640

Ile Tyr Thr Leu His Gln Leu Lys His Val Trp Ser Gln Trp Phe Met
                645                 650                 655

Lys Val Gln Gln Gly Glu Gly Ser Gly Gly Glu Asp Ala Val Leu Leu
            660                 665                 670

Ala Asn Thr Leu Asn Ile Cys Ala Gly Leu Asn Glu Asp Val Leu Ser
        675                 680                 685

Asn Asn Glu Tyr Thr Ala Leu Ser Thr Leu Thr Asn Lys Ile Cys Asn
    690                 695                 700

Arg Leu Ala Gln Ile Gln Asp Asn Lys Ile Leu Gln Val Val Asp Gly
705                 710                 715                 720
```

```
Ser Ile Lys Asp Lys Glu Leu Glu Gln Asp Met Gln Ala Leu Val Lys
                725                 730                 735

Leu Val Leu Gln Glu Asn Gly Gly Ala Val Asp Arg Asn Ile Arg His
            740                 745                 750

Thr Phe Leu Ser Val Ser Lys Thr Phe Tyr Tyr Asp Ala Tyr His Asp
        755                 760                 765

Asp Glu Thr Thr Asp Leu His Ile Phe Lys Val Leu Phe Arg Pro Val
    770                 775                 780

Val
785
```

<210> 25
<211> 784
<212> PRT
<213> *Salvia sclarea*

<400> 25

```
Met Thr Ser Val Asn Leu Ser Arg Ala Pro Ala Ala Ile Ile Arg Arg
1               5                  10                  15

Arg Leu Gln Leu Gln Pro Glu Phe His Ala Glu Cys Ser Trp Leu Lys
            20                  25                  30

Ser Ser Ser Lys His Ala Pro Phe Thr Leu Ser Cys Gln Ile Arg Pro
        35                  40                  45

Lys Gln Leu Ser Gln Ile Ala Glu Leu Arg Val Thr Ser Leu Asp Ala
    50                  55                  60

Ser Gln Ala Ser Glu Lys Asp Ile Ser Leu Val Gln Thr Pro His Lys
65                  70                  75                  80

Val Glu Val Asn Glu Lys Ile Glu Glu Ser Ile Glu Tyr Val Gln Asn
                85                  90                  95

Leu Leu Met Thr Ser Gly Asp Gly Arg Ile Ser Val Ser Pro Tyr Asp
            100                 105                 110

Thr Ala Val Ile Ala Leu Ile Lys Asp Leu Lys Gly Arg Asp Ala Pro
        115                 120                 125

Gln Phe Pro Ser Cys Leu Glu Trp Ile Ala His His Gln Leu Ala Asp
    130                 135                 140
```

```
Gly Ser Trp Gly Asp Glu Phe Phe Cys Ile Tyr Asp Arg Ile Leu Asn
145                 150                 155                 160

Thr Leu Ala Cys Val Val Ala Leu Lys Ser Trp Asn Leu Gln Ser Asp
                165                 170                 175

Ile Ile Glu Lys Gly Val Thr Tyr Ile Lys Glu Asn Val His Lys Leu
            180                 185                 190

Lys Gly Ala Asn Val Glu His Arg Thr Ala Gly Phe Glu Leu Val Val
        195                 200                 205

Pro Thr Phe Met Gln Met Ala Thr Asp Leu Gly Ile Gln Gly Leu Pro
    210                 215                 220

Tyr Asp His Pro Leu Ile Lys Glu Ile Ala Asp Thr Lys Gln Gln Arg
225                 230                 235                 240

Leu Lys Glu Ile Pro Lys Asp Leu Val Tyr Gln Met Pro Thr Asn Leu
                245                 250                 255

Leu Tyr Ser Leu Glu Gly Leu Gly Asp Leu Glu Trp Glu Arg Leu Leu
            260                 265                 270

Lys Leu Gln Ser Gly Asn Gly Ser Phe Leu Thr Ser Pro Ser Ser Thr
        275                 280                 285

Ala Ala Val Leu Met His Thr Lys Asp Glu Lys Cys Leu Lys Tyr Ile
    290                 295                 300

Glu Asn Ala Leu Lys Asn Cys Asp Gly Gly Ala Pro His Thr Tyr Pro
305                 310                 315                 320

Val Asp Ile Phe Ser Arg Leu Trp Ala Ile Asp Arg Leu Gln Arg Leu
                325                 330                 335

Gly Ile Ser Arg Phe Phe Gln His Glu Ile Lys Tyr Phe Leu Asp His
            340                 345                 350

Ile Glu Ser Val Trp Glu Glu Thr Gly Val Phe Ser Gly Arg Tyr Thr
        355                 360                 365

Lys Phe Ser Asp Ile Asp Asp Thr Ser Met Gly Val Arg Leu Leu Lys
    370                 375                 380
```

```
Met His Gly Tyr Asp Val Asp Pro Asn Val Leu Lys His Phe Lys Gln
385                 390                 395                 400

Gln Asp Gly Lys Phe Ser Cys Tyr Ile Gly Gln Ser Val Glu Ser Ala
                405                 410                 415

Ser Pro Met Tyr Asn Leu Tyr Arg Ala Ala Gln Leu Arg Phe Pro Gly
            420                 425                 430

Glu Glu Val Leu Glu Glu Ala Thr Lys Phe Ala Phe Asn Phe Leu Gln
        435                 440                 445

Glu Met Leu Val Lys Asp Arg Leu Gln Glu Arg Trp Val Ile Ser Asp
    450                 455                 460

His Leu Phe Asp Glu Ile Lys Leu Gly Leu Lys Met Pro Trp Tyr Ala
465                 470                 475                 480

Thr Leu Pro Arg Val Glu Ala Ala Tyr Tyr Leu Asp His Tyr Ala Gly
                485                 490                 495

Ser Gly Asp Val Trp Ile Gly Lys Ser Phe Tyr Arg Met Pro Glu Ile
            500                 505                 510

Ser Asn Asp Thr Tyr Lys Glu Leu Ala Ile Leu Asp Phe Asn Arg Cys
        515                 520                 525

Gln Thr Gln His Gln Leu Glu Trp Ile Gln Met Gln Glu Trp Tyr Asp
    530                 535                 540

Arg Cys Ser Leu Ser Glu Phe Gly Ile Ser Lys Arg Glu Leu Leu Arg
545                 550                 555                 560

Ser Tyr Phe Leu Ala Ala Ala Thr Ile Phe Glu Pro Glu Arg Thr Gln
                565                 570                 575

Glu Arg Leu Leu Trp Ala Lys Thr Arg Ile Leu Ser Lys Met Ile Thr
            580                 585                 590

Ser Phe Val Asn Ile Ser Gly Thr Thr Leu Ser Leu Asp Tyr Asn Phe
        595                 600                 605

Asn Gly Leu Asp Glu Ile Ile Ser Ala Asn Glu Asp Gln Gly Leu Ala
    610                 615                 620
```

```
Gly Thr Leu Leu Ala Thr Phe His Gln Leu Leu Asp Gly Phe Asp Ile
625                 630                 635                 640

Tyr Thr Leu His Gln Leu Lys His Val Trp Ser Gln Trp Phe Met Lys
                645                 650                 655

Val Gln Gln Gly Glu Gly Ser Gly Gly Glu Asp Ala Val Leu Leu Ala
            660                 665                 670

Asn Thr Leu Asn Ile Cys Ala Gly Leu Asn Glu Asp Val Leu Ser Asn
        675                 680                 685

Asn Glu Tyr Thr Ala Leu Ser Thr Leu Thr Asn Lys Ile Cys Asn Arg
    690                 695                 700

Leu Ala Gln Ile Gln Asp Asn Lys Ile Leu Gln Val Val Asp Gly Ser
705                 710                 715                 720

Ile Lys Asp Lys Glu Leu Glu Gln Asp Met Gln Ala Leu Val Lys Leu
                725                 730                 735

Val Leu Gln Glu Asn Gly Gly Ala Val Asp Arg Asn Ile Arg His Thr
            740                 745                 750

Phe Leu Ser Val Ser Lys Thr Phe Tyr Tyr Asp Ala Tyr His Asp Asp
        755                 760                 765

Glu Thr Thr Asp Leu His Ile Phe Lys Val Leu Phe Arg Pro Val Val
    770                 775                 780
```

<210> 26
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<220>
<221> misc_feature
<222> (9) .. (9)
<223> n es a, c, g, o t

<220>
<221> misc_feature
<222> (12) .. (12)
<223> n es a, c, g, o t

<220>
<221> misc_feature
<222> (15) .. (15)
<223> n es a, c, g, o t

<400> 26

tatgatacng cngtnatdgc 20

<210> 27
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 27

tatgacacgg cagtgatcgc 20

<210> 28
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<220>
<221> misc_feature
<222> (18) .. (18)
<223> n es a, c, g, o t

<400> 28

tatgacacgg cakkgrtngc 20

<210> 29
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<220>
<221> misc_feature
<222> (15) .. (15)
<223> n es a, c, g, o t

<220>
<221> misc_feature
<222> (18) .. (18)
<223> n es a, c, g, o t

<400> 29

caactggctg atggntcntg ggg 23

<210> 30
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 30

caactggctg atggctcatg ggg 23

<210> 31

<211> 26
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 31

gatcctccaa crtcrwarar rtcrtc 26

<210> 32
<211> 25
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 32

gatcctccac gtcgwagaag tcgtc 25

<210> 33
<211> 81
<212> ADN
<213> *Salvia sclarea*

<400> 33

ctgatgtttc tgtctacggc gccgccattt tcttgaagca ctaacttcac caacgcctgc 60

atatcctgtt ctagctcctt a 81

<210> 34
<211> 73
<212> ADN
<213> *Salvia sclarea*

<400> 34

attcctgcat atggatccac tccaactgat gttgtgtttg gcatctgttg aaatccaata 60

tcgcaagctc ctt 73

<210> 35
<211> 52
<212> ADN
<213> *Salvia sclarea*

<400> 35

tattattgaa aaaggagtga cgtacatcaa ggagaatgtg cataaactta aa 52

<210> 36
<211> 857
<212> ADN
<213> *Salvia sclarea*

<400> 36

atgtcaccac aaacagagac taaagcaagt gttggattca aagcgggtgt taaagagtac 60

aaattgactt attatactcc tgaatacgaa accaaagata ctgatatctt ggcagcattc 120

cgagtaactc ctcaacccgg agttccgcct gaagaagcag gggccgcggt agctgccgaa 180

tcttctactg gtacatggac aactgtgtgg accgatggac ttaccagcct tgatcgttac 240

aaagggcgat gctaccacat tgagcccgtt cctggagaaa aagatcaata tatctgttat 300

gtagcttacc ctttagacct ttttgaagaa ggttctgtta ctaacatgtt tacttccatt 360

gtaggaaatg tatttggatt caaagcccta cgtgctctac gtctggaaga tctgcgaatt 420

cctgttgctt atgttaaaac tttccaaggc ccgcctcatg ggatccaagt tgagagagat 480

aaattgaaca agtacggtcg tcctctgctg ggatgtacta ttaaacctaa attggggtta 540

tctgctaaaa actatggtag agcggtttat gaatgtcttc gcggtggact tgattttacc 600

aaagatgatg agaacgtgaa ctcccagcca tttatgcgtt ggagagaccg cttcttattt 660

tgtgccgaag caatttataa agcacaggct gaaacaggtg aaatcaaagg gcattacttg 720

aatgctactg cgggtacatg cgaagagatg atgaaaagag ctatatttgc tagagaattg 780

ggagttccta tcgtaatgca cgactactta acaggaggat tcaccgcaaa taccagtttg 840

gctcattatt gccgaga 857

<210> 37
<211> 540
<212> ADN
<213> *Salvia sclarea*

<400> 37

aaagtatcca ctgctttaaa ttcaaacttg atttctttcc atacctcaca agcggcagct 60

agttcaggac tccatttgca agcttcacgg ataattgcat taccttcagc agcaagatca 120

cgtccttcat tacgagcttt tacacacgct tctacagcta ctcggttagc tacagcacct 180

ggtgcattac cccaagggtg tcctaaagtt cctccaccga actgtagtac ggaatcgtct 240

ccaaagatct cggtcagagc aggcatatgc caaacgtgaa taccccctga agccacagga 300

ataacacccg gcagggagac ccaatcttga gtgaaataaa taccgcgact tcggtctttt 360

tcaataaaat catcacgcag taaatcaaca aaacctaaag taatgtctct ctctccttca 420

agtttaccta ctacggtacc agagtgaata tgatctccac cggacagacg taacgcttta 480

gctagtacac ggaagtgcat accgtgattc ttctgtctat caataactgc atgcattgca 540

<210> 38
<211> 60
<212> ADN
<213> *Salvia sclarea*

<400> 38

atgttcgtct cgcatttgcc aaaacatctg tgcttgtaac aattatggat gatttttcg 60

<210> 39
<211> 35
<212> ADN
<213> *Salvia sclarea*

<400> 39

cttctacctt ggcctgcatt cttgctctta aaaaa 35

<210> 40
<211> 35
<212> ADN
<213> *Salvia sclarea*

<400> 40

aaaaaaaata tgatctaaaa aatggatcag tttaa 35

<210> 41
<211> 72
<212> ADN
<213> *Salvia sclarea*

<400> 41

actactcatg caatggcatt taggcttttg cgagtgaaag gatacgaagt ttcatcagag 60

gagttggcct ca 72

<210> 42
<211> 35
<212> ADN
<213> *Salvia sclarea*

<400> 42

gcaactgatg attttgtgga tgttgggggc agctc 35

<210> 43
<211> 35
<212> ADN
<213> *Salvia sclarea*

<400> 43

gccaaaataa ttccttgcat ggctttggaa ggaga 35

<210> 44
<211> 11
<212> PRT
<213> *Salvia sclarea*

<400> 44

Ser Thr Leu Ala Cys Ile Leu Ala Leu Lys Lys
1 5 10

<210> 45
<211> 10
<212> PRT
<213> *Salvia sclarea*

<400> 45

Lys Lys Tyr Asp Leu Lys Asn Gly Ser Val
1 5 10

<210> 46
<211> 24
<212> PRT
<213> *Salvia sclarea*

<400> 46

Thr Thr His Ala Met Ala Phe Arg Leu Leu Arg Val Lys Gly Tyr Glu
1 5 10 15

Val Ser Ser Glu Glu Leu Ala Ser
20

<210> 47
<211> 11
<212> PRT
<213> *Salvia sclarea*

<400> 47

Ala Thr Asp Asp Phe Val Asp Val Gly Gly Ser
1 5 10

<210> 48
<211> 11
<212> PRT
<213> *Salvia sclarea*

<400> 48

Leu Leu Pro Lys Pro Cys Lys Glu Leu Phe Trp
1 5 10

<210> 49
<211> 27
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 49

cttctacctt ggcctgcatt cttgctc 27

<210> 50
<211> 27
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 50

gagcaagaat gcaggccaag gtagaag 27

<210> 51
<211> 31
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 51

ctactcatgc aatggcattt aggcttttgc g 31

<210> 52
<211> 33
<212> ADN
<213> *Salvia sclarea*

<400> 52

gtgaaaggat acgaagtttc atcagaggag ttg 33

<210> 53
<211> 32
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 53

gaggccaact cctctgatga aacttcgtat cc 32

<210> 54
<211> 30
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 54

cactcgcaaa agcctaaatg ccattgcatg 30

<210> 55
<211> 29
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 55

ctgatgattt tgtggatgtt ggggcagc 29

<210> 56

<211> 29
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 56

gctgccccca acatccacaa aatcatcag 29

<210> 57
<211> 33
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 57

ctccttccaa agccatgcaa ggaattattt tgg 33

<210> 58
<211> 33
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 58

ccaaaataat tccttgcatg gctttggaag gag 33

<210> 59
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 59

gcgtatctga ttcctgccct ttgc 24

<210> 60
<211> 28
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 60

gtttaattcc ataggggatt tcttcaag 28

<210> 61
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 61

atgtcgcttc ctctctccac ttgc 24

<210> 62
<211> 32
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 62

cgaatgggtt tctcttttat atatagatac ag 32

<210> 63
<211> 28
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 63

ttatatcctt gctcctgttt gttccttg 28

<210> 64
<211> 32
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 64

tcaaaagaca aaggatttca tatcttcctt gg 32

<210> 65
<211> 26
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 65

caccctaata tcacccgtac caaacg 26

<210> 66
<211> 36
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 66

cctatagtgt caaaagacaa aggatttcat atcttc 36

<210> 67
<211> 31
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 67

gcttattgag aaagatggta gtccaagcaa g 31

<210> 68
<211> 32
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 68

gctctataaa gctcaataat catcggcagg tc 32

<210> 69
<211> 39
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 69

ccacaacctg aatgtgttat ctagaatagt gttgatttc 39

<210> 70
<211> 32
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 70

ggaagaattg atcgaccccc aacctttgaa gg 32

<210> 71
<211> 34
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 71

ggcaatgtta ctactcatgc aatggcattt aggc 34

<210> 72
<211> 34
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 72

gcctaaatgc cattgcatga gtagtaacat tgcc 34

<210> 73
<211> 28
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 73

ggtggtcaca ggaaatcctt ggactgga 28

<210> 74
<211> 25
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 74

cctgctccca actctttggc cggag 25

<210> 75
<211> 28
<212> ADN
<213> Secuencia artificial

<400> 75

cgtaactccg acgttgaagg cgagcgac 28

<210> 76
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 76

caccatgtcg ctcgccttca acg 23

<210> 77
<211> 33
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 77

caccatggcg aaaatgaaag agaatttcaa gag 33

<210> 78
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 78

caccatgtcg cttcctctct ccac 24

<210> 79
<211> 27
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 79

caccatggaa actgggcttc aaactgc 27

<210> 80
<211> 30
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador

<400> 80

ttatatcctt gctcctgttt gttccttgag 30

<210> 81
<211> 547
<212> ADN
<213> *Salvia sclarea*

<400> 81

agcggccgct gaattctaga atttggatgt ggaatagaat tgacgaatgg cgtgcaagcg 60

cgcctctctc ctctctcctc tctctagaaa atatgattgt gcagttgagt tggcaaaagc 120

gtatctgatt cctgcccttt gctaactttc ccaaattttg tcccgtttaa ttccataggg 180

gatttcttca aggccgccat gtcgcttcct ctctccactt gcaatggatc acattttcgg 240

agataccgct tgtctcctgc ttcagcagct tctatggaaa ctgggcttca aactgctact 300

tcagcaaaaa tcgcctctat gccagcgtgc tttgaggaga cgagagggag gatagcaaag 360

ttgtttcata aggatgaact ttctgtgtcg acatatgata cagcatgggt tgccatggtc 420

ccttctccaa cttcgttaga ggaaccttgc ttccccgatt gtctaaactg gttgctcgag 480

aaccagtgcc atgatggttc gtgggcccgt ccccaccatc actctttgct aatgaaagat 540

gtccttt 547

<210> 82
<211> 1473
<212> ADN
<213> *Salvia sclarea*

<400> 82

ctccatatgg taaccaagag gctgttagcc agcaaacaaa tgacctgccg atgattattg 60

agctttatag agcagcaaat gagagaatat atgaagaaga gaggagtctt gaaaaaattc 120

ttgcttggac taccatcttt ctcaataagc aagtgcaaga taactcaatt cccgacaaaa 180

aactgcacaa actggtggaa ttctacttga ggaattacaa aggcataacc ataagattgg 240

gagctagacg aaacctcgag ctatatgaca tgacctacta tcaagctctg aaatctacaa 300

acaggttctc taatttatgc aacgaagatt ttctagtttt cgcaaagcaa gatttcgata 360

tacatgaagc ccagaaccag aaaggacttc aacaactgca aaggtggtat gcagattgta 420

ggttggacac cttaaacttt ggaagagatg tagttattat tgctaattat ttggcttcat 480

taattattgg tgatcatgcg tttgactatg ttcgtctcgc atttgccaaa acatctgtgc 540

ttgtaacaat tatggatgat tttttcgact gtcatggctc tagtcaagag tgtgacaaga 600

tcattgaatt agtaaaagaa tggaaggaga atccggatgc agagtacgga tctgaggagc 660

ttgagatcct ttttatggcg ttgtacaata cagtaaatga gttggcggag agggctcgtg 720

ttgaacaggg gcgtagtgtc aaagagtttc tagtcaaact gtgggttgaa atactctcag 780

ctttcaagat agaattagat acatggagca atggcacgca gcaaagcttc gatgaataca 840

tttcttcgtc gtggttgtcg aacggttccc ggctgacagg tctcctgacg atgcaattcg 900

tcggagtaaa attgtccgat gaaatgctta tgagtgaaga gtgcactgat ttggctaggc 960

atgtctgtat ggtcggccgg ctgctcaacg acgtgtgcag ttctgagagg gagcgcgagg 1020

aaaatattgc aggaaaaagt tatagcattc tactagcaac tgagaaagat ggaagaaaag 1080

ttagtgaaga tgaagccatt gcagagatca atgaaatggt tgaatatcac tggagaaaag 1140

tgttgcagat tgtgtataaa aaagaaagca ttttgccaag aagatgcaaa gatgtatttt 1200

tggagatggc taagggtacg ttttatgctt atgggatcaa cgatgaattg acttctcctc 1260

agcaatccaa ggaagatatg aaatcctttg tcttttgaca ctataggctc gtttggtacg 1320

ggtgatatta gggtgtgtaa tacaattatg acactgtaat atttttattt gtacaaaaca 1380

cgtggttctt tgcatatcaa aaatttgaaa atgttataag gatttgtatc cactataaga 1440

aattgttgga taaaaaaaaa aaaaaaaaaa aaa 1473

<210> 83
<211> 207
<212> ADN
<213> *Salvia sclarea*

<400> 83

cacatgtgct ccgtggtcca tctattctac agtaaagacg acggattcac ctcgcaggat 60

ttgattcaag ttgtaaatgc aattattcac aaacctattg tcctcaagga acaaacagga 120

gcaaggatat aatttttta atctgtatct atatataaaa gagaaaccca ttcgttaaaa 180

taaaaaaaaa aaaaaaaaaa aaaaaaa 207

<210> 84
<211> 464
<212> ADN
<213> *Salvia sclarea*

<400> 84

aaattaatta ggaataaaaa aaattggact ttatatttat tagaaacggc cgccgccgca 60

aaaaaatgtc gctcgccttc aacgtcggag ttacgccttt ctccggccaa agagttggga 120

gcaggaaaga aaaatttcca gtccaaggat ttcctgtgac caccccaat aggtcacgtc 180

tcatcgttaa ctgcagcctt actacaatag atttcatggc gaaaatgaaa gagaatttca 240

agagggaaga cgataaattt ccaacgacaa cgactcttcg atccgaagat ataccctcta 300

atttgtgtat aatcgacacc cttcaaaggt tgggggtcga tcaattcttc caatatgaaa 360

tcaacactat tctagataac acattcaggt tgtggcaaga aaaacacaaa gttatatatg 420

gcaatgttac tactcatgca atggcattta ggcttttgcg agtg 464

<210> 85
<211> 2388
<212> ADN
<213> *Salvia sclarea*

<400> 85

atgtcgcttc ctctctccac ttgcaatgga tcacattttc ggagataccg cttgtctcct 60

gcttcagctt ctatggaaac tgggcttcaa actgctactt cagcaaaaat cgcctctatg 120

ccagcgtgct ttgaggagac gagagggagg atagcaaagt tgtttcataa ggatgaactt 180

tctgtgtcga catatgatac agcatgggtt gccatggtcc cttctccaac ttcgttagag 240

gaaccttgct tccccgattg tctaaactgg ttgctcgaga accagtgcca tgatggttcg 300

tgggcccgtc cccaccatca ctctttgcta atgaaagatg tcctttcttc taccttggcc 360

tgcattcttg ctcttaaaaa atggggagtt ggtgaaaaac agattaacag gggcttgcat 420

tttatggagt tgaattttgc ttcagctact gagaagtgtc agattactcc catgggattt 480

gatattgtat ttcctgccat gcttgattat gccagagact tctctttgga catgcattta 540

gagccaacta cgttgaatga tttgatacat aagagggatt tggagcttaa aagcaagcca 600

gatttttcat cggatgggga agcctattgg gcatatatag ctgaaggaat ggggaattta 660

cggaactggg aatcagttat gaaatatcaa agaaggaatg gatctctttt caactgtcct 720

tccacgacag cagctgcttt tgttgcactg ggcaattctg actgcctcaa ctacctgcat 780

tcagccttaa agaagtttgg gaatgcagtt cctgcagttt atcctctaga tatatattct 840

cacctgtgca tagttgacaa tcttgaaagg ttggggatca gccgttattt tttgactgag 900

attcaaagcg tgttagatga aacacacaga tgttggatgc agggcaatga agagatcttc 960

atggatgcct caacttgtgc tttagctttc cggatattgc gattgaacgg atacgatgta 1020

acttcagatc cggttacaaa aattcaacac gagtgctttt cgagttcctt tcatggaaat 1080

gtgatggaca ttaacacgac tcttgaatta tatagggcat ctgaactcat actatatcca 1140

gatgaaagag atctagtgag acaaaattta aggcttaaac aaatactaga gcaagagcta 1200

tccaatggtt ttattcaatc atgtcaactt ggaagaagtg ttaatgcaga ggtgaaccag 1260

gctatcgagt atccatttta tgcaattatg gacagggttg caaaacggaa aaatatagag 1320

aactacaact ttgataatac aagaattctg aaaacttcat attgttcacc aaattttggc 1380

aacaaggatt ttctttttct gtccgtagag gacttcaatc tgtgtcaagc cacacatcgc 1440

gaagaactca gggaacttga aagatgggtc gtagagaata gattggacga gctgcagttt 1500

gcaaggagta agtctgcata ttgttatttt tctgcagcag caacctttc tgctccagaa 1560

ctacgtgatg cacgcatgtc gtgggccaaa ggtggtgttc tgactacagt gattgatgac 1620

ttttttgacg tcggaggttc tatggaagaa ttgaagaact taattcattt ggttgaaaaa 1680

tgggatgtgg atgttagcac agaatgctct tcccataatg tccagataat attttcagca 1740

cttaagagca caatccgtga aattggatac aaagggttga agctacaagg gcgttgtatt 1800

actaaccata taattggcat ttggttagat ttgctgaatt ctatgatgaa agaaactgaa 1860

tgggctagag acaactatgt cccaacaatt gatgaatata tgagcaatgc atatgtgtca 1920

tttgctctgg ggccaattgt tctgccaact ctatatcttg ttgggccgaa gctctcagaa 1980

gagatggcaa accaccccga gtactataaa ctattcaaat tgatgagcac atgcggacgc 2040

cttttaaatg acatccgtgg ttatgagaga gaactcaaag atggtaaact gaacgcgcta 2100

tctttgtaca tggctaatca tggtggtgaa gtaagtaaag aagcagccat ttcagagatc 2160

aaaagttgga ttgagagcag taggagagaa ttactgagat tagttttgga ggggaagaag 2220

agtgtccttc caaagccatg caaggaatta ttttggcaca tgtgctctgt ggtccatcta 2280

ttctacagta aagacgacgg attcacctcg caggatttga ttcaagttgt aaatgcaatt 2340

attcacaaac ctattgtcct caaggaacaa acaggagcaa ggatataa 2388

<210> 86
<211> 795
<212> PRT
<213> *Salvia sclarea*

<400> 86

```
Met Ser Leu Pro Leu Ser Thr Cys Asn Gly Ser His Phe Arg Arg Tyr
1               5                  10                  15

Arg Leu Ser Pro Ala Ser Ala Ser Met Glu Thr Gly Leu Gln Thr Ala
            20                  25                  30

Thr Ser Ala Lys Ile Ala Ser Met Pro Ala Cys Phe Glu Glu Thr Arg
        35                  40                  45

Gly Arg Ile Ala Lys Leu Phe His Lys Asp Glu Leu Ser Val Ser Thr
    50                  55                  60

Tyr Asp Thr Ala Trp Val Ala Met Val Pro Ser Pro Thr Ser Leu Glu
```

65 70 75 80

Glu Pro Cys Phe Pro Asp Cys Leu Asn Trp Leu Leu Glu Asn Gln Cys
85 90 95

His Asp Gly Ser Trp Ala Arg Pro His His His Ser Leu Leu Met Lys
100 105 110

Asp Val Leu Ser Ser Thr Leu Ala Cys Ile Leu Ala Leu Lys Lys Trp
115 120 125

Gly Val Gly Glu Lys Gln Ile Asn Arg Gly Leu His Phe Met Glu Leu
130 135 140

Asn Phe Ala Ser Ala Thr Glu Lys Cys Gln Ile Thr Pro Met Gly Phe
145 150 155 160

Asp Ile Val Phe Pro Ala Met Leu Asp Tyr Ala Arg Asp Phe Ser Leu
165 170 175

Asp Met His Leu Glu Pro Thr Thr Leu Asn Asp Leu Ile His Lys Arg
180 185 190

Asp Leu Glu Leu Lys Ser Lys Pro Asp Phe Ser Ser Asp Gly Glu Ala
195 200 205

Tyr Trp Ala Tyr Ile Ala Glu Gly Met Gly Asn Leu Arg Asn Trp Glu
210 215 220

Ser Val Met Lys Tyr Gln Arg Arg Asn Gly Ser Leu Phe Asn Cys Pro
225 230 235 240

Ser Thr Thr Ala Ala Ala Phe Val Ala Leu Gly Asn Ser Asp Cys Leu
245 250 255

Asn Tyr Leu His Ser Ala Leu Lys Lys Phe Gly Asn Ala Val Pro Ala
260 265 270

Val Tyr Pro Leu Asp Ile Tyr Ser His Leu Cys Ile Val Asp Asn Leu
275 280 285

Glu Arg Leu Gly Ile Ser Arg Tyr Phe Leu Thr Glu Ile Gln Ser Val
290 295 300

Leu Asp Glu Thr His Arg Cys Trp Met Gln Gly Asn Glu Glu Ile Phe
305 310 315 320

```
Met Asp Ala Ser Thr Cys Ala Leu Ala Phe Arg Ile Leu Arg Leu Asn
                325                 330                 335

Gly Tyr Asp Val Thr Ser Asp Pro Val Thr Lys Ile Gln His Glu Cys
            340                 345                 350

Phe Ser Ser Ser Phe His Gly Asn Val Met Asp Ile Asn Thr Thr Leu
        355                 360                 365

Glu Leu Tyr Arg Ala Ser Glu Leu Ile Leu Tyr Pro Asp Glu Arg Asp
    370                 375                 380

Leu Val Arg Gln Asn Leu Arg Leu Lys Gln Ile Leu Glu Gln Glu Leu
385                 390                 395                 400

Ser Asn Gly Phe Ile Gln Ser Cys Gln Leu Gly Arg Ser Val Asn Ala
                405                 410                 415

Glu Val Asn Gln Ala Ile Glu Tyr Pro Phe Tyr Ala Ile Met Asp Arg
            420                 425                 430

Val Ala Lys Arg Lys Asn Ile Glu Asn Tyr Asn Phe Asp Asn Thr Arg
        435                 440                 445

Ile Leu Lys Thr Ser Tyr Cys Ser Pro Asn Phe Gly Asn Lys Asp Phe
    450                 455                 460

Leu Phe Leu Ser Val Glu Asp Phe Asn Leu Cys Gln Ala Thr His Arg
465                 470                 475                 480

Glu Glu Leu Arg Glu Leu Glu Arg Trp Val Val Glu Asn Arg Leu Asp
                485                 490                 495

Glu Leu Gln Phe Ala Arg Ser Lys Ser Ala Tyr Cys Tyr Phe Ser Ala
            500                 505                 510

Ala Ala Thr Phe Ser Ala Pro Glu Leu Arg Asp Ala Arg Met Ser Trp
        515                 520                 525

Ala Lys Gly Gly Val Leu Thr Thr Val Ile Asp Asp Phe Phe Asp Val
    530                 535                 540

Gly Gly Ser Met Glu Glu Leu Lys Asn Leu Ile His Leu Val Glu Lys
545                 550                 555                 560
```

```
Trp Asp Val Asp Val Ser Thr Glu Cys Ser Ser His Asn Val Gln Ile
                565                 570                 575

Ile Phe Ser Ala Leu Lys Ser Thr Ile Arg Glu Ile Gly Tyr Lys Gly
            580                 585                 590

Leu Lys Leu Gln Gly Arg Cys Ile Thr Asn His Ile Ile Gly Ile Trp
        595                 600                 605

Leu Asp Leu Leu Asn Ser Met Met Lys Glu Thr Glu Trp Ala Arg Asp
    610                 615                 620

Asn Tyr Val Pro Thr Ile Asp Glu Tyr Met Ser Asn Ala Tyr Val Ser
625                 630                 635                 640

Phe Ala Leu Gly Pro Ile Val Leu Pro Thr Leu Tyr Leu Val Gly Pro
                645                 650                 655

Lys Leu Ser Glu Glu Met Ala Asn His Pro Glu Tyr Tyr Lys Leu Phe
            660                 665                 670

Lys Leu Met Ser Thr Cys Gly Arg Leu Leu Asn Asp Ile Arg Gly Tyr
        675                 680                 685

Glu Arg Glu Leu Lys Asp Gly Lys Leu Asn Ala Leu Ser Leu Tyr Met
    690                 695                 700

Ala Asn His Gly Gly Glu Val Ser Lys Glu Ala Ala Ile Ser Glu Ile
705                 710                 715                 720

Lys Ser Trp Ile Glu Ser Ser Arg Arg Glu Leu Leu Arg Leu Val Leu
                725                 730                 735

Glu Gly Lys Lys Ser Val Leu Pro Lys Pro Cys Lys Glu Leu Phe Trp
            740                 745                 750

His Met Cys Ser Val Val His Leu Phe Tyr Ser Lys Asp Asp Gly Phe
        755                 760                 765

Thr Ser Gln Asp Leu Ile Gln Val Val Asn Ala Ile Ile His Lys Pro
    770                 775                 780

Ile Val Leu Lys Glu Gln Thr Gly Ala Arg Ile
785                 790                 795
```

<210> 87
<211> 536
<212> ADN
<213> *Salvia sclarea*

<400> 87

aaattaatta ggaataaaaa aaattggact ttatatttat tagaaacggc cgccgccgca 60

aaaaaatgtc gctcgccttc aacgtcggag ttacgccttt ctccggccaa agagttggga 120

gcaggaaaga aaaatttcca gtccaaggat ttcctgtgac caccccaat aggtcacgtc 180

tcatcgttaa ctgcagcctt actacaatag atttcatggc gaaaatgaaa gagaatttca 240

agagggaaga cgataaattt ccaacgacaa cgactcttcg atccgaagat ataccctcta 300

atttgtgtat aatcgacacc cttcaaaggt tgggggtcga tcaattcttc caatatgaaa 360

tcaacactat tctagataac acattcaggt tgtggcaaga aaaacacaaa gttatatatg 420

gcaatgttac tactcatgca atggcattta ggcttttgcg agtgaaagga tacgaagttt 480

catcagagga gttggctcca tatggtaacc aagaggctgt tagccagcaa acaaat 536

<210> 88
<211> 95
<212> ADN
<213> *Salvia sclarea*

<400> 88

ttgcatcttc ttggcaaaat gctttctttt ttatacacaa tctgcaacac ttttctccag 60

tgatattcaa ccatttcatt gatctctgca atggc 95

<210> 89
<211> 55
<212> ADN
<213> *Salvia sclarea*

<400> 89

tgaagcccag aaccagaaag gacttcaaca actgcaaagg tggtatgcag attgt 55

<210> 90
<211> 53
<212> ADN
<213> *Salvia sclarea*

<400> 90

tactctgcat ccggattctc cttccattct tttactaatt caatgatctt gtc 53

<210> 91
<211> 2391
<212> ADN
<213> *Salvia sclarea*

<400> 91

atgtcgcttc ctctctccac ttgcaatgga tcacattttc ggagataccg cttgtctcct 60

gcttcagctt ctatggaaac tgggcttcaa actgctactt cagcaaaaat cgcctctatg 120

ccagcgtgct ttgaggagac gagagggagg atagcaaagt tgtttcataa ggatgaactt 180

tctgtgtcga catatgatac agcatgggtt gccatggtcc cttctccaac ttcgttagag 240

gaaccttgct tccccgattg tctaaactgg ttgctcgaga accagtgcca tgatggttcg 300

tgggcccgtc cccaccatca ctctttgcta atgaaagatg tcctttcttc taccttggcc 360

tgcattcttg ctcttaaaaa atggggagtt ggtgaaaaac agattaacag gggcttgcat 420

tttatggagt tgaattttgc ttcagctact gagaagtgtc agattactcc catgggattt 480

gatattgtat ttcctgccat gcttgattat gccagagact tctctttgga catgcattta 540

gagccaacta cgttgaatga tttgatacat aagagggatt tggagcttaa aagcaagcca 600

gatttttcat cggatgggga agcctattgg gcatatatag ctgaaggaat ggggaattta 660

cggaactggg aatcagttat gaaatatcaa agaaggaatg gatctctttt caactgtcct 720

tccacgacag cagctgcttt tgttgcactg ggcaattctg actgcctcaa ctacctgcat 780

tcagccttaa agaagtttgg gaatgcagtt cctgcagttt atcctctaga tatatattct 840

cacctgtgca tagttgacaa tcttgaaagg ttggggatca gccgttattt tttgactgag 900

attcaaagcg tgttagatga aacacacaga tgttggatgc agggcaatga agagatcttc 960

atggatgcct caacttgtgc tttagctttc cggatattgc gattgaacgg atacgatgta 1020

acttcagatc cggttacaaa aattcaacac gagtgctttt cgagttcctt tcatggaaat 1080

gtgatggaca ttaacacgac tcttgaatta tatagggcat ctgaactcat actatatcca 1140

gatgaaagag atctagtgag acaaaattta aggcttaaac aaatactaga gcaagagcta 1200

tccaatggtt ttattcaatc atgtcaactt ggaagaagtg ttaatgcaga ggtgaaccag 1260

gctatcgagt atccatttta tgcaattatg gacagggttg caaaacggaa aaatatagag 1320

aactacaact ttgataatac aagaattctg aaaacttcat attgttcacc aaattttggc 1380

aacaaggatt ttctttttct gtccgtagag gacttcaatc tgtgtcaagc cacacatcgc 1440

gaagaactca gggaacttga aagatgggtc gtagagaata gattggacga gctgcagttt 1500

gcaaggagta agtctgcata ttgttatttt tctgcagcag caacctttc tgctccagaa 1560

ctacgtgatg cacgcatgtc gtgggccaaa ggtggtgttc tgactacagt gattgatgac 1620

ttttttgacg tcggaggttc tatggaagaa ttgaagaact taattcattt ggttgaaaaa 1680

tgggatgtgg atgttagcac agaatgctct tcccataatg tccagataat attttcagca 1740

cttaagagca caatccgtga aattggatac aaagggttga agctacaagg gcgttgtatt 1800

actaaccata taattggcat ttggttagat ttgctgaatt ctatgatgaa agaaactgaa 1860

tgggctagag acaactatgt cccaacaatt gatgaatata tgagcaatgc atatgtgtca 1920

tttgctctgg ggccaattgt tctgccaact ctatatcttg ttgggccgaa gctctcagaa 1980

gagatggcaa accaccccga gtactataaa ctattcaaat tgatgagcac atgcggacgc 2040

cttttaaatg acatccgtgg ttatgagaga gaactcaaag atggtaaact gaacgcgcta 2100

tctttgtaca tggctaatca tggtggtgaa gtaagtaaag aagcagccat ttcagagatc 2160

aaaagttgga ttgagagcag taggagagaa ttactgagat tagttttgga ggggaagaag 2220

agtgtccttc caaagccatg caaggaatta ttttggcaca tgtgctctgt ggtccatcta 2280

ttctacagta aagacgacgg attcacctcg caggatttga ttcaagttgt aaatgcaatt 2340

attcacaaac ctattgtcct caaggaacaa acaggagcaa ggatataaaa g 2391

<210> 92
<211> 2316
<212> ADN
<213> *Salvia sclarea*

<400> 92

atggaaactg ggcttcaaac tgctacttca gcaaaaatcg cctctatgcc agcgtgcttt 60

gaggagacga gagggaggat agcaaagttg tttcataagg atgaactttc tgtgtcgaca 120

tatgatacag catgggttgc catggtccct tctccaactt cgttagagga accttgcttc 180

cccgattgtc taaactggtt gctcgagaac cagtgccatg atggttcgtg ggcccgtccc 240

caccatcact ctttgctaat gaaagatgtc ctttcttcta ccttggcctg cattcttgct 300

cttaaaaaat ggggagttgg tgaaaaacag attaacaggg gcttgcattt tatggagttg 360

aattttgctt cagctactga gaagtgtcag attactccca tgggatttga tattgtattt 420

cctgccatgc ttgattatgc cagagacttc tctttggaca tgcatttaga gccaactacg 480

ttgaatgatt tgatacataa gagggatttg gagcttaaaa gcaagccaga tttttcatcg 540

gatggggaag cctattgggc atatatagct gaaggaatgg ggaatttacg gaactgggaa 600

tcagttatga aatatcaaag aaggaatgga tctcttttca actgtccttc cacgacagca 660

gctgcttttg ttgcactggg caattctgac tgcctcaact acctgcattc agccttaaag 720

aagtttggga atgcagttcc tgcagtttat cctctagata tatattctca cctgtgcata 780

gttgacaatc ttgaaaggtt ggggatcagc cgttattttt tgactgagat tcaaagcgtg 840

ttagatgaaa cacacagatg ttggatgcag ggcaatgaag agatcttcat ggatgcctca 900

acttgtgctt tagctttccg gatattgcga ttgaacggat acgatgtaac ttcagatccg 960

gttacaaaaa ttcaacacga gtgcttttcg agttcctttc atggaaatgt gatggacatt 1020

aacacgactc ttgaattata tagggcatct gaactcatac tatatccaga tgaaagagat 1080

ctagtgagac aaaatttaag gcttaaacaa atactagagc aagagctatc caatggtttt 1140

attcaatcat gtcaacttgg aagaagtgtt aatgcagagg tgaaccaggc tatcgagtat 1200

ccattttatg caattatgga cagggttgca aaacggaaaa atatagagaa ctacaacttt 1260

gataatacaa gaattctgaa aacttcatat tgttcaccaa attttggcaa caaggatttt 1320

ctttttctgt ccgtagagga cttcaatctg tgtcaagcca cacatcgcga agaactcagg 1380

gaacttgaaa gatgggtcgt agagaataga ttggacgagc tgcagtttgc aaggagtaag 1440

tctgcatatt gttatttttc tgcagcagca acctttttctg ctccagaact acgtgatgca 1500

cgcatgtcgt gggccaaagg tggtgttctg actacagtga ttgatgactt ttttgacgtc 1560

ggaggttcta tggaagaatt gaagaactta attcatttgg ttgaaaaatg ggatgtggat 1620

gttagcacag aatgctcttc ccataatgtc cagataatat tttcagcact taagagcaca 1680

atccgtgaaa ttggatacaa agggttgaag ctacaagggc gttgtattac taaccatata 1740

attggcattt ggttagattt gctgaattct atgatgaaag aaactgaatg ggctagagac 1800

aactatgtcc caacaattga tgaatatatg agcaatgcat atgtgtcatt tgctctgggg 1860

ccaattgttc tgccaactct atatcttgtt gggccgaagc tctcagaaga gatggcaaac 1920

caccccgagt actataaact attcaaattg atgagcacat gcggacgcct tttaaatgac 1980

atccgtggtt atgagagaga actcaaagat ggtaaactga acgcgctatc tttgtacatg 2040

gctaatcatg gtggtgaagt aagtaaagaa gcagccattt cagagatcaa aagttggatt 2100

gagagcagta ggagagaatt actgagatta gttttggagg ggaagaagag tgtccttcca 2160

aagccatgca aggaattatt ttggcacatg tgctctgtgg tccatctatt ctacagtaaa 2220

gacgacggat tcacctcgca ggatttgatt caagttgtaa atgcaattat tcacaaacct 2280

attgtcctca aggaacaaac aggagcaagg atataa 2316

<210> 93
<211> 1578
<212> ADN
<213> *Salvia sclarea*

<400> 93

atggcgaaaa tgaaagagaa tttcaagagg gaagacgata aatttccaac gacaacgact 60

cttcgatccg aagatatacc ctctaatttg tgtataatcg acacccttca aaggttgggg 120

gtcgatcaat tcttccaata tgaaatcaac actattctag ataacacatt caggttgtgg 180

caagaaaaac acaaagttat atatggcaat gttactactc atgcaatggc atttaggctt 240

ttgcgagtga aaggatacga agtttcatca gaggagttgg ctccatatgg taaccaagag 300

gctgttagcc agcaaacaaa tgacctgccg atgattattg agctttatag agcagcaaat 360

gagagaatat atgaagaaga gaggagtctt gaaaaaattc ttgcttggac taccatcttt 420

ctcaataagc aagtgcaaga taactcaatt cccgacaaaa aactgcacaa actggtggaa 480

ttctacttga ggaattacaa aggcataacc ataagattgg gagctagacg aaacctcgag 540

ctatatgaca tgacctacta tcaagctctg aaatctacaa acaggttctc taatttatgc 600

aacgaagatt ttctagtttt cgcaaagcaa gatttcgata tacatgaagc ccagaaccag 660

aaaggacttc aacaactgca aaggtggtat gcagattgta ggttggacac cttaaacttt 720

ggaagagatg tagttattat tgctaattat ttggcttcat taattattgg tgatcatgcg 780

tttgactatg ttcgtctcgc atttgccaaa acatctgtgc ttgtaacaat tatggatgat 840

tttttcgact gtcatggctc tagtcaagag tgtgacaaga tcattgaatt agtaaaagaa 900

tggaaggaga atccggatgc agagtacgga tctgaggagc ttgagatcct ttttatggcg 960

ttgtacaata cagtaaatga gttggcggag agggctcgtg ttgaacaggg gcgtagtgtc 1020

aaagagtttc tagtcaaact gtgggttgaa atactctcag ctttcaagat agaattagat 1080

acatggagca atggcacgca gcaaagcttc gatgaataca tttcttcgtc gtggttgtcg 1140

aacggttccc ggctgacagg tctcctgacg atgcaattcg tcggagtaaa attgtccgat 1200

gaaatgctta tgagtgaaga gtgcactgat ttggctaggc atgtctgtat ggtcggccgg 1260

ctgctcaacg acgtgtgcag ttctgagagg gagcgcgagg aaaatattgc aggaaaaagt 1320

tatagcattc tactagcaac tgagaaagat ggaagaaaag ttagtgaaga tgaagccatt 1380

gcagagatca atgaaatggt tgaatatcac tggagaaaag tgttgcagat tgtgtataaa 1440

aaagaaagca ttttgccaag aagatgcaaa gatgtatttt tggagatggc taagggtacg 1500

ttttatgctt atgggatcaa cgatgaattg acttctcctc agcaatccaa ggaagatatg 1560

aaatcctttg tcttttga 1578

<210> 94
<211> 795
<212> PRT
<213> *Salvia sclarea*

<400> 94

```
Met Ser Leu Pro Leu Ser Thr Cys Asn Gly Ser His Phe Arg Arg Tyr
1               5                  10                  15

Arg Leu Ser Pro Ala Ser Ala Ser Met Glu Thr Gly Leu Gln Thr Ala
            20                  25                  30

Thr Ser Ala Lys Ile Ala Ser Met Pro Ala Cys Phe Glu Glu Thr Arg
        35                  40                  45
```

```
Gly Arg Ile Ala Lys Leu Phe His Lys Asp Glu Leu Ser Val Ser Thr
    50                  55                  60

Tyr Asp Thr Ala Trp Val Ala Met Val Pro Ser Pro Thr Ser Leu Glu
65                  70                  75                  80

Glu Pro Cys Phe Pro Asp Cys Leu Asn Trp Leu Leu Glu Asn Gln Cys
                85                  90                  95

His Asp Gly Ser Trp Ala Arg Pro His His His Ser Leu Leu Met Lys
            100                 105                 110

Asp Val Leu Ser Ser Thr Leu Ala Cys Ile Leu Ala Leu Lys Lys Trp
        115                 120                 125

Gly Val Gly Glu Lys Gln Ile Asn Arg Gly Leu His Phe Met Glu Leu
    130                 135                 140

Asn Phe Ala Ser Ala Thr Glu Lys Cys Gln Ile Thr Pro Met Gly Phe
145                 150                 155                 160

Asp Ile Val Phe Pro Ala Met Leu Asp Tyr Ala Arg Asp Phe Ser Leu
                165                 170                 175

Asp Met His Leu Glu Pro Thr Thr Leu Asn Asp Leu Ile His Lys Arg
            180                 185                 190

Asp Leu Glu Leu Lys Ser Lys Pro Asp Phe Ser Ser Asp Gly Glu Ala
        195                 200                 205

Tyr Trp Ala Tyr Ile Ala Glu Gly Met Gly Asn Leu Arg Asn Trp Glu
    210                 215                 220

Ser Val Met Lys Tyr Gln Arg Arg Asn Gly Ser Leu Phe Asn Cys Pro
225                 230                 235                 240

Ser Thr Thr Ala Ala Ala Phe Val Ala Leu Gly Asn Ser Asp Cys Leu
                245                 250                 255

Asn Tyr Leu His Ser Ala Leu Lys Lys Phe Gly Asn Ala Val Pro Ala
            260                 265                 270

Val Tyr Pro Leu Asp Ile Tyr Ser His Leu Cys Ile Val Asp Asn Leu
        275                 280                 285
```

```
Glu Arg Leu Gly Ile Ser Arg Tyr Phe Leu Thr Glu Ile Gln Ser Val
    290                 295                 300

Leu Asp Glu Thr His Arg Cys Trp Met Gln Gly Asn Glu Glu Ile Phe
305                 310                 315                 320

Met Asp Ala Ser Thr Cys Ala Leu Ala Phe Arg Ile Leu Arg Leu Asn
                325                 330                 335

Gly Tyr Asp Val Thr Ser Asp Pro Val Thr Lys Ile Gln His Glu Cys
            340                 345                 350

Phe Ser Ser Ser Phe His Gly Asn Val Met Asp Ile Asn Thr Thr Leu
        355                 360                 365

Glu Leu Tyr Arg Ala Ser Glu Leu Ile Leu Tyr Pro Asp Glu Arg Asp
    370                 375                 380

Leu Val Arg Gln Asn Leu Arg Leu Lys Gln Ile Leu Glu Gln Glu Leu
385                 390                 395                 400

Ser Asn Gly Phe Ile Gln Ser Cys Gln Leu Gly Arg Ser Val Asn Ala
                405                 410                 415

Glu Val Asn Gln Ala Ile Glu Tyr Pro Phe Tyr Ala Ile Met Asp Arg
            420                 425                 430

Val Ala Lys Arg Lys Asn Ile Glu Asn Tyr Asn Phe Asp Asn Thr Arg
        435                 440                 445

Ile Leu Lys Thr Ser Tyr Cys Ser Pro Asn Phe Gly Asn Lys Asp Phe
    450                 455                 460

Leu Phe Leu Ser Val Glu Asp Phe Asn Leu Cys Gln Ala Thr His Arg
465                 470                 475                 480

Glu Glu Leu Arg Glu Leu Glu Arg Trp Val Val Glu Asn Arg Leu Asp
                485                 490                 495

Glu Leu Gln Phe Ala Arg Ser Lys Ser Ala Tyr Cys Tyr Phe Ser Ala
            500                 505                 510

Ala Ala Thr Phe Ser Ala Pro Glu Leu Arg Asp Ala Arg Met Ser Trp
        515                 520                 525

Ala Lys Gly Gly Val Leu Thr Thr Val Ile Asp Asp Phe Phe Asp Val
```

530 535 540

Gly Gly Ser Met Glu Glu Leu Lys Asn Leu Ile His Leu Val Glu Lys
545 550 555 560

Trp Asp Val Asp Val Ser Thr Glu Cys Ser Ser His Asn Val Gln Ile
565 570 575

Ile Phe Ser Ala Leu Lys Ser Thr Ile Arg Glu Ile Gly Tyr Lys Gly
580 585 590

Leu Lys Leu Gln Gly Arg Cys Ile Thr Asn His Ile Ile Gly Ile Trp
595 600 605

Leu Asp Leu Leu Asn Ser Met Met Lys Glu Thr Glu Trp Ala Arg Asp
610 615 620

Asn Tyr Val Pro Thr Ile Asp Glu Tyr Met Ser Asn Ala Tyr Val Ser
625 630 635 640

Phe Ala Leu Gly Pro Ile Val Leu Pro Thr Leu Tyr Leu Val Gly Pro
645 650 655

Lys Leu Ser Glu Glu Met Ala Asn His Pro Glu Tyr Tyr Lys Leu Phe
660 665 670

Lys Leu Met Ser Thr Cys Gly Arg Leu Leu Asn Asp Ile Arg Gly Tyr
675 680 685

Glu Arg Glu Leu Lys Asp Gly Lys Leu Asn Ala Leu Ser Leu Tyr Met
690 695 700

Ala Asn His Gly Gly Glu Val Ser Lys Glu Ala Ala Ile Ser Glu Ile
705 710 715 720

Lys Ser Trp Ile Glu Ser Ser Arg Arg Glu Leu Leu Arg Leu Val Leu
725 730 735

Glu Gly Lys Lys Ser Val Leu Pro Lys Pro Cys Lys Glu Leu Phe Trp
740 745 750

His Met Cys Ser Val Val His Leu Phe Tyr Ser Lys Asp Asp Gly Phe
755 760 765

Thr Ser Gln Asp Leu Ile Gln Val Val Asn Ala Ile Ile His Lys Pro
770 775 780

Ile Val Leu Lys Glu Gln Thr Gly Ala Arg Ile
785 790 795

<210> 95
<211> 771
<212> PRT
<213> *Salvia sclarea*

<400> 95

Met Glu Thr Gly Leu Gln Thr Ala Thr Ser Ala Lys Ile Ala Ser Met
1 5 10 15

Pro Ala Cys Phe Glu Glu Thr Arg Gly Arg Ile Ala Lys Leu Phe His
20 25 30

Lys Asp Glu Leu Ser Val Ser Thr Tyr Asp Thr Ala Trp Val Ala Met
35 40 45

Val Pro Ser Pro Thr Ser Leu Glu Glu Pro Cys Phe Pro Asp Cys Leu
50 55 60

Asn Trp Leu Leu Glu Asn Gln Cys His Asp Gly Ser Trp Ala Arg Pro
65 70 75 80

His His His Ser Leu Leu Met Lys Asp Val Leu Ser Ser Thr Leu Ala
85 90 95

Cys Ile Leu Ala Leu Lys Lys Trp Gly Val Gly Glu Lys Gln Ile Asn
100 105 110

Arg Gly Leu His Phe Met Glu Leu Asn Phe Ala Ser Ala Thr Glu Lys
115 120 125

Cys Gln Ile Thr Pro Met Gly Phe Asp Ile Val Phe Pro Ala Met Leu
130 135 140

Asp Tyr Ala Arg Asp Phe Ser Leu Asp Met His Leu Glu Pro Thr Thr
145 150 155 160

Leu Asn Asp Leu Ile His Lys Arg Asp Leu Glu Leu Lys Ser Lys Pro
165 170 175

Asp Phe Ser Ser Asp Gly Glu Ala Tyr Trp Ala Tyr Ile Ala Glu Gly
180 185 190

```
Met Gly Asn Leu Arg Asn Trp Glu Ser Val Met Lys Tyr Gln Arg Arg
        195                 200                 205

Asn Gly Ser Leu Phe Asn Cys Pro Ser Thr Thr Ala Ala Ala Phe Val
    210                 215                 220

Ala Leu Gly Asn Ser Asp Cys Leu Asn Tyr Leu His Ser Ala Leu Lys
225                 230                 235                 240

Lys Phe Gly Asn Ala Val Pro Ala Val Tyr Pro Leu Asp Ile Tyr Ser
                245                 250                 255

His Leu Cys Ile Val Asp Asn Leu Glu Arg Leu Gly Ile Ser Arg Tyr
            260                 265                 270

Phe Leu Thr Glu Ile Gln Ser Val Leu Asp Glu Thr His Arg Cys Trp
        275                 280                 285

Met Gln Gly Asn Glu Glu Ile Phe Met Asp Ala Ser Thr Cys Ala Leu
    290                 295                 300

Ala Phe Arg Ile Leu Arg Leu Asn Gly Tyr Asp Val Thr Ser Asp Pro
305                 310                 315                 320

Val Thr Lys Ile Gln His Glu Cys Phe Ser Ser Ser Phe His Gly Asn
                325                 330                 335

Val Met Asp Ile Asn Thr Thr Leu Glu Leu Tyr Arg Ala Ser Glu Leu
            340                 345                 350

Ile Leu Tyr Pro Asp Glu Arg Asp Leu Val Arg Gln Asn Leu Arg Leu
        355                 360                 365

Lys Gln Ile Leu Glu Gln Glu Leu Ser Asn Gly Phe Ile Gln Ser Cys
    370                 375                 380

Gln Leu Gly Arg Ser Val Asn Ala Glu Val Asn Gln Ala Ile Glu Tyr
385                 390                 395                 400

Pro Phe Tyr Ala Ile Met Asp Arg Val Ala Lys Arg Lys Asn Ile Glu
                405                 410                 415

Asn Tyr Asn Phe Asp Asn Thr Arg Ile Leu Lys Thr Ser Tyr Cys Ser
            420                 425                 430

Pro Asn Phe Gly Asn Lys Asp Phe Leu Phe Leu Ser Val Glu Asp Phe
```

435 440 445

Asn Leu Cys Gln Ala Thr His Arg Glu Glu Leu Arg Glu Leu Glu Arg
450 455 460

Trp Val Val Glu Asn Arg Leu Asp Glu Leu Gln Phe Ala Arg Ser Lys
465 470 475 480

Ser Ala Tyr Cys Tyr Phe Ser Ala Ala Ala Thr Phe Ser Ala Pro Glu
485 490 495

Leu Arg Asp Ala Arg Met Ser Trp Ala Lys Gly Gly Val Leu Thr Thr
500 505 510

Val Ile Asp Asp Phe Phe Asp Val Gly Gly Ser Met Glu Glu Leu Lys
515 520 525

Asn Leu Ile His Leu Val Glu Lys Trp Asp Val Asp Val Ser Thr Glu
530 535 540

Cys Ser Ser His Asn Val Gln Ile Ile Phe Ser Ala Leu Lys Ser Thr
545 550 555 560

Ile Arg Glu Ile Gly Tyr Lys Gly Leu Lys Leu Gln Gly Arg Cys Ile
565 570 575

Thr Asn His Ile Ile Gly Ile Trp Leu Asp Leu Leu Asn Ser Met Met
580 585 590

Lys Glu Thr Glu Trp Ala Arg Asp Asn Tyr Val Pro Thr Ile Asp Glu
595 600 605

Tyr Met Ser Asn Ala Tyr Val Ser Phe Ala Leu Gly Pro Ile Val Leu
610 615 620

Pro Thr Leu Tyr Leu Val Gly Pro Lys Leu Ser Glu Glu Met Ala Asn
625 630 635 640

His Pro Glu Tyr Tyr Lys Leu Phe Lys Leu Met Ser Thr Cys Gly Arg
645 650 655

Leu Leu Asn Asp Ile Arg Gly Tyr Glu Arg Glu Leu Lys Asp Gly Lys
660 665 670

Leu Asn Ala Leu Ser Leu Tyr Met Ala Asn His Gly Gly Glu Val Ser
675 680 685

```
Lys Glu Ala Ala Ile Ser Glu Ile Lys Ser Trp Ile Glu Ser Ser Arg
    690                 695                 700

Arg Glu Leu Leu Arg Leu Val Leu Glu Gly Lys Lys Ser Val Leu Pro
705                 710                 715                 720

Lys Pro Cys Lys Glu Leu Phe Trp His Met Cys Ser Val Val His Leu
                725                 730                 735

Phe Tyr Ser Lys Asp Asp Gly Phe Thr Ser Gln Asp Leu Ile Gln Val
            740                 745                 750

Val Asn Ala Ile Ile His Lys Pro Ile Val Leu Lys Glu Gln Thr Gly
        755                 760                 765

Ala Arg Ile
    770
```

<210> 96
<211> 525
<212> PRT
<213> *Salvia sclarea*

<400> 96

```
Met Ala Lys Met Lys Glu Asn Phe Lys Arg Glu Asp Asp Lys Phe Pro
1               5                   10                  15

Thr Thr Thr Thr Leu Arg Ser Glu Asp Ile Pro Ser Asn Leu Cys Ile
            20                  25                  30

Ile Asp Thr Leu Gln Arg Leu Gly Val Asp Gln Phe Phe Gln Tyr Glu
        35                  40                  45

Ile Asn Thr Ile Leu Asp Asn Thr Phe Arg Leu Trp Gln Glu Lys His
    50                  55                  60

Lys Val Ile Tyr Gly Asn Val Thr Thr His Ala Met Ala Phe Arg Leu
65                  70                  75                  80

Leu Arg Val Lys Gly Tyr Glu Val Ser Ser Glu Glu Leu Ala Pro Tyr
                85                  90                  95

Gly Asn Gln Glu Ala Val Ser Gln Gln Thr Asn Asp Leu Pro Met Ile
            100                 105                 110
```

```
Ile Glu Leu Tyr Arg Ala Ala Asn Glu Arg Ile Tyr Glu Glu Glu Arg
        115                 120                 125

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

Val Gln Asp Asn Ser Ile Pro Asp Lys Lys Leu His Lys Leu Val Glu
145                 150                 155                 160

Phe Tyr Leu Arg Asn Tyr Lys Gly Ile Thr Ile Arg Leu Gly Ala Arg
                165                 170                 175

Arg Asn Leu Glu Leu Tyr Asp Met Thr Tyr Tyr Gln Ala Leu Lys Ser
            180                 185                 190

Thr Asn Arg Phe Ser Asn Leu Cys Asn Glu Asp Phe Leu Val Phe Ala
        195                 200                 205

Lys Gln Asp Phe Asp Ile His Glu Ala Gln Asn Gln Lys Gly Leu Gln
    210                 215                 220

Gln Leu Gln Arg Trp Tyr Ala Asp Cys Arg Leu Asp Thr Leu Asn Phe
225                 230                 235                 240

Gly Arg Asp Val Val Ile Ile Ala Asn Tyr Leu Ala Ser Leu Ile Ile
                245                 250                 255

Gly Asp His Ala Phe Asp Tyr Val Arg Leu Ala Phe Ala Lys Thr Ser
            260                 265                 270

Val Leu Val Thr Ile Met Asp Asp Phe Phe Asp Cys His Gly Ser Ser
        275                 280                 285

Gln Glu Cys Asp Lys Ile Ile Glu Leu Val Lys Glu Trp Lys Glu Asn
    290                 295                 300

Pro Asp Ala Glu Tyr Gly Ser Glu Glu Leu Glu Ile Leu Phe Met Ala
305                 310                 315                 320

Leu Tyr Asn Thr Val Asn Glu Leu Ala Glu Arg Ala Arg Val Glu Gln
                325                 330                 335

Gly Arg Ser Val Lys Glu Phe Leu Val Lys Leu Trp Val Glu Ile Leu
            340                 345                 350

Ser Ala Phe Lys Ile Glu Leu Asp Thr Trp Ser Asn Gly Thr Gln Gln
```

```
        355                 360                 365

Ser Phe Asp Glu Tyr Ile Ser Ser Trp Leu Ser Asn Gly Ser Arg
    370                 375                 380

Leu Thr Gly Leu Leu Thr Met Gln Phe Val Gly Val Lys Leu Ser Asp
385                 390                 395                 400

Glu Met Leu Met Ser Glu Glu Cys Thr Asp Leu Ala Arg His Val Cys
                405                 410                 415

Met Val Gly Arg Leu Leu Asn Asp Val Cys Ser Ser Glu Arg Glu Arg
            420                 425                 430

Glu Glu Asn Ile Ala Gly Lys Ser Tyr Ser Ile Leu Leu Ala Thr Glu
        435                 440                 445

Lys Asp Gly Arg Lys Val Ser Glu Asp Glu Ala Ile Ala Glu Ile Asn
    450                 455                 460

Glu Met Val Glu Tyr His Trp Arg Lys Val Leu Gln Ile Val Tyr Lys
465                 470                 475                 480

Lys Glu Ser Ile Leu Pro Arg Arg Cys Lys Asp Val Phe Leu Glu Met
                485                 490                 495

Ala Lys Gly Thr Phe Tyr Ala Tyr Gly Ile Asn Asp Glu Leu Thr Ser
            500                 505                 510

Pro Gln Gln Ser Lys Glu Asp Met Lys Ser Phe Val Phe
        515                 520                 525
```

<210> 97
<211> 805
<212> PRT
<213> *Salvia sclarea*

<400> 97

```
Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro
1               5                   10                  15

Arg Gly Ser His Met Thr Ser Val Asn Leu Ser Arg Ala Pro Ala Ala
            20                  25                  30

Ile Thr Arg Arg Arg Leu Gln Leu Gln Pro Glu Phe His Ala Glu Cys
        35                  40                  45
```

```
Ser Trp Leu Lys Ser Ser Ser Lys His Ala Pro Leu Thr Leu Ser Cys
    50                  55                  60

Gln Ile Arg Pro Lys Gln Leu Ser Gln Ile Ala Glu Leu Arg Val Thr
65                  70                  75                  80

Ser Leu Asp Ala Ser Gln Ala Ser Glu Lys Asp Ile Ser Leu Val Gln
                85                  90                  95

Thr Pro His Lys Val Glu Val Asn Glu Lys Ile Glu Glu Ser Ile Glu
            100                 105                 110

Tyr Val Gln Asn Leu Leu Met Thr Ser Gly Asp Gly Arg Ile Ser Val
        115                 120                 125

Ser Pro Tyr Asp Thr Ala Val Ile Ala Leu Ile Lys Asp Leu Lys Gly
    130                 135                 140

Arg Asp Ala Pro Gln Phe Pro Ser Cys Leu Glu Trp Ile Ala His His
145                 150                 155                 160

Gln Leu Ala Asp Gly Ser Trp Gly Asp Glu Phe Phe Cys Ile Tyr Asp
                165                 170                 175

Arg Ile Leu Asn Thr Leu Ala Cys Val Val Ala Leu Lys Ser Trp Asn
            180                 185                 190

Leu His Ser Asp Ile Ile Glu Lys Gly Val Thr Tyr Ile Lys Glu Asn
        195                 200                 205

Val His Lys Leu Lys Gly Ala Asn Val Glu His Arg Thr Ala Gly Phe
    210                 215                 220

Glu Leu Val Val Pro Thr Phe Met Gln Met Ala Thr Asp Leu Gly Ile
225                 230                 235                 240

Gln Asp Leu Pro Tyr Asp His Pro Leu Ile Lys Glu Ile Ala Asp Thr
                245                 250                 255

Lys Gln Gln Arg Leu Lys Glu Ile Pro Lys Asp Leu Val Tyr Gln Met
            260                 265                 270

Pro Thr Asn Leu Leu Tyr Ser Leu Glu Gly Leu Gly Asp Leu Glu Trp
        275                 280                 285
```

```
Glu Arg Leu Leu Lys Leu Gln Ser Gly Asn Gly Ser Phe Leu Thr Ser
    290                 295                 300

Pro Ser Ser Thr Ala Ala Val Leu Met His Thr Lys Asp Glu Lys Cys
305                 310                 315                 320

Leu Lys Tyr Ile Glu Asn Ala Leu Lys Asn Cys Asp Gly Gly Ala Pro
                325                 330                 335

His Thr Tyr Pro Val Asp Ile Phe Ser Arg Leu Trp Ala Ile Asp Arg
            340                 345                 350

Leu Gln Arg Leu Gly Ile Ser Arg Phe Phe Gln His Glu Ile Lys Tyr
        355                 360                 365

Phe Leu Asp His Ile Glu Ser Val Trp Glu Glu Thr Gly Val Phe Ser
    370                 375                 380

Gly Arg Tyr Thr Lys Phe Ser Asp Ile Asp Asp Thr Ser Met Gly Val
385                 390                 395                 400

Arg Leu Leu Lys Met His Gly Tyr Asp Val Asp Pro Asn Val Leu Lys
                405                 410                 415

His Phe Lys Gln Gln Asp Gly Lys Phe Ser Cys Tyr Ile Gly Gln Ser
            420                 425                 430

Val Glu Ser Ala Ser Pro Met Tyr Asn Leu Tyr Arg Ala Ala Gln Leu
        435                 440                 445

Arg Phe Pro Gly Glu Glu Val Leu Glu Glu Ala Thr Lys Phe Ala Phe
    450                 455                 460

Asn Phe Leu Gln Glu Met Leu Val Lys Asp Arg Leu Gln Glu Arg Trp
465                 470                 475                 480

Val Ile Ser Asp His Leu Phe Asp Glu Ile Lys Leu Gly Leu Lys Met
                485                 490                 495

Pro Trp Tyr Ala Thr Leu Pro Arg Val Glu Ala Ala Tyr Tyr Leu Asp
            500                 505                 510

His Tyr Ala Gly Ser Gly Asp Val Trp Ile Gly Lys Ser Phe Tyr Arg
        515                 520                 525

Met Pro Glu Ile Ser Asn Asp Thr Tyr Lys Glu Leu Ala Ile Leu Asp
```

530 535 540

Phe Asn Arg Cys Gln Thr Gln His Gln Leu Glu Trp Ile His Met Gln
545 550 555 560

Glu Trp Tyr Asp Arg Cys Ser Leu Ser Glu Phe Gly Ile Ser Lys Arg
565 570 575

Glu Leu Leu Arg Ser Tyr Phe Leu Ala Ala Ala Thr Ile Phe Glu Pro
580 585 590

Glu Arg Thr Gln Glu Arg Leu Leu Trp Ala Lys Thr Arg Ile Leu Ser
595 600 605

Lys Met Ile Thr Ser Phe Val Asn Ile Ser Gly Thr Thr Leu Ser Leu
610 615 620

Asp Tyr Asn Phe Asn Gly Leu Asp Glu Ile Ile Ser Ser Ala Asn Glu
625 630 635 640

Asp Gln Gly Leu Ala Gly Thr Leu Leu Ala Thr Phe His Gln Leu Leu
645 650 655

Asp Gly Phe Asp Ile Tyr Thr Leu His Gln Leu Lys His Val Trp Ser
660 665 670

Gln Trp Phe Met Lys Val Gln Gln Gly Glu Gly Ser Gly Gly Glu Asp
675 680 685

Ala Val Leu Leu Ala Asn Thr Leu Asn Ile Cys Ala Gly Leu Asn Glu
690 695 700

Asp Val Leu Ser Asn Asn Glu Tyr Thr Ala Leu Ser Thr Leu Thr Asn
705 710 715 720

Lys Ile Cys Asn Arg Leu Ala Gln Ile Gln Asp Asn Lys Ile Leu Gln
725 730 735

Val Val Asp Gly Ser Ile Lys Asp Lys Glu Leu Glu Gln Asp Met Gln
740 745 750

Ala Leu Val Lys Leu Val Leu Gln Glu Asn Gly Gly Ala Val Asp Arg
755 760 765

Asn Ile Arg His Thr Phe Leu Ser Val Ser Lys Thr Phe Tyr Tyr Asp
770 775 780

```
Ala Tyr His Asp Asp Glu Thr Thr Asp Leu His Ile Phe Lys Val Leu
785                 790                 795                 800

Phe Arg Pro Val Val
                805
```

<210> 98
<211> 804
<212> PRT
<213> *Salvia sclarea*

<400> 98

```
Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro
1               5                  10                  15

Arg Gly Ser His Met Thr Ser Val Asn Leu Ser Arg Ala Pro Ala Ala
            20                  25                  30

Ile Ile Arg Arg Arg Leu Gln Leu Gln Pro Glu Phe His Ala Glu Cys
        35                  40                  45

Ser Trp Leu Lys Ser Ser Ser Lys His Ala Pro Phe Thr Leu Ser Cys
    50                  55                  60

Gln Ile Arg Pro Lys Gln Leu Ser Gln Ile Ala Glu Leu Arg Val Thr
65                  70                  75                  80

Ser Leu Asp Ala Ser Gln Ala Ser Glu Lys Asp Ile Ser Leu Val Gln
                85                  90                  95

Thr Pro His Lys Val Glu Val Asn Glu Lys Ile Glu Glu Ser Ile Glu
            100                 105                 110

Tyr Val Gln Asn Leu Leu Met Thr Ser Gly Asp Gly Arg Ile Ser Val
        115                 120                 125

Ser Pro Tyr Asp Thr Ala Val Ile Ala Leu Ile Lys Asp Leu Lys Gly
    130                 135                 140

Arg Asp Ala Pro Gln Phe Pro Ser Cys Leu Glu Trp Ile Ala His His
145                 150                 155                 160

Gln Leu Ala Asp Gly Ser Trp Gly Asp Glu Phe Phe Cys Ile Tyr Asp
                165                 170                 175
```

```
Arg Ile Leu Asn Thr Leu Ala Cys Val Val Ala Leu Lys Ser Trp Asn
            180                 185                 190

Leu Gln Ser Asp Ile Ile Glu Lys Gly Val Thr Tyr Ile Lys Glu Asn
        195                 200                 205

Val His Lys Leu Lys Gly Ala Asn Val Glu His Arg Thr Ala Gly Phe
    210                 215                 220

Glu Leu Val Val Pro Thr Phe Met Gln Met Ala Thr Asp Leu Gly Ile
225                 230                 235                 240

Gln Gly Leu Pro Tyr Asp His Pro Leu Ile Lys Glu Ile Ala Asp Thr
                245                 250                 255

Lys Gln Gln Arg Leu Lys Glu Ile Pro Lys Asp Leu Val Tyr Gln Met
            260                 265                 270

Pro Thr Asn Leu Leu Tyr Ser Leu Glu Gly Leu Gly Asp Leu Glu Trp
        275                 280                 285

Glu Arg Leu Leu Lys Leu Gln Ser Gly Asn Gly Ser Phe Leu Thr Ser
    290                 295                 300

Pro Ser Ser Thr Ala Ala Val Leu Met His Thr Lys Asp Glu Lys Cys
305                 310                 315                 320

Leu Lys Tyr Ile Glu Asn Ala Leu Lys Asn Cys Asp Gly Gly Ala Pro
                325                 330                 335

His Thr Tyr Pro Val Asp Ile Phe Ser Arg Leu Trp Ala Ile Asp Arg
            340                 345                 350

Leu Gln Arg Leu Gly Ile Ser Arg Phe Phe Gln His Glu Ile Lys Tyr
        355                 360                 365

Phe Leu Asp His Ile Glu Ser Val Trp Glu Glu Thr Gly Val Phe Ser
    370                 375                 380

Gly Arg Tyr Thr Lys Phe Ser Asp Ile Asp Asp Thr Ser Met Gly Val
385                 390                 395                 400

Arg Leu Leu Lys Met His Gly Tyr Asp Val Asp Pro Asn Val Leu Lys
                405                 410                 415

His Phe Lys Gln Gln Asp Gly Lys Phe Ser Cys Tyr Ile Gly Gln Ser
```

```
                420                 425                 430

Val Glu Ser Ala Ser Pro Met Tyr Asn Leu Tyr Arg Ala Ala Gln Leu
        435                 440                 445

Arg Phe Pro Gly Glu Glu Val Leu Glu Glu Ala Thr Lys Phe Ala Phe
    450                 455                 460

Asn Phe Leu Gln Glu Met Leu Val Lys Asp Arg Leu Gln Glu Arg Trp
465                 470                 475                 480

Val Ile Ser Asp His Leu Phe Asp Glu Ile Lys Leu Gly Leu Lys Met
                485                 490                 495

Pro Trp Tyr Ala Thr Leu Pro Arg Val Glu Ala Ala Tyr Tyr Leu Asp
            500                 505                 510

His Tyr Ala Gly Ser Gly Asp Val Trp Ile Gly Lys Ser Phe Tyr Arg
        515                 520                 525

Met Pro Glu Ile Ser Asn Asp Thr Tyr Lys Glu Leu Ala Ile Leu Asp
    530                 535                 540

Phe Asn Arg Cys Gln Thr Gln His Gln Leu Glu Trp Ile Gln Met Gln
545                 550                 555                 560

Glu Trp Tyr Asp Arg Cys Ser Leu Ser Glu Phe Gly Ile Ser Lys Arg
                565                 570                 575

Glu Leu Leu Arg Ser Tyr Phe Leu Ala Ala Ala Thr Ile Phe Glu Pro
            580                 585                 590

Glu Arg Thr Gln Glu Arg Leu Leu Trp Ala Lys Thr Arg Ile Leu Ser
        595                 600                 605

Lys Met Ile Thr Ser Phe Val Asn Ile Ser Gly Thr Thr Leu Ser Leu
    610                 615                 620

Asp Tyr Asn Phe Asn Gly Leu Asp Glu Ile Ile Ser Ala Asn Glu Asp
625                 630                 635                 640

Gln Gly Leu Ala Gly Thr Leu Leu Ala Thr Phe His Gln Leu Leu Asp
                645                 650                 655

Gly Phe Asp Ile Tyr Thr Leu His Gln Leu Lys His Val Trp Ser Gln
            660                 665                 670
```

```
Trp Phe Met Lys Val Gln Gln Gly Glu Gly Ser Gly Gly Glu Asp Ala
        675                 680                 685

Val Leu Leu Ala Asn Thr Leu Asn Ile Cys Ala Gly Leu Asn Glu Asp
    690                 695                 700

Val Leu Ser Asn Asn Glu Tyr Thr Ala Leu Ser Thr Leu Thr Asn Lys
705                 710                 715                 720

Ile Cys Asn Arg Leu Ala Gln Ile Gln Asp Asn Lys Ile Leu Gln Val
                725                 730                 735

Val Asp Gly Ser Ile Lys Asp Lys Glu Leu Glu Gln Asp Met Gln Ala
            740                 745                 750

Leu Val Lys Leu Val Leu Gln Glu Asn Gly Gly Ala Val Asp Arg Asn
        755                 760                 765

Ile Arg His Thr Phe Leu Ser Val Ser Lys Thr Phe Tyr Tyr Asp Ala
    770                 775                 780

Tyr His Asp Asp Glu Thr Thr Asp Leu His Ile Phe Lys Val Leu Phe
785                 790                 795                 800

Arg Pro Val Val
```

**REIVINDICACIONES**

1. Un polipéptido aislado que tiene una actividad de esclareol sintasa y que comprende una secuencia de aminoácidos de al menos un 70 % de identidad con la secuencia de aminoácidos completa de SEC ID Nº: 1.

2. El polipéptido aislado de la reivindicación 1, que comprende una secuencia de aminoácidos de al menos un 80 %, más preferentemente al menos un 90 %, de identidad con SEC ID Nº: 1.

3. El polipéptido aislado de la reivindicación 2, que comprende la secuencia de aminoácidos definida en SEC ID Nº: 1.

4. El polipéptido aislado de la reivindicación 3, que consiste en la secuencia de aminoácidos definida en SEC ID Nº: 1.

5. El polipéptido aislado de cualquiera de las reivindicaciones 1 a 4, derivado de *Salvia Sclarea.*

6. Un ácido nucleico aislado que codifica el polipéptido de cualquiera de las reivindicaciones 1 a 5.

7. El ácido nucleico aislado de la reivindicación 6, que comprende una secuencia de nucleótidos de al menos un 70 % de identidad con SEC ID Nº: 2 o el complemento de la misma.

8. El ácido nucleico aislado de la reivindicación 7, que comprende una secuencia de nucleótidos de al menos un 80 %, preferentemente al menos un 90 %, de identidad con SEC ID Nº: 2 o el complemento de la misma.

9. El ácido nucleico aislado de la reivindicación 8, que comprende una secuencia de nucleótidos idéntica a SEC ID Nº: 2 o el complemento de la misma.

10. El ácido nucleico aislado de la reivindicación 9, que consiste en la secuencia de nucleótidos definida en SEC ID Nº: 2 o en el complemento de la misma.

11. El ácido nucleico aislado de cualquiera de las reivindicaciones 6 a 10, derivado de *Salvia sclarea*.

12. Un vector de expresión que comprende el ácido nucleico de cualquiera de las reivindicaciones 6 a 11.

13. El vector de expresión de la reivindicación 12, en la forma de un vector viral, un bacteriófago o un plásmido.

14. El vector de expresión de la reivindicación 12 o 13, que incluye el ácido nucleico de la invención unido operativamente al menos a una secuencia reguladora que controla la transcripción, traducción, iniciación o terminación, tal como un promotor, operador o potenciador transcripcional o un sitio de unión ribosomal de ARNm y, opcionalmente, que incluye al menos un marcador de selección.

15. Un organismo no humano transformado con el vector de expresión de cualquiera de las reivindicaciones 12 a 14, de modo que albergue el ácido nucleico de cualquiera de las reivindicaciones 6 a 11 y exprese o sobreexprese heterólogamente el polipéptido de cualquiera de las reivindicaciones 1 a 5.

16. El organismo no humano de la reivindicación 15, en el que dicho organismo no humano es una planta, un procariota o un hongo.

17. El organismo no humano de la reivindicación 15, en el que dicho organismo no humano es un microorganismo.

18. El organismo no humano de la reivindicación 17, en el que dicho microorganismo es una bacteria o levadura.

19. El organismo no humano de la reivindicación 18, en el que dicha bacteria es *E. coli* y dicha levadura es *Saccharomyces cerevisiae.*

20. Una célula eucariota superior transformada con el vector de expresión de cualquiera de las reivindicaciones 12 a 14, de modo que albergue el ácido nucleico de cualquiera de las reivindicaciones 6 a 11 y exprese el polipéptido de cualquiera de las reivindicaciones 1 a 5.

21. La célula eucariota superior de la reivindicación 20, en la que dicha célula eucariota superior es una célula de planta o una célula fúngica.

22. Un procedimiento de producción de al menos un polipéptido que tiene una actividad de esclareol sintasa que comprende:

a) cultivar un organismo o célula no humano transformado con el vector de expresión de cualquiera de las reivindicaciones 12 a 14, de modo que albergue un ácido nucleico de acuerdo con cualquiera de las reivindicaciones 6 a 11 y exprese o sobreexprese un polipéptido codificado por dicho ácido nucleico y que tenga una actividad de esclareol sintasa;
b) aislar los polipéptidos que tienen una actividad de esclareol sintasa del organismo o célula no humano

cultivado en la etapa a).

23. El procedimiento de la reivindicación 22, que comprende además, antes de la etapa a), transformar un organismo o célula hospedador no humano con el vector de expresión de cualquiera de las reivindicaciones 12 a 14, de modo que albergue un ácido nucleico de acuerdo con cualquiera de las reivindicaciones 6 a 11 y exprese o sobreexprese el polipéptido codificado por dicho ácido nucleico.

24. Un procedimiento de producción de un polipéptido variante que tiene una actividad de esclareol sintasa que comprende las etapas de:

a) seleccionar un ácido nucleico de acuerdo con cualquiera de las reivindicaciones 6 a 11;
b) modificar el ácido nucleico seleccionado para obtener al menos un ácido nucleico mutante;
c) transformar las células o los organismos unicelulares hospedadores con la secuencia de ácido nucleico mutante para expresar un polipéptido codificado por la secuencia de ácido nucleico mutante;
d) cribar sistemáticamente el polipéptido para al menos una propiedad modificada; y,
e) opcionalmente, si el polipéptido no tiene ninguna actividad de esclareol sintasa variante deseada, repetir las etapas de procedimiento (a) a (d) hasta que se obtenga un polipéptido con una actividad de esclareol sintasa variante deseada; opcionalmente, si se identificó un polipéptido que tiene una actividad de esclareol sintasa variante deseada en la etapa d), aislar el correspondiente ácido nucleico mutante obtenido en la etapa (c).

25. Un procedimiento de producción de esclareol que comprende

a) poner en contacto difosfato de labdenodiol (LPP) con un polipéptido que tiene una actividad de esclareol sintasa y que comprende una secuencia de aminoácidos de acuerdo con cualquiera de las reivindicaciones 1 a 4; y
b) opcionalmente, aislar el esclareol producido en la etapa a).

26. El procedimiento de la reivindicación 25, en el que la etapa a) se realiza por cultivo de un organismo o célula no humano capaz de producir LPP y transformar para expresar dicho polipéptido en condiciones que conducen a la producción de esclareol.

27. El procedimiento de la reivindicación 26, que comprende además, antes de la etapa a), transformar el organismo o célula no humano capaz de producir LPP con un ácido nucleico que codifica dicho polipéptido, de modo que dicho organismo exprese dicho polipéptido.

28. El procedimiento de la reivindicación 26 a 27, en el que dicho organismo no humano es una planta, un procariota o un hongo.

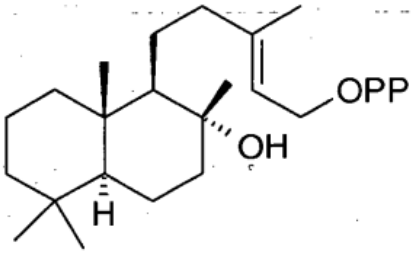
29. El procedimiento de la reivindicación 26 o 27, en el que dicho organismo no humano es un microorganismo.

30. El procedimiento de la reivindicación 29, en el que dicho microorganismo es una bacteria o levadura.

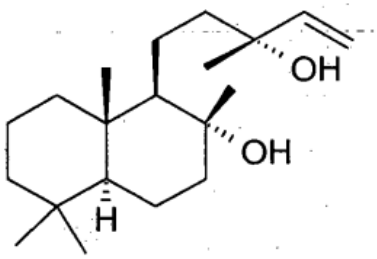
31. El procedimiento de la reivindicación 30, en el que dicha bacteria es *E. coli* y dicha levadura es *Saccharomyces cerevisiae.*

32. El procedimiento de la reivindicación 26 o 27, en el que dicha célula no humana es una célula eucariota superior seleccionada entre células de planta o células fúngicas.

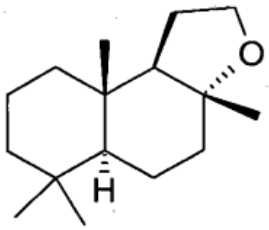
**Figura 1**



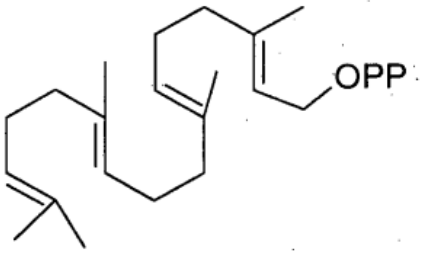

difosfato de labdenodiol



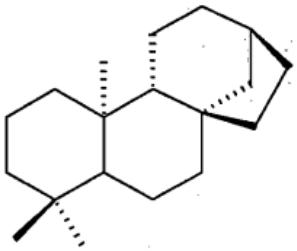

esclareol
(-)-(13R)-14-labdeno-8,13-diol



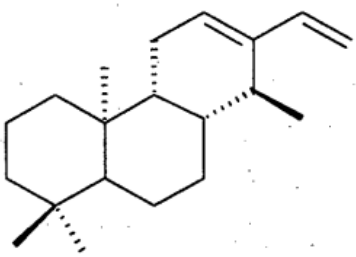

(-)-Ambrox




difosfato de geranilgeranilo (GGPP)



ent-kaureno



ent-cassa-12,15-dieno

**Figura 2**

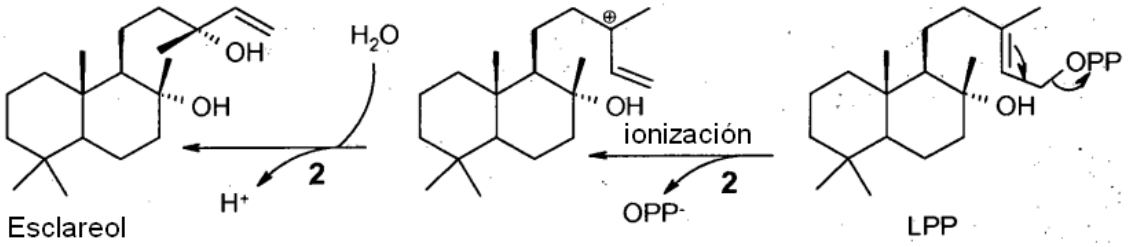


**Figura 3 (1ª parte)**

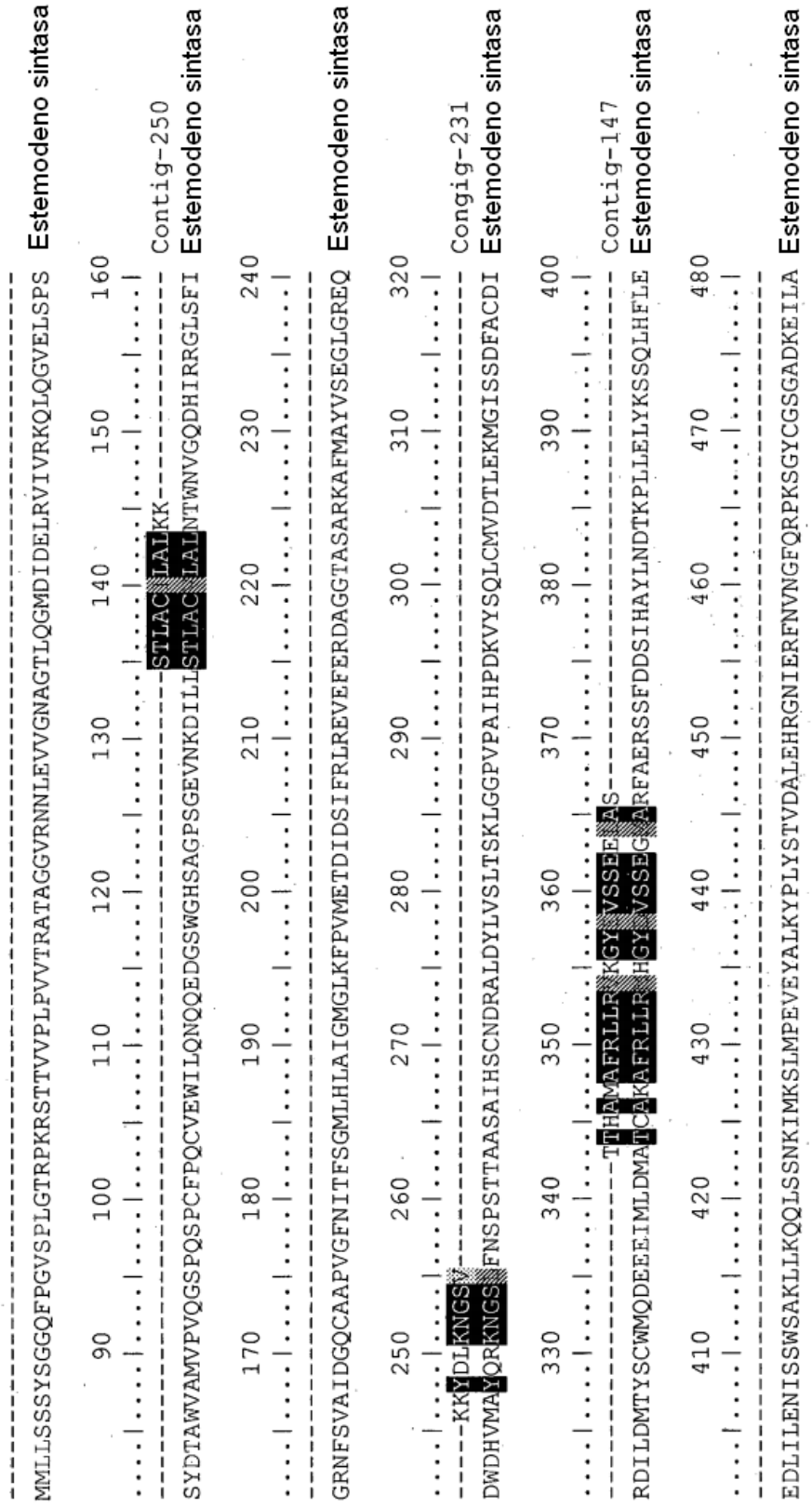


**Figura 3 (2ª parte)**

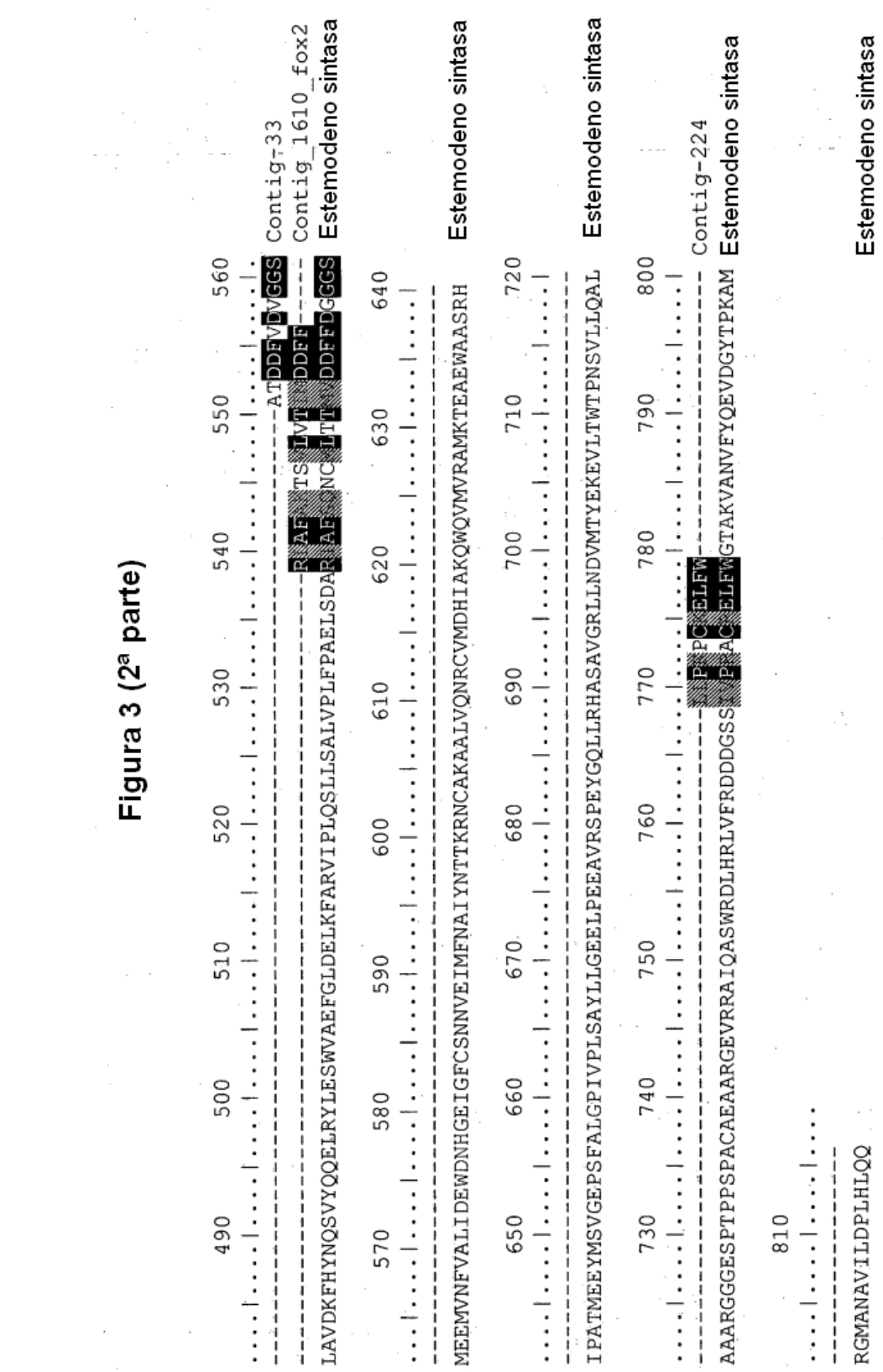


Figura 4 (1ª parte)

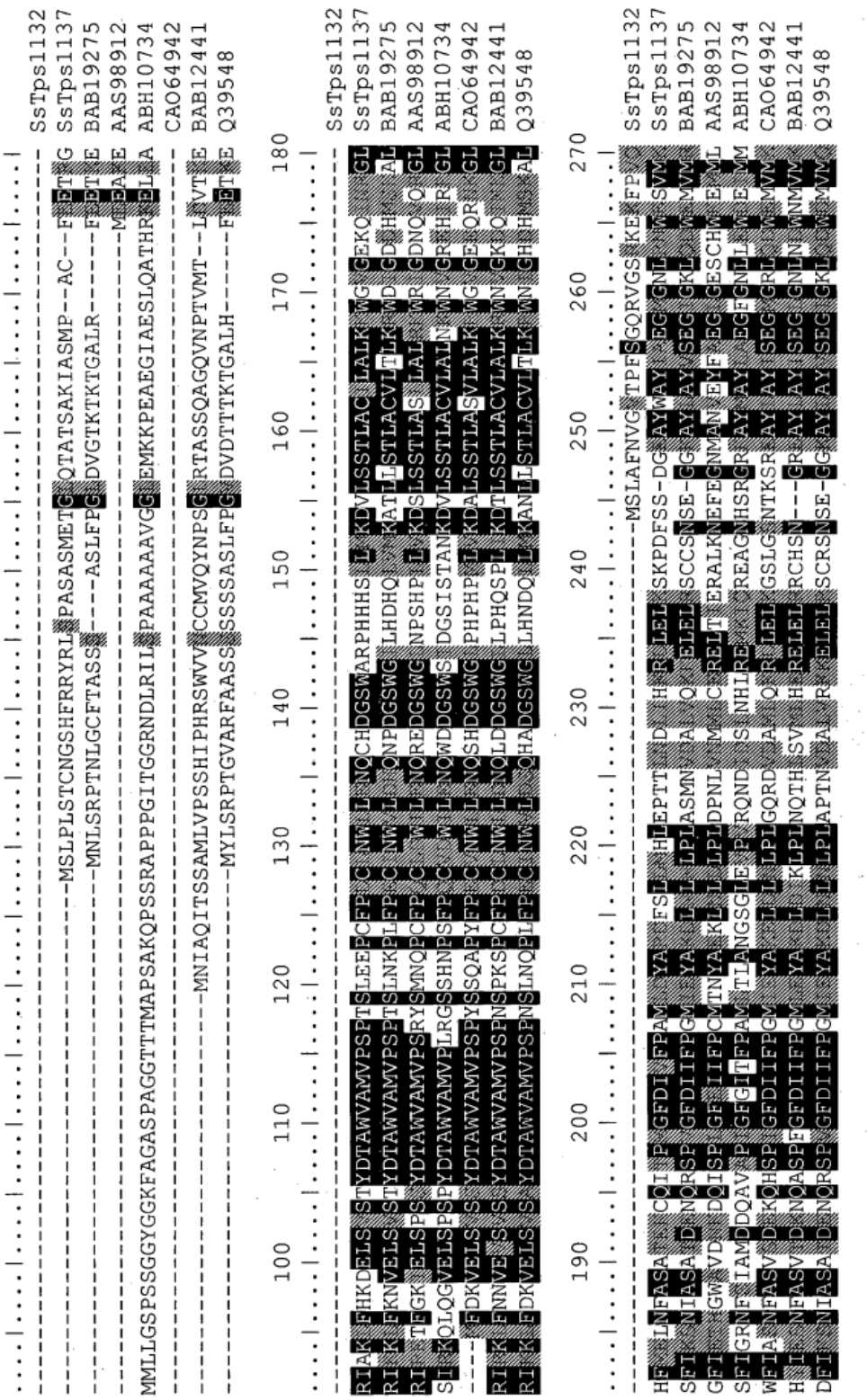


Figura 4 (2ª parte)

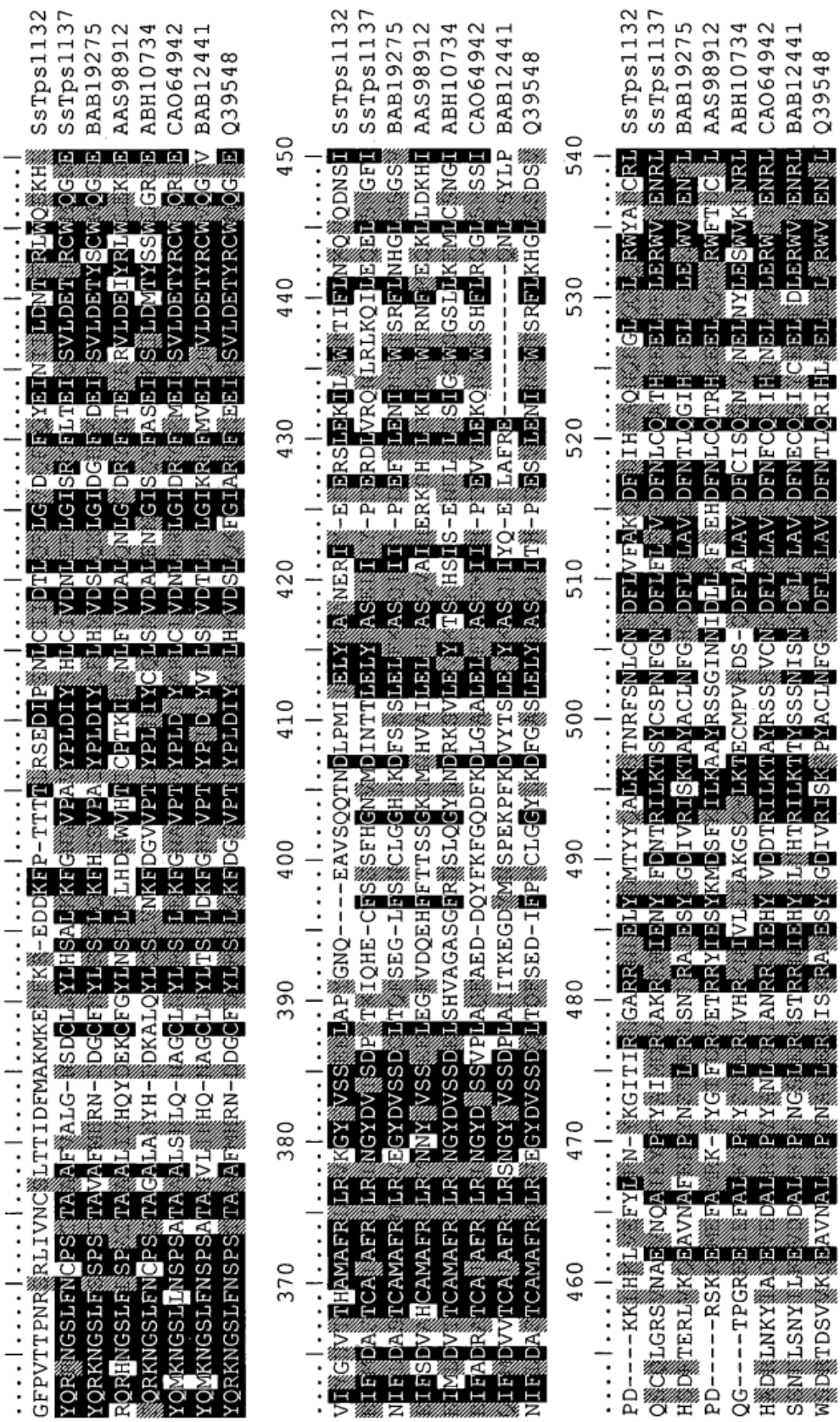


Figura 4 (3ª parte)

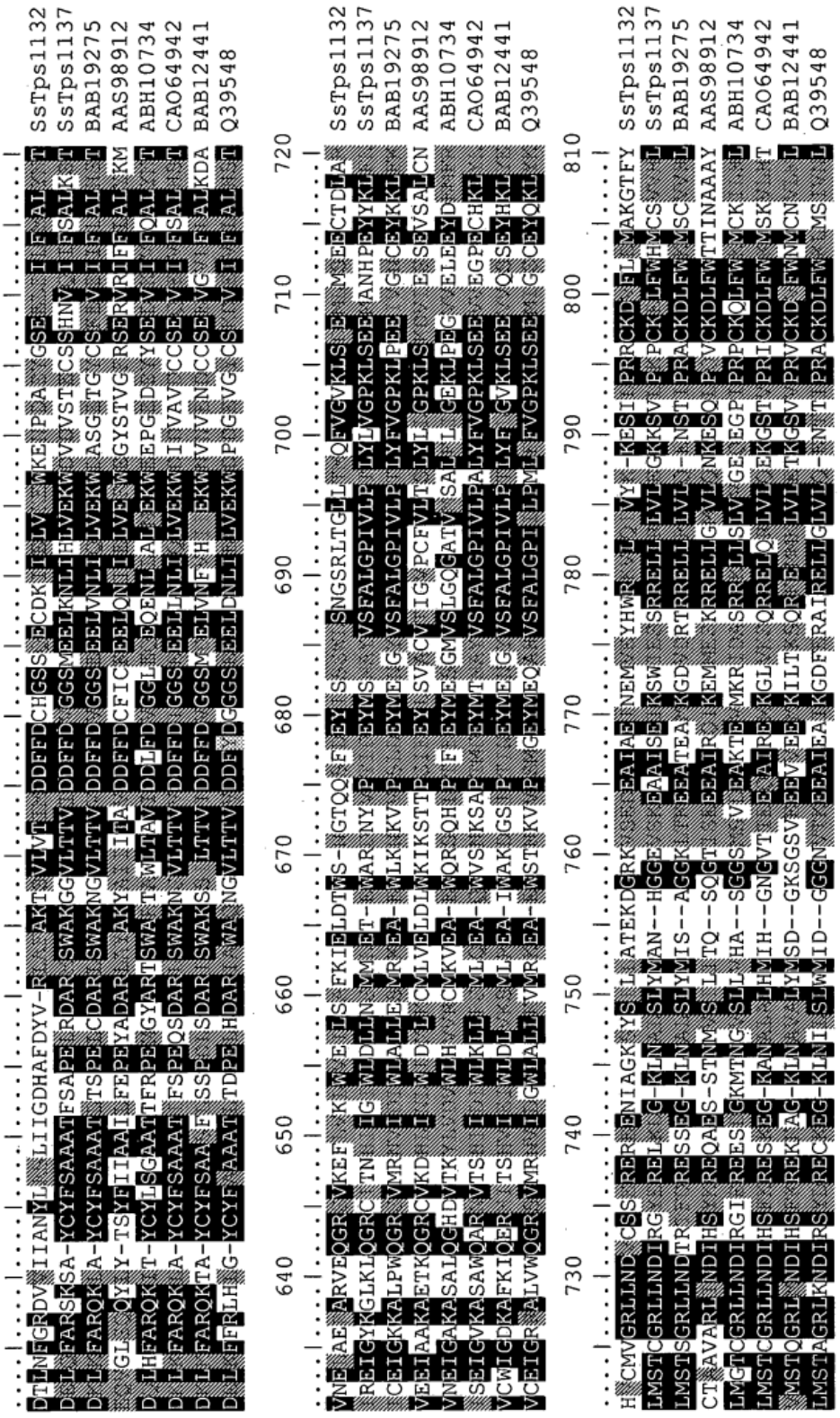
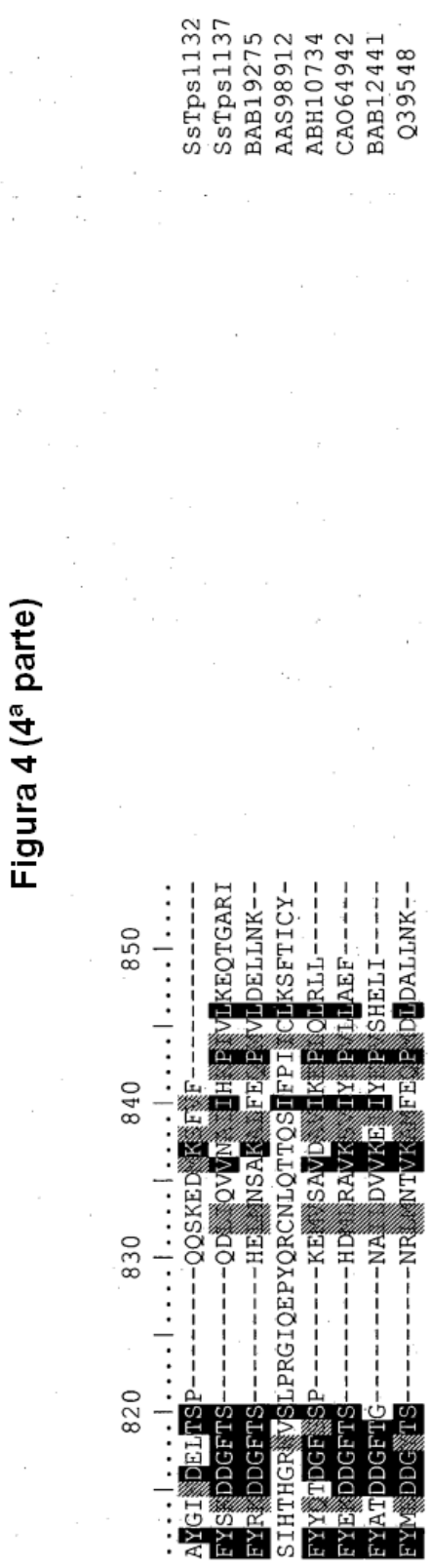


Figura 4 (4ª parte)



# Figura 5

```
                   10        20        30        40        50        60        70        80        90       100       110       120
SsTps1132  MSLAFNVGVTPFSGQRVGSRKEKFPVQGFPVTTPNRSRLIVNCSLTTIDFMAKMKENFKREDDKFPTTTTLRSEDIPSNLCIIDTLQRLGVDQFFQYEINTILDNTFRLWQEKHKVIYGN
1132-2-5   --------------------------------------------------MAKMKENFKREDDKFPTTTTLRSEDIPSNLCIIDTLQRLGVDQFFQYEINTILDNTFRLWQEKHKVIYGN

                  130       140       150       160       170       180       190       200       210       220       230       240
SsTps1132  VTTHAMAFRLLRVKGYEVSSEELAPYGNQEAVSQQTNDLPMIIELYRAANERIYEEERSLEKILAWTTIFLNKQVQDNSIPDKKLHKLVEFYLRNYKGITIRLGARRNLELYDMTYYQAL
1132-2-5   VTTHAMAFRLLRVKGYEVSSEELAPYGNQEAVSQQTNDLPMIIELYRAANERIYEEERSLEKILAWTTIFLNKQVQDNSIPDKKLHKLVEFYLRNYKGITIRLGARRNLELYDMTYYQAL

                  250       260       270       280       290       300       310       320       330       340       350       360
SsTps1132  KSTNRFSNLCNEDFLVFAKQDFDIHEAQNQKGLQQLQRWYADCRLDTLNFGRDVVIIANYLASLIIGDHAFDYVRLAFAKTSVLVTIMDDFFDCHGSSQECDKIIELVKEWKENPDAEYG
1132-2-5   KSTNRFSNLCNEDFLVFAKQDFDIHEAQNQKGLQQLQRWYADCRLDTLNFGRDVVIIANYLASLIIGDHAFDYVRLAFAKTSVLVTIMDDFFDCHGSSQECDKIIELVKEWKENPDAEYG

                  370       380       390       400       410       420       430       440       450       460       470       480
SsTps1132  SEELEILFMALYNTVNELAERARVEQGRSVKEFLVKLWVEILSAFKIELDTWSNGTQQSFDEYISSSWLSNGSRLTGLLTMQFVGVKLSDEMLMSEECTDLARHVCMVGRLLNDVCSSER
1132-2-5   SEELEILFMALYNTVNELAERARVEQGRSVKEFLVKLWVEILSAFKIELDTWSNGTQQSFDEYISSSWLSNGSRLTGLLTMQFVGVKLSDEMLMSEECTDLARHVCMVGRLLNDVCSSER

                  490       500       510       520       530       540       550       560       570
SsTps1132  EREENIAGKSYSILLATEKDGRKVSEDEAIAEINEMVEYHWRKVLQIVYKKESILPRRCKDVFLEMAKGTFYAYGINDELTSPQQSKEDMKSFVF
1132-2-5   EREENIAGKSYSILLATEKDGRKVSEDEAIAEINEMVEYHWRKVLQIVYKKESILPRRCKDVFLEMAKGTFYAYGINDELTSPQQSKEDMKSFVF
```

**Figura 6**

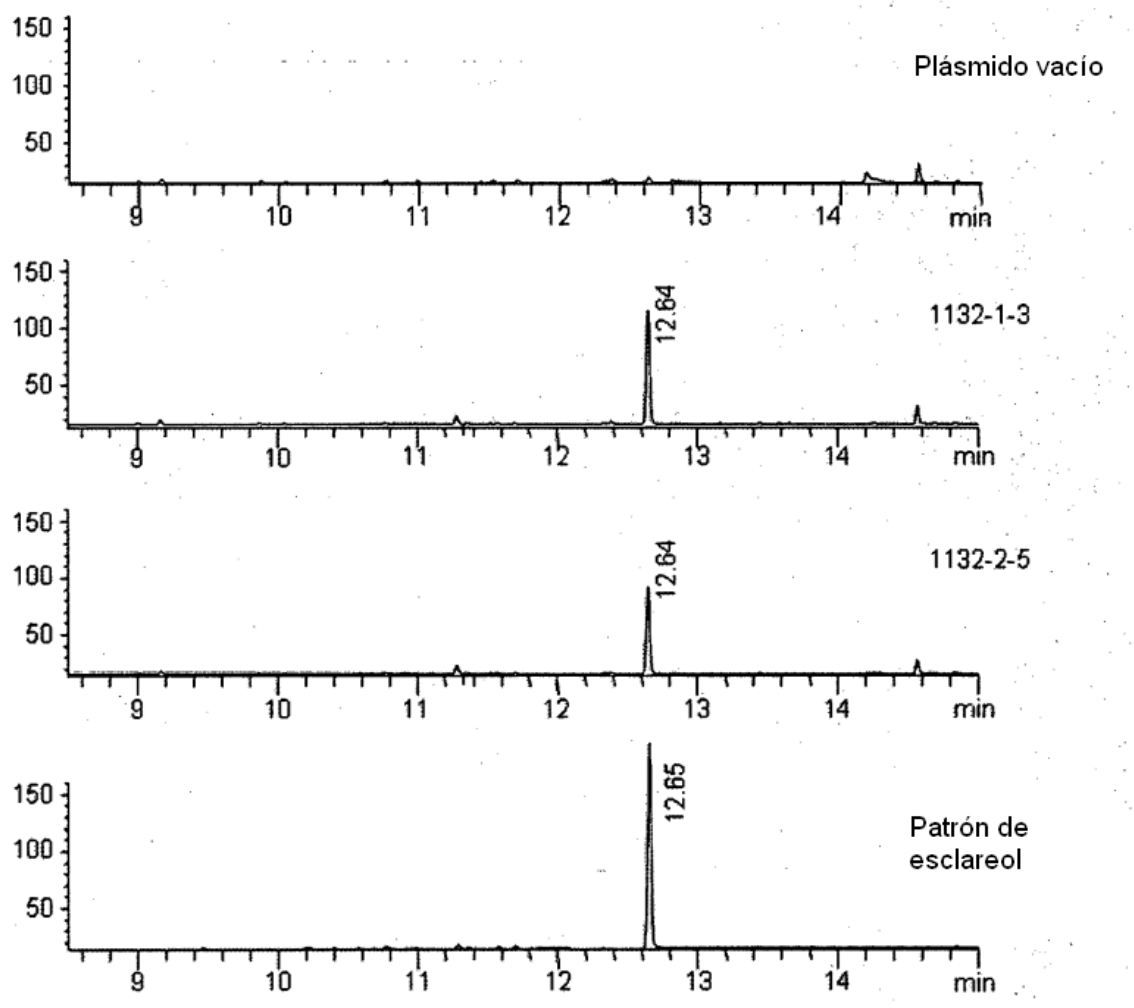

150
100
50
9
10
11
12
13
14
min
Plásmido vacío
12.64
1132-1-3
12.64
1132-2-5
12.65
Patrón de esclareol

**Figura 7**

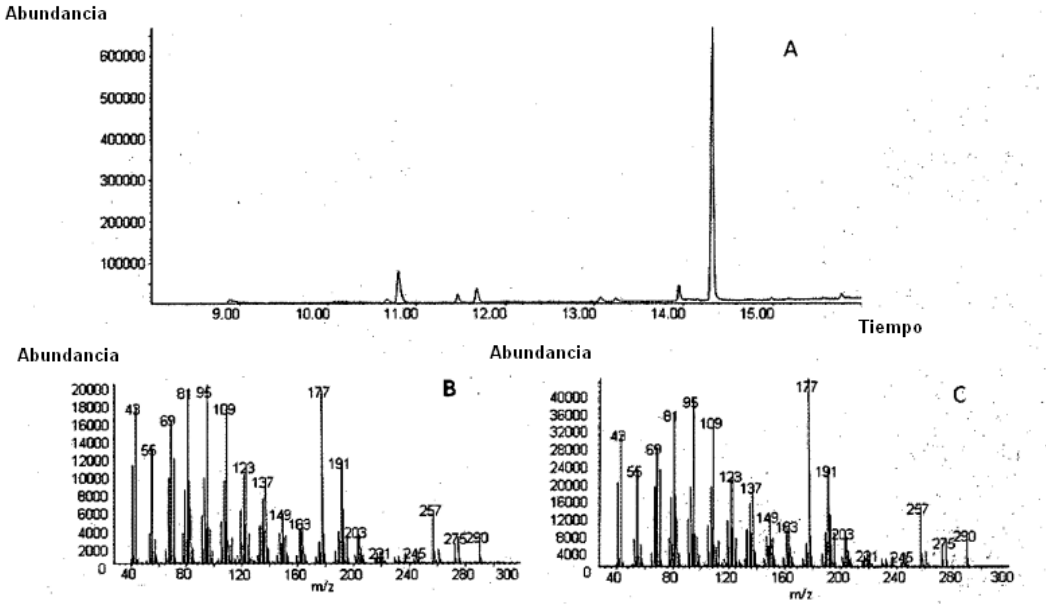
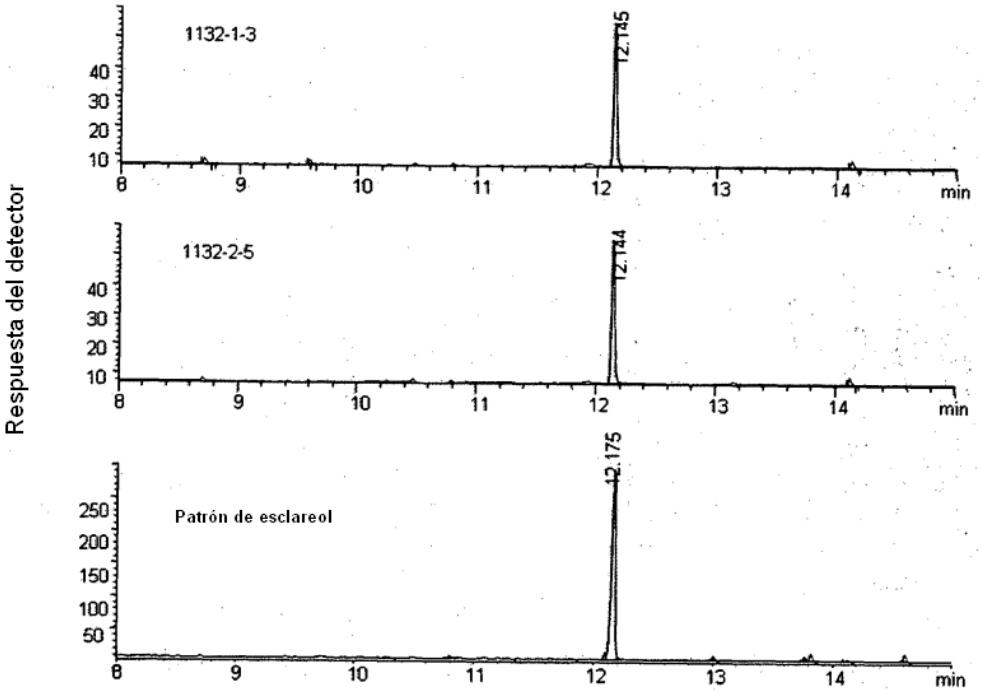


**Figura 8**


Respuesta del detector
1132-1-3
12.145
40
30
20
10
8
9
10
11
12
13
14
min
1132-2-5
12.144
40
30
20
10
8
9
10
11
12
13
14
min
Patrón de esclareol
12.175
250
200
150
100
50
8
9
10
11
12
13
14
min