



República Federativa do Brasil

Ministério do Desenvolvimento, Indústria,
Comércio e Serviços

Instituto Nacional da Propriedade Industrial

(11) PI 0906468-0 B1

(22) Data do Depósito: 12/02/2009

(45) Data de Concessão: 06/06/2023

(54) Título: MÉTODO PARA PRODUZIR ESCLAREOL

(51) Int.Cl.: C12N 15/52; C12N 15/62; C07D 307/92.

(30) Prioridade Unionista: 15/02/2008 EP 08101655.2.

(73) Titular(es): FIRMENICH SA.

(72) Inventor(es): MICHEL SCHALK.

(86) Pedido PCT: PCT EP2009051620 de 12/02/2009

(87) Publicação PCT: WO 2009/101126 de 20/08/2009

(85) Data do Início da Fase Nacional: 29/07/2010

(57) Resumo: MÉTODO PARA PRODUZIR ESCLAREOL. A presente invenção fornece um método de produção de esclareol, método esse compreendendo de contatar um polipeptídeo particular tendo uma atividade de síntese de esclareol com difosfato de labdenediol (LPP). Em particular, o referido método pode ser realizado in vitro ou in vivo para produzir esclareol, um composto muito útil nos campos de perfumaria e aromatizante. A presente invenção também fornece a sequência de aminoácido do polipeptídeo usado no método. Um ácido nucléico derivado da *Salvia sclarea* e que codifica o polipeptídeo da invenção, um vetor de expressão, contendo o referido ácido nucléico, bem como um organismo ou uma célula hospedeira não-humana transformada para hospedar o mesmo ácido nucléico, também são partes da presente invenção.

MÉTODO PARA PRODUZIR ESCLAREOL

Campo técnico

(01) A presente invenção fornece um método de produção de esclareol, método esse compreendendo de contatar, no mínimo, um polipeptídio com geranyl geranyl pirofosfato (GGPP). Em particular, o referido método pode ser realizado in vitro ou in vivo para produzir esclareol, um composto muito útil nos campos de perfumaria e aromatizante. A presente invenção também fornece as sequências de aminoácido de um polipeptídio usado no método da invenção. Um ácido nucléico que codifica o polipeptídio da invenção e um vetor de expressão que contém o referido ácido nucléico são parte da presente invenção. Um organismo hospedeiro não-humano ou uma célula transformada a ser usada no método de produção de esclareol também é um objeto da presente invenção.

Estado da técnica

(02) Os terpenos são encontrados em muitos organismos (microorganismos, animais e plantas). Esses compostos são constituídos de cinco unidades de carbono, chamadas unidades de isopreno e são classificados pela quantidade dessas unidades presentes em sua estrutura. Dessa forma, monoterpenos, sesquiterpenos e diterpenos são terpenos contendo 10, 15 e 20 átomos de carbono, respectivamente. Diterpenos, por exemplo, são amplamente encontrados no reino das plantas e mais de 2500 estruturas de diterpeno foram descritas (Connolly e Hill, Dicionário de terpenóides, 1991, Chapman & Hall, Londres). As moléculas de terpeno foram importantes durante centenas de anos devido a suas propriedades aromatizantes e de fragrância e seus efeitos cosméticos, medicinais e antimicrobianos. Extratos de planta obtidos por diferentes meios, como destilação a vapor ou extração de solvente são usados como fonte de terpenos. As moléculas de

terpeno são, muitas vezes, usadas como tal, mas em alguns casos, reações químicas são usadas para transformar os terpenos em outras moléculas de alto valor.

(03) A produção biossintética de terpenos envolve enzimas chamadas terpeno-sintases. A síntese de um diterpeno pode ser realizada por uma enzima única ou por duas ou mais enzimas. Quando duas enzimas estão envolvidas, a primeira catalisa a síntese de um diterpeno de difosfato éster, que, por sua vez, é convertido ao produto final pela segunda enzima.

(04) Dois tipos de mecanismos de ciclização ocorrem na natureza e são relacionados com dois tipos de diterpeno-sintases, que podem ser classificados em diterpeno-sintases de classe I e classe II (Wendt e Schulz, 1998, *Estrutura*. 6(2):127-33). Para alguns diterpenos, o mecanismo de ciclização é iniciado pela ionização da função de difosfato éster de GGPP, seguida pela reação de carbocation resultante com uma ligação dupla interna. As diterpeno-sintases que catalisam esse tipo de ciclização são diterpeno-sintases de classe I. O segundo modo de ciclização na biossíntese de diterpenos, catalisados por diterpeno-sintases de classe II, é iniciado pela protonação da ligação dupla terminal de GGPP e leva, após rearranjo interno e eliminação de próton, a um intermediário de diterpeno difosfato cíclico.

(05) Genes e cDNAs que codificam diterpeno-sintases de cada uma das duas classes foram clonados e as enzimas recombinantes caracterizadas. A disponibilidade de genes que codificam diferentes tipos de diterpeno-sintases fornece informações sobre as estruturas primárias das enzimas. Alguns motivos de aminoácido são conservados em diterpeno-sintases e são relacionados com a protonação ou a ionização dependente da ciclização. Um motivo DDxxD (onde x representa qualquer aminoácido) é encontrado em diversas diterpeno-sintases de

classe I. O referido motivo provavelmente está envolvido na ligação e na ionização da metade do difosfato. Nas sintases de classe II, um motivo DxDD conservado (onde x representa qualquer aminoácido) é encontrado, no qual o segundo resíduo de aspartato está envolvido como doador de próton.

(06) O esclareol é uma molécula de diterpeno que ocorre naturalmente, extensivamente usada como material primário para a síntese de moléculas de fragrância com notas de âmbar-gris. Essas sínteses foram desenvolvidas para fornecer uma alternativa para o âmbar-gris, uma substância cerácea, secretada pelos intestinos da cachalote. O âmbar-gris é altamente apreciado por seu odor prazeroso e foi historicamente usado como um componente de perfume. Devido a seu alto preço e a elevada demanda de âmbar-gris, e particularmente devido à proteção das espécies de baleia, a síntese química de constituintes e moléculas de âmbar-gris com caráter de âmbar-gris foram desenvolvidas. Entre essas moléculas, Ambrox® (marca registrada da Firmenich SA, Suíça) é o substituto mais amplamente apreciado para o Âmbar-gris. O material primário mais amplamente usado para a síntese de Ambrox® é o diterpeno-diol esclareol.

(07) Geralmente, o preço e a disponibilidade de extratos naturais de planta são dependentes da abundância, do rendimento de óleo e da origem geográfica das plantas. Além disso, a disponibilidade e a qualidade de extratos naturais são muito mais dependentes do clima e de outras condições locais, que levam à variabilidade de ano a ano, tornando o uso desses componentes na perfumaria de alta qualidade muito difícil ou mesmo impossível alguns anos. Portanto, seria uma vantagem fornecer uma fonte de esclareol, que seja menos sujeito a flutuações na disponibilidade e na qualidade. A síntese química pareceria ser uma opção evidente para a preparação do esclareol. Entretanto, dada a sua estrutura altamente

complexa, um processo sintético econômico para a preparação de esclareol ainda é difícil. Uma via bioquímica que leva à síntese de esclareol seria, portanto, de grande interesse.

(08) A biossíntese de terpenos nas plantas e em outros organismos foi extensivamente estudada e não é muito detalhada no presente, mas referência é feita para Dewick, *Nat. Prod. Rep.*, 2002, 19, 181-222, que revisa o estado da técnica das vias biossintéticas de terpeno.

(09) Diversas diterpeno-sintases já foram identificadas. Em particular, terpeno-sintases, que possuem uma certa porcentagem de identificação de sequência com as sequências da presente invenção, também foram encontradas nos bancos de dados de sequências. Todavia, a porcentagem de identificação entre as diterpeno-sintases conhecidas e os polipeptídios da invenção é muito baixa.

(010) As sintases mais próximas às LPP (Labdenediol difosfato) -sintases da invenção são duas copalil difosfato-sintases (uma da *Solanum lycopersicum* (BAA84918) e uma da *Cucurbita maxima* (AAD04293 e AAD04292)), uma copalil difosfato-sintase putativa da *Scoparia dulcis* (BAD91286) e uma proteína hipotética da *Vitis vinifera*. As sequências dessas proteínas compartilham apenas 41% de identificação com as LPP-sintases da invenção e nenhuma dessas sequências é descrita como sendo útil na produção de esclareol, e como tendo atividade de LPP-sintase.

(011) As sintases mais próximas à esclareol-sintase da invenção são uma terpenóide ciclase de função não definida (Número de acesso NCBI AAS98912) tendo 36% de identificação com o polipeptídio da invenção, uma ent-caureno-sintase de *Cucumis sativus* (número de acesso BAB19275) tendo 32% de identificação com o polipeptídio da invenção, uma ent-cassadieno-sintase de *Oryza sativa* (número de

acesso ABH10734 e publicada em Xu, Wilderman, Morrone Xu, Roy, Margis-Pinheiro, Upadhyaya, Coates e Peters, Caracterização funcional da família genética semelhante à caureno-sintase do arroz, *Fitoquímica* 68(3), 2007, 312-326) tendo 32% de identificação com o polipeptídio da invenção e uma ent-caureno-sintase de *Oryza sativa* (número de acesso AAQ72559 e publicada em Margis-Pinheiro, Zhou, Zhu, Dennis e Upadhyaya, Isolamento e caracterização de arroz marcado com DS (*Oryza sativa* L.) Mutante anão responsivo a GA, defeituoso em uma etapa inicial da via de biossíntese de giberelinas, *Plant Cell Rep.*, 23(12), 2005, 819-833) tendo 32% de identificação com a esclareol-sintase fornecida na presente invenção. Além disso, nenhuma dessas sequências é descrita como sendo útil na produção de esclareol e, em particular, como sendo capaz de catalisar a transformação de LPP em esclareol.

(012) Além da diferença entre as próprias sequências, também tem de ser salientado que a estrutura e as propriedades dos produtos sintetizados pelas enzimas acima mencionadas são muito diferentes daquelas do esclareol e LPP. As propriedades do copalil difosfato são muito diferentes daquelas do labdenediol difosfato (LPP). Em particular, diferente de LPP, o copalil difosfato não é usado como um produto intermediário da biossíntese de esclareol. Ent-caureno e ent-cassadieno também são muito diferentes do esclareol. Ent-caureno é um diterpeno tricíclico, que não contém nenhum grupo funcional de álcool, diferente do esclareol, que é um diol bicíclico. Além disso, ent-caureno, que é um precursor de um hormônio de planta regulador de crescimento, está fora de uso no campo da perfumaria e aromatização, considerando-se que o esclareol é de alto interesse nesses campos técnicos, conforme explicado acima.

(013) Um documento do estado da técnica relaciona-se especificamente com uma esclareol-sintase (Banthorpe, Brown e Morris, Purificação parcial de farnesil

pirofosfato: Drimenol ciclase e geranil geranil pirofosfato: Esclareol ciclase, usando cultura celular como uma fonte de material, Fitoquímica 31, 1992, 3391-3395). Nessa referência, uma proteína parcialmente purificada de *Nicotiana glutinosa* é identificada como uma esclareol-sintase, mas nenhuma indicação é fornecida, com relação à sequência de aminoácido dessa proteína, a sequência de nucleotídeo do ácido nucléico que a codifica ou o uso dessa proteína em um método para a biossíntese de esclareol in vitro ou in vivo. Além disso, esse documento não ensina ou mesmo sugere que duas proteínas (uma diterpeno-sintase de classe I e uma de classe II) estejam envolvidas na catálise da transformação de GGPP em esclareol. Ao contrário, este ensina que uma proteína única parcialmente purificada é responsável pela síntese de esclareol.

(014) A WO 2008/007031 divulga uma proteína que tem uma atividade de síntese de *syn*-copalil-8-ol difosfato, a sequência de nucleotídeo que codifica a referida proteína, bem como um vetor e um organismo não-humano transgênico que compreende o referido ácido nucléico. Essa *syn*-copalil-8-ol difosfato-sintase, entretanto, é muito diferente do polipeptídeo da invenção, porque a proteína lá divulgada tem uma sequência de aminoácido apenas 44% idêntica à presente LPP-sintase, usada nos métodos da presente invenção no polipeptídeo de fusão da invenção.

(015) É um objetivo da presente invenção fornecer métodos para criar o esclareol em uma maneira econômica, conforme indicado acima. Conformemente, a presente invenção tem o objetivo de produzir esclareol ao mesmo tempo em que tem pouco resíduo, um processo eficiente de mais energia e recurso e ao mesmo tempo em que reduz a dependência de combustíveis fósseis. É um objetivo adicional fornecer enzimas capazes de sintetizar o esclareol, que é útil como componentes de

perfumaria e/ou aroma.

(016) Nenhum documento do estado da técnica divulga um processo da produção biossintética de esclareol a partir do precursor diterpeno acíclico GGPP.

Abreviações Usadas

bp	par de base
kb	quilo base
BSA	albumina sérica bovina
DMAPP	dimetilalil difosfato
DNA	ácido desoxirribonucléico
cDNA	DNA complementar
dT	deoxi timina
dNTP	deoxi nucleotídeo trifosfato
DTT	ditiotreitól
FPP	farnesil pirofosfato
GC	cromatografia gasosa
GGPP	Geranil geranil pirofosfato:
idi	isopentenil difosfato isomerase
IPP	isopentenil difosfato
IPTG	isopropil-D-tiogalactopiranosídeo
LB	caldo de lisogenia
LPP	labdenediol difosfato
MOPSO	ácido sulfônico 3-(N-morfolino)-2-hidroxipropano
MS	espectrômetro de massa
mvaK1	mevalonato quinase
mvaK2	mevalonato difosfato quinase

PCR	reação em cadeia da polimerase
3’-/5’-RACE	amplificação rápida de 3’ e 5’ de extremidades de cDNA
RMCE	troca de cassete mediada por recombinase
RT-PCR	reação em cadeia da polimerase via transcrição reversa
RNA	ácido ribonucléico
mRNA	ácido ribonucléico mensageiro
RuBisCO	ribulose-1,5-bifosfato carboxilase
SDS-PAGE	eletroforese em gel de poliacrilamida - SDS
SsLPPs	Salvia sclarea labdenediol difosfato-sintase

Descrição da invenção

(017) A presente invenção fornece um método para produzir biossinteticamente esclareol em uma maneira econômica, confiável e reprodutível.

(018) Conforme proposto no presente pedido, esclareol, GGPP, LPP e todos os outros compostos citados no presente pedido são definidos pela maneira de suas fórmulas, conforme representado na figura 1.

(019) Uma “diterpeno-sintase” ou “um polipeptídio tendo uma atividade de síntese de diterpeno” é proposto para o propósito do presente pedido como um polipeptídio capaz de catalisar a síntese de uma molécula de terpeno do precursor de terpeno acíclico GGPP ou de um diterpeno de difosfato éster, como LPP, ou capaz de catalisar a síntese de um diterpeno de difosfato éster do precursor de diterpeno acíclico GGPP.

(020) Como uma "LPP-sintase" ou como um “polipeptídio tendo uma atividade de síntese de LPP”, referimo-nos aqui a um polipeptídio capaz de catalisar a síntese de LPP começando do GGPP.

(021) Como uma "esclareol-sintase" ou como um “polipeptídio tendo uma atividade

de síntese de esclareol”, referimo-nos aqui a um polipeptídio capaz de catalisar a síntese de esclareol começando do LPP.

(022) Como um “polipeptídio capaz de catalisar a transformação de GGPP em esclareol”, referimo-nos aqui a qualquer polipeptídio que seja capaz de catalisar a referida transformação e, em particular, polipeptídios que catalisam um mecanismo de duas etapas, durante o qual o LPP é sintetizado como produto intermediário.

(023) A capacidade de um polipeptídio catalisar a síntese de um diterpeno particular (por exemplo, esclareol) e/ou de um diterpeno de difosfato éster particular (por exemplo, LPP) pode ser simplesmente confirmada pela realização do ensaio enzimático, conforme detalhado no Exemplo 3.

(024) De acordo com a presente invenção, os polipeptídios também devem incluir polipeptídios truncados, considerando-se que eles mantenham sua atividade de síntese de diterpeno, conforme definido em qualquer uma das configurações acima e que eles compartilhem, no mínimo, a porcentagem definida de identificação com o fragmento correspondente de SEQ ID NO:1 ou 2 ou de SEQ ID NO:3. Particularmente, polipeptídios truncados úteis são aqueles com uma deleção no terminal N do sinal de direcionamento de plastídeo.

(025) A porcentagem de identificação entre duas sequências peptídicas ou nucleotídicas é uma função da quantidade de resíduos de aminoácidos ou ácidos nucléicos que são idênticos nas duas sequências, quando um alinhamento dessas duas sequências foi gerado. Resíduos idênticos são definidos como resíduos que são os mesmos nas duas sequências em uma dada posição do alinhamento. A porcentagem de identificação de sequência, conforme usada no presente, é calculada a partir do alinhamento favorável, pegando a quantidade de resíduos idênticos entre as duas sequências, dividindo-a pela quantidade total de resíduos na

sequência mais curta e multiplicando por 100. O alinhamento favorável é o alinhamento no qual a porcentagem de identificação é a maior possível. Lacunas podem ser introduzidas em uma ou ambas as sequências em uma ou mais posições do alinhamento para obter o alinhamento favorável. Essas lacunas são, então, levadas em consideração como resíduos não-idênticos para o cálculo da porcentagem de identificação de sequência.

(026) O alinhamento para fins de determinação da porcentagem de identificação da sequência de aminoácido ou ácido nucléico pode ser atingido de várias maneiras, usando programas de computador e, por exemplo, programas de computador publicamente disponíveis na rede mundial. Preferivelmente, o programa BLAST (Tatiana *et al*, FEMS Microbiol Lett., 1999, 174:247-250, 1999) estabelecido para parâmetros padrão, disponível a partir do Centro Nacional de Informações de Biotecnologia (NCBI) em <http://www.ncbi.nlm.nih.gov/BLAST/bl2seq/wblast2.cgi>, pode ser usado para obter um alinhamento favorável de sequências peptídicas ou nucleotídicas e calcular a porcentagem de identificação de sequência.

(027) Um objetivo da presente invenção é, portanto, um método de produção do esclareol compreendendo

contato de GGPP com, no mínimo, um polipeptídio tendo uma atividade de síntese de LPP e compreendendo uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:1 ou 2;

contato do produto intermediário produzido na etapa a) com, no mínimo, um polipeptídio tendo uma atividade de síntese de esclareol e compreendendo uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:3; e
opcionalmente, isolamento do esclareol produzido na etapa b).

(028) De acordo com uma configuração preferida, as etapas a) e b) do método de

produção de esclareol são realizadas simultaneamente pelo contato do GGPP com o referido, no mínimo um, polipeptídio tendo uma atividade de síntese de LPP e com o referido, no mínimo um, polipeptídio tendo uma atividade de síntese de esclareol em conjunto.

(029) Para o propósito do presente pedido, dizendo que as etapas a) e b) são realizadas simultaneamente, referimo-nos que apenas uma ação é necessária para a pessoa, que quer realizar a invenção, atingir o resultado de ambas as etapas, isto é, contato do GGPP com, no mínimo, dois polipeptídios ou com, no mínimo, um polipeptídio de fusão, conforme descrito abaixo. Todavia, a produção de esclareol ainda será realizada em um mecanismo de duas etapas, conforme ilustrado pela figura 2. LPP é primeiramente sintetizado in situ a partir do GGPP pela LPP-sintase ou pela parte do polipeptídio de fusão, tendo a sequência da LPP-sintase na presença da esclareol-sintase. O LPP assim produzido é, dessa forma, contatado diretamente com a esclareol-sintase ou com a parte do polipeptídio de fusão, tendo a sequência da esclareol-sintase, enzima essa que catalisa a transformação desse precursor em esclareol tão logo este for produzido.

(030) De acordo com uma configuração preferida, as etapas a) e b) do método de produção de esclareol são realizadas simultaneamente através do contato do GGPP com, no mínimo, um polipeptídio de fusão, capaz de catalisar a transformação de GGPP em esclareol e que compreende uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:1 ou 2 e uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:3.

(031) De acordo com uma configuração preferida, o polipeptídio de fusão capaz de catalisar a transformação de GGPP em esclareol compreende a sequência de um polipeptídio tendo uma atividade de síntese de LPP e uma sequência de

aminoácido, no mínimo, 50% idêntica à SEQ ID NO:1 ou 2 e a sequência de um polipeptídio, tendo uma atividade de síntese de esclareol e uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:3.

(032) O método pode ser realizado in vitro, bem como in vivo, já que será explicado em detalhes posteriormente.

(033) O polipeptídio que entrará em contato com LPP in vitro pode ser obtido por extração de qualquer organismo que o expresse, usando tecnologias de extração de proteína ou enzima padrão. Se o organismo hospedeiro for um organismo unicelular ou célula que libera o polipeptídio da invenção no meio de cultura, o polipeptídio pode simplesmente ser coletado do meio de cultura, por exemplo, por centrifugação, opcionalmente seguida por etapas de lavagem e ressuspensão em soluções tampão adequadas. Se o organismo ou a célula acumular o polipeptídio em suas células, o polipeptídio pode ser obtido por ruptura ou lise das células e, ainda, extração do polipeptídio do lisado celular.

(034) O polipeptídio, que tem uma atividade de síntese de esclareol e/ou o polipeptídio, que tem uma atividade de síntese de LPP, ou em uma forma isolada ou juntamente com outras proteínas, por exemplo, em um extrato de proteína crua, obtido de células ou microorganismos cultivados, podem, então, ser suspensos em uma solução tampão em pH favorável. Se adequado, sais, BSA, DTT e outros tipos de cofatores enzimáticos podem ser adicionados a fim de otimizar a atividade enzimática. Condições adequadas são descritas em mais detalhes nos Exemplos posteriormente.

(035) O precursor GGPP ou LPP pode, então, ser adicionado à suspensão ou solução, que é, então, incubada em temperatura favorável, por exemplo, entre 15 e 40°C, preferivelmente entre 25 e 35°C, mais preferivelmente em 30°C. Após a

incubação, o LPP ou o esclareol produzido pode ser isolado da solução incubada por procedimentos de isolamento padrão, como extração de solvente e destilação, opcionalmente após remoção dos polipeptídios da solução.

(036) De acordo com outra configuração preferida, o método de produção de esclareol é realizado in vivo. Nesse caso, as etapas a) e b) do método acima descrito são realizadas simultaneamente e compreendem o cultivo de um organismo ou célula hospedeira não-humana, capaz de produzir GGPP e transformada para expressar, no mínimo, um polipeptídio que compreende uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:1 ou 2 e tendo uma atividade de síntese de LPP e, no mínimo, um polipeptídio que compreende uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:3 e tendo uma atividade de síntese de esclareol sob condições que levam à produção de esclareol.

(037) De acordo com uma configuração mais preferida, o método ainda compreende, antes da etapa a) e b), a transformação de um organismo ou célula não-humana, capaz de produzir GGPP com, no mínimo, um ácido nucléico, que codifica um polipeptídio, que compreende uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:1 ou 2 e tendo uma atividade de síntese de LPP e com, no mínimo, um ácido nucléico, codificando um polipeptídio, que compreende uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:3 e tendo uma atividade de síntese de esclareol, de maneira que o referido organismo expresse os referidos polipeptídios.

(038) Essas configurações da invenção são particularmente vantajosas já que é possível realizar o método in vivo sem previamente isolar o polipeptídio. A reação ocorre diretamente dentro do organismo ou da célula transformada para expressar o referido polipeptídio.

(039) O organismo ou célula hospedeira não-humana pode ser transformada com ambos os ácidos nucléicos ao mesmo tempo ou separadamente. Quando o organismo ou célula hospedeira não-humana é transformada com ambos os ácidos nucléicos ao mesmo tempo, esses ácidos nucléicos podem ser incorporados em um vetor único ou em diferentes vetores.

(040) De acordo com uma configuração particular da invenção, no mínimo, o ácido nucléico que codifica a LPP-sintase compreende uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:4, 5 ou ao complemento da mesma. De acordo com uma configuração mais preferida, o referido ácido nucléico compreende a sequência de nucleotídeo SEQ ID NO:4, 5 ou o complemento da mesma. De acordo com uma configuração ainda mais preferida, o referido ácido nucléico consiste de SEQ ID NO:4, 5 ou o complemento da mesma.

(041) De acordo com uma configuração particular da invenção, no mínimo, o ácido nucléico que codifica a esclareol-sintase compreende uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:6 ou ao complemento da mesma. De acordo com uma configuração mais preferida, o referido ácido nucléico compreende a sequência de nucleotídeo SEQ ID NO:6 ou o

complemento da mesma. Em uma configuração ainda mais preferida, o referido ácido nucléico consiste de SEQ ID NO:6 ou o complemento da mesma.

(042) Em outra configuração particular da invenção, o organismo ou célula hospedeira não-humana pode também ser transformada com, no mínimo, um ácido nucléico, que codifica um polipeptídio de fusão, capaz de catalisar a transformação de GGPP em esclareol e compreendendo uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:1 ou 2 e uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:3.

(043) De acordo com uma configuração preferida, o ácido nucléico que codifica um polipeptídio de fusão compreende uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:4, 5 ou ao complemento da mesma. De acordo com uma configuração mais preferida, o referido ácido nucléico compreende a sequência de nucleotídeo SEQ ID NO:4, 5 ou o complemento da mesma.

(044) De acordo com outra configuração preferida, o ácido nucléico que codifica um polipeptídio de fusão compreende uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:6 ou ao complemento da mesma. De acordo com uma configuração mais preferida, o referido ácido nucléico compreende a sequência de

nucleotídeo SEQ ID NO:6 ou o complemento da mesma.

(045) De acordo com uma configuração ainda mais preferida, o ácido nucléico que codifica um polipeptídeo de fusão consiste de SEQ ID NO:4, 5 ou o complemento da mesma e uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:6 ou ao complemento da mesma. Alternativamente, o ácido nucléico que codifica o polipeptídeo de fusão consiste de SEQ ID NO:6 ou o complemento da mesma e de uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, preferivelmente, no mínimo, 95%, mais preferivelmente, no mínimo, 98% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:4, 5 ou ao complemento da mesma. Alternativamente, o ácido nucléico que codifica o polipeptídeo de fusão consiste de uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, ainda mais preferivelmente, no mínimo, 95% e mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:4, 5 ou ao complemento da mesma e de uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente,

no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, preferivelmente, no mínimo, 95%, mais preferivelmente, no mínimo, 98% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:6 ou ao complemento da mesma. Em uma configuração mais preferida, o ácido nucléico que codifica o polipeptídio de fusão consiste de SEQ ID NO:4, 5 ou o complemento da mesma e de SEQ ID NO:6 ou o complemento da mesma.

(046) De acordo com outra configuração preferida, o ácido nucléico que codifica o polipeptídio de fusão compreende uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% idêntica e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:87. De acordo com uma configuração ainda mais preferida, o ácido nucléico que codifica o polipeptídio de fusão compreende a sequência de nucleotídeo SEQ ID NO:87. De acordo com uma configuração mais preferida, o ácido nucléico que codifica o polipeptídio de fusão consiste de SEQ ID NO:87.

(047) O organismo ou célula não-humana pode vantajosamente ser ainda transformada com, no mínimo, um gene que codifica um polipeptídio envolvido no metabolismo da produção de GGPP, como, por exemplo, enzimas da via do MEP, da via do MVA e/ou prenil transferases. A transformação de um organismo ou célula não-humana, capaz de produzir GGPP com uma LPP-sintase e uma esclareol-sintase, ou com um polipeptídio de fusão, conforme descrito em qualquer uma das

configurações da invenção, é suficiente para a produção de esclareol. Todavia, a transformação adicional com, no mínimo, uma enzima envolvida na produção de GGPP e/ou de um precursor de GGPP, isto é, isopentenil difosfato (IPP) e dimetilalil difosfato (DMAPP), tem a vantagem de aumentar a quantidade de precursor disponível para a conversão em esclareol.

(048) O organismo ou a célula deve “expressar” um polipeptídio, considerando-se que o organismo ou a célula seja transformada para hospedar um ácido nucléico que codifica o referido polipeptídio, esse ácido nucléico é transcrito para mRNA e o polipeptídio é encontrado no organismo ou célula hospedeira. O termo “expressar” abrange “expressar heterologicamente” e “super-expressar”, o último referindo-se aos níveis de mRNA, polipeptídio e/ou atividade de enzima sobre e acima do que é medido em um organismo ou uma célula não-transformados. Uma descrição mais detalhada dos métodos adequados para transformar um organismo ou uma célula não-humana será descrita posteriormente na parte da especificação, que é dedicada a esses organismos ou células hospedeiros não-humanos, transformados como objetos específicos da presente invenção e nos Exemplos.

(049) Um organismo ou célula particular deve ser “capaz de produzir GGPP”, quando este produz GGPP naturalmente ou quando não produz GGPP naturalmente, mas é transformado para produzir GGPP, ou antes da transformação com um ácido nucléico, conforme descrito na presente ou juntamente com o referido ácido nucléico. Organismos ou células transformados para produzir uma quantidade maior de GGPP do que o organismo ou célula que ocorre naturalmente também são abrangidos pelos “organismos ou células capazes de produzir GGPP”. Métodos para transformação de organismos, por exemplo, microorganismos, de maneira que eles produzam GGPP já são conhecidos na técnica. Esses métodos podem, por

exemplo, ser encontrados em Huang, Roessner, Croteau e Scott, Engenharia de *Escherichia coli* para a síntese de taxadieno, um intermediário importante na biossíntese de taxol, *Bioorg Med Chem.*, 9(9), 2001, 2237-2242.

(050) De acordo com uma configuração preferida, o organismo acumula GGPP naturalmente ou é transformado para acumular esse precursor.

(051) Para realizar a invenção in vivo, o organismo ou a célula hospedeira é cultivada sob condições que levam à produção de esclareol. Conformemente, se o hospedeiro for uma planta transgênica, condições de crescimento favorável são fornecidas, como condições de luz, água e nutriente favoráveis, por exemplo. Se o hospedeiro for um organismo unicelular, condições que levam à produção de esclareol podem incluir a adição de cofatores adequados ao meio de cultura do hospedeiro. Além disso, um meio de cultura pode ser selecionado, de maneira a maximizar a síntese de esclareol. Condições de cultura favoráveis são descritas em uma maneira mais detalhada nos seguintes Exemplos.

(052) Organismos hospedeiros não-humanos adequados para realizar o método da invenção in vivo podem ser quaisquer organismos multicelulares ou unicelulares não-humanos. Em uma configuração preferida, o organismo hospedeiro não-humano, usado para realizar a invenção in vivo é uma planta, um procarionte ou um fungo. Qualquer planta, procarionte ou fungo pode ser usado. Particularmente, plantas úteis são aquelas que naturalmente produzem altas quantidades de terpenos. Em uma configuração mais preferida, a planta é selecionada da família das Solanaceae, Poaceae, Brassicaceae, Fabaceae, Malvaceae, Asteraceae ou Lamiaceae. Por exemplo, a planta é selecionada dos gêneros *Nicotiana*, *Solanum*, *Sorghum*, *Arabidopsis*, *Brassica* (rapé), *Medicago* (alfalfa), *Gossypium* (algodão), *Artemisia*, *Salvia* e *Mentha*. Preferivelmente, a planta pertence às espécies de

Nicotiana tabacum.

(053) Em uma configuração mais preferida, o organismo hospedeiro não-humano, usado para realizar a invenção in vivo é um microorganismo. Qualquer microorganismo pode ser usado, mas de acordo com uma configuração ainda mais preferida, o referido microorganismo é uma bactéria ou um fungo. Preferivelmente, o referido fungo é uma levedura. Mais preferivelmente, a referida bactéria é *E. coli* e a referida levedura é *Saccharomyces cerevisiae*.

(054) Diversos desses organismos não produzem GGPP naturalmente. Para serem adequados para realizar o método da invenção, esses organismos tem de ser transformados para produzir o referido precursor. Eles podem ser transformados antes da modificação com o ácido nucléico descrito, de acordo com qualquer uma das configurações acima ou simultaneamente, conforme explicado acima.

(055) Células eucarióticas maiores isoladas podem também ser usadas, em vez de organismos completos, como hospedeiros para realizar o método da invenção in vivo. Células eucarióticas adequadas podem ser qualquer célula não-humana, mas são preferivelmente células de planta.

(056) De acordo com uma configuração preferida, no mínimo, um polipeptídio tendo uma atividade de síntese de LPP, usado em qualquer uma das configurações acima descritas ou codificado pelo ácido nucléico, usado em qualquer uma das configurações acima descritas compreende uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:1 ou 2. De acordo com uma configuração mais

preferida, o referido polipeptídio compreende a sequência de aminoácido SEQ ID NO:1 ou 2. Em uma configuração ainda mais preferida, o referido polipeptídio consiste de SEQ ID NO:1 ou 2.

(057) De acordo com outra configuração preferida, no mínimo, um polipeptídio tendo uma atividade de síntese de esclareol, usado em qualquer uma das configurações acima descritas ou codificado pelo ácido nucléico, usado em qualquer uma das configurações acima descritas compreende uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:3. De acordo com uma configuração mais preferida, o referido polipeptídio compreende a sequência de aminoácido SEQ ID NO:3. Em uma configuração ainda mais preferida, o referido polipeptídio consiste de SEQ ID NO:3.

(058) De acordo com uma configuração preferida adicional, o polipeptídio de fusão, usado em qualquer uma das configurações acima descritas ou codificado pelo ácido nucléico de qualquer uma das configurações acima descritas compreende uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:1 ou 2. De acordo com uma configuração mais preferida, o referido polipeptídio de fusão compreende a sequência de aminoácido SEQ ID NO:1 ou 2.

(059) De acordo com outra configuração preferida, o polipeptídio de fusão, usado em qualquer uma das configurações acima descritas ou codificado pelo ácido nucléico de qualquer uma das configurações acima descritas compreende uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:3. De acordo com uma configuração mais preferida, o referido polipeptídio de fusão compreende a sequência de aminoácido SEQ ID NO:3.

(060) De acordo com uma configuração ainda mais preferida, o polipeptídio de fusão, usado em qualquer uma das configurações acima descritas ou codificado pelo ácido nucléico de qualquer uma das configurações acima descritas consiste de SEQ ID NO:1 ou 2 e de uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:3. Alternativamente, este consiste de SEQ ID NO:3 e de uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:1 ou 2.

Alternativamente, o polipeptídio de fusão consiste de uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:1 ou 2 e de uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:3. Em uma configuração mais preferida, o polipeptídio de fusão consiste de SEQ ID NO:1 ou 2 e de SEQ ID NO:3.

(061) De acordo com outra configuração particularmente preferida, o polipeptídio de fusão, usado em qualquer uma das configurações acima descritas ou codificado pelo ácido nucléico de qualquer uma das configurações acima descritas compreende uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:88. De acordo com uma configuração mais preferida, o polipeptídio de fusão compreende a sequência de aminoácido SEQ ID NO:88. De acordo com uma configuração ainda mais preferida, o polipeptídio de fusão consiste de SEQ ID NO:88.

(062) De acordo com uma configuração específica adicional da invenção, o polipeptídio tendo uma atividade de síntese de LPP, conforme proposto em qualquer configuração do método da invenção, é um polipeptídio que compreende uma sequência de aminoácido, no mínimo, 50% idêntica a qualquer uma das SEQ ID NO:36 ou 39, que são formas truncadas da SEQ ID NO:1. Preferivelmente, o referido polipeptídio compreende uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:36 a 39. De acordo com uma configuração mais preferida, o referido polipeptídio compreende de qualquer uma das SEQ ID NO:36 a 39. De acordo com uma configuração ainda mais preferida, o referido polipeptídio consiste de qualquer uma das SEQ ID NO:36 a 39.

(063) De acordo com outra configuração específica da invenção, o polipeptídio tendo uma atividade de síntese de esclareol, conforme proposto em qualquer configuração do método da invenção, é um polipeptídio que compreende uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:74, que é uma forma truncada da SEQ ID NO:3. Preferivelmente, o referido polipeptídio compreende uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:74. De acordo com uma configuração mais preferida, o referido polipeptídio

compreende a SEQ ID NO:74. De acordo com uma configuração ainda mais preferida, o referido polipeptídio consiste da SEQ ID NO:74.

(064) De acordo com outra configuração particularmente preferida, o polipeptídio de fusão, usado em qualquer uma das configurações acima descritas ou codificado pelo ácido nucléico de qualquer uma das configurações acima descritas compreende uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:36 a 39. Em uma configuração ainda mais preferida, este compreende qualquer uma das SEQ ID NO:36 a 39.

(065) De acordo com outra configuração particularmente preferida, o polipeptídio de fusão, usado em qualquer uma das configurações acima descritas ou codificado pelo ácido nucléico de qualquer uma das configurações acima descritas compreende uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:74. Em uma configuração ainda mais preferida, este compreende a SEQ ID NO:74.

(066) Em uma configuração ainda mais preferida, o polipeptídio de fusão, usado em qualquer uma das configurações acima descritas ou codificado pelo ácido nucléico de qualquer uma das configurações acima descritas consiste de qualquer uma das

SEQ ID NO:36 a 39 e de uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:74. Alternativamente, o polipeptídeo de fusão, usado em qualquer uma das configurações acima descritas ou codificado pelo ácido nucléico de qualquer uma das configurações acima descritas consiste da SEQ ID NO:74 e de uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:36 a 39. Alternativamente, o polipeptídeo de fusão consiste de uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:36 a 39 e de uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda

mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:77. Em uma configuração mais preferida, o polipeptídeo de fusão, usado em qualquer uma das configurações acima descritas ou codificado pelo ácido nucléico de qualquer uma das configurações acima descritas consiste de qualquer uma das SEQ ID NO:36 a 39 e da SEQ ID NO:74.

(067) De acordo com uma configuração adicional, o ácido nucléico que codifica uma LPP-sintase, usado em qualquer uma das configurações acima descritas compreende uma sequência de nucleotídeo, no mínimo, 50% idêntica a qualquer uma das SEQ ID NO:28 a 31, que são formas truncadas da SEQ ID NO:4 ou ao complemento da mesma. Preferivelmente, o referido ácido nucléico compreende uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:28 a 31 ou ao complemento das mesmas. De acordo com uma configuração mais preferida, o referido ácido nucléico compreende qualquer uma das SEQ ID NO:28 a 31 ou o complemento das mesmas. De acordo com uma configuração ainda mais preferida, o ácido nucléico consiste de qualquer uma das SEQ ID NO:28 a 31 ou do complemento das mesmas.

(068) De acordo com outra configuração específica da invenção, o ácido nucléico que codifica uma esclareol-sintase, usado em qualquer uma das configurações acima descritas compreende uma sequência de nucleotídeo, no mínimo, 50% idêntica à SEQ ID NO:73, que é uma forma truncada da SEQ ID NO:6 ou ao complemento da mesma. Preferivelmente, o referido ácido nucléico compreende

uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:73 ou ao complemento da mesma. De acordo com uma configuração mais preferida, o referido ácido nucléico compreende a SEQ ID NO:73 ou o complemento da mesma. De acordo com uma configuração ainda mais preferida, o ácido nucléico consiste da SEQ ID NO:73 ou do complemento da mesma.

(069) De acordo com outra configuração específica, o ácido nucléico, que codifica um polipeptídeo de fusão, conforme usado em qualquer uma das configurações acima descritas compreende uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:28 a 31 ou ao complemento das mesmas. Em uma configuração ainda mais preferida, este compreende a sequência de nucleotídeo de qualquer uma das SEQ ID NO:28 a 31 ou ao complemento das mesmas.

(070) De acordo com outra configuração específica, o ácido nucléico, que codifica um polipeptídeo de fusão, conforme usado em qualquer uma das configurações acima descritas compreende uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo,

75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:73 ou o complemento da mesma. Em uma configuração ainda mais preferida, este compreende a sequência de nucleotídeo de SEQ ID NO:73 ou ao complemento da mesma.

(071) Em uma configuração ainda mais preferida, o ácido nucléico que codifica um polipeptídeo de fusão, conforme usado em qualquer uma das configurações acima descritas consiste de qualquer uma das SEQ ID NO:28 a 31 ou do complemento das mesmas e de uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:73 ou ao complemento da mesma. Alternativamente, o ácido nucléico que codifica um polipeptídeo de fusão, conforme usado em qualquer uma das configurações acima descritas consiste de qualquer uma das SEQ ID NO:73 a 50% ou do complemento das mesmas e de uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:28 a 31 ou ao complemento das mesmas. Alternativamente, o ácido nucléico que codifica o polipeptídeo de fusão consiste de uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente,

no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, ainda mais preferivelmente, no mínimo, 95%, e mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:28 a 31 ou ao complemento das mesmas e de uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:73 ou ao complemento da mesma. Em uma configuração mais preferida, o ácido nucléico que codifica o polipeptídeo de fusão, conforme usado em qualquer uma das configurações acima descritas, consiste de qualquer uma das SEQ ID NO:28 a 31 ou do complemento das mesmas e da SEQ ID NO:73 ou do complemento da mesma.

(072) De acordo com outra configuração preferida, o polipeptídeo ou o ácido nucléico usado no método de qualquer uma das configurações acima é derivado da *Salvia sclarea*.

(073) Uma ferramenta importante para realizar o método da invenção é o próprio polipeptídeo de fusão. Um polipeptídeo de fusão, capaz de catalisar a transformação de GGPP em esclareol e compreendendo uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:1 ou 2 e uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:3 é, portanto, outro objeto da presente invenção.

(074) De acordo com uma configuração preferida, o polipeptídeo de fusão compreende uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no

mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:1 ou 2. De acordo com uma configuração mais preferida, o polipeptídio de fusão compreende a sequência de aminoácido SEQ ID NO:1 ou 2.

(075) De acordo com outra configuração preferida, o polipeptídio de fusão compreende uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:3. De acordo com uma configuração mais preferida, o referido polipeptídio de fusão compreende a sequência de aminoácido SEQ ID NO:3.

(076) Em uma configuração ainda mais preferida, o polipeptídio de fusão consiste da SEQ ID NO:1 ou 2 e de uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:3. Alternativamente, o polipeptídio de fusão consiste da SEQ ID NO:3 e de uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente,

no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:1 ou 2. Em uma configuração mais preferida, o polipeptídio de fusão consiste da SEQ ID NO:1 ou 2 e da SEQ ID NO:3.

(077) De acordo com uma configuração particular da invenção, o polipeptídio de fusão compreende a sequência de uma forma truncada da SEQ ID NO:1, como SEQ ID NO:36 a 39 e/ou a sequência de uma forma truncada da SEQ ID NO:3, como SEQ ID NO:74.

(078) Portanto, de acordo com uma configuração particular da invenção, o polipeptídio de fusão compreende uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:36 a 39. Em uma configuração ainda mais preferida, este compreende qualquer uma das SEQ ID NO:36 a 39.

(079) De acordo com outra configuração específica, o polipeptídio de fusão compreende uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:74. Em uma configuração ainda mais preferida, este compreende a SEQ ID NO:74.

(080) Em uma configuração ainda mais preferida, o polipeptídio de fusão consiste de qualquer uma das SEQ ID NO:36 a 39 e de uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:74. Alternativamente, o polipeptídio de fusão consiste da SEQ ID NO:74 e de uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:36 a 39. Em uma configuração mais preferida, o polipeptídio de fusão consiste de qualquer uma das SEQ ID NO:36 a 39 e da SEQ ID NO:74.

(081) De acordo com outra configuração particularmente preferida, o polipeptídio de fusão da invenção compreende uma sequência de aminoácido, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, ainda mais preferivelmente, no mínimo, 95% e mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:88. De acordo com uma configuração mais preferida, o polipeptídio de fusão compreende a sequência de aminoácido SEQ ID NO:88. De acordo com uma configuração ainda mais

preferida, o polipeptídio de fusão consiste da SEQ ID NO:88.

(082) Conforme usado no presente, o polipeptídio é proposto como um polipeptídio ou fragmento de peptídeo que inclui as sequências de aminoácidos aqui identificadas, bem como polipeptídios truncados ou variantes, considerando-se que eles mantenham sua atividade, conforme definido acima e que compartilhem, no mínimo, a porcentagem definida de identificação com o fragmento correspondente da SEQ ID NO:1, 2 ou 3.

(083) Exemplos de polipeptídios variantes são proteínas que ocorrem naturalmente, que resultam de eventos de combinação de mRNA alternativos ou formam a divisão proteolítica dos polipeptídios aqui descritos. Variações atribuíveis à proteólise incluem, por exemplo, diferenças nos terminais N ou C mediante a expressão em tipos diferentes de células hospedeiras, devido à remoção proteolítica de um ou mais aminoácidos terminais dos polipeptídios da invenção. Polipeptídios codificados por um ácido nucléico, obtido por mutação natural ou artificial de um ácido nucléico da invenção, conforme descrito posteriormente, também são abrangidos pela invenção.

(084) Conforme mencionado acima, o ácido nucléico que codifica o polipeptídio de fusão da invenção é uma ferramenta útil para modificar os organismos ou células hospedeiros não-humanos, propostos para serem usados quando o método for realizado in vivo.

(085) Um ácido nucléico que codifica um polipeptídio de fusão, de acordo com qualquer uma das configurações acima descritas, portanto, também é um objeto da presente invenção.

(086) De acordo com uma configuração preferida, o ácido nucléico compreende uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%,

preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:4, 5 ou ao complemento da mesma. De acordo com uma configuração mais preferida, o ácido nucléico compreende a sequência de nucleotídeo SEQ ID NO:4, 5 ou o complemento da mesma.

(087) De acordo com outra configuração preferida, o ácido nucléico compreende uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:6 ou ao complemento da mesma. De acordo com uma configuração mais preferida, o referido polipeptídeo de fusão compreende a sequência de aminoácido SEQ ID NO:6 ou o complemento da mesma.

(088) Em uma configuração ainda mais preferida, o ácido nucléico consiste da SEQ ID NO:4, 5 ou o complemento da mesma e de uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:6 ou ao complemento da mesma. Alternativamente, o ácido nucléico consiste da SEQ ID NO:6 ou do complemento da mesma e de uma sequência de nucleotídeo, no mínimo, 55%, no mínimo, 60%,

preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, preferivelmente, no mínimo, 95%, mais preferivelmente, no mínimo, e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:4, 5 ou ao complemento da mesma. Alternativamente, o ácido nucleico consiste de uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95%, e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:4, 5 ou ao complemento da mesma e de uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:6 ou ao complemento da mesma. Em uma configuração mais preferida, o ácido nucleico consiste da SEQ ID NO:4, 5 ou do complemento da mesma e da SEQ ID NO:6 ou do complemento da mesma.

(089) Particularmente, ácidos nucleicos úteis são aqueles que codificam polipeptídeos de fusão, compreendendo formas truncadas da SEQ ID NO:1 e/ou 3. Portanto, ácido nucleico que compreende uma forma truncada da SEQ ID NO:4, como SEQ ID NO:28 a 31 ou o complemento das mesmas e/ou que compreende uma forma truncada da SEQ ID NO:6, como SEQ ID NO:73, ou o complemento da mesma são configurações particularmente úteis da invenção.

(090) Portanto, de acordo com outra configuração específica, o ácido nucleico

compreende uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:28 a 31 ou do complemento das mesmas. Em uma configuração ainda mais preferida, este compreende a sequência de nucleotídeo de qualquer uma das SEQ ID NO:28 a 31 ou o complemento das mesmas.

(091) De acordo com outra configuração específica, o ácido nucléico compreende uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:73 ou ao complemento da mesma. Em uma configuração ainda mais preferida, este compreende a sequência de nucleotídeo da SEQ ID NO:73 ou o complemento da mesma.

(092) Em uma configuração ainda mais preferida, o ácido nucléico consiste de qualquer uma das SEQ ID NO:28 a 31 ou do complemento das mesmas e de uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95%, e ainda mais preferivelmente, no mínimo, 98%

idêntica à SEQ ID NO:73 ou ao complemento da mesma. Alternativamente, o ácido nucléico consiste da SEQ ID NO:73 ou do complemento da mesma e de uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95%, e ainda mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:28 a 31 ou ao complemento das mesmas. Em uma configuração mais preferida, o ácido nucléico consiste de qualquer uma das SEQ ID NO:28 a 31 ou do complemento das mesmas e da SEQ ID NO:73 ou do complemento da mesma.

(093) De acordo com outra configuração particularmente preferida, o ácido nucléico compreende uma sequência de nucleotídeo, no mínimo, 50%, preferivelmente, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:87. De acordo com uma configuração mais preferida, o ácido nucléico compreende a sequência de nucleotídeo SEQ ID NO:87. De acordo com uma configuração ainda mais preferida, o ácido nucléico consiste da SEQ ID NO:87.

(094) O ácido nucléico da invenção pode ser definido como incluindo polímeros de desoxirribonucleotídeo ou ribonucleotídeo na forma com filamento simples ou duplo (DNA e/ou RNA). O termo “sequência de nucleotídeo” deve também ser compreendido como incluindo uma molécula de polinucleotídeo ou uma molécula de

oligonucleotídeo na forma de um fragmento separado ou como um componente de um ácido nucléico maior. Ácidos nucléicos da invenção também incluem certas sequências de nucleotídeo isoladas, incluindo aquelas que estão substancialmente livres de conter material endógeno. O ácido nucléico da invenção pode ser truncado, considerando-se que este codifique um polipeptídio abrangido pela presente invenção, conforme descrito acima.

(095) Os ácidos nucléicos, que compreendem uma sequência obtida pela mutação da SEQ ID NO:4, 5 ou do complemento da mesma e/ou uma sequência obtida pela mutação da SEQ ID NO:6 ou do complemento da mesma, também são abrangidos pela invenção, considerando-se que as sequências que eles compreendem compartilhem, no mínimo, a porcentagem definida da identificação com os fragmentos correspondentes da SEQ ID NO: 4, 5 ou 6 ou com o complemento das mesmas e considerando-se que eles codifiquem um polipeptídio de fusão, capaz de catalisar a transformação do GGPP em esclareol conforme definido acima. Mutações podem ser qualquer tipo de mutações desses ácidos nucléicos, como mutações de ponto, mutações de deleção, mutações de inserção e/ou mutações de troca de estrutura. Um ácido nucléico variante pode ser preparado a fim de adaptar sua sequência de nucleotídeo a um sistema de expressão específico. Por exemplo, os sistemas de expressão bacteriana são conhecidos por expressar mais eficazmente os polipeptídios se os aminoácidos forem codificados por um códon preferido. Devido à degeneração do código genético, onde mais de um códon pode codificar o mesmo aminoácido, múltiplas sequências de DNA podem codificar o mesmo polipeptídio, todas essas sequências de DNA sendo abrangidas na invenção.

(096) Outra ferramenta importante para a transformação de organismos ou células

hospedeiros adequados para realizar o método da invenção in vivo é um vetor de expressão, compreendendo um ácido nucléico, de acordo com qualquer configuração da invenção. Esse vetor é, portanto, também um objeto da presente invenção.

(097) Um "vetor da expressão", conforme usado no presente, inclui qualquer vetor recombinante linear ou circular, incluindo, entre outros, vetores virais, bacteriófagos e plasmídeos. O perito é capaz de selecionar um vetor adequado, de acordo com o sistema de expressão. Em uma configuração, o vetor de expressão inclui o ácido nucléico da invenção, operavelmente ligado a, no mínimo, uma sequência regulatória, que controla a transcrição, a tradução, a iniciação e o término, como um promotor, operador ou uma sequência reforçadora transcricional ou um local de ligação ribossomal de mRNA e, opcionalmente, incluindo, no mínimo, um marcador de seleção. Sequências de nucleotídeo são "operavelmente ligadas" quando a sequência regulatória funcionalmente relaciona-se com o ácido nucléico da invenção.

(098) Os vetores de expressão da presente invenção podem ser usados no método para preparação de um organismo e/ou célula hospedeira geneticamente transformada, em organismos e/ou células hospedeiras que abrigam os ácidos nucléicos da invenção e nos métodos de produção ou criação de polipeptídios tendo uma atividade de síntese de esclareol, conforme divulgado mais abaixo.

(099) Organismos e células hospedeiros não-humanos recombinantes, transformados para hospedar, no mínimo, um ácido nucléico que codifica um polipeptídio tendo uma atividade de síntese de LPP e compreendendo uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:1 ou 2 e, no mínimo, um ácido nucléico que codifica um polipeptídio tendo uma atividade de

síntese de esclareol e compreendendo uma sequência de aminoácido, no mínimo, 50% idêntica à SEQ ID NO:3, de maneira que expresse heterologicamente ou super-expresse os referidos polipeptídios também são ferramentas muito úteis para realizar o método da invenção. Esses organismos e células hospedeiras não-humanas são, portanto, outro objeto da presente invenção.

(0100) De acordo com uma configuração preferida, o referido polipeptídio tendo uma atividade de síntese de LPP compreende uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:1 ou 2. De acordo com uma configuração mais preferida, o referido polipeptídio compreende a SEQ ID NO:1 ou 2. De acordo com uma configuração ainda mais preferida, o referido polipeptídio consiste da SEQ ID NO:1 ou 2.

(0101) De acordo com outra configuração preferida, o referido polipeptídio tendo uma atividade de síntese de esclareol compreende uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:3. De acordo com uma configuração mais preferida, o referido polipeptídio compreende a SEQ ID NO:3. De acordo com uma configuração ainda mais preferida, o referido polipeptídio consiste da SEQ ID NO:3.

(0102) De acordo com uma configuração preferida adicional, o ácido nucléico que

codifica um polipeptídio tendo uma atividade de síntese de LPP compreende uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:4, 5 ou ao complemento da mesma. De acordo com uma configuração mais preferida, o referido ácido nucléico compreende a SEQ ID NO:4, 5 ou o complemento da mesma. De acordo com uma configuração ainda mais preferida, o referido ácido nucléico consiste da SEQ ID NO:4, 5 ou do complemento da mesma.

(0103)De acordo com uma configuração preferida adicional, o ácido nucléico que codifica um polipeptídio tendo uma atividade de síntese de esclareol compreende uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:6 ou ao complemento da mesma. De acordo com uma configuração mais preferida, o referido ácido nucléico compreende a SEQ ID NO:6 ou o complemento da mesma. De acordo com uma configuração ainda mais preferida, o referido ácido nucléico consiste da SEQ ID NO:6 ou do complemento da mesma.

(0104)De acordo com outra configuração preferida, o referido polipeptídio tendo uma atividade de síntese de LPP compreende uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente,

no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:36 a 39. De acordo com uma configuração mais preferida, o referido polipeptídio compreende qualquer uma das SEQ ID NO:36 a 39. De acordo com uma configuração ainda mais preferida, o referido polipeptídio consiste de qualquer uma das SEQ ID NO:36 a 39.

(0105) De acordo com outra configuração preferida, o referido polipeptídio tendo uma atividade de síntese de esclareol compreende uma sequência de aminoácido, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:73. De acordo com uma configuração mais preferida, o referido polipeptídio compreende a SEQ ID NO:73. De acordo com uma configuração ainda mais preferida, o referido polipeptídio consiste da SEQ ID NO:73.

(0106) De acordo com uma configuração preferida adicional, o referido ácido nucléico que codifica um polipeptídio tendo uma atividade de síntese de LPP compreende uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica a qualquer uma das SEQ ID NO:28 a 31 ou ao complemento das mesmas. De acordo com uma configuração mais preferida, o referido ácido nucléico compreende qualquer uma das SEQ ID NO:28 a 31 ou o complemento das mesmas. De acordo com uma

configuração ainda mais preferida, o referido ácido nucléico consiste de qualquer uma das SEQ ID NO:28 a 31 ou do complemento das mesmas.

(0107)De acordo com uma configuração preferida adicional, o ácido nucléico que codifica um polipeptídio tendo uma atividade de síntese de esclareol compreende uma sequência de nucleotídeo, no mínimo, 55%, preferivelmente, no mínimo, 60%, preferivelmente, no mínimo, 65%, preferivelmente, no mínimo, 70%, preferivelmente, no mínimo, 75%, preferivelmente, no mínimo, 80%, preferivelmente, no mínimo, 85%, preferivelmente, no mínimo, 90%, mais preferivelmente, no mínimo, 95% e ainda mais preferivelmente, no mínimo, 98% idêntica à SEQ ID NO:73 ou ao complemento da mesma. De acordo com uma configuração mais preferida, o referido ácido nucléico compreende a SEQ ID NO:73 ou o complemento da mesma. De acordo com uma configuração ainda mais preferida, o referido ácido nucléico consiste da SEQ ID NO:73 ou do complemento da mesma.

(0108)De acordo com outra configuração preferida, os organismos e células hospedeiros não-humanos são transformados para hospedar, no mínimo, um ácido nucléico, que codifica um polipeptídio de fusão, conforme descrito em qualquer uma das configurações acima da invenção, de maneira que este expresse heterologicamente ou super-expresse o referido polipeptídio de fusão.

(0109)O organismo ou célula não-humana pode vantajosamente ser ainda transformada com, no mínimo, um gene que codifica um polipeptídio envolvido no metabolismo da produção de GGPP, como, por exemplo, enzimas da via do MEP, da via do MVA e/ou prenil transferases. A transformação de um organismo ou célula não-humana, capaz de produzir GGPP com uma LPP-sintase e uma esclareol-sintase, ou com um polipeptídio de fusão, conforme descrito em qualquer uma das configurações da invenção, é suficiente para a produção de esclareol. Todavia, a

transformação adicional com, no mínimo, uma enzima envolvida na produção de GGPP e/ou de um precursor de GGPP, isto é, isopentenil difosfato (IPP) e dimetilalil difosfato (DMAPP), tem a vantagem de aumentar a quantidade de precursor disponível para a conversão em esclareol.

(0110) Organismos hospedeiros não-humanos da invenção podem ser quaisquer organismos multicelulares ou unicelulares não-humanos. Em uma configuração preferida, o organismo hospedeiro não-humano da invenção é uma planta, um procarionte ou um fungo. Qualquer planta, procarionte ou fungo é adequado para ser transformado, de acordo com a presente invenção. Particularmente, plantas úteis são aquelas que naturalmente produzem altas quantidades de terpenos. Em uma configuração mais preferida, a planta é selecionada da família das *Solanaceae*, *Poaceae*, *Brassicaceae*, *Fabaceae*, *Malvaceae*, *Asteraceae* ou *Lamiaceae*. Por exemplo, a planta é selecionada dos gêneros *Nicotiana*, *Solanum*, *Sorghum*, *Arabidopsis*, *Brassica* (rapé), *Medicago* (alfalfa), *Gossypium* (algodão), *Artemisia*, *Salvia* e *Mentha*. Preferivelmente, a planta pertence às espécies de *Nicotiana tabacum*.

(0111) Em uma configuração mais preferida, o organismo hospedeiro não-humano é um microorganismo. Qualquer microorganismo é adequado para a presente invenção, mas de acordo com uma configuração ainda mais preferida, o referido microorganismo é uma bactéria ou um fungo. Preferivelmente, o referido fungo é levedura. Mais preferivelmente, a referida bactéria é *E. coli* e a referida levedura é *Saccharomyces cerevisiae*.

(0112) Células eurarióticas maiores isoladas podem também ser transformadas, em vez de organismos completos. Com as células eurarióticas maiores, queremos dizer aqui qualquer célula eucariótica não-humana, exceto as células de levedura. As

células eucarióticas maiores preferida são células de planta.

a. O termo “transformado” refere-se ao fato de que o hospedeiro estava sujeito à engenharia genética para abranger uma, duas ou mais cópias de cada um dos ácidos nucléicos exigidos em qualquer uma das configurações acima descritas. Preferivelmente, o termo “transformado” relaciona-se aos hospedeiros que expressam heterologicamente os polipeptídios codificados pelo ácido nucléico com o qual eles são transformados, bem como supere-expressem os referidos polipeptídios. Conformemente, em uma configuração, a presente invenção fornece um organismo transformado, no qual os polipeptídios são expressados em quantidade maior do que no mesmo organismo não transformado.

b. Há diversos métodos conhecidos na técnica para a criação de organismos ou células hospedeiros transgênicos, como plantas, fungos, procariontes ou culturas de células de organismos eucarióticos maiores. Clonagem adequada e vetores de expressão para uso com hospedeiros celulares bacterianos, fúngicos, de levedura, de planta e mamíferos são descritos, por exemplo, em Pouwels *et al.*, Vetores de Clonagem: Um Manual Laboratorial, 1985, Elsevier, New York e Sambrook *et al.*, Clonagem Molecular: Um Manual Laboratorial, 2ª edição, 1989, Cold Spring Harbor Laboratory Press. Vetores de clonagem e de expressão para plantas e/ou células de plantas maiores, em particular, são disponíveis à pessoa qualificada. Vide, por exemplo, Schardl *et al.* Gene 61: 1-11, 1987.

c. Métodos de transformação de organismos ou células hospedeiros para hospedar ácidos nucléicos transgênicos são familiares ao perito. Para a criação de plantas transgênicas, por exemplo, métodos atuais incluem: eletroporação de protoplastos de planta, transformação mediada por lipossomo, transformação mediada por agrobactéria, transformação mediada por polietileno glicol,

bombardeamento de partícula, microinjeção de células de planta e transformação usando vírus.

d. Em uma configuração, DNA transformado é integrado em um cromossomo de um organismo e/ou célula hospedeira não-humana, tal que resulte em um sistema recombinante estável. Qualquer método de integração cromossômica conhecido na técnica pode ser usado na prática da invenção, incluindo, entre outros, troca de cassete mediada por recombinase (RMCE), inserção cromossômica viral de local específico, adenovírus e injeção pronuclear.

(0113) A fim de realizar o método de produção de esclareol in vitro, conforme exposto acima, é muito vantajoso fornecer um método de criação de, no mínimo, um polipeptídeo tendo uma atividade de síntese de diterpeno, conforme descrito em qualquer uma das configurações da invenção. Portanto, a invenção fornece um método para produção de, no mínimo, um polipeptídeo de fusão, capaz de catalisar a transformação de GGPP em esclareol compreendendo

(0114) o cultivo de organismo ou célula hospedeiros, não-humanos, transformados com o vetor de expressão da invenção, de maneira que hospede um ácido nucléico, de acordo com a invenção e expresse ou super-expresse um polipeptídeo, codificado pelo referido ácido nucléico e capaz de catalisar a transformação do GGPP em esclareol;

(0115) isolamento do polipeptídeo capaz de catalisar a transformação de GGPP em esclareol a partir do organismo ou célula hospedeiros, não-humanos, cultivados na etapa a).

(0116) De acordo com uma configuração preferida, o referido método ainda compreende, antes da etapa a), a transformação de um organismo ou célula hospedeiros, não-humanos com, no mínimo, um vetor de expressão da invenção, de

maneira que hospede, no mínimo, um ácido nucléico, de acordo com a invenção e expresse ou super-expresse o polipeptídio codificado pelo referido ácido nucléico.

(0117)A transformação e o cultivo do organismo ou célula hospedeiros, não-humanos podem ser realizados conforme descrito acima para o método de produção de esclareol in vivo. A etapa b) pode ser realizada usando qualquer técnica bem conhecida na técnica para isolar um polipeptídio particular de um organismo ou uma célula.

(0118)Uma “variante de polipeptídio”, conforme referido ao presente, significa um polipeptídio de fusão, capaz de catalisar a transformação de GGPP em esclareol e sendo substancialmente homólogo ao polipeptídio, de acordo com qualquer uma das configurações acima, mas tendo uma sequência de aminoácido diferente daquela codificada por qualquer uma das sequências de ácido nucléico da invenção devido a uma ou mais deleções, inserções ou substituições.

(0119)Variantes podem abranger sequências conservativamente substituídas, significando que um dado resíduo de aminoácido é substituído por um resíduo tendo características físico-químicas semelhantes. Exemplos de substituições conservativas incluem substituição de um resíduo alifático por outro, como Ile, Val, Leu ou Ala por um outro, ou substituições de um resíduo polar por outro, como entre Lys e Arg; Glu e Asp; ou Gln e Asn. Vide Zubay, Bioquímica, Addison-Wesley Pub. Co., (1983). Os efeitos dessas substituições podem ser calculados usando matrizes de escore de substituição como uma PAM-120, PAM-200 e PAM-250, conforme discutido em Altschul, (*J. Mol. Biol.* 219:555-65, 1991). Outras substituições conservadoras, por exemplo, substituições de regiões inteiras tendo características de hidrofobicidade semelhantes, são bem conhecidas.

(0120)Variantes de peptídeo que ocorrem naturalmente também são abrangidos pela invenção. Exemplos dessas variantes são proteínas que resultam de eventos

de combinação de mRNA alternativos ou da divisão proteolítica dos polipeptídios aqui descritos. Variações atribuíveis à proteólise incluem, por exemplo, diferenças nos terminais N ou C mediante a expressão em tipos diferentes de células hospedeiras, devido à remoção proteolítica de um ou mais aminoácidos terminais dos polipeptídios codificados pelas sequências da invenção. Variantes dos polipeptídios da invenção podem ser usadas para obter, por exemplo, atividade enzimática elevada ou reduzida desejada, regioquímica ou estereoquímica modificada ou utilização de substrato ou distribuição de produto alteradas, uma afinidade elevada para o substrato, uma especificidade melhorada para a produção de um ou mais compostos desejados, uma atividade enzimática diferente, um aumento da velocidade da reação enzimática, uma maior atividade ou estabilidade em um meio-ambiente específico (pH, temperatura, solvente, etc.) ou um nível de expressão melhorado em um sistema de expressão desejado. Um mutante direcionado por variante ou local pode ser criado por qualquer método conhecido na técnica. Variantes e derivados de polipeptídios nativos podem ser obtidos por isolamento das variantes que ocorrem naturalmente ou da sequência de nucleotídeo de variantes, de outras linhagens ou espécies de planta ou as mesmas, ou artificialmente programando mutações de sequências de nucleotídeo que codificam os polipeptídios de fusão da invenção. Alterações da sequência de aminoácido nativo podem ser realizadas por qualquer um de diversos métodos convencionais.

(0121)Variantes de polipeptídios, resultantes de uma fusão de sequências de peptídeo adicionais nas extremidades de terminal amino e carboxila dos polipeptídios da invenção, podem ser usadas para aumentar a expressão dos polipeptídios, ser úteis na purificação da proteína ou melhorar a atividade enzimática do polipeptídio em um ambiente ou sistema de expressão desejado. Essas

sequências adicionais de peptídeo podem ser peptídeos de sinal, por exemplo. Conformemente, a presente invenção abrange variantes dos polipeptídios da invenção, como aqueles obtidos por fusão com outros oligo ou polipeptídios e/ou aqueles que são ligados a peptídeos de sinal.

(0122) Portanto, em uma configuração, a presente invenção fornece um método para preparação de um polipeptídio de fusão variante, capaz de catalisar a transformação de GGPP em esclareol e abrangendo as etapas de:

- (a) seleção de um ácido nucléico de acordo com qualquer uma das configurações expostas acima;
- (b) modificação do ácido nucléico selecionado para obter, no mínimo, um ácido nucléico mutante;
- (c) transformação de células ou organismos unicelulares hospedeiros com a sequência de ácido nucléico mutante para expressar um polipeptídio codificado pela sequência de ácido nucléico mutante;
- (d) avaliação do polipeptídio com relação a, no mínimo, uma propriedade modificada; e,
- (e) opcionalmente, se o polipeptídio não tiver atividade de síntese de esclareol variante desejada, repetir as etapas (a) a (d) do processo até um polipeptídio com uma atividade de síntese de esclareol variante desejada ser obtido;
- (f) opcionalmente, se um polipeptídio tendo uma atividade de síntese de esclareol variante desejada tiver sido identificado na etapa d), isolamento do ácido nucléico mutante correspondente obtido na etapa (c).

(0123) Na etapa (b), uma grande quantidade de sequências de ácido nucléico mutante pode ser criada, por exemplo, por mutagênese randômica, mutagênese específica de local ou reorganização de DNA. Os procedimentos

detalhados de reorganização genéticos são encontrados em Stemmer, Reorganização de DNA por fragmentação randômica e remontagem: recombinação *in vitro* para evolução molecular. *Proc Natl Acad Sci USA.*, 1994, 91 (22): 10747-1075. Resumindo, Reorganização de DNA refere-se a um processo de recombinação randômica de sequências conhecidas *in vitro*, envolvendo, no mínimo, dois ácidos nucléicos selecionados para recombinação. Por exemplo, mutações podem ser introduzidas em locais particulares por síntese de oligonucleotídeos contendo uma sequência mutante, flanqueado por locais de restrição que permitem a ligação com fragmentos da sequência nativa. Após a ligação, a sequência reconstruída resultante codifica um análogo que possui a inserção de aminoácido desejada, substituição ou deleção. Alternativamente, procedimentos de mutagênese específicos de local, direcionados por oligonucleotídeo podem ser empregados para fornecer um gene alterado onde códons pré-determinados podem ser alterados por substituição, deleção ou inserção.

(0124) Conformemente, o polipeptídeo de fusão que compreende a SEQ ID NO:4 ou 5 e a SEQ ID NO:6 pode ser re combinado com qualquer outra diterpeno-sintase que codifica ácidos nucléicos, por exemplo, isolados de um organismo diferente da *Salvia sclarea*. Dessa forma, ácidos nucléicos mutantes podem ser obtidos e separados, os quais podem ser usados para transformar uma célula hospedeira, de acordo com procedimentos padrão, por exemplo, conforme divulgado nos presentes Exemplos.

(0125) Na etapa (d), o polipeptídeo obtido na etapa (c) é avaliado com relação a, no mínimo, uma propriedade modificada, por exemplo, uma atividade enzimática modificada desejada. Exemplos de atividades enzimáticas desejadas, para as quais um polipeptídeo expressado pode ser avaliado, incluem atividade

enzimática elevada ou reduzida, conforme medido pelo valor K_M ou V_{max} , regioquímica ou estereoquímica modificada e utilização de substrato alterado ou distribuição de produto. A avaliação da atividade enzimática pode ser realizada de acordo com os procedimentos familiares ao perito e aqueles divulgados nos presentes Exemplos.

(0126) A etapa (e) prepara-se para repetição das etapas (a)-(d) do processo, que pode preferivelmente ser realizada em paralelo. Conformemente, através da criação de uma quantidade significativa de ácidos nucléicos mutantes, muitas células hospedeiras podem ser transformadas com diferentes ácidos nucléicos mutantes ao mesmo tempo, permitindo a avaliação subsequente de uma quantidade elevada de polipeptídios. As chances de obtenção de um polipeptídio variante desejado podem, dessa forma, ser elevadas a critério do perito.

(0127) Todas as publicações mencionadas neste pedido são incorporadas por referência à divulgação e descrevem os métodos e/ou materiais em conexão com o que as publicações são citadas.

Descrição dos desenhos

(0128) Figura 1: Estruturas dos diversos compostos citados na descrição.

(0129) Figura 2: Mecanismo da biossíntese de esclareol a partir de GGPP. As etapas 1 e 2 enzimáticas podem ser catalisadas por duas proteínas distintas (uma LPP-sintase e uma esclareol-sintase) ou por uma enzima bi-funcional única (polipeptídio de fusão).

(0130) Figura 3: Alinhamento deduzido de SsLPPs3 (SEQ ID NO:4) e SsLPPs9 (SEQ ID NO:5) das sequências de amino ácido SEQ ID NO:1 e 2, duas diterpeno-sintases estritamente relacionadas, que codificam cDNAs, isoladas para o propósito da presente invenção. Resíduos idênticos estão em letras brancas e resíduos diferentes

entre as duas sequências estão em letras pretas.

(0131)Figura 4: Análise de SDS-PAGE dos extratos de proteína solúvel crua de células de *E. coli*, que expressam as proteínas SsLPPs3 e SsLPPs9 (SEQ ID NO: 1 e 2). Linhas 1 e 8: padrões de peso molecular; linhas 2 e 7: proteínas de controle, obtidas de células transformadas com o plasmídeo sem inserção; linhas 3 e 4: proteínas das células transformadas com pETDuet-SsLPPs3; linhas 5 e 6: proteínas das células transformadas com pETDuet-SsLPPs9. O gel foi colorido para proteína total, usando Azul de Coomassie.

(0132)Figura 5: Análise de SDS-PAGE da afinidade de SsLPPs3 (SEQ ID NO:1) de diterpeno-sintase de salvia recombinante, purificada, expressada em *E. coli*. Linha M, padrão de peso molecular; linha 1, extrato de proteína solúvel crua de células de controle; linha 2, extrato de proteína solúvel crua de células transformadas com pET28-SsLPPs3; linha 3, frações de fluxo passante; linhas 4 a 7, frações de lavagem; linhas 8 a 10, frações de eluição com 250mM de L-histidina. O gel foi colorido para proteína total usando o Azul de Coomassie.

(0133)Figura 6: Análise de CG dos produtos obtidos após incubação da SsLPPs3 (SEQ ID NO:1) de afinidade purificada com GGPP. (A) extrato de solvente direto, (B) Extrato de solvente da mesma amostra após tratamento com fosfatase alcalina.

(0134)Figura 7: (A) Sequências de N terminal das diterpeno-sintases recombinantes de extensão total e de SsLPPs3 truncada (SEQ ID NO:1 e SEQ ID NO:36 a 39). (B) Análise de SDS-PAGE das versões de extensão total e truncada das diterpeno-sintases de SsLPPs3, expressadas em *E. coli*. Linha M, padrão de peso molecular; linha 1, extrato de proteína solúvel crua das células de controle; linha 2, extrato de proteína solúvel crua de células transformadas com pET28-SsLPPs3; linha 3, SsLPPs3 marcada com histidina purificada; linha 4 e 5: respectivamente 1 e 0,5µL

de extrato de proteína solúvel crua de células transformadas com pETDuet-SsLPPs3; linhas 5 a 9, 0,5µL de extrato de proteína solúvel crua de células transformadas com pETDuet contendo as quatro deleções sequenciais. O gel foi colorido para proteína total usando o Azul de Coomassie.

(0135)Figura 8: Alinhamento das sequências de aminoácidos dos fragmentos tipo diterpeno-sintase de classe II com a sequência da estemodeno-sintase de *Oriza sativa* (Nº de Acesso AAZ76733).

(0136)Figura 9: Alinhamento das sequências de aminoácidos das 1132 construções (SEQ ID NO:3 e 74) para expressão heteróloga em *E. coli*.

(0137)Figura 10: Análise de CG dos produtos obtidos após incubação das diferentes 1132 proteínas recombinantes com LPP. Extratos de proteína crua de *E. coli* expressando as proteínas SsTps1132 (1132-1-3; SEQ ID NO:3) e 1132-2-5 (SEQ ID NO:74) recombinantes foram incubados com LPP em um volume final de 1mL de 50mM de MOPSO, pH 7, suplementado com 15mM de MgCl₂.

(0138)Figura 11: Análise de CG-MS dos produtos gerados a partir de LPP pela proteína 1132-2-5 (SEQ ID NO:74) recombinante. (A) Cromatograma de íon total dos produtos obtidos da incubação de LPP com um extrato de proteína crua de *E. coli*, transformado com pET101-1132-2-5. (B) Espectro de massa do pico no tempo de retenção de 14,3. (C) Espectro de massa de um padrão de esclareol autêntico.

(0139)Figura 12: Análise de CG dos produtos obtidos após coincubação de proteínas recombinantes 1132 (SEQ ID NO:3 e 74) com a proteína recombinante SsLPPs3 (SEQ ID NO:1) na presença de GGPP.

(0140)Figura 13: Estrutura dos plasmídeos usados para a transformação de células de *E. coli* nos Exemplos 10 e 11.

(0141)Configurações específicas da invenção ou Exemplos

(0142) A invenção agora será descrita em detalhes adicionais através dos seguintes Exemplos.

Exemplo 1

(0143) Isolamento de LPP-sintase, que codifica DNAs de *Salvia clarea* por uma abordagem de PCR

A. Extração de matéria de planta e RNA.

(0144) Botões florais em desenvolvimento da *Salvia sclarea* (1,5 a 2cm de extensão, 1-2 dias de idade) foram coletados nos campos de Bassins (Suíça) e diretamente congelados em nitrogênio líquido. O RNA total foi extraído usando o Reagente de RNA de Planta Concert™ da Invitrogen (Carlsbad, CA) e o mRNA foi purificado por cromatografia de afinidade de oligodT-celulose, usando o Kit de isolamento de mRNA FasTrack® 2.0 (Invitrogen, Carlsbad, CA), de acordo com as instruções do fabricante. Uma biblioteca de cDNA foi construída usando o Kit de Amplificação de cDNA Marathon™ (Clontech, Mountain View, CA).

B. Reações em Cadeia da Polimerase para amplificação de cDNAs de diterpeno-sintases

(0145) Sequências de aminoácido de diterpeno-sintases de classe I e II de diferentes plantas foram alinhadas e motivos conservados foram selecionados. Sequências de oligonucleotídeos degenerativos foram deduzidas desses motivos de aminoácido conservados. O Motivo DxDDTAM (x sendo qualquer aminoácido), encontrado na parte central das sequências de aminoácido das diterpeno-sintases e postulado para estar envolvido na interação com a metade de difosfato de GGPP nas diterpeno-sintases de classe II, foi usado para criar o primer inverso DT3F (5'-(SEQ ID NO:7)). Outro motivo, DVW(I/L)GK(T/S), encontrado em algumas diterpeno-sintases, foi usado para criar o primer reverso D/T4R (SEQ ID NO:8)).

(0146) PCR foram realizadas usando esses primers em todas as possíveis combinações de primers inverso e reverso. A mistura da PCR continha 0,4µM de cada primer, 300µM de cada dNTPs, 5µL de tampão de polimerase de DNA 10X HotStartTaq® (Qiagen), 2µL de cDNA dissolvido em 100 vezes, 0,5µL de polimerase de DNA HotStartTaq® em um volume final de 50µL. As condições cíclicas foram: 35 ciclos de 45 seg a 94°C, 45 seg a 50°C e 2 min a 72°C; e 10 min a 72°C. Os tamanhos dos produtos de PCR foram avaliados em um gel de agarose a 1%. As fitas correspondentes ao tamanho esperado foram removidas do gel, purificadas usando o Kit de Extração de Gel QIAquick® (Qiagen) e clonadas no vetor pCR®2.1-TOPO, usando o Kit de clonagem TOPO TA (Invitrogen, Carlsbad, CA). Fragmentos de cDNA inseridos estavam, então, sujeitos à sequenciamento de DNA e a sequência foi comparada contra o banco de dados de proteína não-redundante GenBank (NCBI), usando o algoritmo BLASTX (Altschul *et al*, *J. Mol. Biol.* 215, 403-410, 1990). A partir de diferentes PCR realizadas, apenas a combinação dos primers DT3F (SEQ ID NO:7) e DT4R (SEQ ID NO:8) forneceu um fragmento de DNA com o tamanho esperado e com a homologia de sequência a diterpeno-sintases. Todos os fragmentos emitidos dessa amplificação tiveram a mesma sequência exata. Essa sequência de 354 bp foi denominada FN23 (SEQ ID NO:9).

C. Isolamento de cDNA de extensão total por amplificação rápida por extremidades de cDNA (RACE).

(0147) Oligonucleotídeos específicos para a sequência FN23 (SEQ ID NO:9) foram criados. FN23-F1' (SEQ ID NO:10)), FN23-F2 (SEQ ID NO:11)) e FN23-F3 (SEQ ID NO:12)). Esses primers foram usados em RT-PCR em combinação com primers oligodT, estendidos com uma sequência adaptadora (SEQ ID NO:13). A composição da mistura de reação de RT-PCR foi a seguinte: 10µl de tampão de RT-PCR 5X

Qiagen OneStep, 400µM cada de dNTP, 400nM de cada primer, 2µl de Mistura Enzimática de RT-PCR Qiagen OneStep, 1µl de Inibidor de Ribonuclease RNasin® (Promega Co., Madison, WI) e 1250ng de RNA total em um volume final de 50µl. As condições do termociclador foram: 30 min a 50°C (transcrição reversa); 15 min a 95°C (ativação de polimerase de DNA); 35 ciclos de 45 seg a 94°C, 45 seg a 50°C e 90 seg a 72°C; e 10 min a 72°C. Um segundo ciclo de PCR foi realizado usando os produtos de RT-PCR como template com o primer adapterP (SEQ ID NO:14) em combinação com o mesmo ou primers específicos de FN23 abrigados. Essa abordagem de PCR forneceu um fragmento de cDNA 1271 bp (FN30 (SEQ ID NO:15)) tendo uma sobreposição perfeita de 192 bp com o fragmento FN23 (SEQ ID NO:9) e contendo a extremidade 3', incluindo o códon de interrupção e a sequência sem codificação 3' do cDNA correspondente.

(0148) Para amplificação da extremidade 5' do cDNA, oligonucleotídeos anti-sense, específicos para FN23 (SEQ ID NO:9) foram criados: FN23-R1 (SEQ ID NO:16)), FN23-R2 (SEQ ID NO:17)), FN23-R3 (SEQ ID NO:18)). Esses primers foram usados para RACE de 5', usando a biblioteca de cDNA de *S. sclarea*, após o protocolo de Kit de Amplificação de cDNA Marathon™ (Clontech, Mountain View, CA). As condições cíclicas térmicas foram as seguintes: 1 min a 94°C, 5 ciclos de 30 seg a 94°C e 4 min a 72°C, 5 ciclos de 30 seg a 94°C e 4 min a 70°C, 20 ciclos de 30 seg a 94°C e 4 min a 68°C. Essa RACE de 5' forneceu um fragmento de cDNA 1449 bp (FN40 (SEQ ID NO:19) tendo uma sobreposição perfeita de 227 bp com FN23 (SEQ ID NO:9). Comparação com sequências de diterpeno-sintases conhecidas revelou que o fragmento FN40 (SEQ ID NO:19) continha o códon de iniciação de tradução e uma região sem codificação de 87 bp. A montagem dos três fragmentos de cDNA (FN23, FN30 e FN40 (SEQ ID NO:9, 15 e 19) forneceu uma sequência de cDNA de

extensão total (SaTps1) de 2655 bp com uma estrutura de leitura aberta (SEQ ID NO:20) de codificação 2355 bp para uma proteína de resíduos 785 (SEQ ID NO:21), tendo forte homologia com diterpeno-sintases e, isto é, com copalil difosfato-sintases. O motivo DxDD, envolvido na ciclização iniciada por protonação, estava presente na sequência de aminoácido (posição 372) e o motivo DDxD, envolvido na ciclização iniciada por ionização, não foi encontrado. Dessa forma, essa sequência de proteína tem as características típicas de uma diterpeno-sintase de classe II, que catalisa exclusivamente as ciclizações dependentes de protonação de GGPP. A expressão heteróloga e a caracterização enzimática dessa proteína estão detalhadas nos Exemplos 2-4 a seguir.

Exemplo 2

(0149) Expressão heteróloga da LLP-sintase de *S. sclarea* na *E. coli*

(0150) O pETDuet-1 (Novagen, Madison, WI), criado para expressão sob o controle de um promotor T7, foi usado para expressão em células de *E. coli*. Para construir o plasmídeo de expressão, a estrutura de leitura aberta de SaTps 1 (SEQ ID NO:20) foi amplificada por PCR da biblioteca de cDNA com os primers inverso e reverso SaTps-Nde (SEQ ID NO:22) e SaTps-Kpn (SEQ ID NO:23)), criados para introduzir um local de NdeI imediatamente antes do códon de início e um local de KpnI após o códon de interrupção. Já que a estrutura de leitura aberta com um local NdeI na posição de 1614 da estrutura de leitura aberta, essa amplificação foi realizada em duas etapas por PCR de extensão de sobreposição (Horton *et al*, Gene 78, 68, 1989), usando os primers SaTps-Nde (SEQ ID NO:22) e SaTps-Kpn (SEQ ID NO:23) em combinação com os primers Satps-mut1f (SEQ ID NO:24)) e Satps-mut1r (SEQ ID NO:25)), criados para remover o local NdeI sem alterar a sequência de aminoácido. O cDNA resultante foi primeiramente ligado no plasmídeo PCR2.1-Topo

usando o Kit de Clonagem TOPO TA (Invitrogen, Carlsbad, CA) e as sequências das inserções foram verificadas antes da subclonagem como fragmento NdeI-KpnI no vetor pETDuet-1.

(0151) Análise da sequência de diversos clones obtidos por amplificação da biblioteca de cDNA com os primers específicos SaTps1 mostrou alguma variabilidade em diversas posições da sequência de cDNA. Sete posições foram identificadas, nas quais dois aminoácidos diferentes podem ser encontrados. Uma posição foi encontrada, onde inserção de um resíduo de serina ocorreu em alguns dos clones. Essas posições são listadas na tabela abaixo.

Posições (relativas à sequência de aminoácido)	Aminoácido
34	Ile ou Thr
40	Phe ou Leu
174	Gln ou His
222	Gly ou Asp
538	Gln ou His
560	Arg ou Leu
596	Asn ou Lys
612	Inserção de um Ser

(0152) Essas variações pareceram ocorrer em uma maneira randômica em onze clones diferentes seqüenciados, sugerindo que, no mínimo, duas isoformas relacionadas bem de perto de uma diterpeno-sintase estão presentes no genoma da *S. sclarea* e que a abordagem de amplificação de PCR levou à mistura das sequências. Dois clones, SsLPPs3 (SEQ ID NO:4) e SsLPPs9 (SEQ ID NO:5), representantes da variabilidade das sequências (Figura 3), foram selecionados para

os experimentos de expressão heteróloga e caracterização enzimática.

(0153) Os plasmídeos pETDuet-SsLPPs3 e pETDuet-SsLPPs9 foram transferidos para células de *E. Coli* BL21 (DE3) (Novagen, Madison, WI). Colônias simples de células transformadas foram usadas para inocular 5ml de meio LB. Após 5 a 6 horas de incubação a 37°C, as culturas foram transferidas para uma incubadora a 20°C e deixadas 1 hora para equilíbrio. A expressão da proteína foi, então, induzida pela adição de 1mM de IPTG e a cultura foi incubada durante a noite a 20°C. O dia seguinte, as células foram coletadas por centrifugação, suspensas novamente em volume a 0,1 de 50mM de MOPSO, pH 7, 10% de glicerol e lisados por sonicação. Os extratos foram limpos por centrifugação (30 min a 20.000g) e os sobrenadantes contendo as proteínas solúveis foram usados para experimentos adicionais.

(0154) Os extratos de proteína crua de células transformadas de pETDuet-SsLPPs3 e pETDuet-SsLPPs9 foram analisados por SDS-PAGE e comparados com extratos de proteína, obtidos das células transformadas com o plasmídeo pETDuet vazio. As proteínas SsLPPs3 e SsLPPs9 recombinantes (SEQ ID NO:1 e 2) foram claramente detectadas e o peso molecular aparente estimado a 90 Kda, um valor em conformidade com o peso molecular calculado de 83 KDa (Figura 4).

Exemplo 3

(0155) Purificação da LPP-sintase da *S. sclarea* e atividades enzimáticas

(0156) Para caracterizar adicionalmente as diterpeno-sintases recombinantes, asseguramos purificar as enzimas SsLPPs3 e SsLPPs9 (SEQ ID NO:1 e 2).

(0157) Os plasmídeos PCR2.1-Topo contendo o cDNA de SsLPPs3 e SsLPPs9 (SEQ ID NO:4 e 5) (Exemplo 2) foram digeridos com NdeI e SacI e as inserções foram ligadas no plasmídeo pET28a(+) (Novagen). Os plasmídeos de expressão resultantes (pET28-SsLPPs3 e pET28-SsLPPs9) contem os cDNAs com uma

modificação da extremidade 5' (SEQ ID NO:26 e 27), criados para expressar as proteínas com uma cauda de hexa-histidina de terminal N. A purificação foi realizada sob condições nativas usando o Sistema de Purificação ProBond™ (Invitrogen), seguindo o protocolo do fabricante, com exceção de que, para a eluição, imidazol foi substituído pela L-histidina para minimizar a inibição da enzima. Usando essa abordagem, as enzimas recombinantes SsLPPs3 e SsLPPs9 poderiam ser purificadas para homogeneidade aparente (Figura 5).

a. As enzimas purificadas de afinidade foram incubadas 12 horas a 30°C com 200µM de GGPP e 1mM de DTT em MOPSO, pH 7, 10% de glicerol. Nenhum produto de diterpeno foi observado ao extrair a incubação com pentano e análise do extrato por CG ou CG-MS. O tratamento do mesmo extrato por fosfatase alcalina (Sigma, 6 unidades/ml), seguido por extração com pentano e análise de CG, mostrou a formação de labdenediol (Figura 6) e demonstrou a formação enzimática de labdenediol-difosfato (LPP) como produto único do GGPP pela diterpeno-sintase recombinante.

(0158)A análise de CG foi realizada em um sistema de CG Agilent Série 6890 (Agilent Technologies), equipado com um detector de ionização de chama, usando um diâmetro interno de 0,25mm por coluna capilar SPB-1 de 30mm (Supelco, Bellefonte, PA). O gás carregador foi He em um fluxo constante de 1mL/minuto. A temperatura inicial da estufa foi de 100°C (1 minuto de suspensão), seguida por um gradiente de 10°C/minuto a 300°C.

(0159)A análise de CG-MS foi realizada em um sistema de CG Série 6890 (Agilent), acoplado a um detector de massa 5975 (Agilent Technologies). A coluna foi equipada com um diâmetro interno de 0,25mm por coluna DB-1MS de 30m de extensão (Agilent Technologies). O gás carregador foi He em um fluxo constante de

1mL/minuto. A temperatura inicial da estufa foi de 80°C, seguida por um gradiente de 10°C/minuto a 280°C. Os espectros foram registrados a 70eV com uma voltagem de multiplicador de elétron de 2200V. A identificação do produto foi confirmada pela concordância dos tempos de retenção e a equivalência do espectro de massa com o espectro de padrões autênticos.

Exemplo 4

(0160) Deleções do terminal N da LPP-sintase de *Salvia sclarea*

a. Nas plantas, as diterpeno-sintases estão localizadas nos plastídeos. Essa compartimentalização é controlada por um mecanismo de tratamento que reconhece um sinal de peptídeo de trânsito do terminal N. Dessa forma, as diterpeno-sintases são geralmente expressadas como pré-proteínas e são processadas nos plastídeos pela divisão do sinal de peptídeo resultando em uma proteína madura. Análise da sequência do terminal N de SsLPPs3 e SsLPPs9 (SEQ ID NO:1 e 2), usando o método de ChloroP (Emanuelsson *et al*, *Protein Science* 8, 978-984, 1999; <http://www.cbs.dtu.dk/services/ChloroP/>) não revelou nenhuma evidência clara da presença de um peptídeo de trânsito. Experimentos foram, dessa forma, realizados para avaliar o efeito das deleções no terminal N sobre a atividade enzimática. Quatro cDNAs truncados foram criados para SsLPPs3 resultando em deleção de 17, 37, 53 e 63 aminoácidos, respectivamente, (SEQ ID NO:28 a 31). Cada construção foi criada por PCR usando quatro primers inversos diferentes cada, criados para enrijecer na posição da forma truncada desejada e introduzindo um local de restrição de NdeI, seguido por um códon de iniciação de tradução de ATG (SsLPPs3_del1, (SEQ ID NO:32); SsLPPs3_del2, (SEQ ID NO:33); SsLPPs3_del3, (SEQ ID NO:34), SsLPPs3_del4, (SEQ ID NO:35)). Esses primers foram usados em combinação com o primer SaTps-Kpn (SEQ ID NO:23) (Exemplo 2) e os quatro cDNAs obtidos (SEQ

ID NO:28 a 31) foram ligados no plasmídeo pETDuet-1. Expressão heteróloga das proteínas (SEQ ID NO:36 39) foi realizada em *E. coli*, conforme descrito no Exemplo 2. A Figura 7 mostra uma análise de SDS-PAGE comparando o nível de produção das proteínas heterólogas obtidas com as diferentes construções de extensão total e truncadas. Um nível de expressão melhorada foi claramente observado especialmente para as duas deleções maiores. Esses resultados são típicos para terpeno-sintases localizadas em plastídeo e refletem uma solubilidade e/ou estabilidade melhorada da proteína madura em comparação com a pré-proteína.

Exemplo 5

(0161) Sequenciamento massivamente paralelo de uma biblioteca de cDNA da flor *S. sclarea*

(0162) Usamos a tecnologia de seqüenciamento paralelo massivo de pequenos fragmentos de DNA, desenvolvida pela Illumina (San Diego, Califórnia) para obter informações de sequência de todas as transcrições (transcriptoma) presentes nas flores *Salvia sclarea*. Essa técnica de sequenciamento usa uma química de seqüenciamento com base no terminador reversível e os aparatos de Estação de Aglomerado e Seqüenciador de Genoma, desenvolvidos pela Solexa e Illumina (www.illumina.com).

(0163) A tecnologia e os equipamentos foram preparados na Fasteris SA (Genebra, Suíça) e a preparação das amostras de DNA e o sequenciamento foram realizados pela Fasteris SA. Uma alíquota (1µg) da biblioteca de cDNA, gerada das flores *S. sclarea* em desenvolvimento e usando o Kit de Amplificação de cDNA Marathon™ (Clontech, Mountain View, CA) (Exemplo 1), foi tratada usando o Kit de Preparação de Amostra Genômica (Illumina). Brevemente, o DNA é fragmentado por nebulização, as extremidades são reparadas para gerar extremidades firmes,

adaptadores são ligados às extremidades dos fragmentos de DNA e os fragmentos de DNA modificados por adaptador são amplificados por PCR. Após controlar a qualidade da biblioteca por eletroforese por gel, a geração dos aglomerados de DNA sobre a célula de fluxo e a reação de sequenciamento é realizada nos equipamentos de Estação de Aglomerado e Sequenciador de Genoma. Usando essa tecnologia, 1,9 milhão de curtas sequências (leituras) de 35 bases foram obtidas.

(0164) O *software* Edena (Hernandez *et al*, *Genome res.* 15(5), 802-809, 2008) foi usado para remontar sequências contíguas. As cinco últimas bases foram primeiramente removidas de cada leitura devido às possíveis incorporações erradas, devido à baixa fidelidade nos últimos ciclos do procedimento de sequenciamento. Os parâmetros do software foram estabelecidos para permitir a extensão mínima de 15 bases para as sobreposições com identificação estrita (100%). Os contigs (sequências contíguas) com uma extensão de, no mínimo, 50 bases foram retidos. Nessas condições, 2054 contigs de 50 a 1330 bases em extensão poderiam ser reconstituídos.

(0165) Para avaliar a qualidade da montagem, os contigs foram pesquisados com relação à identificação de sequência com a sequência de DNA de SsLPPs3, as diterpeno-sintases de classe II, primeiramente isoladas da biblioteca de cDNA de *S. sclarea* (SEQ ID NO:4, Exemplo 2). Essa pesquisa foi realizada usando o método de BLASTn (Altschul *et al*, *J. Mol. Biol.* 215, 403-410, 1990). De forma surpreendente, apenas 3 contigs de extensões de 81, 73 e 52 bases (contigs 1 a 3, SEQ ID NO:40 a 42) foram encontrados e apenas quarenta leituras foram usadas pelo *software* Eland para gerar esses contigs. O alinhamento com a sequência de referência SsLPPs3 mostrou que os 3 contigs cobriram apenas 8,7% da sequência de extensão total, embora com uma identificação de 99%.

(0166) Informações de sequência muito limitadas foram relatadas nos bancos de dados públicos para *Salvia sclarea*. A única sequência genética disponível do banco de dados NCBI (<http://www.ncbi.nlm.nih.gov>) foi a sequência da grande subunidade da ribulose-1,5-bifosfato carboxilase (RuBisCO) da *salvia sclarea* (Nº de acesso NCBI Z37450). A pesquisa dos contigs com relação à identificação de DNA com essa sequência de DNA de RuBisCO de *S. sclarea* (Pesquisa de BLASTn) forneceu dois contigs de 870 e 547 bases, respectivamente (contigs 4 e 5, SEQ ID NO:43 e 44). O alinhamento dos dois contigs com a sequência de RuBisCO mostrou cobertura de 98%: apenas 27 bases (entre a posição 858 e 884) das 1420 bases não estavam presentes nos contigs. Além dessa cobertura quase completa, a identificação entre a sequência de referência e os contigs foi de 99,5%, representando uma diferença de apenas 7 nucleotídeos.

(0167) Todas as leituras (dados não-montados) foram, então, pesquisadas com relação à identificação de sequência com a sequência de SsLPPs3 (SEQ ID NO:4). O *software* Eland (Illumina) foi usado para realizar essa pesquisa, permitindo um máximo de 2 divergências com a sequência de referência. Um total de 616 leituras foram recuperadas. O alinhamento dos fragmentos selecionados com a sequência de referência revelou que a sequência de SsLPPs3 foi coberta na extensão total com uma cobertura levemente maior (mais leituras) em direção da extremidade 3'. A mesma manipulação com a sequência de RuBisCO mostrou que 1650 leituras foram obtidas para essa sequência. A cobertura da sequência de referência com as leituras foi muito maior para a RuBisCO do que para a SsLPPs3. Para SsLPPs3 (SEQ ID NO:4), diversas regiões pequenas sem cobertura e regiões com ambiguidade de sequência entre leituras foram encontradas. Essa cobertura incompleta impede a remontagem completa e é certamente a razão para a geração

de apenas alguns contigs muito pequenos.

Exemplo 6

(0168) Extração de sequências semelhantes a diterpeno-sintases de classe I a partir dos dados de sequenciamento

(0169) O algoritmo de Blast (Altschul *et al*, *J. Mol. Biol.* 215, 403-410, 1990) foi usado para pesquisar a homologia das sequências de aminoácido deduzidas com sequências de diterpeno-sintases de classe I.

(0170) Uma pesquisa de Blastx contra um banco de dados de proteína foi primeiramente realizada com os 2054 contigs. Essa pesquisa forneceu apenas um contig (contig 1610, SEQ ID NO:45), apresentando homologia de sequência com diterpeno-sintases de classe I. A sequência de aminoácido deduzida desse contig continha o motivo DDxxD, característico da ciclização iniciada por ionização de prenil-difosfatos.

(0171) Uma fração dos dados brutos, representando aproximadamente 3×10^5 leituras foi, então, pesquisada com relação à homologia com diterpeno-sintases de classe I. As leituras foram pesquisadas usando o algoritmo de tBlastn com cinco sequências de aminoácido de diterpeno-sintase de classe I selecionados (números de acesso NCBI AAC39443, BAB19275, BAB12441, AAD34295, AAS98912). Essa pesquisa selecionou 462 leituras, que foram, então, processadas usando o programa CAP (Huang, *Genomics* 14(1), 18-25, 1992) para identificar sequências de sobreposição. Uma pequena porção das leituras poderia ser montada em curtos contigs de extensão máxima de 111 bases. Esses contigs, bem como as leituras isoladas remanescentes, foram usados para pesquisa de Blastx contra um banco de dados de proteína para confirmar sua identificação com diterpeno-sintases de classe I. Finalmente, 5 fragmentos de DNA foram retidos (SEQ ID NO:46 a 50).

(0172)As sequências de aminoácido (SEQ ID NO:51 a 55) foram deduzidas dos fragmentos selecionados e foram alinhadas com sequências de diterpeno-sintases de referência, permitindo seu posicionamento relativo. A figura 8 mostra um alinhamento dessas sequências com a sequência de diterpeno-sintase de extensão completa da estemodeno-sintase da *Oriza sativa* (Morrone *et al*, 2006; Nº de acesso NCBI AAZ76733) coletado como referência.

Exemplo 7

(0173)Amplificação por PCR de cDNAs de diterpeno-sintases de classe I de extensão completa

(0174)Um conjunto de oligonucleotídeos inverso e reverso foi deduzido das sequências de DNA relacionadas com diterpeno-sintases, selecionadas do sequenciamento da biblioteca de cDNA de *S. sclarea* (Exemplo 6). Esses primers foram usados em combinação com primers adaptadores de cDNA em amplificações por PCR do tipo 3'/5' RACE. As amplificações foram realizadas usando a biblioteca de cDNA de *S. sclarea*, preparada conforme descrito acima no Exemplo 1, seguindo o protocolo do Kit de Amplificação de cDNA Marathon™ (Clontech, Mountain View, CA). As condições cíclicas térmicas foram as seguintes: 1 min a 94°C, 5 ciclos de 30 seg a 94°C e 4 min a 72°C, 5 ciclos de 30 seg a 94°C e 4 min a 70°C, 20 ciclos de 30 seg a 94°C e 4 min a 68°C.

(0175)Usando o primer Cont250-Fwd (SEQ ID NO:56), uma sequência de DNA de 547 bp (1130Cont250, SEQ ID NO:66) foi obtida. Análise da sequência revelou que esta correspondeu à extremidade 5' de um cDNA de diterpeno-sintase e conteve 348 bp da região de codificação. Com o primer Cont147_fw1 (SEQ ID NO:57) e Cont147_fw2 (SEQ ID NO:58), obtivemos uma sequência de 1473 bp (1132Cont147, SEQ ID NO:67), contendo a extremidade 3' e 1293 bp da região de codificação de

um cDNA de diterpeno-sintase. Os primers Cont147_rev1 (SEQ ID NO:59) e Cont147_rev2 (SEQ ID NO:60) permitiram a amplificação de um fragmento de DNA de 464 bp (1134Cont147, SEQ ID NO:68). O aminoácido deduzido mostrou homologia com diterpeno-sintases, mas alinhamento com outras sequências de diterpeno-sintases sugeriu que 200 a 300 códons ainda estavam faltando para atingir a extremidade 5'. Todas as sequências obtidas por essa série de amplificação diferiram significativamente das sequências de SsLPPs previamente isoladas (SsLPPs3 e SsLPPs9, SEQ ID NO:4 e 5).

(0176) Uma abordagem de 5'RACE foi usada para identificar a extremidade 5' da ORF, correspondente à sequência 1132Cont147 (SEQ ID NO:67). Usando os primers 1132_race1 (SEQ ID NO:62) e 1132_race2 (SEQ ID NO:63), uma sequência de 536 bp (1132RACE, SEQ ID NO:69) foi obtida, a qual tinha sobreposição de 41 bases com o fragmento 1132Cont147 (SEQ ID NO:67). O N-term desse produto de RACE foi idêntico à sequência 1134Cont147 previamente obtida (SEQ ID NO:68) e, dessa forma, nenhuma extensão na extremidade 5' foi observada. Conforme observado previamente, essa sequência teve homologia com diterpeno-sintases, mas pareceu mais curta em, no mínimo, 200 códons do que todas as outras sequências de diterpeno-sintases publicadas. Experimentos de 5'RACE foram realizados, a fim de tentar estender a sequência em direção da extremidade 5' da sequência 1132Cont147 (SEQ ID NO:67) e identificar a verdadeira códon de iniciação de tradução. Diversos conjuntos de oligonucleotídeos foram criados, mas nenhuma informação de sequência adicional foi obtida. Isso nos levou a supor que um dos códons ATG na sequência 1134Cont147 (SEQ ID NO:68) foi realmente o códon de iniciação do gene de diterpeno-sintase correspondente. A sequência nucleotídica dessa diterpeno-sintase putativa (denominada SsTps1132, SEQ ID

NO:6) foi reconstituída das sequências 1132Cont147 (SEQ ID NO:67) e 1132RACE (SEQ ID NO:69). Pegando o primeiro ATG, a ORF de 1728 bp da SsTps1132 (SEQ ID NO:6) codificou para uma proteína de 575 aminoácidos (SEQ ID NO:3). Essa proteína conteve o motivo dependente de ionização (DDFFD) e homologia compartilhada, mas relativamente, baixo, com as diterpeno-sintases publicadas; a sequência mais próxima sendo uma terpeno-sintase de *Nicotiana tabacum* (No de acesso NCBI AAS98912), identificação de 37%. Essa proteína também compartilhou apenas 23% de identificação com SsLPPs isoladas de *S. Sclarea* nos Exemplos 1 a 4. SsTps1132 foi alinhada com sequências de diterpeno-sintase selecionadas. Esses alinhamentos mostraram que a SsTps1132 (SEQ ID NO: 3) é truncada na extremidade do terminal N por 150 a 240 aminoácidos, em comparação com outras diterpeno-sintases.

(0177)O método de CloroP (Emanuelsson *et al*, *Protein Science* 8, 978-984, 1999; <http://www.cbs.dtu.dk/services/ChloroP/>) foi usado para prever a presença de um peptídeo de trânsito de cloroplasto em SsTps1132 (SEQ ID NO:3). Um peptídeo de trânsito de cloroplasto de 51 aminoácidos foi previsto, sustentando uma localização de cloroplasto dessa proteína.

(0178)A pesquisa de todas as leituras com relação a sequências idênticas à sequência de DNA SsTps1132 (SEQ ID NO:6) forneceu 425 leituras. O nível de expressão de SsTps1132 (220 leituras/Kb) foi semelhante ao nível de expressão de SsLPPs (SEQ ID NO:4 e 5) (260 leituras/Kb). Com a hipótese de que etapas catalisadoras de enzimas na mesma via metabólica são geralmente expressadas em um nível semelhante, pode-se especular que a SsTps1132 está envolvida na mesma via metabólica que as SsLPPs.

(0179)Os contigs gerados com o *software* Edena (Exemplo 5) foram pesquisados

para sequências de DNA, idênticas às sequências das novas diterpeno-sintases de classe I putativas SsTps1132 (SEQ ID NO:3) Quatro contigs foram encontrados. O contig previamente identificado 1610 (SEQ ID NO:45) e três contigs adicionais (de extensão de 53 a 96 bp, SEQ ID NO:70 a 72), não previamente identificados como fragmento de uma diterpeno-sintase. A pesquisa de Blastx com essas três sequências não mostrou homologia com sequências de proteína conhecidas. A falha em descobrir homologia para esses contigs é devido às extensões curtas desses fragmentos e à baixa homologia de SsTps1132 com as diterpeno-sintases presentes nos bancos de dados.

Exemplo 8

(0180) Expressão heteróloga das diterpeno-sintases de classe I de *S. sclarea* na *E. coli*

(0181) Para determinar uma atividade enzimática para SsTps1132 (SEQ ID NO:3), as proteínas recombinantes foram expressadas na *E. coli*. Os cDNAs de extensão total foram inseridos no vetor pet101/D-TOPO, usando o Kit de Expressão TOPO Direcional Champion pET101.

(0182) Duas construções foram preparadas: uma para expressar a proteína de extensão total e uma para expressar uma proteína truncada, com base na predição do peptídeo de trânsito de cloroplasto. A estrutura de leitura aberta de SsTps1132 de extensão total (SEQ ID NO:6) foi amplificada da biblioteca de cDNA, usando o par de primer 1132_início_1 (SEQ ID NO:64) e 1132-interrupção (SEQ ID NO:61). Os primers 1132_início2 (SEQ ID NO:65) e 1132-interrupção (SEQ ID NO:61) foram usados para preparar a construção para a expressão de SsTps1132 (SEQ ID NO:3) com uma deleção no terminal N de 50 aminoácidos. Todas as amplificações de cDNA para expressão das construções de expressão foram realizadas usando a

polimerase de DNA *Pfu* (Promega), em um volume final de 50µl, contendo 5µg de tampão 10X de polimerase de DNA *Pfu*, 200µM cada de dNTP, 0,4µM cada de primer inverso e reverso, 2,9 unidades de polimerase de DNA *Pfu* e 5µl de cDNA 100 vezes diluído (preparado conforme descrito no presente no Exemplo 1 usando o Kit de Amplificação de cDNA Marathon™ (Clontech)). As condições cíclicas térmicas foram as seguintes: 1,5 min a 95°C; 30 ciclos de 45 seg a 95°C, 30 seg a 58°C e 5 min a 72°C; e 10 min a 72°C.

(0183) Após a ligação no vetor pET101, diversos clones foram selecionados para cada construção e foram sequenciados para assegurar que nenhuma mutação tenha sido introduzida durante a amplificação de PCR. Duas construções foram selecionadas: SsTps1132 (SEQ ID NO:6) e 1132-2-5 (SEQ ID NO:73). O alinhamento das duas sequências de aminoácido deduzidas dessas construções é mostrado na Figura 9.

(0184) Os plasmídeos pET101-SsTps1132 e pET101-1132-2-5 foram transferidos para células de *E. Coli* BL21 (DE3) (Novagene, Madison, WI). Colônias simples de células transformadas foram usadas para inocular 5ml de meio LB. Após 5 a 6 horas de incubação a 37°C, as culturas foram transferidas para uma incubadora a 20°C e deixadas 1 hora para equilíbrio. A expressão da proteína foi, então, induzida pela adição de 1mM de IPTG e a cultura foi incubada durante a noite a 20°C. O dia seguinte, as células foram coletadas por centrifugação, suspensas novamente em volume a 0,1 de 50mM de MOPSO, pH 7, 10% de glicerol e lisados por sonicação. Os extratos foram limpos por centrifugação (30 min a 20.000g) e os sobrenadantes contendo as proteínas solúveis foram usados para experimentos adicionais. Os extratos de proteína crua foram analisados por SDS-PAGE e comparados com extratos de proteína, obtidos das células transformadas com o plasmídeo pET101

vazio. Pareceu que a deleção do sinal de peptídeo melhorou a expressão heteróloga na *E. coli*.

Exemplo 9

(0185) Atividade enzimática das diterpeno-sintases de classe I recombinantes de *S. Sclarea* na *E. coli*

(0186) Os extratos de proteína crua de *E. coli*, contendo as proteínas recombinantes e preparados conforme descrito no Exemplo 8 foram usados para a caracterização das atividades enzimáticas. Os ensaios enzimáticos foram realizados conforme descrito no Exemplo 3. Todos os ensaios foram realizados em 50mM de MOPSO, pH 7, 10% de glicerol, 1mM de DTT.

(0187) As atividades enzimáticas foram primeiramente avaliadas usando um substrato de GGPP ou LPP, o produto de SsLPPs e o intermediário presumido na biossíntese de esclareol (Exemplos 1 a 4). GGPP foi sintetizado conforme descrito por Keller e Thompson (*J. Chromatogr* 645(1), 1993, 161-167) e LPP foi preparado enzimaticamente conforme descrito no Exemplo 3. Os ensaios foram realizados na presença de 10 a 100µM de substrato, 15mM de MgCl₂ e 0,1 a 0,5mg de proteína crua em um volume total de 1mL. Os tubos foram incubados 4 a 12 horas a 30°C e extraídos duas vezes com um volume de pentano. Após concentração sob um fluxo de nitrogênio, os extratos foram analisados por CG e CG/MS (usando as condições descritas no Exemplo 3) e comparados com extratos do ensaio com proteínas de controle (obtidas de células transformadas com o plasmídeo vazio). Com GGPP como substrato, nenhuma atividade foi observada com qualquer uma das proteínas recombinantes (dados não mostrados). Com o LPP como substrato, atividade foi observada com SsTps1132 (SEQ ID NO:3) e 1132-2-5 (SEQ ID NO:74) (Figura 10). As enzimas foram também ativas na ausência de MgCl₂ e os mesmos perfis de

produto foram observados com quase a mesma atividade geral. A identificação de produto foi confirmada pela concordância dos tempos de retenção (Figura 10) e comparação do espectro de massa com o espectro de um padrão autêntico (Figura 11). Em todos os ensaios, um pico único de esclareol foi observado sem traço de produto adicional.

(0188) Ensaios foram, então, realizados com coincubação das diterpeno-sintases de classe II (SsLPPs3, SEQ ID NO:1; Exemplos 1-4) e as diterpeno-sintases de classe I (série 1132, SEQ ID NO:3 e 74). Ensaios foram realizados em 50mM de MOPSO, pH 7, 10% de glicerol, 1mM de DTT, 50µM de GGPP, com 1mM de MgCl₂ e na presença de 50µL dos extratos de proteína crua da *E. coli*, expressando as diferentes construções. Dessa forma, ensaios na presença de 50µL de extratos de proteína crua contendo a enzima recombinante SsLPPs3 (SEQ ID NO:1) e 50µL de extratos contendo SsTps1132 (SEQ ID NO:3) ou 1132-2-5 (SEQ ID NO:74) foram avaliados com relação à produção de produtos de diterpeno. A Figura 12 mostra os perfis de CG de extratos dessas incubações na presença de MgCl₂. O esclareol foi produzido com ambas as construções de 1132 (SEQ ID NO:3 e 74), um resultado resistente com o ensaio descrito acima com LPP como substrato. Nenhuma diferença significativa foi observada ao omitir MgCl₂ das incubações (dados não mostrados).

(0189) Concluindo, a SsTps1132 (SEQ ID NO:6) codifica a esclareol-sintase e catalisa a conversão de LPP para esclareol.

(0190) Dessa forma, mostramos que, na *Salvia sclarea*, o esclareol é sintetizado do GGPP em duas etapas por duas enzimas distintas, uma LPP-sintase e uma esclareol-sintase. Isolamos cDNAs que codificam cada uma das duas diterpeno-sintases. A LPP-sintase, codificada pelos cDNAs de SsLPPs3 e SsLPPs9, catalisa a

conversão de GGPP em LPP e contém os componentes características de diterpeno-sintases de classe II. A esclareol-sintase, codificada pelo cDNA de SsTps1132, catalisa a conversão de LPP em esclareol e é relacionada às diterpeno-sintases de classe I embora com algumas particularidades, isto é, uma deleção grande na extremidade do terminal N.

Exemplo 10

(0191) Produção *in vivo* de esclareol em *E. coli* por coexpressão das duas diterpeno-sintases.

(0192) Para avaliar a produção *in vivo* de esclareol nas células *E. coli*, plasmídeos e células transformadas são preparados para a coexpressão das duas diterpeno-sintases (a LPP-sintase e a esclareol-sintase). Além das duas diterpeno-sintases, uma FPP-sintase e uma GGPP-sintase também são coexpressadas para assegurar um agrupamento suficiente de GGPP nas células. Para aumentar adicionalmente o fluxo de carbono na via e aumentar o nível de GGPP e subsequentemente o nível de esclareol produzido, os genes que codificam uma via do mevalonato parcial também são expressados nas mesmas células. Esses últimos genes codificam uma mevalonato quinase (mvaK1), um fosfomevalonato quinase (mvaK2), a mevalonato difosfato decarboxilase (MvaD) e uma isopentenil difosfato isomerase (idi) e convertem mevalonato para isopentenil difosfato (IPP) e dimetilalil difosfato (DMAPP), os dois substratos da FPP-sintase.

(0193) O gene de FPP-sintase de levedura (Número de acesso J05091) é amplificado do DNA genômico de *S. cerevisiae* usando os primers FPPy_NcoI (SEQ ID NO:75) e FPPY-Eco (SEQ ID NO:6). O DNA genômica é isolado da *S. cerevisiae* usando o Kit Qiagen RNA/DNA Maxi (Qiagen AG, Basiléia, Suíça). A PCR é realizada com a polimerase de DNA *Pfu* (Promega AG, Dubendorf, Suíça) em um

volume final de 50µl contendo 0,4µl de cada primer, 200µM de dNTPs, 0,5µl de polimerase de DNA, 5µl de DNA genômico de *S. cerevisiae*. A condição de ciclo de PCR são as seguintes: 90 seg a 95°C; 28 ciclos de 45 seg a 95°C, 30 seg a 54°C e 4 min a 72°C; e 10 min a 72°C. O DNA amplificado é ligado como fragmento de NdeI-EcoRI no primeiro local de clonagem múltipla (MCS1) do plasmídeo pACYCDuet-1 fornecendo o plasmídeo pACYCDuet-FPPs, que hospeda o gene de FPPs sob o controle de um promotor T7.

(0194) Um operador contendo os genes que codifica mvaK1, mvaK2, MvaD e idi é amplificado do DNA genômico de *Streptococcus pneumoniae* (ATCC BAA-334, Padrões de LGC, Molsheim, França) com os primers MVA-up1-início (SEQ ID NO:77) e MVA-up2-interrupção (SEQ ID NO:78). A PCR é realizada usando a polimerase de DNA de Fusão HS PfuUltra™ II (Stratagene), a composição da mistura de PCR estando de acordo com as instruções do fabricante. A condição de ciclo térmico é 2 min a 95°C; 30 ciclos de 20 seg a 95°C, 20 seg a 58°C e 90 seg a 72°C; e 3 min a 72°C. O fragmento 3,8 Kb é purificado em um gel de agarose e ligado usando o Kit de Clonagem de PCR In-Fusion™ Dry-Down (clontech) no segundo MCS do plasmídeo pACYDuet-FPPs digerido com NdeI e XhoI fornecendo o plasmídeo pACYCDuet-4506 (Figura 13A). As sequências das duas inserções são completamente sequenciadas para excluir qualquer mutação.

(0195) O gene CrtE de *Pantoea agglomerans* que codifica uma GGPP-sintase (Número de acesso M90698) é selecionado e é sintetizada com otimização de códon (DNA2.0, Menlo Park, CA 94025, EUA). O gene CrtE é amplificado com os primers CrtE_Nco (SEQ ID NO:79) e CrtE_Bam (SEQ ID NO:80) a fim de introduzir os locais de restrição NcoI e BamHI na extremidade 5' e extremidade 3', respectivamente. A PCR é realizada com a polimerase de DNA Pfu (Promega AG, Dubendorf, Suíça)

nas mesmas condições conforme descrito acima. O produto dessa amplificação é ligado no primeiro MCS de plasmídeo pETDuet entre os locais de restrição *NcoI* e *BamHI*, fornecendo o plasmídeo pETDuet-CrtE. O plasmídeo pETDuet-SsLPPS3-del4 (contendo o cDNA de SsLPPS3 com uma deleção de 63 códons (SsLPPS3-del4, SEQ ID NO:31, Exemplo 4) é digerido com as enzimas de restrição *NdeI* e *KpnI*. A inserção é recuperada e transferida para os mesmos locais do plasmídeo pETDuet-CrtE, resultando no plasmídeo pETDuet-CrtE-SsLPPS3-del4 contendo ambos os genes SsLPPS e CrtE sob o controle de um promotor T7 (Figura 13B).

(0196) O plasmídeo pET101-1132-2-5 (contendo o cDNA de SsTps1132 com uma deleção de 50 códons (SsTps1132-2-5, SEQ ID NO:73, Exemplo 8)) é digerido com enzimas de restrição *NcoI* e *SacI* e a inserção transferida para os mesmos locais do plasmídeo pRSDuet-1, fornecendo o plasmídeo pRSDuet-1132-2-5 (não representado na figura 13).

(0197) Células *E. coli* BL21 Star™(DE3) (Invitrogen) podem ser cotransformados com os 3 plasmídeos pACYCDuet-4506 (Figura 13A), pETDuet-CrtE-SsLPPS3-del4 (Figura 13B) e pRSDuet-1132-2-5 (cada um hospedando uma origem diferente de gene de réplica e resistência). Células transformadas são selecionadas em placas de carbenicilina (50µg/ml) cloranfenicol (34µg/ml) canamicina (35µg/ml) LB-agarose. Colônias únicas são usadas para inocular 5mL de meio de LB líquido, suplementado com os mesmos antibióticos. A cultura é incubada durante à noite a 37°C. No dia seguinte, 2mL do meio de TB suplementado com os mesmos antibióticos são inoculados com 0,2mL da cultura durante à noite. Após 6 horas de incubação a 37°C, a cultura é resfriada para 28°C e 1mM de IPTG, 2mg/mL de mevalonato (preparado por dissolução de mevalonolactona (Sigma) em 0,5N de NaOH em uma concentração de 1g/mL e incubando a solução durante 30 min a 37°C) e 0,2ml de

decano são adicionados a cada tubo. As culturas são incubadas durante 48 horas a 28°C. As culturas são, então, extraídas duas vezes com 2 volumes de acetato etílico, a fase orgânica é concentrada a 500µL e analisada por CG-MS, conforme descrito acima no Exemplo 3.

(0198) Esse exemplo mostra que uma célula *E. coli* transformada com LPP-sintase e um esclareol-sintase, conforme definido na presente invenção, é capaz de produzir esclareol. As outras enzimas com as quais a célula *E. coli* é transformada não são essenciais para a produção de esclareol. Na verdade, o esclareol também é produzido quando uma célula de *E. coli* é transformada com a LPP-sintase e a esclareol-sintase apenas, mas em quantidades menores. As outras enzimas com as quais a célula de *E. coli* é transformada são adicionadas para o único propósito de aumentar a quantidade de precursor disponível para a LPP-sintase e esclareol-sintase.

Exemplo 11

(0199) Produção *in vivo* de esclareol em *E. coli* pela expressão de uma proteína de fusão contendo uma LPP-sintase e uma esclareol-sintase

(0200) A mesma abordagem é usada para a produção de esclareol com uma proteína de fusão contendo a LPP-sintase e a esclareol-sintase, ligadas em conjunto em um único polipeptídeo. Um plasmídeo é preparado contendo um novo cDNA que compreende a sequência SsLLPS3-del4 (SEQ ID NO:31) na parte 5' e sequência 1132-2-5 (SEQ ID NO:73) na parte 3' (SEQ ID NO:87). Nesse cDNA de fusão, o códon de interrupção de cDNA de SsLPPS3-del4 e o códon de início do cDNA de 1132-2-5 são deletados. Os dois cDNAs são ligados por uma sequência de 15 bp (SEQ ID NO:81) que codifica um peptídeo (SEQ ID NO:82), que constitui o ligador entre os dois domínios de proteína. A sequência de aminoácido do polipeptídeo de

fusão codificado é fornecida na SEQ ID NO:88.

(0201) O cDNA de SsLPPS3-del4 (SEQ ID NO:31) é reamplificado do plasmídeo pETDuet-SsLPPS3-del4 usando os primers Sa3del4-fusão-inf1 (SEQ ID NO:83) e Sa3del-fusão-inf2 (SEQ ID NO:84). O primeiro primer é criado para adicionar na extremidade 5' uma sequência de 15bp, complementar à região *NdeI* do plasmídeo pETDuet-1. O segundo primer é criado para remover o códon de interrupção e adicionar a sequência que codifica o peptídeo ligador. O cDNA de 1132-2-5 (SEQ ID NO:73) é reamplificado do plasmídeo pET101-1132-2-5 com os primers 1132-fusão-inf1 (SEQ ID NO:85) e 1132-inf2 (SEQ ID NO:86). Assim para SsLPPS-del4, esses primers são criados para adicionar a sequência do ligador na extremidade 3' e uma sequência de 15bp, complementar com a sequência da região *KpnI* do plasmídeo pETDuet-1. A PCR é realizada com a polimerase de DNA *Pfu* (Promega AG, Dubendorf, Suíça) nas mesmas condições conforme descrito acima no Exemplo 10. O plasmídeo pETDuet-CrtE-SsLPPS3-del4 é digerido com as enzimas *NdeI* e *KpnI* e o DNA de plasmídeo linear é purificado da inserção. Os dois produtos de PCR e o plasmídeo linear são combinados e ligados juntos usando Kit de Clonagem de PCR In-Fusion™ Dry-Down (clontech). O plasmídeo obtido, pETDuet-CrtE-fusão-SsLPPS-1132 (Figura13C) é DNA controlado por seqüenciamento da inserção.

(0202) Para avaliar a produção de esclareol do GGPP por essa enzima, células de *E. coli* BL21 Star™(DE3) (Invitrogen) podem ser cotransformadas com os plasmídeos pACYCDuet-4506 (Figura 13A) e pETDuet-CrtE-SsLPPS-1132-fusão (Figura 13C). Células transformadas são selecionadas em placas de carbenicilina (50µg/ml) cloranfenicol (34µg/ml) LB-agarose. A cultura, a indução, a suplementação com mevalonato, a extração e a análise são realizadas conforme descrito no Exemplo 10.

(0203) Esse exemplo mostra que uma célula *E. coli* transformada para

expressar uma proteína de fusão com um domínio de LPP-sintase e um domínio de esclareol-sintase, conforme definido na presente invenção, é capaz de produzir esclareol. As outras enzimas com as quais a célula de *E. coli* é transformada não são essenciais para a produção de esclareol. Na verdade, o esclareol também é produzido quando uma célula de *E. coli* é transformada com a LPP-sintase e a esclareol-sintase apenas, mas em quantidades muito menores. As outras enzimas com as quais a célula de *E. coli* é transformada são adicionadas para o único propósito de aumentar a quantidade de precursor disponível para a LPP-sintase e esclareol-sintase.

REIVINDICAÇÕES

1. MÉTODO PARA PRODUÇÃO DE ESCLAREOL, caracterizado pelo fato de que compreende

a) incubação de geranyl geranyl pirofosfato (GGPP) com um polipeptídio isolado compreendendo a sequência de aminoácidos de SEQ ID NO: 1 ou 2, sob condições eficazes para formar labdenediol difosfato (LPP);

b) incubar o LPP produzido na etapa a) com um polipeptídio isolado compreendendo a sequência de aminoácidos da SEQ ID NO: 3, sob condições eficazes para produzir esclareol, produzindo assim esclareol.

2. MÉTODO PARA PRODUÇÃO DE ESCLAREOL, de acordo com a reivindicação 1, **CHARACTERIZADO** pelo referido polipeptídio ter uma atividade de síntese de LPP compreendendo uma sequência de aminoácido SEQ ID NO:1 ou 2 e uma atividade de síntese de esclareol compreendendo uma sequência de aminoácido SEQ ID NO:1.

3. MÉTODO PARA PRODUÇÃO DE ESCLAREOL, de acordo com a reivindicação 2, **CHARACTERIZADO** pelo referido polipeptídio ter uma atividade de síntese de LPP compreendendo a SEQ ID NO:1 ou 2 e em que o referido polipeptídio ter uma atividade de síntese de esclareol compreendendo SEQ ID NO:3.

4. MÉTODO PARA PRODUÇÃO DE ESCLAREOL, de acordo com a reivindicação 3, **CHARACTERIZADO** pelo referido polipeptídio ter uma atividade de síntese de LPP consistindo na SEQ ID NO:1 ou 2 e em que o referido polipeptídio tem uma atividade de síntese de esclareol consistindo na SEQ ID NO:3.

Figura 1

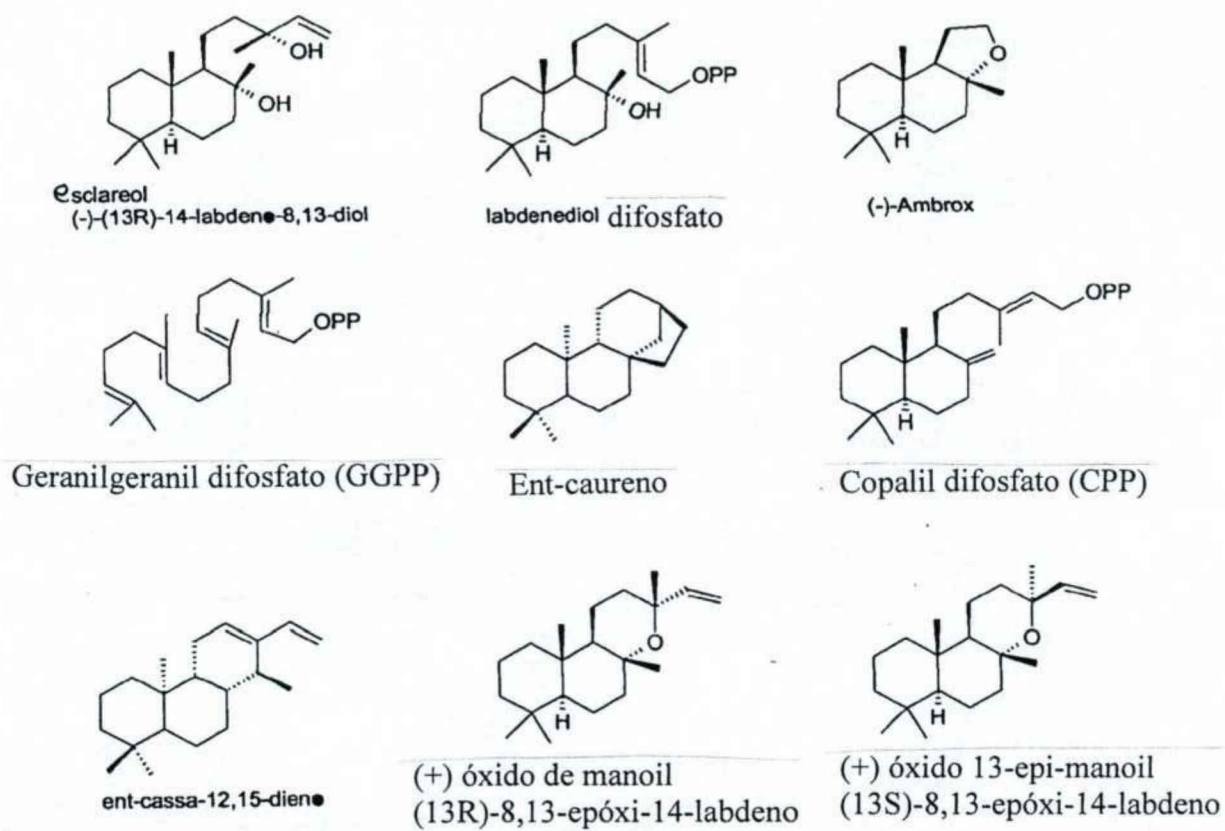


Figura 2

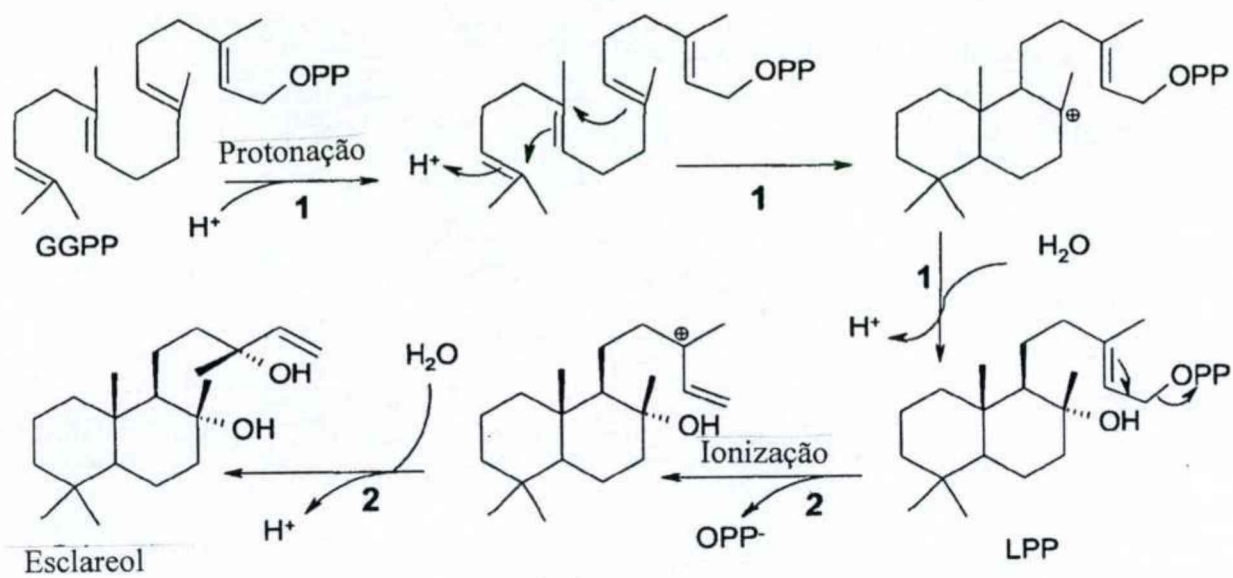


Figura 3

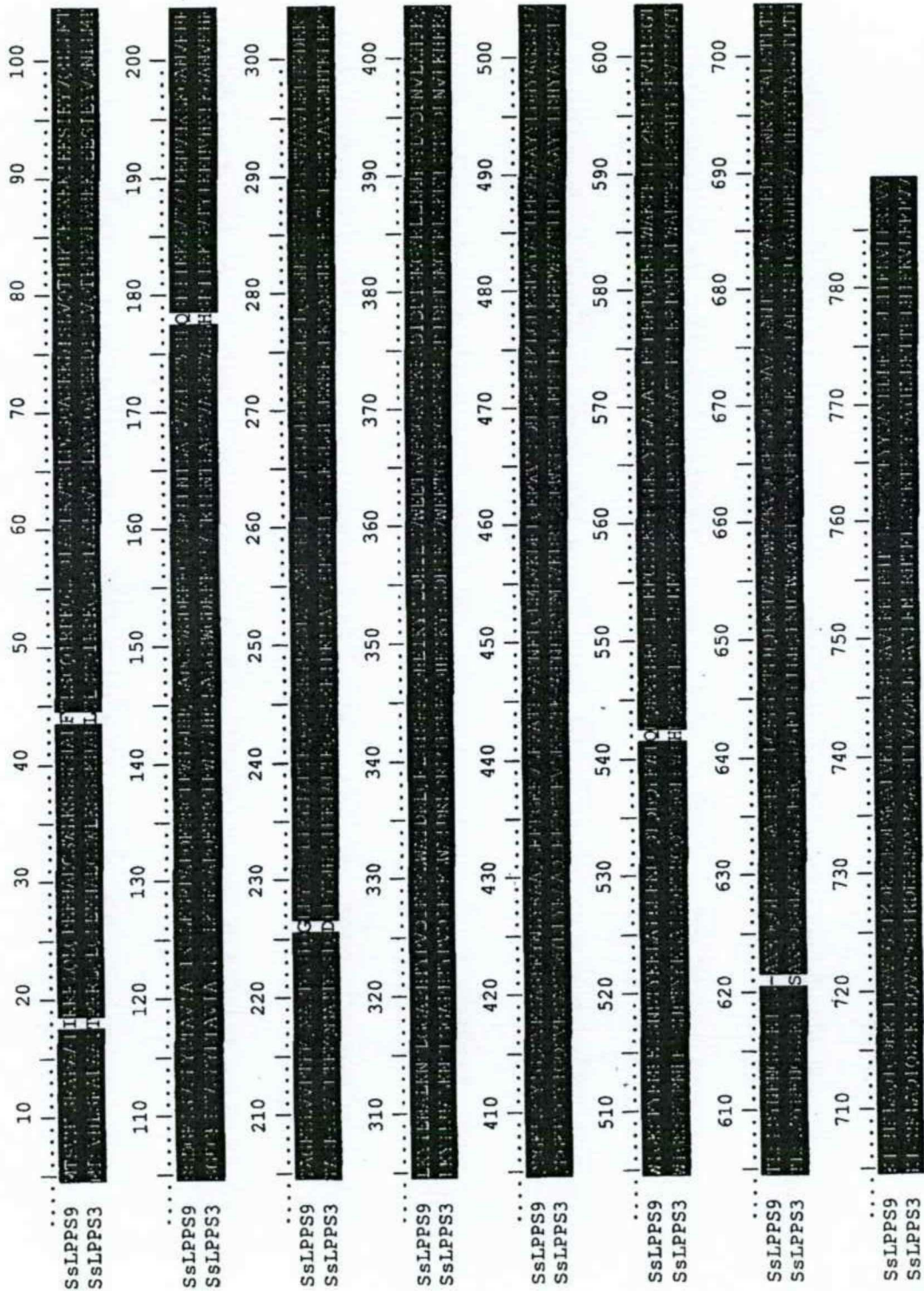


Figura 4

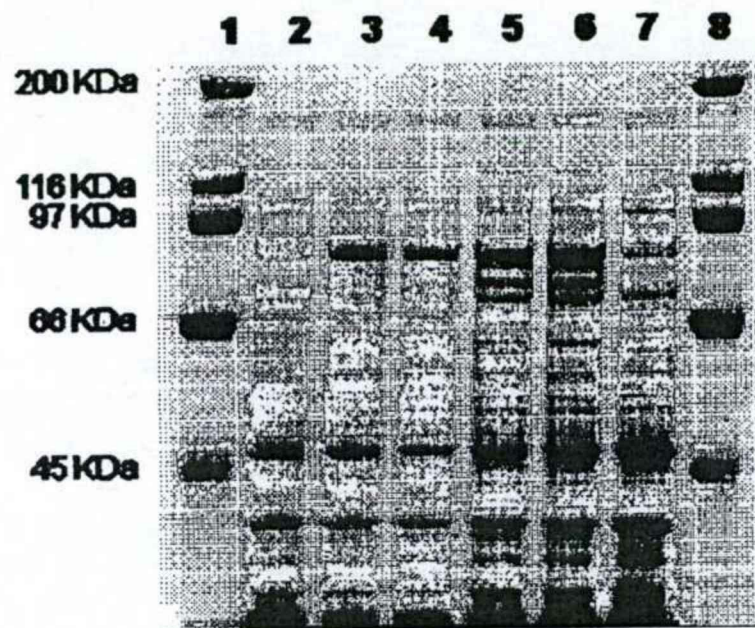


Figura 5

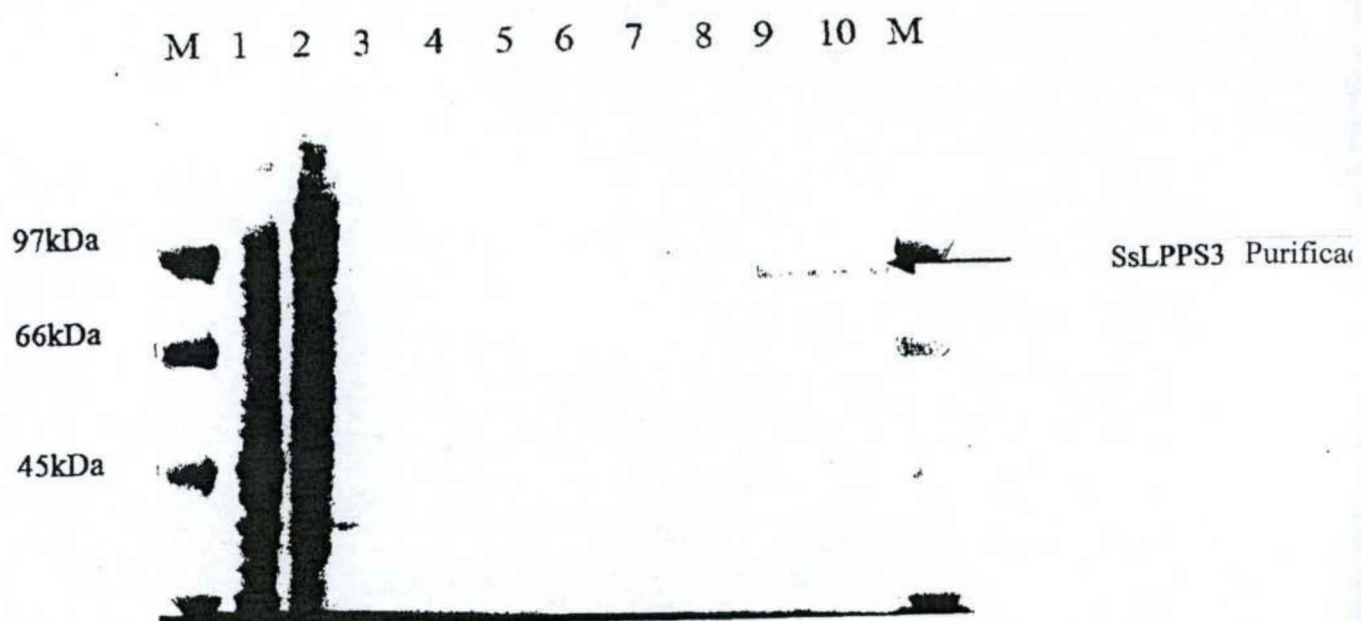


Figura 6

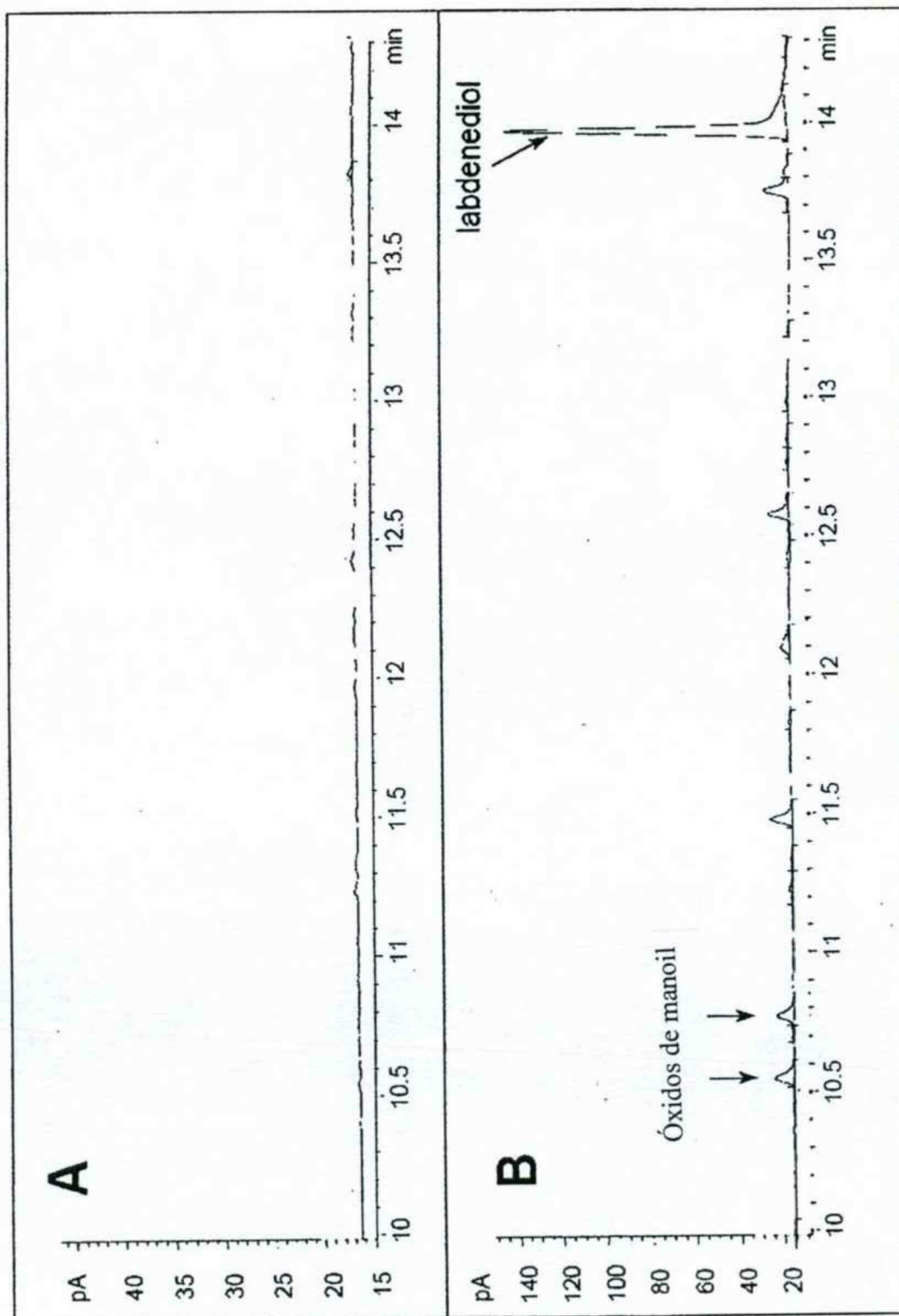


Figura 7

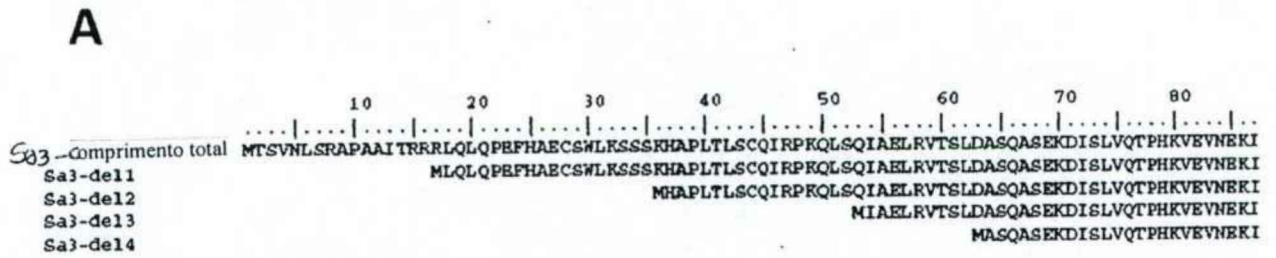
**B**

Figura 8

10 20 30 40 50 60 70 80 90 100
|.....|.....|.....|.....|.....|.....|.....|.....|.....|
 MLLSSSYSGQFPQVSLGTRPKRSTTVPLPVTRATAGGVNNEVVGNAGTLQGMIDELRVIRKQLQGVLSPPSYDTAWVAMPVQGSQSPC estemodeno sintase
 110 120 130 140 150 160 170 180 190 200
|.....|.....|.....|.....|.....|.....|.....|.....|.....|
 FPQCVEMILQNOQEDSGHSGAGSPGEVKNKILLVNTWNVGQDHIRGLSFIGNFESVAIDGQCAAPVGFNITFSGMLHLAIGNGLKFPVME estemodeno sintase
 210 220 230 240 250 260 270 280 290 300
|.....|.....|.....|.....|.....|.....|.....|.....|.....|
 TDIDSIFRLREVEFERDAGGTASAKAFMAYVSEGLGREQDWDHVMARVLFNSPSTTAASAIHSCNDRALDYLVLTSKLGPPVPAIHPDKVYSQ estemodeno sintase
 310 320 330 340 350 360 370 380 390 400
|.....|.....|.....|.....|.....|.....|.....|.....|.....|
 LCMVDTLKMGISSDFACDIRDILDMYSCWQDEEEIMLDMVCAKAPLHMHEIDCHGGRFAERSFPDDSHAYLNDTKPLLELYKSSQLHFLE estemodeno sintase
 410 420 430 440 450 460 470 480 490 500
|.....|.....|.....|.....|.....|.....|.....|.....|.....|
 EDLTILENSSWSAKLLKQLSSNKIMKSLMPEVEYALKYPLYSTVDALHGRNIERNVNGFQRPKSGYCGSGADKEILALAVDKFHYNQSVYQQELRYL estemodeno sintase
 510 520 530 540 550 560 570 580 590 600
|.....|.....|.....|.....|.....|.....|.....|.....|.....|
 ESWVAEFGIDELKFARVIPLEQSLISALVPLFPAELSDAIPQSNQCMTHMVFSGMEEMVNFVALIDEWDNHEIGFCSSNNVEIMFNAYNTTK estemodeno sintase
 610 620 630 640 650 660 670 680 690 700
|.....|.....|.....|.....|.....|.....|.....|.....|.....|
 RNCAKAALVQNRVUMDHIAKQWQVVRAMKTEAEWAASRHIPATMEEYHVSVEPSFALGPVPLSAVLLGEELPEEAVRSPEYQQLLRHASAVGRLNDV estemodeno sintase
 710 720 730 740 750 760 770 780 790 800
|.....|.....|.....|.....|.....|.....|.....|.....|.....|
 MTYEKEVLWTNPNSVLLQALAAARGGESPTPPSPACAEARGEVRRAIQASWRDLHLVFRDDDGSSIVRAIPGKGTAKVANVFQEVGDGTFPKAM estemodeno sintase
 810
|.....|.....|.....|.....|.....|.....|.....|.....|.....|
 RGMANAVILDPLHLQQ estemodeno sintase

Figura 9

[illegible]

Figura 10

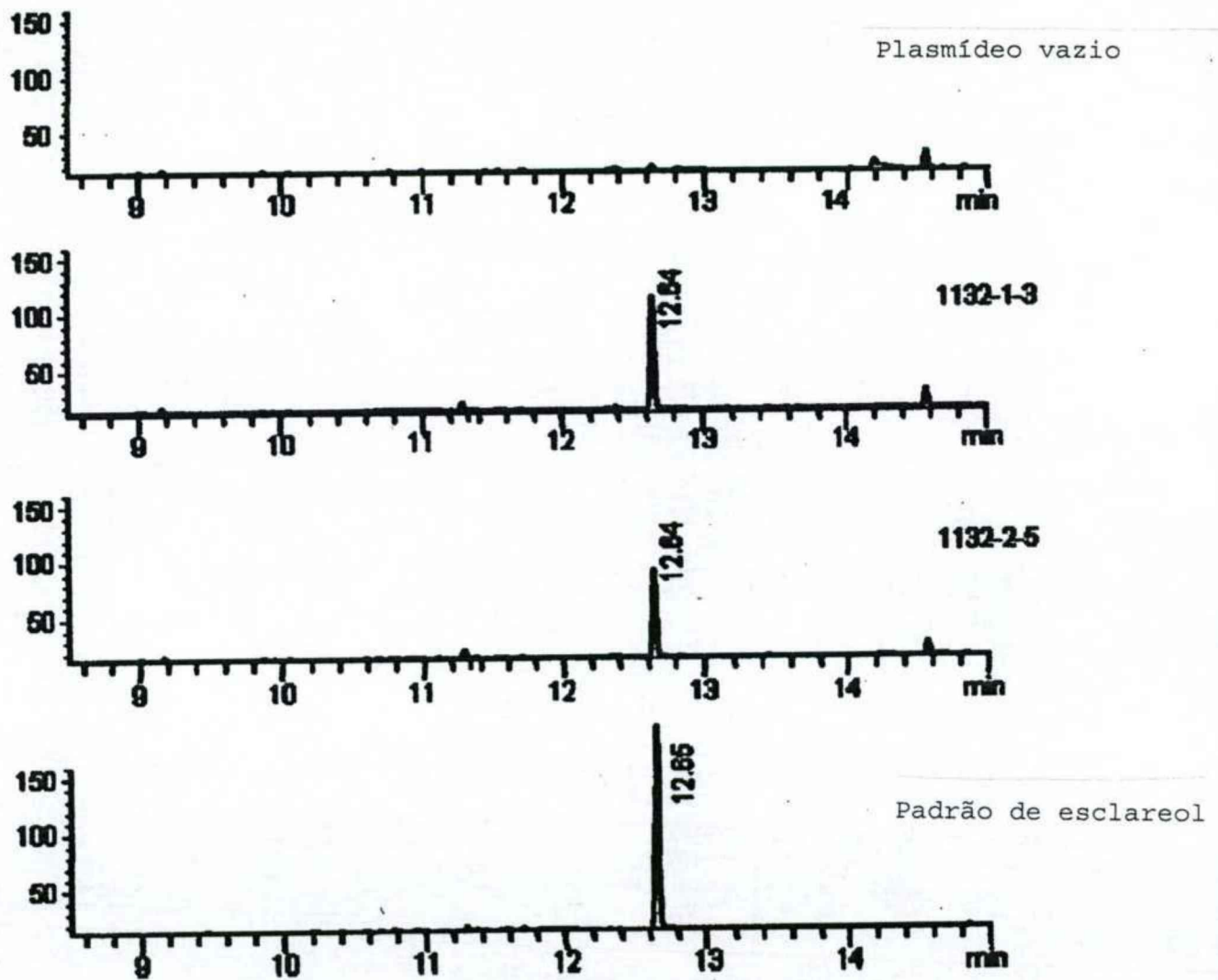


Figura 11

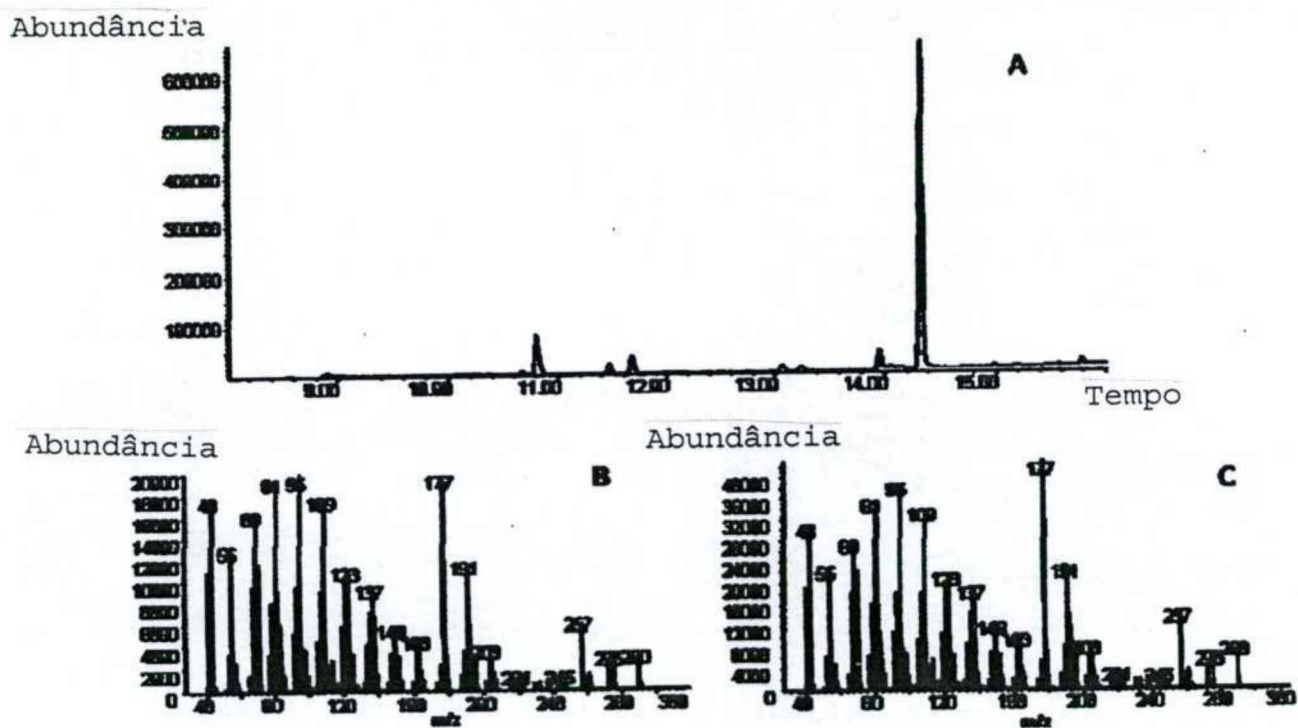


FIGURA 12

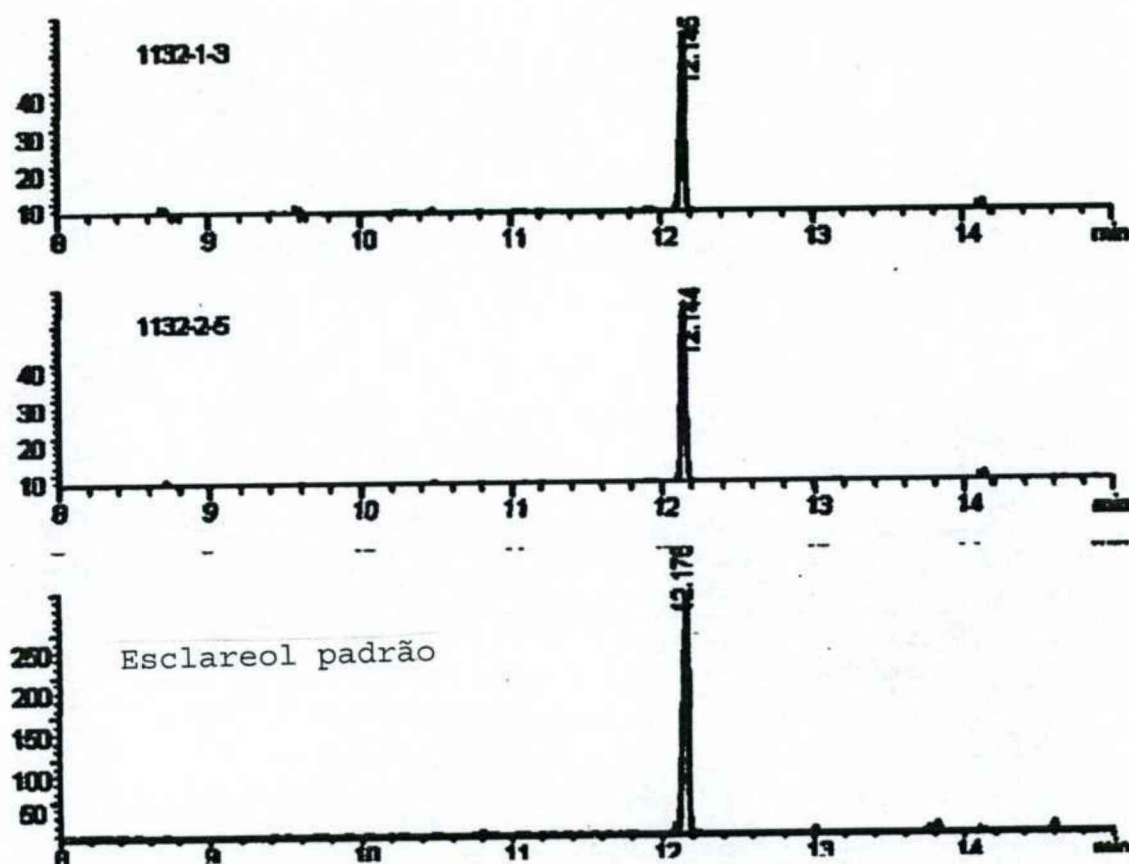
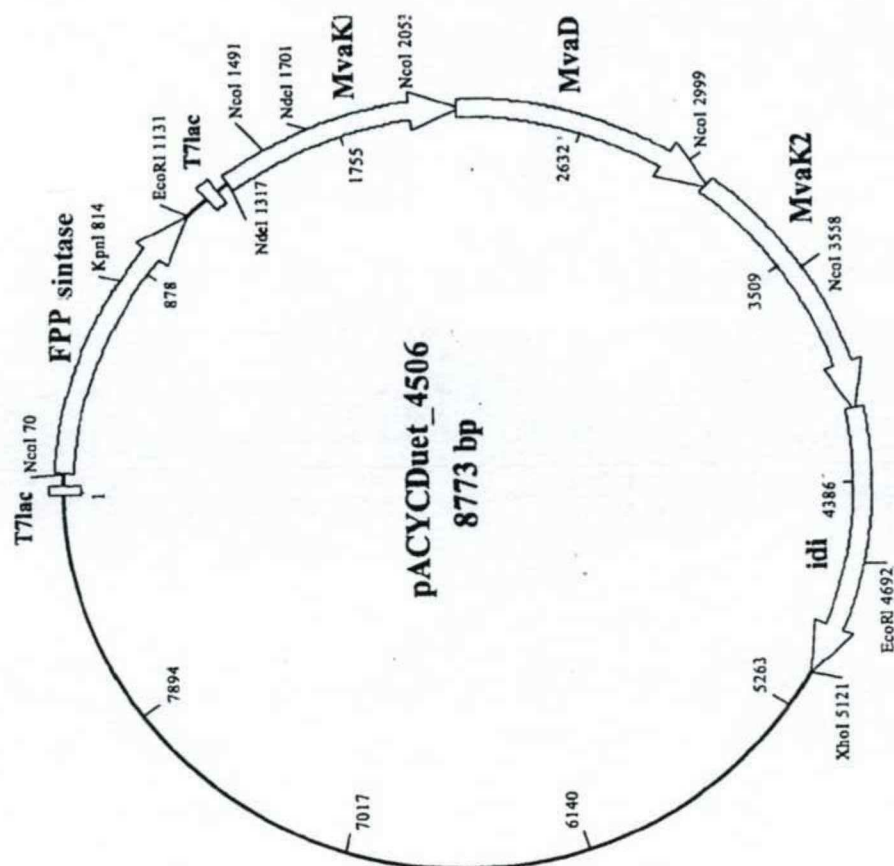


Figura 13 (1ª parte)

A



B

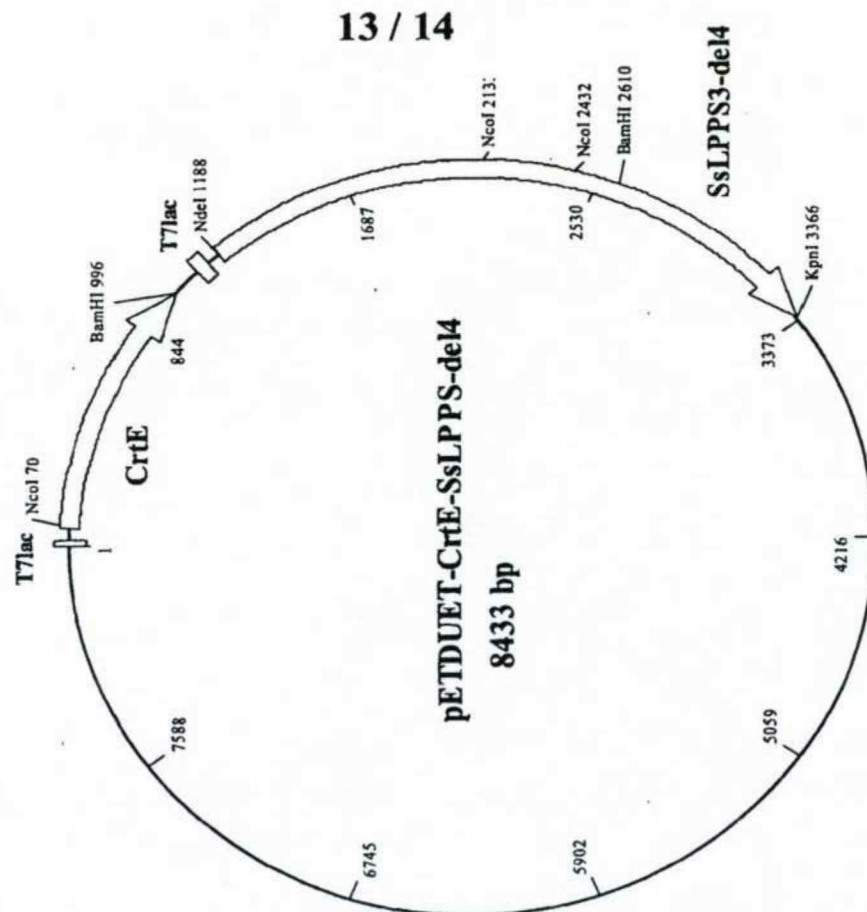


Figura 13 (2ª parte)

C

