

República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e do Comércio Exterior
Instituto Nacional da Propriedade Industrial.

(21) **PI0616883-3 A2**

(22) Data de Depósito: 05/10/2006
(43) Data da Publicação: 05/07/2011
(RPI 2113)



(51) Int.Cl.:
C12N 9/02 2006.01
C12N 1/20 2006.01
C12N 15/00 2006.01
C07H 21/04 2006.01

(54) Título: **ÁCIDOS NUCLEÍCOS QUE CODIFICAM ENZIMA MODIFICADA DE CITOCROMO P450 E MÉTODOS DE USO DOS MESMOS**

(30) Prioridade Unionista: 07/10/2005 US 60/724,525,
27/01/2006 US 60/762,700

(73) Titular(es): THE REGENTS OF THE UNIVERSITY OF CALIFORNIA

(72) Inventor(es): DAE-KYUN RO, JAY D. KEASLING, MICHELLE CHIA-YU CHANG, RACHEL EACHUS, YASUO YOSHIKUNI

(74) Procurador(es): Dannemann, Siemsen, Bigler & Ipanema Moreira

(86) Pedido Internacional: PCT US2006039433 de 05/10/2006

(87) Publicação Internacional: WO 2007/044688 de 19/04/2007

(57) Resumo: ÁCIDOS NUCLEÍCOS QUE CODIFICAM ENZIMA MODIFICADA DE CITOCROMO P450 E MÉTODOS DE USO DOS MESMOS. A presente invenção proporciona ácidos nucleicos compreendendo seqüências de nucleotídeo que codificam enzima modificada de citocromo P450s; bem como vetores recombinantes e células hospedeiras compreendendo os ácidos nucleicos. A presente invenção proporciona adicionalmente métodos de produção de um composto funcionalizado em uma célula hospedeira geneticamente modificada com um ácido nucleico compreendendo seqüências de nucleotídeo que codificam uma enzima modificada de citocromo P450.



Relatório Descritivo da Patente de Invenção para "**ÁCIDOS NUCLEÍCOS QUE CODIFICAM ENZIMA MODIFICADA DE CITOCROMO P450 E MÉTODOS DE USO DOS MESMOS**".

REFEÊNCIA CRUZADA

5 Este pedido reivindica o benefício do Pedido de Patente Provisório dos Estados Unidos No. 60/724.525, depositado em 7 de outubro de 2005, e Pedido de Patente Provisório dos Estados Unidos No. 60/762.700, depositado em 27 de janeiro de 2006, cujos pedidos são incorporados aqui por referência em sua totalidade

10 **CAMPO DA INVENÇÃO**

A presente invenção está no campo de produção de compostos de isoprenóide, e, em particular, células hospedeiras que são geneticamente modificadas com ácidos nucleicos que codificam enzimas de codificação de precursor de isoprenóide.

15 **ANTECEDENTES DA INVENÇÃO**

Os isoprenóides contituem um grupo extremamente grande e diverso de produtos naturais que têm uma origem biosintética comum, isto é, um precursor metabólico simples, isopentenil difosfato (IPP). Os compostos de isoprenóides são também referidos como "terpenos" ou "terpenóides."

20 Mais de 40.000 isoprenóides foram descritos. Por definição, os isoprenóides são produzidos das assim denominadas unidades de isopreno (C5). O número de átomos de carbono presente nos isoprenóides é tipicamente divisível por cinco (C5, C10, C15, C20, C25, C30 e C40), embora isoprenóides irregulares e politerpenos tenham sido reportados. Membros importantes dos

25 isoprenóides incluem os carotenóides, sesquiterpenóides, diterpenóides, e hemiterpenos. Os carotenóides incluem, por exemplo, licopeno, β -caroteno, e similares, muitos dos quais funcionando como antioxidantes. Os sesquiterpenóides incluem, por exemplo, artemisinin, um composto tendo atividade anti-malária. Os diterpenóides incluem, por exemplo, taxol, um agente quimioterápico de câncer.

30

Os isoprenóides compreendem a família mais numerosa e estruturalmente diversa de produtos naturais. Nesta família, os terpenóides

isolados de plantas e outras fontes naturais são usados como compostos de aroma e fragrância comerciais, bem como drogas anti-malária e anti-câncer. Uma maioria dos compostos de terpenóide em uso hoje são produtos naturais ou seus derivados. Os organismos fontes (por exemplo, árvores, invertebrados marinhos) de muitos destes produtos naturais são, nem acessíveis ao cultivo de larga escala para produzir quantidades comercialmente viáveis, nem para manipulação genética para produção aumentada ou derivatização destes compostos. Portanto, os produtos naturais devem ser produzidos semi-sinteticamente de análogos, ou sinteticamente usando-se síntese química convencional. Além disso, muitos produtos naturais têm estruturas complexas, e, como um resultado, são atualmente não-econômicos ou impossíveis de sintetizar. Tais produtos naturais devem ser ou extraídos de suas fontes nativas, tais como árvores, esponjas, corais e micróbios marinhos; ou produzidos sinteticamente ou semi-sinteticamente de precursores mais abundantes. A extração de um produto natural de uma fonte nativa é limitada pela disponibilidade da fonte nativa; e a produção sintética ou semi-sintética de produtos naturais pode sofrer de baixo rendimento e/ou alto custo. Tais problemas de produção e disponibilidade limitada da fonte natural podem restringir o desenvolvimento comercial e clínico de tais produtos.

20 A biosíntese de produtos naturais de isoprenóide em células hospedeiras projetadas (geneticamente modificadas), por exemplo, *in vitro* (por exemplo, em um sistema de fermentação), ou *in vivo* (por exemplo, em um organismo multicelular geneticamente modificado), pode diminuir o potencial comercial e terapêutico não-realizados destas fontes naturais, e produzir química fina e farmacêuticos menos caros e mais amplamente disponíveis. Um obstáculo à produção de isoprenóide ou precursor de compostos de isoprenóide em hospedeiro geneticamente modificado é a produção eficiente de enzimas que modificam os precursores de poliprenil de compostos de isoprenóide, ou que modificam os precursores de isoprenóides.

30 Uma das classes mais importantes de enzimas nas transformações bioquímicas de muitos alvos de produto natural é a superfamília de citocromo P450 (P450), que toma parte em um espectro surpreendentemen-

te amplo de reações metabólicas. Em um exemplo espantoso, P450s catalisa 8 das aproximadamente 20 etapas na biosíntese de taxol de seu precursor, geranil geranil pirofosfato.

Existe uma necessidade na técnica de produção de isoprenóide de aperfeiçoado ou células hospedeiras de produção de precursor de isoprenóide que proporcionam produção de alto nível de compostos de isoprenóide. A presente invenção chama atenção para esta necessidade, e proporciona vantagens relacionadas.

Literatura

- 10 A Publicação de Patente dos Estados Unidos No. 2004/005678; Publicação de Patente dos Estados Unidos No. 2003/0148479; Martin et al. (2003) *Nat. Biotech.* 21(7):796-802; Polakowski et al. (1998) *Appl. Microbiol. Biotechnol.* 49: 67-71; Wilding et al. (2000) *J Bacteriol* 182(15): 4319-27; Publicação de Patente dos Estados Unidos No.
- 15 2004/0194162; Donald et al. (1997) *Appl. Env. Microbiol.* 63:3341-3344; Jackson et al. (2003) *Organ. Lett.* 5:1629-1632; Publicação de Patente dos Estados Unidos No. 2004/0072323; Publicação de Patente dos Estados Unidos No. 2004/0029239; Publicação de Patente dos Estados Unidos No. 2004/0110259; Publicação de Patente dos Estados Unidos No.
- 20 2004/0063182; Patente dos Estados Unidos 5.460.949; Publicação de Patente dos Estados Unidos No. 2004/0077039; Patente dos Estados Unidos No. 6.531.303; Patente dos Estados Unidos 6.689.593; Hamano et al. (2001) *Biosci. Biotechnol. Biochem.* 65:1627-1635; T. Kuzuyama. (2004) *Biosci. Biotechnol. Biochem.* 68(4): 931-934; T. Kazuhiko. (2004) *Biotechnology Letters.*
- 25 26: 1487-1491; Brock et al. (2004) *Eur J. Biochem.* 271: 3227-3241; Choi et al. (1999) *Appl. Environ. Microbio.* 65 4363-4368; Parke et al. (2004) *Appl. Environ. Microbio.* 70: 2974-2983; Subrahmanyam et al. (1998) *J. Bact.* 180: 4596-4602; Murli et al. (2003) *J. Ind. Microbiol. Biotechnol.* 30: 500-509; Starai et al. (2005) *J. Biol. Chem.* 280:26200-26205; and Starai et al. (2004) *J.*
- 30 *Mol. Biol.* 340:1005-1012; Jennewein et al. *Chem. Biol.* 2004, 11, 379-387; Sowden et al. *Org. Biomol. Chem.* 2005, 3, 57-64; Luo et al. *Plant J.* 2001, 28, 95-104; Carter et al. *Phytochem.* 2003, 64, 425-433; Craft et al. *Appl.*

Environ. Microbiol. 2003, 69, 5983–5991; Barnes et al. *Proc. Natl. Acad. Sci. USA* 1991, 88, 5597-5601; Schoch et al. *Plant Physiol.* 2003, 133, 1198-1208; Roosild et al. *Science* 2005, 307, 1317-1321.

SUMÁRIO DA INVENÇÃO

5 A presente invenção proporciona ácidos nucleicos compreendendo seqüências de nucleotídeo que codificam enzimas modificadas de citocromo P450s; bem como vetores recombinantes e células hospedeiras compreendendo os ácidos nucleicos. A presente invenção proporciona adicionalmente métodos de produção de composto funcionalizado em uma célula hospedeira geneticamente modificada com um ácido nucleico compreendendo seqüências de nucleotídeo que codificam uma enzima modificada de citocromo P450.

BREVE DESCRIÇÃO DOS DESENHOS

15 A Figura 1 representa esquematicamente biosíntese de 8-hidroxi- δ -cadinene em *E. coli*.

A Figura 2 representa investigação da cromatografia de gás-espectrometria de massa (GC-MS) de camada orgânica extraída de *E. coli* expressando trajetória biosintética de CadOH.

20 A Figura 3 representa uma investigação de GC-MS da camada orgânica extraída de mevalonato-alimentado *E. coli* expressando trajetória biosintética de CadOH, bem como uma porção da trajetória de mevalonato (pMBIS).

25 As Figura 4A e 4B representam várias modificações N-terminais produzidas para CadH (Figura 4A); e um curso de tempo de produção de CadOH por *E. coli* geneticamente modificada usando várias construções de CadH.

A Figura 5 representa uma seqüência de amino ácido de "mistic".

30 A Figura 6 representa uma seqüência de amino ácido de uma limoneno hidroxilase.

A Figura 7 representa uma seqüência de amino ácido de uma aristolocheno dihidroxilase.

As Figuras 8A-D representam uma seqüência de amino ácido de cadinene hidroxilase com um domínio nativo de transmembrana (sublinhado) (Fig. 8A); cadinene hidroxilase com um domínio heterólogo de transmembrana (texto em negrito) (Fig. 8B); cadinene hidroxilase com um
 5 domínio de solubilização (texto em negrito) (Fig. 8C); e cadinene hidroxilase com um domínio de secreção e um domínio heterólogo de transmembrana (texto em negrito) (Fig. 8D).

As Figuras 9A e 9B representam seqüências de amino ácidos de taxadieno hidroxilases.

10 A Figura 10 representa uma seqüência de amino ácido de *ent*-kaureno oxidase.

A Figura 11A representa uma seqüência de nucleotídeo que codifica cadinene hidroxilase (o início atg é mostrado em negrito); e a Figura 11B representa uma seqüência variante de nucleotídeo que codifica cadi-
 15 ne hidroxilase, códon otimizado para expressão em um procariote.

A Figura 12A representa uma seqüência de amino ácido de um citocromo P450 reductase (CPR) de *Taxus cuspidata*; A Figura 12 B representa uma seqüência de amino ácido de um CPR de *Candida tropicalis*; A Figura 12C representa uma seqüência de amino ácido de um CPR (ATR1)
 20 de *Arabidopsis thaliana*; A Figura 12D representa uma seqüência de amino ácido de um CPR (ATR2) de *Arabidopsis thaliana*; e a Figura 12E representa um amino ácido ATR2 variante que carece de uma seqüência de amino ácido que objetiva cloroplasto.

A Figura 13 representa esquematicamente duas trajetórias biosintéticas "heme".
 25

A Figura 14 representa esquematicamente a biosíntese de produtos de isoprenóide exemplares taxol, artemisinin, e mentol.

A Figura 15 representa esquematicamente o esquema de reação para produção de compostos de isoprenóide exemplares.

30 A Figura 16 é uma representação esquemática de trajetórias metabólicas de isoprenóide que resultam na produção dos intermediários de trajetória biosintética de isoprenóide poliprenil difosfatos geranil difosfato

(GPP), farnesil difosfato (FPP), e geranilgeranil difosfato (GGPPP), de isopentenil difosfato (IPP) e dimetilalil difosfato (DMAPP).

A Figura 17 é uma representação esquemática da trajetória de mevalonato (MEV) para a produção de IPP.

5 A Figura 18 é uma representação esquemática da trajetória de DXP para a produção de IPP e dimetilalil pirofosfato (DMAPP).

As Figuras 19A-C representam seqüências de amino ácido de várias enzimas P450 de modificação de intermediário de trajetória de alcalóide.

10 As Figuras 20A-C representam seqüências de amino ácido de várias enzimas P450 de modificação de intermediário de trajetória de fenilpropanóide.

As Figuras 21A e 21B representam seqüências de amino ácido de várias enzimas P450 de modificação de intermediário de trajetória de polietide.

15 A Figura 22 representa esquematicamente várias construções de amorfadieno oxidase (AMO). (1) nAMO, seqüência AMO nativa conforme isolada de *Artemisia annua*; 2) sAMO, gene AMO sintético de códon otimizado para expressão em *E. coli*; 3) A13-AMO, gene AMO sintético com membrana tipo selvagem substituída com a seqüência A13 N-terminal de *C. tropicalis*; 4) A17-AMO, gene AMO sintético com membrana tipo selvagem substituída com a seqüência A17 N-terminal de *C. tropicalis*; 5) Bov-AMO, gene AMO sintético com membrana tipo selvagem substituída com a seqüência bovina microsomal N-terminal.

25 As Figuras 23A e B representam oxidação de amorfadieno em *E. coli* por várias construções AMO construções.

Figuras 24A e B representam uma seqüência de nucleotídeo que codifica AMO tipo selvagem.

30 A Figura 25 representa um mapa de translação de seqüência de amino ácido da seqüência de nucleotídeo representada na Figura 24.

As Figuras 26 e 27 representam uma seqüência de nucleotídeo que codifica A13-AMO e um mapa de translação da seqüência de amino

ácido, respectivamente.

As Figuras 28 e 29 representam uma sequência de nucleotídeo que codifica A17-AMO e o mapa de translação da sequência de amino ácido, respectivamente.

5 As Figuras 30 e 31 representam uma sequência de nucleotídeo que codifica AMO-bovino e o mapa de translação da sequência de amino ácido, respectivamente.

A Figura 32 representa a produção de CadOH em *E. coli* contendo a trajetória total de mevalonato em adição a um vetor de expressão
10 compreendendo seqüências de nucleotídeo que codificam CadOH, CPR, e CadS.

A Figura 33 representa uma cromatografia e espectro de GC-MS mostrando produção comparativa de ácido artemisínico em *E. coli* expressando a trajetória de amorfadieno total e ou o plasmídeo pDUET-ctAACPR-A13AMO ou o plasmídeo pCWori-A17AMO-ctAACPR.
15

A Figura 34 representa cromatografias de GC-MS mostrando oxidação de álcool artemisínico em aldeído artemisínico em *E. coli* geneticamente modificada com ácidos nucléicos que codificam enzimas de trajetória de mevalonato e amorfadieno sintase, e com o plasmídeo pCWori-A17AMO-ctAACPR.
20

As Figuras 35A e 35B representam seqüências de nucleotídeo que codificam acetoacetyl-CoA tiolase ("atoB"), HMGS, e HMGR truncado (tHMGR).

As Figuras 36A-D representam a seqüência de nucleotídeo de
25 pMBIS.

DEFINIÇÕES

Os termos "isoprenóide," "composto de isoprenóide," "terpeno," "composto de terpeno," "terpenóide," e "composto de terpenóide" são usados intercambiavelmente aqui. Os composto de isoprenóide são compostos de vários números das assim denominadas unidades de isopreno
30 (C5). O número de átomos de C presente nos isoprenóides é tipicamente constantemente divisível por cinco (por exemplo, C5, C10, C15, C20, C25,

C30 e C40). Os isoprenóides e politerpenos irregulares foram reportados, e são também incluídos na definição de "isoprenóide". Os compostos de isoprenóide incluem, mas não estão limitados a, monoterpenos, sesquiterpenos, triterpenos, politerpenos, e diterpenos.

5 Conforme aqui usado, o termo "prenil difosfato" é usado intercambiavelmente com "prenil pirofosfato," e incluem monoprenil difosfatos tendo um grupo prenil simples (por exemplo, IPP e DMAPP), bem como poliprenil difosfatos que incluem 2 ou mais grupos prenil. Os monoprenil difosfatos incluem isopentenil pirofosfato (IPP) e seu isômero dimetilalil pirofosfato (DMAPP).

10 Conforme aqui usado, o termo "terpeno sintase" se refere a qualquer enzima que modifica enzimaticamente IPP, DMAPP, ou um poliprenil pirofosfato, tal que um composto precursor de terpenóide é produzido. O termo "terpeno sintase" inclui enzimas que catalisam a conversão de um

15 prenil difosfato em um isoprenóide ou precursor de isoprenóide.

 A palavra "pirofosfato" é usada intercambiavelmente aqui com "difosfato." Desse modo, por exemplo, os termos "prenil difosfato" e "prenil pirofosfato" são intercambiáveis; os termos "isopentenil pirofosfato" e "isopentenil difosfato" são intercambiáveis; os termos farnesil difosfato" e farnesil pirofosfato" são intercambiáveis; etc.

20 O termo "trajetória de mevalonato" ou "trajetória de MEV" é aqui usado para se referir a trajetória biosintética que converte acetil-CoA em IPP. A trajetória de mevalonato compreende enzimas que canalizam as seguintes etapas: (a) condensação de duas moléculas de acetil-CoA em

25 acetoacetil-CoA; (b) condensação de acetoacetil-CoA com acetil-CoA para formar HMG-CoA; (c) conversão de HMG-CoA em mevalonato; (d) fosforilação de mevalonato em mevalonato 5-fosfato; (e) conversão de mevalonato 5-fosfato em mevalonato 5-pirofosfato; e (f) conversão de mevalonato 5-pirofosfato em isopentenil pirofosfato. A trajetória de mevalonato é ilustrada

30 esquemáticamente na Figura 17. A "metade superior" da trajetória de mevalonato se refere às enzimas responsáveis pela conversão de acetil-CoA em mevalonato através de um intermediário de trajetória de MEV.

O termo "trajetória 1-deoxi-D-xilulose 5-difosfato" ou "trajetória DXP", é usado aqui para se referir à trajetória que converte gliceraldeído-3-fosfato e piruvato em IPP e DMAPP através de um intermediário de trajetória de DXP, onde a trajetória de DXP compreende enzimas que catalisam as reações representadas esquematicamente na Figura 18.

Conforme aqui usado, o termo "prenil transferase" é usado intercambiavelmente com os termos "isoprenil difosfato sintase" e "poliprenil sintase" (por exemplo, "GPP sintase," "FPP sintase," "OPP sintase," etc.) para se referir a uma enzima que catalisa a condensação consecutiva 1'-4 de isopentenil difosfato com substratos primeiros alílicos, resultando na formação de prenil difosfatos de vários comprimentos de cadeia.

Os termos "polinucleotídeo" e "ácido nucléico," usados intercambiavelmente aqui se referem a uma forma polimérica de nucleotídeos de qualquer comprimento, ou ribonucleotídeos ou deoxinucleotídeos. Desse modo, este termo inclui, mas não está limitado a, DNA ou RNA de trançado simples, duplo ou múltiplo, DNA genômico, cDNA, híbridos de DNA-RNA, ou um polímero compreendendo bases de purina e pirimidina, ou outras bases de nucleotídeo naturais, quimicamente ou bioquimicamente modificadas, não-naturais, ou derivatizadas.

Os termos "peptídeo," "polipeptídeo," e "proteína" são usados intercambiavelmente aqui, e se referem a uma forma polimérica de amino ácidos de qualquer comprimento, que podem incluir amino ácidos codificados ou não-codificados, quimicamente ou bioquimicamente modificados, ou amino ácidos derivatizados, e polipeptídeos tendo suportes de peptídeo modificados.

O termo "ocorrendo naturalmente" conforme aqui usado, conforme aplicado a um ácido nucléico, a uma célula, ou a um organismo, se refere a um ácido nucléico, célula, ou organismo que é encontrado na natureza. Por exemplo, um polipeptídeo ou seqüência de polinucleotídeo que está presente em um organismo (incluindo vírus) que podem ser isolados a partir de uma fonte na natureza, e que não tenha sido intencionalmente modificado por um humano no laboratório, está ocorrendo naturalmente.

Conforme aqui usado, o termo "isolado" é significativo para descrever um polinucleotídeo, um polipeptídeo, ou uma célula que está em um ambiente diferente daquele no qual o polinucleotídeo, o polipeptídeo, ou a célula ocorrem naturalmente. Uma célula hospedeira geneticamente modificada pode estar presente em uma população misturada de células hospedeiras geneticamente modificadas.

Conforme aqui usado, o termo "ácido nucléico exógeno" se refere a um ácido nucléico que não é normalmente ou naturalmente encontrado e/ou produzido por uma dada bactéria, organismo, ou célula na natureza.

Conforme aqui usado, o termo "ácido nucléico endógeno" se refere a um ácido nucléico que é normalmente encontrado e/ou produzido por uma dada bactéria, organismo, ou célula na natureza. Um "ácido nucléico endógeno" é também referido como um "ácido nucléico nativo", ou um ácido nucléico que é "nativo" para uma dada bactéria, organismo, ou célula. Por exemplo, os ácidos nucléicos que codificam HMGS, mevalonato kinase, e fosfomevalonato kinase representam ácidos nucléicos exógenos para *E. coli*. Estes ácidos nucléicos de trajetória de mevalonato foram clonados de *Sacchomyces cerevisiae*. Em *S. cerevisiae*, as conseqüências que codificam HMGS, MK, e PMK no cromossomo seriam "ácidos nucléicos endógenos".

O termo "ácido nucléico heterólogo", conforme aqui usado, se refere a um ácido nucléico no qual pelo menos um dos seguintes é verdadeiro: (a) o ácido nucléico é estranho ("exógeno") a (isto é, não naturalmente encontrado) um dado microorganismo hospedeiro ou célula hospedeira; (b) o ácido nucléico compreende uma seqüência de nucleotídeo que é naturalmente encontrada em (por exemplo, é "endógeno a") um dado microorganismo hospedeiro ou célula hospedeira (por exemplo, o ácido nucléico compreende uma seqüência de nucleotídeo que é endógena ao microorganismo hospedeiro, ou célula hospedeira), mas é, ou produzido em uma quantidade não-natural (por exemplo, maior do que esperada ou maior do que naturalmente encontrada) na célula, ou difere em seqüência a partir da seqüência endógena de nucleotídeo, tal que a mesma proteína codificada (tendo a mesma, ou substancialmente a mesma seqüência de amino ácido) conforme

encontrada endogenamente, é produzida em uma quantidade não-natural (por exemplo, maior do que esperada ou maior do que naturalmente encontrada) na célula; (c) o ácido nucléico compreende duas ou mais seqüências de nucleotídeo ou segmentos que não são encontrados no mesmo relacionamento na natureza, por exemplo, o ácido nucléico é recombinante.

O termo "polipeptídeo heterólogo", conforme aqui usado, se refere a um polipeptídeo que não é naturalmente associado com um dado polipeptídeo. Por exemplo, uma enzima de modificação de precursor de isoprenóide que compreende um "domínio heterólogo de transmembrana" se refere a uma enzima de modificação de precursor de isoprenóide que compreende um domínio de transmembrana que não é normalmente associado com (por exemplo, não normalmente contíguo com; não normalmente encontrado na mesma cadeia de polipeptídeo com) a enzima de modificação de precursor de isoprenóide na natureza. Similarmente, uma enzima de modificação de precursor de isoprenóide que compreende um ou mais dentre um "domínio heterólogo de secreção", um "polipeptídeo de inserção de membrana heterólogo", e um "domínio heterólogo de solubilização" é uma enzima de modificação de precursor de isoprenóide que compreende um ou mais dentre um domínio de secreção, um polipeptídeo de inserção de membrana, e um domínio de solubilização que não é normalmente associado com (por exemplo, não normalmente contíguos com; não normalmente encontrado na mesma cadeia de polipeptídeo com) a enzima de modificação de precursor de isoprenóide na natureza.

"Recombinante", conforme aqui usado, significa que um ácido nucléico particular (DNA ou RNA) é o produto de várias combinações de etapas de clonagem, restrição, e/ou ligação, resultando em uma construção tendo uma seqüência de codificação ou não-codificação estrutural distinguível dos ácidos nucléicos endógenos encontrados nos sistemas naturais. Geralmente, as seqüências de DNA que codificam a seqüência de codificação estrutural podem ser montadas a partir de fragmentos de cDNA e ligadores de oligonucleotídeos curtos, ou a partir de uma série de oligonucleotídeos sintéticos, para proporcionar um ácido nucléico sintético que é capaz de ser

expresso de uma unidade transcricional recombinante contida em uma célula, ou em um sistema de transcrição ou translação de célula, ou livre de célula. Tais seqüências podem ser providas na forma de uma estrutura de leitura aberta não-interrompida por seqüências internas não-transladadas, ou introns, que estão tipicamente presentes em genes eucarióticos. O DNA genômico compreendendo as seqüências relevantes pode também ser usado na formação de um gene recombinante, ou unidade transcricional. As seqüências de DNA não-transladado podem estar presentes 5' ou 3' a partir da estrutura de leitura aberta, onde tais seqüências não interferem com manipulação ou expressão das regiões de codificação, e podem, de fato, agirem para modular a produção de um produto desejado por vários mecanismos (ver "seqüências regulatórias de DNA", abaixo).

Desse modo, por exemplo, o termo polinucleotídeo "recombinante" ou ácido nucléico "recombinante" se referem a um que não está ocorrendo naturalmente, por exemplo, é produzido pela combinação artificial de dois segmentos, de outro modo, separados, de seqüência através de intervenção humana. Esta combinação artificial é freqüentemente acompanhada por, ou meio de síntese química, ou pela manipulação artificial de segmentos isolados de ácidos nucléicos, por exemplo, por técnicas de projetos genéticas. Tal é usualmente feito para substituir um códon com um códon redundante que codifica o mesmo, ou um amino ácido conservativo, enquanto tipicamente introduz ou remove um local de reconhecimento de seqüência. Alternativamente, é realizado unir juntos segmentos de ácido nucléico de funções desejadas para gerar uma combinação desejada de funções. Esta combinação artificial é freqüentemente acompanhada por, ou meio de síntese química, ou por manipulação artificial de segmentos isolados de ácidos nucléicos, por exemplo, por técnicas de projetos genéticas.

Similarmente, o termo polipeptídeo "recombinante" se refere a um polipeptídeo que não está ocorrendo naturalmente, por exemplo, é produzido pela combinação artificial de dois segmentos, de outro modo, separados, de seqüência de amino através de intervenção humana. Desse modo, por exemplo, um polipeptídeo que compreende uma seqüência heteróloga

de amino ácido é recombinante.

Por "construção" ou "vetor" é significativo um ácido nucléico recombinante, genialmente DNA recombinante, que tenha sido gerado para a proposta da expressão e/ou propagação de uma seqüência específica de nucleotídeo(s), ou é para ser usado na construção de outra seqüência re-combinante de nucleotídeos.

Conforme aqui usado, os termos "operon" e "unidade de transcrição simples" são usados intercambiavelmente para se referirem a duas ou mais regiões de codificação contíguas (seqüência de nucleotídeos que codifica um produto de gene, tal como um RNA ou uma proteína) que são coordenadamente reguladas por um ou mais elementos de controle (por exemplo, um promotor). Conforme aqui usado, o termo "produto de gene" se refere a RNA codificado por DNA (ou vice-versa), ou proteína que é codificada por um RNA ou DNA, onde um gene tipicamente compreenderá uma ou mais seqüências de nucleotídeo que codificam uma proteína, e pode também incluir introns e outras seqüências de não-codificação de nucleotídeos.

Os termos "seqüências regulatórias de DNA", "elementos de controle", e "elementos regulatórios", usados intercambiavelmente aqui, se referem a seqüências de controle transcricionais e translacionais, tais como promotores, intensificadores, sinais de poliadenilação, terminadores, sinais de degradação de proteína, e similares, que proporcionam e/ou regulam expressão de uma seqüência de codificação e/ou produção de um polipeptídeo codificado em uma célula hospedeira.

O termo "transformação" é usado intercambiavelmente aqui com "modificação genética", e se refere a uma mudança genética permanente ou transiente induzida em uma célula, seguindo introdução de novo ácido nucléico (isto é, DNA exógeno à célula). A mudança genética ("modificação") pode ser acompanhada, ou pela incorporação do novo DNA no genoma da célula hospedeira, ou por manutenção transiente ou estável do novo DNA como um elemento episomal. Onde a célula é uma célula eucariótica, uma mudança genética permanente é geralmente alcançada pela introdução do DNA no genoma da célula. Em células procarióticas, mudanças permanen-

- tes podem ser introduzidas no cromossomo, ou via elementos extracromossomais, tais como plasmídeos e vetores de expressão, que podem conter um ou mais marcadores selecionáveis para auxiliar em sua manutenção na célula hospedeira recombinante. Métodos adequados de modificação genética incluem infecção viral, transfecção, conjugação, fusão de protoplasto, electroporação, tecnologia de canhão de partícula, precipitação de fosfato de cálcio, microinjeção direta, e similares. A escolha do método é geralmente dependente do tipo de célula a ser transformada, e das circunstâncias sob as quais a transformação está ocorrendo (isto é, *in vitro*, *ex vivo*, ou *in vivo*).
- 10 Uma discussão geral destes métodos pode ser encontrada em Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995.

"Operavelmente ligado" se refere a uma justaposição na qual os componentes assim descritos estão em um relacionamento que permite que os mesmos funcionem em sua maneira pretendida. Por exemplo, um promotor é operavelmente ligado a uma sequência de codificação se o promotor afeta sua transcrição ou expressão. Conforme aqui usado, os termos "promotor heterólogo" e "regiões de controle heterólogas" se referem a promotores e outras regiões de controle que não estão normalmente associadas com um ácido nucléico particular na natureza. Por exemplo, uma "região de controle transcricional heteróloga a uma região de codificação" é uma região de controle transcricional que não está normalmente associada com a região de codificação na natureza.

Uma "célula hospedeira", conforme aqui usado, denota uma célula eucariótica *in vivo* ou *in vitro*, uma célula procariótica, ou uma célula de um organismo multicelular (por exemplo, uma linha de célula) cultivada como uma entidade unicelular, cujas células eucarióticas ou procarióticas podem ser, ou tenham sido, usadas como recipientes para um ácido nucléico (por exemplo, um vetor de expressão que compreende uma sequência de nucleotídeo que codifica um ou mais produtos de gene de trajetória biosintética, tais como produtos de gene de trajetória de mevalonato), e incluem a progenia da célula original que tenha sido geneticamente modificada pelo ácido nucléico. É compreendido que a progenia de uma célula simples pode

não ser necessariamente completamente idêntica na morfologia ou no complemento genômico ou de DNA total como fonte original, devido à mutação natural, acidental, ou deliberada. Uma "célula hospedeira recombinante" (também referida como "célula hospedeira geneticamente modificada") é uma célula hospedeira na qual tenha sido introduzido um ácido nucléico heterólogo, por exemplo, um vetor de expressão. Por exemplo, uma célula hospedeira procariótica objeto é uma célula hospedeira procariótica geneticamente modificada (por exemplo, uma bactéria), em virtude da introdução em uma célula hospedeira procariótica adequada de um ácido nucléico heterólogo, por exemplo, um ácido nucléico exógeno que é estranho a (não normalmente encontrado na natureza) célula hospedeira procariótica, ou um ácido nucléico recombinante que não é normalmente encontrado na célula hospedeira procariótica; e uma célula hospedeira eucariótica objeto é uma célula hospedeira eucariótica geneticamente modificada, em virtude da introdução em uma célula hospedeira eucariótica adequada de um ácido nucléico heterólogo, por exemplo, um ácido nucléico exógeno que é estranho à célula hospedeira eucariótica, ou um ácido nucléico recombinante que não é normalmente encontrado na célula hospedeira eucariótica.

O termo "substituição de amino ácido conservativo" se refere à intercambialidade em proteínas de resíduos de amino ácido tendo cadeias laterais similares. Por exemplo, um grupo de amino ácidos tendo cadeias laterais alifáticas consiste de glicina, alanina, valina, leucina, e isoleucina; um grupo de amino ácidos tendo cadeias laterais alifáticas-hidroxil consiste de serina e treonina; um grupo de amino ácidos tendo cadeias laterais contendo amida consiste de asparagina e glutamina; um grupo de amino ácidos tendo cadeias laterais aromáticas consiste de fenilalanina, tirosina, e triptofan; um grupo de amino ácidos tendo cadeias laterais básicas consiste de lisina, arginina, e histidina; e um grupo de amino ácidos tendo cadeias laterais contendo enxofre consiste de cisteína e metionina. Grupos de substituição de amino ácidos conservativos exemplares são: valina-leucina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina, e asparagina-glutamina.

"Ácidos nucléicos sintéticos" podem ser agregados de blocos de construção de oligonucleotídeo que são quimicamente sintetizados usando-se procedimentos conhecidos àqueles técnicos no assunto. Estes blocos de construção são ligados e temperados para formar segmentos de gene que são, em seguida, enzimaticamente agregados para construção do gene total. "Quimicamente sintetizado," conforme relacionado a uma seqüência de DNA, significa que os nucleotídeos do componente foram agregados *in vitro*. A síntese química manual de DNA pode ser acompanhada usando-se procedimentos bem estabelecidos, ou síntese química automática pode ser realizada usando-se um de um número de máquinas comercialmente disponíveis. A seqüência de nucleotídeo dos ácidos nucléicos pode ser modificada para expressão ótima baseada na otimização da seqüência de nucleotídeo para refletir o códon diagonal da célula hospedeira. O técnico no assunto aprecia a probabilidade de expressão bem sucedida se uso do códon é inclinado em direção àqueles códons favorecidos pelo hospedeiro. A determinação de códons preferidos pode estar baseada em uma sobrevivência de genes derivados a partir da célula hospedeira onde informação de seqüência é disponível.

Um polinucleotídeo ou polipeptídeo tem uma certa percentagem de "identidade de seqüência" para outro polinucleotídeo ou polipeptídeo, significando que, quando alinhados, que percentagem de bases ou amino ácidos são as mesmas, e na mesma posição relativa, quando se comparando às duas seqüências. A similaridade da seqüência pode ser determinada em um número de maneiras diferentes. Para determinar identidade de seqüência, as seqüências podem ser alinhadas usando-se os métodos e programas de computador, incluindo BLAST, disponível na web mundial em ncbi.nlm.nih.gov/BLAST. Ver, por exemplo, Altschul *et al.* (1990), *J. Mol. Biol.* 215:403-10. Outro algoritmo de alinhamento é FASTA, disponível em Genetics Computing Group (GCG) package, de Madison, Wisconsin, USA, uma subsidiária totalmente cessionária de Oxford Molecular Group, Inc. Outras técnicas para alinhamento são descritas em *Methods in Enzymology*, vol. 266: Computer Methods for Macromolecular Seqüência Analysis (1996), ed.

- Doolittle, Academic Press, Inc., uma divisão de Harcourt Brace & Co., San Diego, California, USA. De interesse particular são programas de alinhamento que permitem folgas na seqüência. O Smith-Waterman é um tipo de algoritmo que permite folgas nos alinhamentos de seqüência. Ver *Meth. Mol. Biol.* 70: 173-187 (1997). Também, o programa GAP que usa o método de alinhamento Needleman e Wunsch pode ser utilizado para alinhar seqüências. Ver *J. Mol. Biol.* 48: 443-453 (1970).

Um ácido nucléico é "hibridizável" em outro ácido nucléico, tais como cDNA, DNA genômico, ou RNA, quando uma forma trançada simples do ácido nucléico pode se fortalecer ao outro ácido nucléico sob as condições apropriadas de temperatura e resistência iônica de solução. As condições de hibridização e lavagem são bem conhecidas e exemplificadas em Sambrook, J., Fritsch, E. F. e Maniatis, T. *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularmente Capítulo 11 e Tabela 11.1 nesta; e Sambrook, J. and Russell, W., *Molecular Cloning: A Laboratory Manual*, Third Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001). As condições de temperatura e resistência iônica determinam a "estringência" da hibridização. As condições de estringência podem ser ajustadas para classificar fragmentos moderadamente similares, tais como seqüências homólogas de organismos distantemente relacionados, para fragmentos altamente similares, tais como genes que duplicam enzimas funcionais de organismos proximamente relacionados. As condições de hibridização e lavagens de pós-hibridização são úteis para obter as condições de estringência determinadas desejadas da hibridização. Um conjunto de lavagens de pós-hibridização ilustrativo é uma série de lavagens começando com 6 x SSC (onde SSC é 0,15 M de NaCl e 15 mM de tampão citrato), 0,5% de SDS à temperatura ambiente por 15 minutos, em seguida repetido com 2 x SSC, 0,5% de SDS a 45°C por 30 minutos, e, em seguida, repetido duas vezes com 0,2 x SSC, 0,5% de SDS a 50°C por 30 minutos. Outras condições estringentes são obtidas pelo uso de temperaturas mais altas nas quais as lavagens são idênticas àquelas acima, exceto para a temperatura das duas

lavagens finais de 30 minutos em 0,2 x de SSC, 0,5% de SDS, que é aumentada para 60°C. Outro conjunto de condições altamente estridentes usa duas lavagens finais em 0,1 x de SSC, 0,1% de SDS a 65°C. Outro exemplo de condições de hibridização estridentes é hibridização a 50°C, ou mais alta, e 0,1×SSC (15 mM de cloreto de sódio/1,5 mM de citrato de sódio). Outro exemplo de condições de hibridização estridentes é incubação durante toda a noite a 42°C em uma solução: 50% de formamida, 5 x SSC (150 mM de NaCl, 15 mM de citrato de trisódio), 50 mM de fosfato de sódio (pH 7,6), solução 5 x Denhardt, 10% de dextran sulfato, e 20 µg/ml de DNA de esperma de salmão cisalhado desnaturado, seguido por lavagem dos filtros em 0,1 x SSC a cerca de 65°C. Condições de hibridização estridentes e condições de lavagem de pós-hibridização são condições de hibridização e condições de lavagem de pós-hibridização que são pelo menos tão estridentes quanto às condições representativas acima.

A hibridização requer que os dois ácidos nucléicos contenham seqüências complementares, embora, dependendo da estringência da hibridização, não-equiparações entre bases sejam possíveis. A estringência apropriada para hibridizar ácidos nucléicos depende do comprimento dos ácidos nucléicos e do grau de complementação variáveis bem conhecidos na técnica. Quanto maior o grau de similaridade ou homologia entre duas seqüências de nucleotídeos, maior o valor da temperatura de fusão (T_m) para híbridos de ácidos nucléicos tendo aquelas seqüências. A estabilidade relativa (correspondente a T_m mais alta) de hibridizações de ácido nucléico aumenta na seguinte ordem: RNA:RNA, DNA:RNA, DNA:DNA. Para híbridos de mais do que 100 nucleotídeos de comprimento, equações para cálculo da T_m têm sido derivadas (ver Sambrook et al., supra, 9.50-9.51). Para hibridizações com ácidos nucléicos mais curtos, isto é, oligonucleotídeos, a posição de união inadequada torna-se mais importante, e o comprimento do oligonucleotídeo determina sua especificidade (ver Sambrook et al., supra, 11.7-11.8). Tipicamente, o comprimento para um ácido nucléico hibridizável é pelo menos cerca de 10 nucleotídeos. Os comprimentos mínimos ilustrativos para um ácido nucléico hibridizável são: pelo menos cerca de 15 nucleo-

tídeos; pelo menos cerca de 20 nucleotídeos; e pelo menos cerca de 30 nucleotídeos. Além disso, o técnico no assunto reconhecerá que a temperatura e concentração do sal da solução de lavagem podem ser ajustadas conforme necessário de acordo com fatores tais como comprimento da sonda.

5 Antes da presente invenção ser adicionalmente descrita, é para ser compreendido que esta invenção não está limitada às concretizações particulares descritas; como tal podem, naturalmente, variar. É também para ser compreendido que a terminologia aqui usada é para a proposta de descrever concretizações particulares somente, e não é pretendida para ser limitada, visto que o escopo da presente invenção estará limitado somente pelas
10 reivindicações em anexo.

 Onde uma faixa de valores é provida, é compreendido que cada valor interposto, ao décimo da unidade do limite inferior, a menos que o contexto cite claramente de outro modo, entre o limite superior e inferior daquela faixa e qualquer outro valor citado ou interposto naquela faixa citada,
15 está envolvido dentro da invenção. Os limites superior e inferior destas faixas menores podem independentemente estarem incluídos nas faixas menores, e são também envolvidos dentro da invenção, submetido a qualquer limite especificamente excluído na faixa citada. Onde a faixa citada inclui um
20 ou ambos dos limites, as faixas que excluem qualquer ou ambos daqueles limites incluídos estão também incluídas na invenção.

 A menos que de outro modo definido, todos os termos técnicos e científicos aqui usados têm o mesmo significado conforme comumente compreendidos por um técnico no assunto ao qual esta invenção pertence.
25 Embora quaisquer métodos e materiais similares ou equivalentes àqueles aqui descritos possam também serem usados na prática ou teste da presente invenção, os métodos e materiais preferidos são agora descritos. Todas as publicações aqui mencionadas são incorporadas aqui por referência para revelar e descrever os métodos e/ou materiais em conjunto com os quais as
30 publicações são citadas.

 Deve ser notado que conforme aqui usado e nas reivindicações em anexo, as formas singulares "um," "e," e "o" incluem referentes plu-

rais, a menos que o contexto cite claramente de outro modo. Desse modo, por exemplo, referência a "uma enzima de citocromo P450" inclui uma pluralidade de tais enzimas, e referência a "o citocromo P450 reductase" inclui referência a um ou mais citocromos P450 reductase e equivalentes destes conhecidos àquele técnico no assunto, e assim por diante. É adicionalmente notado que as reivindicações podem ser esboçadas para excluir qualquer elemento opcional. Como tal, esta declaração é pretendida para servir como base antecedente para uso de tal terminologia exclusiva como "somente," "apenas" e similares, em conjunto com a recitação de elementos de reivindicação, ou uso de uma limitação "negativa".

As publicações aqui discutidas são providas somente para sua revelação antes da data de depósito do presente pedido. Nada aqui é para ser construído como uma admissão que a presente invenção não é intitulada para antecipar tal publicação em virtude de invenção anterior. Adicionalmente, as datas de publicação providas podem ser diferentes das datas de publicação atuais que podem necessitarem de serem independentemente confirmadas

DESCRIÇÃO DETALHADA DA INVENÇÃO

A presente invenção proporciona ácidos nucleicos compreendendo seqüências de nucleotídeos que codificam enzima modificada de citocromo P450s; bem como vetores recombinantes e células hospedeiras compreendendo os ácidos nucleicos. A presente invenção adicionalmente proporciona métodos de produção de um composto funcionalizado em uma célula hospedeira geneticamente modificada com um ácido nucleico compreendendo seqüência de nucleotídeos que codificam uma enzima modificada de citocromo P450.

A presente invenção adicionalmente proporciona ácidos nucleicos compreendendo seqüências de nucleotídeos que codificam enzimas de modificação de precursor de isoprenóide; bem como vetores recombinantes e células hospedeiras compreendendo os ácidos nucleicos. A presente invenção proporciona métodos de produção de uma enzima de modificação de precursor de isoprenóide enzimaticamente ativo em uma célula hospedei-

ra. A presente invenção adicionalmente proporciona métodos de produção de um composto de isoprenóide em uma célula hospedeira geneticamente modificada com um ácido nucléico compreendendo seqüência de nucleotídeos que codificam uma enzima de modificação de precursor de isoprenóide.

ÁCIDOS NUCLÊICOS, VETORES E CÉLULAS HOSPEDEIRAS

A presente invenção proporciona ácidos nucléicos compreendendo seqüências de nucleotídeo que codificam enzima modificada de citocromo P450s; bem como vetores recombinantes e células hospedeiras compreendendo os ácidos nucléicos. A presente invenção proporciona ácidos nucléicos compreendendo seqüências de nucleotídeo que codificam enzima de modificação de precursor de isoprenóides; bem como vetores recombinantes e células hospedeiras compreendendo os ácidos nucléicos.

O termo "enzima modificada de citocromo P450", conforme aqui usado, se refere a uma enzima que modifica (por exemplo, "funcionaliza") um intermediário em uma trajetória biosintética. A enzima modificada de citocromo P450 codificada por um ácido nucléico objeto catalisa uma ou mais das seguintes reações: hidroxilação, oxidação, epoxidação, desidratação, dehidrogenação, dehalogenação, isomerização, oxidação de álcool, oxidação de aldeído, dealquilação, e clivagem de ligação C-C. Tais reações são aqui referidas genericamente como "modificações intermediárias de trajetória biosintética." Estas reações foram descritas em, por exemplo, Sono et al. ((1996) *Chem. Rev.* 96:2841-2887; ver, por exemplo, Figura 3, de Sono et al., para uma representação esquemática de tais reações).

Em algumas concretizações, uma enzima modificada de citocromo P450 é uma enzima de modificação de precursor de isoprenóide. O termo "enzima de modificação de precursor de isoprenóide," usado intercambieavelmente aqui com "enzima de modificação de isoprenóide," se refere a uma enzima que modifica um composto de precursor de isoprenóide, por exemplo, com um composto de precursor de isoprenóide como substrato, a enzima de modificação de precursor de isoprenóide catalisa uma ou mais das seguintes reações: hidroxilação, epoxidação, oxidação, desidrata-

ção, dehidrogenação, dehalogenação, isomerização, oxidação de álcool, oxidação de aldeído, dealquilação, e clivagem de ligação de C-C. Tais reações são referidas genericamente aqui como "modificações de precursor de isoprenóide." Estas reações foram descritas em, por exemplo, Sono et al. ((1996) *supra*; ver, por exemplo, Figura 3 de Sono et al. para uma representação esquemática de tais reações). As enzimas de modificação de precursor de isoprenóide são, em muitas concretizações, enzimas de citocromo P450s. Ver, por exemplo, Sono et al. (1996) *supra*.

Substratos de uma enzima modificada de citocromo P450

Conforme notado acima, um substrato de uma enzima modificada de citocromo P450 é um intermediário em uma trajetória biosintética. Intermediários exemplares incluem, mas não estão limitados a, precursores de isoprenóide; precursores de alcalóide; precursores de fenilpropanóide; precursores de flavonóide; precursores de esteróide; precursores de policetide; precursores de macrolida; precursores de álcool de açúcar; precursores de composto fenólico; e similares. Ver, por exemplo, Hwang et al. ((2003) *Appl. Environ. Microbiol.* 69:2699-2706; Facchini et al. ((2004) *TRENDS Plant Sci.* 9:116.

Os produtos de trajetória biosintética de interesse incluem, mas não estão limitados a, compostos de isoprenóides, compostos de alcalóide, compostos de fenilpropanóide, compostos de flavonóide, compostos de esteróide, compostos de policetide, compostos de macrolida, compostos de álcool de açúcar, compostos fenólicos, e similares.

Os compostos de alcalóide são um grupo diverso grande de produtos naturais encontrados em cerca de 20% de espécies de planta. Eles são geralmente definidos pela ocorrência de um átomo de nitrogênio em um estado oxidante dentro de um anel heterocíclico. Os compostos de alcanóide incluem compostos benzilisoquinolina alcanóide, compostos indole alcanóide, compostos isoquinolina alcalóide, e similares. Os compostos de alcalóide incluem compostos monocíclicos de alcalóide, compostos dicíclicos de alcalóide, compostos tricíclicos de alcalóide, compostos tetracíclicos de alcalóide, bem como compostos de alcalóide com estruturas de gaiola. Os compos-

tos de alcalóide incluem: 1) Grupo de piridina: piperina, coniina, trigonelina, arecaidina, guvacina, pilocarpina, citisina, esparteína, peletierina; 2) Grupo de pirrolidina: higrina, nicotina, cuscohigrina; 3) Grupo de tropina: atropina, cocaína, ecgonina, peletierina, escopolamina; 4) Grupo de quinolina: quinina, dihidroquinina, quinidina, dihidroquinidina, estriquinina, brucina, e os alcalóides veratrum (por exemplo, veratrina, cevadina); 5) Grupo de isoquinolina: morfina, codeína, tebaína, papaverina, narcotina, narceína, hidrastina, e berberina; 6) Grupo de fenetilamina: metanfetamina, mescalina, efedrina; 7) Grupo de indole: triptaminas (por exemplo, dimetiltryptamina, psilocibin, serotonina), ergolinas (por exemplo, ergina, ergotamina, ácido lisérgico, etc.), e beta-carbolinas (por exemplo, harmina, iohimbina, reserpina, emetina); 8) Grupo de purina: xantinas (por exemplo, cafeína, teobromina, teofilina); 9) Grupo de terpenóide: alcalóides de aconita (por exemplo, aconitina), e esteróides (por exemplo, solanina, samandarin); 10) Grupo de betaína: (compostos de amônia quaternária: por exemplo, muscarina, colina, neurina); e 11) Grupo de pirazole: pirazole, fomepizole. Compostos de alcalóide exemplares são morfina, berberina, vinblastina, vincristina, cocaína, escopolamina, cafeína, nicotina, atropina, papaverina, emetina, quinina, reserpina, codeína, serotonina, etc. Ver, por exemplo, Facchini et al. ((2004) *Trends Plant Science* 9:116).

Substratos de enzima de modificação de isoprenóides

O termo "composto de precursor de isoprenóide" é usado intercambiavelmente com "substrato de precursor de isoprenóide" para se referir a um composto que é um produto da reação de um terpeno sintase em um poliprenil difosfato. O produto de ação de uma reação de terpeno sintase (também referido como um "terpeno ciclase") é o assim denominado "esqueleto de terpeno". Em algumas concretizações, a enzima de modificação de isoprenóide catalisa a modificação de um esqueleto de terpeno, ou um produto a jusante deste. Desse modo, em algumas concretizações, o precursor de isoprenóide é um esqueleto de terpeno. Os substratos de precursor de isoprenóide de uma enzima de modificação de precursor de isoprenóide incluem monoterpenos, diterpenos, triterpenos, e sesquiterpenos.

Os substratos de monoterpeno de uma enzima de modificação de isoprenóide codificada por um ácido nucléico objeto incluem, mas não estão limitados a, qualquer substrato de monoterpeno que produz um produto de oxidação que é um monocomposto de terpeno, ou é um intermediário em uma trajetória biosintética que ocorre em um monocomposto de terpeno. Substratos de monoterpeno exemplares incluem, mas não estão limitados a, substratos de monoterpeno que caem em qualquer das seguintes famílias: monoterpenos acíclicos, Dimetiloctanos, Mentanos, Monoterpenóides Irregulares, Cineóis, Canfanos, Isocanfanes, Monoterpenos monocíclicos, Pina-
 10 nos, Fenchanos, Tujanos, Caranos, Iononos, Iridanos, e Canabanóides. Substratos de monoterpeno exemplares, intermediários, e produtos incluem, mas não estão limitados a, limoneno, citranelol, geraniol, mentol, álcool perillil, linalool, e tujono.

Os substratos de diterpeno de uma enzima de modificação de isoprenóide codificada por um ácido nucléico objeto incluem, mas não estão
 15 limitados a, qualquer substrato de diterpeno que produz um produto de oxidação que é um dicomposto de terpeno, ou é um intermediário em uma trajetória biosintética que ocorre em um dicomposto de terpeno. Substratos de diterpeno exemplares incluem, mas não estão limitados a, substratos de di-
 20 terpeno que caem em qualquer das seguintes famílias: Diterpenóides Acíclicos, Diterpenóides Bicíclicos, Diterpenóides Monocíclicos, Labdanos, Clerodanos, Taxanos, Diterpenóides Tricíclicos, Diterpenóides Tetracíclicos, Kauranos, Beyeranos, Atiseranos, Aphidicolins, Grayanotoxins, Gibberellins, Diterpenos Macrocíclicos, e Elizabethanianos. Substratos de diterpeno exem-
 25 plares, intermediários, e produtos incluem, mas não estão limitados a, casbeno, eleuterobin, paclitaxel, prostratin, e pseudopterosin.

Os substratos de triterpeno de uma enzima de modificação de isoprenóide codificada por um ácido nucléico objeto incluem, mas não estão
 30 limitados a, qualquer substrato de triterpeno que produz um produto de oxidação que é um tricomposto de terpeno, ou é um intermediário em uma trajetória biosintética que ocorre em um tricomposto de terpeno. Substratos de triterpeno exemplares, intermediários, e produtos incluem, mas não estão

limitados a, arbrusideE, bruceantin, testosterona, progesterona, cortisona, e digitoxin.

Os substratos de sesquiterpeno de uma enzima de modificação de isoprenóide codificada por um ácido nucléico objeto incluem, mas não estão limitados a, qualquer substrato de sesquiterpeno que produz um produto de oxidação que é um sesquicomposto de terpeno, ou é um intermediário em uma trajetória biosintética que ocorre em um sesquicomposto de terpeno. Substratos de sesquiterpeno exemplares incluem, mas não estão limitados a, substratos de sesquiterpeno que caem em qualquer das seguintes famílias: Farnesanos, Monociclofarnesanos, Sesquiterpenos monocíclicos, Sesquiterpenos bicíclicos, Biciclofarnesanos, Bisbolanos, Santalanos, Cupranos, Herbertanos, Gimnomitranos, Tricotecanos, Chamigranos, Carotanos, Acoranos, Antisatins, Cadinanos, Oplopananos, Copaanos, Picrotoxanos, Himachalanos, Longipinanos, Longiciclanos, Cariofilanos, Modelfanos, Sifiperfolanos, Humulanos, Intergrifolianos, Lipifolianos, Protoiludanos, Iludanos, Hirsutanos, Lactaranos, Esterpuranos, Fomanosanos, Marasmanos, Germacranos, Elemanos, Eudesmanos, Bakkanos, Quilosifanos, Guaianos, Pseudoguaianos, Sesquiterpenos tricíclicos, Patchoulanos, Trixanos, Aromadendranos, Gorgonanos, Nardosinanos, Brasilanos, Pinguisanos, Sesquipinanos, Sesquicamfanos, Tujopsanos, Bicilcohumulanos, Alliicanos, Esterpuranos, Lactaranos, Africanos, Integrifolianos, Protoiludanos, Aristolanos, e Neolemnanos. Substratos de sesquiterpeno exemplares incluem, mas não estão limitados a, amorfadieno, aloisolongifoleno, (-)- α -trans-bergamoteno, (-)- β -elemeno, (+)-germacreno A, germacreno B, (+)- γ -gurjuneno, (+)-ledeno, neointermoedeol, (+)- β -selineno, e (+)-valenceno.

Modificações

Um ácido nucléico objeto compreende uma seqüência de nucleotídeo que codifica a enzima modificada de citocromo P450, onde a enzima modificada de citocromo P450 codificada por um ácido nucléico objeto terá, em muitas concretizações, uma seqüência de amino ácido não-nativa (tipo não-selvagem, ou que ocorre não-naturalmente, ou variante). A enzima modificada de citocromo P450 codificada terá uma ou mais modificações de

seqüência de amino ácido (anulações, adições, inserções, substituições) que aumentam o nível de atividade da enzima modificada de citocromo P450 em uma célula hospedeira geneticamente modificada com um ácido nucléico objeto, e/ou que aumentam o nível de um dado produto de uma trajetória biosintética produzida por uma célula hospedeira geneticamente modificada com um ácido nucléico objeto.

Em algumas concretizações, um ácido nucléico objeto compreende uma seqüência de nucleotídeo que codifica uma enzima modificada de modificação de precursor de isoprenóide, onde a enzima de modificação de precursor de isoprenóide codificada por um ácido nucléico objeto terá, em muitas concretizações, uma seqüência de amino ácido não-nativa (tipo não-selvagem, ou que ocorre naturalmente, ou variante). A enzima de modificação de precursor de isoprenóide codificada terá uma ou mais modificações de seqüência de amino ácido (anulações, adições, inserções, substituições) que aumentam o nível de atividade da enzima de modificação de precursor de isoprenóide em uma célula hospedeira geneticamente modificada com um ácido nucléico objeto, e/ou que aumentam o nível de um dado composto de isoprenóide produzido por uma célula hospedeira geneticamente modificada com um ácido nucléico objeto. A enzima codificada de modificação de precursor de isoprenóide incluirá, em algumas concretizações, uma ou mais das seguintes modificações em relação a uma enzima tipo selvagem de modificação de precursor de isoprenóide: a) substituição de um domínio nativo de transmembrana com um domínio não-nativo de transmembrana; b) substituição do domínio nativo de transmembrana com um domínio de sinal de secreção; c) substituição do domínio nativo de transmembrana com um domínio de solubilização; d) substituição do domínio nativo de transmembrana com domínio de inserção de membrana; e) mutilação do domínio nativo de transmembrana; e f) uma mudança na seqüência de amino ácido do domínio nativo de transmembrana.

Em muitas concretizações, um ácido nucléico objeto compreende, na ordem de 5' ao 3', e em ligação operável, uma seqüência de nucleotídeo que codifica um primeiro domínio selecionado de um domínio de

transmembrana, um domínio de secreção, um domínio de solubilização, e uma proteína de inserção de membrana; e uma sequência de nucleotídeo que codifica o domínio catalítico de uma enzima P450 modificada (por exemplo, uma enzima de modificação de precursor de isoprenóide), onde o primeiro domínio é heterólogo ao domínio catalítico. Em algumas concretizações, o primeiro domínio compreende ambos um sinal de secreção e um domínio de transmembrana.

Domínio não-nativo de transmembrana

Em algumas concretizações, a enzima modificada de citocromo P450 codificada (por exemplo, uma enzima de modificação de precursor de isoprenóide) compreenderá um domínio de transmembrana não-nativo (por exemplo, um heterólogo). Domínios de transmembrana não-nativos adequados geralmente serão selecionados de domínios de transmembrana que são funcionais em uma dada célula hospedeira. Em algumas concretizações, o domínio de transmembrana não-nativo é um que é funcional em uma célula hospedeira procariótica. Em outras concretizações, o domínio de transmembrana não-nativo é um que é funcional em uma célula hospedeira eucariótica.

Por exemplo, para expressão em *E. coli*, um domínio de transmembrana não-nativo compreenderá, em muitas concretizações, uma das seguintes sequências de amino ácido:

NH₂-MWLLLI~~A~~VFL~~L~~T~~L~~AYLFWP-COOH (SEQ ID NO:1);

NH₂-MALLLAVFLGLSCLLLLSLW-COOH (SEQ ID NO:2);

NH₂-MAILAAIFALVVATATRV-COOH (SEQ ID NO:3);

NH₂-MDASLLLSVALAVVLIPLSLALLN-COOH (SEQ ID NO:4);

e

NH₂-MIEQLLEYWYVVVPVLYIIKQLLAYTK-COOH (SEQ ID NO:5).

Sinal de secreção

Em algumas concretizações, a enzima modificada de citocromo P450 codificada (por exemplo, uma enzima de modificação de precursor de isoprenóide), compreenderá uma sequência de amino ácido não-nativa

que proporciona secreção da proteína de fusão a partir da célula. Aqueles técnicos no assunto estão cientes de seqüências de sinal de secreção. Os sinais de secreção que são adequados para uso na bactéria incluem, mas não estão limitados a, o sinal de secreção de lipoproteína de Braun de *E. coli*, *S. marcescens*, *E. amylosora*, *M. morganii*, e *P. mirabilis*, a proteína TraT de *E. coli* e *Salmonella*; a proteína penicillinase (PenP) de *B. licheniformis* e *B. cereus* e *S. aureus*; proteínas pullulanase de *Klebsiella pneumoniae* e *Klebsiella aerogenes*; lipoproteínas de *E. coli* 1pp-28, Pal, Rp1A, Rp1B, OsmB, NlpB, e Orl17; proteína chitobiase de *V. harseyi*; a proteína β -1,4-endoglucanase de *Pseudomonas solanacearum*, as proteínas Pal e Pcp de *H. influenzae*; a proteína OprI de *P. aeruginosa*; as proteínas MalX e AmiA de *S. pneumoniae*; o antígeno 34 kDa e proteína pTpmA de *Treponema pallidum*; a proteína P37 de *Mycoplasma hyorhinis*; a protease neutra de *Bacillus amyloliquefaciens*; a antígeno 17 kDa de *Rickettsia rickettsii*; a proteína de ligação de maltose malE; a proteína de ligação de ribose rbsB; fosfatase alcalina phoA; e o sinal de secreção OmpA (ver, por exemplo, Tanji et al. (1991) *J. Bacteriol.* 173(6):1997-2005). As seqüências de sinal de secreção adequadas para uso em levedura são conhecidas na técnica, e podem ser usadas. Ver, por exemplo, Patente dos Estados Unidos No. 5.712,113. Os sinais de secreção rbsB, malE, e phoA são discutidos em, por exemplo, Collier (1994) *J. Bacteriol.* 176:3013.

Em algumas concretizações, por exemplo, para expressão em uma célula hospedeira procariótica, tais como *E. coli*, um sinal de secreção compreenderá uma das seguintes seqüências de amino ácido:

- 25 NH₂-MKKTAIAIAVALAGFATVAQA-COOH (SEQ ID NO:6);
 NH₂-MKKTAIAIVVALAGFATVAQA-COOH (SEQ ID NO:7);
 NH₂-MKKTALALAVALAGFATVAQA-COOH (SEQ ID NO:8);
 NH₂-MKIKTGARILALSALTMMFSASALA-COOH (SEQ ID
 NO:9);
- 30 NH₂-MNMKKLATLVSAVALSATVSANAMA-COOH (SEQ ID
 NO:10); e
- NH₂-MKQSTIALALLPLLFTPVTKA-COOH (SEQ ID NO:11).

Em algumas concretizações, a enzima modificada de citocromo P450 codificada (por exemplo, uma enzima de modificação de precursor de isoprenóide) compreenderá ambos uma seqüência de sinal de secreção não-nativa e um domínio heterólogo de transmembrana. Qualquer combinação de seqüência de sinal de secreção e domínio heterólogo de transmembrana pode ser usada.

Como um exemplo não-limitativo, o domínio heterólogo compreendendo uma seqüência de sinal de secreção não-nativa e um domínio heterólogo de transmembrana terá, em algumas concretizações, a seguinte seqüência de amino ácido: NH_2 -MKKTAIAIAVALAGFATVAQALLEYWYVVVPVLYIIKQLLAYTK-COOH (SEQ ID NO:12), onde o domínio de transmembrana é sublinhado, e o sinal de secreção é N-terminal ao domínio de transmembrana.

Domínio de solubilização

Em algumas concretizações, a enzima modificada de citocromo P450 codificada (por exemplo, uma enzima de modificação de precursor de isoprenóide) compreenderá um domínio não-nativo que proporciona a solubilização da proteína.

Em algumas concretizações, um domínio de solubilização compreenderá uma ou mais das seguintes seqüências de amino ácido:

NH_2 -EELLKQALQQAQQLLQQAQELAKK-COOH (SEQ ID NO:13); e

NH_2 -MTVHDIIATYFTKWYVIVPLALIAYRVLDYFY-COOH (SEQ ID NO:14);

NH_2 -GLFGAIAGFIEGGWTGMIDGWYGYGGGKK-COOH (SEQ ID NO:15); e

NH_2 -MAKKTSSKG-COOH (SEQ ID NO:16).

Domínio de inserção de membrana

Em algumas concretizações, a enzima modificada de citocromo P450 modificada (por exemplo, uma enzima de modificação de precursor de isoprenóide) compreenderá uma seqüência não-nativa de amino ácido que proporciona inserção em uma membrana. Em algumas concretizações,

a enzima modificada de citocromo P450 codificada é um polipeptídeo de fusão que compreende um co-participante de fusão heterólogo (por exemplo, uma proteína outra do que uma enzima de citocromo P450) fundido na estrutura, ou no terminal amino, ou no terminal carboxil, onde o co-participante de fusão proporciona a inserção da proteína de fusão em uma membrana biológica.

Em algumas concretizações, o co-participante de fusão é uma proteína "mistic", por exemplo, uma proteína compreendendo a sequência de amino ácido representada na Figura 5 (Nº de Acesso ao GenBank AY874162). Uma sequência de nucleotídeo que codifica a proteína "mistic" é também provida sob o Nº de Acesso ao GenBank AY874162. Outros polipeptídeos que proporcionam inserção em uma membrana biológica são conhecidos na técnica, e são discutidos em, por exemplo, PsbW Woolhead et al. (*J. Biol. Chem.* 276 (18): 14607), descrevendo PsbW; e Kuhn (*FEMS Microbiology Reviews* 17 (1992i) 285), descrevendo proteína de pró-camada M12 e proteína de pró-camada Pf3.

Enzimas de citocromo P450

A enzima de modificação de precursor de isoprenóide codificada será, em muitas concretizações, uma enzima de citocromo P450. A enzima de citocromo P450 codificada efetuará uma ou mais das seguintes reações: hidroxilação, epoxidação, oxidação, desidratação, dehidrogenação, dehalogenação, isomerização, oxidação de álcool, oxidação de aldeído, dealkilação, e clivagem de ligação de C-C. Tais reações são referidas genericamente aqui como "modificações intermediárias de trajetória biosintética" ou, em concretizações particulares, "modificações de precursor de isoprenóide." Estas reações foram descritas em, por exemplo, Sono et al. ((1996) *supra*; ver, por exemplo, Figura 3 de Sono et al., para uma representação esquemática de tais reações). Conforme discutido acima, a enzima modificada de citocromo P450 codificada (por exemplo, enzima de modificação de precursor de isoprenóide) será, em muitas concretizações, um citocromo P450 monooxigenase, um citocromo P450 hidroxilase, um citocromo P450 epoxidase, ou um citocromo P450 dehidrogenase. Uma ampla variedade de

citocromos P450 monooxigenases, hidroxilases, epoxidases, e dehidrogenases (genericamente referidos aqui como "enzimas P450") é conhecida na técnica, e a seqüência de amino ácido de qualquer enzima P450 conhecida, ou uma variante desta, pode ser modificada de acordo com a presente invenção.

Fontes adequadas de ácido nucléicos compreendendo uma seqüência de nucleotídeo que codifica uma enzima de citocromo P450 incluem, mas não estão limitadas a, uma célula ou organismo de qualquer dos seis reinos, por exemplo, Bactéria (por exemplo, Eubacteria); Archaeobacteria; Protista; Fungi; Plantae; and Animalia. Fontes adequadas de ácidos nucléicos exógenos incluem membros similares à planta do reino Protista, incluindo, mas não limitados a, alga (por exemplo, alga verde, alga vermelha, glaucófitos, cianobactéria); membros similares a fungo de Protista, por exemplo, moldes de lama, moldes de água, etc.; membros similares a animal de Protista, por exemplo, flageliformes (por exemplo, Euglena), amebóides (por exemplo, ameba), esporozoários (por exemplo, Apicomplexa, Myxozoa, Microsporidia), e ciliados (por exemplo, Paramecium). Fontes adequadas de ácidos nucléicos exógenos incluem membros do reino Fungi, incluindo, mas não limitado a, membros de qualquer da "phyla": Basidiomycota (club fungi; por exemplo, membros de Agaricus, Amanita, Boletus, Cantherellus, etc.); Ascomycota (sac fungi, incluindo, por exemplo, Saccharomyces); Mycophycophyta (lichens); Zygomycota (fungos de conjugação); e Deuteromycota. Fontes adequadas de ácidos nucléicos exógenos incluem membros do reino Plantae, incluindo, mas não limitado a, membros de qualquer das seguintes divisões: Bryophyta (por exemplo, musgos), Anthocerotophyta (por exemplo, ceratófilos), Hepaticophyta (por exemplo, hepáticas), Lycophyta (por exemplo, musgos), Sphenophyta (por exemplo, cavalinhas), Psilophyta (por exemplo, fetos), Ophioglossophyta, Pterophyta (por exemplo, fetos), Cycadophyta, Ginkgophyta, Pinophyta, Gnetophyta, e Magnoliophyta (por exemplo, plantas florescentes. Fontes adequadas de ácidos nucléicos exógenos incluem membros do reino Animalia, incluindo, mas não limitado a, membros de qualquer da seguinte "phyla": Porifera (sponges); Placozoa; Orthonectida

(parasitas de invertebrados marinhos); Rhombozoa; Cnidaria (corais, anêmonas, medusa, gládios de lula marinhos, amor-perfeito marinho, vespas marinhas); Ctenophora (gelatinas de crista); Platyhelminthes (platielmintes); Nemertina (vermes de fita); Ngathostomulida (vermes de mandíbula); Gastrotricha; Rotifera; Priapulida; Kinorhyncha; Loricifera; Acanthocephala; Entoprocta; Nemotoda; Nematomorpha; Cyclophora; Mollusca (moluscos); Sipuncula (bichos de amendoim); Annelida (vermes segmentados); Tardigrada (bichos de concha); Onychophora (vermes de veludo); Arthropoda (incluindo a "subphyla": Chelicerata, Myriapoda, Hexapoda, e Crustacea, onde os Chelicerata incluem, por exemplo, aracnídeos, Merostomata, e Pycnogonida, onde os Myriapoda incluem, por exemplo, Chilopoda (centopéias), Diplopoda (millipedes), Paropoda, e Symphyla, onde os Hexapoda incluem insetos, e onde os Crustáceos incluem camarão, krill, cracas, etc.; Phoronida; Ectoprocta (animais de musgo); Brachiopoda; Echinodermata (por exemplo, estrela-do-mar, margaridas do mar, comátola, ouriços-do-mar, pepinos do mar, ofiuróide, cestas quebradiças, etc.); Chaetognatha (sagitados); Hemichordata (vermes de glândula); e Chordata. Membros adequados de Chordata incluem qualquer membro da seguinte **subphyla**: Urochordata (esguichos marinhos; incluindo Ascidiacea, Thaliacea, e Larvacea); Cephalochordata (lancelets); Myxini (hagfish); e Vertebrata, onde membros de Vertebrata incluem, por exemplo, membros de Petromyzontida (lampreias), Chondrichthyes (peixe cartilaginoso), Actinopterygii (arraia), Actinista (celacantos), Dipnoi (peixe dipnóico), Reptilia (répteis, por exemplo, serpentes, jacarés, crocodilos, lagartos, etc.), Aves (pássaros); e Mamíferos (mamíferos). Plantas adequadas incluem qualquer **nocotiledon** e qualquer **dicotiledon**.

Desse modo, por exemplo, fontes adequadas incluem células de organismos que incluem, mas não estão limitados a, um protozoário, uma planta, um fungo, uma alga, uma levedura, um réptil, um anfíbio, um mamífero, um microorganismo marinho, um invertebrado marinho, um artrópode, um isópode, um inseto, um aracnídeo, um archabactéria, e uma eubactéria.

Fontes procarióticas adequadas incluem bactéria (por exemplo, Eubactéria) e archaeobactéria. Fontes de archaeobactéria adequadas in-

cluem um metanogen, um halófilo extremo, um termófilo extremo, e similares. Fontes de archaeobactéria adequadas incluem, mas não estão limitadas a, qualquer membro dos grupos Crenarchaeota (por exemplo, *Sulfolobus solfataricus*, *Deinococcus mobilis*, *Pyrodictium occultum*, *Thermophilum*
 5 *pendens*, *Thermoproteus tenax*), Euryarchaeota (por exemplo, *Thermococcus celer*, *Methanococcus thermolithotrophicus*, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Methanobacterium formicicum*, *Methanothermobacter*
 10 *thermautotrophicus*, *Archaeoglobus fulgidus*, *Thermoplasma acidophilum*, *Haloferax volcanii*, *Methanosarcina barkeri*, *Methanosaeta concilli*, *Methanospirillum hungatei*, *Methanomicrobium mobile*), e Korarchaeota. Fontes adequadas de eubactéria incluem, mas não estão limitadas a, qualquer membro de Hidrogenobactéria, Thermotogales, Bactéria de não-enxofre verde, Grupo Denococcus, Cianobactéria, Bactéria púrpura, Planctomyces, Spirochetes, Bactéria de enxofre verde, Cytophagas, e bactéria Gram positiva
 15 (por exemplo, *Mycobacterium* sp., *Micrococcus* sp., *Streptomyces* sp., *Lactobacillus* sp., *Helicobacterium* sp., *Clostridium* sp., *Mycoplasma* sp., *Bacillus* sp., etc.).

Em algumas concretizações, um ácido nucléico que codifica enzima P450 será isolado de um tecido tomado a partir de um organismo; de
 20 uma célula particular ou grupo de células isoladas de um organismo; etc. Por exemplo, onde o organismo é uma planta, o ácido nucléico será, em algumas concretizações, isolado a partir de xilem, do phloem, da camada de câmbio, folhas, raízes, etc. Onde o organismo é um animal, o ácido nucléico será, em algumas concretizações, isolado a partir de um tecido particular
 25 (por exemplo, pulmão, fígado, coração, rim, cérebro, baço, pele, tecido fetal, etc.), ou de um tipo de célula particular (por exemplo, células neuronais, células epiteliais, células endoteliais, astrócitos, macrófagos, células gliais, células "islet", linfócitos T, linfócitos B, etc.).

Em algumas concretizações, um ácido nucléico objeto compreende uma sequência de nucleotídeo que codifica uma enzima P450 que
 30 difere de uma sequência de nucleotídeo tipo selvagem ou que ocorre naturalmente que codifica uma enzima P450, por exemplo, um ácido nucléico

objeto compreende uma seqüência de nucleotídeo que codifica uma enzima P450 variante. Em algumas concretizações, uma variante P450 difere em seqüência de amino ácido por um amino ácido, dois amino ácidos, três amino ácidos, quatro amino ácidos, cinco amino ácidos, seis amino ácidos, sete amino ácidos, oito amino ácidos, nove amino ácidos, ou amino ácidos, ou mais, comparado à seqüência de amino ácido de uma enzima P450 de origem que ocorre naturalmente. Em algumas concretizações, uma enzima P450 variante difere em seqüência de amino ácido por de cerca de 10 amino ácidos a cerca de 15 amino ácidos, de cerca de 15 amino ácidos a cerca de 20 amino ácidos, de cerca de 20 amino ácidos a cerca de 25 amino ácidos, de cerca de 25 amino ácidos a cerca de 30 amino ácidos, de cerca de 30 amino ácidos a cerca de 35 amino ácidos, de cerca de 35 amino ácidos a cerca de 40 amino ácidos, de cerca de 40 amino ácidos a cerca de 50 amino ácidos, ou de cerca de 50 amino ácidos a cerca de 60 amino ácidos, ou mais, comparado à seqüência de amino ácido de uma enzima P450 de origem que ocorre naturalmente.

Em muitas concretizações, conforme discutido acima, a enzima modificada de citocromo P450 codificada compreende uma modificação do N-terminal de uma fonte (por exemplo, tipo selvagem, ou que ocorre naturalmente, ou nativa), por exemplo, uma modificação do domínio de transmembrana e/ou seqüência de amino ácidos N-terminal para o domínio de transmembrana. Em algumas concretizações, a enzima modificada de citocromo P450 codificada incluirá adicionalmente uma ou mais modificações de seqüência de amino ácido na porção catalítica da enzima, comparado à seqüência de amino ácido de uma enzima tipo selvagem de citocromo P450.

Um ácido nucléico compreendendo uma seqüência de nucleotídeo que codifica uma enzima P450 variante (por exemplo, modificada) é um ácido nucléico sintético. Em algumas concretizações, um ácido nucléico sintético compreendendo uma seqüência de nucleotídeo que codifica uma enzima P450 variante é um que hibridiza sob condições de hibridização adequadas a um ácido nucléico compreendendo uma seqüência de nucleotídeo que codifica enzima P450 que ocorre naturalmente. Em algumas con-

cretizações, um ácido nucléico sintético compreendendo uma seqüência de nucleotídeo que codifica uma enzima P450 variante é um que hibridiza sob condições de hibridização estritas a um ácido nucléico compreendendo uma seqüência de nucleotídeo que codifica uma enzima P450 que ocorre naturalmente. Em algumas concretizações, um ácido nucléico sintético compreendendo uma seqüência de nucleotídeo que codifica uma enzima P450 variante compreende uma enzima P450 variante que codifica seqüência de nucleotídeo que tem menos do que cerca de 95% de identidade de seqüência de nucleotídeo para uma enzima P450 que ocorre naturalmente que codifica seqüência de nucleotídeo, por exemplo, a enzima P450 variante que codifica seqüência de nucleotídeo tem não mais do que cerca de 90% a cerca de 95%, de cerca de 85% a cerca de 90%, de cerca de 80% a cerca de 85%, de cerca de 75% a cerca de 80%, de cerca de 70% a cerca de 75%, de cerca de 65% a cerca de 70%, de cerca de 60% a cerca de 65%, de cerca de 55% a cerca de 60%, ou de cerca de 50% a cerca de 55% de identidade de seqüência de nucleotídeo para uma enzima P450 que ocorre naturalmente que codifica seqüência de nucleotídeo.

Em algumas concretizações, a seqüência de nucleotídeo que codifica uma enzima P450 variante codifica uma enzima P450 que tem de cerca de 50% a cerca de 55%, de cerca de 55% a cerca de 60%, de cerca de 60% a cerca de 65%, de cerca de 65% a cerca de 70%, de cerca de 70% a cerca de 75%, de cerca de 75% a cerca de 80%, de cerca de 80% a cerca de 85%, de cerca de 85% a cerca de 90%, ou de cerca de 90% a cerca de 95% de identidade de seqüência de amino ácido para a seqüência de amino ácido de uma enzima P450 que ocorre naturalmente. As seqüências de amino ácido de um número de enzimas P450 são conhecidas na técnica.

Enzimas P450 adequadas que podem ser modificadas e codificadas por uma seqüência de nucleotídeo incluída em um ácido nucléico objeto incluem, mas não estão limitadas a: uma limoneno-6-hidroxilase (ver, por exemplo, Figura 6; e N°s de Acesso ao GenBank AY281025 e AF124815); 5-*epi*-aristolocheno dihidroxilase (ver, por exemplo, Figura 7; e N° de Acesso ao GenBank AF368376); δ -cadineno-8-hidroxilase (ver, por

exemplo, Figura 8A; e N° de Acesso ao GenBank AF332974); taxadieno-5 α -hidroxilase (ver, por exemplo, Figuras 9A e 9B; e N°s de Acesso ao GenBank AY289209, AY959320, e AY364469); *ent*-kaureno oxidase (ver, por exemplo, Figura 10; e N° de Acesso ao GenBank AF047719; ver, por exemplo, Helliwell et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:9019-9024).

As Figuras 8B-D representam variantes P450 exemplares. A Figura 8B representa cadinene hidroxilase com um domínio heterólogo de transmembrana; a Figura 8C representa cadinene hidroxilase com um domínio de solubilização; e a Figura 8C representa cadinene hidroxilase com um domínio de secreção e um domínio heterólogo de transmembrana. A Figura 22 representa adicionalmente variantes P450 exemplares, incluindo amorfadieno oxidase com várias seqüências N-terminais.

As enzimas de citocromo P450 de modificação de intermediário de trajetória de alcalóide são conhecidas na técnica. Ver, por exemplo, Facchini et al. (2004) *supra*; Pauli e Kutchan ((1998) *Plant J.* 13:793-801; Collu et al. ((2001) *FEBS Lett.* 508:215-220; Schroder et al. ((1999) *FEBS Lett.* 458:97-102. Ver também Figuras 19A-C.

As enzimas de citocromo P450 de modificação de intermediário de trajetória de fenilpropanóide são conhecidas na técnica. Ver, por exemplo, Mizutani et al. ((1997) *Plant Physiol.* 113:755-763; e Gang et al. ((2002) *Plant Physiol.* 130:1536-1544. Ver também Figuras 20A-C.

As enzimas de citocromo P450 de modificação de intermediário de trajetória de policetide exemplares são representadas nas Figuras 21A e 21B. Ver também Ikeda et al. ((1999) *Proc. Natl. Acad. Sci. USA* 96:9509-9514; e Ward et al. ((2004) *Antimicrob. Agents Chemother.* 48:4703-4712.

A enzima modificada de citocromo P450 codificada (por exemplo, enzima de modificação de precursor de isoprenóide) é enzimaticamente ativa, por exemplo, a enzima modificada de citocromo P450 (por exemplo, enzima de modificação de precursor de isoprenóide) exibe uma ou mais das seguintes atividades: a) modificação de um intermediário de trajetória biosintética por uma ou mais de: oxidação, hidroxilação, epoxidação, desidratação, dehidrogenação, dehalogenação, isomerização, oxidação de álcool, oxida-

ção de aldeído, dealquilação, ou clivagem de ligação de C-C; b) modificação de um precursor de isoprenóide por uma ou mais de: oxidação, hidroxilação, epoxidação, desidratação, dehidrogenação, dehalogenação, isomerização, oxidação de álcool, oxidação de aldeído, dealquilação, ou clivagem de ligação de C-C. Se um ácido nucléico objeto codifica uma enzima enzimaticamente ativa de citocromo P450 é prontamente determinado pela detecção de um produto da reação da enzima P450 em um substrato, e/ou detectando um produto à jusante da reação da enzima P450 em um substrato. Por exemplo, se um ácido nucléico objeto codifica um terpeno oxidase enzimaticamente ativo, ou um terpeno hidroxilase, pode ser prontamente determinado usando ensaios padrões para estas atividades enzimáticas, usando o substrato apropriado. Os produtos da modificação enzimática são geralmente analisados por espectrometria de massa-cromatografia de gás. Por exemplo, se um ácido nucléico objeto codifica um sesquiterpeno oxidase, ou um sesquiterpeno hidroxilase, pode ser prontamente determinado usando ensaios padrões para estas atividades enzimáticas. Ver, por exemplo, Publicação de Patente dos Estados Unidos No. 20050019882.

Em algumas concretizações, uma seqüência de nucleotídeo que codifica uma enzima modificada de citocromo P450 (por exemplo, uma enzima modificada de modificação de precursor de isoprenóide) é modificada para refletir uma preferência de códon para a célula hospedeira particular. Por exemplo, a seqüência de nucleotídeo será, em algumas concretizações, modificada para preferência de códon de levedura. Ver, por exemplo, Benetzen and Hall (1982) *J. Biol. Chem.* 257(6): 3026-3031. Conforme outro exemplo não-limitativo, a seqüência de nucleotídeo será em outras concretizações modificada para preferência de códon de *E. coli*. Ver, por exemplo, Gouy e Gautier (1982) *Acid nucleics Res.* 10(22):7055-7074; Eyre-Walker (1996) *Mol. Biol. Evol.* 13(6):864-872. Ver também Nakamura et al. (2000) *Acid nucleics Res.* 28(1):292. Conforme um exemplo não-limitativo, a Figura 11A representa uma seqüência tipo selvagem de nucleotídeo que codifica cadinene hidroxilase (códon de partida atg mostrado em negrito); e a Figura 11B representa uma variante otimizada de códon da seqüência representada

na Figura 11A, onde os códons são otimizados para expressão em um procaríote, tal como *E. coli*.

Citocromo P450 reductase

NADPH-citocromo P450 oxidoreductase (CPR, EC 1.6.2.4) é o co-participante redox de muitas P450-monooxygenases. Em algumas concretizações, um ácido nucléico objeto adicionalmente compreende uma sequência de nucleotídeo que codifica um citocromo P450 reductase (CPR). Um ácido nucléico objeto compreendendo uma sequência de nucleotídeo que codifica um CPR é referido como "um ácido nucléico de CPR." Um CPR codificado por um ácido nucléico de CPR objeto transfere elétrons de NADPH para citocromo P450. Por exemplo, em algumas concretizações, um CPR codificado por um ácido nucléico de CPR objeto transfere elétrons de NADPH para uma enzima de modificação de isoprenóide, por exemplo, uma sesquiterpeno oxidase, codificada por uma enzima de modificação de isoprenóide objeto que codifica ácido nucléico.

Em algumas concretizações, um ácido nucléico objeto compreende uma sequência de nucleotídeo que codifica ambos uma enzima modificada de citocromo P450 (por exemplo, uma enzima modificada de modificação de precursor de isoprenóide) e um CPR. Em algumas concretizações, um ácido nucléico objeto compreende uma sequência de nucleotídeo que codifica uma proteína de fusão que compreende uma sequência de amino ácido de enzima modificada de citocromo P450 (por exemplo, uma enzima modificada de modificação de precursor de isoprenóide) que exibe atividade de modificação de precursor de isoprenóide, conforme descrito acima, fundida em um polipeptídeo de CPR. Em algumas concretizações, a proteína de fusão codificada é da fórmula $\text{NH}_2\text{-A-X-B-COOH}$, onde A é a enzima modificada de citocromo P450, X é um ligador opcional, e B é o polipeptídeo de CPR. Em algumas concretizações, a proteína de fusão codificada é da fórmula $\text{NH}_2\text{-A-X-B-COOH}$, onde A é o polipeptídeo de CPR, X é um ligador opcional, e B é a enzima modificada de citocromo P450.

O peptídeo ligador pode ter qualquer de uma variedade de seqüências de amino ácido. As proteínas podem ser unidas por um peptídeo

espaçador, geralmente de uma natureza flexível, embora outras ligações químicas não sejam excluídas. O ligador pode ser um ligador clivável. Sequências de ligador adequadas serão geralmente peptídeos de entre cerca de 5 e cerca de 50 amino ácidos de comprimento, ou entre cerca de 6 e cerca de 25 amino ácidos de comprimento. Ligadores de peptídeo com um grau de flexibilidade geralmente serão usados. Os peptídeos de ligação podem ter virtualmente qualquer sequência de amino ácido, tendo-se em mente que os ligadores preferidos terão uma sequência que resulta em um peptídeo geralmente flexível. O uso de amino ácidos pequenos, tais como glicina e alanina, são de uso na criação de um peptídeo flexível. A criação de tais sequências é rotina para aqueles técnicos no assunto. Uma variedade de ligadores diferentes é comercialmente disponível, e é considerada adequada para uso de acordo com a presente invenção.

Os peptídeos ligadores adequados freqüentemente incluem sequências de amino ácido ricas em resíduos de alanina e prolina, que são conhecidas por conceder flexibilidade a uma estrutura de proteína. Ligadores exemplares têm uma combinação de resíduos de glicina, alanina, prolina e metionine, tais como AAAGGM (SEQ ID NO:17); AAAGGMPPAAAGGM (SEQ ID NO:18); AAAGGM (SEQ ID NO:19); e PPAAAGGM (SEQ ID NO:20). Outros peptídeos ligadores adequados incluem IEGR (SEQ ID NO:21); e GGKGGK (SEQ ID NO:22). Contudo, qualquer ligador flexível geralmente entre cerca de 5 e cerca de 50 amino ácidos de comprimento podem ser usados. Ligadores podem ter virtualmente qualquer sequência que resulta em um peptídeo geralmente flexível, incluindo sequências ricas em alanina-prolina do tipo acima exemplificado.

Em algumas concretizações, um ácido nucléico objeto compreende uma sequência de nucleotídeo que codifica um polipeptídeo de CPR que tem pelo menos cerca de 45%, pelo menos cerca de 50%, pelo menos cerca de 55%, pelo menos cerca de 57%, pelo menos cerca de 60%, pelo menos cerca de 65%, pelo menos cerca de 70%, pelo menos cerca de 75%, pelo menos cerca de 80%, pelo menos cerca de 85%, pelo menos cerca de 90%, pelo menos cerca de 95%, pelo menos cerca de 98%, ou pelo

menos cerca de 99% da identidade de seqüência de amino ácido para um polipeptídeo de CPR que ocorre naturalmente conhecido.

Os polipeptídeos de CPR, bem como ácidos nucléicos que codificam os polipeptídeos de CPR, são conhecidas na técnica, e qualquer ácido nucléico que codifica CPR, ou uma variante deste, podem ser usados na presente invenção. Os ácidos nucléicos que codificam CPR adequados incluem ácidos nucléicos que codificam CPR encontrados nas plantas. Os ácidos nucléicos que codificam CPR adequados incluem ácidos nucléicos que codificam CPR encontrados em fungos. Exemplos de ácidos nucléicos que codificam CPR adequados incluem: Nº de Acesso no GenBank AJ303373 (*Triticum aestivum* CPR); Nº de Acesso no GenBank AY959320 (*Taxus chinensis* CPR); Nº de Acesso no GenBank AY532374 (*Ammi majus* CPR); Nº de Acesso no GenBank AG211221 (*Oryza sativa* CPR); e Nº de Acesso no GenBank AF024635 (*Petroselinum crispum* CPR); *Candida tropicalis* citocromo P450 reductase (Nº de Acesso no GenBank M35199); *Arabidopsis thaliana* citocromo P450 reductase ATR1 (Nº de Acesso no X66016); e *Arabidopsis thaliana* citocromo P450 reductase ATR2 (Nº de Acesso no X66017); e putidaredoxin reductase e putidaredoxin (Nº de Acesso no GenBank J05406).

Em algumas concretizações, um ácido nucléico objeto compreende uma seqüência de nucleotídeo que codifica um polipeptídeo de CPR que é específico para uma dada enzima P450. Conforme um exemplo não-limitativo, um ácido nucléico objeto compreende uma seqüência de nucleotídeo que codifica *Taxus cuspidata* CPR (Figura 12A; Nº de Acesso no GenBank AY571340). Conforme outro exemplo não-limitativo, um ácido nucléico objeto compreende uma seqüência de nucleotídeo que codifica *Candida tropicalis* CPR (Figura 12B). Em outras concretizações, um ácido nucléico objeto compreende uma seqüência de nucleotídeo que codifica um polipeptídeo de CPR que pode servir como um co-participante redox para duas ou mais enzimas P450 diferentes. Um tal CPR é representado na Figura 12C (*Arabidopsis thaliana* citocromo P450 reductase ATR1). Outro tal CPR é representado na Figura 12D (*Arabidopsis thaliana* citocromo P450

reductase ATR2). Também adequado é um modificado ou variante ATR2, por exemplo, conforme representado na Figura 12D, cuja variante ATR2 carece de uma seqüência que tem como alvo cloroplasto.

O CPR codificado compreenderá, em algumas concretizações, uma seqüência heteróloga de amino ácido, ou uma seqüência variante de amino ácido (por exemplo, substituições, anulações, inserções, adições). Em algumas concretizações, o CPR codificado incluirá, em algumas concretizações, uma ou mais das seguintes modificações relativas a um CPR tipo selvagem: a) substituição de um domínio nativo de transmembrana com um domínio não-nativo de transmembrana; b) substituição do domínio nativo de transmembrana com um domínio de sinal de secreção; c) substituição do domínio nativo de transmembrana com um domínio de solubilização; d) substituição do domínio de transmembrana nativo com o domínio de inserção de membrana; e) mutilação do domínio nativo de transmembrana; e f) uma mudança na seqüência de amino ácido do domínio nativo de transmembrana.

Em algumas concretizações, uma seqüência de nucleotídeo que codifica um polipeptídeo de CPR é modificada para refletir a preferência de códon para a célula hospedeira particular. Por exemplo, a seqüência de nucleotídeo será, em algumas concretizações, modificada para preferência de códon de levedura. Ver, por exemplo, Bennetzen e Hall (1982) *J. Biol. Chem.* 257(6): 3026-3031. Conforme outro exemplo não-limitativo, a seqüência de nucleotídeo será, em outras concretizações, modificada para preferência de códon de *E. coli*. Ver, por exemplo, Gouy e Gautier (1982) *Nucleic Acids Res.* 10(22):7055-7074; Eyre-Walker (1996) *Mol. Biol. Evol.* 13(6):864-872. Ver também Nakamura et al. (2000) *Nucleic Acids Res.* 28(1):292.

Construções

A presente invenção adicionalmente proporciona vetores recombinantes ("construções") compreendendo um ácido nucléico objeto. Em algumas concretizações, um vetor recombinante objeto proporciona a amplificação de um ácido nucléico objeto. Em algumas concretizações, um vetor

recombinante objeto proporciona produção de uma enzima modificada de citocromo P450 codificada (por exemplo, uma enzima de modificação de isoprenóide), ou um CPR codificado, em uma célula eucariótica, em uma célula procariótica, ou em um sistema de transcrição/translação livre de célula.

- 5 Vetores de expressão adequados incluem, mas não estão limitados a, vetores de baculovírus, vetores bacteriófago, plasmídeos, fagemídeos, cosmídeos, fosmídeos, cromossomos artificiais bacteriais, vetores virais (por exemplo, vetores virais baseados em vírus de vacina, poliovírus, adenovírus, vírus adeno-associado, SV40, vírus simplex da herpes, e similares), cromossomos artificiais baseados em P1, plasmídeos de levedura, cromossomos artificiais de levedura, e quaisquer outros vetores específicos para hospedeiros específicos de interesse (tais como *E. coli*, levedura, e células de planta).
- 10

- Em algumas concretizações, um vetor recombinante objeto compreende um citocromo P450 modificado objeto que codifica ácido nucléico, e um CPR objeto que codifica ácido nucléico. Em algumas destas concretizações, um vetor recombinante objeto é um vetor de expressão que proporciona produção de ambas as enzima modificada codificada de citocromo P450 (por exemplo, enzima modificada de modificação de isoprenóide), e o CPR codificado em uma célula eucariótica, em uma célula procariótica, ou em um sistema de transcrição/translação livre de célula.
- 15
 - 20

- Certos tipos de vetores permitem que os cassetes de expressão da presente invenção sejam amplificados. Outros tipos de vetores são necessários para introdução eficiente de ácido nucléico objeto nas células e sua expressão estável uma vez introduzido. Qualquer vetor capaz de aceitar um ácido nucléico objeto é contemplado como um vetor recombinante adequado para as propostas da invenção. O vetor pode ser de qualquer comprimento circular ou linear de DNA que, ou se integra no genoma do hospedeiro, ou é mantido na forma episomal. Os vetores podem requerer manipulação adicional ou condições particulares para serem eficientemente incorporados em uma célula hospedeira (por exemplo, quaisquer plasmídeos de expressão), ou podem ser parte de um sistema específico de célula auto-integrante, (por exemplo, um vírus recombinante). O vetor é, em algumas
- 25
 - 30

concretizações, funcional em uma célula procariótica, onde tais vetores funcionam para propagar o vetor recombinante e/ou proporcionar expressão de um ácido nucléico objeto. O vetor é, em algumas concretizações, funcional em uma célula eucariótica, onde o vetor será, em muitas concretizações, um
 5 vetor de expressão.

Numerosos vetores de expressão adequados são conhecidos àqueles técnicos no assunto, e podem ser comercialmente disponíveis. Os seguintes vetores são providos por meio de exemplo; para células hospedeiras bacteriais: pBluescript (Stratagene, San Diego, Calif.), vetores pQE (Qia-
 10 gen), plasmídeos pBluescript, vetores pNH, vetores lambda-ZAP (Stratagene); pTrc (Amann et al., Gene, 69:301-315 (1988)); pTrc99a, pKK223-3, pDR540, e pRIT2T (Pharmacia); para célula hospedeiras eucarióticas: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, e pSVLSV40 (Pharmacia). Contudo, qualquer outro plasmídeo ou outro vetor pode ser usado considerando-
 15 se que ele seja compatível com a célula hospedeira. Em concretizações particulares, o vetor de plasmídeo pSP19g10L é usado para expressão em uma célula hospedeira procariótica. Em outras concretizações particulares, o vetor de plasmídeo pCWori é usado para expressão em uma célula hospedeira procariótica. Ver, por exemplo, Barnes ((1996) *Métodos Enzymol.* 272:1-14)
 20 para uma descrição de pSP19g10L e pCWori.

Em muitas concretizações, um ácido nucléico objeto compreende uma seqüência de nucleotídeo que codifica uma enzima de modificação de isoprenóide, onde a enzima de modificação de isoprenóide que codifica seqüência de nucleotídeo é operavelmente ligada um ou mais elementos
 25 de controle transcricional e/ou translacional. Em muitas concretizações, um ácido nucléico objeto compreende uma seqüência de nucleotídeo que codifica um CPR, onde a seqüência de nucleotídeo que codifica CPR é operavelmente ligada a um ou mais elementos de controle transcricional e/ou translacional.

30 Em algumas concretizações, conforme notado acima, um vetor recombinante objeto compreende uma enzima modificada de citocromo P450 objeto que codifica ácido nucléico, e um ácido nucléico objeto que co-

difica CPR. Em algumas destas concretizações, a enzima modificada de citocromo P450 que codifica seqüência de nucleotídeo, e a seqüência de nucleotídeo que codifica PCR, são operavelmente ligadas a elementos de controle transcricional diferentes. Em outras concretizações, a enzima modificada de citocromo P450 que codifica seqüência de nucleotídeo, e a seqüência de nucleotídeo que codifica PCR, são operavelmente ligadas ao(s) elemento(s) de controle transcricional. Em algumas concretizações, a enzima modificada de citocromo P450 que codifica seqüência de nucleotídeo, e a seqüência de nucleotídeo que codifica CPR, são operavelmente ligadas ao mesmo promotor induzível. Em algumas concretizações, a enzima modificada de citocromo P450 que codifica seqüência de nucleotídeo, e a seqüência de nucleotídeo que codifica PCR, são ambas operavelmente ligadas ao mesmo promotor constitutivo.

Promotores adequados para uso em células hospedeiras procarióticas incluem, mas não estão limitados a, um promotor bacteriófago T7 RNA polimerase; um promotor *trp*; um promotor *lac* operon; um promotor híbrido, por exemplo, um promotor híbrido *lac/tac*, um promotor híbrido *tac/trc*, um promotor *trp/lac*, um promotor T7/*lac*; um promotor *trc*; um promotor *tac*, e similares; um promotor *araBAD*; promotores regulados *in vivo*, tais as um promotor *ssaG* ou um promotor relacionado (ver, por exemplo, Publicação de Patente dos Estados Unidos No. 20040131637), um promotor *pagC* (Pulkkinen and Miller, *J. Bacteriol.*, 1991: 173(1): 86-93; Alpuche-Aranda et al., PNAS, 1992; 89(21): 10079-83), um promotor *nirB* (Harborne et al. (1992) *Mol. Micro.* 6:2805-2813), e similares (ver, por exemplo, Duns-
tan et al. (1999) *Infect. Immun.* 67:5133-5141; McKelvie et al. (2004) *Vaccine* 22:3243-3255; e Chatfield et al. (1992) *Biotechnol.* 10:888-892); um promotor sigma70, por exemplo, um promotor de consenso sigma70 (ver, por exemplo, N^os de Acesso no GenBank AX798980, AX798961, e AX798183); um promotor de fase estacionária, por exemplo, um promotor *dps*, um promotor *spv*, e similares; um promotor derivado de patogenicidade island SPI-2 (ver, por exemplo, WO96/17951); um promotor *actA* (ver, por exemplo, Shetron-Rama et al. (2002) *Infect. Immun.* 70:1087-1096); um promotor *rpsM* (ver,

por exemplo, Valdivia e Falkow (1996). *Mol. Microbiol.* 22:367-378); um promotor tet (ver, por exemplo, Hillen, W. e Wissmann, A. (1989) In Saenger, W. and Heinemann, U. (eds), *Topics in Molecular and Structural Biology, Proteína-Nucleic Acid Interaction*. Macmillan, London, UK, Vol. 10, pp. 143-162); um promotor SP6 (ver, por exemplo, Melton et al. (1984) *Nucl. Acids Res.* 12:7035-7056); e similares.

Exemplos não-limitativos de promotores eucarióticos adequados incluem CMV imediato precoce, HSV timidina kinase, e SV40 precoce e posterior, LTRs de retrovírus, e metalotioneína-I de camundongo. Em algumas concretizações, por exemplo, para expressão em uma célula de levedura, um promotor adequado é um promotor constitutivo tais como um promotor ADH1, um promotor PGK1, um promotor ENO, um promotor PYK1 e similares; ou um promotor regulável, tal como um promotor GAL1, um promotor GAL10, um promotor ADH2, um promotor PHO5, um promotor CUP1, um promotor GAL7, um promotor MET25, um promotor MET3, e similares. A seleção do vetor apropriado e promotor está bem dentro do nível do técnico no assunto. O vetor de expressão pode também conter um local de ligação de ribossomo para iniciação de translação e um terminador de transcrição. O vetor de expressão pode também incluir seqüências apropriadas para amplificação de expressão.

Um vetor recombinante objeto conterá, em muitas concretizações, um ou mais genes marcadores selecionáveis para proporcionar um traço fenotípico para seleção de células hospedeiras transformadas. Marcadores selecionáveis adequados incluem, mas não estão limitados a, dihidrofolato reductase, resistência a neomicin para cultura de célula eucariótica; e resistência à tetraciclina ou ampicilina em células hospedeiras procarióticas, tal como *E. coli*.

Geralmente, os vetores de expressão recombinantes incluirão origens de réplica e marcadores selecionáveis, permitindo transformação da célula hospedeira, por exemplo, o gene de resistência a ampicilina de *E. coli*, o gene de *S. cerevisiae* TRP1, etc.; e um promotor derivado de um gene altamente expresso para dirigir transcrição da seqüência de codificação. Tais

promotores podem ser derivados de operons que codificam enzimas glicolíticas, tais como 3-fosfoglicerato kinase (PGK), fator α , fosfatase ácida, ou proteínas de choque térmico, entre outros.

Em muitas concretizações, uma sequência de nucleotídeo que
 5 codifica uma enzima modificada de citocromo P450 (por exemplo, uma enzima de modificação de isoprenóide modificado) é operavelmente ligada a um promotor induzível. Em muitas concretizações, uma sequência de nucleotídeo que codifica um CPR é operavelmente ligada a um promotor induzível. Os promotores induzíveis são bem conhecidos na técnica. Promotores
 10 induzíveis adequados incluem, mas não estão limitados a, o pL de bacteriófago λ ; Plac; Ptrp; Ptac (promotor híbrido Ptrp-lac); um promotor induzível de isopropila-beta-D-tiogalactopiranoside (IPTG), por exemplo, um promotor *lacZ*; um promotor induzível de tetraciclina; um promotor induzível de arabinose, por exemplo, P_{BAD} (ver, por exemplo, Guzman et al. (1995) *J. Bacteriol.* 177:4121-4130); promotor induzível de xilose, por exemplo, Pxyl (ver, por exemplo, Kim et al. (1996) *Gene* 181:71-76); um promotor GAL1; um promotor de triptofan; um promotor lac; um promotor induzível de álcool, por exemplo, promotor induzível de metanol, um promotor induzível de etanol; promotor induzível de rafinose; um promotor induzível de calor, por exemplo, pro-
 20 motor induzível de calor lambda P_L, um promotor controlado por um represor sensível a calor (por exemplo, vetores de expressão baseados em CI857-reprimido lambda; ver, por exemplo, Hoffmann et al. (1999) *FEMS Microbiol Lett.* 177(2):327-34); e similares.

Em levedura, um número de vetores contendo promotores
 25 constitutivos ou induzíveis podem ser usados. Para uma revisão, ver, Current Protocols in Molecular Biology, Vol. 2, 1988, Ed. Ausubel, et al., Greene Publish. Assoc. & Wiley Interscience, Ch. 13; Grant, et al., 1987, Expression and Secretion Vectors for Yeast, in Métodos in Enzymology, Eds. Wu & Grossman, 31987, Acad. Press, N.Y., Vol. 153, pp.516-544; Glover, 1986,
 30 DNA Cloning, Vol. II, IRL Press, Wash., D.C., Ch. 3; and Bitter, 1987, Heterologous Gene Expression in Yeast, Métodos in Enzymology, Eds. Berger & Kimmel, Acad. Press, N.Y., Vol. 152, pp. 673-684; e The Molecular Biology

of the Yeast *Saccharomyces*, 1982, Eds. Strathern et al., Cold Spring Harbor Press, Vols. I and II. Um promotor de levedura constitutivo, tal como ADH ou LEU2 ou um promotor induzível, tal como GAL, podem ser usados (Cloning in Yeast, Ch. 3, R. Rothstein In: DNA Cloning Vol. 11, A Practical Approach, 5 Ed. DM Glover, 1986, IRL Press, Wash., D.C.). Alternativamente, vetores podem ser usados que promovem integração de seqüências de DNA estranhas no cromossomo de levedura.

Em algumas concretizações, um ácido nucléico objeto ou um vetor objeto compreende um promotor ou outro(s) elemento(s) regulatório(s) 10 para expressão em uma célula de planta. Exemplos não-limitativos de promotores constitutivos adequados que são funcionais em uma célula de planta é o promotor de vírus 35S mosaico de couve-flor, um promotor tandem 35S (Kay et al., *Science* 236:1299 (1987)), um promotor de vírus 19S mosaico de couve-flor, um promotor de gene nopalina sintase (Singer et al., *Plant 15 Mol. Biol.* 14:433 (1990); An, *Plant Physiol.* 81:86 (1986), um promotor de gene octopina sintase, e um promotor de ubiquitin. Promotores induzíveis adequados que são funcionais em uma célula de planta incluem, mas não estão limitados a, um promotor de gene de fenilalanina amônia-liase, um promotor de gene de chalcone sintase, um promotor de gene de proteína 20 relacionada à patogênese, um elemento regulatório induzível de cobre (Mett et al., *Proc. Natl. Acad. Sci. USA* 90:4567-4571 (1993); Furst et al., *Cell* 55:705-717 (1988)); elementos regulatórios induzíveis de tetraciclina e clor-tetraciclina (Gatz et al., *Plant J.* 2:397-404 (1992); Röder et al., *Mol. Gen. Genet.* 243:32-38 (1994); Gatz, *Meth. Célula Biol.* 50:411-424 (1995)); ele- 25 mentos regulatórios de ecdisona (Christopherson et al., *Proc. Natl. Acad. Sci. USA* 89:6314-6318 (1992); Kreutzweiser et al., *Ecotoxicol. Environ. Sa-fety* 28:14-24 (1994)); elementos regulatórios induzíveis de choque (Takaha-shi et al., *Plant Physiol.* 99:383-390 (1992); Yabe et al., *Plant Cell Physiol.* 35:1207-1219 (1994); Ueda et al., *Mol. Gen. Genet.* 250:533-539 (1996)); e 30 elementos de lac operon, que são usados em combinação com um repressor de lac constitutivamente expresso para conferir, por exemplo, expressão induzível de IPTG (Wilde et al., *EMBO J.* 11:1251-1259 (1992); um promotor

induzível de nitrato derivado de gene de nitrito de espinafre reductase (Back et al., *Plant Mol. Biol.* 17:9 (1991)); um promotor induzível leve, tal como aquele associado com a subunidade pequena de RuBP carboxilase ou as famílias de gene LHCP (Feinbaum et al., *Mol. Gen. Genet.* 226:449 (1991);

5 Lam e Chua, *Science* 248:471 (1990)); um elemento regulatório responsiva à luz, conforme descrito na Publicação de Patente dos Estados Unidos No. 20040038400; elementos regulatórios induzíveis por ácido salicílico (Uknes et al., *Plant Cell* 5:159-169 (1993); Bi et al., *Plant J.* 8:235-245 (1995)); elementos regulatórios induzíveis por hormônio (Yamaguchi-Shinozaki et al.,

10 *Plant Mol. Biol.* 15:905 (1990); Kares et al., *Plant Mol. Biol.* 15:225 (1990)); e elementos regulatórios induzíveis por hormônio humano, tais como o elemento de resposta de glucocorticóide humano (Schena et al., *Proc. Natl. Acad. Sci. USA* 88:10421 (1991)).

Elementos regulatórios seletivos de tecido também podem ser

15 incluídos em um ácido nucléico objeto, ou em um vetor objeto. Elementos regulatórios seletivos de tecido adequados, que podem ser usados para expressar ectopicamente um ácido nucléico em um tecido simples, ou em um número limitado de tecidos, incluem, mas não estão limitados a, um elemento regulatório seletivo de xilem, um elemento regulatório seletivo de tracheid,

20 um elemento regulatório seletivo de fibra, um elemento regulatório seletivo de tricoma (ver, por exemplo, Wang et al. (2002) *J. Exp. Botany* 53:1891-1897), um elemento regulatório seletivo tricoma glandular, e similares.

Vetores que são adequados para uso em células de planta são conhecidos na técnica, e quaisquer tais vetores podem ser usados para

25 introduzir um ácido nucléico objeto em uma célula hospedeira de planta. Vetores adequados incluem, por exemplo, um Ti plasmídeo de *Agrobacterium tumefaciens*, ou um Ri₁ plasmídeo de *A. rhizogenes*. O Ti ou Ri₁ plasmídeo é transmitido para células de planta na infecção por *Agrobacterium*, e é totalmente integrado no genoma de planta. J. Schell, *Science*, 237:1176-83

30 (1987). Também adequado para uso é um cromossomo artificial de planta, conforme descrito em, por exemplo, Patente dos Estados Unidos No. 6.900.012.

Composições

A presente invenção adicionalmente proporciona composições compreendendo um ácido nucléico objeto. A presente invenção adicionalmente proporciona composições compreendendo um vetor recombinante objeto. As composições compreendendo um ácido nucléico objeto, ou um vetor objeto de expressão, incluirão, em muitas concretizações, um ou mais de: um sal, por exemplo, NaCl, MgCl, KCl, MgSO₄, etc.; um agente de tamponamento, por exemplo, um tampão Tris, N-(2-Hidroxietil)piperazina-N'-(2-ácido etanosulfônico) (HEPES), 2-(N-Morfolino)ácido etanosulfônico (MES), 2-(N-Morfolino)ácido etanosulfônico sal de sódio (MES), 3-(N-Morfolino)ácido propanosulfônico (MOPS), N-tris[Hidroximetil]metil-3-ácido aminopropanosulfônico (TAPS), etc.; um agente de solubilização; um detergente, por exemplo, um detergente não-iônico, tal como Tween-20, etc.; um inibidor de nuclease; e similares. Em algumas concretizações, um ácido nucléico objeto, ou um vetor recombinante objeto é liofilizado.

Células hospedeiras

A presente invenção proporciona células hospedeiras geneticamente modificadas, por exemplo, células hospedeiras que tenham sido geneticamente modificadas com um ácido nucléico objeto, ou um vetor recombinante objeto. Em muitas concretizações, uma célula hospedeira geneticamente modificada é uma célula hospedeira *in vitro*. Em outras concretizações, uma célula hospedeira geneticamente modificada é uma célula hospedeira *in vivo*. Em outras concretizações, uma célula hospedeira geneticamente modificada é parte de um organismo multicelular.

As células hospedeiras são, em muitas concretizações, organismos unicelulares, ou são desenvolvidas em cultura *in vitro* como células simples. Em algumas concretizações, a célula hospedeira é uma célula eucariótica. As células hospedeiras eucarióticas incluem, mas não estão limitados a, células de levedura, células de inseto, células de planta, células fúngicas, e células de alga. As células hospedeiras eucarióticas adequadas incluem, mas não estão limitadas a, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*,

Pichia thermotolerans, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia* sp., *Saccharomyces cerevisiae*, *Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp., *Fusarium gramineum*, *Fusarium venenatum*, *Neurospora crassa*, *Chlamydomonas reinhardtii*, e similares. Em algumas concretizações, a célula hospedeira é uma célula eucariótica outra do que uma célula de planta.

Em outras concretizações, a célula hospedeira é uma célula de planta. As células de planta incluem células de monocotiledons ("monocots") e dicotiledons ("dicots").

Em outras concretizações, a célula hospedeira é uma célula procariótica. As células procarióticas adequadas incluem, mas não estão limitadas a, qualquer de uma variedade de cepas laboratoriais de *Escherichia coli*, *Lactobacillus* sp., *Salmonella* sp., *Shigella* sp., e similares. Ver, por exemplo, Carrier et al. (1992) *J. Immunol.* 148:1176-1181; Patente dos Estados Unidos No. 6.447.784; e Sizemore et al. (1995) *Science* 270:299-302. Exemplos de cepas de *Salmonella* que podem ser empregadas na presente invenção incluem, mas não estão limitadas a, *Salmonella typhi* e *S. typhimurium*. Cepas de *Shigella* adequadas incluem, mas não estão limitadas a, *Shigella flexneri*, *Shigella sonnei*, e *Shigella dysenteriae*. Tipicamente, a cepa laboratorial é uma que é não-patogênica. Exemplos não-limitativos de outras bactérias adequadas incluem, mas não estão limitados a, *Bacillus subtilis*, *Pseudomonas putida*, *Pseudomonas aeruginosa*, *Pseudomonas mevalonii*, *Rhodobacter sphaeroides*, *Rhodobacter capsulatus*, *Rhodospirillum rubrum*, *Rhodococcus* sp., e similares. Em algumas concretizações, a célula hospedeira é *Escherichia coli*.

Para gerar uma célula hospedeira objeto geneticamente modificada, um ácido nucléico objeto compreendendo seqüências de nucleotídeo que codificam uma enzima modificada de citocromo P450 (por exemplo, uma enzima modificada de modificação de isoprenóide) é introduzido estavelmente ou transientemente em uma célula hospedeira de origem, usando técnicas

estabelecidas, incluindo, mas não limitadas a, eletroporação, precipitação de fosfato de cálcio, transfecção mediada por DEAE-dextran, transfecção mediada por lipossoma, e similares. Para transformação estável, um ácido nucleico incluirá geralmente adicionalmente um marcador selecionável, por exemplo, quaisquer de vários marcadores selecionáveis bem conhecidos, tais como resistência a neomicin, resistência a ampicilina, resistência à tetraciclina, resistência a cloranfenicol, resistência a kanamicin, e similares.

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é uma célula de planta. Uma célula de planta objeto geneticamente modificada é útil para produção de um composto selecionado de isoprenóide em cultura de célula de planta *in vitro*. Orientação com relação à cultura de tecido de planta pode ser encontrada em, por exemplo: Plant Célula and Tissue Culture, 1994, Vasil and Thorpe Eds., Kluwer Academic Publishers; e em: Plant Célula Culture Protocols (Métodos in Molecular Biology 111), 1999, Hall Eds, Humana Press.

Células hospedeiras geneticamente modificadas

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada compreende um vetor objeto de expressão, onde o vetor objeto de expressão compreende uma seqüência de nucleotídeo que codifica uma enzima modificada de citocromo P450. Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada compreende um vetor objeto de expressão, onde o vetor objeto de expressão compreende uma seqüência de nucleotídeo que codifica uma enzima modificada de modificação de precursor de isoprenóide.

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada compreende um primeiro vetor objeto de expressão, onde o primeiro vetor objeto de expressão compreende um ácido nucleico objeto compreendendo uma seqüência de nucleotídeo que codifica uma enzima modificada de citocromo P450; e adicionalmente compreende um segundo vetor objeto de expressão, onde o segundo vetor objeto de expressão compreende um ácido nucleico objeto compreendendo uma seqüência de nucleotídeo que codifica um CPR. Em outras concretizações,

uma célula hospedeira objeto geneticamente modificada compreende um vetor objeto de expressão, no qual o vetor objeto de expressão compreende um ácido nucléico objeto compreendendo uma seqüência de nucleotídeo que codifica uma enzima modificada de citocromo P450 e um ácido nucléico objeto compreendendo uma seqüência de nucleotídeo que codifica um CPR. Em outras concretizações, uma célula hospedeira objeto geneticamente modificada compreende um vetor objeto de expressão, onde o vetor objeto de expressão compreende um ácido nucléico objeto compreendendo uma seqüência de nucleotídeo que codifica um polipeptídeo de fusão (por exemplo, um polipeptídeo que inclui uma enzima modificada de citocromo P450 e um CPR).

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada compreende um primeiro vetor de expressão, onde o primeiro vetor de expressão compreende ácido nucléico objeto compreendendo uma seqüência de nucleotídeo que codifica uma enzima modificada de citocromo P450; e adicionalmente compreende um segundo vetor de expressão, onde o segundo vetor de expressão compreende uma seqüência de nucleotídeo que codifica um CPR. Em outras concretizações, uma célula hospedeira objeto geneticamente modificada compreende um vetor objeto de expressão, no qual o vetor objeto de expressão compreende um ácido nucléico objeto compreendendo uma seqüência de nucleotídeo que codifica uma enzima modificada de citocromo P450 e uma seqüência de nucleotídeo que codifica um CPR.

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é adicionalmente geneticamente modificada para incluir um ou mais ácidos nucléicos compreendendo seqüências de nucleotídeo que codificam uma ou mais enzimas que ocorrem em um substrato para uma enzima de citocromo P450. Exemplos de tais enzimas incluem, mas não estão limitados a, terpeno sintases; prenil transferases; isopentenil difosfato isomerase; uma ou mais enzimas em uma trajetória de mevalonato; e uma ou mais enzimas em uma trajetória DXP. Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é adicionalmente

geneticamente modificada para incluir um ou mais ácidos nucléicos compreendendo seqüências de nucleotídeos que codificam um, dois, três, quatro, cinco, seis, sete, ou oito, ou mais de: um terpeno sintase, uma prenil transferase, um IPP isomerase, uma acetoacetyl-CoA tiolase, um HMGS, um HMGR, um MK, um PMK, e um MPD. Em algumas concretizações, por exemplo, onde uma célula hospedeira objeto geneticamente modificada é adicionalmente geneticamente modificada para incluir um ou mais ácidos nucléicos compreendendo seqüências de nucleotídeo que codificam dois ou mais de um terpeno sintase, uma prenil transferase, uma IPP isomerase, uma acetoacetyl-CoA tiolase, um HMGS, um HMGR, um MK, um PMK, e um MPD, as seqüências de nucleotídeo estão presentes em pelo menos dois operons, por exemplo, dois operons separados, três operons separados, ou quatro operons separados.

Terpeno sintases

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é adicionalmente geneticamente modificada para incluir um ácido nucléico compreendendo uma seqüência de nucleotídeo que codifica um terpeno sintase. Em algumas concretizações, o terpeno sintase é um que modifica FPP para gerar um sesquiterpeno. Em outras concretizações, o terpeno sintase é um que modifica GPP para gerar um monoterpeno. Em outras concretizações, o terpeno sintase é um que modifica GGPP para gerar um diterpeno. O terpeno sintase age em um substrato de poliprenil difosfato, modificando o substrato de poliprenil difosfato por ciclização, rearranjo, ou acoplamento do substrato, produzindo um precursor de isoprenóide (por exemplo, limoneno, amorfadieno, taxadieno, etc.), cujo precursor de isoprenóide é o substrato para uma enzima de modificação de precursor de isoprenóide(s). Pela ação do terpeno sintase em um substrato de poliprenil difosfato, o substrato para uma enzima de modificação de precursor de isoprenóide é produzido.

As seqüências de nucleotídeo que codificam terpeno sintases são conhecidas na técnica, e qualquer seqüência de nucleotídeo que codifica terpeno sintase conhecida pode ser usada para modificar geneticamente

uma célula hospedeira. Por exemplo, as seguintes seqüências de nucleotídeo que codificam terpeno sintase, seguidas por seu números de Acesso no GenBank e os organismos nos quais eles foram identificados, são conhecidas e podem ser usadas: (-)-germacreno D sintase mRNA (AY438099; *Populus balsamifera* subsp. *trichocarpa* x *Populus deltoids*); E,E-alpha-farneseno sintase mRNA (AY640154; *Cucumis sativus*); 1,8-cineolo sintase mRNA (AY691947; *Arabidopsis thaliana*); terpeno sintase 5 (TPS5) mRNA (AY518314; *Zea mays*); terpeno sintase 4 (TPS4) mRNA (AY518312; *Zea mays*); mirceno/ocimeno sintase (TPS10) (At2g24210) mRNA (NM_127982; *Arabidopsis thaliana*); geraniol sintase (GES) mRNA (AY362553; *Ocimum basilicum*); pineno sintase mRNA (AY237645; *Picea sitchensis*); mirceno sintase 1e20 mRNA (AY195609; *Antirrhinum majus*); (E)-β-ocimeno sintase (0e23) mRNA (AY195607; *Antirrhinum majus*); E-β-ocimeno sintase mRNA (AY151086; *Antirrhinum majus*); terpeno sintase mRNA (AF497492; *Arabidopsis thaliana*); (-)-camfeno sintase (AG6.5) mRNA (U87910; *Abies grandis*); (-)-gene 4S-limoneno sintase (por exemplo, seqüência genômica) (AF326518; *Abies grandis*); gene delta-selineno sintase (AF326513; *Abies grandis*); 4,11-dieno-amorfo sintase mRNA (AJ251751; *Artemisia annua*); E-α-bisaboleno sintase mRNA (AF006195; *Abies grandis*); gamma-humuleno sintase mRNA (U92267; *Abies grandis*); δ-selineno sintase mRNA (U92266; *Abies grandis*); pineno sintase (AG3.18) mRNA (U87909; *Abies grandis*); mirceno sintase (AG2.2) mRNA (U87908; *Abies grandis*); etc.

Trajetória de mevalonato

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é uma célula hospedeira que não sintetiza normalmente isopentenil pirofosfato (IPP) ou mevalonato, via uma trajetória de mevalonato. A trajetória de mevalonato compreende: (a) condensação de duas moléculas de acetil-CoA em acetoacetil-CoA; (b) condensação de acetoacetil-CoA com acetil-CoA para formar HMG-CoA; (c) conversão de HMG-CoA em mevalonato; (d) fosforilação de mevalonato em mevalonato 5-fosfato; (e) conversão de mevalonato 5-fosfato em mevalonato 5-pirofosfato; e (f) conversão de mevalonato 5-pirofosfato em isopentenil pirofosfato. A tra-

jetória de enzimas mevalonato requerida para produção de IPP varia, dependendo das condições de cultura.

Conforme notado acima, em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é uma célula hospedeira que não sintetiza normalmente isopentenil pirofosfato (IPP) ou mevalonato via uma trajetória de mevalonato. Em algumas destas concretizações, a célula hospedeira é geneticamente modificada com um vetor objeto de expressão compreendendo um ácido nucléico objeto que codifica uma enzima de modificação de isoprenóide; e a célula hospedeira é geneticamente modificada com um ou mais ácidos nucléicos heterólogos compreendendo seqüências de nucleotídeo que codificam acetoacetil-CoA tiolase, hidroximetilglutaril-CoA sintase (HMGS), hidroximetilglutaril-CoA reductase (HMGR), mevalonato kinase (MK), fosfomevalonato kinase (PMK), e mevalonato pirofosfato decarboxilase (MPD) (e opcionalmente também IPP isomerase). Em muitas destas concretizações, a célula hospedeira é geneticamente modificada com um vetor de expressão compreendendo uma seqüência de nucleotídeo que codifica um CPR. Em algumas destas concretizações, a célula hospedeira é geneticamente modificada com um vetor objeto de expressão compreendendo um ácido nucléico objeto que codifica uma enzima de modificação de isoprenóide; e a célula hospedeira é geneticamente modificada com um ou mais ácidos nucléicos heterólogos compreendendo seqüências de nucleotídeo que codificam MK, PMK, MPD (e, opcionalmente, também IPP isomerase). Em muitas destas concretizações, a célula hospedeira é geneticamente modificada com um vetor de expressão compreendendo uma seqüência de nucleotídeo que codifica um CPR.

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é uma célula hospedeira que não sintetiza normalmente IPP ou mevalonato via uma trajetória de mevalonato; a célula hospedeira é geneticamente modificada com um vetor objeto de expressão compreendendo um ácido nucléico objeto que codifica uma enzima de modificação de isoprenóide; e a célula hospedeira é geneticamente modificada com um ou mais ácidos nucléicos heterólogos compreendendo seqüências

de nucleotídeo que codificam acetoacetyl-CoA tiosase, HMGS, HMGR, MK, PMK, MPD, IPP isomerase, e uma prenil transferase. Em muitas destas concretizações, a célula hospedeira é geneticamente modificada com um vetor de expressão compreendendo uma sequência de nucleotídeo que codifica um CPR. Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é uma célula hospedeira que não sintetiza normalmente IPP ou mevalonato via uma trajetória de mevalonato; a célula hospedeira é geneticamente modificada com um vetor objeto de expressão compreendendo um ácido nucléico objeto que codifica uma enzima de modificação de isoprenóide; e a célula hospedeira é geneticamente modificada com um ou mais ácidos nucléicos heterólogos compreendendo seqüências de nucleotídeo que codificam MK, PMK, MPD, IPP isomerase, e uma prenil transferase. Em muitas destas concretizações, a célula hospedeira é geneticamente modificada com um vetor de expressão compreendendo uma sequência de nucleotídeo que codifica um CPR.

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é uma que não sintetiza normalmente IPP ou mevalonato via uma trajetória de mevalonato, por exemplo, a célula hospedeira é uma que compreende uma trajetória de mevalonato endógena. Em algumas destas concretizações, a célula hospedeira é uma célula de levedura. Em algumas destas concretizações, a célula hospedeira é *Saccharomyces cerevisiae*.

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é adicionalmente geneticamente modificada com um ou mais ácidos nucléicos que compreendem as seqüências de nucleotídeo que codificam uma dehidrogenase ou dehidrogenases, cujas dehidrogenase adicionalmente modifica um composto de isoprenóide. A dehidrogenase codificada pode ser uma que é naturalmente encontrada em uma célula procariótica, ou em uma célula eucariótica, ou pode ser uma variante de tal uma dehidrogenase. Em algumas concretizações, a presente invenção proporciona ácidos nucléicos isolados compreendendo seqüências de nucleotídeo que codificam tais dehidrogenases.

Ácidos nucleicos de trajetória de mevalonato

As seqüências de nucleotídeos que codificam produtos de gene de trajetória de MEV são conhecidas na técnica, e qualquer produto de gene de trajetória de MEV conhecido que codifica seqüência de nucleotídeo

5 pode ser usado para gerar uma célula hospedeira objeto geneticamente modificada. Por exemplo, seqüências de nucleotídeo que codificam acetoacetil-CoA tiolase, HMGS, HMGR, MK, PMK, MPD, e IDI são conhecidas na técnica. Os seguintes são exemplos não-limitativos de seqüências de nucleotídeo conhecidas que codificam produtos de gene de trajetória de MEV, com

10 números de acesso no GenBank e organismo seguinte de cada enzima de trajetória de MEV, em parênteses: acetoacetil-CoA tiolase: (NC_000913 REGION: 2324131..2325315; *E. coli*), (D49362; *Paracoccus denitrificans*), e (L20428; *Saccharomyces cerevisiae*); HMGS: (NC_001145. complemento 19061..20536; *Saccharomyces cerevisiae*), (X96617; *Saccharomyces cerevisiae*), (X83882; *Arabidopsis thaliana*), (AB037907; *Kitasatospora griseola*), e (BT007302; *Homo sapiens*); HMGR: (NM_206548; *Drosophila melanogaster*), (NM_204485; *Gallus gallus*), (AB015627; *Streptomyces* sp. KO-3988), (AF542543; *Nicotiana attenuata*), (AB037907; *Kitasatospora griseola*), (AX128213, proporcionando a seqüência que codifica um HMGR truncado;

15 *Saccharomyces cerevisiae*), e (NC_001145: complemento (115734..118898; *Saccharomyces cerevisiae*)); MK: (L77688; *Arabidopsis thaliana*), e (X55875; *Saccharomyces cerevisiae*); PMK: (AF429385; *Hevea brasiliensis*), (NM_006556; *Homo sapiens*), (NC_001145. complemento 712315..713670; *Saccharomyces cerevisiae*); MPD: (X97557; *Saccharomyces cerevisiae*), (AF290095; *Enterococcus faecium*), e (U49260; *Homo sapiens*); and IDI: (NC_000913, 3031087..3031635; *E. coli*), e (AF082326; *Haematococcus pluvialis*).

20

25

Em algumas concretizações, a região de codificação de HMGR codifica uma forma truncada de HMGR ("tHMGR") que carece do domínio de transmembrana de HMGR tipo selvagem. O domínio de transmembrana de HMGR contém as porções regulatórias da enzima, e não tem atividade catalítica.

30

A seqüência de codificação de qualquer enzima de trajetória de MEV conhecida pode ser alterada em vários modos conhecidos na técnica para gerar mudanças alvos na seqüência de amino ácido da enzima codificada. O amino ácido de uma enzima de trajetória de MEV variante será usualmente substancialmente similar à seqüência de amino ácido de qualquer enzima de trajetória de MEV conhecida, isto é, diferirá por pelo menos um amino ácido, e pode diferir por pelo menos dois, pelo menos 5, pelo menos 10, ou pelo menos 20 amino ácidos, mas tipicamente não mais do que cerca de cinquenta amino ácidos. As mudanças de seqüência podem ser substituições, inserções ou anulações. Por exemplo, conforme descrito abaixo, a seqüência de nucleotídeo pode ser alterada para as inclinações de códon de uma célula hospedeira particular. Em adição, uma ou mais diferenças de seqüência de nucleotídeo podem ser introduzidas, que resultam em mudanças de amino ácido conservativas na proteína codificada.

15 Prenil transferases

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é geneticamente modificada para incluir um ácido nucléico compreendendo uma seqüência de nucleotídeo que codifica uma enzima de modificação de isoprenóide; e, em algumas concretizações, é também geneticamente modificada para incluir um ou mais ácidos nucléicos compreendendo uma seqüência de nucleotídeo(s) que codifica uma ou mais enzimas de trajetória de mevalonato, conforme descrito acima; e um ácido nucléico compreendendo uma seqüência de nucleotídeo que codifica uma prenil transferase.

25 As preniltransferases constituem um amplo grupo de enzimas que catalisam a condensação consecutiva de IPP, resultando na formação de prenil difosfatos de vários comprimentos de cadeia. As preniltransferases adequadas incluem enzimas que catalisam a condensação de IPP com substratos primeiros alílicos para formar composto de isoprenóides com de
30 cerca de 2 unidades de isopreno a cerca de 6000 unidades de isopreno ou mais, por exemplo, 2 unidades de isopreno (Geranil pirofosfato sintase), 3 unidades de isopreno (Farnesil pirofosfato sintase), 4 unidades de isopreno

(geranilgeranil pirofosfato sintase), 5 unidades de isopreno, 6 unidades de isopreno (hexadecilpirofosfato sintase), 7 unidades de isopreno, 8 unidades de isopreno (fitoeno sintase, octaprenil pirofosfato sintase), 9 unidades de isopreno (nonaprenil pirofosfato sintase, 10 unidades de isopreno (decaprenyl pirofosfato sintase), de cerca de 10 unidades de isopreno a cerca de 15 unidades de isopreno, de cerca de 15 unidades de isopreno a cerca de 20 unidades de isopreno, de cerca de 20 unidades de isopreno a cerca de 25 unidades de isopreno, de cerca de 25 unidades de isopreno a cerca de 30 unidades de isopreno, de cerca de 30 unidades de isopreno a cerca de 40 unidades de isopreno, de cerca de 40 unidades de isopreno a cerca de 50 unidades de isopreno, de cerca de 50 unidades de isopreno a cerca de 100 unidades de isopreno, de cerca de 100 unidades de isopreno a cerca de 250 unidades de isopreno, de cerca de 250 unidades de isopreno a cerca de 500 unidades de isopreno, de cerca de 500 unidades de isopreno a cerca de 1000 unidades de isopreno, de cerca de 1000 unidades de isopreno a cerca de 2000 unidades de isopreno, de cerca de 2000 unidades de isopreno a cerca de 3000 unidades de isopreno, de cerca de 3000 unidades de isopreno a cerca de 4000 unidades de isopreno, de cerca de 4000 unidades de isopreno a cerca de 5000 unidades de isopreno, ou de cerca de 5000 unidades de isopreno a cerca de 6000 unidades de isopreno.

Preniltransferases adequadas incluem, mas não estão limitadas a, uma *E*-isoprenil difosfato sintase, incluindo, mas não limitada a, geranil difosfato (GPP) sintase, farnesil difosfato (FPP) sintase, geranilgeranil difosfato (GGPP) sintase, hexaprenil difosfato (HexPP) sintase, heptaprenil difosfato (HepPP) sintase, octaprenil (OPP) difosfato sintase, solanesil difosfato (SPP) sintase, decaprenil difosfato (DPP) sintase, chicle sintase, e guttapercha sintase; e uma *Z*-isoprenil difosfato sintase, incluindo, mas não limitada a, nonaprenil difosfato (NPP) sintase, undecaprenil difosfato (UPP) sintase, dehidrodolichil difosfato sintase, eicosaprenil difosfato sintase, sintase de borracha natural, e outras *Z*-isoprenil difosfato sintases.

As seqüências de nucleotídeo de numerosas prenil transferases de uma variedade de espécies são conhecidas, e podem ser usadas ou

modificadas para uso na geração de uma célula hospedeira objeto geneticamente modificada. As seqüências de nucleotídeo que codificam prenil transferases são conhecidas na técnica. Ver, por exemplo, Human farnesil pirofosfato sintetase mRNA (Nº de Acesso no GenBank J05262; *Homo sapiens*); gene de farnesil difosfato sintetase (FPP) (Nº de Acesso no GenBank J05091; *Saccharomyces cerevisiae*); isopentenil gene de difosfato:dimetilalil difosfato isomerase (J05090; *Saccharomyces cerevisiae*); Wang e Ohnuma (2000) *Biochim. Biophys. Acta* 1529:33-48; Patente dos Estados Unidos No. 6.645.747; *Arabidopsis thaliana* farnesil pirofosfato sintetase 2 (FPS2) / FPP sintetase 2 / farnesil difosfato sintase 2 (At4g17190) mRNA (Nº de Acesso no GenBank NM_202836); *Ginkgo biloba* geranilgeranil difosfato sintase (ggpps) mRNA (Nº de Acesso no GenBank AY371321); *Arabidopsis thaliana* geranilgeranil pirofosfato sintase (GGPS1) / GGPP sintetase / farnesiltransferase (At4g36810) mRNA (Nº de Acesso no GenBank NM_119845); gene de *Synechococcus elongatus* para farnesil, geranilgeranil, geranilfarnesil, hexaprenil, heptaprenil difosfato sintase (Self-HepPS) (Nº de Acesso no GenBank No. AB016095); etc.

Uso de códon

Em algumas concretizações, uma seqüência de nucleotídeo usada para gerar uma célula hospedeira objeto geneticamente modificada é tal que a seqüência de nucleotídeo reflete a preferência de códon para a célula hospedeira particular. Por exemplo, a seqüência de nucleotídeo será, em algumas concretizações, modificada para preferência de códon de levedura. Ver, por exemplo, Bennetzen e Hall (1982) *J. Biol. Chem.* 257(6): 3026-3031. Conforme outro exemplo não-limitativo, a seqüência de nucleotídeo será, em outras concretizações, modificada para preferência de códon de *E. coli*. Ver, por exemplo, Gouy e Gautier (1982) *Nucleic Acids Res.* 10(22):7055-7074; Eyre-Walker (1996) *Mol. Biol. Evol.* 13(6):864-872. Ver também Nakamura et al. (2000) *Nucleic Acids Res.* 28(1):292.

Modificações genéticas adicionais

Em algumas concretizações, a célula hospedeira objeto geneticamente modificada é adicionalmente geneticamente modificada para uma

que é geneticamente modificada para incluir um ou mais ácidos nucleicos compreendendo uma sequência de nucleotídeo(s) que codifica uma enzima modificada de citocromo P450 (por exemplo, uma enzima modificada de modificação de isoprenóide); e que é adicionalmente geneticamente modificada para alcançar produção de "heme" aumentada, e/ou alcançar produção aumentada de um intermediário de trajetória biosintética de terpeno, e/ou que é adicionalmente geneticamente modificada tal que um gene de trajetória biosintética de terpeno endógeno é funcionalmente incapacitado. O termo "funcionalmente incapacitado", conforme aqui usado no contexto de um gene de trajetória biosintética de terpeno exógeno, se refere a uma modificação genética de um gene de trajetória biosintética de terpeno, cuja modificação resulta na produção de um produto de gene codificado pelo gene que é produzido em níveis abaixo do normal, e/ou é não-funcional.

Produção aumentada de "heme"

Em algumas concretizações, a célula hospedeira objeto geneticamente modificada compreende uma ou mais modificações genéticas adicionais que proporcionam produção aumentada de "heme", por exemplo, para alcançar um pelo menos cerca de 10%, pelo menos cerca de 15%, pelo menos cerca de 20%, pelo menos cerca de 25%, pelo menos cerca de 30%, pelo menos cerca de 40%, pelo menos cerca de 50%, pelo menos cerca de 60%, pelo menos cerca de 70%, pelo menos cerca de 80%, pelo menos cerca de 90%, pelo menos cerca de 2 vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 10 vezes, pelo menos cerca de 15 vezes, pelo menos cerca de 20 vezes, ou pelo menos cerca de 25 vezes, ou mais, aumento na produção de "heme", comparado a uma célula hospedeira que não compreende as uma ou mais modificações genéticas adicionais.

A etapa limitante na produção de "heme" em uma célula é a biosíntese de ácido aminolevulínico (ALA). Conforme representado na Figura 13, existem duas trajetórias distintas para biosíntese de ALA envolvendo ou uma trajetória C₄ ou uma trajetória C₅. Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é adicionalmente geneti-

camente modificada para sobre-expressar glutamyl-tRNA reductase (GTR reductase). Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é adicionalmente geneticamente modificada para produzir um nível de atividade de GTR reductase que é pelo menos cerca de

5 10%, pelo menos cerca de 15%, pelo menos cerca de 20%, pelo menos cerca de 25%, pelo menos cerca de 30%, pelo menos cerca de 40%, pelo menos cerca de 50%, pelo menos cerca de 60%, pelo menos cerca de 70%, pelo menos cerca de 80%, pelo menos cerca de 90%, pelo menos cerca de 2 vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 5 vezes, pelo

10 menos cerca de 10 vezes, pelo menos cerca de 15 vezes, pelo menos cerca de 20 vezes, ou pelo menos cerca de 25 vezes, ou mais, mais alto do que o nível de atividade de GTR reductase produzida em uma célula hospedeira de controle.

O aumento do nível de atividade de GTR reductase em uma

15 célula é alcançado em um número de modos, incluindo, mas não limitado a :

1) aumento da resistência do promotor do promotor ao qual a região de codificação de GTR reductase é operavelmente ligada; 2) aumento do número de cópia do plasmídeo compreendendo uma seqüência de nucleotídeo que codifica GTR reductase; 3) aumento da estabilidade de um mRNA de GTR

20 reductase (onde um "mRNA de GTR reductase" é um mRNA compreendendo uma seqüência de nucleotídeo que codifica GTR reductase); 4) modificação do uso de códon de GTR reductase, tal que o nível de translação do mRNA de GTR reductase é aumentado; 5) aumento da estabilidade de enzima de GTR reductase; 6) aumento da atividade específica (atividade de

25 unidades por proteína de unidade) de GTR reductase; e 7) redução da regulação de realimentação negativa de GTR reductase.

Em algumas concretizações, uma modificação genética que resulta em nível aumentado de GTR reductase é uma modificação genética que reduz a regulação de realimentação negativa de GTR reductase. A redução da regulação de realimentação negativa de GTR reductase é, em

30 algumas concretizações, reduzida pela inserção de uma seqüência KK positivamente carregada em ou perto do terminal-N.

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é adicionalmente geneticamente modificada para sobreexpressar ALA sintase. Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é adicionalmente geneticamente modificada para produzir um nível de ALA sintase que é pelo menos cerca de 10%, pelo menos cerca de 15%, pelo menos cerca de 20%, pelo menos cerca de 25%, pelo menos cerca de 30%, pelo menos cerca de 40%, pelo menos cerca de 50%, pelo menos cerca de 60%, pelo menos cerca de 70%, pelo menos cerca de 80%, pelo menos cerca de 90%, pelo menos cerca de 2 vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 10 vezes, pelo menos cerca de 15 vezes, pelo menos cerca de 20 vezes, ou pelo menos cerca de 25 vezes, ou mais, mais alto do que o nível de atividade de ALA sintase produzida em uma célula hospedeira de controle.

O aumento do nível de atividade de ALA sintase em uma célula é alcançado em um número de modos, incluindo, mas não limitado a: 1) aumento da resistência do promotor do promotor ao qual a região de codificação de ALA sintase é operavelmente ligada; 2) aumento do número de cópia do plasmídeo compreendendo uma sequência de nucleotídeo que codifica ALA sintase; 3) aumento da estabilidade de um mRNA de ALA sintase (onde um "mRNA de ALA sintase" é um mRNA compreendendo uma sequência de nucleotídeo que codifica ALA sintase); 4) modificação do uso de códon de ALA sintase, tal que o nível de translação do mRNA de ALA sintase é aumentado; 5) aumento da estabilidade de enzima de ALA sintase; 6) aumento da atividade específica (atividade de unidades por proteína de unidade) de ALA sintase.

Produção aumentada de um intermediário de trajetória biosintética de terpeno endógeno

As modificações genéticas que aumentam a produção de um intermediário de trajetória biosintética de terpeno endógeno incluem, mas não estão limitadas a, modificações genéticas que resultam em um nível reduzido e/ou atividade de uma fosfotransacetilase na célula hospedeira. A

concentração intracelular de um intermediário de trajetória biosintética de terpeno é aumentada pelo aumento da concentração intracelular de acetil-CoA. *E. coli* secreta uma fração significativa de acetil-CoA intracelular na forma de acetato no meio. A anulação do gene que codifica fosfotransacetilase, *pta*, a primeira enzima responsável pela transformação de acetil-CoA em acetato, reduz a secreção de acetato. As modificações genéticas que reduzem o nível e/ou atividade de fosfotransacetilase em uma célula hospedeira procariótica são particularmente úteis onde a célula hospedeira geneticamente é uma que é geneticamente modificada com um ácido nucléico compreendendo seqüências de nucleotídeo que codificam um ou mais produtos de gene de trajetória de MEV.

Em algumas concretizações, uma modificação genética que resulta em um nível reduzido de fosfotransacetilase em uma célula hospedeira procariótica é uma mutação genética que incapacita funcionalmente o gene *pta* endógeno procariótico de célula hospedeira que codifica a fosfotransacetilase. O gene *pta* pode ser funcionalmente incapacitado em qualquer de uma variedade de modos, incluindo inserção de um elemento genético móvel (por exemplo, um transposon, etc.); anulação de todo ou parte do gene, tal que o produto de gene não é produzido, ou é truncado, e é não-funcional na conversão de acetil-CoA em acetato; a mutação do gene tal que o produto de gene não é produzido, ou é truncado, e é não-funcional na conversão de acetil-CoA em acetato; anulação ou mutação de um ou mais elementos de controle que controlam a expressão do gene *pta*, tal que o produto de gene não é produzido; e similares.

Em algumas concretizações, o gene endógeno *pta* de uma célula hospedeira geneticamente modificada é anulado. Qualquer método de anulação de um gene pode ser usado. Um exemplo não-limitativo de um método para anulação de um gene *pta* é pelo uso do sistema de recombinação $\square$ Vermelho. Datsenko e Wanner (2000) *Proc Natl Acad Sci U S A* 97(12): p. 6640-5. O gene *pta* será, em algumas concretizações, anulado de uma célula hospedeira (por exemplo, *E. coli*) que é geneticamente modificada com um ácido nucléico compreendendo seqüências de nucleotídeo que codificam

MK, PMK, MPD, e IDI. O gene *pta* será, em algumas concretizações, anulado de uma célula hospedeira (por exemplo, *E. coli*) que é geneticamente modificada com um ácido nucléico compreendendo seqüências de nucleotídeo que codificam MK, PMK, MPD, e IPP. O gene *pta* será, em algumas concretizações, anulado de uma célula hospedeira (por exemplo, *E. coli*) que é geneticamente modificada com um ácido nucléico compreendendo seqüências de nucleotídeo que codificam MK, PMK, MPD, IPP, e uma prenil transferase.

Trajetória de DXP funcionalmente incapacitada

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é uma que é geneticamente modificada para incluir um ou mais ácidos nucléico compreendendo uma seqüência de nucleotídeo(s) que codifica produto(s) de gene de trajetória biosintética codificada de MEV; e que é adicionalmente geneticamente modificada tal que um gene de trajetória biosintética de DXP endógeno é funcionalmente incapacitado. Em outras concretizações, uma célula hospedeira objeto geneticamente modificada é uma que é geneticamente modificada para incluir um ou mais ácidos nucléicos compreendendo uma seqüência de nucleotídeo(s) que codifica produto(s) de gene de trajetória biosintética de DXP; e que é adicionalmente geneticamente modificada tal que um gene de trajetória biosintética codificada de MEV endógeno é funcionalmente incapacitado.

Em algumas concretizações, onde a célula hospedeira objeto geneticamente modificada é uma célula hospedeira procariótica que é geneticamente modificada com ácido nucléico(s) compreendendo seqüências de nucleotídeo que codificam um ou mais produtos de gene de trajetória de MEV, a célula hospedeira será adicionalmente geneticamente modificada, tal que um ou mais genes de trajetória DXP endógenos são funcionalmente incapacitados. Os genes de trajetória DXP que podem ser funcionalmente incapacitados incluem um ou mais dos genes que codificam qualquer dos seguintes produtos de gene DXP: 1-deoxi-D-xilulose-5-fosfato sintase, 1-deoxi-D-xilulose-5-fosfato reductoisomerase, 4-difosfocitidil-2-C-metil-D-eritritol sintase, 4-difosfocitidil-2-C-metil-D-eritritol kinase, 2C-metil-D-eritritol 2,4-

ciclodifosfato sintase, e 1-hidroxi-2-metil-2-(*E*)-butenil 4-difosfato sintase.

Um gene de trajetória DXP endógeno pode ser funcionalmente incapacitado em qualquer de uma variedade de modos, incluindo inserção de um elemento genético móvel (por exemplo, um transposon, etc.); anulação de todo ou parte do gene, tal que o produto de gene não é produzido, ou é truncado, e é enzimaticamente inativo; mutação do gene tal que o produto de gene não é produzido, ou é truncado, e é enzimaticamente não-funcional; anulação ou mutação de um ou mais elementos de controle que controlam a expressão do gene tal que o produto de gene não é produzido; e similares.

Em outras concretizações, onde a célula hospedeira objeto geneticamente modificada é uma célula hospedeira procariótica que é geneticamente modificada com ácido(s) nucléico(s) compreendendo seqüências de nucleotídeo que codificam um ou mais produtos de gene de trajetória DXP, a célula hospedeira será adicionalmente geneticamente modificada, tal que um ou mais genes de trajetória de MEV endógenos é funcionalmente incapacitado. Os genes de trajetória de MEV endógenos que podem ser funcionalmente incapacitados incluem um ou mais genes que codificam qualquer dos seguintes produtos de gene de MEV: HMGS, HMGR, MK, PMK, MPD, e IDI. Um gene de trajetória de MEV endógeno pode ser funcionalmente incapacitado em qualquer de uma variedade de modos, incluindo inserção de um elemento genético móvel (por exemplo, um transposon, etc.); anulação de todo ou parte do gene, tal que o produto de gene não é produzido, ou é truncado, e é enzimaticamente inativo; mutação do gene, tal que o produto de gene não é produzido, ou é truncado, e é enzimaticamente não-funcional; anulação ou mutação de um ou mais elementos de controle que controlam a expressão do gene, tal que o produto de gene não é produzido; e similares.

Composições compreendendo a célula hospedeira objeto geneticamente modificada

A presente invenção adicionalmente proporciona composições compreendendo uma célula hospedeira objeto geneticamente modificada. Uma composição objeto compreende uma célula hospedeira objeto geneti-

camente modificada, e compreenderá, em algumas concretizações, um ou mais componentes adicionais, cujos componentes são selecionados baseados em parte no uso pretendido da célula hospedeira geneticamente modificada. Componentes adequados incluem, mas não estão limitados a, sais; 5 tampões; estabilizadores; agentes de inibição de protease; agentes de inibição de nuclease; compostos de preservação de membrana celular e/ou de preservação de parede, por exemplo, glicerol, dimetilsulfóxido, etc.; meio nutricional apropriado à célula; e similares. Em algumas concretizações, as células são liofilizadas.

10 Plantas Transgênicas

Em algumas concretizações, um ácido nucléico objeto, ou um vetor objeto de expressão (por exemplo, um ácido nucléico de enzima modificada de citocromo P450 objeto, ou um vetor objeto de expressão compreendendo o ácido nucléico de enzima modificada de citocromo P450) é usado 15 como um transgene para gerar uma planta transgênica que produz a enzima modificada de citocromo P450 codificada. Desse modo, a presente invenção adicionalmente proporciona uma planta transgênica (ou uma parte de planta, semente, tecido, etc.), cuja planta compreende um transgene compreendendo um ácido nucléico objeto compreendendo uma sequência de nucleotídeo 20 que codifica uma enzima modificada de citocromo P450, conforme descrito acima. Em algumas concretizações, o genoma da planta transgênica compreende um ácido nucléico objeto. Em algumas concretizações, a planta transgênica é homozigótica para a modificação genética. Em algumas concretizações, a planta transgênica é heterozigótica para a modificação genética. 25 ca.

Em algumas concretizações, uma planta transgênica objeto produz um citocromo P450 transgene codificado modificado, e produz um produto do citocromo P450 modificado em uma quantidade que é pelo menos cerca de 50%, pelo menos cerca de 2 vezes, pelo menos cerca de 5 30 vezes, pelo menos cerca de 10 vezes, pelo menos cerca de 25 vezes, pelo menos cerca de 50 vezes, ou pelo menos cerca de 100 vezes, ou mais alta, do que a quantidade do produto produzido por uma planta de controle, por

exemplo, uma planta não-transgênica (uma planta que não inclui o transgene que codifica o polipeptídeo) da mesma espécie.

Em algumas concretizações, uma planta transgênica objeto é uma versão transgênica de uma planta não-transgênica de controle que normalmente produz um composto de isoprenóide que é gerado por, ou é um produto a jusante de uma enzima transgene codificada modificada de modificação de precursor de isoprenóide; onde a planta transgênica produz o composto de isoprenóide em uma quantidade que é pelo menos cerca de 50%, pelo menos cerca de 2 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 10 vezes, pelo menos cerca de 25 vezes, pelo menos cerca de 50 vezes, ou pelo menos cerca de 100 vezes, ou mais alta, do que a quantidade do composto de isoprenóide produzido pela planta de controle, por exemplo, um planta não-transgênica (uma planta que não inclui o transgene que codifica o polipeptídeo) da mesma espécie.

Os métodos de introdução de ácidos nucléicos exógenos em células de planta são bem conhecidos na técnica. Tais células de planta são consideradas "transformadas", conforme definido acima. Métodos adequados incluem infecção viral (tais como vírus de DNA de trançado duplo), transfecção, conjugação, fusão de protoplasto, eletroporação, tecnologia de pistola de partícula, precipitação de fosfato de cálcio, microinjeção direta, tecnologia de carbeto de silício, transformação mediada por *Agrobacterium*, e similares. A escolha do método é geralmente dependente do tipo de célula sendo transformada, e das circunstâncias sob as quais a transformação está ocorrendo (isto é, *in vitro*, *ex vivo*, ou *in vivo*).

Os métodos de transformação baseados na bactéria de solo *Agrobacterium tumefaciens* são particularmente úteis para introdução de uma molécula de ácido nucléico exógena em uma planta vascular. A forma de tipo selvagem de *Agrobacterium* contém um plasmídeo Ti (indução de tumor) que direciona produção de crescimento de bile de coroa tumorigênico em plantas hospedeiras. A transferência da região de T-DNA de indução de tumor do plasmídeo Ti para um genoma de planta requer os genes de virulência codificados por plasmídeo Ti, bem como limites de T-DNA, que são

um conjunto de repetições diretas de DNA que esboçam a região a ser transferida. Um vetor baseado em *Agrobacterium* é uma forma modificada de um plasmídeo Ti, em que o tumor que induz funções é substituído pela seqüência de ácido nucléico de interesse a ser introduzida no hospedeiro da planta.

A transformação mediada por *Agrobacterium* geralmente emprega vetores co-integrados ou, preferivelmente, sistemas de vetor binários, em que os componentes do plasmídeo Ti são divididos entre um vetor auxiliar, que reside permanentemente no hospedeiro *Agrobacterium*, e conduz os genes de virulência, e um vetor secundário, que contém o gene de interesse ligado pelas seqüências de T-DNA. Uma variedade de vetores binários é bem conhecida na técnica, e são comercialmente disponíveis, por exemplo, de Clontech (Palo Alto, Calif.). Métodos de co-cultura de *Agrobacterium* com células de planta cultivadas ou tecidos ofendidos, tais com tecido de folha, explantes de raiz, hipocotiledons, peças de caule ou tubérculos, por exemplo, também são bem conhecidos na técnica. Ver., por exemplo, Glick e Thompson, (eds.), *Métodos in Plant Molecular Biology e Biotechnology*, Boca Raton, Fla.: CRC Press (1993).

A transformação mediada por *Agrobacterium* é útil para produção de uma variedade de plantas vasculares transgênicas (Wang et al., supra, 1995), incluindo pelo menos uma espécie de *Eucalyptus* e legumes de forragem, tais como alfafa (lucerna); trevo, trevo branco, *Stylosanthes*, *Lotononis bainesii* e sainfoin.

A transformação mediada por microprojétil também pode ser usada para produzir uma planta transgênica objeto. Este método, primeiro descrito por Klein et al. (*Nature* 327:70-73 (1987)), conta com microprojéteis, tais como ouro ou tungstênio, que são revestidos com a molécula de ácido nucléico desejada por precipitação com cloreto de cálcio, espermidina ou polietileno glicol. As partículas de micro-projétil são aceleradas em alta velocidade em um tecido de angiosperma usando um dispositivo tal como o BIOLISTIC PD-1000 (Biorad; Hercules Calif.).

Um ácido nucléico objeto pode ser introduzido em uma planta

de uma maneira tal que o ácido nucléico é capaz de entrar em uma célula(s) de planta, por exemplo, via um protocolo *in vivo* ou *ex vivo*. Por "*in vivo*," é significativo que o ácido nucléico é administrado a um corpo vivo de uma planta, por exemplo, infiltração. Por "*ex vivo*" é significativo que células ou

5 explantes são modificados fora da planta, e, em seguida, tais células ou órgãos são regenerados em uma planta. Um número de vetores adequados para transformação estável de células de planta, ou para o estabelecimento de plantas transgênicas, foram descritos, incluindo aqueles descritos em Weissbach e Weissbach, (1989) *Methods for Plant Molecular Biology* Academic Press, e Gelvin et al., (1990) *Plant Molecular Biology Manual*, Kluwer Academic Publishers. Exemplos específicos incluem aqueles derivados de uma plasmídeo Ti de *Agrobacterium tumefaciens*, bem como aqueles revelados por Herrera-Estrella et al. (1983) *Nature* 303: 209, Bevan (1984) *Nucl Acid Res.* 12: 8711-8721, Klee (1985) *Bio/Technolo* 3: 637-642. Alternativa-

15 mente, vetores de não-Ti podem ser usados para transferir o DNA em plantas e células pelo uso de técnicas de distribuição de DNA livre. Pelo uso destes métodos, plantas transgênicas, tais como trigo, arroz (Christou (1991) *Bio/Technology* 9:957-962) e milho (Gordon-Kamm (1990) *Plant Célula* 2: 603-618), podem ser produzidas. Um embrião imaturo pode também ser um

20 bom tecido alvo para monocots para técnicas de distribuição de DNA direta pelo uso de pistola de partícula (Weeks et al. (1993) *Plant Physiol* 102: 1077-1084; Vasil (1993) *Bio/Technolo* 10: 667-674; Wan e Lemeaux (1994) *Plant Physiol* 104: 37-48, e para transferência de DNA mediada por *Agrobacterium* (Ishida et al. (1996) *Nature Biotech* 14: 745-750). Métodos exemplares para

25 introdução de DNA em cloroplastos são bombardeio biolístico, transformação de polietileno glicol de protoplastos, e microinjeção (Danieli et al *Nat. Biotechnol* 16:345-348, 1998; Staub et al *Nat. Biotechnol* 18: 333-338, 2000; O'Neill et al *Plant J.* 3:729-738, 1993; Knoblauch et al *Nat. Biotechnol* 17: 906-909; Patentes dos Estados Unidos Nos. 5.451.513, 5.545.817, 5.545.818 e 5.576.198; Pedido de Patente Internacional No. WO 95/16783; e em Boynton et al., *Methods in Enzymology* 217: 510-536 (1993), Svab et al., *Proc. Natl. Acad. Sci. USA* 90: 913-917 (1993), e McBride et al., *Proc. Natl.*

30

Acad. Sci. USA 91: 7301-7305 (1994)). Qualquer vetor adequado para os métodos de bombardeio biolístico, transformação de polietileno glicol de protoplastos e microinjeção serão adequados como um vetor alvo para transformação de cloroplasto. Qualquer vetor de DNA de trançado duplo pode ser
5 usado como um vetor de transformação, especialmente quando o método de introdução não utiliza *Agrobacterium*.

Plantas que podem ser geneticamente modificadas incluem grãos, colheitas de forragem, frutas, vegetais, colheitas de semente de óleo, palmas, silvicultura, e videiras. Exemplos específicos de plantas que podem
10 ser modificadas se seguem: milho, banana, amendoim, ervilhas do campo, girassol, tomate, canola, tabaco, trigo, cevada, aveias, batata, feijão-soja, algodão, cravos, sorgo, tremoço e arroz. Outros exemplos incluem *Artemisia annua*, ou outras plantas conhecidas para produzir composto de isoprenóides de interesse.

15 Também providas pela presente invenção são células de planta transformadas, tecidos, sementes, plantas, e produtos que contêm as células de planta transformadas. Uma característica das células transformadas objetos, e tecidos e produtos que incluem as mesmas, é a presença de um ácido nucléico objeto integrado no genoma, e produção por células de planta
20 de uma enzima modificada de citocromo P450. Células de planta recombinantes da presente invenção são úteis como populações de células recombinantes, ou como um tecido, semente, planta total, caule, fruto, folha, raiz, flor, caule, tubérculo, grão, alimentação de animal, um campo de plantas, e similares.

25 Também provido pela presente invenção é material reprodutivo de uma planta transgênica objeto, onde o material reprodutivo inclui sementes, plantas de progenia e material clonal.

MÉTODOS DE PRODUÇÃO DE UM PRODUTO DE UMA TRAJETÓRIA BIOSINTÉTICA

30 A presente invenção proporciona métodos de produção de um produto de trajetória biosintética. Os métodos geralmente envolvem cultura de uma célula hospedeira objeto geneticamente modificada em um meio a-

dequado. A célula hospedeira objeto geneticamente modificada é uma que foi geneticamente modificada com um ácido nucléico compreendendo a sequência de nucleotídeo que codifica uma enzima modificada de citocromo P450 operavelmente ligada a um domínio selecionado de um domínio de transmembrana, um domínio de secreção, um domínio de solubilização, e uma proteína de inserção de membrana, para produzir uma enzima modificada de citocromo P450. Na presença de um intermediário de trajetória biosintética, a produção da enzima modificada de citocromo P450 resulta em modificação enzimática do intermediário, e produção de um produto de trajetória biosintética. Em outras concretizações, os métodos geralmente envolvem manutenção de uma planta transgênica objeto sob condições que favorecem a produção da enzima modificada de citocromo P450 codificada. A produção da enzima modificada de citocromo P450 resulta na produção do produto de trajetória biosintética. Tipicamente, o método é efetuado *in vitro* (por exemplo, em uma célula viva cultivada *in vitro*), embora produção *in vivo* de um produto de trajetória biosintética seja também contemplada. Em algumas destas concretizações, a célula hospedeira é uma célula eucariótica, por exemplo, uma célula de levedura. Em outras concretizações, a célula hospedeira é uma célula procariótica. Em algumas destas concretizações, a célula hospedeira é uma célula de planta. Em algumas concretizações, o método é efetuado em uma planta transgênica objeto.

Uma célula hospedeira objeto geneticamente modificada proporciona produção aumentada de um produto de trajetória biosintética, comparada a uma célula hospedeira de origem de controle. Desse modo, por exemplo, a produção de um produto de trajetória biosintética é aumentada por pelo menos cerca de 10%, pelo menos cerca de 20%, pelo menos cerca de 50%, pelo menos cerca de 2 vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 10 vezes, pelo menos cerca de 20 vezes, pelo menos cerca de 30 vezes, pelo menos cerca de 40 vezes, pelo menos cerca de 50 vezes, pelo menos cerca de 75 vezes, pelo menos cerca de 100 vezes, pelo menos cerca de 200 vezes, pelo menos cerca de 300 vezes, pelo menos cerca de 400 vezes, ou pelo menos cerca de 500

vezes, ou mais, na célula hospedeira geneticamente modificada, comparada ao nível do produto produzido em uma célula hospedeira de origem de controle. Uma célula hospedeira de origem de controle não compreende a(s) presentes modificação(ões) genética(s) presente(s) na célula hospedeira geneticamente modificada.

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada proporciona produção aumentada de um produto de trajetória biosintética, comparada a uma célula hospedeira de controle. Desse modo, por exemplo, a produção de um produto de trajetória biosintética é aumentada por pelo menos cerca de 10%, pelo menos cerca de 20%, pelo menos cerca de 50%, pelo menos cerca de 2 vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 10 vezes, pelo menos cerca de 20 vezes, pelo menos cerca de 30 vezes, pelo menos cerca de 40 vezes, pelo menos cerca de 50 vezes, pelo menos cerca de 75 vezes, pelo menos cerca de 100 vezes, pelo menos cerca de 200 vezes, pelo menos cerca de 300 vezes, pelo menos cerca de 400 vezes, ou pelo menos cerca de 500 vezes, ou mais, na célula hospedeira geneticamente modificada, comparada ao nível do produto produzido em uma célula hospedeira de controle. Em algumas destas concretizações, a célula hospedeira de controle não compreende modificação(ões) genética(s) presente(s) na célula hospedeira geneticamente modificada, por exemplo, a enzima de modificação de isoprenóide que codifica ácido nucléico (por exemplo, a enzima de citocromo P450 que codifica ácido nucléico) na célula hospedeira de controle é operavelmente ligada a um ou mais de um domínio nativo de transmembrana, um domínio nativo de secreção, um domínio nativo de solubilização, e um polipeptídeo de inserção de membrana nativa, enquanto a célula hospedeira geneticamente modificada compreende uma enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um ou mais domínio não-nativo (por exemplo, heterólogo) de transmembrana, um domínio não-nativo de secreção, um domínio não-nativo de solubilização, e um domínio não-nativo de inserção de membrana. Como um exemplo, onde a célula hospedeira geneticamente modificada compreende uma

enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a uma enzima não-nativa de modificação de isoprenóide que codifica ácido nucléico, uma célula hospedeira de controle adequada compreende a enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um domínio nativo de transmembrana. Como outro exemplo, onde a célula hospedeira geneticamente modificada compreende uma enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um sinal heterólogo de domínio de secreção, uma célula hospedeira de controle adequada compreende a enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um domínio nativo de transmembrana. Como outro exemplo, onde a célula hospedeira geneticamente modificada compreende uma enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um domínio heterólogo de solubilização, uma célula hospedeira de controle adequada compreende a enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um domínio nativo de transmembrana. Como outro exemplo, onde a célula hospedeira geneticamente modificada compreende uma enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um domínio de inserção de membrana heterólogo, uma célula hospedeira de controle adequada compreende a enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um domínio nativo de transmembrana. Como outro exemplo, onde célula hospedeira geneticamente modificada compreende uma enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um domínio variante de transmembrana (por exemplo, uma mutilação do domínio nativo de transmembrana; um domínio de transmembrana compreendendo uma mudança na seqüência de amino ácido comparada à seqüência de amino ácido do domínio nativo de transmembrana), uma célula hospedeira de controle adequada compreende a enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a domínio nativo de transmembrana.

A presente invenção proporciona métodos de produção de

composto de isoprenóide. Os métodos geralmente envolvem cultura de uma célula hospedeira objeto geneticamente modificada em um meio adequado, onde a célula hospedeira objeto geneticamente modificada é uma que foi geneticamente modificada com um ácido nucléico compreendendo uma sequência de nucleotídeo que codifica uma enzima de modificação de precursor de isoprenóide operavelmente ligada a um domínio selecionado de um domínio de transmembrana, um domínio de secreção, um domínio de solubilização, e uma proteína de inserção de membrana, para produzir uma enzima de modificação de precursor de isoprenóide. Na presença de um precursor de composto de isoprenóide, a produção da enzima de modificação de precursor de isoprenóide resulta em modificação enzimática do precursor de isoprenóide, e produção do composto de isoprenóide. Em outras concretizações, os métodos geralmente envolvem manutenção de uma planta transgênica objeto sob condições que favorecem a produção da enzima de modificação de precursor de isoprenóide codificada. A produção da enzima de modificação de precursor de isoprenóide resulta na produção do composto de isoprenóide. Por exemplo, em algumas concretizações, os métodos geralmente envolvem cultura de uma célula hospedeira geneticamente modificada em um meio adequado, no qual referida célula hospedeira é geneticamente modificada com um ácido nucléico objeto compreendendo uma sequência de nucleotídeo que codifica uma enzima de modificação de terpeno, por exemplo, um terpeno oxidase, um terpeno hidroxilase, etc. A produção do terpeno oxidase resulta na produção do composto de isoprenóide. Tipicamente, o método é efetuado *in vitro* (por exemplo, em uma célula viva cultivada *in vitro*), embora produção *in vivo* de um composto de isoprenóide seja também contemplada. Em algumas destas concretizações, a célula hospedeira é uma célula eucariótica, por exemplo, uma célula de levedura. Em outras concretizações, a célula hospedeira é uma célula procariótica. Em algumas destas concretizações, a célula hospedeira é uma célula de planta. Em algumas concretizações, o método é efetuado em uma planta transgênica objeto.

Uma célula hospedeira objeto geneticamente modificada pro-

porciona produção aumentada de um composto de isoprenóide, comparada a uma célula hospedeira de origem de controle. Desse modo, por exemplo, a produção de um isoprenóide ou precursor de isoprenóide é aumentada por pelo menos cerca de 10%, pelo menos cerca de 20%, pelo menos cerca de 50%, pelo menos cerca de 2 vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 10 vezes, pelo menos cerca de 20 vezes, pelo menos cerca de 30 vezes, pelo menos cerca de 40 vezes, pelo menos cerca de 50 vezes, pelo menos cerca de 75 vezes, pelo menos cerca de 100 vezes, pelo menos cerca de 200 vezes, pelo menos cerca de 300 vezes, pelo menos cerca de 400 vezes, ou pelo menos cerca de 500 vezes, ou mais, na célula hospedeira geneticamente modificada, comparada a uma célula hospedeira de origem de controle. Uma célula hospedeira de origem de controle não compreende a(s) modificação(ões) genética(s) presente(s) na célula hospedeira geneticamente modificada.

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada proporciona produção aumentada de um composto de isoprenóide, comparada a uma célula hospedeira de controle. Desse modo, por exemplo, a produção de um isoprenóide ou precursor de isoprenóide é aumentada por pelo menos cerca de 10%, pelo menos cerca de 20%, pelo menos cerca de 50%, pelo menos cerca de 2 vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 10 vezes, pelo menos cerca de 20 vezes, pelo menos cerca de 30 vezes, pelo menos cerca de 40 vezes, pelo menos cerca de 50 vezes, pelo menos cerca de 75 vezes, pelo menos cerca de 100 vezes, pelo menos cerca de 200 vezes, pelo menos cerca de 300 vezes, pelo menos cerca de 400 vezes, ou pelo menos cerca de 500 vezes, ou mais, na célula hospedeira geneticamente modificada, comparada a uma célula hospedeira de controle. Em algumas destas concretizações, a célula hospedeira de controle não compreende a(s) modificação(ões) genética(s) presente(s) na célula hospedeira geneticamente modificada, por exemplo, a enzima de modificação de isoprenóide que codifica ácido nucléico (por exemplo, a enzima de citocromo P450 que codifica ácido nucléico) na célula hospedeira de controle é operavelmen-

te ligada a um ou mais de um domínio nativo de transmembrana, um domínio nativo de secreção, um domínio nativo de solubilização, e um polipeptídeo de inserção de membrana nativo, enquanto a célula hospedeira geneticamente modificada compreende uma enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um ou mais de um

5 domínio não-nativo (por exemplo, heterólogo) de transmembrana, um domínio não-nativo de secreção, um domínio não-nativo de solubilização, e um domínio não-nativo de inserção de membrana. Como um exemplo, onde célula hospedeira geneticamente modificada compreende uma enzima de mo-

10 dificação de isoprenóide que codifica ácido nucléico operavelmente ligada a uma enzima não-nativa de modificação de isoprenóide que codifica ácido nucléico, uma célula hospedeira de controle adequada compreende a enzima de modificação de isoprenóide que codifica ácido nucléico operavelmen-

15 te ligado a um domínio nativo de transmembrana. Como outro exemplo, onde a célula hospedeira geneticamente modificada compreende uma enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um sinal heterólogo de domínio de secreção, uma célula hospedeira de controle adequada compreende a enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um domínio nativo de transmembrana. Como outro exem-

20 plo, onde a célula hospedeira geneticamente modificada compreende uma enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um domínio heterólogo de solubilização, uma célula hospedeira de controle adequada compreende a enzima de modificação de isoprenóide que codifica ácido nucléico opera-

25 velmente ligado a um domínio nativo de transmembrana. Como outro exemplo, onde célula hospedeira geneticamente modificada compreende uma enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um domínio de inserção de membrana heterólogo, uma célula hospedeira de controle adequada compreende a enzima de modificação

30 de isoprenóide que codifica ácido nucléico operavelmente ligado a um domínio nativo de transmembrana. Como outro exemplo, onde a célula hospedeira geneticamente modificada compreende uma enzima de modificação de

isoprenóide que codifica ácido nucléico operavelmente ligado a um domínio variante de transmembrana (por exemplo, uma mutilação do domínio nativo de transmembrana; um domínio de transmembrana compreendendo uma mudança na seqüência de amino ácido comparada à seqüência de amino ácido do domínio nativo de transmembrana), uma célula hospedeira de controle adequada compreende a enzima de modificação de isoprenóide que codifica ácido nucléico operavelmente ligado a um domínio nativo de transmembrana.

Desse modo, em algumas concretizações, uma célula hospedeira objeto geneticamente modificada produz, em uma célula base, um nível de um composto de isoprenóide que é pelo menos cerca de 10%, pelo menos cerca de 15%, pelo menos cerca de 20%, pelo menos cerca de 25%, pelo menos cerca de 30%, pelo menos cerca de 35%, pelo menos cerca de 40%, pelo menos cerca de 45%, pelo menos cerca de 50%, pelo menos cerca de 60%, pelo menos cerca de 70%, pelo menos cerca de 80%, pelo menos cerca de 90%, pelo menos cerca de 2 vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 10 vezes, pelo menos cerca de 20 vezes, pelo menos cerca de 30 vezes, pelo menos cerca de 40 vezes, pelo menos cerca de 50 vezes, pelo menos cerca de 75 vezes, pelo menos cerca de 100 vezes, pelo menos cerca de 200 vezes, pelo menos cerca de 300 vezes, pelo menos cerca de 400 vezes, ou pelo menos cerca de 500 vezes, ou mais, mais alto do que o nível do composto de isoprenóide produzido em uma célula hospedeira de controle não compreende a uma ou mais modificações genéticas que a célula hospedeira geneticamente modificada compreende. O crescimento de células hospedeiras geneticamente modificadas é prontamente determinado usando-se métodos bem conhecidos, por exemplo, medição de densidade ótica (OD) a cerca de 600 nm (OD_{600}) de culturas líquidas de bactéria; tamanho de colônia; taxa de crescimento; e similares.

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada produz um composto de isoprenóide em uma quantidade recuperável de pelo menos cerca de 1mg/L, pelo menos cerca

de 5 mg/L, pelo menos cerca de 10 mg/L, pelo menos cerca de 15 mg/L, pelo menos cerca de 20 mg/L, pelo menos cerca de 25 mg/L, pelo menos cerca de 30 mg/L, pelo menos cerca de 35 mg/L, pelo menos cerca de 40 mg/L, pelo menos cerca de 50 mg/L, pelo menos cerca de 75 mg/L, pelo menos cerca de 100 mg/L, pelo menos cerca de 125 mg/L, pelo menos cerca de 150 mg/L, pelo menos cerca de 200 mg/L, pelo menos cerca de 300 mg/L, pelo menos cerca de 500 mg/L, pelo menos cerca de 1000 mg/L, ou pelo menos cerca de 5000 mg/L.

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada produz um composto de isoprenóide em uma quantidade recuperável de cerca de 1 mg/L a cerca de 5000 mg/L, por exemplo, de cerca de 1 mg/L a cerca de 2 mg/L, de cerca de 2 mg/L a cerca de 5 mg/L, de cerca de 5 mg/L a cerca de 10 mg/L, de cerca de 10 mg/L a cerca de 15 mg/L, de cerca de 15 mg/L a cerca de 20 mg/L, de cerca de 20 mg/L a cerca de 25 mg/L, de cerca de 25 mg/L a cerca de 50 mg/L, de cerca de 50 mg/L a cerca de 75 mg/L, de cerca de 75 mg/L a cerca de 100 mg/L, de cerca de 100 mg/L a cerca de 150 mg/L, de cerca de 150 mg/L a cerca de 200 mg/L, de cerca de 200 mg/L a cerca de 250 mg/L, de cerca de 250 mg/L a cerca de 300 mg/L, de cerca de 300 mg/L a cerca de 350 mg/L, de cerca de 350 mg/L a cerca de 400 mg/L, de cerca de 400 mg/L a cerca de 450 mg/L, de cerca de 450 mg/L a cerca de 500 mg/L, de cerca de 500 mg/L a cerca de 1000 mg/L, de cerca de 1000 mg/L a cerca de 2000 mg/L, de cerca de 2000 mg/L a cerca de 3000 mg/L, de cerca de 3000 mg/L a cerca de 4000 mg/L, ou de cerca de 4000 mg/L a cerca de 5000 mg/L. Os isoprenóides produzidos podem ser recuperados a partir do meio, ou a partir da célula hospedeira, por exemplo, a partir do meio de cultura, ou a partir de uma célula lisato, ou uma fração de uma célula lisato. Os métodos de recuperação podem variar, dependendo de uma variedade de fatores, por exemplo, da natureza dos isoprenóides específicos que são produzidos.

As Figuras 14 e 15 representam esquematicamente a biosíntese de produtos de isoprenóide exemplares. A conversão de poliprenil difosfatos lineares é catalisada por terpeno sintases; e os produtos da conversão

são os substratos de uma enzima de modificação de precursor de isoprenóide (por exemplo, uma enzima P450). A funcionalização específica, em seguida, ocorre por reação do esqueleto de carbono do precursor, catalisado por um P450 e seu co-participante redox, um CPR.

5 Em algumas concretizações, a célula hospedeira geneticamente modificada é adicionalmente geneticamente modificada com um ácido nucléico compreendendo uma seqüência de nucleotídeo que codifica um terpeno sintase, que pode ser um terpeno sintase heterólogo (por exemplo, um terpeno sintase não normalmente produzido na célula hospedeira). Des-

10 se modo, por exemplo, a célula hospedeira é, em algumas concretizações, geneticamente modificada com um ou mais ácidos nucléicos compreendendo seqüências de nucleotídeo que codificam um terpeno sintase, e uma enzima de modificação de isoprenóide (por exemplo, um sesquiterpeno oxidase). A cultura de tal célula hospedeira em um meio de cultura adequado pro-

15 porciona produção do terpeno sintase, e a enzima de modificação de isoprenóide (por exemplo, um sesquiterpeno oxidase). Por exemplo, o terpeno sintase modifica um farnesil pirofosfato para gerar um sesquiterpeno substrato para referido sesquiterpeno oxidase.

 Em algumas concretizações, a célula hospedeira é adicional-

20 mente geneticamente modificada com um ácido nucléico compreendendo uma seqüência de nucleotídeo que codifica um citocromo P450 reductase (CPR). Uma ampla variedade de seqüências de nucleotídeos de CPR é conhecida, e qualquer ácido nucléico que codifica CPR conhecido pode ser usado, considerando-se que o CPR codificado exibe atividade na transferên-

25 cia de elétrons a partir de NADPH. Em algumas concretizações, o ácido nucléico que codifica CPR codifica um CPR que transfere elétrons de NADPH para uma enzima de modificação de isoprenóide, por exemplo, um sesquiterpeno oxidase, codificado por uma enzima de modificação de isoprenóide que codifica ácido nucléico.

30 Em algumas concretizações, uma célula hospedeira é adicionalmente geneticamente modificada para produzir uma prenil transferase e/ou uma ou mais enzimas em uma trajetória biosintética para produzir iso-

pentenil pirofosfato. As células tipicamente usam uma de duas trajetórias para gerar isoprenóides ou precursores de isoprenóide (por exemplo, IPP, poliprenil difosfatos, etc.). As Figuras 16-18 servem para ilustrar as trajetórias usadas pelas células para gerar compostos de isoprenóide, ou precursores, tais como poliprenil difosfatos.

A Figura 16 representa trajetórias de isoprenóide envolvendo modificação de isopentenil difosfato (IPP) e/ou seu isômero dimetilalil difosfato (DMAPP) por prenil transferases para gerar os poliprenil difosfatos geranyl difosfato (GPP), farnesil difosfato (FPP), e geranylgeranyl difosfato (GGPP). GPP e FPP são adicionalmente modificados por terpeno sintases para gerar monoterpenos e sesquiterpenos, respectivamente; e GGPP é adicionalmente modificado por terpeno sintases para gerar diterpenos e carotenóides. IPP e DMAPP são gerados por uma de duas trajetórias: a trajetória de mevalonato (MEV) e a trajetória de 1-deoxi-D-xilulose-5-fosfato (DXP).

A Figura 17 representa esquematicamente a trajetória de MEV, onde acetil CoA é convertida via uma série de reações para IPP.

A Figura 18 representa esquematicamente a trajetória de DXP, na qual piruvato e D-gliceraldeído-3-fosfato são convertidos via uma série de reações a IPP e DMAPP. Células eucarióticas outras do que células de planta usam a trajetória de isoprenóide de MEV exclusivamente para converter acetil-coenzima A (acetil-CoA) em IPP, que é subsequëntemente isomerizado a DMAPP. As plantas usam ambos o MEV e o mevalonato independente, ou trajetórias de DXP para síntese de isoprenóide. As procariótes, com algumas exceções, usam a trajetória de DPX para produzir IPP e DMAPP separadamente através de um ponto de desvio.

Dependendo do meio de cultura no qual a célula hospedeira é cultivada, e dependendo se a célula hospedeira sintetiza IPP, via uma trajetória de DXP, ou via uma trajetória de mevalonato, a célula hospedeira incluirá, em algumas concretizações, adicionalmente modificações genéticas. Por exemplo, em algumas concretizações, a célula hospedeira é uma que não tem uma trajetória endógena de mevalonato, por exemplo, a célula hospedeira é uma que normalmente não sintetiza IPP ou mevalonato via uma tra-

jetória de mevalonato. Por exemplo, em algumas concretizações, a célula hospedeira é uma que não sintetiza normalmente IPP, via uma trajetória de mevalonato, e a célula hospedeira é geneticamente modificada com um ou mais ácidos nucléicos compreendendo seqüências de nucleotídeo que codificam duas ou mais enzimas na trajetória de mevalonato, um IPP isomerase, uma preniltransferase, um terpeno sintase, e uma enzima de modificação de isoprenóide (por exemplo, uma enzima de modificação de isoprenóide codificada por um ácido nucléico objeto). A cultura de tal célula hospedeira proporciona produção da trajetória de enzimas de mevalonato, o IPP isomerase, a preniltransferase, o terpeno sintase, e a enzima de modificação de isoprenóide (por exemplo, um sesquiterpeno oxidase). A produção da trajetória de enzimas de mevalonato, do IPP isomerase, da preniltransferase, do terpeno sintase, e da enzima de modificação de isoprenóide (por exemplo, um sesquiterpeno oxidase) resulta na produção de um composto de isoprenóide. Em muitas concretizações, a preniltransferase é uma FPP sintase, que gera um substrato de sesquiterpeno para um sesquiterpeno oxidase codificado por um ácido nucléico objeto; e a produção do sesquiterpeno oxidase resulta na oxidação do substrato de sesquiterpeno na célula hospedeira. Quaisquer ácidos nucléicos que codificam as enzimas de trajetória de mevalonato, o IPP isomerase, a preniltransferase, e o terpeno sintase, são adequados para uso. Por exemplo, ácidos nucléicos adequados são descritos em, por exemplo, Martin et al. (2003) *supra*.

Em algumas das concretizações acima descritas, onde a célula hospedeira é geneticamente modificada com um ou mais ácidos nucléicos compreendendo seqüências de nucleotídeo que codificam duas ou mais enzimas de trajetória de mevalonato, as duas ou mais enzimas de trajetória de mevalonato incluem MK, PMK, e MPD, e a célula hospedeira é cultivada em meio que inclui mevalonato. Em outras concretizações, as duas ou mais enzimas de trajetória de mevalonato incluem acetoacetyl CoA tiolase, HMGS, HMGR, MK, PMK, e MPD.

Em algumas concretizações, a célula hospedeira é uma que não sintetiza normalmente IPP via trajetória de mevalonato, a célula hospedeira

deira é geneticamente modificada conforme descrito acima, e a célula hospedeira adicionalmente compreende uma trajetória funcionalmente incapacitada de DXP.

Um método objeto é útil para produção de uma variedade de composto de isoprenóides, incluindo, mas não limitado a, ácido artemisínico (por exemplo, onde o substrato de sesquiterpeno é 4,11-dieno amorfo), aloisolongifoleno álcool (por exemplo, onde o substrato é aloisolongifoleno), (*E*)-*trans*-bergamota-2,12-dien-14-ol (por exemplo, onde o substrato é (-)- α -*trans*-bergamotene), (-)-elema-1,3,11(13)-trien-12-ol (por exemplo, onde o substrato é (-)- β -elemene), germacra-1(10),4,11(13)-trien-12-ol (por exemplo, onde o substrato é (+)-germacrene A), germacrene B álcool (por exemplo, onde o substrato é germacreno B), 5,11(13)-guaiadieno-12-ol (por exemplo, onde o substrato é (+)- γ -gurjuneno), ledeno álcool (por exemplo, onde o substrato é (+)-ledeno), 4 β -H-eudesm-11(13)-ene-4,12-diol (por exemplo, onde o substrato é neointermoeol), (+)- β -costol (por exemplo, onde o substrato é (+)- β -selineno, e similares; e adicionalmente derivados de qualquer dos precedentes.

Uma célula hospedeira objeto geneticamente modificada é, em muitas concretizações, cultivada *in vitro* em um meio adequado e a uma temperatura adequada. A temperatura na qual as células são cultivadas é geralmente de cerca de 18°C a cerca de 40°C, por exemplo, de cerca de 18°C a cerca de 20°C, de cerca de 20°C a cerca de 25°C, de cerca de 25°C a cerca de 30°C, de cerca de 30°C a cerca de 35°C, ou de cerca de 35°C a cerca de 40°C (por exemplo, a cerca de 37°C).

Em algumas concretizações, uma célula hospedeira objeto geneticamente modificada é cultivada em um meio adequado (por exemplo, caldo de Luria-Bertoni, opcionalmente suplementado com um ou mais agentes adicionais, tais como um indutor (por exemplo, onde a enzima de modificação de isoprenóide que codifica seqüência de nucleotídeo está sob o controle de um promotor induzível), etc.); e o meio de cultura é sobreposto com um solvente orgânico, por exemplo, dodecano, formando uma camada orgânica. O composto de isoprenóide produzido pela célula hospedeira geneti-

camente modificada se divide na camada orgânica, a partir da qual ele pode ser purificado. Em algumas concretizações, onde a enzima de modificação de isoprenóide que codifica seqüência de nucleotídeo é operavelmente ligada a um promotor induzível, um indutor é adicionado ao meio de cultura; e, após um tempo adequado, o composto de isoprenóide é isolado a partir da camada orgânica sobreposta no meio de cultura.

Em algumas concretizações, o composto de isoprenóide será separado a partir de outros produtos que podem estar presentes na camada orgânica. A separação do composto de isoprenóide de outros produtos que podem estar presentes na camada orgânica é prontamente efetuada usando-se, por exemplo, técnicas de cromatografia padrões.

Em algumas concretizações, um composto de isoprenóide sintetizado por um método objeto é adicionalmente quimicamente modificado em uma reação de célula livre. Por exemplo, em algumas concretizações, ácido artemisínico é isolado do meio de cultura e/ou uma célula lisato, e o ácido artemisínico é adicionalmente quimicamente modificado em uma reação de célula livre para gerar artemisinin.

Em algumas concretizações, o composto de isoprenóide é puro, por exemplo, pelo menos cerca de 40% puro, pelo menos cerca de 50% puro, pelo menos cerca de 60% puro, pelo menos cerca de 70% puro, pelo menos cerca de 80% puro, pelo menos cerca de 90% puro, pelo menos cerca de 95% puro, pelo menos cerca de 98%, ou mais do que 98% puro, onde "puro" no contexto de um composto de isoprenóide se refere a um composto de isoprenóide que é livre de outros compostos de isoprenóide, macromoléculas, contaminantes, etc.

EXEMPLOS

Os seguintes exemplos são colocados de modo a proporcionar àqueles técnicos no assunto com uma revelação completa e descrição de como produzir e usar a presente invenção, e não são pretendidos para limitar o escopo de o que os inventores pretendem com sua invenção, nem são eles pretendidos para representar que os experimentos abaixo são todos ou somente experimentos realizados. Esforços foram feitos para assegurar pre-

cisão com relação a números usados (por exemplo, quantidades, temperatura, etc.), mas alguns erros experimentais e desvios devem ser considerados. A menos que de outro modo indicado, partes são partes por peso, peso molecular é peso molecular médio de peso, temperatura é em graus Celsius, e pressão é em ou perto de atmosférica. Abreviações padrões podem ser usadas, por exemplo, bp, par(es) base(s); kb, kilobase(s); pl, picolitro(s); s ou seg, segundo(s); min, minuto(s); h ou hr, hora(s); aa, amino ácido(s); kb, kilobase(s); bp, par(es) base(s); nt, nucleotídeo(s); i.m., intramuscular(mente); i.p., intraperitoneal(mente); s.c., subcutânea(mente); e similares.

10 Exemplo 1: Produção de 8-hidroxi- δ -cadineno em *Escherichia coli*

Este exemplo descreve a produção de um substrato produzido *in vivo* em níveis altos (até 30 mg L⁻¹) usando-se o P450 nativo que participa na trajetória biosintética. $\square$ -Cadineno-8-hidroxilase (CadH) é uma P450 ligada a uma membrana derivada de planta que hidroxilata o sesquiterpeno, $\square$ -cadineno (cad), em 8-hidroxi- $\square$ -cadineno (CadOH) na biosíntese de gossipol, um composto de defesa de planta. A biosíntese de CadOH em *E. coli* é representada esquematicamente na Figura 1. O substrato (Cad) é produzido de farnesil pirofosfato endógeno (FPP) em *E. coli* por terpeno sintase CadS.

15 Cad é adicionalmente hidroxilado em produto (CadOH) pela ação de CadH junto com seu co-participante redox(CPR).

O vetor de expressão de CadH inclui ambos o gene de CadH, bem como o gene que codifica um citocromo P450 reductase (CPR) redox co-participante de *Candida tropicalis*. Esta construção foi co-transformado em *E. coli* junto com um vetor de expressão compatível com $\square$ -cadineno sintase (CadS), proporcionando, desse modo, o substrato para CadH. Esta cepa foi desenvolvida em meio rico, e induzida na presença de suplementos "heme" por 48 horas a 20°C antes da extração do meio com solvente orgânico. Os resultados, representados na Figura 2, mostram uma quantidade claramente detectável de CadOH (~100 $\square$ g L⁻¹) produzida neste sistema medido por GC-MS (cromatografia de gás-espectrometria de massa).

25

30

Figura 2. Traço de GC-MS de camada orgânica extraída de *E.*

coli expressando trajetória biosintética de CadOH. A inserção mostra ampliação da região mostrando CadOH (pico 1), e as espécies de cetona putativas (pico 2). A linha superior corresponde a amostras que expressam CadS, CadH, e CPR, enquanto a linha inferior corresponde aos controles negativos que expressam CadS e CPR somente, sem CadH.

Em adição, uma pequena quantidade de um produto de cetona putativo ($[M^+]$: $m/z = 218$) foi também observada (inserção da Figura 2, linha superior, pico 2), significando que circulações múltiplas pela mesma enzima podem ser possíveis. O plasmídeo de controle negativo contendo o CPR somente e não CadH, não exibiu pico de produto no traço de GC-MS (inserção da Figura 2, linha inferior). O espectro de massa do CadOH produzido *in vivo* por *E. coli* usando este sistema e o espectro de literatura de CadOH são muito similares [4]. Tentativas anteriores de usar P450s nativas para produção *in vivo* de produtos naturais funcionalizados em uma família similar de compostos foram mal sucedidas, e apontaram problemas de acesabilidade do substrato.

A produção de CadOH foi significativamente aumentada pelo aumento da quantidade de FPP produzido em *E. coli* usando o plasmídeo de pMBIS, que permite que *E. coli* produza FPP a partir de mevalonato [6]. A sequência de nucleotídeo de pMBIS está representada nas Figuras 36A-D (SEQ ID NO:62). pMBIS é também descrito nas Publicações de Patente dos Estados Unidos Nos. 2003/0148479; e 2004/0005678; e compreende sequências de nucleotídeo que codificam mevalonato kinase, fosfomevalonato kinase, mevalonato pirofosfato decarboxilase, IPP isomerase, e FPP sintase. Nestes estudos, *E. coli* foi transformada com três plasmídeos de expressão: (1) pMBIS, (2) CadS, e (3) CadH/CPR. 20 mM de mevalonato foi adicionado sob indução. A adição de pMBIS aumentou a produção de CadOH em 74 vezes comparada àquela produzida por células sem pMBIS (Figura 3). Novamente, o controle negativo (nenhum CadH) não mostrou formação de produto. Estes resultados indicam que a circulação de P450 pode estar limitada pela produção de substrato *in vivo* (por exemplo, em uma célula viva em cultura *in vitro*). Estas células podem também serem desenvolvidas em um

meio aquoso/orgânico misturado; a cepa foi desenvolvida e induzida na presença de uma sobreposição de dodecano sem alterar significativamente a produtividade de CadOH (~2 vezes menos).

Figura 3. Traço de GC-MS de camada orgânica extraída de *E. coli* alimentada de mevalonato expressando trajetória biosintética de CadOH, bem como pMBIS. Cad e CadOH são indicados no traço.

Foi adicionalmente mostrado que a produtividade é aumentada pelo manejo do P450 sem perda de especificidade de produto. A produção *in vivo* com o gene nativo (nCadH), versus um gene sintético (sCadH) com uso de códon otimizado para expressão em *E. coli*, foi comparada (Figura 4B). Esta comparação indica que o gene sintético se realiza levemente melhor do que o gene nativo.

O domínio N-terminal tipo selvagem de transmembrana (TM) foi substituído com seqüências que são conhecidas para funcionarem em *E. coli* (Figura 4A). Entre as seqüências N-terminais testadas estavam dois condutores P450 N-terminais derivados de *C. tropicalis* - CYP52A13 (A13) que não contém domínio de TM prognosticado, e CYP52A17 (A17) que não contém um domínio de TM [7] – bem como um condutor microssomal bovino (bovino) [8].

O domínio de TM tipo selvagem foi removido totalmente (truncado), e foi substituído com uma etiqueta de secreção (OmpA), domínio de solubilização (PD1) [9], ou uma proteína de inserção de membrana (mistic) [10]. O CadH bovino realizou o CadOH tipo selvagem por aproximadamente 2 vezes, produzindo ~30 mg L⁻¹ (Figura 4B).

Referências

1. M. Sono, M. P. Roach, E. D. Coulter, and J. H. Dawson, *Chem. Rev.* **1996**, *96*, 2841-2887.
2. S. Jennewein, R. M. Long, R. M. Williams, and R. Croteau, *Chem. Biol.* **2004**, *11*, 379–387.
3. R. J. Sowden, S. Yasmin, N. H. Rees, S. G. Bell and L.-L. Wong, *Org. Biomol. Chem.* **2005**, *3*, 57-64.
4. P. Luo, Y.-H. Wang, G.-D. Wang, M. Essenberg, and X.-Y. Chen,

- Plant J.* **2001**, *28*, 95-104.
5. O. A. Carter, R. J. Peters, and R. Croteau, *Phytochem.* **2003**, *64*, 425–433.
 6. V. J. J. Martin, D. J. Pitera, S. T. Withers, J. D. Newman, and J. D. Keasling, *Nature Biotech.* **2003**, *21*, 796-801
 7. D. L. Craft, K. M. Madduri, M. Eshoo, and C. R. Wilson, *Appl. Environ. Microbiol.* **2003**, *69*, 5983–5991.
 8. H. J. Barnes, M. P. Arlotto, and M. R. Watermoan, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5597-5601.
 - 10 9. G. A. Schock, R. Attias, M. Belghazi, P. M. Dansette, and D. Werck-Reichert, *Plant Physiol.* **2003**, *133*, 1198-1208.
 10. T. P. Roosild, J. Greenwald, M. Vega, S. Castronovo, R. Riek, and S. Choe *Science* **2005**, *307*, 1317-1321.

Este exemplo descreve a oxidação *in vivo* (por exemplo, em uma célula viva em cultura de célula) de amorfadieno por amorfadieno oxidase (AMO), também denominado CYP71AV1, isolado de *Artemisia annua*. Várias construções compreendendo uma sequência de nucleotídeo que codifica AMO foram geradas e testadas de modo a otimizar a produção de produto oxidado. A Figura 22 representa esquematicamente as várias construções de AMO. (1) nAMO, sequência de AMO nativa conforme isolada de *A. annua*. (2) sAMO, gene de AMO sintético de códon otimizado para expressão em *E. coli*. (3) A13-AMO, gene de AMO sintético com transmembrana tipo selvagem substituída com a sequência A13 N-terminal de *C. tropicalis*. (4) A17-AMO, gene de AMO sintético com Transmenbrana tipo selvagem substituída com a sequência A17 N-terminal de *C. tropicalis*. (5) Bov-AMO, gene de AMO sintético com transmembrana tipo selvagem substituída com sequência N-terminal microsomal bovina. Os nucleotídeo e sequência de amino ácidos de várias construções são representados nas Figuras 24-31.

As várias construções de ANO foram co-expressas com: a) um CPR; b) amorfadieno sintase (ADS); e c) plasmídeo pMBIS. Na presença de mevalonato, amorfadieno foi observado para ser oxidado na posição C-12 ao álcool correspondente. A Figura 23A mostra a quantidade relativa de álcool

artemisínico produzida *in vivo*. A comparação a um padrão autêntico de álcool artemisínico confirma a identidade do produto (Figura 23A, painel de fundo e Figura 23B).

Figuras 23A e 23B. Oxidação *in vivo* de amorfadieno em *E. coli* por várias construções de AMO. (A) Traço de GC-MS mostrando produção de álcool artemisínico produzido em *E. coli* (painel superior) por sAMO, A13-AMO, A17-AMO, e AMO bovino, comparado ao padrão autêntico (painel de fundo). (B) EI-MS do álcool artemisínico produzido em *E. coli* (painel superior) comparado ao padrão autêntico (painel inferior).

Exemplo 3: Oxidação de substrato em células expressando a trajetória total de mevalonato

A oxidação de substrato foi também efetuada em células expressando a trajetória total de mevalonato de acetil-CoA. O exemplo seguinte para produção de CadOH utilizou 3 plasmídeos: (1) pMevT contendo AtoB, HMGR, e HMGS, (2) pMBIS (contendo seqüências de nucleotídeo que codificam MK, PMK, PMD, IDI (IPP isomerase), e IspA (FPP sintase)), e (3) um vetor de expressão contendo CadH, CPR, e CadS. As células foram cultivadas a 20 °C em TB glicerol com a adição do suplemento de "heme", ácido δ -aminolevulínico. As células produziram CadOH até titulação de 60 mg/L. Os dados são mostrados na Figura 32.

Em um segundo exemplo, ácido artemisínico foi produzido usando 2 plasmídeos: (1) um vetor de expressão contendo seqüências de nucleotídeo que codificam o MevT (AtoB, HMGR, e HMGS) (ver Figuras 35A and B), MBIS (MK, PMK, PMD, IDI, e IspA), e ADS operons, e (2) um vetor de expressão contendo seqüências de nucleotídeo que codificam AMO e um co-participante redox de CPR de *A. annua* (AACPR). Após cultura das células de *E. coli* a 20 °C em TB glicerol com a adição do suplemento de "heme", quantidades de traço de ácido artemisínico foram observadas usando um vetor baseado em promotor T7 (Figura 33). Após mudança do vetor para pCWOri, AMO pode ser usado para as 3 etapas de oxidação de amorfadieno para produzir ácido artemisínico em *E. coli* em titulações de 20 mg/L (Figura 33). Em adição, oxidação gradual do álcool a aldeído foi observada com o

aldeído produzido em titulações de 40-80 mg/L (Figura 34).

Conquanto a presente invenção tenha sido descrita com referência às concretizações específicas desta, deve ser compreendido por aqueles técnicos no assunto que várias mudanças podem ser feitas e equivalentes podem ser substituídos sem fugir do espírito verdadeiro e escopo da invenção. Em adição, muitas modificações podem ser feitas para se adaptar a uma situação particular, material, composição de matéria, processo, etapa ou etapas de processo, ao objetivo, espírito e escopo da presente invenção. Todas tais modificações são pretendidas para estarem dentro do escopo das reivindicações em anexo.

Listagem de Sequência

<110> CHANG, MICHELLE CHIA-YU

EACHUS, RACHEL

RO, DAE-KYUN

YOSHIKUNI, YASUO

KEASLING, JAY D.

<120> Ácidos nucleicos que codificam enzima modificada de citocromo p450 e métodos de uso dos mesmos

<130> BERK-053WO

<150> 60/724,525

<151> 2005-10-07

<150> 60/762,700

<151> 2006-01-27

<160> 63

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 19

<212> PRT

<213> Sequência Artificial

<220>

<223> Peptídeo sintético

<400> 1

Met Trp Leu Leu Leu Ile Ala Val Phe Leu Leu Thr Leu Ala Tyr Leu

1 5 10 15

Phe Trp Pro

<210> 2

<211> 20

<212> PRT

<213> Sequência Artificial

<220>

<223> Peptídeo sintético

<400> 2

Met Ala Leu Leu Leu Ala Val Phe Leu Gly Leu Ser Cys Leu Leu Leu

1 5 10 15

Leu Ser Leu Trp

20

<210> 3

<211> 18
 <212> PRT
 <213> Sequência Artificial

<220>
 <223> Peptídeo sintético

<400> 3
 Met Ala Ile Leu Ala Ala Ile Phe Ala Leu Val Val Ala Thr Ala Thr
 1 5 10 15
 Arg Val

<210> 4
 <211> 24
 <212> PRT
 <213> Sequência Artificial

<220>
 <223> Peptídeo sintético

<400> 4
 Met Asp Ala Ser Leu Leu Leu Ser Val Ala Leu Ala Val Val Leu Ile
 1 5 10 15
 Pro Leu Ser Leu Ala Leu Leu Asn
 20

<210> 5
 <211> 27
 <212> PRT
 <213> Sequência Artificial

<220>
 <223> Peptídeo sintético

<400> 5
 Met Ile Glu Gln Leu Leu Glu Tyr Trp Tyr Val Val Val Pro Val Leu
 1 5 10 15
 Tyr Ile Ile Lys Gln Leu Leu Ala Tyr Thr Lys
 20 25

<210> 6
 <211> 21
 <212> PRT
 <213> Sequência Artificial

<220>
 <223> Peptídeo sintético

<400> 6
 Met Lys Lys Thr Ala Ile Ala Ile Ala Val Ala Leu Ala Gly Phe Ala
 1 5 10 15
 Thr Val Ala Gln Ala
 20

<210> 7
 <211> 21
 <212> PRT
 <213> Sequência Artificial

<220>
 <223> Peptídeo sintético

<400> 7

Met Lys Lys Thr Ala Ile Ala Ile Val Val Ala Leu Ala Gly Phe Ala
 1 5 10 15
 Thr Val Ala Gln Ala
 20

<210> 8

<211> 21

<212> PRT

<213> Sequência Artificial

<220>

<223> Peptídeo sintético

<400> 8

Met Lys Lys Thr Ala Leu Ala Leu Ala Val Ala Leu Ala Gly Phe Ala
 1 5 10 15
 Thr Val Ala Gln Ala
 20

<210> 9

<211> 26

<212> PRT

<213> Sequência Artificial

<220>

<223> Peptídeo sintético

<400> 9

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
 1 5 10 15
 Thr Met Met Phe Ser Ala Ser Ala Leu Ala
 20 25

<210> 10

<211> 25

<212> PRT

<213> Sequência Artificial

<220>

<223> Peptídeo sintético

<400> 10

Met Asn Met Lys Lys Leu Ala Thr Leu Val Ser Ala Val Ala Leu Ser
 1 5 10 15
 Ala Thr Val Ser Ala Asn Ala Met Ala
 20 25

<210> 11

<211> 21

<212> PRT

<213> Sequência Artificial

<220>

<223> Peptídeo sintético

<400> 11

Met Lys Gln Ser Thr Ile Ala Leu Ala Leu Leu Pro Leu Leu Phe Thr
 1 5 10 15
 Pro Val Thr Lys Ala
 20

<210> 12
 <211> 44
 <212> PRT
 <213> Sequência Artificial

<220>
 <223> Peptídeo sintético

<400> 12
 Met Lys Lys Thr Ala Ile Ala Ile Ala Val Ala Leu Ala Gly Phe Ala
 1 5 10 15
 Thr Val Ala Gln Ala Leu Leu Glu Tyr Trp Tyr Val Val Val Pro Val
 20 25 30
 Leu Tyr Ile Ile Lys Gln Leu Leu Ala Tyr Thr Lys
 35 40

<210> 13
 <211> 24
 <212> PRT
 <213> Sequência Artificial

<220>
 <223> Peptídeo sintético

<400> 13
 Glu Glu Leu Leu Lys Gln Ala Leu Gln Gln Ala Gln Gln Leu Leu Gln
 1 5 10 15
 Gln Ala Gln Glu Leu Ala Lys Lys
 20

<210> 14
 <211> 32
 <212> PRT
 <213> Sequência Artificial

<220>
 <223> Peptídeo sintético

<400> 14
 Met Thr Val His Asp Ile Ile Ala Thr Tyr Phe Thr Lys Trp Tyr Val
 1 5 10 15
 Ile Val Pro Leu Ala Leu Ile Ala Tyr Arg Val Leu Asp Tyr Phe Tyr
 20 25 30

<210> 15
 <211> 29
 <212> PRT
 <213> Sequência Artificial

<220>
 <223> Peptídeo sintético

<400> 15
 Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly Gly Trp Thr Gly
 1 5 10 15
 Met Ile Asp Gly Trp Tyr Gly Tyr Gly Gly Gly Lys Lys
 20 25

<210> 16
 <211> 9
 <212> PRT
 <213> Sequência Artificial

<220>

<223> Peptídeo sintético

<400> 16

Met Ala Lys Lys Thr Ser Ser Lys Gly

1

5

<210> 17

<211> 6

<212> PRT

<213> Sequência Artificial

<220>

<223> Peptídeo sintético

<400> 17

Ala Ala Ala Gly Gly Met

1

5

<210> 18

<211> 14

<212> PRT

<213> Sequência Artificial

<220>

<223> Peptídeo sintético

<400> 18

Ala Ala Ala Gly Gly Met Pro Pro Ala Ala Ala Gly Gly Met

1

5

10

<210> 19

<211> 6

<212> PRT

<213> Sequência Artificial

<220>

<223> linker peptide

<400> 19

Ala Ala Ala Gly Gly Met

1

5

<210> 20

<211> 8

<212> PRT

<213> Sequência Artificial

<220>

<223> linker peptide

<400> 20

Pro Pro Ala Ala Ala Gly Gly Met

1

5

<210> 21

<211> 4

<212> PRT

<213> Sequência Artificial

<220>

<223> linker peptide

<400> 21

Ile Glu Gly Arg

1

<210> 22

<211> 6

<212> PRT

<213> Sequência Artificial

<220>

<223> linker peptide

<400> 22

Gly Gly Lys Gly Gly Lys

1

5

<210> 23

<211> 23

<212> PRT

<213> Sequência Artificial

<220>

<223> Peptídeo sintético

<400> 23

Met Leu Phe Pro Val Ala Leu Ser Phe Leu Val Ala Ile Leu Gly Ile

1

10

15

Ser Leu Trp His Val Trp Thr

20

<210> 24

<211> 110

<212> PRT

<213> Bacillus subtilis

<400> 24

Met Phe Cys Thr Phe Phe Glu Lys His His Arg Lys Trp Asp Ile Leu

1

5

10

15

Leu Glu Lys Ser Thr Gly Val Met Glu Ala Met Lys Val Thr Ser Glu

20

25

30

Glu Lys Glu Gln Leu Ser Thr Ala Ile Asp Arg Met Asn Glu Gly Leu

35

40

45

Asp Ala Phe Ile Gln Leu Tyr Asn Glu Ser Glu Ile Asp Glu Pro Leu

50

55

60

Ile Gln Leu Asp Asp Asp Thr Ala Glu Leu Met Lys Gln Ala Arg Asp

65

70

75

80

Met Tyr Gly Gln Glu Lys Leu Asn Glu Lys Leu Asn Thr Ile Ile Lys

85

90

95

Gln Ile Leu Ser Ile Ser Val Ser Glu Glu Gly Glu Lys Glu

100

105

110

<210> 25

<211> 496

<212> PRT

<213> Mentha x gracilis

<400> 25

Met Glu Leu Asp Leu Leu Ser Ala Ile Ile Ile Leu Val Ala Thr Tyr

1

5

10

15

Ile Val Ser Leu Leu Ile Asn Gln Trp Arg Lys Ser Lys Ser Gln Gln

20

25

30

Asn Leu Pro Pro Ser Pro Pro Lys Leu Pro Val Ile Gly His Leu His

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

<210> 26

<211> 473

<212> PRT

<213> Nicotiana tabacum

<400> 26

Met Gln Phe Phe Ser Leu Val Ser Ile Phe Leu Phe Leu Ala Phe Leu
1 5 10 15

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

<210> 27

<211> 536

<212> PRT

<213> *Gossypium arboreum*

<400> 27

Met Leu Gln Ile Ala Phe Ser Ser Tyr Ser Trp Leu Leu Thr Ala Ser

1	5	10	15
Asn Gln Lys Asp Gly Met Leu Phe Pro Val Ala Leu Ser Phe Leu Val			
20	25	30	
Ala Ile Leu Gly Ile Ser Leu Trp His Val Trp Thr Ile Arg Lys Pro			
35	40	45	
Lys Lys Asp Ile Ala Pro Leu Pro Pro Gly Pro Arg Gly Leu Pro Ile			
50	55	60	
Val Gly Tyr Leu Pro Tyr Leu Gly Thr Asp Asn Leu His Leu Val Phe			
65	70	75	80
Thr Asp Leu Ala Ala Tyr Gly Pro Ile Tyr Lys Leu Trp Leu Gly			
85	90	95	
Asn Lys Leu Cys Val Val Ile Ser Ser Ala Pro Leu Ala Lys Glu Val			
100	105	110	
Val Arg Asp Asn Asp Ile Thr Phe Ser Glu Arg Asp Pro Pro Val Cys			
115	120	125	
Ala Lys Ile Ile Thr Phe Gly Leu Asn Asp Ile Val Phe Asp Ser Tyr			
130	135	140	
Ser Ser Pro Asp Trp Arg Met Lys Arg Lys Val Leu Val Arg Glu Met			
145	150	155	160
Leu Ser His Ser Ser Ile Lys Ala Cys Tyr Gly Leu Arg Arg Glu Gln			
165	170	175	
Val Leu Lys Gly Val Gln Asn Val Ala Gln Ser Ala Gly Lys Pro Ile			
180	185	190	
Asp Phe Gly Glu Thr Ala Phe Leu Thr Ser Ile Asn Ala Met Met Ser			
195	200	205	
Met Leu Trp Gly Gly Lys Gln Gly Gly Glu Arg Lys Gly Ala Asp Val			
210	215	220	
Trp Gly Gln Phe Arg Asp Leu Ile Thr Glu Leu Met Val Ile Leu Gly			
225	230	235	240
Lys Pro Asn Val Ser Asp Ile Phe Pro Val Leu Ala Arg Phe Asp Ile			
245	250	255	
Gln Gly Leu Glu Lys Glu Met Thr Lys Ile Val Asn Ser Phe Asp Lys			
260	265	270	
Leu Phe Asn Ser Met Ile Glu Glu Arg Glu Asn Phe Ser Asn Lys Leu			
275	280	285	
Ser Lys Glu Asp Gly Asn Thr Glu Thr Lys Asp Phe Leu Gln Leu Leu			
290	295	300	
Leu Asp Leu Lys Gln Lys Asn Asp Ser Gly Ile Ser Ile Thr Met Asn			
305	310	315	320
Gln Val Lys Ala Leu Leu Met Asp Ile Val Val Gly Gly Thr Asp Thr			
325	330	335	
Thr Ser Thr Met Met Glu Trp Thr Met Ala Glu Leu Ile Ala Asn Pro			
340	345	350	
Glu Ala Met Lys Lys Val Lys Gln Glu Ile Asp Asp Val Val Gly Ser			
355	360	365	
Asp Gly Ala Val Asp Glu Thr His Leu Pro Lys Leu Arg Tyr Leu Asp			
370	375	380	
Ala Ala Val Lys Glu Thr Phe Arg Leu His Pro Pro Met Pro Leu Leu			
385	390	395	400
Val Pro Arg Cys Pro Gly Asp Ser Ser Asn Val Gly Gly Tyr Ser Val			
405	410	415	
Pro Lys Gly Thr Arg Val Phe Leu Asn Ile Trp Cys Ile Gln Arg Asp			
420	425	430	
Pro Gln Leu Trp Glu Asn Pro Leu Glu Phe Lys Pro Glu Arg Phe Leu			
435	440	445	
Thr Asp His Glu Lys Leu Asp Tyr Leu Gly Asn Asp Ser Arg Tyr Met			
450	455	460	
Pro Phe Gly Ser Gly Arg Met Cys Ala Gly Val Ser Leu Gly Glu			
465	470	475	480
Lys Met Leu Tyr Ser Ser Leu Ala Ala Met Ile His Ala Tyr Asp Trp			
485	490	495	
Asn Leu Ala Asp Gly Glu Glu Asn Asp Leu Ile Gly Leu Phe Gly Ile			
500	505	510	
Ile Met Lys Lys Lys Pro Leu Ile Leu Val Pro Thr Pro Arg Pro			
515	520	525	
Ser Asn Leu Gln His Tyr Met Lys			
530	535		

<210> 28
 <211> 512
 <212> PRT
 <213> Sequência Artificial

 <220>
 <223> Modified P450 monooxygenase

<400> 28
 Met Ala Leu Leu Leu Ala Val Phe Leu Gly Leu Ser Cys Leu Leu Leu
 1 5 10 15
 Leu Ser Leu Trp Ile Arg Lys Pro Lys Lys Asp Ile Ala Pro Leu Pro
 20 25 30
 Pro Gly Pro Arg Gly Leu Pro Ile Val Gly Tyr Leu Pro Tyr Leu Gly
 35 40 45
 Thr Asp Asn Leu His Leu Val Phe Thr Asp Leu Ala Ala Tyr Gly
 50 55 60
 Pro Ile Tyr Lys Leu Trp Leu Gly Asn Lys Leu Cys Val Val Ile Ser
 65 70 75 80
 Ser Ala Pro Leu Ala Lys Glu Val Val Arg Asp Asn Asp Ile Thr Phe
 85 90 95
 Ser Glu Arg Asp Pro Pro Val Cys Ala Lys Ile Ile Thr Phe Gly Leu
 100 105 110
 Asn Asp Ile Val Phe Asp Ser Tyr Ser Ser Pro Asp Trp Arg Met Lys
 115 120 125
 Arg Lys Val Leu Val Arg Glu Met Leu Ser His Ser Ser Ile Lys Ala
 130 135 140
 Cys Tyr Gly Leu Arg Arg Glu Gln Val Leu Lys Gly Val Gln Asn Val
 145 150 155 160
 Ala Gln Ser Ala Gly Lys Pro Ile Asp Phe Gly Glu Thr Ala Phe Leu
 165 170 175
 Thr Ser Ile Asn Ala Met Met Ser Met Leu Trp Gly Gly Lys Gln Gly
 180 185 190
 Gly Glu Arg Lys Gly Ala Asp Val Trp Gly Gln Phe Arg Asp Leu Ile
 195 200 205
 Thr Glu Leu Met Val Ile Leu Gly Lys Pro Asn Val Ser Asp Ile Phe
 210 215 220
 Pro Val Leu Ala Arg Phe Asp Ile Gln Gly Leu Glu Lys Glu Met Thr
 225 230 235 240
 Lys Ile Val Asn Ser Phe Asp Lys Leu Phe Asn Ser Met Ile Glu Glu
 245 250 255
 Arg Glu Asn Phe Ser Asn Lys Leu Ser Lys Glu Asp Gly Asn Thr Glu
 260 265 270
 Thr Lys Asp Phe Leu Gln Leu Leu Leu Asp Leu Lys Gln Lys Asn Asp
 275 280 285
 Ser Gly Ile Ser Ile Thr Met Asn Gln Val Lys Ala Leu Leu Met Asp
 290 295 300
 Ile Val Val Gly Gly Thr Asp Thr Thr Ser Thr Met Met Glu Trp Thr
 305 310 315 320
 Met Ala Glu Leu Ile Ala Asn Pro Glu Ala Met Lys Lys Val Lys Gln
 325 330 335
 Glu Ile Asp Asp Val Val Gly Ser Asp Gly Ala Val Asp Glu Thr His
 340 345 350
 Leu Pro Lys Leu Arg Tyr Leu Asp Ala Ala Val Lys Glu Thr Phe Arg
 355 360 365
 Leu His Pro Pro Met Pro Leu Leu Val Pro Arg Cys Pro Gly Asp Ser
 370 375 380
 Ser Asn Val Gly Gly Tyr Ser Val Pro Lys Gly Thr Arg Val Phe Leu
 385 390 395 400
 Asn Ile Trp Cys Ile Gln Arg Asp Pro Gln Leu Trp Glu Asn Pro Leu
 405 410 415
 Glu Phe Lys Pro Glu Arg Phe Leu Thr Asp His Glu Lys Leu Asp Tyr
 420 425 430
 Leu Gly Asn Asp Ser Arg Tyr Met Pro Phe Gly Ser Gly Arg Arg Met
 435 440 445

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

<210> 29

<211> 516

<212> PRT

<213> Sequência Artificial

<220>

<223> Modified P450 monooxygenase

<400> 29

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

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

<210> 30

<211> 536

<212> PRT

<213> Sequência Artificial

<220>

<223> Modified P450 monooxygenase

<400> 30

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

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

<210> 31
 <211> 502
 <212> PRT
 <213> *Taxus cuspidata*

<400> 31
 Met Asp Ala Leu Tyr Lys Ser Thr Val Ala Lys Phe Asn Glu Val Thr
 1 5 10 15
 Gln Leu Asp Cys Ser Thr Glu Ser Phe Ser Ile Ala Leu Ser Ala Ile
 20 25 30
 Ala Gly Ile Leu Leu Leu Leu Leu Phe Arg Ser Lys Arg His Ser
 35 40 45
 Ser Leu Lys Leu Pro Pro Gly Lys Leu Gly Ile Pro Phe Ile Gly Glu
 50 55 60
 Ser Phe Ile Phe Leu Arg Ala Leu Arg Ser Asn Ser Leu Glu Gln Phe
 65 70 75 80
 Phe Asp Glu Arg Val Lys Lys Phe Gly Leu Val Phe Lys Thr Ser Leu
 85 90 95
 Ile Gly His Pro Thr Val Val Leu Cys Gly Pro Ala Gly Asn Arg Leu
 100 105 110
 Ile Leu Ser Asn Glu Glu Lys Leu Val Gln Met Ser Trp Pro Ala Gln
 115 120 125
 Phe Met Lys Leu Met Gly Glu Asn Ser Val Ala Thr Arg Arg Gly Glu
 130 135 140
 Asp His Ile Val Met Arg Ser Ala Leu Ala Gly Phe Phe Gly Pro Gly
 145 150 155 160
 Ala Leu Gln Ser Tyr Ile Gly Lys Met Asn Thr Glu Ile Gln Ser His
 165 170 175
 Ile Asn Glu Lys Trp Lys Gly Lys Asp Glu Val Asn Val Leu Pro Leu
 180 185 190
 Val Arg Glu Leu Val Phe Asn Ile Ser Ala Ile Leu Phe Phe Asn Ile
 195 200 205
 Tyr Asp Lys Gln Glu Gln Asp Arg Leu His Lys Leu Leu Glu Thr Ile
 210 215 220
 Leu Val Gly Ser Phe Ala Leu Pro Ile Asp Leu Pro Gly Phe Gly Phe

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

<210> 32
 <211> 502
 <212> PRT
 <213> Sequência Artificial

<220>
 <223> Modified taxadiene-5a-hydroxylase

<400> 32
 Met Asp Ala Leu Tyr Lys Ser Thr Val Ala Lys Phe Asn Glu Val Thr
 1 5 10 15
 Gln Leu Asp Cys Ser Thr Glu Ser Phe Ser Ile Ala Leu Ser Ser Ile
 20 25 30
 Ala Gly Ile Leu Leu Leu Leu Leu Phe Arg Ser Lys Arg His Ser
 35 40 45
 Ser Leu Lys Leu Pro Pro Gly Lys Leu Gly Ile Pro Phe Ile Gly Glu
 50 55 60
 Ser Phe Ile Phe Leu Arg Ala Leu Arg Ser Asn Ser Leu Glu Gln Phe
 65 70 75 80
 Phe Asp Glu Arg Val Lys Lys Phe Gly Leu Val Phe Lys Thr Ser Leu
 85 90 95
 Ile Gly His Pro Thr Val Val Leu Cys Gly Pro Ala Gly Asn Arg Leu
 100 105 110
 Ile Leu Ser Asn Glu Glu Lys Leu Val Gln Met Ser Trp Pro Ala Gln
 115 120 125
 Phe Met Lys Leu Met Gly Glu Asn Ser Val Ala Thr Arg Arg Gly Glu
 130 135 140
 Asp His Ile Val Met Arg Ser Ala Leu Ala Gly Phe Phe Gly Pro Gly
 145 150 155 160
 Ala Leu Gln Ser Tyr Ile Gly Lys Met Asn Thr Glu Ile Gln Asn His

```

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

```

```

<210> 33
<211> 509
<212> PRT
<213> Arabidopsis thaliana

```

```

<400> 33
Met Ala Phe Phe Ser Met Ile Ser Ile Leu Leu Gly Phe Val Ile Ser
  1      5      10      15
Ser Phe Ile Phe Ile Phe Phe Phe Lys Lys Leu Leu Ser Phe Ser Arg
      20      25      30
Lys Asn Met Ser Glu Val Ser Thr Leu Pro Ser Val Pro Val Val Pro
      35      40      45
Gly Phe Pro Val Ile Gly Asn Leu Leu Gln Leu Lys Glu Lys Lys Pro
      50      55      60
His Lys Thr Phe Thr Arg Trp Ser Glu Ile Tyr Gly Pro Ile Tyr Ser
      65      70      75      80
Ile Lys Met Gly Ser Ser Ser Leu Ile Val Leu Asn Ser Thr Glu Thr
      85      90      95
Ala Lys Glu Ala Met Val Thr Arg Phe Ser Ser Ile Ser Thr Arg Lys
      100      105      110
Leu Ser Asn Ala Leu Thr Val Leu Thr Cys Asp Lys Ser Met Val Ala
      115      120      125

```

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

<210> 34

<211> 1896

<212> DNA

<213> *Gossypium arboreum*

<400> 34

ccacttcgca gcaatattat tgcagttcct gggttggtac ctctgagttt tcaacttaaa 60
 atttcttggt ttctctcaag aaggaagaag atgttgcaaa tagctttcag ctcgtattca 120
 tggctgttga ctgctagcaa ccagaaagat ggaatgttgt tcccagtagc ttgttcattt 180
 ttggtagcca tattgggaat ttcactgtgg cacgtatgga ccataaggaa gccaaagaaa 240
 gacatcgccc cattaccgcc gggtcctcgt ggggttgccaa tagtgggata tcttccatat 300
 cttggaactg ataactttca cttggtgttt acagatttgg ctgcagctta cgggtcccatc 360
 tacaagcttt ggctaggaaa caaattatgc gtagtcatta gtcggcacc actggcgaaa 420
 gaagtgggtc gtgacaacga catcacattt tctgaaaggg atcctcccggt ttgtgcaaaag 480
 attattacct ttggcctcaa tgatattgta tttgattctt acagtagtcc agattggaga 540
 atgaagagaa aagtgtgtgt acgtgaaatg cttagccata gtagcattaa agcttggtat 600
 ggtctaagga gggaacaagt gcttaaaggc gtacaaaatg ttgctcaaag tgctggcaag 660

```

ccaattgatt ttggtgaaac ggcattttta acatcaatca atgcatgat gagcatgctg 720
tggggtggca aacagggagg agagcggaaa ggggcccacg tttggggcca atttcgagat 780
ctcataaccg aactaatggt gatacttggg aaaccaaacg tttctgatat tttcccgggtg 840
cttgcaaggt ttgacataca gggattggag aaggaaatga ctaaaatcgt taattctttc 900
gataagcttt tcaactccat gattgaagaa agagagaact ttagcaacaa attgagcaaa 960
gaagatggaa acactgaaac aaaagacttc ttgcagcttc tgttgacct caagcagaag 1020
aacgataagc gaatatcgat aacaatgaat caagtcaagg ccttgctcat ggacattgtg 1080
gtcgggtggaa ctgatacaac atcaaccatg atggaatgga caatggctga actaattgca 1140
aatcctgaag caatgaaaaa ggtgaagcaa gaaatagacg atgttgtcgg ttcggatggc 1200
gccgtcgatg agactcaact gcctaagttg cgctatctag atgctgcagt aaaggagacc 1260
ttccgattgc acccaccgat gccactcctt gtaccccggt gcccgggcga ctcaagcaac 1320
gttggtggct atagcgtacc aaaggcacc agggctctct taaacatttg gtgtattcag 1380
agggatccac agctttggga aaatccttta gaattcaagc ctgagagggt cttgatgat 1440
catgagaagc tcgattattt aggaaaacgat tcccggtaaca tgccgttttg ttctggaagg 1500
agaatgtgtg ccggagtatc tctcggtgaa aagatgttgt attcctcctt ggcagcaatg 1560
atccatgctt atgattggaa cttggccgac ggtgaagaaa atgacttgat tggcttattt 1620
ggaattatta tgaagaaaaa gaagccttta attcttgttc ctacaccaag accatcaaatt 1680
ctccagcact atatgaagta actttactat tgtatttctt ttataccact ttattgcctc 1740
tttgtcatgt ttaggcaaca attctaagta ataagtttg ctatatggtg aacaataatg 1800
tgtttattat acatcataag caatgagctc ttcccacc tagggcaata caatgatact 1860
gtgtattaag tgaaatcaac aaatctttta ttctaa 1896

```

<210> 35

<211> 1677

<212> DNA

<213> Sequência Artificial

<220>

<223> Synthetic polynucleotide

<400> 35

```

atgctgcaga ttgctttttc ttcttattct tggctgctga ccgcttctaa ccagaaagac 60
ggcatgctgt tcccgggtggc gctgagcttc ctgggtggcaa tcctgggcat tagcctgtgg 120
cacgtgtgga ctatccgtaa accgaagaaa gatatcgac cgctgccacc gggctccgctg 180
ggcctgccga tcgttggcta cctgccgtat ctgggcaacc acaacctgca cctggtgttc 240
accgacctgg cagccgcgta cggctccgat taaaaactgt ggctgggcaa taaactgtgc 300
gtagtattct cctctgctcc tctggcgaag gaggtggttc gcgacaacga catcaccttc 360
tccgaacgtg acccaccggt ctgtgctaaa atcatcacct tcggcctgaa cgacatcgta 420
ttcgactcct atagctctcc tgactggcgt atgaaacgta aggttctggt acgcgagatg 480
ctgtcccaca gctccattaa ggcattgctac ggctgcgctc gcgaacaggt actgaaaggc 540
gtacaaaacg tagcgcagtc cgcgggcaaa ccgatcgatt tcggcgaaac ggccttcctg 600
actagatca acgctatgat gtccatgctg tggggtggtg aacagggcgg cgagcgtaaa 660
ggcgcgcgac tatggggcca gtttcgtgac ctgatcaccc aactgatggg gattctgggc 720
aaaccgaacg tcagcgacat cttcccgggt ctggctcgct tcgacatcca gggcctggaa 780
aaagaaatga ccaagatcgt caactcttcc gacaaactgt ttaactccat gatcgaagaa 840
cgcgaaaatt tcttaacaaa actgagcaaa gaagatggca acaccgaaac taaagatttc 900
ctgcagctgc tgcaggacct gaaacaaaag aacgattctg gtatctccat taccatgaac 960
caagtgaag cgctgctgat ggacattggt tgggtggtg ctgacaccac ttctaccatg 1020
atggaatgga cgtatggcga actgattgct aatccggaag cgatgaagaa agtgaacaaa 1080
gaaattgatg atgtagtggg ctctgatggt gcggtagacg agacgcacct gcctaagctg 1140
cgttatctgg acgcagccgt gaaagaaacc ttccgtctgc atccgcctat gccgctgctg 1200
gttccacgtt gccagggcga ttccagcaac gttggtggct atagcgtacc gaagggtacc 1260
cgtgtgttcc tgaatatctg gtgcattcag cgcgaccgcg agctgtggga aaaccgctg 1320
gagttcaaac ctgaacgctt cctgaccgac catgaaaagc tggactacct gggcaacgat 1380
tcccgttaca tgccgttcgg ttctggccgt cgtatgtgcg caggcgtctc cctgggcgag 1440
aaaatgctgt atctagcct ggctgccatg atccacgctt acgactggaa cctggcagat 1500
ggtgaagaga acgacctgat cggcctgttc ggcattatta tgaaaaagaa aaagccgctg 1560
atcctggtgc cgactccggt tccaagcaac ctgcagcact acatgaaact ggtgccgctg 1620
ggctctaag aaaccgctgc tgcaaaattc gaacgtcagc acatggacag ctaataa 1677

```

<210> 36

<211> 717

<212> PRT

<213> *Taxus cuspidata*

<400> 36

Met Gln Ala Asn Ser Asn Thr Val Glu Gly Ala Ser Gln Gly Lys Ser

1	5	10	15
Leu Leu Asp	Ile Ser Arg Leu Asp	His Ile Phe Ala Leu Leu Leu Asn	
	20	25	30
Gly Lys Gly	Gly Asp Leu Gly Ala Met Thr Gly Ser Ala Leu Ile Leu		
	35	40	45
Thr Glu Asn	Ser Gln Asn Leu Met Ile Leu Thr Thr Ala Leu Ala Val		
	50	55	60
Leu Val Ala	Cys Val Phe Phe Phe Val Trp Arg Arg Gly Gly Ser Asp		
65	70	75	80
Thr Gln Lys	Pro Ala Val Arg Pro Thr Pro Leu Val Lys Glu Glu Asp		
	85	90	95
Glu Glu Glu	Glu Asp Asp Ser Ala Lys Lys Val Thr Ile Phe Phe		
	100	105	110
Gly Thr Gln	Thr Gly Thr Ala Glu Gly Phe Ala Lys Ala Leu Ala Glu		
	115	120	125
Glu Ala Lys	Ala Arg Tyr Glu Lys Ala Val Phe Lys Val Val Asp Leu		
	130	135	140
Asp Asn Tyr	Ala Ala Asp Asp Glu Gln Tyr Glu Glu Lys Leu Lys Lys		
145	150	155	160
Glu Lys Leu	Ala Phe Met Leu Ala Thr Tyr Gly Asp Gly Glu Pro		
	165	170	175
Thr Asp Asn	Ala Ala Arg Phe Tyr Lys Trp Phe Leu Glu Gly Lys Glu		
	180	185	190
Arg Glu Pro	Trp Leu Ser Asp Leu Thr Tyr Gly Val Phe Gly Leu Gly		
	195	200	205
Asn Arg Gln	Tyr Glu His Phe Asn Lys Val Ala Lys Ala Val Asp Glu		
	210	215	220
Val Leu Ile	Glu Gln Gly Ala Lys Arg Leu Val Pro Val Gly Leu Gly		
	225	230	235
Asp Asp Asp	Gln Cys Ile Glu Asp Asp Phe Thr Ala Trp Arg Glu Gln		
	245	250	255
Val Trp Pro	Glu Leu Asp Gln Leu Leu Arg Asp Glu Asp Asp Glu Pro		
	260	265	270
Thr Ser Ala	Thr Pro Tyr Thr Ala Ala Ile Pro Glu Tyr Arg Val Glu		
	275	280	285
Ile Tyr Asp	Ser Val Val Ser Val Tyr Glu Glu Thr His Ala Leu Lys		
	290	295	300
Gln Asn Gly	Gln Ala Val Tyr Asp Ile His His Pro Cys Arg Ser Asn		
	305	310	315
Val Ala Val	Arg Arg Glu Leu His Thr Pro Leu Ser Asp Arg Ser Cys		
	325	330	335
Ile His Leu	Glu Phe Asp Ile Ser Asp Thr Gly Leu Ile Tyr Glu Thr		
	340	345	350
Gly Asp His	Val Gly Val His Thr Glu Asn Ser Ile Glu Thr Val Glu		
	355	360	365
Glu Ala Ala	Lys Leu Leu Gly Tyr Gln Leu Asp Thr Ile Phe Ser Val		
	370	375	380
His Gly Asp	Lys Glu Asp Gly Thr Pro Leu Gly Gly Ser Ser Leu Pro		
	385	390	395
Pro Pro Phe	Pro Gly Pro Cys Thr Leu Arg Thr Ala Leu Ala Arg Tyr		
	405	410	415
Ala Asp Leu	Leu Asn Pro Pro Arg Lys Ala Ala Phe Leu Ala Leu Ala		
	420	425	430
Ala His Ala	Ser Asp Pro Ala Glu Ala Glu Arg Leu Lys Phe Leu Ser		
	435	440	445
Ser Pro Ala	Gly Lys Asp Glu Tyr Ser Gln Trp Val Thr Ala Ser Gln		
	450	455	460
Arg Ser Leu	Leu Glu Ile Met Ala Glu Phe Pro Ser Ala Lys Pro Pro		
	465	470	475
Leu Gly Val	Phe Phe Ala Ala Ile Ala Pro Arg Leu Gln Pro Arg Tyr		
	485	490	495
Tyr Ser Ile	Ser Ser Pro Arg Phe Ala Pro Ser Arg Ile His Val		
	500	505	510
Thr Cys Ala	Leu Val Tyr Gly Pro Ser Pro Thr Gly Arg Ile His Lys		
	515	520	525
Gly Val Cys	Ser Asn Trp Met Lys Asn Ser Leu Pro Ser Glu Glu Thr		
	530	535	540

His Asp Cys Ser Trp Ala Pro Val Phe Val Arg Gln Ser Asn Phe Lys
 545 550 555 560
 Leu Pro Ala Asp Ser Thr Thr Pro Ile Val Met Val Gly Pro Gly Thr
 565 570 575
 Gly Phe Ala Pro Phe Arg Gly Phe Leu Gln Glu Arg Ala Lys Leu Gln
 580 585 590
 Glu Ala Gly Glu Lys Leu Gly Pro Ala Val Leu Phe Phe Gly Cys Arg
 595 600 605
 Asn Arg Gln Met Asp Tyr Ile Tyr Glu Asp Glu Leu Lys Gly Tyr Val
 610 615 620
 Glu Lys Gly Ile Leu Thr Asn Leu Ile Val Ala Phe Ser Arg Glu Gly
 625 630 635 640
 Ala Thr Lys Glu Tyr Val Gln His Lys Met Leu Glu Lys Ala Ser Asp
 645 650 655
 Thr Trp Ser Leu Ile Ala Gln Gly Gly Tyr Leu Tyr Val Cys Gly Asp
 660 665 670
 Ala Lys Gly Met Ala Arg Asp Val His Arg Thr Leu His Thr Ile Val
 675 680 685
 Gln Glu Gln Glu Ser Val Asp Ser Ser Lys Ala Glu Phe Leu Val Lys
 690 695 700
 Lys Leu Gln Met Asp Gly Arg Tyr Leu Arg Asp Ile Trp
 705 710 715

<210> 37
 <211> 680
 <212> PRT
 <213> Candida tropicalis

<400> 37
 Met Ala Leu Asp Lys Leu Asp Leu Tyr Val Ile Ile Thr Leu Val Val
 1 5 10 15
 Ala Ile Ala Ala Tyr Phe Ala Lys Asn Gln Phe Leu Asp Gln Gln Gln
 20 25 30
 Asp Thr Gly Phe Leu Asn Thr Asp Ser Gly Asp Gly Asn Ser Arg Asp
 35 40 45
 Ile Leu Gln Ala Leu Lys Lys Asn Asn Lys Asn Thr Leu Leu Leu Phe
 50 55 60
 Gly Ser Gln Thr Gly Thr Ala Glu Asp Tyr Ala Asn Lys Leu Ser Arg
 65 70 75 80
 Glu Leu His Ser Arg Phe Gly Leu Lys Thr Met Val Ala Asp Phe Ala
 85 90 95
 Asp Tyr Asp Phe Glu Asn Phe Gly Asp Ile Thr Glu Asp Ile Leu Val
 100 105 110
 Phe Phe Ile Val Ala Thr Tyr Gly Glu Gly Glu Pro Thr Asp Asn Ala
 115 120 125
 Asp Glu Phe His Thr Trp Leu Thr Glu Glu Ala Asp Thr Leu Ser Thr
 130 135 140
 Leu Lys Tyr Thr Val Phe Gly Leu Gly Asn Ser Thr Tyr Glu Phe Phe
 145 150 155 160
 Asn Ala Ile Gly Arg Lys Phe Asp Arg Leu Leu Gly Glu Lys Gly Gly
 165 170 175
 Asp Arg Phe Ala Glu Tyr Gly Glu Gly Asp Asp Gly Thr Gly Thr Leu
 180 185 190
 Asp Glu Asp Phe Leu Ala Trp Lys Asp Asn Val Phe Asp Ser Leu Lys
 195 200 205
 Asn Asp Leu Asn Phe Glu Glu Lys Glu Leu Lys Tyr Glu Pro Asn Val
 210 215 220
 Lys Leu Thr Glu Arg Asp Asp Leu Ser Gly Asn Asp Pro Asp Val Ser
 225 230 235 240
 Leu Gly Glu Pro Asn Val Lys Tyr Ile Lys Ser Glu Gly Val Asp Leu
 245 250 255
 Thr Lys Gly Pro Phe Asp His Thr His Pro Phe Leu Ala Arg Ile Val
 260 265 270
 Lys Thr Lys Glu Leu Phe Thr Ser Glu Asp Arg His Cys Val His Val
 275 280 285
 Glu Phe Asp Ile Ser Glu Ser Asn Leu Lys Tyr Thr Thr Gly Asp His

```

      290                      295                      300
Leu Ala Ile Trp Pro Ser Asn Ser Asp Glu Asn Ile Lys Gln Phe Ala
305                      310                      315                      320
Lys Cys Phe Gly Leu Glu Asp Lys Leu Asp Thr Val Ile Glu Leu Lys
      325                      330                      335
Ala Leu Asp Ser Thr Tyr Ser Ile Pro Phe Pro Asn Pro Ile Thr Tyr
      340                      345                      350
Gly Ala Val Ile Arg His His Leu Glu Ile Ser Gly Pro Val Ser Arg
      355                      360                      365
Gln Phe Phe Leu Ser Ile Ala Gly Phe Ala Pro Asp Glu Glu Thr Lys
      370                      375                      380
Lys Ser Phe Thr Arg Ile Gly Gly Asp Lys Gln Glu Phe Ala Ser Lys
385                      390                      395                      400
Val Thr Arg Arg Lys Phe Asn Ile Ala Asp Ala Leu Leu Phe Ala Ser
      405                      410                      415
Asn Asn Arg Pro Trp Ser Asp Val Pro Phe Glu Phe Leu Ile Glu Asn
      420                      425                      430
Val Gln His Leu Thr Pro Arg Tyr Tyr Ser Ile Ser Ser Ser Ser Leu
      435                      440                      445
Ser Glu Lys Gln Thr Ile Asn Val Thr Ala Val Val Glu Ala Glu Glu
      450                      455                      460
Glu Ala Asp Gly Arg Pro Val Thr Gly Val Val Thr Asn Leu Leu Lys
465                      470                      475                      480
Asn Ile Glu Ile Glu Gln Asn Lys Thr Gly Glu Thr Pro Met Val His
      485                      490                      495
Tyr Asp Leu Asn Gly Pro Arg Gly Lys Phe Ser Lys Phe Arg Leu Pro
      500                      505                      510
Val His Val Arg Arg Ser Asn Phe Lys Leu Pro Lys Asn Ser Thr Thr
      515                      520                      525
Pro Val Ile Leu Ile Gly Pro Gly Thr Gly Val Ala Pro Leu Arg Gly
      530                      535                      540
Phe Val Arg Glu Arg Val Gln Gln Val Lys Asn Gly Val Asn Val Gly
545                      550                      555                      560
Lys Thr Val Leu Phe Tyr Gly Cys Arg Asn Ser Glu Gln Asp Phe Leu
      565                      570                      575
Tyr Lys Gln Glu Trp Ser Glu Tyr Ala Ser Val Leu Gly Glu Asn Phe
      580                      585                      590
Glu Met Phe Asn Ala Phe Ser Arg Gln Asp Pro Thr Lys Lys Val Tyr
      595                      600                      605
Val Gln Asp Lys Ile Leu Glu Asn Ser Ala Leu Val Asp Glu Leu Leu
      610                      615                      620
Ser Ser Gly Ala Ile Ile Tyr Val Cys Gly Asp Ala Ser Arg Met Ala
625                      630                      635                      640
Arg Asp Val Gln Ala Ala Ile Ala Lys Ile Val Ala Lys Ser Arg Asp
      645                      650                      655
Ile His Glu Asp Lys Ala Ala Glu Leu Val Lys Ser Trp Lys Val Gln
      660                      665                      670
Asn Arg Tyr Gln Glu Asp Val Trp
      675                      680

```

<210> 38

<211> 692

<212> PRT

<213> Arabidopsis thaliana

<400> 38

```

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

```

Gly	Lys	Thr	Arg	Val	Ser	Ile	Phe	Phe	Gly	Thr	Gln	Thr	Gly	Thr	Ala	
				85					90					95		
Glu	Gly	Phe	Ala	Lys	Ala	Leu	Ser	Glu	Glu	Ile	Lys	Ala	Arg	Tyr	Glu	
			100					105					110			
Lys	Ala	Ala	Val	Lys	Val	Ile	Asp	Leu	Asp	Asp	Tyr	Ala	Ala	Asp	Asp	
			115				120					125				
Asp	Gln	Tyr	Glu	Glu	Lys	Leu	Lys	Lys	Glu	Thr	Leu	Ala	Phe	Phe	Cys	
			130			135					140					
Val	Ala	Thr	Tyr	Gly	Asp	Gly	Glu	Pro	Thr	Asp	Asn	Ala	Ala	Arg	Phe	
145					150					155					160	
Ser	Lys	Trp	Phe	Thr	Glu	Glu	Asn	Glu	Arg	Asp	Ile	Lys	Leu	Gln	Gln	
			165					170						175		
Leu	Ala	Tyr	Gly	Val	Phe	Ala	Leu	Gly	Asn	Arg	Gln	Tyr	Glu	His	Phe	
			180					185					190			
Asn	Lys	Ile	Gly	Ile	Val	Leu	Asp	Glu	Glu	Leu	Cys	Lys	Lys	Gly	Ala	
			195				200					205				
Lys	Arg	Leu	Ile	Glu	Val	Gly	Leu	Gly	Asp	Asp	Asp	Gln	Ser	Ile	Glu	
			210			215					220					
Asp	Asp	Phe	Asn	Ala	Trp	Lys	Glu	Ser	Leu	Trp	Ser	Glu	Leu	Asp	Lys	
225					230					235					240	
Leu	Leu	Lys	Asp	Glu	Asp	Asp	Lys	Ser	Val	Ala	Thr	Pro	Tyr	Thr	Ala	
			245					250						255		
Val	Ile	Pro	Glu	Tyr	Arg	Val	Val	Thr	His	Asp	Pro	Arg	Phe	Thr	Thr	
			260					265					270			
Gln	Lys	Ser	Met	Glu	Ser	Asn	Val	Ala	Asn	Gly	Asn	Thr	Thr	Ile	Asp	
			275			280						285				
Ile	His	Pro	Cys	Arg	Val	Asp	Val	Ala	Val	Gln	Lys	Glu	Leu	His		
			290			295				300						
Thr	His	Glu	Ser	Asp	Arg	Ser	Cys	Ile	His	Leu	Glu	Phe	Asp	Ile	Ser	
305					310					315					320	
Arg	Thr	Gly	Ile	Thr	Tyr	Glu	Thr	Gly	Asp	His	Val	Gly	Val	Tyr	Ala	
			325					330						335		
Glu	Asn	His	Val	Glu	Ile	Val	Glu	Glu	Ala	Gly	Lys	Leu	Leu	Gly	His	
			340					345					350			
Ser	Leu	Asp	Leu	Val	Phe	Ser	Ile	His	Ala	Asp	Lys	Glu	Asp	Gly	Ser	
			355				360					365				
Pro	Leu	Glu	Ser	Ala	Val	Pro	Pro	Pro	Phe	Pro	Gly	Pro	Cys	Thr	Leu	
			370			375					380					
Gly	Thr	Gly	Leu	Ala	Arg	Tyr	Ala	Asp	Leu	Leu	Asn	Pro	Pro	Arg	Lys	
385					390					395					400	
Ser	Ala	Leu	Val	Ala	Leu	Ala	Ala	Tyr	Ala	Thr	Glu	Pro	Ser	Glu	Ala	
			405					410						415		
Glu	Lys	Leu	Lys	His	Leu	Thr	Ser	Pro	Asp	Gly	Lys	Asp	Glu	Tyr	Ser	
			420					425					430			
Gln	Trp	Ile	Val	Ala	Ser	Gln	Arg	Ser	Leu	Leu	Glu	Val	Met	Ala	Ala	
			435			440						445				
Phe	Pro	Ser	Ala	Lys	Pro	Pro	Leu	Gly	Val	Phe	Phe	Ala	Ala	Ile	Ala	
			450			455					460					
Pro	Arg	Leu	Gln	Pro	Arg	Tyr	Tyr	Ser	Ile	Ser	Ser	Cys	Gln	Asp	Trp	
465					470					475					480	
Ala	Pro	Ser	Arg	Val	His	Val	Thr	Ser	Ala	Leu	Val	Tyr	Gly	Pro	Thr	
			485					490						495		
Pro	Thr	Gly	Arg	Ile	His	Lys	Gly	Val	Cys	Ser	Thr	Trp	Met	Lys	Asn	
			500					505					510			
Ala	Val	Pro	Ala	Glu	Lys	Ser	His	Glu	Cys	Ser	Gly	Ala	Pro	Ile	Phe	
			515				520					525				
Ile	Arg	Ala	Ser	Asn	Phe	Lys	Leu	Pro	Ser	Asn	Pro	Ser	Thr	Pro	Ile	
			530			535					540					
Val	Met	Val	Gly	Pro	Gly	Thr	Gly	Leu	Ala	Pro	Phe	Arg	Gly	Phe	Leu	
545					550					555					560	
Gln	Glu	Arg	Met	Ala	Leu	Lys	Glu	Asp	Gly	Glu	Glu	Leu	Gly	Ser	Ser	
			565					570						575		
Leu	Leu	Phe	Phe	Gly	Cys	Arg	Asn	Arg	Gln	Met	Asp	Phe	Ile	Tyr	Glu	
			580					585					590			
Asp	Glu	Leu	Asn	Asn	Phe	Val	Asp	Gln	Gly	Val	Ile	Ser	Glu	Leu	Ile	
			595				600					605				
Met	Ala	Phe	Ser	Arg	Glu	Gly	Ala	Gln	Lys	Glu	Tyr	Val	Gln	His	Lys	

610 615 620
 Met Met Glu Lys Ala Ala Gln Val Trp Asp Leu Ile Lys Glu Glu Gly
 625 630 635 640
 Tyr Leu Tyr Val Cys Gly Asp Ala Lys Gly Met Ala Arg Asp Val His
 645 650 655
 Arg Thr Leu His Thr Ile Val Gln Glu Gln Glu Gly Val Ser Ser Ser
 660 665 670
 Glu Ala Glu Ala Ile Val Lys Lys Leu Gln Thr Glu Gly Arg Tyr Leu
 675 680 685
 Arg Asp Val Trp
 690

<210> 39
 <211> 712
 <212> PRT
 <213> Arabidopsis thaliana

<400> 39
 Met Ser Ser Ser Ser Ser Ser Ser Thr Ser Met Ile Asp Leu Met Ala
 1 5 10 15
 Ala Ile Ile Lys Gly Glu Pro Val Ile Val Ser Asp Pro Ala Asn Ala
 20 25 30
 Ser Ala Tyr Glu Ser Val Ala Ala Glu Leu Ser Ser Met Leu Ile Glu
 35 40 45
 Asn Arg Gln Phe Ala Met Ile Val Thr Thr Ser Ile Ala Val Leu Ile
 50 55 60
 Gly Cys Ile Val Met Leu Val Trp Arg Arg Ser Gly Ser Gly Asn Ser
 65 70 75 80
 Lys Arg Val Glu Pro Leu Lys Pro Leu Val Ile Lys Pro Arg Glu Glu
 85 90 95
 Glu Ile Asp Asp Gly Arg Lys Lys Val Thr Ile Phe Phe Gly Thr Gln
 100 105 110
 Thr Gly Thr Ala Glu Gly Phe Ala Lys Ala Leu Gly Glu Glu Ala Lys
 115 120 125
 Ala Arg Tyr Glu Lys Thr Arg Phe Lys Ile Val Asp Leu Asp Asp Tyr
 130 135 140
 Ala Ala Asp Asp Asp Glu Tyr Glu Glu Lys Leu Lys Lys Glu Asp Val
 145 150 155 160
 Ala Phe Phe Phe Leu Ala Thr Tyr Gly Asp Gly Glu Pro Thr Asp Asn
 165 170 175
 Ala Ala Arg Phe Tyr Lys Trp Phe Thr Glu Gly Asn Asp Arg Gly Glu
 180 185 190
 Trp Leu Lys Asn Leu Lys Tyr Gly Val Phe Gly Leu Gly Asn Arg Gln
 195 200 205
 Tyr Glu His Phe Asn Lys Val Ala Lys Val Val Asp Asp Ile Leu Val
 210 215 220
 Glu Gln Gly Ala Gln Arg Leu Val Gln Val Gly Leu Gly Asp Asp Asp
 225 230 235 240
 Gln Cys Ile Glu Asp Asp Phe Thr Ala Trp Arg Glu Ala Leu Trp Pro
 245 250 255
 Glu Leu Asp Thr Ile Leu Arg Glu Glu Gly Asp Thr Ala Val Ala Thr
 260 265 270
 Pro Tyr Thr Ala Ala Val Leu Glu Tyr Arg Val Ser Ile His Asp Ser
 275 280 285
 Glu Asp Ala Lys Phe Asn Asp Ile Thr Leu Ala Asn Gly Asn Gly Tyr
 290 295 300
 Thr Val Phe Asp Ala Gln His Pro Tyr Lys Ala Asn Val Ala Val Lys
 305 310 315 320
 Arg Glu Leu His Thr Pro Glu Ser Asp Arg Ser Cys Ile His Leu Glu
 325 330 335
 Phe Asp Ile Ala Gly Ser Gly Leu Thr Met Lys Leu Gly Asp His Val
 340 345 350
 Gly Val Leu Cys Asp Asn Leu Ser Glu Thr Val Asp Glu Ala Leu Arg
 355 360 365
 Leu Leu Asp Met Ser Pro Asp Thr Tyr Phe Ser Leu His Ala Glu Lys
 370 375 380

Glu Asp Gly Thr Pro Ile Ser Ser Ser Leu Pro Pro Pro Phe Pro Pro
 385 390 395 400
 Cys Asn Leu Arg Thr Ala Leu Thr Arg Tyr Ala Cys Leu Leu Ser Ser
 405 410 415
 Pro Lys Lys Ser Ala Leu Val Ala Leu Ala Ala His Ala Ser Asp Pro
 420 425 430
 Thr Glu Ala Glu Arg Leu Lys His Leu Ala Ser Pro Ala Gly Lys Asp
 435 440 445
 Glu Tyr Ser Lys Trp Val Val Glu Ser Gln Arg Ser Leu Leu Glu Val
 450 455 460
 Met Ala Glu Phe Pro Ser Ala Lys Pro Pro Leu Gly Val Phe Phe Ala
 465 470 475 480
 Gly Val Ala Pro Arg Leu Gln Pro Arg Phe Tyr Ser Ile Ser Ser Ser
 485 490 495
 Pro Lys Ile Ala Glu Thr Arg Ile His Val Thr Cys Ala Leu Val Tyr
 500 505 510
 Glu Lys Met Pro Thr Gly Arg Ile His Lys Gly Val Cys Ser Thr Trp
 515 520 525
 Met Lys Asn Ala Val Pro Tyr Glu Lys Ser Glu Lys Leu Phe Leu Gly
 530 535 540
 Arg Pro Ile Phe Val Arg Gln Ser Asn Phe Lys Leu Pro Ser Asp Ser
 545 550 555 560
 Lys Val Pro Ile Ile Met Ile Gly Pro Gly Thr Gly Leu Ala Pro Phe
 565 570 575
 Arg Gly Phe Leu Gln Glu Arg Leu Ala Leu Val Glu Ser Gly Val Glu
 580 585 590
 Leu Gly Pro Ser Val Leu Phe Phe Gly Cys Arg Asn Arg Arg Met Asp
 595 600 605
 Phe Ile Tyr Glu Glu Glu Leu Gln Arg Phe Val Glu Ser Gly Ala Leu
 610 615 620
 Ala Glu Leu Ser Val Ala Phe Ser Arg Glu Gly Pro Thr Lys Glu Tyr
 625 630 635 640
 Val Gln His Lys Met Met Asp Lys Ala Ser Asp Ile Trp Asn Met Ile
 645 650 655
 Ser Gln Gly Ala Tyr Leu Tyr Val Cys Gly Asp Ala Lys Gly Met Ala
 660 665 670
 Arg Asp Val His Arg Ser Leu His Thr Ile Ala Gln Glu Gln Gly Ser
 675 680 685
 Met Asp Ser Thr Lys Ala Glu Gly Phe Val Lys Asn Leu Gln Thr Ser
 690 695 700
 Gly Arg Tyr Leu Arg Asp Val Trp
 705 710

<210> 40

<211> 667

<212> PRT

<213> Arabidopsis thaliana

<400> 40

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

130	135	140
Asp Arg Gly Glu Trp	Leu Lys Asn Leu Lys Tyr	Gly Val Phe Gly Leu
145	150	155
Gly Asn Arg Gln Tyr	Glu His Phe Asn Lys Val	Ala Lys Val Val Asp
165	170	175
Asp Ile Leu Val Glu	Gln Gly Ala Gln Arg Leu	Val Gln Val Gly Leu
180	185	190
Gly Asp Asp Asp Gln	Cys Ile Glu Asp Asp Phe	Thr Ala Trp Arg Glu
195	200	205
Ala Leu Trp Pro Glu	Leu Asp Thr Ile Leu Arg	Glu Glu Gly Asp Thr
210	215	220
Ala Val Ala Thr Pro	Tyr Thr Ala Ala Val Leu	Glu Tyr Arg Val Ser
225	230	235
Ile His Asp Ser Glu	Asp Ala Lys Phe Asn Asp	Ile Asn Met Ala Asn
245	250	255
Gly Asn Gly Tyr Thr	Val Phe Asp Ala Gln His	Pro Tyr Lys Ala Asn
260	265	270
Val Ala Val Lys Arg	Glu Leu His Thr Pro Glu	Ser Asp Arg Ser Cys
275	280	285
Ile His Leu Glu Phe	Asp Ile Ala Gly Ser Gly	Leu Thr Tyr Glu Thr
290	295	300
Gly Asp His Val Gly	Val Leu Cys Asp Asn Leu	Ser Glu Thr Val Asp
305	310	315
Glu Ala Leu Arg Leu	Leu Asp Met Ser Pro Asp	Thr Tyr Phe Ser Leu
325	330	335
His Ala Glu Lys Glu	Asp Gly Thr Pro Ile Ser	Ser Ser Leu Pro Pro
340	345	350
Pro Phe Pro Pro Cys	Asn Leu Arg Thr Ala Leu	Thr Arg Tyr Ala Cys
355	360	365
Leu Leu Ser Ser Pro	Lys Lys Ser Ala Leu Val	Ala Leu Ala Ala His
370	375	380
Ala Ser Asp Pro Thr	Glu Ala Glu Arg Leu Lys	His Leu Ala Ser Pro
385	390	395
Ala Gly Lys Asp Glu	Tyr Ser Lys Trp Val Val	Glu Ser Gln Arg Ser
405	410	415
Leu Leu Glu Val Met	Ala Glu Phe Pro Ser	Ala Lys Pro Pro Leu Gly
420	425	430
Val Phe Phe Ala Gly	Val Ala Pro Arg Leu Gln	Pro Arg Phe Tyr Ser
435	440	445
Ile Ser Ser Ser Pro	Lys Ile Ala Glu Thr Arg	Ile His Val Thr Cys
450	455	460
Ala Leu Val Tyr Glu	Lys Met Pro Thr Gly	Arg Ile His Lys Gly Val
465	470	475
Cys Ser Thr Trp Met	Lys Asn Ala Val Pro	Tyr Glu Lys Ser Glu Asn
485	490	495
Cys Ser Ser Ala Pro	Ile Phe Val Arg Gln	Ser Asn Phe Lys Leu Pro
500	505	510
Ser Asp Ser Lys Val	Pro Ile Ile Met Ile	Gly Pro Gly Thr Gly Leu
515	520	525
Ala Pro Phe Arg Gly	Phe Leu Gln Glu Arg	Leu Ala Leu Val Glu Ser
530	535	540
Gly Val Glu Leu Gly	Pro Ser Val Leu Phe	Phe Gly Cys Arg Asn Arg
545	550	555
Arg Met Asp Phe Ile	Tyr Glu Glu Glu Leu	Gln Arg Phe Val Glu Ser
565	570	575
Gly Ala Leu Ala Glu	Leu Ser Val Ala Phe	Ser Arg Glu Gly Pro Thr
580	585	590
Lys Glu Tyr Val Gln	His Lys Met Met Asp	Lys Ala Ser Asp Ile Trp
595	600	605
Asn Met Ile Ser Gln	Gly Ala Tyr Leu Tyr	Val Cys Gly Asp Ala Lys
610	615	620
Gly Met Ala Arg Asp	Val His Arg Ser Leu	His Thr Ile Ala Gln Glu
625	630	635
Gln Gly Ser Met Asp	Ser Thr Lys Ala Glu	Gly Phe Val Lys Asn Leu
645	650	655
Gln Thr Ser Gly Arg	Tyr Leu Arg Asp Val	Trp
660	665	

<210> 41
 <211> 560
 <212> PRT
 <213> *Eschscholzia stolonifera*

<400> 41

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

465 470 475 480
 Phe Arg Gly Gln His Phe Glu Tyr Ile Pro Phe Gly Ser Gly Arg Arg
 485 490 495
 Met Cys Pro Gly Val Asn Phe Ala Thr Leu Ile Leu His Met Thr Leu
 500 505 510
 Ala Arg Leu Leu Gln Ala Phe Asp Leu Ser Thr Pro Ser Ser Pro
 515 520 525
 Val Asp Met Thr Glu Gly Ser Gly Leu Thr Met Pro Lys Val Thr Pro
 530 535 540
 Leu Lys Val Leu Leu Thr Pro Arg Leu Pro Leu Pro Leu Tyr Asp Tyr
 545 550 555 560

<210> 42
 <211> 493
 <212> PRT
 <213> Catharanthus roseus

<400> 42
 Met Asp Tyr Leu Thr Ile Ile Leu Thr Leu Leu Phe Ala Leu Thr Leu
 1 5 10 15
 Tyr Glu Ala Phe Ser Tyr Leu Ser Arg Arg Thr Lys Asn Leu Pro Pro
 20 25 30
 Gly Pro Ser Pro Leu Pro Phe Ile Gly Ser Leu His Leu Leu Gly Asp
 35 40 45
 Gln Pro His Lys Ser Leu Ala Lys Leu Ser Lys Lys His Gly Pro Ile
 50 55 60
 Met Ser Leu Lys Leu Gly Gln Ile Thr Thr Ile Val Ile Ser Ser Ser
 65 70 75 80
 Thr Met Ala Lys Glu Val Leu Gln Lys Gln Asp Leu Ala Phe Ser Ser
 85 90 95
 Arg Ser Val Pro Asn Ala Leu His Ala His Asn Gln Phe Lys Phe Ser
 100 105 110
 Val Val Trp Leu Pro Val Ala Ser Arg Trp Arg Ser Leu Arg Lys Val
 115 120 125
 Leu Asn Ser Asn Ile Phe Ser Gly Asn Arg Leu Asp Ala Asn Gln His
 130 135 140
 Leu Arg Thr Arg Lys Val Gln Glu Leu Ile Ala Tyr Cys Arg Lys Asn
 145 150 155 160
 Ser Gln Ser Gly Glu Ala Val Asp Val Gly Arg Ala Ala Phe Arg Thr
 165 170 175
 Ser Leu Asn Leu Leu Ser Asn Leu Ile Phe Ser Lys Asp Leu Thr Asp
 180 185 190
 Pro Tyr Ser Asp Ser Ala Lys Glu Phe Lys Asp Leu Val Trp Asn Ile
 195 200 205
 Met Val Glu Ala Gly Lys Pro Asn Leu Val Asp Phe Phe Pro Leu Leu
 210 215 220
 Glu Lys Val Asp Pro Gln Gly Ile Arg His Arg Met Thr Ile His Phe
 225 230 235 240
 Gly Glu Val Leu Lys Leu Phe Gly Gly Leu Val Asn Glu Arg Leu Glu
 245 250 255
 Gln Arg Arg Ser Lys Gly Glu Lys Asn Asp Val Leu Asp Val Leu Leu
 260 265 270
 Thr Thr Ser Gln Glu Ser Pro Glu Glu Ile Asp Arg Thr His Ile Glu
 275 280 285
 Arg Met Cys Leu Asp Leu Phe Val Ala Gly Thr Asp Thr Thr Ser Ser
 290 295 300
 Thr Leu Glu Trp Ala Met Ser Glu Met Leu Lys Asn Pro Asp Lys Met
 305 310 315 320
 Lys Lys Thr Gln Asp Glu Leu Ala Gln Val Ile Gly Arg Gly Lys Thr
 325 330 335
 Ile Glu Glu Ser Asp Ile Asn Arg Leu Pro Tyr Leu Arg Cys Val Met
 340 345 350
 Lys Glu Thr Leu Arg Ile His Pro Pro Val Pro Phe Leu Ile Pro Arg
 355 360 365
 Lys Val Glu Gln Ser Val Glu Val Cys Gly Tyr Asn Val Pro Lys Gly
 370 375 380

Ser Gln Val Leu Val Asn Ala Trp Ala Ile Gly Arg Asp Glu Thr Val
 385 390 395 400
 Trp Asp Asp Ala Leu Ala Phe Lys Pro Glu Arg Phe Met Glu Ser Glu
 405 410 415
 Leu Asp Ile Arg Gly Arg Asp Phe Glu Leu Ile Pro Phe Gly Ala Gly
 420 425 430
 Arg Arg Ile Cys Pro Gly Leu Pro Leu Ala Leu Arg Thr Val Pro Leu
 435 440 445
 Met Leu Gly Ser Leu Leu Asn Ser Phe Asn Trp Lys Leu Glu Gly Gly
 450 455 460
 Met Ala Pro Lys Asp Leu Asp Met Glu Glu Lys Phe Gly Ile Thr Leu
 465 470 475 480
 Gln Lys Ala His Pro Leu Arg Ala Val Pro Ser Thr Leu
 485 490

<210> 43

<211> 496

<212> PRT

<213> Catharanthus roseus

<400> 43

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

355					360					365					
Arg	Glu	Cys	Arg	Glu	Gln	Cys	Lys	Ile	Lys	Gly	Tyr	Thr	Ile	Pro	Ser
370					375					380					
Lys	Ser	Arg	Val	Ile	Val	Asn	Ala	Trp	Ala	Ile	Gly	Arg	Asp	Pro	Asn
385					390					395					
Tyr	Trp	Ile	Glu	Pro	Glu	Lys	Phe	Asn	Pro	Asp	Arg	Phe	Leu	Glu	Ser
405					410					415					
Lys	Val	Asp	Phe	Lys	Gly	Asn	Ser	Phe	Glu	Tyr	Leu	Pro	Phe	Gly	Gly
420					425					430					
Gly	Arg	Arg	Ile	Cys	Pro	Gly	Ile	Thr	Phe	Ala	Leu	Ala	Asn	Ile	Glu
435					440					445					
Leu	Pro	Leu	Ala	Gln	Leu	Leu	Phe	His	Phe	Asp	Trp	Gln	Ser	Asn	Thr
450					455					460					
Glu	Lys	Leu	Asn	Met	Lys	Glu	Ser	Arg	Gly	Val	Thr	Val	Arg	Arg	Glu
465					470					475					
Asp	Asp	Leu	Tyr	Leu	Thr	Pro	Val	Asn	Phe	Ser	Ser	Ser	Pro	Ala	
485					490					495					

```
<210> 44
<211> 505
<212> PRT
<213> Arabidopsis thaliana
```

<400> 44																
Met	Asp	Leu	Leu	Leu	Leu	Glu	Lys	Ser	Leu	Ile	Ala	Val	Phe	Val	Ala	
1				5					10					15		
Val	Ile	Leu	Ala	Thr	Val	Ile	Ser	Lys	Leu	Arg	Gly	Lys	Lys	Leu	Lys	
		20						25					30			
Leu	Pro	Pro	Gly	Pro	Ile	Pro	Ile	Pro	Ile	Phe	Gly	Asn	Trp	Leu	Gln	
		35					40					45				
Val	Gly	Asp	Asp	Leu	Asn	His	Arg	Asn	Leu	Val	Asp	Tyr	Ala	Lys	Lys	
		50				55					60					
Phe	Gly	Asp	Leu	Phe	Leu	Leu	Arg	Met	Gly	Gln	Arg	Asn	Leu	Val	Val	
65					70					75					80	
Val	Ser	Ser	Pro	Asp	Leu	Thr	Lys	Glu	Val	Leu	Leu	Thr	Gln	Gly	Val	
				85					90					95		
Glu	Phe	Gly	Ser	Arg	Thr	Arg	Asn	Val	Val	Phe	Asp	Ile	Phe	Thr	Gly	
			100					105					110			
Lys	Gly	Gln	Asp	Met	Val	Phe	Thr	Val	Tyr	Gly	Glu	His	Trp	Arg	Lys	
		115					120					125				
Met	Arg	Arg	Ile	Met	Thr	Val	Pro	Phe	Phe	Thr	Asn	Lys	Val	Val	Gln	
	130					135					140					
Gln	Asn	Arg	Glu	Gly	Trp	Glu	Phe	Glu	Ala	Ala	Ser	Val	Val	Glu	Asp	
145					150					155					160	
Val	Lys	Lys	Asn	Pro	Asp	Ser	Ala	Thr	Lys	Gly	Ile	Val	Leu	Arg	Lys	
			165						170					175		
Arg	Leu	Gln	Leu	Met	Met	Tyr	Asn	Asn	Met	Phe	Arg	Ile	Met	Phe	Asp	
			180					185					190			
Arg	Arg	Phe	Glu	Ser	Glu	Asp	Asp	Pro	Leu	Phe	Leu	Arg	Leu	Lys	Ala	
		195					200					205				
Leu	Asn	Gly	Glu	Arg	Ser	Arg	Leu	Ala	Gln	Ser	Phe	Glu	Tyr	Asn	Tyr	
	210					215					220					
Gly	Asp	Phe	Ile	Pro	Ile	Leu	Arg	Pro	Phe	Leu	Arg	Gly	Tyr	Leu	Lys	
225					230					235					240	
Ile	Cys	Gln	Asp	Val	Lys	Asp	Arg	Arg	Ile	Ala	Leu	Phe	Lys	Lys	Tyr	
			245						250					255		
Phe	Val	Asp	Glu	Arg	Lys	Gln	Ile	Ala	Ser	Ser	Lys	Pro	Thr	Gly	Ser	
			260					265					270			
Glu	Gly	Leu	Lys	Cys	Ala	Ile	Asp	His	Ile	Leu	Glu	Ala	Glu	Gln	Lys	
		275					280					285				
Gly	Glu	Ile	Asn	Glu	Asp	Asn	Val	Leu	Tyr	Ile	Val	Glu	Asn	Ile	Asn	
	290					295					300					
Val	Ala	Ala	Ile	Glu	Thr	Thr	Leu	Trp	Ser	Ile	Glu	Trp	Gly	Ile	Ala	
305					310					315					320	
Glu	Leu	Val	Asn	His	Pro	Glu	Ile									

Asp Thr Val Leu Gly Pro Gly Val Gln Val Thr Glu Pro Asp Leu His
 340 345 350
 Lys Leu Pro Tyr Leu Gln Ala Val Val Lys Glu Thr Leu Arg Leu Arg
 355 360 365
 Met Ala Ile Pro Leu Leu Val Pro His Met Asn Leu His Asp Ala Lys
 370 375 380
 Leu Ala Gly Tyr Asp Ile Pro Ala Glu Ser Lys Ile Leu Val Asn Ala
 385 390 395 400
 Trp Trp Leu Ala Asn Asn Pro Asn Ser Trp Lys Lys Pro Glu Glu Phe
 405 410 415
 Arg Pro Glu Arg Phe Phe Glu Glu Glu Ser His Val Glu Ala Asn Gly
 420 425 430
 Asn Asp Phe Arg Tyr Val Pro Phe Gly Val Gly Arg Arg Ser Cys Pro
 435 440 445
 Gly Ile Ile Leu Ala Leu Pro Ile Leu Gly Ile Thr Ile Gly Arg Met
 450 455 460
 Val Gln Asn Phe Glu Leu Leu Pro Pro Pro Gly Gln Ser Lys Val Asp
 465 470 475 480
 Thr Ser Glu Lys Gly Gly Gln Phe Ser Leu His Ile Leu Asn His Ser
 485 490 495
 Ile Ile Val Met Lys Pro Arg Asn Cys
 500 505

<210> 45

<211> 512

<212> PRT

<213> Ocimum basilicum

<400> 45

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

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

<210> 46
 <211> 521
 <212> PRT
 <213> Glycine max

<400> 46
Met Leu Leu Glu Leu Ala Leu Gly Leu Phe Val Leu Ala Leu Phe Leu
1 5 10 15
His Leu Arg Pro Thr Pro Ser Ala Lys Ser Lys Ala Leu Arg His Leu
20 25 30
Pro Asn Pro Pro Ser Pro Lys Pro Arg Leu Pro Phe Ile Gly His Leu
35 40 45
His Leu Leu Lys Asp Lys Leu His Tyr Ala Leu Ile Asp Leu Ser
50 55 60
Lys Lys His Gly Pro Leu Phe Ser Leu Ser Phe Gly Ser Met Pro Thr
65 70 75 80
Val Val Ala Ser Thr Pro Glu Leu Phe Lys Leu Phe Leu Gln Thr His
85 90 95
Glu Ala Thr Ser Phe Asn Thr Arg Phe Gln Thr Ser Ala Ile Arg Arg
100 105 110
Leu Thr Tyr Asp Asn Ser Val Ala Met Val Pro Phe Gly Pro Tyr Trp
115 120 125
Lys Phe Val Arg Lys Leu Ile Met Asn Asp Leu Leu Asn Ala Thr Thr
130 135 140
Val Asn Lys Leu Arg Pro Leu Arg Thr Gln Gln Ile Arg Lys Phe Leu
145 150 155 160
Arg Val Met Ala Gln Ser Ala Glu Ala Gln Lys Pro Leu Asp Val Thr
165 170 175
Glu Glu Leu Leu Lys Trp Thr Asn Ser Thr Ile Ser Met Met Leu
180 185 190
Gly Glu Ala Glu Glu Ile Arg Asp Ile Ala Arg Glu Val Leu Lys Ile
195 200 205
Phe Gly Glu Tyr Ser Leu Thr Asp Phe Ile Trp Pro Leu Lys Tyr Leu
210 215 220
Lys Val Gly Lys Tyr Glu Lys Arg Ile Asp Asp Ile Leu Asn Lys Phe
225 230 235 240
Asp Pro Val Val Glu Arg Val Ile Lys Lys Arg Arg Glu Ile Val Arg
245 250 255

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

<210> 47

<211> 456

<212> PRT

<213> *Streptomyces avermitilis*

<400> 47

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

195	200	205
Arg Glu Ala Ile Thr Ala Tyr	Arg Ala Asp Gly Arg Asp His Gly Asp	
210	215	220
Leu Leu Ser Thr Met Leu Arg Ala Thr Asp Ala Glu Gly Ala Ser Met		
225	230	235
Thr Asp Gln Glu Val His Asp Glu Val Met Gly Ile Ala Val Ala Gly		
245	250	255
Ile Gly Gly Pro Ala Ala Ile Thr Ala Trp Ile Phe His Glu Leu Gly		
260	265	270
Gln Asn Ala Glu Ile Glu Ser Arg Leu His Ala Glu Leu Asp Thr Val		
275	280	285
Leu Gly Gly Arg Leu Pro Thr His Glu Asp Leu Pro Arg Leu Pro Tyr		
290	295	300
Thr Gln Asn Leu Val Lys Glu Ala Leu Arg Lys Tyr Pro Gly Trp Val		
305	310	315
Gly Ser Arg Arg Thr Val Arg Pro Val Arg Leu Gly Gly His Asp Leu		
325	330	335
Pro Ala Asp Val Glu Val Met Tyr Ser Ala Tyr Ala Ile Gln Arg Asp		
340	345	350
Pro Arg Trp Tyr Pro Glu Pro Glu Arg Leu Asp Pro Gly Arg Trp Glu		
355	360	365
Thr Lys Gly Ser Ser Arg Gly Val Pro Lys Gly Ala Trp Val Pro Phe		
370	375	380
Ala Leu Gly Thr Tyr Lys Cys Ile Gly Asp Asn Phe Ala Leu Leu Glu		
385	390	395
Thr Ala Val Thr Val Ala Val Val Ala Ser His Trp Arg Leu His Ala		
405	410	415
Leu Pro Gly Asp Glu Val Arg Pro Lys Thr Lys Ala Thr His Val Phe		
420	425	430
Pro Asn Arg Leu Arg Met Ile Ala Glu Pro Arg Ser Val Val Arg Leu		
435	440	445
Glu Glu Pro Ala Ala Met Gly Ala		
450	455	

<210> 48
 <211> 401
 <212> PRT
 <213> Streptomyces bikiniensis

<400> 48
Met Gly Leu Pro Leu Thr Ser Thr Lys Thr Ala Pro Val Ser Tyr Pro
1 5 10 15
Phe Gly Arg Pro Glu Gly Leu Asp Leu Asp Glu Ala Tyr Glu Gln Ala
20 25 30
Arg Lys Ser Glu Gly Leu Leu Trp Val His Met Pro Tyr Gly Glu Pro
35 40 45
Gly Trp Leu Val Ser Arg Tyr Asp Asp Ala Arg Phe Val Leu Gly Asp
50 55 60
Arg Arg Phe Ser His Ala Ala Glu Ala Glu Asn Asp Ala Pro Arg Met
65 70 75 80
Arg Glu Leu Arg Thr Pro Asn Gly Ile Ile Gly Met Asp Ala Pro Asp
85 90 95
His Thr Arg Leu Arg Gly Leu Val Thr Lys Ala Phe Thr Pro Arg Arg
100 105 110
Val Glu Ala Met Arg Pro His Val Arg Arg Met Thr Ala Ser Leu Leu
115 120 125
Arg Asp Met Thr Ala Leu Gly Ser Pro Val Asp Leu Val Asp His Tyr
130 135 140
Ala Val Pro Leu Pro Val Ala Val Ile Cys Gly Leu Leu Gly Val Pro
145 150 155 160
Glu Glu Asp Arg Asp Leu Phe Arg Gly Trp Cys Glu Ile Ala Met Ser
165 170 175
Thr Ser Ser Leu Thr Ala Glu Asp His Val Arg Leu Ala Gly Glu Leu
180 185 190
Thr Gly Tyr Leu Ala Asp Leu Ile Thr Ala Arg Arg Ala Ala Pro Arg
195 200 205

Asp Asp Leu Val Ser Ala Leu Val Glu Ala Arg Asp Ala Gln Gly Arg
 210 215 220
 Leu Ser Gln Glu Glu Leu Val Asp Leu Ile Val Phe Leu Leu Phe Ala
 225 230 235 240
 Gly His Glu Thr Thr Ala Ser Gln Ile Ser Asn Phe Val Leu Val Leu
 245 250 255
 Leu Glu Gln Pro Asp Gln Leu Ala Leu Leu Arg Asp Arg Pro Asp Leu
 260 265 270
 Leu Asp Asn Ala Val Glu Glu Leu Thr Arg Phe Val Pro Leu Gly Ser
 275 280 285
 Gln Ala Gly Phe Pro Arg Tyr Ala Thr Glu Asp Val Glu Val Gly Gly
 290 295 300
 Thr Leu Val Arg Ala Gly Asp Pro Val Leu Val Gln Met Asn Ala Ala
 305 310 315 320
 Asn Arg Asp Ala Leu Arg Phe Arg Ser Pro Gly Val Leu Asp Ile Thr
 325 330 335
 Arg Asp Asp Ala Gly Arg His Leu Gly Tyr Gly His Gly Pro His His
 340 345 350
 Cys Leu Gly Ala Ser Leu Ala Arg Leu Glu Leu Gln Glu Ala Leu Arg
 355 360 365
 Thr Leu Leu Asp Glu Leu Pro Gly Leu His Leu Ala Gln Pro Val Glu
 370 375 380
 Trp Lys Thr Glu Met Val Val Arg Gly Pro Arg Thr Met Leu Val Gly
 385 390 395 400
 Trp

<210> 49
 <211> 1497
 <212> DNA
 <213> *Artemisia annua*

<400> 49
 catatgaagt ctattctgaa agcaatggct ctgtctctga ccactagcat cgccctggcg 60
 actatcctgc tgtttgtgta caaattcgcg acccggttcta aaagcactaa gaaatctctg 120
 ccggaaccgt ggcgtctgcc aatcatcggc cacatgcacc acctgatcgg caccaccccg 180
 caccgtggcg tacgcgacct ggcgcgtaag tacggctctc tgatgcatct gcagctgggc 240
 gaggtacct ctatcgtcgt ttctccccg aagtgggcca aagaaatcct gactacctat 300
 gacatcactt tcgccaaccg cccggaaacg ctgaccggcg aaattgtcct gtaccataac 360
 acggatgtgg ttctggcccc gtacggtgag tactggcgcc agctgcgcaa aatttgtact 420
 ctggaactgc tgagcggttaa aaagggttaa tccttcaga gcctgcgtga agaggaatgc 480
 tggaacctgg tgcaggagat taaagcgtct ggcagcggtc gtccagttaa cctgtctgag 540
 aatgttttta aactgatcgc tactatcctg tctcgcgcgg cattcggtaa aggtatcaaa 600
 gatcagaaag aactgaccga aatcggttaag gaaatcctgc gccagactgg tggcttcgac 660
 gttgcggaca tcttcccgtc caaaaagttc ctgcaccatc tgtctggcaa acgcgctcgt 720
 ctgacctccc tgcgtaagaa aattgataac ctgattgaca acctggtcgc tgagcacact 780
 gtgaacacct cttctaaaac caacgaaacc ctgctggacg tactgctgcg cctgaaggac 840
 tctgccgaat ttccactgac tagcgacaat atcaaagcaa tcactcctgga catgttcggc 900
 gccggtaccg atacgtcctc ttccacgatt gagtgggcta ttccgaact gatcaaatgc 960
 ccgaaggcga tggaaaaagt gcaggcggaa ctgcgtaaaag cgctgaacgg taaagagaaa 1020
 attcatgaag aggacatcca ggaactgtcc tacctgaata tggtaatcaa agaaactctg 1080
 cgtctgcatc cgccgctgcc actggttctg ccgcgtgaat gccgtcagcc ggttaacctg 1140
 gccggctaca acattccgaa caaaacgaag ctgatcgtca acgttttcgc gatcaaccgc 1200
 gatcctgaat actggaagaa cgcggaagcg ttcatcccg aacgctttga gaactcctct 1260
 gccaccgtta tgggcgctga atacgagtac ctgccgttcg gtgcggttcg ccgtatgtgc 1320
 ccgggtgctg cactgggcct ggcgaacgtt caactgccac tggcgaacat cctgtaccac 1380
 ttcaactgga aactgcctaa cggcgatatc tatgatcaaa tcgacatgac cgaaagctcc 1440
 ggcgcgacca tgcagcgtaa aaccgaactg ctgctgggtc cgtcctttta acctagg 1497

<210> 50
 <211> 1497
 <212> DNA
 <213> Sequência Artificial

<220>
 <223> Modified amorphadiene oxidase nucleic acid

<400> 50

```

catatgaagt ctattctgaa agcaatggct ctgtctctga ccactagcat cgccttggcg 60
actatcctgc tgtttgtgta caaatctcgc acccgttcta aaagcactaa gaaatctctg 120
ccggaaccgt ggcgtctgcc aatcatcggc cacatgcacc acctgatcgg caccaccccg 180
cacctgtggc tacgcgacct ggcgcgtaag tacggctctc tgatgcactc gcagctgggc 240
gagggtacct ctatcgtcgt ttccctcccc aagtggggcca aagaaatcct gactacctat 300
gacatcactt tcgccaaccg cccggaaacg ctgaccggcg aaattgtcct gtaccataac 360
acggatgtgg ttctggcccc gtacggtgag tactggcgcc agctgcgcaa aatttgtact 420
ctggaactgc tgagcgttaa aaagggttaa tccttcaga gcctgcgtga agaggaatgc 480
tggaacctgg tgcaggagat taaagcgtct ggcagcggtc gtccagttaa cctgtctgag 540
aatgttttta aactgatcgc tactatcctg tctcgcgcgg cattcggtaa aggtatcaaa 600
gatcagaaag aactgaccga aatcgttaa gaaatcctgc gccagactgg tggcttcgac 660
gttgccgaca tcttccccgc caaaaagttc ctgcaccatc tgtctggcaa acgcgtcgt 720
ctgacctccc tgcgtaagaa aattgataac ctgattgaca acctggtcgc tgagcacact 780
gtgaaacact cttctaaaac caacgaaacc ctgctggacg tactgtcgc cctgaaggac 840
tctgcccgaat ttccactgac tagcgacaat atcaaagcaa tcactcctgga catgttcggc 900
gccgggtacc atacgtctc ttccacgatt gagtgggcta ttccgaact gatcaaatgc 960
ccgaaggcga tggaaaaagt gcaggcggaa ctgcgtaaag cgctgaacgg taaagagaaa 1020
attcatgaag aggacatcca ggaactgtcc tactgaata tggtaatcaa agaaactctg 1080
cgtctgcac cgcgcgtgcc actggttctg ccgcgtgaat gccgtcagcc ggttaacctg 1140
gccggctaca acattccgaa caaacgaag ctgatcgtca acgttttcgc gatcaaccgc 1200
gatcctgaat actggaaaga cgcggaagc ctgattcgcg aacgctttga gaactcctc 1260
gccaccgtta tggcgctga atacgagtac ctgccgttcg gtgcgggtcg ccgtatgtgc 1320
ccgggtgctg cactgggcct ggcaacgtt caactgccac tggcgaacat cctgtaccac 1380
ttcaactgga aactgcctaa cggcgtatct tatgatcaaa tcgacatgac cgaaagctcc 1440
ggcgcgacca tgcagcgtaa aaccgaactg ctgctgggtc cgtcctttta acctagg 1497

```

<210> 51

<211> 1488

<212> DNA

<213> Sequência Artificial

<220>

<223> Modified amorphadiene oxidase nucleic acid

<400> 51

```

atgaagtcta ttctgaaagc aatggctctg tctctgacca ctagcatcgc cctggcgact 60
atcctgctgt ttgtgtacaa attcgcgacc cgttctaaaa gactaagaa atctctgccg 120
gaaccgtggc gtctgccaat catcggtcac atgcaccacc tgatcggcac caccgcgcac 180
cgtggcgtag gcgacctggc gcgtaagtag ggctctctga tgcatctgca gctggcgtag 240
gtacctacta tcgtcgtttc ctccccgaag tgggccaag aaatcctgac tacctatgac 300
atcactttcg ccaaccgccc ggaacgcgtg accggcgaaa ttgtcctgta ccataaacg 360
gatgtggttc tggccccgta cggtagtagc tggcgccagc tgcgcaaat ttgtactctg 420
gaactgctga gcgttaaaaa ggttaaatcc ttccagagcc tgcgtgaaga ggaatgctgg 480
aacctggtgc aggagattaa agcgtctggc agcggtcgtc cagttaacct gtctgagaat 540
gttttttaac tgatcgtac taccctgtct cgcgcggcat tcggtaaagg tatcaaagat 600
cagaaagaac tgaccgaaat cgtaaaggaa atcctgcgcc agactggtg cttcgacgtt 660
gcggacatct tcccgcccaa aaagtctctg caccatctgt ctggcaaacg cgctcgtctg 720
acctcctgc gtaagaaaat tgataacctg attgacaacc tggctcgtga gcacactgtg 780
aacacctctt ctaaaaccaa cgaaaccctg ctggacgtac tgcgcgcct gaaggactct 840
gccgaatttc cactgactag cgacaatata aaagcaatca tctggacat gttcggcgcc 900
ggtaccgata cgtcctcttc cagcattgag tgggctattt ccgaactgat caaatgccc 960
aaggcgatgg aaaaagtgc ggcggaactg cgtaaagcgc tgaacggtaa agagaaaatt 1020
catgaagagg acatccagga actgtcctac ctgaatatgg taatcaaaga aactctgcgt 1080
ctgcatccgc cgctgccact ggttctgcgc cgtgaatgcc gtcagccggt taacctggcc 1140
ggctacaaca ttccgaacaa aacgaagctg atcgtcaacg ttttcgcgat caaccgcgat 1200
cctgaatact ggaaagacgc ggaagcgttc atccggaac gctttgagaa ctctctgcc 1260
accgttatgg gcgctgaata cgagtacctg ccgttcgggt cgggtcgccg tatgtgccc 1320
ggtgctgcac tgggcctggc gaacgttcaa ctgccactgg cgaacatcct gtaccacttc 1380
aactggaac tgcctaacgg cgtatcttat gatcaaatcg acatgaccga aagctccggc 1440
gcgaccatgc agcgtaaaa cgaactgctg ctggttccgt ctttttaa 1488

```

<210> 52

<211> 495

<212> PRT

<213> Sequência Artificial

<220>

<223> Modified amorphadiene oxidase

<400> 52

```

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

```

<210> 53
 <211> 3018
 <212> DNA
 <213> Sequência Artificial

<220>
 <223> Modified amorphadiene oxidase nucleic acid

<400> 53
 catatgaccg tacacgacat catcgcaacg tacttacta aatggtacgt aattgtgccc 60
 ctggcactga ttgcgtatcg cgtgctggat tatttctacg cgaccggttc taaaagcact 120
 aagaaatctc tgccggaacc gtggcgtctg ccaatcatcg gtcacatgca ccacctgac 180
 ggcaccaccc cgcaccgtgg cgtacgcgac ctggcgctga agtacggctc tctgatgcat 240
 ctgcagctgg gcgaggtacc tactatcgtc gtttctctcc cgaagtgggc caaagaaatc 300
 ctgactacct atgacatcac ttctcgcaac cgcccggaac cgctgaccgg cgaaattgtc 360
 ctgtaccata acacggatgt ggttctggcc cgtacggtg agtactggcg ccagctgcgc 420
 aaaatttgta ctctggaact gctgagcgtt aaaaagggtta aatccttcca gagcctgcgt 480
 gaagaggaaat gctggaacct ggtgcaggag attaaagcgt ctggcagcgg tcgtccagtt 540
 aacctgtctg agaattgttt taaactgatc gctactatcc tgtctcgcgc ggcattcggg 600
 aaaggtatca aagatcagaa agaactgacc gtaaatcgta aggaaatcct gcgcagact 660
 ggtggcctcg acgttgcgga catcttcccg tccaaaaagt tcctgcacca tctgtctggc 720
 aaacgcgctc gtctgacctc cctgcgtaag aaaattgata acctgattga caacctgggtc 780
 gctgagcaca ctgtgaacac ctcttctaaa accaacgaaa ccctgctgga cgtactgctg 840
 cgcctgaagg actctgcgga atttccactg actagcgaca atatcaaagc aatcatcctg 900
 gacatgttcg gcgcgcgtac cgatacgtcc tcttccacga ttgagtgggc tatttccgaa 960
 ctgatcaaat gcccgaaggc gatggaaaaa gtgcaggcgg aactgcgtaa agcgtgaac 1020
 ggtaaaagaga aaattcatga agaggacatc caggaaactgt cctacctgaa tatggtaatc 1080
 aaagaaactc tgctgtgca tccgcgcgtg ccactggttc tgccgcgtga atgccgtcag 1140
 ccggttaacc tggccggcta caacattccg aacaaaacga agctgatcgt caacgttttc 1200
 gcgatcaacc gcgatcctga atactggaaa gacgcggaag cgttcattcc ggaacgcttt 1260
 gagaactcct ctgccaccgt tatggcgct gaatacagat acctgccgtt cgggtgcgggt 1320
 cgccgtatgt gcccgggtgc tgcactgggc ctggcgaaac ttcaactgcc actggcgaaac 1380
 atcctgtacc acttcaactg gaaactgcct aacggcgtat cttatgatca aatcgacatg 1440
 accgaaagct ccggcgcgac catgcagcgt aaaaccgaac tgctgctggg tccgtccttt 1500
 taacctaggc atatgaccgt acacgacatc atcgcaactg acttcaacta atggtagcga 1560
 attgtgccgc tggcactgat tgcgtatcgc gtgctggatt atttctacgc gaccggttct 1620
 aaaagcacta agaaatctct gccggaaccg tggcgtctgc caatcatcgg tcacatgcac 1680
 cacctgatcg gcaccacccc gcaccgtggc gtacgcgacc tggcgcgtaa gtacggctct 1740
 ctgatgcac tgcagctggg cgaggtacct actatcgtcg tttctctccc gaagtgggccc 1800
 aaagaaatcc tgactaccta tgacatcact ttgcgcaacc gcccggaac gctgaccggc 1860
 gaaattgtcc tgtaccataa cacggatgtg gttctggccc cgtacggtga gtactggcg 1920
 cagctgcgca aaatttgtac tctggaactg ctgagcgtaa aaaagggtta atccttccag 1980
 agcctgcgtg aagaggaatg ctggaacctg gtgcaggaga ttaaagcgtc tggcagcggg 2040
 cgtccagtta acctgtctga gaatgttttt aaactgatcg ctactatcct gtctcgcgcg 2100
 gcattcggta aaggtatcaa agatcagaaa gaactgaccg aaatcggtta ggaaatcctg 2160
 cgccagactg gtggccttga cgttgccggac atcttcccg tccaaaaagt cctgcaccat 2220
 ctgtctggga aacgcgctcg tctgacctcc ctgcgtaaga aaattgataa cctgattgac 2280
 aacctggtcg ctgagcacac tctgaacacc tcttctaaaa ccaacgaaac cctgctggac 2340
 gtactgctgc gcctgaagga ctctgccgaa ttctcactga ctagcgacaa tatcaaagca 2400
 atcatcctgg acatgttcgg cgccggtacc gatacgtcct cttccacgat tgagtgggct 2460
 atttccgaac tgatcaaatg cccgaaggcg atggaaaaag tgcaggcgga actgcgtaaa 2520
 gcgctgaacg gtaaagagaa aattcatgaa gaggacatcc aggaactgtc ctacctgaat 2580
 atggtaatca aagaaactct gcgtctgcat ccgcgcgtgc cactggttct gccgcgtgaa 2640
 tgccgtcagc cggttaacct ggccggctac aacattccga acaaaacgaa gctgatcgtc 2700
 aacgttttgc cgatcaaccg cgatcctgaa tactggaaag acgcggaagc gttcattccg 2760
 gaacgctttg agaactctc tgccaccgtt atggcgctg aatacagta cctgccgttc 2820
 ggtgcgggtc gccgtatgtg cccgggtgct gcaactgggc tggcgaaact tcaactgcca 2880
 ctggcgaaac tctgtacca cttcaactgg aaactgccta acggcgatc ttatgatcaa 2940
 atcgacatga ccgaaagctc cggcgcgacc atgcagcgt aaaccgaact gctgctgggt 3000
 ccgtcctttt aacctagg 3018

<210> 54
 <211> 1500
 <212> DNA
 <213> Sequência Artificial

<220>

<223> Modified amorphadiene oxidase nucleic acid

<400> 54

```

atgaccgtac acgacatcat cgcaacgtac ttactaaat ggtacgtaat tgtgccgctg 60
gcactgattg cgtatcgctg gctggattat ttctacgcga cccgttctaa aagcactaag 120
aaatctctgc cggaaccgtg gctgtctgcc atcatcggtc acatgcacca cctgatcggc 180
accaccccg cccgtggcgt acgcgacctg gcgcgtaagt acggctctct gatgcatctg 240
cagctgggcg aggtacctac tatcgctcgt ttctccccga agtgggcca agaaatcctg 300
actacctatg acatcacttt cgccaaccgc ccggaacgc tgaccggcga aattgtcctg 360
taccataaca cggatgtggt tctggccccg tacggtgagt actggcgcca gctgcgcaaa 420
atgtgtactc tggaactgct gagcggtaaa aagggttaa ccttccagag cctgcgtgaa 480
gaggaatgct ggaacctggt gcaggagatt aaagcgtctg gcagcggctg tccagttaac 540
ctgtctgaga atgtttttaa actgatcgct actatcctgt ctgcgcggc attcggtaaa 600
gggtatcaaaa atcagaaaaga actgaccgaa atcggttaag aaatcctgcg ccagactggt 660
ggcttcgacg ttgcggacat cttcccgctc aaaaagtctc tgcaccatct gtctggcaaa 720
cgcgctcgtc tgacctccct gcgtaagaaa attgataacc tgattgacaa cctggctcgt 780
gagcacactg tgaacacctc ttctaaaacc aacgaaaccc tgctggacgt actgctgcgc 840
ctgaaggact ctgccgaatt tccactgact agcgacaata tcaaagcaat catcctggac 900
atgttcggcg ccggtaccga tacgtcctct tccacgattg agtgggctat ttccgaactg 960
atcaaatgcc cgaaggcgat ggaaaaagtg caggcggaac tgcgtaaaag gctgaacggt 1020
aaagagaaaa ttcataaaga ggacatccag gaactgtcct acctgaatat ggtaatcaaa 1080
gaaactctgc gtctgcatcc gccgctgcc ctggttctgc cgcgtgaatg ccgctcagcg 1140
gttaacctgg ccggctacaa cattccgaac aaaacgaagc tgatcgtaa cgttttcgcg 1200
atcaaccgcg atcctgaata ctggaaagac gcggaagcgt tcattccgga acgctttgag 1260
aactcctctg ccaccgttat gggcgctgaa tacgagtacc tgccgttcg tgccgggtcgc 1320
cgtatgtgcc cgggtgctgc actgggcctg gcgaacgttc aactgccact ggcaacatc 1380
ctgtaccact tcaactggaa actgcctaac ggcgtatctt atgatcaaat cgacatgacc 1440
gaaagctccg gcgcgacct gcagcgtaaa accgaactgc tgctgggtcc gtccttttaa 1500

```

<210> 55

<211> 499

<212> PRT

<213> Sequência Artificial

<220>

<223> Modified amorphadiene oxidase

<400> 55

```

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

```

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

<210> 56

<211> 2988

<212> DNA

<213> Sequência Artificial

<220>

<223> Modified amorphadiene oxidase

<400> 56

```

catatgatcg aacaactgct ggaatactgg tacgtggttg tgcctgttct gtatattatc 60
aaacagctgc tggcgtagac taaagcgacc cggtctaaaa gactaagaa atctctgccg 120
gaaccgtggc gtctgccaat catcggtcac atgcaccacc tgatcggcac caccgccgac 180
cgtggcgtag gcgacctggc gcgtaagtac ggctctctga tgcactctga gctgggagag 240
gtacctacta tcgtcgtttc ctccccgaag tgggcaaaag aaatcctgac tacctatgac 300
atcactttcg ccaaccgccc ggaacgctg accggcgaaa ttgtcctgta ccataacacg 360
gatgtggttc tggccccgta cggtgagtag tggcgccagc tgcgcaaaat ttgtactctg 420
gaactgctga gcgttaaaaa ggttaaatcc ttccagagcc tgcgtgaaga ggaatgctgg 480
aacctggtgc aggagattaa agcgtctggc agcgtctgct cagttaacct gtctgagaat 540
gtttttaaac tgatcgctac tatcctgtct cgcgcggcat tcggtaaagg tatcaaagat 600
cagaaagaac tgaccgaaat cgtaaggaa atcctgcgcc agactggtgg cttcgacgtt 660
gcggacatct tcccgtccaa aaagtctctg caccatctgt ctggcaaacg cgctcgtctg 720
acctcctgac gtaagaaaat tgataacctg attgacaacc tggctcgtga gcacactgtg 780
aacacctctt ctaaaaccaa cgaaccctg ctggacgtac tgctgcgcct gaaggactct 840
gccgaatttc cactgactag cgacaatatc aaagcaatca tcctggacat gttcggcgcc 900
ggtagcgata cgtctcttc cagattgag tgggtattt ccgaactgat caaatgccc 960
aaggcgatgg aaaaagtgca ggcggaactg cgtaaagcgc tgaacggtaa agagaaaatt 1020
catgaagagg acatccagga actgtcctac ctgaatatgg taatcaaaga aactctgcgt 1080
ctgcatccgc cgctgccact ggttctgccg cgtgaatgcc gtcagccggt taacctggcc 1140

```

```

ggctacaaca ttccgaacaa aacgaagctg atcgtcaacg ttttcgcat caaccgcat 1200
cctgaataact ggaagacgc ggaagcggtc attccggaac gctttgagaa ctccctctgcc 1260
accgttatgg gcgtgaata cgagtacctg ccgttcgggtg cgggtcgccg tatgtgcccg 1320
ggtgctgcac tgggcctggc gaacgttcaa ctgccactgg cgaacatcct gtaccacttc 1380
aactggaaac tgcctaacgg cgtatcttat gatcaaatcg acatgaccga aagctccggc 1440
gcgaccatgc agcgtaaaaac cgaactgctg ctgggtccgt ccttttaacc taggcatatg 1500
atcgaacaac tgctggaata ctggtacgtg gttgtgcctg ttctgtatat tatcaaacag 1560
ctgctggcgt aactaaagc gaccggttct aaaagcacta agaaatctct gccggaaccg 1620
tggcgtctgc caatcatcgg tcacatgcac cacctgatcg gcaccacccc gcaccgtggc 1680
gtacgcgacc tggcgcgtaa gtacggctct ctgatgcac tgcagctggg cgagggtacct 1740
actatcgctg tttcctcccc gaagtgggccc aaagaaatcc tgactaccta tgacatcact 1800
ttcgccaacc gcccggaac gctgaccggc gaaattgtcc tgtaccataa cacggatgtg 1860
gttctggccc cgtacggtga gtactggcgc cagctgcgca aaatttgtac tctggaactg 1920
ctgagcggtta aaaagggtta atccttccag agcctgcgtg aagaggaatg ctggaacctg 1980
gtgcaggaga ttaaagcgtc tggcagcggt cgtccagtta acctgtctga gaatgttttt 2040
aaactgatcg ctactatcct gtctcgcgcg gcattcggta aaggatcaa agatcagaaa 2100
gaactgaccg aaatcggtta ggaaatcctg cgccagactg gtggcttcga cgttgccggc 2160
atcttcccgt ccaaaaagtt cctgcacat ctgtctggca aacgcgctcg tctgacctcc 2220
ctgcgtaaga aaattgataa cctgattgac aacctggctg ctgagcacac tgtgaacacc 2280
tcttctaata ccaacgaaac cctgctggac gtactgctg gccggaagg ctctgccgaa 2340
tttccactga ctagcgacaa tatcaaagca atcatcctgg acatgttcgg cgccggtacc 2400
gatacgtcct cttccacgat tgagtgggct atttccgaac tgatcaaag cccgaaggcg 2460
atggaaaaag tgcaggcgga actgcgtaaa gcgtgaaac gtaaagagaa aattcatgaa 2520
gaggacatcc aggaactgtc ctacctgaat atggtaatca aagaaactct gcgtctgcat 2580
ccgcgctgc cactggttct gccgcgtgaa tgcgctcagc cggttaacct ggccggctac 2640
aacattccga acaaaacgaa gctgatcgtc aacgttttcg cgatcaaccg cgatcctgaa 2700
tactggaaag acgcggaagc gttcattccg gaacgcttg agaactcctc tgccaccggt 2760
atggcgctg aatacagta cctgccgttc ggtgcgggtc gccgtatgtg cccgggtgct 2820
gcactgggccc tggcgaacgt tcaactgcca ctggcgaaca tcctgtacca cttcaactgg 2880
aaactgccta acggcgatc ttatgatcaa atcgacatga ccgaaagctc cggcgcgacc 2940
atgcagcgta aaaccgaact gctgctggtt ccgtcctttt aacctagg 2988

```

<210> 57

<211> 1485

<212> DNA

<213> Sequência Artificial

<220>

<223> Modified amorphadiene oxidase nucleic acid

<400> 57

```

atgatcgaac aactgctgga atactggtac gtggttgtgc ctgttctgta tattatcaaa 60
cagctgctgg cgtacactaa agcgaccggt tctaaaagca ctaagaaatc tctgccggaa 120
ccgtggcgctc tgccaatcat cggtcacatg caccacctga tcggcaccac ccgcaccgt 180
ggcgtagcgc acctggcgcg taagtacggc tctctgatgc atctgcagct gggcgaggta 240
cctactatcg tcgtttcctc ccggaagtgg gccaaagaaa tcctgactac ctatgacatc 300
actttcgcca accgcccgga aacgctgacc ggcgaaattg tcctgtacca taacacggat 360
gtggttcttg ccccgtagcg tgagtactgg cgccagctgc gcaaaatttg tactctggaa 420
ctgctgagcg ttaaaaagggt taaatccttc cgagcctgc gtgaagagga atgctggaac 480
ctggtgcagg agattaaagc gtctggcagc ggtcgtccag ttaacctgtc tgagaatgtt 540
tttaaaactga tcgctactat cctgtctcgc gcggcattcg gtaaagggtat caaagatcag 600
aaagaactga ccgaaatcgt taaggaaatc ctgcgccaga ctggtggctt cgacgttgcg 660
gacatcttcc cgtccaaaaa gttctcgcac catctgtctg gcaaacgcgc tcgtctgacc 720
tccttgcgta agaaaattga taacctgatt gacaacctgg tcgctgagca cactgtgaac 780
acctcttcta aaaccaacga aacctgctg gacgtactgc tgcgcctgaa ggactctgcc 840
gaatttccac tgactagcga caatatcaaa gcaatcatcc tggacatgtt cggcgccggt 900
accgatacgt cctcttccac gattgagtg gctatttccg aactgatcaa atgccgaag 960
gcgatggaaa aagtgcaggc ggaactgcgt aaagcgctga acggtaaaga gaaaattcat 1020
gaagaggaca tccaggaaact gtcctacctg aatatggtaa tcaaagaaac tctgcgtctg 1080
catccgccgc tgccactggt tctgcccggt gaatgccgtc agccggttaa cctggccggc 1140
tacaacattc cgaacaaaac gaagtgcac gtcaacggtt tcgcatcaa ccgcgatcct 1200
gaatactgga aagacgcgga agcgttcatt ccggaacgct ttgagaactc ctctgccacc 1260
gttatgggcg ctgaatacga gtacctgcgt ttcggtgcgg gtcgccgtat gtgcccggt 1320
gctgcactgg gcctggcgaa cgttcaactg ccactggcga acatcctgta ccacttcaac 1380
tggaaactgc ctaacggcgt atcttatgat caaatcgaca tgaccgaaag ctccggcgcg 1440
accatgcagc gtaaaaccga actgctgctg gttccgtcct ttttaa 1485

```

<210> 58
 <211> 494
 <212> PRT
 <213> Sequência Artificial

<220>
 <223> Modified amorphadiene oxidase

<400> 58

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

Asn Gly Val Ser Tyr Asp Gln Ile Asp Met Thr Glu Ser Ser Gly Ala
 465 470 475 480
 Thr Met Gln Arg Lys Thr Glu Leu Leu Leu Val Pro Ser Phe
 485 490

<210> 59

<211> 2946

<212> DNA

<213> Sequência Artificial

<220>

<223> Modified amorphadiene oxidase nucleic acid

<400> 59

```

catatggctc tgcctgctggc tgtcttctctg ggtctgtcct gcctgctgct gctgtccctg 60
tgggcgaccc gttctaaaag cactaagaaa tctctgccgg aaccgtggcg tctgccaatc 120
atcggtcaca tgcaccacct gatcggcacc accccgcacc gtggcgtagc cgacctggcg 180
cgtaagtacg gctctctgat gcatctgcag ctggcgaggg tacctactat cgtcgtttcc 240
tccccgaagt gggccaaaga aatcctgact acctatgaca tcactttcgc caaccgcccc 300
gaaacgctga cggcgaaaat tgtcctgtac cataacacgg atgtggttct ggccccgtac 360
ggtgagtact ggcgccagct gcgcaaaatt tgtactctgg aactgctgag cgttaaaaag 420
gttaaatcct tccagagcct gcgtgaagag gaatgctgga acctgggtgca ggagattaaa 480
gcgtctggca cgggtcgtcc agttaacctg tctgagaatg tttttaaact gatcgctact 540
atcctgtctc gcgcggcatt cggtaaaggt atcaaagatc agaaagaact gaccgaaatc 600
gttaaggaaa tccctgcgcca gactggtggc ttcgacgttg cggacatctt cccgtccaaa 660
aagttcctgc accatctgtc tggcaaacgc gctcgtctga cctccctgcg taagaaaatt 720
gataacctga ttgacaacct ggtcgttgag cacactgtga acactcttc taaaaccaac 780
gaaacctgc tggcgtact cttgcgcctg aaggactctg ccgaatttcc actgactagc 840
gacaatatca aagcaatcat cctggacatg ttcggcgccg gtaccgatac gtcctcttcc 900
acgattgagt gggctatttc cgaactgatc aaatgcccga aggcgatgga aaaagtgcag 960
gcggaactgc gtaaagcgct gaacggtaaa gagaaaattc atgaagagga catccaggaa 1020
ctgtcctacc tgaatatggt aatcaaagaa actctgcgtc tgcattccgc gctgccactg 1080
gttctgccgc gtgaatgccg tcagccggtt aacctggccg gctacaacat tccgaacaaa 1140
acgaactga tcgtcaacgt tttcgcgatc aaccgcgatc ctgaatactg gaaagacgcg 1200
gaagcgttca tcccggaacg ctttgagaac tccctgccca ccgttatggg cgtgaatac 1260
gagtacctgc cgttcggtgc gggctgccgt atgtgcccg gtgctgact gggcctggcg 1320
aacgttcaac tgccactggc gaacatcctg taccatttca actggaaact gcctaaccggc 1380
gtatcttatg atcaaatcga catgaccgaa agctccggcg cgaccatgca gcgtaaaacc 1440
gaactgctgc tggttccgtc cttttgacct aggcataatg ctctgctgct ggctgtcttc 1500
ctgggtctgt cctgcctgct gctgctgtcc ctgtgggcga cccgttctaa aagcactaag 1560
aaatctctgc cggaaaccgt gcgtctgcc atcatcggtc acatgcacca cctgatcggc 1620
accaccccg accgtggcgt acgcgacctg gcgcgtaagt acggctctct gatgcatctg 1680
cagctggcg aggtacctac tatcgtcgtt tccctcccga agtgggcca agaaatcctg 1740
actacctatg acatcacttt cgccaaccgc cgggaaacgc tgaccggcga aattgtcctg 1800
taccataaca cggatgtggt tctggccccg tacggtgagt actggcgcca gctgcgcaaa 1860
atgtgtactc tggaaactgct gagcgtaaaa aagggttaaat ccttccagag cctgcgtgaa 1920
gaggaatgct ggaacctggt gcaggagatt aaagcgtctg gcagcggctg tccagttaa 1980
ctgtctgaga atgtttttaa actgatcgt actatcctgt ctgcgcggc attcggtaaa 2040
ggtatcaaa atcagaaaaga actgaccgaa atcgtaagg aaatcctgcg ccagactggt 2100
ggcttcgacg ttgaggacat cttcccgtcc aaaaagttcc tgcaccatct gtctggcaaa 2160
cgcgctcgtc tgacctccct gcgtaagaaa attgataacc tgattgacaa cctggctcgt 2220
gagcacactg tgaacacctc ttctaaaacc aacgaaaccc tgctggacgt actgctgcgc 2280
ctgaaggact ctgccgaatt tccactgact agcgacaata tcaaagcaat catcctggac 2340
atgttcggcg ccggtaccga tacgtcctct tccacgattg agtgggctat ttccgaactg 2400
atcaaatgcc cgaaggcgat ggaaaaagtg caggcggaac tgcgtaaagc gctgaacgtg 2460
aaagagaaaa ttcattgaaga ggacatccag gaactgtcct acctgaatat ggtaatcaaa 2520
gaaactctgc gtctgcatcc gccgctgcca ctggttctgc cgcgtgaatg ccgtcagccg 2580
gttaacctgg ccggctacaa cattccgaac aaaacgaagc tgatcgtcaa cgttttcgcg 2640
atcaaccgcg atcctgaata ctggaaagac gcggaagcgt tcattccgga acgctttgag 2700
aactcctctg ccaccgttat gggcgctgaa tacgagtacc tgccgttcgg tcggggtcgc 2760
cgtatgtgcc cgggtgctgc actgggcctg gcgaacgttc aactgccact ggcaacatc 2820
ctgtaccact tcaactggaa actgcctaac ggcgtatctt atgatcaaat cgacatgacc 2880
gaaagctccg gcgcgaccat gcagcgtaaa accgaactgc tgctggttcc gtccttttaa 2940
cctagg

```

<210> 60

<211> 1464

<212> DNA

<213> Sequência Artificial

<220>

<223> Modified amorphadiene oxidase nucleic acid

<400> 60

```

atggctctgc tgctggctgt cttcctgggt ctgtcctgcc tgctgctgct gtccctgtgg 60
gcgaccggtt ctaaaagcac taagaaatct ctgccggaac cgtggcgtct gccaatcatc 120
ggtcacatgc accacctgat cggcaccacc cgcaccgtg gcgtacgcga cctggcgcggt 180
aagtacggct ctctgatgca tctgcagctg ggcgaggtac ctactatcgt cgtttcctcc 240
ccgaagtggg ccaaagaaat cctgactacc tatgacatca ctttcgccaa ccgcccggaa 300
acgctgaccg gcgaaattgt cctgtaccat aacacggatg tggttctggc ccgtacgggt 360
gagtactggc gccagctgcg caaaatttgt actctggaac tgctgagcgt taaaaagggt 420
aaatccttcc agagcctgcy tgaagaggaa tgctggaacc tgggtgcagga gattaaagcg 480
tctggcagcg gtcgtccagt taacctgtct gagaatgttt ttaaactgat cgctactatc 540
ctgtctcgcg cggcattcgg taaaggatc aaagatcaga aagaactgac cgaaatcggt 600
aaggaaatcc tgcgccagac tgggtggcttc gacgttgcyg acatcttccc gtccaaaaag 660
ttcctgcacc atctgtctgg caaacgcgct cgtctgacct ccctgcgtaa gaaaattgat 720
aacctgattg acaacctggg cgctgagcac actgtgaaca cctcttctaa aaccaacgaa 780
accctgctgg acgtactgct gcgcctgaag gactctgccg aatttccact gactagcgac 840
aatatcaaag caatcatcct ggacatgttc ggcgccggta ccgatacgtc ctcttcacg 900
attgagtggg ctatttccga actgatcaaa tgcccgaagg cgatggaaaa agtgcaggcg 960
gaactgcgta aagcgctgaa cggtaaagag aaaattcatg aagaggacat ccagggaactg 1020
tctacactga atatggtaat caaagaaact ctgcgtctgc atccgcgct gccactgggt 1080
ctgccgcgtg aatgccgtca gccggttaac ctggccggct acaacattcc gaacaaaacg 1140
aagctgatcg tcaacgtttt cgcgatcaac cgcgatcctg aatactggaa agacgcggaa 1200
gcgttcattc cggaacgctt tgagaactcc tctgccaccg ttatgggcgc tgaatacgag 1260
tacctgcgtt cgggtgcggg tcgccgtatg tgcccgggtg ctgcaactggg cctggcggaac 1320
gttcaactgc cactggcgaa catcctgtac cacttcaact ggaaactgcc taacggcgta 1380
tcttatgatc aaatcgacat gaccgaaagc tccggcgcgga ccatgcagcg taaaaccgaa 1440
ctgctgctgg ttccgtcctt ttaa                                     1464

```

<210> 61

<211> 487

<212> PRT

<213> Sequência Artificial

<220>

<223> Modified amorphadiene oxidase

<400> 61

```

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

```

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

<210> 62
 <211> 10633
 <212> DNA
 <213> Sequência Artificial

<220>
 <223> Recombinant plasmid

<400> 62
 accttcggga ggcctgaag cccgttctgg acgccttggg gccgttgaat cgggatatgc 60
 aggccaaggc cgccgcgac atcaaggccg tggcgaaaa gctgctgacg gaacagcggg 120
 aagtcacgag ccagaaacag gccacgcgc agcaggaacg cggcgcgca catttccccg 180
 aaaagtgcc cctgacgtct aagaaacat tattatcatg acattaacct ataaaaatag 240
 gcgtatcacg aggccttttc gtcttcaaga attctcatgt ttgacagctt atcatcgata 300
 agctttaatg cggtagttaa tcacagttaa attgctaacg cagtcaggca ccgtgtatga 360
 aatctaaca tgcgctcatc gtcatcctcg gcaccgtcac cctggatgct gtaggcatag 420
 gcttggttat gccgtactg ccgggcctct tgcgggatat cgtccattcc gacagcatcg 480
 ccagtcacta tggcgtgctg ctacgcgtat atgcgttgat gcaatttcta tgcgcacccg 540
 ttctcggagc actgtccgac cgctttggcc gccgccagc cctgctcgct tcgctacttg 600
 gagccactat cgactacgag atcatggcga ccacacccgt cctgtggatc ctctacgccg 660
 gacgcatcgt ggccggcatc accggcgcca caggtgcggt tgctggcgcc tatatcgccg 720
 acatcaccca tggggaagat cgggctcgcc acttcgggct catgagcgct tgtttcggcg 780
 tgggtatggt ggccggcccc gtggcgggg gactgttggg cgccatctcc ttgcatgcac 840
 cattccttgc ggcggcggtg ctcaacggcc tcaacctact actgggctgc ttctaatgc 900
 aggagtgcga taaggagag cgtcgaccga tgcccttgag agccttcaac ccagtcagct 960
 ccttcgggtg ggcgcggggc atgactatcg tcgccgact tatgactgtc ttctttatca 1020
 tgcaactcgt aggacaggtg ccggcagcgc tctgggtcat tttcggcgag gaccgctttc 1080

gctggagcgc gacgatgatc ggccctgtcgc ttgcgggtatt cggaatcttg cagccctcgc 1140
 ctcaagcctt cgtcactggc cccgccacca aacgtttcgc cgagaagcag gccattatcg 1200
 ccggcatggc ggccgacgcg ctgggctacg tcttgctggc gttcgcgcag cgaggctgga 1260
 tggccttccc cattatgatt cttctcgtt cggcgccat cgggatgcc gcgttgacag 1320
 ccatgtgtc caggcaggta gatgacgacc atcagggaca gcttcaagga tcgctcgcgc 1380
 ctcttaccag cctaacttcg atcactggac cgttgatcgt cacggcgatt tatgccgcct 1440
 cggcgagcac atggaacggg ttggcatgga ttgtaggcgc cgccctatac cttgtctgcc 1500
 tccccgcgtt gcgtcgcggc gcatggagcc gggccacctc gacctgaatg gaagccggcg 1560
 gcacctcgct aacggattca ccactccaag aattggagcc aatcaattct tgcggagaac 1620
 tgtgaatgcy caaatgcgcc caatacgcaa accgcctctc cccgcgcgtt ggccgattca 1680
 ttaatgcagc tggcacgaca gggttcccga ctggaaagcg ggcagtgaac gcaacgcaat 1740
 taatgtgagt tagctcactc attaggcacc ccaggcttta cactttatgc ttccggctcg 1800
 tatgttggtg cgaattgtga gcgataaca ggttcaaca ggaacacgct atgaccgcct 1860
 ttacgccaaag cgcgcaatta accctcacta aagggaacaa aagctgggta cggggcccc 1920
 cctcgaggtc gacggatcg ataagcttga tatcgaattc ctgcagttag aggaattaac 1980
 catgtcatta ccgttcttaa cttctgcacc gggaaagggt attatttttg gtgaacactc 2040
 tgctgtgtac aacaagcctg ccgtcgtcgc tagtgtgtct gcgttgagaa cctacctgct 2100
 aataagcgag tcatctgcac cagatactat tgaattggac ttcccgaca ttagctttaa 2160
 tcataagtgg tccatcaatg atttcaatgc catcaccgag gatcaagtaa actcccaaaa 2220
 attggccaag gctcaacaag ccaccgatgg cttgtctcag gaactcgtta gtcttttggg 2280
 tccgttggtta gctcaactat ccgaatcctt ccactaccat gcagcgtttt gtttctgtga 2340
 tatgtttggt tgcctatgcc cccatgccaa gaatattaag ttttctttaa agtctacttt 2400
 acccatcggt gctgggttgg gctcaagcgc ctctatttct gtatcactgg ccttagctat 2460
 ggcctacttg ggggggttaa taggatctaa tgacttgga aagctgtcag aaaacgataa 2520
 gcatatagtg aatcaatggg ccttcatagg tgaagggtgt attcacggtt ccccttcagg 2580
 aatagataac gctgtggcca cttatggtaa tgccctgcta ttgaaaaag actcacataa 2640
 ttggaacaata aacacaaaca attttaagtt cttagatgat tcccagcca ttccaatgat 2700
 cctaacttat actagaattc caaggtctac aaaagatctt gttgctcgcg ttcgtgtgtt 2760
 ggtcaccgag aaatttctcg aagttatgaa gccaatctta gatgccatgg gtgaatgtgc 2820
 cctacaaggc ttagagatca tgactaagtt aagtaaatgt aaaggcaccg atgacgagcg 2880
 tgtagaaact aataatgaac tgtatgaaca actattggaa ttgataagaa taaatcatgg 2940
 actgcttgct tcaatcgggt tttctcatcc tggattagaa cttattaaaa atctgagcga 3000
 tgatttgaga attggtccca caaaacttac cgtgtcgtgt ggcggcggtt gctctttgac 3060
 tttgttacga agagacatta ctcaagacga aattgacagc ttcaaaaaga aattgcaaga 3120
 tgattttagt tacgagacat ttgaaacaga cttgggtggg actggctgct gttgtttaa 3180
 cgcaaaaaat ttgaataaag atcttaaaat caaatcccta gtattccaat tatttgaaaa 3240
 taaaactacc acaaaagcaac aaattgacga tctattattg ccaggaaaca cgaatttacc 3300
 atggacttca taggaggcag atcaaatgtc agagttagaa gccttcagtg cccaggggaa 3360
 agcgttacta gctggtggat atttagtttt agatacaaaa tatgaagcat ttgtagtcgg 3420
 attatcgga agaattgcag ctgtagccca tccctacggt tcattgcaag ggtctgataa 3480
 gtttgtaagt cgtgtgaaaa gtaaaccaatt aaatgggaag gagtggctgt accatataag 3540
 tcctaaaagt ggcttcattc ctgtttcgat aggcggatct aagaaccctt tcattgaaaa 3600
 agttatcgct aacgtattta gctactttaa acctaacatg gacgactact gcaatagaaa 3660
 cttgttcgtt attgatattt tctctgatga tgcctaccat tctcaggagg atagcgttac 3720
 cgaacatcgt ggcaacagaa gattgagttt tcattcgcac agaattgaag aagttcccaa 3780
 aacagggctg ggctcctcgc caggtttagt cacagtttta actacagctt tggcctcctt 3840
 tttgtatcg gacctggaat ataattgtaga caaatataga gaagttattc ataatttagc 3900
 acaagtgtgt cattgtcaag ctcaagggtaa aattgggaag ggggttgatg tagcggcggc 3960
 agcatatgga tctatcagat atagaagatt cccaccgcga ttaattctta atttgccaga 4020
 tattggaagt gctacttacg gcagtaaaact ggcgcatttg gttgatgaag aagactggaa 4080
 tattacgatt aaaagtaacc atttaccttc gggattaact ttatggatgg gcgatattaa 4140
 gaatggttca gaaacagtaa aactgggtcca gaaggtaaaa aattggtatg attcgcatat 4200
 gccagaaagc ttgaaaatat atacagaact ccatgacgat tacagcgatc agatatttga 4320
 actatctaaa ctagatcgct tacacgagac taatgacgat gaaatcacag aagtttaga 4380
 gtctcttgag aggaatgact gtacctgtca aaagtatcct gaaatcacag aagtttaga 4440
 tgcagttgcc acaattagac gttccttttag aaaaataact aaagaatctg gtgccgatat 4500
 cgaacctccc gtacaaacta gcttatttga tgattgccag accttaaaag gagttcttac 4560
 ttgcttaata cctggtgctg gtggttatga cgccattgca gtgattacta agcaagatgt 4620
 tgatcttagg gctcaaacgc ctaatgacaa aagattttct aaggttcaat ggctggatgt 4680
 aactcaggct gactgggggt ttaggaaaga aaaagatccg gaaacttatc ttgataaata 4740
 ggaggttaata ctcatgaccg tttacacagc atccgttacc gcaccgcgta acatcgcaac 4800
 ccttaagtat tgggggaaaa gggacacgaa gttgaactct cccaccaatt cgtccatatc 4860
 agtgacttta tcgcaagatg acctcagaac gttgacctct gcggctactg cactcgattt 4920
 tgaacgcgac actttgtggt taaatggaga accacacagc atcgacaatg aaagaactca 4980
 aaattgtctg cgcgacctac gccaatgaag aaaggaaatg gaatcgagg acgcctcatt 5040
 gccacatta tctcaatgga aactccacat tgtctccgaa aataactttc ctacagcagc 5100
 tgggttagct tctcgcgtc ctggctttgc tgcattgggt tctgcaattg ctaagttata

ccaattacca	cagtcaactt	cagaaatata	tagaatagca	agaaaggggt	ctggttcagc	5160
ttgtagatcg	ttgtttggcg	gatacgtggc	ctgggaaatg	ggaaaagctg	aagatgggtca	5220
tgattccatg	gcagtacaaa	tcgcagacag	ctctgactgg	cctcagatga	aagcttgtgt	5280
cctagttgtc	agcgatatta	aaaaggatgt	gagttccact	cagggtatgc	aattgaccgt	5340
ggcaacctcc	gaactattta	aagaaagaat	tgaacatgtc	gtaccaaaga	gatttgaagt	5400
catgcgtaaa	gccattgttg	aaaaagattt	cgccaccttt	gcaaaggaaa	caatgatgga	5460
ttccaactct	ttccatgccca	catgtttgga	ctctttccct	ccaatattct	acatgaatga	5520
cacttccaag	cgtatcatca	gttggtgccca	caccattaat	cagttttacg	gagaaacaat	5580
cgttgcatatc	acgtttgatg	cagggtccaaa	tgctgtgttg	tactacttag	ctgaaaatga	5640
gtcgaaactc	tttgcattta	tctataaatt	gtttggctct	gttcctggat	gggacaagaa	5700
atttactact	gagcagcttg	aggctttcaa	ccatcaattt	gaatcatcta	actttactgc	5760
acgtgaattg	gatcttgagt	tgcaaaagga	tgttgccaga	gtgattttaa	ctcaagtcgg	5820
ttcaggccca	caagaaacaa	acgaatcttt	gattgacgca	aagactgggtc	taccaaagga	5880
ataactggag	caggggagga	ggattactat	atgcaaacgg	aacacgtcat	tttattgaat	5940
gcacagggag	ttcccacggg	tacgctggaa	aagtatgccg	cacacacggc	agacaccgcg	6000
ttacatctcg	cgttctccag	ttggctgttt	aatgccaaag	gacaattatt	agttaccgcg	6060
cgcgactga	gcaaaaaagc	atggcctggc	gtgtggacta	actcggtttg	tgggcacca	6120
caactgggag	aaagcaacga	agacgcagtg	atccgcggtt	gccgttatga	gcttggcggtg	6180
gaaattacgc	ctcctgaatc	tatctatcct	gactttcgct	accgcgccac	cgatcccgagt	6240
ggcattgttg	aaaatgaagt	gtgtccggta	tttgccgcac	gcaccactag	tcggttacag	6300
atcaatgatg	atgaagtgat	ggattatcaa	ttggtgtgatt	tagcagatgt	attacacggt	6360
attgatgccca	cgccgtgggc	gttcagtcgg	tggatgggtga	tgcaggcgac	aaatcgcgaa	6420
gccagaaaac	gattatctgc	atttaccag	cttaataaac	ccgggggatc	cactagttct	6480
agagcggccg	ccaccgcgga	ggaggaatga	gtaatggact	ttccgcagca	actcgaagcc	6540
tgcgtaagc	aggccaacca	ggcgctgagc	cgttttatcg	ccccactgcc	ctttcagaac	6600
actcccgttg	tcgaaaccat	gcagtatggc	gcattattag	gtggtaagcg	cctgcgacct	6660
ttcctgggtt	atgccaccgg	tcatatgttc	ggcgtttagca	caaacacgct	ggacgcaccc	6720
gctgcgcg	ttgagtgtat	ccacgcttac	tcattaatc	atgatgattt	accggcaatg	6780
gatgatgacg	atctgcgtcg	cggtttgcca	acctgccatg	tgaagtttg	cgaagcaaac	6840
gcgattctcg	ctggcgacgc	tttacaacg	ctggcggttct	cgattttaag	cgatgccgat	6900
atgccggaag	tgctggaccg	cgacagaatt	tcgatgattt	ctgaactggc	gagcgccagt	6960
ggtattgccc	gaatgtgcgg	tggtcaggca	ttagatttag	acgcggaagg	caaacacgta	7020
cctctggacg	cgcttgagcg	tattcatcgt	cataaaaccg	gcgcattgat	tcgcgcgcgc	7080
gttcgccttg	gtgcattaag	cgccggagat	aaaggacgtc	gtgctctgcc	ggtactcgac	7140
aagatgcgag	caagcatcgg	ccttgccctc	caggttcagg	atgacatcct	ggatgtggtg	7200
ggagatactg	caacgttggg	aaaacgccag	gggtccgacc	agcaacttgg	taaaagtacc	7260
taccctgcac	ttctgggtct	tgagcaagcc	cggaagaaag	cccgggatct	gatcgacgat	7320
gcccgtcagt	cgctgaaaca	actggctgaa	cagtcactcg	atacctcggc	actggaagcg	7380
ctagcggact	acatcatcca	gcgtaataaa	taagagctcc	aattcgccct	atagttagtc	7440
gtattacgcg	cgctcactgg	ccgtcgtttt	acaacgtcgt	gactgggaaa	accctggcgt	7500
taccacaact	aatgccttg	cagcacatcc	ccctttcgcc	agctggcgta	atagcgaaga	7560
ggcccgacc	gatcgccctt	cccaacagtt	cgcgagcctg	aatggcgaat	ggaaattgta	7620
agcgtaata	ttttgttaaa	atcgcgtta	aatttttgtt	aaatcagctc	attttttaac	7680
caataggccg	actgcgatga	gtggcagggc	ggggcgtaat	ttttttaagg	cagttatttg	7740
tgcccttaaa	cgctgtgtgc	tacgcctgaa	taagtataaa	taagcggatg	aatggcagaa	7800
attcgaaagc	aaattcgacc	cggctcgtcg	ttcagggcag	ggctgttaaa	tagccgctta	7860
tgtctattgc	tggtttaccg	gtttattgac	taccggaagc	agtgtgaccg	tgtgcttctc	7920
aaatgctctg	ggcagtttg	ctcaggctct	ccccgtggag	gtaataattg	acgatattgat	7980
caattattct	ggcctccaga	gcctgataaa	aacggtgaat	ccgttagcga	ggtgcgcgcg	8040
gcttccattc	aggtcgaggt	ggcccggtc	catgcaccgc	gacgcaacgc	ggggaggcag	8100
acaaggtata	ggggcgcgag	gcggtacag	ccgatagtct	ggaacagcgc	acttacgggt	8160
tgctgcgcaa	cccaagtgtc	accggcgcg	cagcgtgacc	cgtgtcggcg	gctccaacgg	8220
ctcgccatcg	tcagaaaac	acggctcatc	gggcatcgcc	aggcgctgct	gcccgcgcgc	8280
ttcccattec	tcggtttcgg	tcaaggctgg	caggctctgg	tccatgcccg	gaatgccggg	8340
ctggctgggc	ggctcctcgc	cggggccggt	cggtagtgtc	tgctcgcccg	gatacagggt	8400
cgggagtcgg	cgcaggtcgc	catgccccaa	cagcgattcg	tctgtgtcgt	cgtgatcaac	8460
caccacggcg	gcactgaaca	ccgacaggcg	caactggtcg	cggggctggc	cccacgccac	8520
gcggtcattg	accacgtagg	ccgacacggt	gccggggccg	ttgagcttca	cgacggagat	8580
ccagcgctcg	gccaccaagt	ccttgactgc	gtattggacc	gtccgcaaag	aacgtccgat	8640
gagcttgga	agtgtcttct	ggctgaccac	cacggcgctt	tggtggccca	tctgcgccac	8700
gaggtgatgc	agcagcattg	ccgcctggg	tttccctcga	ataagcccg	cccacgcctc	8760
atgcgctttg	cgttcggttt	gcacccagtg	accgggcttg	ttcttggtct	gaatgccgat	8820
ttctctggag	tgcttgccca	tgcttatctc	catcggtag	ggtgcgcgac	ggttgccgca	8880
ccatgcgcaa	tcagctgcaa	cttttcggca	cgcgacaac	aattatgcgt	tgcgtaaaag	8940
tggcagtcga	ttacagattt	tctttaacct	acgcaatgag	ctattgcggg	gggtgcgcga	9000
atgagctggt	gcgtaccccc	cttttttaag	ttgttgattt	ttaagtcttt	cgcatttcgc	9060
cctatatcta	gttctttggt	gccccaaaga	gggcaccctt	gcggggttcc	cccacgcctt	9120

cggcgcggtt cccctccgg caaaaagtgg cccctccggg gcttgttgat cgactgcgcg 9180
 gccttcggcc ttgcccagg tggcgctgcc ccttggaac ccccgactc gccgcccgtga 9240
 ggctcggggg gcaggcgggc gggcttcgcc ttcgactgcc cccactcgca taggcttggg 9300
 tcgttccagg cgcgtcaagg ccaagccgct gcgcggtcgc tgcgcgagcc ttgaccgcgc 9360
 ttccacttgg tgtccaaccg gcaagcgaag cgcgcaggcc gcaggccgga ggcttttccc 9420
 cagagaaaat taaaaaatt gatggggcaa ggccgcaggc cgcgcagttg gagccggtgg 9480
 gtatgtggtc gaaggctggg tagccggtgg gcaatccctg tggtaagct cgtgggcagg 9540
 cgcagcctgt ccatcagctt gtccagcagg gttgtccacg ggccgagcga agcgagccag 9600
 ccggtggccg ctgcgggcca tgcgtccacat atccacgggc tggcaaggga gcgcagcgac 9660
 cgcgcagggc gaagcccgga gagcaagccc gtagggcgcc gcagccgccg taggcggtca 9720
 cgactttgcg aagcaaagt tagtgagtat actcaagcat tgagtggccc gccggaggga 9780
 ccgcttggcg ctgccccgt cgagccggtt ggacaccaa agggaggggc aggcattggcg 9840
 gcatacgcga tcatgcatg caagaagctg gcgaaaatgg gcaacgtggc ggccagtcctc 9900
 aagcacgcct accgcgagcg cgagacgccc aacgctgacg ccagcaggac gccagagaac 9960
 gagcactggg cggccagcag caccgatgaa gcgatgggccc gactgcgcga gttgctgcca 10020
 gagaagcggc gcaaggacgc tgtgttggcg gtcgagtacg tcatgacggc cagcccggaa 10080
 tgggtgaagt cggccagcca agaacagcag gcggcggtct tgcagaaggc gcacaagtgg 10140
 ctggcggaca agtacggggc ggatcgcacg gtgacggcca gcatccaccg tgacgaaacc 10200
 agcccgcaca tgaccgcgtt cgtggtgccg ctgacgcagg acggcaggct gtcggccaag 10260
 gaggttcatcg gcaacaaagc gcagatgacc cgcgaccaga ccacgtttgc ggccgcttgc 10320
 gccgatctag ggctgcaacg gggcatcgag ggcagcaagg cactgcacac gcgcattcag 10380
 gcgttctacg aggccctgga gcggccacca gtgggcccacg tcaccatcag cccgcaagcg 10440
 gtcgagccac gcgcctatgc accgcaggga ttggccgaaa agctgggaat ctcaaagcgc 10500
 gttgagacgc cggaagccgt ggccgaccgg ctgacaaaag cggttcggca ggggtatgag 10560
 cctgccttac aggcccgccg aggagcgcg gtgatgcgca agaaggccga tcaagcccaa 10620
 gagacggccc gag 10633

<210> 63

<211> 4263

<212> DNA

<213> Sequência Artificial

<220>

<223> Polinucleotídeo recombinante

<400> 63

cttgatatcg aattcctgca gcccggggat cctctagagt cgactaggag gaatataaaa 60
 tgaataattg tgtcatcgct agtgcggtac gtactgctat cggtagtttt aacggttcac 120
 tcgcttcac cagcgccatc gacctggggg cgacagtaat taaagccgcc attgaacgtg 180
 caaaaatcga ttcaaacac gttgatgaag tgattatggg taacgtgtta caagccgggc 240
 tggggcaaaa tccggcgctg caggcactgt taaaaagcgg gctggcagaa acggtgtgctg 300
 gattcacggt caataaagta tgtggttcgg gtcttaaaag tgtggcgctt gccgcccagg 360
 ccattcaggc aggtcaggcg cagagcattg tggcgggggg tatggaaaat atgagtttag 420
 cccctactt actcgatgca aaagcacgct ctggttatcg tcttgagac ggacaggttt 480
 atgacgtaat cctgcgcgat ggctgatgt gcgccacca tggttatcat atggggatta 540
 ccgcccgaaca cgtggctaaa gactacggaa ttaccgctga aatgcaggat gaactggcgc 600
 taccattaca cgtaaaagcg gcagccgcaa ttagtccgg tgcttttaca gccgaaatcg 660
 tcccggtaaa tgtgtcact cgaagaaaaa ccttcgtctt cagtcaagac gaattcccga 720
 aagcgaattc aacggctgaa gcgttaggtg cattgcgccc ggcttcgat aaagcaggaa 780
 cagtcaccgc tgggaacgcg tctggtatta acgacggtgc tgccgctctg gtgattatgg 840
 aagaatctgc ggcgctggca gcaggcctta cccctcggc tcgcattaaa agttatgcca 900
 gcggtggcgt gcccccgca ttgatgggta tggggccagt acctgccacg caaaaagcgt 960
 tacaactggc ggggctgcaa ctggcggata ttgatctcat tgaggctaat gaagcatttg 1020
 ctgcacagtt cctgcccgtt gggaaaaacc tgggcttga ttctgagaaa gtgaattgca 1080
 acggcggggc catcgcgctc gggcatccta tgggtgccag tgggtgctctg attctggtca 1140
 cactattaca tgccatgcag gcacgcgata aaacgctggg gctggcaaca ctgtgcattg 1200
 gcggcgggtca ggaattgctg atggtgattg aacggttgaa ttaaggagga cagctaaatg 1260
 aaactctcaa ctaaaactttg ttggtgtggt attaaaggaa gacttaggcc gcaaaagcaa 1320
 caacaattac acaatacaaa cttgcaaatg actgaactaa aaaaacaaa gaccgctgaa 1380
 caaaaaacca gacctcaaaa tgtcgggtatt aaaggtatcc aaatttacct ccaactcaa 1440
 tgtgtcaacc aatctgagct agagaaattt gatggcgttt ctcaaggtaa atacacaatt 1500
 ggtctggggc aaaccaacat gtcttttgtc aatgcagag aagatatcta ctcgatgtcc 1560
 ctaactgttt tgtctaagtt gatcaagagt tacaacatcg acaccaacaa aattggtaga 1620
 ttagaagtcg gtactgaaac tctgattgac aagccaagt ctgtcaagtc tgtcttgatg 1680
 caattgtttg gtgaaaacac tgacgtcgaa ggtattgaca cgcttaatgc ctgttacggt 1740

```

ggtaccaacg cggtgtttcaa ctctttgaac tggattgaat ctaacgcatg ggatggtaga 1800
gacgccattg tagttttgcgg tgatattgcc atctacgata aggggtgccgc aagaccaacc 1860
ggtggtgccg gtactgttgc tatgtggatc ggtcctgatg ctccaattgt atttgactct 1920
gtaagagcctt cttacatgga acacgcctac gattttttaca agccagattt caccagcgaa 1980
tatecttacg tcgatgggtca tttttcatta acttggttacg tcaaggctct tgatcaagtt 2040
tacaagagtt attccaagaa ggctatttct aaagggttg ttagcgatcc cgctgggttcg 2100
gatgctttga acgttttgaa atatttcgac tacaacgttt tccatgttcc aacctgtaaa 2160
ttggtcacaa aatcatcacg tagattacta tataacgatt tcagagccaa tcctcaattg 2220
ttcccagaag ttgacgccga attagctact cgcgattatg acgaatcttt aaccgataag 2280
aacattgaaa aaacttttgt taatgttgct aagccattcc acaaagagag agttgcccaa 2340
tctttgattg ttccaacaaa cacaggtaac atgtacaccg catctgttta tgccgccttt 2400
gcatctctat taaactatgt tggatctgac gacttacaag gcaagcgtgt tggtttattt 2460
tcttacgggt cgggttttagc tgcattctta tattcttgca aaattgttg tgacgtccaa 2520
catattatca aggaattaga tattactaac aaattagcca agagaatcac cgaaactcca 2580
aagattacg aagctgccat cgaattgaga gaaaatgccc atttgaagaa gaacttcaaa 2640
cctcaagggt ccattgagca ttgcaaaagt ggtgtttact acttgacca catcgatgac 2700
aaatttagaa gatcttacga tgttaaaaaa taaggaggat tacactatgg ttttaacca 2760
taaaacagtc atttctggat cgaaagtcaa aagtttatca tctgcgcaat cgagctcatc 2820
aggaccttca tcatctagtg aggaagatga ttcccgcat attgaaagct tggataagaa 2880
aatacgtcct ttagaagaat tagaagcatt attaagtagt ggaaatacaa acaattgaa 2940
gaacaaagag gtcgctgcct tggttattca cggtaagtta ctttgttacg ctttggagaa 3000
aaaattaggt gatactacga gagcggttgc ggtacgtagg aaggctcttt caattttggc 3060
agaagctcct gtattagcat ctgatcgttt accatataaa aattatgact acgaccgct 3120
atltggcgct tgttgtgaaa atgttatagg ttacatgcct ttgcccgttg gtgttatagg 3180
ccccttggtt atcgatggta catcttatca tataccaatg gcaactacag aggggtgttt 3240
ggtagcttct gccatgcgtg gctgtaaggc aatcaatgct ggcggtggtg caacaactgt 3300
tttaactaag gatggtatga caagaggccc agtagtccgt ttcccaactt tgaaaagatc 3360
tgggtgctgt aagatatggt tagactcaga agagggacaa aacgcaatta aaaaagcttt 3420
taactctaca tcaagatttg cactctgca acatattcaa acttgtctag caggagattt 3480
actcttcatg agatttagaa caactactgg tgacgcaatg ggtatgaata tgatttctaa 3540
agggtgctgaa tactcattaa agcaaatggt agaagagtat ggctgggaag atatggaggt 3600
tgtctccgtt tctggtaact actgtaccga caaaaaacca gctgccatca actggatcga 3660
aggctcgtgt aagagtgtcg tcgcagaagc tactattcct ggtgatgttg tcagaaaagt 3720
gttaaaaagt gatgtttccg cattgggtga gttgaacatt gctaagaatt tggttggatc 3780
tgcaatggct gggctctgtg gtggatttaa cgcacatgca gctaatttag tgacagctgt 3840
tttcttggca ttaggacaag atcctgcaca aaatgttgaa agttccaact gtataacatt 3900
gatgaaagaa gtggacggtg atttgagaat ttccgtatcc atgccatcca tcgaagtagg 3960
taccatcgtt ggtggtactg ttctagaacc acaagggtgcc atgttggtact tattaggtgt 4020
aagaggcccg catgctaccg ctctgtgtac caacgcacgt caattagcaa gaatagttgc 4080
ctgtgccgtc ttggcagggtg aattatcctt atgtgtgtcc ctagcagccg gccatttgggt 4140
tcaaagtc atgaccaca acaggaaacc tgctgaacca acaaaaccta acaatttggg 4200
cgccactgat ataaatcgtt tgaaagatgg gtccgtcacc tgcattaaat cctaagtcga 4260
cct

```

REIVINDICAÇÕES

1. Ácido nucléico compreendendo, na ordem de 5' a 3' e em ligação operável, uma seqüência de nucleotídeo que codifica um domínio selecionado a partir de um domínio de transmembrana, um domínio de secreção, um domínio de solubilização, e uma proteína de inserção de membrana, e uma seqüência de nucleotídeo que codifica uma enzima de citocromo P450, no qual referido domínio é heterólogo à enzima de citocromo P450.
2. Ácido nucléico, de acordo com a reivindicação 1, no qual o domínio de transmembrana é funcional em uma célula hospedeira procariótica.
3. Ácido nucléico, de acordo com a reivindicação 1, no qual a enzima de citocromo P450 que codifica seqüência de nucleotídeo é de códon otimizado para expressão em uma célula hospedeira procariótica.
4. Ácido nucléico, de acordo com a reivindicação 1, no qual a enzima de citocromo P450 é uma enzima de modificação de precursor de isoprenóide que catalisa modificação de um precursor de isoprenóide.
5. Ácido nucléico, de acordo com a reivindicação 4, no qual a modificação é selecionada de oxidação, hidroxilação, e epoxidação.
6. Ácido nucléico, de acordo com a reivindicação 1, compreendendo adicionalmente uma seqüência de nucleotídeo que codifica um citocromo P450 reductase.
7. Vetor de expressão compreendendo o ácido nucléico de acordo com a reivindicação 1.
8. Célula hospedeira compreendendo o vetor de expressão de acordo com a reivindicação 7.
9. Célula hospedeira, de acordo com a reivindicação 8, no qual referida célula hospedeira é uma que não produz normalmente isopentenil pirofosfato via uma trajetória de mevalonato.
10. Célula hospedeira, de acordo com a reivindicação 9, no qual referida célula hospedeira é procariótica.
11. Célula hospedeira, de acordo com a reivindicação 8, no qual

a célula hospedeira compreende adicionalmente um ácido nucléico compreendendo uma seqüência de nucleotídeo que codifica um terpeno sintase heterólogo.

5 12. Célula hospedeira, de acordo com a reivindicação 8, no qual a célula hospedeira compreende adicionalmente um ácido nucléico compreendendo uma seqüência de nucleotídeo que codifica um citocromo P450 reductase.

10 13. Célula hospedeira, de acordo com a reivindicação 9, no qual referida célula hospedeira é geneticamente modificada com um ou mais ácidos nucléicos compreendendo seqüências de nucleotídeo que codificam duas ou mais enzimas de trajetória de mevalonato.

14. Método de produção de um produto de trajetória biosintética em uma célula hospedeira, o método compreendendo:

15 cultura de uma célula hospedeira geneticamente modificada em um meio adequado, no qual referida célula hospedeira é geneticamente modificada com um ácido nucléico compreendendo uma seqüência de nucleotídeo que codifica enzima de citocromo P450 operavelmente ligada a um domínio selecionado a partir de um domínio de transmembrana, um domínio de secreção, um domínio de solubilização, e uma proteína de inserção de membrana, para produzir uma enzima de citocromo P450 modificada, enzimaticamente ativa,

20

no qual, na presença de um intermediário de trajetória biosintética, referida produção de referida enzima de citocromo P450 modificada resulta em modificação enzimática do intermediário de trajetória biosintética e produção do produto de trajetória biosintética.

25

15. Método, de acordo com a reivindicação 14, no qual referida enzima de citocromo P450 é uma enzima de modificação de precursor de isoprenóide e no qual, na presença de um composto de precursor de isoprenóide, referida produção de referida enzima de modificação de precursor de isoprenóide resulta em modificação enzimática do precursor de isoprenóide, e produção do composto de isoprenóide.

30

16. Método, de acordo com a reivindicação 14, no qual referida

célula hospedeira é uma célula hospedeira eucariótica.

17. Método, de acordo com a reivindicação 16, no qual referida célula hospedeira é um célula de levedura.

18. Método, de acordo com a reivindicação 16, no qual referida
5 célula hospedeira é uma célula de planta.

19. Método, de acordo com a reivindicação 14, no qual referida célula hospedeira é uma célula procariótica.

20. Método, de acordo com a reivindicação 15, no qual referida célula hospedeira é adicionalmente geneticamente modificada com um ácido
10 nucléico compreendendo uma seqüência de nucleotídeo que codifica um terpeno sintase heterólogo, no qual referida cultura proporciona produção de referido terpeno sintase, no qual referido terpeno sintase modifica um poliprenil pirofosfato para gerar um substrato para referida enzima de modificação de isoprenóide.

21. Método, de acordo com a reivindicação 20, no qual referido poliprenil pirofosfato é selecionado de farnesil pirofosfato, geranil pirofosfato,
15 and geranilgeranil pirofosfato.

22. Método, de acordo com a reivindicação 14, no qual referida célula hospedeira é adicionalmente geneticamente modificada com um ácido
20 nucléico compreendendo uma seqüência de nucleotídeo que codifica um citocromo P450 reductase (CPR).

23. Método, de acordo com a reivindicação 15, no qual referida célula hospedeira é uma que não sintetiza normalmente isopentenil pirofosfato (IPP) via uma trajetória de mevalonato, e no qual a célula hospedeira é
25 geneticamente modificada com um ou mais ácidos nucléicos compreendendo as seqüências de nucleotídeo que codificam duas ou mais enzimas na trajetória de mevalonato, uma isomerase de IPP, uma preniltransferase, e um terpeno sintase, referida cultura proporcionando produção das enzimas de trajetória de mevalonato, no qual referida produção de referidas duas ou
30 mais enzimas de trajetória de mevalonato, referida isomerase de IPP, referida preniltransferase, referido terpeno sintase, e referida enzima de modificação de precursor de isoprenóide resulta na produção de um composto de

isoprenóide.

24. Método, de acordo com a reivindicação 23, no qual referidas duas ou mais enzimas de trajetória de mevalonato compreendem mevalonato kinase, fosfomevalonato kinase, e mevalonato pirofosfato decarboxilase, e
5 no qual a célula hospedeira é cultivada na presença de mevalonato.

25. Método, de acordo com a reivindicação 23, no qual referidas duas ou mais enzimas de trajetória de mevalonato compreendem acetoacetil-CoA tiolase, hidroximetilglutaril-CoA sintase, hidroximetilglutaril-CoA redutase, mevalonato kinase, fosfomevalonato kinase, e mevalonato pirofosfato
10 decarboxilase.

26. Método, de acordo com a reivindicação 14, no qual referida enzima de citocromo P450 que codifica seqüência de nucleotídeo é operavelmente ligada a um promotor indizível.

27. Método, de acordo com a reivindicação 15, no qual o composto de isoprenóide é produzido em uma quantidade de pelo menos cerca
15 de 10 mg por litro.

28. Método, de acordo com a reivindicação 14, compreendendo adicionalmente isolar o produto de trajetória biosintética.

FIG. 1

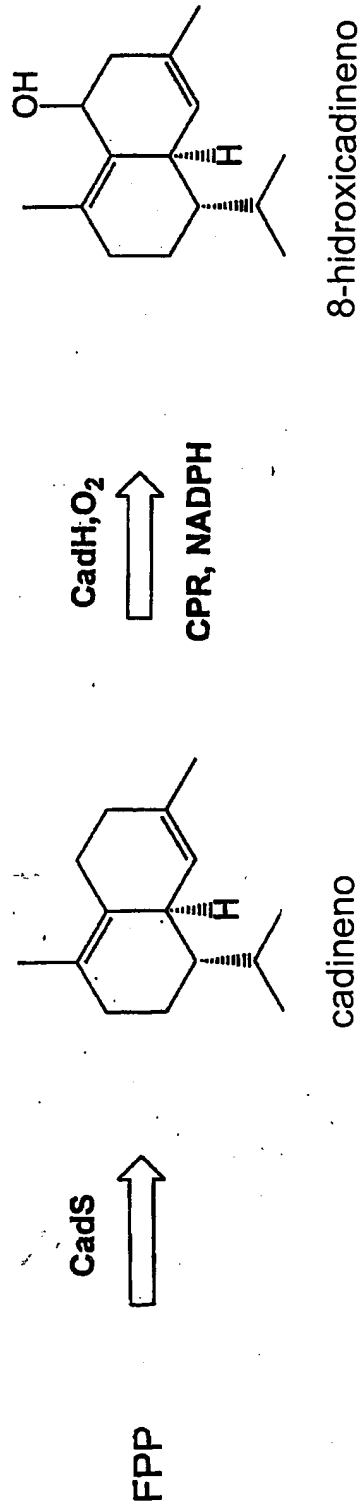


FIG. 2

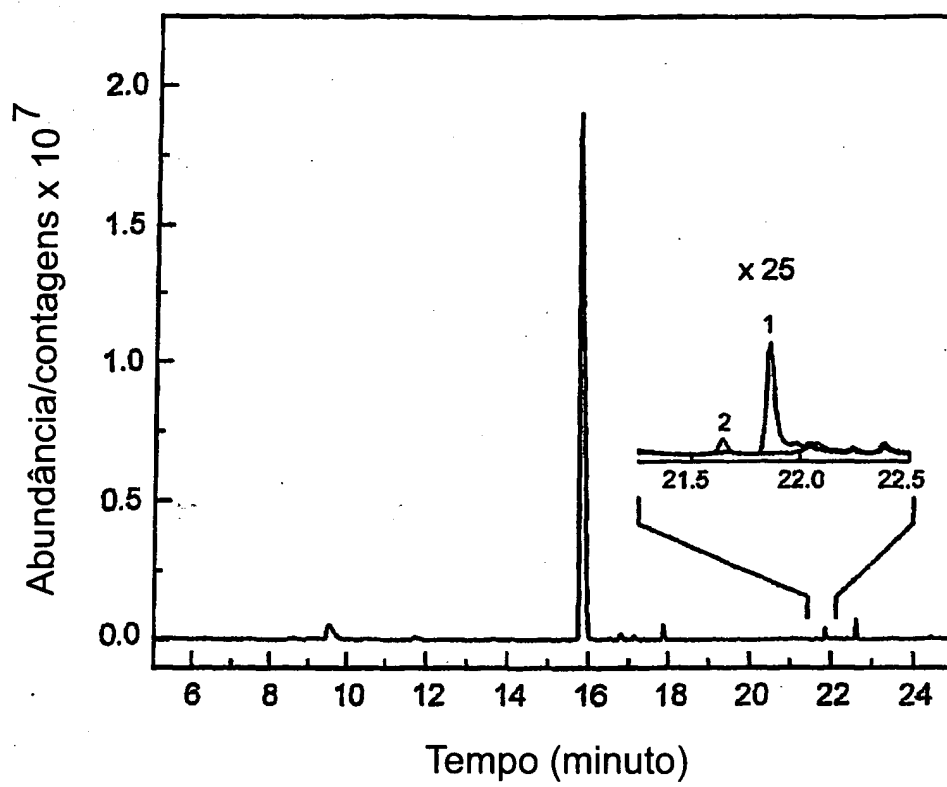


FIG. 3

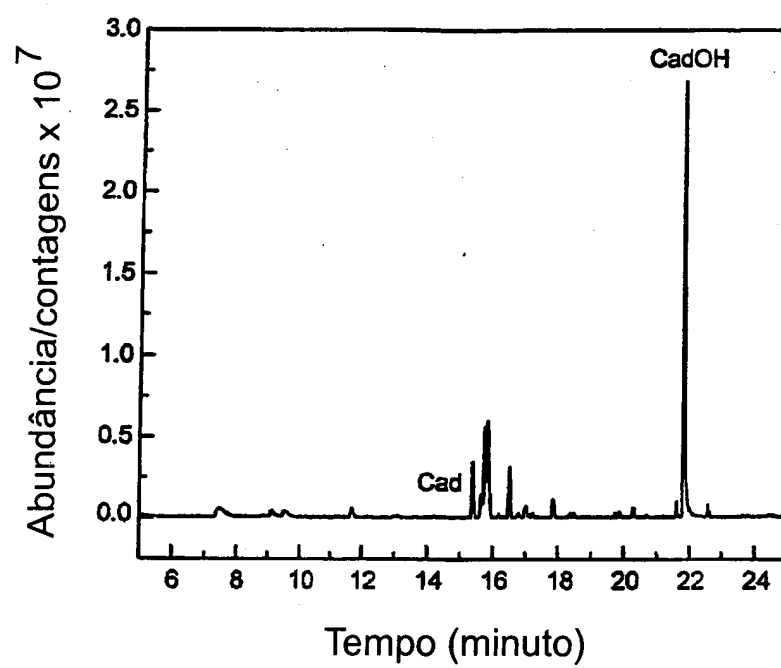


FIG. 4A

tipo selvagem (domínio TM)
 CYP52A13 (*C. tropicalis*, A13)
 CYP52A17 (*C. tropicalis*, A17)
 P45017 α (*Bos taurus*, bovino)
 truncado
 OmpA-A17
 Peptitergent (PD1):
 mistic

MLFPVALSFLVAILGISLWHVWT (SEQ ID NO:23)
 MTVHDIIATYFTKQYVIVPLALIAVRVLDYFY (SEQ ID NO:14)
 MIEQLLEYWYVVPVLYIIRQLLAYTK (SEQ ID NO:5)
 MALLLAVFLGLSCLLLLSLW (SEQ ID NO:2)
 -
 MKKTAIAIAVALAGFATVAQA-A17 TM domain (SEQ ID NO:6)
 EELLKQALQQAQQLLQQAQELAKK (SEQ ID NO:13)
 Fusão de proteína

FIG. 4B

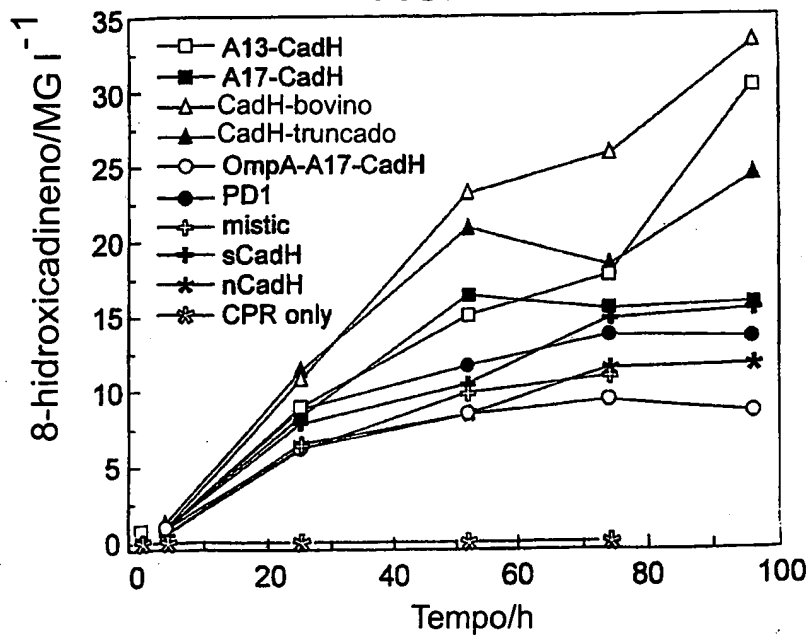


FIG. 5

Nº de Acesso ao Banco de Gene. AAX20121 (AY874162)

MFCTFFEKHRKWDILLEKSTGVMEAMKVTSEEKEQLSTAIDRMNEGLDAFIQLYNESE
IDEPLIQLDDDTAELMKQARDMYGQEKLNKLNNTIIKQILSISVSEEKEKE (SEQ ID
NO:24)

FIG. 6

Mentha x gracilis Limonene-6-Hydroxylase (N° de Acesso ao Banco de Gene. AAQ18706; and AY281025)

MELDLLSAILLVATYIVSLLINQWRKSKSQONLPPSPPKLPVIGHLHFLWGGLPQHVFERSIAQKYGPAH
VOLGEVYSVVLSSAEAAQAMKVLDPNFADRFDGIGSRTMWYDKDDIIFSPYNDEWRQMRRICVTELLSPK
NVRSGYIRQEEIERLIRLLGSSGGAPVDVTEEVSKMSCVVVCRAAFGSVLKDQGS LAELVKESLALASGF
ELADLYPSSWLLNLLSLNKYRLQMRRLDHILDGFL EEHREKKSGEFGGEDIVDVLFRMQKGS DIKIPIT
SNCKGFIFDTFSAGAETSSTTISWALSELMRNPAKMAKVQAEVREALKGKTVVDLSEVQELKYLRSVLKE
TLRLHPPFPLIPRQSREECEVNGYTIPAKTRIFINVWAIGRDPQYWEDPDTRPERFDEVSRDFMGNDFEF
IPFGAGRRICPGLHFGLANVEIPLAQLLYHFDWKLPQGMTDADLDMTETPGLSGPKKKNVCLVPTLYKSP

(SEQ ID NO:25)

FIG. 7

Nicotiana tabacum 5-*epi*-aristolochene dihydroxylase (N° de Acesso ao Banco de Gene
AAK62342; and AF368376)

MQFFSLVSIFLFLAFLFLLRKWKNSNSQSKKLPPGPWKIPILGSMLHMIGGEPHHVLRD
LAKKYGPLMHLQLGEISAVVVTSRDMKEVLKTHDVVFASRPKIVAMDIICYNQSDIAF
SPYGDHWRQMRKICVMELLNAKNVRSFSSIRRDEVVRLIDSIRSDSSSGELVNFTQRII
WFASSMTCRSAFGQVLKGQDIFAKKIREVIGLAEGFDVVDIFPTYKFLHVLSGMKRKL
NAHLKVDAIVEDVINEHKKNLAAAGKSNGALEDMFAAGTETSSTTTVWAMAEMMKNPVSF
TKAQAEVREAFRDKVSFDENDVEELKYLKLVIKETLRLHPPSPLLVPRECREDTDINGY
TIPAKTKVMVNVWALGRDPKYWDDAESFKPERFEQCSVDFFGNNEFLPFGGGGRRICPG
MSEGLANLYLPLAQLLYHFDWKLPTGIMPRDLDTELSGITIARKGDLYLNATPYQPSR
E (SEQ ID NO:26)

FIG. 8A

Gossypium arboreum native (Tipo selvagem) P450 monooxygenase
(δ -cadinene-8-hydroxylase; CadH); Banco de Gene AF332974 and
AAK60517

MLQIAFSSYSWLLTASNQKDGMLFPVALSFLVAILGISLWHVWTIRKPKKDIAPLPPGP
RGLPIVGYPYLGTDLNLHLVFTDLAAAYGPIYKLWLGKLCVVISSAPLAKEVVRDNDI
TFSERDPPVCAKIITFGLNDIVFDSYSSPDWRMKRKVLVREMLSHSSIKACYGLRREQV
LKGVQNVQAQSAGKPIDFGETAFLTSINAMMSMLWGGKQGGGERKGADVWGQFRDLITELM
VILGKPNVSDIFPVLARFDIQGLEKEMTKIVNSFDKLFNSMIEERENFSNKLSKEDGNT
ETKDFLQLLLDLKQKNDSGISITMNQVKALLMDIVVGGTDTTSTMMEWMAELIANPEA
MKKVQKEIDDVVGSDGAVDETHLPKRLRYLDAAVKETFRLHPPMPLLVP RCPGDSSNVGG
YSVPKGTRVFLNIWCIQRDPQLWENPLEFKPERFLT DHEKLDYLGND SRYMPFGSGRRM
CAGVSLGEKMLYSSLAAMIHAYDWNLADGEENDLIGLFGIIMKKKKPLILVPTPRPSNL
QHMYK (SEQ ID NO:27)

FIG. 8B

MALLAVFLGLSCLLLSLWIRKPKKDIAPLPPGP RGLPIVGYPYLGTDLNLHLVFTDL
AAAYGPIYKLWLGKLCVVISSAPLAKEVVRDNDITFSERDPPVCAKIITFGLNDIVFD
SYSSPDWRMKRKVLVREMLSHSSIKACYGLRREQVLKG VQNVQAQSAGKPIDFGETAFLT
SINAMMSMLWGGKQGGGERKGADVWGQFRDLITELMVILGKPNVSDIFPVLARFDIQGLE
KEMTKIVNSFDKLFNSMIEERENFSNKLSKEDGNTETKDFLQLLLDLKQKNDSGISITM
NQVKALLMDIVVGGTDTTSTMMEWMAELIANPEAMKKVKQEI DDVVGSDGAVDETHLP
KRLRYLDAAVKETFRLHPPMPLLVP RCPGDSSNVGGYSVPKGTRVFLNIWCIQRDPQLWE
NPLEFKPERFLT DHEKLDYLGND SRYMPFGSGRRMCAGVSLGEKMLYSSLAAMIHAYDWN
NLADGEENDLIGLFGIIMKKKKPLILVPTPRPSNLQHMYK (SEQ ID NO:28)

FIG. 8C

EELLKQALQQAQQLLQQAQELAKKIRKPKKDIAPLPPGP RGLPIVGYPYLGTDLNLHLV
FTDLAAAYGPIYKLWLGKLCVVISSAPLAKEVVRDNDITFSERDPPVCAKIITFGLND
IVFDSYSSPDWRMKRKVLVREMLSHSSIKACYGLRREQVLKG VQNVQAQSAGKPIDFGET
AFLTSINAMMSMLWGGKQGGGERKGADVWGQFRDLITELMVILGKPNVSDIFPVLARFDI
QGLEKEMTKIVNSFDKLFNSMIEERENFSNKLSKEDGNTETKDFLQLLLDLKQKNDSGI
SITMNQVKALLMDIVVGGTDTTSTMMEWMAELIANPEAMKKVKQEI DDVVGSDGAVDE
THLPKRLRYLDAAVKETFRLHPPMPLLVP RCPGDSSNVGGYSVPKGTRVFLNIWCIQRDP
QLWENPLEFKPERFLT DHEKLDYLGND SRYMPFGSGRRMCAGVSLGEKMLYSSLAAMIH
AYDWNLADGEENDLIGLFGIIMKKKKPLILVPTPRPSNLQHMYK (SEQ ID NO:29)

FIG. 8D

**MKKTAAIAVALAGFATVAQALLEYWYVVVPVLYIIKQLLAYTKIRKPKKDIAPLP
RGLPIVGYLPYLGTDLNLHLVFTDLAAAYGPIYKLWLGNKLCVVISSAPLAKEVVRNDI
TFSERDPPVCAKIITFGLNDIVFDSYSSPDWRMKRKVLVREMLSHSSIKACYGLRREQV
LKGVQNVASAGKPIDFGETAFLTSINAMMSMLWGGKQGGGERKGADVWGQFRDLITELM
VILGKPNVSDIFPVLARFDIQGLEKEMTKIVNSFDKLFNSMIEERENFSNKLSKEDGNT
ETKDFLQLLLDLKQKNDSGISITMNQVKALLMDIVVGGTDTTSTMEWTMAELIANPEA
MKKVKEIDDVVGSDGAVDETHLPKRLRYLDAAVKETFRLHPPMPLLVPKPGDSSNVGG
YSVPKGT RVFLNIWCIQRDPQLWENPLEFKPERFLTDHEKLDYLGNDSTRYPFGSGRRM
CAGVSLGEKMLYSSLAAMIHAYDWNLADGEENDLIGLFGIIMKKKKPLILVPTPRPSNL
QHVMK (SEQ ID NO:30)**

FIG. 9A

Taxus cuspidata taxadiene-5 α -hydroxylase (N° de Acesso ao Banco de Gene . AY289209 and AAQ56240)

MDALYKSTVAKFNEVTQLDCSTESFSIALSAIAGILLLLLLLFRSKRHSSLKLPPGKLG
PFIGESFIFLRALRSNSLEQFFDERVKKFGLVFKTSLIGHPTVVLCGPAGNRLILSNEE
KLVQMSWPAQFMKLMGENSVATRAGEDHIVMRSALAGFFGPGALQSYIGKMNTEIQSHI
NEKWKGKDEVNVLPLVRELVFNISAILFFNIYDKQEQDRLHKLLETILVGSFALPIDLP
GFGFHRALQGRAKLNKIMLSLIKRRKEDLQSGSATATQDLLSVLLTFRDDKGTPLTND
ILDNFSSLLHASDYDTTSPMALIFKLLSSNPECYQKVVOEQLEILSNKEEGEEITWKDL
KAMKYTWQVAQETLRMFPPVFGTFRKAITDIQYDGYTIPKGWKLWTTYSTHPKDLYFN
EPEKFMPSRFDQEGKHVAPYTFLPFGGGQRCVGEFSKMEILLFVHHFVKTFSSYTPV
DPDEKISGDPLPPLPSKGFSIKLFPETIVN (SEQ ID NO:31)

FIG. 9B

Taxus canadensis taxadiene-5 α -hydroxylase (N° de Acesso ao Banco de Gene AY364469)

MDALYKSTVAKFNEVTQLDCSTESFSIALSSIAGILLLLLLLFRSKRHSSLKLPPGKLG
PFIGESFIFLRALRSNSLEQFFDERVKKFGLVFKTSLIGHPTVVLCGPAGNRLILSNEE
KLVQMSWPAQFMKLMGENSVATRAGEDHIVMRSALAGFFGPGALQSYIGKMNTEIQNHI
NEKWKGKDEVNVLPLVRELVFNISAILFFNIYDKQEQDRLHKLLETILVGSFALPIDLP
GFGFHRALQGRATLNKIMLSLIKRRKEDLQSGSATATQDLLSVLLTFRDDKGTPLTND
ILDNFSSLLHASDYDTTSPMALIFKLLSSNPECYQKVVOEQLEILSNKEEGEEITWKDL
KAMKYTWQVAQETLRMFPPVFGTFRKAITDIQYDGYTIPKGWKLWTTYSTHPKDLYFS
EPEKFMPSRFDQEGKHVAPYTFLPFGGGQRCVGEFSKMEILLFVHHFVKTFSSYTPV
DPDEKISGDPLPPLPSKGFSIKLFPETIVN (SEQ ID NO:32)

FIG. 10

Arabidopsis thaliana ent-kaurene oxidase (N° de Acesso ao Banco de Gene
AAC39505 and AF047719)

MAFFSMISILLGFVISSFIFIFFFKLLSFSRKNMSEVSTLPSV
PVVPGFPVIGNLLQLKEKKPHKTFTRWSEIYGPIYSIKMGSSSLIVLNSTETAKEAMV
TRFSSISTRKLSNALTVLTCDKSMVATSDYDDFHKLVKRCLLNGLLGANAQKRKRHYR
DALIENVSSKLHAHARDHPQEPVNFRAIFEHELFGVALKQAFGKDVESIYVKELGVTL
SKDEIFKVLVHDMMEGAIDVDWRDFFPYLKWIPNKSFEARIQQKHKRR LAVMNALI QD
RLKQNGSESDDDCYLNFLMSEAKTLTKEQIAILVWETIIETADTTLVTTEWAIYELAK
HPSVQDRLCKEIQNVCGGEKFKEEQLSQVPYLNQVFHETLRKYSPAPLVPIRYAHEDT
QIGGYHVPAGSEIAINIYGCNMDKKRWERPEDWWPERFLDDGKYETSDLEKTMAFGAG
KRV CAGALQASLMAGIAIGRLVQEF EWKLRDGE EENVDTYGLTSQKLYPLMAIINPRR
S (SEQ ID NO:33)

FIG. 11A

Gossypium arboreum native (Tipo-silvestre) P450 monooxygenase
(δ -cadinene-8-hydroxylase; CadH); Banco de Gene AF332974

```

1 ccacttcgca gcaatattat tgcagttcct ggttggctac ctctgagttt tcaacttaaa
61 atttcttggg ttctctcaag aaggaaagaag atgttgcaaa tagctttcag ctctgattca
121 tggctgttga ctgctagcaa ccagaaagat ggaatgttgt tcccagtagc tttgtcattt
181 ttggtagcca tattgggaat ttcactgtgg cactgatgga ccataaggaa gccaaagaaa
241 gacatcgccc cattaccgcc ggtccccgtt gggttgcaa tagtgggata tcttccatat
301 cttggaactg ataactctca cttggtgttt acagatttgg ctgcagctta cggccccatc
361 tacaagcttt ggttaggaaa caaattatgc gtagtcatta gctcggcacc actggcgaaa
421 gaagtgggtc gtgacaacga catcacattt tctgaaaggg atcctcccgt ttgtgcaaa
481 attattacct ttggcctcaa tgatattgta tttgattctt acagtagtcc agattggaga
541 atgaagagaa aagtgcctgt acgtgaaatg cttagccata gtagcattaa agcttggtat
601 ggtctaagga gyyaacaagt gcttaaaggc gtacaaaatg ttgctcaaag tgctggcaag
661 ccaattgatt ttggtgaaac ggcattttta acatcaatca atgcgatgat gacgatgctg
721 tgggggtggc aacagggagg agagcggaaa ggggccgacg tttggggcca atttcgagat
781 ctcataaccg aactaatggt gatacttggg aaaccasaag tttctgatat tttcccgtg
841 cttgcaaggt ttgacataca gggattggag aaggaaatga ctaaaatcgt taattctttc
901 gataagcttt tcaactccat gattgaagaa agagagaact ttagcaacaa attgagcaaa
961 gaagatggaa aactgaaac aaaagacttc ttgcagcttc tgttggacct caagcagaag
1021 aacgatagcg gaatatcgat aacaatgaat caagtcaagg ccttgctcat ggacattgtg
1081 gtcggtggaa ctgatacaac atcaaccatg atggaatgga caatggctga actaattgca
1141 aatcctgaag caatgaaaaa ggtgaagcaa gaaatagacg atgttgtcgg ttcggatggc
1201 gccgtcgatg agactcactt gcctaagttg cgctatctag atgctgcagt aaaggagacc
1261 ttccgattgc acccaccgat gccactcctt gtaccccggt gcccgggcga ctcaagcaac
1321 gttggtggct atagcgtacc aaagggcaco aggggtctct taaacatttg gtgtattcag
1381 agggatccac agctttggga aaatccttta gaattcaagc ctgagaggtt cttgactgat
1441 catgagaagc tcgattattt aggaacgat tcccggtaac tgccgtttgg ttctggaagg
1501 agaattgttg cggagatgc tctcggtgaa aagatgttgt attcctcctt ggcagcaatg
1561 atccatgctt atgattggaa cttggccgac ggtgaagaaa atgacttgat tggcttattt
1621 ggaattatta tgaagaaaaa gaagccttta attcttgttc ctacaccaag accatcaaat
1681 ctccagcact atatgaagta actttactat tgtatttctt ttataccact ttattgcttc
1741 tttgtcatgt ttaggcaaca attctaagta ataagtttgg ctatatgttg aacaataatg
1801 tgtttattat acatcataag caatgagctc ttcccgaccc tagggcaata caatgatact
1861 gtgtattaag tgaatcaac aaatctttaa ttctaa (SEQ ID NO:34)

```

FIG. 11B

Sequência de nucleotídeo de cadineno hidroxilase de códon otimizado

atgctgcagattgctttttcttcttattcttggctgctgaccgcttctaaccagaaagacggcatgctgtt
cccggtggcgctgagcttcctgggtggcaatcctgggcattagcctgtggcacgtgtggactatccgtaaac
cgaagaaagatatcgcaccgctgccaccgggtccgcgtggcctgccgatcgttggctacctgccgtatctg
ggcaccgacaacctgcacctgggtgttcaccgacctggcagccgcgtacggtccgatctacaaactgtggct
gggcaataaactgtgcgtagtattctcctctgctcctctggcgaaggaggtggttcgcgacaacgacatca
ccttctccgaacgtgacccaccggtctgtgctaaaaatcatcaccttcggcctgaacgacatcgatttcgac
tcctatagctctcctgactggcgtatgaaacgtaagggttctggtagcgcgagatgctgtcccacagctccat
taaggcatgctacggcctgcgtcgcgaacagggtactgaaaggcgtaaaaaactagcgcagtcgcggggca
aaccgatcgatttcggcgaaaacggccttcctgactagcatcaacgctatgatgtccatgctgtgggtggt
aaacagggcgcgagcgtaaaaggcgccgacgtatggggccagtttcgtgacctgatcccgaaactgatggt
gattctgggcaaacggaacgtcagcgacatcttcccggttctggctcgcttcgacatccagggcctggaaa
aagaaatgaccaagatcgtcaactcttccgacaaaactgtttaactccatgatcgaagaacgcgaaaatttc
tctaacaaactgagcaaagaagatggcaacacggaaactaaagatttcctgcagctgctgctggacctgaa
acaaaagaacgattctggtatctccattaccatgaaccaagtgaagcgctgctgatggacattgttgtgg
gtggtactgacaccacttctaccatgatggaatggacgatggcagaactgattgctaataccggaagcgatg
aagaaagtgaacaagaaattgatgatgtagtgggtctctgatggtgcggtagacgagacgcacctgacctaa
gctgcgttatctggacgcagccgtgaaagaaaccttcgctctgcatccgcctatgccgctgctggttccac
gttgcccaggcgattccagcaacgttgggtggctatagcgtaccgaagggtaccggtgtgttcctgaatatc
tggtgcattcagcgcgacccgcagctgtgggaaaacccgctggagttcaaacctgaacgcttcctgaccga
ccatgaaaagctggactacctgggcaacgattcccggtacatgccgttcggttctggccgctcgatgtgcg
caggcgtctccctgggagagaaaatgctgtactctagcctggctgccatgatccaacgcttacgactggaac
ctggcagatggtgaagagaacgacctgatcggcctgttcggcatcattatgaaaaagaaaaagccgctgat
cctggtgccgactccgcgtccaagcaacctgcagcactacatgaaactggtgccgctggctctaaagaaa
ccgctgctgcaaaattcgaacgtcagcacatggacagctaataa (SEQ ID NO:35)

FIG. 12A

Banco de Gene AY571340

Fonte CPR Taxus

cuspidata

MQANSNTVEGASQGSLLDISRLDHIFALLLNGKGGDLGAMTGS
 ALILTENSQNLMIITLALVLVACVFFVWRRGSDTQKPAVRPTPLVKEEDEEEEDD
 SAKKKVTIFFGTQTGTAEQFAKALAEAKARYEKAVFKVVDLDNYAADDEQYEEKLKK
 EKLAFFMLATYGDGEPTDNaARFYKWFLEKEREPLWSDLTYGVFGLGNRQYEHFNKV
 AKAVDEVLIQGAkRLVPVGLGDDDDQCIEDDFTAWREQVWPELDQLLRDEDEPTSAT
 PYTAAIPEYRVEIYDSVVSVEETHALKQNGQAVYDIHPCRSNVAVRRELHTPLSDR
 SCIHLEFDISDTGLIYETGDHVGVTENSietVEEAAKLLGYQLDTIFSVEGDKEDGT
 PLGGSSSLPPFPGPCTLRtALARYADLLNPPRKAFLALAAHASDPAAERLKLFLSSP
 AGKDEYSQWVTASQSRSLLEIMAEFPSAKPPLGVFFAAIAPRLQPRYSSISSSPRFAPS
 RIHVTCALVYGPSPTGRIHKGVCsNWMKNSLPSEETHDCSWAPVFRQSNFKLPADST
 TPIVMVGPGTGFAPFRGFLQERAKLQEAGEKLGPAVLEFGCRNRQMDYIYEDELKGYV
 ERGILTNLIVAFSREGATKEYVQHKMLEKASDTSLIAQGGYLYVCGDAKGMARDVHR
 TLETIVQEQESVDSSKAFLVKKLQMDGRYLrdIW (SEQ ID NO:36)

FIG. 12B

CPR from *Candida tropicalis* N° de Acesso ao Banco de Gene M35199)

MALDKLDLYVIIITLVVAIAAYFAKNQFLDQQQDTGFLNTDSGDG
 NSRDILQALKKNNKNTLLLFSGSQTGTAEYANKLSRELHSRFGKTMVADFADYDFEN
 FGDITEDILVFFIVATYGEGEPTDNaDEFHTWLTTEEADTLSTLKYTVFGLGNSTYEFF
 NAIGRKFDRLIGEGGDRFAEYGEgDDGTGLDEDFLAWKDNVFDSLKNDLNFEEKEL
 KYEPNVKLTERDDLsgNDPVDVSLGEPNVKYIKSEGVDLTGKPFDETHPFLARIVKtKE
 LFTSEDRHCvHVEFDISESNLKYTTGDHLAIWPSNSDENIKQFAKCFGLEDKLDTVIE
 LKALDSTYSIPFPNPITYGAVIRHHLAISGPVSRQFFLSIAGFAPDEETKKSFTTRIGG
 DKQEFASKVTRRKFNiADALLFASNRPWSDVPFEFLIENVQHLTPRYSSISSSSLSSE
 KQTINVTAVVEAEEDGRPVTGVVTNLLKNIEIEQNKtGETPMVHYDLNGPRGKFSK
 FRLPVHVRNSNFKLPKNSTTPVILIGPGTGVAPLRGfVRERVQQVKNGVNVGKTVLFY
 GCRNSEQDFLYKQEWSEYASVLGENFEMFNAFSRQDPTKKVYVQDKILENSALVDELL
 SSGAIIYVCGDASRMARDVQAAIAKIVAKSRDIHEDKAAELVKSWKVQNRyQEDVW (SEQ ID NO:37)

FIG. 12C

ATR1

>gi|6088150|emb|CAB58575.1| NADPH CYTOCHROME P450 REDUCTASE [*Arabidopsis thaliana*] (N° de Acesso ao Banco de Gene: X66016)

MTSALYASDLFKQLKSIMGTDsLSDDVVLVIATTSALVAGFVVLWKKTTADRSGELKPLMIPKSLMAK
 DEDDDLDLGSGKTRVSIFFGTQTGTAEQFAKALSEEIKARYEKAAVKVIDLDDYAADDQYEEKLKKETL
 AFFCVATYGDGEPTDNaARFSKWfTEENERDIKLLQQLAYGVFALGNRQYEHFNKIGIVLDEELCKKGAKR
 LIEVGLGDDDBQSIEDDFNAWKESLWSELdkLLKDEDDKSVATPYTAVIPEYRVVTHDPRFTTQKSMESNV
 ANGNTTIDIHPCRVdVAVQKELHThESDRSCIHLEFDISRTGITYETGDHVGvYAENHVEIVEEAGKLL
 GHSLDLVFSIHADKEDGSPLESaVPPFPGPCTLTGLARYADLLNPPRKSALVALAAAYATEPSEAEKLLK
 ELTSPDGKDEYSQWIVASQSRSLLEVMAAFPSAKPPLGVFFAAIAPRLQPRYSSISSCQDWAPSRVHVtSA
 LVYGPtPTGRIHKGVCSTWMKNAVPAEKSHecSGAPfIRASNFKLPSNPSTPIVMVGPGTGLAPFRGFL
 QERMALKEDGEELGSSLLFFGCRNRQMDfIYBDELNNFVDQGVISELIMAFSREGAQKEYVQHKMMEKAA
 QVWDLIkeEGYLYVCGDAKGMARDVHTLHTIVQEQEGVSSSRAEAIvKKLQTEGRYLrdVW (SEQ ID
 NO:38)

FIG. 12D

Arabidopsis thaliana cytochrome P450 reductase ATR2 (Nº de Acesso ao Banco de Gene No. X66017)

MSSSSSSSTSMIDLMAAIKGEPIVSDPANASAYESVAAELSS
MLIENRQFAMIVTTSIAVLIGCIVMLVWRRSGSGNSKRVEPLKPLVIKPREEEIDDGR
KKVTIFFGTQTGTAEAGFAKALGEEAKARYEKTRFKIVDLDDYAADDDEYEKLLKEDV
AFFFLATYGDGEPTDNAARFYKWFTEGNDRGEWLKLNLYGVFGLGNRQYEHFNKVAKV
VDDILVEQGAQRLVQVGLGDDDDQCIEDDFTAWREALWPELDTILREEGDTAVATPYTA
AVLEYRVSIIHSEDAKFNDITLANGNGYTVFDAQHPYKANVAVKRELHTPESDRSCIH
LEFDIAGSGLTMKLGDBHVGVLCDNLSETVDEALRLLDMSPDTYFSLHAEKEDGTPISS
SLPPPPPPPCNLRTALTRYACLLSSPKKSALVALAAHASDPTAEERLKHLSAPAGKDEY
SKWVVESQRSLLLEVMAEFPSAKPPLGVFFAGVAPRLQPRFYSISSSPKIAETRIHVTC
ALVYEKMPGTGRIHKGVCSTWMKNAVPEYKSEKFLGRPIFVRQSNFKLPSDSKVPIIM
IGPGTGLAPFRGFLQERLALVESGVELGPSVLFFGCRNRRMDFIYEEELQRFVESGAL
AELSVAFSREGPTKEYVQHKMMDKASDIWNMISQAYLYVCGDAKGMARDVHRSLSHTI
AQEQGSMDSTKAEGFVKNLQTSGRYL RDVW (SEQ ID NO:39)

FIG. 12E

ATR2mod

>gi|13272461|gb|AAK17169.1|AF325101_1 NADPH-ferrihemoproteína redutase (ATR2) [*Arabidopsis thaliana*]

MLIENRQFAMIVTTSIAVLIGCIVMLVWRRSGSGNSKRVEPLKPLVIKPREEEIDDGRKKVTIFFGTQTGT
AEGFAKALGEEAKARYEKTRFKIVDLDDYAADDDEYEKLLKEDVAFFFLATYGDGEPTDNAARFYKWFTE
GNDRGEWLKLNLYGVFGLGNRQYEHFNKVAKVDDILVEQGAQRLVQVGLGDDDDQCIEDDFTAWREALWPE
LDTILREEGDTAVATPYTA AVLEYRVSIIHSEDAKFNDINMANGNGYTVFDAQHPYKANVAVKRELHTPES
DRSCIHLEFDIAGSGLTYETGDHVGVLCDNLSETVDEALRLLDMSPDTYFSLHAEKEDGTPISSSLPPPPFP
PCNLRTALTRYACLLSSPKKSALVALAAHASDPTAEERLKHLSAPAGKDEYSKWVVESQRSLLLEVMAEFPS
AKPPLGVFFAGVAPRLQPRFYSISSSPKIAETRIHVTCALVYEKMPGTGRIHKGVCSTWMKNAVPEYKSENC
SSAPIFVRQSNFKLPSDSKVPIIMIGPGTGLAPFRGFLQERLALVESGVELGPSVLFFGCRNRRMDFIYEE
ELQRFVESGALAELSVAFSREGPTKEYVQHKMMDKASDIWNMISQAYLYVCGDAKGMARDVHRSLSHTIAQ
EQGSMDSTKAEGFVKNLQTSGRYL RDVW (SEQ ID NO:40)

FIG. 13

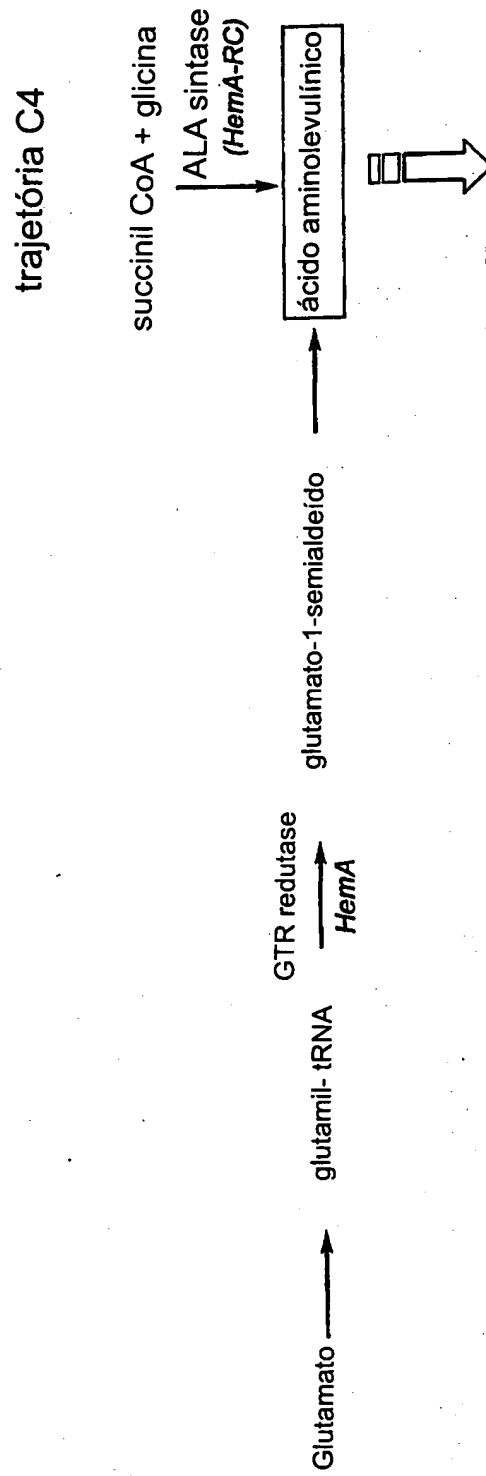


FIG. 14

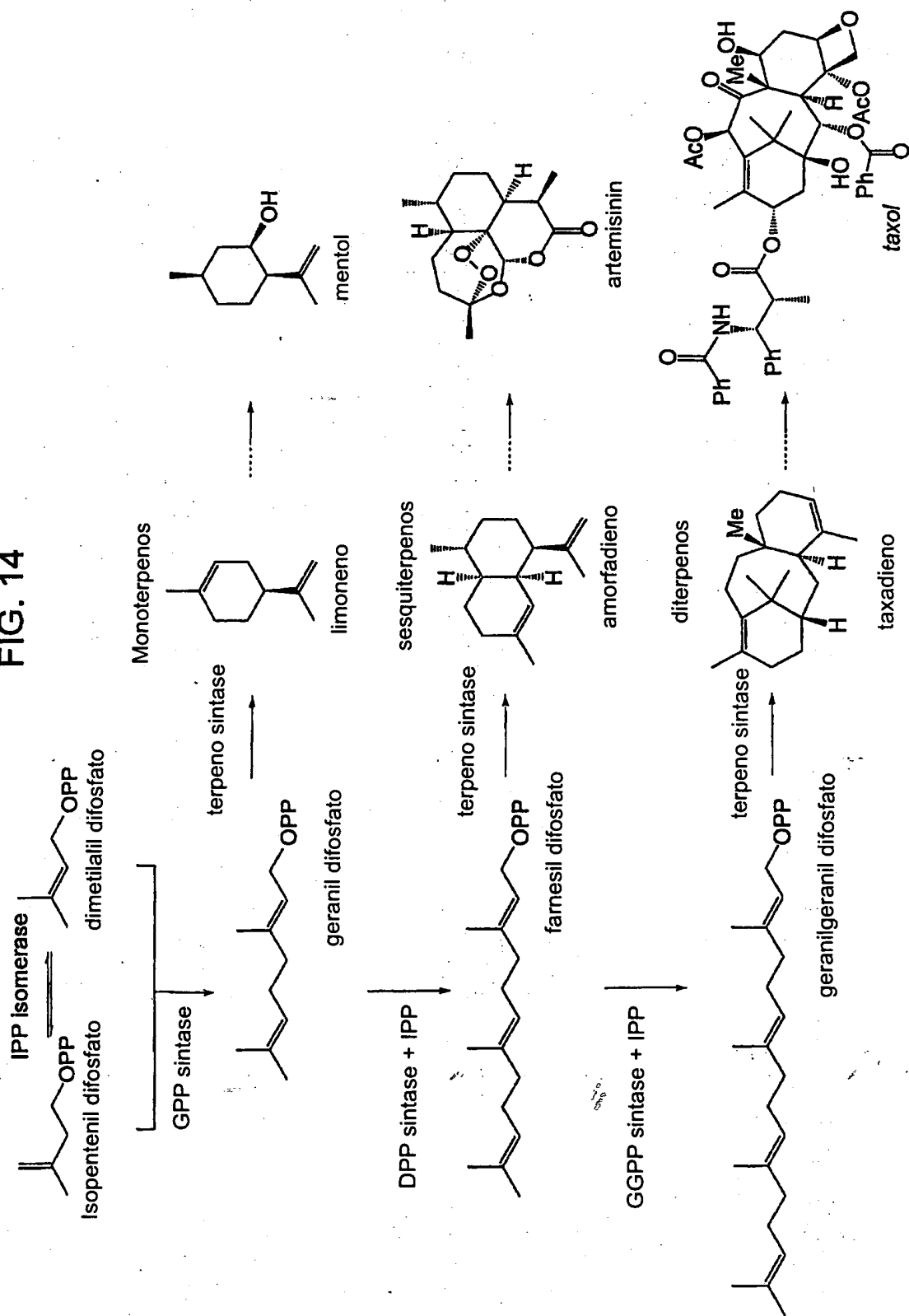


FIG. 15

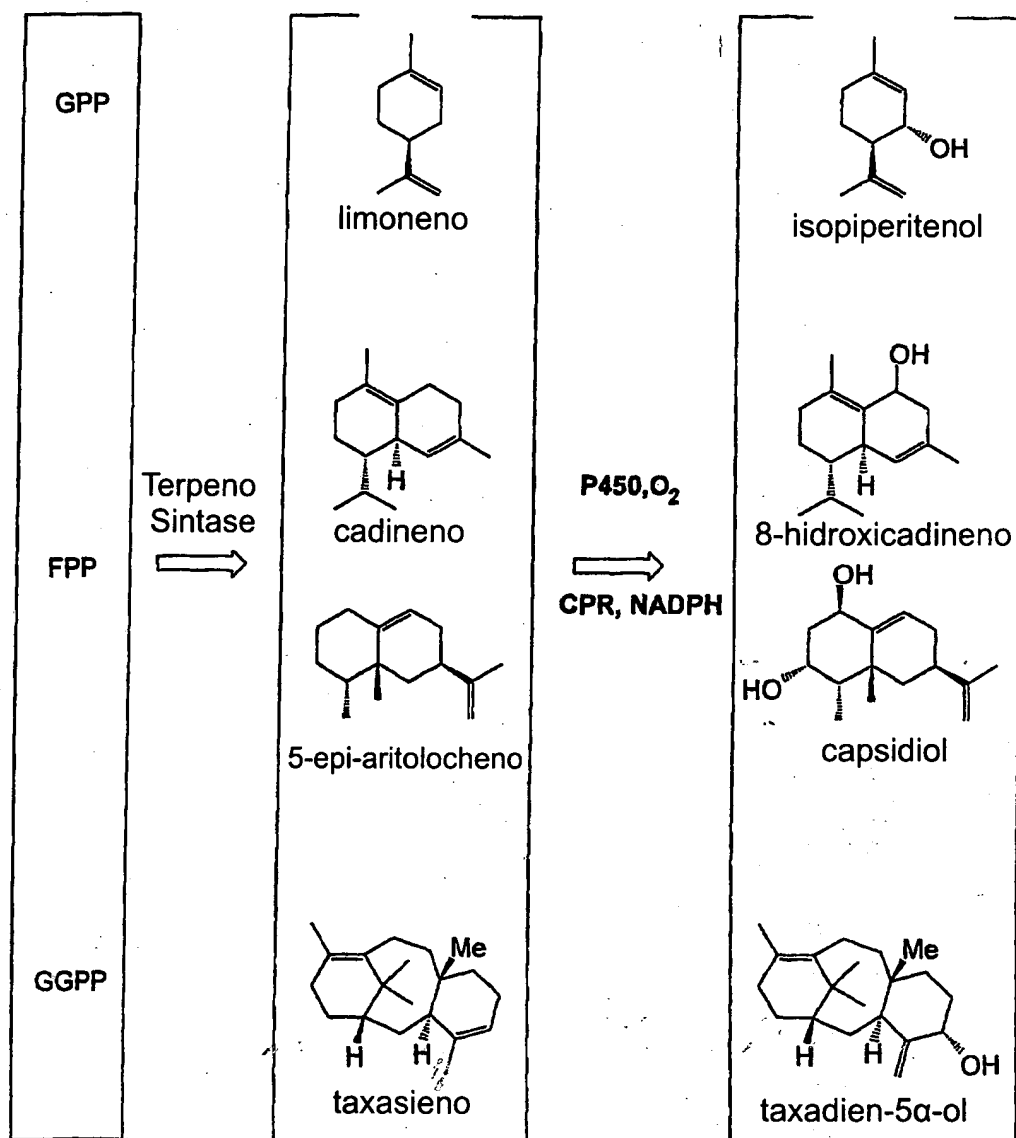


FIG. 16

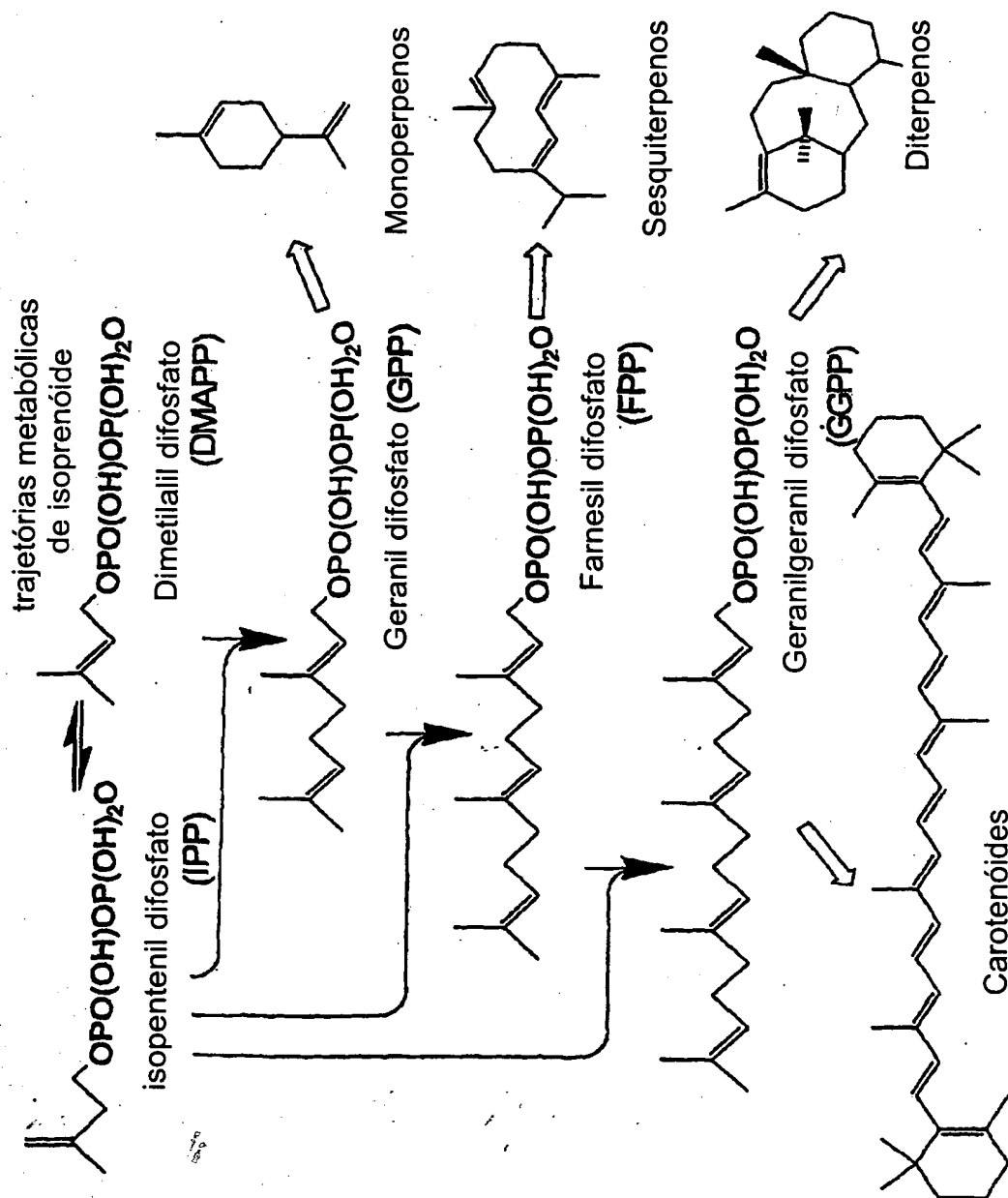
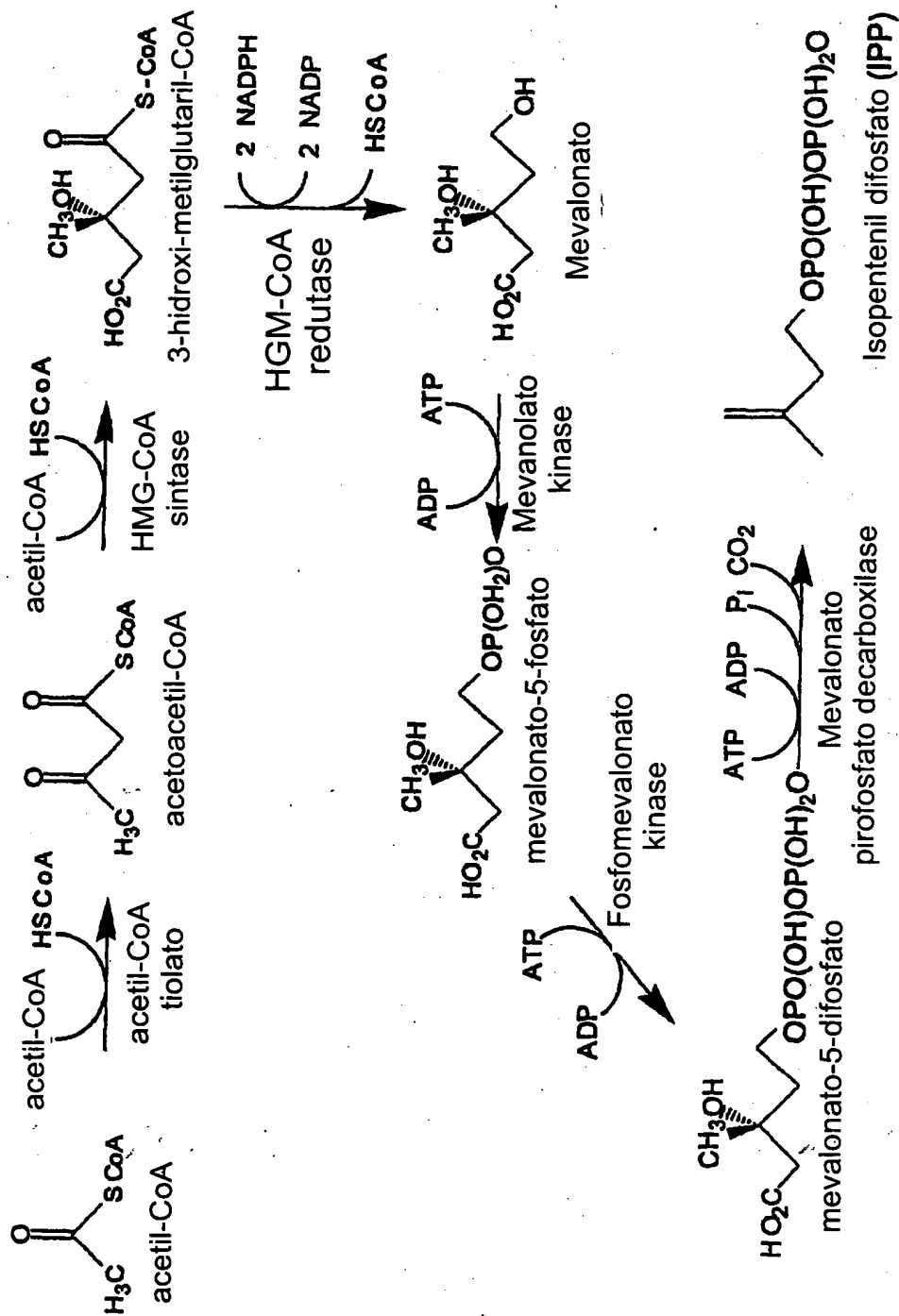


FIG. 17

Tragetória de Mevalonato



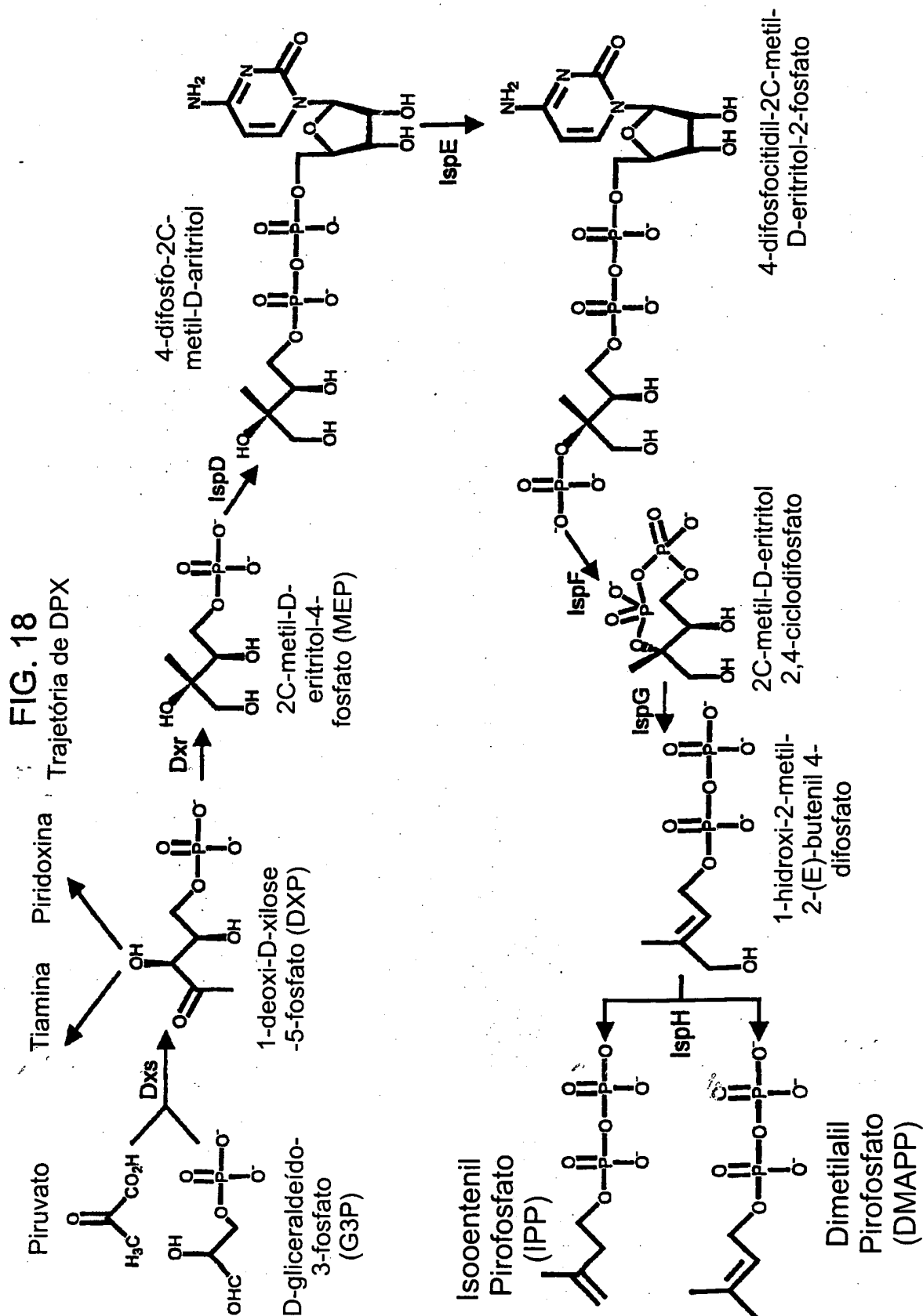


FIG. 19A

Eschscholzia stolonifera Metilcoclaurina 3'-hidroxilase (Acesso: AF014802)
 MEKPILLQLQAGILGLLALICFLYYVIKVSLSSTRNCNQLVKHPPEAAGSWPIVGHLPQL
 VGSGKPLFRVLGDMADKFGPIFMVRFVGYPTLVVSTWEMAKECFTSNDKFLASRPPSAA
 SSYMITYDHAMFGFSFYGPYWREIRKISTLHLLSHRRLELLKHVPHTTEIHNFIKGLFGIW
 KDHQKQQQPTGREDRDSVMLEMSQLFGYLTNLVVLSTLVGKRVCNYHADGHLDDGEEAG
 QGQKLEQTITDFFKLSGVSVASDALPLLGLFDLGGKKESMKRVAKEMDFFAERWLQDKK
 LSLSLSSETNNKQNDAGEGDGDDFMDVLSILPDDDDSLFTKYSRDTVIKATSLSMVVA
 ASDTTSVSLTWALSLLLNNIQVLRKAQDELDTKVGRDRHVEEKDIDNLVYLQAIVKETL
 RMPAGPLSVPHEAIEDCNVGGYHIKTGTRLLVNIWKLQRPVWSNPSEFRPERFLDN
 QSNGTLLDFRGQHFYIPFGSGRRMCPGVNFATLILHMTLARLLQAFDLSTPSSSPVDM
 TEGSGLTMPKVTPKVLTPRLPLPLYDY (SEQ ID NO:41)

FIG. 19B

Catharanthus roseus Geraniol-10-hidroxilase (Acesso: AJ251269)
 MDYLTIILTLLFALTLYEAFSYLSRRTKNLPPGPSPLPFFIGSLHLLGDQPHKSLAKLSK
 KHGPIMSLKLGQITTIVISSSTMAKEVLQKQDLAFSSRSVPNALHAHQFKFSVVWLPLV
 ASRWRSLRKVLNSNIFSGNRDANQHLRTRKVQELIAYCRKNSQSGEAVDVGRAAFRTS
 LNLISNLIFSKDLTDPYSDSAKEFKDLVWNIMVEAGKPNLVDFPFLLEKVDPOGIRHRM
 TIHFGEVLKLFGGVLNERLEQRRSKGEKNDVLDVLLTTSQESPEEIDRTHIERMCLDLF
 VAGTDTTSSTLEWAMSEMLKNPDKMKKTQDELAQVIGRGKTIEESDINRLPYLRVCMKE
 TLRHPPVPFLIPRKVEQSVEVCGYNVPGKSQVLNAWAIGRDETVDWDDALAFKPERFM
 ESELDIRGRDFELIPFGAGRRICPGLPLALRTVPLMLGSLNLSFNWKLEGGMAPKDLDL
 EEKFGITLQKAHPLRAVPSTL (SEQ ID NO:42)

FIG. 19C

Catharanthus roseus Tabersonina-16-hidroxilase (Acesso: AJ238612)
 MLLFCFILSKTTKFGQNSQYSNHDELPPGPPQIPILGNAHQLSGGHTHILRDLAKKY
 GPLMHLKIGEVSTIVASSPQIAEEIFRTHDILFADRPNSNLESFKIVSYDFSMDVVSYPG
 NYWRQLRKISMMELLSQKSVQSFRSIREEEVLNFIKSIKSKEGTRINLSKEISLLIYGI
 TTRAAFGEKNKNTTEEFIRLLDQLTKAVAEPNIADMFPPLKFLQLISTSKYKIEKIHQOF
 DVIVETILKGHKEKINKPLSQENGEKKEDLVDVLLNIQRRNDFEAPLGDKNIKAIIFNI
 FSAGTETSSTTVDWAMCEMIKNPTVMKKAQEEVRKVFNEEGNVDETKLHQLKYLQAVIK
 ETTLRLHPPVPLLLPRECREQCKIKGYTIPSKSRVIVNAWAIGRDPNYWIEPEKFNPDRE
 LESKVDFKGSFEYLPFGGRRICPGITFALANIELELAQLLFHFDWQSNTEKLNMKES
 RGVTVRREDDLYLTPVNFSSSSPA (SEQ ID NO:43)

FIG. 20A

Arabidopsis thaliana cinnamate-4-hydroxylase (Acesso: NM 128601)

MDLLLLLEKSLIAVFVAVILATVISKLRGKKLKLPPGPIPIPIFGNWLQVGDDLNRNL
VDYAKKFGDLFLLRMGQRNLVVVSSPDLTKEVLLTQGVFEFGSRTNRNVFDIFTGKGQDM
VFTVYGEHWRKMRRIMTVPFFTNKVVQQNREGWEFEAASVVEDVKKNPDSATKGIVLRK
RLQLMMYNNMFRIMFDRRFESEDDPLFLRLKALNGERSRLAQSFYNYGDFIPILRPFL
RGYLKICQDVKDRRIALFKKYFVDERKQIASSKPTGSEGLKCAIDHILEAEQKEINED
NVLYIVENINVAAIETTLWSIEWGIAELVNHPEIQSKLRNELDTVLGPGVQVTEPDLHK
LPYLQAVVKETLRLMAIPLLPHMNLHDAKLAGYDIPAESKILVNAWWLANNPNSWKK
PEEFRPERFFEEESHVEANGNDFRYVPPFGVGRRSCPGIILALPILGITIGRMVQNFELL
PPPGQSKVDTSKGGQFSHLNHSIIVMKPRNC (SEQ ID NO:44)

FIG. 20B

Ocimum basilicum p-coumaroil shikimate-3'-hidroxilase (Acesso: AY082611)

MAALLLLLLLPLLLPAIFLLHLYRLRFLPPGPRPLPIVGNLYDVKPVRFRCFADWA
QSYGPIISVWFGSTLNVIVSNTELAKEVLKEKDQQLADRHRSSAAKFSRDGQDLIWD
YGPYVVKRVCTLELFSPKRLEALRPIREDEV TAMVESIYHCTAPDNAGKSLLVKKY
LGAVAFNNITRLAFGKRNVNSEGIIDKQGLEFKAIVSNGKLKGLASLAMAELHPSLRWMF
PLDEDAFAKHGARRDQLTREIMEEHTRAREESGGAKQHFFDALLTLKDKYDLSEDTIIG
LLWDMITAGMDTTAISVEWAMAEIKNPRVQQAQEELDRVIGYERVMTELD FSNLPYL
QCVAKEALRLHPPTPLMLPHRSNSNVKIGGYDIPKGSNVHVNVWAVARDPAVWKNPCEF
RPERFLEEDVDMKGHDFRLLPFGAGRRVCPGAQLGINLVTSMIGHLLHFNWAPPSGVS
SDELDMGENPGLVTYMRTPLEAVPTPLRPSDLYKRIAVDL (SEQ ID NO:45)

FIG. 20C

Glycine max Isoflavona Sintase (Acesso: AF195798)

MLELALGLFVLALFLHLRPTPSAKSKALRHLNPPSPKPRLPFIHGLHLLKDKLLHYA
LIDLKSKKHGPLFSLSFGSMPTVVASTPELFLFLQTHEATSFNTRFQTSAIRRLTYDNS
VAMVPFGPYWKFRKLIMNDLLNATTVNKLRLRTOQIRKFLRVMAQSAEAQKPLDVTE
ELLKWTNSTISMMMLGEAAEIRDIAREVLKIFGEYSLTDFIWPLKYLKVGKYEKRIDDI
LNKFDPVVERVIKKRREIVRRRNKEVVEGEASGVFLDTLLEFAEDETMEIKITKEQIK
GLVVDFFSAGTDSTAVATEWALAEILNNPRVLQKAREEVYSVVGKDRLVDEVDTQNL PY
IRAIVKETFRMHPPPLPVVKRKCTEECEINGYVIPEGALVLFNVWQVGRDPKYWDRPSEF
RPERFLETGAEGEAGPLDLRGQHFLLPFGSGRRMCPGVNLATSGMATLLASLIQCFDL
QVLGPQGQILKGDDAKVSMEERAGLTVPRAHSLVCVPLARIGVASKLLS (SEQ ID NO:46)

FIG. 21A

Streptomyces avermitilis aveE (Acesso: AB032367)

MMSQSTSSIEAPGAWPVVGHVPPLMRQPLEFLRSAADHGDLKLRLGPKTAYLATHPD
LVRTMLVSSGSGDFTRSKGAQGASRFIGPILVAVSGETHRRQRRRMQPGFHRQRLSYV
ATMAAAAQETADSWAGQVVDVEQAACDLSLAMITKTLFFSDLGAKAEALRKTGHDIL
KVARLSALAPTLYEVLPTAGKRSVGRSATSITREAITAYRADGRDHGDLSTMLRATDAE
GASMTDQEVHDEVMGIAVAGIGGPAAITAWIFHELQONAEIESRLHAELDTVLGGRLPT
HEDLPRLPYTQNLVKEALRKYPGWGSRRTVRPVRLGGHDLPADVEVMYSAYAIQDPR
WYPEPERLDPGRWETKGSRRGVPKGAWVPFALGTYKCIGDNFALLETAVTVAVVASHWR
LHALPGDEVPRPKTKATHVFPNRLRMIAEPRSVVRLEEPAAMGA (SEQ ID NO:47)

FIG. 21B

MGLPLTSTKTAPVSYPFGRPEGLDLDEAYEQARKSEGLLWVHMPYGEPEGWLVSRYDDAR
FVLGDRRFSHAEEAENDAPRMRELRTPNGIIGMDAPDHTLRGLVTKAFTPRRVEAMRP
HVRMTASLLRDMTALGSPVDLVDEYAVPLPVAVICGLLGVPEDRDLFRGWCEIAMST
SSLTAEDHVRLAGELTGYLADLITARRAAPRDDLVLSALVEARDAQGRLSQEELVDLIVE
LLFAGHETTASQISNFVLVLEQPDQLALLRDRPDLLDNAMEELTRFVPLGSQAGFPY
ATEDVEVGCTLVRAGDPVLVQMNANRDALFRSPGVLDITRDDAGRHLGYGHGPHHCL
GASLARLELQEALRTLLELPGHLAQPVWKTENVVRGPRTMLVGW (SEQ ID NO:48)

FIG. 22

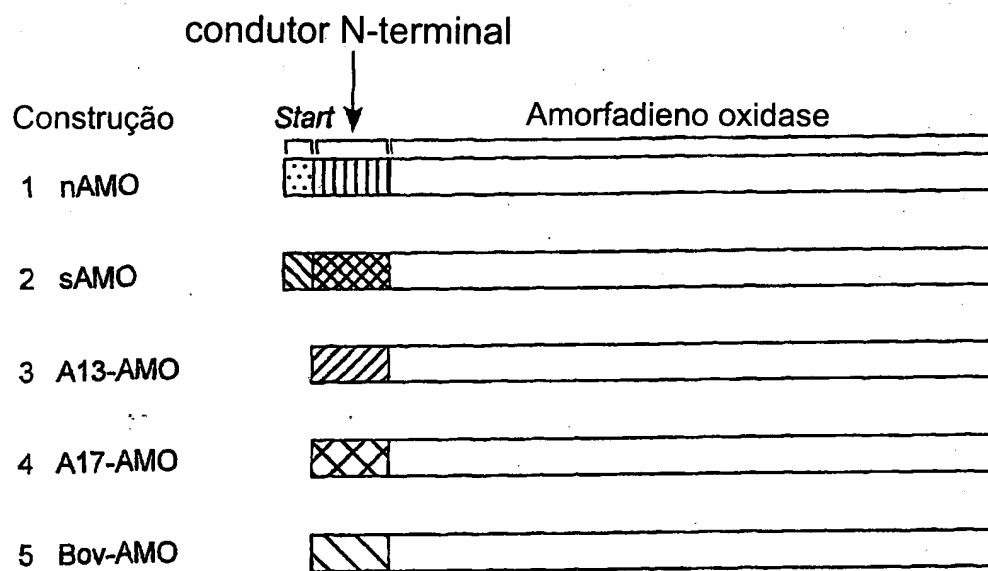


FIG. 23A

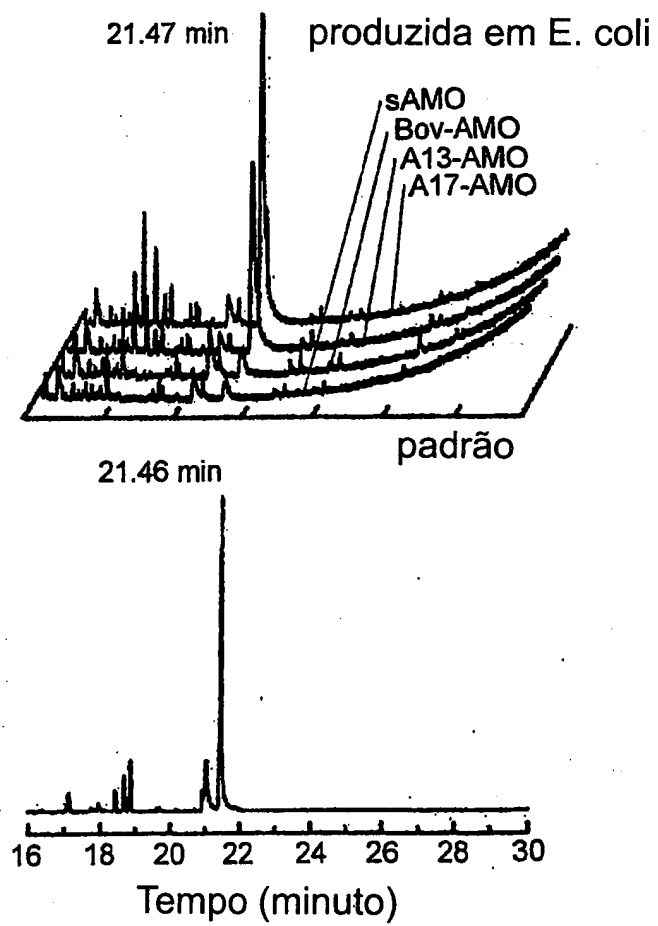


FIG. 23B

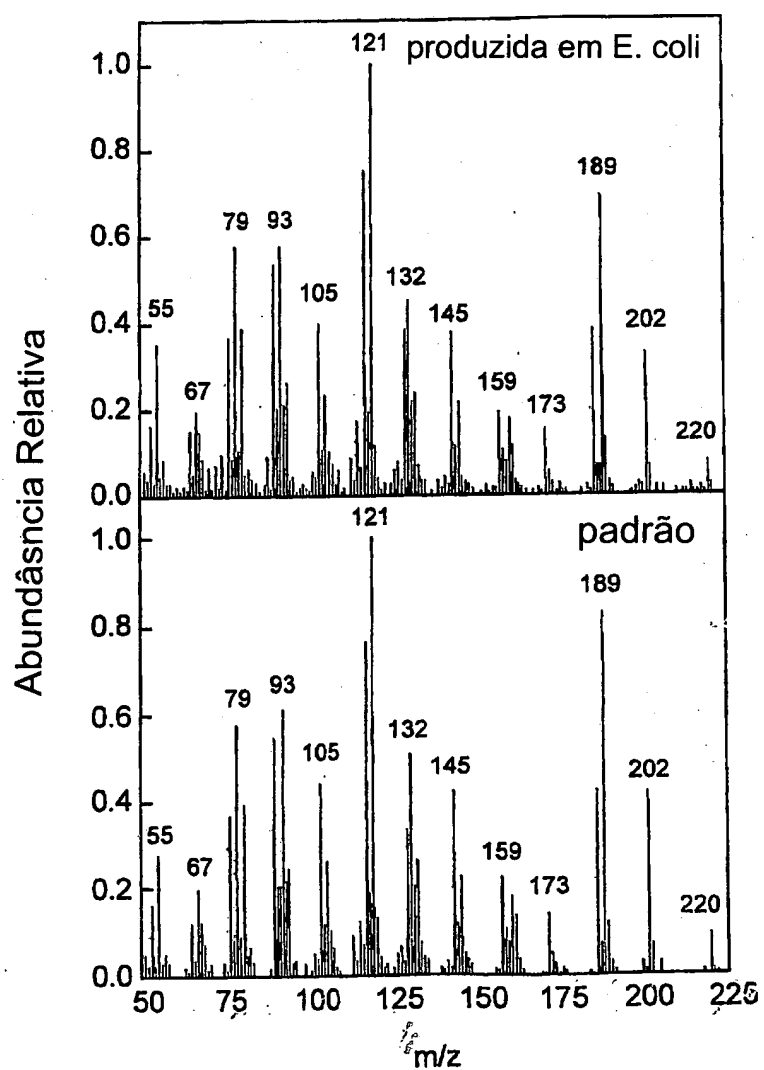


FIG. 24A

Amorfadieno oxidase tipo selvagem

CATATGAAGTCTATTCTGAAAGCAATGGCTCTGTCTCTGACCACTAGCATCGCCCTGGCGACTATCCTGCTGTTTGTGT
 ACAAATTCGCGACCCCGTTCTAAAAGCACTAAGAAATCTCTGCCGGAACCGTGGCGTCTGCCAATCATCGGTACATGCA
 CCACCTGATCGGCACCACCCCGCACCGTGGCGTACGCGACCTGGCGCGTAAGTACGGCTCTCTGATGCATCTGCAGCTG
 GCGGAGGTACCTACTATCGTCTGTTTCTCCCGAAGTGGGCGCAAAGAAATCTGACTACCTATGACATCATTTCGCCA
 ACCGCCCCGAAACGCTGACCGGCGAAATTTGCTCTGTACCATAACACGGATGTGGTTCTGGCCCCGTACGGTGAGTACTG
 GCGCCAGCTGCGCAAAATTTGTACTCTGGAACCTGTGAGCGTTAAAAGGTTAAATCCTTCAGAGCCGTGCGTGAAGAG
 GAATGCTGGAACCTGGTGCAGGAGATTAAAGCGTCTGGCAGCGGTCTCCAGTTAACCTCTCTGAGAATGTTTTAAAC
 TGATCGCTACTATCCTGTCTCGCGCGGCATTCTGGTAAAGGTATCAAAGATCAGAAAGAACTGACCGAAATCGTTAAGGA
 AATCCTGCGCCAGACTGGTGGCTTCGACGTTGCGGACATCTTCCCGTCCAAAAGTTCTCTGCACCATCTGTCTGGCAAA
 CGCGCTCGTCTGACCTCCCTGCGTAAGAAAATTGATAACCTGATTGACAACCTGGTCTGAGCACACTGTGAACACCT
 CTTCTAAAACCAACGAAACCTGCTGGACGTACTGCTGCGCTGAAGGACTCTGCCGAATTTCCACTGACTAGCGACAA
 TATCAAAGCAATCATCTCTGGACATGTTCCGCGCCGGTACCGATACGTCCTCTTCCACGATTGAGTGGGCTATTTCCGAA
 CTGATCAAATGCCGAAGGCGATGGAAAAAGTGCAGGCGGAACTCGCTAAAGCGCTGAACGGTAAAGAGAAAATTCATG
 AAGAGGACATCCAGGAACCTGTCTTACCTGAATATGGTAATCAAAGAACTCTGCGTCTGCATCCGCGCGTGCCTGCT
 TCTGCGCGTGAATGCCGTGAGCGGTTAACCTGGCGCGGTACAAACATTCGGAACAAAACGAAGCTGATCGTCAACGTT
 TTCGCGATCAACCGCGATCTGAATACGTAAGACGCGGAAGCGTTCATTCCGGAACGCTTTGAGAACTCCTCTGCCA
 CGGTTATGGGCGCTGAATACGAGTACCTGCCGTTCCGTTGCGGGTCCGCGTATGTGCCCGGGTGTGCACTGGGCGCTGGC
 GAACGTTCAACTGCCACTGGCGAACATCCTGTACCACTTCAACTGGAACCTGCCTAACGGCGTATCTTATGATCAAATC
 GACATGACCGAAAGCTCCGCGCGACCATGCAGCGTAAACCGAACTGCTGCTGGTTCCGTCCTTTTAAACCTAGG
 (SEQ ID NO:49)

FIG. 24B

Nde ← Condutor → Transmembrana
 CATATGAAGTCTATTCTGAAAGCAATGGCTCTGTCTCTGACCACTAGCATCGCCCTGGCGACTATCCTGCTGTTTGTGT

→ >Amorphadiene Oxidase
 ACAAATTCGCGACCCCGTTCTAAAAGCACTAAGAAATCTCTGCCGGAACCGTGGCGTCTGCCAATCATCGGTACATGCA
 CCACCTGATCGGCACCACCCCGCACCGTGGCGTACGCGACCTGGGCGTAAGTACGGCTCTCTGATGCATCTGCAGCTG
 GCGGAGGTACCTACTATCGTCTGTTTCTCCCGAAGTGGGCGCAAAGAAATCTGACTACCTATGACATCATTTCGCCA
 ACCGCCCCGAAACGCTGACCGGCGAAATTTGCTCTGTACCATAACACGGATGTGGTTCTGGCCCCGTACGGTGAGTACTG
 GCGCCAGCTGCGCAAAATTTGTACTCTGGAACCTGTGAGCGTTAAAAGGTTAAATCCTTCAGAGCCGTGCGTGAAGAG
 GAATGCTGGAACCTGGTGCAGGAGATTAAAGCGTCTGGCAGCGGTCTCCAGTTAACCTGTCTGAGAATGTTTTAAAC
 TGATCGCTACTATCCTGTCTCGCGCGGCATTCTGGTAAAGGTATCAAAGATCAGAAAGAACTGACCGAAATCGTTAAGGA
 AATCCTGCGCCAGACTGTGGCTTCGACGTTGCGGACATCTTCCCGTCCAAAAGTTCTGACCATCTGTCTGGCAAA
 CGCGCTCGTCTGACCTCCCTGCGTAAGAAAATTGATAACCTGATTGACAACCTGGTCTGAGCACACTGTGAACACCT
 CTTCTAAAACCAACGAAACCTGCTGGACGTACTGCTGCGCGCTGAAGGACTCTGCCGAATTTCCACTGACTAGCGACAA
 TATCAAAGCAATCATCTCTGGACATGTTCCGCGCCGGTACCGATACGTCCTCTTCCACGATTGAGTGGGCTATTTCCGAA
 CTGATCAAATGCCCGAAGGCGATGGAAAAAGTGCAGGCGGAACTGCGTAAAGCGCTGAACGGTAAAGAGAAAATTCATG
 AAGAGGACATCCAGGAACCTGTCTTACCTGAATATGGTAATCAAAGAACTCTGCGTCTGCATCCGCGCGTGCCTGCTGGT
 TCTGCGCGTGAATGCCGTGAGCGGTTAACCTGGCGCGGTACAAACATTCGGAACAAAACGAAGCTGATCGTCAACGTT
 TTCGCGATCAACCGCGATCTGAATACGTAAGACGCGGAAGCGTTCATTCCGGAACGCTTTGAGAACTCCTCTGCCA
 CGGTTATGGGCGCTGAATACGAGTACCTGCCGTTCCGTTGCGGGTCCGCGTATGTGCCCGGGTGTGCACTGGGCGCTGGC
 GAACGTTCAACTGCCACTGGCGAACATCCTGTACCACTTCAACTGGAACCTGCCTAACGGCGTATCTTATGATCAAATC

GACATGACCGAAAGCTCCGCGCGACCATGCAGCGTAAACCGAACTGCTGCTGGTTCCGTCCTTTTAACTTAGG
 Stop
 >AvrII
 (SEQ ID NO:50)

FIG. 25

Mapa de Translação
Condutor

1 ATGAAGTCTATTCTG
1 M K S I L

Transmembrana

1 AAAGCAATGGCTCTGTCTCTGACCACTAGCATCGCCCTGGCGACTATCCTGCTGTTTGTG
1 K A M A L S L T T S I A L A T I L L F V
61 TACAAATTC
21 Y K F

Amorphadiene Oxidase

1 GCGACCCGTTCTAAAAGCACTAAGAAATCTCTGCCGGAACCGTGGCGTCTGCCAATCATC
1 A T R S K S T K K S L P E P W R L P I I
61 GGTACATGCACCACCTGATCGGCACCACCCCGCACCGTGGCGTACGCGACCTGGCGCGT
21 G H M H H L I G T T P H R G V R D L A R
121 AAGTACGGCTCTCTGATGCATCTGCAGCTGGGCGAGGTACCTACTATCGTCGTTTCCTCC
41 K Y G S L M H L Q L G E V P T I V V S S
181 CCGAAGTGGGCCAAAGAAATCCTGACTACCTATGACATCACTTTCGCCAACCGCCCGGAA
61 P K W A K E I L T T Y D I T F A N R P E
241 ACGCTGACCGGCGAAATTGTCCTGTACCATAACACGGATGTGGTTCTGGCCCCGTACGGT
81 T L T G E I V L Y H N T D V V L A P Y G
301 GACTACTGGCGCCAGCTGCGCAAAATTTGTACTCTGGAACCTGCTGAGCGTTAAAAAGGTT
101 E Y W R Q L R K I C T L E L L S V K K V
361 AAATCCTTCCAGAGCCTGCGTGAAGAGGAATGCTGGAACCTGGTGCAGGAGATTAAAGCG
121 K S F Q S L R E E E C W N L V Q E I K A
421 TCTGGCAGCGGTCTGTCAGTTAACCTGTCTGAGAATGTTTTTAAACTGATCGCTACTATC
141 S G S G R P V N L S E N V F K L I A T I
481 CTGTCTCGCGCGGCATTCGGTAAAGGTATCAAAGATCAGAAAGAACTGACCGAAATCGTT
161 L S R A A F G K G I K D Q K E L T E I V
541 AAGGAAATCCTGCGCCAGACTGGTGGCTTCGACGTTGCGGACATCTTCCCGTCCAAAAAG
181 K E I L R Q T G G F D V A D I F P S K K
601 TTCCTGCACCATCTGTCTGGCAAACGCGCTCGTCTGACCTCCCTGCGTAAGAAAATTGAT
201 F L H H L S G K R A R L T S L R K K I D
661 AACCTGATTGACAACCTGGTCTGAGCACACTGTGAACACCTCTTCTAAAACCAACGAA
221 N L I D N L V A E H T V N T S S K T N E
721 ACCCTGCTGGACGTACTGCTGCGCCTGAAGGACTCTGCCGAATTTCCACTGACTAGCGAC
241 T L L D V L L R L K D S A E F P L T S D
781 AATATCAAAGCAATCATCTGGACATGTTCCGCGCGGTACCGATACGTCCTCTTCCACG
261 N I K A I I L D M F G A G T D T S S S T
841 ATTGAGTGGGCTATTTCCGAACCTGATCAAATGCCCGAAGGCGATGGAAAAAGTGCAGGCG
281 I E W A I S E L I K C P K A M E K V Q A
901 GAACTGCGTAAAGCGCTGAACGGTAAAGAGAAAATTCATGAAGAGGACATCCAGGAACCTG
301 E L R K A L N G K E K I H E E D I Q E L
961 TCCTACCTGAATATGGTAATCAAAGAACTCTGCGTCTGCATCCGCCGCTGCCACTGGTT
321 S Y L N M V I K E T L R L H P P L P L V
1021 CTGCCGCGTGAATGCCGTCAGCCGTTAACCTGGCCGCTACAACATTCGGAACAAAACG
341 L P R E C R Q P V N L A G Y N I P N K T
1081 AAGCTGATCGTCAACGTTTTTCGCGATCAACCGCGATCCTGAATACTGGAAAGACGCGGAA
361 K L I V N V F A I N R D P E Y W K D A E
1141 GCGTTCATTCCGGAACGCTTTGAGAACTCCTCTGCCACCGTTATGGGCGCTGAATACGAG
381 A F I P E R F E N S S A T V M G A E Y E
1201 TACCTGCCGTTCCGGTGGCGGTGCGCGTATGTGCCGGGTGCTGCACTGGGCGTGGCGAAC
401 Y L P F G A G R R M C P G A A L G L A N
1261 GTTCAACTGCCACTGGCGAACATCCTGTACCCTCAACTGGAAACTGCCTAACCGCGTA
421 V Q L P L A N I L Y H F N W K L P N G V
1321 TCTTATGATCAAATCGACATGACCGAAAGCTCCGGCGCGACCATGCAGCGTAAACCGAA
441 S Y D Q I D M T E S S G A T M Q R K T E
1381 CTGCTGCTGGTTCCGTCCTTT TAA (SEQ ID NO:51)
461 L L L V P S F (SEQ ID NO:52)

FIG. 26

A13 Conductor

CATATGACCGTACACGACATCATCGCAACGTACTTCACTAAATGGTACGTAATTGTGCCGCTGGCACTGATTGCGTATC
 GCGTGCTGGATTATTTCTACGCGACCCGTTCTAAAGCACTAAGAAATCTCTGCCGGAACCGTGGCGTCTGCCAATCAT
 CGGTACATGCACCACCTGATCGGCACCAACCCGACCCGTGGCGTACGCGACCTGGCGGTAAGTACGGCTCTCTGATG
 CATCTGCAGCTGGGCGAGGTACCTACTATCTGCTGTTCTCTCCCGAAGTGGGCCAAAGAAATCCTGACTACCTATGACA
 TCACTTTCCGCCAACCGCCCGAAACGCTGACCGCGGAAATTTGTCTGTACCATTAACACGGATGTGGTTCTGGCCCCGTA
 CGGTGAGTACTGGCGCCAGCTGCGCAAAATTTGTACTCTGGAACCTGCTGAGCGTTAAAGGTTAAATCCTTCCAGAGC
 CTGCGTGAAGAGGAATGCTGGAACCTGGTGCAGGAGATTAAAGCGTCTGGCAGCGGTCTCCAGTTAACCTGTCTGAGA
 ATGTTTTTAAACTGATCGCTACTATCTCTCTCGCGCGGCATTCCGTAAGGTATCAAAGATCAGAAAGAACTGACCGA
 AATCGTTAAGGAAATCCTGCGCCAGACTGGTGGCTTCGACGTTGCGGACATCTTCCCGTCCAAAAAGTTCTGCAACCAT
 CTGCTGCGCAACGCGCTCTGCTGACCTCCCTGCGTAAGAAAATTGATAACCTGATTGACAACCTGGTGGTGGAGCACA
 CTGTGAACACCTCTTCTAAACCAACGAAACCTGCTGGACGTACTGCTGCGCCTGAAGGACTCTGCCGAATTTCCACT
 GACTAGCGCAATATCAAAGCAATCATCTGGACATGTTTCGGCGCGGTACCGGATACGTCCTCTTCCACGATTGAGTGG
 GCTATTTCCGAACCTGATCAAATGCCGAAGGCGATGGAAAAAGTGCAGGCGGAACCTGCGTAAAGCGCTGAACGGTAAAG
 AGAAAATTCATGAAGAGGACATCCAGGAATCTGCTACCTGAATATGGTAATCAAAGAACTCTGCGTCTGCTACCGCC
 GCTGCCACTGGTTCTGCGCGCTGAATGCCGTCAGCCGGTTAACCTGGCCGGCTACAACATTCGGAACAAAACGAAGCTG
 ATCGTCAACGTTTTTCGCGATCAACCGCGATCTCTGAATACCTGGAAGACGCGGAAGCGTTCACTCCGGAACGCTTTGAGA
 ACTCTCTGCGCAACCGTTATGGGCGCTGAATACGAGTACCTGCCGTTTCGGTGGCGGTGCGCGTATGTGCCCGGGTGTCTG
 ACTGGGCTGGCGAACGTTCAACTGCCACTGGCGAACATCTGTACCACTTCAACTGGAACCTGCCTAACGGCGTATCTT
 TATGATCAAAATCGACATGACCGAAAGCTCCGGCGGACCATGCAGCGTAAACCGAACTGCTGCTGGTCCGCTCTTTT

>Nde ← A13N-term
 AACCTAGGCAATATGACCGTACACGACATCATCGCAACGTACTTCACTAAATGGTACGTAATTGTGCCGCTGGCACTGAT
 Start

→ >Amorphadiene Oxidase
 TGGTATCGCGTCTGGATTATTTCTACGCGACCCGTTCTAAAGCACTAAGAAATCTCTGCCGGAACCGTGGCGTCTG
 CCAATCATCGGTACATGCACCACCTGATCGGCACCAACCCGACCCGTGGCGTACGCGACCTGGCGGTAAGTACGGCT
 CTCTGATGCATCTGCAGCTGGGCGAGGTACCTACTATCTGCTGTTCTCTCCCGAAGTGGGCCAAAGAAATCCTGACTAC
 CTATGACATCACTTTCCGCAACCGCCCGGAAACGCTGACCGCGGAAATTTGTCTGTACCATTAACACGGATGTGGTTCTG
 GCCCGTACGGTGAGTACTGGCGCCAGCTGCGCAAAATTTGTACTCTGGAACCTGCTGAGCGTTAAAGGTTAAATCCT
 TCCAGAGCCTGCGTGAAGAGGAATGCTGGAACCTGGTGCAGGAGATTAAAGCGTCTGGCAGCGGTCTCCAGTTAACCT
 GTCTGAGAAATGTTTTTAAACTGATCGCTACTATCTGTCTCGCGCGGCATTCCGTAAGGTATCAAAGATCAGAAAGAA
 CTGACCGAAATCGTTAAGGAAATCCTGCGCCAGACTGGTGGCTTCGACGTTGCGGACATCTTCCCGTCCAAAAAGTTCC
 TGCACCATCTGTCTGGCAACCGCTCGTCTGACCTCCCTGCGTAAGAAAATTGATAACCTGATTGACAACCTGGTCTGC
 TGAGCACACTGTGAACACCTCTTCTAAACCAACGAAACCTGCTGGACGTACTGCTGCGCCTGAAGGACTCTGCCGAA
 TTTCCACTGACTAGCGACAATATCAAAGCAATCATCTGGACATGTTTCGGCGCCGGTACCGATACGTCCTCTTCCACGA
 TTGAGTGGGCTATTTCCGAACCTGATCAAATGCCGAAGGCGATGGAAAAAGTGCAGGCGGAACCTGCGTAAAGCGCTGAA
 CGGTAAGAGAAAAATTCATGAAGAGGACATCCAGGAACCTGTCTACCTGAATATGGTAATCAAAGAACTCTGCGTCTG
 CATCCGCGCTGCCACTGGTTCTGCGCGCTGAATGCCGCTGAGCCGTTAACCTGGCCGGCTACAACATTCGGAACAAA
 CGAAGCTGATCGTCAACGTTTTTCGCGATCAACCGCGATCTCTGAATACTGGAAGACGCGGAAGCGTTCACTCCGGAACG
 CTTTGAGAACTCTCTGCGCAACCGTTATGGGCGCTGAATACGAGTACCTGCCGTTTCGGTGGCGGTGCGCGTATGTGCCCG
 GGTGCTGCACTGGGCTGGCGAACGTTCAACTGCCACTGGCGAACATCTGTACCACTTCAACTGGAACCTGCCTAACG
 CGGTATCTTATGATCAAAATCGACATGACCGAAAGCTCCGGCGGACCATGCAGCGTAAACCGAACTGCTGCTGGTCTTCC

>Stop
 GTCCTTTTAA CCTAGG (SEQ ID NO:53)
 >AvrII

FIG. 27

Mapa de Translação
Início

```

1 ATG
1 M
A13N-term
1 ACCGTACACGACATCATCGCAACGTACTTCACTAAATGGTACGTAATTGTGCCGCTGGCA
1 T V H D I I A T Y F T K W Y V I V P L A
61 CTGATTGCGTATCGCGTGTGGATTATTTCTAC
21 L I A Y R V L D Y F Y
Amorphadiene Oxidase
1 GCGACCCGTTCTAAAAGCACTAAGAAATCTCTGCCGGAACCGTGGCGTCTGCCAATCATC
1 A T R S K S T K K S L P E P W R L P I I
61 GGTACATGCACCACCTGATCGGCACCACCCCGCACCGTGGCGTACGCGACCTGGCGCGT
21 G H M H H L I G T T P H R G V R D L A R
121 AAGTACGGTCTCTGATGCATCTGCAGCTGGGCGAGGTACCTACTATCGTCGTTTCCTCC
41 K Y G S L M H L Q L G E V P T I V V S S
181 CCGAAGTGGGCCAAAGAAATCCTGACTACCTATGACATCACTTTCCGCCAACCGCCCGGAA
61 P K W A K E I L T T Y D I T F A N R P E
241 ACGCTGACCGCGCAAATGTCTGTACCATAACACGGATGTGGTTCTGGCCCCGTACGGT
81 T L T G E I V L Y H N T D V V L A P Y G
301 GAGTACTGGCGCCAGCTGCGCAAAATTTGTACTCTGGAACCTGCTGAGCGTTAAAAAGGTT
101 E Y W R Q L R K I C T L E L L S V K K V
361 AAATCCTTCCAGAGCCTGCGTGAAGAGGAATGCTGGAACCTGGTGCAGGAGATTAAAGCG
121 K S F Q S L R E E E C W N L V Q E I K A
421 TCTGGCAGCGTCTGTCAGTTAACCTGTCTGAGAATGTTTTTAACTGATCGCTACTATC
141 S G S G R P V N L S E N V F K L I A T I
481 CTGTCTCGCGCGCATTCGGTAAAGGTATCAAAGATCAGAAAGAACTGACCGAAATCGTT
161 L S R A A F G K G I K D Q K E L T E I V
541 AAGGAAATCCTGCGCCAGACTGGTGGCTTCGACGTTGCGGACATCTCCCGTCCAAAAG
181 K E I L R Q T G G F D V A D I F P S K K
601 TTCCTGCACCATCTGTCTGGCAAACGCGCTCGTCTGACCTCCCTGCGTAAGAAAATTGAT
201 F L H H L S G K R A R L T S L R K K I D
661 AACCTGATTGACAACCTGGTCTGAGCACACTGTGAACACCTCTTCTAAACCAAGGAA
221 N L I D N L V A E H T V N T S S K T E
721 ACCCTGCTGGACGTACTGCTGCGCCTGAAGGACTCTGCCGAATTTCCACTGACTAGCGAC
241 T L L D V L L R L K D S A E F P L T S D
781 AATATCAAAGCAATCATCCTGGACATGTTCCGGCGCGGTACCGATACGTCCTCTTCCACG
261 N I K A I I L D M F G A G T D T S S S T
841 ATTGAGTGGGCTATTTCCGAACCTGATCAAATGCCCGAAGGCGATGGAAAAAGTGCAGGCG
281 I E W A I S E L I K C P K A M E K V Q A
901 GAACTGCGTAAAGCGCTGAACGGTAAAGAGAAAATTCATGAAGAGGACATCCAGGAACGT
301 E L R K A L N G K E K I H E E D I Q E L
961 TCCTACCTGAATATGGTAATCAAAGAACTCTGCGTCTGCATCCGCGCTGCCACTGGTT
321 S Y L N M V I K E T L R L H P P L P L V
1021 CTGCCGCGTGAATGCCGTCAGCCGGTTAACCTGGCCGGCTACAACATTCGGAACAAAACG
341 L P R E C R Q P V N L A G Y N I P N K T
1081 AAGCTGATCGTCAACGTTTTCGCGATCAACCGCGATCCTGAATACTGGAAAGACGCGGAA
361 K L I V N V F A I N R D P E Y W K D A E
1141 GCGTTCATTCCGGAACGCTTTGAGAACTCCTCTGCCACCGTTATGGGCGCTGAATACGAG
381 A F I P E R F E N S S A T V M G A E Y E
1201 TACCTGCCGTTTCGGTGCGGGTGCGCGTATGTGCCCGGGTGTGCACTGGGCGTGGCGAAC
401 Y L P F G A G R R M C P G A A L G L A N
1261 GTTCAACTGCCACTGGCGAACATCCTGTACCACTTCAACTGGAACTGCCTAACGGCGTA
421 V Q L P L A N I L Y H F N W K L P N G V
1321 TCTTATGATCAAATCGACATGACCGAAAGCTCCGGCGCGACCATGCAGCGTAAAACCGAA
441 S Y D Q I D M T E S S G A T M Q R K T E
1381 CTGCTGCTGGTTCCGTCCTTTTAA (SEQ ID NO:54)
461 L L L V P S F (SEQ ID NO:55)

```

FIG. 28

A17 Conditor

CATATGATCGAACAACTGCTGGAATACTGGTACGTGGTTGTGCCTGTTCTGTATATTATCAAACAGCTGCTGGCGTACA
 CTAAAGCGACCCGTTCTTAAAGCACTAAGAAATCTCTGCCGGAACCGTGGCGTCTGCCAATCATCGGTGACATGCACCA
 CCTGATCGGCACCAACCCGACCGTGGCGTACGCGACCTGGCGCGTAAGTACGGCTCTCTGATGCATCTGCAGCTGGGC
 GAGGTACCTACTATCGTCTGTTCTCTCCCGAAGTGGGCCAAAGAAATCTGACTACCTATGACATCACTTTCGCCAACC
 GCGCGGAAACGCTGACCGGCGAAATTGTCTGTACCATAAACACGGATGTGGTTCTGGCCCGTACGGTGAAGTACTGGCG
 CCAGCTGCGCAAAATTTGTACTCTGGAATCTGAGCGTTAAAGGTTAAATCCTTCCAGAGCCTGCGTGAAGAGGAA
 TGCTGGAACCTGGTGCAGGAGATTAAAGCGTCTGGCAGCGGTCTCCAGTTAACCTGTCTGAGAAATGTTTTAACTGA
 TCGCTACTATCCTGTCTCGCGCGCATTCGGTAAAGGTATCAAAGATCAGAAAGAACTGACCGAAATCGTTAAGGAAAT
 CCTGCCCGACAGTGGTGGCTTCGACGTTGCGGACATCTTCCCGTCCAAAAAGTTCTGCACCATCTGTCTGGCAACGC
 GCTCGTCTGACCTCCCTGCGTAAGAAATTTGATAACCTGATTGACAACCTGGTCTGCTGAGCAGCTGTGAACACCTCTT
 CTAACCAACGAAACCTGCTGGACGTACTGCTGCGCTGAAGGACTCTGCCGAATTTCCACTGACTAGCGACAATAT
 CAAAGCAATCATCTGGACATGTTCCGCGCGGTACCGATACGTCCTCTTCCACGATTGAGTGGGCTATTTCCGAACTG
 ATCAAATGCCGGAAGCGATGGAAAAAGTGCAGGCGGAACCTGCGTAAAGCGCTGAACGGTAAAGAGAAAAATTCATGAAG
 AGGACATCCAGGAACCTGCTTACCTGAATATGGTAATCAAAGAACTCTGCGTCTGCATCCGCCGCTGCCACTGGTTCT
 GCGCGTGAATGCCGTGAGCGGTTAACTGGCGCGGTACAACATTCGGAACAAACGAAGCTGATCGTCAACGTTTTCT
 GCGATCAACCGCGATCTGAATACTGGAAGACGCGGAAGCGTTCATTCCGGAACGCTTTGAGAACTCCTCTGCCACCG
 TTATGGCGCTGAATACGAGTACCTGCCGTTCCGTGCGGGTCCCGGTATGTGCCCGGGTGTGCACTGGGCGTGGCGAA
 CGTTCAACTGCCACTGGCGAACATCTGTACCACTTCACTGGAACCTGCCTAACGGCGTATCTTATGATCAAATCGAC

>Nde

ATGACCGAAAGCTCCGGCGGACCATGCGAGCGTAAACCGAACTGCTGCTGGTTCCGTCCTTTTAACTAGGCATATG

>Start

A17

ATCGAACAACTGCTGGAATACTGGTACGTGGTTGTGCCTGTTCTGTATATTATCAAACAGCTGCTGGCGTACACTAA

>Amorphadiene Oxidase

GCGACCCGTTCTAAAGCACTAAGAAATCTCTGCCGGAACCGTGGCGTCTGCCAATCATCGGTGACATGCACCACCTGA
 TCGGCACCAACCCCGCACCGTGGCGTACGCGACCTGGCGCGTAAGTACGGCTCTCTGATGCATCTGCAGCTGGGCGAGGT
 ACCTACTATCGTCTGTTCTCTCCCGAAGTGGGCCAAAGAAATCTGACTACCTATGACATCACTTTCGCCAACCGCCCG
 GAAACGCTGACCGGCGAAATTTCTCTGTACCATAAACCGGATGTGGTTCTGGCCCGTACGGTGAAGTACTGGCGCCAGC
 TGCGCAAAATTTGTACTCTGGAATCTGTAGCGTTAAAGGTTAAATCCTTCCAGAGCCTGCGTGAAGAGGAATGCTG
 GAACCTGTGTCAGGAGATTAAAGCGTCTGGCAGCGGTCTCCAGTTAACCTGTCTGAGAAATGTTTTAACTGATCGCT
 ACTATCCTGTCTCGCGCGCATTCGGTAAAGGTATCAAAGATCAGAAAGAACTGACCGAAATCGTTAAGGAAATCCTGC
 GCCAGACTGGTGGCTTCGACGTTGCGGACATCTTCCCGTCCAAAAAGTTCTGCACCATCTGTCTGGCAACCGCGCTCG
 TCTGACCTCCTGCGTAAGAAATTTGATAACCTGATTGACAACCTGGTCTGCTGAGCAGCTGTGAACACCTCTTCTAAA
 ACCAACGAAACCCCTGCTGGACGTACTGCTGCGCTGAAGGACTCTGCCGAATTTCCACTGACTAGCGACAATATCAAAG
 CAATCATCTGGACATGTTTCGGCGCCGGTACCGATACGTCCTTCCACGATTGAGTGGGCTATTTCGAACTGATCAA
 ATGCCCGAAGGCGATGGAAAAAGTGCAGGCGGAACCTGCGTAAAGCGCTGAACGGTAAAGAGAAAAATTCATGAAGAGGAC
 ATCCAGGAATGTCCTACCTGAATATGGTAATCAAAGAACTCTGCGTCTGCATCCGCCGCTGCCACTGGTTCTGCCGC
 GTGAATGCCGTGACCGGTTAACCTGGCCGCTACAACTTCCGAACAAACGAAGCTGATCGTCAACGTTTTCTCGCAT
 CAACCGCATCTCTGAATACTGGAAGACGCGGAAGCGTTCAATCCGGAACGCTTTGAGAACTCCTCTGCCACCGTTATG
 GGCCTGAATACGAGTACCTGCCGTTCCGTGCGGGTCCCGGTATGTGCCCGGGTGTGCACTGGGCGTGGCGAACGTTT
 AACTGCCACTGGCGAACATCTGTACCACTTCACTGGAACCTGCCTAACGGCGTATCTTATGATCAAATCGACATGAC

Stop

CGAAAGCTCCGGCGGACCATGCGAGCGTAAACGGAAGTCTGCTGGTTCCGTCCTTTTAACTAGG

>AvrII

(SEQ ID NO:56)

FIG. 29

Mapa de Translação
Início

1 ATG
1 M

A17

1 ATCGAACAACCTGCTGGAATACTGGTACGTGGTGTGCCTGTTCTGTATATTATCAAACAG
1 I E Q L L E Y W Y V V V P V L Y I I K Q
61 CTGCTGGCGTACACTAAA
21 L L A Y T K

Amorphadiene Oxidase

1 GCGACCCGTTCTAAAAGCACTAAGAAATCTCTGCCGGAACCGTGGCGTCTGCCAATCATC
1 A T R S K S T K K S L P E P W R L P I I
61 GGTACATGCACCACCTGATCGGCACCACCCCGCACCGTGGCGTACGCGACCTGGCGCGT
21 G H M H H L I G T T P H R G V R D L A R
121 AAGTACGGCTCTCTGATGCATCTGCAGCTGGGCGAGGTACCTACTATCGTCGTTTCCTCC
41 K Y G S L M H L Q L G E V P T I V V S S
181 CCGAAGTGGGCCAAAGAAATCCTGACTACCTATGACATCACTTTCCGCAACCGCCCGGAA
61 P K W A K E I L T T Y D I T F A N R P E
241 ACGCTGACCGGCGAAATTGTCCTGTACCATAACACGGATGTGGTTCTGGCCCCGTACGGT
81 T L T G E I V L Y H N T D V V L A P Y G
301 GAGTACTGGCGCCAGCTGCGCAAAATTTGTACTCTGGAACCTGCTGAGCGTTAAAAAGGTT
101 E Y W R Q L R K I C T L E L L S V K K V
361 AAATCCTTCCAGAGCCTGCGTGAAGAGGAATGCTGGAACCTGGTGCAGGAGATTAAAGCG
121 K S F Q S L R E E E C W N L V Q E I K A
421 TCTGGCAGCGGTCTCCAGTTAACCTGCTGAGAATGTTTTTAACTGATCGCTACTATC
141 S G S G R P V N L S E N V F K L I A T I
481 CTGTCTCGCGCGCATTCGGTAAAGGTATCAAAGATCAGAAAGAACTGACCGAAATCGTT
161 L S R A A F G K G I K D Q K E L T E I V
541 AAGGAAATCCTGCGCCAGACTGGTGGCTTCGACGTTGCGGACATCTTCCGTTCCAAAAAG
181 K E I L R Q T G G F D V A D I F P S K K
601 TTCCTGCACCATCTGTCTGGCAAACCGCTCGTCTGACCTCCCTGCGTAAGAAAATTGAT
201 F L H H L S G K R A R L T S L R K K I D
661 AACCTGATTGACAACCTGGTCTGCTGAGCACACTGTGAACACCTCTTCTAAAACCAACGAA
221 N L I D N L V A E H T V N T S S K T N E
721 ACCCTGCTGGACGTACTGCTGCGCCTGAAGGACTCTGCCGAATTTCCACTGACTAGCGAC
241 T L L D V L L R L K D S A E F P L T S D
781 AATATCAAAGCAATCATCCTGGACATGTTCCGGCGCCGGTACCGATACGTCTCTTCCACG
261 N I K A I I L D M F G A G T D T S S S T
841 ATTGAGTGGGCTATTTCCGAACTGATCAAATGCCCGAAGGCGATGGAAAAAGTGCAGGCG
281 I E W A I S E L I K C P K A M E K V Q A
901 GAACTGCGTAAAGCGCTGAACGGTAAAGAGAAAATTTCATGAAGAGGACATCCAGGAACTG
301 E L R K A L N G K E K I H E E D I Q E L
961 TCCTACCTGAATATGGTAATCAAAGAAACTCTGCGTCTGCATCCGCCGCTGCCACTGGTT
321 S Y L N M V I K E T L R L H P P L P L V
1021 CTGCCGCGTGAATGCCGTGAGCCGGTTAACCTGGCCGGCTACAACATTCCGAACAAAACG
341 L P R E C R Q P V N L A G Y N I P N K T
1081 AAGCTGATCGTCAACGTTTTCGCGATCAACCGCGATCCTGAATACTGGAAAGACGCGGAA
361 K L I V N V F A I N R D P E Y W K D A E
1141 GCGTTCATTCCGGAACGCTTTGAGAACTCCTCTGCCACCGTTATGGGCGCTGAATACGAG
381 A F I P E R F E N S S A T V M G A E Y E
1201 TACCTGCCGTTTCGGTGGGGTTCGCGTATGTGCCCGGGTCTGCACTGGGCCTGGCGAAC
401 Y L P F G A G R R M C P G A A L G A L A N
1261 GTTCAACTGCCACTGGCGAACATCCTGTACCCTTCAACTGGAAACTGCCTAACGGCGTA
421 V Q L P L A N I L Y H F N W K L P N G V
1321 TCTTATGATCAAATCGACATGACCGAAAGCTCCGGCGCGACCATGCAGCGTAAACCGGAA
441 S Y D Q I D M T E S S G A T M Q R K T E
1381 CTGCTGCTGGTTCCGTCTTTTAA (SEQ ID NO:57)
461 L L L V P S F (SEQ ID NO:58)

>Stop
TGGTCCGTCCTTTTAACTAGG (SEQ ID NO:59)
>AvrII

FIG. 31

Mapa de Translação

1 ATG
1 M

Bovino

1 GCTCTGCTGCTGGCTGTCTTCCTGGGTCTGTCCTGCCTGCTGCTGCTGTCCCTGTGG
1 A L L L A V F L G L S C L L L L S L W

Amorphadiene Oxidase

1 GCGACCCGTTCTAAAAGCACTAAGAAATCTCTGCCGGAACCGTGGCGCTCTGCCAATCATC
1 A T R S K S T K K S L P E P W R L P I I
61 GGTACATGCACCACCTGATCGGCACCACCCCGCACCGTGGCGGTACGCGACCTGGCGCGT
21 G H M H H L I G T T P H R G V R D L A R
121 AAGTACGGCTCTCTGATGCATCTGCAGCTGGGCGAGGTACCTACTATCGTCTGTTTCTCC
41 K Y G S L L M H L Q L G E V P T I V V S S
181 CCGAAGTGGGCCAAAGAAATCCTGACTACCTATGACATCACTTTTCGCCAACCGCCCGGAA
61 P K W A K E I L T T Y D I T F A N R P E
241 ACGCTGACCGGCGAAATTTGCTCTGTACCATAACACGGATGTGGTTCTGGCCCCGTACGGT
81 T L T G E I V L Y H N T D V V L A P Y G
301 GAGTACTGGCGCCAGCTGCGCAAAATTTGTACTCTGGAACCTGCTGAGCGTTAAAAAGGTT
101 E Y W R Q L R K I C T L E L L S V K K V
361 AAATCCTTCCAGAGCCTGCGTGAAGAGCAATGCTGGAACCTGGTGCAGGAGATTAAAGCG
121 K S F Q S L R E E E C W N L V Q E I K A
421 TCTGGCAGCGGTCTGCCAGTTAACTGTCTGAGAATGTTTTTAACTGATCGCTACTATC
141 S G S G R P V N L S E N V F K L I A T I
481 CTGTCTCGCGCGGCATTCGGTAAAGGTATCAAAGATCAGAAAGAACTGACCGAAATCGTT
161 L S R A A F G K G I K D Q K E L T E I V
541 AAGGAAATCCTGCGCCAGACTGGTGGCTTCGACGTTGCGGACATCTTCCCGTCCAAAAAG
181 K E I L R Q T G T G G G F D V A D C I F P S K K
601 TTCCTGACCATCTGTCTGGCAACCGCTCGTCTGACCTTCCGTAAGAAAATTTGAT
201 F L H H L S G K R A R L T S L R K K I D
661 AACCTGATTGACAACCTGGTGCCTGAGCACACTGTGAACACCTCTTCTAAACCAACGAA
221 N L I D N L V A E H T V N T S S K T N E
721 ACCCTGCTGGACGTACTGCTGCGCCTGAAGGACTCTGCCGAATTTCCACTGACTAGCGAC
241 T L L D V L L R L K D S A E F P L T S D
781 AATATCAAAGCAATCATCTGACATGTTTCGGCGCCGGTACCGATACGTCTCTTCCACG
261 N I K A I I L A D M F A G T G T D T S S S T
841 ATTGACTGGGCTATTTCGAACTGATCAAATGCCGAAGGCGATGGAAAAGTGCAGCGG
281 I E W A I S E L I K C P K A M E K V Q A
901 GAACTGCGTAAAGCGCTGAACGGTAAAGAGAAAATTCATGAAGAGGACATCCAGGAACTG
301 E L R K A L N G K E K I H E E D I Q E L
961 TCCTACCTGAATATGTAATCAAAGAACTCTGCGTCTGCATCCGCCGCTGCCACTGGTT
321 S Y L N M V I K E T L R L H P P L P L V
1021 CTGCCGCGTGAATGCCGTCAGCCGGTTAACTGGCCGGCTACAACATTTCCGAACAAAACG
341 L P R E C R Q P V N L A G Y N I P N K T
1081 AAGCTGATCGTCAACGTTTTTCGCGATCAACCGCATCTCTGAATACTGGAAAGACGCGGAA
361 K L I V N V F A I N R D P E Y ⁸¹ K D A E
1141 GCGTTCATTCCGGAACGCTTTGAGAACTCCTCTGCCACCGTTATGGGCGCTGAATACGAG
381 A F I P E R F E N S S A T V M G A E Y E
1201 TACCTGCCGTTCCGTGCGGGTTCGCCGTATGTGCCGGGTGCTGCACTGGGCGCTGGCGAAC
401 Y L P F G A G R R M C P G A A L G L A N
1261 GTTCAACTGCCACTGGCGAACATCCTGTACCATTCAACTGGAACTGCCTAACGGCGTA
421 V Q L P L A N I L Y H F N W K L P N G V
1321 TCTTATGATCAAATCGACATGACCGAAAGCTCCGGCGCGACCATGCAGCGTAAACCGGAA
441 S Y D Q I D M T E S S G A T M Q R K T E
1381 CTGCTGCTGGTTCCGTCCTTTTAA (SEQ ID NO:60)
461 L L L V P S F (SEQ ID NO:61)

FIG. 32

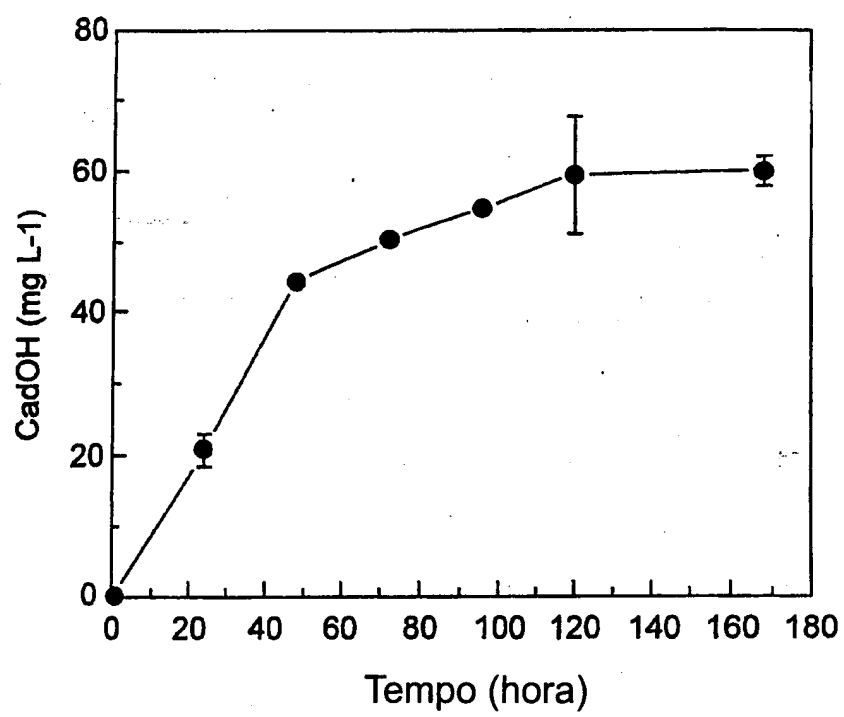


FIG. 33

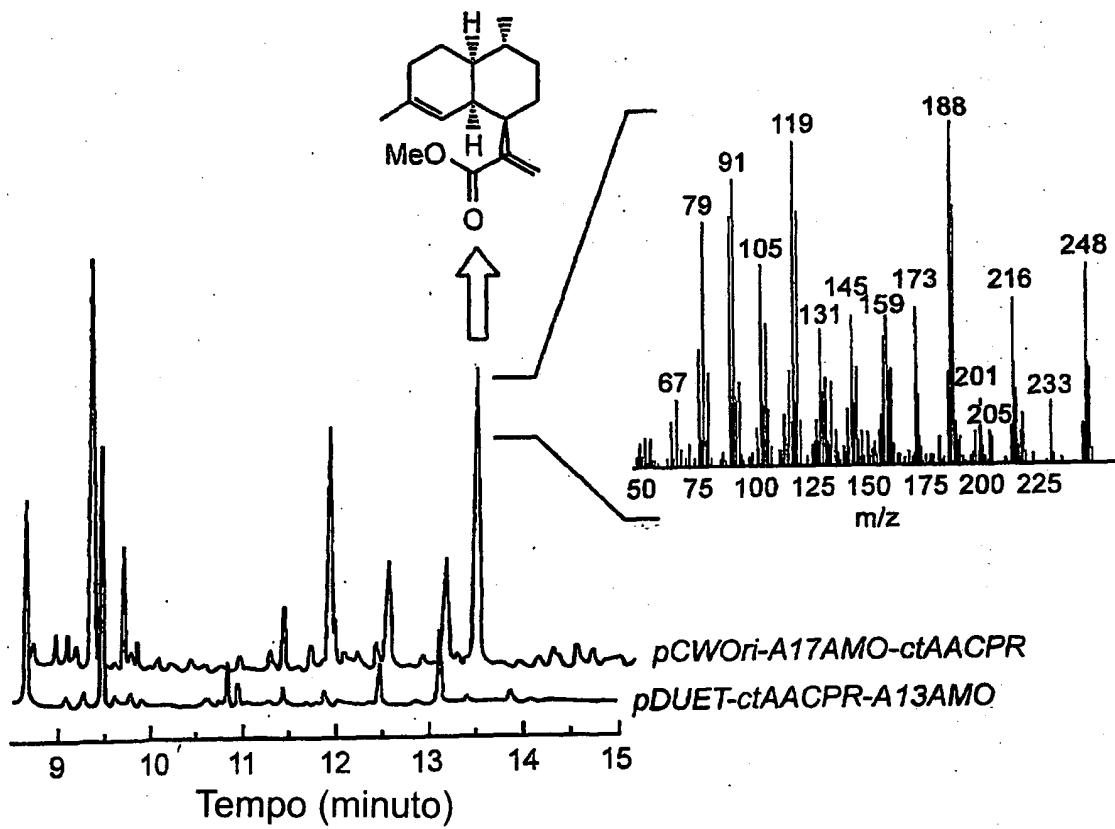


FIG. 34

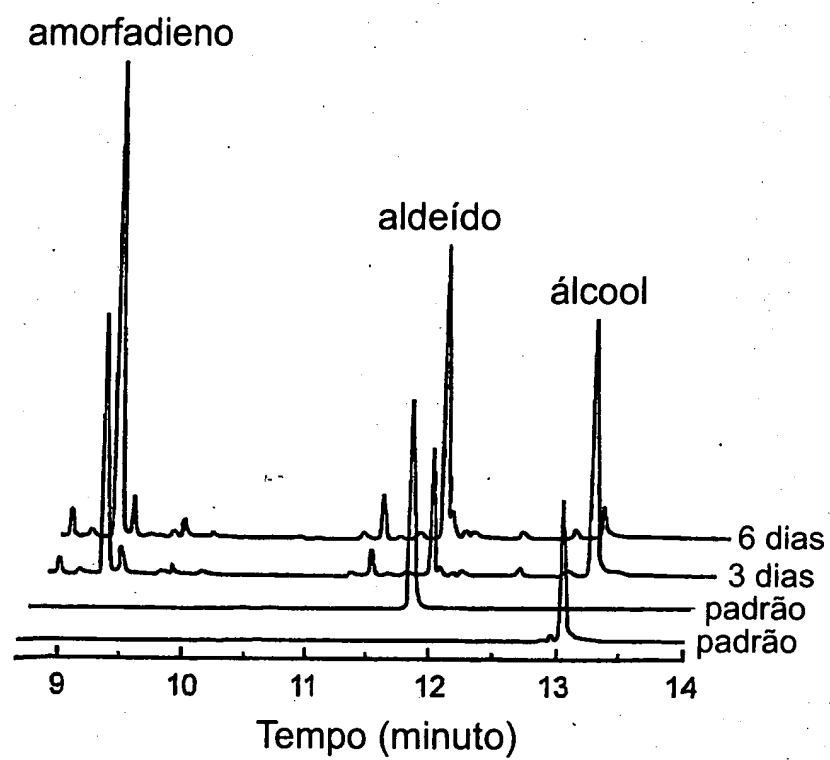


FIG. 35A

MevT

→ *atoB*

CTTGATATCGAATTCCTGCAGCCCGGGGATCCTCTAGAGTCGACTAGGAGGAATATAAAATGAAAAATTGTGTCA
 TCGTCAGTGCGGTACGTACTGCTATCGGTAGTTTTAACGGTTCACTCGCTTCCACCAGCGCCATCGACCTGGGGG
 CGACAGTAATTAAGCCGCCATTGAACGTGCAAAAATCGATTACAAACACGTTGATGAAGTGATTATGGGTAACG
 TGTTACAAGCCCGGCTGGGGCAAAATCCGGCGCGTCAGGCACTGTTAAAAAGCGGGCTGGCAGAAACGGTGTGCG
 GATTACGGTCAATAAAGTATGTGGTTCGGGTCTTAAAGTGTGGCGCTTGCCGCCAGGCCATTAGGCAGGTC
 AGGCGCAGAGCATTGTGGCGGGGGGTATGGAATAATGAGTTTAGCCCCCTACTTACTCGATGCAAAAGCACGCT
 CTGGTTATCGTCTTGGAGACGGACAGGTTTATGACGTAATCCTGCGCGATGGCCTGATGTGCGCCACCCATGGTT
 ATCATATGGGGATTACCGCGAAACGTGGCTAAAGAGTACGGAATTACCCGTGAATGCAGGATGAAGTGGCGC
 TACATTACAGCGTAAGCGCGCAGCCGAATTGAGTCCGGTGTCTTTACAGCCGAAATCGTCCCGGTAAATGTTG
 TCACTCGAAAGAAAACCTTCGTCTTCAGTCAAGACGAATCCCGAAAGCGAATCAACGGCTGAAGCGTTAGGTG
 CATTGCGCCCGCCTTCGATAAAGCAGGAACAGTCACCGCTGGGAACGCGTCTGGTATTAACGACGGTGTGCGG
 CTCCTGGTGATTATGGAAGAATCTGCGGCGCTGGCAGCAGGCCTTACCCCCCTGGCTCGCATTAAAGTTATGCCA
 GCGGTGGCGTGCCCCCGCATTTGATGGGTATGGGGCCAGTACCTGCCACGCAAAAAGCTTACAACCTGGCGGGG
 TGCAACTGGCGGATATTGATCTCATTGAGGCTAATGAAGCATTGCTGACAGTTCCTTGCCGTTGGGAAAAACC
 TGGGCTTTGATTCTGAGAAAGTGAATGTCAACGGCGGGGCCATCGCGCTCGGGCATCTATCGGTGCCAGTGGTG
 CTCGTATTCTGGTCACACTATTACATGCCATGCAGGCACGCGATAAAACGCTGGGGCTGGCAACACTGTGCATTG

atoB ← | → *HMGS*

GCGGCGGTCAGGGAATTGCGATGGTGATTGAACGGTTGAATTAAGGAGGACAGCTAAATGAAACTCTCAACTAAA
 CTTTGTGGTGTGGTATTAAAGGAAGACTTAGGCCGCAAAAGCAACAATTACACAATACAACTTGGCAAAATG
 ACTGAACTAAAAAACAAGACCGCTGAACAAAAACCAGACCTCAAAATGTGGTATTAAAGGTATCCAAAT
 TACATCCCACTCAATGTGTCAACCAATCTGAGCTAGAGAAATTTGATGGCGTTTCTCAAGGTAAATACAAAT
 GGTCTGGGCCAAACCAACATGTCTTTTGTCAATGACAGAGAAGATATCTACTCGATGTCCCTAACTGTTTGTCT
 AAGTTGATCAAGGTTACAACATCGACACCAACAAAATGGTAGATTAGAAGTCGGTACTGAAACTCTGATTGAC
 AAGTCCAAGTCTGTCAAGTCTGTCTTGATGCAATTGTTTGGTGAAAACTGACGTCGAAGGTATTGACACGCTT
 AATGCTGTACGGTGGTACCAACGCGTTGTCAACTCTTTGAACTGGATTGAATCTAACGCATGGGATGGTAGA
 GACGCTATTGTAGTTTGGGTGATATTGCCATCTACGATAAGGGTGCCGCAAGACCAACCGGTGGTGCCGTTACT
 GTTGCTATGTGGATCGGTCTGATGCTCCAATTGTATTTGACTCTGTAAGAGCTTCTTACATGGAACACGCCCTAC
 GATTTTACAAGCCAGATTTACCAGCGAATATCCTTACGTCGATGGTCATTTTTCATTAACTTGTTACGTCAG
 GCTCTTGATCAAGTTTACAAGATTATTTCAAGAAGGCTATTTCTAAAGGGTTGGTTAGCGATCCCGCTGGTTCTG
 GATGCTTTGAACGTTTTTGAATATTTGACTACAACGTTTTCCATGTTCCAACCTGTAAATTGGTCAAAAATCA
 TACGGTAGATTACTATATAACGATTTTCAGAGCCAATCCTCAATTGTTCCAGAAGTTGACGCCGAATTAGCTACT
 CGCGATTATGACGAATCTTTAACCGATAAGAACATTGAAAAAATCTTTGTTAATGTTGCTAAGCCATTCCACAAA
 GAGAGAGTTGCCAATCTTTGATTGTTTCCAACAAACACAGGTAACATGTACACCGCATCTGTTTATGCCGCTTT
 GCATCTCTATTAAACTATGTGGATCTGACGACTTACAAGGCAAGCGTGTGGTTTATTTTCTTACGGTTCCGGT
 TTAGCTGCATCTCTATATTCTTGCAAAATGTTGGTGACGTCCAACATATTATCAAGGAATTAGATATTACTAAC
 AAATTAGCCAAGAGAATCACCGAACTCCAAAGGATTACGAAGCTGCCATCGAATTGAGAGAAAATGCCCATTTG
 AAGAAGAATCTCAACCTCAAGTTCCATTGAGCATTGTGCAAGTGGTGTCTTACTACTTGACCAACATCGATGAC
 AAATTTAGAAGATCTTACGA

HMGS ← | → *HMGR truncado*

TGTTAAAAATAAGGAGGATTACACTATGGTTTAAACCAATAAAACAGTCATTTCTGGATCGAAAGTCAAAAGTT
 TATCATCTGCGCAATCGAGCTCATCAGGACCTTCATCATCTAGTGAGGAAGATGATTCCCGCGATATTGAAAGCT
 TGGATAAGAAAATACGTCCTTTAGAAGAATTAGAAGCATTATTAAGTAGTGGAAATACAAAAAATTGAAGAACA
 AAGAGGTGCTGCCT

FIG. 35B

TGGTTATTACCGGTAAGTTACCTTTGTACGCTTTGGAGAAAAAATTAGGTGATACTACGAGAGCGGTTGCGGTAC
GTAGGAAGGCTCTTTCAATTTTGGCAGAAGCTCCTGTATTAGCATCTGATCGTTTACCATATAAAAAATTATGACT
ACGACCGCGTATTGGCGCTTGTGTGAAAATGTTATAGGTTACATGCCCTTGCCCGTTGGTGTATTAGGCCCTT
TGGTTATCGATGGTACATCTTATCATATACCAATGGCAACTACAGAGGGTTGTTTGGTAGCTTCTGCCATGCGTG
GCTGTAAGGCAATCAATGCTGGCGGTGGTGCAACAACCTGTTTTAACTAAGGATGGTATGACAAGAGGCCCAGTAG
TCCGTTTCCCAACTTTGAAAAGATCTGGTGCCTGTAAGATATGGTTAGACTCAGAAGAGGGACAAAACGCAATTA
AAAAAGCTTTTAACTCTACATCAAGATTTGCACGTCTGCAACATATTCAAACCTTGTCTAGCAGGAGATTTACTCT
TCATGAGATTTAGAACAACACTACTGGTGACGCAATGGGTATGAATATGATTTCTAAAGGTGTCGAATACFCATTAA
AGCAAATGGTAGAAGAGTATGGCTGGGAAGATATGGAGGTTGTCTCCGTTTCTGGTAACTACTGTACCGACAAAA
AACCAGCTGCCATCAACTGGATCGAAGGTGCTGGTAAGAGTGTCTCGTCGCAGAAGCTACTATTCTGGTGATGTTG
TCAGAAAAGTGTAAAAAGTGATGTTTCCGCATTGGTTGAGTTGAACATTGCTAAGAATTTGGTTGGATCTGCAA
TGGCTGGGTCTGTTGGTGGATTTAACGCACATGCAGCTAATTTAGTGACAGCTGTTTTCTTGGCATTAGGACAAG
ATCCTGCACAAAATGTTGAAAGTTCCAACGTATAACATTGATGAAAGAGTGGACGGTGATTTGAGAAATTTCCG
TATCCATGCCATCCATCGAAGTAGGTACCATCGGTGGTGGTACTGTTCTAGAACCACAAGGTGCCATGTTGGACT
TATTAGGTGTAAGAGGCCCGCATGCTACCGCTCCTGGTACCAACGCACGTCATTTAGCAAGAATAGTTGCCTGTG
CCGTCTTGGCAGGTGAATTATCCTTATGTGCTGCCCTAGCAGCCGGCCATTTGGTTCAAAGTCATATGACCCACA
ACAGGAAACCTGCTGAACCAACAAAACCTAACAAATTTGGACGCCACTGATATAAATCGTTTGAAGATGGGTCCG
TCACC

HMGR truncado



TGCATTAAATCCTAAGTCGACCT (SEQ ID NO:63)

FIG. 36A

Sequência de plasmídeo pMB15

ACCTTCGGGAGCGCCTGAAGCCCGTTCTGGACGCCCTGGGGCCGTTGAATCGGGATATGCAGGCCAAGGCCGCCG
CGATCATCAAGGCCGTGGGCGAAAAGCTGCTGACGGAACAGCGGGAAGTCCAGCGCCAGAAAACAGGCCCGAGCGCC
AGCAGGAACGGCGGCGCGCACATTTCCCCGAAAAGTGCCACCTGACGCTAAGAAACATTATTATCATGACATT
AACCTATAAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCTTATCATCGATA
AGCTTTAATGCGGTAGTTTA

→ gene de resistência de tetraciclina

TCACAGTTAAATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCACTCCTCGGCACC
GTCACCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCTCTTGGCGGATATCGTCCATTCC
GACAGCATGCCAGTCAGTATGGCGTGTCTAGCGCTATATGCGTTGATGCAATTTCTATGCCGACCCGTTCTC
GGAGCACTGTCCGACCGCTTTGGCCGCGCCAGTCTGCTCGCTTCTGCTACTTGGAGCCACTATCGACTACGCG
ATCATGGCGACCACACCCGTCCTGTGGATCCTCTACGCGGACGCATCGTGGCCGGCATCACCGCGCCACAGGT
GCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCT
TGTTTCGGCGTGGGTATGGTGGCAGGCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCAFTC
CTTGGCGCGCGGTGCTCAACCGCCTCAACCTACTACTGGGCTGCTTCTTAATGCAGGAGTCGCATAAGGGAGAG
CGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCGTACTATCGTCGCC
GCATTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCAATTTTCGGCGAG
GACCGCTTTTCGCTGGAGCGCGACGATGATCGGCCTGTGCTTTCGGGTATTTCGGAATCTTGCACGCCCTCGCTCAA
GCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCG
CTGGGCTACGTTCTGTCTGGCTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGC
GGCATCGGGATGCCCGCTTCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGA
TCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGATTATGCGCGCTCGGGC
AGCACATGGAACGGGTTGGCATGGATTGTAGGCGCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTTCGGGT
GCATGGAGCCCGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTACCACTCCAAGAATTG
GAGCCAATCAATTCTTGGGAGAACTGTGAATGCGCAAATGCGCCCAATACGCAAAACCGCCTCTCCCCGCGCGTT
GGCCGATTCA

→ promotor Plac Modificado

TTAATGCAGCTGGCAGCAGAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCT
CACTCATTAGGCACCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACA
ATTTACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCTCACTAAAGGGAACAAAAGCT
GGGTACCGGGCCCC

→ MK

CCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGTAGGAGGAATTAACCATGTCATTACCGTT
CTTAACCTCTGCACCGGAAAGGTTATTATTTTGGTGAACACTCTGCTGTGTACAACAAGCCTGCCGTGCTGCTG
TAGTGTGCTGCGTTGAGAACCTACCTGCTAATAAGCGAGTCATCTGCACCAGATACTATTGAATTGGACTTCCC
GGACATTAGCTTTAATCATAAGTGGTCCATCAATGATTTCAATGCCATCACCAGGATCAAGTAAACTCCAAAA
ATTGGCCAAGGCTCAACAAGCCACCGATGGCTTGTCTCAGGAACCTCGTTAGTCTTTTGGATCCGTTGTTAGCTCA
ACTATCCGAATCCTTCCACTACCATGCAGCGTTTTGTTTCTGTATATGTTGTTTGGCTATGCCCCCATGCCAA
GAATATTAAGTTTTCTTTAAAGTCTACTTTACCCATCGGTGCTGGGTTGGGCTCAAGCGCCTCTATTTCTGTATC
ACTGGCCTTAGCTATGGCCTACTTGGGGGGTTAATAGGATCTAATGACTTGGAAAAGCTGTGAGAAAACGATAA
GCATATAGTGAATCAATGGGCCTTCATAGGTGAAAAGTGATTACCGGTACCCCTTCAGGAATAGATAACGCTGT
GGCCACTTATGGTAATGCCCTGCTATTTGA: AAAGACTCACATAATGGAACAATAACACAAACAATTTTAAGTT
CTTAGATGATTTCCAGCCATTCCAATGATCTAACCCTATACTAGAATTCGAAGGTCTACAAAAGATCTTGTGTG
TCCGCTTCGTGTGTTGGTCACCGAGAAATTTCTGAAGTTATGAAGCCAATTCAGATGCCATGGGTGAATGTGC
CCTACAAGGCTTAGAGATCATGACTAAGTTAAGTAAATGTAAGGCACCGATGACGAGGCTGTAGAAACTAATAA
TGAAGTGTATGAACAATATTGGAATTGATAAGAATAAATCATGGACTGCTTGTCTCAATCGGTGTTTCTCATCC
TGGATTAGAACTTATTAATAATCTGACCGATGATTTGAGAATTGGCTCCACAAAACCTACCGGTGCTGGTGGCGG
CGGTTGCTCTTTGACTTTGTTACGAAGAGACATTACTCAAGAGCAAATGACAGCTTCAAAAAGAAATTGCAAGA
TGATTTTAGTTACGAGACATTTGAAACAGACTTGGGTGGGACTGGCTGCTGTTTGTTAAGCGCAAAAATTTGAA
TAAAGATCTTAAATCAATCCCTAGTATTCCAATTATTTGAAAATAAACTACCACAAAGCAACAAATTGACGA
TCTATTATTG

FIG. 36B

MK ← | → PMK

CCAGGAAACACGAATTTACCATGGACTTCATAGGAGGCAGATCAAATGTCAGAGTTGAGAGCCCTTCAGTGCCCCA
 GGGAAAGCGTTACTAGCTGGTGGATATTTAGTTTGTAGATACAAAATATGAAGCATTGTTAGTCGGATTATCGGCA
 AGAATGCATGCTGTAGCCCATCCTTACGGTTCATTGCAAGGGTCTGATAAGTTTGAAGTGCGTGTGAAAAGTAAA
 CAATTTAAAGATGGGGAGTGGCTGTACCATATAAGTCCATAAAGTGGCTTCATTCTGTTTCGATAGGCGGATCT
 AAGAACCCTTTTCATTGAAAAGTTATCGCTAACGTATTTAGCTACTTTAAACCTAACATGGACGACTACTGCAAT
 AGAACTTGTTCGTTATTGATATTTCTCTGATGATGCCTACCATTCTCAGGAGGATAGCGTTACCGAACATCGT
 GGCAACAGAAAGATTGAGTTTTTCATTGCGACAGAAATTGAAGAAGTCCCAAAACAGGGCTGGGCTCCTCGGCAGGT
 TTAGTCACAGTTTTAACTACAGCTTTGGCCTCCTTTTTTGTATCGGACCTGGAAAATAATGTAGACAAATATAGA
 GAAGTTATTATAATTTAGCACAAAGTTGCTCATGTCAAGCTCAGGGTAAAATTGGAAGCGGGTTTGATGTAGCG
 GCGGCAGCATATGGATCTATCAGATATAGAAGATTCCCACCCGCATTAATCTCTAATTTGCCAGATATTGGAAGT
 GCTACTTACGGCAGTAACTGGCGCATTGTTGATGAAGAAGACTGGAATATTACGATTAAAAGTAACCATTTA
 CCTTCGGGATTAACCTTATGGATGGCGGATATTAAGAATGGTTCAAGAACAGTAAACCTGGTCCAGAAGGTAAAA
 AATTGGTATGATTCGCATATGCCAGAAAGCTTGAATAATATACAGAACTCGATCATGCAAAATCTAGATTTATG
 GATGGACTATCTAACTAGATCGCTTACACGAGACTCATGACGATTACAGCGATCAGATATTGAGTCTCTTGAG
 AGGAATGACTGTACCTGTCAAAAGTATCCTGAAATCACAGAACTTAGAGATGCAGTTGCCACAATTAGACGTTCC
 TTAGAAAAATACTAAAGAATCTGGTGCCGATATCGAACCTCCCGTACAACTAGCTTATTGGATGATTGCCAG
 ACCTTAAAGGAGTTCTTACTTGCTTAATACCTGGTGCTGGTGGTTATGACGCCATTGCGAGTGAATTACTAAGCAA
 GATGTTGATCTTAGGGCTCAAACCGCTAATGACAAAAGATTTCTAAGGTTCAATGGCTGGATGTAATCAGGCT
 GACTGGGGTG

PMK ← | → MPD

TTAGGAAAGAAAAAGATCCGGAACTTATCTTGATAAATAGGAGGTAAATACTCATGACCGTTTACACAGCATCCG
 TTACCGCACCCGTCACATCGCAACCTTAAGTATTGGGGGAAAAAGGACACGAAGTTGAATCTGCCACCAATT
 CGTCCATATCAGTGACTTTATCGCAAGATGACCTCAGAAGCTTGACCTCTGCGGCTACTGCACCTGAGTTTGAAC
 GCGACACTTTGTGGTTAAATGAGAACCACACAGCATCGACAATGAAAGAACTCAAAATGTCTGCGCGACCTAC
 GCCAATTAAGAAAGGAAATGGAATCGAAGGACGCCTCATTTGCCACATTATCTCAATGGAAGCTCCACATTGTCT
 CCGAAAATACTTTCTACAGCAGCTGGTTAGCTTCTCCGCTGCTGGCTTTGCTGCATTGGTCTCTGCAATTG
 CTAAGTTATACCAATTACCACAGTCACTTCAGAAATATCTAGAATAGCAAGAAAGGGGTCTGGTTGAGCTTGTG
 GATCGTTGTTTGGCGGATACGTGGCCTGGGAAATGGGAAAGCTGAAGATGGTCAATGATTCCATGGCAGTACAAA
 TCGCAGACAGCTCTGACTGGCTCAGATGAAAGCTTGTGTCTAGTTGTGACGGATATTAAAAAGGATGTGAGTT
 CCACTCAGGGTATGCAATTGACCGTGGCAACCTCCGAATATTAAAGAAAGAAATGAACATGTCGTACCAAGA
 GATTTGAAGTCATGCGTAAAGCCATTGTTGAAAAGATTTCGCCACCTTTGCAAGGAAACAATGATGGATTCCA
 ACTCTTTCCATGCCACATGTTTGGACTCTTTCCCTCCATATTCTACATGAATGACACTTCCAAGCGTATCATCA
 GTTGGTGCCACACCATTAATCAGTTTTACGGAGAAACAATCGTTGCATACACGTTTGTATGCAGGTCCAAATGCTG
 TGTGTACTACTTAGCTGAAAATGAGTCGAAACTCTTTGCATTATCTATAAATGTTTGGCTCTGTTCTCTGGAT
 GGGACAAGAAATTTACTACTGAGCAGCTTGAGGCTTTCAACCATCAATTTGAATCATCTAATTTACTGACGCTG
 AATTGGATCTTGAGTTGCAAAAGGATGTTGCCAGAGTGATTTTAACTCAAGTCGTTTCAAGGCCACAGAAACAA

MPD ← | → idi

ACGAATCTTTGATTGACGCAAAGACTGGTCTACCPAAGGAATTAATGCAGCCCGGGAGGAGGATTACTATATGCA
 AACGGAAACACGTCCTTTTATTGAATGCACAGGGAGTTCCACGGGTACGCTGGAAAAGTATGCCGCACACACGGC
 AGACACCCGCTTACATCTCGCGTTCTCCAGTTGGCTGTTTAAATGCCAAAGGACAAATATTAGTTACCCGCCGCGC
 ACTGAGCAAAAAAGCATGGCCTGGCGTGTGGACTAATCTCGGTTTGTGGGCACCCACAATGGGAGAAAGCAACGA
 AGACCGAGTGATCCGCCGTTGCCGTTATGAGCTTGGCGTGGAAATTACGCCTCTGAAATCTATCTATCTGACTT
 TCGCTACCGCGCCACCGATCCGAGTGGCATTGTGGAAAATGAAGTGTGTCCGGTATTTGCCGCACGCACCACTAG
 TCGGTTACAGATCAATGATGATGAAGTATGATTATCAATGGTGTGATTTAGCAGATGTATTACACGGTATTGA
 TGCCACCCCGTGGGCGTTAGTCCGTTGGATGGTGTGATGCAGGCGACAAATCGCGAAGCCAGAAAACGATTATCTGC
 ATTTACCCAGCTTAAATAACCCGGGGGATCCACTAGTCTCT

FIG. 36C

ispA

AGAGCGGCCGCCACCGCGGAGGAGGAATGAGTAATGGACTTTCCGCAGCAACTCGAAGCCTGCGTTAAGCAGGCC
 AACCAGGCGCTGAGCCGTTTTATCGCCCCACTGCCCTTTCAGAACAATCCCGTGGTCGAAACCATGCAGTATGGC
 GCATTATTAGGTGGTAAGCGCCTGCGACCTTTCCTGGTTTATGCCACCGGTATATGTTCCGCGTTAGCACAAAC
 ACGCTGGACGCACCCGCTGCCCGCGTTGAGTGATCCACGCTTACTCATTAAATTCATGATGATTTACCGGCAATG
 GATGATGACGATCTGCGTCGCGGTTTGCCAACCTGCCATGTGAAGTTTGGCGAAGCAACCGGATTCTCGCTGGC
 GACGCTTACAAACGCTGGCGTTCTCGATTTTAAGCGATGCCGATATGCCGGAAGTGTGCGACCGCGACAGAATT
 TCGATGATTTCTGAACCTGGCGAGCGCAGTGGTATTGCCGGAATGTGCGGTGGTCAGGCATTAGATTTAGACGCG
 GAAGGCAAAACAGTACCTCTGGACGCGCTTGAGCGTATTATCGTCAATAAACCGCGCATTGATTTCGCGCCGCC
 GTTCGCTTGGTGCAATTAAGCGCGGAGATAAAGGACGTCGTGCTCGCCGGTACTCGACAAGTATGCAGAGAGC
 ATCGGCTTGCCTTCCAGGTTCCAGGATGACATCTGGATGTGGTGGGAGATACTGCAACGTTGGGAAAACGCCAG
 GTGCGGACCGCAACTTGGTAAAGTACCTACCTGCACTTCTGGGTCTTGAGCAAGCCCGGAAGAAAGCCCGG
 GATCTGATCGACGATGCCCGTCAGTCGCTGAAACAACCTGGCTGAACAGTCACTCGATACCTCGGCACTGGAAGCG
 CTAGCGGACTACATCATCCAGCGTAATAATAA

GAGCTCCAATTGCCCCATAGTGAGTCGATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGA
 AAACCTGGCGTTACCCAACCTAATCGCCTTGCGACACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGC
 CCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGAAATTGTAACCGTTAATATTTTGTTA
 AAATTGCGCTTAATATTTTGTAAATCAGCTCATTTTTTAACCAATAGGCCGACTGCGATGAGTGGCAGGGCGGG
 GCGTAATTTTTTAAAGCAGTTATTGGTGCCCTTAAACGCTGGTGCTACGCTGAATAAGTGATAATAAGCGGA
 TGAATGGCAGAAATCGAAAGCAAATTCGACCCGGTCGTGGTTTCAGGGCAGGGTCGTAAATAGCCGCTTATGT
 CTATTGCTGGTTTACCGGTTTATTGACTACCGAAGCAGTGTGACCGTGTGCTTCTCAATGCTGAGGCCAGTT
 TGCTCAGGCTCTCCCGTGGAGGTAATAATTGACGATATGATCATTTATTCTGCCTCCAGAGCCTGATAAAAC
 GGTGAATCCGTTAGCGAGGTGCCCGCGCTTCCATTGAGGTGAGGTGGCCCGGCTCCATGCACCGCGACGCAAC
 CGCGGGAGGCAGACAAGGTATAGGCGGCGAGCGGCTACAGCCGATAGTCTGGAACACGCGCACTTACGGGT

Origem de réplica

TGCTGCGCAACCAAGTGCTACCGCGCGGCGAGCGTGACCCGTGTCGGCGGGCTCCAACGGCTCGCCATCGTCCAG
 AAAACACGGCTCATCGGGCATCGGCAGGCGCTGCTGCCCGCGCGCTTCCCATTCCTCCGTTTCGGTCAAGGCTGG
 CAGGTCTGGTTCCATGCCCCGAATGCCGGGCTGGCTGGGCGGCTCCTCGCCGGGGCCGGTCGGTAGTTGCTGCTC
 GCCCGGATCAGGGTCGGGATCGGGCGAGGTGCGCCATGCCCAACAGCGATTGCTTCCTGGTTCGTGATGCAAC
 CACCACGGCGGCACTGAACACCGACAGGCGCAACTGGTTCGGGGGCTGGCCCCACGCCACGCGGTATTGACCAC
 GTAGGCCGACACGGTGCCGGGCGGCTGAGCTTCAGCAGCGAGATCCAGCGCTCGGCCACCAAGTCTTGACTGC
 GTATTGGACCGTCCGCAAGACGTCGATGAGCTTGGAAAGTGCTTCTGCGTGAACACCACGGCGTTCTGGTG
 GCCCATCTGCGCCAGAGGTGATGACGAGCATTGCCGCGCTGGGTTTCTCGCAATAGCCCGGCCACGCCTC
 ATGCGCTTTGCGTTCCGTTTGCAACCCAGTGACCGGGCTTGTCTTGGCTTGAATGCCGATTCTCTGGAAGTGGT
 GGCCATGCTTATCTCCATGCGGTAGGGTGCCGACGGTTGCGGCACCATGCGCAATCAGCTGCAACTTTTCGGCA
 GCGCGACAACAATTATGCGTTGCGTAAAGTGCGAGTCAATTACAGATTTTCTTTAACCTACGCAATGAGCTATT
 GCGGGGGGTGCCGAATGAGCTGTTGCGTACCCCCCTTTTAAAGTTGTTGATTTTAAAGTCTTTTCGATTTTCGC
 CCTATATCTAGTTCTTTGGTGCCAAAGAAGGACCCCCGCGGGGTCCCCACGCCCTTCGGCGCGGCTCCCCC
 TCCGGCAAAAGTGCCCCCTCCGGGGCTTGTGATCGACTGCGCGGCTTCGGCCTTGCCCAAGGTGGCGCTGCC
 CCTTGGAACCCCCGCACTCGCCGCGTGAGGCTCGGGGGCAGGCGGGCGGGCTTCGCCTTCGACTGCCCCAC
 TCGCATAGGCTTGGGTGCTTCCAGGCGCGTCAAGGCCAAGCCGCTGCGCGGTGCTGCGCGAGCCTTGACCCGCC
 TTCCACTTGGTGTCCKACCGCAAGCGAAGCGCGCAGGCCGCGAGGCCGAGGCTTTTCCCCAGAGAAAATTAAAA
 AAATTGATGGGGCAAGGCCGAGGCCGCGCAGTTGGAGCCGGTGGGTATGTGGTCAAGGCTGGGTAGCCGGTGG
 GCAATCCCTGTGGTCAAGCTCGTGGGACGGCAGCCTGTCATCAGCTTGTCCAGCAGGGTTGTCCACGGGCGG
 AGCGAAGCGCAGCCAGCGGTGGCGCTCGCGGCATCGTCCACATATCCACGGGCTGGCAAGGGAGCGCAGCGAC
 CGCGCAGGGCGAAGCCCGAGAGCAAGCCGTAGGGCGCGCAGCCGCGTAGGCGGTACGACTTTGCGAAGCA
 AAGTCTAGTGAGTATACTCAAGCATTGAGTGGCCCGCGGAGGACCGCCTTGGCTGCCCCCGTCGAGCCGGTT
 GGACACCAAAAGGGAGGGGAGGCGATGGCGGCATACGCGATCATGCGATGCAAGAAGCTGGCGAAAATGGGCAAC
 GTGGCGGCCAGTCTCAAGCACGCTACCGCGAGCGCGAGACGCCAACGCTGACGCCAGCAGGACGCCAGAGAAC
 GAGCACTGGCGGCCAGCAGCACCGATGAAGCGATGGGCCGACTGCGCGAGTTGCTGCCAGAGAAGCGGCGCAAG
 GACGCTGTGTTGGCGGTGAGTACGTCATGACGGCCAGCCCGGAATGGTGGAAGTCGGCCAGCCAAGAACAGCAG
 GCGGCTTCTTCGAGAAGGCGCACAAGTGCTGGCGGACAAGTACGGGGCGGATCGCATCGTGACGGCCAGCATC
 CACCGTGACGAAACAGCCCGCACATGACCGGCTTCTGGTGCCGCTGACGCAGGACGGCAGGCTGTGGGCCAAG

FIG. 36D

GAGTTCATCGGCAACAAAGCGCAGATGACCCGCGACCAGACCACGTTTGCGGCCGCTGTGGCCGATCTAGGGCTG
CAACGGGGCATCGAGGGCAGCAAGGCACGTCACACGCGCATTTCAGGCGTTCTACGAGGCCCTGGAGCGGGCCACCA
GTGGGCCACGTCACCATCAGCCCGCAAGCGGTGAGCCACGCGCCTATGCACCGCAGGGATTGGCCGAAAAGCTG
GGAATCTCAAAGCGCGTTGAGACGCCGGAAGCCGTGGCCGACCGGCTGACAAAAGCGGTTCCGGCAGGGGTATGAG
CCTGCCCTACAGGCCGCCGAGGAGCGCGTGAGATGCGCAAGAAGGCCGATCAAGCCCAAGAGACGGCCCCGAG
(SEQ ID NO:62)

RESUMO

Patente de Invenção: **"ÁCIDOS NUCLEÍCOS QUE CODIFICAM ENZIMA MODIFICADA DE CITOCROMO P450 E MÉTODOS DE USO DOS MESMOS"**.

- 5 A presente invenção proporciona ácidos nucleicos compreendendo seqüências de nucleotídeo que codificam enzima modificada de citocromo P450s; bem como vetores recombinantes e células hospedeiras compreendendo os ácidos nucleicos. A presente invenção proporciona adicionalmente métodos de produção de um composto funcionalizado em uma
- 10 célula hospedeira geneticamente modificada com um ácido nucleico compreendendo seqüências de nucleotídeo que codificam uma enzima modificada de citocromo P450.