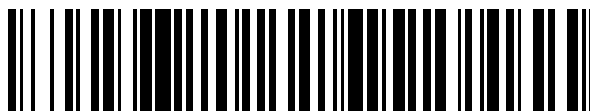


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 531 290**

51 Int. Cl.:

C12N 9/10 (2006.01)
C12N 9/02 (2006.01)
C12Q 1/48 (2006.01)
C12Q 1/26 (2006.01)
C07K 14/38 (2006.01)
C07K 16/40 (2006.01)
C07K 9/00 (2006.01)
A61K 38/44 (2006.01)
A61K 38/45 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **31.12.2008 E 08869698 (4)**

97 Fecha y número de publicación de la concesión europea: **26.11.2014 EP 2238242**

54 Título: **Transferasas y oxidorreductasas, ácidos nucleicos que las codifican y métodos para prepararlas y usarlas**

30 Prioridad:

03.01.2008 US 18868 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
12.03.2015

73 Titular/es:

**BASF ENZYMES LLC (100.0%)
3550 John Hopkins Court
San Diego, CA 92121, US**

72 Inventor/es:

**WEINER, DAVID;
LUGINBUHL, PETER;
BUENO, ANALIA;
CUENCA, JOSLIN y
MARASCO, ERIN**

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

ES 2 531 290 T3

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

DESCRIPCIÓN

Transferasas y oxidorreductasas, ácidos nucleicos que las codifican y métodos para prepararlas y usarlas.

Campo de la invención

5 La invención se relaciona en general con enzimas, polinucleótidos que codifican las enzimas, el uso de tales polinucleótidos y polipéptidos y más específicamente con enzimas que tienen actividad de d-aminoácido transferasa codificados por un ácido nucleico que comprende una secuencia que tiene al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia con la secuencia de ácido nucleico de SEQ ID NO: 219. Así, la invención provee enzimas, composiciones y/o métodos para la producción de péptidos, enzimas, alimentos y aditivos para alimentos, bebidas, y aditivos para bebidas, piensos y aditivos para piensos y suplementos dietéticos que comprenden los polipéptidos o polinucleótidos de acuerdo con la invención.

Antecedentes

15 Las transferasas y/o oxidorreductasas catalizan la transferencia de un grupo químico, catalizan la transaminación, catalizan la reacción: D-alanina + 2-oxoglutarato \rightleftharpoons piruvato + D-glutamato, y/o catalizan una reacción de oxidación-reducción, catalizan la remoción de átomos de hidrógeno y/o catalizan la reacción: D-aminoácido + H₂O + aceptor \rightleftharpoons un 2-oxoácido + NH₃ + aceptor reducido. Las transferasas, por ejemplo, las transaminasas, por ejemplo las d-aminoácidos transferasas (también denominadas como "d-aminotransferasas" o "D-ATs"), y/o las oxidorreductasas, por ejemplo, deshidrogenasas por ejemplo, d-aminoácido deshidrogenasas son de valor comercial considerable, siendo utilizadas en la industria farmacéutica, en las industrias de alimentos, piensos y bebidas, por ejemplo, para la producción de endulzantes, en la industria de madera/papel y en la industria de los combustibles. El documento EP 1590 268 A1 divulga una D-aminotransferasa modificada para producir (2R, 4R)-monatina que tiene alta intensidad de dulzor a partir de ácido 4(indol-3-ilmetil)-4-hidroxi-2-oxoglutarico. La entrada en la base de datos A6LX33 divulga una aminotransferasa clase IV de Clostridium beijerinckii que tiene 282 aminoácidos.

Resumen de la invención

25 Esta invención provee enzimas que tienen actividad de aminoácido transferasa codificadas por un ácido nucleico que comprende una secuencia que contiene al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia con la secuencia de ácido nucleico de SEQ ID NO: 219. La invención provee adicionalmente enzimas que tienen actividad de d-aminoácido transferasa codificadas por un ácido nucleico que comprende una secuencia que tiene al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia con la secuencia de ácidos nucleicos de SEQ ID NO: 219 y ácidos nucleicos que las codifican, vectores y células que las comprenden, sondas para amplificar e identificar estos ácidos nucleicos que codifican transferasa, y métodos para hacer y optimizar estos polipéptidos y péptidos.

30 La invención provee enzimas, composiciones y/o métodos para la producción de enzimas peptídicas, alimentos y aditivos para alimentos, bebidas, y aditivos para bebidas, piensos y aditivos para piensos, suplementos dietéticos, que comprenden los polipéptidos o polinucleótidos de acuerdo con la invención. Estas composiciones pueden ser formuladas en una variedad de formas tales como tabletas, geles, píldoras, implantes, líquidos, aspersiones, películas, micelas, polvos, pellas de alimentos y piensos o como cualquier tipo de forma encapsulada.

En algunas realizaciones, las d-aminoácido transferasas y/o composiciones de las mismas pueden ser útiles en contextos farmacéuticos, industriales y/o agrícolas.

40 En algunas realizaciones la d-aminoácido transferasas y/o composiciones de las mismas pueden ser utilizadas para catalizar una reacción entre un aminoácido y un α -cetoácido. En algunas realizaciones, las d-aminoácido transferasas y/o composiciones de las mismas pueden ser útiles para catalizar una reacción de transaminación. En algunas realizaciones, las d-aminoácido transferasas y/o composiciones de las mismas pueden ser útiles para catalizar una reacción que retira el grupo amino del aminoácido dejando un α -cetoácido, y transfiriendo el grupo amino a un α -cetoácido reactivo que lo convierte en un aminoácido. En realizaciones alternativas, las d-aminoácido transferasas y/o composiciones de las mismas pueden ser útiles en la producción de aminoácidos.

En algunas realizaciones se proveen d-aminoácido transferasas que facilitan la producción de indol-3-piruvato. En algunas realizaciones se proveen d-aminoácido transferasas que facilitan la producción de RR-Monatina.

50 En realizaciones alternativas, la invención provee polipéptidos (y los ácidos nucleicos que los codifican) que tienen actividad de d-aminoácido transferasa que comprende una secuencia de aminoácidos que tiene un 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 220, o una secuencia de aminoácidos como se establece en SEQ ID NO: 220 con una, varias o todas las modificaciones de aminoácidos de la Tabla 46, o una secuencia de aminoácidos como se establece en SEQ ID NO: 220 que tiene al menos una de la combinación de modificaciones de aminoácidos como se establecen en la Tabla 55, o codificadas por un ácido nucleico que comprende una secuencia que tiene al menos 90%, 91%, 92%,

93%,94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia a la secuencia de ácidos nucleicos de SEQ ID NO: 219. En realizaciones alternativas, la invención provee polipéptidos (y los ácidos nucleicos que los codifican) que tienen una actividad de d-aminoácido transferasa como se definió anteriormente pero que carece de una secuencia de señalización, un preprodominio, un dominio de enlazamiento; y en un aspecto, consiste de un dominio de enlazamiento de NAD, NAD(P), calcio, tiamina, FAD, zinc, ADN y/o a lipoilo.

En realizaciones alternativas, la invención provee polipéptidos (y los ácidos nucleicos que los codifican) que tienen una actividad de d-aminoácido transferasa que comprende adicionalmente una secuencia heteróloga; y en un aspecto, la secuencia heteróloga comprende, o consiste de una secuencia que codifica: (i) una secuencia de señalización heteróloga, un dominio heterólogo, un dominio de enlazamiento heterólogo, un dominio de ordenamiento heterólogo, un dominio catalítico heterólogo (CD), o una combinación de los mismos; (ii) la secuencia de (i), en donde la secuencia de señalización heteróloga, el dominio de enlazamiento o el dominio catalítico (CD) son derivados de una enzima heteróloga; o (iii) una etiqueta, un epítipo, un péptido de direccionamiento, una secuencia escindible, una unidad estructural detectable o una enzima; y en un aspecto, el dominio de enlazamiento heterólogo comprende, o consiste de, un dominio del enlazamiento de NAD, NAD(P), calcio, tiamina, FAD, zinc, ADN y/o un lipoilo; y en un aspecto, la secuencia de señal heteróloga direcciona la proteína codificada a una vacuola, al retículo endoplasmático, un cloroplasto o un gránulo de almidón.

La invención provee ácidos nucleicos aislados, sintéticos o recombinantes que comprenden

(a) un ácido nucleico (polinucleótido) que codifica al menos un polipéptido, en donde el ácido nucleico comprende una secuencia que tiene al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia a la secuencia de ácido nucleico (polinucleótido) de SEQ ID NO:219 en donde el ácido nucleico codifica al menos un polipéptido que tiene una actividad de d-aminoácido transferasa (b) el ácido nucleico (polinucleótido) de (a), en donde las identidades de secuencia son determinadas: (A) por análisis con un algoritmo de comparación de secuencias o por inspección visual, o (B) sobre una región de al menos 20, 30, 40, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150 o más residuos, o sobre la longitud completa del ADNc, (ARNm) transcrito o gen; (c) el ácido nucleico (polinucleótido) de (a) o (b), en donde el algoritmo de comparación de secuencias es un algoritmo BLAST versión 2.2.2 en donde se fijan unos parámetros de filtración a blastall -p blastp -d "nr pataa" -F F, y todas las demás opciones son fijadas por sistema.

La invención describe adicionalmente un ácido nucleico (polinucleótido) que codifica al menos un polipéptido o péptido, en donde el ácido nucleico comprende una secuencia que hibrida bajo condiciones restrictivas a un ácido nucleico que comprende la secuencia de SEQ ID NO: 219 y las condiciones restrictivas comprenden una etapa de lavado que comprende un lavado de 0.2X SSC a una temperatura de aproximadamente 65°C durante aproximadamente 15 minutos;

También se proveen (e) el ácido nucleico (polinucleótido) de cualquiera de (a) a (d) que tienen la longitud completa de un gen o un transcrito;

(f) un ácido nucleico (polinucleótido) que codifica al menos un polipéptido que tiene una actividad de d-aminoácido transferasa, en donde el polipéptido comprende la secuencia de SEQ ID NO:220, SEQ ID NO:220 con una, varias o todas las modificaciones de la tabla 46 o la tabla 55;

La invención describe adicionalmente (g) el ácido nucleico (polinucleótido) de cualquiera de (a) a (f) y la codificación de un polipéptido que tiene al menos una sustitución de aminoácido conservadora y retiene su actividad de d-aminoácido transferasa, en donde la al menos una sustitución de aminoácido conservadora comprende sustituir un aminoácido con otro aminoácido de características similares; o, una sustitución conservadora comprende: remplazo de un aminoácido alifático por otro aminoácido alifático; remplazo de una serina con una treonina o viceversa; remplazo de un residuo ácido con otro residuo ácido; remplazo de un residuo que porta un grupo amida con otro residuo que porta un grupo amida; intercambio de un residuo básico con otro residuo básico; o remplazo de un residuo aromático con otro residuo aromático.

También se proveen (h) el ácido nucleico (polinucleótido) de cualquiera de (a) a (g) que codifica un polipéptido que tiene una actividad de d-aminoácido transferasa que carece de una secuencia de señalización, un preprodominio, un dominio de enlazamiento.

(i) el ácido nucleico (polinucleótido) de (h), en donde el dominio de enlazamiento comprende, o consiste de, un dominio de enlazamiento de NAD, NAD(P), calcio, tiamina, FAD, zinc, ADN y/o lipoilo;

(j) el ácido nucleico (polinucleótido) de cualquiera de (a) a (i) que codifica un polipéptido que tiene una actividad de d-aminoácido transferasa que comprende adicionalmente una secuencia heteróloga;

(k) el ácido nucleico (polinucleótido) de (j), en donde la secuencia heteróloga comprende, o consiste de una secuencia que codifica: (a) una secuencia de señalización heteróloga, un dominio heterólogo, un dominio de enlazamiento heterólogo, un dominio de ordenamiento heterólogo, un dominio catalítico heterólogo (CD), o una

combinación de los mismos; (B) la secuencia de (1) en donde la secuencia de señalización heteróloga, el dominio de enlazamiento o el dominio catalítico (CD) son derivados de una enzima heteróloga; o (C), una etiqueta, un epítipo, un polipéptido de direccionamiento, una secuencia escindible, una unidad estructural detectable o una enzima;

5 (I) el ácido nucleico (polinucleótido) de (k), en donde el dominio de enlazamiento heterólogo comprende, o consiste de, un dominio de enlazamiento de NAD, NAD(P), calcio, tiamina, FAD, zinc, ADN y/o lipoilo; y

(m) el ácido nucleico (polinucleótido) de (1), en donde la secuencia de señalización heteróloga direcciona la proteína codificada a una vacuola, el retículo endoplasmático, un cloroplasto, o un granulo de almidón.

También se provee (n) una secuencia de ácido nucleico (polinucleótido) del todo (completamente) complementaria a la secuencia de (a) a (m).

10 La invención provee ácidos nucleicos aislados, sintéticos o recombinantes que comprenden un ácido nucleico que codifica al menos un polipéptido que tiene una actividad de d-aminoácido transferasa en donde el polipéptido tiene una secuencia como la establecida en SEQ ID NO:220, SEQ ID NO:220, con una, varias o todas las modificaciones de la Tabla 46 o la Tabla 55 y variantes de las mismas (abarcando todas estas secuencias de polipéptidos y secuencias de péptidos de la invención).

15 La invención provee ácidos nucleicos aislados, sintéticos o recombinantes que comprenden secuencias completamente complementarias a todas estas secuencias de ácidos nucleicos de la invención (secuencias complementarias (no codificantes) y codificantes también denominadas de aquí en adelante colectivamente como secuencias de ácidos nucleicos de la invención).

20 En un aspecto, la identidad de secuencia es al menos identidad de secuencia (homología) de 90 % , 91 % , 92 % , 93 % , 94 % , 95 % , 96 % , 97 % , 98 % , 99 % , o 100 % (completa). En un aspecto, la identidad de secuencia está sobre la longitud completa de un gen o un transcripto. Por ejemplo, la invención provee ácidos nucleicos aislados, sintéticos o recombinantes que comprenden una secuencia de ácidos nucleicos de SEQ ID NO:219, por ejemplo, como se describe en las tablas 1,2 y 3 en el Listado de Secuencias.

25 La invención provee ácidos nucleicos aislados, sintéticos o recombinantes que codifican un polipéptido que tiene una actividad de d-aminoácido transferasa, en donde el ácido nucleico tiene al menos una modificación de secuencia de una secuencia de ejemplo de la invención, o, cualquier secuencia de la invención.

30 En un aspecto (opcionalmente), los ácidos nucleicos aislados, sintéticos o recombinantes de la invención tienen una actividad de d-aminoácido transferasa en donde la actividad comprende catalizar la transferencia de un grupo químico, catalizar la transaminación, catalizar la reacción: D-alanina + 2-oxoglutarato \rightleftharpoons piruvato + D-glutamato, y/o catalizar una reacción de oxidación-reducción, catalizar la remoción de átomos de y/o catalizar la reacción: D-aminoácido + H₂O +aceptor \rightleftharpoons un 2-oxo ácido + NH₃ + aceptor reducido.

35 En un aspecto, la actividad de la d-aminoácido transferasa es termoestable, por ejemplo, cuando el polipéptido retiene la actividad de aminoácido transferasa bajo condiciones que comprenden un rango de temperatura desde aproximadamente -100°C hasta aproximadamente -80°C ,aproximadamente -80°C hasta aproximadamente -40°C, aproximadamente - 40°C hasta aproximadamente -20°C, aproximadamente -20°C hasta aproximadamente 0°C, aproximadamente 0°C hasta aproximadamente 5°C, aproximadamente 5°C hasta aproximadamente 15°C, aproximadamente 15°C hasta aproximadamente 25°C, aproximadamente 25°C hasta aproximadamente 37°C, aproximadamente 37°C hasta aproximadamente 45°C, aproximadamente 45°C hasta aproximadamente 55°C, aproximadamente 55°C hasta aproximadamente 70°C, aproximadamente 70°C hasta aproximadamente 75°C, aproximadamente 75°C hasta aproximadamente 85°C, aproximadamente 85°C hasta aproximadamente 90°C, aproximadamente 90°C hasta aproximadamente 95°C, aproximadamente 95°C hasta aproximadamente 100°C, aproximadamente 100°C hasta aproximadamente 105°C, aproximadamente 105°C hasta aproximadamente 110°C, aproximadamente 110°C hasta aproximadamente 120°C, o 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 104°C, 105°C, 106°C, 107°C, 108°C, 109°C, 110°C, 111°C, 112°C, 113°C, 114°C, 115°C o más. En algunas realizaciones, los polipéptidos termoestables de acuerdo con la invención retienen la actividad de la d-aminoácido transferasa a una temperatura en los rangos descritos anteriormente, aproximadamente pH 3.0, aproximadamente pH 3.5, aproximadamente pH 4.0, aproximadamente pH 4.5, aproximadamente pH 5.0, aproximadamente pH 5.5, aproximadamente pH 6.0, aproximadamente pH 6.5, aproximadamente pH 7.0, aproximadamente pH 7.5, aproximadamente pH 8.0, aproximadamente pH 8.5, aproximadamente pH 9.0, aproximadamente pH 9.5, aproximadamente pH 10.0, aproximadamente pH 10.5, aproximadamente pH 11.0, aproximadamente pH 11.5, aproximadamente pH 12.0 o más.

55 En un aspecto, la actividad de d-aminoácido transferasa es termotolerante, por ejemplo, en donde el polipéptido retiene una actividad de d-aminoácido transferasa después de exposición a una temperatura en el rango desde aproximadamente -100°C hasta aproximadamente -80°C, aproximadamente -80°C hasta aproximadamente -40°C, aproximadamente -40°C hasta aproximadamente -20°C, aproximadamente -20°C hasta aproximadamente 0°C, aproximadamente 0°C hasta aproximadamente 5°C, aproximadamente 5°C hasta aproximadamente 15°C,

aproximadamente 15°C hasta aproximadamente 25°C, aproximadamente 25°C hasta aproximadamente 37°C, aproximadamente 37°C hasta aproximadamente 45°C, aproximadamente 45°C hasta aproximadamente 55°C, aproximadamente 55°C hasta aproximadamente 70°C, aproximadamente 70°C hasta aproximadamente 75°C, aproximadamente 75°C hasta aproximadamente 85°C, aproximadamente 85°C hasta aproximadamente 90°C, aproximadamente 90°C hasta aproximadamente 95°C, aproximadamente 95°C hasta aproximadamente 100°C, aproximadamente 100°C hasta aproximadamente 105°C, aproximadamente 105°C hasta aproximadamente 110°C, aproximadamente 110°C hasta aproximadamente 120°C, ir 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 104°C, 105°C, 106°C, 107°C, 108°C, 109°C, 110°C, 111°C, 112°C, 113°C, 114°C, 115°C o más. Los polipéptidos termotolerantes de acuerdo con la invención pueden retener la actividad de d-aminoácido transferasa después de exposición a una temperatura en el rango desde aproximadamente -100°C hasta aproximadamente -80°C, aproximadamente -80°C hasta aproximadamente -40°C, aproximadamente -40°C hasta aproximadamente -20°C, aproximadamente -20°C hasta aproximadamente 0°C, aproximadamente 0°C hasta aproximadamente 5°C, aproximadamente 5°C hasta aproximadamente 15°C, aproximadamente 15°C hasta aproximadamente 25°C, aproximadamente 25°C hasta aproximadamente 37°C, aproximadamente 37°C hasta aproximadamente 45°C, aproximadamente 45°C hasta aproximadamente 55°C, aproximadamente 55°C hasta aproximadamente 70°C, aproximadamente 70°C hasta aproximadamente 75°C, aproximadamente 75°C hasta aproximadamente 85°C, aproximadamente 85°C hasta aproximadamente 90°C, aproximadamente 90°C hasta aproximadamente 95°C, aproximadamente 95°C hasta aproximadamente 100°C, aproximadamente 100°C hasta aproximadamente 105°C, aproximadamente 105°C hasta aproximadamente 110°C, aproximadamente 110°C hasta aproximadamente 120°C, o 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 104°C, 105°C, 106°C, 107°C, 108°C, 109°C, 110°C, 111°C, 112°C, 113°C, 114°C, 115°C o más. En algunas realizaciones, los polipéptidos termotolerantes de acuerdo con la invención retienen actividad de d-aminoácido transferasa después de exposición a una temperatura en los rangos descritos anteriormente, a aproximadamente pH 3.0, aproximadamente pH 3.5, aproximadamente pH 4.0, aproximadamente pH 4.5, aproximadamente pH 5.0, aproximadamente pH 5.5, aproximadamente pH 6.0, aproximadamente pH 6.5, aproximadamente pH 7.0, aproximadamente pH 7.5, aproximadamente pH 8.0, aproximadamente pH 8.5, aproximadamente pH 9.0, aproximadamente pH 9.5, aproximadamente pH 10.0, aproximadamente pH 10.5, aproximadamente pH 11.0, aproximadamente pH 11.5, aproximadamente pH 12.0 o más.

En un aspecto, la actividad de d-aminoácido transferasa de polipéptidos codificados por los ácidos nucleicos de la invención retienen actividad bajo condiciones ácidas que comprenden aproximadamente pH 6.5, pH 6, pH 5.5, pH 5, pH 4.5, pH 4.0, pH 3.5, pH 3.0 o menos (más ácido), o, retiene una actividad de d-aminoácido transferasa después de exposición a condiciones ácidas que comprenden aproximadamente pHs 6.5, pH 6, pH 5.5, pH 5, pH 4.5, pH 4.0, pH 3.5, pH 3.0 o menos (más ácidos); o retienen actividad bajo condiciones básicas que comprenden aproximadamente pH 7, pH 7.5 pH 8.0, pH 8.5, pH 9, pH 9.5, pH 10, pH 10.5, pH 11, pH 11.5, pH 12, pH 12.5 o más (más básicos) o, retienen una actividad de d-aminoácido transferasa después de exposición a condiciones básicas que comprenden aproximadamente pH 7, pH 7.5 pH 8.0, pH 8.5, pH 9, pH 9.5, pH 10, pH 10.5, pH 11, pH 11.5, pH 12, pH 12.5 o más (más básico). En un aspecto, la actividad de d-aminoácido transferasa de los polipéptidos codificados por los ácidos nucleicos de la invención retienen actividad a una temperatura de al menos 80°C, 81°C, 82°C, 83°C, 84°C, 85°C, 86°C, 87°C, 88°C, 89°C, 90°C, 91°C, 92°C, 93°C, 94°C, 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 103.5°C, 104°C, 105°C, 107°C, 108°C, 109°C o 110°C, o más, y un pH básico de al menos pH 7.5 pH 8.0, pH 8.5, pH 9, pH 9.5, pH 10, pH 10.5, pH 11, pH 11. pH 12, pH 12.5 o más (más básico).

La invención provee casetes de expresión, vehículos de clonación o un vector (por ejemplo vectores de expresión) que comprenden un ácido nucleico que comprende una secuencia de la invención. El vehículo de clonación puede comprender un vector viral, un plásmido, un fago, un fagémido, un cósmido, un fósido, un bacteriófago o un cromosoma artificial. El vector viral puede comprender un vector de adenovirus, un vector retroviral o un vector viral adenoasociado. El vehículo de clonación puede comprender un cromosoma artificial que comprende un cromosoma artificial bacteriano (BAC), un vector de bacteriófago derivado de P1 (PAC), un cromosoma artificial de levadura (YAC), o un cromosoma artificial de mamífero (MAC).

La invención describe sondas de ácidos nucleicos para identificar un ácido nucleico que codifica un polipéptido con una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo una actividad de d-aminoácido transferasa o una actividad de omega-transaminasa y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, en donde la sonda comprende al menos aproximadamente 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300 o más bases consecutivas de un ácido nucleico que comprende una secuencia de ejemplo de la invención, o, cualquier secuencia de la invención (tal como se define aquí), en donde en un aspecto (opcionalmente) la sonda comprende un oligonucleótido que comprende entre al menos aproximadamente 10 a 300, aproximadamente 25 a 250, aproximadamente 10 a 50, aproximadamente 20 a 60, aproximadamente 30 a 70, aproximadamente 40 a 80, aproximadamente 60 a 100, o aproximadamente 50 a 150 o más bases consecutivas.

La invención describe pares de cebadores de amplificación para amplificar un ácido nucleico que codifica un polipéptido que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa o una actividad de omega-transaminasa, y/o una actividad de

oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, en donde el par de cebadores es capaz de amplificar un ácido nucleico que comprende una secuencia de ejemplo de la invención, o, cualquier secuencia de la invención (tal como es definida aquí), o una subsecuencia de la misma, en donde opcionalmente un miembro del par de secuencias de cebador de amplificación comprende un oligonucleótido que comprende al menos aproximadamente 10 a 50 bases consecutivas de la secuencia, o aproximadamente, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35 o más bases consecutivas de la secuencia. La invención describe pares de cebadores de amplificación en donde el par de cebadores comprende un primer miembro que tiene una secuencia tal como es definida por aproximadamente el primero (el 5') 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35 o más residuos de una secuencia de ejemplo de la invención, o, cualquier secuencia de la invención (tal como se define aquí), y un segundo miembro que tiene una secuencia tal como es definida para aproximadamente la primera (la 5') 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35 o más residuos de la cadena complementaria del primer miembro.

La invención provee ácidos nucleicos que codifican una d-aminoácido transferasa generada por la amplificación de un polinucleótido tal como se definió más arriba utilizando un par de cebadores de amplificación descritos en la invención, en donde opcionalmente la amplificación es por reacción en cadena de polimerasa (PCR). En un aspecto, el ácido nucleico es generado por amplificación de una biblioteca de genes, en donde en un aspecto (opcionalmente) la biblioteca de genes es una biblioteca ambiental. La invención provee transferasas aisladas, sintéticas o recombinantes codificadas por una d-aminoácido transferasa-codificación de ácido nucleico como se describió anteriormente generado por amplificación de un polinucleótido utilizando un par de cebadores de amplificación descrito en la invención. La invención describe métodos para amplificar un ácido nucleico que codifica un polipéptido que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, o una actividad de omega-transaminasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, comprendiendo los métodos de la etapa de amplificación de un ácido nucleico de plantilla con un par de secuencias cebadoras de amplificación capaz de amplificar una secuencia de ejemplo de la invención, o, cualquier secuencia de la invención (tal como se define aquí), o una subsecuencia de la misma.

La invención provee un casete de expresión, un vector o un vehículo de clonación que comprende un ácido nucleico que comprende una secuencia de la invención, en donde opcionalmente el vehículo de clonación comprende un vector viral, un plásmido, un fago, una fagémido, un cósmido, un fósido, un bacteriófago o un cromosoma artificial. El vector viral puede comprender un vector de adenovirus, un vector retroviral o un vector viral adenoasociado, o, el cromosoma artificial comprende un cromosoma artificial bacteriano (BAC), un vector de bacteriófago derivado de P1 (PAC), un cromosoma artificial de levadura (YAC), o un cromosoma artificial de mamífero (MAC).

La invención provee células transformadas que comprenden un ácido nucleico o vector de la invención, o un casete de expresión o vehículo de clonación de la invención, en donde la célula transformada es una célula bacteriana, una célula de mamífero, una de célula fúngica, una célula de levadura, una célula de insecto o una célula vegetal.

La invención describe animales transgénicos no humanos que comprenden una secuencia de la invención. El animal transgénico no humano puede ser un ratón, una rata, un conejo, una oveja, un cerdo, un pollo, una cabra, un pez, un perro, o una vaca. La invención describe plantas transgénicas que comprenden una secuencia de la invención, por ejemplo, en donde la planta es una planta de maíz, una planta de sorgo, una planta de patata, una planta de tomate, una planta de trigo, una planta de oleaginosa, una planta de colza, una planta de soja, una planta de arroz, una planta de cebada, un pasto, o una planta de tabaco. La invención describe semillas transgénicas que comprenden una secuencia de la invención, por ejemplo, en donde la semilla es una semilla de maíz, una grano de trigo, una oleaginosa, una colza, semilla de soja, un grano de palma, una semilla de girasol, una semilla de sésamo, un arroz, una cebada, un cacahuate o una semilla de planta de tabaco.

La invención describe oligonucleótidos antisentido que comprenden una secuencia de ácido nucleicos complementaria a o capaz de hibridar bajo condiciones restrictivas a una secuencia de la invención (incluyendo, por ejemplo, secuencias de ejemplo de la invención), o una subsecuencia de las mismas, en donde opcionalmente el oligonucleótido antisentido tiene entre aproximadamente 10 a 50, aproximadamente 20 a 60, aproximadamente 30 a 70, aproximadamente 40 a 80 o aproximadamente 60 a 100 bases de longitud, y en un aspecto (opcionalmente) las condiciones restrictivas comprenden una etapa de lavado que comprende un lavado en 0.2X SSC a una temperatura de aproximadamente 65°C durante aproximadamente 15 minutos.

La invención describe métodos para inhibir la traducción de una transferasa, por ejemplo, una transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, una deshidrogenasa, por ejemplo una d-aminoácido deshidrogenasa mensajera en una célula que comprende administrar a la célula o expresar en la célula un oligonucleótido antisentido que comprende una secuencia de ácidos nucleicos complementaria a o capaz de hibridar bajo condiciones astringentes a una secuencia de la invención (incluyendo, por ejemplo, secuencias de ejemplo de la invención).

- La invención describe moléculas de ARN inhibidoras (ARNi) de cadena doble, que comprenden una subsecuencia de una secuencia de la invención (incluyendo, por ejemplo, secuencias de ejemplo de la invención). La molécula de ARN inhibidora (ARNi) de doble cadena puede tener aproximadamente 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, o 30, o más de nucleótidos dúplex de longitud. La invención describe métodos para
- 5 inhibir la expresión de una transferasa, por ejemplo, una transaminasa, una d-aminoácido transferasa y/o una oxidoreductasa, por ejemplo, una deshidrogenasa, por ejemplo una d-aminoácido deshidrogenasa en una célula que comprende administrar a la célula o expresar en la célula un ARN inhibidor (ARNi), de doble cadena en donde el ARN comprende una subsecuencia de una secuencia de la invención (incluyendo, por ejemplo, secuencias de ejemplo de la invención).
- 10 La invención provee polipéptidos aislados, sintéticos o recombinantes que tienen una actividad d-aminoácido transferasa:
- (a) que comprenden una secuencia de aminoácidos que tiene al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia a SEQ ID NO: 220, SEQ ID NO: 220 con una, varias o todas las modificaciones de la Tabla 46 o la Tabla 55 en donde el polipéptido o péptido de (i) o (ii) tiene una actividad de d-
- 15 aminoácido transferasa. (b) La identidad de secuencia del polipéptido o péptido de (a), puede ser determinada por: (A) por análisis con un algoritmo de comparación de secuencias o por inspección visual, o (B) sobre una región de al menos aproximadamente 20, 25, 30, 35, 40, 45, 50, 55, 60, 75, 100, 150, 200, 250, 300 o más residuos de aminoácidos, o sobre la longitud completa del polipéptido o péptido o enzima, y/o subsecuencias (fragmentos) activos enzimáticamente de los mismos,
- 20 (c) La identidad de secuencia del polipéptido o péptido de (a) o (b), puede ser determinada por análisis con un algoritmo de comparación de secuencias o por inspección visual, y opcionalmente el algoritmo de comparación de secuencias es un algoritmo BLAST versión 2.2.2 en donde se fija un parámetro de filtración a blastall blastp -p -d "nr pataa" -FF, y todas las otras opciones son fijadas por sistema;
- 25 (d) La presente invención provee adicionalmente una secuencia de aminoácidos codificada por un ácido nucleico que comprende una secuencia que tiene al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia a la secuencia de ácidos nucleico de SEQ ID NO: 219, en donde el polipéptido tiene (i) una actividad de d-aminoácido transferasa (e). La invención describe adicionalmente la secuencia de aminoácidos de cualquiera de (a) a (d), y que comprende al menos una sustitución conservadora de un residuo de aminoácido, y el polipéptido o péptido retiene la actividad de transferasa, por ejemplo, actividad de transaminasa, por ejemplo,
- 30 actividad de d-aminoácido transferasa, y/o actividad de oxidoreductasa, por ejemplo, actividad de deshidrogenasa, por ejemplo, actividad de d- aminoácido deshidrogenasa.
- (e) La invención describe adicionalmente la secuencia de aminoácidos de (d), en donde la sustitución conservadora comprende el remplazo de un aminoácido alifático por otro aminoácido alifático; el remplazo de una serina con una treonina o viceversa; el remplazo de un residuo ácido con otro residuo ácido; al remplazo de un residuo que porta un
- 35 grupo amida con otro residuo que porta un grupo amida; intercambio de un residuo básico con otro residuo básico; o, remplazo de un residuo aromático con otro residuo aromático, o una combinación de los mismos.
- (f) La invención describe adicionalmente una secuencia de aminoácidos de (e), en donde el residuo alifático comprende Alanina, Valina, Leucina, Isoleucina o un sintético equivalente de los mismos; el residuo ácido comprende ácido aspártico, ácido glutámico o un equivalente sintético de los mismos; el residuo que comprende un
- 40 grupo amida comprende ácido aspártico, ácido glutámico o un equivalente sintético de los mismos; el residuo básico comprende Lisina, Arginina o un equivalente sintético de los mismos; o, el residuo aromático comprende Fenilalanina, Tirosina o un equivalente sintético de los mismos;
- (g) La invención provee adicionalmente el polipéptido de cualquiera de (a) a (f) que tiene una actividad de d-
- 45 aminoácido transferasa, pero carece de una secuencia de señalización, un preprodominio, un dominio de enlazamiento,
- (h) el polipéptido de (g) en donde el dominio de enlazamiento comprende, o consiste de un dominio de enlazamiento de NAD, NAD (P), calcio, tiamina, FAD, zinc, ADN y/o lipoilo;
- (i) La invención provee adicionalmente el polipéptido de cualquiera de (a) a (h) que tiene actividad de d-aminoácido transferasa comprendiendo adicionalmente una secuencia heteróloga;
- 50 (j) el polipéptido de (i), en donde la secuencia heteróloga comprende, o consiste de : (A) una secuencia de señalización heteróloga, un dominio heterólogo, un dominio de enlazamiento heterólogo, un dominio de ordenamiento heterólogo, un dominio catalítico heterólogo (CD), o una combinación de los mismos; (B) la secuencia de (A), en donde la secuencia de señalización heteróloga, el dominio de enlazamiento o el dominio catalítico (CD) son derivados de una enzima heteróloga; y/o (C) una etiqueta, un epítipo, una péptido de direccionamiento, una
- 55 secuencia escindible, una unidad estructural detectable o una enzima;

(k) el polipéptido de (i) o (j), en donde la secuencia heteróloga o el dominio de enlazamiento heterólogo comprende, o consiste de, un dominio de enlazamiento de NAD, NAD (P), calcio, tiamina, FAD, zinc, ADN y/o lipilo;

(l) polipéptido de (i), en donde la secuencia de señal heteróloga direcciona la proteína codificada a una vacuola, el retículo endoplasmático, un cloroplasto o un gránulo de almidón.

- 5 En un aspecto, la actividad de la d-aminoácido transferasa comprende catalizar la transferencia de un grupo químico, catalizar la transaminación, catalizar la reacción: D-alanina + 2-oxoglutarato \rightleftharpoons piruvato + D-glutamato, y/o catalizar una reacción de oxidación-reducción, catalizar la remoción de átomos de hidrógeno, y/o catalizar la reacción: D-amino ácido + H₂O + aceptor \rightleftharpoons un 2-oxo ácido + NH₃ + aceptor reducido.

- 10 La invención provee polipéptidos aislados, sintéticos o recombinantes que comprenden un polipéptido de la invención y carece de una secuencia de señalización o de una preprosecuencia. La invención provee polipéptidos aislados, sintéticos o recombinantes que comprenden un polipéptido de la invención y tienen una secuencia de señalización heteróloga o una preprosecuencia heteróloga.

- 15 En un aspecto, un polipéptido de la invención tiene actividad de d-aminoácido transferasa que comprende una actividad específica a aproximadamente 37°C en el rango de aproximadamente 100 hasta aproximadamente 1.000 unidades por miligramo de proteína, desde aproximadamente 500 a aproximadamente 750 unidades por miligramo de proteína, o desde aproximadamente 500 hasta aproximadamente 1200 unidades por miligramo de proteína. En aspectos alternativos, los polipéptidos de la invención, tienen actividad de d-aminoácido transferasa en el rango de entre
20 aproximadamente 0.05 a 20 unidades por gramo, o 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 o más unidades por gramo, en donde una unidad es igual a una micromol de producto liberada por minuto por mg de enzima. En una realización, para transaminasas, una unidad de actividad es igual a una umol de alfa-cetoácido o cetona producido por minuto por mg de enzima (formado a partir del respectivo alfa-aminoácido o amina). En una realización alternativa, para transaminasas, una unidad de actividad es igual a una umol de alfa-aminoácido o amina producida
25 por minuto por mg de enzima (formada a partir del respectivo alfa-cetoácido o cetona).

En un aspecto, los polipéptidos de la invención comprenden al menos un sitio de glicosilación o adicionalmente comprenden un polisacárido. La glicosilación puede ser una glicosilación con enlazamiento a N, por ejemplo, en donde el polipéptido es glicosado después de ser expresado en *P. pastoris* o en *S. pombe*.

- 30 La invención provee preparación de proteínas que comprenden un polipéptido de la invención, en donde la preparación de proteína comprende un líquido, una pasta, un sólido o un gel. La invención provee heterodímeros que comprenden un polipéptido de la invención y un segundo dominio. El segundo dominio puede ser un polipéptido y el heterodímero es una proteína de fusión. El segundo dominio puede ser un epítipo o una etiqueta.

- 35 La invención provee homodímeros o heterodímeros que comprenden un polipéptido de la invención. La invención provee polipéptidos inmovilizados, en donde el polipéptido comprende una secuencia de la invención, o una subsecuencia de la misma, o un polipéptido codificado por un ácido nucleico de la invención, o un polipéptido que comprende un polipéptido de la invención y un segundo dominio, por ejemplo, en donde el polipéptido es inmovilizado sobre o dentro de una célula, una vesícula, un liposoma, una película, una membrana, un metal, una resina, un polímero, una cerámica, un vidrio, un microelectrodo, una partícula de grafito, una perla, un gel, una placa, una disposición, un tubo capilar, un cristal, un tableta, una píldora, una cápsula, un polvo, un aglomerado, una
40 superficie, un estructura porosa, o materiales tales como virutas de madera, palos, pulpa, papel y materiales derivados de los mismos.

- 45 Las transferasas de la invención pueden ser utilizadas o formuladas solas o como mezcla (un "cóctel") de transferasas y oxidoreductasas y otras enzimas hidrolíticas tales como celulasas, mananasas, proteasas, lipasas, amilasas o enzimas rédox tales como lacasas, peroxidasas, catalasas, oxidasas o reductasas. Pueden ser utilizadas formuladas en forma sólida tal como un polvo, una preparación liofilizada, un gránulo, una tableta, una barra, un cristal, una cápsula, una píldora, un pella, o en una forma líquida tal como en una solución acuosa, un aerosol, un gel, una pasta, una suspensión, una emulsión acuosa/oleosa, una crema, una cápsula, o en suspensión vesicular o micelar. Las formulaciones de la invención pueden comprender cualquiera o una combinación de los siguientes ingredientes: polioles tales como polietilenglicol, un polivinílico alcohol, un glicerol, un azúcar tal como una sacarosa,
50 un sorbitol, una trehalosa, una glucosa, una fructosa, una maltosa, una manosa, un agente gelificante tal como una goma guar, una carragenina, un alginato, un dextrano, un derivado celulósico, una pectina, una sal tal como cloruro de sodio, un sulfato de sodio, un sulfato de amonio, un cloruro de calcio, un cloruro de magnesio, un cloruro de zinc, un sulfato de zinc, una sal de un ácido graso y un derivado de un ácido graso, un quelante de metales tal como un EDTA, un EGTA, un citrato de sodio, un agente antimicrobiano tal como un ácido graso o un derivado de un ácido graso,
55 un parabeno, un sorbato, un benzoato, un compuesto modulador adicional para bloquear el impacto de una enzima tal como una proteasa, una proteína voluminosa, tal como una BSA, un hidrolizado de trigo, un compuesto borato, un aminoácido o un péptido, un compuesto modulador apropiado de pH o temperatura, un emulsificante tal

como un detergente no iónico y/o iónico, un agente rédox tal como un cistina/cisteína, una glutatona, una glutatona oxidada, un compuesto reducido o antioxidante tal como ácido ascórbico, o un dispersante. El entrecruzamiento y la modificación de la proteína tal como pegilación, modificación de ácido graso, glicosilación también pueden ser utilizados para mejorar la estabilidad de la proteína.

La invención describe arreglos que comprenden polipéptidos y/o ácidos nucleicos de la invención inmovilizados, y los arreglos comprenden un oligonucleótido inmovilizado de la invención. Las enzimas, fragmentos de las mismas y ácidos nucleicos que codifican las enzimas, o sondas de la invención, y fragmentos de las mismas, pueden ser fijados a un soporte sólido; y estas realizaciones pueden ser económicas y eficientes en el uso de enzimas y ácidos nucleicos de la invención en el procesamiento y otras aplicaciones y procesos industriales, médicos, de investigación, farmacéuticos, de alimentos y piensos y suplementos de alimentos y piensos. Por ejemplo, un consorcio o cóctel de enzimas (o fragmentos activos de las mismas), que son utilizados en una reacción química específica, pueden ser unidos a un soporte sólido y embebidos en una cuba de proceso. La reacción enzimática puede ocurrir. Entonces el soporte sólido puede ser tomado fuera de la cuba junto con las enzimas fijadas al mismo, para uso repetido. En una realización de la invención, el ácido nucleico aislado, sintético o recombinante es fijado a un soporte sólido. En otra realización de la invención, el soporte sólido es seleccionado del grupo de un gel, una resina, un polímero, una cerámica, un vidrio, un microelectrodo y cualquier combinación de los mismos.

Por ejemplo, soportes sólidos útiles en esta invención incluyen geles. Algunos ejemplos de geles incluyen sefarsa, gelatina, glutaraldehído, glutaraldehído tratado con quitosano, albúmina-glutaraldehído, quitosano xantano, gel de perla Toyo (gel polimérico), alginato, alginato-polilisina, carragenina, agarosa, glioxil agarosa, agarosa magnética, dextrano-agarosa, poli(carbamoil sulfonato) en hidrogel, hidrogel de BSA-PEG, alcohol polivinílico fosforilado (PVA), monoaminoetil-N-aminoetil (MANA), amino, o cualquier combinación de los mismos. Otro soporte sólido útil en la presente invención son resinas o polímeros. Algunos ejemplos de resinas o polímeros incluyen celulosa, acrilamida, nailon, rayón, poliéster, resinas de intercambio iónico, AMBERLITE™ XAD-7, AMBERLITE™ XAD-8, AMBERLITE™ IRA-94, AMBERLITE IRC-50™, polivinilo, poliacrílico, polimetacrilato o cualquier combinación de los mismos. Otro tipo de soporte sólido útil en la presente invención es cerámica. Algunos ejemplos incluyen cerámicas no porosas, cerámicas porosas, SiO₂, Al₂O₃. Otro tipo de soporte sólido útil en la presente invención es vidrio. Algunos ejemplos incluyen vidrio no poroso, vidrio poroso, vidrio de aminopropilo o cualquier combinación de los mismos. Otro tipo de soporte sólido que puede ser utilizado es un microelectrodo. Un ejemplo es una magnetita recubierta con polietilenimina. Las partículas de grafito pueden ser utilizadas como soporte sólido. Otro ejemplo de un soporte sólido es una célula, tal como un glóbulo rojo.

Hay muchos métodos que serían conocidos para una persona experimentada en la técnica para inmovilizar enzimas o fragmentos de las mismas, o ácidos nucleicos, sobre un soporte sólido. Algunos ejemplos de tales métodos incluyen generación electrostática de gotitas, medios electroquímicos, a través de adsorción, a través de enlazamiento covalente, a través de entrecruzamiento, a través de reacción o proceso químico, a través de encapsulación, a través de atrapamiento, a través de alginato de calcio, o a través de poli(2-hidroxietil metacrilato). Métodos similares están descritos en *Methods in Enzymology, Immobilized Enzymes and Cells*, Part C. 1987. academic Press. Edited by S. P. Colowick and N. O. Kaplan. Volume 136; and *Immobilization of Enzymes and Cells*. 1997. Humana Press. Edited by G. F. Bickerstaff. Series: *Methods in Biotechnology*, Edited by J. M. Walker.

La invención describe anticuerpos aislados, sintéticos o recombinantes que se enlazan específicamente a un polipéptido de la invención. El anticuerpo puede ser un anticuerpo monoclonal o policlonal, o es anticuerpo de cadena sencilla. La invención describe hibridomas que comprenden un anticuerpo que se enlaza específicamente a un polipéptido de la invención.

La invención describe métodos para aislar o identificar un polipéptido con una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa o una actividad de omega-transaminasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo una actividad de d-aminoácido deshidrogenasa que comprende las etapas de: (a) proveer un anticuerpo de la invención; (b) proveer una muestra que comprende polipéptidos; y (c) poner en contacto la muestra de la etapa (b) con el anticuerpo de la etapa (a) bajo condiciones en donde el anticuerpo puede enlazarse específicamente al polipéptido, aislando o identificando por lo tanto un polipéptido que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo una actividad de d-aminoácido deshidrogenasa. La invención describe métodos para hacer una antitransferasa, por ejemplo, un anticuerpo de una anti-transaminasa, por ejemplo anti-d-aminoácido transferasa, y/o un anti-oxidorreductasa, por ejemplo, anti-deshidrogenasa, por ejemplo anti d-aminoácido deshidrogenasa, que comprende administrar a un animal no humano de un ácido nucleico de la invención o una subsecuencia del mismo en una cantidad suficiente para generar una respuesta inmune humoral, haciendo por lo tanto un anticuerpo de anti-transferasa, por ejemplo, anti-transaminasa, por ejemplo anti d-aminoácido transferasa y/o anti-oxidorreductasa, por ejemplo, anti-deshidrogenasa, por ejemplo anti-d-aminoácido deshidrogenasa. La invención describe métodos para hacer un anticuerpo de anti-transferasa, por ejemplo, anti-transaminasa, por ejemplo anti d-amino-ácido transferasa, y/o anti-oxidorreductasa, por ejemplo, anti-deshidrogenasa, por ejemplo, anti d-aminoácido deshidrogenasa que comprende administrar a un animal no humano

un polipéptido de la invención o una subsecuencia del mismo en una cantidad suficiente para generar una respuesta inmune humoral, haciendo por lo tanto un anticuerpo anti-transferasa, por ejemplo anti-transaminasa, por ejemplo, anti d-aminoácido transferasa y/o anti-oxidorreductasa, por ejemplo anti-deshidrogenasa, por ejemplo anti-d-aminoácido deshidrogenasa.

- 5 La invención provee métodos para producir un polipéptido recombinante que comprende las etapas de: (a) proveer un ácido nucleico enlazado operativamente a un promotor, en donde el ácido nucleico comprende una secuencia de la invención; y (b) expresar el ácido nucleico de la etapa (a) bajo condiciones que permitan la expresión del polipéptido, produciendo por lo tanto un polipéptido recombinante. El método puede comprender adicionalmente la transformación de una célula anfitriona con el ácido nucleico de la etapa (a) seguido por la expresión del ácido nucleico de la etapa (a), produciendo por lo tanto un polipéptido recombinante en una célula transformada.

- 10 La invención describe métodos para identificar un polipéptido que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo una actividad de d-aminoácido deshidrogenasa que comprende: (a) proveer un polipéptido de la invención; (b) proveer un sustrato de una transferasa, por ejemplo una transaminasa, por ejemplo una d-aminoácido transferasa y/o una oxidorreductasa, por ejemplo una deshidrogenasa, por ejemplo de d-aminoácido deshidrogenasa; y (c) poner en contacto el polipéptido con el sustrato de la etapa (b) y detectar un descenso en la cantidad de sustrato o un incremento en la cantidad de un producto de reacción, en donde un descenso en el cantidad del sustrato o un incremento en la cantidad del producto de reacción detecta un polipéptido que tiene una actividad de transferasa, por ejemplo una actividad de transaminasa, por ejemplo una actividad de d- aminoácido transferasa y/o una actividad de oxidorreductasa, por ejemplo una actividad de deshidrogenasa, por ejemplo una actividad de d- aminoácido deshidrogenasa.

- 15 La invención describe métodos para identificar un sustrato de una transferasa, por ejemplo una transaminasa, por ejemplo una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo una deshidrogenasa, por ejemplo una d-aminoácido deshidrogenasa que comprende: (a) proveer un polipéptido de la invención; (b) proveer un sustrato de prueba; y (c) poner en contacto el polipéptido de la etapa (a) con el sustrato de prueba de la etapa (b) y detectar un descenso en la cantidad de sustrato o un incremento en la cantidad de un producto de reacción, en donde un descenso en la cantidad del sustrato o un incremento en la cantidad de un producto de reacción identifica el sustrato de prueba como una transferasa, por ejemplo, una transaminasa, por ejemplo una d-aminoácido transferasa y/o una oxidorreductasa, por ejemplo una deshidrogenasa, por ejemplo un sustrato de d-aminoácido deshidrogenasa.

- 20 La invención describe métodos para determinar si un compuesto de pruebas se enlaza específicamente a un polipéptido que comprende: (a) expresar un ácido nucleico o un vector que comprende el ácido nucleico bajo condiciones permisivas para traducción del ácido nucleico a un polipéptido, en donde el ácido nucleico tiene una secuencia de la invención; (b) proveer un compuesto de prueba; (c) poner en contacto el polipéptido con el compuesto de prueba; y (d) determinar si el compuesto de prueba de la etapa (b) se enlaza específicamente al polipéptido.

- 25 La invención describe métodos para determinar si un compuesto de prueba se enlaza específicamente a un polipéptido que comprende: (a) proveer un polipéptido de la invención (b) proveer un compuesto de prueba; (c) poner en contacto el polipéptido con el compuesto de prueba; y (d) determinar si el compuesto de prueba de la etapa (b) se enlaza específicamente al polipéptido.

- 30 La invención describe métodos para identificar un modulador de una actividad de transferasa, por ejemplo actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, que comprende: (a) proveer un polipéptido de la invención; (b) proveer un compuesto de prueba; (c) poner en contacto el polipéptido de la etapa (a) con el compuesto de prueba de la etapa (b) y medir una actividad de la transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, en donde un cambio en la actividad de transferasa, por ejemplo, la actividad de transaminasa, actividad de d-aminoácido transferasa y/o actividad de oxidorreductasa, por ejemplo, actividad de deshidrogenasa, por ejemplo, actividad de d-aminoácido deshidrogenasa, y/o actividad de oxidorreductasa, por ejemplo, actividad de deshidrogenasa, por ejemplo, actividad de d-aminoácido deshidrogenasa medida en la presencia del compuesto de prueba en comparación con la actividad en la ausencia del compuesto de prueba provee una determinación de que el compuesto de prueba modula la actividad de la transferasa, por ejemplo, actividad de transaminasa, por ejemplo, actividad de d-aminoácido transferasa y/o actividad de oxidorreductasa, por ejemplo, actividad de deshidrogenasa, por ejemplo actividad de d-aminoácido deshidrogenasa. La actividad de transferasa, por ejemplo, actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, puede medirse mediante la provisión de una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa en calidad de sustrato y detectar un descenso en la cantidad del sustrato o un incremento en la cantidad de un producto de reacción, o, un incremento en la cantidad del sustrato o un descenso en el cantidad de un producto de reacción. En

un aspecto, un descenso en la cantidad del sustrato o un incremento en la cantidad de el producto de reacción con el compuesto de prueba en comparación con la cantidad de sustrato o producto de reacción sin el compuesto de prueba identifica el compuesto de prueba como un activado de una actividad de transferasa, por ejemplo actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa. En un aspecto, un incremento en la cantidad del sustrato o un descenso en la cantidad del producto de reacción con el compuesto de prueba en comparación con la cantidad del sustrato o producto de reacción sin el compuesto de prueba identifica el compuesto de prueba como un inhibidor de una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa.

La invención describe sistemas de ordenador que comprenden un procesador y un dispositivo de almacenamiento de datos en donde dicho dispositivo de almacenamiento de datos ha almacenado en sí una secuencia de polipéptidos o una secuencia de ácidos nucleicos, en donde la secuencia de polipéptidos comprende una secuencia de la invención, un polipéptido codificado por un ácido nucleico de la invención. Los sistemas de ordenador pueden comprender adicionalmente un algoritmo de comparación de secuencias y un dispositivo de almacenamiento de datos que tiene al menos una secuencia de referencia almacenada en el mismo. En otro aspecto, el algoritmo de comparación de secuencia comprende un programa de ordenador que indica polimorfismos. En un aspecto, el sistema de ordenador puede comprender adicionalmente un identificador que identifica una o más características en dicha secuencia. La invención describe medios legibles por ordenador que tienen almacenados en sí mismos una secuencia de polipéptidos o de una secuencia de ácidos nucleicos de la invención. La invención describe métodos para identificar una característica en una secuencia que comprende las etapas de: (a) leer la secuencia utilizando un programa de ordenador que identifica una o más características en una secuencia, en donde la secuencia comprende una secuencia de polipéptidos o una secuencia de ácidos nucleicos de la invención; y (b) identificar una o más características en la secuencia con el programa de ordenador. La invención describe métodos para comparar una primera secuencia con una segunda secuencia que comprenden las etapas de: (a) leer la primera secuencia y la segunda secuencia a través del uso de un programa de ordenador el cual compara secuencias, en donde la primera secuencia comprende una secuencia de polipéptidos o una secuencia de ácidos nucleicos de la invención; y (b) determinar diferencias entre la primera secuencia y la segunda secuencia con el programa de ordenador. La etapa de determinar inferencias entre la primera secuencia y la segunda secuencia pueden comprender adicionalmente la etapa de identificar polimorfismos. En un aspecto, el método puede comprender adicionalmente un identificador que identifica una o más características en una secuencia. En otro aspecto, el método puede comprender leer la primera secuencia dejando un programa de ordenador e identificar una o más características en la secuencia.

La invención describe métodos para aislar o recuperar un ácido nucleico que codifica un polipéptido que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa a partir de una muestra ambiental que comprende las etapas de: (a) proveer un par de secuencias de cebador de amplificación para amplificar un ácido nucleico que codifica una polipéptido que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, en donde el par cebador es capaz de amplificar un ácido nucleico de la invención; (b) aislar un ácido nucleico de la muestra ambiental o tratar la muestra ambiental de tal manera que el ácido nucleico en la muestra es accesible para hibridación al par cebador de amplificación; y, (c) combinar el ácido nucleico de la etapa (b) con el par cebador de amplificación de la etapa (a) y amplificar el ácido nucleico de la muestra ambiental, por lo tanto aislando o recuperando un ácido nucleico que codifica un polipéptido que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa a partir de una muestra ambiental. Uno o cada miembro del par de secuencias cebadoras de amplificación puede comprender un oligonucleótido que comprende al menos 10 a 50 bases consecutivas de una secuencia de la invención. En un aspecto, el par de secuencias cebadoras de amplificación es un par de amplificación de la invención.

La invención describe métodos para aislar o recuperar un ácido nucleico que codifica un polipéptido que tiene una actividad de transferasa, por ejemplo a actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa a partir de una muestra ambiental que comprenda las etapas: (a) proveer una sonda de polinucleótido que comprende un ácido nucleico de la invención o una subsecuencia del mismo; (b) aislar un ácido nucleico de la muestra ambiental o tratar la muestra ambiental de tal manera que el ácido nucleico en la muestra es accesible para hibridación a una sonda de polinucleótido de la etapa (a); (c) combinar el ácido nucleico aislado, sintético o recombinante o la muestra ambiental tratada de la etapa (b) con la sonda de polinucleótido de la etapa (a); y (d) aislar un ácido nucleico que hibrida específicamente con la sonda de polinucleótido de la etapa (a), por lo tanto aislando o recuperando un ácido nucleico que codifica un polipéptido que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa,

y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa a partir de una muestra ambiental. La muestra ambiental puede comprender una muestra de agua, una muestra líquida, una muestra de suelo, una muestra de aire o una muestra biológica. En un aspecto, la muestra biológica puede ser derivada de una célula bacteriana, una célula de protozoo, una célula de insecto, una célula de levadura, una célula vegetal, una célula fúngica o una célula de mamífero.

La invención provee métodos para generar una variante de un ácido nucleico que comprende una secuencia que tiene al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia con la secuencia de ácidos nucleicos de SEQ ID NO: 219 que codifica un polipéptido que tiene

una actividad de d-aminoácido transferasa

que comprende las etapas de: (a) proveer un ácido nucleico de plantilla que comprende la SEQ ID NO: 219; y (b) modificar, eliminar o agregar uno o más nucleótidos en la secuencia de plantilla o una combinación de los mismos para generar una variante del ácido nucleico de plantilla. En un aspecto, el método puede comprender adicionalmente expresar la relación del ácido nucleico variante para generar un polipéptido variante de d-aminoácido transferasa. Las modificaciones, adiciones o eliminaciones pueden ser introducidas por un método que comprende PCR propenso a error, mezcla, mutagénesis dirigida por oligonucleótidos, PCR de ensamblaje, mutagénesis por PCR sexual, mutagénesis *in vivo*, mutagénesis en casete, mutagénesis de conjunto recursivo, mutagénesis de conjunto exponencial, mutagénesis específica para el sitio, reensamblaje del gen (por ejemplo, reensamblaje del gen, véase, por ejemplo, la Patente de Estados Unidos No. 6,537,776), en mutagénesis de saturación del sitio del gen (GSSM), reensamblaje de ligazón sintética (SLR) o una combinación de los mismos. En otro aspecto, la mutaciones, adiciones o eliminaciones son introducidas por un método que comprende recombinación, recombinación de secuencias recursivas, mutagénesis de ADN modificada por fosfotoato, mutagénesis de plantilla que contiene uracilo, mutagénesis dúplex con brechas, mutagénesis de reparación de errores puntuales, mutagénesis de cepa anfitrión deficiente en reparaciones, mutagénesis química, mutagénesis radiogénica, mutagénesis por eliminación, mutagénesis por restricción-selección, mutagénesis por restricción-purificación, síntesis de genes artificiales, mutagénesis de conjunto, creación de multímeros de ácidos nucleicos quiméricos y una combinación de los mismos.

En un aspecto, el método puede ser repetido iterativamente hasta que una transferasa por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa que tiene una actividad alterada o diferente o una estabilidad alterada o diferente de la del polipéptido codificado por el ácido nucleico de plantilla se produce. En un aspecto el polipéptido variante de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa es termotolerante, y retiene alguna actividad después de ser expuesto a una temperatura elevada. En otro aspecto, el polipéptido variante de transferasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa tienen glicosilación incrementada en comparación con la transferasa por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidorreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa codificadas por un ácido nucleico de plantilla. Alternativamente, el polipéptido variante de transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa bajo una temperatura alta, en donde la transferasa por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidorreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa codificada por el ácido nucleico de plantilla no es activa bajo la alta temperatura. En un aspecto, el método puede ser repetido iterativamente hasta que una secuencia que codifica una transferasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa que tiene un uso de codón alterado en comparación con el del ácido nucleico de plantilla se produce. En otro aspecto, el método puede ser repetido iterativamente hasta que **se produce** un gen de transferasa, por ejemplo a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa que tiene un nivel más alto o más bajo de expresión de mensaje o estabilidad en comparación con el del ácido nucleico de plantilla. En otro aspecto, la formulación del producto final de transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa permite un incremento o modulación del rendimiento de la transferasa por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidorreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa en el producto.

La invención provee métodos para modificar codones en un ácido nucleico que codifica un polipéptido que tiene una actividad de d-aminoácido transferasa para incrementar su expresión en una célula anfitriona, comprendiendo el método: (a) proveer un ácido nucleico de la invención que codifica un polipéptido que tiene una actividad de d-aminoácido transferasa; y (b) identificar un codón no preferido o menos preferido en el ácido nucleico de la etapa (a)

y reemplazarlo con un codón preferido o neutralmente utilizado que codifica el mismo aminoácido que el codón reemplazado, en donde un codón preferido es un codón sobrerrepresentado en secuencias de codificación en genes en la célula anfitriona y un codón no preferido o menos preferido es un codón subrepresentado en las secuencias de codificación en genes en la célula anfitriona, modificando por lo tanto el ácido nucleico para incrementar su expresión en una célula anfitriona.

La invención provee métodos para modificar codones en un ácido nucleico que codifica un polipéptido que tiene una actividad de d-aminoácido transferasa; el método comprende: (a) proveer un ácido nucleico de la invención; y (b) identificar un codón en el ácido nucleico de la etapa (a) y reemplazarlo con un codón diferente que codifica el mismo aminoácido que el codón reemplazado, modificando por lo tanto los codones en un ácido nucleico que codifica una d-aminoácido transferasa.

La invención provee métodos para modificar codones en un ácido nucleico que codifica un polipéptido que tiene una actividad de d-aminoácido transferasa para incrementar su expresión en una célula anfitriona, comprendiendo el método: (a) proveer un ácido nucleico de la invención que codifica una d-aminoácido transferasa y, (b) identificar un codón no preferido o uno menos preferido en el ácido nucleico de la etapa (a) y reemplazarlo con un codón preferido o neutralmente utilizado que codifica el mismo aminoácido que el codón reemplazado, en donde el codón preferido es un codón sobrerrepresentado en secuencias de codificación en genes en la célula anfitriona y un codón no preferido o menos preferido es un codón subrepresentado en secuencias de codificación en genes en la célula anfitriona, modificando por lo tanto el ácido nucleico para implementar su expresión en una célula anfitriona.

La invención provee métodos para modificar un codón en un ácido nucleico que codifica un polipéptido que tiene una actividad de d-aminoácido transferasa en una célula anfitriona, comprendiendo el método: (a) proveer un ácido nucleico de la invención; y (b) identificar al menos un codón preferido en el ácido nucleico de la etapa (a) y reemplazarlo con un codón no preferido o menos preferido que codifica el mismo aminoácido que el codón reemplazado, en donde el codón preferido es un codón sobrerrepresentado en secuencias de codificación en genes en una célula anfitriona y un codón no preferido o menos preferido es un codón subrepresentado en secuencias de codificación en genes en la célula anfitriona, modificando por lo tanto el ácido nucleico para disminuir su expresión en una célula anfitriona. En un aspecto, la célula anfitriona puede ser una célula bacteriana, una célula fúngica, una célula de insecto, una célula de levadura, una célula vegetal o una célula de mamífero.

La invención describe métodos para producir una biblioteca de ácidos nucleicos que codifican una pluralidad de sitios activos modificados de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa o sitios de enlazamiento del sustrato, en donde los sitios activos modificados o los sitios de enlazamiento al sustrato son derivados de un primer ácido nucleico que comprende una secuencia que codifica un primer sitio activo o un primer sitio de enlazamiento del sustrato comprendiendo el método: (a) proveer un primer ácido nucleico que codifica un primer sitio activo o primer sitio de enlazamiento del sustrato, en donde la secuencia del primer ácido nucleico comprende una secuencia que hibrida bajo condiciones restrictivas a una secuencia de la invención, o una subsecuencia del mismo, y el ácido nucleico codifica un sitio activo de una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa o un sitio de enlazamiento de sustrato de transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa; (b) proveer un conjunto de oligonucleótidos mutagénicos que codifican variantes de aminoácidos de origen natural a una pluralidad de codones direccionados en el primer ácido nucleico; y (c) utilizar el conjunto de oligonucleótidos mutagénicos para generar un conjunto de ácidos nucleicos variantes que codifican sitio activo o que codifican sitio de enlazamiento al sustrato codificando un rango de variaciones de aminoácidos en cada codón de aminoácido que fue mutagenizado, produciendo por lo tanto una biblioteca de ácidos nucleicos que codifican una pluralidad de sitios activos o sitios de enlazamiento al sustrato de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. En un aspecto, el método comprende la mutagenización del primer ácido nucleico de la etapa (a) por un método que comprende un sistema de evolución dirigido optimizado, mutagénesis de saturación al sitio en gen (GSSM), o un reensamblaje de ligazón sintética (SLR). En un aspecto, el método comprende mutagenizar el primer ácido nucleico de la etapa (a) o variantes por un método que comprende PCR propenso a error, mezcla, mutagénesis dirigida a oligonucleótidos, PCR de ensamblaje, mutagénesis por PCR sexual, mutagénesis *in vivo*, mutagénesis por casete, mutagénesis por ensamble recursivo, mutagénesis por ensamble exponencial, mutagénesis específica para el sitio, reensamblaje del gen (reensamblaje del gen, Patente de los Estados Unidos No. 6,537,776), mutagénesis de saturación en el sitio del gen (GSSM), reensamblaje de ligazón sintética (SLR) y una combinación de los mismos. En un aspecto, el método comprende mutagenizar el primer ácido nucleico de la etapa (a) o variantes por un método que comprende recombinación, recombinación de secuencia recursiva, mutagénesis de ADN modificada por fosfotioato, mutagénesis de plantilla que contiene uracilo, mutagénesis dúplex con brechas, mutagénesis de reparación de errores puntuales, mutagénesis de cepa anfitriona deficiente en reparación, mutagénesis química, mutagénesis radiogénica, mutagénesis por eliminación, mutagénesis por restricción-selección, mutagénesis por restricción-purificación, síntesis de gen artificial, mutagénesis de ensamble, creación de multímeros de ácidos nucleicos quiméricos y una combinación de los mismos.

La invención describe métodos para hacer una molécula pequeña que comprende: (a) proveer una pluralidad de enzimas biosintéticas capaces de sintetizar o modificar una molécula pequeña, en donde una de las enzimas comprende una enzima transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa, codificada por un ácido nucleico de la invención; (b) proveer un sustrato para al menos una de las enzimas de la etapa (a); y (c) hacer reaccionar el sustrato de la etapa (b) con las enzimas bajo condiciones que facilitan una pluralidad de reacciones biocatalíticas para generar una molécula pequeña mediante una serie de reacciones biocatalíticas. La invención describe métodos para modificar una molécula pequeña que comprende: (a) proveer una transferasa, por ejemplo, una enzima transferasa por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa, en donde la enzima comprende un polipéptido de la invención, o, un polipéptido codificado por un ácido nucleico de la invención, o una subsecuencia del mismo; (b) proveer una molécula pequeña; (c) hacer reaccionar la enzima de la etapa (a) con la molécula pequeña de la etapa (b) bajo condiciones que faciliten una reacción enzimática catalizada por la enzima transferasa, por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidorreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa, por lo tanto modificando una molécula pequeña mediante una reacción enzimática de transferasa por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa. En un aspecto, el método puede comprender una pluralidad de sustratos de moléculas pequeñas para la enzima de la etapa (a), generando por lo tanto una biblioteca de moléculas pequeñas modificadas producidas por al menos una reacción enzimática catalizada por la enzima transferasa, por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidorreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa. En un aspecto, el método puede comprender una pluralidad de enzimas adicionales bajo condiciones que facilitan una pluralidad de reacciones biocatalíticas por parte de las enzimas para formar una biblioteca de moléculas pequeñas modificadas producidas por la pluralidad de reacciones enzimáticas. En otro aspecto, el método puede comprender adicionalmente la etapa de probar la biblioteca para determinar si una molécula pequeña modificada en particular que exhibe una actividad deseada está presente dentro de la biblioteca. La etapa de probar la biblioteca puede comprender adicionalmente las etapas de eliminar sistemáticamente todas excepto una de las reacciones biocatalíticas utilizadas para producir una porción de la pluralidad de las moléculas pequeñas modificadas dentro de la biblioteca probando la porción de la molécula pequeña modificada en cuanto a la presencia o ausencia de la molécula pequeña modificada en particular con una actividad deseada, e identificar al menos una reacción biocatalítica específica que produce la molécula pequeña modificada en particular de actividad deseada.

La invención describe métodos para determinar un fragmento funcional de una enzima transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa que comprende las etapas de: (a) proveer una enzima transferasa por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa, en donde la enzima comprende un polipéptido de la invención, o un polipéptido codificado por un ácido nucleico de la invención, o una subsecuencia del mismo; y (b) eliminar una pluralidad de residuos de aminoácidos de la secuencia de la etapa (a) y probar la subsecuencia remanente en cuanto a una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, determinando por lo tanto un fragmento funcional de una enzima transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa. En un aspecto, la actividad de transferasa, por ejemplo, por ejemplo, actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, se mide proveyendo un sustrato para transferasa por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa y detectando un descenso en la cantidad de sustrato o un incremento en la cantidad de un producto de reacción.

La invención describe métodos para manipulación de una célula entera de fenotipos nuevos o modificados utilizando análisis de flujo metabólico en tiempo real, comprendiendo el método: (a) hacer una célula modificada modificando la composición genética de una célula, en donde la composición genética es modificada por la adición a la célula de un ácido nucleico de la invención; (b) cultivar la célula modificada para generar una pluralidad de células modificadas; (c) medir al menos un parámetro metabólico de la célula monitoreando el cultivo celular de la etapa (b) en tiempo real; y, (d) analizar los datos de la etapa (c) para determinar si el parámetro medido difiere de una medición comparable en una célula no modificada bajo condiciones similares, identificando por lo tanto un fenotipo manipulado en la célula utilizando análisis de flujo metabólico en tiempo real. En un aspecto, la composición genética de la célula puede ser modificada mediante un método que comprende la eliminación de una secuencia o modificación de una secuencia en la célula, o, la anulación de la expresión de un gen. En un aspecto, el método puede comprender adicionalmente la selección de una célula que comprende un fenotipo recién manipulado. En otro aspecto, el método puede comprender cultivar la célula seleccionada, generando por lo tanto una nueva cepa celular que comprende un fenotipo recién manipulado.

La invención describe secuencias de señalización aisladas, sintéticas o recombinantes que consisten de, o que comprende, una secuencia como se establece en los residuos 1 a 12, 1 a 13, 1 a 14, 1 a 15, 1 a 16, 1 a 17, 1 a 18, 1 a 19, 1 a 20, 1 a 21, 1 a 22, 1 a 23, 1 a 24, 1 a 25, 1 a 26, 1 a 27, 1 a 28, 1 a 28, 1 a 30, 1 a 31, 1 a 32, 1 a 33, 1 a 34, 1 a 35, 1 a 36, 1 a 37, 1 a 38, 1 a 40, 1 a 41, 1 a 42, 1 a 43 o 1 a 44, de un polipéptido de la invención, incluyendo secuencias de polipéptidos de ejemplo de la invención.

La invención describe polipéptidos quiméricos que comprenden al menos un primer dominio que comprende un péptido de señalización (SP) y al menos un segundo dominio que comprende un polipéptido heterólogo o un péptido que comprende una secuencia de la invención, o una subsecuencia del mismo, en donde el polipéptido o péptido heterólogo no está asociado de manera natural con el péptido de señalización (SP). En un aspecto, el péptido de señalización (SP) no es derivado de una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidoreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa. El polipéptido o péptido heterólogo puede ser amino terminal, o carboxi terminal, en ambos extremos del péptido de señalización (SP) o un dominio catalítico (CD) de transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidoreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa. La invención provee ácidos nucleicos aislados, sintéticos o recombinantes que codifican un polipéptido quimérico, en donde el polipéptido quimérico comprende al menos un primer dominio que comprende un péptido de señalización (SP) y al menos un segundo dominio que comprende un polipéptido o péptido heterólogo que comprende una secuencia de la invención, o una subsecuencia del mismo, en donde el polipéptido o péptido heterólogo no está asociado de manera natural con el péptido de señalización (SP).

La invención describe métodos para incrementar la termotolerancia o termoestabilidad de un polipéptido de transferasa, por ejemplo por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidoreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa, comprendiendo el primer método glicosilar un polipéptido de transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidoreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa, en donde el polipéptido comprende al menos treinta aminoácidos contiguos de un polipéptido de la invención; o un polipéptido codificado por una secuencia de ácidos nucleicos de la invención, incrementando por lo tanto la termotolerancia o termoestabilidad del polipéptido de la transferasa, por ejemplo por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidoreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa. En un aspecto, la actividad específica de una transferasa por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidoreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa puede ser termoestable o termotolerante a una temperatura del rango que va de más de aproximadamente 0°C hasta aproximadamente 20°C, aproximadamente 20°C hasta aproximadamente 37°C; aproximadamente 37°C hasta aproximadamente 50°C, aproximadamente 50°C hasta aproximadamente 70°C, aproximadamente 70°C hasta aproximadamente 75°C, aproximadamente 75°C hasta aproximadamente 80°C, aproximadamente 80°C hasta aproximadamente 85°C, aproximadamente 85°C hasta aproximadamente 90°C, aproximadamente 90°C hasta aproximadamente 95°C, aproximadamente 95°C hasta aproximadamente 100°C, aproximadamente 100°C hasta aproximadamente 110°C, o mayor.

La invención provee métodos para sobreexpresar un polipéptido recombinante de transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa en una célula que comprende la expresión de un vector que comprende un ácido nucleico que comprende un ácido nucleico de la invención o una secuencia de ácidos nucleicos de la invención, en donde las identidades de secuencia son determinadas por análisis con un algoritmo de comparación de secuencia o por inspección visual, en donde se efectúa sobreexpresión mediante el uso de un promotor de alta actividad, un vector dicistrónico o por una amplificación genética del vector.

La invención describe métodos para hacer una planta transgénica y semillas que comprenden: (a) introducir una secuencia de ácidos nucleicos heterólogos en la célula, en donde la secuencia de ácidos nucleicos heterólogos comprende una secuencia de ácidos nucleicos de la invención, produciendo por lo tanto una célula de planta o semilla transformada; y (b) producir una planta transgénica a partir de la célula o semilla transformada. En un aspecto, la etapa (a) puede comprender adicionalmente la introducción de la secuencia de ácido nucleico heterólogo por electroporación o microinyección de protoplastos de células vegetales. En otro aspecto, la etapa (a) puede comprender la introducción de la secuencia de ácidos nucleicos heterólogos directamente en el tejido vegetal mediante bombardeo con partículas de ADN. Alternativamente, la etapa (a) puede comprender adicionalmente la introducción de la secuencia de ácidos nucleicos heterólogos en el ADN de la célula vegetal utilizando un anfitrión de *Agrobacterium tumefaciens*. En un aspecto, la célula vegetal puede ser una célula de patata, maíz, arroz, trigo, tabaco, o cebada.

La invención describe métodos para expresar una secuencia de ácidos nucleicos heterólogos en una célula vegetal que comprenden: (a) transformar la célula vegetal con una secuencia de ácidos nucleicos heterólogos enlazados operativamente a un promotor, en donde la secuencia de ácidos nucleicos heterólogos comprende un ácido nucleico de la invención; (b) cultivar la planta bajo condiciones en donde la secuencia de ácidos nucleicos heterólogos es expresada en la célula vegetal. La invención describe métodos para expresar una secuencia de ácidos nucleicos

heterólogos en una célula vegetal que comprenden: (a) transformar la célula vegetal con secuencias de ácidos nucleicos heterólogos enlazados operativamente a un promotor, en donde la secuencia de ácidos nucleicos heterólogos comprende una secuencia de la invención; (b) cultivar la planta bajo condiciones en donde la secuencia de ácidos nucleicos heterólogos es expresada en la célula vegetal.

- 5 La invención describe composiciones detergentes que comprenden un polipéptido de la invención, o un polipéptido codificado por un ácido nucleico de la invención, en donde el polipéptido tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa. La transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa puede no tener actividad de superficie o tener actividad de superficie. La transferasa, por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidorreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa puede ser formulada en una composición líquida no acuosa, una fundición sólida, una forma granular, una forma en partículas, una tableta comprimida, una forma de gel, una pasta o una forma de suspensión.
- 10 La invención describe métodos para lavar un objeto que comprenden: (a) proveer una composición que comprende un polipéptido de la invención que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, o un polipéptido codificado por un ácido nucleico de la invención; (b) proveer un objeto; y (c) poner en contacto el polipéptido de la etapa (a) y el objeto de la etapa (b) bajo condiciones en donde la composición puede lavar el objeto.
- 15
- 20

- La invención describe textiles o telas, incluyendo, por ejemplo, hilados, que comprenden un polipéptido de la invención, o un polipéptido codificado por un ácido nucleico de la invención. La invención provee métodos para tratar un textil o tela (por ejemplo, removiendo una mancha de una composición) que comprenden: (a) proveer una composición que comprende un polipéptido de la invención que tiene una actividad de transferasa, por ejemplo, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, o un polipéptido codificado por un ácido nucleico de la invención; (b) proveer un textil o tela; y (c) poner en contacto el polipéptido de la etapa (a) y la composición de la etapa (b) bajo condiciones en donde la transferasa, por ejemplo, por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidorreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa pueden tratar el textil o tela (por ejemplo, remover la mancha). La invención describe métodos para mejorar el acabado de una tela que comprenden: (a) proveer una composición que comprende un polipéptido de la invención que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, o un polipéptido codificado por un ácido nucleico de la invención; (b) proveer una tela; y (c) poner en contacto el polipéptido de la etapa (a) y la tela de la etapa (b) bajo condiciones en donde el polipéptido puede tratar la tela mejorando por lo tanto el acabado de la tela. En un aspecto, la tela es una lana o una seda. En otro aspecto, la tela es una fibra de celulosa o una mezcla de una fibra natural y una fibra sintética.
- 25
- 30
- 35

- 40 La invención provee piensos, alimentos, suplementos para piensos, suplementos alimenticios, composiciones dietarias o auxiliares dietarios que comprenden un polipéptido de la invención, o un polipéptido codificado por un ácido nucleico de la invención. El alimento o el pienso pueden ser, por ejemplo, un cereal, un grano, un maíz y similares.

- La invención describe productos de masa, panadería u horneados y/o precursores de productos de masa, panadería, u horneados que comprenden un polipéptido que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, en donde el polipéptido comprende una secuencia de la invención, o el polipéptido es codificado por un ácido nucleico que comprende una secuencia de la invención, o un fragmento enzimáticamente activo del mismo.
- 45
- 50

- La invención provee bebidas y precursores de bebidas que comprenden un polipéptido, o un fragmento enzimáticamente activo del mismo, que tiene una actividad de d-aminoácido transferasa, en donde el polipéptido comprende una secuencia de la invención como se definió anteriormente, o el polipéptido es codificado por un ácido nucleico que comprende una secuencia de la invención como se definió más arriba. La invención describe métodos para producción de bebidas que comprenden la administración de al menos un polipéptido que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, en donde el polipéptido comprende una secuencia de la invención, o el polipéptido es codificado por un ácido nucleico que comprende una secuencia de la invención, o un fragmento enzimáticamente
- 55

activo del mismo, a una bebida o un precursor de una bebida, en donde en un aspecto (opcionalmente) la bebida o precursor de bebida es una infusión o una cerveza.

La invención provee suplementos para alimentos, piensos o nutricionales, por ejemplo, para un humano o un animal que comprenden un polipéptido de la invención, por ejemplo un polipéptido codificado por un ácido nucleico de la invención. En un aspecto, el polipéptido en el alimento o suplemento nutricional puede ser glicosilado. La invención provee matrices de administración de enzimas comestibles que comprenden un polipéptido de la invención, por ejemplo, un polipéptido codificado por el ácido nucleico de la invención. En un aspecto, la matriz de administración comprende una pella. En un aspecto, el polipéptido puede ser glicosilado. En un aspecto, la actividad de la d-aminoácido transferasa es termotolerante. En otro aspecto, la actividad de la d-aminoácido transferasa es termoestable.

En un aspecto, la enzima d-aminoácido transferasa puede ser preparada por expresión de un polinucleótido que codifica la d-aminoácido transferasa en un organismo seleccionado del grupo que consistente de una bacteria, una levadura, una planta, un insecto, un hongo y un animal. El organismo puede ser seleccionado del grupo consistente de un *S. pombe*, *S. cerevisiae*, *Pichia pastoris*, *Pseudomonas* sp., *E. coli*, *Streptomyces* sp., *Bacillus* sp. y *Lactobacillus* sp.

La invención provee una matriz de administración de enzimas comestible que comprende un polipéptido de d-aminoácido transferasa recombinante termoestable de la invención. La invención describe métodos para administrar un suplemento de transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidoreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa a un animal, comprendiendo el método: preparar una matriz de administración de enzima comestible en la forma de pellas que comprenden un vehículo comestible granulado y una enzima termoestable recombinante de transferasa, por ejemplo una enzima por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, en donde en las pellas se dispersa fácilmente la enzima transferasa, por ejemplo, la enzima transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidoreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa contenida en ella en medios acuosos, y administrar la matriz de administración de enzima comestible al animal. La enzima transferasa recombinante transferasa, por ejemplo, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, pueden comprender un polipéptido de la invención. El vehículo comestible granulado puede comprender un vehículo seleccionado del grupo que consiste de un germen de grano, un germen de grano que está libre de aceite, un heno, una alfalfa, un fleo, una cáscara de soja, una torta de semilla de girasol y una torta de trigo. El vehículo comestible puede comprender un germen de grano que está libre de aceite. La enzima transferasa, por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidoreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa pueden ser glicosilados para proveer termoestabilidad en las condiciones de pelletización. La matriz de administración puede ser formada pelletizando una mezcla que comprende un germen de grano y una transferasa por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidoreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa. Las condiciones de pelletización pueden incluir aplicación de vapor. Las condiciones de pelletización pueden comprender aplicaciones a temperatura por encima de aproximadamente 80°C durante aproximadamente 5 minutos y la enzima retiene una actividad específica de al menos 350 hasta aproximadamente 900 unidades por miligramo de enzima.

La invención describe métodos para tratar, por ejemplo, mejorar textura y sabor de un producto lácteo que comprenden: (a) proveer un polipéptido de la invención que tiene una actividad de transferasa, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidoreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, o una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidoreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa, codificada por un ácido nucleico de la invención; (b) proveer un producto lácteo; y (c) poner en contacto el polipéptido de la etapa (a) y el producto lácteo de la etapa (b) bajo condiciones en donde la transferasa, por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidoreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa puede tratar, como por ejemplo, mejorar la textura o sabor del producto lácteo. En un aspecto, el producto lácteo comprende un queso o un yogurt. La invención provee productos lácteos que comprenden una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidoreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa de la invención, o es codificado por un ácido nucleico de la invención.

La invención describe métodos para mejorar la extracción de aceite a partir de un material vegetal rico en aceite que comprenden: (a) proveer un polipéptido de la invención que tiene una actividad de transferasa, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidoreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, o una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidoreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa codificadas por un ácido nucleico de la

invención; (b) proveer un material vegetal, rico en aceite; y (c) poner en contacto el polipéptido de la etapa (a) y el material vegetal rico en aceite. En un aspecto, el material vegetal rico en aceite comprende una semilla rica en aceite. El aceite puede ser un aceite de soja, un aceite de oliva, un aceite de colza (canola) o un aceite de girasol.

La invención describe métodos para preparar un zumo, jarabe, puré o extracto de fruta o vegetal que comprende: (a) proveer el polipéptido de la invención que tiene una actividad de transferasa, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, o una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa codificadas por un ácido nucleico de la invención; (b) proveer una composición o un líquido que comprende un material de fruta o vegetal; y (c) poner en contacto el polipéptido de la etapa (a) y la composición, preparando por lo tanto el zumo, jarabe, puré o extracto de fruta o vegetal.

La invención describe métodos para tratar una madera, un producto de madera, un papel, un producto de papel, una pulpa, un producto de pulpa, un residuo de papel o una composición de reciclaje de papel que comprenden: (a) proveer un polipéptido de la invención que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, o una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa codificadas por un ácido nucleico de la invención; (b) proveer una composición que comprende una madera, un producto de madera, un papel, un producto de papel, una pulpa, un producto de pulpa, un residuo de papel o una composición de reciclaje de papel; y (c) poner en contacto el polipéptido de la etapa (a) y la composición, tratando por lo tanto la madera, productos de madera, papel, productos de papel, pulpa, producto de pulpa, residuo de papel o composición de reciclaje de papel. En un aspecto de la invención, el tratamiento comprende reducir o solubilizar la lignina (deslignificación), blanqueo o decoloración, y/o destintado.

La invención describe papeles o productos de papel o pulpa de papel que comprenden una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa de la invención, o un polipéptido codificado por un ácido nucleico de la invención. La invención describe métodos para tratar un papel o una pulpa de papel o madera que comprenden: (a) proveer un polipéptido de la invención que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, o una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa codificadas por un ácido nucleico de la invención; (b) proveer una composición que comprende papel o una pulpa de papel o madera; y (c) poner en contacto el polipéptido de la etapa (a) y la composición de la etapa (b) bajo condiciones en donde la transferasa, por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidorreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa pueden tratar el papel o pulpa de papel o madera.

La invención describe métodos para blanquear un hilado, tela, fibra, vestimenta o textil que comprende poner en contacto la tela, hebra, vestimenta o textil con una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa en condiciones adecuadas para producir un blanqueamiento del textil, en donde la transferasa, por ejemplo, por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidorreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa comprende un polipéptido de la invención, o un fragmento enzimáticamente activo del mismo. El hilado, tela, fibra, vestimenta o textil puede comprender un hilado, tela, fibra, vestimenta o textil celulósico sin algodón. La invención describe telas, fibras, vestimentas o textiles que comprenden un polipéptido que tiene una secuencia de la invención, o un polipéptido codificado por un ácido nucleico que comprende una secuencia de la invención, o un fragmento enzimáticamente activo del mismo, en donde en un aspecto (opcionalmente) la tela, fibra, vestimenta o textil comprenden una tela, fibra, vestimenta o textil celulósicos sin algodón.

La invención describe madera, trozos de madera, pulpa de madera, productos de madera, pulpas de papel, productos de papel, periódicos o residuos de papel que comprenden un polipéptido de la invención, o un fragmento enzimáticamente activo del mismo. La invención describe un hilado, tela, fibra, vestimenta o textil que comprende un polipéptido de la invención, o un fragmento enzimáticamente activo del mismo.

La invención describe métodos para producir etanol que comprenden poner en contacto un material orgánico, por ejemplo, una biomasa, con un polipéptido que tiene una actividad de transferasa por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, en donde el

polipéptido tiene una secuencia de la invención, o el polipéptido es codificado por un ácido nucleico que comprende una secuencia de la invención, o un fragmento enzimáticamente activo del mismo. La invención describe composiciones que comprenden un etanol y un polipéptido que tiene una actividad de transferasa por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, en donde el polipéptido tiene una secuencia de la invención, o el polipéptido es codificado por un ácido nucleico que comprende una secuencia de la invención, o un fragmento enzimáticamente activo del mismo. La invención describe métodos para hacer etanol que comprenden: (a) proveer al menos un polipéptido que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, o un fragmento enzimáticamente activo del mismo; (b) proveer una composición orgánica; y (c) poner contacto la composición de la etapa (b) con el polipéptido de la etapa (a).

La invención describe composiciones farmacéuticas que comprenden un polipéptido que tiene una actividad de transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, en donde el polipéptido comprende una secuencia de la invención, o el polipéptido es codificado por un ácido nucleico que comprende una secuencia de la invención, o un fragmento enzimáticamente activo del mismo. En un aspecto, la composición farmacéutica actúa como una auxiliar digestivo o es usado para el diagnóstico, seguimiento o tratamiento de cualquier condición y/o enfermedad, por ejemplo, daño/enfermedad del hígado o infartos del miocardio. En un aspecto, el tratamiento es profiláctico.

En un aspecto, la invención describe productos de cuidado oral que comprenden un polipéptido de la invención que tiene una actividad de transferasa por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, o una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa, codificada por un ácido nucleico de la invención. El producto de cuidado oral puede comprender pasta dental, una crema dental, un gel o un polvo dental, un odontico, un enjuague bucal, una formulación de enjuague pre o postcepillado, una goma de mascar, una tableta o una golosina. La invención describe composiciones para limpieza de lentes de contacto que comprenden un polipéptido de la invención que tiene una por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa, o a transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa codificada por un ácido nucleico de la invención.

La invención describe transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas quiméricas que comprenden una secuencia de polipéptidos de la invención y al menos un dominio de enlazamiento heterólogo, en donde en un aspecto (opcionalmente), el dominio de enlazamiento comprende un dominio de enlazamiento de NAD, NAD(P), calcio, tiamina, FAD, zinc, ADN y/o lipoilo. La invención describe métodos para diseñar una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa quimérica que tiene una nueva especificidad de enlazamiento o una especificidad de enlazamiento potenciada, que comprende insertar un heterólogo o un dominio de enlazamiento endógeno adicional en una transferasa por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa, en donde el dominio de enlazamiento comprende una subsecuencia de enlazamiento de una secuencia de una transferasa por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa de la invención, o alternativamente un dominio de enlazamiento heterólogo, o un dominio de enlazamiento endógeno adicional, es insertado en una secuencia de transferasa por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa de la invención.

La invención describe mezclas de enzimas, o "cócteles" que comprenden al menos una enzima de la invención y una o más otras enzimas, las cuales pueden ser otra transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, o cualquier otra enzima; Por ejemplo, los "cócteles" de la invención, además de al menos una enzima de la invención, pueden comprender cualquier otra enzima, tal como xilanasas, celulasas, lipasas, esterasas, proteasas, o endoglicosidasas, endo-Beta-1, 4-glucanasas, beta-glucanasas, endo-beta-1,3 (4)-glucanasas, cutinasas, peroxidadas, catalasas, lacasas, amilasas, glucoamilasas, pectinasas, racemasas, isomerasas, epimerasas, deshidrogenasas, oxidorreductasas, reductasas, oxidadas, fenoloxidasas, ligninasas, pululaninas, arabinanasas, hemicelulasas, mananasas, xiloglucanasas, pectina acetil esterasas, ramnogalacturonan acetil esterasas, poligalacturonasas, ramnogalacturonasas, galactanasas, pectina liasas, pectina metilesterasas, celobiohidrolasas y/o transglutaminasas, para nombrar sólo unos pocos ejemplos. En realizaciones alternativas,

estas mezclas de enzimas o "cócteles" que comprenden al menos una enzima de la invención pueden ser utilizadas en cualquier proceso o método de la invención, o composición de la invención, por ejemplo, en alimentos o piensos, suplementos para alimentos o piensos, textiles, papeles, maderas procesadas, etc., y métodos para hacerlos, y en composiciones y métodos para tratar papel, pulpa, madera, papel, residuos o subproductos de pulpa o madera, y similares, y en los productos finales de los mismos.

La invención provee métodos para hacer un piruvato y/o un D-glutamato que comprende (a) proveer una D-alanina y un 2-oxoglutarato; (b) proveer el polipéptido de la invención como se definió anteriormente; y (c) poner en contacto el polipéptido de (b) con la D-alanina + 2-oxoglutarato bajo condiciones en donde el polipéptido cataliza la reacción D-alanina + 2-oxoglutarato \rightleftharpoons piruvato + D-glutamato.

La invención provee métodos para hacer un 2-oxo ácido que comprende (i) (a) proveer un D-aminoácido + H₂O + un aceptor; (b) proveer el polipéptido de transaminasa de la invención como se definió anteriormente; y (c) poner en contacto el polipéptido de (b) con el D-aminoácido + H₂O + aceptor, bajo condiciones en donde el polipéptido cataliza la reacción: D-amino ácido + H₂O + aceptor \rightleftharpoons un 2-oxo ácido + NH₃ + aceptor reducido; o (ii) el método de (i), en donde el aceptor es una benzoquinona, es (iii) el método de (ii) en el donde la benzoquinona es un 1,2-benzoquinona o una 1,4-benzoquinona, o ubiquinona, ubidecarenona o coenzima Q.

La invención provee métodos para transferir un grupo amino de un aminoácido a un alfa-cetoácido que comprende (i) (a) proveer un aminoácido; (b) proveer el polipéptido de transaminasa de la invención como se definió más arriba; y (c) poner en contacto el polipéptido de (b) con el aminoácido bajo condiciones en donde el polipéptido cataliza la conversión del aminoácido a un alfa-cetoácido; o (ii) el método de (i), en donde la actividad de transaminasa comprende catalizar la conversión de una mezcla racémica de aminoácidos a un alfa-cetoácido sustancialmente puro ópticamente.

La invención provee métodos para hacer un aminoácido a partir de un alfa-cetoácido que comprenden (i) (a) proveer un alfa-cetoácido (b) proveer el polipéptido de transaminasa de la invención como se definió anteriormente; y (c) poner en contacto el polipéptido de (b) con el alfa-cetoácido bajo condiciones en donde el polipéptido cataliza la conversión del alfa-cetoácido a un aminoácido; (ii) el método de (i), en donde la actividad de transaminasa comprende catalizar la conversión de una mezcla racémica alfa-ceto a un aminoácido D- o L- sustancialmente puro ópticamente; o (iii) el método de (i) o (ii), en donde el oxaloacetato es convertido a un aspartato, o alfa-cetoglutarato es convertido a glutarato, o un alfa-cetoisovalerato es convertido a un L-valina; o la actividad de transaminasa es una actividad de transaminasa omega que cataliza la conversión de isobutilamina a isobutiraldehído.

La invención provee métodos para catalizar la conversión de una amina a una cetona que comprende (i) (a) proveer una amina; (b) proveer el polipéptido de transaminasa de la invención como se definió más arriba; y (c) poner en contacto el polipéptido de (b) con la amina bajo condiciones en donde el polipéptido cataliza la conversión de una amina a una cetona, en donde la amina viene o no viene de un triptófano (con la condición de que el segundo aminoácido no es triptófano, o con la condición de que la amina no está en o proviene un triptófano); (ii) el método de (i), en donde la actividad de la transaminasa comprende catalizar la conversión de una amina quiral a una cetona; o (iii) el método de (i) o (ii), en donde la amina es una ω -amina.

La invención provee métodos para catalizar la síntesis de un aminoácido que comprende (i) (a) proveer un aminoácido y un cetoácido, en donde el aminoácido no es triptófano; (b) proveer el polipéptido de transaminasa de la invención tal como se definió más arriba; y (c) poner en contacto el polipéptido de (b) con el aminoácido y el cetoácido bajo condiciones en donde se producen un segundo aminoácido y un piruvato, en donde el segundo aminoácido no es triptófano (con la condición de que el segundo aminoácido no es triptófano); o (ii) el método de (i), que comprende adicionalmente hacer reaccionar el piruvato con una enzima acetolactato sintasa bajo condiciones apropiadas para producir un compuesto que no reacciona con la enzima transaminasa; (iii) el método de (ii), en donde el compuesto que no reacciona con la enzima transaminasa es acetolactato o acetoina; o, el primer aminoácido es alanina o L-aspartato; o, el cetoácido es 2-cetobutirato o tri-metil piruvato; o, el segundo aminoácido es 2-aminobutirato o tert-leucina.

Los detalles de una o más realizaciones de la invención están establecidos en los dibujos acompañantes y en la descripción que sigue. Otras características, objetos y ventajas de la invención serán evidentes a partir de la descripción y dibujos, y de las reivindicaciones.

Breve descripción de los dibujos

Los siguientes dibujos son ilustrativos de aspectos de la invención y no pretenden limitar el alcance de la invención tal como es abarcado por las reivindicaciones.

La Figura 1 es un diagrama de bloques de un sistema de ordenador.

La Figura 2 es un diagrama de flujo que ilustra un aspecto de un proceso para comparar una nueva secuencia de nucleótidos o proteínas con una base de datos de secuencias con el fin de determinar los niveles de homología entre la nueva secuencia y las secuencias en la base de datos.

5 La Figura 3 es un diagrama de flujo que ilustra un aspecto de un proceso en un ordenador para determinar si dos secuencias son homólogas.

La Figura 4 es un diagrama de flujo que ilustra un aspecto del proceso identificador 300 para detectar la presencia de una característica en una secuencia.

10 La Figura 5 ilustra una comparación de secuencias de esta invención, que muestra las regiones de consenso de la SEQ ID NO: 894 como proteínas, tal como está resaltado en la figura, como se describe en detalle en el Ejemplo 9, más adelante.

La Figura 6 ilustra una comparación de la secuencia de la enzima de esta invención SEQ ID NO: 910, con otros DATS publicados, y como está resaltado en la figura, pueden verse los residuos que hacen que esta enzima sea única y puede lograr su actividad superior, como se describe en detalle en el Ejemplo 10, más adelante.

15 La Figura 7 ilustra una comparación de D-aminotransferasas relacionadas y de los motivos de secuencias de núcleos que tienen en común, tal como se describe en detalle en el Ejemplo 14, más adelante.

La Figura 8 es un modelo de 3DAA-D-amino ácido aminotransferasa, con residuos numerados que indican aquellos sitios seleccionados para evolución de TMCASM, tal como se describe en detalle en el Ejemplo 29, más adelante.

Símbolos de referencias similares en los diversos dibujos indican elementos similares

Descripción detallada de la invención

20 La invención provee d-aminoácidos transferasas (también denominadas como "D-aminotransferasas" o "D-ATs" o "DATs"), y polinucleótidos que los codifican y métodos para prepararlos y utilizarlos. Las D-aminoácido transferasas, de los polipéptidos de la invención abarcan enzimas que tienen actividad de d-aminoácido transferasa y/o catalizan la transferencia de un grupo químico, catalizan la transaminación, catalizan la reacción: D-alanina + 2-oxoglutarato
25 \rightleftharpoons piruvato + D-glutamato, y/o catalizan una reacción de oxidación-reducción, catalizan la remoción de átomos de hidrógeno, y/o catalizan la reacción: D-aminoácido + H₂O + aceptor \rightleftharpoons un 2-oxo ácido + NH₃ + aceptor reducido. Las transferasas de la invención pueden ser utilizadas para hacer y/o procesar enzimas peptídicas de alimentos y aditivos de alimentarios, bebidas y aditivos de bebidas, piensos y aditivos para piensos, suplementos dietarios y similares.

30 En un aspecto, una enzima de la invención es termotolerante y/o tolerante de condiciones de alto y/o bajo pH. Por ejemplo, en un aspecto, una d-aminoácido transferasa de la invención retiene su actividad bajo condiciones que comprenden una temperatura de al menos aproximadamente 80°C, 85°C, 86°C, 87°C, 88°C, 89°C, 90°C, 91°C, 92°C, 93°C, 94°C, 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 103.5°C, 104°C, 105°C, 107°C, 108°C, 109°C o 110°C, o más, y un pH básico de al menos pH 11, o más.

35 La invención provee ácidos nucleicos aislados, sintéticos o recombinantes que comprenden un ácido nucleico que codifica al menos un polipéptido que tiene una actividad de d-aminoácido transferasa como se describió aquí, en donde el ácido nucleico comprende una secuencia que tiene al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia (homología) con SEQ ID NO: 219 y tal como se describe aquí y en las Tablas 1, 2 y 3, y en el Listado de Secuencias a lo largo de la longitud completa del ADNc, transcripto (ARNm) o gen. Los ácidos nucleicos de la invención incluyen aquellos que codifican un polipéptido de la invención, que tiene al
40 menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con el polipéptido de SEQ ID NO: 220, SEQ ID NO: 220 con una, varias o todas las modificaciones de la Tabla 46 o de la Tabla 55.

Las Tablas 1, 2 y 3, más adelante, son cuadros que describen características seleccionadas de ácidos nucleicos y polipéptidos de ejemplos de la invención, que incluyen comparación de la identidad de secuencia de las secuencias de ejemplo frente a bases de datos públicas.

45 La Tabla 1, mas adelante, describe la actividad asignada (tal como fue determinada por datos experimentales, véanse Ejemplos 1 a 23 de los polipéptidos de ejemplos (codificados por los polinucleótidos de ejemplos) de la invención. La Tabla 1 indica adicionalmente si el polinucleótido (que codifica un polipéptido) de la invención es un clon (una secuencia genómica aislada de la fuente original, tal como se describe en la Tabla 2) o es un subclón (donde el clon es manipulado, por ejemplo, por remoción de una secuencia de señalización natural, adición de una metionina de inicio, adición de una etiqueta, etc.). La Tabla 1 también indica la relación entre el clon y subclón, por
50 ejemplo, cuál subclón fue derivado a partir de cuál clon. Para ayudar en la lectura de la Tabla 1, por ejemplo, las columnas 1 y 4, filas 1 y 2, indican que SEQ ID NO: 32 (codificada por SEQ ID NO: 31) es un clon siendo el subclón correspondiente SEQ ID NO: 868 (codificada por SEQ ID NO: 867), denotado como "par 1 Clon/subclón".

La Tabla 2, mas abajo, indica la fuente de la cual fueron derivados primeramente los ácidos nucleicos y polipéptidos de ejemplo de la invención. La Tabla 2, más abajo, también indica el "sitio de escisión Signalp" para la secuencia de señales de la enzima de ejemplo (o "péptido de señalización", o SP), tal como fue determinado mediante el paradigma SignalP, tal como se discute más abajo (véase Nielsen (1997), infra); la "Secuencia de señalización predicha" es listada a partir del amino terminal hacia el carboxi terminal, por ejemplo, para el polipéptido SEQ ID NO: 258, el péptido de señalización es "MKSAIVLGAGMVGIAAVHL".

La Tabla 3, más abajo, describe características seleccionadas de ácidos nucleicos y polipéptidos de ejemplo de la invención, incluyendo la comparación de identidad de secuencia de las secuencias de ejemplo frente a bases de datos públicas. Para ayudar adicionalmente en la lectura de la Tabla 3, por ejemplo, en la primera fila, marcada "SEQ ID NO:", los números "1, 2" representan el polipéptido de ejemplo de la invención que tiene una secuencia tal como se define en SEQ ID NO: 2, codificada por, por ejemplo, SEQ ID NO: 1. Todas las secuencias descritas en la Tabla 2 (todas las secuencias de ejemplo de la invención) han sido sometidas a la búsqueda BLAST (tal como se describe en detalle, mas adelante) contra dos conjuntos de bases de datos. El primer conjunto de bases de datos está disponible a través del NCBI (National Center for Biotechnology Information). Todos los resultados de búsquedas contra estas bases de datos se encuentran en las columnas tituladas "Descripción NR", "Código de acceso NR", "NR Evalúe" o "organismo NR". "NR" se refiere a la base de datos de nucleótidos No Redundante mantenida por el NCBI. Esta base de datos está compuesta de GenBank, actualizaciones de GenBank, y actualizaciones de EMBL. Las entradas en la columna "Descripción NR" se refieren a la línea de definición en cualquier registro de NCBI, el cual incluye una descripción de la secuencia, tal como el organismo fuente, nombre del gen/nombre de la proteína, o alguna descripción de la función de la secuencia. Las entradas en la columna "Código de acceso NR" se refieren a un identificador único dado a un registro de secuencia. Las entradas en la columna "NR Evalúe" se refieren al valor esperado (Evalúe), el cual representa la probabilidad de un marcador de alineamiento tan bueno como el encontrado entre las secuencia de búsqueda (las secuencias de la invención) y una secuencia de las bases de datos encontraría en el mismo número de comparaciones entre secuencias aleatorias tal como se hace en la actual búsqueda BLAST. Las entradas en la columna "Organismo NR" se refieren al organismo fuente de la secuencia identificada como el acierto BLAST más cercano. El segundo conjunto de bases de datos es conocido colectivamente como las base de datos GENESEQ™, el cual está disponible a través de Thomson Derwent (Philadelphia, PA). Todos los resultados de las búsquedas contra esta base de datos se encuentran en las columnas tituladas "Descripción de Proteína GENESEQ™", "Código de Acceso de Proteína GENESEQ™", "Evalúe", "Descripción de ADN GENESEQ™", "Código de acceso de ADN GENESEQ™" o "Evalúe". La información encontrada en estas columnas es comparable con la información encontrada en las columnas NR descritas más arriba, excepto en que fue derivado a partir de búsquedas BLAST contra la base de datos GENESEQ™ en vez de las bases de datos del NCBI. Además, esta tabla incluye la columna "No. EC Predicho". Un número EC es el número asignado a un tipo de enzima de acuerdo con un esquema de nomenclatura de enzimas estandarizado desarrollado por la Enzyme Commission of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB). Los resultados en la columna "No. EC Predicho" está determinado mediante una búsqueda BLAST contra la base de datos Kegg (Kyoto Encyclopedia of Genes and Genomes). Si la coincidencia BLAST máxima tiene un Evalúe igual o menor a e^{-6} , el Número EC asignado a la coincidencia máxima es introducido en la tabla. El número EC del acierto máximo es utilizado como una guía hacia qué número EC de la secuencia de la invención podría ser. Las columnas "Longitud de ADN Buscado" y "Longitud de Proteína Buscada" se refiere al número de nucleótidos o el número de aminoácidos, respectivamente, en la secuencia de la invención que fue buscada o perseguida contra las bases de datos NCBI o GENESEQ™. Las columnas "Longitud de ADN Sujeto" y "Longitud de Proteína Sujeto" se refieren al número de nucleótidos o de aminoácidos, respectivamente, en la secuencia de las coincidencias máximas a partir de las búsquedas BLAST. Los resultados provistos en estas columnas provienen de la búsqueda que generó el Evalúe menor, bien sea partir de las bases de datos NCBI o de la base de datos Geneseq. Las columnas "% ID Proteína" y "% ID ADN" se refieren al porcentaje de identidad de secuencia entre la secuencia de la invención y la secuencia de la coincidencia máxima BLAST. Los resultados provistos en estas columnas son de la búsqueda que produjeron el Evalúe inferior, bien sea a partir de las bases de datos NCBI o de la base de datos GENESEQ™.

Tabla 1

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
1	31, 32	D-AT	Clon
1	867, 868	D-AT	Subclón
2	955, 956	D-AT	Clon
2	929, 930	D-AT	Subclón
3	957, 958	D-AT	Clon
3	931, 932	D-AT	Subclón

ES 2 531 290 T3

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
4	959, 960	D-AT	Clon
4	935, 936	D-AT	Subclón
5	41, 42	D-AT	Clon
5	869, 870	D-AT	Subclón
6	7, 8	D-AT	Clon
6	943, 944	D-AT	Subclón
7	11, 12	D-AT	Clon
7	941, 942	D-AT	Subclón
8	83, 84	D-AT	Clon
8	879, 880	D-AT	Subclón
9	151, 152	D-AT	Clon
9	913, 914	D-AT	Subclón
10	951, 952	D-AT	Clon
10	933, 934	D-AT	Subclón
11	75, 76	D-AT	Clon
11	881, 882	D-AT	Subclón
12	87, 88	D-AT	Clon
12	883, 884	D-AT	Subclón
13	163, 164	D-AT	Clon
13	921, 922	D-AT	Subclón
14	145, 146	D-AT	Clon
14	919, 920	D-AT	Subclón
15	149, 150	D-AT	Clon
15	925, 926	D-AT	Subclón
16	147, 148	D-AT	Clon
16	915, 916	D-AT	Subclón
17	15, 16	D-AT	Clon
17	947, 948	D-AT	Subclón
18	17, 18	D-AT	Clon
18	949, 950	D-AT	Subclón
19	3, 4	D-AT	Clon
19	937, 938	D-AT	Subclón
20	5,6	D-AT	Clon
20	939, 940	D-AT	Subclón
21	161, 162	D-AT	Clon
21	923, 924	D-AT	Subclón
22	953, 954	D-AT	Clon
22	927, 928	D-AT	Subclón

ES 2 531 290 T3

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
23	19, 20	D-AT	Clon
23	885, 886	D-AT	Subclón
24	21, 22	D-AT	Clon
24	891, 892	D-AT	Subclón
25	23, 24	D-AT	Clon
25	893, 894	D-AT	Subclón
26	13, 14	D-AT	Clon
26	945, 946	D-AT	Subclón
27	143, 144	D-AT	Clon
27	917, 918	D-AT	Subclón
28	43, 44	D-AT	Clon
28	871, 872	D-AT	Subclón
29	45, 46	D-AT	Clon
29	873, 874	D-AT	Subclón
30	49, 50	D-AT	Clon
30	897, 898	D-AT	Subclón
31	51, 52	D-AT	Clon
31	875, 876	D-AT	Subclón
32	37, 38	D-AT	Clon
32	877, 878	D-AT	Subclón
33	25, 26	D-AT	Clon
33	889, 890	D-AT	Subclón
34	27, 28	D-AT	Clon
34	887, 888	D-AT	Subclón
35	131, 132	D-AT	Clon
35	909, 910	D-AT	Subclón
36	53, 54	D-AT	Clon
36	865, 866	D-AT	Subclón
37	29, 30	D-AT	Clon
37	895, 896	D-AT	Subclón
38	125, 126	D-AT	Clon
38	907, 908	D-AT	Subclón
39	133, 134	D-AT	Clon
39	911, 912	D-AT	Subclón
40	127, 128	D-AT	Clon
40	899, 900	D-AT	Subclón
41	137, 138	D-AT	Clon
41	901, 902	D-AT	Subclón

ES 2 531 290 T3

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
42	139, 140	D-AT	Clon
42	903, 904	D-AT	Subclón
43	129, 130	D-AT	Clon
43	905, 906	D-AT	Subclón
44	33, 34	D-AT	Clon
44	969, 970	D-AT	Subclón
45	219, 220	D-AT	Clon
45	973, 974	D-AT	Subclón
46	39, 40	D-AT	Clon
46	971, 972	D-AT	Subclón
47	1, 2	D-AT	Clon
47	975, 976	D-AT	Subclón
48	253, 254	Deshidrogenasa	Clon
48	961, 962	Deshidrogenasa	Subclón
48	963, 964	Deshidrogenasa	Subclón
48	965, 966	Deshidrogenasa	Subclón
48	967, 968	Deshidrogenasa	Subclón
	35, 36	D-AT	Clon
	9, 10	D-AT	Clon
	85, 86	D-AT	Clon
	77, 78	D-AT	Clon
	153, 154	D-AT	Clon
	155, 156	D-AT	Clon
	201, 202	D-AT	Clon
	221, 222	D-AT	Clon
	235, 236	D-AT	Clon
	203, 204	D-AT	Clon
	237, 238	D-AT	Clon
	239, 240	D-AT	Clon
	159, 160	D-AT	Clon
	165, 166	D-AT	Clon
	211, 212	D-AT	Clon
	249, 250	D-AT	Clon
	177, 178	D-AT	Clon
	223, 224	D-AT	Clon
	169, 170	D-AT	Clon
	179, 180	D-AT	Clon
	181, 182	D-AT	Clon

ES 2 531 290 T3

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
	63, 64	D-AT	Clon
	107, 108	D-AT	Clon
	109, 110	D-AT	Clon
	111, 112	D-AT	Clon
	113, 114	D-AT	Clon
	115, 116	D-AT	Clon
	123, 124	D-AT	Clon
	225, 226	D-AT	Clon
	227, 228	D-AT	Clon
	247, 248	D-AT	Clon
	217, 218	D-AT	Clon
	205, 206	D-AT	Clon
	183, 184	D-AT	Clon
	185, 186	D-AT	Clon
	241, 242	D-AT	Clon
	243, 244	D-AT	Clon
	229, 230	D-AT	Clon
	231, 232	D-AT	Clon
	187, 188	D-AT	Clon
	189, 190	D-AT	Clon
	191, 192	D-AT	Clon
	207, 208	D-AT	Clon
	99, 100	D-AT	Clon
	55, 56	D-AT	Clon
	57, 58	D-AT	Clon
	193, 194	D-AT	Clon
	233, 234	D-AT	Clon
	215, 216	D-AT	Clon
	195, 196	D-AT	Clon
	199, 200	D-AT	Clon
	197, 198	D-AT	Clon
	209, 210	D-AT	Clon
	141, 142	D-AT	Clon
	157, 158	D-AT	Clon
	245, 246	D-AT	Clon
	59, 60	D-AT	Clon
	61, 62	D-AT	Clon
	47, 48	D-AT	Clon

ES 2 531 290 T3

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
	213, 214	D-AT	Clon
	171, 172	D-AT	Clon
	167, 168	D-AT	Clon
	173, 174	D-AT	Clon
	175, 176	D-AT	Clon
	65, 66	D-AT	Clon
	67, 68	D-AT	Clon
	69, 70	D-AT	Clon
	71, 72	D-AT	Clon
	73, 74	D-AT	Clon
	79, 80	D-AT	Clon
	81, 82	D-AT	Clon
	93, 94	D-AT	Clon
	91, 92	D-AT	Clon
	95, 96	D-AT	Clon
	97, 98	D-AT	Clon
	117, 118	D-AT	Clon
	119, 120	D-AT	Clon
	121, 122	D-AT	Clon
	101, 102	D-AT	Clon
	103, 104	D-AT	Clon
	105, 106	D-AT	Clon
	89, 90	D-AT	Clon
	135, 136	D-AT	Clon
	259, 260	Deshidrogenasa	Clon
	261, 262	Deshidrogenasa	Clon
	263, 264	Deshidrogenasa	Clon
	327, 328	Deshidrogenasa	Clon
	335, 336	Deshidrogenasa	Clon
	353, 354	Deshidrogenasa	Clon
	355, 356	Deshidrogenasa	Clon
	321, 322	Deshidrogenasa	Clon
	341, 342	Deshidrogenasa	Clon
	265, 266	Deshidrogenasa	Clon
	287, 288	Deshidrogenasa	Clon
	267, 268	Deshidrogenasa	Clon
	269, 270	Deshidrogenasa	Clon
	301, 302	Deshidrogenasa	Clon

ES 2 531 290 T3

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
	413, 414	Deshidrogenasa	Clon
	433, 434	Deshidrogenasa	Clon
	423, 424	Deshidrogenasa	Clon
	303, 304	Deshidrogenasa	Clon
	425, 426	Deshidrogenasa	Clon
	393, 394	Deshidrogenasa	Clon
	297, 298	Deshidrogenasa	Clon
	299, 300	Deshidrogenasa	Clon
	567, 568	Deshidrogenasa	Clon
	515, 516	Deshidrogenasa	Clon
	465, 466	Deshidrogenasa	Clon
	387, 388	Deshidrogenasa	Clon
	409, 410	Deshidrogenasa	Clon
	411, 412	Deshidrogenasa	Clon
	375, 376	Deshidrogenasa	Clon
	407, 408	Deshidrogenasa	Clon
	391, 392	Deshidrogenasa	Clon
	485, 486	Deshidrogenasa	Clon
	603, 604	Deshidrogenasa	Clon
	605, 606	Deshidrogenasa	Clon
	517, 518	Deshidrogenasa	Clon
	543, 544	Deshidrogenasa	Clon
	429, 430	Deshidrogenasa	Clon
	443, 444	Deshidrogenasa	Clon
	365, 366	Deshidrogenasa	Clon
	445, 446	Deshidrogenasa	Clon
	431, 432	Deshidrogenasa	Clon
	449, 450	Deshidrogenasa	Clon
	467, 468	Deshidrogenasa	Clon
	379, 380	Deshidrogenasa	Clon
	367, 368	Deshidrogenasa	Clon
	405, 406	Deshidrogenasa	Clon
	383, 384	Deshidrogenasa	Clon
	357, 358	Deshidrogenasa	Clon
	415, 416	Deshidrogenasa	Clon
	395, 396	Deshidrogenasa	Clon
	385, 386	Deshidrogenasa	Clon
	369, 370	Deshidrogenasa	Clon

ES 2 531 290 T3

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
	397, 398	Deshidrogenasa	Clon
	435, 436	Deshidrogenasa	Clon
	453, 454	Deshidrogenasa	Clon
	469, 470	Deshidrogenasa	Clon
	447, 448	Deshidrogenasa	Clon
	473, 474	Deshidrogenasa	Clon
	389, 390	Deshidrogenasa	Clon
	427, 428	Deshidrogenasa	Clon
	451, 452	Deshidrogenasa	Clon
	399, 400	Deshidrogenasa	Clon
	455, 456	Deshidrogenasa	Clon
	417, 418	Deshidrogenasa	Clon
	403, 404	Deshidrogenasa	Clon
	419, 420	Deshidrogenasa	Clon
	251, 252	Deshidrogenasa	Clon
	371, 372	Deshidrogenasa	Clon
	475, 476	Deshidrogenasa	Clon
	457, 458	Deshidrogenasa	Clon
	459, 460	Deshidrogenasa	Clon
	461, 462	Deshidrogenasa	Clon
	463, 464	Deshidrogenasa	Clon
	477, 478	Deshidrogenasa	Clon
	479, 480	Deshidrogenasa	Clon
	481, 482	Deshidrogenasa	Clon
	629, 630	Deshidrogenasa	Clon
	519, 520	Deshidrogenasa	Clon
	521, 522	Deshidrogenasa	Clon
	589, 590	Deshidrogenasa	Clon
	359, 360	Deshidrogenasa	Clon
	361, 362	Deshidrogenasa	Clon
	381, 382	Deshidrogenasa	Clon
	363, 364	Deshidrogenasa	Clon
	625, 626	Deshidrogenasa	Clon
	549, 550	Deshidrogenasa	Clon
	551, 552	Deshidrogenasa	Clon
	501, 502	Deshidrogenasa	Clon
	571, 572	Deshidrogenasa	Clon
	601, 602	Deshidrogenasa	Clon

ES 2 531 290 T3

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
	573, 574	Deshidrogenasa	Clon
	575, 576	Deshidrogenasa	Clon
	577, 578	Deshidrogenasa	Clon
	611, 612	Deshidrogenasa	Clon
	579, 580	Deshidrogenasa	Clon
	523, 524	Deshidrogenasa	Clon
	553, 554	Deshidrogenasa	Clon
	503, 504	Deshidrogenasa	Clon
	505, 506	Deshidrogenasa	Clon
	507, 508	Deshidrogenasa	Clon
	509, 510	Deshidrogenasa	Clon
	525, 526	Deshidrogenasa	Clon
	373, 374	Deshidrogenasa	Clon
	421, 422	Deshidrogenasa	Clon
	483, 484	Deshidrogenasa	Clon
	511, 512	Deshidrogenasa	Clon
	527, 528	Deshidrogenasa	Clon
	555, 556	Deshidrogenasa	Clon
	529, 530	Deshidrogenasa	Clon
	591, 592	Deshidrogenasa	Clon
	557, 558	Deshidrogenasa	Clon
	559, 560	Deshidrogenasa	Clon
	593, 594	Deshidrogenasa	Clon
	581, 582	Deshidrogenasa	Clon
	613, 614	Deshidrogenasa	Clon
	595, 596	Deshidrogenasa	Clon
	597, 598	Deshidrogenasa	Clon
	599, 600	Deshidrogenasa	Clon
	583, 584	Deshidrogenasa	Clon
	615, 616	Deshidrogenasa	Clon
	619, 620	Deshidrogenasa	Clon
	621, 622	Deshidrogenasa	Clon
	561, 562	Deshidrogenasa	Clon
	563, 564	Deshidrogenasa	Clon
	587, 588	Deshidrogenasa	Clon
	513, 514	Deshidrogenasa	Clon
	531, 532	Deshidrogenasa	Clon
	533, 534	Deshidrogenasa	Clon

ES 2 531 290 T3

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
	535, 536	Deshidrogenasa	Clon
	585, 586	Deshidrogenasa	Clon
	617, 618	Deshidrogenasa	Clon
	627, 628	Deshidrogenasa	Clon
	623, 624	Deshidrogenasa	Clon
	537, 538	Deshidrogenasa	Clon
	565, 566	Deshidrogenasa	Clon
	539, 540	Deshidrogenasa	Clon
	541, 542	Deshidrogenasa	Clon
	569, 570	Deshidrogenasa	Clon
	607, 608	Deshidrogenasa	Clon
	609, 610	Deshidrogenasa	Clon
	487, 488	Deshidrogenasa	Clon
	545, 546	Deshidrogenasa	Clon
	495, 496	Deshidrogenasa	Clon
	497, 498	Deshidrogenasa	Clon
	499, 500	Deshidrogenasa	Clon
	489, 490	Deshidrogenasa	Clon
	491, 492	Deshidrogenasa	Clon
	493, 494	Deshidrogenasa	Clon
	547, 548	Deshidrogenasa	Clon
	401, 402	Deshidrogenasa	Clon
	377, 378	Deshidrogenasa	Clon
	437, 438	Deshidrogenasa	Clon
	439, 440	Deshidrogenasa	Clon
	441, 442	Deshidrogenasa	Clon
	271, 272	Deshidrogenasa	Clon
	273, 274	Deshidrogenasa	Clon
	293, 294	Deshidrogenasa	Clon
	275, 276	Deshidrogenasa	Clon
	277, 278	Deshidrogenasa	Clon
	279, 280	Deshidrogenasa	Clon
	281, 282	Deshidrogenasa	Clon
	283, 284	Deshidrogenasa	Clon
	285, 286	Deshidrogenasa	Clon
	291, 292	Deshidrogenasa	Clon
	289, 290	Deshidrogenasa	Clon
	255, 256	Deshidrogenasa	Clon

ES 2 531 290 T3

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
	295, 296	Deshidrogenasa	Clon
	257, 258	Deshidrogenasa	Clon
	471, 472	Deshidrogenasa	Clon
	323, 324	Deshidrogenasa	Clon
	325, 326	Deshidrogenasa	Clon
	305, 306	Deshidrogenasa	Clon
	331, 332	Deshidrogenasa	Clon
	343, 344	Deshidrogenasa	Clon
	345, 346	Deshidrogenasa	Clon
	337, 338	Deshidrogenasa	Clon
	309, 310	Deshidrogenasa	Clon
	307, 308	Deshidrogenasa	Clon
	347, 348	Deshidrogenasa	Clon
	329, 330	Deshidrogenasa	Clon
	339, 340	Deshidrogenasa	Clon
	311, 312	Deshidrogenasa	Clon
	313, 314	Deshidrogenasa	Clon
	315, 316	Deshidrogenasa	Clon
	317, 318	Deshidrogenasa	Clon
	349, 350	Deshidrogenasa	Clon
	351, 352	Deshidrogenasa	Clon
	333, 334	Deshidrogenasa	Clon
	319, 320	Deshidrogenasa	Clon
	675, 676	Oxidorreductasa	Clon
	671, 672	Oxidorreductasa	Clon
	673, 674	Oxidorreductasa	Clon
	643, 644	Oxidorreductasa	Clon
	645, 646	Oxidorreductasa	Clon
	647, 648	Oxidorreductasa	Clon
	649, 650	Oxidorreductasa	Clon
	663, 664	Oxidorreductasa	Clon
	807, 808	Oxidorreductasa	Clon
	697, 698	Oxidorreductasa	Clon
	699, 700	Oxidorreductasa	Clon
	837, 838	Oxidorreductasa	Clon
	715, 716	Oxidorreductasa	Clon
	717, 718	Oxidorreductasa	Clon
	701, 702	Oxidorreductasa	Clon

ES 2 531 290 T3

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
	687, 688	Oxidorreductasa	Clon
	703, 704	Oxidorreductasa	Clon
	719, 720	Oxidorreductasa	Clon
	709, 710	Oxidorreductasa	Clon
	711, 712	Oxidorreductasa	Clon
	713, 714	Oxidorreductasa	Clon
	689, 690	Oxidorreductasa	Clon
	681, 682	Oxidorreductasa	Clon
	705, 706	Oxidorreductasa	Clon
	721, 722	Oxidorreductasa	Clon
	809, 810	Oxidorreductasa	Clon
	811, 812	Oxidorreductasa	Clon
	813, 814	Oxidorreductasa	Clon
	815, 816	Oxidorreductasa	Clon
	839, 840	Oxidorreductasa	Clon
	695, 696	Oxidorreductasa	Clon
	707, 708	Oxidorreductasa	Clon
	791, 792	Oxidorreductasa	Clon
	745, 746	Oxidorreductasa	Clon
	747, 748	Oxidorreductasa	Clon
	855, 856	Oxidorreductasa	Clon
	857, 858	Oxidorreductasa	Clon
	859, 860	Oxidorreductasa	Clon
	861, 862	Oxidorreductasa	Clon
	863, 864	Oxidorreductasa	Clon
	773, 774	Oxidorreductasa	Clon
	723, 724	Oxidorreductasa	Clon
	775, 776	Oxidorreductasa	Clon
	633, 634	Oxidorreductasa	Clon
	635, 636	Oxidorreductasa	Clon
	749, 750	Oxidorreductasa	Clon
	725, 726	Oxidorreductasa	Clon
	825, 826	Oxidorreductasa	Clon
	751, 752	Oxidorreductasa	Clon
	777, 778	Oxidorreductasa	Clon
	683, 684	Oxidorreductasa	Clon
	685, 686	Oxidorreductasa	Clon
	753, 754	Oxidorreductasa	Clon

ES 2 531 290 T3

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
	755, 756	Oxidorreductasa	Clon
	779, 780	Oxidorreductasa	Clon
	827, 828	Oxidorreductasa	Clon
	781, 782	Oxidorreductasa	Clon
	757, 758	Oxidorreductasa	Clon
	829, 830	Oxidorreductasa	Clon
	851, 852	Oxidorreductasa	Clon
	849, 850	Oxidorreductasa	Clon
	817, 818	Oxidorreductasa	Clon
	831, 832	Oxidorreductasa	Clon
	793, 794	Oxidorreductasa	Clon
	823, 824	Oxidorreductasa	Clon
	843, 844	Oxidorreductasa	Clon
	845, 846	Oxidorreductasa	Clon
	819, 820	Oxidorreductasa	Clon
	759, 760	Oxidorreductasa	Clon
	783, 784	Oxidorreductasa	Clon
	785, 786	Oxidorreductasa	Clon
	833, 834	Oxidorreductasa	Clon
	835, 836	Oxidorreductasa	Clon
	795, 796	Oxidorreductasa	Clon
	797, 798	Oxidorreductasa	Clon
	799, 800	Oxidorreductasa	Clon
	801, 802	Oxidorreductasa	Clon
	761, 762	Oxidorreductasa	Clon
	803, 804	Oxidorreductasa	Clon
	805, 806	Oxidorreductasa	Clon
	763, 764	Oxidorreductasa	Clon
	765, 766	Oxidorreductasa	Clon
	767, 768	Oxidorreductasa	Clon
	769, 770	Oxidorreductasa	Clon
	771, 772	Oxidorreductasa	Clon
	787, 788	Oxidorreductasa	Clon
	789, 790	Oxidorreductasa	Clon
	841, 842	Oxidorreductasa	Clon
	729, 730	Oxidorreductasa	Clon
	731, 732	Oxidorreductasa	Clon
	733, 734	Oxidorreductasa	Clon

Par Clon/subclón	SEQ ID NO:	Actividad	Tipo de secuencia (clon o subclón)
	735, 736	Oxidorreductasa	Clon
	727, 728	Oxidorreductasa	Clon
	691, 692	Oxidorreductasa	Clon
	693, 694	Oxidorreductasa	Clon
	847, 848	Oxidorreductasa	Clon
	853, 854	Oxidorreductasa	Clon
	651, 652	Oxidorreductasa	Clon
	653, 654	Oxidorreductasa	Clon
	655, 656	Oxidorreductasa	Clon
	657, 658	Oxidorreductasa	Clon
	659, 660	Oxidorreductasa	Clon
	661, 662	Oxidorreductasa	Clon
	637, 638	Oxidorreductasa	Clon
	631, 632	Oxidorreductasa	Clon
	639, 640	Oxidorreductasa	Clon
	641, 642	Oxidorreductasa	Clon
	737, 738	Oxidorreductasa	Clon
	739, 740	Oxidorreductasa	Clon
	821, 822	Oxidorreductasa	Clon
	741, 742	Oxidorreductasa	Clon
	743, 744	Oxidorreductasa	Clon
	679, 680	Oxidorreductasa	Clon
	665, 666	Oxidorreductasa	Clon
	667, 668	Oxidorreductasa	Clon
	669, 670	Oxidorreductasa	Clon
	677, 678	Oxidorreductasa	Clon

Tabla 2

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
1, 2	Desconocido		
3, 4	Desconocido		
5, 6	Desconocido		
7, 8	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
9, 10	Desconocido		
11, 12	Desconocido		
13, 14	Desconocido		
15, 16	Desconocido		
17, 18	Desconocido		
19, 20	Desconocido		
21, 22	Desconocido		
23, 24	Desconocido		
25, 26	Desconocido		
27, 28	Desconocido		
29, 30	Desconocido		
31, 32	Desconocido		
33, 34	Desconocido		
35, 36	Desconocido		
37, 38	Desconocido		
39, 40	Desconocido		
41, 42	Desconocido		
43, 44	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
45, 46	Desconocido		
47, 48	Desconocido		
49, 50	Desconocido		
51, 52	Desconocido		
53, 54	Desconocido		
55, 56	Desconocido		
57, 58	Desconocido		
59, 60	Desconocido		
61, 62	Desconocido		
63, 64	Desconocido		
65, 66	Desconocido		
67, 68	Desconocido		
69, 70	Desconocido		
71, 72	Desconocido		
73, 74	Desconocido		
75, 76	Desconocido		
77, 78	Desconocido		
79, 80	Desconocido		

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
81, 82	Desconocido		
83, 84	Desconocido		
85, 86	Desconocido		
87, 88	Desconocido		
89, 90	Desconocido		
91, 92	Desconocido		
93, 94	Desconocido		
95, 96	Desconocido		
97, 98	Desconocido		
99, 100	Desconocido		
101, 102	Desconocido		
103, 104	Desconocido		
105, 106	Desconocido		
107, 108	Desconocido		
109, 110	Desconocido		
111, 112	Desconocido		
113, 114	Desconocido		

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
115, 116	Desconocido		
117, 118	Desconocido		
119, 120	Desconocido		
121, 122	Desconocido		
123, 124	Desconocido		
125, 126	Desconocido		
127, 128	Desconocido		
129, 130	Desconocido		
131, 132	Desconocido		
133, 134	Desconocido		
135, 136	Desconocido		
137, 138	Desconocido		
139, 140	Desconocido		
141, 142	Desconocido		
143, 144	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
145, 146	Desconocido		
147, 148	Desconocido		
149, 150	Desconocido		
151, 152	Desconocido		
153, 154	Desconocido		
155, 156	Desconocido	Probabilidad: 0.991 AA1: 19 AA2: 20	KNSPIIAAYRAATPGSAAA
157, 158	Desconocido		
159, 160	Desconocido		
161, 162	Desconocido		
163, 164	Desconocido		
165, 166	Desconocido		
167, 168	Desconocido		
169, 170	Rhodococcus erythropolis DSMZ 44522		
171, 172	Desconocido		
173, 174	Desconocido		

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
175, 176	Desconocido		
177, 178	Desconocido		
179, 180	Desconocido		
181, 182	Desconocido		
183, 184	Desconocido		
185, 186	Desconocido		
187, 188	Desconocido		
189, 190	Desconocido		
191, 192	Desconocido		
193, 194	Desconocido		
195, 196	Desconocido		
197, 198	Desconocido		
199, 200	Desconocido		
201, 202	Desconocido		
203, 204	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
205, 206	Desconocido		
207, 208	Desconocido		
209, 210	Desconocido		
211, 212	Desconocido		
213, 214	Desconocido		
215, 216	Desconocido		
217, 218	Desconocido		
219, 220	Desconocido		
221, 222	Desconocido		
223, 224	Desconocido		
225, 226	Desconocido		
227, 228	Desconocido		
229, 230	Desconocido		
231, 232	Desconocido		
233, 234	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
235, 236	Desconocido		
237, 238	Desconocido		
239, 240	Desconocido		
241, 242	Desconocido		
243, 244	Desconocido		
245, 246	Desconocido		
247, 248	Desconocido		
249, 250	Desconocido		
251, 252	Desconocido		
253, 254	Desconocido		
255, 256	Desconocido		
257, 258	Desconocido	Probabilidad: 0.612 AA1: 20 AA2: 21	MKSAIVLGAGMVGIATAVHL
259, 260	Desconocido		
261, 262	Desconocido		
263, 264	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
265, 266	Desconocido		
267, 268	Desconocido		
269, 270	Desconocido		
271, 272	Desconocido	Probabilidad: 0.836 AA1: 22 AA2: 23	MKPTSILVLGAGMVGCTALHL
273, 274	Desconocido		
275, 276	Desconocido		
277, 278	Desconocido		
279, 280	Desconocido		
281, 282	Desconocido		
283, 284	Desconocido	Probabilidad: 0.549 AA1: 20 AA2: 21	MKAIVLGSGVLGTTTAYYLA
285, 286	Desconocido	Probabilidad: 0.957 AA1: 24 AA2: 25	MARPRSVIICGGGIIGLCTAYSLA
287, 288	Desconocido		
289, 290	Desconocido		
291, 292	Desconocido		
293, 294	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
295, 296	Desconocido		
297, 298	Desconocido	Probabilidad: 0.898 AA1: 21 AA2: 22	MQSIAVIGGGITGVTSAYALA
299, 300	Desconocido	Probabilidad: 0.898 AA1: 21 AA2: 22	MQSIAVIGGGITGVTSAYALA
301, 302	Desconocido		
303, 304	Desconocido	Probabilidad: 0.945 AA1: 18 AA2: 19	MKVLVLGAGVVGTTATALA
305, 306	Desconocido		
307, 308	Desconocido		
309, 310	Desconocido		
311, 312	Desconocido		
313, 314	Desconocido		
315, 316	Desconocido		
317, 318	Desconocido		
319, 320	Desconocido	Probabilidad: 0.725 AA1: 25 AA2: 26	MKSARPVKTVGIAGAGTMGRGIAAA
321, 322	Desconocido		
323, 324	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
325, 326	Desconocido	Probabilidad: 0.552 AA1: 20 AA2: 21	MRVLVLGSGVIGTASAYYLA
327, 328	Desconocido		
329, 330	Desconocido		
331, 332	Desconocido		
333, 334	Desconocido	Probabilidad: 0.725 AA1: 25 AA2: 26	MKSARPVKTVGIAGAGTMGRGIAAA
335, 336	Desconocido		
337, 338	Desconocido		
339, 340	Desconocido		
341, 342	Desconocido		
343, 344	Desconocido		
345, 346	Desconocido		
347, 348	Desconocido		
349, 350	Desconocido		
351, 352	Desconocido		
353, 354	Desconocido	Probabilidad: 0.791 AA1: 24 AA2: 25	MRQSRSVIICGGGVIGLSCAYYLA

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
355, 356	Desconocido		
357, 358	Desconocido		
359, 360	Rhodococcus ruber DSMZ 44319		
361, 362	Rhodococcus ruber DSMZ 44319		
363, 364	Rhodococcus ruber DSMZ 44319		
365, 366	Desconocido		
367, 368	Desconocido		
369, 370	Desconocido		
371, 372	Desconocido		
373, 374	Desconocido		
375, 376	Desconocido		
377, 378	Desconocido		
379, 380	Desconocido		
381, 382	Rhodococcus ruber DSMZ 44319		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
383, 384	Desconocido	Probabilidad: 0.657 AA1: 21 AA2: 22	MMKIMVLGGGVIGVTTAYYLA
385, 386	Desconocido		
387, 388	Desconocido	Probabilidad: 0.930 AA1: 20 AA2: 21	MRIVVLGAGVVGTTAAYCLA
389, 390	Desconocido		
391, 392	Desconocido	Probabilidad: 0.711 AA1: 20 AA2:21	MSSTRRVIVIGGGVIGAASA
393, 394	Desconocido	Probabilidad: 0.968 AA1: 18 AA2: 19	MKILVIGAGVIGVATAWA
395, 396	Desconocido	Probabilidad: 0.638 AA1: 20 AA2: 21	MTKDIVVLGAGVVGVTALA
397, 398	Desconocido		
399, 400	Desconocido		
401, 402	Desconocido	Probabilidad: 0.999 AA1: 18 AA2: 19	MKTLVLGGGIAGLSSAFA
403, 404	Desconocido	Probabilidad: 0.959 AA1: 20 AA2: 21	MSKKGTSVIIGGGISGLASA
405, 406	Desconocido		
407, 408	Desconocido		
409, 410	Desconocido		
411, 412	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
413, 414	Desconocido		
415, 416	Desconocido		
417, 418	Desconocido	Probabilidad: 0.709 AA1: 20 AA2: 21	MKITILGAGVIGVTSAYLA
419, 420	Desconocido		
421, 422	Desconocido		
423, 424	Desconocido		
425, 426	Desconocido		
427, 428	Desconocido		
429, 430	Desconocido		
431, 432	Desconocido		
433, 434	Desconocido		
435, 436	Desconocido	Probabilidad: 0.738 AA1: 22 AA2: 23	MPGTVDIAVLGAGIVGVSAALA
437, 438	Desconocido		
439, 440	Desconocido		
441, 442	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
443, 444	Desconocido		
445, 446	Desconocido		
447, 448	Desconocido		
449, 450	Desconocido	Probabilidad: 0.823 AA1: 23 AA2: 24	MKRDVIVLGAGMVGVCALHLQA
451, 452	Desconocido		
453, 454	Desconocido		
455, 456	Desconocido		
457, 458	Desconocido		
459, 460	Desconocido	Probabilidad: 0.950 AA1: 21 AA2: 22	MQRIAVIGGGITGITSAYALA
461, 462	Desconocido	Probabilidad: 0.557 AA1: 18 AA2: 19	MPSVLITGATSGFGKAAA
463, 464	Desconocido		
465, 466	Desconocido		
467, 468	Desconocido		
469, 470	Desconocido		
471, 472	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
473, 474	Desconocido		
475, 476	Desconocido		
477, 478	Desconocido		
479, 480	Desconocido		
481, 482	Desconocido		
483, 484	Desconocido		
485, 486	Desconocido		
487, 488	Desconocido		
489, 490	Desconocido		
491, 492	Desconocido	Probabilidad: 1.000 AA1: 21 AA2: 22	MKISIVGAGLAGLCAAHALVA
493, 494	Desconocido		
495, 496	Desconocido		
497, 498	Desconocido		
499, 500	Desconocido	Probabilidad: 0.852 AA1: 23 AA2: 24	MKFDVAVLGAGIVGISTALHLQA
501, 502	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
503, 504	Desconocido		
505, 506	Desconocido		
507, 508	Desconocido		
509, 510	Desconocido		
511, 512	Desconocido	Probabilidad: 0.696 AA1: 28 AA2:29	MTEASRTSRQTEVIVLGAGIVGVSTAL A
513, 514	Desconocido		
515, 516	Desconocido	Probabilidad: 0.898 AA1: 21 AA2: 22	MQSIAVIGGGITGVTSAYALA
517, 518	Desconocido	Probabilidad: 0.798 AA1: 21 AA2: 22	MKSVIIIGGGIIGLCSAYYLA
519, 520	Desconocido		
521, 522	Rhodococcus erythropolis DSMZ 44522		
523, 524	Desconocido	Probabilidad: 0.601 AA1: 22 AA2: 23	MKKKILVIGGGAIGLFCAYYLR
525, 526	Desconocido		
527, 528	Desconocido		
529, 530	Desconocido		
531, 532	Desconocido		

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
533, 534	Desconocido	Probabilidad: 0.716 AA1: 24 AA2: 25	MRNSKSVVVCGGGIVGLCTAYLA
535, 536	Desconocido		
537, 538	Desconocido		
539, 540	Desconocido		
541, 542	Desconocido		
543, 544	Desconocido		
545, 546	Desconocido		
547, 548	Desconocido	Probabilidad: 0.696 AA1: 24 AA2: 25	MTDKRRVVVCGGGVIGLCCADSLA
549, 550	Desconocido		
551, 552	Desconocido		
553, 554	Desconocido		
555, 556	Desconocido		
557, 558	Desconocido		
559, 560	Desconocido		
561, 562	Desconocido	Probabilidad: 0.765 AA1: 22 AA2: 23	MDPHVVIAGCGFGGLFAARALA

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
563, 564	Desconocido		
565, 566	Desconocido		
567, 568	Desconocido		
569, 570	Desconocido		
571, 572	Desconocido		
573, 574	Desconocido	Probabilidad: 0.982 AA1: 24 AA2: 25	MSRPRSVIICGGGIVGLCTAYSLA
575, 576	Desconocido		
577, 578	Desconocido		
579, 580	Desconocido	Probabilidad: 0.772 AA1: 20 AA2: 21	MKITILGAGVIGVTSAYYLA
581, 582	Desconocido		
583, 584	Desconocido		
585, 586	Desconocido		
587, 588	Desconocido		
589, 590	Rhodococcus erythropolis DSMZ 44522		
591, 592	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
593, 594	Desconocido		
595, 596	Desconocido		
597, 598	Desconocido	Probabilidad: 0.672 AA1: 20 AA2: 21	MKVLVLGGGVIGVSSAYFLA
599, 600	Desconocido	Probabilidad: 0.873 AA1: 20 AA2: 21	MKVIVLGAGV VGVTSAYQLA
601, 602	Desconocido	Probabilidad: 0.773 AA1: 20 AA2: 21	MKITILGAGVIGVTSAYYLA
603, 604	Desconocido	Probabilidad: 0.781 AA1: 21 AA2: 22	MKRVIVIGSGALGLCSAYFLQ
605, 606	Desconocido		
607, 608	Desconocido		
609, 610	Desconocido		
611, 612	Desconocido	Probabilidad: 0.772 AA1: 20 AA2: 21	MKITILGAGVIGVTSAYYLA
613, 614	Desconocido	Probabilidad: 0.995 AA1: 18 AA2: 19	MKITIIGAGIAGVSTAWA
615, 616	Desconocido		
617, 618	Desconocido		
619, 620	Desconocido		
621, 622	Desconocido	Probabilidad: 0.710 AA1: 24 AA2: 25	MRTSKSVIVCGGGIVGLCTAYYLA

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
623, 624	Desconocido		
625, 626	Flavobacterium sp. ATCC 27551		
627, 628	Desconocido		
629, 630	Desconocido		
631, 632	Desconocido		
633, 634	Desconocido	Probabilidad: 0.648 AA1: 20 AA2: 21	MKVIVLGAGVIGTTTAYYLA
635, 636	Desconocido	Probabilidad: 0.651 AA1: 20 AA2:21	MKVIVLGAGVIGTTTAYYLA
637, 638	Desconocido		
639, 640	Desconocido		
641, 642	Desconocido		
643, 644	Desconocido	Probabilidad: 0.993 AA1: 24 AA2: 25	MTRARHVVVIGAGVVGSTAQALA
645, 646	Desconocido		
647, 648	Desconocido		
649, 650	Desconocido		
651, 652	Desconocido	Probabilidad: 0.598 AA1: 21 AA2: 22	MAREVIVLGAGIVGVSTAAHL

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
653, 654	Desconocido		
655, 656	Desconocido		
657, 658	Desconocido		
659, 660	Desconocido		
661, 662	Desconocido	Probabilidad: 0.944 AA1: 39 AA2:40	MKRSVQRRQVVLGSGAALLVGALDG CAGSIRTLQSPAPA
663, 664	Desconocido		
665, 666	Desconocido	Probabilidad: 0.910 AA1: 17 AA2: 18	MKSVAIIGAGLAGLATA
667, 668	Desconocido		
669, 670	Desconocido		
671, 672	Desconocido		
673, 674	Desconocido		
675, 676	Desconocido		
677, 678	Desconocido		
679, 680	Desconocido		
681, 682	Desconocido	Probabilidad: 0.781 AA1: 22 AA2: 23	MRIVVIGAGLPGVTTACFLAQA

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
683, 684	Desconocido	Probabilidad: 0.549 AA1: 18 AA2: 19	MNVLVVGAGVVGSTALS
685, 686	Desconocido	Probabilidad: 0.957 AA1: 24 AA2: 25	MNKRTPERVVVIGGGVVGATTALA
687, 688	Desconocido		
689, 690	Desconocido		
691, 692	Desconocido	Probabilidad: 0.984 AA1: 19 AA2: 20	MKTIAVLGAGVTGTTTAYA
693, 694	Desconocido		
695, 696	Desconocido		
697, 698	Desconocido		
699, 700	Desconocido		
701, 702	Desconocido	Probabilidad: 0.999 AA1: 22 AA2: 23	MNRSVAIIGAGVSGLTGCVVFA
703, 704	Desconocido		
705, 706	Desconocido	Probabilidad: 0.935 AA1: 19 AA2: 20	MKSAIVLGAGMVGVSTALA
707, 708	Desconocido		
709, 710	Desconocido		
711, 712	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
713, 714	Desconocido		
715, 716	Desconocido	Probabilidad: 0.528 AA1: 18 AA2: 19	MKVIVIGAGVVGATTALS
717, 718	Desconocido		
719, 720	Desconocido		
721, 722	Desconocido		
723, 724	Desconocido	Probabilidad: 0.857 AA1: 20 AA2: 21	MHTIVIGAGVVGASTALSLA
725, 726	Desconocido		
727, 728	Desconocido	Probabilidad: 0.785 AA1: 18 AA2: 19	MHIVVIGAGVMGVTTAYA
729, 730	Desconocido		
731, 732	Desconocido		
733, 734	Desconocido	Probabilidad: 0.903 AA1: 18 AA2: 19	MHVIVIGAGVVGSTTALA
735, 736	Desconocido	Probabilidad: 0.983 AA1: 21 AA2: 22	KEFGTSSISAATLALAARPAQS
737, 738	Desconocido		
739, 740	Desconocido		
741, 742	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
743, 744	Desconocido		
745, 746	Desconocido	Probabilidad: 0.754 AA1: 23 AA2:24	MTHSDILIIGGGIAGMSAAFFLA
747, 748	Desconocido		
749, 750	Desconocido		
751, 752	Desconocido		
753, 754	Desconocido		
755, 756	Desconocido		
757, 758	Desconocido	Probabilidad: 0.911 AA1: 19 AA2: 20	MQDILVLGAGMVGSTALA
759, 760	Desconocido		
761, 762	Desconocido		
763, 764	Desconocido	Probabilidad: 0.857 AA1: 20 AA2: 21	MHTIVIGAGVVGASTALSLA
765, 766	Desconocido		
767, 768	Desconocido	Probabilidad: 0.682 AA1: 22 AA2: 23	MSLHVIVIGAGVVGASTVLSLA
769, 770	Desconocido		
771, 772	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
773, 774	Desconocido		
775, 776	Desconocido	Probabilidad: 0.527 AA1: 15 AA2: 16	MRVLVIGAGLAGLTA
777, 778	Desconocido		
779, 780	Desconocido		
781, 782	Desconocido		
783, 784	Desconocido		
785, 786	Desconocido	Probabilidad: 0.706 AA1: 20 AA2: 21	MHTIVIGAGVVGSTALSLA
787, 788	Desconocido	Probabilidad: 0.639 AA1: 20 AA2: 21	MHAIVIGAGVVGASTALSLA
789, 790	Desconocido	Probabilidad: 0.953 AA1: 19 AA2: 20	MKEVVVLGAGMVGTTALA
791, 792	Desconocido		
793, 794	Desconocido		
795, 796	Desconocido		
797, 798	Desconocido		
799, 800	Desconocido		
801, 802	Desconocido	Probabilidad: 0.903 AA1: 18 AA2: 19	MHVIVIGAGVVGSTTALA

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
803, 804	Desconocido	Probabilidad: 0.763 AA1: 22 AA2: 23	MPPHVIVVGAGVVGASTALSLA
805, 806	Desconocido		
807, 808	Desconocido	Probabilidad: 0.785 AA1: 18 AA2: 19	MHIVVIGAGVMGVTTAYA
809, 810	Desconocido		
811, 812	Desconocido		
813, 814	Desconocido		
815, 816	Desconocido		
817, 818	Desconocido	Probabilidad: 0.973 AA1: 18 AA2: 19	MKILVLGAGVVGTTATALA
819, 820	Desconocido	Probabilidad: 0.926 AA1: 20 AA2: 21	MHVVLGAGVVGTTTALALA
821, 822	Desconocido		
823, 824	Desconocido		
825, 826	Desconocido		
827, 828	Desconocido		
829, 830	Desconocido	Probabilidad: 0.973 AA1: 24 AA2: 25	MNKRTPERVVVIGGGVVGASTALA
831, 832	Desconocido	Probabilidad: 0.995 AA1: 33 AA2: 34	MYSETKTTRNVDCIVIGAGMAGASA AASLSAEA

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
833, 834	Desconocido	Probabilidad: 0.903 AA1: 18 AA2: 19	MHVIVIGAGVVGSTTALA
835, 836	Desconocido		
837, 838	Desconocido		
839, 840	Desconocido		
841, 842	Desconocido		
843, 844	Desconocido		
845, 846	Desconocido		
847, 848	Desconocido		
849, 850	Desconocido	Probabilidad: 0.930 AA1: 18 AA2: 19	MRVLVLGAGVVGTTATA
851, 852	Desconocido		
853, 854	Desconocido		
855, 856	Desconocido	Probabilidad: 0.553 AA1: 23 AA2: 24	MQKDIWDFVIVGAGMAGASTAWQ
857, 858	Desconocido	Probabilidad: 0.553 AA1: 23 AA2: 24	MQKDIWDFVIVGAGMAGASTAWQ
859, 860	Desconocido	Probabilidad: 0.553 AA1: 23 AA2: 24	MQKDIWDFVIVGAGMAGASTAWQ
861, 862	Desconocido	Probabilidad: 0.672 AA1: 23 AA2: 24	MAHYDAVVVGAGVVGLTTAVSLA

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
863, 864	Desconocido	Probabilidad: 0.544 AA1: 20 AA2: 21	MRVLVLGSGVIGTASAYYLA
865, 866	Desconocido		
867, 868	Desconocido		
869, 870	Desconocido		
871, 872	Desconocido		
873, 874	Desconocido		
875, 876	Desconocido		
877, 878	Desconocido		
879, 880	Desconocido		
881, 882	Desconocido		
883, 884	Desconocido		
885, 886	Desconocido		
887, 888	Desconocido		
889, 890	Desconocido		
891, 892	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
893, 894	Desconocido		
895, 896	Desconocido		
897, 898	Desconocido		
899, 900	Desconocido		
901, 902	Desconocido		
903, 904	Desconocido		
905, 906	Desconocido		
907, 908	Desconocido	Probabilidad: 0.665 AA1: 18 AA2: 19	MAADVVLNGAVVPAAEA
909, 910	Desconocido		
911, 912	Desconocido		
913, 914	Desconocido		
915, 916	Desconocido		
917, 918	Desconocido		
919, 920	Desconocido		
921, 922	Desconocido		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
923, 924	Desconocido		
925, 926	Desconocido		
927, 928	Desconocido		
929, 930	Desconocido		
931, 932	Desconocido		
933, 934	Desconocido	Probabilidad: 0.607 AA1: 30 AA2: 31	MARVSRRFLEDSSSGATTMAFAQLAS EAKR
935, 936	Desconocido		
937, 938	Desconocido		
939, 940	Desconocido		
941, 942	Desconocido		
943, 944	Desconocido		
945, 946	Desconocido		
947, 948	Desconocido		
949, 950	Desconocido		
951, 952	Pyrolobus fumarius		

ES 2 531 290 T3

SEQ ID NO:	Fuente	Sitio de escisión Signalp	Secuencia de señalización predicha
953, 954	Aquifex aeolicus		
955, 956	Desconocido		
957, 958	Desconocido		
959, 960	Desconocido		

Tabla 3

SEQ ID NO:	Descripción NR		Código de Acceso NR	NR Evalúe	Organismo NR	Descripción de Proteína Geneseq		Código de Acceso Proteína Geneseq
219, 220	D-amino ácido aminotransferasa [Clostridium beijerinckii NCIMB 8052] gi82726488 glEAP61226.1 D-amino ácido aminotransferasa [Clostridium beijerinckii NCIMB 8052]		82745661	3.00E-98	Clostridium beijerinckii NCIMB 8052	Proteína de Bacillus Mutante sphaericus dat.		ABB08244
Código de Acceso ADN Geneseq	Evalue	Numero EC Predicho	Longitud ADN Buscado	Buscada Proteína Longitud	Longitud de ADN Sujeto	Longitud de Proteína Sujeto	% ID Proteína	SEQ ID NO:
ABZ36033	0.16	2.6.1.21	861	286	0	282	62	219, 220

La invención provee variantes de polinucleótidos o polipéptidos de la invención, los cuales comprenden secuencias modificadas en uno o más pares de bases, codones, intrones, exones, o residuos de aminoácidos (respectivamente), que aun todavía retienen la actividad biológica de una d-aminoácido transferasa de la invención. Las variantes pueden ser producidas por cualquier número de medios incluyendo métodos tales como, por ejemplo,

- 5 PCR propenso a error, mezcla, mutagénesis dirigida por oligonucleótidos, PCR de ensamble, mutagénesis por PCR sexual, mutagénesis *in vivo*, mutagénesis por casete, mutagénesis de ensamble recursivo, mutagénesis por ensamble exponencial, mutagénesis específica para el sitio, reensamblaje del gen (por ejemplo, Reensamblaje de Gen, véase, por ejemplo, Patente de los Estados Unidos No. 6,537,776), GSSM y cualquier combinación de los mismos.
- 10 El término "mutagénesis por saturación", "mutagénesis por saturación en el sitio del gen" o "GSSM" incluye un método que utiliza cebadores de oligonucleótidos degenerados para introducir mutaciones puntuales en un polinucleótido, tal como se describe en detalle más adelante.

- El término "sistema de evolución dirigida a optimizar" o "evolución dirigida a optimizar incluye un método para reensamblar fragmentos de secuencias de ácidos nucleicos relacionadas, por ejemplo, genes relacionados, y que se explica en detalle más adelante.
- 15

El término "reensamblaje por ligazón sintética" o "SLR" incluye un método para ligar fragmentos de oligonucleótidos en una forma no estocástica y que se explica en detalle más adelante.

Generación y manipulación de ácidos nucleicos

- La invención provee ácidos nucleicos (por ejemplo, ácidos nucleicos que codifican polipéptidos que tienen una actividad transferasa, por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa; incluyendo enzimas que tienen al menos una modificación de secuencia de una secuencia de ácidos nucleico de ejemplo de la invención (como se define más arriba), en donde la modificación de secuencia comprende uno o más cambios en residuos de nucleótidos (o los equivalentes del mismo), incluyendo casetes de expresión tales como vectores de expresión, que codifican los polipéptidos de la invención.
- 20
- 25

- La invención también describe métodos para descubrir nuevas secuencias de transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa utilizando los ácidos nucleicos de la invención. La invención también describe métodos para inhibir la expresión de genes, transcriptos y polipéptidos de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa usando los ácidos nucleicos de la invención. También se proveen métodos para modificar los ácidos nucleicos de la invención, por ejemplo, por reensamble de ligazón sintética, sistema de evolución dirigida optimizado y/o mutagénesis por saturación.
- 30

- Los ácidos nucleicos de la invención pueden ser hechos, aislados y/o manipulados por ejemplo, por clonación y expresión de bibliotecas de ADNc, amplificación de ADN mensajero o genómico por PCR, y similares.
- 35

- En un aspecto, la invención también describe ácidos nucleicos que codifican transferasas, por ejemplo, transaminasa-, por ejemplo, d-aminoácido transferasa-, y/o oxidorreductasa-, por ejemplo, deshidrogenasa-, por ejemplo, d-aminoácido deshidrogenasa con una novedad común en que ellos son derivados de una fuente ambiental, o una fuente bacteriana, o una fuente arqueal. En la práctica de los métodos de la invención, los genes homólogos se pueden modificar manipulando un ácido nucleico de plantilla, tal como se describe aquí. La invención puede ser puesta en práctica en conjunción con cualquier método o protocolo o dispositivo conocidos en la técnica, los cuales están bien descritos en la literatura científica y de patentes.
- 40

- Un aspecto de la invención es un ácido nucleico aislado, sintético o recombinante que comprende una secuencia que tiene al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia con la secuencia de ácidos nucleicos SEQ ID NO: 219 y secuencias sustancialmente idénticas a la misma, secuencias complementarias a la misma. Los ácidos nucleicos aislados pueden comprender ADN, incluyendo ADNc, ADN genómico y ADN sintético. El ADN puede ser de cadena doble o cadena sencilla y si es de cadena sencilla puede ser la cadena codificadora o la cadena no codificadora (antisentido). Alternativamente, los ácidos nucleicos aislados, sintéticos o recombinantes de la invención pueden comprender ARN.
- 45
- 50

- De acuerdo con lo anterior, otro aspecto de la invención es un ácido nucleico aislado, sintético o recombinante que codifica uno de los polipéptidos de la invención. Las secuencias de codificación de estos ácidos nucleicos pueden ser idénticas a una de las secuencias de codificación de uno de los ácidos nucleicos de la invención, o un fragmento de los mismos o pueden ser diferentes secuencias de codificación que codifican uno de los polipéptidos de la invención, secuencias sustancialmente idénticas a las mismas y fragmentos que tienen al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 de aminoácidos consecutivos de uno de los polipéptidos de la invención, como
- 55

resultado de la redundancia o degeneración del código genético. El código genético es bien conocido para los experimentados en la técnica y puede ser obtenido, por ejemplo, en la página 214 de B. Lewin, Genes VI, Oxford University Press, 1997.

El ácido nucleico aislado, sintético o recombinante que codifica uno de los polipéptidos de la invención y secuencias sustancialmente idénticas al mismo, puede incluir, pero no se limita a: sólo la secuencia de codificación de un ácido nucleico de la invención y secuencias sustancialmente idénticas al mismo y secuencias de codificación adicionales, tales como secuencias guía o secuencias propeptina y secuencias no codificadoras, tales como intrones o secuencias no codificadoras 5' y/o 3' de la secuencia de codificación. Así, tal como se utiliza aquí, el término "polinucleótido que codifica un polipéptido" abarca un polinucleótido que incluye solamente la secuencia de codificación para el polipéptido así como un polinucleótido que incluye secuencias de codificación y/o no codificación adicionales.

Alternativamente, las secuencias de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, pueden ser mutagenizadas utilizando técnicas convencionales, tal como mutagénesis, u otras técnicas familiares para los experimentados en la técnica, para introducir cambios silenciosos en los polinucleótidos de la invención y secuencias sustancialmente idénticas a los mismos. Tal como se utiliza aquí, "cambios silenciosos" incluyen, por ejemplo, cambios que no alteran la secuencia de aminoácidos codificada por el polinucleótido. Tales cambios pueden ser deseables con el fin de incrementar el nivel del polipéptido producido por células anfitrionas que contienen un vector que codifica el polipéptido introduciendo codones o pares de codones que se presentan frecuentemente en el organismo anfitrión.

La invención también se relaciona con polinucleótidos que tienen cambios en nucleótidos que dan como resultado sustituciones, adiciones, eliminaciones, fusiones y truncamientos de aminoácidos, en los polipéptidos de la invención y secuencias sustancialmente idénticas a los mismos. Tales cambios en nucleótidos pueden ser introducidos utilizando técnicas tales como mutagénesis dirigida al sitio, mutagénesis química aleatoria, eliminación de exonucleasa III y otras técnicas de ADN recombinante. Alternativamente, tales cambios en nucleótidos pueden ser variantes alélicas de origen natural que son aisladas identificando ácidos nucleicos que hibridan específicamente a sondas que comprenden al menos 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, o 500 bases consecutivas de una de las secuencias de la invención y secuencias sustancialmente idénticas a las mismas (o las secuencias complementarias para las mismas) bajo condiciones de alta, moderada o baja restricción tal como se proveen aquí.

Técnicas generales

Los ácidos nucleicos utilizados para poner en práctica esta invención, bien sea ARN, ARNi, ácido nucleico antisentido, ADNc, ADN genómico, vectores, virus o híbridos de los mismos, pueden ser aislados a partir de una variedad de fuentes, manipuladas genéticamente, amplificadas y/o expresadas/ generadas por vía recombinante. Los polipéptidos recombinantes (por ejemplo transaminasas y oxidorreductasas de la invención) generados a partir de estos ácidos nucleicos pueden ser aislados o clonados y probados individualmente para una actividad deseada. Cualquier sistema de expresión recombinante puede ser utilizado, incluyendo sistemas de expresión bacterianos, mamíferos, de levaduras, de insectos o células vegetales.

Alternativamente, estos ácidos nucleicos pueden ser sintetizados *in vitro* por técnicas de síntesis química bien conocidas, como se describe en, por ejemplo, Adams (1983) J. Am. Chem. Soc. 105: 661; Belousov (1997) Nucleic Acids Res. 25: 3440-3444; Frenkel (1995) Free Radic. Biol. Med. 19: 373-380; Blommers (1994) Biochemistry 33: 7886 a 7.896; Narang (1979) Meth. Enzymol. 68:90; Brown (1979) Meth. Enzymol. 68: 109; Beaucage (1981) Tetra. Lett. 22: 1859; Patente de los Estados Unidos. Nº 4.458.066.

Técnicas para la manipulación de ácidos nucleicos, tales como, por ejemplo, subclonación, sondas de marcación (por ejemplo, marcación con cebador aleatorio utilizando la polimerasa Klenow, traducción de muescas, amplificación), secuenciación, hibridación y similares están bien descritos en la literatura científica y de patentes, véase, por ejemplo, Sambrook, ed., MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. John Wiley & Sons, Inc., New York (1997); LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY: HYBRIDIZATION WITH NUCLEIC ACID PROBES, Part I. Theory and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y. (1993).

Otro medio útil para obtener y manipular ácidos nucleicos utilizados para poner en práctica los métodos de la invención es clonarlo a partir de muestras genómicas, y, si se desea, seleccionar y reclonar insectos aislados o amplificados a partir de, por ejemplo, clones genómicos o clones de ADNc. Fuentes de ácidos nucleicos usados en los métodos de esta invención incluyen bibliotecas genómicas o de ADNc contenidas en, por ejemplo, cromosomas artificiales de mamíferos (MAC), véase, por ejemplo las patentes de los Estados Unidos números. 5,721,118; 6,025,155; cromosomas artificiales humanos, véase, por ejemplo Rosenfeld (1997) Nat. Genet. 15: 333-335; cromosomas artificiales de levadura (YAC); cromosomas artificiales bacterianos (BAC); cromosomas artificiales P1,

véase por ejemplo Woon (1998) *Genomics* 50:306-316; vectores derivados de P1 (PACs), véase, por ejemplo, Kern (1997) *Biotechniques* 23: 120-124; cósmidos, virus recombinantes, fagos o plásmidos.

En un aspecto, un ácido nucleico que codifica un polipéptido de la invención es ensamblado en fase apropiada con una secuencia guía capaz de dirigir la secreción del polipéptido traducido o un fragmento del mismo.

- 5 La invención describe proteínas de fusión y ácidos nucleicos que las codifican. Un polipéptido de la invención puede ser fusionado a un péptido o polipéptido heterólogo, tal como péptidos de identificación en N-terminal los cuales imparten características deseadas, tales como estabilidad incrementada o purificación simplificada. Los péptidos y polipéptidos de la invención también pueden ser sintetizados y expresados como proteínas de fusión con uno o más dominios adicionales enlazados a los mismos, por ejemplo, produciendo un péptido más inmunogénico, para aislar
- 10 más fácilmente un péptido sintetizado por vía recombinante, para identificar y aislar anticuerpos y células B que expresan anticuerpos, y similares. Dominios que facilitan la detección y purificación incluyen, por ejemplo, péptidos quelantes de metales tales como trectos de polihistidina y módulos de histidina-triptófano que permiten la purificación sobre metales inmovilizados, dominios de proteína A que permiten la purificación sobre inmunoglobulina inmovilizada, y el dominio utilizado en el sistema de purificación por extensión/afinidad FLAGS (Immunex Corp,
- 15 Seattle WA). La inclusión de secuencias de enlazamiento escindible tales como el Factor Xa o la enteroquinasa (Invitrogen, San Diego CA) entre un dominio de purificación y el péptido o polipéptido que comprende el motivo para facilitar la purificación. Por ejemplo, un vector de expresión puede incluir una secuencia de ácido nucleico que codifica un epítipo enlazada a seis residuos de histidina seguida por una tiorredoxina y un sitio de escisión de enteroquinasa (véase, por ejemplo, Williams (1995) *Biochemistry* 34: 1787-1797; Dobeil (1998) *Protein Expr. Purif.* 12: 404-414). Los residuos de histidina facilitan la detección y purificación, mientras que el sitio de escisión de enteroquinasa provee un medio para purificar el epítipo a partir del resto de la proteína de fusión. La tecnología pertinente a vectores que codifican las proteínas de fusión y la aplicación de las proteínas de fusión están bien descritas en la literatura científica y de patentes, véase, por ejemplo, Kroll (1993) *DNA Cell. Biol.*, 12: 441-53.

- 25 Las expresiones "ácido nucleico" o "secuencia de ácidos nucleicos" tal como se utilizan aquí, se refieren a un oligonucleótido, nucleótidos, polinucleótidos, o un fragmento de cualquiera de estos, a ADN o ARN de origen genómico o sintético que pueden ser de cadena sencilla o cadena doble y pueden representar una cadena en sentido o antisentido, a un ácido nucleico peptídico (PNA), o a cualquier material similar a ADN o similar a ARN, naturales o sintético en su origen. Las expresiones "ácido nucleico" o "secuencia de ácidos nucleicos" Incluidos oligonucleótidos, nucleótido, polinucleótido o a un fragmento de cualquiera de éstos, ADN o ARN (por ejemplo,
- 30 ARNm, ARNr, ARNt, ARNi) de origen genómico o sintético que pueden ser de cadena sencilla o cadena doble y pueden representar una cadena en sentido o antisentido, ácidos nucleicos peptídicos (PNA), o cualquier material similar a ADN o similar a ARN, natural o sintético en su origen, incluyendo, por ejemplo, ARNi, ribonucleoproteínas (ARNi de cadena doble, por ejemplo, sPNRi). El término abarca ácidos nucleicos, por ejemplo, oligonucleótidos, que contienen análogos conocidos de nucleótidos naturales. El término también abarca estructuras similares a ácidos
- 35 nucleicos con esqueletos sintéticos, véase, por ejemplo, Mata (1997) *Toxicol. Appl. Pharmacol.* 144:189-197; Strauss- Soukup (1997) *Biochemistry* 36:8692-8698; Samstag (1996) *Antisense Nucleic Acid Drug Dev* 6:153-156. "Oligonucleótido" incluye bien sea un polidesoxinucleótido de cadena sencilla o dos cadenas de polidesoxinucleótido complementarias que pueden ser sintetizadas químicamente. Tales oligonucleótidos sintéticos no tienen fosfato en 5' y por lo tanto no se ligan con otro oligonucleótido sin agregar un fosfato con un ATP en la presencia de una quinasas.
- 40 Un oligonucleótido sintético puede ligarse a un fragmento que no ha sido desfosforilado.

Una "secuencia de codificación de" o una "secuencia de nucleótidos que codifica" un polipéptido o proteína en particular, es una secuencia de ácidos nucleicos que es transcrita y traducida en un polipéptido o proteína cuando se coloca bajo el control de secuencias reguladoras apropiadas.

- 45 En un aspecto, el término "gen" significa el segmento de ADN involucrado en la producción de una cadena de polipéptidos. Incluye regiones precedentes y posteriores a la región de codificación (guía y arrastre), así como, cuando es aplicable, secuencias de intervención (intrones) entre segmentos de codificación individuales (exones). En un aspecto, "enlazado operativamente" tal como se utiliza aquí, se refiere a una relación funcional entre dos o más segmentos de ácidos nucleicos (por ejemplo, ADN). En un aspecto, se refiere a la relación funcional de secuencias reguladoras transcripcionales con una secuencia transcrita. Por ejemplo, un promotor está enlazado
- 50 operativamente a una secuencia codificadora, tal como un ácido nucleico de la invención, si estimula o modula la transcripción de la secuencia de codificación en una célula anfitriona apropiada u otro sistema de expresión. En un aspecto, secuencias reguladoras de transcripción promotoras están enlazadas operativamente a una secuencia transcrita y son físicamente contiguas a la secuencia transcrita, esto es, actúan en cis. En un aspecto las secuencias reguladoras transcripcionales, tales como potenciadores, pueden ser físicamente continuas o localizadas en
- 55 proximidad cercana a las secuencias de codificación cuya transcripción potencian.

En un aspecto el término "casete de expresión" se refiere a una secuencia de nucleótidos que es capaz de afectar la expresión de un gen estructural (esto es, una secuencia de codificación de proteínas, tal como una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención) en un anfitrión compatible con tales secuencias. Los

casetes de expresión incluyen al menos un promotor enlazado operativamente con la secuencia que codifica el polipéptido; y, en un aspecto, con otras secuencias, por ejemplo señales de terminación de la transcripción. Factores adicionales necesarios o útiles en efectuar la expresión pueden ser utilizados también, por ejemplo, potenciadores. Así, los casetes de expresión también incluyen plásmidos, vectores de expresión, virus recombinantes, y cualquier forma de vectores de "ADN desnudo", y similares. En un aspecto, un "vector" comprende un ácido nucleico que puede infectar, transfectar, transducir transiente o permanentemente una célula. En aspectos alternativos un vector puede ser un ácido nucleico desnudo, o un ácido nucleico complejo con una proteína o un lípido. El vector en un aspecto comprende ácidos nucleicos y/o proteínas virales y bacterianos, y/o membranas (por ejemplo, una membrana celular, una envoltura lipídica, viral, etc.). Los vectores incluyen, pero no se limitan a replicones (por ejemplo, replicones de ARN, bacteriófagos) a los cuales pueden unirse fragmentos de ADN y replicarse así. Los vectores incluyen por lo tanto, pero no se limitan a ARN, a ADN o ARN circular o lineal auto-replicante autónomo (por ejemplo, plásmidos, virus, y similares, véase, por ejemplo, la patente de los Estados Unidos. No. 5,217,879), e incluyen tanto los plásmidos de expresión como de no expresión. Cuando un microorganismo recombinante o un cultivo celular es descrito como que aloja un "vector de expresión" esto incluye tanto ADN circular y lineal extra cromosómico como ADN que ha sido incorporado en el cromosoma del anfitrión. Cuando un vector está siendo mantenido por una célula anfitriona, el vector puede ser replicado bien sea de manera estable por las células durante la mitosis como una estructura autónoma, o es incorporado dentro del genoma del anfitrión.

En un aspecto, el término "promotor" incluye todas las secuencias capaces de guiar la transcripción de una secuencia de codificación en una célula, por ejemplo una célula vegetal. Así, los promotores usados en los constructos de la invención incluyen elementos de control transcripcional que actúan en cis y secuencias reguladoras que están involucradas en la regulación o modulación de la temporización y/o tasa de la transcripción de un gen. Por ejemplo, un promotor puede ser un elemento de control transcripcional que actúa en cis, incluyendo un potenciador, un promotor, un terminador de transcripción, un origen de replicación, una secuencia de integración cromosómica, regiones no traducidas 5' y 3', o una secuencia intrónica, la cual está involucrada en la regulación transcripcional. Estas secuencias que actúan en cis interactúan típicamente con proteínas u otras biomoléculas para llevar a cabo (iniciar/ detener, regular, modular, etc.) la transcripción. Promotores "constitutivos" son aquellos que guían la expresión de manera continua bajo la mayoría de las condiciones ambientales y estados de desarrollo o diferenciación celular. Promotores "Inducibles" o "regulables" dirigen la expresión del ácido nucleico de la invención bajo la influencia de las condiciones ambientales o condiciones de desarrollo. Ejemplos de condiciones ambientales que pueden afectar la transcripción por promotores inducibles incluyen condiciones anaeróbicas, temperatura elevada, sequía o la presencia de luz.

En realizaciones alternativas, promotores "específicos para tejido" son elementos de control transcripcional que son activos solamente en células o tejidos u órganos particulares, por ejemplo, en plantas o animales. La regulación específica para tejidos puede ser lograda mediante ciertos factores intrínsecos que aseguran que los genes que codifican las proteínas específicas para un tejido dado están expresados. Se sabe que tales factores existen en mamíferos y plantas de tal manera que permiten el desarrollarlo de tejidos específicos.

En realizaciones alternativas, el término "aislado" significa que el material (por ejemplo, un ácido nucleico, un polipéptido, una célula) es removido de su ambiente original (por ejemplo, el ambiente natural si es de origen natural). Por ejemplo, un polinucleótido o polipéptido de origen natural presente en animales vivos no es aislado, pero el mismo polinucleótido o polipéptido separado de parte o todos los materiales coexistentes en el sistema natural es aislado. Tales polinucleótidos podrían ser parte de un vector y/o tales polinucleótidos o polipéptidos podrían ser parte de una composición y todavía estar aislados en cuanto tal vector o composición no es parte de su ambiente natural. En realizaciones alternativas, el término "purificado" no requiere pureza absoluta; en vez de esto, se entiende como una definición relativa. Ácidos nucleicos individuales obtenidos a partir de una biblioteca han sido purificados convencionalmente mediante homogeneidad electroforética. Las secuencias obtenidas de estos clones no podrían ser obtenidas directamente ni de la biblioteca ni del ADN humano total. Los ácidos nucleicos purificados de la invención han sido purificados a partir del resto del ADN genómico en el organismo en al menos 10^4 – 10^6 veces. Sin embargo, el término "purificado" también incluye ácidos nucleicos que han sido purificados a partir del resto del ADN genómico o a partir de otras secuencias en una biblioteca u otro ambiente por al menos un orden de magnitud, típicamente dos o tres órdenes y más típicamente de cuatro a cinco órdenes de magnitud.

En realizaciones alternativas, el término "recombinante" significa que el ácido nucleico es adyacente a un ácido nucleico "de esqueleto" el cual no es adyacente en su ambiente natural. En realizaciones alternativas, para estar "enriquecido" los ácidos nucleicos representarán el 5% o más del número de insertos de ácidos nucleicos en una población de moléculas de esqueleto de ácidos nucleicos. Las moléculas del esqueleto de acuerdo con la invención incluyen ácidos nucleicos tales como vectores de expresión, ácidos nucleicos auto replicantes, virus, ácidos nucleicos integrantes y otros vectores o ácidos nucleicos utilizados para mantener o manipular un inserto de ácidos nucleicos de interés. En realizaciones alternativas, los ácidos nucleicos enriquecidos representan 15% o más del número de insertos de ácidos nucleicos en la población de moléculas de esqueleto recombinantes. En realizaciones alternativas, los ácidos nucleicos enriquecidos representan 50% o más del número de insertos de ácido nucleico en la población de moléculas de esqueleto recombinantes. En un aspecto, los ácidos nucleicos enriquecidos

representan 90% o más del número de insertos de ácidos nucleicos en la población de moléculas de esqueleto recombinantes.

Los "plásmidos" son designados por una minúscula "p" precedida y/o seguida por letras mayúsculas y/o números. Los planes de partida aquí son disponibles comercialmente, disponibles públicamente sobre una base no restringida, o pueden ser contruidos a partir de plásmidos disponibles de acuerdo con procedimientos publicados. Además, los plásmidos equivalentes a los descritos aquí son conocidos en la técnica y será evidente para personas de experiencia ordinaria en la técnica. Los "plásmidos" pueden estar disponibles comercialmente, disponibles públicamente sobre una base restringida, o pueden ser contruidos a partir de plásmidos disponibles de acuerdo con procedimientos publicados. Los plásmidos equivalentes a los descritos aquí son conocidos en la técnica y serán evidentes para la persona con experiencia normal.

En realizaciones alternativas, la "digestión" de ADN se refiere a la escisión catalítica del ADN con una enzima de restricción que actúa solamente en ciertas secuencias en el ADN. Las diversas enzimas de restricción utilizadas aquí están disponibles comercialmente y sus condiciones de reacción, cofactores y otros requerimientos fueron usados tal como es sabido por la persona de experiencia normal. Para propósitos analíticos, típicamente 1 µg de plásmido o fragmento de ADN se utiliza con aproximadamente 2 unidades de enzima en aproximadamente 20 µl de solución reguladora. Para el propósito de aislar los fragmentos de ADN para la construcción de plásmidos, típicamente 5 a 50 µg de ADN son digeridos con 200 a 250 unidades de una enzima en un volumen mayor. Los reguladores y cantidades de sustrato apropiados para enzimas de restricción apropiadas son especificados por el fabricante. Los tiempos de incubación de aproximadamente 1 hora a 37°C se utilizan ordinariamente, pero puede variar de acuerdo con las instrucciones del proveedor. Después de la digestión, puede llevarse a cabo una electroforesis por gel para aislar el fragmento deseado.

En realizaciones alternativas, "hibridación" se refiere a un proceso mediante el cual la cadena de ácido nucleico se une con una cadena complementaria a través del apareamiento de bases. Las reacciones de hibridación pueden ser sensibles y selectivas de tal manera que una secuencia particular de interés puede ser identificada incluso en muestras en donde está presente en bajas concentraciones. Pueden definirse concentraciones adecuadamente restrictivas por ejemplo, mediante las concentraciones de sal o formamida en la soluciones de prehibridación o hibridación, o mediante la temperatura de hibridación y son bien conocidas en la técnica. En realizaciones alternativas la restricción puede ser incrementada reduciendo la concentración de sal, incrementando la concentración de formamida, o elevando la temperatura de hibridación. En aspectos alternativos, los ácidos nucleicos de la invención están definidos por su capacidad para hibridar bajo diversas condiciones de restricción (por ejemplo, altas, medias y bajas), como se define aquí.

Por ejemplo, la hibridación bajo condiciones de alta restricción pueden ocurrir en aproximadamente 50% de formamida a aproximadamente 37°C a 42°C. La hibridación podría ocurrir bajo condiciones de restricción reducidas en aproximadamente 35% a 25% de formamida a aproximadamente 30°C a 35°C. En particular, la hibridación podría ocurrir bajo condiciones de alta restricción a 42°C en formamida a 50%, 5X SSPE, SDS 0.3% y 200 µg/ml de ADN de esperma de salmón separado y desnaturalizado. La hibridación podría ocurrir bajo condiciones de restricción reducida tal como se describió más arriba, pero en 35% de formamida a una temperatura reducida de 35°C. El rango de temperatura correspondiente a un nivel particular de restricción puede ser estrechado adicionalmente calculando la relación purina a pirimidina del ácido nucleico de interés y ajustando la temperatura de acuerdo con ello. A variaciones en los rangos y condiciones anteriores son bien conocidos en la técnica.

Secuencias de control transcripcional y de traducción

La invención provee secuencias de ácidos nucleicos (por ejemplo, ADN) de la invención enlazado operativamente a secuencias de control de expresión (por ejemplo, transcripcional o translacional), por ejemplo, promotores o potenciadores, para dirigir o modular la síntesis/ expresión de ARN. La secuencia de control de expresión puede estar en un vector de expresión. Promotores bacterianos de ejemplo incluyen lacI, lacZ, T3, T7, gpt, lambda PR, PL y trp. Promotores eucariotas de ejemplo incluyen CMV inmediata temprana, HSV timidina quinasa, SV40 temprana y tardía, LTR de retrovirus y metalotionina I de ratón. Una secuencia promotora está "enlazada operativamente" a una secuencia de codificación cuando la ARN polimerasa que inicia la transcripción en el promotor transcribe la secuencia de codificación en ARNm.

Los promotores adecuados para expresar un polipéptido en bacterias incluyen los promotores de E. coli lac o trp, el promotor lacI, el promotor lacZ, el promotor T3, el promotor T7, el promotor gpt, el promotor lambda PR, el promotor lambda PL, promotores de operones que influyen en enzimas glicolíticas tales como 3-fosfoglicerato quinasa (PGK), y el promotor de fosfatasa ácida. Los promotores eucariotas incluyen el promotor temprano CMV intermediario, el promotor HSV timidina quinasa, promotores de tratamiento por calor, siendo el último el promotor SV40 temprano y tardío, LTR de retrovirus, y el promotor de metalotionina-1 de ratón. Otros promotores conocidos para controlar la expresión de genes en células procariotas o eucariotas o sus virus también pueden ser utilizados. Promotores adecuados para expresar los polipéptidos o fragmentos de los mismos en bacterias incluyen la E. coli lac o promotores trp, el promotor lacI, (el promotor lacZ, el promotor T3, el promotor T7, el promotor gpt, el promotor

lambda P_R, el promotor lambda P_L, o promotores de operones que codifican enzimas glicolíticas tales como 3-fosfoglicerato quinasa (PGK) y el promotor de la ácido fosfatasa. Promotores fúngicos incluyen el promotor de factor V, promotores eucariotas incluyen el promotor temprano inmediato CMV, el promotor de timidina quinasa HSV, promotores con choque por calor, el primero y último promotor SV40, los LTR de retrovirus y el promotor metalotioneina-I de ratón. Otros promotores conocidos por controlar la expresión de genes en células procariotas o eucariotas de los virus también pueden ser usados.

Promotores vegetales específicos para tejido

La invención provee casetes de expresión que pueden ser expresados de una manera específica para un tejido, por ejemplo, pueden expresar una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, de la invención en una forma específica del tejido. La invención también provee plantas o semillas que expresan una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, de la invención en una forma específica del tejido. La especificidad de tejido puede ser específica para semillas, específica para tallo, específica para hoja, específica para raíz, específica para fruta y similares.

En un aspecto, un promotor constitutivo tal como un promotor CaMV 35S puede ser utilizado para expresión en partes específicas de la planta o semilla o lo largo de la planta. Por ejemplo, para sobreexpresión, un fragmento promotor vegetal puede ser empleado de forma que dirija la expresión directa del ácido nucleico en alguno o todos los tejidos de la planta, por ejemplo, una planta regenerada. Tales promotores se denominan aquí, promotores "constitutivos" y son activos bajo la mayoría de las condiciones ambientales y estados de desarrollo o diferenciación celular. Ejemplos de promotores constitutivos incluyen la región de iniciación de la transcripción del virus mosaico de la coliflor (CaMV) 35S, el promotor 1' o 2' derivado de T-ADN de *Agrobacterium tumefaciens*, y otras regiones de iniciación de la transcripción de diversos genes vegetales conocidos para los expertos. Tales genes incluyen por ejemplo, ACT11 de *Arabidopsis* (Huang (1996) Plant Mol Biol 33: 125-139.); Cat3 de *Arabidopsis* (GenBank N° U43147, Zhong (1996) Mol. Gen. Genet. 251: 196-203); el gen que codifica la desaturasa de la proteína portadora de estearoil-acilo de *Brassica napus* (GenBank No. X74782, Solcombe (1994) Plant Physiol 104: 1167-1176); GPC1 de maíz (GenBank No. X15596; Martínez (1989) J. Mol. Biol 208: 551-565); el gPC2 de maíz (GenBank N° U45855, Manjunath (1997) Plant Mol. Biol. 33: 97-112); promotores vegetales descritos en las patentes de los Estados Unidos Nos. 4,962,028; 5,633,440.

La invención utiliza promotores específicos de tejido o constitutivos derivados de virus, los cuales pueden incluir, por ejemplo, el promotor subgenómico del tobamovirus (Kumagai (1995) Proc. Natl. Acad. Sci. USA 92:1679-1683; el virus baciliforme de tungro de arroz (RTBV), el cuál replica solamente en células de floema en plantas de arroz infectadas, con su promotor el cuál guía la expresión del gen reportado específico de floema fuerte; el promotor del virus mosaico de la vena de cassava (CVMV), con su actividad mayor en elementos vasculares, en células mesófilas de hojas, y en puntas de raíces, (Verdaguer (1996) Plant Mol. Biol. 31:1129-1139).

Alternativamente, el promotor vegetal puede dirigir la expresión de los ácidos nucleicos que expresan transferasa de por ejemplo, transaminasa-, por ejemplo, d-aminoácido transferasa-, y/o oxidoreductasa-, por ejemplo, deshidrogenasa-, por ejemplo, d-aminoácido deshidrogenasa en un tejido, órgano o tipo de célula específicos (esto es, promotores específicos de tejido) o pueden de alguna otra manera estar bajo condiciones ambientales más precisas o control de desarrollo o bajo el control de un promotor inducible. Ejemplos de condiciones ambientales que pueden afectar la transcripción incluyen condiciones anaeróbicas, temperatura elevada, la presencia de luz, o aspersión con agentes químicos/hormonas. Por ejemplo, la invención incorpora el promotor inducible por sequía del maíz (Busk (1997) supra); el promotor inducible por frío, sequía, y alta salinidad de patata (Kirch (1997) Plant Mol. Biol. 33:897 909).

Los promotores específicos de tejido pueden promover la transcripción solamente dentro de un cierto marco de tiempo de la etapa de desarrollo dentro de este tejido. Véase, por ejemplo Blazquez (1998) Plant Cell 10:791-800, caracteriza el promotor del gen LEAFY de *Arabidopsis*. Véase también Cardon (1997) Plant J 12:367-77, que describe el factor de transcripción SPL3, el cual reconoce un motivo de secuencia conservado de la región promotora del gen AP1 de identidad del meristema floral de *A. thaliana*; y Mandel (1995) Plant Molecular Biology, Vol. 29, pp 995-1004, que describe el promotor eIF4 del meristema. Pueden usarse promotores específicos de tejidos que son activos a lo largo del ciclo de vida de un tejido particular. En otro aspecto, los ácidos nucleicos de la invención están enlazados operativamente a un promotor activo primariamente solamente en células de fibra de algodón. En un aspecto, los ácidos nucleicos de la invención están enlazados operativamente a un promotor activo primariamente durante las etapas de elongación de las células de fibra de algodón, por ejemplo, tal como lo describe Rinehart (1996) supra. Los ácidos nucleicos pueden estar enlazados operativamente al promotor del gen Fb12A para ser expresados de manera preferencial en células de fibra de algodón (Ibid). Véase también, John (1997) Proc. Natl. Acad. Sci. USA 89:5769-5773; John, et al, patentes de los Estados Unidos Nos. 5,608,148 y 5,602,321, que describe los promotores específicos de fibras de algodón y métodos para la construcción de plantas de algodón transgénicas. Promotores de específicos para la raíz también se pueden ser utilizados para expresar los ácidos

nucleicos de la invención. Ejemplos de promotores específicos para la raíz incluyen el promotor del gen deshidrogenasa de alcohol (DeLisle (1990) Int. Rev. Cytol. 123:39-60). Otros promotores que pueden ser utilizados para expresar los ácidos nucleicos de la invención incluyen, por ejemplo, promotores específicos para óvulo, específicos para embrión, específicos para endosperma, específicos para integumento, específicos para recubrimiento de semillas, o alguna combinación de los mismos; un promotor específico para hojas (véase, por ejemplo, Busk (1997) Plant J. 11:1285-1295, que describe un promotor específico para hojas en maíz); el promotor ORF13 de *Agrobacterium rhizogenes* (el cual exhibe actividad en raíces, véase por ejemplo, Hansen (1997) supra); un promotor específico para el polen de maíz (véase, por ejemplo, Guerrero (1990) Mol. Gen. Genet. 224:161-168); puede ser utilizado un promotor de tomate activo durante la maduración de la fruta, senescencia y abscisión, y hasta un grado mayor de flores (véase, por ejemplo, Blume (1997) Plant J. 12:731-746); un promotor específico del pistilo del gen SK2 de patata (véase, por ejemplo, Ficker (1997) Plant Mol. Biol. 35:425-431); el gen Blec4 de guisante, el cual es activo en tejido epidérmico de ápices de brotes vegetativos y florales de alfalfa transgénica convirtiéndola en una herramienta útil para direccionar la expresión de genes foráneos en la capa epidérmica de brotes o fibras crecientes activamente; el gen BEL1 específico de óvulo (véase, por ejemplo, Reiser (1995) Cell 83:735-742, GenBank No. U39944); y/o, el promotor en Klee, patente de los Estados Unidos No. 5,589,583, que describe una región promotora vegetal que es capaz de conferir altos niveles de transcripción en tejido meristemático y/o células que se dividen rápidamente.

Alternativamente, promotores vegetales que son inducibles por exposición a hormonas vegetales, tales como auxinas, son utilizados para expresar los ácidos nucleicos de la invención. Por ejemplo, la invención puede utilizar el fragmento promotor E1 de elementos de respuesta a la auxina (AuxREs) en la soja (*Glycine max* L.) (Liu (1997) Plant Physiol. 115:397-407). En el promotor GST6 de *Arabidopsis* que responde a auxina (también que responde a ácido salicílico y peróxido de hidrógeno) (Chen (1996) Plant J. 10: 955-966); el promotor parC inducible por auxina de tabaco (Sakai (1996) 37:906-913); un elemento de respuesta a biotina vegetal (Streit (1997) Mol. Plant Microbe Interact. 10:933-937); y, el promotor que responde a la hormona del estrés del ácido abscísico (Sheen (1996) Science 274:1900-1902).

Los ácidos nucleicos de la invención también pueden estar enlazados operativamente a promotores vegetales que son inducibles por exposición a reactivos químicos los cuales pueden ser aplicados a la planta, tales como herbicidas o antibióticos. Por ejemplo, el promotor In2-2 de maíz, activado por aseguradores del herbicida benceno sulfonamida, puede ser utilizado (De Veylder (1997) Plant Cell Physiol. 38:568-577); la aplicación de diferentes asegurados de herbicidas induce distintos patrones de expresión genética, incluyendo expresión en la raíz, hidátodos, y el meristema apical de brote. La secuencia de codificación puede estar bajo el control de, por ejemplo, un promotor inducible por tetraciclina, por ejemplo, como se describe con las plantas de tabaco transgénicas que contienen el gen de la arginina descarboxilasa de avena sativa L. (avena) (Masgrau (1997) Plant J. 11:465-473); o un elemento que responde al ácido salicílico, (Stange (1997) Plant J. 11:1315-1324). Utilizando promotores inducidos químicamente (por hormonas o pesticidas), esto es, promotores que responden a un agente químico que puede ser aplicado a la planta transgénica en el campo, la expresión de un polipéptido de la invención puede ser inducida en una etapa particular de desarrollo de la planta. Así, la invención también provee plantas transgénicas que contienen un gen inducible que codifica para polipéptidos de la invención cuyo rango de anfitrión está limitado a especies de la planta objetivo, tales como maíz, arroz, cebada, trigo, patata u otros cultivos, inducibles en cualquier etapa del desarrollo del cultivo.

Una persona experimentada reconocerá que un promotor vegetal específico para un tejido puede guiar la expresión de secuencias enlazadas operativamente en tejidos diferentes al tejido objetivo. Así, un promotor específico para tejido es aquel que guía la expresión preferencialmente en el tejido objetivo o tipo de célula, pero también lleva alguna expresión en otros tejidos.

Los ácidos nucleicos de la invención también pueden estar operativamente enlazados a promotores vegetales que son inducibles por exposición a reactivos químicos. Estos reactivos incluyen, por ejemplo, herbicidas, auxinas sintéticas, o antibióticos que pueden ser aplicados, por ejemplo, asperjados sobre plantas transgénicas. La expresión inducible de los ácidos nucleicos que producen transferasa, por ejemplo, transaminasa-, por ejemplo, d-aminoácido transferasa-, y/o oxidoreductasa-, por ejemplo, deshidrogenasa-, por ejemplo, d-aminoácido deshidrogenasa de la invención permitirán el crecimiento de plantas seleccionadas con la expresión y/o actividad óptima de la transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. El desarrollo de partes de la planta puede ser controlado de esta manera. De esta forma la invención provee medios para facilitar la recolección de plantas y partes de plantas. Por ejemplo, se utiliza en diversas realizaciones en el promotor In2-2 de maíz, activado por aseguradores de herbicidas de bencenosulfonamida, (De Veylder (1997) Plant Cell Physiol. 38:568-577); la aplicación de diferentes aseguradores de herbicidas induce patrones de expresión genética distintos, incluyendo la expresión en la raíz, hidátodos, y el meristema apical de brote. Las secuencias de codificación de la emisión también están bajo el control de un promotor inducible por tetraciclina, como por ejemplo, como se describe con plantas de tabaco transgénicas que contienen el gen de arginina descarboxilasa de Avena Sativa L. (Avena) (Masgrau (1997) Plant J. 11:465-473); o un elemento que transponde al ácido salicílico (Stange (1997) PlantJ. 11:1315-1324).

Algunos aspectos, la expresión aplicada de polipéptidos puede requerir la región de poliadenilación en el extremo 3' de la región de codificación. La región de poliadenilación puede ser derivada del gen natural, de una variedad de genes de otra planta (o animal u otros) o de genes de el T-ADN de agro bacteria.

El término "planta" (como por ejemplo, como en planta transgénica o en semilla de planta de esta invención, o un promotor de planta utilizado en un vector de la invención) incluye plantas completas, partes de plantas (por ejemplo, hojas, tallos, flores, raíces, etc.), protoplastos de plantas, células de semillas y plantas y progenie de las mismas. Las clases de plantas que pueden ser usadas para poner en la práctica esta invención (incluyendo composiciones y métodos) puede ser tan amplia como lo es la clase de las plantas superiores, incluyendo partes propicias para las técnicas de transformación incluyendo angiospermas, (plantas monocotiledóneas y dicotiledóneas), así como gimnospermas; también se incluye una variedad de niveles ploidies, incluyendo estados poliploides, diploides, aploides y hemizigóticos. Tal como se utiliza aquí "plantas transgénicas" incluye plantas o células de plantas en las cuales se ha insertado una secuencia de un ácido nucleico heterólogo, por ejemplo, los ácidos nucleicos y diversos constructos recombinantes (por ejemplo casetes de expresión, tales como vectores) de la invención. Las plantas transgénicas de la invención también se discuten, más adelante.

15 Vectores de expresión y vehículos de clonación

La invención provee vectores de expresión y vehículos de clonación que comprenden ácidos nucleicos de la invención, por ejemplo, secuencias que codifican las transferasas por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas de la invención. Los vectores de expresión y vehículos de clonación de la invención pueden comprender partículas virales, báculos virus, fagos, plásmidos, fagémidos, cósmidos, fósmodos, cromosomas artificiales bacterianos, ADN viral (por ejemplo, vacunas, adenovirus, viruela aviar, seudorrabias y derivados de SV40), cromosomas artificiales basados en P-1, plásmidos de lavaduras, cromosomas artificiales de levadura, y cualquier otros vectores específicos para anfitriones específicos de interés (tales como *Bacillus*, *Aspergillus* y levaduras). Los vectores de la invención pueden incluir secuencias de ADN cromosómicas, no cromosómicas y sintéticas. Se conoce un gran número de vectores adecuados por parte de los experimentados en la técnica, y están disponibles comercialmente. Vectores de ejemplo incluyen: bacterianos: vectores pQE (Qiagen), plásmidos pBluescript, vectores pNH, (vectores lambda-ZAP (Stratagene); ptrc99a, pKK223-3, pDR540, pRIT2T (Farmacia); Eucariotas: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, pSVLSV40 ((Pharmacia). Sin embargo, cualquier otro plásmido o vector puede ser utilizado en tanto sea replicable y viable en el anfitrión. Con la presente invención puede emplearse un número de copias bajo o un número de copias alto de vectores.

La expresión vector puede comprender un promotor, un sitio de enlazamiento en ribosoma para iniciación de la traducción y un terminador de la transcripción. El vector también puede incluir secuencias apropiadas para amplificar la expresión. Los sectores de expresión en mamíferos pueden comprender un origen de replicación, cualquier sitio de enlazamiento de ribosoma necesario, un sitio de poliadenilación, sitios donores y aceptores de acoplamiento, secuencia de terminación transcripcional, y secuencias no transcritas que flanquean 5'. En algunos aspectos, las secuencias de ADN derivados del acoplamiento SV40 y los sitios de poliadenilación pueden ser utilizados para proveer los elementos genéticos no transcritos requeridos.

En un aspecto, los vectores de expresión contienen uno o más genes marcadores seleccionadores para permitir la selección de células anfitrionas que contienen el vector. Tales marcadores seleccionables incluyen genes que codifican dihidrofolato reductasa o genes que confieren resistencia a la neomicina para un cultivo de células eucariotas, genes que confieren resistencia a la tetraciclina o ampicilina en *E.coli*, y el gen TRP1 de *S. cerevisiae*. A las regiones promotoras pueden ser seleccionadas de cualquier gen deseado utilizando vectores de cloranfenicol transferasa (CAT) u otros vectores con marcadores seleccionables.

Los vectores para expresar el polipéptido o fragmento del mismo en células eucariotas también puede contener potenciadores para incrementar los niveles de expresión, los potenciadores son elementos que actúan en CIS de ADN, usualmente desde aproximadamente 10 hasta aproximadamente 300 bp en longitud que actúan sobre un promotor para incrementar su transcripción. Ejemplos incluyen el potenciador SV40 en el lado tardío del origen de la replicación bp 100 a 270, el potenciador del promotor temprano del citomegalovirus, el potencializador de polioma en el lado tardío del origen de la replicación, y los potenciadores de adenovirus.

Una secuencia de ácidos nucleicos puede ser insertada en un vector mediante una variedad de procedimientos. En general, la secuencia está ligada a la posición deseada en el vector después de la digestión del inserto y el vector con endonucleasas de restricción apropiada. Alternativamente, los extremos romos tanto en el inserto como en el vector pueden ser ligados. En la técnica se conoce una variedad de técnicas de clonación, por ejemplo, como se describe en Ausubel y Sambrook. Se considera que tales procedimientos y otros están dentro del alcance de los experimentados en la técnica.

El vector puede estar en la forma de un plásmido, una partícula viral, o un fago. Otros vectores incluyen secuencias de ADN cromosómicas, no cromosómicas y sintéticas, derivados de SV40; plásmidos bacterianos, ADN de fagos,

baculovirus, plásmidos de levadura, vectores derivados de combinaciones de plásmidos y ADN de fagos, ADN viral tales como vacunas, adenovirus, virus de viruela aviar y seudorrabias. Una variedad de vectores de clonaciones y expresión para uso con anfitriones procariotas y eucariotas están descritos por, por ejemplo, Sambrook.

Los vectores bacterianos particulares que pueden ser utilizados incluyen los plásmidos disponibles comercialmente comprenden elementos genéticos del vector de clonación bien conocido pBR322 (ATCC 37017), pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden), GEM1 (Promega Biotec, Madison, WI, USA) pQE70, pQE60, pQE-9 (Qiagen), pD10, psiX174 pBluescript II KS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene), ptrc99a, pKK223-3, pKK233-3, DR540, pRIT5 (Pharmacia), pKK232-8 y pCM7. Vectores eucariotas en partícula incluyen pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, y pSVL (Pharmacia). Sin embargo, puede utilizarse cualquier otro vector en tanto sea replicable y viable en la célula anfitriona.

Los ácidos nucleicos de la invención pueden ser expresados en casetes de expresión, vectores o virus y expresados transientemente o de manera estable en células y semilla vegetales. Un sistema de expresión transiente de ejemplo utiliza sistemas de expresión episómicos, por ejemplo, el ARN viral del virus mosaico de la coliflor (CaMV) generado en el núcleo por transcripción de un minicromosoma episómico que contiene ADN súper enrollado, véase, por ejemplo, Covey (1990) Proc. Natl. Acad. Sci. USA 87:1633-1637. Alternativamente, pueden insertarse secuencias de codificación, esto es todas o subfragmentos de las secuencias de la invención en el genoma de una célula de planta anfitriona que se convierte en una parte integral del ADN cromosómico anfitrión. Los transcritos en sentido o antisentido pueden ser expresados de esta manera. Un vector que comprende las secuencias (por ejemplo, promotores o regiones de codificación) de ácidos nucleicos de la invención puede comprender un gen marcador que confiere un fenotipo seleccionable sobre una célula o semilla de una planta. Por ejemplo, el marcador puede codificar resistencia a un biosida, particularmente resistencia a un antibiótico, tal como resistencia a la Kanamicina, G418, bleomicina, higromicina, o resistencia a herbicidas, tales como resistencia a clorosulfurón o Basta.

Los vectores de expresión capaces de expresar ácidos nucleicos y proteínas en plantas son bien conocidos en la técnica, y pueden incluir, por ejemplo, vectores de *Agrobacterium spp*, virus X de la patata (véase, por ejemplo, Angell (1997) EMBO J. 16:3675-3684), virus mosaico del tabaco (véase, por ejemplo, Casper (1996) Gene 173:69-73), virus de atrofia de ramas de tomate (véase por ejemplo, Hillman (1989) Virology 169:42-50), virus de grabado del tabaco (véase, por ejemplo, Dolja (1997) Virology 234:243-252), virus mosaico dorado de judías (véase, por ejemplo, Morinaga (1993) Microbiol Immunol. 37:471-476), virus mosaico de la coliflor (véase, por ejemplo, Cecchini (1997) Mol. Plant Microbe Interact. 10:1094-1101), elemento transportable Ac/Ds del maíz (véase, por ejemplo, Rubin (1997) Mol. Cell. Biol. 17:6294-6302; Kunze (1996) Curr. Top. Microbiol. Immunol. 204:161-194), y el elemento transponible supresor-mutador del maíz (Spm) (véase, por ejemplo, Schlappi (1996) Plant Mol. Biol. 32:717-725); y derivados de los mismos.

En un aspecto, el vector de expresión puede tener dos sistemas de replicación para permitir que se mantenga en dos organismos, por ejemplo en células de mamíferos o en insectos para la expresión y en un anfitrión procariota para clonación y amplificación. Además, para vectores de expresión integradores, el vector de expresión puede contener al menos una secuencia homóloga al genoma de la célula anfitriona. Puede contener dos secuencias homólogas que flanquean en constructo de expresión. El vector de integración puede ser dirigido a un locus específico en la célula anfitriona seleccionando la secuencia homóloga apropiada para inclusión en el vector. Los constructos para vectores de integración son bien conocidos en la técnica.

Los vectores de expresión de la invención también pueden incluir un gen marcador seleccionable para permitir la selección de cepas bacterianas que hayan sido transformadas, por ejemplo, genes que hacen a las bacterias resistentes a fármacos tales como la ampicilina, cloranfenicol, eritromicina, kanamicina, neomicina y tetraciclina. Los marcadores seleccionables también pueden incluir genes biosintéticos, tales como los de las rutas biosintéticas de histidina, triptófano y leucina.

La secuencia de ADN en el vector de expresión está enlazado operativamente a secuencias de expresión apropiadas (promotoras) para dirigir la síntesis de ARN. Promotores bacterianos citados en particular incluyen *lacI*, *lacZ*, *T3*, *T7*, *gpt*, *lambda PR*, *PL* y *trp*. Los promotores eucariotas incluyen CMV inmediato temprano, HSV timidina quinasa, SV40 temprano y tardío, LTRs de retrovirus y metalotioneína-L de ratón. La selección del vector y promotor apropiados cae bien dentro del nivel de experiencia en la técnica. El vector de expresión también contiene un sitio de enlazamiento en ribosoma para la iniciación de la traducción y un terminador de la transcripción. El vector también puede incluir secuencias apropiadas para amplificar la expresión. Las regiones promotoras pueden ser seleccionadas a partir de cualquier gen deseado utilizando los vectores de cloranfenicol transferasa (CAT) y otros vectores con marcadores seleccionables. Además los vectores de expresión contienen preferiblemente uno o más genes marcadores seleccionables para proveer una característica fenotípica para la selección de células anfitrionas transformadas tales como dihidrofolato reductasa o resistencia a la neomicina para cultivo celular eucariota, o tales como resistencia a la tetraciclina o ampicilina en *E. coli*.

Los vectores de expresión en mamíferos también pueden comprender un origen de replicación, cualquier sitio de enlazamiento a ribosomas necesario, un sitio de poliadenilación, sitios de acoplamiento de donante y aceptor,

secuencias de terminación transcripcional y secuencias no transcritas que flanquean 5'. En algunos aspectos las secuencias de ADN derivadas del acoplamiento de SV40 y los sitios de poliadenilación pueden ser utilizadas para proveer los elementos genéticos no transcritos requeridos.

- 5 Los vectores para expresar el polipéptido o fragmento del mismo en células eucariotas también pueden contener potenciadores para incrementar los niveles de expresión. Los potenciadores son elementos que actúan en cis del ADN, usualmente desde aproximadamente 10 hasta aproximadamente 300 bp de longitud que actúan sobre un promotor para aumentar su transcripción. Ejemplos incluyen el potencializador de SV40 en el lado tardío del origen de replicación de bp 100 a 270, el potenciador promotor temprano de citomegalovirus, el potenciador de polioma en el lado tardío del origen de la replicación y los potenciadores de adenovirus.
- 10 Además, los vectores de expresión contienen típicamente uno o más genes marcadores seleccionables para permitir la selección de células anfitrionas que contiene el vector. Tales marcadores seleccionables incluyen genes que codifican dihidrofolato reductasa o genes que confieren resistencia a la neomicina para cultivo de células eucariotas, genes que confieren resistencia a la tetraciclina o a la ampicilina en el punto *E. coli* y el gen *TRP1S. cerevisiae*.
- 15 En algunos aspectos, el ácido nucleico que codifica uno de los polipéptidos de la invención y sus secuencias sustancialmente idénticas a los mismos, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos de los mismos está ensamblado en fase apropiada con una secuencia guía capaz de dirigir la secreción del polipéptido traducido o en fragmentos del mismo. El ácido nucleico puede codificar un polipéptido de fusión en el cual uno de los polipéptidos de la invención y la secuencia sustancialmente idénticas del mismo, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos de los mismos se fusiona en péptidos o polipéptidos heterólogos, tales como péptidos con identificación en terminal N que imparten características deseadas, tales como estabilidad incrementada o purificación simplificada.
- 20 La secuencia de ADN apropiada puede ser insertada en el vector mediante una variedad de procedimientos. En general, la secuencia de ADN está ligada a la posición deseada en el vector después de la digestión del inserto y el vector con endonucleasas de restricción apropiadas. Alternativamente, los extremos romos tanto en el inserto como en el vector pueden ser ligados. Se divulga una variedad de técnicas de clonación en Ausubel et al. *Current Protocols in Molecular Biology*, John Wiley 503 Sons, Inc. 1997 y Sambrook et al., *Molecular Cloning: A Laboratory Manual* 2nd Ed., Cold Spring Harbor Laboratory Press (1989). Se considera que tales procedimientos y otros están dentro del alcance de los experimentados en la técnica.
- 25 El vector puede estar, por ejemplo, en la forma de un plásmido, una partícula viral, o un fago. Otros vectores incluyen secuencias de ADN cromosómicas, no cromosómicas y sintéticas, derivados de SV40; plásmidos bacterianos, ADN de fago, baculovirus, plásmidos de levadura, vectores derivados a partir de combinaciones de plásmidos y ADN de fagos, ADN viral tal como vacunas, adenovirus, virus aviar y seudorrabias. Una variedad de vectores de clonación y expresión para uso con anfitriones procarionta y eucariotas es descrita por Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor, N.Y., (1989).
- 30
- 35

Células anfitrionas y células transformadas

- La invención también provee células transformadas que comprenden una secuencia de ácidos nucleicos de la invención, por ejemplo, una secuencia que codifica una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa de la invención o un vector de la invención. La célula anfitriona puede ser cualquiera de las células anfitrionas familiares para los expertos en la técnica, incluyendo células procariotas, células eucariotas, tales como células bacterianas, células fúngicas, células de levadura, células de mamíferos, células de insectos, o células vegetales. Células bacterianas de ejemplo incluyen cualquier especie dentro de los géneros *Escherichia*, *Bacillus*, *Streptomyces*, *Salmonella*, *Pseudomonas* y *Staphylococcus*, incluyendo, por ejemplo, *Escherichia coli*, *Lactococcus lactis*, *Bacillus subtilis*, *Bacillus cereus*, *Salmonella typhimurium*, *Pseudomonas fluorescens*. Células fúngicas de ejemplo incluyen cualquier especie de *Aspergillus*. Células de levadura de ejemplo incluyen cualquier especies de *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, o *Schwanniomyces*, incluyendo *Pichia pastoris*, *Saccharomyces cerevisiae*, o *Schizosaccharomyces pombe*. Células de insectos de ejemplo incluyen cualquier especie de *Spodoptera* o *Drosophila*, incluyendo *Drosophila S2* y *Spodoptera Sf9*. Células animales de ejemplo incluyen CHO, COS o melanoma de Bowes o cualquier línea celular de ratón o humano. La selección de una anfitriona apropiada está dentro de las capacidades de los experimentados en la técnica. Las técnicas para transformar una variedad amplia de especies de plantas superiores son bien conocidas y están descritas en la literatura técnica y científica. Véase, por ejemplo, Weising (1988) *Ann. Rev. Genet.* 22:421-477; patente de los Estados Unidos No. 5,750,870.
- 40
- 45
- 50
- 55 El vector puede ser introducido en las células anfitriona usando cualquiera de una variedad de técnicas incluyendo transformación, transducción, infección viral, pistolas genéticas o transferencia de genes mediada por Ti. Métodos particulares de transfección con fosfato de calcio, transfección mediada DEAE-dextrano, lipofección, o electroporación (Davis, L., Dibner, M., Battey, I., *Basic Methods in Molecular Biology*, (1986)).

- En un aspecto, los ácidos nucleicos o vectores de la invención son introducidos en las células para selección, así los ácidos nucleicos entran en las células de una manera adecuada para la expresión subsiguiente del ácido nucleico. El método de introducción está dictado principalmente por el tipo de célula objetivo, los métodos de ejemplo incluyen precipitación con CaPO_4 , fusión de liposomas, lipofección (por ejemplo, LIPOFECTIN™), electroporación, infección viral, etc. Los ácidos nucleicos candidatos pueden integrarse de manera estable en el genoma de la célula anfitriona (por ejemplo, con introducción retroviral) o pueden existir bien sea de manera transiente o de manera estable en el citoplasma (esto es, a través del uso de plásmidos tradicionales como utilizando secuencias reguladoras estándar, marcadores de selección, etc.). Como muchas selecciones farmacéuticamente importante requieren objetivos celulares humanos o mamíferos modelo, pueden usarse vectores retrovirales capaces de transfectar tales objetivos.
- 10 Cuando sea apropiado las células anfitrionas manipuladas pueden ser cultivadas en medios nutritivos convencionales modificados según sea apropiado para activar los promotores, seleccionar los transformantes o amplificar los genes de la invención. Después de la transformación de una cepa anfitriona adecuada y el crecimiento de la cepa anfitriona hasta una densidad celular apropiada, el promotor seleccionado puede ser inducido por medios apropiados (por ejemplo, desplazamiento de temperatura o inducción química) y las células pueden ser cultivadas
- 15 durante un periodo adicional para permitirles producir el polipéptido deseado o el fragmento del mismo.
- Las células pueden ser recolectadas por centrifugación, deshechas por medios físicos o químicos y el extracto crudo resultante es retenido para purificación posterior. Las células microbianas empleadas para expresión de proteínas pueden ser deshechas mediante cualquier método conveniente, incluyendo ciclos de congelamiento-descongelamiento, sonicación, ruptura mecánica o el uso de agentes parálisis celular. Tales métodos son bien
- 20 conocidos para los experimentados en la técnica. El polipéptido o fragmento del mismo expresado puede ser recuperado y purificado a partir de los cultivos de células recombinantes por métodos que incluyen precipitación con sulfato de amonio o con etanol, extracción ácida, cromatografía de intercambio de aniones o cationes, cromatografía en fosfocelulosa, cromatografía de interacción hidrófoba, cromatografía de afinidad, cromatografía en hidroxiapatita y cromatografía con lectina. Pueden utilizarse etapas de repliegue de proteínas, según sea necesario para completar
- 25 la configuración del polipéptido. Si se desea, puede emplearse cromatografía líquida de alto rendimiento (HPLC) para etapas de purificación final.
- Los constructos en la células anfitrionas pueden ser utilizados de manera convencional para producir el producto genético codificado por la secuencia recombinante. Dependiendo del anfitrión empleado en un procedimiento de producción recombinante, los polipéptidos producidos por las células anfitrionas que contienen el vector pueden ser
- 30 glicosilados o pueden ser no glicosilados. Los polipéptidos de la invención pueden o pueden no incluir también un residuo de aminoácido metionina inicial.
- Los sistemas de traducción libres de células también pueden ser empleados para producir un polipéptido de la invención. Los sistemas de traducción libres de células pueden utilizar ARNm transcritas a partir de un constructo de ADN que comprende un promotor enlazado operativamente a un ácido nucleico que modifica el polipéptido o un
- 35 fragmento del mismo. En algunos aspectos, el constructo de ADN puede ser linealizado antes de llevar cabo una reacción de transcripción *in vitro*. El ARNm transcrito es incubado entonces con un extracto de traducción libre de células, tal como un extracto de reticulocitos de conejo para producir el polipéptido o fragmento del mismo deseados.
- Los vectores de expresión pueden contener uno o más genes marcadores seleccionables para proveer una característica fenotípica para la selección de células anfitrionas transformadas tales como dihidrofolato reductasa o
- 40 resistencia a la neomicina para cultivos celulares eucariotas, o tal como resistencia a la tetraciclina o ampicilina en *E. coli*.
- Células anfitrionas que contienen los polinucleótidos de interés, como por ejemplo, ácidos nucleicos de la invención pueden ser cultivadas en medios nutritivos convencionales modificados según sea apropiado para la activación de los promotores, la selección de los transformantes o la amplificación de los genes. Las condiciones de cultivo, tales
- 45 como la temperatura, pH y similares, son las usadas previamente con las células anfitrionas seleccionadas para la expresión y serán evidentes para la persona con experiencia normal. Los clones que son identificados por tener la actividad enzimática especificada pueden ser secuenciados entonces para identificar la secuencia de polinucleótidos que codifica una enzima que tiene la actividad potenciada.
- La invención provee un método para sobreexpresar una d-aminoácido transferasa recombinante en una célula que
- 50 comprende expresar un vector que comprende un ácido nucleico de la invención, por ejemplo un ácido nucleico que comprende una secuencia de ácidos nucleicos con al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más identidades de secuencia con una secuencia de SEQ ID NO: 219 a lo largo de una región de al menos aproximadamente 100 residuos, en donde las identidades de secuencia son determinadas mediante análisis con un
- 55 algoritmo de comparación de secuencias o mediante inspección visual, o, un ácido nucleico que hibrida bajo condiciones restrictivas a una secuencia de ácidos nucleicos de la invención, o una subsecuencia del mismo. La sobre expresión puede ser efectuada por cualquier medio, por ejemplo, mediante el uso de un promotor de alta actividad, o un vector dicistrónico o mediante amplificación genética del vector.

- Loa ácidos nucleicos de la invención pueden ser expresados, o sobreexpresados, en cualquier sistema de expresión *in vitro* o *in vivo*. Cualquier sistema de cultivo celular puede ser empleado para expresar, o sobreexpresar, proteínas recombinantes, incluyendo cultivos bacterianos, de insectos, de levaduras, fúngicos o de mamíferos. La sobreexpresión puede ser efectuada mediante la selección apropiada de promotores, potenciadores, vectores (por ejemplo, el uso de vectores replicón, vectores dicistrónicos (véase, por ejemplo, Gurtu (1996) *Biochem. Biophys. Res. Commun.* 229:295-8)), medios, sistemas de cultivo y similares. En un aspecto, la amplificación de genes utilizando marcadores de selección, por ejemplo glutamina sintetasa (véase, por ejemplo, Sanders (1987) *Dev. Biol. Stand.* 66:55-63), en sistemas celulares se utiliza para sobreexpresar los polipéptidos de la invención.
- Detalles adicionales con respecto a esta metodología están en la literatura pública y/o son conocidos por la persona experimentada. Una ejemplificación particular no limitante, tal literatura disponible públicamente incluye: *Journal of Bacteriology*. 1998 August. 180(16):4319-4323; *Applied Microbiology and Biotechnology*. 2003 June. 61(5-6):463-471; *Gene*. 1996 Oct. 177(1):217-222; *J Bacteriol.* 1994 June; 176(12): 3552-3558; *JBC*. 1997 Sept. 272(37):23303-23311, aunque estas referencias no enseñan las enzimas de la invención de la presente solicitud.
- La célula anfitriona puede ser cualquiera de las células anfitrionas familiares para los experimentados en la técnica, incluyendo células procariotas, células eucariota, células de mamíferos, células de insectos o células de plantas. Como ejemplos representativos de anfitriones apropiados, pueden mencionarse: células bacterianas como *E. coli*, *Streptomyces*, *Bacillus subtilis*, *Bacillus cereus*, *Salmonella typhimurium* y diversas especies dentro de los géneros *Pseudomonas*, *Streptomyces* y *Staphylococcus*, células fúngicas, tales como, *Aspergillus*, levaduras tales como cualquier especie de *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, *Schwanniomyces*, incluyendo *Pichia pastoris*, *Saccharomyces cerevisiae*, o *Schizosaccharomyces pombe*, células de insectos tales como *Drosophila S2* and *Spodoptera*, Sf9, células animales tales como CHO, COS o melanoma de Bowes o adenovirus. La selección de un anfitrión apropiado está dentro de las capacidades de los experimentados en la técnica.
- El vector puede ser introducido dentro de las células anfitrionas utilizando cualquiera de una variedad de técnicas, incluyendo transformación, transfección, transducción, infección viral, pistolas genéticas, o transferencia genética mediada por Ti. Los métodos particulares incluyen transfección con fosfato de calcio, transfección mediada por DEAE-Dextrano, lipofección, o electroporación (Davis, L., Dibner, M., Battey, I., *Basic Methods in Molecular Biology*, (1986)).
- Cuando sea apropiado, las células anfitrionas manipuladas pueden ser cultivadas en medios nutritivos convencionales modificados según sea apropiado para activar promotores, seleccionar transformantes o amplificar los genes de la invención. Después de la transformación de una cepa anfitriona adecuada y el crecimiento de la cepa anfitriona hasta una densidad celular apropiada, el promotor seleccionado puede ser inducido por medios apropiados (por ejemplo, desplazamiento de temperatura o inducción química) y las células pueden ser cultivadas durante un periodo adicional para permitirles producir el polipéptido o fragmento del mismo deseado.
- Las células son recolectadas típicamente por centrifugación, deshechas por medios físicos o químicos y el extracto crudo resultante puede ser retenido para purificación posterior. Las células microbianas entradas para la inspección de proteínas pueden ser deshechas por cualquier método conveniente, incluyendo ciclos de congelamiento-descongelamiento, sonicación, ruptura mecánica o uso de agentes de lisis celular. Tales métodos son bien conocidos para los experimentados en la técnica. El polipéptido o fragmento del mismo expresado puede ser recuperado y purificado a partir de los cultivos celulares recombinantes por métodos que incluyen precipitación con sulfato de amonio o con etanol, extracción ácida, cromatografía de intercambio de aniones o cationes, cromatografía de fosfocelulosa, cromatografía de interacción hidrófoba, cromatografía de afinidad, cromatografía en hidroxipatita y cromatografía con lectina. Pueden utilizarse etapas de repliegue de proteínas, según sea necesario, para completar la configuración del polipéptido. Si se desea, puede emplearse cromatografía líquida de alto rendimiento (HPLC) para las etapas de purificación final.
- Pueden emplearse también diversos sistemas de cultivo celular de mamíferos para expresar proteínas recombinantes. Ejemplos de sistemas de expresión de mamíferos incluyen las líneas COS-7 de fibroblastos de riñón de mono (descritos por Gluzman, *Cell*, 23: 175, 1981) y otras líneas celulares capaces de expresar proteínas a partir de un vector compatible, tales como las líneas celulares C127, 3T3, CHO, HeLa y BHK.
- Los constructos en la células anfitrionas pueden ser utilizados en una forma convencional para producir el producto genético codificado por la secuencia recombinante. Dependiendo del anfitrión empleado en un procedimiento de producción recombinante, los polipéptidos producidos por células anfitrionas que contienen el vector pueden ser glicosilados o pueden ser no glicosilados. Los polipéptidos de la invención pueden o pueden no incluir también un residuo de aminoácido metionina inicial.
- Alternativamente, los polipéptidos de las secuencias de aminoácidos de la invención, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos de los mismos pueden ser producidos sintéticamente por sintetizadores de péptidos convencionales. En otros aspectos, los fragmentos o porciones de los polipéptidos pueden ser empleados para producir el correspondiente polipéptido de longitud

completa mediante síntesis de péptidos; por lo tanto, los fragmentos pueden ser empleados como intermediarios para producir los polipéptidos de longitud completa.

Los sistemas de traducción libres de células también pueden ser empleados para producir unos de los polipéptidos de las secuencias de aminoácidos de la invención, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 de aminoácidos consecutivos del mismo utilizando ARNm transcritos a partir de un constructo de ADN que comprende un promotor enlazado operativamente a un ácido nucleico que codifica el polipéptido o fragmento del mismo. En algunos aspectos el constructo de ADN puede ser linealizado antes de llevar a cabo una reacción de transcripción *in vitro*. El ARNm transcrito es incubado entonces con un extracto de traducción libre de células apropiado tal como extracto de reticulocitos de conejo, para producir el polipéptido o fragmento del mismo deseado.

Amplificación de ácidos nucleicos

En la práctica, de la invención, los ácidos nucleicos de la invención son ácidos nucleicos que codifican las d-aminoácidos transferasas de la invención, o ácidos nucleicos modificados de la invención, pueden ser reproducidos por amplificación. La amplificación también puede ser utilizada para clonar o modificar los ácidos nucleicos de la invención. Así, la invención provee pares de secuencias de cebadores de amplificación para amplificar ácidos nucleicos de la invención. Una persona de experiencia en la técnica puede diseñar pares de secuencia de cebadores de amplificación para cualquier parte o para la longitud completa de estas secuencias.

En un aspecto, la invención provee un ácido nucleico amplificado por un par de cebadores de la invención, por ejemplo, un par de cebadores tal como se establece por aproximadamente los primeros (el 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 residuos de un ácido nucleico de la invención, de aproximadamente el primero (el 5') 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 residuos de la cadena complementaria.

La invención provee un par de secuencias de cebadores de amplificación para amplificar un ácido nucleico que codifica un polipéptido que tiene una actividad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, en donde el primer par es capaz de amplificar un ácido nucleico que comprende una secuencia de la invención o fragmentos o subsecuencias del mismo. Uno o cada miembro del par de secuencias de cebadores de amplificación pueden comprender un oligonucleótido que comprende al menos aproximadamente 10 a 50 bases consecutivas de la secuencia, o aproximadamente 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 bases consecutivas de la secuencia. La invención provee pares de cebadores de amplificación, en donde el primer par comprende un primer miembro que tiene una secuencia tal como se establece mediante aproximadamente el primero (el 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 residuos de un ácido nucleico de la invención, y un segundo miembro que tiene una secuencia como se establece por aproximadamente el primero (el 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 residuos de la cadena complementaria del primer miembro. La invención provee d-aminoácido transferasas, generadas por amplificación, por ejemplo por reacción en cadena de polimerasa (PCR), utilizando un par de cebadores de amplificación de la invención. La invención provee métodos para hacer una transferasa, por ejemplo por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa por amplificación, por ejemplo, por reacción en cadena de polimerasa (PCR) utilizando un par de cebadores de amplificación de la invención. En un aspecto, el par de cebadores de amplificación amplifica un ácido nucleico a partir de una biblioteca, por ejemplo, una biblioteca de genes, tal como una biblioteca ambiental.

Las reacciones de amplificación también pueden ser utilizadas para cuantificar la cantidad de ácido nucleico en una muestra (tal como la cantidad de mensaje en una muestra celular), marcar el ácido nucleico (por ejemplo aplicarlo a un arreglo o a una siembra), detectar el ácido nucleico, o cuantificar la cantidad de un ácido nucleico específico en una muestra. En un aspecto de la invención, los mensajes aislados a partir de una célula o de una biblioteca de ADNc son amplificados.

La persona experimentada puede seleccionar y diseñar cebadores de amplificación oligonucleótidos adecuados. Los métodos de amplificación son bien conocidos en la técnica e incluyen, por ejemplo, la reacción en cadena de polimerasa, PCR (véase, por ejemplo, PCR PROTOCOLS, AGUIDE TO METHODS AND APPLICATIONS, ed. Innis, Academic Press, N.Y. (1990) and PCR STRATEGIES (1995), ed. Innis, Academic Press, Inc., N.Y., reacción en cadena de ligasa (LCR) (véase, por ejemplo, Wu (1989) Genomics 4:560; Landegren (1988) Science 241:1077; Barringer (1990) Gene 89:117); amplificación de transcripción (véase, por ejemplo, Kwok (1989) Proc. Natl. Acad. Sci. USA 86:1173); y, replicación de secuencia auto sostenida (véase, por ejemplo, Guatelli (1990) Proc. Natl. Acad. Sci. USA 87:1874); amplificación de Q Beta replicasa (véase, por ejemplo, Smith (1997) J. Clin. Microbiol. 35:1477-1491), ensayo de amplificación de Q-Beta replicasa automatizado (véase, por ejemplo, Burg (1996) Mol. Cell. Probes 10:257-271) y otras técnicas mediadas por ARN polimerasa (por ejemplo, NASBA, Cangene, Mississauga, Ontario); véase también Berger (1987) Methods Enzymol. 152:307-316; Sambrook; Ausubel; patente de los Estados Unidos Nos. 4,683,195 y 4,683,202; Sooknanan (1995) Biotechnology 13:563-564.

Determinación del grado de identidad de secuencia

La invención provee ácidos nucleicos aislados, recombinantes y/o sintéticos que comprenden secuencias que tienen al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia con la SEQ ID NO: 219 a lo largo de una región de al menos 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550 o más, residuos. La invención provee polipéptidos que comprenden secuencias que tienen al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia con SEQ ID NO: 220. El grado de identidad de secuencia (homología) puede ser determinado utilizando cualquier programa de ordenador y parámetros asociados, incluyendo los descritos aquí, tales como BLAST 2.2.2. o FASTA versión 3.0t78, con los parámetros por sistema.

En realizaciones alternativas los términos “ordenador”, “programa de ordenador” y “procesador” se utilizan en sus contextos generales más amplios e incorporan todos los dichos dispositivos, tal como se describe en detalle más adelante. Una “secuencia de codificación de” o una “secuencia que codifica” en un polipéptido o proteína en particular, es una secuencia de ácidos nucleicos que es transcrita y traducida en un polipéptido o proteína cuando se coloca bajo el control de secuencias reguladoras apropiadas.

En realizaciones alternativas, la expresión “sustancialmente idénticos” en el contexto de dos ácidos nucleicos o polipéptidos se refiere a dos o más secuencias que tiene, por ejemplo, al menos un 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más de identidad en residuos de nucleótidos o aminoácidos (secuencia) como cuando se comparan y alinean para máxima correspondencia, tal como se miden utilizando uno de los algoritmos de comparación de secuencia conocidos o por inspección visual. Típicamente, la identidad sustancial existe a lo largo de una región de al menos aproximadamente 100 residuos y lo más comúnmente las secuencias son sustancialmente idénticas a lo largo de al menos 150-200 residuos. En algunos aspectos, las secuencias son sustancialmente idénticas a lo largo de la longitud completa de las regiones de codificación.

En realizaciones alternativas, una secuencia de aminoácidos “sustancialmente idéntica” es una secuencia que difiere de una secuencia de referencia en una o más sustituciones, eliminaciones o reinserciones de aminoácidos conservadoras o no conservadoras, particularmente cuando tal sustitución ocurre en un sitio que no es un sitio activo de la molécula y con la condición de que el polipéptido retenga esencialmente sus propiedades funcionales. En realizaciones alternativas una sustitución de aminoácidos conservadora, por ejemplo, sustituye un aminoácido por otro de la misma clase (por ejemplo, la sustitución de un aminoácido hidrófobo, tal como isoleucina, valina, leucina o metionina, por otro, o sustitución de un aminoácido polar por otro, tal como una sustitución de arginina por lisina, ácido glutámico por ácido aspártico o glutamina por asparagina). Uno o más aminoácidos pueden ser eliminados, por ejemplo, de una transferasa, por ejemplo, de un polipéptido de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, dando como resultado una modificación de la estructura del polipéptido, sin alterar significativamente su actividad biológica. Por ejemplo los aminoácidos amino o carboxilo terminales que no son requeridos para la actividad biológica de la transferasa, por ejemplo transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa pueden ser removidos. Las secuencias de polipéptido modificadas de la invención pueden ser ensayadas en cuanto a la actividad biológica de transferasa, por ejemplo, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa por cualquier número de métodos, incluyendo poner en contacto la secuencia de polipéptidos modificada con un sustrato de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa y determinar si el polipéptido modificado disminuye la cantidad de sustrato específico en la prueba o incrementa los bioproductos de la reacción enzimática funcional de una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. La secuencia de ácidos nucleicos subsecuencias sustancialmente idénticas a las mismas, con referencia a una secuencia que tiene al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más identidad de secuencia (homología) a estas secuencias. La homología puede ser determinada utilizando cualquiera de los programas de ordenador y parámetros descritos aquí, incluyendo FASTA versión 3.0t78 con los parámetros por sistema. Las secuencias homólogas también pueden incluir secuencias de ARN en las cuales uridinas remplazan las timinas en las secuencias de ácidos nucleicos de la invención. Las secuencias homólogas pueden ser obtenidas utilizando cualquiera de los procedimientos descritos aquí, o pueden ser el resultado de la corrección de un error de secuenciación. Será evidente que las secuencias de ácidos nucleicos de la invención y las secuencias sustancialmente idénticas a las mismas pueden ser representadas en el formato de caracteres individuales tradicionalmente (véase la cubierta interna de Stryer, Lubert. Biochemistry, 3rd Ed., W. H Freeman & Co., New York.) o en cualquier otro formato que registre la identidad de los nucleótidos en una secuencia.

Diversos programas de comparación de secuencias identificados en cualquier lugar de esta especificación de patente se contemplan particularmente para su uso en este aspecto de la invención. Las homologías en las secuencias de proteínas y/o ácidos nucleicos pueden ser evaluados utilizando cualquiera de una variedad de

algoritmos de comparación de secuencias y programas conocidos en la técnica. Tales algoritmos y programas incluyen, pero de ninguna manera se limitan a, TBLASTN, BLASTP, FASTA, TFASTA y CLUSTALW (Pearson and Lipman, Proc. Natl. Acad. Sci. USA 85(8):2444-2448, 1988; Altschul et al., J. Mol. Biol. 215(3):403-410, 1990; Thompson et al., Nucleic Acids Res. 22(2):4673-4680, 1994; Higgins et al., Methods Enzymol. 266:383-402, 1996; Altschul et al., J. Mol. Biol. 215(3):403-410, 1990; Altschul et al., Nature Genetics 3:266-272, 1993).

La homología o identidad se mide frecuentemente utilizando un software de análisis de secuencia (por ejemplo, el Sequence Analysis Software Package del Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). Tal software compara secuencias similares asignando grados de homología a diversas eliminaciones, sustituciones y otras modificaciones. Los términos "homología" e "identidad" en el contexto de dos o más secuencias de ácidos nucleicos o polipéptidos, se refiere a dos o más secuencias o subsecuencias que son la misma o tienen un porcentaje especificado de residuos de aminoácidos o nucleótidos que son los mismos cuando se comparan y se alinean para una máxima correspondencia a lo largo de una ventana de comparación o una región designada según se mide utilizando cualquier número algoritmos de comparación de secuencias o mediante alineamiento manual e inspección visual.

Para la comparación de las secuencias, típicamente una secuencia actúa como una secuencia de referencia, contra la cual se comparan las secuencias de prueba. Cuando se utiliza un algoritmo de comparación de secuencias, las secuencias de prueba de referencias son introducidas en un ordenador, se designan las coordenadas de subsecuencia, si es necesario y se designan los parámetros del programa de algoritmo de secuencias. Pueden ser utilizados los parámetros del programa por sistema, o pueden designarse parámetros alternativos. El algoritmo de comparación de secuencias calcula entonces el porcentaje de identidad en secuencias para las secuencias de prueba con respecto a la secuencia de referencia, con base en los parámetros del programa.

En realizaciones alternativas, una "ventana de comparación" incluye referencia a un segmento de cualquiera de un cierto número de posiciones contiguas seleccionadas del grupo consistente desde 20 a 600, usualmente de manera aproximada de 50 a aproximadamente 200, más usualmente de manera aproximada de 100 a aproximadamente 150 en la cual una secuencia puede ser comparada con una secuencia de referencia del mismo número de posiciones contiguas después de que las dos secuencias son alineadas de manera óptima. Métodos de alineamiento de secuencias para comparación son bien conocidos en la técnica. El alineamiento óptimo de secuencias para comparación puede llevarse a cabo, por ejemplo, mediante el algoritmo de homología local de Smith & Waterman, Adv. Appl. Math. 2:482, 1981, mediante el algoritmo de alineamiento de homología de Needleman & Wunsch, J. Mol. Biol. 48:443, 1970, por la búsqueda por el métodos de similitud de Person & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444, por implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI) o por alineamiento manual e inspección visual. Otros algoritmos para determinar la homología o identidad incluyen, por ejemplo, además de un programa BLAST (Basic Local Alignment Search Tool at the National Center for Biological Information), ALIGN, AMAS (Analysis of Multiply Aligned Sequences), AMPS (Protein Multiple Sequence Alignment), ASSET (Aligned Segment Statistical Evaluation Tool), BANDS, BESTSCOR, BIOSCAN (Biological Sequence Comparative Analysis Node), BLIMPS (Blocks IMProved Searcher), FASTA, Intervals & Points, BMB, CLUSTAL V, CLUSTAL W, CONSENSUS, LCONSENSUS, WCONSENSUS, Smith-Waterman algorithm, DARWIN, Las Vegas algorithm, FNAT (Forced Nucleotide Alignment Tool), Framealign, Framesearch, DYNAMIC, FILTER, FSAP (Fristensky Sequence Analysis Package), GAP (Global Alignment Program), GENAL, GIBBS, GenQuest, ISSC (Sensitive Sequence Comparison), LALIGN (Local Sequence Alignment), LCP (Local Content Program), MACAW (Multiple Alignment Construction & Analysis Workbench), MAP (Multiple Alignment Program), MBLKP, MBLKN, PIMA (Pattern-Induced Multi-sequence Alignment), SAGA (Sequence Alignment by Genetic Algorithm) y WHAT-IF. Tales programas de alineamiento también pueden ser utilizados para seleccionar bases de datos de genomas para identificar secuencias de polinucleótidos que tienen secuencias sustancialmente idénticas. Hay disponible un cierto número de bases de datos de genomas, por ejemplo, una porción sustancial del genoma humano está disponible como parte del Proyecto de Secuenciación del Genoma Humano. Al menos otros veintidós genomas han sido secuenciados ya, incluyendo, por ejemplo, *M. genitalium* (Fraser et al., 1995), *M. jannaschii* (Bult et al., 1996), *H. influenzae* (Fleischmann et al., 1995), *E. coli* (Blattner et al., 1997) y levadura (*S. cerevisiae*) (Mewes et al., 1997) y *D. melanogaster* (Adams et al., 2000). Se han hecho también progresos significativos en secuencia a los genomas de un organismo modelo tal como ratón, *C. elegans* y *Arabidopsis* sp. Varias bases de datos que contienen información genética anotada con alguna información funcional son mantenidas por diferentes organizaciones y están accesibles a través del internet.

Un ejemplo es los algoritmos BLAST y BLAST 2.0, los cuales están descritos en Altschul et al., Nuc. Acids Res. 25:3389-3402, 1977 y Altschul et al., J. Mol. Biol. 215:403-410, 1990, respectivamente. El software para llevar a cabo análisis BLAST está disponible públicamente a través del National Center for Biotechnology Information. Este algoritmo involucra primero identificar pares de secuencia de alta calificación (HSPs) identificando palabras cortas de longitud W en la secuencia de búsqueda, que bien coinciden o satisfacen algunos marcadores de umbral de valor positivo cuando se alinean con una palabra de la misma longitud de una secuencia de una base de datos. T es conocido como el umbral de marcación de palabra vecina (Altschul et al., supra). Esta palabra vecina inicial actúa como semilla para iniciar búsquedas para encontrar HSPs más largos que las contiene. Los aciertos de palabras son

extendidos en ambas direcciones a lo largo de cada secuencia hasta tanto pueda incrementarse el marcador de alineamiento acumulativo. Los marcadores acumulativos son calculados utilizando, para secuencias de nucleótidos, los parámetros M (marcador de recompensa para un par de residuos coincidentes; siempre > 0) para secuencias de aminoácidos, se utiliza una matriz de marcación para calcular el marcador acumulativo. La extensión de los aciertos de palabra en cada dirección se tiene cuando: la marcación del alineamiento acumulativo cae en la cantidad X a partir de su valor alcanzado máximo; el marcador acumulativo llega a cero o por debajo de cero, debido a la acumulación de uno o más alineamientos de residuos con marcación negativa; o cuando el final de cada secuencia es alcanzado. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad de velocidad del alineamiento. El programa BLASTN para secuencias de nucleótidos utiliza como sistema una longitud de palabra (W) de 11, una expectativa (E) de 10, $M = 5$, $N = -4$ y una comparación de ambas cadenas. Para secuencias de aminoácidos, el programa BLASTP utiliza como sistema una longitud de palabra de tres y expectativas (E) de 10 y la matriz de marcación BLOSUM62 (véase Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915, 1989) alineamientos (B) de 50, expectativa (E) de 10, $M = 5$, $N = -4$ y una comparación de ambas cadenas.

El algoritmo BLAST también lleva a cabo análisis estadístico de similitud entre dos secuencias (véase, por ejemplo, Karlin & Altschul, Proc. Natl. Acad. Sci. USA 90:5873, 1993). Una medida de la similitud provista por el algoritmo BLAST es la probabilidad de suma más pequeña ($P(N)$), lo cual provee una indicación de la probabilidad con la cual una coincidencia entre dos secuencias de nucleótidos o aminoácidos ocurriría por casualidad. Por ejemplo, un ácido nucleico es considerado similar a la secuencia de referencia si la probabilidad de la suma más pequeña en una comparación del ácido nucleico de prueba y el ácido nucleico de referencia es menor de aproximadamente 0.2 más preferiblemente menor de aproximadamente 0.01 y lo más preferiblemente menor de aproximadamente 0.001.

En un aspecto, las homologías de secuencias de proteínas y ácidos nucleicos se evalúan utilizando el Basic Local Alignment Search Tool ("BLAST"). En particular, se utilizan cinco programas BLAST específicos para llevar a cabo la siguiente tarea:

(1) BLASTP y BLAST3 comparan una secuencia de búsqueda de aminoácidos contra una base de datos de secuencias de proteínas;

(2) BLASTN compara una secuencia de búsqueda de nucleótidos contra una base de datos de secuencia de nucleótidos;

(3) BLASTX compara los cinco productos de traducción conceptual de seis marcos de una secuencia de nucleótidos de búsqueda (ambas cadenas) contra una base de datos de secuencias de proteínas;

(4) TBLASTN compara una secuencia de proteínas de búsqueda contra una base de datos de secuencias de nucleótidos traducida en todos los seis marcos de lectura (ambas cadenas); y

(5) TBLASTX compara las traducciones en seis marcos de una secuencia de búsqueda de nucleótidos contra las traducciones en seis marcos de una base de datos de secuencias de nucleótidos.

Los programas BLAST identifican secuencia homóloga identificando segmento similares, los cuales se denominan aquí como "pares de segmentos de alta marcación", entre una secuencia de búsqueda de aminoácidos o ácidos nucleicos y una secuencia de prueba que se obtienen preferiblemente a partir de una base de datos de secuencias de proteínas o ácidos nucleicos. Los pares de segmentos de alta marcación son identificados preferiblemente (esto es alineados) por medio de una matriz de marcación, muchas de las cuales son conocidas en la técnica. Preferiblemente, la matriz de marcación utilizada es la matriz BLOSUM62 (Gonnet *et al.*, Science 256:1443-1445, 1992; Henikoff and Henikoff, Proteins 17:49-61, 1993). Menos preferiblemente, pueden utilizarse también las matrices PAM o PAM250, (véase, por ejemplo, Schwartz y Dayhoff, eds., 1978, *Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure*, Washington: National Biomedical Research Foundation). Los programas BLAST son accesibles a través de la U.S. National Library of Medicine.

Los parámetros usados con los algoritmos anteriores pueden ser adaptados dependiendo de la longitud de la secuencia y grado de homología estudiado. En algunos aspectos los parámetros pueden ser parámetros por sistema utilizados por los algoritmos en la ausencia de soluciones de parte del usuario.

Sistemas de ordenador y productos de programa de ordenador

Para determinar e identificar identidades en secuencias, homologías estructurales motivos y similares *in silico*, una secuencia de un ácido nucleico o un polipéptido de la invención puede ser almacenada, registrada y manipulada sobre cualquier medio que pueda ser leído y al cual se tenga acceso mediante un ordenador.

De acuerdo con lo anterior, la invención provee ordenadores, sistemas de ordenador, medios legibles por ordenador, productos de programas para ordenador y similares teniendo grabados o almacenados en los mismos las secuencias de ácidos nucleicos y polipéptidos de la invención. Tal como se utilizan aquí las palabras "grabados" y "almacenados" se refiere a un proceso para almacenar información sobre un medio de ordenador.

Una persona experimentada puede adoptar fácilmente cualquier método conocido para grabar información sobre un medio legible en un ordenador para generar productos que comprenden una o más de las secuencias del ácido nucleico y/o del polipéptido de la invención.

- 5 Los polipéptidos de la invención excluyen las secuencias de ejemplo de la invención y secuencias sustancialmente idénticas al mismo, y fragmentos de cualquiera de las secuencias precedentes. Secuencias sustancialmente idénticas, u homólogas, de polipéptidos se refieren a una secuencia de polipéptidos que tienen al menos 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de identidad de secuencia con SEQ ID NO: 220.

- 10 La homología puede ser determinada utilizando cualquiera de los programas de ordenador y parámetros descritos aquí, incluyendo FASTA versión 3.0t78 con los parámetros de fondo o con cualquier parámetro modificado. Las secuencias homólogas pueden ser obtenidas utilizando cualquiera de los procedimientos descritos aquí o pueden resultar de la corrección de un error de secuenciación. Los fragmentos de polipéptidos comprenden al menos 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500 o más aminoácidos consecutivos de los polipéptidos de la invención y secuencias sustancialmente idénticas de los mismos. Será evidentemente que los
- 15 códigos de polipéptidos de las secuencias de aminoácidos de la invención y secuencias sustancialmente idénticas a los mismos, pueden ser representados por el formato tradicional de carácter individual o el formato de tres letras (véase Stryer, Lubert. Biochemistry, 3rd Ed., supra) o cualquier otro formato que relacione la identidad de los polipéptidos en una secuencia.

- 20 Una secuencia de ácidos nucleicos o polipéptidos de la invención puede ser almacenada, grabada y manipulada en cualquier medio que pueda ser leído y al cual se pueda tener acceso mediante un ordenador. Tal como se utilizan aquí las palabras "grabado" y "almacenados" se refieren a un proceso para almacenar información sobre un medio de ordenador. Una persona experimentada puede adoptar fácilmente cualquiera de los métodos actualmente conocidos para grabar información sobre un medio legible en ordenador para generar productos que comprendan una o más de las secuencias de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las
- 25 mismas, o una o más de las secuencias de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas. Otro aspecto de la invención es un medio legible en ordenador que ha grabado sobre el mismo al menos 2, 5, 10, 15, o 20 o más secuencias de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas.

- 30 La invención describe adicionalmente un medio legible en ordenador que tiene grabadas sobre sí una o más de las secuencias de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas. También se describe un medio legible en ordenador que tiene registrado sobre sí mismo una o más de las secuencias de polipéptido de la invención y secuencias sustancialmente idénticas a las mismas. También se describe un medio legible en ordenador que tiene registradas en sí mismo al menos 2, 5, 10, 15, o 20 o más de las secuencias tal como se estableció anteriormente.

- 35 Los medios legible en ordenador incluyen medios legibles magnéticamente, medios legibles ópticamente, medio legibles electrónicamente y medios magnéticos/ópticos. Por ejemplo, los medios legibles por el ordenador pueden ser un disco duro, un disco flexible, una cinta magnética, CD-ROM, un disco digital versátil (DVD), Memoria de Acceso Aleatorio (RAM) o Memoria Solo de Lectura (ROM) así como otros tipos de otros medios conocidos para los experimentados en la técnica.

- 40 También se describe sistemas (por ejemplo, sistemas basados en internet), particularmente sistemas de ordenador que almacenan y manipulan la información de secuencias descrita aquí. Un ejemplo de un sistema 100 de un ordenador es ilustrado en el diagrama de bloques de la figura 1. Tal como se utiliza aquí, un "sistema de ordenador" se refiere a los componentes de hardware, a los componentes de software y componentes para almacenamiento de datos utilizados para analizar una secuencia de nucleótidos de una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o una secuencia de polipéptidos tal como se establece en las
- 45 secuencias de aminoácidos de la invención. El sistema 100 de ordenador incluye típicamente un procesador para procesar, tener acceso y manipular los datos de las secuencias. El procesador 105 puede ser cualquier tipo bien conocido de unidad de procesamiento central, tal como, por ejemplo, el Pentium 3 de Intel Corporation o un procesador similar de Sun, Motorola, Compaq, AMD o International Business Machines.

- 50 Típicamente el sistema 100 de ordenador es un sistema de propósito general que comprende el procesador 105 y uno o más componentes 110 de almacenamiento de datos internos para almacenar datos y uno o más dispositivos para recuperación de datos para recuperar los datos almacenados en los componentes de almacenamiento de datos. Una persona experimentada puede apreciar fácilmente que cualquiera de los sistemas de ordenadores disponibles actualmente son adecuados.

- 55 En un aspecto particular, el sistema 100 de ordenar incluye un procesador 105 conectado a un bus el cual está conectado a una memoria principal 115 (implementada preferiblemente como RAM) y uno o más dispositivos 110 de almacenamiento de datos internos, tales como un disco duro y/o otros medios legibles por ordenador que tienen datos registrados sobre sí mismos. En algunos aspectos, el sistema 100 de ordenador incluye adicionalmente uno o

más dispositivos 118 para recuperación de datos para leer los datos almacenados en los dispositivos 110 de almacenamiento de datos internos.

El dispositivo 118 de recuperación de datos puede representar, por ejemplo, un lector de discos flexibles, un lector de discos compactos, un lector de cintas magnéticas, o un módem capaz de conectarse a un sistema de almacenamiento de datos remoto (por ejemplo, a través del internet) etc.. En algunos aspectos, el dispositivo 110 de almacenamiento de datos interno es un medio legible por ordenador removible tal como un disco flexible, un disco compacto, una cinta magnética, etc., que contiene programas de control y/o datos grabados en el mismo. El sistema 100 de ordenador puede incluir ventajosamente o ser programado mediante software apropiado para leer los programas de control y/o datos del componente de almacenamiento de datos una vez insertado en el dispositivo de recuperación de datos.

El sistema 100 de ordenador incluye una pantalla 120 que se utiliza para presentar datos al usuario del ordenador. Ha de notarse también que el sistema 100 de ordenador puede estar conectado con otros sistemas de ordenador 125a-c en una red o en una red de área amplia para proveer acceso centralizado al sistema 100 de ordenador.

El software para tener acceso y procesar las secuencias de nucleótidos de una secuencias de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o una secuencia de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas, (tales como herramientas de búsqueda, herramientas de comparación y herramientas de modelación, etc.) pueden recibir en la memoria principal 115 durante la ejecución.

En algunos aspectos, el sistema 100 de ordenador puede comprender adicionalmente un algoritmo de comparación de secuencias para comparar una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o una secuencia de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas, almacenadas en un medio legible por un ordenador a un nucleótido de referencia o a secuencias de referencia de nucleótidos o polipéptidos almacenadas en un medio legible por ordenador. Un "algoritmo de comparación de secuencias" se refiere a uno o más programas que son implementados (por vía local o remota) sobre el sistema 100 de ordenador para comparar una secuencia de nucleótidos con otras secuencias de nucleótidos y/o compuestos almacenados dentro de un medio de almacenamiento de datos. Por ejemplo, el algoritmo de comparación de secuencias puede comparar las secuencias de nucleótidos de una secuencias de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o una secuencia de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas, almacenadas sobre un medio legible por ordenador contra secuencias de referencia almacenadas en un medio legible en ordenador para identificar homologías o motivos estructurales.

La figura 2 es un diagrama de flujo que ilustra un aspecto del proceso 200 para comparar una nueva secuencia de nucleótidos o proteínas con una base de datos de secuencias con el fin de determinar los niveles de homología entre la nueva secuencia y las secuencias en la bases de datos. La base de datos de secuencias puede ser una base de datos privada almacenada con el sistema de ordenador 100, o una base de datos pública tal como GENBANK que está disponible a través de internet.

El proceso 200 comienza en un estado 201 de inicio y luego se mueve a un estado 202 en donde las nuevas secuencias que van a ser comparadas se almacenan en una memoria en un sistema 100 de ordenador. Tal como se discutió anteriormente, la memoria podría ser cualquier tipo de memoria, incluyendo RAM o un dispositivo de almacenamiento interno.

El proceso 200 se mueve entonces a un estado 204 en donde una base de datos de secuencias es abierta para análisis y comparación. El proceso 200 se mueve entonces a un estado 206 en donde la primera secuencia almacenada en la base de datos es leída en una memoria del ordenador. Una comparación se lleva a cabo entonces en el estado 210 para determinar si la primera secuencia es la misma que la segunda secuencia. Es importante anotar que estas etapas no están limitadas para llevar a cabo una comparación exacta entre la nueva secuencia y la primera secuencia en la base de datos. Se conocen métodos bien conocidos para los experimentados en la técnica para comparar dos secuencias de nucleótidos o proteínas, incluso si no son idénticas. Por ejemplo, pueden introducirse brechas en una secuencia con el fin de elevar el nivel de homología entre las dos secuencias probadas. Los parámetros que controlan si las brechas u otras características son introducidas en una secuencia durante la comparación son introducidas normalmente por el usuario del sistema de ordenador.

Una vez que la comparación de las dos secuencias ha sido llevada a cabo en el estado 210, se hace una determinación en el estado de decisión 210 si las dos secuencias son la misma. Desde luego, el término "misma" no está limitado a secuencias que sean absolutamente idénticas. Secuencias que están dentro de los parámetros de homología introducidos por el usuario serán marcadas como "la misma" en el proceso 200.

Si se determina que las dos secuencias son la misma, el proceso 200 se mueve al estado 214 en donde el nombre de la secuencia para la base de datos es desplegado al usuario. Este estado notifica al usuario de que la secuencia con el nombre desplegado satisface las restricciones de homología que fueron introducidas. Una vez que el nombre

de la secuencia almacenada es desplegado al usuario, el proceso 200 se mueve a un estado de decisión 218 en donde se hace una determinación de si existen más secuencias en las bases de datos. Si no existen más secuencias en la base de datos, entonces el proceso 200 termina en un estado final 220. Sin embargo, si existen más secuencias en la base de datos, entonces el proceso 200 se mueve a un estado 224 en donde se mueve un
 5 señalador a la siguiente secuencia en la base de datos de tal manera que pueda ser comparada con la nueva secuencia. De esta manera, la nueva secuencia es alineada y comparada con cada secuencia en la base de datos.

Debe anotarse que si la determinación ha sido hecha en el estado de decisión 212 de que las secuencias no eran homólogas, entonces el proceso 200 se movería inmediatamente al estado de decisión 218 con el fin de determinar si cualquier otra secuencia está disponible en la base de datos para comparación.

De acuerdo con lo anterior, la invención describe un sistema de ordenador que comprende un procesador, un dispositivo de almacenamiento de datos que tiene almacenados en sí mismo una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o una secuencia de ácidos de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas, teniendo un dispositivo de almacenamiento de
 10 datos secuencias de nucleótidos o secuencias de polipéptidos en el mismo almacenadas de manera recuperable para ser comparadas con una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas de las mismas, o una secuencia de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas y un comparador de secuencias para llevar a cabo la comparación. El comparador de secuencias puede
 15 indicar un nivel de homología entre las secuencias comparadas o identificar motivos estructurales en el código de ácidos nucleicos descrito más arriba de secuencias de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o una secuencia de un polipéptido de la invención y secuencias
 20 sustancialmente idénticas a las mismas, o puede identificar motivos estructurales en secuencias que son comparadas con estos códigos de ácidos nucleicos y códigos de polipéptidos. En algunos aspectos, el dispositivo de almacenamiento de datos puede tener almacenados en sí mismo las secuencias de al menos 2, 5, 10, 15, 20, 25, 30 o 40 o más de las secuencias de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las
 25 mismas, o las secuencias de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas.

La invención describe adicionalmente un método para determinar el nivel de homología entre una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o una secuencia de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas y una secuencia de nucleótidos de
 30 referencia. El método incluye leer el código de ácidos nucleicos o el código de polipéptidos y la secuencia de nucleótidos o polipéptidos de referencia a través del uso de un programa de ordenador que determina los niveles de homología y determina la homología entre el código de ácidos nucleicos o el código de polipéptidos y la secuencia de nucleótidos o polipéptidos de referencia con el programa de ordenador. El programa de ordenador puede ser
 35 cualquiera de un cierto número de programas de ordenador para determinar niveles de homología, incluyendo los enumerados específicamente aquí, (por ejemplo, BLAST2N con los parámetros de fondo o con cualquier parámetro modificado), el método puede ser implementado utilizando sistemas de ordenador como se describieron anteriormente. El método también puede ser ejecutado leyendo al menos 2, 5, 10, 15, 20, 25, 30 o 40 o más de las
 40 secuencias de ácidos nucleicos descritos de la invención, o la secuencias de polipéptidos de la invención a través del uso del programa de ordenador y determinando la homología entre los códigos de ácidos nucleicos o los códigos de polipéptidos y las secuencias de nucleótidos o secuencias de polipéptidos de referencia.

La Figura 3 es un diagrama de flujo que ilustra un aspecto de un proceso 250 en un ordenador para determinar si dos secuencias son homólogas. El proceso 250 comienza en un estado de inicio 252 y luego se mueve a un estado 254 en donde una primera secuencia que va a ser comparada es almacenada en una memoria. La segunda
 45 secuencia que va a ser comparada es almacenada entonces en una memoria en un estado 256. El proceso 250 sumado entonces al estado 260 en donde el primer carácter en la primera secuencia es leído y luego a un estado 262 en donde el primer carácter de la segunda secuencia es leído, debe entenderse que si la secuencia es una secuencia de nucleótidos, entonces el carácter nuevamente será A, T, C, G o U. Si la secuencia es una secuencia de proteínas, entonces es preferible el código de aminoácidos de letra individual de tal manera que la primera y segunda secuencias puedan ser comparadas fácilmente.

Se hace entonces una determinación en el estado de decisión 264 en donde los dos caracteres son los mismos. Si son los mismos entonces el proceso 250 se mueve a un estado 268 donde los siguientes caracteres en la primera y segundas secuencias son leídos. Se hace entonces una determinación de si los siguientes caracteres son los
 50 mismos. Si lo son, entonces el proceso 250 continúa este bucle hasta que dos caracteres no sean el mismo. Si se hace una determinación de que los siguientes dos caracteres no son el mismo, el proceso 250 se mueve a un estado de decisión 274 para determinar si hay más caracteres en cualquier secuencia para leer.

Si no hay más caracteres para leer, entonces el proceso 250 se mueve a un estado 276 en donde el nivel de homología entre la primera y segunda secuencias es desplegado al usuario. El nivel de homología es determinando calculando la proporción de caracteres entre las secuencias que fueron la misma del número total de secuencias en la primera secuencia. Así, si cada carácter en una primera secuencia de 100 nucleótidos alineada con cada carácter
 55 en una segunda secuencia, el nivel de homología sería del 100%.

Alternativamente, el programa de ordenador puede ser un programa de ordenador que compara las secuencias de nucleótidos de una secuencia de ácidos nucleicos como se establece en la invención, con una o más secuencias de nucleótidos de referencia con el fin de determinar si el código de ácido nucleico de una secuencia de ácidos nucleicos de la invención y la secuencias sustancialmente idénticas a las mismas, difieren de una secuencia de ácidos nucleicos de referencia en una o más posiciones. En un aspecto tal programa graba la longitud de identidad de nucleótidos insertados, eliminados o sustituidos con respecto a la secuencia del polinucleótido de referencia o una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas. En un aspecto, el programa de ordenador puede ser un programa que determina si una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, contienen un polimorfismo de nucleótidos individual (SNP) con respecto a una secuencia de nucleótidos.

Se describe adicionalmente un método para determinar si una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a la misma, difieren en uno o más nucleótidos con respecto a una secuencia de nucleótidos de referencia que comprende las etapas de leer el código de ácidos nucleicos y la secuencia de nucleótidos de referencia a través del uso de un programa de ordenador que identifica diferencias entre las secuencias de ácidos nucleicos e identifica diferencias entre el código de ácidos nucleicos y las secuencias de nucleótidos de referencia con el programa de ordenador. En algunos aspectos, el programa de ordenar es un programa que identifica polimorfismos de nucleótidos individuales. El método puede ser implementado por el sistema de ordenador descrito más arriba y el método ilustrado en la Figura 3. El método también puede ser llevado a cabo leyendo el menos 2, 5, 10, 15, 20, 25, 30, o 40 o más de las secuencias de los ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas e identificar diferencias entre los códigos de ácidos nucleicos y las secuencias de nucleótidos de referencia con el programa de ordenador.

En otros aspectos el sistema basado en ordenador puede comprender adicionalmente un identificador para identificar características dentro de una secuencia de ácidos nucleicos de la invención a una secuencia de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas.

Un "identificador" se refiere a uno o más programas que identifican ciertas características dentro de una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o una secuencia de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas. En un aspecto, el identificador puede comprender un programa que identifica un marco de lectura abierto en una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas.

La Figura 4 es un diagrama de flujo que ilustra un aspecto del proceso de identificación 300 para detectar la presencia de una característica en una secuencia. El proceso 300 comienza en un estado de inicio 302 y luego se mueve a un estado 304 en donde una primera secuencia que va a ser verificada en cuanto a características es almacenada en una memoria 115 en el sistema 100 de ordenador. El proceso 300 se mueve entonces a un estado 306 en donde una base de datos de características de secuencias es abierta. Tal base de datos incluiría una lista de cada una de los atributos junto con el nombre de la característica. Por ejemplo, un nombre de característica podría ser "codón de iniciación" y el atributo podría ser "ATG". Otro ejemplo sería el nombre de característica "TAATAA Box" y el atributo de la característica sería "TAATAA". Un ejemplo de tales bases de datos es producido por la University of Wisconsin Genetics Computer Group. Alternativamente, las características pueden ser motivos de polipéptidos estructurales tales como hélices alfa, laminas beta o motivos de polipéptidos funcionales tales como sitios de actividad enzimática, motivos de hélice-giro-hélice u otros motivos conocidos para los experimentados en la técnica.

Una vez que la base de datos de características es abierta en el estado 306, el proceso 300 se mueve a un estado 308 en donde la primera característica es leída a partir de la base de datos. Una comparación del atributo de la primera característica con la primera secuencia se hace entonces en un estado 310. Se hace entonces la determinación en un estado de decisión 316 en donde el atributo de la característica fue encontrado en la primera secuencia. Si el atributo fue encontrado, entonces el proceso 300 se mueve a un estado 318 en donde el nombre de la característica encontrada es desplegado al usuario.

El proceso 300 se mueve entonces a un estado 320 de decisiones en donde se hace una determinación de si existen más características en la base de datos. Si no existen más características, entonces el proceso 300 termina en un estado de terminación 324. Sin embargo, si existen más características en la base de datos, entonces el proceso 300 lee la siguiente característica de secuencia en un estado 326 y retorna al estado 316 en donde el atributo de la siguiente característica es comparado contra la primera secuencia. Debe anotarse, que si el atributo de característica no es encontrado en la primera secuencia en el estado de decisión 316, el proceso 300 se mueve directamente al estado de decisión 320 para determinar si existen más características en la base de datos.

De acuerdo con lo anterior, se describe adicionalmente un método para identificar una característica dentro de una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o una secuencia de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas, que comprende al leer los códigos de los ácidos nucleicos o los códigos de los polipéptidos a través del uso de un programa de

ordenador que identifica características en los mismos e identifica características dentro de los códigos de ácidos nucleicos con el programa de ordenador. En un aspecto, el programa de ordenador comprende un programa de ordenador que identifica marcos de lectura abiertos. El método puede ser llevado a cabo leyendo una secuencia individual o al menos 2, 5, 10, 15, 20, 25, 30, o 40 de las secuencias de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o las secuencias de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas, a través del uso del programa de ordenador a identificar características dentro de los códigos de ácidos nucleicos o códigos de polipéptidos con el programa de ordenador.

Una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o una secuencia de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas, pueden ser almacenadas y manipuladas en una variedad de programas de procesamiento de datos en una variedad de formatos. Por ejemplo, una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o una secuencia de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas, pueden ser almacenadas como texto en un archivo de procesamiento de palabra, tal como Microsoft WORD™ o WORDPERFECT™ o un archivo ASCII en una variedad de programas familiares para los experimentados en la técnica, tales como DB2™, SYBASE™, o ORACLE™. Además, muchos programas de ordenador y bases de datos pueden ser utilizados como algoritmos de comparación de secuencias, identificadores o fuentes de secuencias de nucleótidos o secuencias de polipéptidos de referencia para ser comparadas con una secuencia de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o una secuencia de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas. La siguiente lista no pretende limitar la invención sino proveer una guía para los programas y bases de datos que son útiles con las secuencias de ácidos nucleicos de la invención y secuencias sustancialmente idénticas a las mismas, o las secuencias de polipéptidos de la invención y secuencias sustancialmente idénticas a las mismas.

Los programas y bases de datos que pueden ser utilizados incluyen, pero no se limitan a: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul et al, J. Mol. Biol. 215: 403, 1990), FASTA (Pearson and Lipman, Proc. Natl. Acad. Sci. USA, 85: 2444, 1988), FASTDB (Brutlag et al. Comp. App. Biosci. 6:237-245, 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius2.DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMM (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), la base de datos MDL Available Chemicals Directory, la base de datos MDL Drug Data Report, la base de datos Comprehensive Medicinal Chemistry, la base de datos Derwent's World Drug Index, la base de datos BioByteMasterFile, la base de datos Genbank y la base de datos Genseq. Muchos otros programas y bases de datos serían evidentes para una persona de experiencia en la técnica dada la presente divulgación.

Los motivos que pueden ser detectados utilizando los programas anteriores incluyen secuencias que codifican cremalleras de leucina, motivos hélice-giro-hélice, sitios de glicosilación, sitios de ubiquitinación, hélices alfa y láminas beta, secuencias de señalización que codifican péptidos de señalización que dirigen la secreción de las proteínas codificadas, secuencias implicadas en la regulación de la transcripción tales como homeocajas, tensionamientos ácidos, sitios activos enzimáticos, sitios de enlazamiento de sustrato y sitios de escisión enzimática.

Hibridación de ácidos nucleicos

La invención describe ácidos nucleicos aislados, sintéticos o recombinantes que pueden hibridar bajo condiciones de restricción a una secuencia de ejemplo de la invención. Las condiciones de restricción pueden ser condiciones altamente restrictivas, condiciones de restricción media y bajo condiciones de restricción baja, incluyendo las condiciones de restricción alta y reducidas descritas aquí. En un aspecto, es la restricción de las condiciones de lavado la que define las condiciones que determinan si un ácido nucleico está dentro del alcance de la invención, como se discutió más adelante.

En aspectos alternativos los ácidos nucleicos de la invención tal como se definen por su capacidad de hibridar bajo condiciones de restricción pueden ser entre aproximadamente cinco residuos y la longitud completa del ácido nucleico de la invención; por ejemplo, pueden ser al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 55, 60, 65, 70, 75, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, o más, residuos de longitud. Los ácidos nucleicos más cortos de la longitud completa también están incluidos. Estos ácidos nucleicos pueden ser útiles, por ejemplo, como sondas de hibridación, sondas de marcación, sondas de oligonucleótidos para PCR, ARNi "cadena sencilla o doble", secuencias antisentido que codifican péptidos de enlazamiento de anticuerpos (epítomos), motivos, sitios activos y similares.

En un aspecto, los ácidos nucleicos tal como se describen están definidos por su capacidad para hibridar bajo alta restricción que comprende condiciones de aproximadamente 50% de formamida a aproximadamente 37°C hasta 42°C. En un aspecto los ácidos nucleicos de la invención están definidos por su capacidad para hibridar bajo condiciones que comprenden restricción reducida en aproximadamente 35% a 25% de formamida a aproximadamente 30°C hasta 35°C.

Alternativamente, los ácidos nucleicos tal como se describen son definidos por su capacidad de hibridar bajo alta restricción que comprende condiciones a 42°C en 50% de formamida, 5X SSPE, 0.3% de SDS, y un ácido nucleico de bloqueo de secuencia repetitiva, tal como ADN de COT-1 o esperma de salmón (por ejemplo, 200 ug/ml de ADN de esperma de salmón separado y desnaturalizado). En un aspecto, los ácidos nucleicos de la invención están definidos por su capacidad para hibridar bajo condiciones de restricción reducida que comprende 35% de formamida a una temperatura reducida de 35°C.

En las reacciones de hibridación de ácidos nucleicos, las condiciones usadas para alcanzar un nivel particular de restricción varían, dependiendo de la naturaleza de los ácidos nucleicos que están siendo hibridados, por ejemplo, la longitud, grado de complementariedad, composición de secuencias de nucleótidos (por ejemplo, contenido de GC versus AT) y tipo de ácido nucleico (por ejemplo, ARN versus ADN) de las regiones de hibridación de los ácidos nucleicos pueden considerarse en la selección de las condiciones de hibridación. Una consideración adicional es si uno de los ácidos nucleicos es inmovilizado, por ejemplo, sobre un filtro.

La hibridación puede llevarse a cabo bajo condiciones de baja restricción, restricción moderada o alta restricción. Como ejemplo de hibridación de ácidos nucleicos, una membrana polimérica que contiene ácidos nucleicos desnaturalizados inmovilizados es prehibridada primero durante 30 minutos a 45°C en una solución que consiste de NaCl 0.9 M, NaH₂PO₄ 50 mM, pH 7.0, Na₂EDTA 5.0 mM, SDS al 0.5%, 10X Denhardt's y 0.5 mg/ml de ácido polirribonucleico. Aproximadamente 2×10^7 cpm (actividad específica $4-9 \times 10^8$ cpm/ug) de una sonda de oligonucleótido marcada en extremo con ³²P se agregan entonces a la solución. Después de 12-16 horas de incubación, la membrana es lavada durante 30 minutos a temperatura ambiente en 1 X SET (NaCl 150 mM, Trisclorhidrato 20 mM, pH 7.8, Na₂EDTA 1 mM) que contiene 0.5% de SDS, seguido por un lavado durante 30 minutos en 1 X SET fresco a Tm-10°C para la sonda de oligonucleótido. La membrana es expuesta entonces a la película autorradiográfica para la detección de señales de hibridación.

Todas las hibridaciones anteriores serán consideradas como condiciones de alta restricción.

Después de la hibridación, un filtro puede ser lavado para retirar cualquier sonda detectable no específicamente enlazada. La restricción utilizada para lavar el filtro puede variarse dependiendo de la naturaleza de los ácidos nucleicos que están siendo hibridados, la longitud de los ácidos nucleicos que están siendo hibridados, el grado de complementariedad, la composición de secuencias de nucleótidos (por ejemplo, contenido de GC versus AT), y el tipo de ácido nucleico (ARN versus ADN). Ejemplos de lavados en condiciones de restricción progresivamente más altas son como sigue: 2X SSC, SDS al 0.1 % a temperatura ambiente durante 15 minutos (baja restricción); 0.1 X SSC, SDS al 0.5% a temperatura ambiente durante 30 minutos hasta 1 hora (restricción moderada); 0.1 X SSC, SDS al 0.5 % durante 15 a 30 minutos a entre la temperatura de hibridación y 68°C (alta restricción); y NaCl 0.15 M durante 15 minutos a 72°C (restricción muy alta). Un lavado final de restricción baja puede ser llevado a cabo en 0.1 X SSC a temperatura ambiente. Los ejemplos anteriores son solamente ilustrativos de un conjunto de condiciones que pueden ser utilizadas para lavar filtros. Una persona experimentada en la técnica sabrá que hay numerosas recetas para diferentes lavados de restricción. Abajo se dan algunos otros ejemplos.

Los ácidos nucleicos que han sido hibridados a la sonda son identificados por autorradiografía u otras técnicas convencionales.

El procedimiento anterior puede ser modificado para identificar ácidos nucleicos que tienen niveles decrecientes de homología con respecto a la secuencia de sonda, por ejemplo, para obtener ácidos nucleicos de homología decreciente con respecto a la sonda detectable, pueden utilizarse condiciones menos restrictivas. Por ejemplo, la temperatura de hibridación puede ser disminuida en incrementos de 5°C desde 68°C hasta 42°C en un regulador de hibridación que tiene una concentración de Na⁺ de aproximadamente 1M. Después de la hibridación, el filtro puede ser lavado con 2X SSC, SDS al 0.5% a la temperatura de hibridación. Estas condiciones se consideran como condiciones "moderadas" por encima de 50°C y condiciones "bajas" por debajo de 50°C. Un ejemplo específico de condiciones de hibridación "moderadas" es cuando la hibridación anterior es llevada a cabo a 55°C. Un ejemplo específico de condiciones de hibridación de "baja restricción" es cuando la hibridación anterior es llevada a cabo a 45°C.

Alternativamente, la hibridación puede ser llevada a cabo en reguladores, tales como, 6X SSC, que contienen formamida a una temperatura de 42°C. En este caso, la concentración de formamida en el regulador de hibridación puede ser reducida en incrementos de 5% desde 50% a 0% para identificar clones que tienen niveles decrecientes de homología con respecto a la sonda. Después de la hibridación, el filtro puede ser lavado con 6X SSC, SDS al 0.5% a 50°C. Estas condiciones se consideran como condiciones "moderadas" por encima de 25% de formamida y

condiciones “bajas” por debajo de 25% de formamida. Un ejemplo específico de condiciones de hibridación “moderadas” es cuando la hibridación anterior se lleva a cabo a 30% de formamida. Un ejemplo específico de condiciones de hibridación en “baja restricción” es cuando la hibridación anterior es llevada a cabo a 10% de formamida.

- 5 Sin embargo, la selección de un formato de hibridación no es crítica – es la restricción de las condiciones de lavado la que establece las condiciones que determinan si un ácido nucleico está dentro del alcance de la invención. Las condiciones de lavado utilizadas para identificar ácidos nucleicos dentro del alcance de la invención incluyen, por ejemplo, una concentración de sal de aproximadamente 0.02 molar a pH 7 y a una temperatura de al menos aproximadamente 50°C o aproximadamente 55°C hasta aproximadamente 60°C; o una concentración de sal de aproximadamente 0.15M de NaCl a 72°C durante aproximadamente 15 minutos; o, una concentración de sal de aproximadamente 0.2X SSC a una temperatura de al menos aproximadamente 50°C o aproximadamente 55°C hasta aproximadamente 60°C durante aproximadamente 15 hasta aproximadamente 20 minutos; o, el complejo de hibridación es lavado dos veces con una solución con una concentración de sal de aproximadamente 2X SSC que contiene 0.1% de SDS a temperatura ambiente durante 15 minutos y luego lavado dos veces con 0.1X SSC que contiene 0.1% de SDS a 68°C durante 15 minutos; o, condiciones equivalentes. Véase Sambrook, Tijssen y Ausubel para la descripción del regulador SSC y condiciones equivalentes.

- 20 Estos métodos pueden ser utilizados para aislar ácidos nucleicos de la invención. Por ejemplo, los métodos presentes pueden ser utilizados para aislar ácidos nucleicos que tienen una secuencia con al menos 97%, al menos 95%, al menos 90%, al menos 85%, al menos 80%, al menos 75%, al menos 70%, al menos 65%, al menos 60%, al menos 55%, o al menos 50% de homología con respecto a una secuencia de ácidos nucleicos seleccionada del grupo consistente de una de las secuencias de la invención y secuencias sustancialmente idénticas a las mismas, o fragmentos que contienen al menos aproximadamente 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, o 500 bases consecutivas de las mismas y secuencias complementarias a las mismas. La homología puede ser medida utilizando el algoritmo de alineamiento. Por ejemplo, los polinucleótidos homólogos pueden tener una secuencia de codificación que es una variante alélica de origen natural de una de las secuencias de codificación descritas aquí. Tales variantes alélicas pueden tener una sustitución, eliminación o adición de uno o más nucleótidos en comparación con los ácidos nucleicos de la invención o las secuencias complementarias a los mismos.

- 30 Adicionalmente, los procedimientos anterior pueden ser utilizados para aislar ácidos nucleicos que codifican polipéptidos que tiene al menos aproximadamente 99%, 95%, al menos 90%, al menos 85%, al menos 80%, al menos 75%, al menos 70%, al menos 65%, al menos 60%, al menos 55%, o al menos 50% de homología de un polipéptido que tiene la secuencia de una de las secuencias de aminoácidos de la invención, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos de las mismas según se determina utilizando un algoritmo de alineamiento de secuencias (por ejemplo, tal como el algoritmo FASTA versión 3.0t78 con los parámetros por sistema).

- 35 Sondas de oligonucleótidos y métodos para usarlas

- La invención describe sondas de ácidos nucleicos que pueden ser utilizadas, por ejemplo, para identificar ácidos nucleicos que codifican un polipéptido con una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa o fragmentos de los mismos o para identificar genes de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. En un aspecto, la sonda comprende al menos 10 bases consecutivas de un ácido nucleico de la invención. Alternativamente, una sonda descrita puede tener al menos aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 150 o aproximadamente de 10 a 50, aproximadamente de 20 a aproximadamente 60, o aproximadamente 30 a 70, bases consecutivas de una secuencia tal como se establece en un ácido nucleico de la invención. Las sondas identifican un ácido nucleico enlazándose por enlazamiento y/o hibridación. Las sondas pueden ser utilizadas en arreglos de la invención, véase la discusión de más adelante, incluyendo, por ejemplo, arreglos capilares. Las sondas descritas también pueden ser utilizadas para aislar otros ácidos nucleicos o polipéptidos.

- 50 Los ácidos nucleicos aislados, sintéticos o recombinantes tal como se describen y las secuencias sustancialmente idénticas a los mismos, las secuencias complementarias a los mismos, o un fragmento que comprende al menos 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, o 500 bases consecutivas de una de las secuencias de la invención y las secuencias sustancialmente idénticas a las mismas, o las secuencias complementarias a las mismas también pueden ser utilizadas como sondas para determinar si una muestra biológica, tal como una muestra de suelo, contiene un organismo que tiene una secuencia de ácidos nucleicos de la invención o un organismo del cual fue obtenido el ácido nucleico. En tales procedimientos, se obtiene una muestra biológica que potencialmente aloja el organismo del cual fue aislado el ácido nucleico y los ácidos nucleicos son obtenidos de la muestra. Los ácidos nucleicos son puestos en contacto con la sonda bajo condiciones que permiten que la sonda hibride específicamente cualquier secuencia complementaria de las cuales se presentan aquí.

Cuando es necesario, pueden determinarse las condiciones que permiten que la sonda hibride específicamente a secuencias complementarias colocando la sonda en contacto con secuencias complementarias de muestras conocidas por contener la secuencia complementaria así como secuencias de control que no contienen la secuencia complementaria. Las condiciones de hibridación, tales como la concentración de sal del regulador de hibridación, la concentración en formamida del regulador de hibridación, o la temperatura de hibridación, pueden variarse para identificar condiciones que permiten que la sonda hibride específicamente a ácidos nucleicos complementarios.

Si la muestra contiene el organismo desde el cual fue aislado el ácido nucleico, la hibridación específica de la sonda se detecta entonces. La hibridación puede ser detectada por marcación de la sonda con un agente detectable tal como un isótopo radiactivo, un colorante fluorescente o una enzima capaz de catalizar la formación de un producto detectable.

Muchos métodos para utilizar las sondas marcadas para detectar la presencia de ácidos nucleicos complementarios en una muestra son familiares para los experimentados en la técnica. Incluyen inmunoprecipitación Southern, inmunoprecipitación Northern, procedimientos de hibridación en colonias y siembras de puntos. Los protocolos para cada uno de estos procedimientos están provistos en Ausubel et al. *Current Protocols in Molecular Biology*, John Wiley 503 Sons, Inc. 1997) and Sambrook et al., *Molecular Cloning: A Laboratory Manual* 2nd Ed., Cold Spring Harbor Laboratory Press (1989).

Alternativamente, más de una sonda (al menos una de las cuales es capaz de hibridar específicamente a cualquier secuencia complementaria que está presente en la muestra de ácido nucleico), puede ser utilizada en una reacción de amplificación para determinar si la muestra contiene un organismo que contiene una secuencia de ácidos nucleicos de la invención (por ejemplo, un organismo del cual fue aislado el ácido nucleico). Típicamente, las sondas comprenden oligonucleótidos. En un aspecto, la reacción de amplificación puede comprender una reacción de PCR. Los protocolos de PCR están descritos en Ausubel y Sambrook, *supra*. Alternativamente, la amplificación puede comprender una reacción en cadena de ligasa, 3SR, o reacción de desplazamiento de cadena. (Véase Barany, F., "The Ligase Chain Reaction in a PCR World", *PCR Methods and Applications* 1:5-16, 1991; E. Fahy et al., "Self-sustained Sequence Replication (3SR): An Isothermal Transcription-based Amplification System Alternative to PCR", *PCR Methods and Applications* 1:25-33, 1991; y Walker G.T. et al., "Strand Displacement Amplification-an Isothermal *in vitro* DNA Amplification Technique", *Nucleic Acid Research* 20:1691-1696, 1992). En tales procedimientos, los ácidos nucleicos en la muestra son puestos en contacto con las sondas, la reacción de amplificación se lleva a cabo y cualquier producto de amplificación resultante es detectado. El producto de amplificación puede ser detectado llevando a cabo electroforesis en gel sobre los productos de reacción y tinción del gel con un intercalador tal como bromuro de etidio. Alternativamente, pueden marcarse una o más de las sondas con un isótopo radiactivo y puede detectarse la presencia de un producto de amplificación radiactivo por autorradiografía después de la electroforesis en gel.

Las sondas derivadas a partir de secuencias cerca de los extremos de las secuencias de la invención y secuencias sustancialmente idénticas a las mismas, pueden ser utilizadas en procedimientos de desplazamiento de cromosomas para identificar clones que contienen secuencias genómicas localizadas adyacentes a las secuencias de la invención y secuencias sustancialmente idénticas a las mismas. Tales métodos permiten el aislamiento de genes que codifican proteínas adicionales a partir del organismo anfitrión.

Los ácidos nucleicos aislados, sintéticos o recombinantes de la invención y secuencias sustancialmente idénticas de los mismos, las secuencias complementarias a los mismos, o un fragmento que contiene al menos 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, o 500 bases consecutivas de una de las secuencias de la invención y secuencias sustancialmente idénticas a las mismas, o las secuencias complementarias a las mismas, pueden ser utilizadas como sondas para identificar y aislar ácidos nucleicos relacionados. En algunos aspectos, los ácidos nucleicos relacionados pueden ser ADNc o ADN genómico de organismos diferentes a aquel del cual fue aislado el ácido nucleico. Por ejemplo, los otros organismos pueden ser organismos relacionados. En tales procedimientos, una muestra de ácido nucleico es puesta en contacto con la sonda bajo condiciones que permiten que la sonda hibride específicamente a secuencias relacionadas. La hibridación de la sonda a ácidos nucleicos a partir de los organismos relacionados se detecta entonces utilizando cualquiera de los métodos descritos más arriba.

Variando la restricción de las condiciones de hibridación utilizadas para identificar los ácidos nucleicos, tales como ADNc o ADN genómico, que hibridizan a la sonda detectable, los ácidos nucleicos que tienen diferentes niveles de homología a la sonda pueden ser identificados y aislados. La restricción puede variarse llevando a cabo la hibridación a temperaturas variables por debajo de las temperaturas de fusión de las sondas. La temperatura de fusión, T_m , es la temperatura (bajo fuerza iónica y pH definido) en la cual el 50% de la secuencia objetivo hibrida a una sonda perfectamente complementaria. Condiciones muy restrictivas se seleccionan para igualar a o aproximadamente 5°C por debajo de la T_m para una sonda particular. La temperatura de fusión de la sonda puede ser calculada utilizando las siguientes fórmulas:

Para sondas entre 14 y 70 nucleótidos de longitud la temperatura de fusión (T_m) se calcula utilizando la fórmula:

$T_m = 81,5 + 16,6(\log [Na^+]) + 0,41(\text{fracción } G + C) - (600 / N)$, donde N es la longitud de la sonda.

Si la hibridación se lleva a cabo en una solución que contiene formamida, la temperatura de fusión puede ser calculada utilizando la ecuación: $T_m = 81,5 + 16,6(\log [Na^+]) + 0,41(\text{fracción } G + C) - (0,63\% \text{ formamida}) - (600/N)$ en donde N es la longitud de la sonda.

- 5 La prehibridación puede ser llevada a cabo en 6X SSC, 5X de reactivo de Denhardt, SDS al 0.5%, ADN de esperma de salmón fragmentado desnaturalizado 100 µg/ml o 6X SSC, 5X, de reactivo de Denhardt, SDS al 0.5%, ADN de esperma de salmón fragmentado desnaturalizado de 100 µg/ml, formamida al 50%. De las fórmulas para las soluciones SSC y de Denhardt están listadas en Sambrook et al, *supra*.

- 10 La hibridación se lleva a cabo agregando la sonda detectable a las soluciones de prehibridación listadas más arriba. Cuando la sonda comprende ADN de doble cadena, es desnaturalizada antes de la adición de la solución de hibridación. El filtro es puesto en contacto con la solución de hibridación durante un período de tiempo suficiente para permitir que la sonda hibride a ADNc o ADN genómico que contienen secuencias complementarias a las mismas u homólogas a la mismas. Para sondas por encima de 200 nucleótidos de longitud, la hibridación puede ser llevada a cabo a 15-25°C por debajo de T_m . Para sondas más cortas, tales como sondas de oligonucleótidos, la hibridación puede ser llevada a cabo a 5-10°C por debajo de la T_m . Típicamente, para hibridaciones en 6x SSC, la hibridación se lleva a cabo a aproximadamente a 68°C. Usualmente, para hibridaciones en soluciones que contienen formamida al 50%, la hibridación se lleva a cabo a aproximadamente 42°C.

Inhibición de la expresión de transaminasas y/o oxidorreductasas.

- 20 La invención provee de ácidos nucleicos complementarios a (por ejemplo secuencias antisentido a) los ácidos nucleicos de la invención, por ejemplo, ácidos nucleicos que codifican d-aminoácido transferasa. En realizaciones las secuencias antisentido de la invención son capaces de inhibir el transporte, acoplamiento o transcripción de los genes que codifican transferasa, por ejemplo, transaminasa-, por ejemplo, d-aminoácido transferasa-, y/o oxidorreductasa-, por ejemplo, deshidrogenasa-, por ejemplo, d-aminoácido deshidrogenasa. La inhibición puede ser efectuada a través del direccionamiento del ADN genómico o ARN mensajero. La transcripción o función del ácido nucleico direccionado puede ser inhibida, por ejemplo, por hibridación y/o escisión. En realizaciones alternativas los inhibidores descritos por la presente invención incluyen oligonucleótidos que son capaces bien sea de enlazar un gen o mensajero de transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, en cualquier caso previniendo o inhibiendo la producción o funcionamiento de una transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. La asociación puede ser a través de hibridación específica de secuencias. Otra clase útil de inhibidores incluye oligonucleótidos que producen la inactivación o escisión de un mensajero de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. El oligonucleótido puede tener actividad enzimática que hace tal escisión, tal como ribozimas. El oligonucleótido puede ser modificado químicamente o conjugado con una enzima o composición capaz de escindir el ácido nucleico complementario. Un conjunto de muchos diferentes de tales oligonucleótidos puede ser seleccionado para aquellos con la actividad deseada. Así, la invención describe diversas composiciones para la inhibición de la expresión de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa en un nivel de ácido nucleico y/o proteína, por ejemplo, ARNi antisentido y ribozimas que comprenden secuencias de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención y los anticuerpos anti-transferasa, por ejemplo, antitransaminasa, por ejemplo, anti-d-aminoácido transferasa, y/o anti-oxidorreductasa, por ejemplo, anti-deshidrogenasa, por ejemplo, anti-d-aminoácido deshidrogenasa de la invención.
- 45 Las composiciones tal como están descritas para la inhibición de la expresión de transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa (por ejemplo antisentido, ARNi, micro ARN, ribozimas, anticuerpos) pueden ser utilizadas como composiciones farmacéuticas (fármacos).

Oligonucleótidos antisentido

- 50 La presente invención describe oligonucleótidos antisentido capaces de enlazarse a un mensajero de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa los cuales pueden inhibir, por ejemplo, la actividad de transferasa y/o deshidrogenasa direccionando ARNm. Las estrategias para diseñar oligonucleótidos antisentido están bien descritas en la literatura científica y de patentes, y la persona experimentada puede diseñar tales oligonucleótidos de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa utilizando los reactivos novedosos de la invención. Por ejemplo, los protocolos de desplazamiento de genes/mapeo de ARN para

seleccionar oligonucleótidos antisentido efectivos son bien conocidos en la técnica, véase, por ejemplo, Ho (2000) *Methods Enzymol.* 314: 168-183, que describe un ensayo de mapeo de ARN, el cual está basado en técnicas moleculares estándar para proveer un método fácil y confiable para la selección de secuencias antisentido potentes. Véase también Smith (2000) *Eur. J. Pharm.Sci.* 11: 191-198.

- 5 Los ácidos nucleicos de origen natural son usados como oligonucleótidos antisentido. Los oligonucleótidos antisentido pueden ser de cualquier longitud. Por ejemplo, en aspectos alternativos, los oligonucleótidos antisentido están entre aproximadamente 5 hasta 100, aproximadamente 10 a 80, aproximadamente 15 a 60, aproximadamente 18 a 40. La longitud óptima puede ser determinada por selecciones de rutina. Los oligonucleótidos antisentido pueden estar presentes en cualquier concentración. La concentración óptima puede ser determinada por selección de rutina. Se conoce una amplia variedad de nucleótidos y análogos de ácidos nucleicos de origen sintético, no naturales que pueden abordar este problema potencial. Por ejemplo, pueden usarse ácidos nucleicos peptídicos (PNA) que contienen esqueletos no iónicos, tales como unidades N-(2-aminoetil) glicina. Los oligonucleótidos antisentido que tienen enlaces fosforotioato también pueden ser utilizados, como se describe en WO 97/03211; WO 96/39154; Mata (1997) *Toxicol Appl Pharmacol* 144:189-197; *Antisense Therapeutics*, ed. Agrawal (Humana Press, Totowa, N.J., 1996). Los oligonucleótidos antisentido que tienen análogos de esqueletos de ADN sintéticos provistos por la invención también puede incluir ácidos nucleicos fosforoditioato, metilfosfonato, fosforamidato, alquil fosfortriéster, sulfamato, 3'-tioacetal, metileno (metilimino), 3'-N-carbamato y morfolino carbamato, como se describió más arriba.

- 20 La metodología de química combinatorial puede ser utilizada para crear vastos números de oligonucleótidos, que pueden ser seleccionados rápidamente para oligonucleótidos específicos que tienen afinidades y especificidades de enlazamiento apropiadas hacia cualquier objetivo, tales como las secuencias de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa en sentido y antisentido de la invención (véase, por ejemplo, Gold (1995) *J. of Biol. Chem.* 270:13581-13584).

25 Ribozimas inhibidoras

- La invención describe ribozimas capaces de enlazar un mensajero de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. Estas ribozimas pueden inhibir la actividad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, por ejemplo, apuntando a ARNm. Las estrategias para designar ribozimas y seleccionar la secuencia antisentido específica para transferasa, por ejemplo, transaminasa-, por ejemplo, d-aminoácido transferasa-, y/o oxidorreductasa-, por ejemplo, deshidrogenasa-, por ejemplo, d-aminoácido deshidrogenasa para direccionar están bien descritas en la literatura científica y de patentes, y la persona experimentada puede diseñar tales ribozimas utilizando los reactivos novedosos (por ejemplo, ácidos nucleicos) de la invención. Las ribozimas actúan enlazando un ARN objetivo a través de la porción de enlazamiento del ARN de una ribozima la cual se mantiene en cercana proximidad a una porción enzimática del ARN que escinde el ARN objetivo. Así, la ribozima reconoce y se enlaza a un ARN objetivo a través de apareamiento de bases complementarias, y una vez enlazado al sitio correcto, actúa enzimáticamente para escindir e inactivar el ARN objetivo. La escisión de un ARN objetivo de tal manera destruirá su capacidad para dirigir la síntesis de una proteína codificada si la escisión ocurre en la secuencia de codificación. Después de que una ribozima se ha enlazado y ha escindido su ARN objetivo, puede ser liberada de ese ARN para enlazar y escindir nuevos objetivos repetidamente.

- En algunas circunstancias, la naturaleza enzimática de una ribozima puede ser ventajosa sobre otras tecnologías, tales como la tecnología antisentido (en donde una molécula de ácido nucleico simplemente se enlaza a un ácido nucleico objetivo para bloquear su transcripción, traducción o asociación con otra molécula) puesto que la concentración efectiva de ribozima necesaria para efectuar un tratamiento terapéutico puede ser inferior que la de un oligonucleótido antisentido. Esta ventaja potencial refleja la capacidad de la ribozima para actuar enzimáticamente. Así, una molécula de ribozimas individuales es capaz de escindir muchas moléculas de ARN objetivo. Además, una ribozima es típicamente un inhibidor altamente específico, cuya especificidad de inhibición depende no solamente del mecanismo de enlazamiento de pares de bases, sino también del mecanismo mediante el cual la molécula inhibe la expresión del ARN al cual se enlaza. Esto es, la inhibición es producida por la escisión del ARN objetivo y así la especificidad está definida como la relación de la tasa de escisión del ARN objetivo sobre la tasa de escisión del ARN no objetivo. Este mecanismo de escisión depende de factores adicionales a los involucrados en el apareamiento de bases. Así, la especificidad de acción de una ribozima puede ser mayor que la de un oligonucleótido antisentido que se enlaza al mismo sitio de ARN.

- 55 La ribozima tal como se describe, por ejemplo, una molécula de ARN ribozima enzimática, puede ser formada en un motivo de cabeza de martillo, un motivo de horquilla, un motivo de virus delta de la hepatitis, un motivo de intrón grupo I, y/o un ARN similar a RNasaP en asociación con una secuencia de guía de ARN. Ejemplos de motivos de cabeza de martillo son descritos, por ejemplo, por Rossi (1992) *Aids Research and Human Retroviruses* 8:183; motivos en horquilla por Hampel (1989) *Biochemistry* 28:4929, y Hampel (1990) *Nuc. Acids Res.* 18:299; el motivo

del virus de delta de hepatitis por Perrotta (1992) Biochemistry 31:16; el motivo de RNasaP de Guerrier-Takada (1983) Cell 35:849; y el intrón de grupo I por Cech Patente de los Estados Unidos No. 4,987,071. La citación de estos motivos específicos no pretende ser limitante. Los experimentados en la técnica reconocerán que una ribozima de la invención, por ejemplo, una molécula de ARN enzimática de esta invención, puede tener un sitio de enlazamiento a sustrato específico complementario a una o más de las regiones de ARN del gen objetivo. Una ribozima de la invención puede tener una secuencia de nucleótidos dentro de o circundante al sitio de enlazamiento al sustrato que imparte una actividad de escisión de ARN a la molécula.

ARN de interferencia (ARNi)

En un aspecto, la invención describe una molécula inhibidora de ARN, una así llamada molécula "ARNi", que comprende una secuencia de enzima transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención. Una molécula de ARN de cadena doble (ARNds), por ejemplo, moléculas de ARNsi, una microARN (ARNmi) y/o ARN de horquilla corta (ARNsh). La molécula de ARNi, por ejemplo, ARNsi (ARN inhibidor pequeño) puede inhibir la expresión de un gen de enzima transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, y/o la ARNmi (microARN) para inhibir la traducción de un mensajero de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa en un aspecto, la molécula de ARNi, por ejemplo ARNsi y/o ARNmi, tiene aproximadamente 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o más nucleótidos dúplex de longitud. El ARNi puede entrar a una célula y producir la degradación de un ARN de cadena sencilla (ARNss) de secuencias similares o idénticas, incluyendo ARNm endógeno. Cuando una célula es expuesta a un ARN de cadena doble (ARNds), el ARNm del gen homólogo es degradado selectivamente por un proceso denominado interferencia de ARN (iARN). Un mecanismo posible básico detrás de la iARN es la ruptura de un ARN de cadena doble (ARNds) coincidiendo con una secuencia genética específica en trozos cortos denominada interferencia de ARN corta, la cual dispara la degradación de ARNm que coincide con su secuencia. En un aspecto, la iARN tal como se describe es usada en terapias de silenciamiento de genes, por ejemplo, Shuey (2002) Drug Discov. Today 7:1040-1046. En un aspecto, la invención describe métodos para degradar selectivamente ARN utilizando moléculas de ARNi, por ejemplo, ARNsi y/o ARNmi, de la invención. El proceso puede ser puesto en práctica *in vitro*, *ex vivo* o *in vivo*. En un aspecto, las moléculas de ARNi tal como están descritas pueden ser utilizadas para generar una mutación de pérdida de función en una célula, un órgano, o un animal.

En un aspecto, la introducción intracelular de los iARN es por internalización de un ligando específico para la célula objetivo enlazado a una proteína de enlazamiento ARN que comprende un ARNi (por ejemplo, microARN) es adsorbida. El ligando específico a un antígeno de superficie celular objetivo único. El ligando puede ser internalizado espontáneamente después de enlazarse al antígeno de la superficie celular. Si el antígeno de la superficie celular no es internalizado de manera natural después de enlazamiento a su ligando, la internalización puede ser promovida por la incorporación de un péptido rico en arginina, u otro péptido permeable de la membrana, en la estructura del ligando o proteína de enlazamiento del ARN o unión de tal péptido al ligando o a la proteína de enlazamiento de ARN. Véase, por ejemplo, las solicitudes de patente de los Estados Unidos publicaciones números. 20060030003; 20060025361; 20060019286; 20060019258. En un aspecto, la invención provee formulaciones basadas en lípidos para administrar, por ejemplo, introducir ácidos nucleicos de la invención como partículas lipídicas de ácido nucleico que comprende una molécula de ARNi en una célula, véase, por ejemplo, la publicación de solicitud de patente de los Estados Unidos número. 20060008910.

Los métodos para hacer y utilizar moléculas de ARNi, por ejemplo, ARNsi y/o ARNmi, para degradar selectivamente a ARN son bien conocidos en la técnica, véase, por ejemplo, las patentes de los Estados Unidos No. 6,506,559; 6,511,824; 6,515,109; 6,489,127.

Modificación de ácidos nucleicos

La invención provee métodos para generar variantes de los ácidos nucleicos de la invención, por ejemplo, aquellos que codifican una transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. Estos métodos pueden ser repetidos o utilizados en diversas combinaciones para generar d-aminoácido transferasas que tienen una actividad alterada o diferente o una estabilidad alterada o diferente de la de una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, codificadas por el ácido nucleico plantilla. Estos métodos también pueden ser repetidos o utilizados en diversas combinaciones, por ejemplo, para generar variaciones en la expresión gen/mensaje, traducción de mensajes o estabilidad de mensajes. En otro aspecto, la composición genética de una célula es alterada, por ejemplo, por modificación de un gen homólogo *ex vivo*, seguida por su inserción en la célula.

En realizaciones alternativas, un ácido nucleico de la invención puede ser alterado por cualquier medio. Por ejemplo, métodos aleatorios o estocásticos, o métodos no estocásticos, o "evolución dirigida", véase, por ejemplo, la Patente

de los Estados Unidos número 6,361,974. Los métodos por mutación aleatoria de genes son bien conocidos en la técnica, véase, por ejemplo, la Patente de los Estados Unidos No. 5,830,696. Por ejemplo, los mutágenos pueden ser utilizados para mutar aleatoriamente un gen. Los mutágenos incluyen, por ejemplo, luz ultravioleta o de radiación gamma, o un mutágeno químico, por ejemplo, mitomicina, ácido nitroso, psoralenos fotoactivados, solos o en combinación, para inducir rupturas de ADN propensas de reparación por recombinación. Otros mutágenos químicos incluyen, por ejemplo, bisulfito de sodio, ácido nitroso, hidroxilamina, hidrazina o ácido fórmico. Otros mutágenos son análogos de los precursores de nucleótidos, por ejemplo, nitrosoguanidina, 5-bromouracilo, 2-aminopurina, o acridina. Estos agentes pueden ser agregados a una reacción de PCR en lugar del nucleótido precursor haciendo mutar por lo tanto la secuencia. También pueden utilizarse agentes intercalables tales como proflavina, acriflavina, quinacrina y similares.

Puede utilizarse cualquier técnica de biología molecular, por ejemplo, mutagénesis por PCR aleatorio, véase, por ejemplo, Rice (1992) *Proc. Natl. Acad. Sci. USA* 89:5467-5471; o, mutagénesis por casetes múltiples combinacionales, véase, por ejemplo, Crameri (1995) *Biotechniques* 18:194-196. Alternativamente, los ácidos nucleicos, por ejemplo genes, pueden ser reensamblados después de fragmentación aleatoria o "estocástica", véase, por ejemplo, las patentes de los Estados Unidos Nos. 6,291,242; 6,287,862; 6,287,861; 5,955,358; 5,830,721; 5,824,514; 5,811,238; 5,605,793. En aspectos alternativos, se introducen modificaciones, adiciones o eliminaciones por PCR propenso a error, mezcla, mutagénesis dirigida por oligonucleótidos, PCR de ensamblaje, mutagénesis PCR sexual, mutagénesis *in vivo*, mutagénesis por casete, mutagénesis de ensamblaje recursivo, mutagénesis de ensamblaje exponencial, mutagénesis específica del sitio, reensamble de genes (por ejemplo, reensamble de genes, véase, por ejemplo, la patente de los Estados Unidos No. 6,537,776), mutagénesis por saturación del sitio del gen (GSSM), reensamble por ligazón sintética (SLR), recombinación, recombinación de secuencia recursiva, mutagénesis de ADN modificado por fosfotioato, mutagénesis de plantilla que contiene uracilo, mutagénesis dúplex con brechas, mutagénesis de reparación de errores puntuales, mutagénesis de cepas anfitrionas deficientes en reparación, mutagénesis química, mutagénesis radiogénica, mutagénesis por eliminación, mutagénesis por restricción-selección, mutagénesis por restricción-purificación, síntesis de genes artificiales, mutagénesis de ensamblaje, creación de multímeros de ácidos nucleicos quiméricos, y/o una combinación de estos y otros métodos.

Las siguientes publicaciones describen una variedad de procedimientos de recombinación recursivos y/o métodos que pueden ser incorporados en los métodos de la invención: Stemmer (1999) "Molecular breeding of viruses for targeting and other clinical properties" *Tumor Targeting* 4:1-4; Ness (1999) *Nature Biotechnology* 17:893-896; Chang (1999) "Evolution of a cytokine using DNA family shuffling" *Nature Biotechnology* 17:793-797; Minshull (1999) "Protein evolution by molecular breeding" *Current Opinion in Chemical Biology* 3:284-290; Christians (1999) "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling" *Nature Biotechnology* 17:259-264; Crameri (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature* 391:288-291; Crameri (1997) "Molecular evolution of an arsenate detoxification pathway by DNA shuffling," *Nature Biotechnology* 15:436-438; Zhang (1997) "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening" *Proc. Natl. Acad. Sci. USA* 94:4504-4509; Patten et al. (1997) "Applications of DNA Shuffling to Pharmaceuticals and Vaccines" *Current Opinion in Biotechnology* 8:724-733; Crameri et al. (1996) "Construction and evolution of antibody-phage libraries by DNA shuffling" *Nature Medicine* 2:100-103; Gates et al. (1996) "Affinity selective isolation of ligands from peptide libraries through display on a lac repressor 'headpiece dimer'" *Journal of Molecular Biology* 255:373-386; Stemmer (1996) "Sexual PCR and Assembly PCR" In: *The Encyclopedia of Molecular Biology*. VCH Publishers, New York. pp.447-457; Crameri and Stemmer (1995) "Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wildtype cassettes" *BioTechniques* 18:194-195; Stemmer et al. (1995) "Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides" *Gene*, 164:49-53; Stemmer (1995) "The Evolution of Molecular Computation" *Science* 270: 1510; Stemmer (1995) "Searching Sequence Space" *Bio/Technology* 13:549-553; Stemmer (1994) "Rapid evolution of a protein *in vitro* by DNA shuffling" *Nature* 370:389-391; and Stemmer (1994) "DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution." *Proc. Natl. Acad. Sci. USA* 91:10747-10751.

Métodos de mutación para generar diversidad incluyen, por ejemplo, mutagénesis dirigida al sitio (Ling et al. (1997) "Approaches to DNA mutagenesis: an overview" *Anal Biochem.* 254(2): 157-178; Dale et al. (1996) "Oligonucleotide directed random mutagenesis using the phosphorothioate method" *Methods Mol. Biol.* 57:369-374; Smith (1985) "In vitro mutagenesis" *Ann. Rev. Genet.* 19:423-462; Botstein & Shortle (1985) "Strategies and applications of *in vitro* mutagenesis" *Science* 229:1193-1201; Carter (1986) "Site-directed mutagenesis" *Biochem. J.* 237:1-7; and Kunkel (1987) "The efficiency of oligonucleotide directed mutagenesis" in *Nucleic Acids & Molecular Biology* (Eckstein, F. and Lilley, D. M. J. eds., Springer Verlag, Berlín)); mutagénesis utilizando plantillas que contienen un uracilo (Kunkel (1985) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel et al. (1987) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Methods in Enzymol.* 154, 367-382; and Bass et al. (1988) "Mutant Trp repressors with new DNA-binding specificities" *Science* 242:240-245); oligonucleotidedirected mutagenesis (*Methods in Enzymol.* 100: 468-500 (1983); *Methods in Enzymol.* 154: 329-350 (1987); Zoller (1982) "Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment" *Nucleic Acids Res.* 10:6487-6500;

Zoller & Smith (1983) "Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors" *Methods in Enzymol.* 100:468-500; and Zoller (1987) "Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template" *Methods in Enzymol.* 154:329-350; phosphorothioate-modified DNA mutagenesis (Taylor (1985) "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA" *Nucl. Acids Res.* 13: 8749-8764; Taylor (1985) "The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA" *Nucl. Acids Res.* 13: 8765-8787 (1985); Nakamaye (1986) "Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis" *Nucl. Acids Res.* 14: 9679-9698; Sayers (1988) "Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis" *Nucl. Acids Res.* 16:791-802; and Sayers et al. (1988) "Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide" *Nucl. Acids Res.* 16: 803-814); mutagenesis using gapped duplex DNA (Kramer et al. (1984) "The gapped duplex DNA approach to oligonucleotide-directed mutation construction" *Nucl. Acids Res.* 12: 9441-9456; Kramer & Fritz (1987) *Methods in Enzymol.* "Oligonucleotide-directed construction of mutations via gapped duplex DNA" 154:350-367; Kramer (1988) "Improved enzymatic *in vitro* reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations" *Nucl. Acids Res.* 16: 7207; and Fritz (1988) "Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions *in vitro*" *Nucl. Acids Res.* 16: 6987-6999).

Protocolos adicionales que pueden ser utilizados para poner en práctica la invención incluyen reparación de errores puntuales (Kramer (1984) "Point Mismatch Repair" *Cell* 38:879-887), mutagenesis using repair-deficient host strains (Carter et al. (1985) "Improved oligonucleotide site-directed mutagenesis using M13 vectors" *Nucl. Acids Res.* 13: 4431-4443; and Carter (1987) "Improved oligonucleotide-directed mutagenesis using M13 vectors" *Methods in Enzymol.* 154: 382-403), deletion mutagenesis (Eghtedarzadeh (1986) "Use of oligonucleotides to generate large deletions" *Nucl. Acids Res.* 14: 5115), restriction-selection and restriction-selection and restriction-purification (Wells et al. (1986) "Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin" *Phil. Trans. R. Soc. Lond. A* 317: 415-423), mutagenesis por síntesis total de genes (Nambiar et al. (1984) "Total synthesis and cloning of a gene coding for the ribonuclease S protein" *Science* 223: 1299-1301; Sakamar and Khorana (1988) "Total synthesis and expression of a gene for the α -subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)" *Nucl. Acids Res.* 14: 6361-6372; Wells et al. (1985) "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites" *Gene* 34:315-323; and Grundstrom et al. (1985) "Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis" *Nucl. Acids Res.* 13: 3305-3316), double-strand break repair (Mandecki (1986); Arnold (1993) "Protein engineering for unusual environments" *Current Opinion in Biotechnology* 4:450-455. "Oligonucleotide-directed double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis" *Proc. Natl. Acad. Sci. USA*, 83:7177-7181). Detalles adicionales sobre muchos de los métodos anteriores pueden ser encontrados en *Methods in Enzymology* Volume 154, la cual también describe controles útiles para resolución de problemas con diversos métodos de mutagénesis.

Protocolos que pueden ser utilizados para practicar la invención están descritos en las patentes Nos. 5,605,793 to Stemmer (Feb. 25, 1997), "Methods for *In vitro* Recombination;" Patente de los Estados Unidos No. 5,811,238 de Stemmer et al. (Sep. 22, 1998) "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" Patente de los Estados Unidos No. 5,830,721 de Stemmer et al. (Nov. 3, 1998), "DNA Mutagenesis by Random Fragmentation and Reassembly;" Patente de los Estados Unidos No. 5,834,252 de Stemmer, et al. (Nov. 10, 1998) "End-Complementary Polymerase Reaction;" Patente de los Estados Unidos No. 5,837,458 de Minshull, et al. (Nov. 17, 1998), "Methods and Compositions for Cellular and Metabolic Engineering;" WO 95/22625, Stemmer and Crameri, "Mutagenesis by Random Fragmentation and Reassembly;" WO 96/33207 de Stemmer and Lipschutz "End Complementary Polymerase Chain Reaction;" WO 97/20078 de Stemmer and Crameri "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" WO 97/35966 de Minshull y Stemmer, "Methods and Compositions for Cellular and Metabolic Engineering;" WO 99/41402 by Punnonen et al. "Targeting of Genetic Vaccine Vectors;" WO 99/41383 de Punnonen et al. "Antigen Library Immunization;" WO 99/41369 de Punnonen et al. "Genetic Vaccine Vector Engineering;" WO 99/41368 de Punnonen et al. "Optimization of Immunomodulatory Properties of Genetic Vaccines;" EP 752008 de Stemmer and Crameri, "DNA Mutagenesis by Random Fragmentation and Reassembly;" EP 0932670 de Stemmer "Evolving Cellular DNA Uptake by Recursive Sequence Recombination;" WO 99/23107 de Stemmer et al., "Modification of Virus Tropism and Host Range by Viral Genome Shuffling;" WO 99/21979 de Apt et al., "Human Papillomavirus Vectors;" WO 98/31837 de del Cardayre et al. "Evolution of Whole Cells and Organisms by Recursive Sequence Recombination;" WO 98/27230 de Patten and Stemmer, "Methods and Compositions for Polypeptide Engineering;" WO 98/27230 de Stemmer et al., "Methods for Optimization of Gene Therapy by Recursive Sequence Shuffling and Selection;" WO 00/00632, "Methods for Generating Highly Diverse Libraries;" WO 00/09679, "Methods for Obtaining *in vitro* Recombined Polynucleotide Sequence Banks and Resulting Sequences;" WO 98/42832 de Arnold et al., "Recombination of Polynucleotide Sequences Using Random or Defined Primers;" WO 99/29902 de Arnold et al., "Method for Creating Polynucleotide and Polypeptide Sequences;" WO 98/41653 de Vind, "An *in vitro* Method for Construction of a DNA Library;" WO 98/41622 de Borchert et al., "Method for Constructing a Library Using DNA Shuffling;" y WO 98/42727 de Pati y Zarling, "Sequence Alterations using Homologous Recombination."

Protocolos que pueden ser utilizados para poner en práctica la invención (que proveen detalles concernientes a diversos métodos de generación de diversidad) están descritos, por ejemplo, en la Solicitud de Patente de los Estados Unidos serie no. (USSN) 09/407,800, "SHUFFLING OF CODON ALTERED GENES" de Patten et al. Presentada el 28 de septiembre de 1999; "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION" de del Cardayre et al., Patente de los Estados Unidos No. 6,379,964; "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" de Crameri et al., Patente de los Estados Unidos Nos. 6,319,714; 6,368,861; 6,376,246; 6,423,542; 6,426,224 y PCT/US00/01203; "USE OF CODON-VARIED OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING" de Welch et al., Patente de los Estados Unidos No. 6,436,675; "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" de Selifonov et al., presentada el 18 de enero de 2000, (PCT/US00/01202) y, por ejemplo "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" de Selifonov et al., presentada el 18 de julio de 2000 (U.S. Ser. No. 09/618,579); "METHODS OF POPULATING DATA STRUCTURES FOR USE IN EVOLUTIONARY SIMULATIONS" de Selifonov and Stemmer, presentada el 18 de enero de 2000 (PCT/US00/01138); y "SINGLE-STRANDED NUCLEIC ACID TEMPLATE-MEDIATED RECOMBINATION AND NUCLEIC ACID FRAGMENT ISOLATION" de Affholter, presentada el 6 de septiembre de 2000 (U.S. Ser. No. 09/656,549); y Patente de los Estados Unidos Nos. 6,177,263; 6,153,410.

Los métodos no estocásticos, o "de evolución dirigida", incluyen, por ejemplo, mutagénesis por saturación (GSSM), reensamble por ligazón sintética (SLR), o una combinación de los mismos y son usados para modificar los ácidos nucleicos de la invención para generar transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas con propiedades nuevas o alteradas (por ejemplo, actividad bajo condiciones altamente ácidas o alcalinas, altas o bajas temperaturas, y similares). Los polipéptidos codificados por los ácidos nucleicos modificados pueden ser seleccionados para una actividad. Puede utilizarse cualquier modalidad o protocolo de prueba, por ejemplo, utilizando una plataforma de arreglos capilares. Véase, por ejemplo, las Patentes de los Estados Unidos Nos. 6,361,974; 6,280,926; 5,939,250. .

Mutagénesis por saturación en el sitio del gen, o, GSSM

La invención también provee métodos para hacer enzimas utilizando mutagénesis por saturación en el sitio del gen, o, GSSM, tal como se describe aquí, y también en las patentes de los Estados Unidos Nos 6,171,820 y 6,579,258. En un aspecto, cebadores de codón que contienen una secuencia N,N,G/T degenerada, se utiliza para introducir mutaciones puntuales en un polinucleótido, por ejemplo, una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa o un anticuerpo de la invención, de tal manera que se genera un conjunto de polipéptidos de progenie en los cuales un rango completo de sustituciones de aminoácidos está representado en cada posición de aminoácido, por ejemplo, un residuo de aminoácido en un sitio activo de una enzima o un sitio de enlazamiento del ligando destinado para ser modificado. Estos oligonucleótidos pueden comprender una primera secuencia homóloga continua, una secuencia N,N,G/T degenerada, y, en un aspecto, una secuencia homóloga. Los productos de traducción de progenie corriente abajo del uso de tales oligonucleótidos incluyen todos los posibles cambios de aminoácidos en cada sitio de aminoácido a lo largo del polipéptido, porque la degeneración de la secuencia N,N,G/T incluye codones para todos los 20 aminoácidos. En un aspecto, uno de tales oligonucleótidos degenerados (compuesto de, por ejemplo, un casete N,N,G/T degenerado) se utiliza para someter cada codón original en una plantilla de polinucleótido progenitora a un rango completo de sustituciones por codones. En otro aspecto, se usan al menos dos casetes degenerados- bien sea en el mismo oligonucleótido o no, para someter al menos dos codones originales en una plantilla de polinucleótido progenitor hasta un rango completo de sustituciones por codón. Por ejemplo, más de una secuencia N,N,G/T puede estar contenida en un oligonucleótido para introducir mutaciones en aminoácidos en más de un sitio. Esta pluralidad de secuencias N,N,G/T puede ser directamente contigua, o separada por una o más secuencias de nucleótidos adicionales. En otro aspecto, pueden utilizarse oligonucleótidos útiles para introducir adiciones y eliminaciones bien sea solas o en combinación con los codones que contiene una secuencia N,N,G/T, para introducir cualquier combinación o permutación de adiciones, eliminaciones y/o sustituciones de aminoácidos.

En un aspecto, la mutagénesis simultánea de dos o más posiciones de aminoácidos contiguas se hace utilizando un oligonucleótido que contiene tripletes contiguos N,N,G/T, esto es una secuencia (N,N,G/T)_n degenerada. En otro aspecto, se usan casetes degenerados que tienen menos degeneración que la secuencia N,N,G/T. Por ejemplo, puede ser deseable en algunos casos utilizar (por ejemplo en un oligonucleótido) una secuencia de triplete degenerada que comprende solamente una N, en donde dicha N puede ser una primera o segunda o tercera posición del triplete. Cualquier otra base que incluye cualquier combinación y permutación de la misma puede ser utilizada en las dos posiciones restantes del triplete. Alternativamente, puede ser deseable en algunos casos utilizar (por ejemplo, en un oligo) una secuencia de triplete N,N,N degenerada.

En un aspecto, el uso de tripletes degenerados (por ejemplo, tripletes N,N,G/T) permite una generación sistemática y fácil de un rango completo de posibles aminoácidos naturales (para un total de 20 aminoácidos) en todas y cada una

de las posiciones de aminoácidos en un polipéptido (en aspectos alternativos, los métodos también incluyen generación de menos que todas las posibles sustituciones por posición de residuo de aminoácido, o codón). Por ejemplo, para un polipéptido de 100 aminoácidos, pueden generarse 2000 especies distintas (esto es, 20 posibles aminoácidos por posición X 100 posiciones de aminoácidos) a través del uso de un oligonucleótido o conjunto de oligonucleótidos que contienen un triplete N,N,G/T degenerado, 32 secuencias individuales pueden codificar para todos los posibles 20 aminoácidos naturales. Así, en un recipiente de reacción en el cual se somete una secuencia de polinucleótidos progenitora a mutagénesis por saturación utilizando al menos uno de tales oligonucleótidos, se generan 32 polinucleótidos de progenie distintos que codifican 20 polipéptidos distintos. En contraste, el uso de un oligonucleótido no degenerado en la mutagénesis dirigida al sitio lleva a solamente un polipéptido de progenie producido por recipiente de reacción. Los oligonucleótidos no degenerados pueden ser usados en un aspecto en combinación con cebadores degenerados divulgados; por ejemplo, los oligonucleótidos no degenerados pueden ser utilizados para generar mutaciones puntuales específicas en un polinucleótido de trabajo. Esto provee un medio para generar mutaciones puntuales silenciosas específicas, mutaciones puntuales que llevan a cambios correspondientes en aminoácidos, y mutaciones puntuales que producen la generación de codones de detención y la correspondiente expresión de los fragmentos de polipéptido.

En un aspecto, cada recipiente de reacción de mutagénesis por saturación contiene polinucleótidos que codifican al menos 20 moléculas de polipéptidos de progenie por ejemplo, (por ejemplo, transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas) de tal manera que todos los 20 aminoácidos naturales están representados en la posición de aminoácidos específica correspondiente a la posición de codón mutagenizado en el polinucleótido progenitor (otros aspectos usan menos de todas las 20 combinaciones naturales). Los polipéptidos de progenie degenerada 32 veces generados a partir de cada recipiente de reacción de mutagénesis por saturación pueden ser sometidos a amplificación clonal (por ejemplo, clonados en un anfitrión adecuado, por ejemplo, un anfitrión de *E. coli*, utilizando, por ejemplo, un vector de expresión) y sometidos a selección de expresión. Cuando un polipéptido de progenie individual es identificado por selección para desplegar un cambio favorable en propiedades (cuando se compara con un polipéptido progenitor, tal como una actividad por ejemplo, actividad de transaminasa, por ejemplo, una actividad de d-aminoácido transferasa, y/o una actividad de oxidorreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de d-aminoácido deshidrogenasa o una actividad de ω -transaminasa incrementada bajo condiciones alcalinas o ácidas) puede ser secuenciado para identificar la sustitución de aminoácidos correspondientemente favorable contenida en el mismo.

En un aspecto, al mutagenizar todas y cada una de las posiciones de aminoácidos en un polipéptido progenitor utilizando mutagénesis por saturación tal como se divulga aquí, pueden identificarse cambios en aminoácidos favorables en más de una posición de aminoácidos. Una o más moléculas progenitoras pueden ser generadas de manera que contengan una combinación de todas o parte de estas sustituciones favorables de aminoácidos. Por ejemplo, si se identifican 2 cambios favorables específicos de aminoácidos en cada una de las 3 posiciones de aminoácidos en un polipéptido, las permutaciones incluyen 3 posibilidades en cada posición (sin cambio a partir del aminoácido original, y cada uno de dos cambios favorables) y 3 posiciones. Así, hay $3 \times 3 \times 3$ o 27 posibilidades totales, incluyendo 7 que fueron examinadas previamente - 6 mutaciones en puntos individuales (esto es 2 en cada una de tres posiciones) y sin cambios en ninguna posición.

En aún otro aspecto, la mutagénesis por saturación de sitio puede ser utilizada junto con mezcla, quimerización, recombinación u otros procesos de mutagénesis, junto con selección. Esta invención provee el uso de cualquier proceso de mutagénesis, incluyendo mutagénesis por saturación, de manera reiterativa. En una ejemplificación, el uso reiterativo de cualquier proceso de mutagénesis es utilizado en combinación con la selección.

La invención también provee el uso de cebadores de codones originales (que contienen una secuencia N,N,N degenerada) para introducir mutaciones puntuales en un polinucleótido, de tal manera que generen un conjunto de polipéptidos de progenie en el cual un rango completo de sustituciones de aminoácidos individuales es representando en cada posición de aminoácido (mutagénesis por saturación del sitio del gen (GSSM)). Los oligos usados están comprendidos contiguos a una primera secuencia homóloga, una secuencia N,N,N degenerada y preferiblemente pero no necesariamente a una segunda secuencia homóloga. Los productos translacionales de la progenie corriente abajo del uso de tales oligos incluyen todos los posibles cambios en aminoácidos, en cada sitio de aminoácidos a lo largo del polipéptido, porque la degeneración de la secuencia N,N,N incluye codones para todos los 20 aminoácidos.

En un aspecto, un tal oligo degenerado (compuesto de un casete N,N,N degenerado) se utiliza para someter cada codón original en una plantilla de nucleótidos progenitora a un rango completo de sustituciones de codones. En otro aspecto, se usan al menos dos casetes de N,N,N degenerados - bien sea en el mismo oligo o no, para someter al menos dos codones originales en una plantilla de polinucleótidos progenitora a un rango completo de sustituciones de codones. Así, más de una secuencia de N,N,N puede estar contenida en el oligo para introducir mutaciones en aminoácidos en más de un sitio. La pluralidad de secuencias N,N,N puede ser directamente contigua, o separada por una o más secuencias de nucleótidos adicionales. En otro aspecto, los oligos útiles para introducir adiciones y eliminaciones pueden ser utilizados bien sea solos o en combinación con los codones que contienen una secuencia

N,N,N, para introducir cualquier combinación o permutación de adiciones, eliminaciones y/o sustituciones de aminoácidos.

En una ejemplificación particular, es posible mutagenizar simultáneamente dos o más posiciones contiguas de aminoácidos utilizando un oligo que contenga tripletes N,N,N contiguos, esto es, una secuencia (N,N,N)_n degenerada.

En otro aspecto, la presente invención provee el uso de casetes degenerados que tienen menos degeneración que la secuencia N,N,N. Por ejemplo, puede ser deseable en algunos casos usar (por ejemplo en un oligo) una secuencia triplete degenerada compuesta de solamente un N, en donde el N puede ser la primera, segunda o tercera posición del triplete. Cualquier otra base que incluya cualquier combinación y permutación de las mismas puede ser utilizada en las dos posiciones restantes del triplete. Alternativamente, puede ser deseable en algunos casos usar (por ejemplo, en un oligo) una secuencia triplete N,N,N, N,N,G/T degenerada o una secuencia triplete N,N, G/C.

Es evidente, sin embargo, que el uso de un triplete degenerado (tal como una secuencia triplete N,N,G/T o N,N, G/C) tal como se divulga en la presente invención es ventajoso por varias razones. En un aspecto, esta invención provee un medio para generar sistemáticamente y con bastante facilidad la sustitución del rango completo de posibles aminoácidos (para un total de 20 aminoácidos) en todas y cada una de las posiciones de aminoácidos en un polipéptido. Así, para un polipéptido de 100 aminoácidos, la invención provee una manera de generar sistemáticamente y de manera bastante fácil 2000 especies distintas (esto es, 20 posibles aminoácidos por posición por 100 posiciones de aminoácidos). Es evidente que se provee, a través del uso de un oligo que contiene una secuencia N,N,G/T o N,N, G/C de triplete degenerada, 32 secuencias individuales que codifican por 20 posibles aminoácidos. Así, en un recipiente de reacción en el cual una secuencia de polinucleótidos progenitora es sometida a mutagénesis por saturación utilizando tal oligo, hay 32 polinucleótidos de progenie distinta que codifican 20 polipéptidos distintos. En contraste, el uso de un oligo no degenerado en mutagénesis dirigida al sitio lleva a solamente un producto de polipéptido de progenie por recipiente de reacción.

Esta invención también provee el uso de oligos no degenerados, los cuales pueden en un aspecto ser utilizados en combinación con los cebadores degenerados divulgados. Es evidente que en algunas situaciones, es ventajoso utilizar oligos no degenerados para generar mutaciones puntuales específicas en un polinucleótido de trabajo. Esto provee un medio para generar mutaciones puntuales silenciosas específicas, mutaciones que llevan a cambios correspondientes a aminoácidos y mutaciones puntuales que producen la generación de codones de detención y la correspondiente expresión de fragmentos de polipéptidos.

Así, en un aspecto de esta invención, cada recipiente de reacción de mutagénesis por saturación contiene polinucleótidos que codifican al menos 20 moléculas de polipéptidos de progenie de tal forma que todos los 20 aminoácidos están representados en la única posición de aminoácido específica correspondiente a la posición del codón mutagenizado en el polinucleótido progenitor. Los polipéptidos de la progenie degenerada 32 veces generados a partir de dicho recipiente de reacción de mutagénesis por saturación pueden ser sometidos a amplificación clonal (por ejemplo, clonados en un anfitrión *E. coli* adecuado usando un vector de expresión) y sometidos a selección de expresión. Cuando se identifica un polipéptido de progenie individual seleccionando para desplegar un cambio favorable en propiedades (cuando se compara con el polipéptido progenitor) puede ser secuenciado para identificar la sustitución de aminoácido favorable correspondientemente contenida en el mismo.

Es evidente que al mutagenizar todas y cada una de las posiciones de aminoácidos en un polipéptido progenitor utilizando mutagénesis por saturación tal como se divulga aquí, pueden identificarse cambios favorables en aminoácidos en más de una posición de aminoácido. Pueden generarse una o más nuevas moléculas de progenie que contienen una combinación de todas o parte de estas sustituciones favorables de aminoácidos. Por ejemplo, si se identifican dos cambios específicos favorables de aminoácidos en cada una de 3 posiciones de aminoácidos en un polipéptido, las permutaciones incluyen 3 posibilidades en cada posición (sin cambio desde el aminoácido original y cada uno de los dos cambios favorables) y 3 posiciones. Así, hay 3 x 3 x 3 o 27 posibilidades totales, incluyendo 7 que han sido examinadas previamente - 6 mutaciones puntuales individuales (esto es 2 en cada una de tres posiciones) y sin cambios en ninguna posición.

Así, en una ejemplificación no limitante, esta invención provee el uso de mutagénesis por saturación en combinación con procesos de mutagenización adicional, tales como procesos en donde dos o más polinucleótidos relacionados son introducidos en una célula anfitriona adecuada de tal manera que se genera un polinucleótido híbrido por recombinación y reordenamiento reductor.

Además de llevar a cabo la mutagénesis a lo largo de la secuencia completa de un gen, la presente invención provee que la mutagénesis puede ser de uso para remplazar cada uno de cualquier número de bases en una secuencia de polinucleótidos, en donde el número de bases que van a ser mutagenizadas es preferiblemente cada entero desde 15 a 100,000. Así, en vez de mutagenizar cada posición a lo largo de una molécula, se puede someter cada o un número discreto de bases (preferiblemente un subconjunto que totaliza desde 15 a 100,000) a la

mutagénesis. Preferiblemente, un nucleótido separado se utiliza para mutagenizar cada posición o grupo de posiciones a lo largo de una secuencia de polinucleótidos. Un grupo de 3 posiciones que van a ser mutagenizadas puede ser un codón. Las mutaciones son introducidas preferiblemente utilizando un cebador mutagénico, que contiene un casete heterólogo, también denominado como casete mutagénico. Casetes de ejemplo pueden tener de 1 a 500 bases. Cada posición de nucleótidos en tales casetes heterólogos pueden ser N, A, C, G, T, A/C, A/G, A/T, C/G, C/T, G/T, C/G/T, A/G/T, A/C/T, A/C/G, o E, donde E es cualquier base que no es A, C, G o T (A puede ser denominada como un oligo de diseño).

En un sentido general, la mutagénesis por saturación está compuesta de mutagenizar un conjunto completo de casetes mutagénicos (en donde cada casete tiene aproximadamente 1-500 bases de longitud) en secuencias de polinucleótidos definidas para ser mutagenizadas (en donde la secuencia que va a ser mutagenizada tiene preferiblemente 15,000 a 100,000 bases de longitud). Así, un grupo de mutaciones (que varía de 1 a 100 mutaciones) es introducida en cada casete que va a ser mutagenizado. Un agrupamiento de mutaciones que va a ser introducido en un casete puede ser diferente o el mismo de un segundo agrupamiento de mutaciones que van a ser introducidas en un segundo casete, durante la aplicación de una ronda de mutagénesis por saturación. Tales agrupamientos son ejemplificados por eliminaciones, adiciones, agrupamientos de codones particulares y agrupaciones de casetes de nucleótidos particulares.

Secuencias definidas que van a ser mutagenizadas incluyen el gen completo, rutas, ADNc, un marco de lectura abierto completo (ORF) y un promotor, potenciador, represor/transactivador completos, origen de replicación, intrones, operadores o cualquier grupo de polinucleótidos funcional. En general, una "secuencia definida" para este propósito puede ser cualquier polinucleótido que tiene una secuencia de polinucleótidos de 15 bases y secuencias de polinucleótidos de longitudes entre 15 bases y 15.000 bases (esta invención cita específicamente cada entero entre ellos). Las consideraciones en la selección de agrupamientos de codones incluyen tipos de aminoácidos codificados por un casete mutagénico degenerado.

En una ejemplificación de un agrupamiento de mutaciones que pueden ser introducidas en un casete mutagénico, esta invención provee específicamente sustituciones de codones degenerados (usando oligos degenerados) que codifican para 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 y 20 aminoácidos en cada posición y una biblioteca de polipéptidos codificados por los mismos.

Reensamble por ligazón sintética (SLR)

La invención provee un sistema de modificación genética no estocástico denominado "reensamble por ligazón sintética" o, simplemente "SLR", un "proceso de evolución dirigido", para generar polipéptidos, por ejemplo, transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas, o anticuerpos de la invención, con propiedades nuevas o alteradas.

El SLR es un método para ligar fragmentos de oligonucleótidos entre sí de manera no estocástica. Este método difiere de la mezcla de oligonucleótidos estocástica en que los bloques de construcción de los ácidos nucleicos no son mezclados, concatenados o quimerizados aleatoriamente, sino que son ensamblados de manera no estocástica (Véase, por ejemplo, las patentes de los Estados Unidos Nos. 6,773,900; 6,740,506; 6,713,282; ,635,449; 6,605,449; 6,537,776. En un aspecto, el SLR comprende: (a) proveer un polinucleótido de plantilla, en donde el polinucleótido de plantilla comprende secuencias que codifican un gen homólogo; (b) proveer una pluralidad de polinucleótidos de bloques de construcción, en donde los polinucleótidos del bloque de construcción están diseñados para reensamblarse por cruzamiento con el polinucleótido de plantilla en una secuencia predeterminada, y el polinucleótido de bloque de construcción comprende una secuencia que es una variante del gen homólogo y una secuencia homóloga al polinucleótido de plantilla que flanquea la secuencia variante; (c) combinar un polinucleótido de bloque de construcción con un polinucleótido de plantilla de tal manera que el polinucleótido de bloque de construcción se reensamble por cruzamiento con el polinucleótido de plantilla para generar polinucleótidos que comprenden variaciones de secuencia genética homólogas.

El SLR no depende de la presencia de altos niveles de homología entre los polinucleótidos que van a ser reordenados. Así, este método puede ser utilizado para generar de manera no estocástica bibliotecas (o conjuntos) de moléculas de progeie compuestas de más de 10^{100} quimeras diferentes. El SLR puede ser utilizado para generar bibliotecas compuestas de más de 10^{1000} quimeras de progeie diferentes. Así, aspectos de la presente invención incluyen métodos no estocásticos para producir un conjunto de moléculas de ácido nucleico quiméricas finalizadas que comparten un orden de ensamble global que es escogido por diseño. Este método incluye las etapas de generar por diseño una pluralidad de bloques de construcción de ácidos nucleicos específicos que tienen extremos ligables compatibles mutuamente utilizables, y el ensamble de estos bloques de construcción de ácidos nucleicos, de tal manera que se logra un orden de ensamble global diseñado.

Los extremos ligables compatibles mutuamente de los bloques de construcción de ácidos nucleicos que van a ser ensamblados se consideran como "utilizables" para este tipo de ensamble ordenado si son capaces de construir

bloques que van a ser acoplados en órdenes predeterminados. Así, el orden de ensamble en global en el cual los bloques de construcción de ácidos nucleicos pueden ser acoplados está especificado por el diseño de los extremos ligables. Si se va a usar más de una etapa de ensamble, entonces el orden de ensamble global en el cual los bloques de construcción de ácidos nucleicos pueden ser acoplados también está especificado por el orden secuencial de las etapas de ensamblaje. En un aspecto, las piezas de construcción fusionadas son tratadas con una enzima, tal como una ligasa (por ejemplo, T4 ADN ligasa), para alcanzar un enlace covalente de las piezas de construcción.

En un aspecto, el diseño de los bloques de construcción de oligonucleótidos es obtenido analizando un conjunto de plantillas de secuencias de ácidos nucleicos progenitores que sirven como base para producir un conjunto de progenie de polinucleótidos quiméricos finalizados. Estas plantillas de oligonucleótidos progenitoras sirven así como una fuente de información de secuencia que ayuda en el diseño de los bloques de construcción de ácidos nucleicos que van a ser mutagenizados, por ejemplo, quimerizados o mezclados. En un aspecto de este método, las secuencias de una pluralidad de plantillas de ácidos nucleicos progenitores son alineadas con el fin de seleccionar uno o más puntos de demarcación. Los puntos de demarcación pueden estar localizados en un área de homología, y están compuestos de uno o más nucleótidos. Estos puntos de demarcación son compartidos preferiblemente por al menos dos de las plantillas progenitoras. Los puntos de demarcación pueden por lo tanto ser utilizados para delinear las fronteras de los bloques de construcción de oligonucleótidos que van a ser generadas con el fin de reordenar los polinucleótidos progenitores. Los puntos de demarcación identificados y seleccionados en las moléculas progenitoras sirven como puntos de quimerización potenciales en el ensamble de las moléculas de progenie quimérica finales. Un punto de demarcación puede ser un área de homología (compuesta de al menos una base de nucleótidos homólogos) compartida por al menos dos secuencias de polinucleótidos progenitores. Alternativamente, un punto de demarcación puede ser un área de homología que es compartida por al menos la mitad de las secuencias de polinucleótidos progenitores, o, puede ser un área de homología que es compartida por al menos dos tercios de las secuencias de polinucleótidos progenitores. Incluso más preferiblemente un punto de demarcación utilizable es un área de homología que es compartida por al menos tres cuartas partes de las secuencias de polinucleótidos progenitoras, o, pueden ser compartidas por casi todas las secuencias de polinucleótidos progenitoras. En un aspecto, un punto de demarcación es un área de homología que es compartida por todas las secuencias de polinucleótidos progenitoras.

En un aspecto, un proceso de reensamble por ligazón se lleva a cabo exhaustivamente con el fin de generar una biblioteca exhaustiva de polinucleótidos quiméricos de progenie. En otras palabras, todas las posibles combinaciones de los bloques de construcción de ácidos nucleicos son representados en el conjunto de moléculas de ácidos nucleicos quiméricos finalizados. Al mismo tiempo, en otro aspecto, el orden de ensamble (esto es, el orden de ensamble de cada bloque de construcción en la secuencia de 5' a 3' de cada ácido nucleico quimérico finalizado) en cada combinación es por diseño (o no estocástico) tal como se describió anteriormente. Debido a la naturaleza no estocástica de esta invención, la posibilidad de productos colaterales no deseados se reduce grandemente.

En otro aspecto, el método de reensamble por ligazón se lleva a cabo sistemáticamente. Por ejemplo, el método es llevado a cabo con el fin de generar una biblioteca compartimentalizada sistemáticamente de moléculas de progenie, con compartimientos que pueden ser seleccionados sistemáticamente, por ejemplo, uno por uno. Estas palabras y esta invención provee que, a través del uso selectivo y juicioso de los bloques de construcción de ácidos nucleicos específicos, acoplado con el uso selectivo y juicioso de reacciones de avance secuencial pueda alcanzarse un diseño en donde conjuntos específicos de productos de progenie se hacen en cada uno de varios recipientes de reacción. Esto permite llevar a cabo un procedimiento de examen y selección sistemático. Así, estos métodos permiten que un número potencialmente muy grande de moléculas de progenie sea examinado sistemáticamente en grupos más pequeños. Debido a su capacidad para formar quimerizaciones en una forma que es altamente flexible a la vez que exhaustiva y sistemática, particularmente cuando hay un bajo nivel de homología entre las moléculas progenitoras, estos métodos proveen la generación de una biblioteca (o conjunto) compuesto de un gran número de moléculas de progenie. Debido a la naturaleza no estocástica de la presente invención de reensamblaje por ligazón, las moléculas de progenie generadas preferiblemente comprenden una biblioteca de moléculas de ácidos nucleico quimérico finalizadas que tienen un orden de ensamblaje global que es escogido por el diseño. Los métodos de mutagénesis por saturación y de evolución dirigida optimizada también pueden ser utilizados para generar diferentes especies moleculares de progenie. Es evidente que la invención provee libertad de selección y control con respecto a la selección de puntos de demarcación, el tamaño y número de los bloques de construcción de ácidos nucleicos, y el tamaño y diseño de los acoplamientos. Es evidente, además, que el requerimiento para homología intermolecular está altamente relajado por la operabilidad de esta invención. En efecto, los puntos de demarcación incluso pueden ser escogidos en áreas de poca o ninguna homología intermolecular por ejemplo, debido al bamboleo del codón, esto es la degeneración de codones, pueden introducirse sustituciones de nucleótidos en los bloques de construcción de ácidos nucleicos sin alterar el aminoácido codificado generalmente en la plantilla progenitora correspondiente. Alternativamente, puede alterarse un codón de tal manera que la codificación para un aminoácido original es alterada. Esta invención provee que tales sustituciones pueden ser introducidas en el bloque de construcción de ácidos nucleicos con el fin de incrementar la incidencia de los puntos de demarcación homólogos

intramoleculares y así permitir que un número incrementado de acoplamientos se logre entre los bloques de construcción, lo cual a su vez permite generar un número más grande de moléculas quiméricas de progenie.

Reensamble de genes sintéticos

5 En un aspecto, la presente invención provee un método no estocástico denominado reensamble de genes sintéticos (por ejemplo, reensamble de genes, véase Patente de los Estados Unidos No. 6,537,776), que difiere de la mezcla estocástica en que los bloques de construcción de ácidos nucleicos no son mezclados o concatenados o quimerizados aleatoriamente sino que son ensamblados de manera no estocástica.

10 El reensamble de genes sintéticos no depende de la presencia de un alto nivel de homología entre polinucleótidos que van a ser mezclados. La invención puede ser utilizada para generar de manera no estocástica bibliotecas (o conjuntos) de moléculas de progenie que comprenden 10^{100} quimeras diferentes. De manera concebible, el reensamble de genes sintéticos puede ser utilizado para generar bibliotecas que comprenden más de 10^{1000} diferentes quimeras de progenie.

15 Así, en un aspecto, la invención provee un método no estocástico para producir un conjunto de moléculas de ácidos nucleicos quiméricas finalizadas que tienen un orden de ensamble global que es escogido por diseño, método que comprende las etapas de generar por diseño una pluralidad de bloques de construcción de ácidos nucleicos específicos que tienen extremos ligables compatibles mutuamente utilizables y ensamble de estos bloques de construcción de ácidos nucleicos, de tal manera que se alcanza el orden de ensamble global diseñado.

20 En un aspecto, el reensamble de genes sintéticos comprende un método de: 1) preparar una generación de progenie de moléculas que incluye una molécula que comprende una secuencia de polinucleótidos, por ejemplo, una molécula que comprende una secuencia que codifica un polipéptido); que es mutagenizada para alcanzar al menos una mutación, adición, eliminación y/o quimerización puntual, a partir de una o más plantillas de generación ancestrales o progenitoras; 2) seleccionar la molécula de generación de progenie, utilizando por ejemplo un método de alto rendimiento, para al menos una propiedad de interés (tal como una mejora en una actividad enzimática); 3) En un aspecto obtener y/o catalogar información estructural y/o funcional concerniente a las moléculas de
25 generación de progenitores y/o progenie; y 4) en un aspecto repetir cualquiera de las etapas 1) a 3). En un aspecto, se genera (por ejemplo, a partir de una plantilla de polinucleótidos progenitora), en lo que se denomina "mutagénesis por saturación del sitio de codón" una generación de progenie de polinucleótidos, que tiene cada una al menos un conjunto de hasta tres mutaciones puntuales contiguas (esto es, bases diferentes que comprenden un nuevo codón), de tal manera que cada codón (o cada familia de codones degenerados que codifican el mismo aminoácido) es representada en cada posición del codón. Correspondiendo con, y codificado por esta generación de progenie de polinucleótidos, también se genera un conjunto de polipéptidos de progenie, que tienen cada uno al menos una mutación puntual de aminoácidos. En un aspecto, se genera, en lo que se denomina "mutagénesis por saturación en el sitio de aminoácidos", un tal polipéptido mutante para cada una de las 19 sustituciones de alfa-aminoácidos formadores de polipéptidos codificados de manera natural en todas y cada una de las posiciones de aminoácidos a lo largo del polipéptido. Esto produce, para todas y cada una de las posiciones del aminoácido a lo largo del polipéptido progenitor, un total de 20 polipéptidos de progenie distinta incluyendo el aminoácido original, o potencialmente más de 21 polipéptidos de progenie distinta si se utilizan aminoácidos adicionales bien sea en vez de o además de los 20 aminoácidos codificados de manera natural.

40 Así, en otro aspecto, esta metodología también es utilizable para generar mutantes que contienen, además de y/o en combinación con los 20 alfa-aminoácidos formadores de polipéptidos codificados de manera natural, otros aminoácidos y derivados de aminoácidos raros y/o no codificados de manera natural. En aún otro aspecto, esta metodología también es utilizable para generar mutantes mediante el uso de, además de y/o en combinación con sistemas de reconocimiento de codones naturales o inalterados de anfitriones adecuados, de sistemas de reconocimiento de codones alterados, mutagenizados y/o diseñados (tales como en una célula anfitriona con una o más moléculas de ARNt alteradas).

45 En aún otro aspecto, esta invención se relaciona con recombinación y más específicamente con un método para preparar polinucleótidos que codifican un polipéptido mediante un método de reordenamiento *in vivo* de secuencias de polinucleótidos que contienen regiones de homología parcial, ensamblando los polinucleótidos para formar al menos un polinucleótido y seleccionando los polinucleótidos para la producción de polipéptidos que tienen una propiedad útil.

50 En aún otro aspecto, esta invención es utilizable para analizar y catalogar, con respecto a cualquier propiedad molecular (por ejemplo, una actividad enzimática) o combinación de propiedades permitidas por la tecnología actual, los defectos de cualquier cambio mutacional logrado (incluyendo particularmente mutagénesis por saturación). Así, se provee un método amplio para determinar el efecto de cambiar cada aminoácido en un polipéptido progenitor en cada una de las al menos 19 posibles sustituciones. Esto permite que cada aminoácido en un polipéptido progenitor sea caracterizado y catalogado de acuerdo con su espectro de efectos potenciales sobre una propiedad medible del polipéptido.

55

- En un aspecto, un intrón puede ser introducido en una molécula de progenie quimérica mediante un bloque de construcción de ácidos nucleicos. Los intrones tienen frecuentemente secuencias consenso en ambos terminales con el fin de hacerlos operacionales. Además de permitir el acoplamiento de genes, los intrones pueden servir para cualquier propósito adicional proveyendo sitios de homología en otros ácidos nucleicos para permitir la recombinación homóloga. Para este propósito, y potencialmente para otros, puede ser deseable algunas veces generar un bloque de construcción de ácidos nucleicos grande para introducir un intrón. Si el tamaño es sobradamente grande generando fácilmente por síntesis química directa de dos oligos de cadena individual, tal bloque de construcción de ácidos nucleicos especializado, puede ser generado también por síntesis química de más de dos oligos de cadena individual o utilizando una reacción de amplificación basada en polimerasa.
- Los extremos ligables compatibles mutuamente de los bloques de construcción de ácidos nucleicos que van a ser ensamblados se consideran como "utilizables" para este tipo de ensamble ordenado si permiten que los bloques de construcción sean acoplados en órdenes predeterminados. Así, en un aspecto, el orden de ensamble global en el cual bloques de construcción de ácidos nucleicos pueden ser acoplados está especificado por el diseño de los extremos ligables y, si se va a utilizar mediante una etapa de ensamble, entonces el orden de ensamble global en el cual los bloques de construcción de ácidos nucleicos pueden ser acoplados también está especificado por el orden secuencial de las etapas de ensamble. En un aspecto de la invención, las piezas de construcción fusionadas son tratadas con una enzima, tal como una ligasa (por ejemplo, T4 ADN ligasa) para lograr un enlace covalente de las piezas de construcción.
- El acoplamiento puede ocurrir de una manera que no haga uso de cada nucleótido en un excedente participante. El acoplamiento tiene particularmente posibilidades de sobrevivir (por ejemplo en un anfitrión transformado) si el acoplamiento es reforzado por tratamiento con una enzima ligasa para formar lo que puede denominarse como una "ligazón de brecha" o una "ligazón con brechas". Este tipo de acoplamiento puede contribuir a la generación de productos de fondo no deseados, pero también puede ser utilizado ventajosamente para incrementar la diversidad de la biblioteca de progenie generada por el reensamble por ligazón diseñada. Ciertos excedentes son capaces de experimentar autoacoplamiento para formar un acoplamiento palindrómico. Un acoplamiento es fortalecido sustancialmente si es reforzado por tratamiento con una enzima ligasa. La carencia de fosfatos en 5' en estos excedentes puede ser utilizada ventajosamente para prevenir este tipo de autoligazón palindrómica. De acuerdo con lo anterior esta invención provee que los bloques de construcción de ácidos nucleicos que puedan hacerse (o ordenarse) químicamente carezcan de un grupo fosfato 5'. Alternativamente, pueden ser retirados, por ejemplo, por tratamiento con una enzima fosfatasa, tal como fosfatasa alcalina intestinal de ternera (CIAP), con el fin de prevenir las autoligaciones palindrómicas en los procesos de reensamble por ligazón.
- En otro aspecto, el diseño de los bloques de construcción de ácidos nucleicos se obtiene por análisis de las secuencias de un conjunto de plantillas de ácidos nucleicos progenitores que sirven como base para la producción de un conjunto de progenie de moléculas de ácidos nucleicos quiméricas finalizadas. Estas plantillas de ácidos nucleicos progenitores sirven así como fuente de información de secuencia que ayudan en el diseño de bloques de construcción de ácidos nucleicos que van a ser mutagenizados, estos es, quimerizados o mezclados.
- En una ejemplificación la invención provee la quimerización de una familia de genes relacionados y su familia codificada de productos relacionados. En una ejemplificación particular, los productos codificados son enzimas. Las transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas de la presente invención pueden ser mutagenizados de acuerdo con los métodos descritos aquí.
- Así de acuerdo con un aspecto de la invención, las secuencias de una pluralidad de plantillas de ácidos nucleicos progenitores (por ejemplo, polinucleótidos de la invención) están alineadas con el fin de seleccionar uno o más puntos de demarcación que pueden estar localizados en un área de homología. Los puntos de demarcación pueden ser utilizados para delinear las fronteras de los bloques de construcción de ácidos nucleicos que van a ser generados. Así, los puntos de demarcación identificados y seleccionados en las moléculas progenitoras sirven como puntos de quimerización potenciales en el ensamble de las moléculas de progenie.
- Típicamente un punto de demarcación utilizable es un área de homología (que comprende al menos una base de nucleótidos homóloga) compartida por al menos dos plantillas progenitoras, pero el punto de demarcación puede ser un área de homología que es compartida por al menos la mitad de las plantillas progenitoras, al menos dos tercios de las plantillas progenitoras, al menos tres cuartos de las plantillas progenitoras y preferiblemente al menos la totalidad de las plantillas progenitoras. Incluso más preferiblemente todavía un punto de demarcación utilizable es un área de homología que es compartida por la totalidad de las plantillas progenitoras.
- En un aspecto, el proceso de reensamble de genes se lleva a cabo exhaustivamente con el fin de generar una biblioteca exhaustiva. En otras palabras, todas las combinaciones ordenadas posibles de los bloques de construcción de ácidos nucleicos están representadas en el conjunto de moléculas de ácidos nucleicos quiméricas finalizadas. Al mismo tiempo, el orden de ensamblaje (esto es, el orden de ensamblaje de cada bloque de construcción en la secuencia de 5' a 3' de cada ácido nucleico quimérico finalizado) en cada combinación es por

diseño (o no estocástica). Debido a la naturaleza no estocástica del método, la posibilidad de productos colaterales no deseados se reduce grandemente.

En otro aspecto, el método provee que el proceso de reensamblaje de genes se lleve a cabo sistemáticamente, por ejemplo para generar una biblioteca compartimentalizada sistemáticamente, con compartimientos que pueden ser seleccionados de manera sistemática, por ejemplo, uno por uno. En otras palabras la invención provee que, a través del uso selectivo y juicioso de bloques de construcción de ácidos nucleicos específicos, acoplados con el uso selectivo y juicioso de reacciones de ensamblaje por etapas secuenciales, puede alcanzarse un diseño experimental en donde los conjuntos específicos de productos de progenie están hechos en cada uno de varios recipientes de reacción. Esto permite llevar a cabo un procedimiento de examen y selección sistemático. Así, permite que un número potencialmente muy grande de moléculas de progenie sea examinado sistemáticamente en grupos más pequeños.

Debido a su capacidad para llevar a cabo quimerizaciones de una manera que es altamente flexible y a la vez exhaustiva y sistemática, particularmente cuando hay un bajo nivel de homología entre las moléculas progenitoras, la presente invención provee la generación de una biblioteca (o conjunto) compuesta de un gran número de moléculas de progenie. Debido a la naturaleza no estocástica del reensamble de genes de la presente invención, las moléculas de progenie generadas comprenden preferiblemente una biblioteca de moléculas de ácido nucleico quiméricas finalizadas que tienen un orden de ensamblaje global que es escogido por diseño. En un aspecto particular, tal biblioteca generada está compuesta de más de 10^3 hasta más elevado a la 10^{1000} diferentes especies moleculares de progenie.

En un aspecto, un conjunto de moléculas de ácidos nucleicos quiméricos finalizados, producidas como se describe está compuesto de una polinucleótido que codifica un polipéptido. De acuerdo con un aspecto, este polinucleótido es un gen, el cual puede ser un gen hecho por el hombre. De acuerdo con otro aspecto, este polinucleótido es una ruta genética, la cual puede ser una ruta genética hecha por el hombre. La invención provee que más de uno o más genes hechos por el hombre generados por la invención puedan ser incorporados en una ruta genética hecha por el hombre, tal como una ruta operable en un organismo eucariota (incluyendo una planta).

En otra ejemplificación, la naturaleza sintética de la etapa en la cual los bloques de construcción se generan permite el diseño e introducción de nucleótidos (por ejemplo, uno o más nucleótidos, los cuales pueden ser, por ejemplo, codones o intrones o secuencias reguladoras) que pueden ser mas tarde en una aspecto retiradas en un proceso *in vitro* (por ejemplo, por mutagénesis) o en un proceso *in vivo* (por ejemplo, utilizando la capacidad de acoplamiento de genes de un organismo anfitrión). Es evidente que en muchos casos la introducción de estos nucleótidos también puede ser deseable por muchas otras razones además del beneficio potencial de crear un punto de demarcación utilizable.

Así, de acuerdo con otro aspecto, la invención provee que un bloque de construcción de ácidos nucleicos puede ser usado para introducir un intrón. Así, la invención provee que pueden introducirse intrones funcionales en un gen hecho por el hombre de la invención. La invención también provee que pueden introducirse intrones funcionales en una ruta genética hecha por el hombre de la invención. De acuerdo con lo anterior, la invención provee la generación de un polinucleótido quimérico que es un gen hecho por el hombre que contiene uno (o más) intrones introducidos artificialmente.

De acuerdo con lo anterior, la invención también provee la generación de un polinucleótido quimérico que es una ruta genética hecha por el hombre que contiene uno (o más) intrones introducidos artificialmente. Preferiblemente, los intrones introducidos artificialmente son funcionales en una o más células anfitrionas para el acoplamiento de genes principalmente en la manera en que los intrones de origen natural sirven funcionalmente en el acoplamiento de genes. La invención provee un proceso para producir polinucleótidos que contienen intrones hechos por el hombre para ser introducidos en organismos anfitriones para recombinación y/ o acoplamiento.

Un gen producido por el hombre que utiliza la invención también sirve como sustrato para la recombinación con otro ácido nucleico. De la misma manera, una ruta genética hecha por el hombre, producida usando la invención también puede servir como sustrato para recombinación con otro ácido nucleico. En un aspecto, la recombinación es facilitada por, u ocurre en, áreas de homología entre el gen que contiene el intrón, hecho por el hombre y un ácido nucleico, el cual sirve como un asociado de recombinación. En un aspecto, el asociado de recombinación puede ser también un ácido nucleico generado por la invención, incluyendo un gen hecho por el hombre o una ruta genética hecha por el hombre. La recombinación puede ser facilitada por o puede ocurrir en áreas de homología que existen en el uno (o más) intrón introducido artificialmente en el gen hecho por el hombre.

El método de reensamble genético de la invención utiliza una pluralidad de bloques de construcción de ácidos nucleicos, cada uno de los cuales tiene preferiblemente dos extremos ligables. Los dos extremos ligables de cada bloque de construcción de ácidos nucleicos pueden ser dos extremos romos (esto es, teniendo cada uno un excedente de cero nucleótidos), o preferiblemente un extremo romo y uno con excedente, o más preferiblemente todavía dos excedentes.

Un excedente útil para este propósito puede ser un excedente en 3' o un excedente en 5'. Así, un bloque de construcción de ácidos nucleicos puede tener un excedente en 3' o alternativamente un excedente en 5' o alternativamente dos excedentes en 3' o alternativamente dos excedentes en 5'. El orden global en el cual los bloques de construcción de ácidos nucleicos son ensamblados para formar una molécula de ácido nucleico quimérica finalizada está determinado por el diseño experimental propuesto y no es aleatorio.

En un aspecto, un bloque de construcción de ácidos nucleicos es generado por síntesis química de dos ácidos nucleicos de cadena sencilla (también denominados como oligos de cadena sencilla) y poniéndolos en contacto de tal manera que se les permita fusionarse para formar un bloque de construcción de ácidos nucleicos de cadena doble.

Un bloque de construcción de ácidos nucleicos de cadena doble puede ser de tamaño variable. Los tamaños de estos bloques de construcción pueden ser pequeños o grandes. Tamaños de ejemplo para bloques de construcción varían desde 1 par de bases (no incluyendo excedentes) hasta 100.000 pares de bases (no incluyendo excedentes). También se proveen otros rangos de tamaños de ejemplo, los cuales tienen límites inferiores desde 1 bp hasta 10.000 bp (incluyendo cada valor entero entre ellos) y límites superiores desde 2 bp hasta 100.000 bp (incluyendo cada valor entero entre ellos).

Existen muchos métodos mediante los cuales puede generarse un bloque de construcción de ácidos nucleicos de cadena doble que es utilizable para la invención; y estos son conocidos en la técnica y pueden ser ejecutados fácilmente por la persona experimentada.

De acuerdo con un aspecto, se genera un bloque de construcción de ácidos nucleicos de cadena doble generando primero dos ácidos nucleicos de cadena sencilla y permitiéndoles fusionarse para formar un bloque de construcción de ácidos nucleicos de cadena doble. La dos cadenas de un bloque de construcción de ácidos nucleicos de cadena doble puede ser complementaria en cada nucleótido aparte de cualquiera que forme un excedente; así sin contener errores, aparte de cualquier excedente. De acuerdo con otro aspecto, las dos cadenas de un bloque de construcción de ácidos nucleicos de cadena doble son complementarias en poco menos que cualquier nucleótido aparte de cualquiera que forma un excedente. Así, de acuerdo con este aspecto, un bloque de construcción de ácidos nucleicos de cadena doble puede ser utilizado para introducir degeneración en codones. La degeneración en codones puede ser introducida utilizando la mutagénesis por saturación en el sitio descrita aquí, utilizando uno o más casetes N,N,G/T o alternativamente utilizando uno o más casetes N,N,N.

El método de recombinación *in vivo* de la invención puede ser llevado de manera ciega en un conjunto de híbridos o alelos desconocidos de un polinucleótido o secuencia específicos. Sin embargo, no es necesario conocer la secuencia de ADN o ARN real del polinucleótido específico.

La metodología de utilizar la recombinación dentro de una población mixta de genes puede ser útil para la generación de cualquier proteína útil, por ejemplo, interleucina I, anticuerpos, tPA y hormona de crecimiento. Esta metodología puede ser utilizada para generar proteínas que tienen especificidad o actividad alterada. La metodología también puede ser útil para generación de secuencias de ácidos nucleicos híbridas, por ejemplo, regiones promotoras, intrones, exones, secuencias potenciadoras, regiones 3' no traducidas o regiones 5' no traducidas de genes. Así esta metodología puede ser utilizada para generar genes que tienen ratas incrementadas de expresión. Esta metodología también puede ser útil en el estudio de secuencias de ADN repetitivas. Finalmente, esta metodología puede ser útil para hacer mutar ribozimas o aptámeros.

En un aspecto la invención descrita aquí está dirigida al uso de ciclos repetidos de reordenamiento, recombinación y selección reductores, lo que permite la evolución molecular dirigida de secuencias lineales altamente complejas, tales como ADN, ARN o proteínas a través de recombinación.

Sistema de evolución dirigida optimizado

La invención provee un sistema de modificación de genes no estocástico denominado "sistema de evolución dirigida optimizado" para generar polipéptidos, por ejemplo, transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas, o anticuerpos de la invención, con propiedades nuevas o alteradas. La evolución dirigida optimizada es dirigida al uso de ciclos repetidos de reordenamiento, recombinación y selección reductores que permiten la evolución molecular dirigida de ácidos nucleicos a través de recombinación. La evolución dirigida optimizada permite la generación de una gran población de secuencias quiméricas evolucionadas, en donde la población generada es enriquecida significativamente en secuencias que tienen un número predeterminado de eventos de intercambio.

Un evento de intercambio es un punto en una secuencia quimérica, en donde un desplazamiento en la secuencia ocurre desde una variante progenitora hasta otra variante progenitora. Tal punto está normalmente en la unión en donde los oligonucleótidos de dos progenitores se ligan entre sí para formar una secuencia sencilla. Este método permite el cálculo de las concentraciones correctas de secuencias de oligonucleótidos de tal manera que la

población quimérica final de la secuencia es enriquecida en el número escogido de eventos de intercambio. Esto provee más control sobre la selección de las variantes quiméricas que tienen un número predeterminado de eventos de intercambio.

Además, este método provee un medio conveniente para explorar una cantidad tremenda del posible espacio variante de proteína en comparación con otros sistemas. Previamente, si se generaban, por ejemplo, 10^{13} moléculas quiméricas durante una reacción, sería extremadamente difícil probar tan alto número de variantes quiméricas para una actividad particular. Además, una porción significativa de la población de progenie tendría un número muy alto de eventos de intercambio lo que daría como resultado proteínas que iban a tener probablemente menos niveles incrementados de una actividad particular. Utilizando estos métodos, la población de moléculas quiméricas puede ser enriquecida para aquellas variantes que tengan un número particular de eventos de intercambio. Así, aunque solamente se pueden generar todavía 10^{13} moléculas quiméricas durante una reacción, cada una de las moléculas escogida para análisis posterior tiene lo más probablemente, por ejemplo, solamente tres eventos de intercambio. Debido a que la población de progenie resultante puede ser sesgada para tener un número predeterminado de eventos de intercambio, las fronteras de la variedad funcional entre las moléculas quiméricas es reducida. Esto provee un número más manejable de variables cuando se calcula cuáles de los oligonucleótidos de los polinucleótidos progenitores originales podría ser responsable por afectar una característica particular.

Un método para crear una secuencia de polinucleótidos de progenie quimérica es crear oligonucleótidos correspondientes a fragmentos o porciones de cada secuencia progenitora. Cada oligonucleótido incluye preferiblemente una región única de superposición de tal manera que al mezclar los oligonucleótidos entre sí da como resultado una nueva variante que tiene cada fragmento de oligonucleótido ensamblado en el orden correcto. Puede encontrarse también información adicional, por ejemplo, en USSN 09/332.835; patente de los Estados Unidos No 6,361,974.

El número de oligonucleótidos generados por cada variante progenitora lleva una relación al número total de intercambios resultantes en la molécula quimérica que se crea finalmente. Por ejemplo, tres variantes de secuencia de nucleótidos progenitores puede ser provista para experimentar una reacción de ligazón con el fin de encontrar una variante quimérica que tiene, por ejemplo, una actividad mayor a alta temperatura. Como ejemplo, puede generarse un conjunto de 50 secuencias de oligonucleótidos correspondientes a cada porción de cada variante progenitora. De acuerdo con lo anterior, durante el proceso de reensamblaje por ligazón, podría haber hasta 50 eventos de intercambio dentro de cada una de las secuencias quiméricas. La probabilidad de que cada uno de los polinucleótidos quiméricos generados contenga oligonucleótidos de cada variante progenitora en orden alternante es muy baja. Si cada fragmento de oligonucleótido está presente en la reacción de ligazón en la misma cantidad molar es probable que en algunas posiciones los oligonucleótidos del mismo polinucleótido progenitor se ligan a otro y así no haya un evento de intercambio. Si la concentración de cada oligonucleótido de cada progenitor se mantiene constante durante cualquier etapa de ligazón en este ejemplo, hay una posibilidad de $1/3$ (asumiendo 3 progenitores) de que un oligonucleótido de la misma variante progenitora se ligue dentro de la secuencia quimérica y no produzca intercambio.

De acuerdo con lo anterior, puede determinarse una función de densidad de probabilidad (PDF) para predecir la población de eventos de intercambio que probablemente ocurran durante cada etapa en una reacción de ligazón dado un número de conjuntos de variantes progenitoras, un número de oligonucleótidos correspondientes a cada variante, y las concentraciones de cada variante durante cada etapa en la reacción de ligazón. La estadística y matemática detrás de la determinación de PDF se describe a continuación. Utilizando estos métodos, se puede calcular tal función de densidad de probabilidad, y enriquecer así la población de progenie quimérica para un número predeterminado de eventos de intercambio resultante de una reacción de ligazón particular. Además, un número objetivo de eventos de intercambio puede ser predeterminado, y el sistema puede ser programado entonces para calcular las cantidades de partida de cada oligonucleótido progenitor durante cada etapa de la reacción de ligazón para dar como resultado una función de densidad de probabilidad que se centre en el número predeterminado de eventos de intercambio. Estos métodos están dirigidos al uso de ciclos repetidos de reordenamiento, recombinación y selección reductores que permiten la evolución molecular dirigida de un ácido nucleico que codifica un polipéptido a través de recombinación. Este sistema permite la generación de una gran población de secuencias quiméricas evolucionadas, en donde la población generada es significativamente enriquecida para secuencias que tienen un número predeterminado de eventos de intercambio. Un evento de intercambio es un punto en una secuencia quimérica en donde ocurre un desplazamiento en la secuencia de una variante progenitora a otra variante progenitora. Tal punto está normalmente en la unión en donde los oligonucleótidos de dos progenitores se ligan entre sí para formar una secuencia sencilla. El método permite el cálculo de las concentraciones correctas de las secuencias de oligonucleótidos de tal manera que la población quimérica final de secuencias es enriquecida en el número escogido de eventos de intercambio. Esto provee más control en la selección de variantes quiméricas que tienen un número predeterminado de eventos de intercambio.

Además, estos métodos proveen un medio conveniente para explorar una cantidad tremenda de los posibles espacios variantes en proteínas en comparación con otros sistemas. Utilizando los métodos descritos aquí, la población de moléculas quiméricas puede ser enriquecida para aquellas variantes que tengan un número particular

de eventos de intercambio. Así, aunque se puedan generar todavía 10^{13} moléculas quiméricas durante una reacción, cada una de las moléculas escogidas para análisis posterior lo más probablemente tiene, por ejemplo, sólo tres eventos de intercambio. Debido a que la población de progenie resultante puede ser sesgada para tener un número predeterminado de eventos de intercambio, las fronteras de la variedad funcional entre las moléculas quiméricas se reduce. Esto provee un número más manejable de variables cuando se calcula cuáles oligonucleótidos de los polinucleótidos progenitores originales puede ser responsable por afectar una característica particular.

En un aspecto, el método crea una secuencia de polinucleótidos de progenie quimérica creando oligonucleótidos correspondientes a fragmentos o porciones de cada secuencia progenitora. Cada oligonucleótido incluye preferiblemente una región única de superposición de tal manera que al mezclar los oligonucleótidos entre sí da como resultado una nueva variante que tiene cada fragmento de oligonucleótido ensamblado en el orden correcto. Véase también USSN 09/332,835.

Determinación de eventos de intercambio

Aspectos de la invención incluyen un sistema y software que recibe una función de densidad de probabilidad de intercambio deseada (PDF), el número de genes progenitores que van a ser reensamblados, y el número de fragmentos en el reensamblaje como entradas. El resultado de este programa es un "PDF de fragmentos" que puede ser utilizado para determinar una receta para producir genes reensamblados, y el PDF de intercambio estimado para estos genes. El procesamiento descrito aquí se lleva a cabo preferiblemente en un lenguaje de programación MATLAB™ (The Mathworks, Natick, Massachusetts) y un ambiente de desarrollo para cómputo técnico.

Proceso iterativo

En la práctica de esta invención, estos procesos pueden ser repetidos de manera iterativa. Por ejemplo, un ácido nucleico (o, el ácido nucleico) responsable para un fenotipo alterado o nuevo de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa es identificado, reaislado, modificado de nuevo, probado de nuevo para su actividad. Este proceso puede ser repetido iterativamente hasta que se diseñe un fenotipo deseado. Por ejemplo, una ruta anabólica o catabólica bioquímica completa puede ser diseñada en una célula, incluyendo, por ejemplo, actividad de transferasa por ejemplo, transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa.

De la misma manera, si se determina que un oligonucleótido particular no tiene efecto total sobre la característica deseada (por ejemplo, un nuevo fenotipo de transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa), puede ser retirado como una variable sintetizando oligonucleótidos progenitores más grandes que incluyen la secuencia que va a ser removida. Puesto que la incorporación de la secuencia dentro de una secuencia más grande evita cualquier evento de intercambio, no habrá más variación alguna de esta secuencia en los polinucleótidos de progenie. Esta práctica iterativa de determinar cuáles oligonucleótidos están más relacionados con la característica deseada, y cuáles están no relacionados, permite una exploración más eficiente de las variantes proteínicas posibles que podrían proveer una característica o actividad particular.

Mezcla *in vivo*

La mezcla *in vivo* de moléculas es usada en métodos de la invención que proveen variantes de polipéptidos de la invención, por ejemplo, anticuerpos, transferasas por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas, y similares. La mezcla *in vivo* puede ser llevada a cabo utilizando la propiedad natural de las células de recombinar multímeros. Mientras que la recombinación *in vivo* ha provisto la principal ruta natural para la diversidad molecular, la recombinación genética sigue siendo un proceso relativamente complejo que involucra 1) el reconocimiento de homologías; 2) escisión de cadenas, invasión de cadenas, y etapas metabólicas que llevan a la producción de quiasma recombinante; y finalmente 3) la resolución de quiasma en moléculas recombinadas discretas. La formación del quiasma requiere el reconocimiento de secuencias homólogas. En otro aspecto, la invención incluye un método para producir un polinucleótido híbrido a partir de al menos un primer polinucleótido y un segundo polinucleótido. La invención puede ser utilizada para producir un polinucleótido híbrido introduciendo al menos un primer polinucleótido y un segundo polinucleótido que comparten al menos una región de homología de secuencia parcial en una célula anfitriona adecuada. Las regiones de homología de secuencia parcial promueven procesos que dan como resultado la reorganización de secuencias produciendo un polinucleótido híbrido. El término "polinucleótido híbrido", tal como se utiliza aquí, es cualquier secuencia de nucleótidos que resulta del método de la presente invención y contiene secuencias a partir de al menos dos secuencias de polinucleótidos originales. Tales polinucleótidos híbridos pueden resultar de eventos de recombinación intermoleculares que promueven la integración de secuencias entre moléculas de ADN. Además, tales polinucleótidos híbridos pueden resultar de procesos de reordenamiento reductor intramolecular que utilizan secuencias repetidas para alterar una secuencia de nucleótidos dentro de una molécula de ADN.

El reordenamiento *in vivo* está enfocado en procesos “intermoleculares” denominados colectivamente como “recombinación” los cuales en bacterias, se ven generalmente como un fenómeno “dependiente de RecA”. La invención puede basarse en procesos de recombinación de una célula anfitriona para recombinar y reordenar secuencias, o la capacidad de las células para mediar en procesos reductores para disminuir la complejidad de la secuencias cuasirrepetidas en las células por eliminación. Este proceso de “reordenamiento reductor” ocurre mediante un proceso “intramolecular” independiente de RecA.

Por lo tanto, en otro aspecto de la invención, pueden generarse polinucleótidos novedosos mediante el proceso de reordenamiento reductor. El método involucra la generación de constructos que contienen secuencias consecutivas (secuencias codificadoras originales) su inserción en un vector apropiado y su introducción subsecuente en una célula anfitriona apropiada. El reordenamiento de las identidades moleculares individuales ocurre por procesos combinatoriales entre las secuencias consecutivas en las regiones de posesión del constructo de homología, o entre unidades cuasirrepetidas. El proceso de reordenamiento recombina y/o reduce la complejidad y grado de las secuencias repetidas y da como resultado la producción de especies moleculares novedosas. Pueden aplicarse diversos tratamientos para potenciar la rata de reordenamiento. Estos podrían incluir tratamiento con luz ultravioleta, o agentes químicos que dañan el DNA y/o el uso de líneas celulares anfitrionas que despliegan niveles potenciados de “inestabilidad genética”. Así el proceso de reordenamiento puede involucrar recombinación homóloga o la propiedad natural de secuencias cuasirrepetidas para dirigir su propia evolución.

Las secuencias repetidas o “cuasirrepetidas” juegan un papel en la inestabilidad genética. En la presente invención, “cuasirrepeticiones” son repeticiones que no están restringidas a la estructura unitaria original. Las unidades cuasirrepetidas pueden ser presentadas como un arreglo de secuencias en un constructo; unidades consecutivas de secuencias similares. Una vez ligadas, las uniones entre las secuencias consecutivas se hacen esencialmente invisibles y la naturaleza cuasirrepetitiva del constructo resultante es ahora continua al nivel molecular. El proceso de eliminación que la célula lleva a cabo para reducir la complejidad de los constructos resultantes opera entre las secuencias cuasirrepetidas. Las unidades cuasirrepetidas proveen un repertorio prácticamente ilimitado de plantillas sobre las cuales pueden ocurrir eventos de disminución. Los constructos que contienen las cuasirrepeticiones proveen así efectivamente elasticidad molecular suficiente que los eventos de eliminación (y potencialmente de inserción) puedan ocurrir virtualmente en cualquier lugar dentro de las unidades cuasirrepetitivas.

Cuando las secuencias cuasirrepetidas son ligadas todas en la misma orientación, por ejemplo, de cabeza a cola o viceversa, la célula no puede distinguir las unidades individuales. Consecuentemente, el proceso reductor puede ocurrir a lo largo de las secuencias. En contraste, cuando por ejemplo, las unidades son presentadas cabeza a cabeza más que de cabeza a cola, la inversión delinea los puntos finales de las unidades adyacentes de tal manera que la formación de eliminación favorecerá la pérdida de unidades discretas. Así, es preferible con el presente método que las secuencias estén en la misma orientación. La orientación aleatoria de secuencias cuasirrepetidas dará como resultado la pérdida de eficiencia en el reordenamiento, mientras que la orientación consistente de las secuencias ofrecerá la mayor eficiencia. Sin embargo, a la vez que tener pocas de las secuencias contiguas en la misma orientación hace disminuir la eficiencia, puede todavía proveer suficiente elasticidad para la recuperación efectiva de las moléculas novedosas. Los constructos pueden hacerse con las secuencias cuasirrepetidas en la misma orientación para permitir una eficiencia mayor.

Las secuencias pueden ser ensambladas en una orientación de cabeza a cola utilizando cualquiera de una variedad de métodos, incluyendo los siguientes:

a) Podrían utilizarse cebadores que incluyen una cabeza poli-A y una cola poli-T los cuales cuando se hacen de cadena sencilla proveerían orientación. Esto se logra teniendo las primeras pocas bases de los cebadores hechas a partir de ARN y por lo tanto el ARNseH es fácilmente removido.

b) Pueden utilizarse cebadores que incluyen sitios de escisión de restricción únicos. Se requerirían sitios múltiples, una batería de secuencias únicas y etapas de síntesis y ligazón repetidas.

c) Las pocas bases internas del cebador podrían ser tioladas y podría utilizarse una exonucleasa para producir moléculas con colas apropiadas.

La recuperación de las secuencias reordenadas se basa en la identificación de vectores de clonación con un índice repetitivo reducido (RI). Las secuencias de codificación reordenadas pueden ser recuperadas entonces por amplificación. Los productos son reclonados y expresados. La recuperación de los vectores de clonación con RI reducido puede ser afectada por:

1) El uso de vectores mantenidos solo de manera estable cuando el constructo es reducido en complejidad.

2) La recuperación física de vectores acortados por procedimientos físicos, en este caso, el vector de clonación sería recuperado utilizando procedimientos estándar de aislamientos de plásmidos y fraccionados por tamaño bien sea sobre gel de agarosa, o en columna con un corte de peso molecular bajo utilizando procedimientos estándar.

3) La recuperación de vectores que contienen genes interrumpidos los cuales pueden ser seleccionados cuando disminuye el tamaño del inserto.

4) El uso de técnicas de selección directa con un vector de expresión y la selección apropiada.

5 Las secuencias codificadoras (por ejemplo, genes) a partir de organismos relacionados puede demostrar un alto grado de homología y codificar productos proteínicos bastante diversos. Estos tipos de secuencias son particularmente útiles en la invención presente como cuasirrepeticiones. Sin embargo, mientras que los ejemplos ilustrados más adelante demuestran el reordenamiento de secuencias de codificación originales cercanamente idénticas (cuasirrepeticiones), este proceso no está limitado a tales repeticiones cercanamente idénticas.

10 El siguiente ejemplo demuestra un método de la invención. Se describen secuencias de ácidos nucleicos codificadores (cuasirrepeticiones) derivadas de tres (3) únicas. Cada secuencia codifica una proteína con un conjunto distinto de propiedades. Cada una de las secuencias difiere por un par de bases único o unos pocos en una posición única en la secuencia. Las secuencias cuasirrepetidas son amplificadas separada o colectivamente y ligadas en ensamblajes aleatorios de tal manera que todas las posibles permutaciones y combinaciones están disponibles en la población de moléculas ligadas, el número de unidades cuasirrepetidas puede ser controlado mediante las condiciones de ensamblaje. El número promedio de unidades cuasirrepetidas en un constructo está definido como el índice repetitivo (RI).

20 Una vez formados, los constructos pueden, o pueden no ser fraccionados por tamaños sobre un gel de agarosa de acuerdo con protocolos publicados, insertados en un vector de clonación y transfectados en una célula anfitriona apropiada. Las células son propagadas entonces y se efectúa el "reordenamiento reductor". La rata del proceso de reordenamiento reductor puede ser estimulada por la introducción de un daño de ADN si se desea. Si la reducción en RI es mediada por la formación de eliminación entre secuencias repetidas mediante un mecanismo "intra molecular", o mediada por eventos similares a recombinación a través de mecanismos "inter moleculares" no tiene importancia. El resultado final es un reordenamiento de las moléculas en todas las posibles combinaciones.

25 En un aspecto (opcionalmente), el método comprende la etapa adicional de seleccionar los mismos de la biblioteca del conjunto mezclado para identificar miembros de la biblioteca mezclados individuales que tiene la capacidad de enlazarse o de alguna otra manera interactuar, o catalizar una reacción particular (por ejemplo, tal como un dominio catalítico de una enzima) con una molécula predeterminada, tal como por ejemplo un receptor proteína, un oligosacárido, un virión u otro compuesto o estructura predeterminados.

30 Los polipéptidos que son identificados a partir de tales bibliotecas pueden ser utilizados para propósitos terapéuticos, diagnósticos, de investigación o relacionados (por ejemplo, catalizadores, solutos para osmolaridad creciente de una solución acuosa y similares) y/o pueden ser sometidos a uno o más ciclos adicionales de mezcla y/o selección.

35 En otro aspecto se prevé que antes de o durante la recombinación o reordenamiento, los polinucleótidos generados por el método de la invención pueden ser sometidos a agentes o procesos que promueven la introducción de mutaciones en los polinucleótidos originales. La introducción de tales mutaciones incrementaría la diversidad de los polinucleótidos y polipéptidos híbridos resultantes codificados a partir de las mismas. Los agentes o procesos que promueven la mutagénesis pueden incluir, pero no limitarse a: (+)-CC-1065, o un análogo sintético tal como (+)-CC-1065-(N3-Adenina) (Véase Sun and Hurley, (1992); un aducto 4'-fluoro-4-aminobifenilo N-acetilido o desacetilado capaz de inhibir la síntesis de ADN (véase, por ejemplo, van de Poll *et al.* (1992)); o un aducto de 4-aminofenilo N-acetilado o desacetilado capaz de inhibir la síntesis de ADN (véase también, van de Poll *et al.* (1992), pp. 751-758); cromo trivalente, una sal de cromo trivalente, un hidrocarburo aromático policíclico, como aducto de ADN (PAH) capaz de inhibir la replicación de ADN, tal como 7-bromometilbenz[a]antraceno ("BMA"), tris(2,3-dibromopropil)fosfato ("Tris-BP"), 1,2-dibromo-3-cloropropano ("DBCP"), 2-bromoacroleína (2BA), benzo[a]piren-7,8-dihidrodiol-9-10-epóxido ("BPDE"), una sal de halógeno de platino(II), N-hidroxi-2-amino-3-metilimidazo[4,5-f]-quinolina ("N-hidroxi-IQ") y N-hidroxi-2-amino-1-metil-6-fenilimidazo[4,5-f]-piridina ("N-hidroxi-PhIP"). Medios de ejemplo para hacer más lenta o detener la amplificación por PCR consiste de luz ultravioleta (+)-CC-1065 y (+)-CC-1065-(N3-Adenina). Medios particularmente abarcados son aductos de ADN o polinucleótidos que comprenden los aductos de ADN de los polinucleótidos o conjunto de polinucleótidos, que pueden ser liberados o removidos mediante un proceso que incluye calentar la solución que comprende los polinucleótidos antes del procesamiento posterior.

50 En otro aspecto la invención está dirigida a un método para producir proteínas recombinantes que tienen actividad biológica tratando una muestra que comprende polinucleótidos de plantilla de cadena doble que codifican una proteína tipo silvestre bajo condiciones de acuerdo con la invención que provee la producción de polinucleótidos híbridos o reordenado.

55 Producción de variantes de secuencia

La invención también provee métodos adicionales para hacer variantes de secuencias de ácidos nucleicos d-aminoácido transferasas de la invención. La invención también provee métodos adicionales para aislar transferasas, por ejemplo por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas usando los ácidos nucleicos y polipéptidos de la invención. En un aspecto la invención provee variantes de una d-aminoácido transferasa (por ejemplo, un gen, un ADNc o mensajero) de la invención, los cuales pueden ser alterados por cualquier medio incluyendo, por ejemplo, métodos aleatorios o estocásticos, o métodos no estocásticos o "de evolución dirigida", como se describió más arriba.

Las variantes aisladas pueden ser de origen natural. Las variantes también pueden ser creadas *in vitro*. Las variantes pueden ser creadas utilizando técnicas de ingeniería genética tales como la mutagénesis dirigida al sitio, mutagénesis química aleatoria, procedimientos de eliminación de Exonucleasa III, y técnicas de clonación estándar. Alternativamente, tales variantes, fragmentos, análogos o derivados pueden ser creados utilizando procedimientos de síntesis o modificación química. Otros métodos para hacer variantes también son familiares para los experimentados en la técnica. Estos incluyen procedimientos en los cuales las secuencias de ácidos nucleicos obtenidas a partir de aislados naturales, son modificadas para generar nuevos ácidos nucleicos que codifican polipéptidos que tiene características que potencian su valor en otras aplicaciones y procesos industriales, médicos, de laboratorio (investigación), farmacéuticos, de procesamiento de alimentos y piensos y de suplementos para alimentos y piensos y otros. En tales procedimientos, un gran número de secuencias variantes que tienen una o más diferencia en nucleótidos con respecto a la secuencia obtenida en el aislado natural es generado y caracterizado. Estas diferencias en nucleótidos pueden dar como resultado cambios en aminoácidos con respecto a los polipéptidos codificados por los ácidos nucleicos a partir de los aislados naturales.

Por ejemplo, pueden crearse variantes utilizando PCR propenso al error. En la PCR propensa al error, la PCR se lleva a cabo bajo condiciones en donde la fidelidad de copiado de la ADN polimerasa es baja, de tal manera que se obtiene una alta rata de mutaciones puntuales a lo largo de la longitud completa del producto de PCR. La PCR propensa al error está descrita, por ejemplo, en Leung, DW, et al, Technique, 1:11-15., 1989) y Caldwell, RC & Joyce GF, métodos de PCR Aplic., 2:28-33, 1992. En resumen, en tales procedimientos, los ácidos nucleicos que van a ser mutagenizados son mezclados con cebadores de PCR, reguladores de reacción, $MgCl_2$, $MnCl_2$, Taq polimerasa y una concentración apropiada de dNTPs para alcanzar una alta rata de mutación puntual a lo largo de la longitud completa del producto de PCR. Por ejemplo, la reacción puede ser llevada a cabo utilizando 20 fmoles de ácido nucleico para hacer mutagenizado, 30 pmoles de cada cebador de PCR, un regulador de reacción que comprende KCl 50 mM, Tris HCl 10 mM (pH 8.3) y 0.01% de gelatina, $MgCl_2$ 7 mM, $MnCl_2$ 0.5 mM, 5 unidades de Taq polimerasa, dGTP 0.2 mM, dATP 0.2 mM, dCTP 1 mM, y dTTP 1 mM. La PCR puede ser llevada a cabo durante 30 ciclos de 94°C durante 1 minuto, 45°C durante 1 minuto, y 72°C durante 1 minuto. Sin embargo, será evidente que estos parámetros pueden ser variados según sea apropiado. Los ácidos nucleicos mutagenizados son clonados en un vector apropiado y se avalúan las actividades de los polipéptidos codificados por los ácidos nucleicos mutagenizados.

Pueden crearse también variantes utilizando mutagénesis dirigida al oligonucleótido para generar mutaciones específicas en el sitio y cualquier ADN clonado de interés. La mutagénesis de oligonucleótidos esta descrita, por ejemplo, en Reidhaar-Olson (1988) Science 241:53-57. En resumen, en tales procedimientos una pluralidad de oligonucleótidos de cadena doble que portan una o más mutaciones para ser introducidas en el ADN clonado son sintetizados e insertados en el ADN clonado que va a ser mutagenizado. Los clones que contienen el ADN mutagenizado son recuperados y las actividades de los polipéptidos que codifican son establecidas.

Otro método para generar variantes es el PCR de ensamblaje. El PCR de ensamblaje involucra el ensamblaje de un producto de PCR a partir de una mezcla de fragmentos pequeños de ADN. Un gran número de diferentes reacciones por PCR ocurre en paralelo en el mismo vial, primando los productos de una reacción sobre los de otra reacción. El PCR de ensamblaje está descrito, por ejemplo, en la patente de los Estados Unidos No. 5,965,408.

Todavía otro método para generar variantes es la mutagénesis PCR sexual. En la mutagénesis por PCR sexual, ocurre una recombinación homóloga forzada entre moléculas de ADN de diferente pero altamente relacionadas secuencia de ADN *in vitro*, como resultado de una fragmentación aleatoria de la molécula de ADN con base en homología de secuencia, seguida por la fijación del intercambio por la extensión del cebador en una reacción de PCR. La mutagénesis por PCR sexual está descrita, por ejemplo, en Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747-10751. En resumen, en tales procedimientos una pluralidad de ácidos nucleicos que van a ser recombinados son digeridos con DNasa para generar fragmentos que tienen un tamaño promedio de 50-200 nucleótidos. Los fragmentos del tamaño promedio del cebador son purificados y resuspendidos en una mezcla para PCR. Se lleva a cabo la PCR bajo condiciones que facilitan la recombinación entre los fragmentos de ácidos nucleicos. La PCR puede ser ejecutada resuspendiendo los fragmentos purificados en una concentración de 10-30 ng/μl en una solución de 0.2 mM de cada uno de dNTP, $MgCl_2$ 2 mM, KCl 50mM, Tris HCl 10mM, pH 9.0, y 0.1% de Triton X-100. Se agregan 2.5 unidades de Taq polimerasa por 100:1 de mezcla de reacción y se lleva a cabo el PCR utilizando el siguiente régimen: 94°C durante 60 segundos, 94°C durante 30 segundos, 50-55°C durante 30 segundos, 72°C durante 30 segundos (30-45 veces) y 72°C durante 5 minutos. Sin embargo, será evidente que

estos parámetros pueden variarse según sea apropiado. En algunos aspectos, los oligonucleótidos pueden ser incluidos en las reacciones de PCR. En otros aspectos, el fragmento Klenow de ADN polimerasa 1 puede ser utilizado en un primer conjunto de reacciones de PCR y la Taq polimerasa puede ser utilizada en un subsecuente conjunto de reacciones de PCR. Las secuencias recombinantes son aisladas y se establecen las actividades de los polipéptidos que ellas codifican.

Las variantes también pueden ser creadas por mutagénesis *in vivo*. En algunos aspectos, las mutaciones aleatorias en una secuencia de interés son generadas propagando la secuencia de interés en una cepa bacteriana, tal como una cepa de *E. coli*, la cual porta las mutaciones en una o más de las rutas de reparación de ADN. Tales cepas "mutantes" tienen una rata de mutación aleatoria más alta que la de un progenitor tipo silvestre. La propagación del ADN en una de estas cepas generará eventualmente mutaciones aleatorias dentro del ADN. Las cepas mutadoras adecuadas para uno en la mutagénesis *in vivo* están descritas en la publicación PCT No. WO 91/16427, publicada el 31 de Octubre de 1991, titulada "Métodos para la Creación de Genotipos a partir de Poblaciones de Genes Múltiples".

Las variantes también pueden ser generadas utilizando mutagénesis en casete. En las mutagénesis en casete una pequeña región de la molécula de ADN de cadena doble es remplazada con un "casete" de oligonucleótido sintético que difiere de la secuencia nativa. El oligonucleótido contiene frecuentemente secuencias nativas completas y/o parcialmente aleatorizadas.

La mutagénesis de ensamble recursivo puede ser utilizada también para generar variantes. La mutagénesis de ensamble recursivo es un algoritmo para la ingeniería de proteínas (mutagénesis de proteínas) desarrollado para producir diversas poblaciones de mutantes fenotípicamente relacionados cuyos miembros difieren en secuencias de aminoácidos. Este método utiliza un mecanismo de retroalimentación para controlar rondas sucesivas de mutagénesis de casetes combinacionales. La mutagénesis de ensamble recursivo está descrita en Arkin, A.P. y Youvan, D.C., PNAS, USA, 89:7811-7815, 1992.

En algunos aspectos, se crean variantes utilizando mutagénesis de ensamble exponencial. La mutagénesis de ensamble exponencial es un proceso para generar bibliotecas combinacionales con un alto porcentaje de mutantes únicos y funcionales, en donde pequeños grupos de residuos son aleatorizados en paralelo para identificar, en cada posición alterada, aminoácidos que llevan a proteínas funcionales. La mutagénesis de ensamble exponencial está descrita en Delegrave, S. and Youvan, D.C., Biotechnology Research, 11:1548-1552, 1993. Las mutagénesis aleatoria y dirigidas al sitio están descritas en Arnold, F.H., Current Opinion in Biotechnology, 4:450-455, 1993.

En algunos aspectos, las variantes son creadas utilizando procedimientos de mezcla en donde porciones de una pluralidad de ácidos nucleicos que codifican distintos polipéptidos son funcionadas entre sí para crear secuencias de ácidos nucleicos quiméricas que codifican polipéptidos quiméricos tal como se describe en la patente de los Estados Unidos No. 5,965,408, presentada el 9 de Julio de 1996, titulada, "Method of DNA Reassembly by Interrupting Synthesis" y la patente de los Estados Unidos No. 5,939,250, presentada el 22 de Mayo de 1996 titulada como "Production of Enzymes Having Desired Activities by Mutagenesis".

Las variantes de los polipéptidos de la invención pueden ser variantes en las cuales uno o más residuos de aminoácidos de los polipéptidos de la invención están sustituidos con un residuo de aminoácidos conservado o no conservado (preferiblemente un residuo de aminoácido conservado) y tal residuo de aminoácido conservado puede o puede no ser codificado por el código genético.

Las sustituciones conservadoras son aquellas que sustituyen un aminoácido dado en un polipéptido por otro aminoácido de características similares. Típicamente vistas como sustituciones conservadoras están los siguientes remplazos de un aminoácido alifático tal como alanina, valina, leucina e isoleucina con otro aminoácido alifático; remplazo de una serina con una treonina o viceversa; remplazo de un residuo ácido tal como ácido aspártico y ácido glutámico con otro residuo ácido; remplazo de un residuo que porta un grupo amida, tal como asparagina y glutamina, con otro residuo que porta un grupo amida; es intercambio de un residuo básico tal como lisina y arginina con otro residuo básico; y remplazo de un residuo aromático tal como fenilalanina, tirosina, con otro residuo aromático.

Otras variantes son aquellas en las cuales uno o más de los residuos de aminoácidos de los polipéptidos de la invención incluyen un grupo sustituyente.

Todavía otras variantes son aquella en las cuales el polipéptido está asociado con otro compuesto, tal como un compuesto para incrementar la vida media del polipéptido (por ejemplo, polietilen glicol).

Variantes adicionales son aquella en las cuales los aminoácidos adicionales son fusionados al polipéptido, tal como una secuencia de guía, una secuencia secretora, una secuencia proproteína o una secuencia que facilita la purificación, enriquecimiento o estabilización del polipéptido.

En algunos aspectos, los fragmentos derivados análogos retienen la nueva función o actividad biológica que los polipéptidos de la invención y secuencias sustancialmente idénticas a los mismos. En otros aspectos, el fragmento, derivado o análogo incluye una proproteína, de tal forma que el fragmento, derivado o análogo pueden ser derivados por escisión de la porción proproteínica para producir un polipéptido activo.

5 Codones de optimización para alcanzar altos niveles de expresión de proteína en células anfitrionas

La invención provee métodos para modificar ácidos nucleicos que codifican d-aminoácido transferasa para modificar el uso de codones. En un aspecto, la invención provee métodos para modificar codones en un ácido nucleico que codifica una d-amino transferasa para incrementar o disminuir su expresión en una célula anfitriona. La invención también provee ácidos nucleicos que codifican una transferasa, por ejemplo, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa para incrementar su expresión en una célula anfitriona, enzimas así modificadas, y métodos para hacer las enzimas modificadas. El método comprende identificar un codón "no preferido" o "menos preferido" en ácidos nucleicos que codifican transferasa por ejemplo, transaminasa-, por ejemplo, d-aminoácido transferasa-, y/o oxidorreductasa-, por ejemplo, deshidrogenasa-, por ejemplo, d-aminoácido deshidrogenasa y remplazando uno o más de estos codones no preferidos o menos preferidos con un "codón preferido" que codifica el mismo aminoácido que el codón remplazado y al menos un codón no preferido o menos preferido en el ácido nucleico ha sido remplazado por un codón preferido que codifica el mismo aminoácido. Un codón preferido es un codón sobrerrepresentado en secuencias de codificación en genes en la célula anfitriona y un codón no preferido o menos preferido en un codón subrepresentado en secuencias en codificación en genes de la célula anfitriona.

Las células anfitrionas para expresar los ácidos nucleicos, casetes de expresión y vectores de la invención incluyen bacterias, levaduras, hongos, células vegetales, células de insectos y células de mamíferos. Así, la invención provee métodos para optimizar el uso de codones en todas estas células, ácidos nucleicos alterados por codones y polipéptidos hechos por los ácidos nucleicos alterados por codones. Células anfitrionas de ejemplo incluyen bacterias gram negativas tales como *Escherichia coli* y *Pseudomonas fluorescens*; bacterias gram positivas tales como *Lactobacillus gasserii*, *Lactococcus lactis*, *Lactococcus cremoris*, *Bacillus subtilis*. Células anfitrionas de ejemplo también incluyen organismos eucariotas, por ejemplo, diversas levaduras, tales como *Saccharomyces* sp., incluyendo *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, y *Kluyveromyces lactis*, *Hansenula polymorpha*, *Aspergillus niger* y células y líneas celulares de mamíferos y células y líneas celulares de insectos. Otra células anfitrionas de ejemplo incluyen células bacterianas, tales como *E. coli*, *Streptomyces*, *Bacillus subtilis*, *Bacillus cereus*, *Salmonella typhimurium* y diversas especies con los géneros *Pseudomonas*, *Streptomyces* y *Staphylococcus*, células fúngicas, tales como *Aspergillus*, levaduras tales como cualquier especies de *Pichia*, *Saccharomyces*, *chizosaccharomyces*, *Schwanniomyces*, incluyendo *Pichia pastoris*, *Saccharomyces cerevisiae*, o *chizosaccharomyces pombe* células de insectos tales como *Drosophila S2* y *Spodoptera Sf9*, células animales tales como CHO, COS o melanoma de Bowes y adenovirus. La selección de un anfitrión apropiado está dentro de las habilidades de las personas experimentadas en la técnica. Así, la invención también incluye ácidos nucleicos y polipéptidos optimizados para expresión en estos organismos y especies.

Por ejemplo, los codones de un ácido nucleico que codifica una d-aminoácido transferasa aislada a partir de una célula bacteriana son modificados de tal manera que el ácido nucleico es expresado de manera óptima en una célula bacteriana diferente de la bacteria de la cual fue derivada la transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, una levadura, un hongo, una célula vegetal, una célula de insecto o una célula de mamífero. Métodos para optimizar codones son bien conocidos en la técnica, véase, por ejemplo, la patente de los Estados Unidos No. 5,795,737; Baca (2000) Int. J. Parasitol. 30:113-118; Hale (1998) Protein Expr. Purif. 12:185-188; Narum (2001) Infect. Immun. 69:7250-7253. Véase también, Narum (2001) Infect. Immun. 69:7250-7253, que describe la optimización de codones en sistemas en ratones; Outchkourov (2002) Protein Expr. Purif. 24:18-24, que describe optimización de codones en levadura; Feng (2000) Biochemistry 39:15399-15409, que describe la optimización de codones en *E. coli*; Humphreys (2000) Protein Expr. Purif. 20:252-264. que describe el uso de la optimización de codones que afecta la secreción en *E. coli*.

Animales transgénicos no humanos

La invención describe animales transgénicos no humanos que comprenden un ácido nucleico, un polipéptido (por ejemplo, una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa), un casete de expresión o vector o una célula transfectada o transformada de la invención. La invención también describe métodos para hacer y utilizar estos animales transgénicos no humanos.

Los animales transgénicos no humanos pueden ser, por ejemplo, cabras, conejos, ovejas, cerdos, vacas, ratas, caballos, perros, peces y ratones, que comprenden los ácidos nucleicos de la invención. Estos animales pueden ser utilizados, por ejemplo, como modelos *in vivo* para estudiar la actividad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-

aminoácido deshidrogenasa, o como modelos para seleccionar agentes que cambian la actividad de transferasa, por ejemplo la transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa *in vivo*. Las secuencias de codificación para los polipéptidos que van a ser expresados en los animales transgénicos no humanos pueden ser diseñadas para ser constitutivas o, bajo el control de factores reguladores transcripcionales específicos para tejidos o específicos del desarrollo o inducibles. Animales transgénicos no humanos pueden ser diseñados y generados utilizando cualquier método conocido en la técnica. Véase, por ejemplo, las patentes de los Estados Unidos Nos. 6,211,428; 6,187,992; 6,156,952; 6,118,044; 6,111,166; 6,107,541; 5,959,171; 5,922,854; 5,892,070; 5,880,327; 5,891,698; 5,639,940; 5,573,933; 5,387,742; 5,087,571, que describen como hacer y utilizar células y ovarios transformados y ratones, ratas, conejos, ovejas, cerdos, pollos, cabras, peces y vacas transformados. Véase también, por ejemplo, Pollock (1999) J. Immunol. Methods 231:147-157, que describe la producción de proteínas recombinantes en la leche de animales transgénicos productores de lácteos; Baguisi (1999) Nat. Biotechnol. 17:456-461, que demuestra la producción de cabras transgénicas; La patente de los Estados Unidos No. 6,211,428, que describe como hacer y usar en mamíferos transgénicos no humanos que expresan en sus cerebros un constructo de ácido nucleico que comprende una secuencia de ADN. La patente de los Estados Unidos No. 5,387,742, que describe la inyección de secuencias de ADN recombinante o sintético clonado en ovarios de ratón fertilizados, implantando los ovarios inyectados en hembras pseudoembarazadas y cultivándolos hasta terminar en ratones transgénicos cuyas células expresan proteínas relacionadas con la patología de la enfermedad de Alzheimer. La patente de los Estados Unidos No. 6,187,992, describe hacer y usar un ratón transgénico cuyo genoma comprende una perturbación del gen que codifica la proteína precursora de amiloides (APP).

Los "Animales anulados" también pueden ser utilizados para practicar los métodos de la invención. Por ejemplo, en un aspecto, los animales transgénicos o modificados de la invención comprenden un "animal anulado", por ejemplo, un "ratón anulado," manipulado para que no exprese un gen endógeno, el cual es remplazado con un gen que expresa una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención, o, una proteína de fusión que comprende una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención.

Plantas y semillas transgénicos

La invención describe plantas y semillas transgénicos que comprenden un ácido nucleico, un polipéptido (por ejemplo, a transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa), un casete de expresión o un vector o una célula transfectada o transformada de la invención. La invención también describe productos o subproductos vegetales, por ejemplo, frutas, aceites, semillas, hojas, extractos y similares, que incluyen cualquier parte de la planta, que comprende un ácido nucleico y/o un polipéptido (por ejemplo una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa) de la invención, por ejemplo, en donde el ácido nucleico o polipéptido de la invención es heterólogo a la planta, parte de la planta, semilla etc. La planta transgénica (que incluye partes de la planta, frutas, semillas, etc.) puede ser dicotiledónea (una dicotiledónea) o monocotiledónea (una monocotiledónea). La invención también provee métodos para hacer y usar estas plantas y semillas transgénicas. La planta o célula de planta transgénica que expresa un polipéptido de la presente invención puede ser construida de acuerdo con cualquier método conocido en la técnica. Véase, por ejemplo, la Patente de los Estados Unidos No. 6,309,872.

Los ácidos nucleicos y los constructos de expresión de la invención pueden ser introducidos en cualquier célula vegetal por cualquier medio. Por ejemplo, los ácidos nucleicos o constructos de expresión pueden ser introducidos en el genoma de un anfitrión vegetal deseado, o, los ácidos nucleicos o constructos de expresión pueden ser episomas. La introducción en el genoma de una planta deseada puede ser tal que la producción de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa del huésped es regulada por elementos de control de transcripción o traducción endógenos. La invención también provee "plantas anuladas" en donde la inserción de la secuencia genética por, por ejemplo, recombinación homóloga, ha perturbado la expresión del gen endógeno. Los medios para generar plantas "anuladas" son bien conocidos en la técnica, véase, por ejemplo, Strepp (1998) Proc Natl. Acad. Sci. USA 95:4368-4373; Miao (1995) Plant J 7:359-365. Véase la discusión de plantas transgénicas más abajo.

Los ácidos nucleicos de la invención pueden ser utilizados para conferir características deseadas sobre esencialmente cualquier planta, por ejemplo, sobre plantas productoras de almidón, tales como patata, trigo, arroz, cebada y similares. Los ácidos nucleicos de la invención pueden ser utilizados para manipular rutas metabólicas de una planta con el fin de optimizar o alterar la expresión en la anfitriona de una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. Pueden cambiar la actividad de transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa en una planta. Alternativamente, una transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido

deshidrogenasa de la invención puede ser utilizada en la producción de una planta transgénica para producir un compuesto que no es producido de manera natural por esa planta. Esto puede disminuir los costes de producción o crear un producto novedoso.

En un aspecto, la primera etapa en la producción de una planta transgénica involucra hacer un constructo de expresión para expresión en una célula vegetal. Estas técnicas son bien conocidas en el arte. Pueden incluir seleccionar y clonar un promotor, una secuencia codificadora para facilitar el enlazamiento eficiente de ribosomas a ARNm y seleccionar las secuencias terminadoras del gen apropiadas. Un promotor constitutivo de ejemplo es CaMV35S, del virus mosaico de la coliflor, el cual generalmente da como resultado un alto grado de expresión en plantas. Otros promotores son más específicos y responde a pistas en el ambiente interno o externo de la planta. Un promotor de ejemplo inducible por la luz es el promotor del gen *cab*, que codifica la principal proteína de enlazamiento a/b de la clorofila.

En un aspecto, el ácido nucleico es modificado para alcanzar una expresión mayor en una célula vegetal. Por ejemplo, una secuencia de la invención probablemente tenga un porcentaje más alto de pares de nucleótido AT, en comparación con el visto en una planta, algunas de las cuales prefieren los pares de nucleótidos G-C. Por lo tanto, los nucleótidos A-T en la secuencia de codificación pueden ser sustituidos con nucleótidos G-C sin cambiar significativamente la secuencia de aminoácidos para potenciar la producción del gen del producto genético en las células vegetales.

Un gen marcador seleccionable puede ser agregado al constructo genético con el fin de identificar células o tejidos vegetales que tengan integrado el transgén exitosamente. Esto puede ser necesario porque alcanzar la incorporación y expresión de genes en células vegetales es un evento raro, que ocurre solamente en un bajo porcentaje de los tejidos o células objetivo. Los genes marcadores seleccionables codifican proteínas que proveen resistencia a agentes que normalmente son tóxicos para las plantas, tales como antibióticos o herbicidas. Solamente las células vegetales que tienen integrado el gen marcador seleccionable sobrevivirán cuando crecen sobre un medio que contiene el antibiótico o herbicida apropiados. Como sucede para otros genes insertados, los genes marcadores también requieren secuencias promotoras y de terminación para funcionar apropiadamente.

En un aspecto, hacer plantas o semillas transgénicas comprende incorporar secuencias de la invención y, en un aspecto (opcionalmente) genes marcadores en un constructo de expresión objetivo (por ejemplo, un plásmido), junto con el posicionamiento de las secuencias promotoras y terminadoras. Esto puede involucrar transferir el gen modificado hacia la planta a través de un método adecuado. Por ejemplo, un constructo puede ser introducido directamente en el ADN genómico de la planta utilizando técnicas tales como electroporación y microinyección de protoplastos de células vegetales, o los constructos pueden ser introducidos directamente en tejido vegetal utilizando métodos balísticos, tales como bombardeo con partículas de ADN. Por ejemplo, véase Christou (1997) *Plant Mol. Biol.* 35:197-203; Pawlowski (1996) *Mol. Biotechnol.* 6:17-30; Klein (1987) *Nature* 327:70-73; Takumi (1997) *Genes Genet. Syst.* 72:63-69, que discuten el uso de bombardeo de partículas para introducir transgenes en trigo; y Adam (1997) *supra*, para el uso de bombardeo de partículas para introducir YACs en células vegetales. Por ejemplo, Rinehart (1997) *supra* usó el bombardeo con partículas para generar plantas de algodón transgénicas. El aparato para acelerar partículas está descrito en las patentes de los Estados Unidos No. 5,015,580; y el instrumento de aceleración de partículas comercialmente disponible en BioRad (Biolistics) PDS-2000; véase también, John, patente de los Estados Unidos No. 5,608,148; y Ellis, patente de los Estados Unidos No. 5,681,730, que describen transformación de gimnospermas mediada por partículas.

En un aspecto, los protoplastos pueden ser inmovilizados e inyectados con un ácido nucleico, por ejemplo, un constructo de expresión. Aunque la regeneración de plantas a partir de protoplastos no es fácil con cereales, la regeneración de plantas es posible en legumbres utilizando embriogénesis somática a partir de callus derivados de protoplasto. Los tejidos organizados pueden ser transformados con ADN desnudo utilizando la técnica de pistola de genes, en donde el ADN es colocado como recubrimiento sobre microproyectiles de tungsteno, disparado a 1/100 el tamaño de las células, pudiendo llevar el ADN profundamente dentro de las células y sus organelos. El tejido transformado es inducido entonces para regenerarse, usualmente mediante embriogénesis somática. Esta técnica ha sido exitosa en varias especies de cereales incluyendo maíz y arroz.

Los ácidos nucleicos, por ejemplo los constructos de expresión, también pueden ser introducidos en células vegetales utilizando virus recombinantes. Las células vegetales pueden ser transformadas utilizando vectores virales, tales como, por ejemplo, vectores, derivados del virus mosaico del tabaco (Rouwendal (1997) *Plant Mol. Biol.* 33:989-999), véase Porta (1996) "Use of viral replicons for the expression of genes in plants," *Mol. Biotechnol.* 5:209-221.

Alternativamente, los ácidos nucleicos, por ejemplo, un constructo de expresión, pueden ser combinados con regiones que flanquean T-ADN e introducidos en un vector anfitrión convencional de *Agrobacterium tumefaciens*. Las funciones de virulencia del anfitrión de *Agrobacterium tumefaciens* dirigirán la inserción del constructo y el marcador adyacente hacia adentro del ADN de la célula vegetal cuando la célula es infectada por la bacteria. Las técnicas de transformación mediadas por *Agrobacterium tumefaciens*, incluyendo el desarme y el uso de vectores

binarios, está bien descrita en la literatura científica. Véase, por ejemplo, Horsch (1984) *Science* 233:496-498; Fraley (1983) *Proc. Natl. Acad. Sci. USA* 80:4803 (1983); *Gene Transfer to Plants*, Potrykus, ed. (Springer-Verlag, Berlin 1995). El ADN en la célula de *A. tumefaciens* está contenido en el cromosoma bacteriano así como en otra estructura conocida como un plásmido Ti (inductor de tumor). El plásmido Ti contiene un estiramiento de ADN denominado T-ADN (~20kb de longitud) que es transferido a la célula vegetal en el proceso de infección y una serie de genes vir (virulencia) que dirigen el proceso de infección. El *A. tumefaciens* puede infectar a una planta solamente a través de heridas: cuando una raíz o tallo de una planta es herido puede generar ciertas señales químicas, en respuesta a las cuales los genes vir del *A. tumefaciens* se activan y dirigen una serie de eventos necesarios para transferir el T-ADN del plásmido Ti al cromosoma de la planta. El T-ADN entra entonces en la célula vegetal a través de la herida. Una especulación es que el T-ADN espera hasta que el ADN de la planta esté siendo replicado o transcrito, luego se inserta a sí mismo en el ADN expuesto de la planta. Con el fin de utilizar *A. tumefaciens* como vector transgénico, la sección de inducción de tumor del T-ADN tiene que ser retirada, a la vez que se retiene las regiones frontera del T-ADN y los genes vir. El transgén es insertado entonces entre las regiones frontera del T-ADN, en donde es transferido a la célula vegetal y se integra en los cromosomas de la planta.

La invención describe la transformación de plantas monocotiledóneas utilizando los ácidos nucleicos de la invención, incluyendo cereales importantes, véase Hiei (1997) *Plant Mol. Biol.* 35:205-218. véase también, por ejemplo, Horsch, *Science* (1984) 233:496; Fraley (1983) *Proc. Natl. Acad. Sci. USA* 80:4803; Thykjaer (1997) *supra*; Park (1996) *Plant Mol. Biol.* 32:1135-1148, donde se discute la integración de T-ADN en el ADN genómico. Véase también D'Halluin, patente de los Estados Unidos No. 5,712,135, que describe un proceso para la integración estable de un ADN que comprende un gen que es funcional en una célula de un cereal, u otra planta monocotiledónea.

En un aspecto, la tercera etapa puede involucrar la selección y regeneración de plantas completas capaces de transmitir el gen objetivo incorporado a la siguiente generación. Tales técnicas de regeneración se basan en la manipulación de ciertas fitohormonas en un medio de crecimiento de cultivo de tejidos, y puede basarse en un marcador biocida y/o herbicida que ha sido introducido junto con las secuencias de nucleótido deseadas. La regeneración de plantas a partir de protoplastos cultivados está descrita en Evans et al., *Protoplasts Isolation and Culture*, Handbook of Plant Cell Culture, pp. 124-176, MacMillan Publishing Company, New York, 1983; y Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. La regeneración también puede ser obtenida a partir de callus de plantas, explantes, órganos o partes de las mismas. Tales técnicas de regeneración están descritas en general en Klee (1987) *Ann. Rev. of Plant Phys.* 38:467-486. Para obtener plantas completas a partir de tejidos transgénicos tales como embriones inmaduros, pueden crecer bajo condiciones ambientales controladas en una serie de medios que contienen nutrientes y hormonas, un proceso conocido como cultivo de tejidos. Una vez que las plantas completas son generadas y producen semillas, comienza la evaluación de la progenie.

Después de que el casete de expresión es incorporado de manera estable en las plantas transgénicas, puede ser introducido en otras plantas por cruce sexual. Cualquier número de técnicas de cruzamiento estándar pueden ser utilizadas, dependiendo de la especie que se va a cruzar. Puesto que la expresión transgénica de los ácidos nucleicos de la invención lleva a cambios fenotípicos, las plantas que comprenden los ácidos nucleicos que comprenden los ácidos nucleicos de la invención fueron cruzadas sexualmente con una segunda planta para obtener un producto final. Así, la semilla de la invención puede ser derivada de un cruzamiento entre dos plantas transgénicas de la invención, o un cruzamiento entre una planta de la invención y otra planta. Los efectos deseados (por ejemplo, expresión de los polipéptidos de la invención para producir una planta en la cual se altera el comportamiento de floración) puede potenciarse cuando ambas plantas progenitoras expresan los polipéptidos (por ejemplo una transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa) de la invención. Los efectos deseados pueden ser pasados a generaciones futuras de plantas mediante medios de propagación estándar.

Los ácidos nucleicos y polipéptidos de la invención son expresados en o insertados en cualquier planta o semilla. Las plantas transgénicas de la invención pueden ser dicotiledóneas o monocotiledóneas. Ejemplos de plantas transgénicas monocotiledóneas de la invención son pastos, tales como pasto poa (pasto azul, poa), pasto para forraje tal como festuca, lolium, pasto templado, tales como *Agrostis*, y cereales, por ejemplo, trigo, avena, centeno, cebada, arroz, sorgo, y maíz. Ejemplos de plantas transgénicas dicotiledóneas de la invención son tabaco, legumbres, tales como lupinos, patata, remolacha de azúcar, guisantes, judías y soja, y plantas crucíferas (familia *Brassicaceae*), tales como coliflor, colza, y el organismo modelo cercanamente relacionado a *Arabidopsis thaliana*. Así, las plantas y semillas transgénicas de la invención incluyen un rango amplio de plantas, incluyendo, pero no limitándose a, especies de los géneros *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*, *Oryza*, *Panicum*, *Pennisetum*, *Persea*, *Phaseolus*, *Pistachia*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Ricinus*, *Secale*, *Senecio*, *Sinapis*, *Solanum*, *Sorghum*, *Theobromus*, *Trigonella*, *Triticum*, *Vicia*, *Vitis*, *Vigna*, y *Zea*. Las plantas y semillas transgénicas de la invención pueden ser cualquier monocotiledónea o dicotiledónea, por ejemplo, monocotiledóneas como maíz, caña de azúcar, arroz, trigo, cebada, pasto de aguja o

Miscanthus; o dicotiledóneas como cultivos de leguminosas, soja, canola, colza, linaza, algodón, palma de aceite, remolacha de azúcar, cacahuete, árbol de álamo o lupino.

En realizaciones alternativas, los ácidos nucleicos de la invención son expresados en plantas (y/o sus semillas), las cuales contienen células de fibra, incluyendo, por ejemplo, algodón, árbol de algodón sedoso (Kapok, Ceiba pentandra), sauce del desierto, arbusto de creosota, alarba, balso, ramio, kenaf, cáñamo, hibisco, yute, sisal albaca y linaza. En realizaciones alternativas, las plantas transgénicas de la invención pueden ser miembros del género *Gossypium*, incluyendo miembros de cualquier especie de *Gossypium*, tales como *Gossypium*, *G. arboreum*, *G. herbaceum*, *G. barbadense*, y *G. hirsutum*.

La invención también describe plantas transgénicas (y/o sus semillas) para ser usadas para la producción de grandes cantidades de los polipéptidos (por ejemplo, a transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa o anticuerpos) de la invención. Por ejemplo, véase Palmgren (1997) Trends Genet. 13:348; Chong (1997) Transgenic Res. 6:289-296 (que producen la proteína de leche humana beta-caseína en plantas de patata transgénicas utilizando un promotor de manopina sintasa (mas1',2') bidireccional inducible por auxina, con métodos de transformación de disco de hoja mediados por *Agrobacterium tumefaciens*).

Utilizando procedimientos conocidos, una persona de experiencia puede seleccionar plantas (y/o sus semillas) de la invención detectando el incremento o decremento de ARNm o proteína transgénicos en plantas transgénicas. Los medios para detectar y cuantificar los ARNm o proteínas son bien conocidos en la técnica.

Polipéptidos y péptidos

En un aspecto, la invención provee polipéptidos y péptidos aislados, sintéticos o recombinantes que tienen actividad de d-aminoácido transferasa incluyendo las secuencias de aminoácidos de la invención, que incluyen aquellos que tienen al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con SEQ ID NO:220, SEQ ID NO:220 con una, o varias o todas las modificaciones de la tabla 46 o tabla 55.

En un aspecto, la invención provee enzimas quiméricas, incluyendo una d-aminoácido transferasa que tiene dominios de enlazamiento heterólogos, por ejemplo, para el uso en los procesos de la invención y en diversas aplicaciones industriales, médicas, farmacéuticas, de investigación, de procesamiento de alimentos y piensos y suplementos para alimentos y piensos y otras aplicaciones. Por ejemplo, en un aspecto, la invención describe enzimas, por ejemplo, transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas que comprenden uno o más dominios de enlazamiento de una enzima de la invención. Como se describe posteriormente, los dominios de enlazamiento, entre diferentes enzimas de la invención pueden ser intercambiados; o, alternativamente, uno o más dominios de enlazamiento de una o más enzimas de la invención pueden ser intercambiados en una enzima, por ejemplo, una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. En un aspecto de la invención, los dominios de enlazamiento son seleccionados de un dominio de enlazamiento de NAD, NAD(P), calcio, tiamina, FAD, zinc, ADN y/o lípido.

La invención provee adicionalmente enzimas quiméricas que tienen sustratos heterólogos, no naturales; incluyendo enzimas quiméricas que tienen sustratos múltiples por naturaleza de sus dominios de enlazamientos heterólogos "intercambiados" - dando así a la enzima quimérica nueva especificidad o enlazamiento potenciado. Los dominios de enlazamiento heterólogos de las enzimas quiméricas de la invención pueden ser diseñados para ser modulares, esto es, para ser anexados a un módulo catalítico o dominio catalítico (por ejemplo, un sitio activo), en los cuales también pueden ser heterólogos o pueden ser homólogos a la enzima.

La utilización de solamente el módulo catalítico de una transferasa, por ejemplo, una transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa (por ejemplo, una enzima de la invención) han mostrado ser efectivos. Así, la invención describe péptidos y polipéptidos que consisten de, o comprenden, dominios de enlazamiento modular/módulos de sitios activos, los cuales pueden ser apareados o unidos de manera homóloga como pares de módulos de enlazamiento de sitios activos quiméricos (heterólogos). Así, estos polipéptidos/péptidos quiméricos tal como se describen aquí pueden ser utilizados para mejorar o alterar el comportamiento de una enzima individual, por ejemplo, una enzima transferasa, por ejemplo, a transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa. Un módulo catalítico quimérico de la invención (que comprende, por ejemplo, al menos, un dominio de enlazamiento de la invención) puede ser diseñado para direccionar la enzima a regiones particulares de un sustrato. Por ejemplo, en un aspecto, esto se logra haciendo fusiones de la d-aminoácido transferasa, y diversos dominios de enlazamiento (bien sea una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención con un dominio de enlazamiento heterólogo, o, un dominio de enlazamiento de la invención con otra enzima, por ejemplo, por ejemplo,

a transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa.

Así, la invención describe transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas, por ejemplo, una fusión de una transaminasa y/o a deshidrogenasa, quiméricas, por ejemplo, una fusión de una transaminasa y/o una deshidrogenasa con diferentes dominios de enlazamiento (por ejemplo heterólogos). En un aspecto, las transferasas por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas, quiméricas, por ejemplo d-amino deshidrogenasas comprenden una enzima de la invención. En un aspecto, la enzima quimérica comprende funciones de diferentes dominios de enlazamiento. La invención también provee métodos que comprenden la recombinación de diferentes dominios de enlazamiento con diferentes transferasas por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas (por ejemplo, dominios de enlazamiento de la invención y/o transferasa, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas por ejemplo, d-aminoácido deshidrogenasas de la invención) y seleccionando los quiméricos resultantes para encontrar la mejor combinación para una aplicación o sustrato particular.

También se describen otras variaciones, por ejemplo, en donde uno, dos, tres, cuatro o cinco o más residuos son removidos de los extremos terminales carboxi o amino de un polipéptido de la invención. Otra variación incluye la modificación de cualquier residuo para incrementar o disminuir el pI de un polipéptido, por ejemplo, removiendo o modificando (por ejemplo, a otro aminoácido) un glutamato. Este método fue utilizado como esquema general para mejorar las propiedades de la enzima sin crear problemas regulatorios puesto que ningún aminoácido sufre mutación; y este esquema general puede ser utilizado con cualquier polipéptido de la invención.

La invención provee polipéptidos aislados, sintéticos o recombinantes que tienen actividad de d-aminoácido transferasa, en donde el polipéptido tiene una modificación de secuencia de cualquier polipéptido de SEQ ID NO:220, SEQ ID NO:220 con una, varias o todas las modificaciones de la tabla 46 o tabla 55. Los cambios en la secuencia también pueden comprender cualquier modificación en aminoácidos para cambiar el pI de un polipéptido, por ejemplo, eliminación o modificación de un glutamato, o cambiar de un glutamato a otro residuo.

La invención provee adicionalmente polipéptidos aislados, sintéticos o recombinantes que tienen una identidad de secuencia de 90% 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% a una secuencia de la invención.

En un aspecto, el polipéptido tiene una actividad de d-aminoácido transferasa y/o cataliza la transferencia de un grupo químico, cataliza la transaminación, cataliza la reacción: D-alanina + 2-oxoglutarato \rightleftharpoons piruvato + D-glutamato, y/o cataliza una reacción de oxidación-reducción, cataliza la remoción de átomos de hidrógeno, y/o cataliza la reacción: D-aminoácido + H₂O + aceptor \rightleftharpoons un 2-oxo ácido + NH₃ + aceptor reducido.

Cualquier ensayo de actividad de transferasa, por ejemplo, actividad de transaminasa, por ejemplo, actividad de d-aminoácido transferasa conocido en la técnica puede ser utilizado para determinar si un polipéptido tiene actividad de transferasa, por ejemplo, actividad de transaminasa, por ejemplo, actividad de d-aminoácido transferasa, y está dentro del alcance de la invención, por ejemplo, Lee, et. al. (AEM. 2006 February. 72(2): 1588-1594) describe el cálculo de la rata de formación de piruvato utilizando un ensayo de acoplamiento con lactato deshidrogenasa. En otro ejemplo, Mayer (Journal of Biomolecular Screening, Vol. 7, No. 2, 135-140 (2002)) describe un ensayo colorimétrico que utiliza nitroazul de tetrazolio (NBT) y metasulfato de fenazina (PMS) en reacción con NADPH producido por deshidrogenasas para producir un formazano azul-púrpura insoluble.

Los polipéptidos de la invención incluyen d-aminoácidos transferasas, en una forma activa o inactiva. Por ejemplo, los polipéptidos de la invención incluyen proproteínas antes de la "maduración" o procesamiento de preprosecuencias, por ejemplo, mediante una enzima procesadora de proproteínas tal como una proproteína convertasa para generar una proteína madura "activa". Los polipéptidos de la invención incluyen d-aminoácido transferasas inactivas por otras razones, por ejemplo, antes de la "activación" por un evento de procesamiento postraducción, por ejemplo, una acción de una endo o exo peptidasa o proteinasa, un evento de fosforilación, una amidación, una glicosilación o una sulfatación, un evento de dimerización y similares. Los polipéptidos de la invención incluyen todas las formas activas, incluyendo subsecuencias activas, por ejemplo, dominios catalíticos o sitios activos de la d-aminoácido transferasa.

Los métodos para identificar secuencias de dominio "prepro" y secuencias de señalización son bien conocidos en la técnica, véase, por ejemplo, Van de Ven (1993) Crit. Rev. Oncog. 4(2):115-136. Por ejemplo, para identificar una preprosecuencia, la proteína es purificada a partir del espacio extracelular y la secuencia de proteína N-terminal es determinada y comparada con la forma no procesada.

La invención incluye polipéptidos con o sin una secuencia de señalización y/o una preprosecuencia. La invención incluye polipéptidos y secuencias de señalización heterólogas y/o preprosecuencias. La preprosecuencia

(incluyendo una secuencia de la invención usada como un dominio preproheterólogo) puede estar localizada en el extremo amino terminal o carboxi terminal de la proteína. La invención también incluye secuencias de señalización aisladas, sintéticas o recombinantes, secuencias prepro y dominios catalíticos (por ejemplo, "sitios activos") que comprenden secuencias de la invención.

- 5 El porcentaje de identidad de secuencia puede estar por encima de la longitud completa del polipéptido. Los polipéptidos de la invención pueden también ser más cortos que la longitud completa de polipéptidos de ejemplo.

Los péptidos tal como se describen aquí, (por ejemplo, una subsecuencia de un polipéptidos de ejemplo de la invención) pueden ser utilizados, por ejemplo, como sondas de marcación, antígenos, tolerágenos, motivos, sitios activos (por ejemplo, "dominios catalíticos"), secuencias de señalización y/o preprodominios.

- 10 Los polipéptidos y péptidos de la invención pueden ser aislados a partir de fuentes naturales, pueden ser sintéticos, o pueden ser polipéptidos generados por vía recombinante. Los polipéptidos y proteínas pueden ser expresados de manera recombinante *in vitro* o *in vivo*. Los péptidos y polipéptidos de la invención pueden ser hechos y aislados utilizando cualquier método conocido en la técnica. Los polipéptidos y péptidos de la invención también pueden ser sintetizados, solos o en parte, utilizando métodos químicos bien conocidos en la técnica. Véase, por ejemplo, Caruthers (1980) Nucleic Acids Res. Symp. Ser. 215-223; Horn (1980) Nucleic Acids Res. Symp. Ser. 225-232; Banga, A.K., Therapeutic Peptides and Proteins, Formulation, Processing and Delivery Systems (1995) Technomic Publishing Co., Lancaster, PA. Por ejemplo, puede llevarse a cabo la síntesis de péptidos utilizando diversas técnicas en fase sólida (véase, por ejemplo, Roberge (1995) Science 269:202; Merrifield (1997) Methods Enzymol. 289:3-13) y puede lograrse síntesis automatizada, por ejemplo utilizando el ABI 431A Peptide Synthesizer (Perkin Elmer) de acuerdo con las instrucciones provistas por el fabricante.

- 20 Los péptidos y polipéptidos de la invención también pueden ser glicosilados. La glicosilación puede ser agregada postraducción bien sea químicamente o mediante mecanismos biosintéticos celulares, en donde estos últimos incorporan el uso de motivos de glicosilación conocidos, los cuales pueden ser nativos para la secuencia o pueden ser agregados como un péptido o agregados en la secuencias de codificación del ácido nucleico. La glicosilación puede ser enlazada a O ó enlazada a N.

- 25 "Aminoácidos" o "secuencias de aminoácidos" tal como se utilizan aquí se refieren a una secuencia de un oligopéptido, péptido, polipéptido o proteína, o a un fragmento, o porción o subunidad de cualquiera de estos y a moléculas de origen natural o sintéticas. "Aminoácidos" o "secuencia de aminoácidos" incluyen una secuencia de oligopéptidos, péptidos, polipéptidos o proteína, o un fragmento porción o subunidad de cualquiera de estos, y a moléculas de origen natural o sintéticas. El término "polipéptidos" tal como se utiliza aquí, se refiere a aminoácidos unidos uno a otro por enlaces peptídicos o enlaces peptídicos modificados, esto es, isoésteres peptídicos y pueden contener aminoácidos modificados diferentes a los veinte aminoácidos codificados genéticamente. Los polipéptidos pueden modificados bien sea por procesos naturales, tales como procesamientos postraducción, o por técnicas de modificación química que son bien conocidas en el arte. Las modificaciones pueden ocurrir en cualquier lugar del polipéptido, incluyendo el esqueleto peptídico, las cadenas laterales de los aminoácidos y los terminales amino o carboxilo. Será evidente que el mismo tipo de modificación puede estar presente en el mismo o en grados variables en diferentes sitios en un polipéptido dado. También un polipéptido dado puede tener muchos tipos de modificaciones. Las modificaciones incluyen acetilación, acilación, ribosilación con ADP, amidación, unión covalente de flavina, unión covalente de una unidad estructural heme, unión covalente de un nucleótido o derivado de nucleótido, unión covalente de un lípido o derivado de lípido, unión covalente de un fosfatidilinositol, ciclación por entrecruzamiento, formación de puentes disulfuro, desmetilación, formación de enlaces cruzados covalentes, formación de cisteína, formación de piroglutamato, formilación, gama-carboxilación, glicosilación, formación de ancla GPI, hidroxilación, yodación, metilación, miristilación, oxidación, pegilación, fosforilación, prenilación, racemización, selenoilación, sulfatación y adición mediada por ARN de transferencia de aminoácidos a proteínas tales como arginilación. Véase Creighton, T.E., Proteins - Structure and Molecular Properties 2nd Ed., W.H. Freeman and Company, New York (1993) 1993; Posttranslational Covalent Modifications of Proteins, B.C. Johnson, Ed., Academic Press, New York, pp. 1-12 (1983)). Los péptidos y polipéptidos de la invención también incluyen todas las formas "miméticas" y "peptidomiméticas" tal como se describe en más detalle a continuación.

- 30 Los polipéptidos o proteínas "recombinantes" se refieren a polipéptidos o proteínas producidos por técnicas de ADN recombinante; esto es, producidos a partir de células transformadas por un constructo de ADN exógeno que codifica el polipéptido o proteína deseados. Ácidos nucleicos "sintéticos" (incluyendo oligonucleótidos) polipéptidos o proteínas de la invención incluyen aquellos preparados por cualquier síntesis química, por ejemplo, como se describe más adelante. Los métodos de síntesis química de péptidos en fase sólida pueden ser utilizados también para sintetizar el polipéptido o fragmentos de la invención. Tal método ha sido conocido en la técnica de el inicio de la década de 1960 (Merrifield, R. B., J. Am. Chem. Soc., 85:2149-2154, 1963) (véase también Stewart, J. M. and Young, J. D., Solid Phase Peptide Synthesis, 2nd Ed., Pierce Chemical Co., Rockford, Ill., pp. 11-12)) que han sido empleados recientemente en el diseño de péptidos de laboratorio disponibles comercialmente y kits de síntesis (Cambridge Research Biochemicals); tales kits de laboratorio disponibles comercialmente han utilizados siguiendo en general las enseñanzas de H. M. Geysen et al, Proc. Natl. Acad. Sci., USA, 81:3998 (1984) y proveen síntesis de

péptidos bajo la guía de una multitud de “varillas” o “pasadores” todos los cuales están conectados a una placa individual. Cuando se utiliza tal sistema, una placa de varillas o pasadores es invertida e insertada en una segunda placa de pozos o reservorios correspondientes, los cuales contienen soluciones para unir o anclar un aminoácido particular a las puntas del pasador o la varilla. Repitiendo tal etapa de proceso, esto es como invirtiendo e insertando las puntas de las varillas y pasadores en las soluciones apropiadas, se conforman con los aminoácidos los péptidos deseados. Además, hay disponible un cierto número de sistemas de síntesis de péptidos FMOC. Por ejemplo, el ensamblaje de un polipéptido o un fragmento puede llevarse a cabo en un soporte sólido utilizando un sintetizador de péptidos automatizado Applied Biosystems, Inc. Model 431A. Tales equipos proveen fácil acceso a los péptidos de la invención, bien sea por síntesis directa o por síntesis de una serie de fragmentos que pueden ser acoplados utilizando otras técnicas conocidas.

“Fragmentos” tal como se utiliza aquí son una porción de una proteína de origen natural que pueden existir en al menos dos conformaciones diferentes. Los fragmentos pueden tener la misma o sustancialmente la misma secuencia de aminoácidos que la proteína de origen natural. “Sustancialmente la misma” significa que una secuencia de aminoácidos es grandemente, pero no completamente, la misma, pero retiene al menos una actividad funcional de la secuencia con la cual está relacionada. En general dos secuencias de aminoácidos son “sustancialmente la misma” o “sustancialmente homólogas” si son al menos 85% idénticas. Los fragmentos que tienen diferentes estructuras tridimensionales que las proteínas de origen natural también están incluidos. Un ejemplo de estos, es una molécula “proforma”, tal que una proproteína de baja actividad puede ser modificada por escisión para producir una encima madura con actividad significativamente mayor.

Los péptidos y polipéptidos de la invención, tal como se definió anteriormente, incluyen todas las formas “miméticas” y “péptido miméticas”. Los términos “meticas” y “péptido miméticas” se refieren a un compuesto químico sintético que tiene sustancialmente las mismas características estructurales y/o funcionales de los polipéptidos de la invención. Los miméticos pueden estar compuestos bien sea completamente de análogos sintéticos no naturales de aminoácidos, o, es una molécula quimérica de aminoácidos peptídicos parcialmente naturales y análogos parcialmente no naturales de los aminoácidos. Los miméticos también pueden incorporar cualquier partida de sustituciones conservadoras de aminoácidos naturales en tanto tales sustituciones no alteren sustancialmente la estructura y/o actividad mimética. Como sucede con los polipéptidos de la invención que son variantes conservadoras, la experimentación de rutina determinará si un mimético está dentro del alcance de la invención, esto es, que su estructura y/o función no se altera sustancialmente, así, en un aspecto, una composición mimética está dentro del alcance de la invención si tiene una actividad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa.

Las composiciones miméticas de polipéptidos de la invención pueden contener cualquier combinación de componentes estructurales no naturales. En un aspecto alternativo, las composiciones miméticas de la invención incluyen uno o todos de los siguientes tres grupos estructurales: a) grupos de enlace de residuos diferentes al enlace amida natural (“enlace peptídico”), b) residuos no naturales en lugar de residuos de aminoácidos de origen natural; o c) residuos que inducen imitación estructural secundaria, esto es, que inducen o estabilizan una estructura secundaria, por ejemplo, un giro beta, un giro gama, una lámina beta, una conformación de hélice alfa y similares. Por ejemplo, un polipéptido de la invención puede ser caracterizado como mimético cuando todos o algunos de sus residuos están unidos por medios químicos diferentes a los enlaces peptídicos naturales. Los residuos peptidomiméticos individuales pueden ser unidos por enlaces químicos, otros enlaces químicos o medios de acoplamiento, tales como, por ejemplo, glutaraldehído, ésteres de N-hidroxisuccinimida, maleimidas bifuncionales, N,N'-diciclohexilcarbodiimida (DCC) o N,N'-diisopropilcarbodiimida (DIC). Los grupos de enlace que pueden ser una alternativa a las uniones tradicionales de enlace amida (“enlace peptídico”) incluyen, por ejemplo, cetometileno (por ejemplo, -C(=O)-CH₂- para -C(=O)-NH-), aminometileno (CH₂-NH), etileno, olefina (CH=CH), éter (CH₂-O), tioéter (CH₂-S), tetrazol (CN₄-), tiazol, retroamida, tioamida, o éster (véase, por ejemplo, Spatola (1983) in Chemistry and Biochemistry of Amino Acids, , Vol. 7, pp 267-357, "Peptide Backbone Modifications," Marcell Dekker, NY).

Un polipéptido de la invención también puede ser caracterizado como mimético por contener todos o algunos de los residuos no naturales en lugar de residuos de aminoácidos de origen natural. Los residuos no naturales están bien descritos en la literatura científica y de patentes. Unas pocas composiciones no naturales de ejemplo útiles como miméticos de residuos y guías de aminoácidos naturales se describen a continuación. Miméticos de aminoácidos aromáticos pueden ser generados reemplazando, por ejemplo, D- o L- naftilalanina; D- o L- fenilglicina; D- o L-2 tienilalanina; D- o L-1, -2, 3-, o 4- pirenilalanina; D- o L-3 tieneilalanina; D- o L-(2-piridinil)-alanina; D- o L-(3-piridinil)-alanina; D- o L-(2-pirazinil)-alanina; D- o L-(4-isopropil)- fenilglicina; D-(trifluorometil)- henilglicina; D-(trifluorometil)-fenilalanina; D-p-fluoro-fenilalanina; D o L-p-bifenilfenilalanina; D- o L-p-metoxi-bifenilfenilalanina; D- o L-2-indol (alquil)alaninas; y, D- o L alquilalaninas, en donde alquilo puede ser, metilo, etilo, propilo, hexilo, butilo, pentilo, isopropilo, isobutilo, sec-isotilo, iso-pentilo, o un aminoácido no ácido sustituidos o no sustituidos. Anillos aromáticos de aminoácidos no naturales incluyen, por ejemplo, anillos aromáticos de tiazolilo, tiofenilo, pirazolilo, bencimidazolilo, naftilo, furanilo, pirrolilo y piridilo.

Miméticos de aminoácidos ácidos pueden ser generados por sustitución, por ejemplo, aminoácidos no carboxilados mientras que mantienen una carga negativa; (fosfono) alanina, treonina sulfatada. Grupos laterales carboxilo (por ejemplo, aspártico o glutámico) también pueden ser modificados selectivamente por reacción con carbodiimidas (R'-N-C-N-R') tales como por ejemplo, 1-ciclohexil-3-(2-morfolinil-(4-etil) carbodiimida o 1-etil-3-(4-azonia-4,4-dimetolpentil) carbodiimida. Aspártico o glutámico también pueden ser convertidos en residuos asparaginilo y glutaminilo por reacción con iones amonio. Los miméticos de aminoácidos básicos pueden ser generados por sustitución, por ejemplo, (además de lisina y arginina) por los aminoácidos ornitina, citrulina, o ácido (guanidino) -acético, o ácido (guanidino) -acético, en donde alquilo es como se definió más arriba. El derivado nitrilo (por ejemplo, que contiene la unidad estructural CN- en lugar de COOH) puede ser sustituido por asparagina o glutamina. Los residuos de asparaginilo y glutaminilo pueden ser desaminados a los correspondientes residuos aspártico o glutámico. Los imitadores de residuos de arginina pueden ser generados por reacción de arginilo con, por ejemplo, uno o más reactivos convencionales, incluyendo, por ejemplo, fenilgloxal, 2,3-butanodiona, 1,2- ciclohexanodiona o ninhidrina, preferiblemente bajo condiciones alcalinas. Los imitadores de residuo de tirosina pueden ser generados por reacción de tirosilo, por ejemplo, compuestos de diazonio aromáticos o tetranitrometano. El N-acetilimidazol y tetranitrometano pueden ser utilizados para formar especies de O-acetil tirosilo y derivados 3-nitro, respectivamente. Las imitaciones de residuos de cisteína pueden ser generadas haciendo reaccionar residuos de cisteinilo con, por ejemplo, alfa-haloacetatos tales como ácido 2-cloroacético o cloroacetamida y aminos correspondientes; para dar derivados de carboximetilo o carboxiamidometilo. Las imitaciones de residuos de cisteína pueden ser generadas también haciendo reaccionar residuos de cisteinilo con, por ejemplo, bromo-trifluoroacetona, ácido alfa-bromo-beta-(5-imidozoi) propiónico; fosfato de cloroacetilo, N-alquilmaleimidas, disulfuro de 3-nitro-2-piridilo; disulfuro de metil 2-piridilo; p-cloromercuribenzoato; 2-cloromercuri-4-nitrofenol; o, cloro-7-nitrobenzo-oxa-1,3-diazol. Las imitaciones de lisina pueden ser generadas (y los residuos terminales amino pueden ser alterados) por reacción de lisinilo con, por ejemplo anhídridos succínico u otro anhídridos de ácidos carboxílicos. Las imitaciones de lisina de otros residuos que contienen alfa amino también pueden ser generadas por reacción con imidoésteres, tales como picolinimidato de metilo, fosfato de piridoxal, piridoxal, cloroborohidruro, ácido trinitro-bencenosulfónico, O-metilisourea, 2,4,pentanodiona, y reacciones catalizadas por transamidasa con glicoxilato. Las imitaciones de metionina pueden ser generadas por reacción, por ejemplo, con sulfóxido de metionina. Las imitaciones de prolina incluyen, por ejemplo, ácido pipercolico, y ácido tiazolidin carboxílico, 3- o 4-hidroxiprolina, deshidro prolina, 3- o 4-metilprolina, o 3,3,-dimetilprolina. Las imitaciones del residuo de histidina pueden ser generadas haciendo reaccionar histidilo con, por ejemplo, dietilprocarbonato o bromuro de para-bromofenacilo. Otras imitaciones incluyen, por ejemplo, las generadas por hidroxilación de prolina y lisina; la fosforilación de los grupos hidroxilo o serilo o residuos de treonilo; metilación de los grupos alfa-amino de lisina, arginina e histidina; acetilación de la amina N-terminal; metilación de los residuos principales de cadena amida o sustitución con aminoácidos N-metilo; o amidación de los grupos carboxilo C-terminales.

Un residuo, por ejemplo, un aminoácido, de un polipéptido de la invención también puede ser reemplazado por un aminoácido (o un residuo peptidomimético) de quiralidad opuesta. Así, cualquier aminoácido de origen natural en la configuración L (la cual también puede ser denominada como R o S, dependiendo de la estructura de la entidad química), puede ser reemplazado con el amino ácido del mismo tipo estructural químico o un peptidomimético, pero de quiralidad opuesta, denominado como D-aminoácido, pero también puede ser denominada como la forma R o S.

La invención también provee métodos para modificar los polipéptidos de la invención bien sea por procesos naturales, tales como procesamiento postranslacional (por ejemplo, fosforilación, acilación, etc.) o por técnicas de modificación química, y los polipéptidos modificados resultantes. Las modificaciones pueden ocurrir en cualquier lugar de un polipéptido, incluyendo el esqueleto peptídico, las cadenas de aminoácidos laterales y los terminales amino o carboxilo. Será evidente que el mismo tipo de modificación puede estar presente en el mismo o variables grados en varios sitios en un polipéptido dado. También un polipéptido dado puede estar presente en muchos tipos de modificaciones. Las modificaciones incluyen acetilación, acilación, ADP-ribosilación, amidación, enlazamiento covalente de flavina, enlazamiento covalente de una unidad heme, enlazamiento covalente de una nucleótido o derivado de nucleótido, enlazamiento covalente de un lípido o derivado de lípido, enlazamiento covalente de un fosfatidilinositol, ciclización por entrecruzamiento, formación de enlaces disulfuro, desmetilación, formación de enlaces cruzados covalentes, formación de cisteína, formación de piroglutamato, formilación, gamma-carboxilación, glicosilación, formación de ancla GPI, hidroxilación, yodación, metilación, miristilación, oxidación, pegilación, procesamiento proteolítico, fosforilación, prenilación, racemización, selenoilación, sulfatación, y adición de aminoácidos a proteínas mediadas por ARN de transferencia tal como arginilación. Véase, por ejemplo, Creighton, T.E., *Proteins - Structure and Molecular Properties* 2nd Ed., W.H. Freeman and Company, New York (1993); *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York, pp. 1-12 (1983).

Los métodos de síntesis química de péptidos en fase sólida también pueden ser utilizados para sintetizar el polipéptido o fragmentos de la invención. Tal método ha sido conocido en la técnica desde el inicio de la década de 1960 (Merrifield, R. B., *J. Am. Chem. Soc.*, 85:2149-2154, 1963) (Véase también Stewart, J. M. and Young, J. D., *Solid Phase Peptide Synthesis*, 2nd Ed., Pierce Chemical Co., Rockford, Ill., pp. 11-12)) y han sido empleados recientemente en kits de diseño y síntesis de péptidos en laboratorio disponibles comercialmente (Cambridge Research Biochemicals) tales kits de laboratorio comercialmente disponibles han utilizado en general las enseñanzas de H. M. Geysen et al, *Proc. Natl. Acad. Sci., USA*, 81:3998 (1984) y proveen la síntesis de péptidos

sobre las puntas de una multitud de "Varillas" o "Pasadores" todos los cuales están conectados a una placa individual. Cuando se utiliza tal sistema, una placa de varillas o pasadores es invertida e insertada en una segunda placa de pozos o reservorios correspondientes, los cuales contienen soluciones para unir o anclar un aminoácido apropiado a la punta del pasador o varilla. Repitiendo tal etapa del proceso, esto es, invirtiendo e insertando las puntas de las varillas y pasadores en soluciones apropiadas, se construyen los aminoácidos en los péptidos deseados. Además, hay disponible un cierto número de sistemas de síntesis de péptidos FMOC. Por ejemplo, el ensamblaje de un polipéptido o fragmento puede llevarse a cabo sobre un soporte sólido utilizando un sintetizador de péptidos automatizado Applied Biosystems Inc. Model 431A™. Tal equipo provee fácil acceso a los péptidos de la invención, bien sea por síntesis directa o por síntesis de una serie de fragmentos que pueden ser acoplados utilizando otras técnicas conocidas.

La invención incluye d-aminoácido transferasas de la invención con o sin señalización. El polipéptido que comprende una secuencia de señalización de la invención puede ser una d-aminoácido transferasa de la invención, u otra transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa u otra enzima u otro polipéptido.

La invención incluye anticuerpos inmovilizados de transferasas, d-aminoácido transferasas, y anti-d-aminoácido transferasas y fragmentos de los mismos. La invención describe métodos para inhibir la actividad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, por ejemplo, utilizando mutantes negativos dominantes o anticuerpos anti-transferasa, por ejemplo, anti-transaminasa, por ejemplo, anti-d-aminoácido transferasa, y/o anti-oxidorreductasa, por ejemplo, anti-deshidrogenasa, por ejemplo, anti-d-aminoácido deshidrogenasa de la invención. La invención incluye heterocomplejos, por ejemplo, proteínas de fusión, heterodímeros, etc., que comprenden las transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas de la invención.

Los polipéptidos de la invención pueden tener una actividad de d-aminoácido transferasa bajo diversas condiciones, por ejemplo, extremos en pH y/o temperatura, agentes oxidantes, y similares. La invención provee métodos que llevan a preparaciones alternativas de transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa con diferentes eficiencias y estabilidades catalíticas, por ejemplo, si la temperatura, agentes oxidantes y condiciones de lavado son cambiantes. En un aspecto, las variantes de transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa pueden ser producidos utilizando técnicas de mutagénesis dirigida al sitio y/o mutagénesis aleatoria. En un aspecto, la evolución dirigida puede ser utilizada para producir una gran variedad de variantes de transferasa por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa con especificidades y estabilidad alternativas.

Las proteínas de la invención también son útiles como reactivos de investigación para identificar moduladores de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa por ejemplo, activadores o inhibidores de la actividad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. En resumen, muestras de prueba (compuestos, caldos, extractos y similares) son agregados a los ensayos de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa para determinar su capacidad de inhibir la escisión del sustrato. Los inhibidores definidos de esta manera pueden ser utilizados en la industria e investigación para reducir o prevenir proteólisis indeseada. Los inhibidores pueden ser combinados para incrementar el espectro de actividad.

Las enzimas de la invención también son útiles como reactivos de investigación para digerir proteínas o en secuenciación de proteínas. Por ejemplo, las transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas pueden ser utilizadas para romper polipéptidos en fragmentos más pequeños para secuenciación utilizando, por ejemplo, un secuenciador automatizado.

La invención también provee métodos para descubrir nuevas transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas utilizando los ácidos nucleicos, polipéptidos y anticuerpos de la invención. En un aspecto, se seleccionan bibliotecas de fagémidos para el descubrimiento basado en la expresión de transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas. En otro aspecto, se seleccionan bibliotecas de fagos lambda para descubrimiento basado en la expresión de transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas. La selección de las bibliotecas de fagos o fagémidos puede permitir la detección de clones tóxicos; el acceso mejorado al sustrato; necesidad reducida para manipular un anfitrión, omitiendo el potencial para cualquier modificación

resultante de la escisión de masas de la biblioteca; y, crecimiento más rápido a densidades de clones bajas. La selección de bibliotecas de fagos o fagémidos puede hacerse en fase líquida o en fase sólida. En un aspecto, la invención provee la selección en fase líquida. Esto da una mayor flexibilidad en las condiciones de ensayo; flexibilidad de sustrato adicional; sensibilidad más alta para clones débiles; y facilidad de automatización sobre selección en fase sólida.

La invención provee métodos de selección utilizando las proteínas y ácidos nucleicos de la invención y automatización robótica para permitir la ejecución de muchos miles de reacciones biocatalíticas y ensayos de selección en un corto período de tiempo, por ejemplo, por día, así como asegurar un alto nivel de exactitud y reproducibilidad (véase discusión de arreglos, más adelante). Como resultado, una biblioteca de compuestos derivados puede ser producida en materia de semanas. Para enseñanzas adicionales sobre la modificación de moléculas, incluyendo moléculas pequeñas, véase PCT/ US94 /09174.

Otro aspecto de la invención es un polipéptido aislado purificado que comprende la secuencia de la invención y secuencias sustancialmente idénticas al mismo. Como se discutió anteriormente, tales polipéptidos pueden ser obtenidos insertando un ácido nucleico que codifica el polipéptido en un vector de tal manera que la secuencia de codificación está enlazado operativamente a una secuencia capaz de guiar la expresión del polipéptido codificado en una célula anfitriona adecuada. Por ejemplo, el vector de expresión puede comprender un promotor, un sitio de enlazamiento al ribosoma para iniciación de la traducción y un terminador de la transcripción. El vector también puede incluir secuencias apropiadas para amplificar la expresión.

Otro aspecto de la invención es polipéptidos que tienen al menos 90%, al menos aproximadamente 95%, o más de aproximadamente 95% de homología con uno de los polipéptidos de la invención y secuencias sustancialmente idénticas al mismo. La homología puede ser determinada utilizando cualquiera de los programas descritos anteriormente los cuales alinean los polipéptidos o fragmentos que están siendo comparados y determinan el grado de identidad o similitud en aminoácidos entre ellos. Será evidente que la "homología" en aminoácidos incluye sustituciones de aminoácidos conservadoras tales como las descritas más arriba.

Los polipéptidos que tienen homología a uno de los polipéptidos de la invención, pueden ser obtenidos aislando los ácidos nucleicos que los codifican utilizando las técnicas descritas más arriba.

Alternativamente, los polipéptidos o fragmentos homólogos pueden ser obtenidos a través de enriquecimiento bioquímico o procedimientos de purificación. La secuencia de polipéptidos o fragmentos potencialmente homólogos puede ser determinada por electroforesis en gel y/o microsecuenciación. La secuencia del polipéptido o fragmento homólogo prospectivo puede ser comparada con uno de los polipéptidos de la invención, o un fragmento que comprende al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 o 150 de aminoácidos consecutivos utilizando por lo tanto cualquiera de los programas descritos más arriba.

Otro aspecto de la invención es una prueba para identificar fragmentos o variantes de la invención, que retiene la función enzimática de los polipéptidos de la invención. Por ejemplo los fragmentos o variantes de dichos polipéptidos, pueden ser utilizados para catalizar reacciones bioquímicas, las cuales indican que el fragmento variante retiene la actividad enzimática de los polipéptidos de la invención.

El ensayo para determinar si los fragmentos de variantes retienen la actividad enzimática de los polipéptidos de la invención incluye las etapas de poner en contacto el fragmento de polipéptido o variante con una molécula sustrato bajo condiciones que permitan que el fragmento o variante del polipéptido funcione o detecte bien sea un descenso en el nivel de sustrato o un incremento en el nivel de un producto de reacción específico de la reacción entre el polipéptido y el sustrato.

La presente invención explota las propiedades catalíticas únicas de las enzimas. Mientras que el uso de biocatalizadores (por ejemplo, enzimas purificadas o crudas, células no vivas o vivas) en transformaciones químicas requiere normalmente la identificación de un biocatalizador particular que reacciona con un compuesto de partida específico, la presente invención utiliza biocatalizadores seleccionados y condiciones de reacción que son específicos para grupos funcionales que están presentes en muchos compuestos de partida, tales como moléculas pequeñas. Cada biocatalizador es específico para un grupo funcional, o varios grupos funcionales relacionados y puede reaccionar con muchos compuestos de partida que contienen este grupo funcional.

Las reacciones biocatalíticas producen una población de derivados a partir de un compuesto de partida individual. Estos derivados pueden ser sometidos a otra ronda de reacciones biocatalíticas para producir una segunda población de compuestos derivados. Miles de variaciones de la molécula o compuesto pequeño original pueden ser producidas con cada repetición de la derivación biocatalítica.

Las enzimas reaccionan en sitios específicos de un compuesto de partida sin afectar el resto de la molécula, un proceso que es muy difícil de alcanzar utilizando métodos químicos tradicionales. Este alto grado de especificidad biocatalítica provee los medios para identificar un compuesto activo individual dentro de la biblioteca. La biblioteca es caracterizada por la serie de reacciones biocatalíticas utilizadas para producir la, una así llamada "historia

biosintética". Seleccionando la biblioteca para actividades biológicas y trazando la historia biosintética se identifica la secuencia de reacciones específicas que producen el compuesto activo. La secuencia de reacciones se repite y la estructura del compuesto sintetizado se determina. Este modo de identificación, a diferencia de otras metodologías de síntesis y selección, no requieren tecnologías de inmovilización y los compuestos pueden ser sintetizados y probados libres en solución utilizando virtualmente cualquier tipo de ensayo de selección. Es importante anotar, que el alto grado de especificidad de las reacciones enzimáticas sobre grupos funcionales permite el "rastreo" de reacciones enzimáticas específicas que constituyen la biblioteca producida catalíticamente.

Muchas de las etapas del procedimiento se llevan a cabo utilizando automatización robótica lo que permite la ejecución de muchos miles de reacciones catalíticas y de ensayos de selección por día así como asegurar un alto nivel de exactitud y reproducibilidad. Como resultado, puede producirse una biblioteca de compuestos derivados en materia de semanas lo que tomaría años para producir utilizando métodos químicos corrientes.

En un aspecto particular, la invención describe un método para modificar moléculas pequeñas, que comprende poner en contacto un polipéptido codificado por un polinucleótido descrito aquí o fragmentos enzimáticamente activos del mismo con una molécula pequeña para producir una molécula pequeña modificada. Una biblioteca de moléculas pequeñas modificadas es probada para determinar si una molécula pequeña modificada está presente dentro de la librería que exhibe una actividad deseada. Una reacción biocatalítica específica que produce una molécula pequeña modificada de actividad deseada es identificada eliminando sistemáticamente cada una de las reacciones biocatalíticas utilizadas para producir una porción de la biblioteca y luego probando las pequeñas moléculas producidas en la porción de la biblioteca en cuanto a la presencia o ausencia de la molécula pequeña modificada con la actividad deseada. Las reacciones biocatalíticas específicas que producen la molécula pequeña modificada de actividad deseada son repetidas en un aspecto (opcionalmente). Las reacciones biocatalíticas son llevadas a cabo con un grupo de biocatalizadores que reaccionan con distintas unidades estructurales encontradas dentro de la estructura de una molécula pequeña, cada biocatalizador es específico para una medida estructural o un grupo de unidades estructurales relacionadas; y cada biocatalizador reacciona con muchas diferentes moléculas pequeñas que contienen la unidad estructural distintiva.

Secuencias de señalización de transaminasa y óxido reductasa de preprodominios y dominios catalíticos

La invención describe secuencias de señalización de transferasa, por ejemplo, transaminasa, por ejemplo d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo D-aminoácido deshidrogenasa (por ejemplo péptidos de señalización (SPs)), preprodominios y dominios catalíticos (CDs). Los SP, preprodominios y/o CD de la invención pueden ser péptidos aislados, sintéticos o recombinantes o pueden ser parte de una proteína de fusión, por ejemplo, como un dominio heterólogo en una proteína quimérica. La invención describe ácidos nucleicos que codifican estos dominios catalíticos (CD) preprodominios y secuencias de señalización (SPs, por ejemplo, un péptido que tiene una secuencia que comprende/consiste de residuos amino terminales de un polipéptido de la invención). En un aspecto, la invención describe una secuencia de señalización que comprende un péptido que comprende/consiste de una secuencia tal como se define en los residuos 1 a 12, 1 a 13, 1 a 14, 1 a 15, 1 a 16, 1 a 17, 1 a 18, 1 a 19, 1 a 20, 1 a 21, 1 a 22, 1 a 23, 1 a 24, 1 a 25, 1 a 26, 1 a 27, 1 a 28, 1 a 28, 1 a 30, 1 a 31, 1 a 32, 1 a 33, 1 a 34, 1 a 35, 1 a 36, 1 a 37, 1 a 38, 1 a 39, 1 a 40, 1 a 41, 1 a 42, 1 a 43, 1 a 44, 1 a 45, 1 a 46, 1 a 47, 1 a 48, 1 a 49 o 1 a 50, de un polipéptido de la invención.

Las secuencias de transferasa, por ejemplo, transaminasa, por ejemplo d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo d-aminoácido deshidrogenasa (SP) y/o preprosecuencias tal como se describen pueden ser péptidos aislados, o secuencias unidas a otra transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa or a no-transferasa, por ejemplo, no-transaminasa, por ejemplo, no-d-aminoácido transferasa, y/o no-oxidorreductasa, por ejemplo, no-deshidrogenasa, por ejemplo, no-d-aminoácido deshidrogenasa polipéptido, por ejemplo, como una proteína de fusión (quimérica). En un aspecto, la invención describe polipéptidos que comprenden secuencias de señalización de por ejemplo transferasa, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención. En un aspecto, los polipéptidos que comprenden secuencias de señalización de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa SP y/o prepro de la invención comprende secuencias heterólogas a una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa or a no-transferasa, por ejemplo, no-transaminasa, por ejemplo, no-d-aminoácido transferasa, y/o no-oxidorreductasa, por ejemplo, no-deshidrogenasa, por ejemplo, proteína de polipéptido de no-d-aminoácido deshidrogenasa). En un aspecto, la invención describe transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención con SP heterólogos y/o preprosecuencias, por ejemplo, secuencias con una secuencia de señalización de levadura. Una transferasa, por ejemplo, transaminasa, por

ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención puede comprender un SP heterólogo y/o un prepro en un vector, por ejemplo, un vector de serie pPIC (Invitrogen, Carlsbad, CA).

En un aspecto, las secuencias SP y/o preprosecuencias tal como se describen son identificadas siguiendo la identificación de novedosos polipéptidos de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. Los caminos mediante los cuales las proteínas son conducidas y transportadas a sus localizaciones celulares apropiadas se denominan frecuentemente como caminos de direccionamiento de proteínas. Uno de los elementos más importantes en todos estos sistemas de direccionamiento es una secuencia de aminoácidos corta en el terminal amino de un polipéptido recién sintetizado denominado a la secuencia de señalización. La secuencia de señalización dirige una proteína a su localización apropiada en la célula, y es retirada durante el transporte o cuando la proteína alcanza su destino final. La mayoría de las proteínas del lisosoma, membrana o secretadas tienen una secuencia de señalización con terminal amino que las marca para translocación hacia el lumen del retículo endoplasmático. Más de 100 secuencias de señalización para proteínas en este grupo han sido determinadas. Las secuencias de señalización pueden variar en longitud de entre aproximadamente 10 a 50 residuos de aminoácidos. Diversos métodos de reconocimiento de secuencias de señalización son conocidos para los experimentados en la técnica. Por ejemplo, en un aspecto, novedosos péptidos de señalización de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa son identificados por un método denominado como SignalP. SignalP utiliza una red neural combinada que reconoce ambas señales de los péptidos y su sitios de escisión; véase, por ejemplo, Nielsen (1997) "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites." Protein Engineering 10:1-6.

Debe entenderse que en algunos aspectos, las transferasas por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas de la invención pueden no tener secuencias SP y/o preprosecuencias, o "dominios". En un aspecto, la invención provee las transferasas, por ejemplo transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas de la invención que carecen de todos o parte de los dominios SP y/o prepro. En un aspecto, la invención provee una secuencia de ácidos nucleicos que codifica una secuencia de señalización (SP) y/o prepro de una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa enlazados operativamente a una secuencia de ácidos nucleicos de una diferente transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa or, en un aspecto (opcionalmente), puede ser deseable una secuencia de señalización (SP) y/o un dominio prepro para una proteína no transferasa, por ejemplo no transaminasa, por ejemplo no d-aminoácido transferasa, y/o no oxidorreductasa, por ejemplo, no deshidrogenasa, por ejemplo no d-aminoácido deshidrogenasa.

La invención también provee polipéptidos sintéticos o recombinantes aislados que comprenden secuencias de señalización (SPs), preprodominios y/o dominios catalíticos (CD) de la invención y secuencias heterólogas. Las secuencias heterólogas son secuencias que se asocian de manera no natural (por ejemplo a una transferasa, por ejemplo a transaminasa, por ejemplo d-aminoácidos transferasa, y/o oxidorreductasa, por ejemplo deshidrogenasa, por ejemplo d-aminoácido deshidrogenasa) con un preprodominio SP y/o CD. La secuencia a la cual el preprodominio SP y/o CD no está naturalmente asociado puede estar en el preprodominio SP y/o en el terminal amino CD, en el extremo terminal carboxi, o en ambos extremos de SP y/o CD. En un aspecto, la invención provee un polipéptido aislado, sintético o recombinante que comprende (o consiste de) un polipéptido que comprende una secuencia de señalización (SP), un preprodominio y/o un dominio catalítico (CD) de la invención con la condición de que no esté asociado con ninguna secuencia a la cual se asocia naturalmente (por ejemplo, una secuencia de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa). De la misma forma en un aspecto, la invención provee ácidos nucleicos aislados, sintéticos o recombinantes de la invención que codifican la secuencia para una secuencia de señalización (SP), un preprodominio y/o un dominio catalítico (CD) de la invención y en la secuencia heteróloga (esto es, una secuencia que no se asocia de manera natural con una secuencia de señalización (SP), preprodominio y/o dominio catalítico (CD) de la invención). La secuencia heteróloga puede estar en el extremo terminal 3', en el extremo terminal 5', y/o en ambos extremos de la secuencia de codificación SP, preprodominio y/o CD.

Transferasas y/u oxidorreductasas híbridas (quiméricas) y bibliotecas de péptidos

En un aspecto la invención describe transferasas híbridas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas y proteínas de fusión, que incluyen bibliotecas de péptidos, que comprenden secuencias de la invención. Las bibliotecas de péptidos de la invención pueden ser utilizadas para aislar moduladores de péptidos (por ejemplo activadores o inhibidores) de objetivos, tales como sustratos, receptores, enzimas de transferasa, por ejemplo transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por

ejemplo, d-aminoácido deshidrogenasa. Las bibliotecas de péptidos de la invención pueden ser utilizadas para identificar asociados de enlace formales de los objetivos, tales como ligandos, por ejemplo, citoquinas, hormonas y similares. En un aspecto, la invención provee proteínas quiméricas que comprenden una secuencia de señalización (SP), preprodominios y/o dominio catalítico sede de la invención o una combinación de los mismos y una secuencia heteróloga (véase más arriba).

En un aspecto, las proteínas de fusión tal como están descritas (por ejemplo, la unidad estructural péptidos) están estabilizadas conformacionalmente (con respecto a los péptidos lineales) para permitir una afinidad de enlace superior a los objetivos. La invención describe fusiones de transferasas, por ejemplo transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas de la invención y otros péptidos, incluyendo péptidos conocidos y aleatorios. Pueden ser fusionados de tal manera que la estructura de las transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas no es perturbada significativamente y el péptido es metabólicamente o estructuralmente estabilizado en sentido conformacional. Esto permite la creación de una biblioteca de péptidos que es monitorizada fácilmente tanto por su presencia dentro de las células como en su cantidad.

Las variantes en secuencia de aminoácidos de la invención pueden ser caracterizadas por la naturaleza predeterminada de la variación, una característica que las fija separadas de cualquier forma de origen natural, por ejemplo, una secuencia de variación alélica o interespecies de una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. En un aspecto, las variantes de la invención exhiben la misma actividad biológica cualitativa que el análogo de origen natural. Alternativamente, las variantes pueden ser seleccionadas por tener características modificadas. En un aspecto, mientras que el sitio de región para introducir una variación en la secuencia de aminoácidos está predeterminado, la mutación per se no necesita ser predeterminada. Por ejemplo, con el fin de optimizar el rendimiento de una mutación en un sitio dado, la mutagénesis aleatoria puede ser llevada a cabo en el codón o región no objetivo, por ejemplo las variantes expresadas de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa seleccionadas para la combinación óptima de la actividad deseada. Las técnicas para hacer mutaciones por sustitución en sitios predeterminados en el ADN que tienen una secuencia conocida son bien conocidas, tal como se discute aquí por ejemplo, en la mutagénesis de cebador M13 y mutagénesis por PCR. La selección de los mutantes puede hacerse utilizando, por ejemplo, ensayos de actividad de transaminasa y/o oxidorreductasa. En aspectos alternativos, las sustituciones de aminoácidos pueden ser residuos individuales; las inserciones pueden ser en el orden de desde aproximadamente 1 hasta 20 aminoácidos, aunque pueden hacerse inserciones considerablemente más grandes. Las eliminaciones pueden variar desde aproximadamente 1 hasta aproximadamente 20, 30, 40, 50, 60, 70 residuos o más. Para obtener un derivado final con las propiedades óptimas, pueden utilizarse sustituciones, eliminaciones, inserciones o cualquier combinación de los mismos. En general, estos cambios se hacen sobre unos pocos aminoácidos para minimizar alteración de la molécula. Sin embargo, pueden tolerarse cambios más grandes en ciertas circunstancias.

La invención describe transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas en donde la estructura del esqueleto polipeptídico, la estructura secundaria o la terciaria, por ejemplo, una estructura alfa-helicoidal o de lámina beta, ha sido modificada. En un aspecto, la carga de hidrofobicidad ha sido modificada. En otro aspecto, el volumen de una cadena lateral ha sido modificado. Cambios sustanciales en función o en identidad inmunológica se hacen seleccionando sustituciones que sean menos conservadoras. Por ejemplo, pueden hacerse sustituciones que afecten más significativamente: la estructura del esqueleto polipeptídico en el área de la alteración, por ejemplo, una estructura alfa-helicoidal o de lámina beta; una carga o un sitio hidrófobo de la molécula, el cual puede ser un sitio activo; o una cadena lateral. La invención describe sustituciones en polipéptidos de la invención en donde (a) un residuo hidrofílico, por ejemplo serilo o treonilo, es sustituido por (o por) un residuo hidrófobo, por ejemplo, leucilo, isoleucilo, fenilalanilo, valilo, o alanilo; (b) una cisteína o proteína es sustituida por (o por) otro residuo; (c) un residuo que tiene una cadena lateral electropositiva, por ejemplo, lisilo, arginilo o histidilo, es sustituida por (o por) un residuo electronegativo, por ejemplo glutamilo o aspartilo; o (d) un residuo que tiene una cadena lateral voluminosa, por ejemplo fenilalanina, es sustituido por (o por) uno que no tenga una cadena lateral, por ejemplo glicina. Las variantes pueden exhibir la misma actividad biológica cualitativa (por ejemplo, actividad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa) aunque pueden seleccionarse variantes para modificar las características de las transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas según sea necesario.

En un aspecto, las transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas de la invención comprenden epítopos o etiquetas de purificación, secuencias de señalización u otras secuencias de fusión, etc. En un aspecto, las transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas de la invención

pueden ser fusionados en un péptido aleatorio para formar un polipéptido de fusión. Por “fusionado” o “enlazado operativamente” aquí se entiende que el péptido aleatorio y la transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa son enlazados entre sí, de tal manera que se minimice la perturbación en la estabilidad de la estructura, por ejemplo, retiene la actividad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. El polipéptido de fusión (o polinucleótido de fusión que codifica el polipéptido de fusión) puede comprender componentes adicionales también, incluyendo múltiples péptidos y múltiples bucles.

En un aspecto, los péptidos y ácidos nucleicos que los codifican son aleatorizados, bien sea aleatorizados o son desplazados en su aleatorización, por ejemplo, en frecuencia nucleótido/residuo generalmente o por posición. “Aleatorizado” significa que cada ácido nucleico y péptido consiste de nucleótidos y aminoácidos esencialmente aleatorios, respectivamente. En un aspecto, los ácidos nucleicos que dan lugar a los péptidos pueden ser sintetizados químicamente, y así pueden incorporar cualquier nucleótido en cualquier posición. Así, cuando los ácidos nucleicos son expresados para formar péptidos, cualquier residuo de aminoácidos puede ser incorporado en cualquier posición. El proceso sintético puede ser diseñado para generar ácidos nucleicos aleatorizados, para permitir la formación de todos o la mayor parte de las combinaciones posibles a lo largo de la longitud del ácido nucleico, formando así una biblioteca de ácidos nucleicos aleatorizados. La biblioteca puede proveer una población estructuralmente diversa suficiente de productos de expresión aleatorizada para afectar un rango probabilísticamente suficiente de respuestas celulares para proveer una o más células que exhiben una respuesta deseada. Así, la invención describe una biblioteca de interacción lo suficientemente grande de tal forma que al menos uno de sus miembros tendrá una estructura que le da afinidad por alguna molécula, proteína u otro factor.

Las transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas son enzimas multidominio que consisten en un aspecto (opcionalmente) de un péptido de señalización, un dominio de enlazamiento, un dominio catalítico, un dominio enlazante y/u otro dominio catalítico.

La invención describe un medio para generar polipéptidos quiméricos los cuales pueden codificar polipéptidos híbridos biológicamente activos (por ejemplo, transferasas híbridas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas). En un aspecto, los polinucleótidos originales codifican polipéptidos biológicamente activos. El método tal como se describe producen nuevos polipéptidos híbridos utilizando procesos celulares que integran la secuencia de los polinucleótidos originales de tal manera que el polinucleótido híbrido resultante codifica un polipéptido que demuestra actividades derivadas de los polipéptidos biológicamente activos originales. Por ejemplo, los polinucleótidos originales pueden codificar una enzima particular a partir de diferentes microorganismos. Una enzima codificada por un primer polinucleótido de un organismo o variante, puede, por ejemplo, funcionar efectivamente bajo una condición ambiental particular, por ejemplo, alta salinidad. Una enzima codificada por un segundo polinucleótido a partir de un organismo diferente o variante puede funcionar efectivamente bajo una condición ambiental diferente, tal como temperaturas extremadamente altas. Un polinucleótido híbrido que contiene secuencias del primero y segundo polinucleótidos originales puede codificar una enzima que exhibe características de ambas enzimas codificadas por los polinucleótidos originales. Así, la enzima codificada por el polinucleótido híbrido puede funcionar efectivamente bajo condiciones ambientales compartidas por cada una de las enzimas codificadas por el primero y segundo polinucleótidos, por ejemplo, alta salinidad y temperaturas extremas.

Un polipéptido híbrido resultante del método tal como se describe puede exhibir actividad enzimática especializada no desplegada en las enzimas originales. Por ejemplo, después de la recombinación y/o reordenamiento reductor de los polinucleótidos que codifican la actividad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, el polipéptido híbrido resultante codificado por un polinucleótido híbrido puede ser seleccionado para actividades de transferasas, por ejemplo, de transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa obtenidas de cada una de las enzimas originales, esto es, el tipo de enlace sobre el cual la transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa actúa y la temperatura a la cual la transferasa, por ejemplo la transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa funciona. Así, por ejemplo, la transferasa, por ejemplo transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa pueden ser seleccionadas para establecer aquellas funcionalidades químicas que distinguen la transferasa híbrida, por ejemplo transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de las transferasas originales, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas, por ejemplo, diferencias en actividad a diversas temperaturas, pH o concentraciones de sal.

Las fuentes de los polinucleótidos originales pueden ser aisladas a partir de organismos individuales ("aislados"), colecciones de organismos que han sido cultivados en medios definidos ("medios de enriquecimiento") u organismos no cultivados ("muestras ambientales"). El uso de una metodología independiente del cultivo para derivar polinucleótidos que codifican bioactividades novedosas a partir de muestras ambientales es lo más preferible puesto que permite tener acceso a recursos no descubiertos de biodiversidad.

"Bibliotecas ambientales" son generadas a partir de muestras ambientales y representan los genomas colectivos de organismos de origen natural alcanzados en vectores de clonación que pueden ser propagados en huéspedes procariotas adecuadas. Puesto que el ADN clonado es extraído inicialmente de manera directa de muestras ambientales, las bibliotecas no están limitadas a la pequeña fracción de procariotas que pueden cultivarse en cultivo puro. Adicionalmente, una normalización del ADN ambiental presente en estas muestras podría permitir una representación más igual del ADN proveniente de todas las especies presentes en la muestra original. Esto puede incrementar dramáticamente la eficiencia del hallazgo de genes interesantes a partir de constituyentes menores de la muestra que pueden estar subrepresentados por varias órdenes de magnitud en comparación con las especies dominantes.

Por ejemplo, las bibliotecas de genes generadas a partir de uno o más microorganismos no cultivados son seleccionadas en cuanto a una actividad de interés. Los caminos potenciales que codifican moléculas bioactivas de interés son capturados primero en células procariotas en la forma de bibliotecas de expresión genética. Los polinucleótidos que codifican actividades de interés son aislados a partir de tales bibliotecas e introducidos en una célula anfitriona. La célula anfitriona es cultivada bajo condiciones que promueven la recombinación y/o reordenamiento reductivo creando biomoléculas potencialmente activas con actividades novedosas o potenciadas.

Adicionalmente, la subclonación puede llevarse a cabo para aislar adicionalmente secuencias de interés. En la subclonación, se amplifica una porción del ADN, se digiere, generalmente por enzimas de restricción, para cortar la secuencia deseada, la secuencia deseada es ligada en un vector receptor y es amplificada. En cada etapa en la subclonación, la porción es examinada en cuanto a la actividad de interés, con el fin de asegurarse de que el ADN que codifica la proteína estructural no ha sido excluido. El inserto puede ser purificado en cualquier etapa de la subclonación, por ejemplo, por electroforesis en gel antes de la ligación en un vector o cuando las células que contienen el vector receptor y las células que no contienen el vector receptor son colocadas en un medio selectivo que contiene, por ejemplo, un antibiótico, el cual mata las células que no contienen el vector receptor. Métodos específicos para subclonar insertos de ADN en vectores son bien conocidos en la técnica (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press (1989)). En otro aspecto, las enzimas de la invención son subclones. Tales subclones pueden diferir del clon progenitor por ejemplo, en cuanto a longitud, una mutación, una etiqueta o un marcador.

Los microorganismos de los cuales puede ser preparado el polinucleótido incluye microorganismos procariotas, tales como Eubacteria y Archaeobacteria y microorganismos eucariotas inferiores tales como hongos, algunas algas y protozoos. Los polinucleótidos pueden ser aislados a partir de muestras ambientales en cuyo caso el ácido nucleico puede ser recuperado sin cultivar un organismo o puede ser recuperado a partir de uno o más organismos cultivados. En un aspecto, tales microorganismos pueden ser extremófilos, tales como hipertermófilos, psicrófilos, psicrótofos, halófilos, barófilos y acidófilos. Los polinucleótidos que codifican enzimas aislados de microorganismos extremófilos pueden ser utilizados. Tales enzimas pueden funcionar a temperaturas por encima de 100°C en fuentes calientes terrestres y escapes térmicos en el mar profundo, a temperaturas por debajo de 0°C en aguas árticas, en el ambiente salino saturado del Mar Muerto, a valores de pH alrededor de 0 en depósitos de carbón y en fuentes geotérmicas ricas en azufre, o a valores de pH superiores a 11 en lodos de desagüe. Por ejemplo, varias esterasas y lipasas clonadas y expresadas a partir de organismos extremófilos muestran alta actividad a lo largo de un amplio rango de temperaturas y pHs.

Los polinucleótidos seleccionados y aislados como se describió anteriormente son introducidos en una célula anfitriona adecuada. Una célula anfitriona adecuada es una célula que es capaz de promover la recombinación y/o reordenamiento reductivo. Los polinucleótidos seleccionados están preferiblemente en un vector que incluye secuencias de control apropiadas. La célula anfitriona puede ser una célula eucariota superior, tal como una célula de mamífero, o una célula eucariota inferior, tal como una célula de levadura, o preferiblemente, la célula anfitriona puede ser una célula procariota, tal como una célula bacteriana. La introducción del constructo en la célula anfitriona puede ser efectuada por transfección con fosfato de calcio, transfección mediada por DEAE-Dextrano, o electroporación (Davis et al., 1986).

Como ejemplos representativos de anfitriones apropiados, pueden mencionarse: células bacterianas, tales como E.coli, Streptomyces, Salmonella typhimurium; células fúngicas, tales como levadura; células de insecto tales como Drosophila S2 y Spodoptera Sf9; células animales tales como CHO, COS o melanoma de Bowes; adenovirus; y células vegetales. La selección de un anfitrión adecuado se considera dentro del alcance de los experimentados en la técnica a partir de las presentes enseñanzas.

Con referencia particular a diversos sistemas de cultivo de células de mamíferos que pueden ser empleados para expresar proteínas recombinantes, ejemplos de sistemas de expresión en mamíferos incluyen las líneas COS-7 de fibroblastos de riñón de mono, descritas en "SV40-transformed simian cells support the replication of early SV40 mutants" (Gluzman, 1981) y otras líneas celulares capaces de expresar un vector compatible, por ejemplo, las líneas celulares C127, 3T3, CHO, HeLa y BHK. Los vectores de expresión en mamíferos comprenderán un origen de la replicación, un promotor adecuado y un potenciador y también cualquier sitio de enlazamiento a ribosomas necesario, sitio de poliadenilación, sitios de división de donante y aceptor, secuencias de terminación transcripcionales y secuencias no transcritas flanqueantes de 5'. Las secuencias de ADN derivadas de la división de SV40 y los sitios de poliadenilación pueden ser utilizadas para proveer los elementos genéticos no transcritos requeridos.

En otro aspecto, se prevé que el método tal como está descrito puede ser utilizado para generar novedosos polinucleótidos que codifican caminos bioquímicos de uno o más operones o aglomeraciones de genes o porciones de los mismos. Por ejemplo, bacterias y muchos eucariotas tienen un mecanismo coordinado para regular genes cuyos productos están involucrados en procesos relacionados. Los genes son aglomerados, en estructuras denominadas como "aglomeraciones de genes", sobre un cromosoma individual y son transcritos juntos bajo el control de una secuencia reguladora individual, incluyendo un promotor individual el cual inicia la transcripción para la aglomeración completa. Así, una aglomeración de genes es un grupo de genes adyacentes que son bien sea idénticos o relacionados, usualmente en cuanto a su función. Un ejemplo de un camino bioquímico codificado por aglomeraciones de genes son los policétidos.

El ADN de aglomeraciones de genes puede ser aislado a partir de diferentes organismos y ligado en vectores, particularmente vectores que contienen secuencias reguladoras de la expresión las cuales pueden controlar y regular la producción de una proteína o arreglo relacionado con proteína detectable a partir de las aglomeraciones de genes ligados. El uso de vectores que tienen una capacidad excepcionalmente grande para la introducción de ADN exógeno son particularmente apropiados para uso con tales aglomeraciones de genes y son descritos a manera de ejemplo aquí para incluir el factor f (o factor de fertilidad) de E.coli. Este factor f de E. coli es un plásmido que afecta la transferencia a alta frecuencia de sí mismo durante la conjugación y es ideal para alcanzar y propagar de manera estable grandes fragmentos de ADN, tales como aglomeraciones de genes a partir de muestras biológicas mixtas. Un aspecto de la invención es utilizar vectores de clonación, denominados como "fósidos" o vectores de cromosomas artificiales bacterianos (BAC). Estos son derivados del factor de f de E.coli el cual es capaz de integrar de manera estable grandes segmentos del ADN genómico. Cuando se integran con ADN a partir de una muestra ambiental mixta no cultivada, esto hace posible alcanzar grandes fragmentos genómicos en la forma de una "biblioteca de ADN ambiental" estable. Otro tipo de vector para uso de la presente invención es un vector cósmido. Los vectores cósmidos fueron diseñados originalmente para clonar y propagar grandes segmentos de ADN genómico. La clonación en vectores cósmidos esta descrito en detalle en Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press (1989). Una vez ligados en un vector apropiado, dos o más vectores que contienen diferentes aglomeraciones de gen de policétido sintasa pueden ser introducidos en una célula anfitriona adecuada. Las regiones de homología de secuencia parciales compartidas por las aglomeraciones de genes promoverán procesos que dan como resultado la reorganización en secuencias que a su vez da como resultado una aglomeración de genes híbridos. La aglomeración de genes híbridos novedosa puede entonces ser seleccionada en cuanto a actividades potenciadas no encontradas en las aglomeraciones de genes originales.

Por lo tanto, en un aspecto, la invención describe un método para producir un polipéptido híbrido biológicamente activo y selección de tal polipéptido en cuanto a su actividad potenciada mediante:

1) introducción de al menos un primer polinucleótido en enlazamiento operable y un segundo polinucleótido en enlazamiento operable, compartiendo el al menos primer polinucleótido y segundo polinucleótido una región de homología de secuencia parcial, en una célula anfitriona adecuada;

2) cultivar la célula anfitriona bajo condiciones que promueven la reorganización de secuencias dando como resultado un polinucleótido híbrido en enlazamiento operable;

3) expresión de un polipéptido híbrido codificado por un polinucleótido híbrido;

4) selección del polipéptido híbrido bajo condiciones que promueven la identificación de actividad biológica potenciada; y

5) aislar el polinucleótido que codifica el polipéptido híbrido.

Métodos para seleccionar diversas actividades enzimáticas son conocidos para los experimentados en la técnica y se discuten a lo largo de la presente especificación. Tales métodos pueden ser empleados cuando se aíslan los polipéptidos y polinucleótidos de la invención.

Metodología de selección y dispositivos de monitorización "On-line"

En la práctica de los métodos de la invención, puede utilizarse una variedad de aparatos y metodologías en conjunción con los polipéptidos y ácidos nucleicos de la invención, por ejemplo, para seleccionar polipéptidos para actividad de transferasa, por ejemplo transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, para seleccionar compuestos como moduladores potenciales, por ejemplo, activadores o inhibidores, de una actividad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, para anticuerpos que se enlazan a un polipéptido de la invención, para ácidos nucleicos que hibridan a un ácido nucleico de la invención, para seleccionar células que expresan un polipéptido de la invención y similares. Además de los formatos de arreglo descritos en detalle más adelante para seleccionar muestras, también pueden utilizarse formatos alternativos para practicar los métodos de la invención. Tales formatos incluyen, por ejemplo, espectrómetros de masa, cromatógrafos, por ejemplo, HPLC de alto rendimiento y otras formas de cromatografía líquida y formatos más pequeños, tales como placas de 1536 pozos, placas de 384 pozos y así sucesivamente. Aparatos de selección de alto rendimiento pueden ser adaptados y utilizados para practicar los métodos de la invención, véase, por ejemplo la Solicitud de Patente de los Estados Unidos No. 20020001809.

Arreglos capilares

Los ácidos nucleicos o polipéptidos de la invención pueden ser inmovilizados o ser aplicados a un arreglo. Los arreglos pueden ser utilizados para seleccionar o monitorizar bibliotecas de composiciones (por ejemplo, moléculas pequeñas, anticuerpos, ácidos nucleicos, etc.) en cuanto a su habilidad para enlazarse a o modular la actividad de un ácido nucleico o un polipéptido de la invención. Los arreglos capilares, tales como GIGAMATRIZTM, Diversa Corporation, San Diego, CA; y arreglos descritos por ejemplo en la Solicitud de Patente No. 20020080350 A1; WO 0231203 A; WO 0244336 A, proveen un aparato alternativo para sostener y seleccionar muestras. En un aspecto, el arreglo capilar incluye una pluralidad de capilares conformados en un arreglo de capilares adyacentes, en donde cada capilar comprende al menos una pared que define un lumen para retener una muestra. El lumen puede ser cilíndrico, cuadrado, hexagonal o cualquier otra forma geométrica en tanto las paredes formen un lumen para la retención de un líquido o muestra. Los capilares del arreglo capilar pueden ser mantenidos juntos en cercana proximidad para formar una estructura plana. Los capilares pueden ser unidos entre sí, bien fusionados (por ejemplo, cuando los capilares están hechos de vidrio), engomados, pegados o pinzados lado a lado. Adicionalmente, el arreglo de capilares puede incluir material intersticial dispuesto entre capilares adyacentes en el arreglo, formando por lo tanto un dispositivo plano sólido que contiene una pluralidad de orificios que pasan de un lado al otro.

Un arreglo de capilares puede ser formado a partir de cualquier número de capilares individuales, por ejemplo, un rango de 100 a 4.000.000 de capilares. Adicionalmente, un arreglo de capilares que tiene alrededor de 100.000 o más capilares individuales puede ser formado en el tamaño estándar y forma de una placa de Microtiter® para ajuste en equipo de laboratorio estándar. Los lúmenes son llenados manual o automáticamente utilizando bien sea la acción capilar o microinyección utilizando una aguja delgada. Las muestras de interés pueden ser removidas subsecuentemente de los capilares individuales para análisis o caracterización posterior. Por ejemplo, una sonda similar a una aguja delgada es posicionada en comunicación fluida con un capilar seleccionado bien sea para agregar o retirar material del lumen.

En un ensayo de selección de recipiente sencillo, los componentes del ensayo son mezclados produciendo una solución de interés, antes de la inserción en el arreglo capilar. El lumen es llenado por acción capilar cuando al menos una porción del arreglo es sumergida en una solución de interés. Las reacciones químicas o biológicas y/o la inactividad en cada capilar son monitoreadas para eventos detectables. Un evento detectable se denomina frecuentemente como un “acierto”, el cual puede usualmente ser distinguido de los capilares que producen “no aciertos” por detección óptica. Así, los arreglos de capilares permiten una detección masivamente paralela de “aciertos”.

En un ensayo de selección de recipiente múltiple, un polipéptido o ácido nucleico, por ejemplo un ligando, puede ser introducido en un primer componente, el cual es introducido en al menos una porción de un capilar de un arreglo de capilares. Puede introducirse entonces una burbuja de aire dentro del capilar por detrás del primer componente. Puede introducirse entonces un segundo componente en el capilar, en donde el segundo componente está separado del primer componente por la burbuja de aire. El primero y segundo componentes pueden ser mezclados aplicando presión hidrostática a ambos lados del arreglo capilar para colapsar la burbuja. El arreglo de capilares es monitorizado entonces en busca de un evento detectable resultante de la reacción o no reacción de los dos componentes.

En un ensayo de selección de enlazamiento, una muestra de interés puede ser introducida en un primer líquido marcado con una partícula detectable en un capilar de un arreglo de capilares, en donde el lumen del capilar está recubierto con un material de enlazamiento para enlazar la partícula deseable al lumen. El primer líquido puede ser removido entonces del tubo capilar, en donde la partícula detectable enlazada es mantenida dentro del capilar, y un segundo líquido puede ser introducido en el tubo capilar. El capilar es monitorizado entonces en busca de un evento detectable resultante de la reacción o no reacción de la partícula con el segundo líquido.

Arreglos o "Biochips"

Los ácidos nucleicos y/o polipéptidos de la invención pueden ser inmovilizados en o aplicados a un arreglo, por ejemplo, un "biochip". Los arreglos pueden ser utilizados para seleccionar o para monitorizar bibliotecas de composiciones (por ejemplo moléculas pequeñas, anticuerpos, ácidos nucleicos, etc.) por su capacidad para enlazarse o modular la actividad de un ácido nucleico o un polipéptido de la invención. Por ejemplo, en un aspecto de la invención, un parámetro monitorizado es la expresión transcripta de un gen de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. Uno o más, o todos los transcritos de una célula pueden ser medidos por hibridación de una muestra que comprende transcritos de la célula, o ácidos nucleicos, representativos de o complementarios a transcritos de una célula, por hibridación a ácidos nucleicos inmovilizados en un arreglo, o "biochips". Utilizando un "arreglo" de ácidos nucleicos en un microchip, algunos o todos los transcritos de una célula pueden ser cuantificados simultáneamente. Alternativamente, los arreglos que comprenden ácido nucleico genómico también pueden ser utilizados para determinar el genotipo de una cepa recientemente manipulada por lo métodos de la invención. Los arreglos de polipéptidos también pueden ser utilizados para cuantificar simultáneamente una pluralidad de proteínas. La presente invención puede ser puesta en práctica con cualquier "arreglo" conocido, también denominado como "microarreglo" o "arreglo de ácidos nucleicos" o "arreglo de polipéptidos" o "arreglo de anticuerpos" o "biochip" o variaciones de los mismos. Los arreglos son genéricamente una pluralidad de "puntos" o "elementos objetivos", comprendiendo cada elemento objetivo una cantidad definida de una o más moléculas biológicas, por ejemplo, oligonucleótidos, inmovilizadas sobre un área definida de una superficie de sustrato para enlazamiento específico a una molécula de muestra, por ejemplo, transcritos de ARNm.

Los términos "arreglo" o "microarreglo" o "biochip" o "chip", tal como se utilizan aquí son una pluralidad de elementos objetivo, comprendiendo cada elemento objetivo una cantidad definida de uno o más polipéptidos (incluyendo anticuerpos) o ácidos nucleicos inmovilizados sobre un área definida de una superficie de sustrato, tal como se discute en detalle más adelante.

Al practicar los métodos de la invención, cualquier arreglo y/o método conocido para hacer y utilizar arreglos puede ser incorporado en su totalidad o en parte, o variaciones de los mismos, como se describe por ejemplo, en las Patentes de los Estados Unidos Nos. 6,277,628; 6,277,489; 6,261,776; 6,258,606; 6,054,270; 6,048,695; 6,045,996; 6,022,963; 6,013,440; 5,965,452; 5,959,098; 5,856,174; 5,830,645; 5,770,456; 5,632,957; 5,556,752; 5,143,854; 5,807,522; 5,800,992; 5,744,305; 5,700,637; 5,556,752; 5,434,049; véase también, por ejemplo, WO 99/51773; WO 99/09217; WO 97/46313; WO 96/17958; véase también, por ejemplo, Johnston (1998) Curr. Biol. 8:R171-R174; Schummer (1997) Biotechniques 23:1087-1092; Kern (1997) Biotechniques 23:120-124; Solinas-Toldo (1997) Genes, Chromosomes & Cancer 20:399-407; Bowtell (1999) Nature Genetics Supp. 21:25-32. Véase también las Solicitudes de Patente de los Estados Unidos publicadas Nos. 20010018642; 20010019827; 20010016322; 20010014449; 20010014448; 20010012537; 20010008765.

Anticuerpos y métodos de selección basados en anticuerpos

La invención describe anticuerpos aislados, sintéticos o recombinantes que se enlazan específicamente a una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención. Estos anticuerpos pueden ser utilizados para aislar, identificar o cuantificar una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención o polipéptidos relacionados. Estos anticuerpos pueden ser utilizados para aislar otros polipéptidos dentro del alcance de la invención u otras transferasas relacionadas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas. Los anticuerpos pueden ser diseñados para enlazarse a un sitio activo de una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. Así, la invención describe métodos para inhibir transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas usando los anticuerpos tal como se describieron (véase la discusión más arriba concerniente a las aplicaciones para antitransferasa, por ejemplo, anti-transaminasa, por ejemplo, anti-d-aminoácido transferasa, y/o anti-oxidorreductasa, por ejemplo, anti-deshidrogenasa, por ejemplo, anti-d-aminoácido deshidrogenasa de composiciones de la invención).

La invención describe fragmentos de las enzimas de la invención incluyendo fragmentos inmunogénicos de un polipéptido de la invención. La invención provee composiciones que comprenden un polipéptido o péptido de la invención y adyuvantes o vehículos y similares.

Los anticuerpos pueden ser utilizados en inmunoprecipitación, tinción, columnas de inunoafinidad y similares. Si se desea, las secuencias de ácidos nucleicos que codifican antígenos específicos pueden ser generadas por inmunización seguida por aislamiento de polipéptido o ácidos nucleicos, amplificación o clonación e inmovilización de polipéptidos sobre un arreglo de la invención. Alternativamente, los métodos de la invención pueden ser utilizados

para modificar la estructura de un anticuerpo producido por una célula que va a ser modificada, por ejemplo, la afinidad de un anticuerpo puede ser incrementada o disminuida. Adicionalmente, la capacidad para hacer o modificar anticuerpos puede ser un genotipo manipulado en una célula por los métodos de la invención.

El término "anticuerpo" incluye un péptido o polipéptido derivado de, modelado a partir de, o codificado sustancialmente por un gen de inmunoglobulina o genes de inmunoglobulina o fragmentos de los mismos, capaces de enlazarse específicamente a un antígeno o epítipo, véase, por ejemplo, *Fundamental Immunology*, Third Edition, W.E. Paul, ed., Raven Press, N.Y. (1993); Wilson (1994) *J. Immunol. Methods* 175:267-273; Yarmush (1992) *J. Biochem. Biophys. Methods* 25:85-97. El término anticuerpo incluye porciones de enlazamiento a antígeno, esto es, "sitios de enlazamiento a antígeno" (por ejemplo, fragmentos, subsecuencias, regiones determinantes de la complementareidad, (CDRs)) que retienen la capacidad de enlazarse a antígenos, incluyendo (i) un fragmento Fab, un fragmento monovalente consistente de los dominios VL, VH, CL y CH1; (ii) un fragmento F(ab')₂, un fragmento bivalente que comprende dos fragmentos Fab enlazados por un puente de disulfuro en la región bisagra; (iii) un fragmento Fd consistente de los dominios VH y CH1; (iv) un fragmento Fv consistente de los dominios VL y VH de un brazo individual de un anticuerpo, (v) un fragmento dAb (Ward et al., (1989) *Nature* 341:544-546), el cual consiste de un dominio VH; y (vi) una región determinante de complementareidad aislada (CDR). Las cadenas de anticuerpos individuales también están incluidas por referencia al término "anticuerpo".

Los métodos de inmunización, producción y aislamiento de anticuerpos (policlonales y monoclonales) son conocidos para los experimentados en la técnica y están descritos en la literatura científica y de Patentes, véase por ejemplo, Coligan, *CURRENT PROTOCOLS IN IMMUNOLOGY*, Wiley/Greene, NY (1991); Stites (eds.) *BASIC AND CLINICAL IMMUNOLOGY* (7th ed.) Lange Medical Publications, Los Altos, CA ("Stites"); Goding, *MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE* (2d ed.) Academic Press, New York, NY (1986); Kohler (1975) *Nature* 256:495; Harlow (1988) *ANTIBODIES, A LABORATORY MANUAL*, Cold Spring Harbor Publications, New York. Los anticuerpos también pueden ser generados *in vitro*, por ejemplo, utilizando bibliotecas de despliegue de fagos que expresan sitios de enlazamientos de anticuerpos recombinantes, además de los métodos *in vivo* tradicionales usando animales. Véase, por ejemplo, Hoogenboom (1997) *Trends Biotechnol.* 15:62-70; Katz (1997) *Annu. Rev. Biophys. Biomol. Struct.* 26:27-45.

Los polipéptidos de la invención o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 o 150 aminoácidos consecutivos de los mismos, pueden ser utilizados para generar anticuerpos que se enlazan específicamente a los polipéptidos o fragmentos. Los anticuerpos resultantes pueden ser utilizados en procedimientos de cromatografía de inmunoafinidad para aislar o purificar el polipéptido o para determinar si el polipéptido está presente en una muestra biológica. En tales procedimientos, una preparación de proteínas, tal como un extracto, o una muestra biológica se pone en contacto con un anticuerpo capaz de enlazarse específicamente a uno de los polipéptidos de la invención, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 o 150 aminoácidos consecutivos de los mismos.

En procedimientos de inmunoafinidad, el anticuerpo es unido a un soporte sólido, tal como una perla u otra matriz de columna. La preparación de proteínas es puesta en contacto con el anticuerpo bajo condiciones en las cuales el anticuerpo se enlaza específicamente a uno de los polipéptidos de la invención, o a un fragmento del mismo. Después de un lavado para remover proteínas no específicamente enlazadas, los polipéptidos específicamente enlazados son sometidos a elución.

La capacidad de las proteínas en una muestra biológica para enlazarse al anticuerpo puede ser determinado utilizando cualquiera de una variedad de procedimientos familiares para los experimentados en la técnica. Por ejemplo, el enlazamiento puede ser determinado marcando el anticuerpo con un marcador detectable tal como un agente fluorescente, un marcador enzimático, o un radioisótopo. Alternativamente, el enlazamiento del anticuerpo a la muestra puede ser detectado utilizando un anticuerpo secundario que tiene tal marcador detectable sobre sí mismo. Ensayos particulares incluyen ensayos de ELISA, ensayos sándwich, radioinmunoensayos e inmunoprecipitaciones Western.

Anticuerpos policlonales generados contra los polipéptidos de la invención, o fragmentos que comprendan al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 o 150 o más aminoácidos consecutivos pueden ser obtenidos por lo tanto por inyección directa de los polipéptidos en un animal o por administración de los polipéptidos a un animal, por ejemplo, un no humano. El anticuerpo así obtenido se enlazará entonces al polipéptido mismo. De esta manera, incluso una secuencia que codifica solamente un fragmento del polipéptido puede ser usada para generar anticuerpos que pueden enlazarse al polipéptido nativo completo. Tales anticuerpos pueden ser utilizados entonces para aislar el polipéptido de células que expresan ese polipéptido.

Para la preparación de anticuerpos monoclonales puede utilizarse cualquier técnica que provea anticuerpos producidos por cultivos de líneas celulares continuos. Ejemplos incluyen la técnica del hibridoma (Kohler and Milstein, *Nature*, 256:495-497, 1975), la técnica del trioma, la técnica del hibridoma de células B humanas (Kozbor et al., *Immunology Today* 4:72, 1983) y la técnica de EBV-hibridoma (Cole, et al., 1985, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96).

Las técnicas descritas para la producción de anticuerpos de cadena sencilla (Patente de los Estados Unidos No. 4,946,778) pueden ser adaptadas para producir anticuerpos de cadena sencilla a los polipéptidos de la invención, o fragmentos que comprenden al menos 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 o 150 aminoácidos consecutivos de los mismos. Alternativamente, pueden usarse ratones transgénicos para expresar anticuerpos humanizados para estos polipéptidos o fragmentos de los mismos.

Los anticuerpos generados contra los polipéptidos de la invención, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 o 150 aminoácidos consecutivos de los mismos pueden ser utilizados en la selección para polipéptidos similares de otros organismos y muestras. En tales técnicas, los polipéptidos de los organismos son puestos en contacto con el anticuerpo y aquellos polipéptidos que se enlazan específicamente al anticuerpo son detectados. Cualquiera de los procedimientos descritos anteriormente puede ser utilizado para detectar el enlazamiento de anticuerpos. Un tal ensayo de selección está descrito en "Methods for Measuring Cellulase Activities", Methods in Enzymology, Vol 160, pp 87-116.

Kits

La invención describe kits que comprende las composiciones, por ejemplo, ácidos nucleicos, casetes de expresión, vectores, células, semillas o plantas u otras partes transgénicas, polipéptidos (por ejemplo, transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas) y/o anticuerpos de la invención. Los kits también pueden contener material de instrucción que enseñan las metodologías y procesamiento industrial, de investigación, médico, farmacéutico, de alimentos y piensos y suplementos para alimentos y piensos y otras aplicaciones y procesos de la invención, tal como se describen aquí.

Manipulación de células completas y medición de parámetros metabólicos

Los métodos de la invención describen evolución de células completas, o manipulación de células completas, de una célula para desarrollar una nueva cepa celular que tiene un nuevo fenotipo, por ejemplo, una nueva actividad de transferasa, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa, nuevas o modificadas, modificando la composición genética de la célula. La composición genética puede ser modificada por la adición a la célula de un ácido nucleico de la invención, por ejemplo, una secuencia de codificación para una enzima de la invención. Véase, por ejemplo, WO0229032; WO0196551.

Para detectar un nuevo fenotipo, se monitorizan al menos un nuevo parámetro metabólico de una célula modificada en la célula en un marco de tiempo de "tiempo real" o "en línea". En un aspecto, una pluralidad de células, tal como un cultivo celular, es monitorizada en "tiempo real" o "en línea". En un aspecto, una pluralidad de parámetros metabólicos es monitorizada en "tiempo real" o "en línea". Los parámetros metabólicos pueden ser monitorizados utilizando las actividades de las transferasas, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención.

El análisis de flujo metabólico (MFA) está basado en un marco bioquímico conocido. Una matriz metabólica linealmente independiente es construida con base en la ley de conservación de masas y en la hipótesis del estado de pseudobalance (PSSH) sobre los metabolitos intracelulares. En la práctica de los métodos de la invención, se establecen redes metabólicas, incluyendo:

- identidad de todos los sustratos, productos y metabolitos intermediarios de las rutas;
- identidad de todas las reacciones químicas que interconvierten los metabolitos de las rutas, la estequiometría de las reacciones de las rutas;
- identidad de todas las enzimas que catalizan las reacciones, la cinética de la reacción enzimática;
- las interacciones reguladoras entre componentes de la ruta, por ejemplo interacciones alostéricas, interacciones enzima-enzima, etc.,
- compartimentación intracelular de enzimas o cualquier otra organización supramolecular de las enzimas; y
- la presencia de cualquier gradiente de concentración de metabolitos, enzimas o moléculas efectoras o barreras de difusión a su movimiento.

Una vez que la red metabólica para una cepa dada es construida, puede introducirse la presentación matemática por nociones de matrices para estimar los flujos metabólicos intracelulares si los datos de metaboloma en línea están disponibles. El fenotipo metabólico se basa en los cambios de la red metabólica completa dentro de una célula. El fenotipo metabólico se basa en el cambio de la utilización de rutas con respecto a las condiciones ambientales, regulación genética, estado de desarrollo y el genotipo, etc. En un aspecto de los métodos descritos, después del

cálculo de MFA en línea, el comportamiento dinámico de las células, su fenotipo y otras propiedades son analizadas investigando la utilización de la ruta. Por ejemplo, si el suministro de glucosa es incrementado y el oxígeno disminuido durante la fermentación de levadura, la utilización de rutas respiratorias puede ser reducida y/o detenida, y la utilización de rutas fermentadoras dominará. El control del estado fisiológico de los cultivos celulares será posible después del análisis de ruta. Los métodos de la invención pueden ayudar a determinar cómo manipular la fermentación determinando como cambiar el suministro de sustrato, temperatura, uso de inductores, etc., para controlar el estado fisiológico de las células para moverse a lo largo de la dirección deseada. En la práctica de los métodos descritos, los resultados del MFA también pueden ser comparados con los datos del transcriptoma y proteoma para diseñar experimentos y protocolos para manipulación metabólica o mezcla genética, etc.

En la práctica de los métodos descritos, cualquier fenotipo modificado o nuevo puede ser conferido y detectado, incluyendo características nuevas o mejoradas en la célula. Puede monitorizarse cualquier aspecto del metabolismo o crecimiento.

Monitorización de la expresión de un transcrito de ARNm

En un aspecto de la invención, el fenotipo manipulado comprende el incremento o descenso de la expresión de un transcrito de ARNm (por ejemplo un mensaje de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa) o la generación de nuevos transcritos en una célula (por ejemplo, transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa). Esta expresión incrementada o disminuida puede ser seguida probando la presencia de una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención o por ensayos de actividad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. Los transcritos de ARNm o mensajes, también pueden ser detectados y cuantificados por cualquier método conocido en la técnica, incluyendo, por ejemplo, inmunoprecipitaciones de Northern, reacciones de amplificación cuantitativa, hibridación a arreglos, y similares. Las reacciones de amplificación cuantitativa incluyen, por ejemplo PCR cuantitativo, incluyendo, por ejemplo, reacción en cadena de polimerasa de transcripción reversa cuantitativa, o RT-PCR; RT-PCR cuantitativo en tiempo real, o "RT-PCR cinético en tiempo real" (véase, por ejemplo, Kreuzer (2001) Br. J. Haematol. 114:313-318; Xia (2001) Transplantation 72:907-914).

En un aspecto de la invención, el fenotipo manipulado es generado anulando la expresión de un gen homólogo. La secuencia de codificación del gen o uno o más elementos de control transcripcional pueden ser anulados, por ejemplo, promotores o potenciadores. Así, la expresión de un transcrito puede ser eliminada completamente o solamente disminuida.

En un aspecto de la invención, el fenotipo manipulado comprende el incremento de la expresión de un gen homólogo. Esto puede ser efectuado por la anulación de un elemento de control negativo, incluyendo un elemento regulador transcripcional que actúa en cis- o trans-, o mutagenizando un elemento de control positivo. Uno o más, o, todos los transcritos de una célula pueden ser medidos por hibridación de una muestra que comprende transcritos de la célula, o, ácidos nucleicos representativos de o complementarios a los transcritos de una célula, por hibridación a ácidos nucleicos inmovilizados sobre un arreglo.

Monitorización de la expresión de polipéptidos, péptidos y aminoácidos

En un aspecto de la invención, el fenotipo manipulado comprende el incremento o disminución de la expresión de un polipéptido (por ejemplo, a transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa) o generando nuevos polipéptidos en una célula. Esta expresión incrementada o disminuida puede ser seguida determinando la cantidad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa presente o por ensayos de actividad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa. Los polipéptidos, péptidos y aminoácidos también pueden ser detectados y cuantificados por cualquier método conocido en la técnica, incluyendo, por ejemplo resonancia magnética nuclear (RMN), espectrofotometría, radiografía (radiomarcación de proteína), electroforesis, electroforesis capilar, cromatografía líquida de alto rendimiento (HPLC), cromatografía de capa fina (TLC), cromatografía por hiperdifusión, diversos métodos inmunológicos, por ejemplo, inmunoprecipitación, inmunodifusión, inmunoelectroforesis, radioinmunoensayo (RIA), ensayos inmunsorbentes enlazados por enzimas (ELISA), ensayos inmunofluorescentes, electroforesis en gel (por ejemplo, SDS-PAGE), tinción con anticuerpos, selección de células activadas fluorescentes (FACS), espectrometría de masas con pirólisis, espectrometría infrarroja con transformadas de Fourier, espectrometría Raman, espectrometrías de masas con GC-MS, y LC-electroaspiración y electroaspiración cap con LC en tándem, y similares. También pueden seleccionarse bioactividades novedosas utilizando métodos, o variaciones de los mismos, descritos en la Patente de los Estados Unidos No. 6,057,103.

Adicionalmente, como se discute más abajo en detalle, uno o más, o, todos los polipéptidos de una célula vegetal pueden ser medidos utilizando un arreglo de proteínas.

Aplicaciones industriales, en energía, farmacéuticas, médicas, en procesamiento de alimentos y otras

- Los polipéptidos de la invención pueden ser utilizados en cualquier proceso industrial, agrícola, de alimentos y piensos y de procesamiento de suplementos de alimentos y piensos, farmacéutico (fármacos), médicos, investigación (laboratorio) y otros. La invención provee procesos industriales que utilizan enzimas de la invención, por ejemplo, en la industria farmacéutica o de suplementos nutritivos (dietarios), en la industria de la energía (por ejemplo, para hacer biocombustibles "limpios"), en las industrias de alimentos y piensos, por ejemplo en métodos para producir productos alimenticios y piensos y aditivos para alimentos y piensos. En un aspecto, la invención provee procesos que utilizan enzimas de la invención en la industria médica, por ejemplo, para hacer productos farmacéuticos (fármacos), precursores o intermediarios farmacéuticos (de fármacos), o auxiliares o suplementos dietarios, o suplementos y aditivos para alimentos. Además, la invención describe métodos para utilizar las enzimas de la invención en la producción de biocombustibles, incluyendo, por ejemplo, un bioalcohol, tal como bioetanol, biometanol, biobutanol o biopropanol, comprendiendo así una producción de combustible "limpio".
- Las transferasas de la invención, las cuales pueden tener actividad de d-aminoácido transferasa, por ejemplo, catalizar la conversión de ω -aminas a cetonas. Una enzima de la invención puede ser un catalizador altamente selectivo. Pueden catalizar reacciones con exquisitas estéreo-, regio- y quimioselectividades que no tienen paralelo en la química sintética convencional. Por ejemplo, una enzima de la invención puede tener una actividad de ω -transaminasa que puede comprender la catálisis de la conversión de ω -aminas quirales a cetonas.
- Las enzimas de la invención pueden ser notablemente versátiles. Por ejemplo, las enzimas d-aminoácido transferasa de la invención pueden ser "hechas a medida" (por ejemplo modificadas en secuencia, o modificadas de otra manera, por ejemplo, glicosiladas) para funcionar en solventes orgánicos, operar a pHs extremos (por ejemplo pHs altos y pHs bajos), temperaturas extremas (por ejemplo, temperaturas altas y temperaturas bajas), niveles de salinidad extremos (por ejemplo alta salinidad y baja salinidad) y catalizar reacciones con compuestos que están estructuralmente no relacionados con sus sustratos naturales fisiológicos.

- En un aspecto, las enzimas transaminasa de la invención se utilizan en procesos en reacciones de síntesis orgánica en la manufactura de medicamentos, pesticidas o intermediarios de los mismos. En una realización, las enzimas transaminasa y/o aminotransferasa de la invención se utilizan como objetivos para examinar el retorno de proteínas en respuesta a procesos patológicos o biológicos, por ejemplo, en daño/enfermedad del hígado o infarto del miocardio. En una reacción de transaminación de ejemplo utilizando una enzima de la invención, un grupo alfa-amino es transferido a un átomo de carbono alfa de un alfa-cetoglutarato generando el correspondiente alfa-cetoácido análogo del aminoácido.

Síntesis de aminoácidos naturales y no naturales

- Las enzimas de la invención pueden ser utilizadas para producir todos los aminoácidos naturales y diversos no naturales, incluyendo L- y/o D-aminoácidos, los cuales pueden ser utilizados, por ejemplo, en productos industriales o farmacéuticos (fármacos) y/o precursores y/o intermediarios farmacéuticos (de fármacos), tales como endulzantes, antibióticos, enzimas peptídicas y hormonas peptídicas. Por ejemplo, alitame (por ejemplo, ACLAME™), tirocidin A, antinomycin D, penicilina N y cefalosporina C, y ciclosporina A comprenden D-alanina y D-fenilalanina, D-valina, D-alfa amino valerato, D-alfa aminovalerato y D-alanina, respectivamente, son manufacturados utilizando una enzima de la invención. En un aspecto, una enzima de la invención se utiliza en la conversión de cefalosporina C a ácido glutaril-7- aminocefalosporánico (GL-7ACA), el cual es un agente químico farmacéutico altamente valioso para la síntesis de antibióticos cephem. En un método de ejemplo para producir ácido glutaril-7-aminocefalosporánico a partir de cefalosporina C utilizando una aminotransferasa de la invención, una enzima de la invención es agregada a una mezcla de sustrato que comprende cefalosporina C y un aceptor de D-amino, por ejemplo, un piruvato o un alfa-cetoglutarato, para formar ceto-GL-7ACA y D-alanina o D-glutamato, respectivamente. La mezcla se hace reaccionar entonces con peróxido de hidrógeno y se aísla el ácido glutaril-7-aminocefalosporánico como producto. Véase, por ejemplo, la Patente de los Estados Unidos No. 6,337,190.

- La invención provee enzimas y procesos para hacer aminoácidos naturales y no naturales (no naturales). Enzimas de esta invención pueden ser utilizadas con cualquier proceso de síntesis de aminoácidos naturales o no naturales conocido en la técnica; por ejemplo, en realizaciones alternativas, procesos de la invención, que utilizan enzimas de esta invención, sintetizan aminoácidos naturales o no naturales utilizando métodos como los descritos en las Patentes de los Estados Unidos Nos. (USPNs) 6,197,558; 6,180,374; 4,518,692 y 4,826,766. Por ejemplo, en un aspecto la invención provee un método que comprende hacer reaccionar un primer aminoácido, un cetoácido y una enzima transaminasa de esta invención bajo condiciones apropiadas para producir un segundo aminoácido y piruvato. Haciendo reaccionar el piruvato con acetolactato sintasa bajo condiciones apropiadas para producir un compuesto que no reacciona con la enzima transaminasa y separar el segundo aminoácido.

En un aspecto, las enzimas y procesos de la invención se utilizan para hacer tert-leucina, la cual a su vez puede ser utilizada en procesos de síntesis de fármacos/farmacéuticos, por ejemplo, la L-tert-leucina-2-piridilamida es un intermediario útil para la preparación de compuestos inhibidores de la matriz de metaloproteinasa, véase, por ejemplo, la Patente EP EP0822186.

5 Las enzimas y proceso de la invención pueden ser utilizados para ser aminoácidos “no naturales” tales como:

- ◆ B-aminoácido ($\beta 3$ y $\beta 2$)
- ◆ Homo-aminoácidos
- ◆ Aminoácidos cíclicos
- ◆ Aminoácidos aromáticos

10 ◆ N-metil aminoácidos

- ◆ Derivados Pro y Pyr
- ◆ Derivados de alanina sustituidos en 3
- ◆ Derivados de glicina
- ◆ Derivados de Phe y Tyr sustituidos en anillo

15 ◆ Aminoácidos de núcleo lineal

- ◆ Diaminoácidos (diaminas)
- ◆ Ornitina
- ◆ Norleucina
- ◆ Fenilselenocisteina (véase la solicitud de Patente de Estados Unidos publicación 20070238152)

20 ◆ Aminoácidos sustituidos (por ejemplo alaninas, fenilalaninas, tiroxinas, etc., sustituidas en orto, para o meta)

- ◆ DAB (ácido 2,4-diaminobutírico)

Aminoácidos alicíclicos (por ejemplo, ácido (cis)-3-aminobiciclo[2.2.1]heptano-2-carboxílico; ácido 1-amino-1-ciclobutanocarboxílico; ácido 1-aminociclohexanocarboxílico purum; ácido cis-2-amino-1-ciclopentanocarboxílico, y similares). Las enzimas y procesos de la invención pueden ser utilizados en conjunción con métodos para incorporación *in vivo* de aminoácidos no naturales, por ejemplo, como se describe en las Publicaciones de Solicitud de Patente de los Estados Unidos 20070117184; 20060234339; 20060233744; 20050272121; 20050250183; 20030082575.

25

Composiciones detergentes, desinfectantes y de limpieza

30 La invención describe composiciones de limpieza, por ejemplo, composiciones detergentes, desinfectantes o de limpieza (limpiadoras o de limpieza), por ejemplo, para limpiar telas, lavado de vajillas, lavandería, limpieza oral, limpieza de dentaduras y lentes de contacto, que comprende uno o más polipéptidos (por ejemplo, transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas) de la invención, y métodos para hacer y utilizar estas composiciones. La invención incorpora todos los métodos para hacer y utilizar composiciones detergentes, desinfectantes o limpiadoras, véase, por ejemplo, las Patentes de los Estados Unidos No. 6,413,928; 6,399,561; 6,365,561; 6,380,147. Las composiciones detergentes, desinfectante o limpiadoras pueden ser una composición acuosa en una o dos partes, una composición líquida no acuosa, un sólido fundido, una forma granular, una forma en partículas, una tableta comprimida, un gel y/o una pasta y una forma en suspensión. Las transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas de la invención también pueden ser utilizadas como un producto aditivo para detergentes, desinfectantes o limpiadores en una forma sólida o líquida. Tales productos aditivos pretenden complementar o potenciar el rendimiento de composiciones detergentes convencionales y pueden ser agregados en cualquier etapa del proceso de limpieza.

35

40 El contenido de enzima activa real depende del método de manufactura de una composición detergente, desinfectante o limpiadora y no es crítico, asumiendo que la solución detergente tiene la actividad enzimática

45

deseada. En un aspecto, la cantidad de transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa presente en la solución final varía desde de aproximadamente 0.001 mg a 0.5 mg por gramo de la composición detergente. La enzima particular escogida para uso en el proceso y productos de esta invención depende de las condiciones de utilidad final, incluyendo la forma física del producto, el pH de uso, la temperatura de uso, y los tipos de suciedad que van a ser degradados o alterados. La enzima puede ser escogida para proveer actividad y estabilidad óptimas para cualquier conjunto dado de condiciones de utilidad. En un aspecto, las transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas de la presente invención son activas en los rangos de pH que van desde aproximadamente 4 hasta aproximadamente 12 y en el rango de temperatura que va desde aproximadamente 20°C hasta aproximadamente 95°C. Los detergentes de la invención pueden comprender surfactantes catiónicos, semipolares no iónicos o zwitteriónicos, o mezclas de los mismo.

Las transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas de la invención pueden ser formuladas en detergentes, desinfectantes o limpiadores pulverizados y líquidos que tienen pH entre 4.0 y 12.0 a niveles de aproximadamente 0.01 hasta aproximadamente 5% (preferiblemente 0.1% a 0.5%), en peso. Estas composiciones detergentes, desinfectantes o limpiadoras también pueden incluir otras transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas y/o otras enzimas tales como xilanasas, celulasas, lipasas, esterases, proteasas, o endoglicosidasas, endo-beta-1,4-glucanasas, beta-glucanasas, endo-beta-1,3(4)-glucanasas, cutinasas, peroxidasas, catalasas, lacasas, amilasas, glucoamilasas, pectinasas, oxidorreductasas, reductasas, oxidasas, epimerasas, isomerasas, racemasas, fenoloxidasas, ligninasas, pululanases, arabinanasas, hemicelulasas, mananasas, xiloglucanasas, pectina acetil esterases, ramnogalacturonan acetil esterases, poligalacturonasas, ramnogalacturonasas, galactanasas, pectina liasas, pectina metilesterases, celobiohidrolasas y/o otras glutaminasas. Estas composiciones detergentes, desinfectantes o limpiadoras también pueden incluir pigmentos, colorantes, odorizantes, blanqueadores, reguladores, aglomerantes, "agentes potenciadores" de enzima (véase, por ejemplo, la Solicitud de Patente de los Estados Unidos No. 20030096394) y estabilizadores.

En un aspecto, la invención describe un método para limpiar o lavar un objeto que comprende poner en contacto el objeto con un polipéptido de la invención bajo condiciones suficientes para limpiar o lavar. Una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención puede ser incluida como aditivo detergente, desinfectante o limpiador. Una composición suavizadora para telas puede comprender una transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención.

Tratamiento de alimentos y procesamiento de alimentos

Las d-aminoácido transferasas de la invención tienen numerosas aplicaciones en la industria de procesamiento de alimentos. Por ejemplo, en un aspecto, las d-aminoácido transferasas son utilizadas para mejorar la extracción de aceite a partir de material vegetal rico en aceite, por ejemplo, semillas ricas en aceite, por ejemplo, aceite de semillas de soja de semillas de soja, aceite de oliva de olivas, aceite de colza de semillas de colza y/o aceite de girasol de semillas de girasol. En otro aspecto, las d-aminoácido transferasas de la invención pueden ser utilizadas para separación de componentes de materiales celulares vegetales.

Las transferasas, d-aminoácido transferasas de la invención pueden ser utilizadas en la preparación de zumos, jarabes, extractos y similares de frutas o vegetales para incrementar el rendimiento. Las d-aminoácido transferasas de la invención pueden ser utilizadas en el tratamiento enzimático de diversos materiales derivados de la pared celular de plantas o materiales residuales, por ejemplo, de la producción de cereales, granos, vino o zumos o residuos agrícolas tales como cortezas de vegetales, cortezas de judías, pulpa de remolacha de azúcar, pulpa de oliva, pulpa de patata y similares. Las d-aminoácido transferasas de la invención pueden ser utilizadas para modificar la consistencia y apariencia de frutas o vegetales procesados. Las d-aminoácido transferasas de la invención pueden ser utilizadas para tratar material vegetal para facilitar el procesamiento del material vegetal, incluyendo alimentos, facilitar la purificación o extracción de componentes de la planta.

En un aspecto las d-aminoácido transferasas de la invención se utilizan en aplicaciones en horneado, por ejemplo galletas, panes y tostadas. En un aspecto las d-aminoácido transferasas de la invención se utilizan como aditivos en el procesamiento de masas. En otro aspecto de la invención, las d-aminoácido transferasas de la invención también pueden ser utilizadas en el tratamiento de cualquier alimento o bebida o en el proceso de producción de alimentos o bebidas. En otro aspecto de la invención, las d-aminoácido transferasas de la invención pueden ser incluidas en cualquier composición de alimentos o bebidas.

Pienso y alimentos o aditivos para pienso y alimentos

La invención provee métodos para tratar piensos, alimentos, aditivos para piensos o alimentos, suplementos para alimentos o piensos, o auxiliares dietarios, utilizando las d-aminoácido transferasas de la invención, animales incluyendo mamíferos (por ejemplo humanos), aves, peces y similares. La invención provee a piensos, alimentos, aditivos para alimentos o piensos, suplementos para alimentos o piensos, o auxiliares dietarios que comprenden las d-aminoácido transferasas de la invención. En un aspecto, el tratamiento de piensos, alimentos, aditivos para alimentos o piensos, suplementos para alimentos o piensos, o ayudas dietarias utilizando las d-aminoácido transferasas de la invención puede ayudar en la disponibilidad de nutrientes, por ejemplo, almidón, proteínas, azúcares y similares en los piensos, alimentos, aditivos para alimentos o piensos, suplementos para alimentos o piensos, o ayudas dietarias.

Los piensos, alimentos, aditivos para alimentos o piensos, suplementos para alimentos o piensos, o ayudas dietarias de la invención pueden estar en una forma granulada, peletizada o en partículas, las cuales pueden ser recubiertas o no recubiertas. Alternativamente, los piensos, alimentos, aditivos para alimentos o piensos, suplementos para alimentos o piensos, o ayudas dietarias de la invención pueden ser una composición líquida estabilizada. Esta puede ser una suspensión acuosa o basada en aceite. Véase, por ejemplo, la Patente de los Estados Unidos No. 6,245,546.

En otro aspecto las d-aminoácido transferasas de la invención pueden ser suministradas expresando las enzimas directamente en cultivos transgénicos (por ejemplo, como plantas, semillas y similares transgénicos), tales como granos, cereales, maíz, soja, colza, lupino y similares. Como se discutió anteriormente, la invención describe plantas transgénicas, partes de plantas y células vegetales que comprenden una secuencia de ácidos nucleicos que codifican un polipéptido de la invención. En un aspecto, el ácido nucleico es expresado de tal manera que la transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención es producida en cantidades recuperables. Las transferasas, por ejemplo, transaminasas, por ejemplo, d-aminoácido transferasas, y/o oxidorreductasas, por ejemplo, deshidrogenasas, por ejemplo, d-aminoácido deshidrogenasas pueden ser recuperadas a partir de cualquier planta o parte de una planta. Alternativamente, la planta o parte de la planta que contiene el polipéptido recombinante puede ser utilizada como tal para mejorar la calidad de un alimento o pienso, por ejemplo, mejorar el valor nutricional, palatabilidad y propiedades reológicas o destruir cualquier factor antinutritivo.

Puede aplicarse un recubrimiento a los gránulos, pellas, partículas enzimáticas de la invención para muchos diferentes propósitos, tales como para agregar un sabor o un suplemento nutricional, para retardar la liberación de los nutrientes y enzimas en condiciones gástricas, o similares. O, el recubrimiento puede ser aplicado para alcanzar un objetivo funcional, por ejemplo, cuando sea deseable retardar la liberación de la enzima de las partículas de la matriz o para controlar las condiciones bajo las cuales la enzima va a ser liberada. La composición del material de recubrimiento puede ser tal que se deshaga selectivamente por un agente al cual es susceptible (por ejemplo calor, ácido o base, enzimas u otros agentes químicos). Alternativamente, pueden aplicarse consecutivamente dos o más recubrimientos susceptibles de diferentes tales agentes de ruptura a las partículas de la matriz.

La invención describe un proceso para preparar una matriz para la liberación de la enzima. De acuerdo con la invención, el proceso comprende proveer partículas plurales discretas de un sustrato basado en granos de un tamaño de partícula adecuado para uso como una matriz para liberación de enzimas, en donde las partículas comprenden una enzima transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa codificadas por una secuencia de aminoácidos de la invención o al menos 30 aminoácidos consecutivos de la misma. Preferiblemente, el proceso incluye compactar o comprimir las partículas de la matriz para liberación de enzimas en gránulos, lo cual puede ser logrado por peletización. El inhibidor de moldeo y el agente de cohesividad, cuando se usan, pueden ser agregados en cualquier momento adecuado, y pueden ser mezclados con el sustrato basado en granos en las proporciones deseadas antes de la peletización del sustrato basado en granos. El contenido de humedad en la pella alimentada al molino puede estar en los rangos definidos más arriba con respecto al contenido de humedad en el producto terminado, y puede ser aproximadamente 14-15%. En un aspecto, se agrega humedad a la materia prima en la forma de una preparación acuosa de la enzima para llevar la materia prima a este contenido de humedad. La temperatura en el molino de la pella puede llevarse hasta aproximadamente 82°C con vapor. La pella puede ser operada bajo cualquier condición que imparta operatividad suficiente a la materia prima para proveer las pellas. El proceso de peletización en sí mismo es un proceso efectivo en costes para remover agua de la composición que contiene la enzima.

En un aspecto, el molino para las pellas se opera con un molde de 1/8 pulgadas por 2 pulgadas a 100 libras/minuto de presión a 82°C para proveer las pellas, las cuales luego son trituradas en un triturador de molino de pellas para proveer partículas plurales discretas que tienen un tamaño de partícula capaz de pasar a través de un tamiz de malla 8 pero ser retenidas en un tamiz de malla 20.

Las d-aminoácido transferasas termoestables de la invención pueden ser utilizadas en las pellas de la invención. Pueden tener temperaturas altas óptimas y alta resistencia al calor de tal manera que puede alcanzarse una

reacción de la enzima a una temperatura hasta ahora no alcanzada. El gen que codifica las d-aminoácido transferasas de acuerdo con la presente invención (tal como se define en cualquiera de las secuencias de la invención) puede ser utilizado en la preparación de las d-aminoácido transferasas (por ejemplo utilizando GSSM como se describió aquí) con características diferentes a las de las transferasas, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención (en términos de pH óptimo, temperatura óptima, resistencia al calor, estabilidad a solventes, actividad específica, afinidad al sustrato, capacidad de secreción, rata de traducción, control de transcripción y similares).

Tratamiento de residuos

- La invención describe adicionalmente que las transferasas, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención pueden ser utilizadas en una variedad de otras aplicaciones industriales, por ejemplo, en el tratamiento de residuos. Por ejemplo, se describe un proceso de digestión de residuos sólidos utilizando transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención. Los métodos pueden comprender reducir la masa y volumen de residuos sólidos sustancialmente no tratados. El residuo sólido puede ser tratado con un proceso digestivo enzimático en la presencia de una solución enzimática (que incluye transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención) a una temperatura controlada. Esto da como resultado una reacción sin fermentación bacteriana apreciable a partir de microorganismos agregados. El residuo sólido es convertido en un residuo licuado y en algo de residuo sólido residual. El residuo licuado resultante puede ser separado de dicho residuo solidificado residual. Véase, por ejemplo, la Patente de los Estados Unidos No. 5,709, 796.

- La invención también describe que la transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención puede ser utilizada en cualquier proceso de tratamiento de residuos. Adicionalmente, la transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención puede ser incluida en cualquier composición para tratamiento de residuos.

Productos para el cuidado oral

- La invención describe productos para el cuidado oral que comprende transferasas, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa de la invención, incluyendo las mezclas enzimáticas o "cócteles" de la invención. Productos para cuidado oral de ejemplo incluyen pastas dentales, cremas dentales, geles o polvos dentales, odonticos, enjuagues bucales, formulaciones para enjuague pre o postcepillado, gomas de mascar, tabletas o golosinas. Véase, por ejemplo, la Patente de los Estados Unidos No. 6,264,925.

Conversión de biomasa y producción de biocombustibles

- La invención describe métodos y procesos para la conversión de biomasa o cualquier material orgánico en un combustible, por ejemplo, en un combustible, por ejemplo un biocombustible, tal como bioetanol biometanol, biopropanol y/o biobutanol y similares, utilizando enzimas de la invención, incluyendo las mezclas de enzimas o "cócteles" de la invención. Así, la invención describe combustibles, por ejemplo, biocombustibles, tales como bioetanoles, que comprenden un polipéptido de la invención, incluyendo las mezclas enzimáticas o "cócteles" de la invención, o un polipéptido codificado por un ácido nucleico de la invención. En aspectos alternativos, el combustible es derivado de un material vegetal, el cual opcionalmente comprende patatas, soja (colza), cebada, centeno, maíz, avena, trigo, remolachas o caña de azúcar y opcionalmente el combustible comprende un bioetanol o una mezcla gasolina-etanol.

- La invención describe métodos para hacer un combustible que comprende poner en contacto una composición de biomasa o cualquier material orgánico con un polipéptido de la invención, o un polipéptido codificado por un ácido nucleico de la invención, o cualquiera de las mezclas de "cócteles" o productos de manufactura de la invención. En realizaciones alternativas, la composición de biomasa comprende una planta, un producto vegetal o un derivado vegetal, y la planta o producto vegetal puede comprender plantas o productos vegetales de caña de azúcar, remolachas o remolachas de azúcar, trigo, maíz, soja, patata, arroz o cebada. En un aspecto, el combustible comprende un bioetanol o una mezcla gasolina-etanol, o un biopropanol o una mezcla gasolina-propanol o un biobutanol o una mezcla gasolina-butanol, o un biometanol o una mezcla gasolina-metanol, o un biodiesel, o una mezcla gasolina-biodiesel, o cualquier combinación de los mismos.

- La invención describe métodos para hacer bioetanol, biobutanol, biometanol y/o un biopropanol que comprende poner en contacto una composición de biomasa o cualquier material orgánico con un polipéptido de la invención, o

un polipéptido codificado por un ácido nucleico de la invención, o cualquiera de las mezclas o "cócteles" o productos de manufactura de la invención. En realizaciones alternativas, la composición de biomasa comprende una planta, un producto vegetal o un derivado vegetal, y la planta o producto vegetal puede comprender plantas o productos vegetales de caña de azúcar, remolachas o remolachas de azúcar, trigo, maíz, soja, patata, arroz o cebada. En realizaciones alternativas, el material orgánico o biomasa es derivado de cultivos agrícolas (por ejemplo trigo, cebada, patatas, pasto de agujas, madera de álamo), es un subproducto de una producción de alimentos o piensos, es un producto residual lignocelulósico, o es un producto vegetal o un papel residual o producto de papel residual, y opcionalmente el residuo vegetal comprende tallos, hojas, cáscaras, cortezas, maíz o mazorcas de maíz, cubiertas de mazorca, fibras de maíz, heno, paja (por ejemplo paja de arroz o paja de trigo), bagazo de caña de azúcar, pulpa de remolacha de azúcar, pulpa cítrica, cáscaras de cítricos, madera, cortes de madera, trozos de madera, pulpa de madera, residuos de pulpa, residuos de madera, aserrín y polvo de madera, residuos y desechos de construcción y/o demolición (por ejemplo madera, cepillados de madera y aserrín) y opcionalmente el residuo de papel comprende papel de fotocopia descartado o usado, papel de impresora de ordenadores, papel de notas, papel de cuadernos, papel de máquina de escribir, periódicos, revistas, cartón y materiales de empaque basados en papel y materiales de papel reciclados. Además, residuos urbanos, por ejemplo, la fracción de papel de los residuos sólidos municipales, residuos de madera municipales, y residuos verdes municipales pueden ser usados.

La invención describe composiciones (incluyendo productos de manufactura, ensambles enzimáticos o "cócteles") que comprenden una mezcla (o "cóctel") de enzimas transferasa, por ejemplo, transaminasa, por ejemplo, d-aminoácido transferasa, y/o oxidorreductasa, por ejemplo, deshidrogenasa, por ejemplo, d-aminoácido deshidrogenasa.

La invención provee células y/o organismos que expresan enzimas de la invención (por ejemplo, en donde las células de los organismos comprenden como ácidos nucleicos heterólogos, una secuencia de esta invención) para su participación en ciclos químicos que involucran conversión de biomasa natural (por ejemplo plantas). Alternativamente, el polipéptido de la invención puede ser expresado en el material vegetal de biomasa o materia prima misma.

Los métodos tal como son descritos incluyen también tomar el material de biomasa convertido (por ejemplo lignocelulósico) (procesado por enzimas de la invención) y convertirlo en un combustible (por ejemplo un biocombustible tal como un bioetanol, biobutanol, biometanol, un biopropanol, o un biodiesel) por fermentación y/o por síntesis química. En un aspecto, los azúcares producidos son fermentados y/o los productos no fermentables son gasificados.

Los métodos descritos también incluyen convertir algas, aceites vegetales vírgenes, aceites vegetales residuales, grasas y sebos animales (por ejemplo sebo, manteca y grasa amarilla), o aguas negras, utilizando enzimas de la invención, y convirtiéndolas en un combustible (por ejemplo un bioalcohol, por ejemplo, un bioetanol, biometanol, biobutanol o biopropanol o biodiesel) por fermentación y/o por síntesis o conversión química.

Las enzimas de la invención (incluyendo, por ejemplo, organismos, tales como microorganismos, por ejemplo, hongos, levaduras o bacterias, y plantas y células vegetales y partes vegetales, por ejemplo semillas, que hacen y en algunos aspectos secretan enzimas recombinantes de la invención) pueden ser usadas en o incluidas/integradas en cualquier etapa de cualquier proceso de conversión de materia orgánica/biomasa, por ejemplo, en cualquier etapa, varias etapas, o incluidos en todas las etapas, o todos de los siguientes métodos de procesos de conversión de biomasa, o en todas de estas alternativas de biocombustible:

Combustión directa: La quema de material por calor directo y es la tecnología de biomasa más simple; puede ser muy económica si la fuente de biomasa es cercana.

1 Pirólisis: Es la degradación térmica de la biomasa por calor en la ausencia de oxígeno. En un aspecto, la biomasa es calentada a una temperatura de entre aproximadamente 800 y 1400 grados Fahrenheit, pero no se introduce oxígeno para soportar la combustión dando como resultado la creación de gas, aceite combustible y carbón.

2 Gasificación: La biomasa puede ser utilizada para producir metano a través de calentamiento o digestión anaeróbica. El Syngas, una mezcla de monóxido de carbono e hidrógeno, puede ser derivados de la biomasa.

Gas de relleno de tierra: Es generado por la degradación (digestión anaeróbica) de residuos enterrados en terrenos. Cuando el residuo orgánico se descompone, genera gas que consiste de aproximadamente 50% de metano, el componente principal del gas natural.

Digestión anaeróbica: Convierte la materia orgánica en una mezcla de metano, el componente principal del gas natural, y dióxido de carbono. En un aspecto, la biomasa tal como aguas residuales (aguas negras), estiércol, o residuos del procesamiento de alimentos, se mezcla con agua y se alimenta en un tanque de digestión sin aire.

Fermentación

5 Fermentación alcohólica: Alcohol combustible se produce convirtiendo masa celulósica y/o almidón en azúcar, fermentando el azúcar alcohol, luego separando la mezcla de alcohol agua por destilación. Las materias primas tales como cultivos dedicados (por ejemplo, trigo, cebada, patatas, pasto, madera de álamo) residuos y desperdicios agrícolas (por ejemplo paja de arroz, cáscara de maíz, paja de trigo, bagazo de azúcar de caña, cáscaras de arroz, fibra de maíz, pulpa de remolacha de azúcar, pulpa de cítricos y pieles de cítricos), residuos forestales (por ejemplo madera dura y cepillados de madera, residuos de madera blanda de troncos y operaciones de corte de madera, aserrín, y polvo de madera) residuos urbanos (por ejemplo la fracción de papel de residuos sólidos municipales, residuos de madera municipales, residuos verdes municipales), desperdicios de madera (por ejemplo residuos de molienda de madera, residuos de molienda de pulpa, residuos de construcción, residuos de demolición, cepillados de madera y aserrín), y papel residual u otros materiales que contienen azúcar, almidón y/o celulosa pueden ser convertidos en azúcares y luego a alcohol por fermentación con levadura. Alternativamente, los materiales que contienen azúcares pueden ser convertidos directamente en alcohol por fermentación.

15 Transesterificación: Una reacción de ejemplo para convertir aceite a biodiesel es llamada transesterificación. El proceso de transesterificación hace reaccionar un alcohol (como metanol) con los aceites triglicéridos contenidos en aceites vegetales, grasas animales, o grasas recicladas, formando ésteres de alquilo de ácidos grasos (biodiesel) y glicerina. La reacción requiere calor y un catalizador básico fuerte, tal como hidróxido de sodio o hidróxido de potasio.

20 Biodiesel: El biodiesel es una mezcla de ésteres de alquilo de ácidos grasos hechos a partir de aceites vegetales, grasas animales o grasas recicladas. El biodiesel puede ser utilizado como combustible para vehículos en su forma pura, pero usualmente es utilizado como un aditivo diesel de petróleo para reducir niveles de partículas, monóxido de carbono, hidrocarburos y tóxicos en el aire de vehículos alimentados con diesel.

Hidrólisis: Incluye la hidrólisis de un compuesto, por ejemplo, una biomasa, tal como un material lignocelulósico, catalizado utilizando una enzima de la presente invención.

25 Cogeneración: Es la producción simultánea de más de una forma de energía utilizando un combustible e instalación sencillos. En un aspecto, la cogeneración de biomasa tiene más potencial de crecimiento que la generación de biomasa sola porque la cogeneración produce tanto calor como electricidad.

30 Las enzimas de la invención también pueden ser utilizadas en la refinación de glicerina. El subproducto glicerina contiene catalizadores y jabones que no han reaccionado que son neutralizados con un ácido. Se retiran el agua y el alcohol para producir 50% a 80% de glicerina cruda. Los contaminantes restantes incluyen grasas y aceites sin reaccionar, los cuales pueden ser procesados utilizando los polipéptidos de la invención. En una planta de biodiesel grande de la invención, la glicerina puede ser purificada adicionalmente, por ejemplo, hasta 99% o mayor pureza para las industrias farmacéutica y cosmética.

Biocombustibles como líquido o una gasolina gaseosa

35 La invención describe biocombustibles y combustibles sintéticos en la forma de un gas, una gasolina, por ejemplo un syngas. En un aspecto, los métodos de la invención que comprenden el uso de enzimas de la invención para ciclos químicos para la conversión de biomasa natural, por ejemplo, para la hidrólisis de una biomasa para hacer un biocombustible, por ejemplo, un bioetanol, biopropanol, biobutanol o biometanol, o un combustible sintético, en la forma de un líquido o un gas, tal como un "syngas".

40 Por ejemplo, la invención describe métodos para hacer gases biocombustibles y combustibles de gases sintéticos ("syngas") que comprenden un bioetanol, biopropanol, biobutanol y/o un biometanol fabricado utilizando un polipéptido de la invención, o fabricado utilizando un método de la invención; y en un aspecto de este gas biocombustible de la invención es mezclado con un gas natural (también puede ser producido a partir de biomasa), por ejemplo, un gas combustible basado en hidrógeno o un hidrocarburo. En un aspecto, la invención describe métodos para procesar biomasa hasta un combustible sintético, por ejemplo, un syngas, tal como un syngas producido a partir de una biomasa por gasificación. En un aspecto, la invención describe métodos para hacer un gas de etanol, propanol, butanol y/o metanol a partir de caña de azúcar, por ejemplo un bagazo. En un aspecto, este combustible o gas es utilizado como combustible motor, por ejemplo, como un combustible para automóvil, camión, aeroplano, bote, motor pequeño, etc. En un aspecto, la invención provee métodos para hacer un etanol, propanol, butanol y/o metanol a partir de una planta, por ejemplo, maíz, o un producto vegetal, por ejemplo, heno o paja (por ejemplo, paja de arroz o paja de trigo, o cualquiera de los residuos secos de cualquier planta de cereal), o un producto residual agrícola.

55 En un aspecto, el etanol, propanol, butanol y/o metanol hechos utilizando un método de la composición pueden usarse como un aditivo para combustible (por ejemplo, una gasolina) (por ejemplo un oxigenador) o en un uso directo como combustible. Por ejemplo, un etanol, propanol, butanol y/o metanol, incluyendo un combustible, hecho por un método de la invención puede ser mezclado con etil butil terciario éter (ETBE), o una mezcla de ETBE tal como ETBE que contiene 47% de etanol como biocombustible, o con MTBE (metil butil terciario éter). En otro

aspecto, un etanol, propanol, butanol y/o metanol, incluyendo un combustible, hecho por un método de la invención puede ser mezclado con:

Nombre IUPAC	Nombre común
but-1-eno	α -butileno
cis-but-2-eno	cis- β -butileno
trans-but-2-eno	trans- β -butileno
2-metilpropano	isobutileno

5 Un butanol y/o etanol hechos por un método de la invención (por ejemplo, utilizando una enzima de la invención) puede ser procesado adicionalmente utilizando fermentación "A.B.E" (Acetona, Butanol, Etanol); en un aspecto, siendo el butanol el único producto líquido. En un aspecto, este butanol y/o etanol es quemado "directamente" en motores de gasolina existentes (sin modificación al motor o automóvil), produce más energía y es menos corrosivo y menos soluble en agua que el etanol, y puede ser distribuido a través de infraestructuras ya existentes.

10 En un aspecto, uno, varios o todos de estos alcoholes se hace mediante un proceso que utiliza una enzima de la invención, y el proceso puede comprender adicionalmente una tecnología de biomasa a líquido, por ejemplo, un proceso de gasificación para producir syngas seguido por síntesis catalítica, o por una conversión de biomasa a un combustible de alcohol mixto.

15 La invención también describe procesos que comprenden el uso de una enzima de la invención que incorpora (o, incorporado en) procesos "gas a líquido", o GTL; o "carbón a líquido" o CTL; o "biomasa a líquidos" o BTL; o "arenas oleosas a líquidos" o OTL; y en un aspecto estos procesos se utilizan para hacer combustibles sintéticos. En un aspecto, uno de estos procesos de la invención comprende hacer un biocombustible (por ejemplo un syncombustible) a partir de biomasa utilizando, por ejemplo, el así llamado proceso "Fischer Tropsch" (una reacción química catalizada en la cual el monóxido de carbono y el hidrógeno son convertidos en hidrocarburos líquidos de diversa formas; los catalizadores típicos usados están basados en hierro y cobalto; el propósito principal de este proceso es producir un sustituto sintético del petróleo para uso como aceite de lubricación sintético o como combustible sintético). En un aspecto, este biocombustible sintético puede contener oxígeno y puede ser utilizado como aditivo en diesel y petróleo de alta calidad.

25 En aspectos alternativos, los procesos tal como se describen utilizan diversos pretratamientos, los cuales pueden ser agrupados en tres categorías: físicos, químicos y múltiples (físicos + químicos). Puede utilizarse cualquier agente químico como agente de pretratamiento, por ejemplo, ácidos, álcalis, gases, solventes de celulosa, alcoholes, agentes oxidantes y agentes reductores. Entre estos agentes químicos, los álcalis son los agentes de pretratamiento más populares porque son relativamente no costosos y dan como resultado menos degradación de la celulosa. Los álcalis más comunes hidróxido de sodio y caliza también pueden ser utilizados como agentes de pretratamiento. Aunque el hidróxido de sodio incrementa significativamente la digestibilidad de la biomasa, es difícil de reciclar, es relativamente costoso, y es peligroso en su manejo. En contraste, la caliza tiene muchas ventajas: es seguro y muy poco costoso, y puede ser recuperado por carbonatación del agua de lavado con dióxido de carbono.

35 En un aspecto, la invención describe un biocombustible, por ejemplo, un biogás, producido por el proceso de digestión anaeróbica de material orgánico por anaerobios, en donde el proceso comprende el uso de una enzima de la invención o un método de la invención. Este biocombustible, por ejemplo, un biogás, puede ser producido bien sea a partir de materiales residuales biodegradables o mediante el uso de cultivos energéticos alimentados en digestores anaeróbicos para complementar los rendimientos de gas. El producto sólido, digerido, también puede ser utilizado como biocombustible.

40 La invención describe métodos para hacer aceites producidos biológicamente, incluyendo aceites crudos, y gases que pueden ser utilizados en motores diesel, en donde el proceso comprende el uso de una enzima de la invención o un método de la invención. En un aspecto, estos métodos pueden refinar petróleo, por ejemplo, aceites crudos en queroseno, petróleo, diesel y otras fracciones.

La invención describe métodos (que utilizan una enzima de la invención o un método de la invención) para hacer aceites producidos biológicamente a partir de:

- Aceite vegetal directo (SVO).

- Aceite vegetal residual (WVO) – aceites y grasas de cocina residuales producidos en cantidad principalmente por cocinas comerciales.

- Biodiesel obtenido a partir de la transesterificación de grasas animales y aceites vegetales, utilizable directamente en motores de petróleo diesel.

5 ● Aceite crudo derivado biológicamente, junto con biogás y sólidos de carbono a través de despolimerización térmica de materiales orgánicos complejos incluyendo materiales no basados en aceites; por ejemplo, producto residuales tales como neumáticos viejos, offal, madera y plástico.

10 ● Aceite de pirólisis; el cual puede ser producido a partir de biomasa, residuos de madera, etc., utilizando calor solamente en el proceso de pirólisis instantánea (el aceite puede tener que ser tratado antes de utilizarlo en sistemas de combustibles convencionales o en motores de combustión interna).

- Madera, carbón y estiércol.

La invención será descrita adicionalmente con referencia a los siguientes ejemplos; sin embargo debe entenderse que la invención no está limitada a tales ejemplos.

Ejemplos

15 Las aminotransferasas y oxidoreductasas descritas aquí fueron obtenidas utilizando una estrategia de selección. En esta estrategia de selección, se construyeron bibliotecas de ADN ambientales en cepas anfitrionas bacterianas que exhiben auxotrofia al L-triptófano. Los clones de la biblioteca fueron inoculados en medios que contenían D-triptófano (pero que carecen de L-triptófano). Los únicos clones que pudieron crecer son aquellos que expresaban un gen en uno de los fragmentos de ADN ambiental discretos que codificaban una enzima activa sobre el D-triptófano. Por ejemplo, se identificaron clones que expresaban una triptófano racemasa activa y fueron capaces de convertir el D-triptófano en L-triptófano. Adicionalmente, se identificaron clones que expresaban una oxidoreductasa (tal como una aminoácido oxidasa o una deshidrogenasa) que pudieron convertir el D-triptófano en un intermediario que la célula anfitriona pudo, a su vez, convertir en L-triptófano. En el caso de oxidoreductasas tales como aminoácido oxidasas y deshidrogenasas, un tal intermediario es indol-3-piruvato.

25 Los ejemplos en la Parte A describe la metodología usada para la caracterización inicial del candidato DAT y los ácidos nucleicos de oxidoreductasa y los polipéptidos codificados. La caracterización adicional de los ácidos nucleicos y polipéptidos particulares está descrita en la Parte B.

Parte A

Ejemplo 1 – Procedimientos de crecimiento y ensayo # 1

30 Preparación de la enzima

Se utilizaron reservas de glicerol para inocular matraces que contenían 50 mL de medio LB con el antibiótico apropiado. Los cultivos de partida fueron cultivados durante la noche a 37°C con agitación a 230 rpm y se verificó el OD_{600nm}. El cultivo de inicio fue utilizado para inocular 400 ml hasta OD_{600nm} de 0.05. El cultivo fue incubado a 37°C con agitación a 230 rpm.

35 Los cultivos fueron inducidos con IPTG 1 mM cuando la OD_{600nm} estaba entre 0.5 y 0.8 y fueron incubados a 30°C y 230 rpm durante la noche. Los cultivos fueron recolectados pelotizando las células por centrifugación a 4000 rpm durante 15 minutos. El sobrenadante fue vertido y la pella fue bien congelada para uso posterior y resuspendida en 20 mL de regulador de fosfato de sodio 50 mM (pH 7.5) suplementado con 26 U/mL de seADN y sometido a lisis utilizando un microfluidizador (Microfluidics Corporation, Newton, MA) según las instrucciones del fabricante. El lisado clarificado fue recolectado y centrifugado a 11,000 rpm durante 30 minutos. El sobrenadante fue recolectado en un tubo limpio y filtrado a través de un filtro de 0.2 µm. Se colocaron alícuotas de 5 mL del lisado clarificado en un vial y se liofilizaron utilizando un liofilizador (Virtis Company, Gardiner, NY) según las instrucciones del fabricante. Aproximadamente 1 mL del lisado clarificado fue retenido para la cuantificación de proteína utilizando el Bio-Rad Protein Assay Reagent (Bio-Rad, Hercules, CA) y análisis SDS-PAGE. Luego se calculó la cantidad de proteína en cada muestra liofilizada.

Ensayo de actividad

Las enzimas para los ensayos de actividad fueron preparadas en fosfato de sodio 50 mM pH 7.5. Los ensayos de DAT fueron llevados a cabo usualmente utilizando aproximadamente 1 mg/ mL de proteína total.

Ensayo DAT utilizando sustrato de RR-monatina

5 RR-monatina 25 mM, sal de sodio del ácido pirúvico 25 mM, PLP 0.08 mM, fosfato de sodio 90 mM pH 8.0 y 0.8 mg/mL de DAT (proteína total) preparados como se describió anteriormente (bajo la sección "preparación de enzimas") fueron combinados e incubados a 30°C a 300 rpm. En diversos puntos del tiempo (en general 0, 2, 4 y 24 horas), 50 µL del producto de reacción fueron transferidos a 150 mL de acetonitrilo enfriado con hielo, y la muestra fue sometida a vórtex durante 30 segundos. Las muestras fueron centrifugadas a 13,200 rpm durante 10 minutos a 4°C, y el sobrenadante fue pasado a través de un filtro de 0.45 µm. El filtrado fue diluido 10 veces en metanol, y las muestras fueron analizadas por cromatografía líquida/espectrometría de masas/espectrometría de masas (LC-MS/MS) para monitorizar la D-alanina formada (descrito en este ejemplo bajo la sección "método LC/MS/MS para detectar D-alanina o R,R-monatina").

10 Ensayo de DAT utilizando sustrato de triptófano

15 D-triptófano 10 mM, sal de sodio de ácido pirúvico 25 mM, PLP 0.08 mM, fosfato de sodio 90 mM pH 8.0 y 0.8 mg/mL de DAT (proteína total) preparada como se describió anteriormente (bajo la sección "preparación de enzimas") fueron combinados e incubados a 30°C y 300 rpm. En diversos puntos del tiempo (en general 0, 2, 4 y 24 horas), se transfirieron 50 µL del producto de reacción a 150 mL de acetonitrilo enfriado con hielo, se sometió a vórtex durante 30 segundos y se centrifugó a 13,200 rpm durante 10 minutos a 4°C. El sobrenadante fue pasado a través de un filtro de 0.45 µm y el filtrado fue diluido 10 veces en metanol. Las muestras fueron analizadas por LC/MS/MS para monitorizar la D-alanina formada (descrito en este ejemplo bajo la sección "método LC/MS/MS para detectar D-alanina o R,R-monatina").

Método de LC/MS/MS para detectar D-alanina o R,R-monatina

20 El análisis por cromatografía líquida/espectrometría de masas/espectrometría de masas (LC-MS /MS) fue logrado inyectando muestras de placas de 96 pozos utilizando un automuestreador CTCPal (LEAP Technologies, Carrboro, NC) en una mezcla 30/70 de H₂O/Acn (0.1% de ácido fórmico) suministrada por bombas LC-10ADvp (Shimadzu, Kyoto, Japón) a 1.0 mL/minuto a través de una columna Zorbax ECLIPSE XDB-C8™ (2.1 x 50 mm) y en el espectrómetro de masas triple cuádruple API4000 TURBOION-SPRAY™ (Applied Biosystems, Foster City, CA).

25 La aspersión de iones y la monitorización de reacción múltiple (MRM) fueron llevadas a cabo para los analitos de interés en el modo de ión positivo. Alanina: iones progenitores/hijos: 90.12/44.25 monatina: iones progenitores/hijos: 293.11/130.15.

Ejemplo 2 – Actividad de DAT utilizando procedimientos de ensayo # 1

30 El vector pSE420-cHis es un derivado de pSE420 (Invitrogen, Carlsbad, CA). Para pSE420-cHis, el vector fue cortado con NcoI y Hind III, y ligado con C-His. C-His: 5'-CCA TGG GAG GAT CCA GAT CTC ATC ACC ATC ACC ATC ACT AAG CTT (SEQ ID NO:977). La expresión de la etiqueta His en este vector depende de la selección del anfitrión y del codón de detención. Esto es, si se usa un codón de detención TAG y un anfitrión supE, se expresa la etiqueta His; si se utilizan un codón de detención TAG y un anfitrión no supE, no se expresa la etiqueta His. A menos que se indique otra cosa, la etiqueta His no fue expresada en estos experimentos.

35 Los subclones DAT estuvieron en el vector pSE420-cHis /anfitrión E.coli HMS 174 (Novagen, San Diego, CA) con la excepción de los siguientes subclones: SEQ ID NO:930, 932, 936 estaban en el pET101 D-Topo vector/ BL21Star(DE3) anfitrión (Invitrogen, Carlsbad, CA); SEQ ID NO:934 estaba en el pET101 D-Topo vector/ BL21 CODON PLUSRIL™ anfitrión (Stratagene, La Jolla, CA); SEQ ID NO:938, 942, 944, 946 estaban en el pSE420 vector / XL1BLUE™ anfitrión (Stratagene, La Jolla, CA); SEQ ID NO:940, 948, 950, 962 and 966 were in the pSE420-c-His vector / XL1BLUE™ anfitrión (Stratagene, La Jolla, CA); and SEQ ID NO:928 estaba en el pQET1 vector / M15pREP4 anfitrión (pQET1 descrito en las Patentes de los Estados Unidos Nos. 5,814,473 y 6,074,860; M15pREP4 de Qiagen; Valencia, CA).

45 Los subclones fueron cultivados, sometidos a lisis y liofilizados de acuerdo con los procedimientos descritos en el Ejemplo 1. Las muestras fueron probadas en cuanto a la actividad sobre la R,R-monatina así como D-triptófano (tal como se describió en el Ejemplo 1). Para el ensayo de DAT con la monatina, el DAT fue incubado con R,R-monatina 25 mM, sal de sodio de ácido pirúvico 25 mM y PLP 0.08 mM (pH 8) a 30°C. Para el ensayo del D-triptófano, los DAT fueron incubados con D-triptófano 10 mM, sal de sodio del ácido pirúvico 25 mM y PLP 0.08 mM (pH 8) a 30°C. Todos los DAT fueron cargados a 0.8 mg/mL de proteína total en ambos ensayos,

50 En puntos de tiempo indicados, se agregaron 50 µL del producto de reacción a 150 µL de acetonitrilo enfriado con hielo. Las muestras fueron sometidas a vórtex durante 30 segundos y el sobrenadante fue diluido entonces 10 veces en metanol. Las muestras fueron analizadas por LC/MS/MS (tal como se describe en el Ejemplo 1) para monitorizar la D-alanina formada. Las tablas a continuación muestran la actividad de la D-aminotransferasa sobre ambos sustratos.

Tabla 4. Actividad de subclones de D-aminotransferasa sobre R,R-monatina y D-triptófano

SEQ ID NO:	Actividad sobre R,R-monatina	Actividad sobre D- triptófano	Expresión relativa
	µg/mL de D-alanina formados a la hora indicada	µg/mL de D-alanina formada a la hora indicada	
928	30@24hr	NT	+
938	122@24hr	NT	++
940	5@24hr	NT	ND
942	12@24hr	NT	ND
944	75@24hr	NT	ND
946	39@24hr	NT	ND
948	200@0.5hr	441@0.5hr	ND
950	75@0.5hr	452@0.5hr	ND
962	NT	NT	+
964	7@24hr	ND@24hr	++
966	NT	NT	++
968	6.7@24hr	52@24hr	+++
886 /expresado en células XL1Blue)	NT	NT	++
886 (expresado en células E. coli HMS174)	15.4@24hr	143@24hr	+++
888 (expresado en células XL1Blue)	NT	NT	++
888 (expresado en células HMS174)	7@24hr	317@24hr	+++
890 (expresado en células XL1 Blue)	NT	NT	++
890 (expresado en células HMS174)	54@24hr	278@24hr	+++
892 (expresado en células XL1 Blue)	NT	NT	+
892 (expresado en células HMS174)	113@24hr	<5@24hr	++
894 (expresado en células XL1 Blue)	NT	NT	ND
894 (expresado en células HMS174)	16@24hr	116@24hr	+
866	28@24hr	NT	+++

SEQ ID NO:	Actividad sobre R,R-monatina	Actividad sobre D- triptófano	Expresión relativa
	µg/mL de D-alanina formados a la hora indicada	µg/mL de D-alanina formada a la hora indicada	
868	<1@24hr	NT	+++
970	10.8@24hr	NT	+
870	123.5@24hr	NNTT	+++
872	62.3@24hr	NT	+++
874	46.5@hr	NT	+++
876	44@24hr	NT	++
878	37@24	NT	+++
972	<5@24	NT	+
880	72.4@24hr	79.6@24hr	+
882	158.8@24hr	344@2hr	+++
884	290@24hr	363@2hr	++
896	54@24hr	450@2hr	+++
898	466@24hr	300@24hr	+
900	135@24hr	154@24hr	+
902	280@24hr	130@24hr	++
904	170@24hr	140@24hr	+
906	700@24hr	500@24hr	+++
908	55@24hr	45@24hr	Insoluble
910	384@24hr	240@24hr	+++
912	NT	NT	ND
914	NT	NT	ND
916	NT	NT	ND
918	NT	NT	ND
920	NT	NT	ND
922	NT	NT	ND
924	NT	NT	ND
926	NT	NT	ND
974 (expresado en células E. coli HMS174)	NT	NT	ND

SEQ ID NO:	Actividad sobre R,R-monatina	Actividad sobre D- triptófano	Expresión relativa
	µg/mL de D-alanina formados a la hora indicada	µg/mL de D-alanina formada a la hora indicada	
974 (expresado en células XL1 Blue)	NT	NT	ND
930	NT	NT	ND
932	NT	NT	ND
934	NT	NT	ND
936	NT	NT	ND
976 (expresado en células LX1 Blue)	NT	<5@2hr	++
NT, no probado; ND, no detectado; +, baja expresión, ++, expresión moderada, +++, expresión elevada			

Debe anotarse que hay solamente diferencias muy conservadoras entre los subclones listados anteriormente y sus secuencias nativas que también están en el listado de secuencias. Por ejemplo, con bastante frecuencia, un codón de inicio o de tensión fue modificado para ser más eficiente para la expresión en E.coli. Se espera que la clonación de las secuencias tipo silvestre den similares resultados en términos de actividad DAT.

Parte B

Ejemplo 3 – Detección de monatina, MP, triptófano, alanina y HMG

Este ejemplo describe la metodología analítica asociada con la caracterización posterior de enzimas de D-aminotransferasa (DAT) seleccionada.

10 Análisis de monatina y triptófano por monitorización de reacción múltiple (MRM) por LC/MS/MS

El análisis de mezclas para monatina y triptófano derivados de reacciones bioquímicas fue llevado a cabo utilizando un instrumento de cromatografía líquida Waters/MICROMASS® con espectrometría de masas en tándem (LC/MS/MS) incluyendo un cromatógrafo líquido Waters 2795TM con un monitor de absorbancia Waters 996 Photo-Diode Array (PDA) colocado en serie entre el cromatógrafo y un espectrómetro de masas de cuadrupolo triple MCROMASS® QUATTRO ULTIMA®. Las separaciones de LC se hicieron utilizando una columna de cromatografía de fase reversa Xterra MS C8, 2.1 mm x 250 mm a 40°C. La fase móvil para LC consistió de A) agua que contenía 0.3% de ácido fórmico y formiato de amonio 10 mM y B) metanol que contenía 0.3% de ácido fórmico y formiato de amonio 10 mM.

El gradiente de elución fue lineal de 5% de B a 45% de B, 0-8.5 minutos, lineal desde 45% de B a 90% de B, 8.5-9 minutos, isocrático desde 90% de B a 90% de 90, 9-12.5 minutos, lineal de 90% de B a 5% de B, 12.5-13 minutos, con un periodo de reequilibrio de 4 minutos entre recorridos. La rata de flujo fue de 0.27 mL/minuto, y la absorbancia PDA fue monitorizada desde 210 nm hasta 400 nm. Todos los parámetros del ESI-MS fueron optimizados y seleccionados con base en la generación de iones moleculares protonados ($[M + H]^+$) de los analitos de interés, y la producción de iones fragmentarios característicos. Se utilizaron los siguientes parámetros instrumentales para el análisis de Monitorización de Reacción Múltiple (MRM) por LC/MS/MS de monatina y triptófano: Capilar: 3.5 kV; Cono: 40 V; Hex 1: 20 V; Apertura: 0 V; Hex 2: 0 V; Temperatura de fuente: 100° C; Temperatura de desolvatación: 350° C; Gas de desolvatación: 500 L/h; Gas de cono: 50 L/h; Resolución de masa baja (Q1): 12.0; Resolución de masa alta (Q1): 12.0; Energía de ión: 0.2; Entrada: -5 V; Energía de colisión: 8; Salida: 1 V; Resolución de masa baja (Q2): 15; Resolución de masa alta (Q2): 15; Energía de ión (Q2): 3.5; Multiplicador: 650. Se utilizaron cuatro transiciones MRM de progenitora a hija específicas para monatina y una transición de progenitora a hija específica para triptófano para detectar específicamente monatina y triptófano en reacciones *in vitro* e *in vivo*. Las transiciones monitorizadas son 293.08 a 157.94, 293.08 a 167.94, 293.08 a 130.01, y 293.08 a 256.77. El triptófano monitorizado con la transición de MRM de 205.0 a 146.0. Para la cuantificación de estándar interno de monatina y triptófano, se analizaron cuatro estándares de calibración que contenían 4 relaciones diferentes de cada analito a d5-triptófano y d5-monatina. Estos datos fueron sometidos a análisis de mínimos cuadrados lineales para formar una curva de calibración para la monatina y el triptófano. A cada muestra se agregó una cantidad fija de d5-triptófano y d5-monatina (la d5-monatina fue sintetizada a partir del d5-triptófano de acuerdo con los métodos de WO 2003/091396

A2), y las relaciones de respuesta (monatina/d5-monatina; triptófano/d5-triptófano) en conjunción con las curvas de calibración descritas arriba se utilizan para calcular la cantidad de cada analito en las mezclas. Las transiciones de masa de progenitor a hija monitorizadas para d5-triptófano y d5-monatina son 210.0 a 150.0, y 298.1 a 172.0 y 298.1 a 162.00 respectivamente.

5 Medición de monatina (MRM) por LC/MS/MS quiral

La determinación de la distribución de estereoisómeros de monatina a reacciones bioquímicas fue lograda por formación de derivados con 1-flúoro-2-4-dinitrofenil-5-L-alanina amida (FDAA), seguida por medición MRM por LC/MS/MS en fase reversa.

Formación de derivados de monatina con FDAA

10 A 50 µL de muestra o estándar y 10 µL de estándar interno se agregaron 100 µL de una solución al 1% de FDAA en acetona. Se agregaron 20 µL de bicarbonato de sodio 1.0 M, y la mezcla fue incubada durante 1 hora a 40°C con mezcla ocasional. La muestra fue retirada y enfriada, y neutralizada con 20 µL de HCl 2.0 M (puede requerirse más HCl para efectuar la neutralización de una mezcla biológica regulada). Después de que se terminó la desgasificación, las muestras estaban listas para análisis por LC/MS/MS.

15 Monitorización de reacción múltiple por LC/MS/MS para la determinación de la distribución de estereoisómeros de monatina

Los análisis fueron llevados a cabo utilizando la instrumentación de LC/MS/MS descritas en las secciones previas. La separaciones por LC capaces de separar todos los cuatro estereoisómeros de monatina (específicamente FDAA-monatina) fueron llevadas a cabo sobre una columna de cromatografía en fase reversa C18 Phenomenex Luna® 2.0 x 250 mm (3 µm) a 40°C. La fase móvil para LC consistió de A) agua que contenía 0.05% (masa/volumen) de acetato de amonio y B) acetonitrilo. La elusión fue isocrática a 13% de B, 0-2 minutos, lineal de 13% de B a 30% de B, 2-15 minutos, lineal de 30% de B 80% de B, 15-16 minutos, isocrática a 80% de B 16-21 minutos y lineal de 80% de B a 13% de B, 21-22 minutos, con un periodo de reequilibrio de 8 minutos entre recorridos. La rata de flujo fue de 0.23 mL/minuto, y la absorbancia de PDA fue monitorizada desde 200 nm a 400 nm. Todos los parámetros de la ESI-MS fueron optimizados y seleccionados con base en la generación de iones moleculares desprotonados ([M – H]) de FDAA-monatina y la producción de iones fragmentarios característicos.

Se utilizaron los siguientes parámetros instrumentales para el análisis por LC/MS de monatina en el modo de ión negativo en ESI/MS: Capilar: 3.0 kV; Cono: 40 V; Hex 1: 15 V; Apertura: 0.1 V; Hex 2: 0.1 V; Temperatura de fuente: 120°C; Temperatura de desolvatación: 350°C; Gas de desolvatación: 662 L/h; Gas de cono: 42 L/h; Resolución de masa baja (Q1): 14.0; Resolución de masa alta (Q1): 15.0; energía del ión: 0.5; Entrada: 0 V; Energía de colisión: 20; Salida: 0 V; Resolución de masa baja (Q2): 15; Resolución de masa alta (Q2): 14; energía del ión (Q2): 2.0; Multiplicador: 650. Se utilizaron tres transiciones de progenitor a hija específicas de FDAA-monatina para detectar específicamente FDAA-monatina en reacciones *in vitro* e *in vivo*. Las transiciones monitorizadas para monatina fueron 542.97 a 267.94, 542.97 a 499.07, y 542.97 a 525.04. La transición de masas derivadas del estándar interno de monatina fue 548.2 a 530.2. La identificación de los estereoisómeros de FDAA-monatina se basó en el tiempo de retención cromatográfico en comparación con estereoisómeros de monatina purificados, y los datos espectrales de masas. Se utilizó un estándar interno para monitorizar el progreso de la reacción y para la confirmación del tiempo de retención del estereoisómero S,S.

40 Detección por fluorescencia en cromatografía líquida postcolumna de aminoácidos, incluyendo triptófano, monatina, alanina y HMG

Procedimiento para triptófano, monatina y alanina

La cromatografía líquida con detección por fluorescencia postcolumna para la determinación de aminoácidos en reacciones bioquímicas fue llevada a cabo en un sistema Waters 2690 LC o equivalente combinado con un detector de fluorescencia Waters 474, y un módulo de reacción postcolumna Waters (método LC/OPA). Las separaciones por LC fueron llevadas a cabo en una columna de intercambio iónico cargada con interacción-sodio a 60°C. La fase móvil A fue el regulador Pickering Na 328 (Pickering Laboratories, Inc.; Mountain View, CA). La fase móvil B fue el regulador Pickering Na 740. El gradiente de elusión fue de 0% de B a 100% de B, 0-20 minutos, isocrática a 100% de B, 20-30 minutos, y lineal desde 100% de B a 0% de B, 30-31 minutos, con un periodo de reequilibrio de 20 minutos entre recorridos. La rata de flujo para la fase móvil fue de 0.5 mL/minuto. La rata de flujo para la solución de derivación postcolumna OPA fue 0.5 mL/minuto. Los parámetros del detector de fluorescencia fueron EX 338 nm y Em 425 nm. Se empleo Norleucina como estándar interno para el análisis. La identificación de los aminoácidos se basó en los datos de tiempo de retención cromatográfica para estándares purificados.

Procedimiento para HMG

Las muestras de reacciones bioquímicas fueron limpiadas mediante cartuchos de extracción en fase sólida (SPE) que contenían C18 como material de empaque y 0.6% de ácido acético como eluyente. La fracción recolectada del SPE fue llevada a un volumen conocido y analizada utilizando derivación con O-ftalaldehído (OPA) postcolumna de HPLC con un detector de fluorescencia. La separación cromatográfica fue hecha posible utilizando un sistema de cromatografía líquida Waters 2695 y dos columnas Phenomenex AquaC18 en serie; una columna de 2.1 mm x 250 mm con partículas de 5 µm, y una columna de 2.1 mm x 150 mm con partículas de 3 µm. La temperatura de la columna fue de 40°C y la rata de flujo isocrática fue de 0.18 mL/minuto. La fase móvil fue ácido acético al 0.6%. El sistema de derivación postcolumna con OPA y de detección consiste de un Waters Reagent Manager (RMA), una cámara de reacción en espiral, un módulo de control de temperatura para la reacción de la cámara en espiral, y un detector Waters 2847 Florescent. La rata de flujo de OPA fue fijada a 0.16 mL/minuto, y la cámara de reacción en espiral fue fijada a 80°C. El detector de fluorescencia fue fijado con una longitud de onda de excitación de 348 nm y una longitud de onda de emisión de 450 nm. Otros parámetros de control de la sensibilidad del detector, tales como la ganancia y atenuación de la señal, fueron fijados según las necesidades experimentales. La cuantificación de HMG se basó en la respuesta molar del ácido glutámico.

15 Detección de MP por LC/MS

Las separaciones por cromatografía líquida se realizaron utilizando un sistema de cromatografía líquida Waters 2690 y una columna de cromatografía en fase reversa 2.1 mm x 50 mm Agilent Eclipse XDB- C18 1.8 µm con rata de flujo a 0.25 mL/minuto y condiciones de gradiente como siguen:

Tiempo (minuto)	A%	B%
0.00	95	5
0.2	95	5
1.2	5	95
4.5	5	95
5.0	95	5
10	95	5

La fase móvil A fue 0.3% (v/v) de ácido fórmico con formiato de amonio 10 mM, y la fase móvil B fue de 0.3% de ácido fórmico con formiato de amonio 10 mM en metanol/acetonitrilo 50:50. La temperatura de la columna fue de 40°C.

Los parámetros para el espectrómetro de masas de cuadrupolo Micromass ZQ operando en modo de ionización por electroaspersión negativa (-ESI) fueron fijadas como sigue: Capilar: 2.2 kV; Cono: 35 V; Extractor: 4 V; RF lente: 1 V; Temperatura de fuente: 120°C; Temperatura de desolvatación: 380°C; Gas de desolvatación: 600 L/h; Gas de cono: Cerrado; Resolución de masa baja: 15.0; Resolución de masa alta: 15.0; energía del ión: 0.2; Multiplicador: 650. El experimento de MS monitorizando iones sencillos fue configurado para permitir la detección selectivamente para m/z 290.3, 210.3, 184.3, y 208.4. El m/z de 208.4 es el ión molecular desprotonado [M-H]⁻ del estándar interno de d₅-triptófano.

30 Detección de MP por LC/MS/MS

Las separaciones por LC fueron hechas utilizando el sistema de cromatografía líquido HPLC y una columna de cromatografía en fase reversa 2.1mm x 50 mm Agilent Eclipse XDB- C18 1.8 µm con una rata de flujo de 0.25 mL/minuto y las condiciones de gradiente son como sigue:

Tiempo (minuto)	A%	B%
0.00	95	5
0.7	95	5
3.0	5	95

Tiempo (minuto)	A%	B%
4.0	5	95
4.3	95	5
6.0	95	5

La fase A móvil fue 0.3% (v/v) de ácido fórmico con formiato de amonio 10 mM y la B fue 0.3% de ácido fórmico con formiato de amonio 10 mM en metanol/acetronitrilo 50:50. La temperatura de la columna fue de 40°C.

5 Los parámetros del espectrómetro de masas de cuadrupolo triple Waters Premier XE para los experimentos de Monitorización con Reacción Múltiple de LC/MS/MS (MRM) operando en modo de ionización de electroaspersión negativa (-ESI) fueron fijados como sigue: Capilar: 3.0 kV; Cono: 25 V; Extractor: 3 V; RF lente: 0 V; Temperatura de fuente: 120°C; Temperatura de desolvatación: 350°C; Gas de desolvatación: 650 L/hr; Gas de cono: 47 L/hr; Resolución de masa baja (Q1): 13.5; Resolución de masa alta (Q1): 13.5; energía del ión (Q1): 0.5 V; Entrada: 1 V; Energía de colisión: 18 V; Salida 1: 19; Resolución de masa baja (Q2): 15; Resolución de masa alta (Q2): 15; energía del ión (Q2): 2.0; Multiplicador: 650. Se monitorizaron cuatro transiciones de MRM de progenitor a hija para detectar selectivamente el precursor de monatina (MP) y el precursor de d5-monatina (d5-MP); se utilizó d5-MP como estándar interno (I.S.). Las cuatro transiciones MRM fueron 290.1 a 184.1, 290.1 a 210.1, 290.1 a 228.1, y 10 295.1 a 189.1. Dos de estas transiciones, 290.1 a 184.1 para MP, y 295.1 a 189.1 para d5-MP, fueron utilizadas para generar curvas de calibración y para propósitos de cuantificación. Las transiciones de 290.1 a 210.1 y 290.1 a 228.1 15 fueron utilizadas como confirmación cualitativa secundaria de MP.

Producción de monatina y MP para estándares y para ensayos

Producción de monatina

20 Una mezcla racémica de R,R y S,S monatina fue producida sintéticamente como se describe en la Patente de los Estados Unidos No. 5,128,482. Las R,R y S,S monatinas fueron separadas mediante formación de derivados y una etapa de hidrólisis. En resumen, la mezcla racémica de monatina fue esterificada, el grupo amino libre fue bloqueado, con carbamazepina (CBZ), se formó una lactona, y la S,S lactona fue hidrolizada selectivamente utilizando una enzima de proteasa inmovilizada. La monatina también puede ser separada según se describe en Bassoli et al., Eur. J. Org. Chem., 8:1652-1658, (2005).

Producción de MP

25 La R-MP fue producida mediante la transaminación de la R,R monatina utilizando D-aminotransferasa AT-103 de rango amplio (BioCatalytics, Pasadena, CA) en regulador de fosfato de potasio 0.1 M, utilizando piruvato de sodio como aceptor de amino. La S-MP fue producida por la transaminación de S,S monatina utilizando L-aminotransferasa AT-102 (BioCatalytics) en regulador de fosfato de potasio 0.1 M, utilizando piruvato de sodio como aceptor de amino. Ambas reacciones fueron llevadas a cabo a 30°C y a un pH de aproximadamente 8.0-8.3, durante 30 aproximadamente 20 horas. Ambos compuestos fueron purificados utilizando HPLC en escala preparativa con una resina hidrófoba Rohm and Haas (Filadelfia, PA) (XADTM 1600), eluyendo en agua. Las muestras que contenían más de 90% de pureza del precursor de monatina fueron recolectadas y liofilizadas.

Ejemplo 4 – Métodos para preparación de proteínas

35 Este ejemplo describe la metodología utilizada para la clonación, expresión, preparación de extractos celulares, purificación de proteínas y cuantificación de proteínas para caracterización secundaria de DATs seleccionados.

Las personas experimentadas en la técnica entenderán que la presencia de actividad en un polipéptido codificado a partir de un ácido nucleico subclonado (por ejemplo, un fragmento) o modificado de otra manera (por ejemplo, etiquetado) se considera predictiva de la presencia de actividad en el correspondiente polipéptido codificado a partir del ácido nucleico de longitud completa o tipo silvestre.

40 Amplificación de genes que codifican DAT para clonación en plásmidos Topo

Las reacciones de PCR para clonación Topo (bien sea utilizando Pfu turbo o Pfu clonado de Stratagene) fue como sigue: 1X de regulador recomendado para la enzima polimerasa, dNTPs 0.2 mM, 0.5 µM de cada cebador, y 1 µl por 50 µl de reacción de la polimerasa (2.5 unidades). Las reacciones contenían aproximadamente 5-100 ng de ADN de plantilla por reacción. Un inicio en caliente a 94°C durante 2 minutos fue utilizado para PCR, así como una 45 temperatura de fusión de 94°C. La temperatura de fusión fue dependiente del Tm de los cebadores, y fue

bien 30 o 60 segundos. El tiempo de extensión (a 72°C) fue al menos 2 minutos por kb. Los productos de reacción fueron separados normalmente sobre un gel de agarosa 1x TAE al 1%, y las bandas de tamaños apropiados fueron purificadas con el QIAquick Gel Extraction Kit según lo recomienda el fabricante excepto que se utilizó un volumen de dilución de 10 a 50 µl. Se utilizaron volúmenes de 1 a 4 µl del producto de PCR ya purificado para la ligazón con el plásmido PCRII-Topo Blunt (Invitrogen, Carlsbad, CA) según lo recomienda el fabricante.

Clonación de DAT en pET30a para expresión no etiquetada

Los DATs que tienen la secuencia mostrada en SEQ ID NO: 945, 947, 949, 891, 893, 869, 873, 877, 881, 883, y 895 (que codifican los polipéptidos que tienen la secuencia de SEQ ID NO: 946, 948, 950, 892, 894, 870, 874, 878, 882, 884, y 896) fueron amplificados a partir de plásmidos o productos de PCR con Pfu Turbo (Stratagene, La Jolla, CA) y cebadores agregando un Nde I en el extremo 5' y un sitio de restricción Not I o BamH I en el extremo 3'. Los fragmentos de PCR fueron clonados en pCR-Blunt II-Topo (Invitrogen, Carlsbad, CA) según lo recomienda el fabricante. La secuencia fue verificada por secuenciación (Agencourt, Beverly, MA) y los insertos con las secuencias correctas fueron liberadas entonces del vector utilizando las enzimas de restricción apropiadas y ligados en los sitios de restricción Nde I y Not I (o BamH I) de pET30a. Véase la Tabla 5 para cebadores específicos.

El ácido nucleico DAT que tiene la secuencia de SEQ ID NO: 155 (que codifica el polipéptido que tiene SEQ ID NO: 156) fue amplificado con Pfu Turbo (Stratagene) y los cebadores que agregan un sitio de restricción Nde I y Hind III en el extremo 5' y 3', respectivamente. Los fragmentos de PCR fueron digeridos utilizando enzimas de restricción Nde I y Hind III y ligados en los sitios de restricción de Nde I y Hind III de pET30a. Véase la Tabla 5 para cebadores específicos. Debe anotarse que el polipéptido que tiene la secuencia de SEQ ID NO: 156 parecía contener la siguiente secuencia guía con una probabilidad de 0.991 (determinada según SignalP, como se discute en Nielsen, 1997, Protein Engineering, 10:1-6): KNSPIIAAYRAATPGSAAA (SEQ ID NO: 1084). El ácido nucleico que codifica este polipéptido DAT fue clonado con la secuencia guía aparente.

Tabla 5. Cebadores para amplificación

Amplifica SEQ ID NO:	Cebadores para PCR	SEQ ID NO
945	5'-CCGCCCCATATGAACGCACTAGGATATTACAACGGAAAATGG-3'	978
	5'-GGCGGATCCTTATCCAAAGAATTCGGCAGGAGCTGTC-3'	979
947	5'-CCGCCCCATATGCGCGAAATTGTTTTTTGAATGGG -3'	980
	5'-CGGATCCCTAAACCATCTCAAAAACTTTTGCTGAATAAACCGTG -3'	981
949	5'-CCGCCCCATATGTTGGATGAACGGATGGTGTTCATTAAC-3'	982
	5'-GGCGGATCCCTAGTCCACGGCATAGAGCCACTCGG-3'	983
891	5'-GGCCGCATATGGACGCACTGGGATATTACAACGGAAAATG-3'	984
	5'-GGCCGCGGCCGCTATGCCTTTCTCCACTCAGGCGTGTAGC -3'	985
893	5'-GGCCGCATATGGACGCACTGGGATATTACAACGGAAAATG -3'	986
	5'-GGCCGCGGCCGCTATACTGTGCTCCACTCAGGCGTGTAGCC -3'	987
869	5'-CATATGTATTATTATGGAATGATCAATAGTGAAGG -3'	988
	5'-GCGGCCGCTATTTATTCGTAAGGTGTTGGAATTTTCG -3'	989
873	5'-CATATGAGCACCCCGCCGACCAATC-3'	990
	5'-GCGGCCGCTAGGCCGCTTCACTTCACGCTC -3'	991
877	5'-CATATGAGCACCCCGCCAACCAATTC-3'	992
	5'-GCGGCCGCTACGCGCCTTCACTTCGCGC -3'	993

Amplifica SEQ ID NO:	Cebadores para PCR	SEQ ID NO
881	5'-TCCAGGCATATGAGCACAGTATATTTAAATGGCC -3'	994
	5'- CCAGTAGCGGCCGCCTAACACTCAACACTATACTTATGC -3'	995
883	5'- TCTAGGCATATGGTTTATCTGAACGGGCG -3'	996
	5'- ACTGTAGCGGGCGCCTATCCGAGGGACGCGTTGG -3'	997
895	5'- CATATGAAAGAGCTGGGCTATTACAACGGAAAAATC-3'	998
	5'- GCGGCCGCCTATGACCTCCACCCCTGATTTCCAAAATAC-3'	999
155	5'- CTAGGATTCCATATGAAGAATTCGCCGATCATC-3'	1000
	5'- CGAAGCTTCAACAGCGGCCGCTTAAAG -3'	1001

Mutagénesis dirigida al sitio

La mutagénesis dirigida al sitio fue llevada a cabo utilizando el QuikChangeMulti Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) de acuerdo con las instrucciones del fabricante. Para generar el mutante SEQ ID NO: 870 T242N, el constructo no etiquetado pET30a descrito en el Ejemplo 10 fue utilizado como plantilla. Para generar los mutantes SEQ ID NO: 220 G240N y SEQ ID NO: 220 T241N, el constructo pET30a con un terminal C de etiqueta His descrito en el ejemplo 10 fue utilizado como plantilla. Los cebadores mutagénicos usados se alistan más abajo en la Tabla 6. Todas las mutaciones deseadas fueron confirmadas por secuenciación de ADN.

Tabla 6. Secuencia de cebador

Designación de polipéptido mutante (SEQ ID NO:)	Secuencia	SEQ ID NO:
870 T242N	5'- AATTATTTGTTTCATCAACAAATTCTGAAATTACGCCGGTTATTG-3'	1002
220 G240N	5'- CTTGTGTCCAGCAGCAACACACTCGGCCTTAG -3'	1003
220 T241N	5'- GTCCAGCAGCGGCAACCTCGGCCTTAGCGCC -3'	1004

Clonación de productos de DAT por PCR en pET30a para la expresión de una proteína no etiquetada

Los ácidos nucleicos de DAT que tienen la secuencia mostradas en SEQ ID NO:177, 179, 153, 165, 181, 217, 187, 189, 207, 219, 215, 195, 199, 197, 209, 201, 221, 235, 203, 237, 239, 223, 225, 227, 229, 231, 245, 213, 155, 169, 171, 167, 173, y 175 (que codifican polipéptidos DAT que tienen la secuencia mostrada en SEQ ID NO:178, 180, 154, 166, 182, 218, 188, 190, 208, 220, 216, 196, 200, 198, 210, 202, 222, 236, 204, 238, 240, 224, 226, 228, 230, 232, 246, 214, 156, 170, 172, 168, 174, y 176) fueron recibidos como productos de PCR con extremos compatibles Nde I y Not I, así como nucleótidos extraños para mejorar las eficiencias de corte.

Los productos de DAT por PCR contenían un sitio de una enzima de restricción NdeI en el extremo 5' y un sitio NotI en el extremo 3'. Los fragmentos de PCR fueron clonados primero en el vector pCR4 TOPO o pCR-Blunt II-TOPO (Invitrogen). Después de que las secuencias de ADN fueron verificadas por secuenciación, los genes de DAT fueron liberados de los plásmidos TOPO por la digestión de NdeI y NotI y ligados en el vector pET30a el cual había sido cortado utilizando las mismas enzimas de restricción. Los genes de DAT que contenían tanto el sitio NdeI o el NotI fueron amplificados utilizando cebadores con sitios de enzima de restricción compatibles y clonados en pET30a. Por ejemplo, el ácido nucleico de DAT que tenía la SEQ ID NO: 155 (que codifica el polipéptido que tiene la secuencia de SEQ ID NO: 156) fue reamplificado a partir del producto de PCR original utilizando los sitios de restricción NdeI y HindIII para clonar en el pET30a.

Clonación de DATs en pET30a para la expresión como proteína de fusión C-His-etiquetada

Los ácidos nucleicos que codifican DAT 4978 y DAT 4978 T243N (descritos en el Ejemplo 6) SEQ ID NO: 870 (expresados del plásmido pSE420-cHis), y SEQ ID NO: 870 T242N, SEQ ID NO: 176 y SEQ ID NO: 220 (versiones no etiquetadas expresadas a partir de pET30) fueron reamplificados con Pfu Turbo (Stratagene) y cebadores que colocan un sitio XhoI inmediatamente corriente arriba de los codones de detención. Los fragmentos de PCR fueron clonados en pCR-Blunt II-TOPO (Invitrogen, Carlsbad, CA) como lo recomienda el fabricante o fueron clonados directamente en los sitios de restricción Nde I y Xho I de pET30a. La secuencia fue verificada por secuenciación (Agencourt, Beverly, MA) y se liberó entonces un inserto con la secuencia correcta a partir del vector utilizando enzimas de restricción Nde I y Xho I y el inserto fue ligado en los sitios de restricción de Nde I y Xho I de pET30a. Véase la Tabla 7 para cebadores y nombres de plásmidos específicos.

Tabla 7. Secuencias de cebadores

Designación del polipéptido (SEQ ID NO:)	Secuencias	SEQ ID NO
870 y	5'- CATATGTATTTCATTATGGAATGATCAAATAGTGAAGG -3'	1005
870 T242N	5'- CTCGAGTTTATTCGTAAAAGGTGTTGGAATTTTCGTTTC -3'	1006
DAT 4978 y DAT4978T243N	5'- CATATGAGTTATAGCTTATGGAATGACCAAATTGTGAATG -3'	1007
	5'- CTCGAGTGCGCGAATACCTTTTGGGATTTTCGTATC -3'	1008
220	5'- CTAGGATCTCATATGGACGCACTGGGATATTAC -3'	1009
	5'- GCCTCGAGTACCCTGCTCCACTCAGG -3'	1010
176	5'- CTAGGATTCCATATGGACGCGCTTGGCTATTAC -3'	1011
	5'- GCCTCGAGTACCCTGCTCCACGCAG -3'	1012

Clonación de CbDAT y CaDAT

Una D-aminotransferasa de *Clostridium beijerinckii* fue amplificada por PCR utilizando Pfu Turbo (Stratagene) y ADN genómico de *C. beijerinckii* con cebadores de PCR que contenían sitios de restricción en 5' NdeI y en 3' NotI. El ADN genómico fue extraído de *C. beijerinckii* (ATCC 51743) utilizando el kit de purificación de ADN genómico Purrgene (Gentra Systems, Minneapolis, MN) según las instrucciones del fabricante.

El producto de PCR 824 bp fue extraído en gel utilizando un Qiagen Gel Extraction Kit y TOPO clonado en pCR-Blunt II-Topo (Invitrogen). Después de verificar la secuencia, el gen fue ligado a los vectores de corte pET28b y pET30a de Nde I y Not I utilizando un kit de Ligazón Rápida (Roche).

El DAT de *C. acetobutylicum* fue amplificado por PCR utilizando ADN genómico (ATCC 824) y el Stratagene Optiprime PCR Kit con cebadores de PCR que contenían un sitio de restricción en 5' NdeI y en 3' NotI. La reacción de PCR exitosa fue clonada en el vector pCR4 TOPO y los clones TOPO fueron secuenciados. Un clon TOPO positivo fue digerido con enzimas de restricción NdeI y NotI y el fragmento de DAT ligado en el vector pET30a digerido con las mismas enzimas de restricción.

Tabla 8. Secuencias de cebador

Designación	Secuencia	SEQ ID NO:
CbDAT1	5'- GGTTCATATGGAGAATTTAGGTTATTA -3'	1013
	5'- GGAAGCGGCCGCATATTCTACCTCCTATTCTG -3'	1014
CaDAT2	5'- GGTTCATATGAAAGATTTAGGATATTACAATGGAGAATAC-3'	1015
	5'-GGAAGCGGCCGCTTAATTTGTTTCTTCCAAAAATTCATTAAG -3'	1016

Síntesis *in vitro* de LsDAT

El DAT de *Lactobacillus salivarius* fue ensamblado utilizando un método revisado basado en Stemmer et al., 1995, Gene, 164:49-53. En resumen, 43 oligonucleótidos (primariamente 40-meros) fue ordenado a partir de IDT con base en la secuencia genética y su secuencia de ADN complementaria, con 20 superposiciones de pares de bases entre las cadenas en sentido y antisentido. Véase la Tabla 9 para la lista de cebadores. Los cebadores fueron diluidos a 250 µM en agua y 5 µL de cada cebador fueron mezclados juntos en un tubo de microcentrífuga. Se llevó a cabo el PCR como sigue: por 100 µL de reacción, 1.5 µL del conjunto de cebadores, 4 µL dNTP, regulador de PCR 1X XL, acetato de magnesio 1 mM, 2 µL rTth polimerasa (Roche, Indianápolis, IN) y fueron agregados 0.25 µL de Pfu polimerasa (Stratagene, La Jolla, CA). Se hizo un inicio en caliente de 3 minutos a 94°C, seguido por 15 ciclos de 94°C durante 30 segundos, 42°C durante 30 segundos, y 68°C durante 15 segundos. Se realizaron diez ciclos más con un tiempo de extensión de 30 segundos (a 68°C). Se llevaron a cabo diez ciclos más con un tiempo de extensión de 75 segundos. Finalmente, se realizó una etapa de extensión de cadena durante siete minutos a 68°C.

Tabla 9. Oligos utilizados para la síntesis de LsDAT

Designación	Secuencia (5' → 3')	SEQ ID NO:
F1:	ATGAAGCAAG TTGGATACTA CAATGGTACT ATCGCTGATT	1017
F2:	TAAATGAAGT TAAGGTGCCT GCTACTGATC GTGCACTTTA	1018
F3:	TTTTGGTGAT GGTTGCTACG ATGCAACTAC ATTTAAGAAC	1019
F4:	AATGTTGCAT TTGCCTTAGA AGATCATCTT GATCGTTTTT	1020
F5:	ATAATAGTTG TCGCCTACTA GAGATCGATT TCCCTTTAAA	1021
F6:	TCGCGATGAA CTAAAGAAA AGCTTTACGC TGTTATTGAT	1022
F7:	GCTAACGAAG TTGATACTGG TATCCTTTAT TGGCAAACCT	1023
F8:	CACGTGGTTC TGGTTTACGT AACCATATTT TCCCAGAAGA	1024
F9:	TAGCCAACCT AATTTATTAA TTTTACTGCT TCCTTATGGT	1025
F10:	TTAGTTCCAT TTGATACTGA ATATAAACTT ATATCTCGCG	1026
F11:	AAGACACTCG CTTCTTACAT TGCAATATTA AAACCTTGAA	1027
F12:	TTACTTCCA AACGTTATTG CAAGTCAAAA GGCTAATGAA	1028
F13:	AGTCATTGCC AAGAAGTGGT CTTCCATCGC GGTGACAGAG	1029
F14:	TTACAGAATG TGCACACTCT AACATCTTAA TTCTAAAAGA	1030
F15:	TGGCGTTCTT TGCTCCCCAC CTAGAGATAA TTTAATCTTG	1031
F16:	CCAGGAATTA CTTTGAAACA TCTCTTGCAA TTAGCAAAAG	1032
F17:	AAAATAATAT TCCTACTTCC GAAGCACCAT TCACTATGGA	1033
F18:	TGATCTTAGA AATGCTGATG AAGTTATTGT TAGTTCTTCA	1034
F19:	GCTTGTCTAG GTATTCGCGC AGTCGAGCTT GATGGTCAGC	1035
F20:	CTGTTGGTGG AAAAGATGGA AAGACTTTAA AGATCTTGCA	1036
F21:	AGATGCTTAT GCTAAGAAAT ATAATGCTGA AACTGTAAGT CGTTAA	1037
R1:	TAGTATCCAA CTTGCTTCAT	1038

R2:	AGGCACCTTA AGTTCATTTA AATCAGCGAT AGTACCATTG	1039
R3:	CGTAGCAACC ATCACCAAAA TAAAGTGCAC GATCAGTAGC	1040
R4:	TCTAAGGCAA ATGCAACATT GTTCTTAAAT GTAGTTGCAT	1041
R5:	TAGTAGGCGA CAACTATTAT AAAAACGATC AAGATGATCT	1042
R6:	TTTCTTTAAG TTCATCGCGA TTTAAAGGGA AATCGATCTC	1043
R7:	CCAGTATCAA CTTGTTAGC ATCAATAACA GCGTAAAGCT	1044
R8:	ACGTAAACCA GAACCACGTG AAGTTTGCCA ATAAAGGATA	1045
R9:	TTAATAAATT AGGTTGGCTA TCTTCTGGGA AAATATGGTT	1046
R10:	TCAGTATCAA ATGGAAGTAA ACCATAAGGA GCAGTAAAAA	1047
R11:	ATGTAAGAAG CGAGTGTCTT CGCGAGATAT AAGTTTATAT	1048
R12:	CAATAACGTT TGGAAGTAAA TTCAAAGTTT TAATATTGCA	1049
R13:	ACCACTTCTT GGCAATGACT TTCATTAGCC TTTTGACTTG	1050
R14:	AGAGTGTGCA CATTCTGTAA CTCTGTCACC GCGATGGAAG	1051
R15:	GTGGGGAGCA AAGAACGCCA TCTTTTAGAA TTAAGATGTT	1052
R16:	TGTTTCAAAG TAATTCCTGG CAAGATTAAT TTATCTCTAG	1053
R17:	GGAAGTAGGA ATATTATTTT CTTTGTCTAA TTGCAAGAGA	1054
R18:	CATCAGCATT TCTAAGATCA TCCATAGTGA ATGGTGCTTC	1055
R19:	GCGCGAATAC CTAGACAAGC TGAAGAACTA ACAATAACTT	1056
R20:	TCCATCTTTT CCACCAACAG GCTGACCATC AAGCTCGACT	1057
R21:	ATTTCTTAGC ATAAGCATCT TGCAAGATCT TTAAAGTCTT	1058
R22:	TTAACGACTT ACAGTTTCAG CATTAT	1059

Una amplificación secundaria con cebadores de L. sal DAT R Not I y L. sal DAT F Nde I (más abajo) dieron como resultado un banda del peso molecular correcto. Véase la Tabla 10 para estas secuencias de cebadores de amplificación secundarias.

5

Tabla 10. Secuencias de cebadores

Designación	Secuencias (5' → 3')	SEQ ID NO:
L. sal DAT R NotI	TTGGCCAAGCGGCCGCTTAACGACTTACAGTTT	1060
L. sal DAT F NdeI	GGTTCAAGGCATATGAAGCAAGTTGGATACTA	1061

10

La reacción de PCR secundaria fue configurada de la misma forma que la anterior con la excepción de que solamente se agregaron 2 cebadores. Para la plantilla de PCR, se utilizaron 2.5 µl de la reacción de PCR primaria. Se hizo un inicio en caliente de 3 minutos a 94°C., seguidos por 10 ciclos de 94°C durante 30 segundos, 42°C durante 30 segundos, y 68°C durante 15 segundos. Se hicieron 10 ciclos más con una temperatura de fusión incrementada de 48°C durante 30 segundos con un tiempo de extensión de 30 segundos (a 68°C). Finalmente, se hizo una etapa de extensión de cadena durante siete minutos a 68°C.

El fragmento fue clonado en un vector pCR-BluntII-TOPO y los clones TOPO fueron secuenciados. Se cortó un clon TOPO positivo con NdeI y NotI y el fragmento de DAT fue ligado en un vector pET30a digerido con las mismas enzimas de restricción.

Preparación de enzimas

- 5 Se utilizó una cepa de *E. coli* BL21(DE3) como cepa anfitriona para la expresión de DAT a partir de plásmidos derivados de pET. La cepa de *E. coli* TOP10 fue utilizada en todos los otros constructos de DAT. Colonias individuales de constructos deseados fueron inoculadas típicamente en medio Overnight Express II (Novagen) que contenía la cantidad apropiada de antibióticos. Después del cultivo a 30°C durante la noche, las células fueron recolectadas por centrifugación cuando la OD600 fue superior a 10. Alternativamente, los cultivos durante la noche fueron utilizados para inocular cultivos en medio LB que contenían los antibióticos apropiados. Los cultivos se hicieron crecer a 30°C hasta una OD600 de 0.5 a 0.9 y la expresión de la proteína fue inducida con IPTG 1 mM durante 4 horas a la misma temperatura.

- 15 Los extractos celulares fueron preparados agregando 5 mL por g de pella de células ó 5 mL por 50 mL de cultivo nocturno, del Reactivo de Extracción BugBuster® (primario libre de aminas) (EMD Biosciences/Novagen catálogo #70923) con 5 µL/mL de inhibidor de proteasa Cocktail II (EMD Bioscience/Calbiochem catálogo #539132), 1 µl/ml de Benzonase® Nuclease (EMD Biosciences/Novagen catálogo #70746), y 0.033 µl/ml de solución de r-Lisozima™ (EMD Biosciences/Novagen catálogo #71110) a las células. La resuspensión de células fue incubada a temperatura ambiente durante 15 minutos con agitación suave. Después de la centrifugación a 16,100 rcf durante 20 minutos a 4°C, el sobrenadante fue retirado como extracto libre de células.

- 20 Antes de utilizar la preparación enzimática para las reacciones de monatina, se retiraron los detergentes y compuestos de bajo peso molecular del extracto libre de células pasando a través de una columna PD-10 (GE Healthcare, Piscataway, NJ) que fue equilibrada previamente con regulador de fosfato de potasio (100 mM, pH 7.8) o regulador EPPS (100 mM, pH 8.2) que contenía 0.05 mM de PLP. La proteína fue eluida utilizando el regulador de equilibrio. Las concentraciones de proteína fueron determinadas típicamente utilizando el ensayo de placa BioRad Coomassie (también conocido como el ensayo Bradford) de ensayo en placa con BSA (Pierce) como estándar. Ocasionalmente, el ensayo en placa de microtitulación BCA (Pierce) fue utilizado para la determinación de proteína cuando está anotado. Para estimar la concentración de la D-aminotransferasa en los extractos libres de células, se cargaron muestras de 1 mg/mL sobre el sistema de electroforesis Experion (Bio-Rad, Hercules, Ca) y se utilizó el Experion Software (Versión 2.0.132.0) para calcular el porcentaje de la proteína DAT soluble en el extracto libre de células. Alternativamente, se hizo el análisis SDS-PAGE y se utilizó la estimación visual para estimar el porcentaje de expresión.

- 35 Estas proteínas de fusión etiquetadas con His fueron purificadas utilizando bien sea la resina GE Healthcare Chelating Sepharose Fast Flow o columnas Novagen His-Bind. La purificación utilizando la resina de Sefarosa involucró la carga del extracto libre de células sobre una columna que ha sido equilibrada previamente con el regulador de fosfato de potasio (100 mM, pH 7.8) que contenía 200 mM de cloruro de sodio y 0.050 mM de PLP. La columna fue lavada entonces sucesivamente utilizando 3-5 volúmenes de columna del regulador de equilibrio, 3-5 volúmenes de columna del regulador de equilibrio que contenía 25 mM de imidazol y 3-5 volúmenes de columna del regulador de equilibrio que contenía 50-100 de imidazol. Esta proteína etiquetada con His fue eluida de la columna usando 3-5 volúmenes de columna del regulador de equilibrio que contenían 500 mM de imidazol. El eluido fue concentrado utilizando Amicon (Billerica, MA) Centricon-70. El imidazol y las sales de cloruro de sodio de la solución concentrada de proteína fueron removidas pasándola a través de columnas de desalinización PD-10 que fueron equilibradas previamente utilizando el regulador de fosfato de potasio (100 mM, pH 7.8) (para DAT4978 y DAT4978 T243N) o el regulador EPPS (100 mM, pH 8.2) (para SEQ ID NO: 870 y SEQ ID NO: 870 T242N) que contenía 50 µM de PLP. Las concentraciones de proteína fueron determinadas utilizando Bio-Rad Protein Assay (Bio-Rad) y albúmina (Pierce) como estándar. Se almacenaron alícuotas (0.5-1 mL) de la enzima purificada a -80°C hasta su uso. La purificación de la proteína etiquetada con His utilizando las columnas His-Bin se hizo siguiendo las instrucciones del fabricante. El eluido de la columna fue desalinizado utilizando la columna PD10 como se describió anteriormente.

Ejemplo 5 - Procedimientos de ensayo #2 para la actividad de la D-aminotransferasa

- 50 Ensayo de producción de monatina (estándar)

Se combinaron los siguientes componentes: EPPS 100 mM, pH 8.2; piruvato de sodio 200 mM; D-triptófano 100 mM; PLP 50 µM; MgCl₂ 1 mM; Tween-80 0.01%; 50 µg/mL de aldolasa descrita en el Ejemplo 6 (se utilizó el extracto libre de células; la concentración de aldolasa se estimó con base en la lectura de porcentajes del chip Experion) y una cantidad apropiada de DAT (típicamente 0.1-1 mg/mL).

- 55 Excepto por la solución de reserva de PLP y las soluciones de proteína, todos los demás reactivos fueron hechos utilizando agua desionizada libre de oxígeno y almacenados en la cámara anaeróbica. Las reacciones fueron

llevadas a cabo en la cámara anaeróbica a temperatura ambiente con agitación suave constante. Para tomar un punto de tiempo, se agregó ácido fórmico en una alícuota de la mezcla de reacción hasta una concentración final de 2% (v/v). Después de la centrifugación a 16,100 RCF durante 5 minutos utilizando una microcentrífuga de mesa, el sobrenadante fue filtrado a través de un filtro de membrana de nylon de 0.2 µm. Las muestras fueron diluidas entonces de 20 a 100 veces con agua antes del análisis por LC/MS/MS.

Ensayo de transaminación de D-triptófano

Para comparar las actividades de transaminación de D-triptófano de ciertas D-aminotransferasas, se llevaron a cabo los siguientes ensayos. La mezcla de ensayo contenía: 0.5 mg/mL de proteína del extracto celular que contenía D-AT; fosfato de potasio 40 mM pH 8.2; D-triptófano 20 mM; piruvato de sodio 100 mM; y PLP 50 µM. Los ensayos fueron incubados a 37°C durante 30 minutos y luego colocados sobre hielo.

El grado de reacción fue seguido midiendo la cantidad de indol-3-piruvato formado utilizando el siguiente ensayo: a 5 µl, 10 µl y 20 µl de mezcla de reacción, se agregaron 200 µl de la siguiente solución: arsenato de sodio 0.5 mM; EDTA 0.5 mM; y tetraborato de sodio 50 mM (pH 8.5). La absorbancia del complejo indol-3-piruvato enol-borato a 325 nm fue comparada con una curva estándar de indol-3-piruvato preparada en la misma solución.

La formación de alanina también puede ser utilizada para seguir el grado de las reacciones de transaminación de D-triptófano. Las concentraciones de alanina fueron determinadas como se describe en el Ejemplo 3.

Ensayo de transaminación de R,R monatina

Las condiciones del ensayo (volumen final 2 mL) incluyeron: 0.01% de Tween; EPPS 100 mM pH 8.2; piruvato de sodio 100 mM; R,R monatina aproximadamente 3 mM; 0.5 mg/mL de DAT; y PLP 50 µM. El grado de la reacción fue monitorizado por detección de la alanina o de la R-MP formada utilizando los protocolos descritos en el Ejemplo 3.

Ejemplo 6 – Método para obtener DAT y una aldolasa

Este método describe la clonación de la aldolasa utilizada en las reacciones de formación de monatina con las D-aminotransferasas, y D-aminotransferasas aisladas previamente que fueron utilizadas para propósitos comparativos.

Aldolasa

La aldolasa utilizada en los ensayos de producción de monatina a partir de D-triptófano fue aislada y subclonada en vectores pET tal como se describe en WO 2007/103389 (citado en esa solicitud como la aldolasa de SEQ ID NO: 276 codificada por el ácido nucleico de SEQ ID NO: 275).

DAT y DAT4978 T243N

Una D-aminotransferasa de ATCC #4978 (DAT 4978) fue clonada como se describe en la Publicación de los Estados Unidos No. 2006/0252135. Se hizo un mutante T243N utilizando el constructo pET30 (no etiquetado) DAT 4978.

El cebador para mutagénesis fue diseñado siguiendo las sugerencias listadas en el kit Stratagene Multi-Change (La Jolla, CA). El cebador fue fosforilado en 5'. La mutagénesis se hizo utilizando el kit Stratagene Multi-Change siguiendo las instrucciones del fabricante. La secuencia de oligonucleótidos mutagénica se muestra en la Tabla 11.

Tabla 11. Secuencias de oligonucleótidos mutagénica

Nombre del mutante	Cambio de aminoácido	Cebador	SEQ ID NO:
DAT4978T243N	T243N	5'-GTGATTGTTTCATCAACGAATTCAGAAGTAACGCC-3'	1062

Se transformaron células de E. coli XL10-Gold (Stratagene) y las preparaciones de plásmido purificado resultantes fueron secuenciadas para verificar que las mutaciones correctas fueron incorporadas. El plásmido que contenía el DAT 4978 T243N fue transformado entonces en el anfitrión de expresión E. coli BL21 (DE3).

DAT de B. sphaericus

Una D-aminotransferasa de B. sphaericus (número ATCC 10208) fue clonada como se describe en US 2006/0252135. La proteína fue preparada como se describe en la misma referencia.

Ejemplo 7 – Análisis de DATs

Los polipéptidos DAT que tienen las secuencias mostradas en SEQ ID NO: 928, 930, 932, 934, 936, 938, 940, 942, 944, 946, 948, y 950 fueron producidos expresando el ácido nucleico correspondiente en los vectores y en los anfitriones de expresión de *E. coli* compatibles descritos en el Ejemplo 2. Una persona experimentada en la técnica puede sintetizar los genes que codifican estas D-aminotransferasas utilizando técnicas de ensamblaje por PCR tales como las descritas en el Ejemplo 4. Los cultivos durante la noche en medio LB que contenían carbenicilina (100 µg/mL) fueron diluidos 100X en 100 mL del mismo medio y cultivados en un matraz con deflexión de 500 mL. El cultivo fue cultivado a 30°C hasta un OD600 de 0.5 a 0.9, y la expresión de la proteína fue inducida con IPTG 1 mM durante 4 horas a la misma temperatura. Las muestras para la proteína tal fueron tomadas inmediatamente antes de la recolección. Las células fueron recolectadas por centrifugación y lavadas una vez con 10 mL de regulador de fosfato de potasio pH 7.8. Las células fueron congeladas inmediatamente a -80°C hasta que se prepararon los extractos celulares.

Los extractos de células se prepararon y desalaron como se describe en el Ejemplo 4 utilizando fosfato de potasio 100 mM como regulador para eluir y equilibrar la columna PD10. Las concentraciones de proteína total y DAT fueron determinadas según se describió.

La transaminación de la R,R monatina con piruvato como aceptor amino fue llevada a cabo como se describe en el Ejemplo 5 excepto que se utilizó R,R monatina 10 mM. El análisis inicial de los niveles de alanina, monatina y del precursor de monatina no fueron consistentes uno con otro y los resultados se consideraron cualitativos. El polipéptido de DAT que tenía la secuencia de SEQ ID NO: 248 pareció mostrar la formación del precursor de la monatina.

Para confirmación adicional de la actividad, se realizó un ensayo de formación de monatina tal como se describe en los métodos con una concentración de DAT de aproximadamente 0.2 mg/mL. Como control, se evaluó una concentración de 0.2 mg/mL de DAT de *B. sphaericus* purificado. Después de 2 y 21 horas, se tomó una alícuota y se agregó ácido fórmico hasta una concentración final de 2%, y las muestras fueron congeladas. Las muestras fueron entonces descongeladas, agitadas y filtradas. Las muestras fueron analizadas en cuanto a contenido de monatina utilizando la metodología de LC/MS/MS y en cuanto a triptófano y alanina utilizando la metodología de fluorescencia postcolumna de LC/OPA descrita en el Ejemplo 3. Los polipéptidos de DAT que tenían la secuencia de SEQ ID NO: 946 y 950 fueron capaces de formar R,R monatina bajo las condiciones probadas. El polipéptido de DAT que tenía la secuencia de SEQ ID NO: 948 mostró una pérdida de triptófano y un incremento en la formación de alanina, demostrando su actividad como una D-triptófano transaminasa. El polipéptido de DAT que tenía la secuencia de SEQ ID NO: 946 se expresó bien tal como fue determinado según la cantidad de proteína total pero no fue muy soluble, lo cual explica algunos resultados inconsistentes. Los polipéptidos de DAT que tenían las secuencias mostradas en SEQ ID NO: 930, 932, 940, 942 y 944 no produjeron bandas visibles por análisis con SDS-PAGE y, por lo tanto, pueden ser activos si se producen bajo condiciones diferentes. Véase la Tabla 12 para los resultados.

Tabla 12. Actividad de DATs

Polipéptido de DAT (SEQ ID NO:)	Monatina [mM] Tiempo = 2 horas	Monatina [mM] Tiempo = 21 horas
928	nd	nd
930	nd	nd
932	nd	nd
934	nd	nd
936	nd	nd
938	nd	nd
940	nd	nd
942	nd	nd
944	nd	nd

Polipéptido de DAT (SEQ ID NO:)	Monatina [mM] Tiempo = 2 horas	Monatina [mM] Tiempo = 21 horas
946	0.1	0.6
948	nd	nd
950	0.4	3.1
DAT de control de <i>B.sphaericus</i>	0.8	4.4
nd=no detectado bajo las condiciones probadas		

Análisis de los polipéptidos de DAT en pET30a

- Los polipéptidos de DAT que tenían la secuencia de SEQ ID NO: 946, 948 y 950 fueron subclonados en pET30a como se describe en el Ejemplo 4. Se cultivaron cultivos duplicados de la cepa de *E. coli* BL21 DE3 que contenía los DATs en pET30 durante la noche en Overnight Express II (Solución 1-6, Novagen) tanto a 25 como a 30°C. Como control, también se cultivo una cepa que contenía plásmido pET30a sin un inserto. Las células fueron recolectadas a una OD600 de 5-10. Las células fueron recolectadas por centrifugación y lavadas con 10 mL de regulador de fosfato de potasio 100 mM pH 7.8. Las células fueron congeladas a -80°C hasta que fueron procesadas posteriormente.
- Los extractos celulares fueron preparados como se describe en el Ejemplo 4 utilizando fosfato de potasio 100 mM como regulador para eluir y equilibrar las columnas PD10. Las concentraciones de proteína total y de proteína DAT fueron determinadas según se describió. El polipéptido de DAT que tenía la secuencia de SEQ ID NO 946 se expresó bien a 30°C en la fracción de proteína total, pero no fue soluble según se observó por SDS-PAGE. Los polipéptidos de DAT que tenían la secuencia de SEQ ID NO: 948 y 950 se expresaron a la temperatura mayor también, pero eran solubles.
- Se realizó un ensayo de formación de monatina como se describe en el Ejemplo 5, excepto que con una concentración de DAT de 0.1 mg/mL para el polipéptido de SEQ ID NO: 946 (0.5 mg/mL para todos los demás). Como control positivo, se probó también DAT purificado de *B. sphaericus* a una concentración de 0.5 mg/mL. Después de 2 y 21 horas, se tomó una alícuota, se agregó ácido fórmico hasta una concentración final del 2% y las muestras fueron congeladas hasta ser procesadas posteriormente. Las muestras fueron descongeladas entonces, agitadas y filtradas. Las muestras fueron analizadas en cuanto a monatina utilizando LC/MS/MS, y para triptófano y alanina utilizando métodos de detección por fluorescencia postcolumna LC/OPA descrita en el Ejemplo 3. Los resultados se muestran en la Tabla 13. El consumo del D-triptófano y la formación de alanina fueron evidentes para todas las D-aminotransferasas probadas indicando que todas ellas tienen actividad sobre el D-triptófano. Bajo estas condiciones, solamente los polipéptidos de DAT que tienen la secuencia de SEQ ID NO: 946 y 948 parecen tener actividad para la formación de monatina. Es posible que las diferencias de expresión o estabilidad entre los dos sistemas anfitriones es la razón por la cual se ven en algunos casos pero no en otros.

Tabla 13

Polipéptido de DAT (SEQ ID NO:)	Monatina [mM] Tiempo = 2 horas	Monatina [mM] Tiempo = 21 horas
pET30 (control negativo)	nd	nd
946	0.4	1.8
948	nd	0.2
950	nd	nd
Control positivo de <i>B.sphaericus</i>	1.8	8.6
nd, no detectado bajo las condiciones probadas		

Ejemplo 8 – Análisis de DATs en pSE420-cHis

Los polipéptidos de DAT que tienen la secuencia mostrada en los DATs de SEQ ID NO: 886, 888, 890, 892 y 894 fueron producidos a partir del vector pSE420-cHis en *E. coli* HMS174. Una persona experimentada en la técnica puede sintetizar los genes codificando estas D-aminotransferasas utilizando técnicas de ensamblaje por PCR tales como las descritas en el Ejemplo 4. Los cultivos durante la noche de los diversos constructos de DAT fueron cultivados en medio LB que contenía ampicilina (100 µg/mL) a 30°C. 50 mL del mismo medio fueron inoculados al día siguiente con 1 mL de los cultivos de la noche. Los cultivos fueron cultivados a 30°C hasta que la OD_{600nm} alcanzó aproximadamente 0.5 y luego fueron inducidos con IPTG 1 mM. Los cultivos fueron incubados adicionalmente durante 4 horas a 30°C y luego recolectados por centrifugación a 3,800 rcf durante 15 minutos las células fueron lavadas con 1.5 mL de fosfato de potasio 50 mM, pH 7.0 y luego centrifugadas de nuevo. El sobrenadante fue decantado y las pellas de células fueron pesadas.

Los extractos celulares fueron preparados como se describió en los métodos utilizando fosfato de potasio 100 mM como regulador para eludir y equilibrar la columna. Las concentraciones totales y de DAT fueron determinadas como se describe excepto que se utilizó BCA (Pierce) en vez de Bradford para la determinación de proteínas totales. Se cultivaron solamente dos cultivos de vectores diferentes en los mismos anfitriones *E. coli*, como los DAT clonados. Todas las proteínas produjeron bandas visibles en los geles de SDS-PAGE, pero diferentes grados de solubilidad. Los polipéptidos que tenían la secuencia de SEQ ID NO: 892 no eran muy solubles.

Para comparar estas actividades de transaminación de D-triptófano de cada una de las enzimas, se ejecutaron los ensayos de transaminación de D-triptófano y de transaminación de R,R monatina descritos en el Ejemplo 5. La aminotransferasa direccionada a D-triptófano utilizando una concentración final de 0.5 mg/mL de extracto celular que contenía D-aminotransferasa y 0.1 mg/mL del DAT de *B.sphaericus* purificado como control. La cuantificación de los DAT en los extractos celulares fue difícil debido a los bajos niveles de polipéptidos solubles. Los polipéptidos de DAT que tenían la secuencia mostrada en SEQ ID NO: 888, 892 y 894 mostraron buena actividad con D-triptófano como sustrato durante la reacción de 30 minutos. Los polipéptidos de DAT que tenían la secuencia mostrada en SEQ ID NO: 886 y 890 tenían actividad medible por encima del control no enzimático, pero exhibieron poca actividad bajo las condiciones probadas.

Los experimentos de transaminación de monatina fueron llevados a cabo a temperatura ambiente, tomando muestras después de 0.5, 1 y 2 horas direccionando 0.5 mg/mL de cada DAT, incluyendo el control purificado positivo de *B. sphaericus*. Las muestras de transaminación de R,R monatina fueron analizadas entonces en cuanto al contenido de monatina y alanina. La cantidad de monatina restante fue cuantificada por LC/MS/MS. La formación de alanina fue medida utilizando el método de derivación postcolumna del Ejemplo 3. Bajo las condiciones probadas, los polipéptidos de DAT que tenía la secuencia mostrada en SEQ ID NO: 892 y 894 eran activos. El polipéptido de DAT que tenía una secuencia de SEQ ID NO: 894 pareció tener la actividad más alta para la conversión de R, R monatina en R-MP. Las tendencias fueron consistentes cuando se probó la formación de alanina. Los números de producción de alanina (en mM) para los diversos puntos en el tiempo se muestran en la Tabla 14.

Tabla 14. Formación de alanina (mM) a partir de reacciones de transaminación de R,R monatina

Polipéptido de DAT (SEQ ID NO:)	0.5 horas	1 hora	2 horas
Vector de control 1	0.139	0.185	0.215
Vector de control 2	0.179	0.242	0.301
886	0.128	0.203	0.242
888	0.13	0.203	0.275
890	0.112	0.153	0.176
892	1.034	1.587	2.167
894	2.2	2.52	2.663
BsphDAT (purificado)	0.287	0.519	0.894
Sin enzima	0.043	0.035	0.037

Los ácidos nucleicos de DAT que tenían la secuencia mostrada en SEQ ID NO: 891 y 893 fueron subclonados en pET30 tal como se describió en el Ejemplo 4. Estos constructos fueron transformados en una variedad de anfitriones *E. coli* que portaban el lisógeno DE3 para la expresión a partir de un promotor T7, incluyendo las dos cepas K-12 y B de *E. coli*, y una cepa que portaba el plásmido pLysS. Los clones fueron expresados en el Overnight Express System II del Ejemplo 4, con y sin la adición de piridoxina 0.5 mM, y analizados por SDS-PAGE o Experion en cuanto a su expresión. A partir de estos experimentos, se hizo evidente que las proteínas se estaba expresando mayormente en la fracción insoluble. La piridoxina ayudó a mejorar la solubilidad hasta un grado pequeño como lo hizo el descenso de la temperatura de 37 a 30°C para inducción. Se hizo trabajo adicional en los sistemas de clonación diseñado para maximizar la expresión soluble (véanse Ejemplos 16-22).

10 **Ejemplo 9 – Análisis de CaDAT, CbDAT y LsDAT en pET30a**

La secuencia de aminoácidos mostrada en SEQ ID NO: 894 fue utilizada para buscar proteínas similares disponibles en las bases de datos públicas. Se encontraron tres DAT que tenían similitud con SEQ ID NO: 894. Eran de *Lactobacillus salivarius* (47% idéntica al nivel de proteína), *Clostridium beijerinckii* (57% idéntico al nivel de proteína), y *Clostridium acetobutylicum* (60% idéntico al nivel de proteína). Las secuencias genéticas y de proteínas y sus números de acceso se muestran al final de este ejemplo. La figura 5 es un alineamiento (una comparación de secuencias) que muestra las regiones de consenso de esta proteína similares a SEQ ID NO: 894 de esta invención. Como se resalta en las cifras, puede verse un alto grado de regiones de consenso lo que indica similitudes estructurales.

Estos ácidos nucleicos fueron clonados en pET30a, y los polipéptidos correspondientes fueron expresados y probados en cuanto a su actividad como se describe aquí. CbDAT.

La D-aminotransferasa de *Clostridium beijerinckii* (CbDAT) fue clonada en pET30 (no etiquetada) BL21 (DE3) y expresada utilizando OVERNIGHT EXPRESS IITM (Novagen). Las células fueron recolectadas a una densidad óptica a 600 nm de aproximadamente 9 y centrifugadas a 4000 rcf durante 15 minutos. Las células fueron lavadas una vez con fosfato de potasio 100 mM pH 7.8 (frio), y se agitaron de nuevo.

Los extractos celulares fueron preparados como se describe aquí utilizando EPPS 100 mM pH 8.2 como regulador para eluir y equilibrar la columna. Las concentraciones de proteína total y proteína de DAT fueron determinadas como se describe excepto que se utilizó el método BCA (Pierce) en vez del ensayo de Bradford (Coomassie). La CbDAT se expresó bien pero solamente era parcialmente soluble.

Se realizó un ensayo de formación de monatina tal como se describe en el Ejemplo 5 pero la actividad de CbDAT (0.5 mg/mL) fue estudiada también a pH 7.4 (con fosfato de potasio como regulador). Como control, se probó DAT de *B. sphaericus* purificado (1 mg/mL) a pH 8.2. Después de 1, 2, 4, 8 y 23 horas, se tomaron alícuotas y se agregó ácido fórmico hasta una concentración final de 2%. Las muestras fueron congeladas a -80°C hasta su análisis. Las muestras fueron luego descongeladas, agitadas y filtradas. Las muestras fueron analizadas en cuanto al contenido de monatina utilizando la metodología de LC/MS/MS descrita en el Ejemplo 3. Los resultados se muestran en la Tabla 15. La cantidad de monatina producida fue ligeramente superior para los ensayos llevados a cabo a pH 8.2. Se llevaron a cabo experimentos similares con los polipéptidos expresados a partir de pET28 con una etiqueta His N terminal. La actividad de la versión etiquetada pareció ser ligeramente menor que la de la no etiquetada, pero todavía fácilmente detectable.

Tabla 15. Actividad con el tiempo

Enzima de DAT	Monatina (ppm) 1 horas	Monatina (ppm) 2 horas	Monatina (ppm) 4 horas	Monatina (ppm) 8 horas	Monatina (ppm) 23 horas
CbDAT pET30 (7.4 mg/mL)	45	126	280	428	502
CbDAT pET30 (8.2 mg/mL)	67	189	344	436	568
<i>B. sphaericus</i> DAT (1 mg/mL)	531	968	1742	2310	3706

CaDAT y LsDAT

Las D-aminotransferasas de *Lactobacillus salivarius* (LsDAT) y *Clostridium acetobutylicum* (CaDAT) fueron clonadas en pET30a (no etiquetadas) BL21 (DE3) y expresadas utilizando el Overnight Express II (Novagen). Las células

fueron recolectadas cuando el cultivo alcanzó una densidad óptica a 600 nm de aproximadamente 9 por centrifugación a 4000 rcf durante 15 minutos.

Los extractos celulares fueron preparados como se describe aquí utilizando EPPS 100 mM pH 8.2 como regulador para eludir y equilibrar la columna. Las concentraciones de proteína total y proteína de DAT fueron determinadas utilizando el protocolo BCA (Pierce). Ambas enzimas se expresaron bien y eran solubles.

El ensayo fue llevado a cabo a temperatura ambiente bajo condiciones anaeróbicas. Como control, se ensayó D-aminotransferasa de *B. sphaericus* purificada. Se usaron aproximadamente 0.5 mg/mL de cada DAT. Después de 0.5, 1, 2, 4, 6, 8 y 22 horas se tomó una alícuota y se agregó ácido fórmico hasta una concentración final de 2% y las muestras fueron congeladas. Las muestras fueron entonces descongeladas, agitadas y filtradas. Las muestras fueron analizadas en cuanto a monatina utilizando la metodología de LC/MS/MS descrita en el Ejemplo 3. Los resultados se muestran en la Tabla 16.

Tabla 16

Polipéptido de DAT	Monatina (ppm) 0.5 horas	Monatina (ppm) 1 hora	Monatina (ppm) 2 horas	Monatina (ppm) 4 horas	Monatina (ppm) 6 horas	Monatina (ppm) 8 horas	Monatina (ppm) 22 horas
<i>B. sphaericus</i>	76.6	194.4	457.6	860.8	1186	1770	2546
LsDAT	2.8	6.4	14.6	33	52	69.8	173
CaDAT	50.2	141.2	318.6	543.4	612	1144	668

Los homólogos de DAT que tienen la secuencia mostrada en SEQ ID NO: 894 eran activos. Puesto que los homólogos que muestran la secuencia conservada más arriba eran activos en los ensayos de formación de monatina, se espera que cualquier D-aminotransferasa que contiene las secuencias de consenso descritas aquí también sea activo, aunque su identidad de secuencia primaria sea tan baja como 47%. No ha habido evidencia antes de este trabajo que estas D-aminotransferasas únicas, con baja homología con la D-aminotransferasa de *Bacillus* más caracterizada, tuvieran actividad para monatina o fueran enzimas de especificidad amplia.

Secuencia de ADN CaDAT (ACCESSION AE001437 AE007513-AE007868; VERSION AE001437.1 GI: 25168256; nucleótidos 914049. 914891)

```

1  atgaaagatt taggatatta caatggagaa tacgacttaa ttgaaaatat gaaaatacca
61  atgaatgata gtgtatgcta ttttgggtgat ggtgtttatg atgctactta tagtagaac
121 cataatatat ttgcactaga tgagcatatt gaccgatttt ataatagtgc cgagctttta
181 agaattaaaa ttccatatac aaagaaggaa atgaaagagc ttttaaagga tatggttaaa
241 aaggttgata gcggagaaca atttgtatat tggcagggtta ctagagggtac tggcatgcgt
301 aatcatgctt ttttgagtga ggatgttaag gctaataattt ggattgtttt aaagccacta
361 aaggtaaaag atatgtcaaa aaaattaaaa ctaataacat tagaggatac tagattttta
421 cattgtaaca taaaaacctt aaatttgctt cctagtgtaa ttgcagcaca aaaaactgaa
481 gaagcaggct gccaggaagc agtatttcat agaggagata gagttactga atgtgctcat
541 agtaatgttt caattataaa ggatgagatt ttaaaaactg cgccaacaga taatcttatt
601 ttgccgggaa tagcaagggc gcatcttata aaaatgtgca aaaaatttga gatacctgta
661 gatgaaactc cattttacatt aaaggagtta attaattgcg atgaagttat agttacaagt
721 tcagggcaat tttgtatgac tgcttgtgag atagatggaa gacctgtagg cggaaaagcg
781 ccagatatta ttaaaaagct tcagactgcc ttacttaattg aatttttgga agaaacaaat
841 taa (SEQ ID NO:1063)

```

Secuencia de proteínas CaDAT (ACCESSION NP_347428; VERSION NP_347428.1 GI: 15894079)

```

1  MKDLGYNGE YDLIENMKIP MNDRCYFGD GVDATYSRN HNIFALDEHI DRFYNSAELL
61  RIKIPYTKKE MKELLKDMVK KVDSGEQFVY WQVTRGTGMR NHAFLSEVVK ANIWIVLKPL
121 KVKDSMKLLK LITLEDTRFL HCNIKTLNLL PSVIAAQKTE EAGCQEA VFH RGRVTECAH
181 SNVSIKDEI LKTAPTDNLI LPGIARAH LI KMCKKFEIPV DETPFTL KEL INADEVIVTS
241 SGQFCMTACE IDGRPVGGKA PDIKKLQTA LLNEFLEETN (SEQ ID NO:1064)

```

Secuencia de ADN CbDAT (ACCESSION CP000721 AAL001000000 AAL001000001-AAL001000089 VERSION CP000721.1 GI:149901357; nucleótidos 3213484.. 3212636)

```

1  atggagaatt taggttatta taatggaaag tttggattat tagaggaaat gacagtacca
61  atgcttgatc gtgtttgcta ttttggagat ggagtttatg atgctactta tagcagaaat
121 cacaagattt ttgcattgga ggagcatatt gaaagatttt acaacagcgc tggtttatta
181 ggaattaaaa ttctttattc aaaggagcaa gtaaaagaaa tccttaagga gatggtatta
241 aaggttgatt caggagaaca atttgtatat tggcaaatta ctagaggaaac tggaatgaga
301 aatcatgctt ttcttgga tgaggtccct gcaaacttat ggattatggt aaagccttta
361 aatattaagg atatgtcaca aaaattaaag ttaatcactt tagaagacac tagattttta
421 cactgtaata tcaaaacctt aaatttatta ccaagtgtaa ttgcatctca aaaaactgaa
481 gaggcaggat gtcaggaagc tgtatttcat agaggggata gagtaactga atgtgcacat
541 agcaatgtat caattattaa ggatggtata ttaaaaactg ctccaacaga caatttaatt
601 ttaccaggta tagcaagagc tcaccttatt aaaatgtgta aatcctttta tattcctgta
661 gatgaaacag cattttacctt gaaggaatta atggaggcag atgaagttat agttactagt
721 tcaggtcaat tttgtatggc aaccagttaa atagatggaa tacctgtagg gggaaaagca
781 ccagagcttg taaagaaatt acaagatgca ttgttaaatg agttcttaga agaaacaaa
841 acagaatag (SEQ ID NO:1065)

```

Secuencia de proteína CbDAT (ACCESSION YP_001309869 VERSION YP_001309869.1 GI:150017615)

```

1  MENLGYNGK FGLLEEMTVP MLDRVCFYGD GVDATYSRN HKIFALEEHI ERFYNSAGLL
61  GIKIPYSKEQ VKEILKEMVL KVDSEGFVY WQITRGTMGR NHAFPGDEVP ANLWIMLKPL
121 NIKDMSQKLK LITLEDTRFL HCNIKTLNLL PSVIASQKTE EAGCQEAVFH RGDRVTECAH
181 SNVSIIKDKI LKTAPTDNLI LPGIARAHLI KMCKSFNIPV DETAFTLKL MEADEVIVTS
241 SGQFCMATSE IDGIPVGGKA PELVKKLQDA LLNEFLEETK TE (SEQ ID NO:1066)

```

Secuencia de ADN LsDAT (ACCESSION CP000233 VERSION CP000233.1 GI:90820184; nucleótidos 1750082.. 1750927)

```

1  atgaagcaag ttggatacta caatggtact atcgctgatt taaatgaact taaggtgcct
61  gctactgatc gtgcacttta ttttggatgat gggttgctacg atgcaactac atttaagaac
121 aatggtgcat ttgccttaga agatcatctt gatcggtttt ataatagttg tcgcctacta
181 gagatcgatt tccctttaaa tcgcgatgaa cttaaagaaa agctttacgc tgttattgat
241 gctaacgaag ttgatactgg tctcctttat tggcaaactt cacgtggttc tggtttacgt
301 aaccatattt tcccagaaga tagccaacct aatttattaa tttttactgc tccttatggt
361 ttagttccat ttgatactga atataaactt atatctcgcg aagacactcg cttcttacat
421 tgcaatatta aaactttgaa tttacttcca aacgttattg caagtcaaaa ggctaataaa
481 agtcattgcc aagaagtggc cttccatcgc ggtgacagag ttacagaatg tgcacactct
541 aacatcttaa ttctaaaaga tggcggttctt tgctcccccac ctagagataa ttaaatcttg
601 ccaggaatta ctttgaaaca tctcttgcaa ttagcaaaaag aaaataatat tctacttcc
661 gaagcaccat tcaactatgga tgatcttaga aatgctgatg aagttattgt tagttcttca
721 gcttgtctag gtattcgcg agtcgagctt gatggtcagc ctggttggtg aaaagatgga
781 aagactttta agatcttgca agatgcttat gctaagaaat ataatgctga aactgtaagt
841 cggttaa (SEQ ID NO:1067)

```

Secuencia de proteína LsDAT (ACCESSION YP_536555 VERSION YP_536555.1 GI:90962639)

```

1  MKQVGYNGT IADLNELKVP ATDRALYFGD GCYDATTFKN NVAFALEDHL DRFYNSCRLL
61  EIDFPLNRDE LKEKLYAVID ANEVDIGILY WQTSRGSGLR NHIFPEDSQP NLLIFTAPYG
121 LVPFDTEYKL ISREDTRFLH CNIKTLNLLP NVIASQKANE SHCQEVVFHR GDRVTECAHS
181 NILILKDGVL CSPPRDNLIL PGITLKHLLQ LAKENNIPTS EAPFTMDDLR NADEVIVSSS
241 ACLGIRAVEL DGQPVGGKDG KTLKILQDAY AKKYNAETVS R (SEQ ID NO:1068)

```

Ejemplo 10 – Análisis de los DATs

E. coli HMS 174 que contiene los ácidos nucleicos DAT que tiene las secuencias de SEQ ID NO: 865, 867, 869, 871, 873, 875 y 877 en el vector pSE420-cHis fueron obtenidos y sembrados sobre placas de agar que contenían medio LB con ampicilina. Una persona experimentada en la técnica puede sintetizar los genes que codifican estas D-

aminotransferasas utilizando técnicas de ensamblaje por PCR tales como las descritas en el Ejemplo 4. Se utilizaron colonias sencillas para inocular 3 mL de medio LB que contenía ampicilina (100 µg/mL). 500 µl del cultivo nocturno fueron utilizados para inocular 50 mL del mismo medio en 250 mL en matraces con deflexión. Las células fueron cultivadas a 30°C hasta aproximadamente una OD_{600nm} de 0.5. Se agregó IPTG hasta una concentración final de 1 mM. Las células fueron inducidas a 30°C durante 4 horas y recolectadas por centrifugación.

Se prepararon los extractos celulares como se describió en el Ejemplo 4. Las concentraciones totales de proteína y DAT se determinaron como se describió en el Ejemplo 4. Los DATs parecen expresarse todos bien, y la mayoría de ellos mostraron un alto grado de solubilidad.

Se realizó un ensayo de formación de monatina como se describe en el Ejemplo 5 excepto con una concentración de DAT de 0.1 mg/mL y la aldolasa a una concentración de 10 µg/mL. Como control, se probaron 0.1 mg/mL de DAT de *B. sphaericus* purificado. Después de 6 y 22 horas, se tomó un alícuota y se agregó ácido fórmico hasta una concentración final de 2%, y las muestras fueron congeladas. Las muestras fueron entonces descongeladas, agitadas y filtradas. Las muestras fueron analizadas en cuanto a las concentraciones de monatina utilizando la metodología de LC/MS/MS descrita en el Ejemplo 3. Bajo las condiciones probadas, las SEQ ID NO: 870 874 y 878 parecieron tener todas alta actividad en el ensayo de formación de monatina de tres etapas. Los polipéptidos de DAT que tienen la secuencias mostradas en SEQ ID NO: 866, 872 y 876 también tienen actividad en la ruta de formación de monatina pero no hasta el mismo grado como los polipéptidos que tiene la secuencia mostrada en SEQ ID NO: 870, 874 y 878 bajo las condiciones probadas. La Tabla 17 muestra los resultados para la formación de monatina (en ppm).

Tabla 17. Ensayo y formación de monatina

Polipéptido de DAT (SEQ ID NO:)	6 horas	22 horas
866	17.4	76
868	nd	nd
870	132	836
872	13.8	50
874	281.6	798
876	2.4	12
878	223.4	576
DAT de <i>B. sphaericus</i>	175.6	616
nd, no detectado bajo las condiciones probadas		

Ensayos adicionales de polipéptidos que tienen la secuencia de SEQ ID NO: 870, 874 y 878 en pET30a

Cultivos de *E. coli* BL21 DE3 transformados con plásmidos pET30a que contienen ácidos nucleicos que codifican los DAT antes indicados fueron cultivados durante la noche en 50 mL de Overnight Express II (Solución 1-6, Novagen) a 30°C. Como control positivo, una cepa que contenía los DAT de ATCC # 4978 en pET30a también fue cultivada e inducida (descrita en el Ejemplo 6). Las células fueron recolectadas a una OD_{600nm} de 5-10, recolectada por centrifugación y congelada a menos 80°C hasta su posterior procesamiento.

Los extractos celulares fueron preparados como se describió en el Ejemplo 4. Las concentraciones de proteína total y de DAT fueron determinadas como se describió en el Ejemplo 4.

Se realizó un ensayo de formación de monatina como se describe en el Ejemplo 5 excepto con una concentración de polipéptido de DAT de 0.5 mg/mL para SEQ ID NO: 870 una concentración de 0.275 mg/mL para cada uno de SEQ ID NO: 874 Y 878. Como control, se ensayaron DAT 4978 y DAT de *B. sphaericus* purificado a una concentración de 0.5 mg/mL. Después de 0.5, 1, 2, 4, 6.5, 9, 24 y 22 horas, se tomó una alícuota y se agregó ácido fórmico hasta una concentración final de 2% y las muestras fueron congeladas. Las muestras fueron luego descongeladas, agitadas y filtradas. Las muestras fueron analizadas en cuanto a monatina utilizando la metodología de LC/MS/MS descrita en el Ejemplo 3. Los resultados se muestran en la Tabla 18 (en ppm de monatina formada).

Tabla 18

Polipéptidos de DAT (SEQ ID NO:)	0.5 horas	1 hora	2 horas	4 horas	6.5 horas	9 horas	24 horas
4978 DAT	18	85.6	283.4	673.2	890	1226	2020
870	14.4	71	279.4	736	1340	1680	3362
874	63.8	182.6	415.6	674	888	938	1154
878	97.8	244.4	607	912.2	1068	1174	1356
B. sphaericus	44.6	142.8	375.2	813	1294	1382	2746

- 5 Todos los tres DATs subclonados (que codifican polipéptidos que tienen la secuencia de SEQ ID NO: 870, 874 y 878) se expresaron bien en el sistema pET y produjeron proteína soluble. El polipéptido que tenía la secuencia mostrada en SEQ ID NO: 870 dio la cantidad más alta de expresión en la fracción soluble y exhibió alta actividad que no pareció disminuir con el tiempo en comparación con los polipéptidos de DAT que tenía la secuencia de SEQ ID NO: 874 y 878.

Comparación entre polipéptidos tipo silvestre y DAT mutantes

- 10 Un polipéptido mutante en el cual el residuo SEQ ID NO: 870 fue cambiado de una T a una N (SEQ ID NO: 870 T242N) fue construido como se describe en el Ejemplo 4 y se expresó y comparó con DAT4978, DAT4978 T243N (descrito en el Ejemplo 6), B. sphaericus y SEQ ID NO: 870 tipo silvestre.

- 15 Los cultivos de BL21 DE3 en los cuales los polipéptidos de tipo silvestre mutantes que tenían la secuencia de SEQ ID NO: 870, 870 T242N, DAT4978 y DAT4978 T243N fueron expresados a partir del vector pET30a, fueron cultivados en 50 mL de Overnight Express (Novagen) en un matraz con deflexión de 250 mL durante la noche a 30°C y 250 rpm. Las células fueron recolectadas por centrifugación cuando alcanzaron una densidad óptica a 600 nm por encima de 10. Los extractos celulares fueron preparados como se describió en el Ejemplo 4 y se determinaron las concentraciones de proteína total y DAT tal como se describe en el Ejemplo 4. Todos los DATs probados fueron altamente expresados y solubles, todo cerca del 30% según se determinó utilizando el software Experion. El polipéptido que tiene la secuencia de SEQ ID NO: 870 T242N tenía la expresión más alta, la cual fue predicha en 36.3% de la proteína soluble total.

- 25 Se realizó un ensayo de formación de monatina como se describe en el Ejemplo 5 a una concentración de polipéptidos de DAT de 0.5 mL. Como control, se probaron 0.5 mg/mL de DAT purificados de B. sphaericus. Después de 0.5, 1, 2, 4, 6.5, 9 y 23.25 horas, se tomó una alícuota, se agregó ácido fórmico a la alícuota hasta una concentración final del 2% y las muestras fueron congeladas. Las muestras fueron luego descongeladas, agitadas y filtradas. Las muestras fueron analizadas en cuanto a monatina, triptófano, alanina y ácido 4-hidroxi-4-metil glutámico (HMG) como se describe en el Ejemplo 3.

En el último punto de tiempo, se tomó una alícuota adicional para determinar el porcentaje de R,R monatina mediante el método de derivación con FDA descrito en el Ejemplo 3.

- 30 Los números de formación de monatina (ppm) se presentan en la Tabla 19 a continuación. El porcentaje de R,R está dado en la columna de la derecha, para el punto de tiempo de 23.25 horas.

Tabla 19

Polipéptidos de DAT (SEQ ID NO:)	0.5 horas	1 hora	2 horas	4 horas	6.5 horas	9 horas	23.25 horas	% R,R
DAT 4978 tipo silvestre	11	57	216	472	694	942	1616	95.0

Polipéptidos de DAT (SEQ ID NO:)	0.5 horas	1 hora	2 horas	4 horas	6.5 horas	9 horas	23.25 horas	% R,R
4978 T243N	74	237	542.6	1106	1396	1784	2202	99.0
870	15.6	74.4	269.6	702	1250	1522	2788	97.8
870 T242N	49.4	194	655.2	1496	2212	2666	3670	99.5
B. sphaericus	40.6	144	372	800	1090	1458	2434	97.2

La actividad del mutante T242N del polipéptido de SEQ ID NO: 870 fue muy alta, y fue mejor que las de los controles positivos y mayor que la de la forma tipo silvestre de los polipéptidos de DAT. El porcentaje de R,R monatina formado por este mutante también fue superior al de cualquiera de las otras enzimas comparadas. El análisis de la cantidad de HMG (es un subproducto) formado es cualitativo, pero parece ser que, en los puntos de tiempo de 9 horas y 23.25 horas, se formaron cantidades similares de HMG por los polipéptidos de DAT 4978 T243N y los polipéptidos de SEQ ID NO: 870 T242N.

El polipéptido DAT de ejemplo de esta invención que tiene la secuencia mostrada en SEQ ID NO: 870 es una proteína novedosa, que exhibe 76% de identidad en secuencia con la D-aminotransferasa conocida más cercana (D-aminotransferasa del Bacillus YM-1) y 69% de identidad de secuencia de aminoácidos con los DAT de B. sphaericus descritos en el Ejemplo 6. La figura 6 muestra un alineamiento de esta novedosa enzima de la invención con otros DAT publicados, y se puede ver que los residuos son una propiedad que hace que esta enzima sea única y pueda ser tenida en cuenta por su actividad superior.

El polipéptido de DAT altamente activo que tiene la secuencia mostrada en SEQ ID NO: 910 (más similar a los DAT tipo B. sphaericus; véase Ejemplo 12) también se muestran en el alineamiento. Como ejemplo del carácter único del polipéptido de SEQ ID NO: 870, en la región que circunda el recibo de aminoácidos 54-55 (numeración de B. sphaericus) en el alineamiento de la figura 6, es claro que el DAT similar al Bacillus tiene un alto grado de conservación mientras que SEQ ID NO: 870 tiene los residuos EC en vez de los AS. En otro ejemplo, en la región altamente conservada que circunda el residuo 135 del alineamiento mostrado en la figura 6, el polipéptido de SEQ ID NO: 870 tiene un residuo más hidrofílico (T) versus residuos que son predominantemente valina. La secuencia central que representa la enzima de SEQ ID NO: 870, pero que excluye D-aminotransferasa de especificidad amplia previamente conocidas y homólogos altamente relacionados, se muestra como Secuencias de Consenso A y B. se espera que cualquier polipéptido que contenga una de estas secuencias de consenso exhibirá actividad de DAT y será activada en las etapas de la ruta de formación de monatina.

Secuencia de Consenso A

Y.*LWND.*IV.*EDRGYQFGDG.*YEV.*KVY.*G.*FT.*EH.*DR.*YEC AEKI.*PYTK.*H.*LLH.*L.*E.*N.*TG.*YFQ.*TRGVA.*RVHNFPAGN.*Q.*V.*SGT.*K.*F.*R.*N.*KGVKAT.*TED.*RWLRCDIKSLNLLGAVLAKQEAIEKGCYEA.*LHR.*G.*TE.*SS.*N.*GIK.*GTLYTHPA.*N.*ILRGITR.*V.*TCAKEIE.*PV.*Q.*T.*K.*LEMDE.*V.*S.*SE.*TP.*I.*DG.*KI.*NG.*G.*WTR.*LQ.*F.*K.*P (SEQ ID NO:1069).

Secuencia de Consenso B

Y[ST]LWND[QK]IV.[DE].{2}[VI].[IV].{2}EDRGYQFGDG[IV]YEV[IV]KVY[ND]G. [ML]FT.{2}EH[IV]DR.YECAEKI[RK][LIV].[IV]PYTK.{3}H[QK]LLH.L[VI]E.N.[LV]. TG[HN][IV]YFQ[IV]TRGVA.RVHNFPAGN[VI]Q.V[LI]SGT.K.F.R.{3}N.[EQ]KGV KAT.TED[IV]RWLRCDIKSLNLLGAVLAKQEAIEKGCYEA[IV]LHR.G.[VI]TE.SS.N[VI][FY]GIK[DN]GTLYTHPA[ND]N.ILRGITR.V[IV][LI]TCAKEIE[LMI]PV.[EQ]Q. { 2}T.K.{2}LEMDE[LIVM].V[ST]S.[TS]SE[IV]TP[VI][DE][IVL]DG.KI.NG.{2}G[ED] WTR[KQ]LQ.{2}F.{2}K[IL]P. (SEQ ID NO:1070).

De manera similar al lenguaje convencional de expresión regular PERL, “.” indica que cualquier número de residuos de aminoácidos pueden estar presente en la forma de 20 aminoácidos proteinogénicos; [] indica que cualquiera de los aminoácidos en dos paréntesis puede estar presente; “{#}” indica que cualquiera de los 20 aminoácidos proteinogénicos puede estar presente en tanto el número de residuos coincida con el número {#} indicado en los paréntesis.

Con respecto al uso de “.” en la Secuencia de Consenso A, el número de aminoácidos en cualquiera de las posiciones “.” puede variar, por ejemplo, desde 0 hasta aproximadamente 20 residuos (véase, por ejemplo, la Secuencia de Consenso B (SEQ ID NO: 1070)) o desde aproximadamente 20 residuos hasta aproximadamente 100

residuos, o el número de aminoácidos puede ser mucho mayor, por ejemplo, hasta 1000 o más residuos. Sin limitación, una inserción en una o más de las posiciones “.” puede corresponder, por ejemplo, a un dominio tal como (pero no limitándose a) un dominio de enlazamiento de quitinasa (por ejemplo, de *Pyrococcus furiosus* (Accesión No. 2CZN_A) o *P. burkholderia* (Accession No. YP_331531) o un dominio de enlazamiento de celulosa (por ejemplo, de *Cellulomonas fimi* (Accession No. 1 EXH_A) o *Clostridium stercorarium* (Accession No. 1UYI_A). En algunas realizaciones (sin limitación), cinco o menos de las posiciones designadas “.” contienen cada una inserción de, por ejemplo, mayor de aproximadamente 20 residuos (por ejemplo, mayor de aproximadamente 100 residuos). En otras realizaciones (sin limitación), cinco o más de las posiciones designadas “.” contienen cada una inserción de al menos de aproximadamente 100 residuos (por ejemplo, menos de aproximadamente 20 residuos, por ejemplo, 3, 5, 10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90, o 95 residuos). La actividad de un polipéptido que tiene una secuencia que corresponde a uno o más de las secuencias de consenso divulgadas aquí y que contiene cualquier número de residuos insertados en una o más de las posiciones “.” puede ser evaluada utilizando métodos que son descritos aquí.

Polipéptidos representativos no limitantes que contienen la secuencia de consenso mostrada en SEQ ID NO: 1069 incluyen el polipéptido que tiene la secuencia mostrada en SEQ ID NO: 870 y la Secuencia de Consenso B (SEQ ID NO: 1070).

Comparación entre SEQ ID NO: 870, 870 T242N, DAT 4978 y DAT 4978 T243N, etiquetadas y no etiquetadas

Células de *E. coli* BL21 DE3 que expresan los polipéptidos que tienen la secuencia de SEQ ID NO: 870, 870 T242N, DAT 4978 y DAT 4978 T243N fueron cultivadas en 50 mL de Overnight Express (Novagen) en un matraz con deflexión de 250 mL durante la noche a 30°C y 250 rpm. Las células fueron recolectadas por centrifugación cuando la densidad óptica a 600 nm fue superior a 10. Los extractos celulares fueron preparados como se describió en el Ejemplo 4. Las concentraciones de proteína total y DAT fueron determinadas como se describe en el Ejemplo 4.

Los cultivos de *E. coli* BL21 DE3 que expresan los polipéptidos etiquetados C-terminal SEQ ID NO: 870, 870 T242N, DAT 4978 y DAT 4978 T243N fueron cultivados en 200 mL de Overnight Express (Novagen) en un matraz con deflexión de 1000 mL durante la noche a 30°C y 250 rpm. Se cultivaron dos cultivos de cada clon. Las células fueron reunidas y recolectadas por centrifugación cuando se alcanzó una densidad óptica a 600 nm por encima de 10. Los extractos celulares fueron creados mediante la adición de 50 mL de Bug Buster Primary Amine Free (Novagen) con 50 µL de Benzonase Nucleasa (Novagen), 0.75 µL de r-Lisozima (Novagen), y 250 mL del Cocktail II Inhibidor de Proteasa (Calbiochem). Las células fueron incubadas durante 15 minutos a temperatura ambiente con agitación suave. Los extractos fueron centrifugados a 45,000 x g durante 10 minutos.

Las proteínas etiquetadas con His fueron purificadas como se describe en la sección de métodos utilizando resina GE Healthcare (Piscataway, NJ) Chelating Sepharose™ Fast Flow. La proteína purificada fue desalinizada utilizando una columna PD10 en fosfato de potasio 100 mM, pH 7.8 (para los polipéptidos DAT 4978 y DAT 4978 T243N) o EPPS 100 mM pH 8.2 (para polipéptidos SEQ ID NO: 870 y 870 T242N), conteniendo ambos reguladores 50 µM de PLP.

Se llevó a cabo un ensayo de formación de R,R monatina como se describe en el Ejemplo 5, con concentración de DAT de 0.5 mg/mL. Se tomaron alícuotas a 2.25, 4.5, 9 y 24 horas, pH ajustado con ácido fórmico, y congeladas. Se tomó una alícuota extra en el punto de tiempo final sin adición de ácido fórmico para la determinación de la distribución estereoisomérica utilizando el método de derivación con FDAA descrito en el Ejemplo 3. Las muestras fueron descongeladas y centrifugadas durante 5 minutos y el sobrenadante fue filtrado con un filtro de membrana de nylon de 0.2 µm. Las muestras fueron sometidas al análisis de monatina utilizando el método de LC/MS/MS descrito en el Ejemplo 3. Los resultados se muestran en la Tabla 20, en ppm de monatina formada. La columna derecha última es el porcentaje de R,R monatina formado al final del experimento.

Tabla 20

Polipéptidos de DAT (SEQ ID NO:)	2.25 horas	4.5 horas	9 horas	24 horas	% R,R
DAT 4978	330	676	1470	2384	92.3
DAT 4978 T243N	1392	2856	4068	3688	98.2
870	395	952	1896	2998	97.8
870 T242N	1416	2936	3868	3976	99.3
DAT 4978 6xHis etiquetado	362	887	1664	2818	96.5

Polipéptidos de DAT (SEQ ID NO:)	2.25 horas	4.5 horas	9 horas	24 horas	% R,R
DAT 4978 T243N 6xHis etiquetado	1364	2298	3464	4440	98.9
870 6Xhis etiquetado	228	688	1508	3138	98.1
870 T242N 6xHis etiquetado	746	2020	3962	4552	99.5

La actividad global y la estereoespecificidad de las enzimas etiquetadas y no etiquetadas con C terminal fueron muy similares. Además, se espera que la presencia de actividad en un polipéptido codificado a partir de un ácido nucleico subclonado prediga la presencia de actividad en el polipéptido correspondiente codificado a partir del ácido nucleico de longitud completa o tipo silvestre.

Ejemplo 11 – Análisis de DATs

E. coli HMS 174 con contenido de ácidos nucleicos de DAT en el vector pSE420-cHis que codifica los polipéptidos que tienen la secuencia de SEQ ID NO: 880, 882, y 884, fue sembrado sobre placas de agar que contenían medio LB con ampicilina. Una persona experimentada en la técnica puede sintetizar los genes que codifican estas D-aminotransferasas utilizando técnicas de PCR de ensamblaje tales como las descritas en el Ejemplo 4. Se utilizaron colonias individuales para inocular 3 mL de medio LB que contenía ampicilina (100 µg/mL). Se utilizaron 500 µL para inocular 50 mL del mismo medio en un matraz con deflexión de 250. Las células fueron cultivadas a 30°C hasta aproximadamente una OD_{600nm} de 0.4, y se agregó IPTG hasta una concentración final de 1 mM. Las células fueron cultivadas a 30°C durante 4 horas y recolectadas por centrifugación.

Los extractos celulares fueron preparados como se describe en el Ejemplo 4. Las concentraciones de proteína total y polipéptidos de DAT fueron determinadas como se describe en el Ejemplo 4. Las SEQ ID NO: 882 y 884 se expresaron bien y estuvieron presentes a niveles altos en la fracción soluble.

Se llevó a cabo un ensayo de formación de R,R monatina como se describió en el Ejemplo 5 utilizando aproximadamente 0.5 mg/mL de cada polipéptido de DAT (excepto que se utilizaron 0.35 mg/mL del polipéptido de SEQ ID NO: 880). Después de 2, 8, y 23 horas, se tomó una alícuota, se agregó ácido fórmico hasta una concentración final de 2%, y las muestras fueron congeladas. Las muestras fueron luego descongeladas, agitadas y filtradas. Las muestras fueron analizadas en cuanto al contenido de monatina utilizando la metodología de LC/MS/MS descrita en el Ejemplo 3. Los resultados se muestran en la Tabla 21.

En el último punto del tiempo, se tomó una alícuota extra (sin ajuste de pH) para determinar la distribución estereoisomérica de la monatina producida utilizando la metodología de derivación con FDAA descrita en el Ejemplo 3. El porcentaje de R,R producido se muestra en la columna a mano derecha de la Tabla 18 más adelante, siendo el resto predominantemente S,R monatina.

Tabla 21

Polipéptido (SEQ ID NO:)	Monatina ppm (2 horas)	Monatina ppm (8 horas)	Monatina ppm (23 horas)	% R,R (23 horas)
880	31.6	140	176	97.5
882	31.6	872	2790	99.3
884	79.4	644	1610	100
B. sphaericus	337	1518	2538	96.7

Los polipéptidos que tienen la secuencia mostrada en SEQ ID NO: 882 y 884 exhibieron buena actividad en las reacciones de formación de monatina a partir de D- triptófano.

La estereopureza de la monatina producida fue superior cuando se utilizaron estos DAT en comparación con la enzima de control de B. sphaericus. Los ácidos nucleicos de DAT que codifican los polipéptidos de DAT que tienen la secuencia de SEQ ID NO: 882 y 884 fueron subclonados en vectores pET30a como se describe en el Ejemplo 4.

Análisis de péptidos de DAT que tienen la secuencia de SEQ ID NO: 882 y 884 expresados a partir del vector pET30a

Cultivos de *E. coli* BL21 DE3 que contienen ácidos nucleicos que codifican polipéptidos de DAT que tiene la secuencia de SEQ ID NO: 882 y 884 en el vector pET30a fueron cultivados en 50 mL de Overnight Express (Novagen) en un matraz con deflexión de 250 mL durante la noche a 30°C y 250 rpm. Las células fueron recolectadas por centrifugación cuando la densidad óptica a 600 nm fue superior a 10.

Se prepararon extractos celulares como se describe en el Ejemplo 4. Las concentraciones de proteína total y polipéptidos de DAT fueron determinadas como se describió. Las muestras de proteína total y soluble fueron analizadas utilizando un gel de acrilamida con un gradiente de 4-15% así como mediante el sistema Experion. Se predijo que la expresión era aproximadamente 30% mediante el software Experion. Las bandas visibles fueron estudiadas para las fracciones tanto de proteína total como de proteína soluble (extracto libre de células).

Se llevó a cabo un ensayo de formación de monatina como se describe en el Ejemplo 5 con concentraciones de 0.5 y 2 mg/mL del polipéptido de DAT. La DAT purificada de *B. sphaericus* fue utilizada como control. Después de 2, 4.5, 9, 24, 36 y 48 horas, se tomó una alícuota, se agregó ácido fórmico hasta una concentración final de 2% y las muestras fueron congeladas. Las muestras fueron luego descongeladas, agitadas y filtradas. Las muestras fueron analizadas en cuanto a contenido de monatina utilizando la metodología de LC/MS/MS descrita en el Ejemplo 3. Las muestras fueron analizadas cuantitativamente en cuanto a los niveles de HMG. Se tomaron alícuota adicionales (sin ajuste del pH) para el análisis estereoisómero utilizando la metodología de derivación con FDAA descrita en el Ejemplo 3. Los resultados se muestran en las Tablas 22 y 23.

Tabla 22

Polipéptido de DAT (SEQ ID NO:)	Monatina (ppm) 2 horas	Monatina (ppm) 4.5 horas	Monatina (ppm) 9 horas	Monatina (ppm) 24 horas	Monatina (ppm) 36 horas	Monatina (ppm) 48 horas
882 (0.5 mg/mL)	61	274	780	1802	2172	2170
882 (2 mg/mL)	985	2452	3232	3128	3082	3158
884 (0.5 mg/mL)	149	362	656	1394	1756	2158
884 (2 mg/mL)	811	1628	2466	2988	3178	2864
<i>B. sphaericus</i> (0.5 mg/mL)	362	860	1268	2362	2532	2804
<i>B. sphaericus</i> (2 mg/mL)	1335	2344	3154	3866	3842	4008

Tabla 23. Estereopurezas de monatina producida en puntos de tiempos seleccionados

Polipéptido de DAT (SEQ ID NO:)	24 horas (% R,R)	48 horas (% R,R)
882 (2 mg/mL)	95.4	94.1
884 (2 mg/mL)	99.6	99.4
Bs DAT (2 mg/mL)	95.8	93.3

Los polipéptidos que tienen la secuencia mostrada en SEQ ID NO: 882 y 884 exhibieron buena actividad de formación de monatina y esteroespecificidad, y parecían producir menos HMG que el control de *B. sphaericus* durante los puntos de tiempo iniciales. Los polipéptidos que tienen la secuencia de SEQ ID NO: 882 exhibieron tasas de formación de monatina iniciales similares pero parecieron haberse estabilizado en este experimento en un título de monatina inferior.

Análisis de ejemplo de DATs en pSE420-cHis, y de DAT en pET30a

Los marcos de lectura abierta que codifican los polipéptidos DAT que tienen la secuencia de SEQ ID NO: 898, 900, 902, 904, 906, 910, y 896 fueron evaluados. Una persona de experiencia normal en la técnica puede sintetizar los genes que codifican estas D-aminotransferasas utilizando las técnicas de PCR en ensamblaje tales como las descritas en el Ejemplo 4.

- 5 Un cultivo de *E. coli* BL21 DE3 que contiene un ácido nucleico que codifica un polipéptido DAT que tiene la secuencia mostrada en SEQ ID NO: 896 en el vector pET30a (subclonado como se describe en el Ejemplo 4) fue cultivado en 50 mL de Overnight Express (Novagen) en un matraz con deflexión de 250 mL durante la noche a 30°C y 250 rpm. Las células fueron recolectadas por centrifugación cuando la densidad óptica a 600 nm fue superior a 10.

- 10 Células de *E. coli* Top10 (Invitrogen, Carlsbad, CA) fueron transformadas con el plásmido pSE420-cHis que contiene los ácidos nucleicos de DAT que tienen la secuencia mostrada en SEQ ID NO: 897, 899, 901, 903, 905, y 909 y se sembraron en medio LB que contenía ampicilina (100 µg/mL). Se utilizaron 500 µL de un cultivo durante la noche para inocular 50 mL del mismo medio en un matraz con deflexión de 250. Las células fueron cultivadas a 30°C hasta una OD_{600nm} de aproximadamente 0.4 y se agregó IPTG hasta una concentración final de 1 mM. Las células fueron cultivadas a 30°C durante 4 horas y recolectadas por centrifugación.

- 15 Se prepararon extractos celulares como se describe en el Ejemplo 4. Las concentraciones de proteína soluble y DAT estimada fueron determinadas como se describe en el Ejemplo 4.

- 20 Se llevó a cabo un ensayo de formación de R,R monatina tal como se describe en el Ejemplo 5 con concentraciones de polipéptidos de DAT de 0.5 mg/mL, excepto que se utilizaron 0.3 mg/mL del polipéptido que tiene la secuencia de SEQ ID NO: 896; se usaron 0.06 mg/mL del polipéptido que tiene la secuencia de SEQ ID NO: 898; se usaron 0.4 mg/mL del polipéptido que tiene la secuencia mostrada en SEQ ID NO: 900; se usaron 0.1 mg/mL del polipéptido que tiene la secuencia de SEQ ID NO: 902; y se usaron 0.12 mg/mL del polipéptido que tiene la secuencia de SEQ ID NO: 904. Como controles positivos, se probaron SEQ ID NO: 870, 870 T242N, y *B. sphaericus* purificado a concentraciones de polipéptido de DAT de 0.5 mg/mL. Después de 2, 6, y 24 horas, se tomó una alícuota, se agregó ácido fórmico hasta una concentración final de 2% y las muestras fueron congeladas. Las muestras fueron luego descongeladas, agitadas y filtradas. Las muestras fueron analizadas en cuanto al contenido de monatina utilizando la metodología de LC/MS/MS descrita en el Ejemplo 3. Se tomaron alícuotas adicionales para el análisis de distribución estereoisomérica y no fueron tratadas con ácido fórmico. Los resultados para el punto de tiempo de 24 horas se muestran en la Tabla 24. El ácido nucleico de DAT que codifica el polipéptido de DAT que tiene la secuencia mostrada en SEQ ID NO: 908 no fue subclonado y no pudo ser probado.

- 30 Tabla 24

Polipéptido (SEQ ID NO:)	Monatina 2 horas (ppm)	Monatina 6 horas (ppm)	Monatina 24 horas (ppm)	24 horas % R,R
<i>B. sphaericus</i>	261	1203	2604	95.6
870	193	1067	2490	96.8
870 T242N	813	2230	3380	98.8
896	30	127	286	99.8
898	nd	3	15	95.7
900	4	16	56	92.9
902	144	411	1209	96.7
904	nd	1	4	92.3
906	14	18	25	98
910	487	1154	2770	94.5
nd = no detectable bajo las condiciones probadas				

Los polipéptidos de DAT que tiene la secuencia mostrada en SEQ ID NO: 910 y 902 tiene niveles altos de actividad para las reacciones de formación de monatina, y produjeron niveles ligeramente altos de R,R monatina. Los resultados indicaron que el polipéptido de DAT que tiene la secuencia mostrada en SEQ ID NO: 870 exhibe actividad comparable a la del polipéptido tipo silvestre que tiene la secuencia mostrada en SEQ ID NO: 910 bajo las condiciones probadas; sin embargo, la mutación T242N en el polipéptido de SEQ ID NO: 870 produce un mejor agrande en actividad y esteroespecificidad de la enzima.

Ejemplo 13-Análisis de los DATs

Se obtuvieron los plásmidos de DATs (pSE420-cHis) que contienen las secuencias de ácidos nucleicos que codifican SEQ ID NO: 912, 914, 916, 918, 920, 922, 924, y 926. Una persona experimentada en la técnica podría clonar los genes utilizando cualquier número de protocolos de ensamblaje de genes tales como los descritos en el Ejemplo 4.

Células de E. coli Top 10 (Invitrogen) fueron transformadas con los plásmidos pSE420-cHis que contenían los polipéptidos DAT que tienen la secuencia de SEQ ID NO: 912, 914, 916, 918, 920, 922, 924, y 926 y se sembraron sobre medio LB que contenía ampicilina (100 µg/mL). Se utilizaron 500 µl del cultivo durante la noche para inocular 50 mL del mismo medio en matraces con deflexión de 250. Los cultivos fueron cultivados a 30°C hasta una OD_{600nm} de aproximadamente 0.4. Se agregó IPTG hasta una concentración final de 1 mM. Las células fueron cultivadas a 30°C durante 4 horas y recolectadas por centrifugación.

Los extractos celulares fueron preparados como se describe en el Ejemplo 4. Las concentraciones de proteína soluble total y proteína de DAT fueron determinadas como se describe en el Ejemplo 4. La mayor parte de los polipéptidos de DAT que fueron expresados eran solubles excepto para el polipéptido SEQ ID NO: 916, el cual era solamente parcialmente soluble.

Se llevó a cabo un ensayo de formación de R,R monatina como se describió en el Ejemplo 5, con una concentración de polipéptido de DAT direccionada aproximadamente 0.25 mg/mL; excepto que se usaron 0.1 mg/mL del polipéptido de SEQ ID NO: 922 y 0.2 mg/mL del polipéptido de la SEQ ID NO: 926. Como control positivo, se probó el DAT de B. sphaericus purificado a una concentración de 0.25 mg/mL. Después de 2, 8 y 24 horas, se tomó una alícuota y se agregó ácido fórmico hasta una concentración final de 2%, y las muestras fueron congeladas. Las muestras fueron luego descongeladas, agitadas y filtradas. Las muestras fueron analizadas en cuanto al contenido de monatina utilizando la metodología LC/MS/MS. Descrita en el Ejemplo 3. Los resultados se muestran en la Tabla 25.

Tabla 25

Polipéptido (SEQ ID NO:)	Monatina 2 horas (ppm)	Monatina 8 horas (ppm)	Monatina 24 horas (ppm)
B. sphaericus	179	774	1482
912	0.2	2	2
914	4	22	62
916	0.6	2	6
918	158.8	402	496
920	5	40	96
922	0.2	2	2
924	nd	nd	nd
926	1.2	12	30
nd = no detectable bajo las condiciones probadas			

Los ácidos nucleicos de DAT que tienen la secuencia mostrada en SEQ ID NO: 169, 171, 167, 173 y 175 (que codifican los polipéptidos de DAT que tiene la secuencia mostradas en SEQ ID NO: 170, 172, 168, 174 y 176) fueron obtenidos como productos de PCR y fueron subclonados en pET30a como se describe en el Ejemplo 4. Una

persona de experiencia en la técnica podría reconstruir los genes utilizando cualquier número de métodos de ensamblaje de genes tales como el descrito en el Ejemplo 4.

5 Se cultivaron células de *E. coli* BL21 DE3 que contenían los ácidos nucleicos de DAT que tienen las secuencias de SEQ ID NO: 169, 171, 167, 173 y 175 en el vector pET30a en 50 mL de Overnight Express (Novagen) en un matraz con deflexión de 250 mL, durante la noche a 30°C y 250 rpm. Las células fueron recolectadas por centrifugación cuando la densidad óptica a 600 nm fue mayor de 10.

10 Los extractos celulares fueron preparados como se describe en el Ejemplo 4. Las concentraciones de proteína soluble total y proteína de DAT fueron determinadas como se describe en el Ejemplo 4. El polipéptido que tiene la secuencia de SEQ ID NO: 170 (codificado por el ácido nucleico de DAT que tiene la secuencia de SEQ ID NO: 169) no parecía ser soluble, lo cual puede haber impedido los ensayos de actividad.

15 Se llevó a cabo un ensayo de formación de R,R monatina como se describe en el Ejemplo 5 utilizando una concentración de polipéptido de DAT de 0.5 mg/mL, excepto que se usaron 0.25 mg/mL del polipéptido de SEQ ID NO: 170. Como control positivo, se probó DAT de *B. sphaericus* a una concentración de 0.5 mg/mL. Después de 2, 8 y 24 horas, se tomó una alícuota, se agregó ácido fórmico hasta una concentración final de 2%, y las muestras fueron congeladas. Luego las muestras fueron descongeladas, agitadas y filtradas, y analizadas en cuanto al contenido de monatina utilizando la metodología de LC/MS/MS descrita en el Ejemplo 3. Los resultados se muestran en la Tabla 26.

Tabla 26

Polipéptido (SEQ ID NO:)	Monatina 2 horas (ppm)	Monatina 8 horas (ppm)	Monatina 24 horas (ppm)
<i>B. sphaericus</i>	456	1502	2970
170	2	8	14
172	5	20	60
168	15	68	186
174	1	4	8
176	451	1508	2744

20 Se analizaron muestras (sin ajuste de pH) para determinar el porcentaje de R,R utilizando el protocolo de derivación con FDAA descrito en el Ejemplo 3. La monatina producida por el polipéptido DAT que tiene la secuencia mostrada en SEQ ID NO: 176 fue de 99.6% de R,R después de 24 horas en comparación con la producida por *B. sphaericus* que fue de 95.2% de R,R en el mismo punto de tiempo. La actividad y estereopureza resultantes del polipéptido de DAT que tiene la secuencia de SEQ ID NO: 176 fueron ambas bastante altas, y el ácido nucleico correspondiente
25 fue subclonado como una proteína etiquetada con terminal C tal como se describe en el Ejemplo 4 para más estudios cuantitativos.

Caracterización de la proteína de SEQ ID NO: 176 etiquetada en C-His

30 El ácido nucleico que tiene la SEQ ID NO: 175, la cual codifica el polipéptido que tiene la secuencia de SEQ ID NO: 176, fue clonado en pET30a sin un codón de detención de tal manera que podría ser expresado como una proteína de fusión con una etiqueta 6xHis en el terminal C. La proteína fue purificada utilizando la resina de enlazamiento a His descrita en el Ejemplo 4. Cuando la proteína de fusión fue eluida de la columna de desalinización PD-10, se formó un precipitado amarillo en la solución. Un residuo amarillo también fue observado en la columna. El color
35 amarillo usualmente es indicativo de la presencia de una proteína que se enlaza a PLP. En un esfuerzo para prevenir la precipitación de la proteína que se enlaza a PLP en la etapa de desalinización, se utilizaron diferentes reguladores (fosfato 100 mM y EPPS 100 mM con y sin 10% de glicerol) a dos valores de pH (7.8 y 8.2), ninguno de los reguladores tratados pareció evitar completamente la precipitación.

El ensayo de monatina fue llevado a cabo utilizando una solución de proteína heterogénea bien mezclada y una concentración de polipéptido de DAT de 0.5 mg/mL. Los resultados se muestran en la Tabla 27. El polipéptido de DAT purificados de SEQ ID NO: 176 (etiquetado en C) mostró actividad comparable al polipéptido DAT de control

positivo de *B. sphaericus*; sin embargo, la actividad pareció ser inferior a la actividad exhibida por los polipéptidos mutantes que tienen la secuencia de SEQ ID NO: 870 T242N o SEQ ID NO: 870 T242N.

Tabla 27. Producción de monatina (ppm)

Polipéptido (SEQ ID NO:)	2 horas	4 horas	8 horas	24 horas
<i>B.sphaericus</i>	262	676	1044	2150
870	332	678	1346	2826
870 T242N	942	1938	2834	4004
176	208	392	732	1806

5 Ejemplo 14 – Evaluación de DATs

Los marcos de lectura abierta que codifican 29 DATs fueron obtenidos como productos de PCR. Hay que anotar que una persona de experiencia normal en la técnica puede sintetizar los genes que codifican los DATs utilizando técnicas de ensamblaje tales como las descritas en el Ejemplo 4. Los ácidos nucleicos de DAT fueron subclonados en el vector pET30a y expresados como proteínas no etiquetadas como se describe en el Ejemplo 4. Los extractos libres de células desalinizados (preparados como se describe en el Ejemplo 4) fueron utilizados en ensayos de formación de monatina. Una concentración de polipéptido de DAT de 0.5 mg/mL fue utilizada para el ensayo de monatina excepto para los siguientes polipéptidos (cantidades usadas en paréntesis): el polipéptido SEQ ID NO: 156 (0.4 mg/mL), el polipéptido SEQ ID NO: 182 (0.45 mg/mL), el polipéptido SEQ ID NO: 240 (0.47 mg/mL), y el polipéptido SEQ ID NO: 204 (0.42 mg/mL).

La mayor parte de los polipéptidos de DAT mostraron una producción de monatina no detectable hasta baja bajo las condiciones ensayadas en comparación con enzimas de control positivo. La mayor parte de los polipéptidos de DAT fueron solubles según se determinó mediante Experion; sin embargo, los polipéptidos que tenían SEQ ID NO: 204 y 240 fueron expresados a niveles muy bajos y pueden no haber sido muy solubles. Por otro lado, se predijo que el polipéptido que tiene SEQ ID NO: 220 iba a ser el 68% de la proteína total según se juzgó mediante el software Experion.

Los resultados de producción de monatina se muestran en la Tabla 28 y 29. A 24 horas, el polipéptido de DAT que tiene SEQ ID NO: 156 y 214 produjo 40-50% de monatina en comparación con la enzima de control positivo, el DAT de *B. sphaericus*. El polipéptido de DAT más activo fue el polipéptido de SEQ ID NO 220. Aproximadamente 4 horas después de que la reacción se inició, la concentración de monatina alcanzó un máximo. Se espera que la proteína madura de SEQ ID NO: 156 (sin la secuencia de guía predicha) sea el componente activo del polipéptido de DAT, y por lo tanto, la proteína pueda ser producida de manera recombinante con la secuencia de guía ausente.

Tabla 28. Monatina formada (ppm)

Polipéptido (SEQ ID NO:)	2 horas	4 horas	8 horas	24horas
178	11	24	62	194
180	nd	nd	nd	nd
154	52	103	166	178
182	nd	nd	nd	nd
218	1.6	2.8	274	12
188	2.4	5.2	10	22
190	3.8	9.4	22	42
208	1	1.2	nd	nd
220	2418	3563	3812	3882

Polipéptido (SEQ ID NO:)	2 horas	4 horas	8 horas	24horas
196	1	1.8	nd	8
156	64.8	156	296	796
B. sphaericus	422	791	1302	2124
nd = no detectable bajo las condiciones probadas				

Tabla 29. Monatina formada (ppm)

Polipéptido (SEQ ID NO:)	2 horas	4 horas	8 horas	24horas
166	nd	nd	1	nd
216	69	91	109	134
200	nd	nd	1	1
198	nd	nd	nd	nd
210	1.6	3.8	6.2	15.2
202	3.4	6.2	12.2	29.8
222	3.6	7	12.8	25.8
236	nd	nd	nd	nd
204	nd	nd	nd	3.6
238	10.4	21.8	46.4	115.6
240	3	6	12.4	32.2
224	39.8	85	171.8	268.8
226	2.6	5.8	12.4	30.6
228	4.2	9.8	21.4	66.8
230	9.2	21.8	42.2	94.4
232	3.6	9.4	21.6	57
246	nd	nd	nd	nd
214	160	327	694	1346
B. sphaericus	393	986	1624	2597
nd = no detectable bajo las condiciones probadas				

5 La alta actividad del polipéptido de SEQ ID NO: 220 fue confirmada en otro ensayo de formación de monatina donde el polipéptido de SEQ ID NO: 220 fue comparado con los polipéptidos DAT de SEQ ID NO: 870, 870 T242N y B. sphaericus. Se utilizó una concentración de polipéptido de DAT de 0.1 mg/mL y 0.5 mg para cada uno de los polipéptidos de DAT ensayados. Los resultados se muestran en la Tabla 30. Debido al alto grado de actividad de los polipéptidos de DAT probados, las muestras de monatina tuvieron que ser diluidas 100 veces para que estuvieran dentro del rango cuantitativo de los instrumentos usados (en oposición a una dilución típica de 10 o 20 veces).

Tabla 30. Monatina formada (ppm)

Polipéptido de DAT (SEQ ID NO:)	2 horas	4 horas	8 horas	24horas
B. sphaericus (0.1 mg/mL)	74	170	309	728
B. sphaericus (0.5 mg/mL)	510	921	1068	2704
870 (0.1 mg/mL)	28	81	179	706
870 (0.5 mg/mL)	399	847	1466	2916
870 T242N (0.1 mg/mL)	93.2	245.8	582.4	1270
870 T242N (0.5 mg/mL)	1158.8	2026	3202	4126
220 (0.1 mg/mL)	965.8	1512	2330	3788
220 (0.5 mg/mL)	2950	4302	4618	4488

El porcentaje de R,R formado por el polipéptido de DAT que tiene SEQ ID NO: 220 en los experimentos anteriores se determinó utilizando la metodología de derivación con FDAA descrito en el Ejemplo 3. El polipéptido de DAT que tiene la secuencia de SEQ ID NO: 220 es altamente estereoespecífico, produciendo 99.3% de R,R monatina las 24 horas en comparación con el 92.9% de R,R para B. sphaericus. En otro ensayo, el polipéptido de SEQ ID NO: 220 produjo 99.8% de R,R monatina.

El polipéptido de DAT de ejemplo de esta invención que tiene la secuencia de SEQ ID NO: 220 es una proteína novedosa que es 62% idéntica al nivel de proteína del polipéptido DAT de C. beijerinckii mostrado en el Ejemplo 9. El polipéptido de SEQ ID NO: 220 tiene 86%-97% de homología de secuencia primaria (identidad de secuencia) con otros polipéptidos DAT altamente activos (por ejemplo, aquellos que tiene la secuencia mostrada en SEQ ID NO: 892 y 894 (Ejemplo 8), 946 (Ejemplo 7) y 176 (Ejemplo 13)). Estos polipéptidos de DAT altamente activos y novedosos no estaban caracterizados antes de este trabajo, y estas enzimas exhibieron una actividad y estereoespecificidad más altas para las reacciones de producción de R,R monatina que cualquiera de las D-aminotransferasas similares a Bacillus publicadas. La figura 7 muestra un alineamiento de estas D-aminotransferasas relacionadas y los motivos de secuencia de consenso que ellas tienen en común y se describen más abajo.

Secuencia Consenso C

M.*GYYNG.*P.*DR.*FGDG.*YDAT.*N.*FAL.*H.*RF.*NS.*LL.*L.*K.*YWQ.*RG.*G.*R.*H.*F.*N.*P.*KLI.*DTRF.
 *HCNIKTLNL.*P.*VIA.*Q.*E.*C.*E.*VFHRG.*VTECAHSN.*I.*NLIL.*G.*HL.*P.*E.*F.*L.*ADE.*V.*SS.*DG.*GGK.
 *K.*Q.*T (SEQ ID NO: 1071)

Secuencia Consenso D

M.{3}GYYNG.{10}P.{2}DR.{3}FGDG.YDAT.{3}N.{3}FAL.{2}H.{2}RF.NS.{2}LL.I.{9}K.{17}YWQ.{2}RG.G.R.H.F.{5,7}N.{2}I.{3}P.{10}KLI.{3}DTRF.HCNIKTLNL. P.VIA.Q.{3}E.{2}C.E.VFHRG.{2}VTECAHSN.{2}I.{11}DNLIL.G.{4}HL.
 {9}P.{2}E. {2}F.{4}L.{2}ADE.{2}V.SS.{10}DG.{3}GGK.{5}K.{2}Q.{10}T(SEQ ID NO:1072)

Secuencia Consenso E

M.*[LV]GYYNG.*[LI].*[ML].*[VI]P.*DR.*[YF]FGDG.*YDAT.*N.*FAL[DE][ED]H[IL][DE]RF.*NS.*LL.*I.*[KR].*[EQ][L
 MV]K[KE].*[MV].*[DE].*[VL]YWQ.*[TS]RG[TS]G.*R[NS]H.*F.*N[LI].*L*P.*[IVL].*[KE].*KLI[TS].*[ED]DTRF.*HCNIK
 TLNL[I L]P[NS]VIA[SA]Q[RK].*E.*C.*E.*VFHRG[ED].*VTECAHSN[VI].*[I][IL][KR][ND].*[TS].*DNLIL.*G.
 HL[LI][QK].[IV]P.*E.*F[TS][LM].*[ED]L.*ADE[VI][LI]V[ST]SS. *[LMI].*[IL]DG.*GGK.*[LVI]K.*[IL]Q.*[EK][FY].*T
 (SEQ ID NO: 1073)

Similar al lenguaje de convención de expresión regular PERL (perloc.perl.org), “.” indica que cualquier número de residuos de aminoácidos pueden estar presente de cualquiera de los 20 aminoácidos proteinogénicos; [] indica que cualquiera de los aminoácidos en los paréntesis puede estar presente; “.{#}” indica que cualquiera de los 20 aminoácidos proteinogénicos puede estar presente en tanto el número de residuos coincida con el número (#) indicado en los paréntesis.

Con respecto al uso de “.” en la Secuencia Consenso C, el número de aminoácidos en cualquiera de las posiciones “.” puede variar, por ejemplo, desde aproximadamente 0 hasta aproximadamente 20 residuos (véanse, por ejemplo, Secuencias Consenso D (SEQ ID NO: 1072) y E (SEQ ID NO: 1073)) o desde aproximadamente 20 residuos hasta aproximadamente 100 residuos, o el número de aminoácidos puede ser mucho mayor, por ejemplo, hasta 1000 o más residuos. Sin limitación, una inserción en una o más de las posiciones “.” puede corresponder, por ejemplo, a un dominio tal como (pero no limitándose a) un dominio de enlazamiento de quitinasa (por ejemplo, de *Pyrococcus furiosus* (Accesión o. 2CZN_A) o *P. burkholderia* (Accession No. YP_331531) o un dominio de enlazamiento de celulosa (por ejemplo, de *Cellulomonas fimi* (Accession No. 1 EXH_A) o *Clostridium stercorarium* (Accession No. 1UYI_A). En algunas realizaciones (sin limitación), cinco o menos de las posiciones designadas como “.” contienen cada una inserción de, por ejemplo, más de aproximadamente 20 residuos (por ejemplo, más de aproximadamente 100 residuos). En otras realizaciones (sin limitación), cinco o más de las posiciones designadas como “.” contienen cada una inserción de al menos de aproximadamente 100 residuos (por ejemplo, menos de aproximadamente 20 residuos, por ejemplo, 3, 5, 10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90, o 95 residuos). La actividad de un polipéptido que tiene una secuencia que corresponde a uno o más de las Secuencias Consenso divulgadas aquí y que contiene cualquier número de residuos insertados en una o más de las posiciones “.” puede ser evaluada utilizando métodos que se describen aquí.

Polipéptidos representativos no limitantes que contienen la secuencia consenso mostrada en SEQ ID NO: 1071 incluyen SEQ ID NO: 220, 892, 894, 176 y 946 y secuencias consenso D (SEQ ID NO: 1072 y E (SEQ ID NO: 1073)). Se espera que cualquier D-aminotransferasa que exhibe cualquier secuencia de consenso C, D, o E sea activa en las etapas de la ruta de formación de monatina.

Caracterización de un polipéptido etiquetado con c-His que tiene la secuencia de SEQ ID NO: 220

Los ácidos nucleicos de DAT que tienen la secuencia de SEQ ID NO: 219 (que codifica el polipéptido de SEQ ID NO: 220) fueron clonados en un pET30a sin un codón de detección de tal manera que se expresó como una proteína de fusión con una etiqueta 6xHis sobre el terminal C (como se describe en el Ejemplo 4). La proteína de fusión fue purificada utilizando la columna His-Bind (Novagen) descrito en el Ejemplo 4. El eluido de la columna de desalinización PD-10 formó un precipitado amarillo. El residuo amarillo también fue observado sobre la columna. Los ensayos de monatina se hicieron utilizando una solución de proteína heterogénea bien mezclada. Las cantidades de polipéptido DAT usadas están indicadas en paréntesis en la columna de la izquierda. La anotación w/Trp indica que la enzima fue incubada con D-triptófano 10 mM durante la noche sobre hielo. Los resultados se muestran en la Tabla 31.

Tabla 31. Monatina formada (ppm)

Polipéptido (SEQ ID NO:)	2 horas	4 horas	8 horas	24 horas
B. sphaericus (0.5 mg/mL)	382	660	1021	1986
870 T242N (0.1 mg/mL)	69	205	412	1074
870 T242N w/Trp (0.1 mg/mL)	63	163	383	978
870 T242N (0.5 mg/mL)	919	1698	2356	3130
870 T242N w/ Trp (0.5 mg/mL)	772	1519	2294	3023
220 (0.1 mg/mL)	847	1462	2202	3004
220 (0.1 mg/mL)	811	1522	2202	2887
220 w/ Trp (0.1 mg/mL)	537	1080	1590	2401
220 (0.5 mg/mL)	2885	3446	3813	4066
220 w/ Trp (0.5 mg/mL)	1933	3223	3939	3911

La reacción que contiene 0.1 mg/mL del polipéptido de SEQ ID NO: 220 mostró un transcurso en el tiempo de formación de monatina similar a la reacción que contenía 0.5 mg/mL del polipéptido de SEQ ID NO: 870 T242N. La adición de D-triptófano (10 mM) a la solución que contenía la proteína purificada eliminó la precipitación. La pérdida de actividad fue observada para la muestra en la cual SEQ ID NO: 220 fue incubada con D-triptófano (10 mM) durante la noche sobre hielo, pero no se observó efecto negativo cuando el polipéptido de SEQ ID NO: 870 T242N fue tratado con D-triptófano (10 mM). La presencia de HMG fue analizada también cualitativamente para reacciones catalizadas por el polipéptido de SEQ ID NO: 220 comparando las áreas de pico. Cuando ambos polipéptidos de DAT fueron utilizados a una concentración de 0.5 mg/mL, la reacción catalizada por el polipéptido de SEQ ID NO: 220 formó alrededor de 40% del HMG en comparación con la reacción que contenía el polipéptido de SEQ ID NO: 870 T242N. Puntos en el tiempo más temprano mostraron una diferencia aún más pronunciada entre las dos enzimas.

En un intento para prevenir la precipitación de proteína durante la purificación del polipéptido de SEQ ID NO: 220, se incluyó DTT (5 mM) en todos los reguladores incluyendo el reactivo BugBuster, los reguladores para la columna His-Bind y el regulador para la columna PD-10. No se observó precipitación después de la columna de desalinización, y no se observó efecto negativo sobre la actividad del polipéptido de SEQ ID NO: 220 cuando se incluyó DTT durante la purificación o se agregó en la solución de proteína purificada a una concentración de 2 mM. Los datos para los ensayos de formación de monatina se muestran en la Tabla 32. La cantidad de polipéptido de DAT usada está indicada en la columna de la izquierda en paréntesis. "DTT agregado" indica que se agregó DTT 2 mM para resolubilizar la proteína después de la purificación; y "w/DTT" purificado indica que había presente DTT 5 mM a lo largo de la purificación.

Tabla 32. Monatina formada (ppm)

Polipéptido (SEQ ID NO:)	2 horas	4 horas	24 horas
B. sphaericus (0.5 mg/mL)	426	965	2638
870 T242N (0.5 mg/mL)	977	1916	4227
220 (0.1 mg/mL)	1214	2163	3964
220 (0.5 mg/mL)	3534	4246	4415
220 DTT agregado (0.1 mg/mL)	1287	2202	3566
220 DTT agregado (0.4 mg/mL)	3495	4833	5082
220 w/DTT purificado (0.1 mg/mL)	1204	2169	3997
220 w/DTT purificado (0.5 mg/mL)	3562	4110	4353

Las prioridades altamente deseables del polipéptido de SEQ ID NO: 220 lo hacen un excelente candidato para mutagénesis o experimentos de evolución dirigida posteriores.

Mutagénesis dirigida al sitio del polipéptido de SEQ ID NO: 220

Una región bucle del polipéptido de DAT relacionado con el polipéptido DAT de *Bacillus* fue identificada como importante en cuanto a la especificidad por el sustrato y estereoespecificidad de la enzima (Ro et al., 1996, FEBS Lett, 398:141-145; Sugio et al., 1995, Biochemistry 34:9661-9669; and EP 1 580 268). Un residuo clave en esta región es una T en el residuo 242 (en el polipéptido de DAT de ATCC #4978, esta posición corresponde a una T en el residuo 243). Un mutante T242N del polipéptido de SEQ ID NO: 870 mostró un mejoramiento tanto en actividad como en estereoespecificidad, tal como lo hizo el mutante T243N de DAT 4978 (véase Ejemplo 10). El alineamiento de secuencia primario del polipéptido de SEQ ID NO: 220 con el polipéptido de SEQ ID NO: 870 mostró solamente un 35% de identidad en secuencia de aminoácidos y 65% de homología. El residuo T242 en SEQ ID NO: 870 fue alineado con un residuo G240 en SEQ ID NO: 220, el cual es seguido por un residuo T241. Utilizando el software Accelrys DS Modeler para ambas proteínas (con estructura YM-1 de *Bacillus* como plantilla) fue difícil superponer la región bucle del polipéptido de SEQ ID NO: 870 con el polipéptido de SEQ ID NO: 220. Por lo tanto, ambos aminoácidos fueron escogidos como objetivos para mutagénesis dirigida al sitio.

Un polipéptido mutante designado como SEQ ID NO: 220 G240N y SEQ ID NO: 220 T241N fue generado por mutagénesis dirigida al sitio del ácido nucleico correspondiente (SEQ ID NO: 219) como se describe en el Ejemplo 4. Los dos polipéptidos mutantes fueron expresados y purificados como proteínas de fusión etiquetadas con 6xHis en

el ensayo de formación de monatina. La precipitación amarilla fue observada en la etapa de desalinización para ambos polipéptidos mutantes de SEQ ID NO: 220. Los resultados se muestran en las Tablas 33 y 34 para ensayos de formación de monatina. La cantidad de D-aminotransferasa usada está indicada en la columna de la izquierda en paréntesis. Se utilizaron diferentes preparaciones del polipéptido de SEQ ID NO: 220 en el ensayo. "ferm" indica que el polipéptido de SEQ ID NO: 220 utilizado fue producido en un fermentador como se describe en el Ejemplo 15.

5

Tabla 33. Monatina producida (ppm)

Polipéptido (SEQ ID NO:)	2 horas	4 horas	8 horas	24 horas
B. sphaericus (0.5 mg/mL)	440	777	1510	2621
870 T242N (0.5 mg/mL)	961	1913	2793	3904
ferm 220 (0.1 mg/mL)	1396	2379	3217	3770
ferm 220 (0.2 mg/mL)	2301	3277	3789	4328
ferm 220 w/ DTT (0.1 mg/mL)	1434	2384	3109	3730
ferm 220 w/ DTT (0.2 mg/mL)	2423	3568	3859	4755
220 (0.1 mg/mL)	1109	1912	2809	3713
220 T241N (0.1 mg/mL)	554	856	1084	1986

Tabla 34. Monatina formada (ppm)

Polipéptido (SEQ ID NO:)	2 horas	4 horas	8 horas	24 horas
B. sphaericus (0.5 mg/mL)	634	938	1651	2754
870 T242N (0.5 mg/mL)	1422	1922	3211	3793
ferm 220 (0.1 mg/mL)	1976	2505	3442	4211
ferm 220 (0.2 mg/mL)	3198	3430	4452	4639
220 G240N (0.1 mg/mL)	3	5	14	42
220 G240N (0.2 mg/mL)	9	17	46	94

- 10 Una cantidad muy pequeña de monatina (95.7% de R,R monatina) fue formada en la reacción catalizada por el polipéptido mutante SEQ ID NO: 220 G240N. El polipéptido mutante SEQ ID NO: 220 T241N perdió aproximadamente 50% de la actividad, pero todavía mantuvo la estereoespecificidad (99.7% de R,R monatina producida). Estos resultados, junto con la homología (identidad de secuencia) modelación y alineamientos, sugieren que, en la región que circundan los residuos 242-243 (y potencialmente más allá), la estructura del polipéptido de
- 15 SEQ ID NO: 220 no es similar a la estructura del polipéptido SEQ ID NO: 870 ó la estructura del polipéptido de DAT similar a Bacillus en la literatura. Puesto que no hay estructura cristalina por rayos X, la mutagénesis aleatoria, metodologías combinacionales y otras metodologías de evolución dirigida del polipéptido de SEQ ID NO: 220 y polipéptidos de DAT relacionados se esperan que sea altamente productivos en mejoramiento adicional de la actividad de la enzima.

20 **Ejemplo 15 – Producción de un DAT en un fermentador**

Los componentes del medio de crecimiento bacteriano fueron de Dfco, Fisher Scientific, o VWR; otros reactivos fueron de grado analítico o del grado más alto disponible comercialmente. La fermentación fue llevada a cabo en un fermentador New Brunswick Scientific (Edison, NJ) BioFlo 3000®. La centrifugación fue llevada a cabo utilizando una centrífuga Beckman (Fullerton, CA) Avanti® J-251 con un rotor JLA-16.250 o JA-25.50.

El ácido nucleico de DAT que codifica el polipéptido que tiene la secuencia en SEQ ID NO: 220 con una etiqueta His C terminal fue clonado utilizando los sitios de restricción Nde I / Xho I en el vector pMet1a descrito en el Ejemplo 16. El marcador antibiótico (gen bla) puede ser removido posteriormente utilizando digestión con enzima de restricción Psi I, purificación en gel de la banda del vector, autoligazón de los extremos del vector, transformación en el anfitrión E. coli y selección sobre placas de medio mínimo que no contienen metionina. Típicamente, se usa el medio Neidhardt con 15 aminoácidos. Los sitios de clonación fueron NdeI/XhoI para insertar la secuencia de ácidos nucleicos de SEQ ID NO: 220 en pMET1a (véase ejemplo 16).

El polipéptido de DAT de SEQ ID NO: 220 que porta una etiqueta de purificación His en C terminal fue producido en un fermentador a escala de 2.5 L, en un proceso de alimentación-lote que alcanza altas densidades celulares y altos niveles de expresión de la proteína deseada. El protocolo y resultados para el crecimiento de la cepa de E. coli B834 (DE3:: SEQ ID N: 220 cHis pMET1 son descritos como sigue: Partiendo de un placa de cultivo fresca (Neidhardt + 15 aminoácidos, sin metionina), las células fueron cultivadas en 5 mL de medio Neidhardt suplementado con 15 aminoácidos, a 30°C y 225 rpm durante 6-8 horas. Se transfirió 1 mL del cultivo a cada uno de 2 alícuotas de 125 mL del medio de producción suplementado con 5 g/L de glucosa. Los matraces fueron cultivados a 30°C y 225 rpm durante la noche (16-18 horas). Un fermentador fue cargado con 2.5 litros del medio de producción que contenían (por litro) 2.0 g/L (NH₄)₂SO₄; 8.0 g/L K₂HPO₄; 2.0 g/L NaCl; 1.0 g/L Na₃Citrato-2H₂O; 1.0 g/L M₂SO₄· 7H₂O; 0.025 g/L CaCl₂·2H₂O; 0.05 g/L FeSO₄·7H₂O; 0.4 mL/L de micronutrientes de Neidhardt, y 2.0 g/L de glucosa. El fermentador fue inoculado con 10% v/v del cultivo nocturno. Tres horas después de la inoculación, se estableció una alimentación con glucosa exponencial utilizando una solución de glucosa al 60% p/v. La alimentación fue suministrada a la rata requerida para soportar el crecimiento microbiano a una rata exponencial de 0.15 h⁻¹. Cuando la rata de evolución de dióxido de carbono (CER) había alcanzado un valor de 100 mmoles/L/h (aproximadamente 21 horas después de la inoculación; correspondiente a una biomasa celular de 15-16 g de DCW/L), la expresión genética fue inducida con una adición de bolus de 2 g/L de lactosa (suministrada como una solución al 20%). La alimentación fue cambiada de 60% p/v de glucosa a 50% p/v de glucosa + 10% p/v de lactosa mientras que la rata de alimentación fue fijada en la rata en el momento de la inducción. La alimentación de "50% p/v de glucosa + 10% p/v de lactosa" fue mantenida durante 6 horas. Al final de la fermentación, las células fueron recolectadas por centrifugación a 5000-7000 xg durante 10 minutos y congelada como una pasta celular a -80°C. La pasta celular (318 gramos) fue recolectada a partir de 2.8 L de caldo celular.

Para preparar el extracto libre de células que contiene el polipéptido SEQ ID NO: 220, se suspendieron 50 g de la pasta celular húmeda en 150 mL de regulador EPPS 50 mM (pH 8.4) que contenía 50 µM de fosfato de piridoxal (PLP) y luego se deshizo utilizando un homogeneizador Microfluidics (Microfluidics, Newton, MA) (3 pasos a 18,000 psi), manteniendo la temperatura de la suspensión a menos de 15°C. El residuo celular fue retirado por centrifugación (20,000 xg durante 30 minutos). Se agregó DTT 2 mM al extracto celular clarificado.

Para preparar SEQ ID NO: 220 purificado, se cargaron 2 x 25 mL alícuotas de extracto celular clarificado cada una en una columna de 45 mL con resina Chelating Sepharose™ Fast Flow (forma níquel (II)) que había sido equilibrada previamente con EPPS 50 mM (pH 8.4) que contenía PLP 0.05 mM y cloruro de sodio 200 mM. Después de cargar la muestra, la columna fue lavada/eluida sucesivamente con 3-5 volúmenes del regulador de equilibrio, 3-5 volúmenes de regulador de equilibrio que contenían imidazol 25 mM, 3-5 volúmenes del regulador de equilibrio que contenían imidazol 100 mM y 3-5 volúmenes del regulador de equilibrio que contenían imidazol 500 mM. El eluyente imidazol 500 mM fue concentrado 10X con un dispositivo de filtración centrífugo Amicon (Billerica, MA) Centricon-70 (MWCO 10 kDa). El imidazol y el cloruro de sodio fueron removidos pasando a través de columnas de desalinización GE Healthcare PD10 equilibradas previamente con EPPS 50 mM (pH 8.4) que contenía PLP 0.05 mM. La concentración de proteína de la solución desalinizada fue determinada utilizando el kit de ensayo Pierce BCA (Rockford, IL). La pureza de cada fracción y el nivel de expresión en la fracción del extracto libre de células fue determinada mediante SDS-PAGE con geles con 4-15% de gradiente. Aproximadamente 450 mg de proteína que era ~90% pura fueron recuperados a partir de 50 mL del extracto celular clarificado. Se agregó DTT 2 mM a 10 mL de la proteína purificada. La proteína purificada fue dispensada en alícuotas (0.5-5 mL) y almacenada a -80°C.

Se llevaron a cabo reacciones a escala de mesa (250 mL) en fermentadores Sixfors de 0.7 L con agitación (Infors AG, Bottmingen, Suiza) bajo un espacio de cabeza con nitrógeno. La mezcla de reacción contenía fosfato de potasio 10 mM, MgCl₂ 1 mM, PLP 0.05 mM, piruvato de sodio 200 mM y D-triptófano 100 mM. La mezcla de reacción fue ajustada a la temperatura apropiada, y ajustada el pH apropiado con hidróxido de potasio. La aldolasa descrita en el Ejemplo 6 fue agregada como un extracto celular clarificado a 0.02 mg/mL de proteína objetivo. El polipéptido de DAT de SEQ ID NO: 220 fue agregado (bien sea como enzima purificada o como extracto celular clarificado) a 0.25 mg/mL de proteína objetivo.

El progreso de la reacciones fue seguido midiendo la concentración de monatina utilizando la metodología de LC/MS/MS descrita en el Ejemplo 3.

Comenzando con D-triptófano bajo las condiciones probadas el pH óptimo de las reacciones de formación de monatina utilizando el polipéptido de SEQ ID NO: 220 fue determinado como aproximadamente pH 8.0 y la temperatura óptima para las reacciones de formación de monatina utilizando el polipéptido de SEQ ID NO: 220 fue

determinada como aproximadamente 25°C. Estas reacciones tienen dinámicas complejas y las condiciones de reacción óptimas para la reacción de producción de monatina completas pueden no ser las mismas que las condiciones óptimas para las reacciones individuales catalizadas por el polipéptido de DAT.

Ejemplo 16 – La coexpresión de chaperonas para incrementar la expresión soluble de un polipéptido de DAT

- 5 Debido a que la expresión soluble del polipéptido de DAT de SEQ ID NO: 894 fue baja utilizando los protocolos de expresión estándar (tanto en IPTG 1 mM en LB o en el Novagen Overnight Express Autoinduction System2, véase Ejemplo 8), la fue examinada la expresión del polipéptido de SEQ ID NO: 894 y una variedad de chaperonas disponibles comercialmente.

Chaperonas:

- 10 El TaKaRa Chaperone Set (TAKARA BIO catálogo # 3340) consiste de cinco diferentes conjuntos de chaperonas desarrolladas por el HSP Research Institute, Inc. Están diseñadas para permitir una expresión eficiente de múltiples chaperonas moleculares conocidas por trabajar en cooperación en el proceso de plegamiento. El conjunto contenía lo siguiente:

Plásmido	Chaperona	Promotor	Inductor	Marcador de resistencia
pG-KJE8	dnaK-dnaJ-grpE; groES-groEL	araB; Pzt1	L-Arabinosa; tetraciclina	Cloranfenicol
pGro7	groES-groEL	araB	L-Arabinosa	Cloranfenicol
pKJE7	dnaK-dnaJ-grpE	araB	L-Arabinosa	Cloranfenicol
pG-Tf2	groES-groEL-tig	Pzt1	Tetraciclina	Cloranfenicol
pTf16	tig	araB	L-Arabinosa	Cloranfenicol

15 Protocolo de transformación

- Células BL21 (DE3) químicamente competentes (EMD Biosciences/Novagen catálogo # 69450) fueron transformadas con 20 ng de uno de los plásmidos de chaperona TaKaRa y 20 ng del polipéptido de SEQ ID NO: 893/pET30a (que codifica el polipéptido de SEQ ID NO: 894; véase por ejemplo C2 para la construcción del plásmido) por choque con calor durante 30 segundos a 42°C. Las células transformadas fueron recuperadas en 0.5 mL de medio SOC durante 1 hora a 37°C y sembradas sobre placas LB que contenían 50 mg/mL de kanamicina y 25 mg/mL de cloranfenicol. Las placas fueron incubadas durante la noche a 37°C. Las colonias fueron extraídas de las placas cultivadas durante la noche y utilizadas para inocular 5 mL de medio 2xYT que contenía 50 mg/L de kanamicina y 25 mg/L de cloranfenicol. Después de incubación durante la noche a 37°C, los plásmidos fueron aislados a partir de las pellas celulares utilizando un QUIAprep Spin Miniprep Kit (Qiagen catálogo # 27104). Los plásmidos fueron analizados por digestión con restricción con enzimas de un corte de New England Biolabs (Beverly, MA) tanto para el plásmido chaperona como para el plásmido de SEQ ID NO: 893/pET30a siguiendo el protocolo recomendado por el fabricante.

Plásmido	Enzima de restricción
pG-KJE8	XhoI
pGro7	XbaI
pKJE7	NheI
pG-Tf2	XhoI
pTf16	XbaI

El ADN aislado que contenía tanto SEQ ID NO: 893/pET30a como pKJE7 fue digerido con NheI y XbaI.

30 Estudios de expresión

Los matraces del Novagen Overnight Express™ AutoinductionSystem 2 (EMD Biosciences/Novagen catálogo #71366) que contenían soluciones 1-6, 50 mg/L de kanamicina y 25 mg/L de cloranfenicol (25 mL en cada matraz) fueron inoculados a partir de placas frescas de las células cotransformadas con un plásmido chaperona y SEQ ID NO: 893/pET30. En la inoculación los inductores requirieron que los plásmidos de chaperona también fueran agregados.

Plásmido	Concentración de inductor
pG-KJE8	2 mg/mL L-arabinosa; 10 ng/mL tetraciclina
pGro7	2 mg/mL L-arabinosa
pKJE7	2 mg/mL L-arabinosa
pG-Tf2	10 ng/mL tetraciclina
pTf16	2 mg/mL L-arabinosa

Las células fueron incubadas a 30°C durante la noche y recolectadas por centrifugación cuando la OD a 600 nm alcanzó 6 o más. Las células fueron lavadas con regulador EPPS 50 mM frío (pH 8.4), centrifugadas de nuevo, y usadas inmediatamente o bien congeladas a -80°C.

Los extractos celulares fueron preparados agregando 5 mL por g de pella de célula de BugBuster® (libre de aminos primarias) Extraction Reagent (EMD Biosciences/Novagen catálogo # 70923) con 5 µL/mL de inhibidor de proteasa Cocktail II (EMD Bioscience/Calbiochem catálogo #539132), 1 µL/mL de Benzonase® Nuclease (EMD Biosciences/Novagen catálogo # 70746), y 0.033 µL/mL de solución de r-Lisozima (EMD Biosciences/Novagen catálogo # 71110) a las células. Las suspensiones celulares fueron incubadas a temperatura ambiente con mezcla suave durante 15 minutos. Agitación a 14,000 rpm, durante 20 minutos (4°C) y los sobrenadantes fueron removidos cuidadosamente. Las concentraciones de proteína totales fueron determinadas utilizando el kit del ensayo de proteínas Pierce BCA (Pierce catálogo # 23225) con albúmina de suero bovino como estándar y un formato de placa de microtitulación. La expresión de la D-aminotransferasa fue analizada por SDS-PAGE utilizando geles de gradiente de 4-15% de poliacrilamida Bio-Rad Ready Gel® Precast (Bio-Rad Laboratories catálogo #161-1104). Los estándares de rango bajo de Bio-Rad SDS-PAGE (catálogo # 161-0304) fueron corridos como estándares sobre cada gel. Alícuotas de los extractos celulares (15 µg de proteína) fueron mezclados con regulador de carga de proteína que contenía 2% de SDS, 10% de glicerol, 12.5% de 2-mercaptoetanol, 0.1% de azul de bromofenol y Tris-HCl 62.5 mM (pH 8), se incubaron a 95°C durante 5 minutos, se enfriaron y luego cargaron sobre el gel. Además, la expresión de proteína soluble e insoluble combinada (proteína total) fue analizada para cada transformante. Una alícuota de 10 µL de cada suspensión celular antes de la centrifugación fue diluida en 90 µL de regulador de carga de proteína, incubada a 95°C durante 10 minutos y enfriada. Se cargaron 10 µL de cada solución enfriada sobre el gel.

El gel de proteína soluble mostró que la mejor expresión soluble del polipéptido que tiene la secuencia de SEQ ID NO: 894 ocurre cuando se coexpresan las chaperonas GroEL-GroES (pGro7).

La expresión del polipéptido de SEQ ID NO: 894 utilizando el plásmido alternativo en presencia de las chaperonas GroEL-GroES también fue examinada. El ácido nucleico de SEQ ID NO: 893 fue subclonado en el plásmido pMET1a utilizando las enzimas de restricción NdeI y XhoI de New England Biolabs. El plásmido es un derivado de pET23a (EMD Biosciences/Novagen catálogo # 69745-3) y porta el gen metE (insertado en el sitio NgoMIV del plásmido) y puede complementar la auxotrofia de la metionina de cepas de E. coli B834 (DE3) y E. coli BW30384 (DE3) ompTmetE ("EE2D"). La construcción de la cepa "EE2D" está descrita en WO 2006/066072. La construcción de un plásmido análogo a pMET1a que es un derivado de pET23d está descrita en la misma solicitud PCT.

El plásmido de SEQ ID NO: 893/pMET1a (25 ng) fue transformado en células electrocompetentes "EE2D" de manera individual o fue cotransformado con pGro7 (20 ng) utilizando el protocolo estándar de electroporación Bio-Rad para células E. coli con un sistema Bio-Rad Gene Pulser II (catálogo # 165-2111). Las células transformadas fueron recuperadas en 0.5 mL de medio SOC durante 1 hora a 37°C y sembradas sobre placas LB que contenían 100 mg/L de ampicilina o sobre placas que contenían 100 mg/L de ampicilina y 25 mg/L de cloranfenicol (transformantes de doble plásmido). Las placas fueron incubadas durante la noche a 37°C. Una colonia de cada placa fue utilizada para inocular 50 mL de Novagen Overnight Express™ AutoinductionSystem 2 que contenía soluciones 1-6, 100 mg/mL de ampicilina y 2 mg/mL de L-arabinosa. El cultivo inoculado con células que contenían el plásmido pGro7 también contenía 25 mg/mL de cloranfenicol. Las células fueron incubadas a 30°C durante la noche y recolectadas por centrifugación cuando la OD₆₀₀ alcanzó 6 ó más. Las pellas de células fueron lavadas con regulador EPPS frío 50 mM (pH 8.4), centrifugadas de nuevo y usadas inmediatamente o congeladas a -80°C. Los extractos celulares fueron preparados como se describió más arriba utilizando el Novagen BugBuster® (libre de

amina primaria) Extraction Reagent. La expresión de D-aminotransferasa soluble y total fue analizada por SDS-PAGE como se describió más arriba.

El gel mostró que la expresión del polipéptido de SEQ ID NO: 894 soluble era mayor cuando las proteínas GroEL-GroES fueron coexpresadas. Sin embargo la expresión soluble no fue tan alta como el constructo pET30a descrito más arriba.

El efecto de la temperatura de incubación durante la expresión también fue examinado. Un cultivo de 5 mL de LB que contenía 100 mg/L de ampicilina y 25 mg/L de cloranfenicol fue inoculado a partir de una placa fresca de EE2D:: SEQ ID NO: 894 PENT1a+pGro7. El cultivo fue incubado durante la noche a 30°C y luego utilizado para inocular 3 matraces, conteniendo cada uno 50 mL de un Novagen Overnight Express™ Autoinduction System 2 que contenía soluciones 1-6, 100 mg/mL de ampicilina, 25 mg/L de cloranfenicol y 2 mg/mL de L-arabinosa. Un matraz fue incubado a 20°C, el segundo a 25°C y el tercero a 30°C. Las células fueron recolectadas cuando la OD₆₀₀ alcanzó 6 o más. Las células fueron recolectadas y los extractos celulares fueron preparados como se describió más arriba. Las concentraciones de proteína total fueron determinadas utilizando el kit de ensayo de proteínas Pierce BCA (Pierce catálogo # 23225) con albúmina de suero bovino como estándar y un formato de placa de microtitulación. La expresión de la D-aminotransferasa fue analizada utilizando la Bio-Rad Experion Pro 260 Automated Electrophoresis Station siguiendo el protocolo del fabricante con las soluciones de extracto celular diluidas a 1 mg/mL. Los resultados se muestran en la Tabla 35. Parece que la temperatura más baja dio la cantidad máxima de expresión del polipéptido de SEQ ID NO: 894.

Tabla 35

Línea	Muestra	Temp	Expresión de DAT estimada
1	Pro260 Ladder		
2	Extracto celular de EE2D::23463pMET1+pGRO7	20°C	23%
3	Extracto celular de EE2D::23463pMET1+pGRO7	25°C	21%
4	Extracto celular de EE2D::23463pMET1+pGRO7	30°C	19%

Protocolo de ensayo de actividad

La actividad enzimática del DAT de SEQ ID NO: 894 coexpresado con las chaperonas GroEL-GroES fue probada siguiendo el protocolo estándar de reacción de monatina. En resumen, cada tubo de ensayo contenía lo siguiente (en un total de 2 mL): 0.050 mg/mL de aldolasa en extracto celular (asumiendo una expresión soluble de 20%); 1.0 mg/mL de D-aminotransferasa en extracto celular (asumiendo una expresión soluble del 20% para un extracto que contenía el polipéptido de SEQ ID NO: 894) o D-aminotransferasa de *B. sphaericus* purificada; 0.01 % de Tween-80; piruvato de sodio 200 mM, D-triptófano 100 mM; EPPS 100 mM (pH 8.2); MgCl₂ 1 mM; PLP 0.05 mM; y fosfato de potasio 10 mM.

Las reacciones fueron incubadas a temperatura ambiente en una cámara anaeróbica de Coy Laboratory Products, Inc. para minimizar la exposición al oxígeno. Todos los componentes excepto las enzimas fueron mezclados entre sí (el triptófano no se disolvió completamente hasta al menos 1 hora después de la adición de las enzimas). Las reacciones fueron iniciadas mediante la adición de las enzimas. Las muestras fueron extraídas a 1, 4, 8 y 22 horas. También se llevó a cabo una reacción de control utilizando DAT de *B. sphaericus* purificado 1 mg/mL. La reconstrucción, expresión y purificación de este DAT están descritas en el Ejemplo 6. Las concentraciones de los sustratos y los productos fueron medidas como se describe en el Ejemplo 3.

Los resultados se muestran en la Tabla 36. A 22 horas, la concentración de monatina fue 9.2 mM cuando el polipéptido de SEQ ID NO: 894 estaba presente y 12.4 mM cuando se utilizó la enzima de *B. sphaericus*. La concentración del coproducto HMG fue significativamente menor cuando el polipéptido de SEQ ID NO: 894 estaba en la mezcla de ensayo (<1/3 de la concentración cuando se compara con la muestra de ensayo que contiene la enzima de *B. sphaericus*). Las concentraciones de HMG fueron evaluadas comparando las áreas de pico de las muestras derivadas con OPA.

Tabla 36. Formación de monatina (mM)

Polipéptido (SEQ ID NO:)	1 hora	4 horas	8 horas	22 horas
894 DAT + GroEL-ES	1.3	4.5	6.8	9.2
B. sphaericus DAT	1.6	5.7	8.3	12.4

Ejemplo 17 – Uso del sistema ArcticExpress™ para incrementar la expresión del polipéptido de DAT soluble

Debido a que la expresión soluble del polipéptido de SEQ ID NO: 894 fue baja utilizando los protocolos de expresión estándar (bien sea IPTG 1 mM en LB o Novagen Overnight Express™ Autoinduction System 2 – véase Ejemplo 8) la fue examinada la expresión del plásmido de SEQ ID NO: 893/pMET1a en el sistema Stratagene ArcticExpress™.

El sistema Stratagene ArcticExpress™ contiene células competentes de *E. coli* que portan las chaperonas psicrófilas Cpn 10 y Cpn 60. Estas son chaperonas aisladas del organismo psicrófilo *Oleispira antarctica*. La Cpn 10 y la Cpn 60 muestran alta similitud de secuencia a las chaperonas de *E. coli* GroEL y GroES, respectivamente, y tienen altas actividades de plegamiento de proteína a 4-12°C. Las células anfitrionas ArcticExpress™ son derivadas de la cepa de *E. coli* BL21. No solamente estas células carecen de proteasa Lon, sino que también han sido manipuladas para ser deficientes en la proteasa OmpT.

Protocolo de transformación

El plásmido SEQ ID NO: 893/pMET1a (descrito en el Ejemplo 16) fue transformado en células ArcticExpress™ (DE3) químicamente competentes (catálogo # 230192) siguiendo el protocolo del fabricante. Las células transformadas fueron recuperadas en 0.5 mL de medio SOC durante 1 hora a 37°C y sembradas sobre placas LB que contenían 100 mg/L de ampicilina. Las placas fueron incubadas durante la noche a 37°C y luego almacenadas a 4°C.

Protocolo de expresión

Las colonias de las placas de transformación fueron usadas para inocular 5 mL de medio 2xYT que contenía 100 mg/L de ampicilina y 10 mg/L de gentamicina y se incubaron durante la noche a 30°C. Los matraces de Novagen Overnight Express™ Autoinduction System 2 (EMD Biosciences/Novagen catálogo # 71366) que contenía las soluciones 1-6, con 100 mg/L de ampicilina y 12 mg/L de gentamicina fueron inoculadas utilizando los cultivos nocturnos. Después de la incubación a 30°C durante 6 horas los cultivos Overnight Express™ fueron movidos a incubadoras a 15°C o 20°C o 25°C. Las incubaciones fueron continuadas hasta que la OD a 600 nm de los cultivos alcanzó 6 o más. Las células fueron recolectadas por centrifugación, lavadas con EPPS y 50 mM frío, pH 8.4, y las pellas de células fueron congeladas a -80°C.

Los extractos celulares fueron preparados agregando 5 mL por g de pella de células de BugBuster® (libre de amina primaria) Extraction Reagent (EMD Biosciences/Novagen catálogo # 70923), con 5 µL/mL de Protease Inhibitor Cocktail II (EMD Bioscience/Calbiochem catálogo # 539132), 1 µL/mL de Benzonase Nuclease (EMD Biosciences/Novagen catálogo # 70746), y 0.033 µL/mL de r-L-lisozima™ en solución (EMD Biosciences/Novagen catálogo # 71110) a las células. Las suspensiones celulares fueron incubadas a temperatura ambiente con agitación suave durante 15 minutos; agitación a 14,000 rpm durante 20 minutos (4°C) y los sobrenadantes fueron removidos cuidadosamente. Las concentraciones de proteína total fueron determinadas utilizando el kit de ensayo de proteína Pierce BCA (Pierce catálogo # 23225) con albúmina de suero bovino como estándar y un formato de placa de microtitulación. La expresión de la D-aminotransferasa fue analizada utilizando la Bio-Rad Experion Pro260 Automated Electrophoresis Station siguiendo el protocolo del fabricante con las soluciones de extractos celulares diluidas a 1 mg/mL.

Los resultados de electroforesis muestran que el sistema ArcticExpress™ incrementó significativamente la expresión soluble del polipéptido de SEQ ID NO: 894 en comparación con la expresión sin chaperonas o cuando se coexpresó con las chaperonas de *E. coli* GroEL-GroES descritas en el Ejemplo 16. La expresión soluble fue más alta a temperaturas más bajas, pero todavía muy alta a 25°C.

Línea	Muestra	Temperatura de incubación	Expresión estimada de DAT
1	Pro260 Ladder		
2	Extracto celular de ArcticExpress (DE3)::894pMET1	15°C	58%
3	Extracto celular de ArcticExpress (DE3)::894pMET1	20°C	46%
4	Extracto celular de ArcticExpress (DE3)::894pMET1	25°C	47%

Protocolo de ensayo de actividad:

La actividad enzimática del polipéptido de SEQ ID NO: 894 expresado en el sistema ArcticExpress™ fue probada siguiendo la formación de monatina en la presencia de aldolasa descrita en el Ejemplo 6. Cada tubo de ensayo contenía lo siguiente (en un total de 2 mL): 0.010 mg/mL de aldolasa en extracto celular (asumiendo una expresión soluble del 20%); 1.0 o 2.0 de mg/mL del polipéptido de SEQ ID NO: 894 en extracto celular (asumiendo una expresión soluble del 50% para el extracto que contenía el polipéptido de SEQ ID NO: 894) o D-aminotransferasa purificada de *B. sphaericus*; 0.01 % de Tween-80; piruvato de sodio 200 mM, D-triptófano 100 mM; EPPS 100 mM (pH 8.2); MgCl₂ 1 mM; PLP 0.05 mM; y fosfato de potasio 10 mM.

Las reacciones fueron incubadas a temperatura ambiente en una cámara anaeróbica de Coy Laboratory Products, Inc. para minimizar la exposición al oxígeno. Todos los componentes excepto las enzimas fueron mezclados entre sí (el triptófano no se disolvió completamente hasta al menos 1 hora después de la adición de las enzimas). Las reacciones fueron iniciadas mediante la adición de las enzimas. Las muestras fueron extraídas a 1, 4, 7 y 22 horas. Las reacciones de control utilizando 1 o 2 mg/mL de D-aminotransferasa de *B. sphaericus* purificada también fueron llevadas a cabo. Las concentraciones de los sustratos y los productos fueron medidas como se describe en el Ejemplo 3. Los resultados se muestran en la Tabla 37. A las 22 horas, la concentración de monatina fue 8.2 mM cuando el polipéptido de SEQ ID NO: 894 estaba presente a 1 mg/mL y 10.5 mM a 2 mg/mL del polipéptido de DAT. Cuando la enzima de *B. sphaericus* fue agregada a 1 mg/mL, la concentración de monatina a las 22 horas fue de 10.7 mg/mL; a 2 mg/mL, la concentración de monatina fue de 14.5 mM. La estereopureza (determinada según el protocolo de derivación con FDAA del Ejemplo 3) del producto fue >98% de R,R con ambas enzimas y concentraciones de enzimas. La concentración del coproducto HMG fue significativamente menor cuando el polipéptido de SEQ ID NO: 894 fue usado (~1/3 la concentración comparada con las muestras de ensayo que contenían la enzima de *B. sphaericus* a cualquier concentración de enzimas. Las concentraciones de HMG fueron evaluadas comparando las áreas de pico de las muestras derivadas con OPA.

Tabla 37. Formación de monatina (mM)

Polipéptido (SEQ ID NO:)	1 hora	4 horas	7 horas	22 horas
894 DAT (1 mg/mL)	0.9	2.3	3.6	8.3
894 DAT (2 mg/mL)	1.4	3.7	6.3	10.5
<i>B. sphaericus</i> DAT (1 mg/mL)	1.0	4.2	6.1	10.7
<i>B. sphaericus</i> DAT (2 mg/mL)	1.5	6.7	8.2	14.5

Ejemplo 18 – Uso del sistema Stratagene ArcticExpress™ para incrementar la expresión de DAT soluble

Protocolo de transformación

Los plásmidos SEQ ID NO: 891/pET30a (que codifica el polipéptido de SEQ ID NO: 892), SEQ ID NO: 873/pET30a (que codifica el polipéptido de SEQ ID NO: 874) y el DAT de *Clostridium beijerinckii* (CbDAT) en pET30a fueron transformados en células Stratagene ArcticExpress™ (DE3) químicamente competentes (catálogo # 230192) siguiendo el protocolo del fabricante. La clonación de estos genes está descrita en el Ejemplo 4 y los resultados del ensayo están en el Ejemplo 9. Las células transformadas fueron recuperadas en 0.5 mL de medio SOC durante 1 hora a 37°C y sembradas sobre placas LB que contenían 100 mg/L de ampicilina y 13 mg/L de gentamicina. Las placas fueron incubadas a temperatura ambiente durante 2 días y luego almacenadas a 4°C.

Protocolo de expresión

Las colonias de las placas de recuperación de transformación fueron utilizadas para inocular 5 mL de medio 2xYT que contenía 50 mg/L de kanamicina y 13 mg/L de gentamicina; los cultivos líquidos fueron incubados a 6 horas a 30°C. Los matraces de Novagen Overnight Express™ Autoinduction System 2 (EMD Biosciences/Novagen catálogo # 71366) contenían las soluciones 1-6, con 100 mg/L de ampicilina y 13 mg/L de gentamicina y fueron inoculados a partir de los cultivos de 5 mL. Después de la incubación a 30°C durante 5-6 horas, los cultivos fueron movidos a una incubadora a 15°C. Las incubaciones a 15°C fueron continuadas hasta que la OD600 de los cultivos alcanzó 6 o más. Las células fueron recolectadas por centrifugación, lavadas con EPPS 50 mM, pH 8.4, y luego las pellas de células fueron congeladas a -80°C.

- 10 Los extractos celulares fueron preparados agregando 5 ml por g de pella de células de BugBuster® (libre de amina primaria) Extraction Reagent (EMD Biosciences/Novagen catálogo # 70923) con 5 µL/mL de Protease Inhibitor Cocktail II (EMD Bioscience/Calbiochem catálogo # 539132), 1 µL/mL de Benzonase Nuclease (EMD Biosciences/Novagen catálogo # 70746), y 0.033 µL/mL de r-L-lisozima™ en solución (EMD Biosciences/Novagen catálogo # 71110) a las células. Las suspensiones celulares fueron incubadas a temperatura ambiente con agitación suave durante 15 minutos; agitación a 14,000 rpm durante 20 minutos (4°C) y los sobrenadantes fueron removidos cuidadosamente. Las concentraciones de proteína total fueron determinadas utilizando el kit de ensayo de proteína Pierce BCA (Pierce catálogo # 23225) con albúmina de suero bovino como estándar y un formato de placa de microtitulación. La expresión del DAT fue analizada utilizando la Bio-Rad Experion Pro260 Automated Electrophoresis Station siguiendo el protocolo del fabricante con las soluciones de extractos celulares diluidas a 1 mg/mL. Los resultados se muestran en las Tablas 38 y 39.

- 25 El resultado de electroforesis muestra que el polipéptido de SEQ ID NO: 874 se expresó en forma soluble ligeramente mejor que el polipéptido de SEQ ID NO: 892 utilizando el sistema ArcticExpress™. La expresión soluble del CbDAT varió dependiendo de la colonia extraída en la placa de transformación. Ninguno de estos polipéptidos de DAT se expresó en forma soluble utilizando el sistema ArcticExpress™ así como el polipéptido de SEQ ID NO: 894 descrito en el Ejemplo 16 fue expresado.

Tabla 38

Línea	Muestra	Temperatura de incubación	Expresión estimada de DAT
1	Pro260 Ladder		
2	Extracto celular (colonia # 1) ArcticExpress (DE3)::891pET30a	15°C	11%
3	Extracto celular (colonia # 2) ArcticExpress (DE3)::891pET30a	15°C	9%
4	Extracto celular (colonia # 1) ArcticExpress (DE3)::873pET30a	15°C	15%
5	Extracto celular (colonia # 2) ArcticExpress (DE3)::873pET30a	15°C	10%
6	Extracto celular (colonia # 1) ArcticExpress (DE3)::891pET30a	15°C	9%
7	Extracto celular (colonia # 2) ArcticExpress (DE3)::891pET30a	15°C	8%
9	Extracto celular (colonia # 1) ArcticExpress (DE3)::873pET30a	15°C	15%
10	Extracto celular (colonia # 2) ArcticExpress (DE3)::873pET30a	15°C	11%

Tabla 39

Línea	Muestra	Temperatura de incubación	Expresión estimada de DAT
1	Pro260 Ladder		
2	Extracto celular (colonia #1) ArcticExpress (DE3)::CbDAT en pET30a	15°C	9%

Línea	Muestra	Temperatura de incubación	Expresión estimada de DAT
3	Extracto celular (colonia #2) ArcticExpress (DE3)::CbDAT en pET30a	15°C	19%
4	Extracto celular (colonia #3) ArcticExpress (DE3)::CbDAT en pET30a	15°C	13%
5	Extracto celular (colonia #4) ArcticExpress (DE3)::CbDAT en pET30a	15°C	4%

Protocolo de ensayo de actividad

La actividad enzimática de los polipéptidos de DAT expresados en el sistema ArcticExpressTM fue probada siguiendo la formación de la monatina en la presencia de aldolasa descrita en el Ejemplo 6. Cada tubo de ensayo contenía lo siguiente (en un total de 3 mL): 0.050 mg/mL de aldolasa en extracto celular (estimando un 20% de expresión soluble); 0.5 mg/mL de polipéptido de DAT en extracto celular (estimando la expresión celular a partir de los datos Experion) o DAT purificado de *B. sphaericus*; 0.01 % de Tween-80; piruvato de sodio 200 mM, D-triptófano 100 mM; EPPS 50 mM (pH 8.2); MgCl₂ 1 mM; PLP 0.05 mM; y fosfato de potasio 10 mM.

Las reacciones fueron incubadas a temperatura ambiente en una cámara anaeróbica de Coy Laboratory Products, Inc. para minimizar la exposición al oxígeno. Todos los componentes excepto las enzimas fueron mezclados entre sí (el triptófano no se disolvió completamente hasta al menos 1 hora después de la adición de las enzimas). Las reacciones fueron iniciadas mediante la adición de las enzimas. Las muestras fueron retiradas a las 2, 4.5, 9 y 24 horas. Las reacciones de control con D-aminotransferasa purificada de *B. sphaericus* fueron ejecutadas también. Las concentraciones de los sustratos y productos fueron medidas como se describe en el Ejemplo 3. A las 24 horas, los ensayos que contenían el polipéptido de SEQ ID NO: 892 u 894 habían producido aproximadamente el mismo título de monatina que la reacción de control con DAT de *B. sphaericus*. La reacción con DAT de *C. beijerinckii* produjo menos de 1/8 del producto monatina, mientras que el polipéptido de SEQ IDNO: 874 produjo aproximadamente la mitad. La estereopureza del producto a 24 horas (determinado por el protocolo de derivación con FDAA en el Ejemplo 3) fue de 96% de R,R de monatina o más para el polipéptido de SEQ ID NO: 892. La concentración del subproducto HMG (ácido 4-hidroxi-4-metil glutámico) fue medida para las reacciones con los polipéptidos de SEQ ID NO: 892, SEQ ID NO: 894 y DAT de *B. sphaericus*. El ensayo con polipéptidos que tenían la secuencia mostrada en SEQ ID NO: 894 produjo bastante menos del subproducto HMG que los otros dos (aproximadamente 20% del producido por el ensayo que contenía el DAT de *B. sphaericus* y aproximadamente 40% del producido por el ensayo con el polipéptido de SEQ ID NO: 892. Las concentraciones de HMG fueron estimadas comparando las áreas de pico de las muestras derivadas postcolumna con OPA.

Tabla 40. Formación de monatina (mM)

Polipéptido (SEQ ID NO:)	2 horas	4.5 horas	9 horas	24 horas
<i>C. beijerinckii</i> DAT	0.3	0.7	0.8	0.9
874 DAT	1.8	3.2	4.1	4.4
892 DAT	1.6	3.7	5.4	8.4
894 DAT	1.4	2.8	4.5	8.4
<i>B. sphaericus</i> DAT	1.2	2.9	4.3	8.1

Estos resultados indican que cuando se expresan bajo las condiciones apropiadas, los polipéptidos que tienen la secuencia de SEQ ID NO: 892 y 894 pueden ser utilizados en reacciones para producir R,R monatina altamente pura en un título tan alto como la coenzima de control positiva.

Ejemplo 19 – Evaluación de anfitriones con expresión alternativa para incrementar la expresión soluble de un DAT

Debido a que la expresión soluble del polipéptido de SEQ ID NO: 894 era baja cuando el ácido nucleico estaba expresado en BL21 (DE3) (véase Ejemplo 8), se evaluaron anfitriones de expresión alternativos para mejorar la expresión soluble. Los anfitriones OverExresTM C41(DE3) y C43(DE3) contienen mutaciones genéticas

seleccionadas fenotípicamente para conferir tolerancia a la toxicidad y expresar algunas proteínas tóxicas en títulos más altos que otros anfitriones de *E. coli*.

Protocolo de transformación

El plásmido SEQ ID NO: 893/pET30a fue transformado en células electrocompetentes del OverExpress™ C41(DE3) y C43(DE3) (Lucigen catálogo # 60341 y 60345, respectivamente) utilizando un sistema Bio-Rad Gene Pulsar II siguiendo el protocolo del fabricante. Las mezclas de transformación fueron recuperadas en 1 mL de medio SOC durante 1 hora a 37°C y sembrada sobre placas LB que contenían 50 mg/L de kanamicina. Las placas fueron incubadas durante la noche a 37°C. Se sembraron por contacto colonias múltiples sobre placas frescas y se analizaron en cuanto al tamaño del inserto apropiado utilizando PCR de colonia. Una alícuota pequeña de células fue suspendida en 0.025 mL de H₂O e incubada a 95°C durante 10 minutos. Después de enfriar, se utilizaron 2 µL de cada una de las suspensiones como plantilla en reacciones de 0.025 mL, conteniendo cada una 0.5 µL de cebador de promotor T7 (0.1 mM), 0.5 µL de cebador determinador T7 (0.1 mM), 0.5 µL de mezcla de nucleótido de PCR (Roche # 12526720; 10 mM de cada nucleótido), 2.5 µL de regulador de polimerasa Roche Expand DNA # 2 y 0.5 µL de polimerasa Expand DNA (Roche Expand High Fidelity PCR System catálogo # 1732650). El programa del termociclador de 3 etapas fue corrido durante 25-30 ciclos: 1 minuto a 94°C; 1 minutos a 54°C, 1.3 minutos a 72°C con una etapa de pulmimento final de 7 minutos a 72°C.

Estudios de expresión

Los matraces de Novagen Overnight Express™ Autoinduction System 2 (EMD Biosciences/Novagen catálogo # 71366) que contenían soluciones 1-6, y 50 mg/L de kanamicina (40 mL en cada matraz) fueron inoculados a partir de las placas de siembra por contacto de las células C41(DE3) y C43(DE3) transformadas (2 siembras por cada una de las transformaciones). Las células fueron incubadas a 30°C durante la noche y recolectadas por centrifugación cuando la OD₆₀₀ alcanzó 6 o más. Las células fueron lavadas con regulador frío, fueron centrifugadas de nuevo y usadas inmediatamente o congeladas a -80°C.

Los extractos celulares fueron preparados agregando 5 ml por g de pella de célula de BugBuster® (libre de amina primaria) Extraction Reagent (EMD Biosciences/Novagen catálogo # 70923) con 5 µL/mL de Protease Inhibitor Cocktail II (EMD Bioscience/Calbiochem catálogo # 539132), 1 µL/mL de Benzonase® Nuclease (EMD Biosciences/Novagen catálogo # 70746), y 0.033 µL/mL de r-L-lisozima™ en solución (EMD Biosciences/Novagen catálogo # 71110) a las células. Las suspensiones celulares fueron incubadas a temperatura ambiente con agitación suave durante 15 minutos; agitación a 14,000 rpm durante 20 minutos (4°C) y los sobrenadantes fueron removidos cuidadosamente. Las concentraciones de proteína total fueron determinadas utilizando el kit de ensayo de proteína Pierce BCA (Pierce catálogo # 23225) con albúmina de suero bovino como estándar y un formato de placa de microtitulación. La expresión del polipéptido de DAT fue analizada por SDS-PAGE utilizando geles en gradiente de poliacrilamida al 4-15% de Bio-Rad Ready GelR Precast (Bio-Rad Laboratories catálogo # 161-1104). Se corrieron estándares de bajo rango Bio-Rad SDS-PAGE (catálogo # 161-0304) sobre cada gel. Las alícuotas de los extractos celulares (15 µg de proteína) fueron mezcladas con el regulador de carga de proteína que contenía 2% de SDS, 10% de glicerol, 12.5% de 2-mercaptoetanol, 0.1% de azul de bromofenol y Tris-HCl 62.5 mM (pH 8), se incubaron a 95°C durante 5 minutos, se enfriaron y luego se cargaron sobre el gel. Además, la expresión de proteína soluble e insoluble combinadas (proteína total) fue analizada para los transformantes. Una alícuota de 10 µL de cada suspensión celular antes de la centrifugación fue diluida en 90 µL de regulador de carga de proteína, incubada a 95°C durante 10 minutos y enfriada. Se cargaron 10 µL de cada solución enfriada sobre el gel.

El gel de electroforesis muestra que la proteína se expresó mejor en el anfitrión C41(DE3) que en el anfitrión C43(DE3), sin embargo la expresión soluble aparente no fue más alta que cuando se utilizaron las células BL21(DE3).

Ejemplo 20 – Evaluación de la expresión a baja temperatura para incrementar la expresión de DAT soluble

Debido a que la expresión soluble de la D-aminotransferasa de SEQ ID NO: 894 fue baja cuando el gen fue expresado en la cepa de *E. coli* BL21 (DE3) (véase Ejemplo 8), el gen fue insertado en vectores con promotores de proteína A de choque en frío para evaluar la expresión a baja temperatura.

Los vectores Takara pCold Expression son cuatro vectores diferentes que utilizan el promotor de proteína de choque en frío (cspA) para la expresión de proteína recombinante de alta pureza de alto rendimiento en *E. coli*. Estos vectores inducen selectivamente la síntesis de proteína objetivo a bajas temperaturas (15°C) cuando la síntesis de otras proteínas es suprimida y disminuye la actividad de la proteasa. Además del promotor cspA, todos los cuatro vectores contienen un operador lac (para control de expresión), gen de resistencia a la ampicilina (amp^r), origen ColE1 de la replicación, fragmento de M13 IG y sitio de clonación múltiple (MCS). Tres de estos vectores también contienen un elemento potenciador de la traducción (TEE), una secuencia con etiqueta His, y/o un sitio de escisión de factor Xa.

Protocolo de clonación

- El ácido nucleico de DAT de SEQ ID NO: 894 del plásmido SEQ ID NO: 894 pET30a fue subclonado en los vectores Takara pCold en los sitios NdeI y XhoI de los vectores pCOLDII (que contienen un TEE y una secuencia de etiqueta His), pCOLDIII (que contiene un TEE) y pCOLDIV. El vector digerido y las bandas de inserto fueron purificadas por gel utilizando un QIAquick Gel Extraction Kit (Qiagen catálogo # 28704) y ligados utilizando un Roche Rapid DNA Ligation Kit (catálogo # 1635379) siguiendo el protocolo del fabricante. Las muestras de ligazón fueron transformadas en células químicamente competentes Invitrogen OneShot TOP10 (catálogo # C404003) por choque con calor a 42°C. Después de recuperar en 500 µL de medio SOC durante 1 hora a 37°C, las mezclas de transformación fueron sembradas sobre placas LB que contenían 100 µg/L de ampicilina y se incubaron a 37°C durante la noche. Las colonias fueron retiradas de las placas de transformación y utilizadas para inocular cultivos de 5 mL de LB que contenían 100 mg/mL de ampicilina que fueron incubadas durante la noche a 37°C. El ADN del plásmido fue purificado a partir de los cultivos de 5 mL utilizando un QIAprep Spin Miniprep Kit (Qiagen catálogo # 27104). Los insertos fueron verificados en la digestión de la restricción con NdeI y XhoI por secuenciación (Agencourt Bioscience Corp, Beverly, MA).
- Los plásmidos de SEQ ID NO: 894DAT pCOLD fueron transformados en células químicamente competentes Stratagene ArcticExpress™ (DE3) y células Novagen BL21(DE3) siguiendo los protocolos del fabricante. Las mezclas de transformación fueron recuperadas en 0.5-1 mL de medio SOC durante 1 hora a 37°C y sembradas en placas LB que contenían 100 mg/mL de ampicilina y 13 mg/L de gentamicina (ArcticExpress™ (DE3)) o 100 g/mL de ampicilina (BL21(DE3)). Las placas fueron incubadas durante la noche a 37°C.
- Estudios de expresión
- Los matraces de Novagen Overnight Express™ AutoinductionSystem 2 (EMD Biosciences/Novagen catálogo # 71366) que contenían las soluciones 1-6, y 100 mg/L de ampicilina y 13 mg/L de gentamicina (arcticExpress™ (DE3) o 100 mg/mL de ampicilina (BL21(DE3) fueron inoculadas a partir de las placas de siembra por contacto de las células transformadas (2 siembras para cada una de las transformaciones de SEQ ID NO: 893/pCOLDII, SEQ ID NO: 893/pCOLDIII y SEQ ID NO: 893/pCOLDIV).
- Después de la incubación a 30°C durante 3-5 horas los cultivos fueron movidos a una incubadora a 15°C. Las incubaciones a 15°C continuaron hasta que la OD a 600 nm de los cultivos alcanzó 6 o más. Las células fueron recolectadas por centrifugación, lavadas con regulador frío y luego las pellas de células fueron congeladas a -80°C.
- Los extracto celulares fueron preparados agregando 5 ml por g de la pella celular de BugBusterR (libre de amina primaria) Extraction Reagent (EMD Biosciences/Novagen catálogo # 70923) con 5 µL/mL de Protease Inhibitor Cocktail II (EMD Bioscience/Calbiochem catálogo # 539132), 1 µL/mL de BenzonaseR Nuclease (EMD Biosciences/Novagen catálogo # 70746), y 0.033 µL/mL de r-L-lisozima™ en solución (EMD Biosciences/Novagen catálogo # 71110) a las células. Las suspensiones celulares fueron incubadas a temperatura ambiente con mezcla suave durante 15 minutos; agitación a 14,000 rpm durante 20 minutos (4°C) y los sobrenadantes fueron removidos cuidadosamente. Las concentraciones de proteína total fueron determinadas utilizando el kit de ensayo de proteína Pierce BCA (Pierce catálogo # 23225) con albúmina de suero bovino como estándar y un formato de placa de microtitulación. La expresión de la D-aminotransferasa fue analizada utilizando la Bio-Rad Experion™ Pro260 Automated Electrophoresis Station siguiendo los protocolos del fabricante con las soluciones de extractos celulares diluidas a 1 mg/mL. Los resultados se muestran en las Tablas 41 y 42.

Tabla 41

Línea	Muestra	% de expresión de DAT estimado
L	Pro260 Ladder	
1	BL21(DE3)::SEQ ID NO:893/pCOLDII#1	8.7
2	BL21(DE3):: SEQ ID NO:893/pCOLDII#2	8.5
3	ArcticExpress™(DE3):: SEQ ID NO:893/pCOLDII#1	9.8
4	ArcticExpress™(DE3):: SEQ ID NO:893/pCOLDII#2	6.4

Tabla 42

Línea	Muestra	% de expresión de DAT estimado
L	Pro260 Ladder	
1	BL21(DE3):: SEQ ID NO:893/pCOLDIII#1	4.3
2	BL21(DE3):: SEQ ID NO:893/pCOLDIII#2	2.3
3	BL21(DE3):: SEQ ID NO:893/pCOLDIV#1	14.6
4	BL21(DE3):: SEQ ID NO:893/pCOLDIV#2	14.2
5	BL21(DE3):: SEQ ID NO:893/pCOLDIII#1	4.4
6	BL21(DE3):: SEQ ID NO:893/pCOLDIII#2	5.2
7	BL21(DE3):: SEQ ID NO:893/pCOLDIV#1	13.8
8	BL21(DE3):: SEQ ID NO:893/pCOLDIV#2	16.1
9	BL21(DE3):: SEQ ID NO:893/pCOLDIV#1	16.2
10	BL21(DE3):: SEQ ID NO:893/pCOLDIV#2	<u>16.8</u>

Los resultados de Experion Pro260 muestran que la proteína de DAT de SEQ ID NO: 894 se expresa mejor cuando el ácido nucleico es incorporado en el vector pCOLDIV que en el vector pCOLDII o pCOLDIII. A partir de los experimentos mostrados anteriormente, el nivel de expresión promedio para SEQ ID NO: 893/pCOLDII fue aproximadamente 8%, independientemente del anfitrión de expresión usado; el nivel de expresión promedio para SEQ ID NO: 893/pCOLDIII fue aproximadamente 4%, mientras que el nivel de expresión promedio para SEQ ID NO: 893/pCOLDIV fue -15%. Estos niveles de expresión son significativamente menores que los descritos en los Ejemplos 16 cuando el ácido nucleico de SEQ ID NO: 893 fue coexpresado con las chaperonas GroEL-GroES y en el Ejemplo 17 cuando el ácido nucleico fue expresado en el sistema Startagene ArcticExpress™.

Ejemplo 21 – Modificación de codón de un DAT

Un intento para mejorar la solubilidad del polipéptido de SEQ ID NO: 894 expresado en E. coli fue llevado a cabo con la presunción de que, disminuyendo la rata de traducción en E. coli permitiría un mayor tiempo para un plegamiento apropiado del polipéptido de SEQ ID NO: 894, dando por lo tanto una expresión superior de la proteína soluble. Una búsqueda BLAST (NCBI) de la secuencia del polipéptido de SEQ ID NO: 894 reveló que algunas de las secuencias públicas más cercanamente relacionadas eran de especies de Clostridium, específicamente Clostridium beijerinckii. El Ejemplo 9 describe los resultados de clonar, expresar y ensayar el CbDAT y su uso en las reacciones de formación de monatina. Específicamente, la expresión fue alta en la fracción de proteína total pero muy baja en la fracción de proteína soluble.

Las tablas de utilización de codón de C. beijerinckii y E. coli K12 fueron comparadas. Varios codones raramente usados en C. beijerinckii que fueron encontrados como altamente abundantes en E. coli K2 (Tabla 43). Es posible que estos codones produzcan pausas de traducción en C. beijerinckii, mientras que en un anfitrión E. coli K12 pueden no ser una pausa. En la secuencia de SEQ ID NO: 894, se identificaron 4 dobles en los cuales codones raros en tándem para C. beijerinckii se han convertido en "no raros" (esto es, abundante) en E. coli K12. La meta era cambiar estos codones en codones raros para la expresión en el anfitrión E. coli K12 utilizando la tabla de utilización del codón de E. coli K12. Se diseñaron cebadores para cambiar estos dobles. Se utilizó SEQ ID NO: 893/PpET30a (descrito en el Ejemplo 4) como plantilla y se llevó a cabo la mutación de acuerdo con las instrucciones del kit Stratagene QuickChange. Los cebadores utilizados para modificar las secuencias de ácidos nucleicos de SEQ ID NO: 893 se muestran más abajo, junto con el gen nativo (las secuencias de doblete objetivo están subrayadas).

Tabla 43

Codones originales	Utilización de codón (por mil)		Codones alterados	Utilización de codón
	C. beijerinckii	E. coli		
GCC	3.7	25.6	GCT	15.3
CTG	1.4	52.9	CTA	3.9
ACC	2.5	23.5	ACA	7.0
CGC	0.8	22.0	CGA	3.5
GCG	2.9	33.8	GCT	15.3
CCG	1.1	23.3	CCC	5.4

>SEQ ID NO: 893 Secuencia nativa

ATGGACGCACTGGGATATTACAACGGAAAATGGGGGCTCTGGACGAGATGACCGTGCCGATGAAC
GACAGGGGTTGTTTCTTTGGGGACGGAGTGACGACGCTACCATCGCCGCTAACGGAGTGATCTTTGCCCTG
GACGAGCACATTGACCGTTTTTAAACAGCGCAAAGCTCCTGGAAATAGAAATCGGTTTTACAAAAGAGGAA
TTAAAAAACTTTTTTGAATGCACTCCAAAGTGGATAAAGGGGTGTACATGGTTTATTGGCAGGCGACT
CGCGGAACAGGCCGTCGAAGCCATGTATTTCCGGCAGGTCCCTCAAATCTCTGGATTATGATTAAGCCCAAT
CACGTGACGATCTTTATAGAAAAATCAAGCTCATTACCATGGAAGATAACCGCTTCCTCCACTGCAACATC
AAGACCCCTTAACCTTATTCCTCAATGTCAATGCTCCAGCGGGCGCTGGGAAGCGGGCTGCCACGAGGCGGTC
TTTACCGGGGTGAAACAGTAACCGAGTGCGCCACAGCAATGTCCACATCATTAAAAACGGCAGGTTTATC
ACCCACCAGGCGGACAACCTAATCCTTCGGGGCATAGCCCGTAGCCATTTATTGCAAGCCTGTATCAGGCTG
AACATTCCATTTGACGAACGGGAATTTACCCTTTCCGAATTATTTGACGCGGATGAGATTCTTGTTGTCAGC
AGCGGCACACTCGGCCTTAGCGCCAATACAATTGATGGAAAAACGTGGGGGGAAAAAGCGCCGGAAC TGCTA
AAAAAAATTAGGGCGAAGTGTTGAGGGAATTTATCGAAGCGACAGGCTACACGCCTGAGTGGAGCACAGTA
TAG

5

Cebadores para doblete 1 mutante

CTAACGGAGTGATCTTTGCTCTAGACGAGCACATTGAC (SEQ ID NO: 1074)

GTCAATGTGCTCGTCTAGAGCAAAGATCACTCCGTTAG (SEQ ID NO: 1075)

Cebadores para doblete 2 mutante

10 CATGGAAGATACACGATTCCTCCACTGCAACATCAAGAC (SEQ ID NO: 1076)

GTCTTGATGTTGCAGTGGAGGAATCGTGTATCTTCCATG (SEQ ID NO: 1077)

Cebadores para doblete 3 mutante

ATTGCCTCCCAGCGGGCTCTAGAAGCGGGCTGCCACG (SEQ ID NO: 1078)

CGTGGCAGCCCCGCTTCTAGAGCCCCGCTGGGAGGCAAT (SEQ ID NO: 1079)

15 Cebadores para doblete 4 mutante

GGGGGGAAAAGCTCCCGAACTGCTAAAAAAATTCAGG (SEQ ID NO: 1080)

CCTGAATTTTTTTTAGCAGGTCGGGAGCTTTTCCCCC (SEQ ID NO: 1081)

Los clones con la secuencia correctamente alterada fueron transformados en el anfitrión BL21(DE3) para los ensayos de expresión de enzimas. La expresión de enzimas fue determinada cultivando las células durante la noche en Overnight Express II y sometiendo las células a lisis con el reactivo BugBuster seguido por análisis en SDS PAGE del extracto celular crudo y proteína soluble.

Parece ser que hay una mejora ligera en la expresión de proteína soluble con cambios en el codón a los dobles 1, 2 y 3. Los cambios en el codón en el doblete 4 no fueron beneficiosos para la expresión de la proteína soluble. Los cambios en codón para los dobles 1, 2 y 3 fueron combinados en pares usando el kit Stratagene Quick Change y los cebadores diseñados para los cambios de codón inicial. Los clones con la secuencia correctamente alterada fueron transformados en anfitriones BL21(DE3) para los ensayos de expresión enzimática. La expresión enzimática fue determinada cultivando las células durante la noche en Overnight Express II y sometiendo las células a lisis con el reactivo BugBuster seguido por el análisis por SDS PAGE del extracto celular crudo y la proteína soluble. Las combinaciones de las mutaciones a los dobles 1 y 2, 2 y 3 y 1 y 3 produjeron bandas de proteína soluble visibles sobre el gel de SDS-PAGE. Sin embargo, aún parece haber más proteína en las fracciones de proteína total.

Ejemplo 22 – La evaluación de la expresión periplasmática de un polipéptido de DAT

Debido que la expresión soluble del polipéptido de SEQ ID NO: 894 fue baja cuando el producto genético era expresado como una proteína citoplasmática en el anfitrión E. coli BL21(DE3) (véase Ejemplo 8), el gen fue clonado en vectores para generar proteínas de fusión que deberían ser exportadas en el espacio periplasmático. El periplasma provee condiciones que promueven el plegamiento y formación de puentes de disulfuro apropiados y puede potenciar la seguridad de actividad de ciertas proteína objetivo.

La clonación en EMD Biosciences/Novagen PET26b permite la producción de la proteína objetivo con una secuencia de señalización periplasmática. La secuencia de señalización es escindida por la peptidasa de señalización concomitante con la exportación. Los EMD Biosciences/Novagen PET39b y PET40b están diseñados para clonar y para la expresión de proteínas objetivo fusionadas con un aminoácido 208 DsbA-TagTM o un aminoácido 236 DsbC-TagTM. DsbA y DsbC son enzimas periplasmáticas que catalizan la formación e isomerización de puentes disulfuro, respectivamente. Las proteínas de fusión están localizadas típicamente en el periplasma.

Protocolo de clonación

El ácido nucleico de SEQ ID NO: 893 del plásmido SEQ ID NO: 893/pET30a (descrito en el Ejemplo 4) fue clonado en los vectores EMD Biosciences/Novagen pET26b (catálogo #69862-3), pET39b (catálogo #70090-3), y pET40b (catálogo #70091-3) en los sitios EcoRI y NotI de los vectores. El ácido nucleico de DAT con un sitio EcoRI en 5' y un sitio NotI en 3' fue generado utilizando el protocolo de amplificación descrito en el Ejemplo 4 y los siguientes cebadores:

5'-CGCAG**GAATTC**GACGCACTGGGATATTACAAC-3' (SEQ ID NO: 1082)

5'-GTTAG**GCGGCCG**CTATACTGTGCTCCACTCAG-3' (SEQ ID NO: 1083)

Los sitios de restricción están en itálicas en las secuencias de cebadores. El producto de ADN resultante y los vectores pET26b, pET39b, y pET40b fueron digeridos con EcoRI y NotI (New England Biolabs) siguiendo el protocolo sugerido por el fabricante. Las mezclas de reacción del vector fueron tratadas subsecuentemente con fosfatasa alcalina de camarón (Roche catálogo # 1758250). El vector digerido y las bandas del inserto fueron purificados en gel a partir de un gel de agarosa al 1% utilizando un Qiagen QIAquickR Gel Extraction Kit (catálogo # 28704) y ligados utilizando un Roche Rapid Ligation Kit (catálogo # 1635379) siguiendo el protocolo del fabricante. Las mezclas de ligazón fueron transformadas en células químicamente competentes Invitrogen OneShotR TOP10 (catálogo # C404003) por shock con calor a 42°C. Después de la recuperación en 500 µL de medio SOC durante 1 hora a 37°C, las mezclas de transformación fueron sembradas sobre placas LB que contenían 50 mg/mL de kanamicina incubadas a 37°C durante la noche. Las colonias fueron tomadas de las placas de transformación y utilizadas para inocular 5 mL de cultivos de LB que contenían 50 mg/mL de kanamicina que fueron incubadas durante la noche a 37°C. El ADN del plásmido fue purificado a partir de los cultivos de 5 mL utilizando un kit de minipreparación por rotación Qiagen QIApre (catálogo # 27104). Las secuencias de ácidos nucleicos fueron verificadas por secuenciación (Agencourt Biosciences Corp, Beverly, MA). Los plásmidos con las secuencias de inserto correctas fueron transformados en células químicamente competentes EMD Biosciences/Novagen BL21(DE3) (catálogo # 68450) por choque con calor como se describió más arriba.

Estudios de expresión

Matraces de Novagen Overnight ExpressTM AutoinductionSystem 2 (EMD Biosciences/Novagen catálogo # 71366) que contenían soluciones 1-6, y 50 mg/L de kanamicina (50 mL en cada matraz) fueron inoculados a partir de placas frescas de células BL21(DE3) que portaban el ácido nucleico de DAT de SEQ ID NO: 893 (que codifica el polipéptido

de SEQ ID NO: 894) en pET26b, pET39b o pET40b. Las células fueron incubadas a 30°C durante la noche y fueron recolectadas por centrifugación cuando la OD₆₀₀ alcanzó 6 o más. Las células fueron lavadas con regulador frío, fueron centrifugadas de nuevo, y usadas inmediatamente o las pellas celulares fueron congeladas a menos 80°C. Antes de recolectar, se extrajeron alícuotas de cultivo de 2 mL de cada matraz para los análisis de expresión de proteína soluble y total (soluble e insoluble). Los extractos celulares fueron preparados como se describe en el Ejemplo 16. Las muestras de proteína total fueron preparadas suspendiendo una cantidad pequeña de la pella celular en regulador de carga de proteína que contenía 2% de SDS, 10% de glicerol, 12.5 % de 2-mercaptoetanol, 0.1% de azul de bromofenol, y Tris-HCl 62.5 mM, pH 8, e incubando a 95°C durante 10 minutos.

Las fracciones celulares periplasmáticas fueron preparadas a partir del resto de las células de cada cultivo siguiendo el protocolo descrito en el EMD Biosciences/Novagen PET System Manual. Las fracciones resultantes fueron concentradas 30 veces utilizando los dispositivos de filtración centrífuga Amicon Ultracel 10K (Millipore catálogo # UFC901024). Las concentraciones de proteína total de los extractos celulares y las fracciones periplasmáticas fueron determinadas utilizando el kit de ensayo de proteína Pierce BCA (Pierce catálogo # 23225) con albúmina de suero bovino como estándar y un formato de placa de microtitulación. Muestras de 15 µg de proteína de los extractos celulares y muestras de 10 µg de proteína de las fracciones periplasmáticas fueron analizadas por SDS-PAGE utilizando geles de gradiente de poliacrilamida de 4-15% Bio-Rad Ready GelR Precast (Bio-Rad Laboratories catálogo # 161-1104). Además, las muestras de proteína total fueron analizadas por SDS-PAGE. Los estándares de rango bajo Bio-Rad SDS-PAGE (catálogo # 161-0304) fueron corridos como estándares en cada gel.

El análisis del gel de SDS-PAGE de proteína total muestra que las proteínas con los pesos moleculares predichos se expresaron utilizando el Overnight Express™ Autoinduction System 2. Sin embargo, el análisis del gel de SDS-PAGE cargado con las fracciones de extractos celulares o con las fracciones periplasmáticas sugiere que estas proteínas no se expresan de manera soluble ni fueron exportadas hacia el periplasma.

Ejemplo 23 – Producción de monatina a partir de indol-3-piruvato

La concentración máxima de monatina obtenida cuando D-triptófano y ácido pirúvico son la materia prima de partida en el ensayo de formación de monatina descrito en el Ejemplo 5 está limitada por la solubilidad del triptófano. Con el fin de explorar el potencial de utilizar una aldolasa y el polipéptido de DAT de SEQ ID NO: 220 (descrito en el Ejemplo 14) para alcanzar concentraciones de monatina más altas, se analizó la reacción que parte de indol-3-piruvato (I3P) y ácido pirúvico como materia prima. En este caso, fue necesario proveer un donante amino tal como D-alanina y también D-alanina y D-triptófano.

La prueba fue llevada a cabo utilizando el polipéptido de DAT de SEQ ID NO: 220 purificado (producción y purificación descritas en el Ejemplo 15) y una aldolasa (descrita en el Ejemplo 6). La reacción fue fijada como sigue (en un total de 1 mL): indol-3-piruvato (I3P) 200 mM; piruvato de sodio 200 mM; D-alanina 400 mM; EPPS 100 mM, pH 8.0; MgCl₂ 1 mM; PLP 0.05 mM; y fosfato de potasio 10 mM.

Ambas enzimas fueron agregadas en exceso para facilitar la conversión a monatina para minimizar la competencia de reacciones de degradación no enzimáticas. Las reacciones fueron incubadas a temperatura ambiente en una cámara anaeróbica de Coy Laboratory Products, Inc. para minimizar la exposición al oxígeno. Todos los componentes excepto las enzimas fueron mezclados entre sí y el pH fue ajustado a 8.0. Las reacciones fueron iniciadas mediante la adición de enzimas (0.04 mg/mL de aldolasa como extracto celular (asumiendo 20% de expresión) y 0.40 mg/mL del polipéptido de DAT SEQ ID NO: 220).

En algunas pruebas como se indica en la Tabla a continuación, se agregó también D-triptófano bien sea a 50 o 100 mM además de la D-alanina para actuar como un donante de amino y también para limitar la cantidad de indol-3-piruvato consumida en la formación de D-triptófano. La formación de monatina fue medida después de 18 horas utilizando la metodología de LC/MS/MS descrita en el Ejemplo 3, y los resultados se presentan en la Tabla 44 a continuación.

Tabla 44. Formación de monatina a partir de I3P (mM)

Concentraciones iniciales de reactivos (mM)	Concentración de monatina (mM)
200 I3P; 200 pir: 400 D-ala	44.9
200 I3P; 200 pir: 400 D-ala, 50 D-trp	47.8
200 I3P; 200 pir: 400 D-ala, 100 D-trp	61.0

Como se muestra más arriba, las enzimas aminotransferasa y aldolasa fueron muy activas a las concentraciones de reactivo más altas y se alcanzó una concentración de monatina mucho mayor.

5 A las 18 horas, utilizando indol-3-piruvato 200 mM, piruvato de sodio 200 mM y D-triptófano 100 mM, la concentración de monatina fue 61 mM. Se observó un pequeño descenso en la producción de monatina (47.8 mM) bajo las condiciones del ensayo, con la adición de D-triptófano 50 mM en vez de utilizar solamente D-alanina 400 mM.

Ejemplo 24 – Tabla de homología

La Tabla 45 muestra algunos de los polipéptidos de DAT más activos y los homólogos correspondientes más cercanos de las bases de datos publicadas o literatura.

10

Tabla 45

Polipéptido de DAT (SEQ ID NO:)	Acierto más cercano de la base de datos	% identidad de secuencia
896	Bacillus sp. YM-1	76
874	Serina glioxilato transaminasa de Acidiphilium cryptum JF-5 (NR Accession No: 148260372)	51
878	glutamato-1-semialdehído 2,1-aminomutasa putativa de Planctomyces maris DSM 8797 (NR Accession No. 149173540)	43
882	D-alanina transaminasa de Oceanobacter sp. RED65 (NR Accession No: 94500389)	57
910	DAT de B. macerans	91
176	D-amino acid aminotransferasa de Clostridium beijerincki (NCIMB 8052)	62
220	D-amino acid aminotransferasa de Clostridium beijerincki (NCIMB 8052)	62
156	Aminotransferasa clase-III (guiada) de Chloroflexus aggregans DSM 9485 (NR Accession No: 118045454)	46
214	D-amino acid aminotransferasa de Clostridium beijerincki (NCIMB 8052)	61
918	Aminotransferasa clase IV de Robiginitalea biformata HTCC2501 (NR Accession No: 88806011)	57
902	glutamato-1-semialdehído 2;1-aminomutasa putativa de Planctomyces maris DSM 8797 (misma proteína de más arriba)	46
884	D-alanina transaminasa de Thiobacillus denitrificans ATCC 25259 (NR Accession No: 74316285)	40
866	D-alanina aminotransferasa de Lactobacillus salivarius subsp. Salivarius UCC118	49
946	D-amino acid aminotransferasa de Clostridium beijerincki (NCIMB 8052)	63

Como se muestra en el Ejemplo 9, los homólogos de los polipéptidos que tienen la secuencia mostrada en SEQ ID NO: 866, 946, 220 y 176 fueron clonados y también tenían actividad en la producción de R,R monatina, a pesar de la identidad de secuencia entre estos polipéptidos de entre 49% y 63%. De la misma manera, la D-alanina transaminasa predicha de especies de *Oceanobacter* y la aminotransferasa de *Robinginitalea biformata* se espera que tengan una actividad de D-aminoácido transferasa amplia como los polipéptidos de DAT que tienen la secuencia de SEQ ID NO: 882 y 918.

Ejemplo 25 – Construcción y prueba de mutantes GSSMSM

Este ejemplo describe la construcción de ácidos nucleicos y polipéptidos de ejemplo, y describe su actividad enzimática. La secuencia de nucleótidos (SEQ ID NO: 219) fue subclonada en un vector pSE420-C-His y expresada en el anfitrión *E. coli* XL1-BLUE (Stratagene, La Jolla, CA) para producir la enzima D-aminotransferasa (DAT) de ejemplo que tiene la secuencia de aminoácidos mostradas en SEQ ID NO: 220. El vector pSE420-C-His fue creado agregando una etiqueta His en C terminal al vector pSE420 de Invitrogen (Carlsbad, CA). El constructo SEQ ID NO: 220 (en *E. coli* XL1-BLUE) fue utilizado como secuencia de partida en la cual se introdujeron modificaciones y se denomina aquí como secuencia tipo silvestre (WT). Se llevó a cabo una primera ronda de modificación (esto es mutaciones de residuos sencillos) utilizando la tecnología Gene Site Saturated MutagenesisSM (GSSMSM) (véase, por ejemplo, la Patente de los Estados Unidos No. 6,171,820). Los mutantes hechos utilizando la tecnología GSSMSM fueron expresados en el vector pSE420—His en anfitrión *E. coli* XL1-BLUE, dispuesto en placas de 384 pozos y cultivado a 37°C durante la noche. Las muestras fueron subcultivadas y se hicieron crecer a 30°C durante dos noches (36-48 horas). Los cultivos fueron congelados a -20°C hasta que pudieron prepararse los lisados de las células.

Las células fueron sometidas a lisis mediante la adición de 10 µL de BPER II (Thermo Scientific, Rockford, IL) a cada pozo. Las muestras fueron mezcladas tres veces y sometidas a lisis sobre hielo durante una hora. Los lisados crudos fueron probados entonces en la primera selección. 25 µL de lisado crudo fueron probados con R,R monatina 1 mM, sal de sodio de ácido pirúvico 25 mM, PLP 0.08 mM en fosfato de sodio 50 mM (pH 8) a temperatura ambiente. Después de tres horas, se tomó una alícuota y se agregó ácido fórmico hasta una concentración final de 2%. Las muestras fueron diluidas con agua para estar dentro del rango de la curva estándar. Las muestras fueron analizadas en cuanto al consumo de monatina y la formación de alanina usando los métodos de LC/MS/MS descritos en el Ejemplo 1 (método de LC/MS/MS para la detección de D-alanina o R,R monatina). El comportamiento de la muestra fue comparado con el comportamiento del control tipo silvestre (esto es, SEQ ID NO: 220).

Los mutantes que sobrepasaron al control tipo silvestre fueron seleccionados como aciertos a partir de la selección primaria de GSSMSM. Las reservas en glicerol de los aciertos primarios fueron utilizadas para inocular nuevas placas de 384 pozos. Los aciertos fueron dispuestos en cuadruplicado, cultivados y sometidos a lisis como se indica para la selección primaria. Los aciertos primarios fueron probados entonces en una selección secundaria. El método de la selección secundaria fue el mismo que el de la selección primaria excepto que los mutantes fueron probados con sustrato de R,R monatina 1 mM y 15 mM. Las muestras fueron analizadas en cuanto al consumo de monatina y formación de alanina utilizando LC/MS/MS. El comportamiento de la muestra fue comparado con el comportamiento del control tipo silvestre.

El comportamiento de la muestra fue evaluado utilizando el sistema de marcación basado en la producción de alanina y el consumo de monatina. La marcación máxima para una muestra individual fue seis. Se asignó un máximo de tres puntos para la producción de alanina y un máximo de tres puntos para el consumo de monatina. Los criterios de marcación fueron como sigue: 1 punto asignado para un valor entre promedio y una desviación estándar del control positivo; 2 puntos asignados para un valor entre una y dos desviaciones estándar del control positivo; y 3 puntos asignados para un valor más allá de dos desviaciones estándar del control positivo.

La marcación total potencial más alta para un mutante fue 48 (puesto que las muestras fueron seleccionadas en cuadruplicado a monatina 1mM y 15 mM). En general, los mutantes que calificaron con 20 puntos o más fueron seleccionados como aciertos secundarios. Sin embargo, se hicieron algunas excepciones para muestras con marcación menor de 20 puntos. Las muestras con valores de formación de alanina y consumo de monatina sobre el límite de los requerimientos de umbral también fueron seleccionadas como aciertos. Esto evita la eliminación prematura de aciertos y permite una prueba y caracterización adicional en la selección terciaria.

Las muestras identificadas como aciertos en la selección secundaria, utilizando los criterios anteriores, aparecen listadas en la Tabla 46. Los aciertos secundarios fueron sesgados a partir de reservas de glicerol sobre placas de agar LB que contenían carbenicilina 100 µg/mL y fueron cultivadas durante la noche a 37°C. Se utilizaron colonias individuales para inocular 1 mL de LB que contenía carbenicilina (100 µg/mL). Los cultivos se hicieron crecer durante la noche a 37°C. Se aisló el ADN de los cultivos, y luego se preparó y secuenció utilizando secuenciadores automáticos 3730XL (Applied Biosystems, Foster City, CA).

Las mutaciones y la posición de los aminoácidos de la mutación para los aciertos secundarios aparecen listados en la Tabla 46. La numeración de las posiciones de aminoácidos comienza con la metionina N terminal. Por ejemplo, la

primera mutación listada “Y6L” se refiere a cambiar la tiroxina en la posición de aminoácido 6 de la enzima tipo silvestre (SEQ ID NO: 220) a leucina. A nivel de ácidos nucleicos, cualquier codón que codifique el aminoácido deseado (mutado) puede ser usado.

5 Todas las secuencias de aminoácidos descritas en las Tablas 46, 47 y 53, a continuación, son polipéptidos de ejemplo; también se proveen secuencias de ácidos nucleicos que codifican tales polipéptidos.

Ejemplo 26 – Lista de mutaciones por GSSMSM

Tabla 46: Mutantes por GSSMSM identificados como aciertos de la selección secundaria

Nombre del mutante	Mutación	Nombre del mutante	Mutación
1	Y6L	92	D2G
2	Y6C; MUTACION SILENCIOSA AT AA31 (GGC → GGT)	93	D2Q
3	Y6F	94	D2F
4	Y6L	95	D2A
5	Y6H	96	D2T
6	Y6L	97	D2N
7	Y6M	98	D2R
8	N10S	99	D2I
9	N10W	100	D2V;G9A
10	N10T	101	G12A
11	N10R	102	D47W
12	N10T	103	S56S
13	L14V	104	I64H
14	L14L	105	L66L
15	G41G	106	I64C
16	T18W	107	L66G
17	N40N	108	E69Y
18	V19T	109	T74L
19	V42V	110	K73L
20	I62C	111	T74V
21	V82A	112	T74M
22	A57M	113	T74R
23	V42M	114	T74A
24	G41Y	115	N76C
25	A45L	116	E77R

ES 2 531 290 T3

Nombre del mutante	Mutación	Nombre del mutante	Mutación
26	V93Y	117	R156A
27	V93G	118	K72M
28	L46A	119	S205A
29	L46H	120	Q209S
30	G98A	121	V212E
31	P20S	122	R213W
32	V93A	123	I216T
33	V103T	124	P217H
34	P108F	125	P217V
35	V93L	126	D219F
36	S101S	127	E220V
37	A106G	128	R221E
38	S101Q	129	F223C
39	P108T	130	S226P
40	N118G	131	L228F
41	P108G	132	V234A
42	I120L	133	S238S
43	A106W	134	V236T
44	N118R	135	V236T
45	N110A; N118G	136	T241R
46	N118A	137	L242F
47	N118R	138	T241R
48	P117W; N118K	139	T241C
49	D133N	140	E248F
50	K126Q	141	D250E
51	K126R	142	K257V
52	K128S	143	G256K
53	I127M	144	E260G
54	T131T	145	L262R
55	D133L	146	K263M
56	M132A	147	D267G

ES 2 531 290 T3

Nombre del mutante	Mutación	Nombre del mutante	Mutación
57	D133E	148	D267R
58	L129V	149	I265L
59	K126K	150	E268S
60	I130M	151	L270S
61	M132Y	152	L270G
62	K128R	153	L270W
63	M132R	154	R271S
64	L129I	155	I274W
65	K128L; D2D (GAC → GAT)	156	G278S
66	F137W	157	Y279C
67	I152V	158	S284R
68	N55L	159	E282G
69	N150S	160	T280N
70	L138L	161	V286G
71	P149P	162	R285F
72	G161G	163	V286R
73	A165T	164	G240G
74	H163R	165	E61R
75	H163K	166	E61D
76	H168A	167	E61Y
77	E171S	168	G85G
78	E171R	169	G85D
79	E171R	170	S80R
80	T172I	171	Y79R
81	C176G	172	Y79V
82	A177S	173	W283V
83	A177S	174	W283E
84	S80L; R156W	175	W283A
85	H182G	176	W283S
86	N186S	177	W283G
87	K185R	178	W283A

Nombre del mutante	Mutación	Nombre del mutante	Mutación
88	K185T	179	W283R
89	D2H	180	W283T
90	D2T; E250G	181	P281W
91	D2Y	182*	V236T; T241R
*El mutante 182 fue creado utilizando mutagénesis dirigida al sitio, utilizando ADN mutante 136 como plantilla e introduciendo la mutación V236T. Una persona experimentada en la técnica puede sintetizar este gen utilizando técnicas de mutagénesis dirigida al sitio.			

5 Las muestras listadas en la Tabla 46 fueron preparadas entonces para selección terciaria. Las reservas en glicerol fueron utilizadas para inocular 5 mL de LB que contenía carbenicilina 100 µg/mL. Los cultivos fueron hechos crecer durante la noche a 37°C. Los cultivos nocturnos fueron usados entonces para inocular 50 mL de cultivos de LB que
10 contenían 100 µg/mL de carbenicilina en 250 matraces con deflexión hasta OD_{600nm} de 0.05. Se agregó IPTG hasta una concentración final de 1 mM cuando la OD_{600nm} alcanzó a 0.4-0.8. Los cultivos fueron inducidos durante la noche a 30°C. Se recolectaron las pellas celulares por centrifugación a 6,000 rpm durante 20 minutos. Las pellas celulares fueron congeladas a -20°C hasta que pudieron ser preparados los lisados celulares. Las células fueron sometidas a lisis con BPER II (Thermo Scientific, Rockford, IL) sobre hielo durante 1 hora. Los lisados clarificados fueron
15 preparados por centrifugación a 12,000 rpm durante 30 minutos.

La proteína fue cuantificada mediante el Bio-Rad Bradford Protein Assay (Bio-Rad, Hercules, CA) según las instrucciones del fabricante. El análisis por SDS-PAGE y la densitometría fueron utilizados para determinar la cantidad de D-aminotransferasa expresada. Las muestras fueron normalizadas una d-aminotransferasa expresada. D-aminotransferasa de 0.02 mg/mL fue probada en la selección terciaria. El método de selección terciaria fue el
20 mismo que el método de la selección secundaria excepto en que las muestras fueron tomadas a 0, 5, 10, 15, 30, 60, 120 y 210 minutos para desarrollar un transcurso en el tiempo. Los valores de producción de alanina y de consumo de monatina fueron medidos por análisis por LC/MS/MS y comparados con una curva estándar. Las muestras fueron comparadas con el control tipo silvestre.

20 Las muestras con títulos finales más altos o ratas iniciales más rápidas que el control tipo silvestre fueron identificadas como aciertos y se refieren a sobremutantes. Los sobremutantes por GSSMSM identificados en la selección terciaria están listados en la Tabla 47. Estos sobremutantes están descritos adicionalmente en el Ejemplo 27 a continuación.

Ejemplo 27 – Actividad enzimática de polipéptidos sobremutantes

25 Este ejemplo describe datos que demuestran la actividad enzimática de polipéptidos sobremutantes de ejemplo divulgados aquí, por ejemplo, los polipéptidos que tienen secuencias de aminoácidos descritas en la Tabla 47. La Tabla 47 muestra la actividad de los sobremutantes con respecto al control tipo silvestre en el punto de tiempo de 15 minutos en reacciones que utilizan sustrato de R,R-monatina 1 mM y 15 mM. La actividad relativa es la cantidad de alanina producida por la muestra dividida por la cantidad de alanina producida por el control tipo silvestre.

Tabla 47: Actividad de sobremutantes por GSSM en selección terciaria

Mutante	Mutación	Actividad relativa a control tipo silvestre (SEQ ID NO: 220)	
		Reacción con sustrato de monatina 1 mM	Reacción con sustrato de monatina 15 mM
23	V42M	1.28	1.04
24	G41Y	1.37	1.31
27	V93G	1.73	1.98
31	P20S	1.29	1.60

Mutante	Mutación	Actividad relativa a control tipo silvestre (SEQ ID NO: 220)	
		Reacción con sustrato de monatina 1 mM	Reacción con sustrato de monatina 15 mM
35	V93L	1.14	0.96
40	N118G	2.61	1.52
44	N118R	1.55	0.47
45	N110A; N118G	2.50	2.02
46	N118A	2.28	0.69
48	P117W; N118K	2.54	1.12
58	L129V	1.04	0.85
66	F137W	1.25	1.44
67	I152V	1.19	1.24
81	C176G	1.11	1.27
82	A177S	1.24	1.02
104	I64H	1.37	1.07
109	T74L	1.37	1.31
110	K73L	2.83	3.75
111	T74V	1.99	2.19
112	T74M	1.78	2.01
135	V236T	3.44	2.88
136	T241R	2.64	1.79
152	L270G	1.24	0.89
153	L270W	2.00	1.54
174	W283E	1.23	0.84
175	W283A	1.61	1.09
177	W283G	1.71	1.06
6	Y6L	2.52	2.21
88	K185T	1.04	0.95
107	L66G	1.08	1.02

Se identificaron varias muestras que sobrepasaban el rendimiento del control tipo silvestre bajo las condiciones probadas. Se identificaron sobremutantes potenciales K_m y V_{max} . Estos resultados indican que el control tipo silvestre (SEQ ID NO: 220) es evolucionable fácilmente hacia una actividad específica incrementada de D-aminotransferasa sobre la monatina.

5

Ejemplo 28 – Actividad de mutantes por GSSMSM en el proceso de monatina

Análisis de GSSM DATs en pSE420-C-His

Este ejemplo describe datos que demuestran la actividad enzimática de polipéptidos de ejemplo divulgados aquí. El mutante 27, Mutante 44, Mutante 58, Mutante 119, Mutante 135, Mutante 152, Mutante 154 y el control tipo silvestre (en el vector pSE420-C-His en E.coli XL1-Blue, como se describe en los Ejemplos 25 y 26) fueron esparcidos sobre placas de agar que contenían medio LB con ampicilina (100 µg/mL). Se utilizaron colonias individuales para inocular 5 mL de medio LB que contenía ampicilina (100 µg/mL). Se utilizaron 500 µL para inocular 50 mL del mismo medio en un matraz con deflexión de 250 mL. Las células fueron cultivadas a 30°C hasta aproximadamente un OD_{600nm} de 0.4. Se agregó IPTG hasta una concentración final de 1 mM. Las células fueron cultivadas a 30°C durante 4 horas y recolectadas por centrifugación. Las células fueron congeladas inmediatamente a -80°C hasta que se prepararon los extractos celulares.

Los extractos celulares fueron preparados como se describe en el Ejemplo 4. Las concentraciones de proteína fueron determinadas utilizando el ensayo de placa de microtitulación BCA (Pierce, Rockford, IL) con BSA (Pierce, Rockford, IL) como estándar, según las instrucciones del fabricante. Para estimar la concentración de la D-aminotransferasa en los extractos libres de células, se llevó a cabo el análisis por SDS-PAGE y se utilizó estimación visual para estimar el porcentaje de expresión. Las proteínas de DAT fueron solubles en el rango de 10-25 % de expresión como porcentaje de proteína total y este se utilizó para calcular la dosificación de los ensayos.

Se llevó a cabo un ensayo de formación de R,R monatina que contenía regulador EPPS 100 mM pH 7.8, MgCl₂ 1 mM, PLP 0.05 mM, piruvato de sodio 200 mM, fosfato de potasio 10 mM, Tween-80 al 0.01% con aldolasa 0.1 mg/mL y 0.2 mg/mL de DAT en una reacción de 4 mL a temperatura ambiente. El mutante 27 utilizó 0.15 mg/mL de enzima DAT en vez de 0.2 mg/mL. Después de 0.5, 1, 2, 4 y 23 horas, se tomó una alícuota, se agregó ácido fórmico hasta una concentración final de 2%, y las muestras fueron agitadas y filtradas. Las muestras fueron analizadas en cuanto al contenido de monatina utilizando la metodología de LC/MS/MS descrita en el Ejemplo 36. Los resultados se muestran en la Tabla 48.

Tabla 48: Actividad de DATs (clonados en pSE420-C-His)

Polipéptido de DAT	Monatina (mM) 0.5 horas	Monatina (mM) 1 hora	Monatina (mM) 2 horas	Monatina (mM) 4 horas
Control tipo silvestre	2.12	5.26	9.34	13.05
Mutante 27	4.74	9.55	14.72	18.06
Mutante 44	3.73	6.61	10.38	13.23
Mutante 58	3.61	7.51	11.85	14.56
Mutante 135	3.50	7.72	12.50	16.17
Mutante 136	1.40	4.63	6.59	--
Mutante 152	4.79	9.19	13.08	14.85
Mutante 154	3.76	7.66	11.85	14.38

Como puede verse a partir de los datos mostrados en la Tabla 48, un cierto número de mutantes de DAT obtenidos a través de la evolución por GSSMSM mostraron ratas iniciales mejoradas de formación de monatina con respecto al control tipo silvestre bajo las condiciones del ensayo.

Análisis de GSSMSM DATs en pMet1a

Este ejemplo describe datos que demuestran la actividad enzimática de polipéptidos de ejemplo divulgados aquí. Mutante 2, mutante 6, mutante 11, mutante 27, mutante 40, mutante 44, mutante 45, mutante 58, Mutante 110, Mutante 135, y Mutante 136 fueron recreados mediante mutagénesis dirigida al sitio utilizando QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) de acuerdo con las instrucciones del fabricante. Para generar los mutantes, el constructo etiquetado pMET1a descrito en el Ejemplo 16 (pMET1a: SEQ ID NO: 220 (WT)) fue utilizado como plantilla. Los cebadores mutagénicos utilizados se listan abajo en la Tabla 49. Los fragmentos de PCR fueron digeridos con Dpn1 (Invitrogen, Carlsbad, CA) durante 1 hora y transformados en células E. coli Top 10 (Invitrogen, Carlsbad, CA). Las preparaciones de plásmidos purificados resultantes fueron secuenciadas (Agencourt,

Beverly, MA) para verificar que se incorporaron las mutaciones correctas. Los plásmidos fueron transformados entonces en anfitriones de expresión E. coli B834(DE3) (Novagen, San Diego, CA).

Tabla 49. Cebadores para mutagénesis

Mutante producido	Cebadores para PCR	Plantilla
Mutante 2	5'-ATG GAC GCA CTG GGA TGT TAC AAC GGA AAT TGG -3' (SEQ ID NO:1084) 5'-CCA ATT TCC GTT GTA ACA TCC CAG TGC GTC CAT -3' (SEQ ID NO:1085)	SEQ ID NO: 220
Mutante 6	5'- ATG GAC GCA CTG GGA CTT TAC AAC GGA AAT TGG GGG-3' (SEQ ID NO:1086) 5'- CCC CCA ATT TCC GTT GTA AAG TCC CAG TGC GTC CAT-3' (SEQ ID NO:1087)	SEQ ID NO: 220
Mutante 27	5'- TAC CTG GTT TAT TGG CAG GGT ACT CGC GGA ACA GGC CGG-3' (SEQ ID NO:1088) 5'- CCG GCC TGT TCC GCG AGT ACC CTG CCA ATA AAC CAG GTA-3' (SEQ ID NO:1089)	SEQ ID NO: 220
Mutante 40	5'- CTC TGG ATT ATA ATT AAG CCC GGC CAC ATC GAC AAT CTT TAT AG -3' (SEQ ID NO:1090) 5'- CTA TAA AGA TTG TCG ATG TGG CCG GGC TTA ATT ATA ATC CAG AG -3' (SEQ ID NO:1091)	SEQ ID NO: 220
Mutante 44	5'- CTC TGG ATT ATA ATT AAG CCC AGG CAC ATC GAC AAT CTT TAT AG-3' (SEQ ID NO:1092) 5'- CTA TAA AGA TTG TCG ATG TGC CTG GGC TTA ATT ATA ATC CAG AG-3' (SEQ ID NO:1093)	SEQ ID NO: 220
Mutante 45	5'- GTA TTT CCG GCA GGC CCT TCA GCG CTC TGG ATT ATA ATT AAG CC -3' (SEQ ID NO:1094) 5'- GGC TTA ATT ATA ATC CAG AGC GCT GAA GGG CCT GCC GGA AAT AC -3' (SEQ ID NO:1095)	Mutante 40
Mutante 58	5'- CAA TCT TTA TAG AAA AAT CAA GGT TAT TAC CAT GGA TGA TAC CCG C 3' (SEQ ID NO:1096) 5'- GCG GGT ATG ATC CAT GGT AAT AAC CTT GAT TTT TCT ATA AAG ATT G -3' (SEQ ID NO:1097)	SEQ ID NO: 220
Mutante 110	5'- CTT AAC AAA AGA GGA ATT GAA ACT GAC TTT AAA TGA AAT GTA CTC C-3' (SEQ ID NO:1098) 5'- GGA GTA CAT TTC ATT TAA AGT CAG TTT CAA TTC CTC TTT TGT TAA G -3' (SEQ ID NO:1099)	SEQ ID NO: 220
Mutante 135	5'- TTC GAC GCG GAC GAG GTG CTT ACT TCC AGC AGC GGC ACA CTC G-3' (SEQ ID NO:1100) 5'- CGA GTG TGC CGC TGC TGG AAG TAA GCA CCT CGT CCG CGT CGA A-3' (SEQ ID NO:1101)	SEQ ID NO: 220
Mutante 136	5'- TGC TTG TGT CCA GCA GCG GCC GGC TCG GCC TTA GCG CCG-3' (SEQ ID NO:1102) 5'- CGG CGC TAA GGC CGA GCC GGC CGC TGC TGG ACA CAA GCA-3' (SEQ ID NO:1103)	SEQ ID NO: 220

5 El Mutante 2, Mutante 6, Mutante 27, Mutante 40, Mutante 45, Mutante 58, Mutante 110, Mutante 119, Mutante 131, Mutante 135, Mutante 136, Mutante 152, Mutante 154 fueron generados en el vector pMET1a y transformados en el anfitrión de expresión compatible de E. coli B834(DE3) (Novagen, San Diego, Ca) descrito en el Ejemplo 2. Los

cultivos nocturnos en medio LB que contenían carbenicilina (100 µg/mL) fueron diluidos 1:100 en 100 mL del mismo medio y cultivados en un matraz con deflexión de 500 mL. El cultivo fue hecho crecer a 30°C durante la noche hasta una OD_{600nm} de 10 en medio Overnight Express II (Solución 1-6, Novagen). Las muestras para proteína total fueron tomadas inmediatamente antes de la recolección. Las células fueron recolectadas por centrifugación y lavadas con 10 mL de regulador de fosfato de potasio pH 7.8. Las células fueron congeladas inmediatamente a -80°C hasta que se prepararon los extractos celulares. Hay que anotar, que además de la mutagénesis dirigida al sitio, una persona experimentada en la técnica puede sintetizar los genes que codifican estas D-aminotransferasas utilizando técnicas de PCR de mutagénesis multicambio tales como las descritas en el Ejemplo 25.

Los extractos celulares fueron preparados y desalinizados como se describe en el Ejemplo 4 utilizando fosfato de potasio 100 mM como regulador para eludir y equilibrar la columna PD10. Las concentraciones de proteína total y DAT fueron determinadas como se describe.

La transaminación de R,R monatina con piruvato como receptor de amino fue llevada a cabo como se describe en el Ejemplo 5 excepto que se utilizó R,R monatina 15 mM. Los análisis iniciales de los niveles de alanina, monatina y precursor de monatina identificados como Mutante 40, Mutante 135 y Mutante 136 como mutantes superiores dio como resultado los niveles más altos de producción de alanina como se muestra en la Tabla 50. El Mutante 136 de DAT pareció tener la actividad más alta para la conversión de R,R monatina a R-MP. Los números de producción de alanina (en mM) para los diversos puntos del tiempo se muestran en la Tabla 50.

Tabla 50: Formación de alanina (mM) a partir de reacciones de transaminación de R,R monatina a partir de DATs clonados en pMET1a

Polipéptido de DAT	Alanina (mM) 15 minutos	Alanina (mM) 30 minutos	Alanina (mM) 60 minutos	Alanina (mM) 120 minutos
Control tipo silvestre	3.08	5.47	8.19	10.07
Mutante 2	3.38	5.74	8.85	10.52
Mutante 6	3.51	5.97	8.99	10.81
Mutante 27	4.36	8.00	10.72	10.52
Mutante 40	7.89	10.37	11.79	12.50
Mutante 44	2.65	4.58	7.18	--
Mutante 58	3.90	6.95	9.93	10.52
Mutante 110	3.50	6.17	9.53	10.52
Mutante 135	5.35	8.64	10.82	10.91
Mutante 136	6.24	9.46	11.24	11.15
Mutante 152	4.26	7.12	9.83	10.32
Mutante 154	4.16	7.13	10.07	10.76
--: no determinado bajo las condiciones presentes				

Para establecer adicionalmente la actividad, se realiza un ensayo de formación de monatina como se describe en el Ejemplo 1 con una concentración de DAT de aproximadamente 0.2 mg/mL. Como control, se evaluó una concentración de 0.2 mg/mL de DAT tipo silvestre purificado. Después de 0.5, 1, 2 y 4 horas se tomó una alícuota y se agregó ácido fórmico hasta una concentración de 2% y las muestras fueron agitadas y filtradas. Las muestras fueron analizadas en cuanto al contenido de monatina utilizando la metodología de LC/MS/MS descrita aquí y en cuanto al contenido de triptófano y alanina utilizando la metodología de fluorescencia postcolumna LC/OPA descrita en el Ejemplo 36.

Tabla 51: Actividad de DATs en pMET1a

Polipéptido de DAT	Monatina(mM) 0.50 horas	Monatina (mM) 1,00 horas	Monatina (mM) 2,00 horas	Monatina (mM) 4,00 horas
Control tipo silvestre	3.96	7.83	9.70	11.18
Mutante 2	1.56	3.78	8.77	12.68
Mutante 27	4.70	9.70	n.d.	13.80
Mutante 44	3.03	5.61	8.50	12.28
Mutante 45	1.40	4.00	7.70	11.50
Mutante 58	3.83	7.23	11.33	14.12
Mutante 110	2.60	5.90	9.90	12.70
Mutante 119	4.12	7.87	11.37	13.50
Mutante 131	3.75	7.41	11.40	13.90
Mutante 135	6.39	10.65	13.49	13.15
Mutante 136	3.36	8.02	12.86	13.16
Mutante 154	3.00	6.06	10.67	13.17

Todos los DAT mostrados en la Tabla 51 produjeron monatina. Los mutantes de DAT Mutante 58, Mutante 135 y Mutante 136 tuvieron ratas iniciales más rápidas que el control tipo silvestre. El Mutante 136 fue más lento para la reacción uno (conversión de D-Trp a l3P) pero tuvo mejor producción global de monatina que el control tipo silvestre.

Para el punto de tiempo final, se tomó una alícuota adicional (sin la adición de ácido fórmico) para determinar la distribución estereoisomérica de la monatina producida utilizando la metodología de derivación con FDAA descrita en el Ejemplo 36. Para los mutantes seleccionados probados, hubo poco o ningún impacto sobre la estereopureza. En todos los casos, los mutantes produjeron por encima de 98.8% de R,R bajo las condiciones de ensayo probadas. Estos resultados se muestran en la Tabla 52.

Tabla 52: Estereopurezas de monatina producida por mutantes seleccionados a las 4 horas

Polipéptido de DAT	% SS	% RS	% RR	% SR
Control tipo silvestre (pEmt1a)	0.00	0.40	99.30	0.20
Mutante 6	0.00	0.40	99.50	0.10
Mutante 27	0.00	0.80	98.80	0.30
Mutante 40	0.00	0.20	99.80	0.00
Mutante 45	0.00	0.50	99.40	0.10
Mutante 110	0.10	0.40	99.30	0.10
Mutante 135	0.00	0.40	99.50	0.10
Mutante 136	0.02	1.00	99.00	0.03

Ejemplo 29 – Construcción y prueba de mutantes con ensamblaje combinacional de sitios múltiples a la medida (TMCASM)

Este ejemplo describe la construcción de ácidos nucleicos y polipéptidos de ejemplo, y describe su actividad enzimática. Se seleccionó un subconjunto de mutaciones GSSM utilizando la tecnología de ensamblaje combinatorial de sitio múltiple a la medida Tailored Multi-Site Combinatorial AssemblySM (TMCASM). Los diez ejecutores superiores de la evolución GSSM en cada una de las diez reacciones de monatina 1 o 15 mM fueron seleccionados para la evolución TMCASM. La secuencia tipo silvestre (SEQ ID NO: 220) fue encajada en un modelo de D-aminoácido transferasa 3DAA (figura 8). El modelo en la figura 8 es mostrado con piridoxil-5'-fosfato D-alanina, con los residuos numerados indicando aquellos sitios seleccionados para la evolución por TMCASM. La Tabla 53 también lista las mutaciones que fueron seleccionadas para inclusión en la biblioteca TMCA. La evolución TMCA fue llevada a cabo sobre tipo silvestre (SEQ ID NO: 220) y el Mutante 45 utilizando los métodos descritos en la Solicitud PCT No. PCT/US08/071771.

La evolución TMCA esta descrita en la Solicitud PCT número PCT/US08/071771 y comprende un método para producir una pluralidad de polinucleótidos de progenie que tienen diferentes combinaciones de diversas mutaciones en sitios múltiples. El método puede ser llevado a cabo, en parte, por una combinación de al menos una o más de las siguientes etapas:

Obtener información de secuencia de un polinucleótido (“primero” o “plantilla”). Por ejemplo, la primera secuencia o de platilla puede ser un tipo silvestre (por ejemplo, SEQ ID NO: 220) o secuencia mutada (por ejemplo, Mutante 45). La información de secuencia puede ser la del polinucleótido completo (por ejemplo, un gen o un marco de lectura abierto) o de regiones parciales de interés, tales como una secuencia que codifica un sitio para enlazamiento, especificidad de enlazamiento, catálisis o especificidad por sustrato.

Identificar tres o más mutaciones de interés a lo largo de la secuencia de polinucleótido primera o de plantilla. Por ejemplo, las mutaciones pueden estar en 3, 4, 5, 6, 8, 10, 12, 20 o más posiciones dentro de la secuencia primera o de plantilla. Las posiciones pueden ser predeterminadas con absoluta precisión o por el contexto de los residuos circundantes o la homología. Para la TMCA de polipéptidos de DAT, los cambios en el codón superior 10 que dan como resultado un comportamiento mejorado de la enzima fueron incluidos como mutaciones de interés. Las secuencias que flanquean las posiciones de mutación en cada lado pueden ser conocidas. Cada posición de mutación puede contener dos o más mutaciones, tales como para diferentes aminoácidos. Tales mutaciones pueden ser identificadas utilizando la tecnología de mutagénesis por saturación en el sitio del genSM (GSSMSM), tal como se describe aquí en las Patentes de los Estados Unidos Nos. 6,171,820; 6,562,594 y 6,764,835.

Proveer cebadores (por ejemplo, oligonucleótidos sintéticos) que comprenden las mutaciones de interés. En una realización, se provee un cebador para cada mutación de interés. Así, un polinucleótido primero o de plantilla que tiene 3 mutaciones de interés puede utilizar 3 cebadores en esa posición. El cebador también puede ser provisto como un conjunto de cebadores que contiene una posición degenerada de tal manera que la mutación de interés es el rango de cualquier nucleótido o aminoácido de origen natural, o un subconjunto de ese rango. Por ejemplo, un conjunto de cebadores puede ser provisto de manera que favorezca las mutaciones para residuos de aminoácidos alifáticos.

Los cebadores pueden ser preparados como cebadores de avance o reversa, o los cebadores pueden ser preparados como al menos un cebador de avance y al menos un cebador de reversa. Cuando las mutaciones están posicionadas cercanamente juntas, puede ser conveniente utilizar cebadores que contengan mutaciones para más de una posición o diferentes combinaciones de mutación en posiciones múltiples.

Proveer un polinucleótido que contiene la secuencia plantilla. El polinucleótido primero o de plantilla puede ser circular, o puede estar superenrollado, tal como un plásmido o un vector para clonación, secuenciación o expresión. El polinucleótido puede ser de cadena sencilla (“ADNss”), o puede ser de cadena doble (“ADNds”). Por ejemplo, el método TCMA somete la plantilla de ADNds (“sc”) a una etapa de calentamiento a 95°C durante 1 minuto (véase Levy, Nucleic Acid Res., 28(12):e57(i-vii) (2000)).

Agregar los cebadores al polinucleótido de plantilla en una mezcla de reacción. Los cebadores y el polinucleótido de plantilla son combinados bajo condiciones que permiten que los cebadores se fusionen al polinucleótido de plantilla. En una realización del protocolo de TMCA los cebadores son agregados al polinucleótido en una mezcla de reacción sencilla, pero puede ser agregado en reacciones múltiples.

Ejecutar reacciones de extensión de polimerasa. Los productos de extensión (por ejemplo, como una “progenie” o “polinucleótido extendido modificado”) pueden ser amplificados por medios convencionales. Los productos pueden ser analizados en cuanto a longitud, secuencia, propiedades de ácidos nucleicos deseados, o expresados como polipéptidos. Otros métodos de análisis incluyen hibridación in situ, selección de secuencia o selección de expresión. El análisis puede incluir una o más rondas de selección y selección de una propiedad deseada.

Los productos también pueden ser transformados en una célula u otro sistema de expresión, tal como un sistema libre de células. El sistema libre de células puede contener enzimas relacionadas con la replicación, reparación,

recombinación, transcripción o traducción de ADN. Anfitriones de ejemplo incluyen células y líneas celulares bacterianas, de levadura, de plantas y de animales, e incluyen *E. coli*, *Pseudomona fluorescens*, *Pichia pastoris* y *Aspergillus niger*. Por ejemplo, las cepas XL1-Blue o Stb12 de *E. coli* pueden ser utilizadas como anfitrionas.

5 El método de la invención puede ser utilizado con los mismos o diferentes cebadores bajo diferentes condiciones de reacción para promover productos que tienen diferentes combinaciones o números de mutaciones.

Llevando a cabo el método de ejemplo descrito más arriba, este protocolo también provee uno o más polinucleótidos producidos por este método de evolución TMCA, los cuales pueden ser entonces escogidos o seleccionados en cuanto a una propiedad deseada. Uno o más de los polinucleótidos de progenie pueden ser expresados como polipéptidos, y escogidos o seleccionados opcionalmente en cuanto a una propiedad deseada. Así, esta realización de protocolo de evolución TMCA provee polinucleótidos y los polipéptidos codificados, así como bibliotecas de tales polinucleótidos que codifican tales polipéptidos. Esta realización del protocolo de evolución TMCA provee adicionalmente la selección de bibliotecas escogiendo o seleccionando la biblioteca para obtener uno o más polinucleótidos que codifican uno o más polipéptidos que tiene la actividad deseada.

15 Otra realización del protocolo de evolución TMCA descrita en PCT/US08/071771 comprende un método para producir una pluralidad de polinucleótidos modificados. Tales métodos incluyen en general (a) agregar al menos tres cebadores a un polinucleótido de plantilla de cadena doble en una mezcla de reacción sencilla, en donde los al menos tres cebadores no se superponen, y en donde cada uno de los al menos tres cebadores comprende al menos una mutación diferente de los otros cebadores, en donde al menos un cebador es un cebador de avance que puede fusionarse a una cadena menos de la plantilla y al menos un cebador es un cebador de reversa que puede fusionarse a una cadena más de la plantilla, y (b) someter la mezcla de reacción a una reacción de extensión con polimerasa para producir una pluralidad de polinucleótidos modificados extendidos a partir de los al menos tres cebadores.

25 Otra realización del protocolo de evolución TMCA descrito en PCT/US08/071771 comprende un método en donde una célula es transformada con la pluralidad de productos extendidos que no han sido tratados con una ligasa. En otra realización de la invención, la pluralidad de polinucleótidos modificados extendidos es recuperada a partir de la célula. En otra realización, la pluralidad recuperada de polinucleótidos modificados extendidos es analizada, por ejemplo, expresando al menos uno de la pluralidad de polinucleótidos modificados extendidos y analizando el polipéptido expresado a partir del mismo. En otra realización, la pluralidad de polinucleótidos modificados extendidos que comprenden las mutaciones de interés es seleccionada.

30 En otra realización del protocolo de evolución TMCA, se obtiene la información de secuencia concerniente al polinucleótido de plantilla, y tres o más mutaciones de interés a lo largo del polinucleótido de plantilla pueden ser identificadas. En otra realización, los productos obtenidos por la extensión con polimerasa pueden ser analizados antes de transformar la pluralidad de productos modificados extendidos en una célula.

35 En una realización del protocolo de evolución TMCA, los productos obtenidos por la extensión de polimerasa son tratados con una enzima, por ejemplo, una enzima de restricción tal como una enzima de restricción de DpnI, destruyendo por lo tanto la secuencia de polinucleótido de plantilla. Los productos tratados pueden ser transformados en una célula, una célula de *E. coli*.

40 En una realización de protocolo de evolución TMCA, pueden usarse al menos dos, or al menos tres, o al menos cuatro, o al menos cinco, o al menos seis, o al menos siete, o al menos ocho, o al menos nueve, o al menos diez, o al menos once, o al menos doce, o más cebadores. En una realización cada cebador comprende una mutación de punto individual. En otra realización, dos cebadores de avance o dos cebadores de reversa comprenden un cambio diferente en la misma posición sobre el polinucleótido de plantilla. En otra realización, el al menos un cebador comprende al menos dos cambios en diferentes posiciones sobre el polinucleótido de plantilla. En aún otra realización, al menos un cebador comprende al menos dos cambios en diferentes posiciones y al menos dos cebadores de avance o dos de reversa comprenden un cambio diferente en la misma posición sobre el polinucleótido de plantilla.

45 En una realización del protocolo de evolución TMCA, los cebadores de avance están agrupados en un grupo de avance y los cebadores de reversa están agrupados en un grupo de reversa, y los cebadores en el grupo de avance y los cebadores en el grupo de reversa, independientemente uno de otro, son normalizados para tener una concentración igual en el grupo correspondiente independientemente de las posiciones sobre el polinucleótido de plantilla, y en donde después de la normalización una cantidad igual de los cebadores de avance y reversa se agrega a la reacción. Este método de normalización, puede omitirse una combinación de algunas posiciones. La omisión puede deberse, por ejemplo, a una concentración relativamente baja del cebador en una posición que contiene un cebador individual en comparación con una posición que contiene cebadores múltiples. "Omisión de posición" se refiere a polinucleótidos resultantes que muestran una fuerte preferencia por la incorporación de cebadores en una posición individual con respecto a las otras posiciones dentro de su grupo de cebadores de avance o reversa. Esto da como resultado una combinación de polinucleótidos modificados que puede tener un alto

porcentaje de mutaciones dentro de una posición de cebador individual pero un bajo porcentaje de mutaciones en otra posición dentro de su grupo de cebadores de avance o reversa. Esta omisión es desfavorable cuando la meta del TMCA es general por polinucleótidos de progenie que comprenden todas las posibles combinaciones de cambios a la plantilla. La omisión puede ser corregida, por ejemplo, normalizando los cebadores como un conjunto en cada posición para que sean iguales.

En una realización del protocolo de evolución TMCA, la normalización del cebador se lleva a cabo organizando los cebadores en múltiples grupos dependiendo de su localización sobre el polinucleótido de plantilla, en donde los cebadores que cubren la misma región seleccionada sobre la plantilla están en un grupo; normalizando los cebadores agrupados dentro de cada grupo para ser una concentración igual; reuniendo los cebadores de avance dentro de un grupo en un grupo de avance y normalizando la concentración entre cada grupo de los cebadores de avance para que sean iguales; reuniendo los cebadores de reversa dentro de un grupo en un grupo de reversa y normalización de la concentración entre cada grupo de los cebadores de reversa para que sean iguales; y agregar una cantidad igual de los cebadores reunidos de avance y reversa en la reacción. No se ha observado omisión para las combinaciones de posición.

En una realización del protocolo de evolución TMCA, se provee un conjunto de cebadores degenerados comprendiendo cada uno una posición degenerada, en donde la mutación de interés es un rango de nucleótidos diferentes en la posición degenerada. En otra realización, se provee un conjunto de cebadores degenerados que comprenden al menos un codón degenerado que corresponde a al menos un codón del polinucleótido de plantilla y al menos una secuencia adyacente que es homóloga a una secuencia adyacente al codón de la secuencia de polinucleótidos de plantilla. En otra realización, el codón degenerado es N, N, N y codifica cualquiera de los 20 aminoácidos de origen natural. En otra realización, el codón degenerado codifica menos de 20 aminoácidos de origen natural.

Otra realización del protocolo de evolución TMCA descrita en PCT/US08/071771 comprende un método para producir una pluralidad de polinucleótidos modificados que comprende las mutaciones de interés. Tales métodos incluyen en general (a) agregar al menos dos cebadores un polinucleótido de plantilla de cadena doble en una mezcla de reacción sencilla, en donde los al menos dos primeros cebadores no se superponen, y en donde cada uno de los al menos dos cebadores comprende al menos una mutación diferente del otro cebador, en donde al menos un cebador es un cebador de avance que puede fusionarse a una cadena menos de la plantilla y al menos un cebador es un cebador de reversa que puede fusionarse en una cadena más de la plantilla; (b) someter la mezcla de reacción a una reacción de extensión con polimerasa para producir una pluralidad de polinucleótidos modificados extendidos a partir de al menos dos cebadores, (c) tratar la pluralidad de polinucleótidos modificados extendidos con una enzima, destruyendo por lo tanto el polinucleótido de plantilla, (d) transformar los polinucleótidos modificados extendidos tratados que no han sido tratados con una ligasa en una célula, (e) recuperar la pluralidad del polinucleótidos modificados extendidos a partir de la célula, y (f) seleccionar la pluralidad de polinucleótidos modificados extendidos que comprenden las mutaciones de interés.

Tabla 53: Lista de sitios para evolución TMCA

Mutación	Nuevo codón
P20S	AGT
K73L	TTG
T74V	GTG
V93G	GGT
N110A	GCT
P117W	TGG
N118G	GGG
N118A	GCG
V236T	ACT
T241R	CGG
L270W	TGG

5 Los mutantes TMCA fueron cultivados, dispuestos, probados y secuenciados utilizando el mismo método tal como se describió para la evolución GSSM en el Ejemplo 25. El rendimiento de la muestra fue comparado con el rendimiento del candidato superior de evolución GSSM – Mutante 135 – utilizando el mismo sistema de marcación como se describió en el Ejemplo 25. La Tabla 55 lista los aciertos de marcación secundaria TMCA con secuencias de ADN únicas (los mutantes TMCA están designados con caracteres alfabéticos para distinguirlos de los mutantes GSSM, los cuales están designados numéricamente).

Tabla 55: Mutantes TMCA identificados como aciertos de selección secundaria

Nombre de mutante	Mutación	Nombre de mutante	Mutación
A	P20S-N118G	EE	P20S-N110A-N118G
B	T74V-V93G-L270W	FF	N110A-N118G-T241R-L270W
C	P20S-T74V-L270W	GG	P20S-T74V-V93G-N110A-N118G-V236T-L270W
D	T74V-L270W	HH	V93G-N110A-N118G-V236T
E	P20S-K73L-T241R-L270W	II	P20S-T74V-N110A-N118G
F	K73L-V93G-V236T-T241R	JJ	N110A-N118G
G	P20S-T74V-V236T	KK	P20S-V93G-N110A-N118G-T241R
H	P20S-K73L-V93G	LL	N110A-N118A-L270W
I	K73L-V236T	MM	P20S-N110A-N118G-L270W
J	P20S-L270W	NN	N110A-N118A-V236T-T241R
K	2N-P20S-K73L-V93G-N118G	OO	N110A-N118G-L270W
L	P20S-T74V-N118A	PP	V93G-N110A-N18G-T241R
M	P20S-V236T	QQ	P20S-V93G-N110A-N18G
N	P20S-T241R-L270W	RR	V93G-N110A
O	P20S-T241R	SS	P20S-N110A-N118G-V236T
P	T74V-V93G-V236T-T241R	TT	T74V-N110A-N118A-V236T
Q	P20S-K73L-T74V-L270W	UU	P20S-K73L-T74V-N110AN118G-V236T-T241R
R	P20S-V93G-V236T	VV	86E (SILENT GAG → □GAA)-N110A-N118A-V236T
S	P20S-K73L-T74V-T241R-L270W	WW	T74V-N118G
T	P20S-K73L-L270W	XX	P20S-T241R-L270W-277T (SILENT ACA → □ACG)
U	T74V-V93G-N118G-V236T-T241R	YY	T74V-N118A-L270W
V	P20S-K73L-210A (SILENT-GCC → GCT)-V236T	ZZ	P20S-K73L-N118A-L270W
W	N118A-L270W	AAA	P20S-V93G-T241R

Nombre de mutante	Mutación	Nombre de mutante	Mutación
X	P20S-58K (SILENT AAG → AAA)-L270W	BBB	T74V-V93G-N110A-T241R
Y	P20S-V93G-N118G	CCC	V93G-N110A-N118A
Z	P20S-V236T-T241R	DDD	P20S-T74V-V93G-N110A-N118G-T241R
AA	P20S-P117W-N118A-V236T-L270W	EEE	T74V-N110A-N118G-L270W
BB	V93G-V236T	FFF	P20S-231A (SILENT GCG → GCA)-V236T
CC	V236T-L270W	GGG	V93G-V236T-T241R
DD	P20S-N118G-L270W		

Las muestras identificadas en la Tabla 55 fueron cultivadas, normalizadas y probadas en la selección terciaria utilizando el mismo método que se describió para la evolución GSSM en el Ejemplo 26. Los valores de monatina y alanina fueron determinados por LC/MS/MS y comparados con una curva estándar. El rendimiento de la muestra fue comparado con la actividad del Mutante 135 (el de mejor rendimiento de la evolución GSSM). Los sobremutantes TMCA identificados en la selección terciaria están listados en la Tabla 56.

Ejemplo 31 – Actividad de aciertos TMCA

Este ejemplo describe los datos que demuestran la actividad enzimática de los polipéptidos de ejemplo. La Tabla 56 a continuación muestra la actividad de los sobremutantes con respecto al Mutante 135 en el punto de tiempo a 15 minutos en reacciones utilizando sustrato de R,R- monatina 1 mM y 15 mM. La actividad relativa es la cantidad de alanina producida por la muestra dividida por la cantidad de alanina producida por el Mutante 135.

Tabla 56: Actividad de sobremutantes TMCA en selección terciaria

Mutante	Mutación	Actividad relativa a Mutante 135 por GSSM	
		Reacciones con sustrato de monatina 1 mM	Reacciones con sustrato de monatina 15 mM
C	P20S-T74V-L270W	1.02	0.93
E	P20S-K73L-T241R-L270W	1.32	1.31
F	K73L-V93G-V236T-T241R	1.29	0.64
G	P20S-T74V-V236T	1.28	1.30
I	K73L-V236T	1.24	1.29
J	P20S-L270W	0.79	1.01
L	P20S-T74V-N118A	1.62	0.83
M	P20S-V236T	1.27	1.46
O	P20S-T241R	1.33	1.71
R	P20S-V93G-V236T	1.22	1.02

Mutante	Mutación	Actividad relativa a Mutante 135 por GSSM	
		Reacciones con sustrato de monatina 1 mM	Reacciones con sustrato de monatina 15 mM
S	P20S-K73L-T74V-T241R-L270W	1.16	1.18
V	P20S-K73L-210A (SILENT-GCC → GCT)-V236T	1.03	1.00
Z	P20S-V236T-T241R	1.02	0.89
BB	V93G-V236T	1.55	1.98
CC	V236T-L270W	1.24	1.40
DD	P20S-N118G-L270W	1.54	1.78
PP	V93G-N110A-N118G-T241R	1.40	1.53
TT	T74V-N110A-N118A-V236T	1.10	0.42
VV	86E (SILENT GAG → GAA)-N110AN118A-V236T	1.31	0.52
WW	T74V-N118G	1.23	1.49
YY	T74V-N118A-L270W	1.97	1.30
ZZ	P20S-K73L-N118A-L270W	1.01	0.44
AAA	P20S-V93G-T241R	1.86	3.49
CCC	V93G-N110A-N118A	1.26	0.56

Se identificaron varias muestras que superaron al Mutante 135 bajo las condiciones probadas. Se identificaron los sobremutantes de potencial Km y Vmax. Los resultados de las evoluciones GSSM y TMCA indican que la SEQ ID NO. 220 tipo silvestre es evolucionable posteriormente para actividad específica incrementada sobre la monatina.

5 Ejemplo 32 – Evaluación de los DATs de mutante TMCA en pMET1a

Este ejemplo describe datos que demuestran la actividad enzimática de polipéptidos de ejemplo divulgados aquí. El Mutante E, Mutante G, Mutante I, Mutante M, Mutante O, Mutante P, Mutante BB, Mutante PP, Mutante WW, y Mutante AAA (DATs creado utilizando tecnología TMCA, véanse Ejemplos 29 y 30) fueron recreados por mutagénesis dirigida al sitio utilizando el QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) de acuerdo con las instrucciones del fabricante. Para generar los mutantes, los constructos etiquetados de pMET1a descritos en el Ejemplo 16 y Ejemplo 28 fueron utilizados como plantillas. Los cebadores mutagénicos utilizados están listados más adelante en la Tabla 57. Los fragmentos de PCR fueron digeridos con Dpn1 (Invitrogen, Carlsbad, CA) durante 1 hora y transformados en células E. coli XL10-Gold (Stratagene, La Jolla, CA). Las preparaciones de plásmidos purificados resultantes fueron secuenciadas (Agencourt, Beverly, MA) para verificar que las mutaciones correctas fueron incorporadas. Los plásmidos fueron luego transformados en el anfitrión de expresión E. coli B834(DE3) (Novagen, San Diego, CA).

Tabla 57. Cebadores para mutantes en vector pMET1a

Polipéptido mutante producido por TMCA	Cebadores PCR	Plantilla
Mutante E	5'- CTG GAC GAG ATG ACT GTG AGT ATG AAC GAC AGG GGC TGC TAC -3' (SEQ ID NO:1104) 5'- TGC TTG TGT CCA GCA GCG GCC GGC TCG GCC TTA GCG CCG-3' (SEQ ID NO:1105) 5'-CTA AAA AAA ATC CAG GAT GAA GTG TGG AGG GAA TTT ATC GAA GCG ACA GG3' (SEQ ID NO:1106)	Mutante 110
Mutante G	5'- CAA AAG AGG AAT TGA AAA AAG TGT TAA ATG AAA TGT ACT CC -3' (SEQ ID NO:1107) 5'- GGA GTA CAT TTC ATT TAA CAC TTT TTT CAA TTC CTC TTT TG -3' (SEQ ID NO:1108)	Mutante M
Mutante I	5'- CTT AAC AAA AGA GGA ATT GAA ACT GAC TTT AAA TGA AAT GTA CTC C-3' (SEQ ID NO:1109) 5'- GGA GTA CAT TTC ATT TAA AGT CAG TTT CAA TTC CTC TTT TGT TAA G -3' (SEQ ID NO:1110)	Mutante 135
Mutante M	5'- CTG GAC GAG ATG ACT GTG AGT ATG AAC GAC AGG GGC TGC TAC -3' (SEQ ID NO:1111) 5'- GTA GCA GCC CCT GTC GTT CAT ACT CAC AGT CAT CTC GTC CAG -3' (SEQ ID NO:1112)	Mutante 135
Mutante O	5'- CTG GAC GAG ATG ACT GTG AGT ATG AAC GAC AGG GGC TGC TAC -3' (SEQ ID NO:1113) 5'- GTA GCA GCC CCT GTC GTT CAT ACT CAC AGT CAT CTC GTC CAG -3' (SEQ ID NO:1114)	Mutante 136
Mutante P	5'- CAA AAG AGG AAT TGA AAA AAG TGT TAA ATG AAA TGT ACT CC -3' (SEQ ID NO:1115) 5'- TTC GAC GCG GAC GAG GTG CTT ACT TCC AGC AGC GGC ACA CTC G-3' (SEQ ID NO:1116) 5'- TGC TTG TGT CCA GCA GCG GCC GGC TCG GCC TTA GCG CCG-3' (SEQ ID NO:1117)	Mutante 27
Mutante BB	5'- TAC CTG GTT TAT TGG CAG GGT ACT CGC GGA ACA GGC CGG-3' (SEQ ID NO:1118) 5'- CCG GCC TGT TCC GCG AGT ACC CTG CCA ATA AAC CAG GTA-3' (SEQ ID NO:1119)	Mutante 135
Mutante PP	5'- TAC CTG GTT TAT TGG CAG GGT ACT CGC GGA ACA GGC CGG-3' (SEQ ID NO:1120) 5'- TGC TTG TGT CCA GCA GCG GCC GGC TCG GCC TTA GCG CCG-3' (SEQ ID NO:1121)	Mutante 45
Mutante WW	5'- CAA AAG AGG AAT TGA AAA AAG TGT TAA ATG AAA TGT ACT CC -3' (SEQ ID NO:1122) 5'- GGA GTA CAT TTC ATT TAA CAC TTT TTT CAA TTC CTC TTT TG -3' (SEQ ID NO:1123)	Mutante M
Mutante AAA	5'- CTG GAC GAG ATG ACT GTG AGT ATG AAC GAC AGG GGC TGC TAC -3' (SEQ ID NO:1124) 5'- TGC TTG TGT CCA GCA GCG GCC GGC TCG GCC TTA GCG CCG-3' (SEQ ID NO:1125)	Mutante 27

Los cultivos de *E. coli* B834(DE3) (Novagen, San Diego, CA) que expresan las proteínas etiquetadas con His en el terminal carboxi de Mutante 110, Mutante 135, Mutante 136, Mutante E, Mutante G, Mutante I, Mutante M, Mutante O, Mutante P, Mutante BB, Mutante PP, Mutante WW, Mutante AAA y tipo silvestre (SEQ ID NO: 220) fueron cultivados en 200 mL de medio Overnight Express II (Solución 1-6, Novagen) en un matraz con deflexión de 500 mL durante la noche a 30°C hasta una OD₆₀₀ de 10. Las muestras para proteína total fueron tomadas inmediatamente antes de la recolección. Las células fueron recolectadas por centrifugación y congeladas inmediatamente a -80°C hasta que los extractos celulares fueron preparados como se describe en el Ejemplo 4.

Los extractos celulares fueron creados mediante la adición de 50 mL de BugBuster Primary Amine Free (Novagen, San Diego, CA) con 50 µl de Benzonase Nuclease (Novagen, San Diego, CA), 0.75 µl de r-Lisozima (Novagen, San Diego, CA), y 250 µl de Protease Inhibitor Cocktail II (Calbiochem, San Diego, CA). Las células fueron incubadas durante 15 minutos a temperatura ambiente con agitación suave. Los extractos fueron centrifugados a 45,000 x g durante 10 minutos.

Las proteínas etiquetadas con His fueron purificadas según se describe en el Ejemplo 4 utilizando la resina GE Healthcare (Piscataway, NJ) Chelating Sepharose™ Fast Flow. La excepción fue el Mutante 182, el cual fue analizado como CFE tal como se describe en el Ejemplo 4. La proteína purificada fue desalinizada utilizando una columna PD10 fosfato de potasio 100 mM, pH 7.8 con PLP 0.050 mM. Las concentraciones de proteína total y DAT fueron determinadas como se describe en el Ejemplo 4.

Se realizó un ensayo de formación de monatina de 3 etapas como se describe en el Ejemplo 5 con una concentración de DAT de aproximadamente 0.2 mg/mL y la aldolasa a una concentración de 0.1 mg/mL. Como control, se evaluó una concentración de 0.2 mg/mL de DAT tipo silvestre purificada (SEQ ID NO: 220). Después de 0.5, 1, 2, 4 y 24 horas, se tomó una alícuota, se agregó ácido fórmico hasta una concentración final de 2% y las muestras fueron agitadas y filtradas. Las muestras fueron analizadas en cuanto al contenido de monatina utilizando la metodología LC/MS/MS y en cuanto a triptófano y alanina utilizando la metodología de fluorescencia postcolumna LC/OPA descrita en el Ejemplo 36. En el punto de tiempo final, se tomó una alícuota adicional (sin ajuste del pH) para determinar el porcentaje de R,R monatina por el método de derivación con FDAA descrito en el Ejemplo 36. La cantidad de monatina (mM) producida en diversos puntos del tiempo puede encontrarse en la Tabla 58. La estereopureza también fue determinada y el porcentaje del estereoisómero R,R puede encontrarse en la columna del extremo derecho. El estereoisómero R,S constituyó la mayoría del resto.

Tabla 58: Actividad de mutantes de DAT seleccionados

Polipéptidos de DAT	Monatina (mM) 0.25 horas	Monatina (mM) 0.5 horas	Monatina (mM) 1 hora	Monatina (mM) 4 horas	% RR
Control tipo silvestre (SEQ ID NO: 220)	1.60 (±0.42)	2.95 (±0.64)	5.00 (±0.85)	11.40 (±0.42)	99.50 (±0.08)
Mutante 110	1.70 (±0.00)	3.20 (±0.85)	5.75 (±0.21)	12.60 (±0.14)	99.48 (±0.32)
Mutante 135	3.65 (±0.35)	6.17 (±0.65)	10.33 (±0.32)	13.20 (±0.56)	99.42 (±0.11)
Mutante 136	2.60	5.00	8.10	12.90	98.98
Mutante 182	-	3.20	6.80	14.30	99.50
Mutante E	1.80	3.80	8.60	18.60	99.45
Mutante G	3.10	6.50	9.90	12.90	99.05
Mutante I	2.90	5.30	8.50	12.90	99.46
Mutante M	4.20	8.10	11.20	13.80	98.96
Mutante O	2.60	5.70	9.50	14.00	98.59

Polipéptidos de DAT	Monatina (mM) 0.25 horas	Monatina (mM) 0.5 horas	Monatina (mM) 1 hora	Monatina (mM) 4 horas	% RR
Mutante BB	4.20	8.20	11.40	13.70	98.97
Mutante PP	2.40	3.20	6.20	17.40	97.25
Mutante AAA	2.80	6.80	11.80	14.80	97.98
--= no determinado bajo las condiciones probadas					

Las ratas relativas de producción de monatina bajo las condiciones probadas indican la mayor mejora en la actividad inicial del Mutante 135, Mutante 136, Mutante E, Mutante G, Mutante O, Mutante BB y Mutante AAA según se determina comparando la rata de formación de monatina con proteína purificada a lo largo de la primer hora entre los mutantes y el DAT de control tipo silvestre (SEQ ID NO: 220). El Mutante E y el Mutante AAA de los DATs tienen alta actividad pero no fueron bien expresados (menos de 5% de proteína total) ni muy solubles bajo las condiciones probada.

Las muestras de ensayo fueron analizadas en cuanto a los intermediarios tales como precursor de monatina, I3P, y el subproducto ácido 4-hidroxi-4-metil glutámico (HMG) tal como se describe en el Ejemplo 36. El análisis de la cantidad de HMG formada fue determinado para los mutantes Mutante E, Mutante G, Mutante I, Mutante M, Mutante O, Mutante BB, Mutante PP, Mutante AAA y Mutante 110, Mutante 135 y Mutante 136. Parece que en el punto de tiempo a las 4 horas, se formó más HMG por parte de los mutantes Mutante 135, Mutante G, Mutante I, Mutante M y Mutante BB. Estos mutantes contenían todos el cambio V236T. El HMG estaba también presente por encima del control tipo silvestre (SEQ ID NO: 220) con los mutantes Mutante E, Mutante G, Mutante M y Mutante AAA probablemente debido al cambio en el residuo P20S.

Tabla 59: Formación de HMG por mutantes de DAT después de 4 horas

Polipéptido de DAT	HMG (mM) 4 horas
Control tipo silvestre (SEQ ID NO: 220)	nd
Mutante 110	nd
Mutante 135	1.0
Mutante 136	nd
Mutante E	0.2
Mutante G	1.6
Mutante I	0.8
Mutante M	1.6
Mutante O	nd
Mutante BB	1.5
Mutante AAA	0.6
nd=no detectado	

Ensayo de DAT para monitorizar la formación de I3P

La formación de I3P a partir de triptófano fue detectada y monitorizada a una longitud de onda de 340 nm. Las reacciones fueron llevadas a cabo en un volumen de reacción de 1 mL que contenía 900 µL de un D-triptófano 25 mM, sal de sodio y ácido pirúvico 25 mM, PLP 0.05 mM, fosfato de potasio 100 mM (pH 7.8) en solución combinado

con diluciones en 100 µL de DAT (proteína total) preparada como se describió anteriormente. Las enzimas fueron diluidas 1:100 y 1:200 con fosfato de potasio 50 mM frío (pH 7.8) y PLP 50 µM antes de la adición al ensayo. La enzima fue agregada a la mezcla de reacción 1:100 y monitorizada en incrementos de 15 segundos durante 3 minutos. La formación de indol-3-piruvato (I3P) fue monitorizada a una longitud de onda de 340 nm durante 3 minutos en un Bio-Rad Spectrophotometer (GE Healthscience, Piscataway, NJ) y se midieron las ratas dentro del rango dinámico de una curva estándar. La curva estándar fue generada con proteína de DAT tipo silvestre purificada (SEQ ID NO: 220) y la concentración de DAT en el extracto celular fue determinada con base en la ecuación de la línea para la curva estándar. La concentración efectiva de DAT con respecto a la DAT tipo silvestre para la primera reacción se reporta en la Tabla 60.

Tabla 60: Actividad de DAT (Conversión de triptófano a I3P)

Polipéptido de DAT	Rata de formación de I3P (ΔAbs340 nm/minuto)	Concentración de DAT (determinada por actividad) mg/mL	Actividad relativa a tipo silvestre para la primera reacción
Control tipo silvestre (SEQ ID NO: 220)	0.058	0.065	1.0
Mutante 135	0.067	0.075	1.2
Mutante 136	0.017	0.019	0.3
Mutante E	0.000	0.002	0.1
Mutante G	0.027	0.030	0.5
Mutante M	0.050	0.055	0.9
Mutante O	0.031	0.033	0.8
Mutante BB	0.045	0.050	0.8
Mutante AAA	0.002	0.004	0.2

La DAT tipo silvestre (SEQ ID NO: 220) y Mutante 136, E, G, M, O, BB y AAA pueden facilitar la conversión tanto de triptófano a I3P como del precursor de monatina a monatina. La Tabla 60 muestra que estos mutantes tienen actividad más baja para la conversión de triptófano a I3P con respecto al DAT tipo silvestre (SEQ ID NO: 220). Aún así, de acuerdo con la Tabla 58, los mismos mutantes produjeron más monatina total a partir de triptófano de lo que hizo el DAT tipo silvestre (SEQ ID NO: 220). Así, bajo las condiciones del ensayo descrito aquí, parece haber un efecto beneficioso de la producción de monatina a través del control de la conversión del triptófano a I3P en la ruta biosintética de la monatina. Por ejemplo, aunque el Mutante E mostró la actividad relativa más baja para la conversión de triptófano a I3P (véase Tabla 60) también produjo la cantidad más alta de monatina a las 4 horas (véase Tabla 58). Sin estar limitados por la teoría, los efectos beneficiosos de controlar la primera etapa en la reacción podría ser atribuida a una reducción de la constitución de I3P y subsecuente degradación potencial de I3P a productos diferentes a la monatina. En general, parece ser que el control de la rata de una o más de las reacciones involucradas en la producción de la monatina a partir del triptófano, utilizando, por ejemplo, uno o más mutantes DAT, puede tener un efecto beneficioso sobre la cantidad total de monatina producida.

Ejemplo 33 – Evaluación de DATs mutantes a 35°C.

Este ejemplo describe datos que demuestran la actividad enzimática de polipéptidos de ejemplo divulgados aquí. Los cultivos de inicio fueron cultivados durante la noche a 37°C con agitación a 250 rpm hasta que la OD_{600nm} alcanzo 0.05. 200 mL del medio Overnight Express II (Novagen, San Diego, CA) fueron inoculados y cultivados como se describe en el Ejemplo 3. Los cultivos fueron hechos crecer en duplicado y las pellas de células fueron combinadas. Las pellas fueron resuspendidas en 40 mL de regulador de fosfato de sodio 50 mM (pH 7.8) con PLP 0.05 mM y se sometieron a lisis utilizando una French Press (Sim Aminco, Rochester, NY) según las instrucciones del fabricante. El sobrenadante fue recolectado en un tubo limpio y almacenado a -80°C hasta su uso.

Se llevó a cabo un ensayo de formación de monatina en 3 etapas como se describe en los métodos con una concentración de DAT de aproximadamente 0.2 mg/mL y la aldolasa a una concentración de 0.1 mg/mL en viales de vidrio. Se incubaron muestras por duplicado a 25°C o 35°C y después de 1, 3 y 4 horas, se tomó una alícuota y se

- 5 agregó ácido fórmico hasta una concentración final de 2%, y las muestras fueron agitadas y filtradas. Las muestras fueron analizadas en cuanto al contenido de monatina utilizando la metodología del LC/MS/MS y para el triptófano y la alanina utilizando la metodología de fluorescencias postcolumna LC/OPA descritas en el Ejemplo 36. Las muestras fueron analizadas también con respecto a intermediarios tales como precursor de monatina, l3P, y ácido 4-hidroxi-4-metil glutámico (HMG) como se describe en el Ejemplo 36. La cantidad de monatina producida (mM) en diversos puntos del tiempo se muestra en la Tabla 61.

El ensayo de formación de monatina fue repetido para control tipo silvestre (SEQ ID NO: 220), Mutante 135 y Mutante M bajo condiciones similares excepto que las reacciones fueron llevadas a cabo en viales plásticos. La producción de monatina en diversos momentos del tiempo puede encontrarse en la Tabla 61.

10

Tabla 61: Formación de monatina a 25°C y 35°C

Polipéptidos de DAT	Monatina (mM) 1 hora	Monatina (mM) 3 horas	Monatina (mM) 4 horas
25°C			
Control tipo silvestre (pMET1a)	2.0	6.8	8.4
Mutante 135 (V236T)	10.0	14.4	14.2
Mutante 136 (241R)	4.0	10.8	12.4
Mutante E (20S, 73L, 241R, 270W)	0.8	4.2	5.6
Mutante M (20S, 236T)	10.0	13.4	14.2
Mutante O (20S, 241R)	8.0	13.8	13.6
Mutante BB (93G, 236T)	4.0	11.4	12.4
Mutante AAA (20S, 93G, 241R)	0.2	1.0	1.6
35°C			
Control tipo silvestre (pMET1a)	2.2	5.4	6.2
Mutante 135 (V236T)	9.4	9.2	9.8
Mutante 136 (241R)	4.8	9.4	10.4
Mutante E (20S, 73L, 241R, 270W)	0.6	3.4	4.2
Mutante M (20S, 236T)	9.2	13.6	14.6
Mutante O (20S, 241R)	9.6	10.6	10.8
Mutante BB (93G, 236T)	4.6	9.0	9.2
Mutante AAA (20S, 93G, 241R)	0.2	1.6	2.2

- 15 Se observaron títulos más bajos de monatina usando las enzimas DAT descritas aquí a 35°C bajo las condiciones del ensayo. Sin embargo, los mutantes seleccionados Mutante 135, Mutante 136, Mutante M, Mutante O y Mutante BB mostraron ratas de monatina iniciales incrementadas y títulos de monatina superiores a las 4 horas que el control tipo silvestre (SEQ ID NO: 220) a 35°C bajo las condiciones del ensayo.

Ejemplo34-Evaluación de DATs mutantes en biorreactores

Este Ejemplo describe datos que demuestran la actividad enzimática de polipéptidos de ejemplo divulgados aquí, en biorreactores. Se usaron reservas en glicerol del control tipo silvestre (SEQ ID NO: 220), Mutante 135, Mutante 136, Mutante M, Mutante O, y Mutante BB para sembrar placas con colonias individuales. Se utilizaron las colonias

individuales para inocular matraces que contenían 5 mL de medio LB con el antibiótico apropiado. Los cultivos de partida fueron cultivados durante la noche a 37°C con agitación a 250 rpm y se verificó la OD_{600nm}. Cuando la OD_{600nm} alcanzó 0.05, el cultivo de 5 mL fue inoculado en 200 mL de medio Overnight Express II (Novagen, San Diego, CA) y luego incubado a 30°C con agitación a 250 rpm. Cada cultivo fue cultivado en duplicado y las pellas de células fueron combinadas. Los cultivos fueron recolectados formando pellas de células por centrifugación a 4000 rpm durante 15 minutos. El sobrenadante fue vertido y la pella fue congelada para uso posterior o resuspendida en 40 mL de regulador de fosfato de sodio 50 mM (pH 7.8) y sometido a lisis utilizando una French Press (Sim Aminco, Rochester, NY) o un microfluidizador (Microfluidics Corporation, Newton, MA) según las instrucciones del fabricante. El sobrenadante fue recolectado en un tubo limpio y almacenado a -80°C hasta su uso. Aproximadamente 1 mL del lisado clarificado fue retenido para cuantificación de proteína utilizando el ensayo BCA (Pierce, Rockford, IL) y análisis por SDS-PAGE.

Las reacciones a escala de mesón (250 mL) fueron llevadas a cabo en fermentadores agitados Sixfors de 0.7 L (Infors AG, Bottmingen, Suiza) bajo un espacio de cabeza de nitrógeno como se describió en el Ejemplo 15. La mezcla de reacción contenía fosfato de potasio 10 mM, MgCl₂ 1 mM, PLP 0.05 mM, piruvato de sodio 200 mM y D-triptófano 130 mM. La mezcla de reacción fue ajustada a 25°C y ajustada a pH 7.8 con hidróxido de potasio. La aldolasa descrita en el Ejemplo 6 fue agregada como un extracto celular clarificado a 0.02 mg/mL de proteína objetivo. El control tipo silvestre (SEQ ID NO: 220), Mutante 135, Mutante 136, Mutante M, Mutante O y Mutante BB los DAT tuvieron expresiones de proteína soluble que variaban desde 15-35% con base en estimación visual. Los extractos celulares clarificados fueron agregados a 0.2 mg/mL de proteína objetivo.

El progreso de las reacciones fue seguido midiendo la producción de monatina a 1, 2, 4 y 24 horas utilizando la metodología de LC/MS/MS descrita en el Ejemplo 36. Los resultados se muestran la Tabla 62.

Tabla 62: Producción de monatina en fermentadores

Polipéptido de DAT	Expresión de proteína	Monatina (mM) 1 hora	Monatina (mM) 2 horas	Monatina (mM) 4 horas	Monatina (mM) 24 horas
Control tipo silvestre	25%	0.90	2.80	12.40	12.80
Mutante 135	30%	0.50	8.80	12.40	12.40
Mutante 136	35%	3.80	7.80	11.60	12.80
Mutante M	15%	3.40	6.80	12.10	12.20
Mutante O	15%	5.20	8.60	10.90	9.80
Mutante BB	15%	3.40	6.20	10.50	12.60

La rata inicial de producción de monatina observada con los mutantes Mutante 136, Mutante M, Mutante O, y Mutante BB fue más rápida que la rata con el control tipo silvestre (SEQ ID NO: 220). Todos los mutantes mostraron formación de monatina mejorada a las 2 horas bajo las condiciones probadas. El título de monatina más bajo que el esperado a 1 hora para el Mutante 135 fue atribuido a la exposición inadvertida a oxígeno durante la primera hora. Después de 4 horas, el título de monatina fue comparable entre los mutantes y el control bajo las condiciones probadas.

Ejemplo 35 – Evaluación del impacto de la temperatura sobre DATs de mutante en BioReactores

Este ejemplo describe datos que demuestran la actividad enzimática de polipéptidos de ejemplo divulgados aquí bajo diferentes condiciones de temperatura. El control tipo silvestre (SEQ ID NO: 220), Mutante 135 y Mutante M fueron producidos en un fermentador a la escala de 2,5 L como se describió en el Ejemplo 15. Al final de la fermentación, las células fueron recolectadas por centrifugación a 5000-7000 x g durante 10 minutos y congeladas como una pasta celular húmeda a -80°C.

Para preparar un extracto libre de células que contenía las D-aminotransferasas del control tipo silvestre, Mutante 135 y Mutante M, se suspendieron 50 g de la pasta celular húmeda en 150 mL de regulador de fosfato de potasio 50 mM (pH 7.8) que contenía piridoxal fosfato (PLP) 0.05 mM y luego se deshizo utilizando un homogeneizador Microfluidics (Microfluidics, Newton, MA) (3 pases a 18,000 psi), manteniendo la temperatura de la suspensión a menos de 15°C. Los residuos celulares fueron retirados por centrifugación (20,000 x g durante 30 minutos).

La rata de formación de I3P a partir de triptófano fue monitorizada a 340 nm durante 3 minutos como se describe en el Ejemplo 32. La concentración del control tipo silvestre fue determinada como 6.8 mg/mL, la concentración del Mutante 135 fue determinada como 7.0 mg/mL y el Mutante M fue determinado como 5.6 mg/mL con base en la curva estándar generada con DAT purificado de control tipo silvestre. Las concentraciones de DAT determinadas por la formación de I3P fueron utilizadas para dosificar el infors a 0.2 mg/mL de DAT. La aldolasa fue agregada como un extracto libre de células a 0.02 mg/mL de aldolasa. La mezcla de reacción contenía fosfato de potasio 10 mM, MgCl₂ 1 mM, PLP 0.05 mM, piruvato de sodio 200 mM y D-triptófano 130 mM bajo un espacio de cabeza de nitrógeno. Cada uno de los DAT fue evaluado para la producción de monatina en un biorreactor a 35°C y a 25°C.

Se tomaron muestras a 0.5, 1, 3, 4 y 24 horas y se analizaron utilizando la metodología de LC/MS/MS descrita en el Ejemplo 36. Los resultados se muestran en la Tabla 63.

Tabla 63: Fermentadores a 25° y 35°C

Polipéptidos de DAT	Monatina (mM) 0.5 horas	Monatina (mM) 1 horas	Monatina (mM) 3 horas	Monatina (mM) 4 horas	Monatina (mM) 24 horas
25°C					
Control tipo silvestre (SEQ ID NO: 220)	0.9	2.4	5.6	7.9	19.1
Mutante 135	1.6	4.4	10.9	12.1	18.6
Mutante M	2.1	4.5	9.4	12.4	17.4
35°C					
Control tipo silvestre (SEQ ID NO: 220)	2.3	3.9	6.5	7.9	10.7
Mutante 135	4.1	6.1	9.8	11.5	14.9
Mutante M	4.1	6.3	9.9	11.3	14.9

Como se ve en el Ejemplo 34, los DAT seleccionados de mutante produjeron títulos de monatina superiores a 35°C en comparación con los DAT de control tipo silvestre (SEQ ID NO: 220). El DAT de control tipo silvestre tenía una rata inicial más lenta de producción de monatina pero un título final más alto a 25°C bajo las condiciones probadas. Ambos mutantes Mutante 135 y Mutante M mostraron una actividad mejorada sobre el control tipo silvestre a 25°C y 35°C. Los mutantes Mutante 135 y Mutante M mostraron ambos una rata inicial más alta de producción de monatina y un título final más alto a 35°C en comparación con el control bajo las condiciones probadas. Los mutantes probados fueron más estables que el control tipo silvestre a las temperaturas más altas, esto indica las ventajas de las tecnologías GSSM y TMCA en la producción de mutantes con mayor termoestabilidad que el control tipo silvestre. Una persona experimentada en la técnica podría seleccionar estas bibliotecas GSSM o TMCA para mutantes, por ejemplo, con tolerancia incrementada a la temperatura.

Ejemplo 36 – Detección de monatina, MP, triptófano, alanina y HMG

Este ejemplo describe la metodología analítica asociada con la caracterización posterior de enzimas D-aminotransferasa (DAT) divulgadas aquí.

Análisis por UPLC/UV de monatina y triptófano

Los análisis de mezclas de monatina y triptófano derivados de reacciones bioquímicas fueron llevados a cabo utilizando un instrumento Waters Acquity UPLC incluyendo un monitor de absorbancia Waters Acquity Photo-Diode Array (PDA). La separaciones UPLC fueron hecha utilizando una columna Agilent XDB C8 1,8 µm 2,1 x 100 nm (parte # 928700-906) (Milford, MA) a 23°C. La fase móvil de UPLC consistió de A) agua que contenía 0.1% de ácido fórmico B) acetonitrilo que contenía 0.1% de ácido fórmico.

- El gradiente de elución fue lineal de 5% de B a 40% de B, 0-4 minutos, lineal de 40% de B a 90% de B, 4-4.2 minutos, isocrático de 90% de B a 90% de B, 4.2-5.2 minutos, lineal de 90% de B a 5% de B, 5.2-5.3 minutos, con un periodo de reequilibrio de 1.2 minutos entre corrida. La rata de flujo fue de 0.5 mL/minuto y la absorbancia PDA fue monitorizada a 280 nm. Las concentraciones de muestras se calculan a partir de una calibración de mínimos cuadrados lineal del área de pico a 280 nm a concentración conocida, con un coeficiente mínimo de determinación de 99.9%.
- Derivación de intermediarios de monatina (ácido indol-3-pirúvico (I3P), ácido hidroximetiloxiglutarico, precursor de monatina y piruvato) con clorhidrato de O-(4-nitrobencil) hidroxilamina (NBHA)
- En el proceso de la producción de monatina, se forman y utilizan diversos compuestos intermediarios. Estos compuestos incluyen: ácido indol-3-pirúvico (I3P), ácido hidroximetiloxiglutarico, precursor de monatina y piruvato. El grupo funcional cetona en estos compuestos puede ser derivado con clorhidrato de O-(4-nitrobencil) hidroxilamina (NBHA).
- A 20 µL de muestra o estándar, se agregaron 140 µL de NBHA (40 mg/mL en piridina) en un vial ámbar. Las muestras fueron sometidas a sonicación durante 15 minutos en la presencia de calor con mezcla ocasional. Se llevó a cabo una dilución 1:3 en acetonitrilo al 35% en agua.
- Análisis por UPLC/UV de intermediarios de monatina (ácido indol-3-pirúvico, ácido hidroximetiloxiglutarico, precursor de monatina y piruvato)
- Se utilizó un instrumento Waters Acquity UPLC que incluía un monitor de absorbancia Waters Acquity Photo-Diode Array (PDA) (Waters, Milford, MA) para el análisis de los compuestos intermediarios. Las separaciones por UPLC fueron hechas utilizando una columna Waters Acquity HSS T3 1.8 mm x 150 mm (Waters, Milford, MA) a 50°C. La fase móvil para UPLC consistió de A) agua que contenía 0.3% de ácido fórmico y formiato de amonio 10 mM y B) acetonitrilo/metanol 50/50 que contenía 0.3% de ácido fórmico y formiato de amonio 10 mM.
- La elución de gradiente fue línea de 5% de B a 40% de B, 0-1.5 minutos, lineal de 40% de B, a 50% de B, 1.5-4.5 minutos, lineal de 50% de B a 90% de B, 4.5-7.5 minutos, lineal de 90% de B a 95% de B, 7.5-10.5 minutos, con un periodo de reequilibrio de 3 minutos entre corridas. La rata de flujo fue de 0.15 mL/minuto desde 0-7.5 minutos, 0.18 mL/minuto de 7.5-10.5 minutos, 0.19 mL/minuto de 10.5-11 minutos, y 0.15 mL/minuto de 11-13.5 minutos. La absorbancia de PDA fue monitorizada a 270 nm.
- Las concentraciones de muestra fueron calculadas a partir de una calibración de mínimos cuadrados lineal de área de pico a 270 nm a concentración conocida, con un coeficiente mínimo de determinación de 99.9%.
- Medición de monatina quiral por LC/MS/MS (MRM)
- La determinación de la distribución de estereoisómeros de monatina a reacciones bioquímicas fue lograda por derivación con 1-fluoro-2-4-dinitrofenil-5-L-alanina amida 30 (FDAA), seguida por medición en fase reversa por LC/MS/MS MRM.
- Derivación de monatina con FDAA
- 100 µL de una solución al 1% de FDAA en acetona fueron agregados a una muestra de 50 µL de muestra o estándar. Se agregaron 20 µL de bicarbonato de sodio a 1.0 M, y la mezcla fue incubada durante 1 hora a 40°C con agitación ocasional. La muestra fue removida y enfriada, y neutralizada con 20 µL de HCl 2.0 M (puede requerirse más HCl para efectuar la neutralización de una mezcla biológica regulada). Las muestras fueron analizadas por LC/MS/MS.
- Monitorización de reacción múltiple por LC/MS/MS para la determinación de la distribución de estereoisómeros de monatina
- Los análisis fueron llevados a cabo utilizando el instrumento de cromatografía líquida en tándem con espectrometría de masas Waters/Micromass® (LC/MS/MS) incluyendo un cromatógrafo líquido Waters 2795 con un monitor de absorbancia Waters 996 Photo-Diode Array (PDA) (Waters, Milford, MA) colocado en serie entre el cromatógrafo y un espectrómetro de masas de cuadrupolo triple Micromass® Quattro Ultima®. Las separaciones por LC capaces de separar todos los cuatro estereoisómeros de la monatina (específicamente FDAA-monatina) fueron llevadas a cabo sobre una columna de cromatografía de fase reversa Phenomenex Luna® 2.0 x 250 mm (3 µm) C18 a 40°C. La fase móvil para LC consistió de A) agua que contenía 0.05% (masa/volumen) de acetato de amonio y B) acetonitrilo. La elución fue isocrática a 13% de B, 0-2 minutos, lineal desde 13% de B a 30% de B, 2-15 minutos, lineal desde 30% de B a 80% de B, 15-16 minutos, isocrático a 80% de B 16-21 minutos, y lineal desde 80% de B a 13% de B, 21-22 minutos, con un periodo de reequilibrio de 8 minutos entre corridas. La rata de flujo fue de 0.23 mL/minuto, y la absorbancia por PDA fue monitorizada desde 200 nm hasta 400 nm. Todos los parámetros de ESI-MS fueron optimizados y seleccionados con base en la generación de 20 iones moleculares desprotonados ([M - H]⁻) de FDAA-

monatina, y la producción de fragmentos de iones fragmentarios característicos. Se usaron los siguientes parámetros instrumentales para el análisis por LC/MS de monatina en el modo ESI/MS de ión negativo: Capilar: 3.0 kV; Cono: 40 V; Hex 1: 15 V; Apertura: 0.1 V; Hex 2: 0.1 V; Temperatura de fuente: 120°C; Temperatura de desolvatación: 350°C; Gas de desolvatación: 662 L/h; Gas de cono: 42 L/h; Resolución de masa baja (Q1): 14.0; Resolución de masa alta (Q1): 15.0; energía del ión: 0.5; Entrada: 0 V; Energía de colisión: 20; Salida: 0 V; resolución de masa baja (Q2): 15; Resolución de masa alta (Q2): 14; energía del ión (Q2): 2.0; Multiplicador: 650. Se utilizaron tres transiciones progenitora a hija específicas de FDAA-monatina para detectar específicamente la FDAA-monatina en reacciones *in vitro* e *in vivo*. Las transiciones monitorizadas para la monatina fueron 542.97 a 267.94, 542.97 a 499.07, y 542.97 a 525.04. La identificación de estereoisómeros de FDAA-monatina se basó en el tiempo de retención cromatográfico en comparación con estereoisómeros de monatina purificados, y datos de espectros de masas.

Derivación en cromatografía líquida postcolumna con OPA, detección de fluorescencia de aminoácidos, incluyendo hidroximetil glutamato (HMG) y alanina

Los análisis de mezclas para HMG y alanina derivados a partir de reacciones bioquímicas fueron llevados a cabo utilizando un Waters Alliance 2695 y un instrumento Waters 600 configurado con un detector de absorbancia de longitudes de onda duales Waters 2487 y un detector de fluorescencia Waters 2475 como sistema de detección (Waters, Milford, MA). Las separaciones por HPLC se hicieron utilizando DOS? columnas Phenomenex Aqua C18 125A, 150 mm x 2.1 mm, 3 m, Cat #00F-4311B0 en serie como columnas analíticas, y Phenomenex Aqua C18 125A, 30 mm x 2.1 mm, 3 m, Cat # 00A-4311B0 como columna de extracción en fase sólida en línea (SPE). La temperatura de las dos columnas analíticas fue fijada a 55°C, y la columna SPE en línea estaba a temperatura ambiente. La fase móvil para HPLC consistió de A) ácido acético al 0.6% con 1% de metanol. La rata de flujo fue (100% A) 0.2 mL/minuto de 0-3.5 minutos, 0.24 mL/minuto de 3.5-6.5 minutos, 0.26 mL/minutos de 6.5-10.4 minutos, y 0.2 mL/minuto de 10.4-11 minutos. El detector de absorbancia UV-VIS fue fijado para monitorizar a longitud de onda de 336 nm. El detector de fluorescencia fue fijado a 348 nm y 450 nm para monitorizar las longitudes de onda de excitación y emisión respectivamente.

REIVINDICACIONES

1. Un ácido nucleico aislado, sintético o recombinante que codifica un polipéptido que tiene actividad de D-aminoácido transferasa que comprende

5 (a) una secuencia de ácidos nucleicos que tiene al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de ácidos nucleicos de SEQ ID NO: 219;

(b) un ácido nucleico que codifica un polipéptido como se define en SEQ ID NO: 220;

(c) un ácido nucleico que codifica un polipéptido como se define en SEQ ID NO: 220 con una, varias o todas las modificaciones de aminoácidos como se definen en la siguiente Tabla:

Nombre del mutante	Mutación	Nombre del mutante	Mutación
1	Y6L	92	D2G
2	Y6C; MUTACION SILENCIOSA AT AA31 (GGC → GGT)	93	D2Q
3	Y6F	94	D2F
4	Y6L	95	D2A
5	Y6H	96	D2T
6	Y6L	97	D2N
7	Y6M	98	D2R
8	N10S	99	D2I
9	N10W	100	D2V;G9A
10	N10T	101	G12A
11	N10R	102	D47W
12	N10T	103	S56S
13	L14V	104	I64H
14	L14L	105	L66L
15	G41G	106	I64C
16	T18W	107	L66G
17	N40N	108	E69Y
18	V19T	109	T74L
19	V42V	110	K73L
20	I62C	111	T74V
21	V82A	112	T74M
22	A57M	113	T74R
23	V42M	114	T74A
24	G41Y	115	N76C

ES 2 531 290 T3

Nombre del mutante	Mutación	Nombre del mutante	Mutación
25	A45L	116	E77R
26	V93Y	117	R156A
27	V93G	118	K72M
28	L46A	119	S205A
29	L46H	120	Q209S
30	G98A	121	V212E
31	P20S	122	R213W
32	V93A	123	I216T
33	V103T	124	P217H
34	P108F	125	P217V
35	V93L	126	D219F
36	S101S	127	E220V
37	A106G	128	R221E
38	S101Q	129	F223C
39	P108T	130	S226P
40	N118G	131	L228F
41	P108G	132	V234A
42	I120L	133	S238S
43	A106W	134	V236T
44	N118R	135	V236T
45	N110A; N118G	136	T241R
46	N118A	137	L242F
47	N118R	138	T241R
48	P117W; N118K	139	T241C
49	D133N	140	E248F
50	K126Q	141	D250E
51	K126R	142	K257V
52	K128S	143	G256K
53	I127M	144	E260G
54	T131T	145	L262R

Nombre del mutante	Mutación	Nombre del mutante	Mutación
55	D133L	146	K263M
56	M132A	147	D267G
57	D133E	148	D267R
58	L129V	149	I265L
59	K126K	150	E268S
60	I130M	151	L270S
61	M132Y	152	L270G
62	K128R	153	L270W
63	M132R	154	R271S
64	L129I	155	I274W
65	K128L; D2D (GAC → GAT)	156	G278S
66	F137W	157	Y279C
67	I152V	158	S284R
68	N55L	159	E282G
69	N150S	160	T280N
70	L138L	161	V286G
71	P149P	162	R285F
72	G161G	163	V286R
73	A165T	164	G240G
74	H163R	165	E61R
75	H163K	166	E61D
76	H168A	167	E61Y
77	E171S	168	G85G
78	E171R	169	G85D
79	E171R	170	S80R
80	T172I	171	Y79R
81	C176G	172	Y79V
82	A177S	173	W283V
83	A177S	174	W283E
84	S80L; R156W	175	W283A

Nombre del mutante	Mutación	Nombre del mutante	Mutación
85	H182G	176	W283S
86	N186S	177	W283G
87	K185R	178	W283A
88	K185T	179	W283R
89	D2H	180	W283T
90	D2T; E250G	181	P281W
91	D2Y	182*	V236T; T241R

(d) un ácido nucleico que codifica un polipéptido como se define en SEQ ID NO: 220 que tiene al menos una combinación de modificaciones de aminoácidos como se establece en la siguiente Tabla:

Número/designación	Mutación	Número/designación	Mutación
1	P20S-N118G	30	P20S-N110A-N118G
2	T74V-V93G-L270W	31	N110A-N118G-T241R-L270W
3	P20S-T74V-L270W	32	P20S-T74V-V93G-N110A-N118G-V236T-L270W
4	T74V-L270W	33	V93G-N110A-N118G-V236T
5	P20S-K73L-T241R-L270W	34	P20S-T74V-N110A-N118G
6	K73L-V93G-V236T-T241R	35	N110A-N118G
7	P20S-T74V-V236T	36	P20S-V93G-N110A-N118G-T241R
8	P20S-K73L-V93G	37	N110A-N118A-L270W
9	K73L-V236T	38	P20S-N110A-N118G-L270W
10	P20S-L270W	39	N110A-N118A-V236T-T241R
11	2N-P20S-K73L-V93G-N118G	40	N110A-N118G-L270W
12	P20S-T74V-N118A	41	V93G-N110A-N118G-T241R
13	P20S-V236T	42	P20S-V93G-N110A-N118G
14	P20S-T241R-L270W	43	V93G-N110A
15	P20S-T241R	44	P20S-N110A-N118G-V236T
16	T74V-V93G-V236T-T241R	45	T74V-N110A-N118A-V236T
17	P20S-K73L-T74V-L270W	46	P20S-K73L-T74V-N110A-N118G-V236T-T241R
18	P20S-V93G-V236T	47	86E (SILENT GAG → GAA)-N110A-N118A-V236T
19	P20S-K73L-T74V-T241R-L270W	48	T74V-N18G

20	P20S-K73L-L270W	49	P20S-T241R-L270W-277T (SILENT ACA → ACG)
21	T74V-V93G-N118G-V236T-T241R	50	T74V-N118-A-L270W
22	P20S-K73L-210A (SILENT-GCC → GCT)-V236T	51	P20S-K73L-N118A-L270W
23	N118A-L270W	52	P20S-V93G-T241R
24	P20S-58K (SILENT AAG → AAA)-L270W	53	T74V-V93G-N110A-T241R
25	P20S-V93G-N118G	54	V93G-N110A-N118A
26	P20S-V236T-T241R	55	P20S-T74V-V93G-N110A-N118G-T241R
27	P20S-P117W-N118A-V236T-L270W	56	T74V-N110A-N118G-L270W
28	V93G-V236T	57	P20S-231A (SILENT GCG → GCA)-V236T
29	V236T-L270W	58	V93G-V236T-T241R
30	P20S-N118G-L270W		

; 0

(e) una secuencia de ácidos nucleicos completamente complementaria a la secuencia de ácidos nucleicos de cualquiera de (a) a (d).

- 5 2. Un casete de expresión, un vector o un vehículo de clonación que comprende la secuencia de ácidos nucleicos de la reivindicación 1, en donde el vehículo de clonación comprende un vector viral, un plásmido, un fago, un fagélido, un cósmido, un fósido, un bacteriófago o un cromosoma artificial, en donde el vector viral comprende un vector de adenovirus, un vector retroviral o un vector viral adenoasociado, o, el cromosoma artificial comprende un cromosoma artificial bacteriano (BAC), un vector derivado de bacteriófago PI (PAC), un cromosoma artificial de levadura (YAC), o un cromosoma artificial de mamífero (MAC).
- 10 3. Una célula transformada que comprende la secuencia de ácidos nucleicos de la reivindicación 1, o el casete de expresión, vector vehículo de clonación de la reivindicación 2, en donde la célula es una célula bacteriana, una célula de mamífero, una célula fúngica, una célula de levadura, o una célula de insecto, o una célula vegetal.
4. Un polipéptido aislado, sintético o recombinante que tiene una actividad de D-aminoácido transferasa que comprende
- 15 (a) una secuencia de aminoácidos que tiene 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 220;
- (b) una secuencia de aminoácidos como se define en SEQ ID NO: 220 con una, varias o todas las modificaciones de aminoácidos de la siguiente Tabla:

Nombre del mutante	Mutación	Nombre del mutante	Mutación
1	Y6L	92	D2G
2	Y6C; MUTACION SILENCIOSA AT AA31 (GGC → GGT)	93	D2Q
3	Y6F	94	D2F
4	Y6L	95	D2A
5	Y6H	96	D2T
6	Y6L	97	D2N
7	Y6M	98	D2R

Nombre del mutante	Mutación	Nombre del mutante	Mutación
8	N10S	99	D2I
9	N10W	100	D2V;G9A
10	N10T	101	G12A
11	N10R	102	D47W
12	N10T	103	S56S
13	L14V	104	I64H
14	L14L	105	L66L
15	G41G	106	I64C
16	T18W	107	L66G
17	N40N	108	E69Y
18	V19T	109	T74L
19	V42V	110	K73L
20	I62C	111	T74V
21	V82A	112	T74M
22	A57M	113	T74R
23	V42M	114	T74A
24	G41Y	115	N76C
25	A45L	116	E77R
26	V93Y	117	R156A
27	V93G	118	K72M
28	L46A	119	S205A
29	L46H	120	Q209S
30	G98A	121	V212E
31	P20S	122	R213W
32	V93A	123	I216T
33	V103T	124	P217H
34	P108F	125	P217V
35	V93L	126	D219F
36	S101S	127	E220V
37	A106G	128	R221E
38	S101Q	129	F223C

Nombre del mutante	Mutación	Nombre del mutante	Mutación
39	P108T	130	S226P
40	N118G	131	L228F
41	P108G	132	V234A
42	I120L	133	S238S
43	A106W	134	V236T
44	N118R	135	V236T
45	N110A; N118G	136	T241R
46	N118A	137	L242F
47	N118R	138	T241R
48	P117W; N118K	139	T241C
49	D133N	140	E248F
50	K126Q	141	D250E
51	K126R	142	K257V
52	K128S	143	G256K
53	I127M	144	E260G
54	T131T	145	L262R
55	D133L	146	K263M
56	M132A	147	D267G
57	D133E	148	D267R
58	L129V	149	I265L
59	K126K	150	E268S
60	I130M	151	L270S
61	M132Y	152	L270G
62	K128R	153	L270W
63	M132R	154	R271S
64	L129I	155	I274W
65	K128L; D2D (GAC → GAT)	156	G278S
66	F137W	157	Y279C
67	I152V	158	S284R
68	N55L	159	E282G
69	N150S	160	T280N

Nombre del mutante	Mutación	Nombre del mutante	Mutación
70	L138L	161	V286G
71	P149P	162	R285F
72	G161G	163	V286R
73	A165T	164	G240G
74	H163R	165	E61R
75	H163K	166	E61D
76	H168A	167	E61Y
77	E171S	168	G85G
78	E171R	169	G85D
79	E171R	170	S80R
80	T172I	171	Y79R
81	C176G	172	Y79V
82	A177S	173	W283V
83	A177S	174	W283E
84	S80L; R156W	175	W283A
85	H182G	176	W283S
86	N186S	177	W283G
87	K185R	178	W283A
88	K185T	179	W283R
89	D2H	180	W283T
90	D2T; E250G	181	P281W
91	D2Y	182	V236T; T241R

(c) una secuencia de aminoácidos como se define en SEQ ID NO: 220 que tiene al menos una de las combinaciones de modificaciones de aminoácidos como se definen en la siguiente Tabla:

Número/designación	Mutación	Número/designación	Mutación
1	P20S-N118G	30	P20S-N110A-N118G
2	T74V-V93G-L270W	31	N110A-N118G-T241R-L270W
3	P20S-T74V-L270W	32	P20S-T74V-V93G-N110A-N118G-V236T-L270W
4	T74V-L270W	33	V93G-N110A-N118G-V236T
5	P20S-K73L-T241R-L270W	34	P20S-T74V-N110A-N118G

Número/designación	Mutación	Número/designación	Mutación
6	K73L-V93G-V236T-T241R	35	N110A-N118G
7	P20S-T74V-V236T	36	P20S-V93G-N110AN118G-T241R
8	P20S-K73L-V93G	37	N110A-N118A-L270W
9	K73L-V236T	38	P20S-N110A-N118G-L270W
10	P20S-L270W	39	N110A-N118A-V236T-T241R
11	2N-P20S-K73L-V93G-N118G	40	N110A-N118G-L270W
12	P20S-T74V-N118A	41	V93G-N110AN118G-T241R
13	P20S-V236T	42	P20S-V93G-N110AN118G
14	P20S-T241R-L270W	43	V93G-N110A
15	P20S-T241R	44	P20S-N110A-N118G-V236T
16	T74V-V93G-V236T-T241R	45	T74V-N110A-N118A-V236T
17	P20S-K73L-T74V-L270W	46	P20S-K73L-T74V-N110A-N118G-V236T-T241R
18	P20S-V93G-V236T	47	86E (SILENT GAG → GAA)-N110A-N118A-V236T
19	P20S-K73L-T74V-T241R-L270W	48	T74V-N18G
20	P20S-K73L-L270W	49	P20S-T241R-L270W-277T (SILENT ACA → ACG)
21	T74V-V93G-N118G-V236T-T241R	50	T74V-N118-A-L270W
22	P20S-K73L-210A (SILENT-GCC → GCT)-V236T	51	P20S-K73L-N118A-L270W
23	N118A-L270W	52	P20S-V93G-T241R
24	P20S-58K (SILENT AAG → AAA)-L270W	53	T74V-V93G-N110A-T241R
25	P20S-V93G-N118G	54	V93G-N110A-N118A
26	P20S-V236T-T241R	55	P20S-T74V-V93G-N110A-N118G-T241R
27	P20S-P117W-N118A-V236T-L270W	56	T74V-N110A-N118G-L270W
28	V93G-V236T	57	P20S-231A (SILENT GCG → GCA)-V236T
29	V236T-L270W	58	V93G-V236T-T241R
30	P20S-N118G-L270W		

; o

(d) una secuencia de aminoácidos codificada por la secuencia de ácidos nucleicos de la reivindicación 1.

5 5. Una composición que comprende el polipéptido de la reivindicación 4, en donde la preparación de proteína comprende un líquido, un sólido o un gel.

6. Un método para producir un polipéptido recombinante de la reivindicación 4 que comprende las etapas de:

(a) proveer el ácido nucleico de la reivindicación 1; y

(b) expresar el ácido nucleico de la etapa (a) bajo condiciones que permiten la expresión del polipéptido, produciendo por lo tanto un polipéptido recombinante, y opcionalmente el método comprende transformar una célula anfitriona con el ácido nucleico de la etapa (a) seguido por la expresión del ácido nucleico de la etapa (a), produciendo por lo tanto un polipéptido recombinante en una célula transformada.

5 7. Un método para generar una variante del ácido nucleico de la reivindicación 1 que comprende las etapas de:

(a) proveer ácido nucleico plantilla que comprende la secuencia de cualquiera de las reivindicación 1; y

(b) modificar, eliminar o agregar uno o más nucleótidos en la secuencia de plantilla, o una combinación de los mismos, para generar una variante del ácido nucleico de plantilla,

10 en donde opcionalmente el método comprende adicionalmente expresar el ácido nucleico variante para generar un polipéptido variante de transferasa por ejemplo, transaminasa, por ejemplo, D-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, D-aminoácido deshidrogenasa,

y opcionalmente, las modificaciones, adiciones o eliminaciones son introducidas por un método que comprende PCR propenso al error, mezcla, mutagénesis dirigida a oligonucleótidos, PCR de ensamblaje, mutagénesis por PCR sexual, mutagénesis *in vivo*, mutagénesis por casete, mutagénesis de ensamblaje recursivo, mutagénesis de ensamblaje exponencial, mutagénesis dirigida al sitio, reensamblaje de genes, mutagénesis por saturación en el sitio del gen (GSSM), reensamblaje por ligazón sintética (SLR) y una combinación de los mismos, o, las modificaciones, adiciones o eliminaciones son introducidas por un método que comprende recombinación, recombinación de secuencias recursiva, mutagénesis de ADN modificada con fosfotioato, mutagénesis de plantilla que contiene uracilo, mutagénesis de dúplex con brecha, mutagénesis de reparación de punto de no coincidencia, mutagénesis de cepa anfitriona deficiente en reparación, mutagénesis química, mutagénesis radiogénica, mutagénesis por eliminación, mutagénesis por restricción-selección, mutagénesis por restricción-purificación, síntesis de genes artificiales, mutagénesis de ensamble, creación de ácidos nucleicos y multímeros quiméricos y una combinación de los mismos,

25 en donde opcionalmente, el método es repetido iterativamente hasta que una transferasa por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidoreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa que tiene una actividad alterada o diferente o una estabilidad alterada o diferente de la del polipéptido codificado por el ácido nucleico de plantilla se produce,

30 en donde opcionalmente el polipéptido variante de la transferasa por ejemplo, transaminasa, por ejemplo, D-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, D-aminoácido deshidrogenasa es termotolerante, y retiene alguna actividad después de ser expuesto a una temperatura elevada, o, opcionalmente el polipéptido variante de la transferasa por ejemplo, transaminasa, por ejemplo, D-aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, D-aminoácido deshidrogenasa tiene glicosilación incrementada en comparación con la transferasa por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidoreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa codificada por un ácido nucleico de plantilla, u opcionalmente, el polipéptido variante de transferasa por ejemplo, transaminasa, por ejemplo, d- aminoácido transferasa, y/o oxidoreductasa, por ejemplo, deshidrogenasa, por ejemplo, D-aminoácido deshidrogenasa tiene una actividad de transferasa por ejemplo, una actividad de transaminasa, por ejemplo, una actividad de D-aminoácido transferasa or una actividad de w-transaminasa, y/o una actividad de oxidoreductasa, por ejemplo, una actividad de deshidrogenasa, por ejemplo, una actividad de D-aminoácido deshidrogenasa bajo una alta temperatura, en donde la transferasa, por ejemplo, la transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidoreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa codificada por el ácido nucleico de plantilla no es activa bajo la alta temperatura,

45 en donde opcionalmente, el método es repetido iterativamente hasta que una secuencia de codificación de transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidoreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa que tiene un uso de codón alterado con respecto al del ácido nucleico de plantilla es producido,

50 en donde opcionalmente el método es repetido iterativamente hasta que un gen de transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidoreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa que tienen un nivel más alto o más bajo de expresión de mensaje o estabilidad con respecto al ácido nucleico de plantilla es producido.

8. Una bebida o precursor de bebida que comprende el polipéptido de la reivindicación 4.

55 9. Un pienso, alimento, aditivo para alimento o pienso, suplemento para alimento o pienso, o auxiliar dietario o suplemento dietario que comprende el polipéptido de la reivindicación 4, en donde opcionalmente el alimento, pienso, aditivo para alimento o pienso, suplemento para alimento o pienso, o auxiliar dietario comprende

adicionalmente un vehículo seleccionado del grupo consistente de un germen de grano, un germen de grano que carece de aceite, un heno una alfalfa, un fleo, una corteza de soja, una torta de semilla de girasol y un midd de trigo,

- 5 y opcionalmente un vehículo comprende un germen de grano que carece de aceite, u opcionalmente, la enzima transferasa, por ejemplo, the transaminasa, por ejemplo, la d-aminoácido transferasa, y/o la oxidorreductasa, por ejemplo, la deshidrogenasa, por ejemplo, la d-aminoácido deshidrogenasa es glicosilada para proveer termoestabilidad en condiciones de peletización, y opcionalmente la matriz de administración es formada por peletización de una mezcla que comprende un germen de grano y una transferasa, por ejemplo, a transaminasa, por ejemplo, una d-aminoácido transferasa, y/o una oxidorreductasa, por ejemplo, a deshidrogenasa, por ejemplo, una d-aminoácido deshidrogenasa, y opcionalmente las condiciones de peletización incluyen la aplicación de vapor, y
- 10 opcionalmente las condiciones de peletización comprenden la aplicación de una temperatura por encima de aproximadamente 80°C durante aproximadamente 5 minutos y la enzima retiene una actividad específica de al menos 350 hasta aproximadamente 900 unidades por miligramo de enzima.

10. Un método para hacer un piruvato y/o un D-glutamato que comprende

- (a) proveer una D-alanina y un 2-oxoglutarato;
- 15 (b) proveer el polipéptido de la reivindicación 4; y
- (c) poner en contacto el polipéptido de (b) con la D-alanina + 2-oxoglutarato bajo condiciones en donde el polipéptido cataliza la reacción D-alanina + 2-oxoglutarato \rightleftharpoons piruvato + D-glutamato.

11. Un método para hacer un 2-oxo ácido que comprende

- (i)
- 20 (a) proveer un D-aminoácido + H₂O + aceptor;
- (b) proveer el polipéptido de transaminasa de la reivindicación 4; y
- (c) poner en contacto el polipéptido de (b) con el D-aminoácido + H₂O + aceptor, bajo condiciones en donde el polipéptido cataliza la reacción: D-aminoácido + H₂O + aceptor \rightleftharpoons un 2-oxo ácido + NH₃ + aceptor reducido; o
- 25 (ii) el método de (i), en donde el aceptor es una benzoquinona es (iii) el método de (ii) en donde la benzoquinona es una 1,2-benzoquinona o una 1,4-benzoquinona, o ubiquinona, ubidecarenona o coenzima Q.

12. Un método para transferir un grupo amino de un aminoácido a un alfa-cetoácido que comprende

- (i)
- (a) proveer un aminoácido;
- (b) proveer el polipéptido de transaminasa de la reivindicación 4; y
- 30 (c) poner en contacto el polipéptido de (b) con el aminoácido bajo condiciones en donde el polipéptido cataliza la conversión del aminoácido a un alfa-cetoácido; o
- (ii) el método de (i), en donde la actividad de transaminasa comprende catalizar la conversión de una mezcla de aminoácidos racémica a un alfa-cetoácido de manera sustancial ópticamente puro.

13. Un método para hacer un aminoácido a partir de una alfa-cetoácido que comprende

- 35 (i)
- (a) proveer un alfa-cetoácido
- (b) proveer el polipéptido de la reivindicación 4; y
- (c) poner en contacto el polipéptido de (b) con el alfa-cetoácido bajo condiciones en donde el polipéptido cataliza la conversión del alfa-cetoácido a un aminoácido;
- 40 (ii) el método de (i), en donde el polipéptido de la reivindicación 4 comprende catalizar la conversión de una mezcla alfa-ceto racémica a un D- o L-aminoácido sustancialmente puro ópticamente; o

(iii) el método de (i) o (ii), en donde el oxaloacetato es convertido a un aspartato, o el alfa-cetoglutarato es convertido a glutarato, o un alfa-cetoisovalerato es convertido a una L-valina; o la actividad de transaminasa es una actividad de transaminasa omega que cataliza la conversión de isobutilamina a isobutilaldehído.

14. Un método para catalizar la conversión de una amina a una cetona que comprende

5 (i)

(a) proveer una amina;

(b) proveer el polipéptido de la reivindicación 4; y

10 (c) poner en contacto el polipéptido de (b) con la amina bajo condiciones en donde el polipéptido cataliza la conversión de una amina a una cetona, en donde la amina no es o no proviene de un triptófano (con la condición de que el segundo aminoácido no es triptófano, o con la condición de que la amina no es o no proviene de un triptófano);

(ii) el método de (i), en donde la actividad de transaminasa comprende catalizar la conversión de una amina quiral a una cetona; o

(iii) el método de (i) o (ii), en donde la amina es una ω -amina.

15 15. Un método para catalizar la síntesis de un aminoácido que comprende

(i)

(a) proveer un aminoácido y un cetoácido, donde el aminoácido no es triptófano;

(b) proveer el polipéptido de transaminasa de la reivindicación 4; y

20 (c) poner en contacto el polipéptido de (b) con el aminoácido y el cetoácido bajo condiciones en donde se produce un segundo aminoácido y un piruvato, en donde el segundo aminoácido no es triptófano (con la condición de que el segundo aminoácido no es triptófano); o

(ii) el método de (i), que comprende adicionalmente hacer reaccionar el piruvato con una enzima acetolactato sintasa bajo condiciones apropiadas para producir un compuesto que no reacciona con la enzima transaminasa;

25 (iii) el método de (ii), en donde el compuesto que no reacciona con la enzima transaminasa es acetolactato o acetoina; o, el primer aminoácido es alanina o L-aspartato; o, el cetoácido es 2-cetobutirato o trimetil piruvato; o, el segundo aminoácido es 2-aminobutirato o tert-leucina.

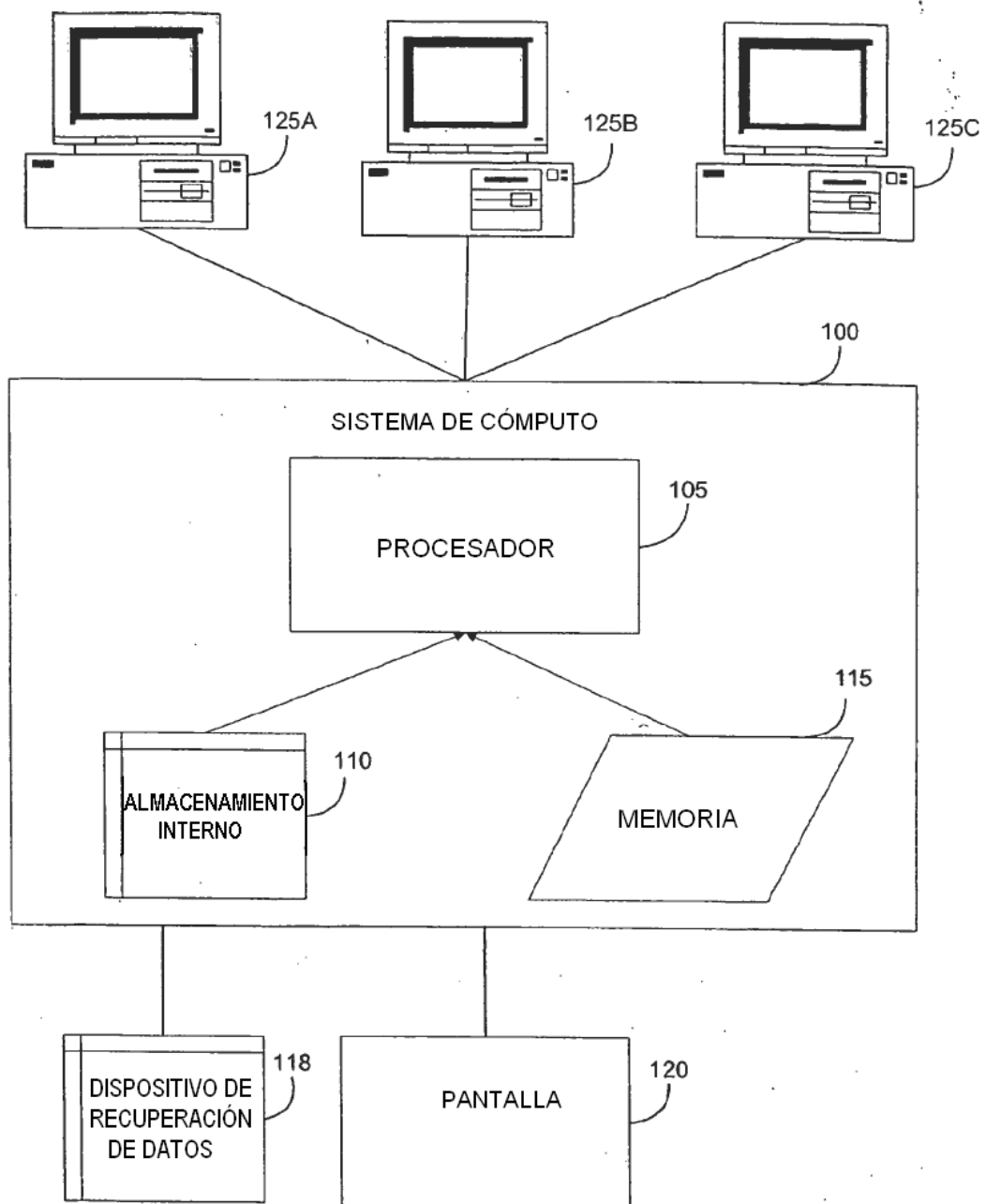
16. El polipéptido de la reivindicación 4, en donde el polipéptido carece de una secuencia de señalización, carece de un preprodominio, o carece de un dominio de enlazamiento.

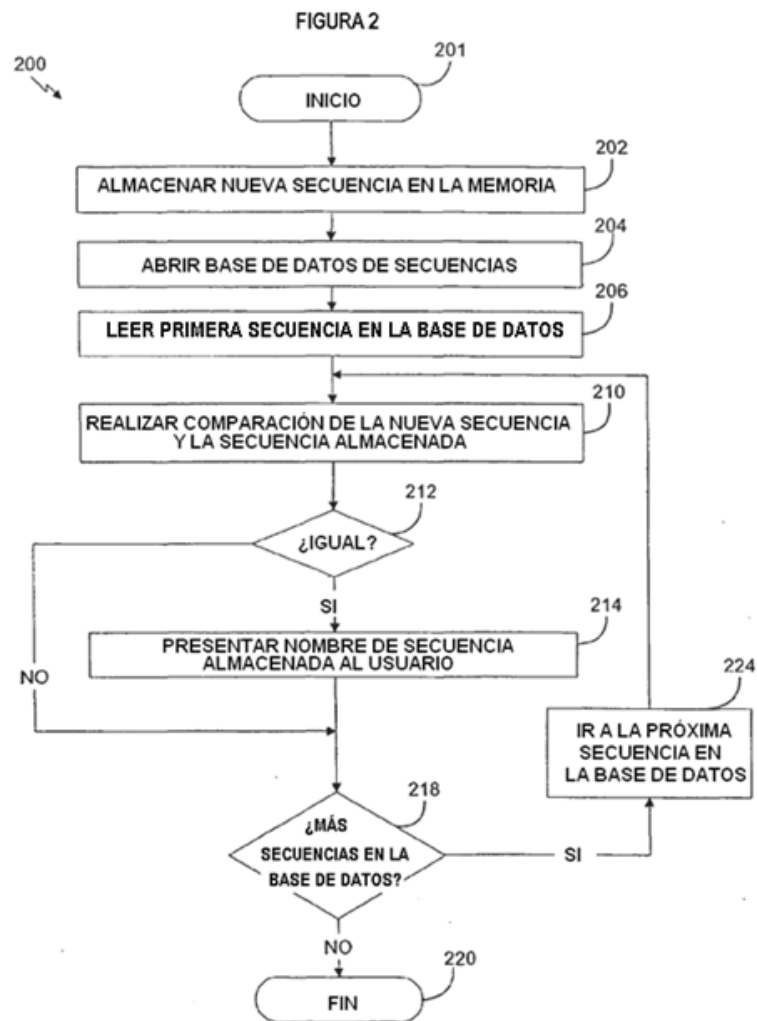
30 17. El polipéptido de la reivindicación 4, en donde el polipéptido comprende adicionalmente una secuencia heteróloga.

18. El polipéptido de la reivindicación 17, en donde el polipéptido comprende al menos un sitio de glicosilación o comprende adicionalmente un polisacárido, en donde opcionalmente la glicosilación es una glicosilación enlazada a N, y opcionalmente el polipéptido es glicosilado después de ser expresado en una *P.pastoris* o en una *S.pombe*.

35 19. El polipéptido de la reivindicación 4, en donde la secuencia de aminoácidos es un fragmento enzimáticamente activo de SEQ ID NO: 220 y el fragmento tiene una actividad de D-aminoácido transferasa.

FIGURA 1





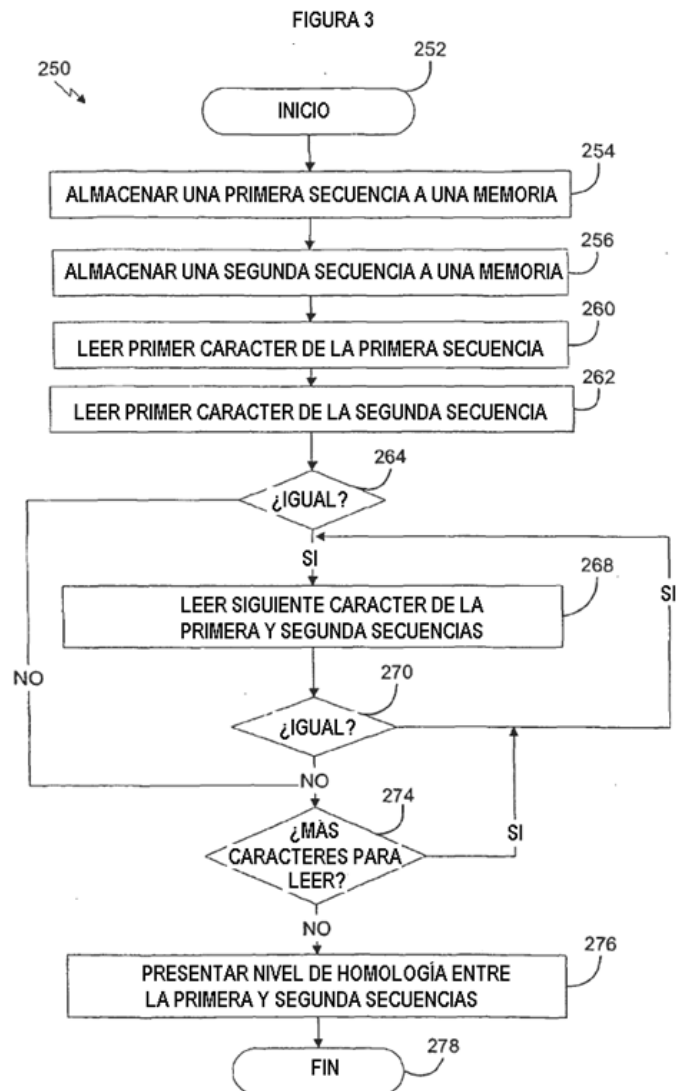


FIGURA 4

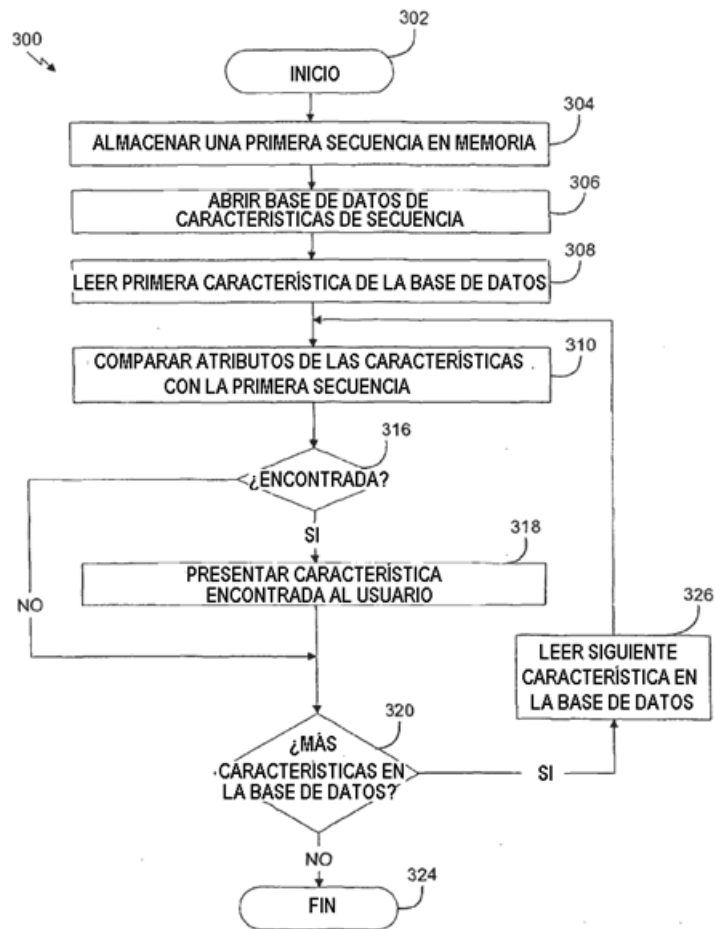


Figura 5

SEQ ID NO: 894	MDALGYNGKKGWGLDEMTPMNDRCGFGDGVYDATIANGVIFALDEHIDRFINSAKLLEIEIGFTKEELKKTFFEM--
SEQ ID NO: 1066	MENLGYNGKFGLLLEMTVPMLDRCVYFGDGVYDATYSRNHKKIFALEEHIDRFYNSAGLLGKIPYSKEQVKEILKEM--
SEQ ID NO: 1064	MKDLYNGGEYDLIENMKIPMNDRCVYFGDGVYDATYSRNHNI FALDEHIDRFYNSAELLRIKIPYTKKEMKELKDM--
SEQ ID NO: 1068	MKQVGYNGGTIADLNLKVPATDRALYFGDGCYDATTFKNNVAFALDEHLDRCFYNSCRLLLEIDFPLNRDELKEKLYAVID
SEQ ID NO: 894	HSKVDKGVYMWYVWQATRGTRRRSHVFPAG--P--S-NLWIMIKPNHVDDLYRKIKLITMEDTRFLHCNICKTLNLLIPNVIA
SEQ ID NO: 1066	VLKVDSEGEQFVYWQITRGTMGRNHAFPGDEVP--S-NLWIMLKLPLNIDMSQKLLITLEDTRFLHCNICKTLNLLPSVIA
SEQ ID NO: 1064	VKKVDSEGEQFVYWQVTRGTGMRNHAFLESE--DKVA-NIWIIVLKLPLKVKDMSKKLLITLEDTRFLHCNICKTLNLLPSVIA
SEQ ID NO: 1068	ANEVDTGI--LYWQTSRSGSLRNHIFPED--S--QPNLLIFTAPYGLVPFDTEYKLI SREDTRFLHCNICKTLNLLIPNVIA
SEQ ID NO: 894	SQRALEAGCHEAVFHRGETVTECAHSNVHIIKNGRFITHQADNLIILRGIAARSHLLQACIRLNI PFDEREFTLSELFDADE
SEQ ID NO: 1066	SQKTEEAGCQAEVFRGDRVTECAHSNVSI IKDGILKTAPTDNLILPGIARAHLIKMKCKSFNIPVDETAFTLKELMEADE
SEQ ID NO: 1064	AQKTEEAGCQAEVFRGDRVTECAHSNVSI IKDEILKTAPTDNLILPGIARAHLIKMKCKFEI PVDETPFTLKELINADE
SEQ ID NO: 1068	SQKANESHCEQEVVFRGDRVTECAHSNIIILKDGVLCSPPRDNLILPGITLKHLLQLAKENNIPTSEAPFTMDDLRNADE
SEQ ID NO: 894	ILVSSSGTGLSANTIDGKNVGGKAPELLKKIQGEVIREFIEATGYTPPEWSTV*
SEQ ID NO: 1066	VIVTSSGQFCMATSEIDGIPVGGKAPELVKKLQDALINEFLEETKTE-----
SEQ ID NO: 1064	VIVTSSGQFCMTACEIDGRPVGGKAPDIIKKLQDALINEFLEETN-----
SEQ ID NO: 1068	VIVSSSACLGI RAVELDGQPVGGKDGKTLKILQDAYAKYNAETVSR-----

Figura 6 (1/2)

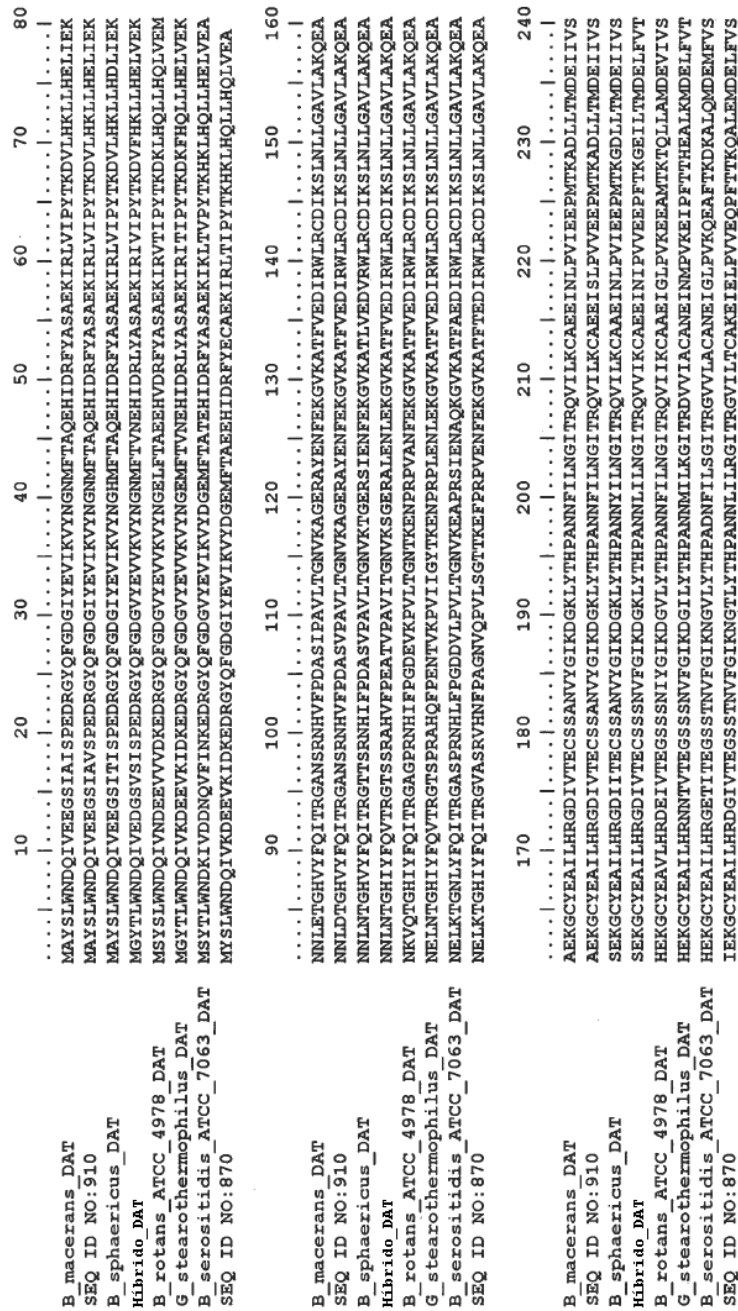


Figura 6 (2/2)

B_macerans_DAT	250	260	270	280	(SEQ ID NO:1126)
SEQ ID NO:910				
B_sphaericus_DAT	SVSSEVTPVIDVDGNQIGAGVPGEWTRQLQQSFEAKLPLSMNTK-----				
Hibrido_DAT	SVSSEVTPVIDVDGNQIGAGVPGEWTRKLQQAFAEAKLPLSLNTK-----				
B_rotans_ATCC_4978_DAT	SVSSEVTPVIDVDGQQIGAGVPGEWTRKLQKAFAEAKLPLSINA-----				(SEQ ID NO:1127)
G_stearothermophilus_DAT	SVTSEITPVIDIDGNQIGAGVPGEWTRKLQKAFAEAKIPLSLNS-----				(SEQ ID NO:1128)
B_serositidis_ATCC_7063_DAT	STTSEVTPIIDIDGTIVIGAGKPGDWTRKLQAQFDTKIPKGIRA-----				(SEQ ID NO:1129)
SEQ ID NO:870	STTSEITPVIDLDGVAINGGEIGEWTRKLQKQFETKIPKPLHI-----				(SEQ ID NO:1130)
	STTSEITPVIDDGVKINNGVIGEWTRKLQKQFETKIP-TPFTNK-----				(SEQ ID NO:1131)

SEQ	ID	NO: 946
SEQ	ID	NO: 894
SEQ	ID	NO: 892
SEQ	ID	NO: 220
SEQ	ID	NO: 176
SEQ	ID	NO: 1064
SEQ	ID	NO: 1066
SEQ	ID	NO: 1068
SEQ	ID	NO: 946
SEQ	ID	NO: 894
SEQ	ID	NO: 892
SEQ	ID	NO: 220
SEQ	ID	NO: 176
SEQ	ID	NO: 1064
SEQ	ID	NO: 1066
SEQ	ID	NO: 1068
SEQ	ID	NO: 946
SEQ	ID	NO: 894
SEQ	ID	NO: 892
SEQ	ID	NO: 220
SEQ	ID	NO: 176
SEQ	ID	NO: 1064
SEQ	ID	NO: 1066
SEQ	ID	NO: 1068

Figura 7 (2/2)

SEQ ID NO: 946	250	260	270	280
SEQ ID NO: 894	SGTLGLSADTIDGKNVGGKAPPELLKKRIQDEVLKEFIEATARAFFG--				
SEQ ID NO: 892	SGTLGLSANTIDGKNVGGKAPPELLKKIQGEVLRREFIEATGYTPWSTV				
SEQ ID NO: 220	SGTFGLSADTIDGKSVGGKAPPELLKKIQDEVLMREFIEATGYTPWRKA				
SEQ ID NO: 176	SGTLGLSAEEIDGKKAGGKAPPELLKKIQDEVLRREFIEATGYTPWSRV				
SEQ ID NO: 1064	SGTLGLSADTIDGKAVGGKAPPELLKKIQDEVLRREFTEVTGYKPAWSRV				
SEQ ID NO: 1066	SGQFCMTACEIDGRPVGGKAPDIKKLQTALLNEFLEETN-----				
SEQ ID NO: 1068	SGQFCMATSEIDGIPVGGKAPELVKKLQDALINEFLEETKTE-----				
	SACLGIRAVELDGQPVGGKDGKTLKILQDAYAKKYNAEVTVSR-----				

Figura 8

