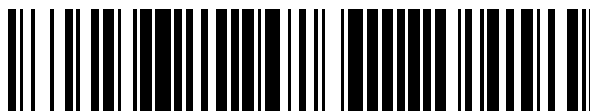


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 383 767**

51 Int. Cl.:  
**C12N 9/10** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **07758102 .3**  
96 Fecha de presentación: **07.03.2007**  
97 Número de publicación de la solicitud: **1999256**  
97 Fecha de publicación de la solicitud: **10.12.2008**

54 Título: **Aldolasas, ácidos nucleicos que las codifican y métodos para producir y usar las mismas**

30 Prioridad:  
**07.03.2006 US 780515 P**

45 Fecha de publicación de la mención BOPI:  
**26.06.2012**

45 Fecha de la publicación del folleto de la patente:  
**26.06.2012**

73 Titular/es:  
**VERENIUM CORPORATION  
INTELLECTUAL PROPERTY DEPARTMENT P.O.  
BOX 910550  
SAN DIEGO CA 92191-0550, US**

72 Inventor/es:  
**BURKE, Ellen;  
HICKS, Paula M.;  
LUGINBUHL, Peter;  
RICHARDSON, Toby;  
WEINER, David P. y  
ZHAO, Lishan**

74 Agente/Representante:  
**de Elzaburu Márquez, Alberto**

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).

ES 2 383 767 T3

**DESCRIPCIÓN**

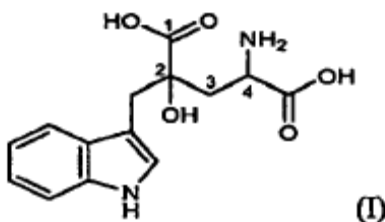
Aldolasas, ácidos nucleicos que las codifican y métodos para producir y usar las mismas

**Campo de acuerdo con la invención**

5 La presente invención se refiere a la biología molecular y celular y a la bioquímica. Más específicamente, la invención se refiere a polipéptidos que tienen actividad de aldolasa, polinucleótidos que codifican estos polipéptidos y métodos para preparar y utilizar polinucleótidos y polipéptidos.

**Antecedentes de la invención**

La monatina es un edulcorante de alta intensidad que tiene la fórmula química:



10 La monatina incluye dos centros quirales que conducen a cuatro posibles configuraciones estereoisoméricas. La configuración R,R (el "estereoisómero R,R" o "la monatina R,R"); la configuración S,S (el "estereoisómero S,S" o "la monatina S,S"); la configuración R,S (el "estereoisómero R,S" o "la monatina R,S"); y la configuración S,R (el "estereoisómero S,R" o "la monatina S,R"). Tal como se utiliza en la presente, a menos que se indique de otra forma, el término "monatina" se usa para referirse a composiciones que incluyen los cuatro estereoisómeros de monatina, composiciones que incluyen cualquier combinación de estereoisómeros de monatina (tal como una composición que incluye solo los estereoisómeros R,R y S,S de monatina), así como también una sola forma isomérica.

15 A los efectos de esta divulgación, la estructura principal de carbono se numerará como se ilustró anteriormente, siendo el carbono unido directa y covalentemente al grupo alcohol identificado como el carbono en la posición 2 y el carbono unido directa y covalentemente al grupo amino identificado como el carbono en la posición 4. Consecuentemente, las referencias en la presente una monatina R,R, monatina S,S, monatina R,S y monatina S,R significan: monatina 2R,4R, monatina 2S,4S, monatina 2R,4S y monatina 2S,4R, respectivamente, a menos que se indique de otra forma.

20 Cabe señalar que, en la bibliografía, la estructura principal de carbono de la monatina también se ha numerado usando una convención alternativa, siendo el carbono unido al grupo alcohol el carbono en la posición 4 y el carbono unido al grupo amino el carbono en la posición 2. Por lo tanto, por ejemplo, las referencias a la monatina 2S,4R en esta divulgación serían iguales a las referencias a la monatina 2R,4S en la bibliografía usando la convención de numeración alternativa.

25 Más aun, debido a las distintas convenciones de designación, la monatina se conoce con varios nombres químicos alternativos, entre los que se incluyen: ácido 2-hidroxi-2-(indol-3-ilmetil)-4-aminoglutarico; ácido 4-amino-2-hidroxi-2-(1H-indol-3-ilmetil)-pentanodioico; ácido 4-hidroxi-4-(3-indolilmetil)glutámico y 3-(1-amino-1,3-dicarboxi-3-hidroxi-but-4-il)indol.

30 Ciertas formas isoméricas de monatina pueden encontrarse en la corteza de las raíces de la planta *Sclerochiton ilicifolius* que se encuentra en la región de Transvaal en Sudáfrica. Las Solicitudes de Patente de los Estados Unidos Nos. 10/422.366 ("la solicitud '366") y 10/979.821 ("la solicitud '821") divulgan, entre otras cosas, polipéptidos, rutas y microorganismos para la producción in vitro e in vivo de monatina.

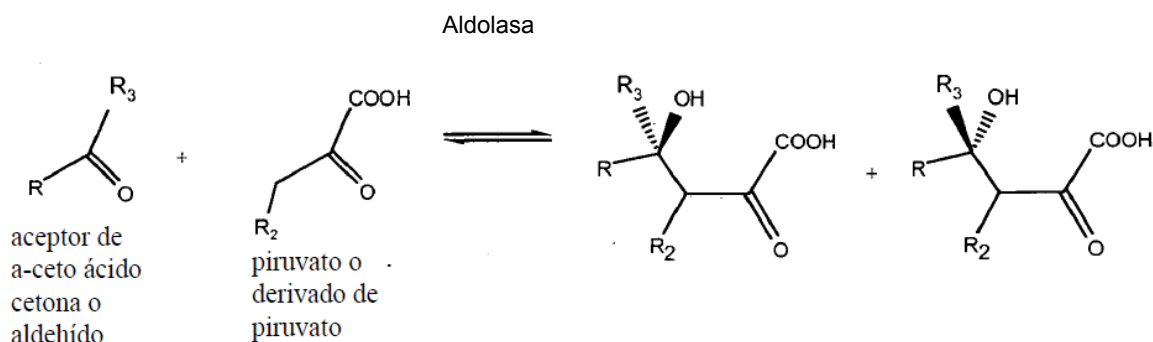
**Resumen**

35 La invención proporciona polipéptidos que tienen actividad de aldolasa (de aquí en adelante "aldolasas"), incluyendo actividad de piruvato aldolasa tal como, a modo no taxativo, actividad de HMG y KHG aldolasa, polinucleótidos que codifican los polipéptidos y métodos para preparar y utilizar los polipéptidos y polinucleótidos. En algunas realizaciones, la invención también proporciona composiciones que comprenden los polipéptidos de acuerdo con la invención. Estas composiciones pueden formularse en una variedad de formas, tales como comprimidos, geles, píldoras, implantes, líquidos, aerosoles, películas, micelas, polvos, alimento, granulados de pienso o cualquier tipo de forma encapsulada.

40 En algunas realizaciones, las aldolasas y/o composiciones de las mismas pueden ser útiles en contextos farmacéuticos, industriales y/o agrícolas.

En algunas realizaciones, las aldolasas y/o composiciones de las mismas pueden ser útiles para formar o escindir enlaces carbono-carbono.

- 5 En algunas realizaciones se proporcionan aldolasas que catalizan reacciones que forman enlaces carbono-carbono entre un aceptor de alfa-cetoácido y un piruvato o un donante derivado de piruvato (ver esquema de reacción más adelante). En algunas realizaciones, el aceptor también puede ser una cetona o un aldehído. En algunas realizaciones se proporcionan aldolasas que tienen actividad 4-hidroxi-2-oxoglutarato aldolasa (tal como 2-ceto-4-hidroxi-2-oxoglutarato aldolasa, 2-oxo-4-hidroxi-2-oxoglutarato aldolasa, KHG-aldolasa, EC 4.1.3.16) y catalizan la siguiente reacción: 4-hidroxi-2-oxoglutarato  $\rightleftharpoons$  piruvato + glioxilato. En algunas realizaciones se proporcionan aldolasas que tienen actividad HMG-aldolasa (tal como 4-hidroxi-4-metil-2-oxoglutarato aldolasa, piruvato aldolasa, aldolasa gamma-metil-gamma-hidroxi-alfa-cetoglutárica, 4-hidroxi-4-metil-2-cetoglutarato aldolasa, EC 4.1.3.17) y catalizan la siguiente reacción: 4-hidroxi-4-metil-2-oxoglutarato  $\rightleftharpoons$  2-piruvato. Una HMG aldolasa también actuará sobre 4-hidroxi-4-metil-2-oxoadipato y 4-carboxi-4-hidroxi-2-oxohexadioato.



- 15 R = H, alquilo, alquilo sustituido, arilo, arilo sustituido, bencilo, bencilo sustituido

R<sub>2</sub> = H, alquilo, alquilo sustituido, arilo, arilo sustituido, bencilo, bencilo sustituido

R<sub>3</sub> = H, alquilo, alquilo sustituido, arilo, arilo sustituido, bencilo, bencilo sustituido, ácido carboxílico.

- 20 En algunas realizaciones se proporcionan aldolasas, tales como una piruvato aldolasa, tal como, a modo no taxativo, una HMG y/o una KHG aldolasa, que facilitan la producción de un 2-ceto-glutarato 3,4-sustituido. En una realización, la invención proporciona un método para preparar un 2-ceto-glutarato 3,4-sustituido que comprende: (a) proporcionar un polipéptido que tiene una actividad de aldolasa, tal como una actividad de piruvato aldolasa, tal como, a modo no taxativo, una actividad de HMG aldolasa y/o KMG aldolasa; (b) proporcionar un donante y un compuesto aceptor; y (c) poner en contacto el polipéptido de la etapa (a) con los compuestos de la etapa (b) en condiciones en las cuales la aldolasa cataliza la síntesis de un 2-ceto-glutarato 3,4-sustituido, en donde
- 25 opcionalmente el donante y el aceptor son un piruvato o un donante de piruvato y un aceptor  $\alpha$ -cetoácido, una cetona y/o un aldehído.

- 30 En algunas realizaciones se proporcionan aldolasas que facilitan la producción de ácido R-2-hidroxi 2-(indol-3-ilmetil)-4-cetoglutárico (R-MP), un precursor de monatina. En algunas realizaciones, una piruvato aldolasa, tal como una HMG y/o una KHG aldolasa, puede utilizarse junto con una D-aminotransferasa para preparar un ácido D-glutámico 4-sustituido o un derivado del mismo. Un ácido D-glutámico 4-sustituido y/o un derivado del mismo pueden utilizarse como un antibiótico, dado que se ha descubierto que estos compuestos inhiben la glutamato racemasa bacteriana (WO0214261A3). En una realización, la invención proporciona un método para preparar un ácido D-glutámico 4-sustituido que comprende: (a) proporcionar un polipéptido que tiene una actividad de aldolasa, tal como una actividad de piruvato aldolasa, tal como, a modo no taxativo, una actividad de HMG aldolasa y/o KMG aldolasa;
- 35 (b) proporcionar un aceptor  $\alpha$ -cetoácido y un piruvato o un donante de piruvato; y (c) poner en contacto el polipéptido de la etapa (a) con los compuestos de la etapa (b) en condiciones en las cuales la aldolasa cataliza la síntesis de un ácido D-glutámico 4-sustituido, en donde opcionalmente el polipéptido tiene actividad de piruvato aldolasa, HMG aldolasa y/o KHG aldolasa y en donde opcionalmente el método además comprende el uso de una D-aminotransferasa.

- 40 La invención proporciona ácidos nucleicos aislados, sintéticos o recombinantes que comprenden una secuencia de ácidos nucleicos que tiene al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más, o 100% (completa) de identidad de secuencia con respecto a un ácido nucleico de SEQ ID NO:275. En algunas realizaciones, uno o más ácidos nucleicos codifican al menos un polipéptido que tiene una actividad de aldolasa, incluyendo actividad de piruvato tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa. En algunas realizaciones, las identidades de secuencia se determinan mediante análisis con un algoritmo de comparación de secuencias o mediante una inspección visual.
- 45

En realizaciones alternativas, uno o más ácidos nucleicos codifican al menos un polipéptido capaz de generar un anticuerpo que puede unirse específicamente a un polipéptido de la invención, o estos ácidos nucleicos pueden utilizarse como sondas para identificar o aislar ácidos nucleicos que codifican aldolasa o para inhibir la expresión de ácidos nucleicos que expresan aldolasa.

5 Ácidos nucleicos de acuerdo con la invención también incluyen ácidos nucleicos aislados, sintéticos o recombinantes que codifican enzimas de acuerdo con la invención, tales como enzimas que incluyen uno o más polipéptidos que tienen una secuencia tal como se establece en la SEQ ID NO:276, y fragmentos enzimáticamente activos de la misma. En algunas realizaciones, el polipéptido tiene una actividad de aldolasa, incluyendo actividad de piruvato tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa.

10 En algunas realizaciones, la invención proporciona ácidos nucleicos que codifican aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa preferiblemente derivados de cultivos mixtos. En algunas realizaciones, la invención proporciona ácidos nucleicos que codifican una enzima que forma o escinde enlaces carbono-carbono aislados de cultivos mixtos que comprenden polinucleótidos de acuerdo con la invención, tales como una secuencia que tiene al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%,  
15 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más, o 100% (completa) de identidad de secuencia con respecto a un ácido nucleico de SEQ ID NO:275, en la longitud total de un ADNc, transcripto (ARNm) o gen.

En algunas realizaciones, la invención proporciona ácidos nucleicos que codifican enzima aldolasa, tal como enzima piruvato aldolasa, enzima HMG y/o KHG, incluyendo secuencias de polinucleótidos de acuerdo con la invención y los polipéptidos que las codifican, incluyendo enzimas de acuerdo con la invención, tales como polipéptidos de  
20 acuerdo con la invención, tales como SEQ ID NO:276, y fragmentos enzimáticamente activos de las misma, preferiblemente derivados de una fuente común, tal como una fuente ambiental. En algunas realizaciones, la invención también proporciona ácidos nucleicos que codifican enzima aldolasa, tal como enzima piruvato aldolasa, enzima HMG y/o KHG preferiblemente derivados de fuentes ambientales, tales como fuentes ambientales mixtas.

En algunas realizaciones, el algoritmo de comparación de secuencias es un algoritmo BLAST versión 2.2.2, en el  
25 que se fija una configuración de filtro de blastall -p blastp -d "nr pataa" -F F y todas las demás opciones se configuran con los valores por defecto.

La presente divulgación además describe ácidos nucleicos aislados, sintéticos o recombinantes que incluyen al  
30 menos 10, 15, 20, 25, 30, 35, 40, 45, 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, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2200, 2250, 2300, 2350, 2400, 2450, 2500 o más bases consecutivas de una secuencia de ácidos nucleicos básicamente idénticas a los mismos y las secuencias complementarias a los mismos.

En algunas realizaciones, los ácidos nucleicos aislados, sintéticos o recombinantes de acuerdo con la invención  
35 codifican un polipéptido que tiene una actividad de aldolasa, incluyendo actividad de piruvato tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa, que es termoestable. El polipéptido termoestable de acuerdo con la invención puede retener una actividad de aldolasa, tal como una actividad de piruvato aldolasa, tal como una actividad de HMG y/o KHG aldolasa, en condiciones que comprenden un rango de temperatura de aproximadamente -100°C a aproximadamente -80°C, aproximadamente -80°C a aproximadamente -40°C,  
40 aproximadamente -40°C a aproximadamente -20°C, aproximadamente -20°C a aproximadamente 0°C, aproximadamente 0°C a aproximadamente 37°C, aproximadamente 0°C a aproximadamente 5°C, aproximadamente 5°C a aproximadamente 15°C, aproximadamente 15°C a aproximadamente 25°C, aproximadamente 25°C a aproximadamente 37°C, aproximadamente 37°C a aproximadamente 45°C, aproximadamente 45°C a aproximadamente 55°C, aproximadamente 55°C a aproximadamente 70°C, aproximadamente 70°C a aproximadamente 75°C, aproximadamente 75°C a aproximadamente 85°C, aproximadamente 85°C a  
45 aproximadamente 90°C, aproximadamente 90°C a aproximadamente 95°C, aproximadamente 95°C a aproximadamente 100°C, aproximadamente 100°C a aproximadamente 105°C, aproximadamente 105°C a aproximadamente 110°C, aproximadamente 110°C a 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. Los polipéptidos termoestables de acuerdo con la invención pueden retener una actividad de aldolasa, tal como una actividad de piruvato aldolasa, tal como una actividad de HMG y/o KHG aldolasa, en temperaturas en el rango de aproximadamente -100°C a aproximadamente -80°C, aproximadamente -80°C a aproximadamente -40°C, aproximadamente -40°C a aproximadamente -20°C, aproximadamente -20°C a aproximadamente 0°C, aproximadamente 0°C a aproximadamente 5°C, aproximadamente 5°C a aproximadamente 15°C, aproximadamente 15°C a aproximadamente 25°C, aproximadamente 25°C a aproximadamente 37°C, aproximadamente 37°C a  
50 aproximadamente 45°C, aproximadamente 45°C a aproximadamente 55°C, aproximadamente 55°C a aproximadamente 70°C, aproximadamente 70°C a aproximadamente 75°C, aproximadamente 75°C a aproximadamente 85°C, aproximadamente 85°C a aproximadamente 90°C, aproximadamente 90°C a aproximadamente 95°C, aproximadamente 95°C a aproximadamente 100°C, aproximadamente 100°C a aproximadamente 105°C, aproximadamente 105°C a aproximadamente 110°C, aproximadamente 110°C a  
55 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,  
60 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 una actividad de aldolasa 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 otras realizaciones, los ácidos nucleicos aislados, sintéticos o recombinantes codifican un polipéptido que tiene una actividad de aldolasa, incluyendo actividad de piruvato tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa, que es termotolerante. Los polipéptidos termotolerantes de acuerdo con la invención pueden retener una actividad de aldolasa, tal como una actividad de piruvato aldolasa, tal como una actividad de HMG y/o KHG aldolasa, luego de la exposición a condiciones que comprenden una temperatura en el rango de aproximadamente -100°C a aproximadamente -80°C, aproximadamente -80°C a aproximadamente -40°C, aproximadamente -40°C a aproximadamente -20°C, aproximadamente -20°C a aproximadamente 0°C, aproximadamente 0°C a aproximadamente 5°C, aproximadamente 5°C a aproximadamente 15°C, aproximadamente 15°C a aproximadamente 25°C, aproximadamente 25°C a aproximadamente 37°C, aproximadamente 37°C a aproximadamente 45°C, aproximadamente 45°C a aproximadamente 55°C, aproximadamente 55°C a aproximadamente 70°C, aproximadamente 70°C a aproximadamente 75°C, aproximadamente 75°C a aproximadamente 85°C, aproximadamente 85°C a aproximadamente 90°C, aproximadamente 90°C a aproximadamente 95°C, aproximadamente 95°C a aproximadamente 100°C, aproximadamente 100°C a aproximadamente 105°C, aproximadamente 105°C a aproximadamente 110°C, aproximadamente 110°C a 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. Los polipéptidos termotolerantes de acuerdo con la invención pueden retener una actividad de aldolasa, tal como una actividad de piruvato aldolasa, tal como una actividad de HMG y/o KHG aldolasa, luego de la exposición a una temperatura en el rango de aproximadamente -100°C a aproximadamente -80°C, aproximadamente -80°C a aproximadamente -40°C, aproximadamente -40°C a aproximadamente -20°C, aproximadamente -20°C a aproximadamente 0°C, aproximadamente 0°C a aproximadamente 5°C, aproximadamente 5°C a aproximadamente 15°C, aproximadamente 15°C a aproximadamente 25°C, aproximadamente 25°C a aproximadamente 37°C, aproximadamente 37°C a aproximadamente 45°C, aproximadamente 45°C a aproximadamente 55°C, aproximadamente 55°C a aproximadamente 70°C, aproximadamente 70°C a aproximadamente 75°C, aproximadamente 75°C a aproximadamente 85°C, aproximadamente 85°C a aproximadamente 90°C, aproximadamente 90°C a aproximadamente 95°C, aproximadamente 95°C a aproximadamente 100°C, aproximadamente 100°C a aproximadamente 105°C, aproximadamente 105°C a aproximadamente 110°C, aproximadamente 110°C a 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 una actividad de aldolasa luego de la 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.
- La invención proporciona ácidos nucleicos aislados, sintéticos o recombinantes que comprenden una secuencia que hibrida en condiciones estrictas un ácido nucleico que comprende la secuencia tal como se establece en SEQ ID NO:275 o fragmentos enzimáticamente activos o subsecuencias de los mismos. En algunas realizaciones, los ácidos nucleicos codifican polipéptidos que tienen una actividad de aldolasa, incluyendo actividad de piruvato tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa. Los ácidos nucleicos pueden ser de al menos aproximadamente 10, 15, 20, 25, 30, 35, 40, 45, 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 o más residuos de longitud o la longitud completa del gen o transcripto. En algunas realizaciones, las condiciones estrictas 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 divulgación proporciona sondas de ácido nucleico para identificar o aislar ácidos nucleicos que codifican polipéptidos que tienen una actividad de aldolasa, incluyendo actividad de piruvato tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa, en donde las sondas comprenden aproximadamente 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000 o más bases consecutivas de una secuencia de acuerdo con la invención, y en donde las sondas identifican el ácido nucleico mediante unión o hibridación. Las sondas pueden comprender un oligonucleótido que comprende entre aproximadamente 10-100 bases consecutivas de una secuencia de acuerdo

con la invención o fragmentos o subsecuencias de la misma, por ejemplo, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 bases o más, o cualquier longitud intermedia que se desee.

5 La divulgación proporciona sondas de ácido nucleico para identificar o aislar ácidos nucleicos que codifican polipéptidos que tienen una actividad de aldolasa, incluyendo actividad de piruvato tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa, en donde las sondas comprenden ácidos nucleicos que comprenden una secuencia de al menos aproximadamente 10, 15, 20, 30, 40, 50, 60, 70, 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 un ácido nucleico de acuerdo con la invención, tal como un polinucleótido que tiene al menos aproximadamente 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más o 100% (completa) de identidad de secuencia con respecto a un ácido nucleico de la invención. En algunos casos, las identidades de secuencia se determinan mediante análisis con un algoritmo de comparación de secuencias o mediante inspección visual. En ciertos casos, las sondas pueden comprender un oligonucleótido que comprende entre al menos aproximadamente 10-100 bases consecutivas de una secuencia de ácidos nucleicos de acuerdo con la invención o una subsecuencia de la misma, por ejemplo 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 bases o más, o cualquier longitud intermedia que se desee.

15 La invención proporciona pares de cebadores de amplificación para amplificar (tal como mediante PCR) ácidos nucleicos que codifican polipéptidos que tienen actividad de aldolasa, incluyendo actividad de piruvato tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa, en donde cada par de cebadores es capaz de amplificar un ácido nucleico que comprende una secuencia de acuerdo con la invención o fragmentos o subsecuencias de la misma (ver el Listado de Secuencias). Uno o cada miembro del par de secuencias cebadoras de amplificación puede comprender un oligonucleótido que comprende al menos aproximadamente 10 a 50, o más, 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, 36 o más bases consecutivas de la secuencia. En algunas realizaciones, la invención proporciona pares de cebadores de amplificación, en donde cada par de cebadores comprende un primer miembro que tiene una secuencia tal como fue establecido en aproximadamente los primeros (5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36 o más residuos de un ácido nucleico de acuerdo con la invención, y un segundo miembro que tiene una secuencia tal como fue establecido en aproximadamente los primeros (5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36 o más residuos de la hebra complementaria del primer miembro.

20 La presente divulgación además describe ácidos nucleicos que codifican aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa generados por amplificación, tal como reacción en cadena de la polimerasa (PCR), utilizando un par de cebadores de amplificación de acuerdo con la invención. En algunos casos, los ácidos nucleicos que codifican aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa se generan por amplificación, tal como reacción en cadena de la polimerasa (PCR), utilizando un par de cebadores de amplificación de acuerdo con la invención. En algunos casos, la invención proporciona métodos para preparar una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa mediante amplificación, tal como reacción en cadena de la polimerasa (PCR), utilizando un par de cebadores de amplificación de acuerdo con la invención. En algunos casos, el par de cebadores de amplificación amplifica un ácido nucleico de una biblioteca, tal como una biblioteca genómica, tal como una biblioteca ambiental.

25 La invención proporciona métodos para amplificar un ácido nucleico que codifica un polipéptido que tiene una actividad de aldolasa, incluyendo actividad de piruvato tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa que comprenden la amplificación de un ácido nucleico plantilla con un par de secuencias cebadoras de amplificación capaces de amplificar una secuencia de ácidos nucleicos de acuerdo con la invención o fragmentos o subsecuencias de la misma.

30 La invención proporciona cassettes de expresión que comprenden un ácido nucleico de acuerdo con la invención o una subsecuencia del mismo. En algunas realizaciones, el cassette de expresión puede comprender el ácido nucleico que se encuentra operativamente ligado a un promotor. El promotor puede ser un promotor viral, bacteriano, mamífero, fúngico, de levadura o vegetal. En algunas realizaciones, el promotor vegetal puede ser un promotor de patata, arroz, maíz, trigo, tabaco o cebada. El promotor puede ser un promotor constitutivo. El promotor constitutivo puede comprender CaMV35S. En otras realizaciones, el promotor puede ser un promotor inducible. En algunas realizaciones, el promotor puede ser un promotor específico de tejido o un promotor regulado por el ambiente o por el desarrollo. De esta forma, el promotor puede ser, por ejemplo, un promotor específico de semillas, específico de hojas, específico de raíces o específico de tallos o un promotor inducido por abscisión. En algunas realizaciones, el cassette de expresión además puede comprender un vector vegetal o de expresión de virus vegetal.

35 La invención proporciona vehículos de clonación que comprenden un cassette de expresión (tal como un vector) de acuerdo con la invención o un ácido nucleico de acuerdo con la invención. El vehículo de clonación puede ser un vector viral, un plásmido, un fago, un fagémido, un cósmido, un fósmido, un bacteriófago o un cromosoma artificial. El vector viral puede comprender un vector de adenovirus, un vector retroviral o un vector viral adeno-asociado. El

vehículo de clonación puede comprender un cromosoma artificial bacteriano (BAC), un plásmido, un vector derivado de bacteriófago P1 (PAC), un cromosoma artificial de levadura (YAC) o un cromosoma artificial de mamífero (MAC).

5 La invención proporciona células transformadas que comprenden ácidos nucleicos de acuerdo con la invención o cassettes de expresión (tales como vectores) de acuerdo con la invención o vehículos de clonación de acuerdo con la invención. En algunas realizaciones, la célula transformada puede ser una célula bacteriana, una célula de mamífero, una célula fúngica, una célula de levadura, una célula de insecto o una célula vegetal. En algunas realizaciones, la célula vegetal puede ser célula de soja, colza, semilla oleaginosa, tomate, caña de azúcar, un cereal, una patata, trigo, arroz, maíz, tabaco o cebada.

10 La invención proporciona animales no humanos transgénicos que comprenden un ácido nucleico de acuerdo con la invención o un cassette de expresión (tal como un vector) de acuerdo con la invención. En algunas realizaciones, el animal es un ratón, una rata, un cerdo, una cabra o una oveja.

15 La invención proporciona plantas transgénicas que comprenden un ácido nucleico de acuerdo con la invención o un cassette de expresión (tal como un vector) de acuerdo con la invención. La planta transgénica puede ser una planta de cereal, una planta de maíz, una planta de patata, una planta de tomate, una planta de trigo, una planta de semilla oleaginosa, una planta de colza, una planta de soja, una planta de arroz, una planta de cebada o una planta de tabaco.

20 La invención proporciona semillas transgénicas que comprenden un ácido nucleico de acuerdo con la invención o un cassette de expresión (tal como un vector) de acuerdo con la invención. La semilla transgénica puede ser una planta de cereal, una semilla de maíz, un grano de trigo, una semilla oleaginosa, una colza, una semilla de soja, un grano de palma, una semilla de girasol, una semilla de sésamo, una semilla vegetal de maní o tabaco.

25 La invención proporciona oligonucleótidos antisentido que comprenden secuencias de ácido nucleico complementarias a o capaces de hibridarse en condiciones rigurosas con ácidos nucleicos de acuerdo con la invención. En algunas realizaciones, la invención proporciona métodos para inhibir la traducción de un mensaje de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa 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 hibridarse en condiciones rigurosas con, un ácido nucleico de acuerdo con la invención.

30 La invención proporciona métodos para inhibir la traducción de un mensaje de enzima aldolasa, tal como piruvato aldolasa, tal como enzima HMG y/o KHG aldolasa en una célula que comprenden administrar a la célula o expresar en la célula un oligonucleótido antisentido que comprende una secuencia de ácidos nucleicos complementaria a, o capaz de hibridarse en condiciones rigurosas con, un ácido nucleico de acuerdo con la invención.

35 La divulgación describe moléculas de ARN inhibitorio (ARNi o ARN de interferencia) de doble hebra (incluido ARN interferente pequeño, o ARNsi, para inhibir la transcripción y microARN o miARN, para inhibir la traducción) que comprenden una subsecuencia de una secuencia de acuerdo con la invención. En algunos casos, el ARNsi es de aproximadamente 21 a aproximadamente 24 residuos o de aproximadamente al menos 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 o más nucleótidos dúplex de largo. La divulgación además proporciona métodos para inhibir la expresión de una enzima aldolasa, tal como la enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa en una célula que comprenden administrar a la célula o expresar en la célula un ARN inhibitorio de doble hebra (ARNsi o miARN), en donde el ARN comprende una subsecuencia de una secuencia de acuerdo con la invención.

40

45 La invención proporciona polipéptidos aislados, sintéticos o recombinantes que comprenden una secuencia de aminoácidos que tiene al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más o 100% (completa) de identidad de secuencia con respecto a un polipéptido o péptido de acuerdo con la invención en la longitud total del polipéptido. En algunas realizaciones, las identidades de secuencia se determinan mediante análisis con un algoritmo de comparación de secuencias o mediante una inspección visual. Las secuencias de polipéptidos o péptidos de acuerdo con la invención incluyen SEQ ID NO:276 y fragmentos enzimáticamente activos de la misma. Polipéptidos de acuerdo con la invención también incluyen fragmentos de al menos aproximadamente 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600 o más residuos de longitud o en la longitud total de una enzima. Las secuencias de polipéptidos o péptidos de acuerdo con la invención incluyen secuencia codificada mediante un ácido nucleico de acuerdo con la invención. Las secuencias de polipéptidos o péptidos de acuerdo con la invención incluyen polipéptidos o péptidos unidos específicamente mediante un anticuerpo de acuerdo con la invención (tal como epítomos) o polipéptidos o péptidos que pueden generar un anticuerpo de acuerdo con la invención (tal como un inmunógeno).

50

55 En algunas realizaciones, un polipéptido de acuerdo con la invención tiene al menos una actividad de enzima aldolasa, tal como, actividad de enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa. En otras realizaciones,

un polinucleótido de acuerdo con la invención codifica un polipéptido que tiene al menos una actividad de enzima aldolasa, tal como, actividad de enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa.

La presente divulgación además describe polipéptidos o péptidos aislados, sintéticos o recombinantes que comprenden al menos 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150 o más bases consecutivas de las secuencias de polipéptidos o péptidos de acuerdo con la invención, secuencias básicamente idénticas a las mismas y las secuencias complementarias de las mismas. El péptido puede ser, por ejemplo, un fragmento inmunogénico, un motivo (tal como un sitio de unión), una secuencia señal, una secuencia prepro o un sitio activo.

La invención proporciona ácidos nucleicos aislados, sintéticos o recombinantes que comprenden una secuencia que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, y una secuencia señal, en donde el ácido nucleico comprende una secuencia de acuerdo con la invención. Una "secuencia señal" significa una señal de secreción u otro dominio que facilita la secreción de la aldolasa de acuerdo con la invención de la célula huésped. La secuencia señal puede derivarse de otra enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, o una enzima no aldolasa, tal como no piruvato aldolasa, tal como no HMG y/o no KHG aldolasa (heteróloga). En algunas realizaciones, la invención proporciona ácidos nucleicos aislados, sintéticos o recombinantes que comprenden una secuencia que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, en donde la secuencia no contiene una secuencia señal y el ácido nucleico comprende una secuencia de acuerdo con la invención. En algunas realizaciones, la invención proporciona polipéptidos aislados, sintéticos o recombinantes que comprenden polipéptidos de acuerdo con la invención que carecen de toda o parte de una secuencia señal. En algunas realizaciones, el polipéptido aislado, sintético o recombinante puede comprender el polipéptido de acuerdo con la invención que comprende una secuencia señal heteróloga, tal como una secuencia señal de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, o secuencia señal de enzima no aldolasa, tal como no piruvato aldolasa, tal como no HMG y/o no KHG aldolasa, heteróloga.

En algunas realizaciones, la invención proporciona proteínas quiméricas que comprenden un primer dominio que comprende una secuencia señal de acuerdo con la invención y al menos un segundo dominio. La proteína puede ser una proteína de fusión. El segundo dominio puede comprender una enzima. La proteína puede ser una no enzima.

La invención proporciona polipéptidos quiméricos que comprenden al menos un primer dominio que comprende péptido señal (SP), una secuencia prepro y/o un dominio catalítico (CD) de acuerdo con la invención y al menos un segundo dominio que comprende un péptido o polipéptido heterólogo, en donde el péptido o polipéptido heterólogo no se asocia naturalmente con el péptido señal (SP), secuencia prepro y/o dominio catalítico (CD). En algunas realizaciones, el péptido o polipéptido heterólogo no es una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa. El péptido o polipéptido heterólogo puede estar en el extremo amino terminal, carboxi terminal o en ambos extremos del péptido señal (SP), la secuencia prepro y/o el dominio catalítico (CD).

La invención proporciona ácidos nucleicos aislados, sintéticos o recombinantes que codifican un polipéptido quimérico, en donde el polipéptido quimérico comprende al menos un primer dominio que comprende péptido señal (SP), un dominio prepro y/o un dominio catalítico (CD) de acuerdo con la invención y al menos un segundo dominio que comprende un péptido o polipéptido heterólogo, en donde el péptido o polipéptido heterólogo no se asocia naturalmente con el péptido señal (SP), dominio prepro y/o dominio catalítico (CD).

La invención proporciona secuencias señal aisladas, sintéticas o recombinantes (tales como péptidos señal) que consisten en o que comprenden una secuencia como se establece en los residuos 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 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, 1 a 44, de un polipéptido de SEQ ID NO:276. En algunas realizaciones, la invención proporciona secuencias señal que comprenden los primeros 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 amino terminales de un polipéptido de acuerdo con la invención.

En algunas realizaciones, la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa comprende una actividad específica de aproximadamente 10 a aproximadamente 12.000 unidades por miligramo de proteína. En otras realizaciones, la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, comprende una actividad específica de aproximadamente 1000 a aproximadamente 10.000 unidades por miligramo de proteína, o de aproximadamente 5000 a aproximadamente 7500 unidades por miligramo de proteína. De forma alternativa, la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, comprende una actividad específica en el rango de aproximadamente 10 a aproximadamente 7500 unidades por miligramo de proteína, o de aproximadamente 5000 a aproximadamente 12.000 unidades por miligramo de proteína. En algunas realizaciones, la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa comprende una actividad específica en el rango de aproximadamente 10 a aproximadamente 5000 unidades por miligramo de proteína, o de aproximadamente 7500 a aproximadamente 10.000 unidades por miligramo de proteína. En otras realizaciones, la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, comprende una actividad específica en el rango de



aproximadamente 10 a aproximadamente 2500 unidades por miligramo de proteína. De forma alternativa, la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, comprende una actividad específica en el rango de aproximadamente 10 a aproximadamente 1000 unidades por miligramo de proteína. Un método ejemplar para medir la actividad de diferentes enzimas aldolasas, tales como piruvato aldolasas, tales como HMG y/o KHG aldolasa utiliza un sustrato general, 4-carboxi-4-hidroxi-2-oxoadipato ("CHA"). Un ensayo típico comprende 50 mM fosfato de sodio pH 7,5, 1 mM MgCl<sub>2</sub>, 1 mM CHA, 10 µg/ml D-lactato deshidrogenasa ("LDH") de *Lactobacillus leichmanii* (Sigma-Aldrich, St. Louis, MO), 0,5 mM NADH. El ensayo se inicia agregando la enzima a ser medida. La liberación de piruvato, junto con la formación de NAD<sup>+</sup>, se monitorea de forma continua en un espectrofotómetro a 340 nm. Una unidad de actividad de enzima se define como la cantidad que libera suficiente piruvato como para disminuir la absorbancia 1 OD por minuto a 340 nm.

En otras realizaciones, la termotolerancia de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, comprende la retención de al menos la mitad de la actividad específica de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa luego de calentarse hasta alcanzar una temperatura elevada, tal como una temperatura de aproximadamente 0°C a aproximadamente 20°C, aproximadamente 20°C a aproximadamente 37°C, aproximadamente 37°C a aproximadamente 50°C, aproximadamente 50°C a aproximadamente 70°C, aproximadamente 70°C a aproximadamente 75°C, aproximadamente 75°C a aproximadamente 80°C, aproximadamente 80°C a aproximadamente 85°C, aproximadamente 85°C a aproximadamente 90°C, aproximadamente 90°C a aproximadamente 95°C, aproximadamente 95°C a aproximadamente 100°C, aproximadamente 100°C a aproximadamente 110°C o más alta. De forma alternativa, la termotolerancia puede comprender la retención de actividad específica de aproximadamente 10 a aproximadamente 12.000 unidades por miligramo de proteína, o de aproximadamente 5000 a aproximadamente 10.000 unidades por miligramo de proteína, luego de calentarse hasta alcanzar una temperatura elevada, como se ha descrito anteriormente. En otras realizaciones, la termotolerancia puede comprender la retención de actividad específica en el rango de aproximadamente 10 a aproximadamente 5000 unidades por miligramo de proteína luego de calentarse hasta alcanzar una temperatura elevada, como se ha descrito anteriormente.

La invención proporciona polipéptidos aislados, sintéticos o recombinantes de acuerdo con la invención, en donde los polipéptidos comprenden al menos un sitio de glicosilación. En algunas realizaciones, la glicosilación puede ser una glicosilación enlazada a N. En algunas realizaciones, el polipéptido puede glicosilarse luego de expresarse en un huésped *P. pastoris* o *S. pombe* o en una célula huésped de mamífero.

En algunas realizaciones, el polipéptido puede retener actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, en condiciones que presentan un pH de aproximadamente 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). En otras realizaciones, el polipéptido puede retener una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, en condiciones que presentan un pH de aproximadamente 7, pH 7,5, pH 8,0, pH 8,5, pH 9, pH 9,5, pH 10, pH 10,5, pH 11,0, pH 11,5, pH 12, pH 12,5 o más (más básico). En algunas realizaciones, el polipéptido puede retener una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, luego de la exposición a condiciones que presentan un pH de aproximadamente 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). En otras realizaciones, el polipéptido puede retener una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, luego de la exposición a condiciones que presentan un pH de aproximadamente 7, pH 7,5, pH 8,0, pH 8,5, pH 9, pH 9,5, pH 10, pH 10,5, pH 11,0, pH 11,5, pH 12, pH 12,5 o más (más básico).

En algunas realizaciones, la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención tiene actividad en condiciones alcalinas, tales como las condiciones alcalinas del intestino, tal como el intestino delgado. En algunas realizaciones, el polipéptido puede retener actividad luego de la exposición al pH ácido del estómago.

La invención proporciona preparaciones de proteína que comprenden un polipéptido (incluyendo péptidos) de acuerdo con la invención, en donde la preparación de proteína comprende un líquido, un sólido o un gel. En algunas realizaciones, la invención proporciona heterodímeros que comprenden un polipéptido de acuerdo con la invención y un segundo miembro, tal como un polipéptido u otro (segundo) dominio. El segundo miembro del heterodímero puede ser una enzima aldolasa diferente, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, una enzima diferente u otra proteína. En algunas realizaciones, el segundo dominio puede ser un polipéptido y el heterodímero puede ser una proteína de fusión. En algunas realizaciones, el segundo dominio puede ser un epítipo o una etiqueta. En algunas realizaciones, la invención proporciona homomultímeros, incluyendo, a modo no taxativo, homodímeros, homotrímeros, homotetrámeros, homopentámeros y homohexámeros, que comprenden un polipéptido de acuerdo con la invención.

La invención proporciona polipéptidos inmovilizados (incluyendo péptidos) que tienen actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, en donde el polipéptido inmovilizado comprende un polipéptido de acuerdo con la invención, un polipéptido codificado por un ácido nucleico de acuerdo con la invención o un polipéptido que comprende un polipéptido de acuerdo con la invención y un segundo dominio. En algunas realizaciones, el polipéptido puede inmovilizarse en una célula, un metal, una resina, un polímero, una cerámica, un vidrio, un microelectrodo, una partícula grafitica, una perla, un gel, una placa, un arreglo o un tubo capilar.

La invención también proporciona ensayos que comprenden un ácido nucleico inmovilizado de acuerdo con la invención, tales como sondas de acuerdo con la invención. En algunas realizaciones, la invención también proporciona ensayos que comprenden un anticuerpo de acuerdo con la invención.

5 La invención proporciona anticuerpos aislados, sintéticos o recombinantes que se unen específicamente a un polipéptido de acuerdo con la invención o a un polipéptido codificado por un ácido nucleico de acuerdo con la invención. Estos anticuerpos de acuerdo con la invención pueden ser un anticuerpo monoclonal o policlonal. En algunas realizaciones, la invención proporciona hibridomas que comprenden un anticuerpo de acuerdo con la invención, tal como un anticuerpo que se une específicamente a un polipéptido de acuerdo con la invención o a un polipéptido codificado por un ácido nucleico de acuerdo con la invención. En algunas realizaciones, la invención  
10 proporciona ácidos nucleicos que codifican estos anticuerpos.

La invención proporciona métodos para aislar o identificar polipéptidos que tienen actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, que comprenden las etapas de: (a) proporcionar un anticuerpo de acuerdo con la invención; (b) proporcionar una muestra que comprende polipéptidos; y (c) poner en contacto la muestra de la etapa (b) con el anticuerpo de la etapa (a) en condiciones en las cuales el anticuerpo  
15 puede unirse específicamente al polipéptido y de esa forma aislar o identificar un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa.

La divulgación además proporciona métodos para preparar un anticuerpo anti-enzima aldolasa, tal como anti-piruvato aldolasa, tal como anti-HMG y/o anti-KHG aldolasa que comprenden administrar a un animal no humano un ácido nucleico de acuerdo con la invención o un polipéptido de acuerdo con la invención o subsecuencias de los  
20 mismos en una cantidad suficiente para generar una respuesta inmune humoral y de esa forma elaborar un anticuerpo anti-enzima aldolasa, tal como anti-piruvato aldolasa, tal como anti-HMG y/o anti-KHG aldolasa. La divulgación proporciona métodos para preparar una respuesta inmune anti-aldolasa, tal como anti-piruvato aldolasa, tal como anti-HMG y/o anti-KHG aldolasa (celular o humoral) que comprende administrar a un animal no humano un ácido nucleico de acuerdo con la invención o un polipéptido de acuerdo con la invención o subsecuencias de los  
25 mismos en una cantidad suficiente para generar una respuesta inmune (celular o humoral).

La invención proporciona métodos para producir un polipéptido recombinante que comprende las etapas de: (a) proporcionar un ácido nucleico de acuerdo con la invención operativamente enlazado a un promotor; y (b) expresar el ácido nucleico de la etapa (a) en condiciones que permiten la expresión del polipéptido y de esa forma producir un polipéptido recombinante. En algunas realizaciones, el método puede comprender además transformar una célula  
30 huésped con el ácido nucleico de la etapa (a) y a continuación expresar el ácido nucleico de la etapa (a) y de esa forma producir un polipéptido recombinante en una célula transformada.

La divulgación proporciona métodos para identificar un polipéptido que tiene actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, que comprenden las siguientes etapas: (a) proporcionar un polipéptido de acuerdo con la invención; o un polipéptido codificado por un ácido nucleico de acuerdo con la  
35 invención; (b) proporcionar sustrato de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa; y (c) poner en contacto el polipéptido o un fragmento o una variante del mismo de la etapa (a) con el sustrato de la etapa (b) y detectar una disminución en la cantidad de sustrato o un aumento en la cantidad de un producto de reacción, en donde una disminución en la cantidad del sustrato o un aumento en la cantidad del producto de reacción detecta un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa,  
40 tal como HMG y/o KHG aldolasa. En algunos casos, el sustrato es un carbohidrato, un compuesto que comprende un carbohidrato y/o un mimético de carbohidrato.

La divulgación además proporciona métodos para identificar sustrato de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa que comprenden las siguientes etapas: (a) proporcionar un polipéptido de acuerdo con la invención o un polipéptido codificado por un ácido nucleico de acuerdo con la invención; (b) proporcionar 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  
45 detectar una disminución en la cantidad de sustrato o un aumento en la cantidad de producto de reacción, en donde una disminución en la cantidad del sustrato o un aumento en la cantidad de un producto de reacción identifica el sustrato de prueba como un sustrato de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa.

La divulgación proporciona métodos para determinar si un compuesto de prueba se une específicamente a un polipéptido que comprende las siguientes etapas: (a) expresar un ácido nucleico o un vector que comprende el ácido nucleico en condiciones permisivas para trasladar el ácido nucleico a un polipéptido, en donde el ácido nucleico  
50 comprende un ácido nucleico de acuerdo con la invención, o proporcionar un polipéptido de acuerdo con la invención; (b) proporcionar 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 une específicamente a el polipéptido.

55 La divulgación proporciona métodos para identificar un modulador de una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, que comprenden las siguientes etapas: (a) proporcionar un polipéptido de acuerdo con la invención o un polipéptido codificado por un ácido nucleico de acuerdo con la invención; (b) proporcionar un compuesto de prueba; (c) poner en contacto el polipéptido de la etapa (a) con el

5 compuesto de prueba de la etapa (b) y medir una actividad de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, en donde un cambio en la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, medido en presencia del compuesto de prueba comparado con la actividad en ausencia del compuesto de prueba determina que el compuesto de prueba modula la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. En algunos casos, la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, puede medirse proporcionando un sustrato de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa y detectar una disminución en la cantidad del sustrato o un aumento en la cantidad de un producto de reacción, o un aumento en la cantidad del sustrato o una disminución en la cantidad de un producto de reacción. Una disminución en la cantidad del sustrato o un aumento en la cantidad del 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 activador de la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. Un aumento en la cantidad del sustrato o una disminución en la cantidad del 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 inhibidor de la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa.

20 La divulgación además proporciona sistemas informáticos que comprenden un procesador y un dispositivo de almacenamiento de datos, en donde dicho dispositivo de almacenamiento de datos ha almacenado una secuencia de polipéptidos o una secuencia de ácidos nucleicos de acuerdo con la invención (tal como un polipéptido o péptido codificado por un ácido nucleico de acuerdo con la invención). En algunos casos, el sistema informático además puede comprender 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 otros casos, el algoritmo de comparación de secuencias comprende un programa informático que indica polimorfismos. En algunos casos, el sistema informático además puede comprender un identificador que identifica una o más características en dicha secuencia. En algunos casos, la divulgación proporciona medios legibles por ordenador que tienen almacenada una secuencia de polipéptidos o una secuencia de ácidos nucleicos de acuerdo con la invención. La divulgación proporciona métodos para identificar una característica en una secuencia que comprenden las etapas de: (a) leer la secuencia utilizando un programa informático 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 acuerdo con la invención; y (b) identificar una o más características en la secuencia con el programa informático. La divulgación proporciona además 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 informático que compara secuencias, en donde la primera secuencia comprende una secuencia de polipéptidos o una secuencia de ácidos nucleicos de acuerdo con la invención; y (b) determinar diferencias entre la primera secuencia y la segunda secuencia con el programa informático. La etapa de determinar diferencias entre la primera secuencia y la segunda secuencia además puede comprender la etapa de identificar polimorfismos. En algunos casos, el método puede comprender además un identificador que identifica una o más características en una secuencia. En otros casos, el método puede comprender leer la primera secuencia utilizando un programa informático e identificar una o más características en la secuencia.

40 La divulgación proporciona además métodos para aislar o recuperar un ácido nucleico que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, de una muestra, tal como una muestra ambiental, que comprenden las etapas de: (a) proporcionar un par de secuencias cebadoras de amplificación para amplificar un ácido nucleico que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, en donde el par de cebadores es capaz de amplificar un ácido nucleico de acuerdo con la invención; (b) aislar un ácido nucleico de la muestra o tratar la muestra de forma tal que el ácido nucleico en la muestra sea accesible para la hibridación con el par de cebadores de amplificación; y (c) combinar el ácido nucleico de la etapa (b) con el par de cebadores de amplificación de la etapa (a) y amplificar ácido nucleico de la muestra y de esa forma aislar o recuperar un ácido nucleico que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, de una muestra. Uno o cada miembro del par de secuencias cebadoras de amplificación pueden comprender un oligonucleótido que comprende un par de secuencias cebadoras de amplificación de acuerdo con la invención, por ejemplo con al menos aproximadamente 10 a 50 bases consecutivas de una secuencia de acuerdo con la invención. En una realización de la invención, la muestra es una muestra ambiental.

55 La divulgación proporciona además métodos para aislar o recuperar un ácido nucleico que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, de una muestra, tal como una muestra ambiental, que comprenden las etapas de: (a) proporcionar una sonda de polinucleótidos que comprende un ácido nucleico de acuerdo con la invención o una subsecuencia del mismo; (b) aislar un ácido nucleico de la muestra o tratar la muestra de forma tal que el ácido nucleico en la muestra sea accesible para la hibridación con una sonda de polinucleótidos de la etapa (a); (c) combinar el ácido nucleico aislado o la prueba tratada de la etapa (b) con la sonda de polinucleótidos de la etapa (a); y (d) aislar un ácido nucleico que se hibrida específicamente con la sonda de polinucleótidos de la etapa (a) y de esa forma aislar o recuperar un ácido nucleico que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, de una muestra. La muestra puede comprender una muestra de agua, una muestra líquida,

una muestra de suelo, una muestra de aire o una muestra biológica. En algunos casos, la muestra biológica puede derivarse de una célula bacteriana, una célula de protozoo, una célula de insecto, una célula de levadura, una célula de planta, una célula fúngica o una célula de mamífero. En un caso de la invención, la muestra es una muestra ambiental.

5 La invención proporciona métodos para generar una variante de un ácido nucleico que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, que comprenden las etapas de: (a) proporcionar un ácido nucleico plantilla que comprende un ácido nucleico de acuerdo con la invención; y (b) modificar, eliminar o agregar uno o más nucleótidos en la secuencia de la plantilla o una combinación de los mismos, para generar una variante del ácido nucleico plantilla. En algunas realizaciones, el  
10 método puede comprender además expresar el ácido nucleico variante para generar una aldolasa variante, tal como polipéptido de enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa. Las modificaciones, adiciones o eliminaciones pueden introducirse mediante un método que comprende PCR susceptible a errores, transposición, mutagénesis dirigida a oligonucleótidos, PCR de ensamblaje, mutagénesis por PCR sexual, mutagénesis in vivo, mutagénesis de cassette, mutagénesis de conjunto recursivo, mutagénesis de conjunto exponencial, mutagénesis  
15 específica de sitio, reensamblaje de genes, Mutagénesis de Saturación de Sitios de Genes (GSSM), reensamblaje de ligación sintética (SLR), Mutagénesis de Saturación Cromosómica (CSM) o una combinación de los mismos. En otras realizaciones, las modificaciones, adiciones o eliminaciones se introducen mediante un método que comprende recombinación, recombinación de secuencia recursiva, mutagénesis de ADN modificado por fosfotioato, mutagénesis de plantilla que contiene uracilo, mutagénesis de dúplex discontinuo, mutagénesis de reparación de no  
20 coincidencia puntual, mutagénesis de cepa huésped de reparación deficiente, mutagénesis química, mutagénesis radiogénica, mutagénesis de eliminación, mutagénesis de selección de restricción, mutagénesis de purificación de restricción, síntesis génica artificial, mutagénesis de conjunto, creación de multímeros de ácidos nucleicos quiméricos y una combinación de los mismos.

En algunas realizaciones, el método puede repetirse hasta que se produzca una enzima aldolasa, tal como piruvato  
25 aldolasa, HMG y/o KHG aldolasa que tenga una actividad alterada o diferente o una estabilidad alterada o diferente de la de un polipéptido codificado por el ácido nucleico plantilla. En algunas realizaciones, el polipéptido de enzima aldolasa, tal como enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa, variante es termotolerante y retiene cierta actividad luego de exponerse a una temperatura elevada. En otras realizaciones, la variante aldolasa, tal como polipéptido de enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa ha aumentado la glicosilación en  
30 comparación con la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa codificada por un ácido nucleico plantilla. De forma alternativa, el polipéptido de aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, variante tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, a una alta temperatura, en donde la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa codificada por el ácido nucleico plantilla no es activa a altas temperaturas. En algunas realizaciones,  
35 el método puede repetirse hasta que se produzca una secuencia de codificación de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa que tenga un uso de codón alterado con respecto al del ácido nucleico plantilla. En otras realizaciones, el método puede repetirse hasta que se produzca un gen de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa que tenga mayor o menor nivel de expresión del mensaje o estabilidad que el ácido nucleico plantilla.

40 La invención proporciona métodos para modificar codones en un ácido nucleico que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, para aumentar su expresión en una célula huésped, comprendiendo dicho método las siguientes etapas: (a) proporcionar un ácido nucleico de acuerdo con la invención que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa; y (b) identificar un codón no preferido o menos preferido  
45 en el ácido nucleico de la etapa (a) y reemplazarlo por un codón preferido o utilizado de forma neutra 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 huésped y un codón no preferido o menos preferido es un codón infrarrepresentado en secuencias de codificación en genes en la célula huésped y de esa forma modificar el ácido nucleico para aumentar su expresión en una célula huésped.

50 La invención proporciona métodos para modificar codones en un ácido nucleico que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, comprendiendo dicho método las siguientes etapas: (a) proporcionar un ácido nucleico de acuerdo con la invención; y (b) identificar un codón en el ácido nucleico de la etapa (a) y reemplazarlo por un codón diferente que codifica el mismo aminoácido que el codón reemplazado y de esa forma modificar codones en un ácido nucleico que codifica una  
55 enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa.

La invención proporciona métodos para modificar codones en un ácido nucleico que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, para aumentar su expresión en una célula huésped, comprendiendo dicho método las siguientes etapas: (a) proporcionar un ácido nucleico de acuerdo con la invención que codifica un polipéptido de enzima aldolasa, tal como piruvato aldolasa,  
60 HMG y/o KHG aldolasa; y (b) identificar un codón no preferido o menos preferido en el ácido nucleico de la etapa (a)

y reemplazarlo por un codón preferido o utilizado de forma neutra 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 huésped y un codón no preferido o menos preferido es un codón infrarrepresentado en secuencias de codificación en genes en la célula huésped y de esa forma modificar el ácido nucleico para aumentar su expresión en una célula huésped.

La invención proporciona métodos para modificar un codón en un ácido nucleico que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, para disminuir su expresión en una célula huésped, comprendiendo dicho método las siguientes etapas: (a) proporcionar un ácido nucleico de acuerdo con la invención; y (b) identificar al menos un codón preferido en el ácido nucleico de la etapa (a) y reemplazarlo por un codón no preferido o menos preferido 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 una célula huésped y un codón no preferido o menos preferido es un codón infrarrepresentado en secuencias de codificación en genes en la célula huésped y de esa forma modificar el ácido nucleico para disminuir su expresión en una célula huésped. En algunas realizaciones, la célula huésped puede ser una célula bacteriana, una célula fúngica, una célula de insecto, una célula de levadura, una célula de planta o una célula de mamífero.

La divulgación proporciona además métodos para producir una biblioteca de ácidos nucleicos que codifican una pluralidad de sitios activos o sitios de unión a sustrato de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, modificados, en donde los sitios activos o sitios de unión a sustrato modificados derivan de un primer ácido nucleico que comprende una secuencia que codifica un primer sitio activo o un primer sitio de unión a sustrato, comprendiendo dicho método las siguientes etapas: (a) proporcionar un primer ácido nucleico que codifica un primer sitio activo o primer sitio de unión a sustrato, en donde la primera secuencia de ácido nucleico comprende una secuencia que se hidrida en condiciones rigurosas con un ácido nucleico de acuerdo con la invención, y el ácido nucleico codifica un sitio activo de una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa o un sitio de unión a sustrato de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa; (b) proporcionar un conjunto de oligonucleótidos mutagénicos que codifican variantes de aminoácido naturales en una pluralidad de codones objetivo en el primer ácido nucleico; y (c) utilizar el conjunto de oligonucleótidos mutagénicos para generar un conjunto de ácidos nucleicos variantes de codificación de sitios activos o de codificación de sitios de unión a sustrato que codifican un rango de variaciones de aminoácidos en cada codón de aminoácido que fue mutagenizado y de esa forma producir una biblioteca de ácidos nucleicos que codifican una pluralidad de sitios activos o sitios de unión a sustrato de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa modificados. En algunos casos, el método comprende mutagenizar el primer ácido nucleico de la etapa (a) mediante un método que comprende un sistema de evolución dirigida optimizado, Mutagénesis de Saturación de Sitios de Genes (GSSM), reensamblaje por ligamiento sintético (SLR), PCR susceptible a errores, transposición, mutagénesis dirigida a oligonucleótidos, PCR de ensamblaje, mutagénesis por PCR sexual, mutagénesis in vivo, mutagénesis de cassette, mutagénesis de conjunto recursivo, mutagénesis de conjunto exponencial, mutagénesis específica de sitio, reensamblaje de genes y una combinación de los mismos. En otros casos, el método comprende mutagenizar el primer ácido nucleico de la etapa (a) o variantes mediante un método que comprende recombinación, recombinación de secuencia recursiva, mutagénesis de ADN modificado por fosfotioato, mutagénesis de plantilla que contiene uracilo, mutagénesis de dúplex discontinuo, mutagénesis de reparación de no coincidencia puntual, mutagénesis de cepa huésped de reparación deficiente, mutagénesis química, mutagénesis radiogénica, mutagénesis de eliminación, mutagénesis de selección de restricción, mutagénesis de purificación de restricción, síntesis génica artificial, mutagénesis de conjunto, creación de multímeros de ácidos nucleicos quiméricos y una combinación de los mismos.

La divulgación proporciona además métodos para obtener una molécula pequeña que comprenden las siguientes etapas: (a) proporcionar una pluralidad de enzimas biosintéticas capaces de sintetizar o modificar una molécula pequeña, en donde una de las enzimas comprende una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa codificada por un ácido nucleico de acuerdo con la invención; (b) proporcionar 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 en condiciones que facilitan una pluralidad de reacciones biocatalíticas para generar una molécula pequeña mediante una serie de reacciones biocatalíticas. La divulgación proporciona además métodos para modificar una molécula pequeña que comprenden las siguientes etapas: (a) proporcionar una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa, en donde la enzima comprende un polipéptido de acuerdo con la invención, o un polipéptido codificado por un ácido nucleico de acuerdo con la invención o una subsecuencia del mismo; (b) proporcionar una molécula pequeña; y (c) hacer reaccionar la enzima de la etapa (a) con la molécula pequeña de la etapa (b) en condiciones que facilitan una reacción enzimática catalizada por la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa y de esa forma modificar una molécula pequeña mediante una reacción enzimática de aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa. En algunos casos, el método puede comprender una pluralidad de sustratos de molécula pequeña para la enzima de la etapa (a) y de esa forma generar una biblioteca de moléculas pequeñas modificadas producidas mediante al menos una reacción enzimática catalizada por la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. En algunos casos, el método puede comprender una pluralidad de enzimas adicionales en 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 otros casos, el método puede comprender además la etapa de evaluar la biblioteca para determinar si una molécula pequeña modificada particular que exhibe una actividad deseada se encuentra presente en la biblioteca. La etapa de evaluación de la biblioteca además puede comprender las etapas de eliminar sistemáticamente todas menos una de las reacciones biocatalíticas utilizadas para producir una porción de la pluralidad de las moléculas pequeñas modificadas dentro de la biblioteca mediante evaluación de la porción de la molécula pequeña modificada para detectar la presencia o ausencia de la molécula pequeña particular modificada con una actividad deseada, e identificar al menos una reacción biocatalítica específica que produzca la molécula pequeña particular modificada de actividad deseada.

La invención proporciona métodos para determinar un fragmento funcional de una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa que comprenden las etapas de: (a) proporcionar una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa, en donde la enzima comprende un polipéptido de acuerdo con la invención o un polipéptido codificado por un ácido nucleico de acuerdo con la invención o una subsecuencia del mismo; y (b) eliminar una pluralidad de residuos de aminoácidos de la secuencia de la etapa (a) y evaluar la subsecuencia remanente para detectar una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa y de esa forma determinar un fragmento funcional de una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa. En algunas realizaciones, la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa se mide proporcionando un sustrato de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa y detectar una disminución en la cantidad del sustrato o un aumento en la cantidad de un producto de reacción.

La invención proporciona métodos para manipulación de células enteras de fenotipos nuevos o modificados utilizando análisis de flujo metabólico en tiempo real, comprendiendo dicho método las siguientes etapas: (a) obtener una célula modificada modificando la composición genética de una célula, en donde la composición genética se modifica por la adición a la célula de un ácido nucleico de acuerdo con 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 en condiciones similares y, de esta forma, identificar un fenotipo manipulado en la célula utilizando análisis de flujo metabólico en tiempo real. En algunas realizaciones, la composición genética de la célula puede modificarse mediante un método que comprende la eliminación de una secuencia o la modificación de una secuencia en la célula o desactivar la expresión de un gen. En algunas realizaciones, el método puede comprender además seleccionar una célula que comprenda un fenotipo recién manipulado. En otras realizaciones, el método puede comprender cultivar la célula seleccionada y generar así una nueva cepa celular que comprende un fenotipo recién manipulado.

La invención proporciona métodos para aumentar la termotolerancia o termoestabilidad de un polipéptido de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa, comprendiendo el método glicosilar un polipéptido de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa, en donde el polipéptido comprende al menos treinta aminoácidos contiguos de un polipéptido de acuerdo con la invención; o un polipéptido codificado por una secuencia de ácidos nucleicos de acuerdo con la invención y de esa forma aumentar la termotolerancia o termoestabilidad del polipéptido de aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. En algunas realizaciones, la actividad específica de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa puede ser termoestable o termotolerante a una temperatura en el rango entre más de aproximadamente 37°C a aproximadamente 95°C.

La invención proporciona métodos para sobreexpresar un polipéptido de aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, recombinante en una célula que comprenden expresar un vector que comprende un ácido nucleico que comprende un ácido nucleico de acuerdo con la invención o una secuencia de ácidos nucleicos de acuerdo con la invención, en donde las identidades de secuencia se determinan mediante análisis con un algoritmo de comparación de secuencias o mediante inspección visual, en donde la sobreexpresión se lleva a cabo mediante el uso de un promotor de alta actividad, un vector dicistrónico o mediante amplificación del gen del vector.

La invención proporciona métodos para preparar una planta transgénica que comprenden las siguientes etapas: (a) introducir una secuencia heteróloga de ácidos nucleicos en la célula, en donde la secuencia nucleica heteróloga comprende una secuencia de ácidos nucleicos de acuerdo con la invención y de esa forma producir una célula de planta transformada; y (b) producir una planta transgénica a partir de la célula transformada. En algunas realizaciones, la etapa (a) además puede comprender introducir la secuencia heteróloga de ácidos nucleicos mediante electroporación o microinyección de protoplastos de célula de planta. En otras realizaciones, la etapa (a) además puede comprender introducir la secuencia heteróloga de ácidos nucleicos directamente al tejido de la planta mediante bombardeo con partículas de ADN. De forma alternativa, la etapa (a) además puede comprender introducir la secuencia heteróloga de ácidos nucleicos en el ADN de la célula de planta utilizando un huésped *Agrobacterium tumefaciens*. En algunas realizaciones, la célula de planta puede ser célula de caña de azúcar, remolacha, soja, tomate, patata, maíz, arroz, trigo, tabaco o cebada.

La invención proporciona métodos para expresar una secuencia heteróloga de ácidos nucleicos en una célula de planta que comprenden las siguientes etapas: (a) transformar la célula de planta con una secuencia heteróloga de

ácidos nucleicos operativamente enlazado a un promotor, en donde la secuencia nucleica heteróloga comprende un ácido nucleico de acuerdo con la invención; (b) cultivar la planta en condiciones en las cuales la secuencia heteróloga de ácidos nucleicos se expresa en la célula de planta. En algunas realizaciones, la invención proporciona métodos para expresar una secuencia heteróloga de ácidos nucleicos en una célula de planta que comprenden las siguientes etapas: (a) transformar la célula de planta con una secuencia heteróloga de ácidos nucleicos operativamente enlazado a un promotor, en donde la secuencia nucleica heteróloga comprende una secuencia de acuerdo con la invención; (b) cultivar la planta en condiciones en las cuales la secuencia heteróloga de ácidos nucleicos se expresa en la célula de planta.

La invención proporciona alimentos o piensos que comprenden un polipéptido de acuerdo con la invención o un polipéptido codificado por un ácido nucleico de acuerdo con la invención. En algunas realizaciones, la invención proporciona alimentos, piensos, líquidos, tales como bebidas (tales como jugos de frutas o cerveza), panes o masas o productos panificados o precursores de bebidas (tales como mosto), que comprenden un polipéptido de acuerdo con la invención. En otras realizaciones, la invención proporciona aditivos para alimentos, piensos o bebidas que comprenden un polipéptido de acuerdo con la invención. En algunas realizaciones, la invención proporciona alimentos o complementos nutricionales, tales como para un humano o un animal, que comprenden un polipéptido de acuerdo con la invención, tal como un polipéptido codificado por el ácido nucleico de acuerdo con la invención.

En algunas realizaciones, el polipéptido en el alimento o complemento nutricional puede ser glicosilado. En algunas realizaciones, la invención proporciona matrices de administración de enzima comestible que comprenden un polipéptido de acuerdo con la invención, tal como un polipéptido codificado por el ácido nucleico de acuerdo con la invención. En algunas realizaciones, la matriz de administración comprende un granulado. En algunas realizaciones, el polipéptido puede ser glicosilado. En algunas realizaciones, la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa es termotolerante. En otras realizaciones, la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa es termoestable.

La invención proporciona alimentos, piensos o complementos nutricionales que comprenden un polipéptido de acuerdo con la invención. En algunas realizaciones, la invención proporciona métodos para utilizar una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa como un complemento nutricional en una dieta animal, comprendiendo dicho método: preparar un complemento nutricional que contiene una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa que comprende al menos treinta aminoácidos contiguos de un polipéptido de acuerdo con la invención; y administrar el complemento nutricional a un animal. El animal puede ser un humano, un rumiante o un animal monogástrico. La enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa puede prepararse mediante expresión de un polinucleótido que codifica la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa en un organismo que se selecciona del grupo que consiste en una bacteria, una levadura, una planta, un insecto, un hongo y un animal. El organismo puede seleccionarse del grupo que consiste en un *S. pombe*, *S. cerevisiae*, *Pichia pastoris*, *E. coli*, *Streptomyces* sp., *Bacillus* sp. *Pseudomonas* sp., *Aspergillus* sp. y *Lactobacillus* sp.

La divulgación proporciona además matrices de administración de enzima comestible que comprenden una enzima aldolasa recombinante termoestable, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, tal como un polipéptido de acuerdo con la invención. En algunas realizaciones, la invención proporciona métodos para administrar un complemento de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa a un animal, comprendiendo dicho método: preparar una matriz de administración de enzima comestible en forma de granulosos que comprenden un portador comestible granulado y una enzima aldolasa recombinante termoestable, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, en donde los granulosos fácilmente dispersan la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa contenida en los mismos en el medio acuoso, y administrar la matriz de administración de enzima comestible al animal. La enzima aldolasa recombinante, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa puede comprender un polipéptido de acuerdo con la invención. La enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa puede ser glucosada para proporcionar termoestabilidad en las condiciones de granulación. La matriz de administración puede formarse mediante granulación de una mezcla que comprende un germen de grano y una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa. Las condiciones de granulación pueden incluir aplicación de vapor. Las condiciones de granulación pueden comprender aplicación de una temperatura superior a aproximadamente 80°C durante aproximadamente 5 minutos y la enzima retiene una actividad específica de al menos 350 a aproximadamente 900 unidades por miligramo de enzima.

La divulgación proporciona composiciones farmacéuticas que comprenden una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de acuerdo con la invención o un polipéptido codificado por un ácido nucleico de acuerdo con la invención. En algunos casos, la composición farmacéutica actúa como un auxiliar digestivo.

En algunas realizaciones, un compuesto que contiene un enlace carbono-carbono está en contacto con un polipéptido de acuerdo con la invención que tiene actividad de enzima aldolasa, tal como, actividad de enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa, a un pH en el rango de aproximadamente pH 3,0 a aproximadamente 9,0, aproximadamente 3,0 a aproximadamente 10,0, aproximadamente 3,0 a aproximadamente 11,0 o más. En otras realizaciones, un compuesto que contiene un enlace carbono-carbono está en contacto con la

enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, a una temperatura de al menos aproximadamente 55°C, 60°C, 65°C, 70°C, 75°C, 80°C, 85°C, 90°C o más.

5 Esta divulgación proporciona, entre otras cosas, polipéptidos que son útiles para facilitar una reacción en procesos para producir monatina, derivados de monatina, y sales de la misma, por ejemplo en la producción de ácido R-2-hidroxi 2-(indol-3-ilmetil)-4-cetoglutárico (también conocido como monatina de R-alfa cetoácido, precursor de R-monatina, R-MP, y la forma alfa-ceto de monatina), un precursor para ciertos estereoisómeros de monatina, tales como monatina R,R y S,R. La divulgación también proporciona métodos para preparar monatina, derivados de monatina y sales y productos de condensación interna de la misma utilizando uno o más polipéptidos de la invención. Los métodos para sintetizar R-MP, estereoisómeros de monatina y/o estereoisómeros de derivados de monatina incluyen el uso de un polipéptido con actividad de aldolasa de SEQ ID NO:276 y fragmentos enzimáticamente activos de la misma.

15 Asimismo, los métodos para sintetizar R-MP, estereoisómeros de monatina y/o estereoisómeros de derivados de monatina pueden incluir el uso de un polipéptido con actividad de aldolasa codificado por una secuencia de ácidos nucleicos que tiene al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más o 100% (completa) de identidad de secuencia con respecto al ácido nucleico de SEQ ID NO:275 en la longitud total de un ADNc, transcripto (ARNm) o gen.

Más aun, los métodos para sintetizar R-MP, estereoisómeros de monatina y/o estereoisómeros de derivados de monatina pueden incluir el uso de un polipéptido con actividad de aldolasa codificado por una secuencia de ácidos nucleicos que se hidrida en condiciones rigurosas con un ácido nucleico de SEQ ID NO:275.

20 La invención proporciona un método que comprende: producir un producto elegido de monatina, derivados de monatina, sales de la misma y combinaciones de la misma en una vía de múltiples etapas, en donde una reacción en la vía es facilitada por un polipéptido aislado o recombinante que comprende la secuencia de aminoácidos de SEQ ID NO:276 o fragmentos o subsecuencias de la misma que tienen actividad de aldolasa. En algunas realizaciones, los fragmentos o subsecuencias de la misma tienen una actividad de aldolasa de al menos 0,2 mg MP/mg proteína/hr. En otras realizaciones, los fragmentos o subsecuencias de la misma tienen una actividad de aldolasa de al menos 0,1 mg MP/mg proteína/hr. En algunas realizaciones, la reacción facilitada por uno o más polipéptidos de acuerdo con la invención se lleva a cabo en aproximadamente 1,0 a aproximadamente 10,0 mM de MgCl<sub>2</sub>. En otras realizaciones, la reacción facilitada por uno o más polipéptidos de acuerdo con la invención se lleva a cabo a aproximadamente pH 7,0 a aproximadamente pH 11,5. Incluso en otras realizaciones, la reacción facilitada por uno o más polipéptidos de acuerdo con la invención se lleva a cabo en aproximadamente 0,005% a aproximadamente 1% de detergente polisorbato.

En algunas realizaciones, la reacción es una reacción entre indol-3-piruvato y una fuente de carbono C3. En algunas realizaciones, la reacción preferiblemente produce ácido R-2-hidroxi-2-(indol-3-il-metil)-4-cetoglutárico en lugar de ácido S-2-hidroxi-2-(indol-3-il-metil)-4-cetoglutárico.

35 En algunas realizaciones, el producto obtenido mediante la vía de múltiples etapas es monatina, sales de la misma y combinaciones de la misma.

En otras realizaciones, al menos una de monatina R,R, monatina R-S o una combinación de las mismas se produce en cantidades mayores que monatina S,S o monatina S,R en la vía de múltiples etapas. En algunas realizaciones, la monatina R,R se produce en cantidades mayores que la monatina R,S, la monatina S,S y la monatina S,R en la vía de múltiples etapas.

40 La invención proporciona un método que comprende: producir un producto elegido de monatina, derivados de monatina, sales de la misma y combinaciones de la misma en una vía de múltiples etapas, en donde una reacción en la vía es facilitada por al menos un polipéptido codificado por una secuencia de ácidos nucleicos que comprende una secuencia que tiene un porcentaje de identidad de secuencia de al menos 80% con respecto a SEQ ID NO:275. En algunas realizaciones, el porcentaje de identidad de secuencia es al menos 95%. En otras realizaciones, el porcentaje de identidad de secuencia es 100%.

45 La invención proporciona un método que comprende una reacción que preferiblemente produce ácido R-2-hidroxi-2-(indol-3-il-metil)-4-cetoglutárico en lugar de ácido S-2-hidroxi-2-(indol-3-il-metil)-4-cetoglutárico en donde al menos un polipéptido codificado por una secuencia de ácidos nucleicos que comprende una secuencia que tiene al menos 95% de identidad de secuencia con respecto a SEQ ID NO:276 se utiliza para facilitar una reacción en una vía de múltiples etapas.

50 La invención proporciona un método que comprende: producir un producto elegido de monatina, derivados de monatina, sales de la misma y combinaciones de la misma en una vía de múltiples etapas, en donde una reacción en la vía es facilitada por al menos un polipéptido codificado por una secuencia de ácidos nucleicos que comprende una secuencia que se hidrida en condiciones rigurosas con un ácido nucleico de SEQ ID NO:275.



La invención también proporciona un método que comprende: producir un producto elegido de precursor de monatina, sales del mismo y combinaciones del mismo en una vía de múltiples etapas, en donde una reacción en la vía es facilitada por un polipéptido elegido de un polipéptido aislado o recombinante que comprende la secuencia de aminoácidos de SEQ ID NO:276 o fragmentos o subsecuencias de la misma que tienen actividad de aldolasa, en donde dicho precursor de monatina, sales del mismo y combinaciones del mismo es dulce.

La invención proporciona además un método que comprende: producir un producto elegido de precursor de monatina, sales del mismo y combinaciones del mismo en una vía de múltiples etapas, en donde una reacción en la vía es facilitada por al menos un polipéptido codificado por una secuencia de ácidos nucleicos que comprende una secuencia que tiene un porcentaje de identidad de secuencia de al menos 80% con respecto a SEQ ID NO:275 en donde dicho precursor de monatina, sales del mismo y combinaciones del mismo es dulce.

La invención además proporciona un método que comprende: producir un producto elegido de precursor de monatina, sales del mismo y combinaciones del mismo en una vía de múltiples etapas, en donde una reacción en la vía es facilitada por al menos un polipéptido codificado por una secuencia de ácidos nucleicos que comprende una secuencia que se hidrida en condiciones rigurosas con un ácido nucleico de SEQ ID NO:275 en donde dicho precursor de monatina, sales del mismo y combinaciones del mismo es dulce.

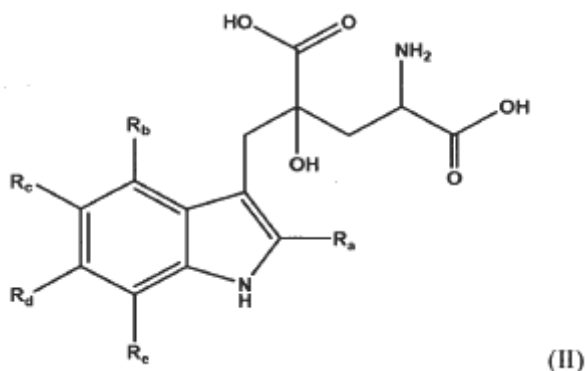
En un esfuerzo por ser concisos, siempre que se identifique que se han formado intermediarios/productos en la memoria descriptiva y reivindicaciones (tales como monatina, precursor de monatina o derivado o derivados de monatina), debe entenderse que debería incluirse la expresión "y/o sales de los mismos" donde corresponda. En otras palabras, por ejemplo, debe entenderse que la frase "indol-3-piruvato se convierte en MP" incluye "ácido indol-3-pirúvico se convierte en MP y/o sales del mismo". El experto en la técnica, de hecho, apreciará que bajo las condiciones de reacción descritas, las sales de los intermediarios/productos están de hecho presentes.

De acuerdo con algunas realizaciones, el método produce una monatina o composición derivada de monatina en donde la monatina o componente derivado de monatina de la composición incluye solo las formas R,R y S,R de monatina o derivado de monatina. El término "solo", cuando se utiliza para indicar que solo se forman ciertos isómeros, significa que la vía produciría solo los isómeros identificados si no se produce la racemización. Por consiguiente, el término "solo" no significa la ausencia de otros isómeros, sino que el experto en la técnica entenderá que otras formas isoméricas pueden estar presentes en una cantidad relativamente pequeña debido a la racemización que puede producirse. De acuerdo con algunas realizaciones, el método produce una composición en donde la monatina o componente derivado de monatina de la composición incluye solo la forma R,R de monatina o derivado de monatina (excepto que se llegue al punto que se produzca la racemización dando como resultado otras formas isoméricas).

Tal como se utiliza en la presente, la frase "composición de monatina" significa composiciones que incluyen uno o más isómeros de monatina; la expresión también puede significar solo una forma isomérica simple de monatina y nada más.

Tal como se utiliza en la presente, la frase "composición derivada de monatina" significa composiciones que incluyen uno o más isómeros de un derivado de monatina; la expresión también puede significar solo una forma isomérica simple del derivado de monatina y nada más.

Tal como se utiliza en la presente, la frase "derivado de monatina" tiene la siguiente estructura:



en donde, R<sub>a</sub>, R<sub>b</sub>, R<sub>c</sub>, R<sub>d</sub> y R<sub>e</sub> representan cada uno independientemente cualquier sustituyente que se seleccione de un átomo de hidrógeno, un grupo hidroxilo, un grupo alquilo C<sub>1</sub>-C<sub>3</sub>, un grupo alcoxi C<sub>1</sub>-C<sub>3</sub>, un grupo amino o un átomo de halógeno, tal como un átomo de yodo, átomo de bromo, átomo de cloro o átomo de flúor. Sin embargo, R<sub>a</sub>, R<sub>b</sub>, R<sub>c</sub>, R<sub>d</sub>, y R<sub>e</sub> no pueden simultáneamente ser todos hidrógeno. De forma alternativa, R<sub>b</sub> y R<sub>c</sub>, y/o R<sub>d</sub> y R<sub>e</sub> juntos pueden formar un grupo alquilenos C<sub>1</sub>-C<sub>4</sub>, respectivamente.

Tal como se utiliza en la presente, "indol-3-piruvato sustituido" significa que uno o más átomos de carbono del anillo indol del indol-3-piruvato está sustituido independientemente por uno o más de los grupos sustituyentes R<sub>a</sub>, R<sub>b</sub>, R<sub>c</sub>, R<sub>d</sub>, y R<sub>e</sub> definidos anteriormente. Sin embargo, R<sub>a</sub>, R<sub>b</sub>, R<sub>c</sub>, R<sub>d</sub>, y R<sub>e</sub> no pueden simultáneamente ser todos hidrógeno. De forma alternativa, R<sub>b</sub> y R<sub>c</sub>, y/o R<sub>d</sub> y R<sub>e</sub> juntos pueden formar un grupo alquileo C<sub>1</sub>-C<sub>4</sub>, respectivamente.

5 Tal como se utiliza en la presente, "triptófano sustituido" significa que uno o más átomos de carbono del anillo indol del triptófano está sustituido independientemente por uno o más de los grupos sustituyentes R<sub>a</sub>, R<sub>b</sub>, R<sub>c</sub>, R<sub>d</sub>, y R<sub>e</sub> definidos anteriormente. Sin embargo, R<sub>a</sub>, R<sub>b</sub>, R<sub>c</sub>, R<sub>d</sub>, y R<sub>e</sub> no pueden simultáneamente ser todos hidrógeno. De forma alternativa, R<sub>b</sub> y R<sub>c</sub>, y/o R<sub>d</sub> y R<sub>e</sub> juntos pueden formar un grupo alquileo C<sub>1</sub>-C<sub>4</sub>, respectivamente. En una  
10 realización, el triptófano sustituido contiene el o los mismos grupos sustituyentes en el anillo indol que el derivado final de monatina.

Más aun, las vías biosintéticas para producir monatina descritas en la presente pueden utilizar un triptófano sustituido para proporcionar derivados de monatina que probablemente sean dulces. En algunas realizaciones, el triptófano sustituido a ser utilizado en las vías biosintéticas descritas en la presente incluyen triptófano y 5-hidroxitriptófano clorado.

15 Por ejemplo, los D-triptófanos clorados, que tienen similitudes estructurales con la monatina R,R, han sido identificados como edulcorantes no nutritivos (particularmente 6-cloro-D-triptófano). De forma similar, se ha descubierto que las formas de monatina halogenadas y sustituidas por hidroxilo son dulces. Ver la Solicitud de Patente de los Estados Unidos Publicada No. 2005/0118317. Grupos halógeno e hidroxilo pueden sustituirse por hidrógeno, particularmente en las posiciones 1-4 del anillo de benceno en el indol de triptófano, sin interferir en las  
20 subsiguientes conversiones a D- o L-triptófano, indol-3-piruvato, MP o monatina. En la bibliografía se ha mostrado que los indoles sustituidos son sustratos adecuados para enzimas PLP y han proporcionado triptófanos sustituidos. Fukuda, D. S., et al., "Production of Substituted L-Tryptophans by Fermentation," *Appl. Environ. Microbiol.*, 21:841-43 (1971). El halógeno no parece dificultar estéricamente el mecanismo catalítico de las subunidades beta de la triptófano sintasa y la enantioespecificidad también permaneció intacta.

25 En algunas realizaciones de la presente invención se proporciona un proceso para producir una composición de monatina que incluye producir indol-3-piruvato a partir de L-triptófano, producir ácido 2-hidroxi 2-(indol-3-ilmetil)-4-cetoglutarico ("precursor de monatina" o "MP") a partir de indol-3-piruvato y producir monatina a partir de MP. La reacción de L-triptófano para producir indol-3-piruvato es facilitada por una enzima que tiene mayor especificidad, mayor actividad o ambas para L-triptófano que para R-MP, monatina R,R o ambos. De acuerdo con ciertas  
30 realizaciones, la reacción de indol-3-piruvato es facilitada por una enzima que tiene actividad de aldolasa R específica y por consiguiente produce R-MP. De acuerdo con ciertas realizaciones, se proporciona una enzima racemasa que puede facilitar la epimerización del producto derivado de aminoácido de la reacción del triptófano de una forma isomérica a otra forma isomérica.

35 En algunas realizaciones de acuerdo con la invención se proporciona un proceso para producir una composición de monatina que incluye producir indol-3-piruvato a partir de L-triptófano, producir ácido 2-hidroxi 2-(indol-3-ilmetil)-4-cetoglutarico ("precursor de monatina" o "MP") a partir de indol-3-piruvato y producir monatina a partir de MP. La reacción de L-triptófano para producir indol-3-piruvato es facilitada por una enzima que tiene mayor especificidad, mayor actividad o ambas para el L-triptófano que para R-MP, monatina R,R o ambos, y la reacción de MP para formar monatina es facilitada por una enzima, que es estereoselectiva para R-MP. El término "estereoselectivo"  
40 significa que una enzima tiene mayor especificidad, mayor actividad o ambas para un isómero, en este caso para R-MP que para S-MP, que otra. En realizaciones preferidas, una enzima estereoselectiva tiene actividad limitada para un isómero en comparación con otra. Actividad "limitada" significa actividad que puede percibirse mínimamente o que no se percibe, por ejemplo como se determina de acuerdo con experimentos proporcionados en la presente.

45 Cabe señalar que, cuando se hace referencia a series de reacciones tales como en los párrafos precedentes, no es necesario que se lleve a cabo de forma explícita cada etapa de la invención; es suficiente que las etapas se lleven a cabo implícitamente. En otras palabras, por ejemplo, el proceso para producir una composición de monatina, que incluye producir indol-3-piruvato a partir de L-triptófano, producir ácido 2-hidroxi 2-(indol-3-ilmetil)-4-cetoglutarico ("precursor de monatina" o "MP") a partir de indol-3-piruvato y producir monatina a partir de MP, en donde cada  
50 reacción es facilitada por una enzima apropiada, puede llevarse a cabo combinando L-triptófano con las enzimas y fijando las condiciones de forma tal que puedan producirse las reacciones enumeradas. En dicho caso, el L-triptófano puede hacerse reaccionar para producir indol-3-piruvato, el indol-3-piruvato producido a partir de la reacción de L-triptófano puede hacerse reaccionar para formar MP y el MP producido a partir de la reacción de indol-3-piruvato puede hacerse reaccionar para formar monatina. El proceso también puede llevarse a cabo, a modo de ejemplo, proporcionando un compuesto que puede producir L-triptófano, en condiciones adecuadas para que ocurra  
55 la producción de L-triptófano y combinar dicho compuesto con enzimas capaces de facilitar la serie de reacciones indicadas en condiciones adecuadas para que se produzcan dichas reacciones. Como otro ejemplo, el proceso puede llevarse a cabo proporcionando un microorganismo manipulado genéticamente para producir monatina de acuerdo con la vía descrita y proporcionando condiciones apropiadas para que se produzca el proceso de fermentación. Por ejemplo, un microorganismo que naturalmente produce grandes cantidades de L-triptófano podría manipularse genéticamente para producir o sobreproducir una o más de las enzimas utilizadas para facilitar  
60

reacciones en la vía hacia monatina y podrían proporcionarse condiciones adecuadas de forma tal que el microorganismo pueda así producir monatina.

5 En otras realizaciones de acuerdo con la invención se proporciona un proceso para producir monatina, en el cual un sustrato forma un L-aminoácido cuando el L-triptófano se convierte en indol-3-piruvato, el indol-3-piruvato reacciona para formar MP (que puede incluir R-MP y S-MP aunque preferiblemente incluye solo o predominantemente R-MP) y el L-aminoácido reacciona para regenerar (también denominado "reciclar") el sustrato cuando la R-MP se convierte en monatina R,R. La reacción de R-MP para formar monatina R,R es facilitada por una aminotransferasa estereoinversora tal como D-metionina aminotransferasa (EC 2.6.1.41) o una enzima que tiene actividad de D-fenilglicina aminotransferasa.

10 En otras realizaciones de acuerdo con la invención se proporciona un proceso para producir una composición de monatina que incluye producir D-triptófano a partir de L-triptófano, producir indol-3-piruvato a partir de D-triptófano, producir R-MP a partir de indol-3-piruvato y producir monatina R,R a partir de R-MP. La producción del D-triptófano a partir de L-triptófano es facilitada por una triptófano racemasa y equivalentes funcionales de la misma. En ciertas realizaciones adicionales, las reacciones de D-triptófano para formar indol-3-piruvato y de MP para formar monatina son facilitadas por la misma enzima. En otras realizaciones adicionales, la reacción de indol-3-piruvato es facilitada por una enzima que tiene actividad de aldolasa R específica y por consiguiente se forma R-MP y las reacciones del D-triptófano para formar indol-3-piruvato y de R-MP para formar monatina R,R son facilitadas por la misma enzima.

20 En algunas realizaciones de acuerdo con la invención se proporciona un proceso para producir un derivado de monatina que incluye producir el derivado de monatina a partir de un indol-3-piruvato sustituido y piruvato, utilizando una enzima que tiene actividad de aldolasa R específica para catalizar la reacción.

Los detalles de una o más realizaciones de la invención quedan establecidos en los dibujos adjuntos y la descripción que figura más adelante. Otras características, objetos y ventajas de acuerdo con la invención serán evidentes a partir de la descripción y los dibujos y a partir de las reivindicaciones.

#### Breve descripción de los dibujos

25 Los siguientes dibujos son ilustrativos de realizaciones de la invención.

La Figura 1 es un diagrama de flujo que muestra un ejemplo de un proceso enzimático para producir monatina R,R a partir de L-triptófano de acuerdo con la invención. En este ejemplo, el proceso incluye utilizar una L-aminotransferasa (ejemplos de los cuales incluyen una L-triptófano aminotransferasa, una aminotransferasa L-aromática, una L-aspartato aminotransferasa y una L-alanina aminotransferasa) en la reacción de L-triptófano que tiene mayor especificidad y/o selectividad para L-triptófano como sustrato que para R-MP y/o el proceso incluye utilizar una L-aminoácido oxidasa con especificidad y/o selectividad limitada para monatina R,R como sustrato. En el ejemplo específico diagramado en la Figura 1, una L-aminotransferasa o L-aminoácido oxidasa convierte L-triptófano en indol-3-piruvato, el indol-3-piruvato se hace reaccionar con una aldolasa R-específica y piruvato para producir monatina de R-alfa cetoácido (R-MP) y la R-MP se convierte en monatina R,R mediante una D-aminotransferasa o una D-aminoácido deshidrogenasa. Como se muestra en la Figura 1, las reacciones son reversibles pero, a los efectos de la invención, no es necesario que las reacciones se lleven a cabo en dirección inversa.

La Figura 2 es un diagrama de flujo que muestra un ejemplo de otro proceso para producir monatina R,R de acuerdo con la invención. En este ejemplo, el proceso incluye utilizar una enzima para convertir R-MP en monatina que es estereoselectiva para R-MP. En el ejemplo específico diagramado en la Figura 2 se muestra la conversión del triptófano en indol-3-piruvato en una reacción inversa. El indol-3-piruvato puede hacerse reaccionar con una aldolasa no estereoespecífica para formar inversamente monatina de alfa cetoácido (R y S-MP). La R-MP se convierte inversamente en monatina R,R mediante una D-aminotransferasa estereoselectiva o una D-aminoácido deshidrogenasa estereoselectiva. Cualquier S-MP que se forme mediante la aldolasa no estereoespecífica puede convertirse nuevamente en indol-3-piruvato si se utiliza una D-aminotransferasa o D-aminoácido deshidrogenasa estereoselectiva. A los efectos de la invención, no es necesario que las reacciones que se muestran como reversibles se lleven a cabo en dirección inversa.

La Figura 3 es un diagrama de flujo que muestra un ejemplo de otro proceso adicional para producir monatina R,R a partir de L-triptófano de acuerdo con la invención. En este ejemplo, el proceso incluye convertir L-triptófano en D-triptófano utilizando una triptófano racemasa y utilizar un producto de D-aminoácido en la reacción acoplado a la reacción de formación de indol-3-piruvato como sustrato en la reacción acoplado a la reacción de formación de monatina R,R. En el ejemplo específico diagramado en la Figura 3, L-triptófano se convierte en D-triptófano mediante una triptófano racemasa en una reacción inversa. El D-triptófano se hace reaccionar con alfa-cetoglutarato ( $\alpha$ -KG) y D-aminotransferasa de amplia especificidad para producir indol-3-piruvato y D-glutamato. El indol-3-piruvato se hace reaccionar con piruvato y una aldolasa R-específica y se convierte en monatina de R-alfa cetoácido (R-MP) y la R-MP se hace reaccionar con D-aminotransferasa de amplia especificidad y D-glutamato para formar monatina R,R y alfa-cetoglutarato ( $\alpha$ -KG). Como se muestra en la Figura 3, cada una de las reacciones son

reversibles pero, a los efectos de la invención, no es necesario que las reacciones se lleven a cabo en dirección inversa.

La Figura 4 es un diagrama de flujo que muestra un ejemplo de otro proceso adicional para producir monatina R,R a partir de L-triptófano de acuerdo con la invención. En este ejemplo, el proceso incluye convertir el L-aminoácido formado en la reacción acoplada a la reacción de L-triptófano en un D-aminoácido; este D-aminoácido actúa como donante de amino para la reacción en la cual la R-MP se convierte en monatina R,R. En el ejemplo específico diagramado en la Figura 4 se hace reaccionar L-triptófano con una L-aminotransferasa y alfa-cetoglutarato para producir indol-3-piruvato y L-glutamato. El indol-3-piruvato se hace reaccionar con piruvato y una aldolasa R-específica y se convierte en monatina de R-alfa cetoácido (R-MP) y la R-MP se hace reaccionar con una D-aminotransferasa de amplia especificidad y D-glutamato para formar monatina R,R y alfa-cetoglutarato. Como se muestra en la Figura 4, las reacciones son reversibles pero, a los efectos de la invención, no es necesario que las reacciones se lleven a cabo en dirección inversa.

La Figura 5 es un diagrama de flujo que muestra un ejemplo de otro proceso adicional para producir monatina R,R a partir de L-triptófano de acuerdo con la invención. En este ejemplo, el proceso incluye facilitar enzimáticamente la conversión de R-MP en monatina R,R utilizando una enzima estereoinversora de forma tal que el L-aminoácido formado mediante la reacción acoplada a la reacción de L-triptófano puede utilizarse como sustrato para la reacción acoplada a la reacción de R-MP con monatina R,R. En el ejemplo específico diagramado en la Figura 5 se hace reaccionar L-triptófano con una L-aminotransferasa y oxaloacetato, piruvato o alfa-cetoglutarato ( $\alpha$ -KG) para producir indol-3-piruvato y L-aspartato (si se utiliza oxaloacetato), L-alanina (si se utiliza piruvato) o L-glutamato (si se utiliza  $\alpha$ -KG). El indol-3-piruvato se hace reaccionar con piruvato y una aldolasa R-específica y se convierte en monatina de R-alfa cetoácido (R-MP) y la R-MP se hace reaccionar con una aminotransferasa estereoinversora y L-aspartato, L-alanina o L-glutamato para formar monatina R,R y oxaloacetato (si se utiliza L-aspartato), piruvato (si se utiliza L-alanina) o alfa-cetoglutarato ( $\alpha$ -KG, si se utiliza L-glutamato). Como se muestra en la Figura 5, las reacciones son reversibles pero, a los efectos de la invención, no es necesario que las reacciones se lleven a cabo en dirección inversa.

La Figura 6 es un diagrama de flujo que muestra un ejemplo de otro proceso adicional para producir monatina R,R de acuerdo con la presente invención. En este ejemplo, el proceso incluye reciclar el L-aminoácido producido en la reacción de formación de indol-3-piruvato con el D-aminoácido utilizado como reactivo con R-MP en la reacción de formación de monatina R,R mediante una serie de reacciones de conversión. En el ejemplo específico diagramado en la Figura 6, el L-triptófano se hace reaccionar reversiblemente con una L-aminotransferasa y oxaloacetato para producir indol-3-piruvato y L-aspartato. El indol-3-piruvato se hace reaccionar reversiblemente con piruvato y una aldolasa R-específica y se convierte en monatina de R-alfa cetoácido (R-MP) y la R-MP se hace reaccionar reversiblemente con una D-aminotransferasa y D-alanina para formar monatina R,R y piruvato. El L-aspartato se convierte en L-alanina y  $\text{CO}_2$  utilizando una aspartato 4-descarboxilasa. La L-alanina se convierte en D-alanina con una alanina racemasa. A los efectos de la invención, no es necesario que las reacciones que se muestran como reversibles se lleven a cabo en dirección inversa.

La Figura 7 es un diagrama de flujo que muestra un ejemplo de otro proceso adicional para producir monatina R,R de acuerdo con la presente invención. En este ejemplo, el proceso incluye hacer avanzar la reacción de L-triptófano (es decir, impulsar la reacción hacia la producción de indol-3-piruvato) mediante la conversión del producto derivado de L-aminoácido de dicha reacción en otro producto. En este ejemplo, el producto derivado L-aspartato de L-aminoácido se convierte en L-alanina en una reacción irreversible utilizando una descarboxilasa. En el ejemplo específico diagramado en la Figura 7 se hace reaccionar L-triptófano reversiblemente con una L-aminotransferasa y con alfa-cetoglutarato ( $\alpha$ -KG) u oxaloacetato para producir indol-3-piruvato y L-glutamato (si se utiliza  $\alpha$ -KG) o L-aspartato (si se utiliza oxaloacetato). Se hace reaccionar indol-3-piruvato reversiblemente con piruvato y una aldolasa R-específica y se convierte en monatina de R-alfa cetoácido (R-MP). La R-MP se hace reaccionar reversiblemente con una D-aminotransferasa y un D-aminoácido para formar monatina R,R y cualquiera de oxaloacetato, piruvato o  $\alpha$ -KG. El L-glutamato o L-aspartato que fue producto de la reacción de la L-aminotransferasa se convierte en 4-aminobutanoato y  $\text{CO}_2$  (si el sustrato es L-glutamato) o en  $\beta$ -alanina y  $\text{CO}_2$  (si el sustrato es L-aspartato) utilizando un ácido glutámico o una aspartato descarboxilasa. A los efectos de la invención, no es necesario que las reacciones que se muestran como reversibles se lleven a cabo en dirección inversa.

La Figura 8 es un diagrama de flujo que muestra un ejemplo de otro proceso adicional para producir monatina R,R de acuerdo con la presente invención. En este ejemplo, el proceso incluye reciclar el producto derivado de aminoácido de la reacción de L-triptófano con el reactivo aminoácido de la reacción de la R-MP mediante una serie de reacciones de conversión. En el ejemplo específico diagramado en la Figura 8 se hace reaccionar L-triptófano reversiblemente con una L-aminotransferasa y con alfa-cetoglutarato ( $\alpha$ -KG) para producir indol-3-piruvato y L-glutamato. Se hace reaccionar indol-3-piruvato reversiblemente con piruvato y una aldolasa R-específica y se convierte en monatina de R-alfa cetoácido (R-MP). La R-MP se hace reaccionar reversiblemente con una D-aminotransferasa y D-alanina para formar monatina R,R y piruvato. Se utilizan una L-alanina aminotransferasa y piruvato para convertir reversiblemente el L-glutamato que fue producto de la reacción de la L-aminotransferasa nuevamente en  $\alpha$ -KG, con L-alanina como un coproducto. Una alanina racemasa reversiblemente convierte la L-

alanina en la D-alanina que es útil en la tercera reacción (la reacción de D-aminotransferasa). A los efectos de la invención, no es necesario que las reacciones que se muestran como reversibles se lleven a cabo en dirección inversa.

La Figura 9 diagrama de bloques de un sistema informático.

- 5 La Figura 10 es un diagrama de flujo que ilustra una realización de un proceso para comparar una nueva secuencia de proteína o nucleótido con una base de datos de secuencias para determinar los niveles de homología entre la nueva secuencia y las secuencias en la base de datos.

La Figura 11 es un diagrama de flujo que ilustra una realización de un proceso en un ordenador para determinar si dos secuencias son homólogas.

- 10 La Figura 12 es un diagrama de flujo que ilustra una realización de un proceso de identificación 300 para detectar la presencia de una característica en una secuencia.

Las Figuras 13 y 14 juntas ilustran las actividades de 58 aldolasas diferentes (cada una identificada con su número específico de SEQ ID) en la formación de precursor de monatina (MP) tal como se midió mediante LC/MS/MS.

- 15 La Figura 15 ilustra el efecto de ditiotreitól en la producción de monatina por parte del polipéptido con actividad de aldolasa de SEQ ID NO:88.

Símbolos de referencia similares en los distintos dibujos indican elementos similares.

### Descripción detallada

Precedentemente se han descrito varias realizaciones que serán descritas más detalladamente más adelante. Las realizaciones de la invención incluyen uno o más de los aspectos descritos.

### 20 Abreviaturas y términos

- Las siguientes explicaciones de términos y métodos se proporcionan para describir mejor la presente divulgación y guiar a los expertos en la materia en la puesta en práctica de la presente divulgación. Tal como se utiliza en la presente, "incluyendo" significa "que comprende". Además, las formas singulares "un", "una" o "el" o "la" incluyen las referencias al plural, a menos que el contexto lo indique claramente de otra manera. Por ejemplo, la referencia a "que comprende una proteína" incluye una o varias de dichas proteínas, y la referencia a "que comprende las células" incluye la referencia a una o más células y equivalentes de las mismas conocidas por los expertos en la materia y así sucesivamente. El término "aproximadamente" abarca el rango de error experimental presente en cualquier medición. A menos que se indique lo contrario, se supone que todos los números correspondientes a mediciones tienen la palabra "aproximadamente" delante de ellos, aun si la palabra "aproximadamente" no se utiliza expresamente.
- 25
- 30

- Sustitución conservadora: una sustitución de un aminoácido por otro aminoácido en un polipéptido, la cual tiene poco a ningún impacto sobre la actividad del polipéptido. La sustitución se considera conservadora independientemente de si los aminoácidos intercambiados tienen una apariencia estructural o funcional similar. Por ejemplo, de forma ideal, un polipéptido de triptófano aminotransferasa que incluye una o más sustituciones conservadoras retiene actividad de triptófano aminotransferasa. Puede producirse un polipéptido para que contenga una o más sustituciones conservadoras mediante manipulación de la secuencia nucleotídica que codifica este polipéptido utilizando, por ejemplo, procedimientos convencionales tales como mutagénesis dirigida a sitios o PCR u otros métodos conocidos en la técnica.
- 35

- Ejemplos no taxativos de aminoácidos que pueden sustituirse por un aminoácido original en una proteína y que se consideran sustituciones conservadoras si hay poco a ningún impacto sobre la actividad del polipéptido incluyen: Ala sustituido por ser o thr; arg sustituido por gln, his o lys; asn sustituido por glu, gln, lys, his, asp; asp sustituido por asn, glu o gln; cys sustituido por ser o ala; gln sustituido por asn, glu, lys, his, asp o arg; glu sustituido por asn, gln lys o asp; gly sustituido por pro; his sustituido por asn, lys, gln, arg, tyr; ile sustituido por leu, met, val, phe; leu sustituido por ile, met, val, phe; lys sustituido por asn, glu, gln, his, arg; met sustituido por ile, leu, val, phe; phe sustituido por trp, tyr, met, ile o leu; ser sustituido por thr, ala; thr sustituido por ser o ala; trp sustituido por phe, tyr; tyr sustituido por his, phe o trp; y val sustituido por met, ile, leu.
- 40
- 45

- Información adicional acerca de sustituciones conservadoras puede encontrarse en, entre otros lugares, Ben-Bassat *et al.*, (*J. Bacteriol.* 169:751-7, 1987), O'Regan *et al.*, (*Gene* 77:237-51, 1989), Sahin-Toth *et al.*, (*Proteína Sci.* 3:240-7, 1994), Hochuli *et al.*, (*Bio/Technology* 6: 1321-5, 1988), el documento WO 00/67796 (Curd *et al.*) y en libros de textos convencionales de genética y de biología molecular.
- 50

Derivado: A los efectos de la memoria descriptiva y las reivindicaciones, una sustancia es "derivada" de un organismo o fuente si una o más de las siguientes afirmaciones es verdadera: 1) la sustancia está presente en el

organismo/fuente; 2) la sustancia se elimina del huésped nativo; o 3) la sustancia se retira del huésped nativo y evoluciona, por ejemplo, mediante mutagénesis.

Aislado: El término “aislado”, tal como se utiliza en la presente, se refiere a cualquier sustancia que se retira de su huésped nativo; la sustancia no necesita ser purificada. Por ejemplo “ácido nucleico aislado” se refiere a un ácido nucleico natural que no está inmediatamente contiguo a ambas secuencias a las cuales está inmediatamente contiguo (una en el extremo 5' y una en el extremo 3') en el genoma natural del organismo del que deriva. Por ejemplo, un ácido nucleico aislado puede ser, a modo no taxativo, una molécula de ADN recombinante de cualquier longitud, siempre que una de las secuencias del ácido nucleico que normalmente se encuentran inmediatamente flanqueando esa molécula de ADN recombinante en un genoma natural se retire o esté ausente. De esta forma, un ácido nucleico aislado incluye, a modo no taxativo, un ADN recombinante que existe como una molécula separada (tal como un ADNc o un fragmento de ADN genómico producido por PCR o tratamiento con endonucleasas de restricción) independiente de otras secuencias, así como ADN recombinante que se incorpora en un vector, un plásmido que se replica de manera autónoma, un virus (por ejemplo, un retrovirus, adenovirus o virus herpético) o en el ADN genómico de un procarionte o eucariote. Además, un ácido nucleico aislado puede incluir una molécula de ADN recombinante que forma parte de una secuencia de ácido nucleico híbrido o de fusión.

Tal como se utiliza en la presente, el término “aislado” significa que el material (tal como una proteína o ácido nucleico de acuerdo con la invención) se retira de su ambiente original (tal como el ambiente natural si es natural). Por ejemplo, un polinucleótido o polipéptido natural presente en un animal vivo no está aislado, pero el mismo polinucleótido o polipéptido, separado de algunos o todos los materiales coexistentes en el sistema natural, está aislado. Dichos polinucleótidos pueden ser parte de un vector y/o dichos polinucleótidos o polipéptidos pueden ser parte de una composición y aún así estar aislados, ya que dicho vector o composición no es parte de su ambiente natural.

El término “aislado”, tal como se utiliza en la presente con relación al ácido nucleico, incluye también cualquier ácido nucleico artificial, ya que las secuencias de ácido nucleico artificiales no se encuentran en la naturaleza y no tienen secuencias inmediatamente contiguas en un genoma natural. Por ejemplo, el ácido nucleico, artificial tal como un ácido nucleico manipulado genéticamente, se considera que es un ácido nucleico aislado. El ácido nucleico manipulado genéticamente puede obtenerse utilizando técnicas de clonación molecular comunes o de síntesis química de ácido nucleico. El ácido nucleico artificial aislado puede ser independiente de otras secuencias o incorporarse en un vector, un plásmido que se replica de manera autónoma, un virus (por ejemplo, un retrovirus, adenovirus o virus herpético) o el ADN genómico de un procarionte o eucariote. Además, un ácido nucleico artificial puede incluir una molécula de ácido nucleico que forma parte de una secuencia de ácido nucleico híbrido o de fusión.

Purificado: El término “purificado”, tal como se utiliza en la presente, no requiere pureza absoluta, sino que pretende ser un término relativo. Así, por ejemplo, una preparación purificada de polipéptido o ácido nucleico puede ser aquella en la que el polipéptido o ácido nucleico en cuestión esté en una concentración mayor que la que el polipéptido o ácido nucleico estaría en su ambiente natural dentro de un organismo o una concentración mayor que la del medio ambiente de cual se retiró.

Se han purificado convencionalmente ácidos nucleicos individuales obtenidos de una biblioteca hasta lograr una homogeneidad electroforética. Las secuencias obtenidas de estos clones no pudieron obtenerse directamente de la biblioteca o del ADN humano total. Los ácidos nucleicos purificados de acuerdo con la invención se purificaron a partir del resto del ADN genómico en el organismo al menos  $10^4$ - $10^6$  veces. En algunas realizaciones, el término “purificado” incluye ácidos nucleicos que fueron purificados a partir del resto de ADN genómico o a partir de otras secuencias en una biblioteca u otro ambiente por al menos una orden de magnitud, tal como, en algunas realizaciones, dos o más órdenes o cuatro o cinco órdenes de magnitud.

Aminoácido: “Aminoácido” o “secuencia de aminoácidos”, tal como se utiliza en la presente, se refiere a una secuencia de oligopéptidos, péptidos, polipéptidos o proteínas o a un fragmento, porción o subunidad de cualquiera de éstos, y a moléculas naturales o sintéticas. “Aminoácido” o “secuencia de aminoácidos” incluye una secuencia de oligopéptidos, péptidos, polipéptidos o proteínas o un fragmento, porción o subunidad de cualquiera de éstos, y moléculas naturales o sintéticas. El término “polipéptido”, tal como se usa en la presente, se refiere a aminoácidos unidos entre sí por enlaces peptídicos o enlaces peptídicos modificados, es decir, isoésteres peptídicos, y pueden contener aminoácidos modificados que no sean los 20 aminoácidos codificados por genes. Los polipéptidos pueden modificarse por procesos naturales, tal como procesamiento post-translacional o por técnicas de modificación química que se conocen bien en la técnica. Pueden ocurrir modificaciones en cualquier parte en el polipéptido, incluida la estructura central del péptido, las cadenas laterales de aminoácidos y los extremos amino o carboxilo. Se apreciará que el mismo tipo de modificación puede estar presente en el mismo grado o en grados diferentes en varios sitios en un polipéptido dado. Asimismo, un polipéptido dado puede tener muchos tipos de modificaciones. Las modificaciones incluyen acetilación, acilación, ADP-ribosilación, amidación, unión covalente de flavina, unión covalente de un resto hemo, unión covalente de un nucleótido o derivado de nucleótidos, unión covalente de un lípido o derivado de lípidos, unión covalente a fosfatidilinositol, ciclización reticulante, formación de enlaces de disulfuro, desmetilación, formación de enlaces intermoleculares covalentes, formación de cisteína, formación de

piroglutamato, formilación, gama-carboxilación, glicosilación, formación de anclajes GPI, hidroxilación, yodación, metilación, miristilación, oxidación, pegilación, procesamiento de glucano hidrolasa, fosforilación, prenilación, racemización, selenoilación, sulfación y adición de aminoácidos mediada por ARN de transferencia a proteína tal como arginilación. (Ver Creighton, T. E., *Proteins - Structure and Molecular Properties* 2da Ed., W.H. Freeman and Company, New York (1993); *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York, págs. 1-12 (1983)). Los péptidos y polipéptidos de acuerdo con la invención también incluyen todas las formas "miméticas" y "peptidomiméticas", como se describe en mayor detalle más adelante.

Polipéptido que tiene una actividad de aldolasa: "Polipéptido que tiene una actividad de aldolasa" significa un polipéptido que solo o en asociación con uno o más polipéptidos adicionales (que tienen la misma secuencia o una secuencia diferente) es una proteína con la actividad enzimática de una aldolasa.

Recombinante: Polipéptidos o proteínas "recombinantes" se refiere a polipéptidos o proteínas producidas por técnicas de ADN recombinante; es decir, producidas a partir de células transformadas por un constructo de ADN exógeno que codifica el polipéptido o proteína deseada. Polipéptidos o proteínas "sintéticas" son aquellas preparadas por síntesis química. Los métodos de síntesis química de péptidos también pueden usarse para sintetizar el polipéptido o los fragmentos de acuerdo con la invención. Dicho método se conoce en la técnica desde principios de los años sesenta (Merrifield, R. B., *J. Am. Chem. Soc.*, **85**:2149-2154, 1963) (Ver también Stewart, J. M. y Young, J. D., *Solid Phase Peptide Synthesis*, 2da Ed., Pierce Chemical Co., Rockford, Ill., págs. 11-12)) y se emplearon recientemente en kits de diseño y síntesis de laboratorio disponibles en el comercio (Cambridge Research Biochemicals). Dichos kits de laboratorio disponibles comercialmente han utilizado generalmente los descubrimientos de H. M. Geysen et al., *Proc. Natl. Acad. Sci., Estados Unidos*, **81**:3998 (1984) y proporcionan péptidos de sintetización sobre las puntas de múltiples "varillas" o "clavijas", todas las cuales están conectadas a una única placa.

Básicamente idénticos: La frase "básicamente idénticos", en el contexto de dos ácidos nucleicos o polipéptidos, se refiere a dos o más secuencias que tienen, por ejemplo, al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más identidad (de secuencia) de residuos de nucleótidos o aminoácidos cuando se comparan y se alinean para una correspondencia máxima, según se mide utilizando uno de los algoritmos de comparación de secuencias conocidos o mediante inspección visual. Las secuencias son básicamente idénticas en la longitud completa de las regiones de codificación.

Además, una secuencia de aminoácidos "básicamente idéntica" es una secuencia que difiere de una secuencia de referencia por una o más sustituciones, eliminaciones o inserciones de aminoácidos conservadoras o no conservadoras. En algunas realizaciones, la sustitución ocurre en un sitio que no es el sitio activo de la molécula o, alternativamente, la sustitución ocurre en un sitio que es el sitio activo de la molécula, siempre que el polipéptido retenga esencialmente sus propiedades funcionales (enzimáticas). Una sustitución de aminoácidos conservadora, por ejemplo, sustituye un aminoácido por otro de la misma clase (tal como la sustitución de un aminoácido hidrófobo, tal como isoleucina, valina, leucina o metionina, por otro, o la sustitución de un aminoácido polar por otro, tal como la sustitución de arginina por lisina, ácido glutámico por ácido aspártico o glutamina por asparagina). Pueden eliminarse uno o más aminoácidos, por ejemplo, de un polipéptido de aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa, lo que resulta en la modificación de la estructura del polipéptido, sin alterar considerablemente su actividad biológica. Por ejemplo, los aminoácidos amino o carboxilo-terminales que no son necesarios para la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, pueden eliminarse. Las secuencias de polipéptidos modificados de acuerdo con la invención pueden evaluarse en busca de actividad biológica de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, mediante diferentes métodos, incluyendo poner en contacto la secuencia de polipéptidos modificados con un sustrato y determinar si el polipéptido modificado disminuye la cantidad de sustrato específico en el ensayo o aumenta los bioproductos de la reacción enzimática de un polipéptido funcional de aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa con el sustrato.

Fragmento: Un "fragmento", tal como se usa en la presente con respecto a una proteína o polipéptido o ácido nucleico, es una porción de la proteína, polipéptido o ácido nucleico, respectivamente. Los fragmentos pueden tener el mismo o básicamente la misma secuencia de aminoácidos o ácidos nucleicos que la secuencia de proteínas, polipéptidos o ácidos nucleicos más larga de la cual deriva el fragmento. También se incluyen los fragmentos que tienen estructuras tridimensionales diferentes en comparación con las de las proteínas, polipéptidos o ácido nucleicos más largos. Un ejemplo de esto es una molécula "pro-forma", tal como una proproteína de baja actividad que puede modificarse por escisión para producir una enzima madura con una actividad considerablemente más alta. Un fragmento de una proteína o polipéptido puede ser una porción enzimáticamente activa de una proteína o polipéptido.

Aminotransferasa estereoinversora: Una "aminotransferasa estereoinversora" es un polipéptido capaz de producir preferiblemente o selectivamente un producto de aminoácido quiral (tal como monatina) mientras que utiliza un sustrato de quiralidad opuesta como donante de amino. Por ejemplo, una aminotransferasa estereoinversora puede ser una D-fenilglicina aminotransferasa (también denominada D-4-hidroxifenilglicina aminotransferasa) que utiliza preferiblemente o selectivamente L-glutamato como sustrato para producir monatina R,R. Ejemplos no taxativos de

aminotransferasas estereoinversoras incluyen D-metionina aminotransferasa (EC 2.6.1.41) y enzimas que tienen actividad de D-fenilglicina aminotransferasa o actividad de D-4-hidroxifenilglicina aminotransferasa.

5 La invención proporciona polipéptidos con actividad de aldolasa, incluyendo actividad de piruvato tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa, polinucleótidos que los codifican y métodos para preparar y utilizar estos polinucleótidos y polipéptidos. En algunas realizaciones, la invención también proporciona enzimas aldolasa, tales como enzimas piruvato aldolasa, tal como HMG y/o KHG aldolasa, polinucleótidos que codifican estas enzimas, el uso de dichos polinucleótidos y polipéptidos.

10 En algunas realizaciones, la invención proporciona aldolasas, tales como piruvato aldolasas, HMG y/o KHG aldolasas, modificadas o evolucionadas con una actividad específica mayor en comparación con las aldolasas no modificadas ni evolucionadas, respectivamente.

15 En algunas realizaciones se proporcionan aldolasas, tales como una piruvato aldolasa, tal como, a modo no taxativo una HMG y/o una KHG aldolasa, que facilitan la producción de un 2-ceto-glutarato 3,4-sustituido. En una realización, la invención proporciona un método para preparar un 2-ceto-glutarato 3,4-sustituido que comprende: (a) proporcionar un polipéptido que tiene una actividad de aldolasa, tal como una actividad de piruvato aldolasa, tal como, a modo no taxativo, una actividad de HMG aldolasa y/o KMG aldolasa; (b) proporcionar un donante y un compuesto aceptor; y (c) poner en contacto el polipéptido de la etapa (a) con los compuestos de la etapa (b) en condiciones en las cuales la aldolasa cataliza la síntesis de un 2-ceto-glutarato 3,4-sustituido, en donde opcionalmente el donante y el aceptor son un piruvato o un donante de piruvato y un aceptor  $\alpha$ -cetoácido, una cetona y/o un aldehído.

20 En otra realización de la invención, una piruvato aldolasa, tal como una HMG y/o una KHG aldolasa, puede utilizarse junto con una D-aminotransferasa para preparar un ácido D-glutámico 4-sustituido o un derivado del mismo. Un ácido D-glutámico 4-sustituido y/o un derivado del mismo puede utilizarse como un antibiótico, dado que se ha descubierto que estos compuestos inhiben la glutamato racemasa bacteriana. En una realización, la invención proporciona un método para preparar un ácido D-glutámico 4-sustituido que comprende: (a) proporcionar un polipéptido que tiene una actividad de aldolasa, tal como una actividad de piruvato aldolasa, tal como, a modo no taxativo, una actividad de HMG aldolasa y/o KMG aldolasa; (b) proporcionar un aceptor  $\alpha$ -cetoácido y un piruvato o un donante de piruvato; y (c) poner en contacto el polipéptido de la etapa (a) con los compuestos de la etapa (b) en condiciones en las cuales la aldolasa cataliza la síntesis de un ácido D-glutámico 4-sustituido, en donde opcionalmente el polipéptido tiene actividad de piruvato aldolasa, HMG aldolasa y/o KHG aldolasa y en donde opcionalmente el método además comprende el uso de una D-aminotransferasa.

30 En algunas realizaciones la invención proporciona composiciones (tales como preparaciones de enzimas, alimentos y aditivos para alimentos, piensos y aditivos para piensos, bebidas y aditivos para bebidas, fármacos y aditivos para fármacos, y complementos alimentarios) que comprenden las enzimas, polipéptidos o polinucleótidos de acuerdo con la invención. Estas composiciones pueden formularse en una variedad de formas, tales como líquidos, geles, píldoras, comprimidos, aerosoles, películas, micelas, polvos, alimento, granulados de pienso o formas encapsuladas, incluyendo formas nanoencapsuladas.

35 Ensayos para medir la actividad de aldolasa, incluyendo actividad de piruvato tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa, tal como para determinar si un polipéptido tiene actividad de aldolasa, incluyendo actividad de piruvato tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa, son bien conocidos en la técnica y están dentro del alcance de acuerdo con la invención; ver E.E. Dekker & R.P. Kitson, J. Biol. Chem. 267, 10507-10514, 1992; Taha T S, Deits TL, Purification and characterization of 2-keto-3-deoxy-6-phosphogluconate aldolase from *Azotobacter vinelandii*: evidence that the enzyme is bifunctional towards 2-keto-4-hydroxy glutarate cleavage, Biochem Biophys Res Commun. 1994 Apr. 15; 200(1):459-66; Dekker EE, Kobes RD, Grady SR, 2-keto-4-hydroxyglutarate aldolase from bovine liver, Methods Enzymol. 1975;42:280-5; Dekker EE, Nishihara H, Grady SR, Methods Enzymol. 1975;42:285-90, 2-keto-4-hydroxyglutarate aldolase from *Escherichia coli*; Nishihara H, Dekker EE, Biochim Biophys Acta. 1969 Jul 8;185(1):255-7, A stereospecific 2-keto-4-hydroxyglutarate aldolase from *Escherichia coli*. Un ejemplo de un ensayo adecuado para determinar si un polipéptido tiene actividad de aldolasa, tal como actividad de piruvato aldolasa, tal como HMG y/o KHG aldolasa se describe en el Ejemplo 3.

40 En algunas realizaciones, las aldolasas de la invención pueden utilizarse de forma efectiva con una variedad de condiciones de pH, incluyendo por ejemplo, de un rango de aproximadamente 3,0 a aproximadamente 12,0. En otras realizaciones, las aldolasas de la invención pueden utilizarse a aproximadamente pH 3,0, 4,0, 4,5, 5,0, 5,5, 6,0, 6,5, 7,0, 7,5, 8,0, 8,5, 9,0, 9,5, 10,0, 10,5, 11,0, 11,5 o aproximadamente 12,0. También pueden resultar ventajosas las condiciones de reacción llevadas a cabo en condiciones ácidas o alcalinas, tales como en ciertas aplicaciones industriales o farmacéuticas de enzimas de acuerdo con la invención.

55 La invención proporciona polipéptidos de aldolasa, tales como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención en una variedad de formas y formulaciones. En los métodos de acuerdo con la invención, se utilizan polipéptidos de aldolasa, tales como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención en una variedad de formas y formulaciones. Por ejemplo, polipéptidos purificados de aldolasa, tales como



- piruvato aldolasa, tal como HMG y/o KHG aldolasa pueden utilizarse en preparaciones de enzimas desplegadas en la producción de ácido R-2-hidroxi 2-(indol-3-ilmetil)-4-cetoglutárico (R-MP) y ciertos estereoisómeros de monatina, tales como monatina R,R y S,R, y sales de los mismos, así como ciertos estereoisómeros de derivados de monatina, tales como las configuraciones de derivados de monatina R,R y S,R, y sales de los mismos o aplicaciones farmacéuticas o de auxiliar alimenticio. De forma alternativa, las enzimas de acuerdo con la invención pueden utilizarse directamente en procesos para producir ácido R-2-hidroxi 2-(indol-3-ilmetil)-4-cetoglutárico (R-MP) y ciertos estereoisómeros de monatina, tales como monatina R,R y S,Ra, y sales de los mismos, así como ciertos estereoisómeros de derivados de monatina, tales como las configuraciones de derivados de monatina R,R y S,R, y sales de los mismos, para procesar alimentos, líquidos, piensos y similares.
- 5
- 10 En algunas realizaciones, polipéptidos de aldolasa, tales como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención pueden expresarse en un microorganismo utilizando procedimientos conocidos en la técnica. En algunas realizaciones, la polipéptidos de aldolasa, tales como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención puede inmovilizarse en un soporte sólido antes de su uso en los métodos de acuerdo con la invención. Métodos para inmovilizar enzimas sobre soportes sólidos son comúnmente conocidos en la técnica, por ejemplo J. Mol. Cat. B: Enzymatic 6 (1999) 29-39; Chivata et al. Biocatalysis: Immobilized cells and enzymes, J. Mol. Cat. 37 (1986) 1-24; Sharma et al., Immobilized Biomaterials Techniques and Applications, Angew. Chem. Int. Ed. Engl. 21 (1982) 837-54; Laskin (Ed.), Enzymes and Immobilized Cells in Biotechnology.
- 15

#### Ácidos nucleicos, sondas y moléculas inhibitoras

- La invención proporciona ácidos nucleicos aislados y recombinantes de la SEQ ID NO: 275; ácidos nucleicos que codifican polipéptidos, incluidas las secuencias de polinucleótidos de acuerdo con la invención; incluidos cassettes de expresión tales como vectores de expresión y varios vehículos de clonación que comprenden ácidos nucleicos de acuerdo con la invención. La divulgación incluye métodos para descubrir, identificar o aislar secuencias de polipéptidos de aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa nuevas utilizando los ácidos nucleicos de acuerdo con la invención. En algunas realizaciones, la invención también incluye métodos para inhibir la expresión de genes y transcriptos que codifican aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa utilizando los ácidos nucleicos de acuerdo con la invención.
- 20
- 25

- Adicionalmente se proporcionan métodos para modificar los ácidos nucleicos de acuerdo con la invención, que incluyen hacer variantes de ácidos nucleicos de acuerdo con la invención mediante, por ejemplo, reensamblaje de ligadura sintética, sistema de evolución dirigida optimizada y/o mutagénesis por saturación tal como mutagénesis por saturación de sitio génico (GSSM). La expresión "mutagénesis por saturación", Mutagénesis por Saturación de Sitio Génico 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. La expresión "sistema de evolución dirigida optimizada" o "evolución dirigida optimizada" incluye un método para reensamblar fragmentos de secuencias de ácido nucleico relacionadas, tales como genes relacionados y se explica en detalle más adelante. La expresión "reensamblaje de ligadura sintética" o "SLR" incluye un método para ligar fragmentos de oligonucleótidos de forma no estocástica y se explica en detalle más adelante. El término "variante" se refiere a polinucleótidos o polipéptidos de acuerdo con la invención modificados en uno o más pares de bases, codones, intrones, exones o residuos de aminoácidos (respectivamente) que aún conservan la actividad biológica de una aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de acuerdo con la invención. Las variantes pueden producirse mediante varios medios incluidos métodos tales como, por ejemplo, PCR propensa a errores, transposición, mutagénesis dirigida mediante oligonucleótidos, PCR de ensamblaje, mutagénesis por PCR sexual, mutagénesis in vivo, mutagénesis por cassette, mutagénesis recursiva de conjunto, mutagénesis exponencial de conjunto, mutagénesis sitio específica, reensamblaje génico, GSSM y combinaciones de los mismos.
- 30
- 35
- 40

- Los ácidos nucleicos de acuerdo con la invención pueden hacerse, aislarse y/o manejarse mediante, por ejemplo, clonación y expresión de bibliotecas de ADNc, amplificación de ADN mensajero o genómico mediante PCR y similares. Por ejemplo, las secuencias de acuerdo con la invención se derivaron inicialmente de fuentes ambientales. Por consiguiente, en algunas realizaciones, la invención proporciona ácidos nucleicos que codifican aldolasa, tales como piruvato aldolasa, tales como enzima HMG y/o KHG aldolasa y los polipéptidos codificados por los mismos, preferiblemente derivados de una fuente común, tal como una ambiental, cultivo mixto o una fuente bacteriana.
- 45

- Al poner en práctica los métodos de acuerdo con la invención, se pueden modificar genes homólogos utilizando un ácido nucleico plantilla, tal como se describe en la presente. En algunas realizaciones, la invención puede ponerse en práctica junto con cualquier método o protocolo o dispositivo conocido en la técnica, los cuales se han descrito en la literatura científica y de patentes.
- 50

- Las frases "ácido nucleico" o "secuencia de ácidos nucleicos", tal como se usan en la presente, se refieren a un oligonucleótido, nucleótido, polinucleótido o a un fragmento de cualquiera de estos, de ADN o ARN de origen genómico o sintético, los cuales pueden ser de una sola hebra o de doble hebra y pueden representar una cadena sentido o antisentido (complementaria) con respecto al ácido nucleico peptídico (PNA) o a cualquier material similar a ADN o similar a ARN de origen natural o sintético. Las frases "ácido nucleico" o "secuencia de ácidos nucleicos" incluyen oligonucleótidos, nucleótidos, polinucleótidos o un fragmento de cualquiera de estos, de ADN o ARN (tal
- 55

5 como ARNm, ARNr, ARNt, ARNi) de origen genómico o sintético, los cuales pueden ser de una sola hebra o de doble hebra y pueden representar una cadena sentido o antisentido con respecto a ácido nucleico peptídico (PNA) o a cualquier material similar a ADN o similar a ARN de origen natural o sintético, incluidos, por ejemplo, ARNi, ribonucleoproteínas (tales como ARNi de doble cadena, tales como RNPI). El término abarca ácidos nucleicos, es  
 10 dicen, oligonucleótidos que contienen análogos conocidos de nucleótidos naturales. El término además abarca estructuras similares a ácido nucleico con cadenas principales sintéticas, ver 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 un polidesoxinucleótido de una sola cadena o dos cadenas de polidesoxinucleótidos complementarias que pueden sintetizarse químicamente. Dichos oligonucleótidos sintéticos no  
 15 tienen fosfato en 5' y por lo tanto no se unirán a otro oligonucleótido sin agregar un fosfato con un ATP en presencia de una cinasa. Un oligonucleótido sintético puede unirse a un fragmento que no ha sido desfosforilado.

Una "secuencia codificadora de" o una "secuencia de nucleótidos que codifica" un polipéptido o proteína específica, es una secuencia de ácidos nucleicos que puede transcribirse y traducirse a un polipéptido o proteína cuando se  
 20 coloca bajo el control de secuencias reguladoras apropiadas. El término "gen" significa el segmento de ADN que participa en la producción de una cadena de polipéptidos; el mismo incluye regiones que preceden y que siguen a la región codificadora (líder y trasera) así como también, cuando corresponde, secuencias intercaladas (intrones) entre segmentos codificadores individuales (exones). Una secuencia promotora está "unida operativamente a" una secuencia codificadora cuando la ARN polimerasa que inicia la transcripción en el promotor transcribirá la secuencia codificadora a ARNm. "Unido operativamente" tal como se usa en la presente se refiere a una relación entre dos o  
 25 más segmentos de ácido nucleico (tales como ADN). Puede referirse a la relación funcional de la secuencia reguladora transcripcional con una secuencia transcrita. Por ejemplo, un promotor está unido operativamente a una secuencia codificadora, tal como un ácido nucleico de acuerdo con la invención, si estimula o modula la transcripción de la secuencia codificadora en una célula huésped apropiada u otros sistema de expresión. Generalmente, las secuencias reguladoras transcripcionales promotoras que están unidas operativamente a una secuencia transcrita están físicamente contiguas a la secuencia transcrita, es decir, actúan en *cis*. Sin embargo, algunas secuencias reguladoras transcripcionales, tales como potenciadores, no deben estar físicamente contiguos o ubicados de forma cercana a las secuencias codificadoras, cuya transcripción potencian.

La expresión "cassette de expresión" tal como se usa en la presente se refiere a una secuencia de nucleótidos que es capaz de afectar la expresión de un gen estructural (es decir, una secuencia codificadora de proteína, tal como una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de acuerdo con la invención) en un huésped compatible con dichas secuencias. Los cassettes de expresión incluyen al menos un promotor unido operativamente con la secuencia codificadora de polipéptido; y, opcionalmente, con otras secuencias, tales como señales de finalización de la transcripción. También se pueden usar factores adicionales necesarios o útiles para producir la expresión, tales como potenciadores, factores alfa. Por consiguiente, los cassettes de expresión también  
 35 incluyen plásmidos, vectores de expresión, virus recombinantes, cualquier forma de vector de "ADN desnudo" recombinante y similares. Un "vector" comprende un ácido nucleico que puede infectar, transfectar, transducir transitoriamente o permanentemente una célula. Se reconocerá que un vector puede ser un ácido nucleico desnudo o un ácido nucleico unido a una proteína o lípido. El vector comprende opcionalmente ácidos nucleicos y/o proteínas y/o membranas virales o bacterianas (tales como una membrana celular, una envoltura lipídica viral, etc.). Los vectores incluyen, a modo no taxativo, replicones (tales como replicones de ARN, bacteriófagos) a los cuales pueden unirse fragmentos de ADN y replicarse. Por consiguiente, los vectores incluyen, a modo no taxativo, ARN, ADN o ARN autónomo autorreplicante circular o lineal (tal como plásmidos, virus y similares, ver Patente de los Estados Unidos No. 5.217.879) e incluyen plásmidos de expresión y no expresión. Cuando un microorganismo o cultivo celular recombinante se describe como huésped de un "vector de expresión", esto incluye ADN extracromosómico circular y lineal y ADN que se ha incorporado al cromosoma(s) huésped. Cuando un vector se mantiene en una célula huésped, el vector puede replicarse establemente por medio de las células durante la mitosis como una estructura autónoma o se incorpora dentro del genoma huésped.

Tal como se usa en la presente, el término "recombinante" abarca ácidos nucleicos adyacentes a un ácido nucleico "de cadena principal" a la cual no son adyacentes en su ambiente natural. En algunas realizaciones, para "enriquecerse" los ácidos nucleicos representarán aproximadamente 5% o más de los insertos de ácido nucleico en una población de moléculas de cadena principal de ácido nucleico. Las moléculas de cadena principal de acuerdo con la invención incluyen ácidos nucleicos tales como vectores de expresión, ácidos nucleicos autorreplicantes, virus, ácidos nucleicos integrantes y otros vectores o ácidos nucleicos usados para mantener o manejar un inserto de ácido nucleico de interés. En algunas realizaciones, los ácidos nucleicos enriquecidos representan aproximadamente 15% o más de los insertos de ácido nucleico en la población de moléculas de cadena principal recombinantes. En algunas realizaciones, los ácidos nucleicos enriquecidos representan aproximadamente 50% o más de los insertos de ácido nucleico en la población de moléculas de cadena principal recombinantes. En algunas realizaciones, los ácidos nucleicos enriquecidos representan aproximadamente 90% o más de los insertos de ácido nucleico en la población de moléculas de cadena principal recombinantes.

60 Una realización de la invención es un ácido nucleico sintético o recombinante aislado que comprende una de las secuencias de acuerdo con la invención, o un fragmento que comprende al menos 10, 15, 20, 25, 30, 35, 40, 50, 75,

100, 150, 200, 300, 400 o 500 o más bases consecutivas de un ácido nucleico de acuerdo con la invención. Los ácidos nucleicos sintéticos o recombinantes aislados pueden comprender ADN, incluido ADNc, ADN genómico y ADN sintético. El ADN puede ser de doble cadena o de una sola cadena y si es de una sola cadena puede ser la cadena codificadora o la cadena no codificadora (antisentido). Alternativamente, los ácidos nucleicos sintéticos o recombinantes aislados comprenden ARN.

Los ácidos nucleicos sintéticos o recombinantes aislados de acuerdo con la invención pueden usarse para preparar uno de los polipéptidos de acuerdo con la invención o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 o 150 o más aminoácidos consecutivos de uno de los polipéptidos de acuerdo con la invención. Por consiguiente, otra realización de la invención es un ácido nucleico sintético o recombinante aislado que codifica uno de los polipéptidos de acuerdo con la invención o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 o 150 o más aminoácidos consecutivos de uno de los polipéptidos de acuerdo con la invención. Las secuencias codificadoras de estos ácidos nucleicos pueden ser idénticas a una de las secuencias codificadoras de uno de los ácidos nucleicos de acuerdo con la invención o pueden ser secuencias codificadoras diferentes que codifican uno de los de acuerdo con la invención que tienen al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 o 150 o más aminoácidos consecutivos de uno de los polipéptidos de acuerdo con la invención, como resultado de la redundancia o degeneración del código genético. Los expertos en la técnica conocen el código genético y el mismo puede obtenerse por ejemplo en la página 214 de B. Lewin, Genes VI, Oxford University Press, 1997.

El ácido nucleico aislado que codifica uno de los polipéptidos de la invención y secuencias sustancialmente idénticas al mismo, puede incluir, a modo no taxativo: la secuencia codificadora de un ácido nucleico de acuerdo con la invención y secuencias codificadoras adicionales, tales como secuencias líderes o secuencias proproteína y secuencias no codificadoras, tales como intrones o secuencias no codificadoras en 5' y/o 3' de la secuencia codificadora. Por lo tanto, tal como se usa en la presente, la expresión "polinucleótido que codifica un polipéptido" abarca un polinucleótido que incluye la secuencia codificadora para el polipéptido, así como también un polinucleótido que incluye secuencia codificadora y/o no codificadora adicional.

Alternativamente, las secuencias de ácido nucleico de la invención y secuencias sustancialmente idénticas a las mismas, pueden mutageneizarse utilizando técnicas convencionales, tales como mutagénesis sitio dirigida u otras técnicas conocidas para los expertos en la técnica, para introducir cambios silenciosos en los polinucleótidos de acuerdo con la invención. Tal como se usa en la presente, "cambios silenciosos" incluye, por ejemplo, cambios que no alteran la secuencia de aminoácidos codificada por el polinucleótido. Dichos cambios pueden ser deseables para aumentar el nivel del polipéptido producido por células huéspedes que contienen un vector que codifica el polipéptido mediante la introducción de codones o pares de codones que ocurren frecuentemente en el organismo huésped.

La invención también se refiere a polinucleótidos que tienen cambios nucleotídicos que resultan en sustituciones, adiciones, eliminaciones, fusiones y truncamientos de aminoácidos en los polipéptidos de acuerdo con la invención. Dichos cambios nucleotídicos pueden introducirse utilizando técnicas tales como mutagénesis sitio dirigida, mutagénesis química aleatoria, eliminación de exonucleasa III y otras técnicas de ADN recombinante. De forma alternativa, dichos cambios nucleotídicos pueden ser variantes alélicas de origen natural que se aíslan mediante la identificación de ácidos nucleicos que se hibridan específicamente con 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 acuerdo con la invención (o las secuencias complementarias a las mismas) en condiciones de rigurosidad alta, moderada o baja tal como se proporciona en la presente.

#### Técnicas generales

Los ácidos nucleicos usados para poner en práctica esta invención, ya sean ARN, ARNsi, miARN, ácido nucleico antisentido, ADNc, ADN genómico, vectores, virus o híbridos de los mismos, pueden aislarse a partir de una variedad de fuentes, genéticamente manipuladas, amplificadas y/o expresadas/generadas recombinantemente. Los polipéptidos recombinantes (tales como enzimas aldolasas, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa) generados a partir de estos ácidos nucleicos pueden aislarse o clonarse y evaluarse individualmente para determinar si tienen una actividad deseada. Se puede usar cualquier sistema de expresión recombinante, incluidos sistemas de expresión celular bacteriana, de mamíferos, levadura, insectos o plantas.

Alternativamente, estos ácidos nucleicos pueden sintetizarse in vitro mediante técnicas de síntesis química conocidas, tales como las descritas, por ejemplo, en 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 33:7886-7896; 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.

Las técnicas para el manejo de ácidos nucleicos, tales como subclonación, sondas de etiquetado (tales como etiquetado con cebador aleatorio utilizando polimerasa Klenow, traducción de muesca, amplificación), secuenciación, hibridación y similares se describen en la literatura científica y de patentes, ver Sambrook, ed., MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Tomos 1-3, Cold Spring Harbor Laboratory, (1989);

CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. John Wiley & Sons, Inc., Nueva York (1997); LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY: HYBRIDIZATION WITH NUCLEIC ACID PROBES, Parte I. Theory and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y. (1993).

5 Otro medio útil para obtener y manejar ácidos nucleicos utilizados para poner en práctica los métodos de acuerdo con la invención es clonar a partir de muestras genómicas y, si se desea, someter a detección y volver a clonar insertos aislados o amplificados a partir de, por ejemplos, clones genómicos o clones de ADNc. Las fuentes de ácidos nucleicos utilizados en los métodos de acuerdo con la invención incluyen las contenidas en bibliotecas genómicas o de ADNc, tales como cromosomas artificiales de mamíferos (MACs), ver las Patentes de los Estados Unidos Nos. 5.721.118; 6.025.155; cromosomas artificiales humanos, ver Rosenfeld (1997) Nat. Genet. 15:333-335; 10 cromosomas artificiales de levadura (YAC); cromosomas artificiales bacterianos (BAC); cromosomas artificiales P1, ver Woon (1998) Genomics 50:306-316; vectores derivados de P1 (PACs), ver Kern (1997) Biotechniques 23:120-124; cósmidos, virus recombinantes, fagos o plásmidos.

15 En algunas realizaciones, un ácido nucleico que codifica un polipéptido de acuerdo con la invención se ensambla en la fase apropiada con una secuencia líder capaz de dirigir la secreción del polipéptido traducido o de un fragmento del mismo.

La invención proporciona proteínas de fusión y ácidos nucleicos que las codifican. Un polipéptido de acuerdo con la invención puede fusionarse con un péptido o polipéptido heterólogo, tales como péptidos de identificación del extremo N-terminal que imparten características deseadas, tales como estabilidad aumentada o purificación simplificada. Los péptidos y polipéptidos de acuerdo con la invención también pueden sintetizarse y expresarse 20 como proteínas de fusión con uno o más dominios adicionales unidos a las mismas, por ejemplo, para producir un péptido más inmunogénico, para aislar más fácilmente un péptido sintetizado recombinantemente, para identificar y aislar anticuerpos y células B que expresan anticuerpos y similares. Los dominios que facilitan la detección y purificación incluyen, por ejemplo péptidos de quelación 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 25 purificación sobre inmunoglobulina inmovilizada y el dominio utilizado en el sistema de purificación mediante extensión/afinidad FLAGS (Immunex Corp, Seattle Wash.). La inclusión de secuencias escindibles por un enlazante tal como Factor Xa o enterocinasa (Invitrogen, Carlsbad, 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 ácidos nucleicos que codifica un epítipo unida a seis residuos de histidina seguidos por una 30 tiorredoxina y un sitio de escisión de enterocinasa (ver por ejemplo Williams (1995) Biochemistry 34:1787-1797; Dobeli (1998) Protein Expr. Purif. 12:404-414). Los residuos de histidina facilitan la detección y purificación mientras el sitio de escisión de enterocinasa proporciona un medio para purificar el epítipo a partir del resto de la proteína de fusión. La tecnología relacionada con vectores que codifican proteínas de fusión y la aplicación de proteínas de fusión se describe en la literatura científica y de patentes, ver por ejemplo Kroll (1993) DNA Cell. Biol., 12:441-53.

### 35 Secuencias de control de transcripción y traducción

La invención proporciona secuencias de ácido nucleico (tales como ADN) de acuerdo con la invención unidas operativamente a secuencia(s) de control de expresión (tales como de transcripción o traducción), tales como 40 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. Los promotores bacterianos ejemplares incluyen *lacI*, *lacZ*, T3, T7, *gpt*, *lambda PR*, *PL* y *trp*. Los promotores eucarióticos ejemplares incluyen el inmediato-temprano de CMV, el de timidina cinasa de HSV, el temprano y tardío de SV40, LTR de retrovirus y el de metalotioneína I de ratón.

Tal como se usa en la presente, el término "promotor" incluye todas las secuencias capaces de activar la transcripción de una secuencia codificadora en una célula, tal como una célula vegetal o animal. Por consiguiente, 45 los promotores usados en los constructos de acuerdo con la invención incluyen elementos de control de transcripción que actúan en *cis* y secuencias reguladoras que participan en la regulación o modulación del tiempo y/o velocidad de transcripción de un gen. Por ejemplo, un promotor puede ser un elemento de control de transcripción que actúa en *cis*, incluido un potenciador, un promotor, un terminador de transcripción, un origen de replicación, una secuencia de integración cromosómica, regiones sin traducir en 5' y 3' o una secuencia intrónica, que participan en la regulación de la transcripción. Estas secuencias que actúan en *cis* pueden interactuar con 50 proteínas u otras biomoléculas para llevar a cabo (activar/desactivar, regular, modular, etc.) la transcripción. Los promotores "constitutivos" son aquellos que activan la expresión continuamente en la mayoría de las condiciones ambientales y estados de desarrollo o diferenciación celular. Los promotores "inducibles" o "regulables" dirigen la expresión del ácido nucleico de acuerdo con la invención bajo la influencia de condiciones ambientales o condiciones de desarrollo. Los ejemplos de condiciones ambientales que pueden afectar la transcripción mediante 55 promotores inducibles incluyen condiciones anaeróbicas, temperatura elevada, sequía o la presencia de luz.

Los promotores "específicos de tejido" son elementos de control de transcripción que son activos solamente en células o tejidos u órganos específicos, tales como en plantas o animales. La regulación específica de tejido puede lograrse mediante ciertos factores intrínsecos que aseguran la expresión de genes que codifican proteínas

específicas para un tejido dado. Se sabe que dichos factores existen en mamíferos y plantas para permitir el desarrollo de tejidos específicos.

5 Los promotores adecuados para expresar un polipéptido en bacterias incluyen los promotores lac o trp de *E. coli*, el promotor *lacI*, el promotor *lacZ*, el promotor T3, el promotor T7, el promotor *gpt*, el promotor *lambda P<sub>R</sub>*, el promotor *lambda P<sub>L</sub>*, promotores a partir de operones que codifican enzimas glucolíticas tales como 3-fosfoglicerato cinasa (PGK) y el promotor de ácido fosfatasa. Los promotores eucarióticos incluyen el promotor inmediato-temprano de CMV, el promotor de la timidina cinasa de HSV, promotores de choque térmico, el promotor temprano y tardío de SV40, LTR de retrovirus y el promotor de la metalotioneína-I de ratón. También se pueden utilizar otros promotores conocidos por controlar la expresión de genes en células procarióticas o eucarióticas o sus virus. Los promotores adecuados para expresar el polipéptido o fragmento del mismo en bacterias incluyen los promotores lac o trp de *E. coli*, el promotor *lacI*, el promotor *lacZ*, el promotor T3, el promotor T7, el promotor *gpt*, el promotor *lambda P<sub>R</sub>*, el promotor *lambda P<sub>L</sub>*, promotores a partir de operones que codifican enzimas glucolíticas tales como 3-fosfoglicerato cinasa (PGK) y el promotor de ácido fosfatasa. Los promotores fúngicos incluyen el promotor de factor  $\alpha$ . Los promotores eucarióticos incluyen el promotor inmediato-temprano de CMV, el promotor de timidina cinasa de HSV, promotores de choque térmico, el promotor temprano y tardío de SV40, LTR de retrovirus y el promotor de metalotioneína-I de ratón. También se pueden utilizar otros promotores conocidos por controlar la expresión de genes en células procarióticas o eucarióticas o sus virus.

Promotores vegetales específicos de tejido

20 La invención proporciona cassettes de expresión que pueden expresarse de forma específica de tejido, tales como estos pueden expresar una aldolasa, tal como piruvato aldolasa, enzima HMG y/o KHG aldolasa de acuerdo con la invención de forma específica de tejido. En algunas realizaciones, la invención también proporciona plantas o semillas que expresan una aldolasa, tal como piruvato aldolasa, enzima HMG y/o KHG aldolasa de acuerdo con la invención de forma específica de tejido. La especificidad de tejido puede ser específico de semilla, específico de tallo, específico de hoja, específico de raíz, específico de fruto y similares.

25 El término "planta" incluye plantas enteras, partes de plantas (tales como hojas, tallos, flores, raíces, etc.), protoplastos vegetales, células de semillas y plantas y progenie de las mismas. La clase de planta que puede utilizarse en el método de acuerdo con la invención es en general tan amplia como la clase de plantas superiores susceptibles a técnicas de transformación, incluidos angiospermas (plantas monocotiledóneas y dicotiledóneas), así como también gimnospermas. Incluye plantas de una variedad de niveles de ploidía, incluidos los estados poliploide, diploide, haploide y hemicigótico. Tal como se usa en la presente, la expresión "planta transgénica" incluye plantas o células vegetales en las que se ha insertado una secuencia de ácidos nucleicos heteróloga, tales como los ácidos nucleicos y varios constructos recombinantes (tales como cassettes de expresión) de acuerdo con la invención.

35 En algunas realizaciones, un promotor constitutivo tal como el promotor CaMV 35S se puede utilizar para expresión en partes específicas de la planta o semilla o en toda la planta. Por ejemplo, para la sobreexpresión, se puede emplear un fragmento de promotor vegetal que dirigirá la expresión de un ácido nucleico en algunos o todos los tejidos de una planta, tal como una planta regenerada. Dichos promotores se denominan en la presente 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 transcripción del virus del mosaico de la coliflor (CaMV) 35S, el promotor 1' o 2' derivado de ADN-T de *Agrobacterium tumefaciens* y otras regiones de iniciación de transcripción de varios genes vegetales conocidos para los expertos. Dichos genes incluyen, por ejemplo, ACT11 de *Arabidopsis* (Huang (1996) *Plant Mol. Biol.* 33:125-139); Cat3 de *Arabidopsis* (GenBank No. U43147, Zhong (1996) *Mol. Gen. Genet.* 251:196-203); el gen que codifica la proteína portadora de estearoil-acilo desaturasa de *Brassica napus* (Genbank No. X74782, Solocombe (1994) *Plant Physiol.* 104:1167-1176); GPC1 de maíz (GenBank No. X15596; Martinez (1989) *J. Mol. Biol.* 208:551-565); el Gpc2 de maíz (GenBank No. 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.

50 La invención utiliza promotores específicos de tejido o constitutivos derivados de virus que pueden incluir, por ejemplo, el promotor subgenómico del tobamovirus (Kumagai (1995) *Proc. Natl. Acad. Sci. Estados Unidos* 92:1679-1683); el virus baciliforme tungro del arroz (RTBV), el cual se replica solo en células del floema de plantas de arroz infectadas, con su promotor que activa la expresión intensa del gen reportero específico del floema; el promotor del virus del mosaico de la nervadura de la yuca (CVMV), con actividad más alta en elementos vasculares, en células del mesófilo de la hoja y en puntas de raíces (Verdaguer (1996) *Plant Mol. Biol.* 31:1129-1139).

55 En algunas realizaciones, el promotor vegetal dirige la expresión del ácido nucleico que expresa la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa en un tejido, órgano o tipo celular específico (es decir, promotores específicos de tejido) o de otra forma bajo un control ambiental o de desarrollo más preciso o bajo el control de un promotor inducible. Los ejemplos de condiciones ambientales que pueden afectar la transcripción incluyen condiciones anaeróbicas, temperatura elevada, la presencia de luz o la pulverización 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 salinidad alta de la patata (Kirch (1997) *Plant Mol. Biol.* 33:897 909).

En algunas realizaciones, los promotores específicos de tejido promueven la transcripción solamente dentro de cierto marco temporal de la etapa de desarrollo dentro de dicho tejido. Ver Blazquez (1998) *Plant Cell* 10:791-800, con la caracterización del promotor del gen LEAFY de *Arabidopsis*. Ver también Cardon (1997) *Plant J* 12:367-77, que describe el factor de transcripción SPL3, el cual reconoce un motivo secuencial conservado en la región promotora del gen AP1 de identidad del meristemo floral de *A. thaliana*; y Mandel (1995) *Plant Molecular Biology*, Tomo 29, pp 995-1004, que describe el promotor del meristemo eIF4. Se pueden usar promotores específicos de tejido que están activos durante todo el ciclo de vida de un tejido específico. En algunas realizaciones, los ácidos nucleicos de acuerdo con la invención están unidos operativamente a un promotor activo principalmente solo en células de fibra de algodón. En algunas realizaciones, los ácidos nucleicos de acuerdo con la invención están unidos operativamente a un promotor activo principalmente durante las etapas de alargamiento celular de la fibra de algodón, tal como lo describe Rinehart (1996) supra. Los ácidos nucleicos pueden unirse operativamente al promotor del gen Fb12A para expresarse preferiblemente en células de fibra de algodón (Ibid). Ver también, John (1997) *Proc. Natl. Acad. Sci. Estados Unidos* 89:5769-5773; John, et al., las Patentes de los Estados Unidos Nos. 5.608.148 y 5.602.321, que describen promotores específicos de fibra de algodón y métodos para la construcción de plantas de algodón transgénicas. Los promotores específicos de raíz también pueden usarse para expresar los ácidos nucleicos de acuerdo con la invención. Los ejemplos de promotores específicos de raíz incluyen el promotor del gen del alcohol deshidrogenasa (DeLisle (1990) *Int. Rev. Cytol.* 123:39-60). Otros promotores que pueden usarse para expresar los ácidos nucleicos de acuerdo con la invención incluyen, por ejemplo promotores específicos de óvulo, específicos de embrión, específicos de endospermo, específicos de tegumento, específicos de cubierta seminal o alguna combinación de los mismos; un promotor específico de hoja (ver Busk (1997) *Plant J.* 11:1285 1295, que describe un promotor específico de hoja de maíz); el promotor ORF13 de *Agrobacterium rhizogenes* (que exhibe actividad alta en raíces, ver Hansen (1997) supra); un promotor específico de polen de maíz (ver Guerrero (1990) *Mol. Gen. Genet.* 224:161 168); puede usarse un promotor de tomate activo durante la maduración del fruto, la senescencia y abscisión de las hojas y, en menor medida, de las flores (ver Blume (1997) *Plant J.* 12:731 746); un promotor específico de pistilo del gen SK2 de la patata (ver Ficker (1997) *Plant Mol. Biol.* 35:425 431); el gen Blec4 del guisante, que es activo en el tejido epidérmico de ápices caulinares vegetativos y florales de alfalfa transgénica, convirtiéndolo en una herramienta útil para dirigir la expresión de genes externos con respecto a la capa epidérmica de los brotes o fibras en crecimiento activo; el gen BEL1 específico de óvulo (ver 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 niveles altos de transcripción en el tejido meristemático y/o dividir rápidamente las células.

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

Los ácidos nucleicos de acuerdo con la invención también pueden unirse operativamente a promotores vegetales que son inducibles tras la exposición a reactivos químicos que pueden aplicarse a la planta, tales como herbicidas o antibióticos. Por ejemplo, puede usarse el promotor In2-2 del maíz, activado por protectores herbicidas de bencenosulfonamida (De Veylder (1997) *Plant Cell Physiol.* 38:568-577); la aplicación de diferentes protectores herbicidas induce patrones de expresión génica diferenciados, incluida la expresión en las raíces, hidatodos y el meristemo apical del brote. Las secuencias codificadoras pueden estar bajo el control de, por ejemplo, un promotor inducible por tetraciclina, tal como el descrito con 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). Con la utilización de promotores inducidos por agentes químicos (tales como hormonas o pesticidas), es decir, promotores que responden a un agente químico que puede aplicarse a la planta transgénica en el campo, se puede inducir la expresión de un polipéptido de acuerdo con la invención en una etapa específica del desarrollo de la planta. Por consiguiente, la invención también proporciona plantas transgénicas que contienen un gen inducible que codifica polipéptidos de acuerdo con la invención, cuyo rango de huéspedes se limita a las especies vegetales objetivo, tales como maíz, arroz, cebada, soja, tomate, trigo, patata y otros cultivos, inducibles en cualquier etapa de desarrollo del cultivo.

El experto reconocerá que un promotor vegetal específico de tejido puede activar la expresión de secuencias unidas operativamente en tejidos distintos a los tejidos objetivo. Por consiguiente, en algunas realizaciones, un promotor específico de tejido es uno que activa la expresión preferiblemente en el tejido o tipo celular objetivo pero también puede provocar la expresión en otros tejidos.

Los ácidos nucleicos de acuerdo con la invención también pueden unirse operativamente a promotores vegetales que son inducibles tras la exposición a reactivos químicos. Estos reactivos incluyen, por ejemplo, herbicidas, auxinas sintéticas o antibióticos que pueden aplicarse, tal como mediante pulverización, sobre las plantas transgénicas. La

expresión inducible de ácidos nucleicos que producen la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención permitirá que el agricultor seleccione las plantas con la expresión y/o actividad óptima de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. Así puede controlarse el desarrollo de las partes de la planta. De esta forma la invención proporciona el medio para facilitar la cosecha de plantas y partes de planta. Por ejemplo, en varias realizaciones, se usa el promotor In2-2 del maíz, activado por protectores herbicidas de bencenosulfonamida (De Veylder (1997) *Plant Cell Physiol.* 38:568-577); la aplicación de diferentes protectores herbicidas induce patrones de expresión génica diferenciados, incluida la expresión en las raíces, hidatodos y el meristemo apical del brote. Las secuencias codificadoras de acuerdo con la invención están también bajo el control de, por ejemplo, un promotor inducible por tetraciclina, tal como el descrito con 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).

En algunas realizaciones, la expresión apropiada del polipéptido puede requerir una región de poliadenilación en el extremo 3' de la región codificadora. La región de poliadenilación puede derivarse del gen natural, de una variedad de distintos genes vegetales o animales o de otro origen o de genes en el ADN-T Agrobacteriano.

#### Vectores de expresión y vehículos de clonación

La invención proporciona vectores de expresión y vehículos de clonación que comprenden ácidos nucleicos de acuerdo con la invención, tales como secuencias que codifican las enzimas aldolasas, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención. Los vectores de expresión y vehículos de clonación de acuerdo con la invención pueden comprender partículas virales, baculovirus, fago, plásmidos, fagómidos, cósmidos, fósidos, cromosomas bacterianos artificiales, ADN viral (tal como variolovacuna, adenovirus, virus de la viruela aviar, seudorrabia y derivados de SV40), cromosomas artificiales en base a P1, plásmidos de levadura y cualquier otro vector específico para huéspedes específicos de interés (tales como *bacillus*, *Aspergillus* y levadura). Los vectores de acuerdo con la invención pueden incluir secuencias cromosómicas, no cromosómicas y de ADN sintético. Los expertos en la técnica conocen una gran cantidad de vectores adecuados y los mismos están disponibles en el mercado. Los vectores ejemplares incluyen: bacterianos: vectores pQE™ (Qiagen, Valencia, CA), plásmidos pBLUESCRIPT™, vectores pNH, vectores lambda-ZAP (Stratagene, La Jolla, CA); ptrc99a, pKK223-3, pDR540, pRIT2T (GE Healthcare, Piscataway, N.J.), vectores pET (Novagen, Madison, Wis.); Eucarióticos: pXT1, pSG5 (Stratagene, La Jolla, CA), pSVK3, pBPV, pMSG, pSVLSV40 (Farmacia). Sin embargo, cualquier otro plásmido u otro vector puede usarse siempre que sea replicable y viable en el huésped. Se pueden emplear vectores con baja cantidad de copias o alta cantidad de copias en la presente invención. Los "plásmidos" pueden estar disponibles en el mercado, disponibles públicamente sin restricciones o pueden construirse a partir de plásmidos disponibles de acuerdo con procedimientos publicados. En la técnica se conocen plásmidos equivalentes a los descritos en la presente y estos serán evidentes para los expertos en la técnica.

El vector de expresión puede comprender un promotor, un sitio de unión a un ribosoma para iniciar la traducción y un terminador de transcripción. El vector también puede incluir secuencias apropiadas para amplificar la expresión. Los vectores de expresión de mamíferos pueden comprender un origen de replicación, todos los sitios de unión a ribosoma necesarios, un sitio de poliadenilación, sitios donadores y aceptores de empalme, secuencias de terminación transcripcional y secuencias no transcritas flanqueadoras en 5'. En algunas realizaciones, las secuencias de ADN derivadas del empalme SV40 y los sitios de poliadenilación pueden utilizarse para proporcionar los elementos genéticos no transcritos necesarios.

En algunas realizaciones, los vectores de expresión contienen uno o más genes marcadores seleccionables para permitir la selección de células huésped que contienen el vector. Dichos marcadores seleccionables incluyen genes que codifican dihidrofolato reductasa o genes que confieren resistencia a neomicina a cultivos de células eucarióticas, genes que confieren resistencia a tetraciclina o ampicilina en *E. coli* y el gen TRP1 de *S. cerevisiae*. Las regiones promotoras pueden seleccionarse a partir de cualquier gen deseado utilizando vectores de cloranfenicol transferasa (CAT) u otros vectores con marcadores seleccionables.

En algunas realizaciones, los vectores para expresar el polipéptido o fragmento del mismo en células eucarióticas contienen potenciadores para aumentar los niveles de expresión. Los potenciadores son elementos de ADN que actúan en *cis* que pueden tener una longitud de aproximadamente 10 a aproximadamente 300 pb. Los mismos pueden actuar sobre un promotor para aumentar su transcripción. Los potenciadores ejemplares incluyen el potenciador SV40 en el lado tardío del origen de replicación pb 100 a 270, el potenciador del promotor temprano del citomegalovirus, el potenciador del poliovirus en el lado tardío del origen de replicación y los promotores del adenovirus.

Una secuencia de ácidos nucleicos puede insertarse en un vector mediante una variedad de procedimientos. En general, la secuencia se liga a la posición deseada en el vector tras la digestión del inserto y el vector con las endonucleasas de restricción apropiadas. Alternativamente, se pueden ligar extremos romos al inserto y al vector. En la técnica se conoce una variedad de técnicas de clonación, tales como las descritas en Ausubel et al. *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. 1997 y Sambrook et al., *Molecular Cloning: A Laboratory*

Manual 2a Ed., Cold Spring Harbor Laboratory Press (1989). Se considera que dichos procedimientos y otros están dentro del alcance de los expertos en la técnica.

El vector puede estar en forma de un plásmido, una partícula viral o un fago. Otros vectores incluyen secuencias cromosómicas, no cromosómicas y de ADN sintético, derivados de SV40; plásmidos bacterianos, ADN fágico, baculovirus, plásmidos de levadura, vectores derivados de combinaciones de plásmidos y ADN fágico, ADN viral tal como variolovacuna, adenovirus, virus de la viruela aviar yseudorrabia. Una variedad de vectores de clonación y expresión para uso con huéspedes procarióticos y eucarióticos se describe, por ejemplo, en Sambrook.

Los vectores bacterianos específicos que pueden utilizarse incluyen los plásmidos disponibles en el mercado que comprenden elementos genéticos del vector de clonación conocido pBR322 (ATCC 37017), pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Suecia), GEM1 (Promega Biotec, Madison, Wis., EE. UU.) pQE70, pQE60, pQE-9 (Qiagen, Valencia, CA), pD10, psiX174 pBLUESCRIPT II KS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene, La Jolla, CA), ptrc99a, pKK223-3, pKK233-3, DR540, pRIT5 (Pharmacia), pKK232-8, pET (Novagen, Madison, Wis.) y pCM7. Los vectores procarióticos específicos incluyen pSV2CAT, pOG44, pXT1, pSG (Stratagene, La Jolla, CA) pSVK3, pBPV, pMSG y pSVL (Pharmacia). Sin embargo, cualquier otro vector puede utilizarse siempre que sea replicable y viable en la célula huésped.

Los ácidos nucleicos de acuerdo con la invención pueden expresarse en cassettes de expresión, vectores o virus y pueden expresarse transitoriamente o establemente en células vegetales y semillas. Un sistema de expresión transitorio ejemplar utiliza sistemas de expresión episómicos, tales como el ARN viral del virus del mosaico de la coliflor (CaMV) generado en el núcleo mediante la transcripción de un minicromosoma episómico que contiene ADN superenrollado, ver Covey (1990) Proc. Natl. Acad. Sci. Estados Unidos 87:1633-1637. Alternativamente, las secuencias codificadoras, es decir, la totalidad o subfragmentos de secuencias de acuerdo con la invención, pueden insertarse en el genoma de una célula huésped vegetal convirtiéndose en parte integral del ADN cromosómico huésped. Los transcritos sentido o antisentido pueden expresarse de esta forma. Un vector que comprende las secuencias (tales como promotores o regiones codificadoras) de ácidos nucleicos de acuerdo con la invención puede comprender un gen marcador que confiere un fenotipo seleccionable en una célula vegetal o una semilla. Por ejemplo, el marcador puede codificar resistencia biocida, tal como resistencia antibiótica, tal como resistencia a kanamicina, G418, bleomicina, higromicina o resistencia herbicida, tal como resistencia a clorosulfurón o Basta.

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

En algunas realizaciones, 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 insectos para expresión y en un huésped procariótico para clonación y amplificación. Adicionalmente, para integrar vectores de expresión, el vector de expresión puede contener al menos una secuencia homóloga al genoma de la célula huésped. Puede contener dos secuencias homólogas que flanquean el constructo de expresión. El vector de integración puede dirigirse a un locus específico en la célula huésped seleccionando la secuencia homóloga apropiada para su inclusión en el vector. Los constructos para integrar vectores se conocen en la técnica.

Los vectores de expresión de acuerdo con la invención también pueden incluir un gen marcador seleccionable para permitir la selección de cepas bacterianas que han sido transformadas, tales como genes que proporcionan las bacterias resistentes a fármacos tales como 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 unida operativamente a una(s) secuencia(s) de control de expresión apropiada(s) (promotor) para dirigir la síntesis de ARN. Los promotores denominados bacterianos específicos incluyen *lacI*, *lacZ*, *T3*, *T7*, *gpt*, *lambda P<sub>R</sub>*, *P<sub>L</sub>* y *trp*. Los promotores eucarióticos incluyen el inmediato-temprano de CMV, el de timidina cinasa de HSV, el temprano y tardío de SV40, el LTR de retrovirus y el de metalotioneína-I de ratón. La selección del vector y promotor apropiados está dentro del nivel de experiencia en la técnica. El vector de expresión también contiene un sitio de unión a un ribosoma para iniciar la traducción y un terminador de transcripción. El vector también puede incluir secuencias apropiadas para amplificar la expresión. Las regiones promotoras pueden seleccionarse a partir de cualquier gen deseado utilizando vectores de cloranfenicol transferasa (CAT) u otros vectores con marcadores seleccionables. Adicionalmente, los vectores de expresión en algunas realizaciones contienen uno o más genes marcadores seleccionables para proporcionar un rasgo fenotípico



para selección de células huésped transformadas tal como dihidrofolato reductasa o resistencia a neomicina para cultivo de células eucarióticas o tal como resistencia a tetraciclina o ampicilina en *E. coli*.

5 Los vectores de expresión de mamíferos también pueden comprender un origen de replicación, todos los sitios de unión a ribosoma necesarios, un sitio de poliadenilación, sitios donadores y aceptores de empalme, secuencias de terminación transcripcional y secuencias no transcritas flanqueadoras en 5'. En algunas realizaciones, las secuencias de ADN derivadas del empalme SV40 y los sitios de poliadenilación pueden utilizarse para proporcionar los elementos genéticos no transcritos necesarios.

10 Los vectores para expresar el polipéptido o fragmento del mismo en células eucarióticas también pueden contener potenciadores para aumentar los niveles de expresión. Los potenciadores son elementos de ADN que actúan en *cis*, generalmente con una longitud de aproximadamente 10 a aproximadamente 300 pb que actúan sobre un promotor para aumentar su transcripción. Los ejemplos incluyen el potenciador SV40 en el lado tardío del origen de replicación pb 100 a 270, el potenciador del promotor temprano del citomegalovirus, el potenciador del poliovirus en el lado tardío del origen de replicación y los promotores del adenovirus.

15 Además, los vectores de expresión pueden contener uno o más genes marcadores seleccionables para permitir la selección de células huésped que contienen el vector. Dichos marcadores seleccionables incluyen genes que codifican dihidrofolato reductasa o genes que confieren resistencia a neomicina a cultivos de células eucarióticas, genes que confieren resistencia a tetraciclina o ampicilina en *E. coli* y el gen TRP1 de *S. cerevisiae*.

20 En algunas realizaciones, el ácido nucleico que codifica uno de los polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas al mismo o fragmentos que comprenden al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 o 150 o más aminoácidos consecutivos del mismo se ensambla en la fase apropiada con una secuencia líder capaz de dirigir la secreción del polipéptido traducido o fragmento del mismo. En algunas realizaciones, el ácido nucleico puede codificar un polipéptido de fusión en el cual uno de los polipéptidos de acuerdo con la invención o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 o 150 o más aminoácidos consecutivos del mismo se fusiona a péptidos o polipéptidos heterólogos, tales como péptidos de identificación de extremo N-terminal que imparten características deseadas, tales como estabilidad aumentada o purificación simplificada.

25 La secuencia de ADN apropiada puede insertarse en el vector mediante una variedad de procedimientos. En general, la secuencia de ADN se liga a la posición deseada en el vector tras la digestión del inserto y el vector con las endonucleasas de restricción apropiadas. Alternativamente, se pueden ligar extremos romos al inserto y al vector. Se divulga una variedad de técnicas de clonación en Ausubel et al. *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. 1997 y Sambrook et al., *Molecular Cloning: A Laboratory Manual 2a Ed.*, Cold Spring Harbor Laboratory Press (1989). Se considera que dichos procedimientos y otros están dentro del alcance de los expertos en la técnica.

30 El vector puede, por ejemplo, estar en forma de un plásmido, una partícula viral o un fago. Otros vectores incluyen secuencias cromosómicas, no cromosómicas y de ADN sintético, derivados de SV40; plásmidos bacterianos, ADN fágico, baculovirus, plásmidos de levadura, vectores derivados de combinaciones de plásmidos y ADN fágico, ADN viral tal como variolovacuna, adenovirus, virus de la viruela aviar yseudorrabia. Una variedad de vectores de clonación y expresión para uso con huéspedes procarióticos y eucarióticos se describe, por ejemplo, en Sambrook, et al., *Molecular Cloning: A Laboratory Manual, 2a Ed.*, Cold Spring Harbor, N.Y., (1989).

40 Células huésped y células transformadas

La invención también proporciona células transformadas que comprenden secuencias de ácido nucleico de acuerdo con la invención, tales como secuencias que codifican enzimas aldolasas, tales como piruvato aldolasas, HMG y/o KHG aldolasa de acuerdo con la invención o vectores de acuerdo con la invención. La célula huésped puede ser cualquiera de las células huéspedes conocidas para los expertos en la técnica, incluidas células procarióticas, células eucarióticas, tales como células bacterianas, células fúngicas, células de levadura, células de mamíferos, células de insectos o células vegetales. Las células bacterianas ejemplares incluyen cualquier especie de *Streptomyces*, *Staphylococcus*, *Pseudomonas* o *Bacillus*, incluidas *E. coli*, *Bacillus subtilis*, *Pseudomonas fluorescens*, *Bacillus cereus* o *Salmonella typhimurium*. Las células fúngicas ejemplares incluyen cualquier especie de *Aspergillus*. Las células de levadura ejemplares incluyen cualquier especie de *Pichia*, *Saccharomyces*, *Schizosaccharomyces* o *Schwanniomyces*, incluidas *Pichia pastoris*, *Saccharomyces cerevisiae* o *Schizosaccharomyces pombe*. Las células de insectos ejemplares incluyen cualquier especie de *Spodoptera* o *Drosophila*, incluidas *Drosophila*S2 y *Spodoptera* Sf9. Las células de animales ejemplares incluyen las líneas celulares CHO, COS o de melanoma de Bowes o cualquier línea celular murina o humana. La selección de un huésped apropiado está dentro de las capacidades de los expertos en la técnica. Las técnicas para transformar una amplia variedad de especies de plantas superiores se conocen y describen en la literatura técnica y científica. Ver Weising (1988) *Ann. Rev. Genet.* 22:421-477; Patente de los Estados Unidos No. 5.750.870.

El vector puede introducirse en las células huésped utilizando cualquiera de una variedad de técnicas, incluidas transformación, transfección, transducción, infección viral, pistolas génicas o transferencia génica mediada por Ti. Los métodos específicos incluyen transfección mediante fosfato de calcio, transfección mediada por DEAE-Dextrano, lipofección o electroporación (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

5 En algunas realizaciones, los ácidos nucleicos o vectores de acuerdo con la invención se introducen en las células para someterlas a detección, por lo tanto, los ácidos nucleicos ingresan a las células de forma adecuada para la posterior expresión del ácido nucleico. El método de introducción depende en gran medida del tipo celular objetivo. Los métodos ejemplares incluyen precipitación de CaPO<sub>4</sub>, fusión de liposomas, lipofección (tal como LIPOFECTIN™), electroporación, infección viral, etc. Los ácidos nucleicos candidatos pueden integrar establemente el genoma de la célula huésped (por ejemplo, con introducción retroviral) o pueden existir transitoriamente o establemente en el citoplasma (es decir, a través del uso de plásmidos tradicionales, utilizando secuencias reguladoras estándares, marcadores de selección, etc.). Dado que muchas detecciones farmacéuticamente importantes requieren células objetivo de mamíferos modelo o humanas, se pueden utilizar vectores retrovirales capaces de transfectar dichos objetivos.

10 Cuando sea apropiado, las células huésped manipuladas pueden cultivarse en medios nutritivos convencionales modificados según sea apropiado para activar promotores, seleccionar transformantes o amplificar los genes de acuerdo con la invención. Tras la transformación de una cepa huésped adecuada y el crecimiento de la cepa huésped hasta alcanzar una densidad celular apropiada, el promotor seleccionado puede inducirse mediante medios apropiados (tales como cambio de temperatura o inducción química) y las células pueden cultivarse durante un período adicional para permitir que produzcan el polipéptido o fragmento del mismo deseado.

15 Las células pueden cosecharse mediante centrifugación, homogeneizarse mediante medios físicos o químicos y el extracto bruto resultante se conserva para someterse a purificación adicional. Las células microbianas empleadas para la expresión de las proteínas pueden homogeneizarse mediante cualquier método conveniente, incluidos ciclo de congelación y descongelación, exposición a ultrasonidos, homogeneización mecánica o utilización de agentes de lisis celular. Dichos métodos son conocidos para los expertos en la técnica. El polipéptido o fragmento del mismo expresado puede recogerse y purificarse de los cultivos celulares recombinantes mediante métodos que incluyen precipitación de sulfato de amonio o etanol, extracción ácida, cromatografía de intercambio aniónico o catiónico, cromatografía en fosfocelulosa, cromatografía de interacción hidrófoba, cromatografía de afinidad, cromatografía de hidroxapatita y cromatografía en lectina. Se pueden utilizar etapas de repliegue de proteínas, según sea necesario, para completar la configuración del polipéptido. Si se desea, se puede emplear cromatografía líquida de alto rendimiento (HPLC) para las etapas finales de purificación.

20 Los constructos en células huésped pueden utilizarse de forma convencional para producir el producto génico codificado por la secuencia recombinante. Dependiendo del huésped empleado en un procedimiento de producción recombinante, los polipéptidos producidos mediante células huésped que contienen el vector pueden ser glicosilados o pueden ser no glicosilados. Los polipéptidos de acuerdo con la invención pueden o no incluir también un residuo de aminoácido de metionina inicial.

25 También se pueden emplear sistemas de traducción in vitro para producir un polipéptido de acuerdo con la invención. Los sistemas de traducción in vitro pueden utilizar ARNm transcrito a partir de un constructo de ADN que comprende un promotor unido operativamente a un ácido nucleico que codifica el polipéptido o fragmento del mismo. En algunas realizaciones, el constructo de ADN puede linearizarse antes de llevar a cabo una reacción de transcripción in vitro. A continuación el ARNm transcrito se incuba con un extracto de traducción in vitro apropiado, tal como un extracto de reticulocitos de conejo, para producir el polipéptido o fragmento del mismo deseado.

30 Los vectores de expresión pueden contener uno o más genes marcadores seleccionables para proporcionar un rasgo fenotípico para selección de células huésped transformadas tal como dihidrofolato reductasa o resistencia a neomicina para cultivo de células eucarióticas o tal como resistencia a tetraciclina o ampicilina en *E. coli*.

35 Las células huésped que contienen los polinucleótidos de interés, tales como ácidos nucleicos de acuerdo con la invención, pueden cultivarse en medios nutritivos convenciones modificados según sea apropiado para activar promotores, seleccionar transformantes o amplificar genes. Las condiciones de cultivo, tales como temperatura, pH y similares, son las utilizadas previamente con la célula huésped seleccionada para expresión y serán evidentes para el experto en la técnica. Los clones identificados por tener la actividad enzimática especificada a continuación pueden secuenciarse para identificar la secuencia de polinucleótidos que codifica una enzima que tiene la actividad potenciada.

40 La invención proporciona métodos para sobreexpresar enzimas aldolasas, tales como piruvato aldolasas, tales como HMG y/o KHG aldolasa recombinantes en células que comprenden y expresan un vector que comprende un ácido nucleico de acuerdo con la invención, tal como un ácido nucleico que comprende una secuencia de ácidos nucleicos con al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o mayor identidad secuencial con una secuencia de acuerdo con la invención, en donde las identidades secuenciales se determinan mediante análisis con un algoritmo de comparación secuencial o

mediante inspección visual, o un ácido nucleico que se hibrida en condiciones rigurosas con una secuencia de ácidos nucleicos de acuerdo con la invención o una subsecuencia de la misma. La sobreexpresión puede producirse mediante cualquier medio, tal como la utilización de un promotor de gran actividad, un vector dicistrónico o mediante amplificación génica del vector.

- 5 Los ácidos nucleicos de acuerdo con la invención pueden expresarse o sobreexpresarse en cualquier sistema de expresión in vitro o in vivo. Se puede emplear cualquier sistema de cultivo celular para expresar o sobreexpresar proteínas recombinantes, incluidos cultivos bacterianos, de insectos, de levadura, fúngicos o de mamíferos. La sobreexpresión puede producirse mediante la elección apropiada de promotores, potenciadores, vectores (tal como el uso de vectores de replicón, vectores dicistrónicos (ver Gurtu (1996) Biochem. Biophys. Res. Commun. 229:295-8)), medios, sistemas de cultivo y similares. En algunas realizaciones, se utiliza amplificación génica utilizando marcadores de selección, tales como glutamina sintetasa (ver Sanders (1987) Dev. Biol. Stand. 66:55-63), en sistemas celulares para sobreexpresar los polipéptidos de acuerdo con la invención.

- 15 Los detalles adicionales en relación a este abordaje se encuentran en la literatura pública y/o son conocidos para los expertos. En un ejemplo específico no taxativo, dicha literatura disponible públicamente incluye EP 0659215 (WO 9403612 A1) (Nevalainen et al.); Lapidot, A., Mechaly, A., Shoham, Y., "Overexpression and single-step purification of a thermostable xylanase from *Bacillus stearothermophilus* T-6", J. Biotechnol. Noviembre 51:259-64 (1996); Lüthi, E., Jasmal, N. B., Bergquist, P. L., "Xylanase from the extremely thermophilic bacterium *Caldocellum saccharolyticum*: overexpression of the gene in *Escherichia coli* and characterization of the gene product", Appl. Environ. Microbiol. Setiembre 56:2677-83 (1990); y Sung, W. L., Luk, C. K., Zahab, D. M., Wakarchuk, W., "Overexpression of the *Bacillus subtilis* and *circulans* xylanases in *Escherichia coli*", Protein Expr. Purif. Junio 4:200-6 (1993), aunque estas referencias no describen las enzimas de la invención divulgada en la presente solicitud.

- 25 La célula huésped puede ser cualquiera de las células huéspedes conocidas para los expertos en la técnica, incluidas células procarióticas, células eucarióticas, células de mamíferos, células de insectos o células vegetales. Como ejemplos representativos de huéspedes apropiados, se pueden mencionar: células bacterianas, tales como *E. coli*, *Streptomyces*, *Bacillus subtilis*, *Bacillus cereus*, *Salmonella typhimurium* y varias especies dentro de los géneros *Pseudomonas*, *Streptomyces* y *Staphylococcus*, células fúngicas, tales como *Aspergillus*, levadura tal como cualquier especie de *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, *Schwanniomyces*, incluidas *Pichia pastoris*, *Saccharomyces cerevisiae*, o *Schizosaccharomyces pombe*, células de insectos tales como *Drosophila* S2 y *Spodoptera* Sf9, células de animales tales como de CHO, COS o de melanoma de Bowes y adenovirus. La selección de un huésped apropiado está dentro de las capacidades de los expertos en la técnica.

- 30 El vector puede introducirse en las células huésped utilizando cualquiera de una variedad de técnicas, incluidas transformación, transfección, transducción, infección viral, pistolas génicas o transferencia génica mediada por Ti. Los métodos específicos incluyen transfección mediante fosfato de calcio, transfección mediada por DEAE-Dextrano, lipofección o electroporación (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

- 35 Cuando sea apropiado, las células huésped manipuladas pueden cultivarse en medios nutritivos convencionales modificados según sea apropiado para activar promotores, seleccionar transformantes o amplificar los genes de acuerdo con la invención. Tras la transformación de una cepa huésped adecuada y el crecimiento de la cepa huésped hasta alcanzar una densidad celular apropiada, el promotor seleccionado puede inducirse mediante medios apropiados (tales como cambio de temperatura o inducción química) y las células pueden cultivarse durante un período adicional para permitir que produzcan el polipéptido o fragmento del mismo deseado.

- 40 Las células pueden cosecharse mediante centrifugación, homogeneizarse mediante medios físicos o químicos y el extracto bruto resultante se conserva para someterse a purificación adicional. Las células microbianas empleadas para la expresión de las proteínas pueden homogeneizarse mediante cualquier método conveniente, incluidos ciclo de congelación y descongelación, exposición a ultrasonidos, homogeneización mecánica o utilización de agentes de lisis celular. Dichos métodos son conocidos para los expertos en la técnica. El polipéptido o fragmento del mismo expresado puede recogerse y purificarse a partir de los cultivos celulares recombinantes mediante métodos que incluyen precipitación de sulfato de amonio o etanol, extracción ácida, cromatografía de intercambio aniónico o catiónico, cromatografía en fosfocelulosa, cromatografía de interacción hidrófoba, cromatografía de afinidad, cromatografía de hidroxipatita y cromatografía en lectina. Se pueden utilizar etapas de repliegue de proteínas, según sea necesario, para completar la configuración del polipéptido. Si se desea, se puede emplear cromatografía líquida de alto rendimiento (HPLC) para las etapas finales de purificación.

- 45 También se pueden emplear varios sistemas de cultivos celulares de mamíferos para expresar la proteína recombinante. Ejemplos de sistemas de expresión de mamíferos incluyen las líneas COS-7 de fibroblastos de riñón de mono (descritas 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.

- 50 Los constructos en células huésped pueden utilizarse de forma convencional para producir el producto génico codificado por la secuencia recombinante. Dependiendo del huésped empleado en un procedimiento de producción recombinante, los polipéptidos producidos mediante células huésped que contienen el vector pueden ser glicosilados

o pueden ser no glicosilados. Los polipéptidos de acuerdo con la invención pueden o no incluir también un residuo de aminoácido de metionina inicial.

Alternativamente, los polipéptidos de acuerdo con la invención o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 o 150 o más aminoácidos consecutivos de los mismos pueden producirse sintéticamente mediante sintetizadores peptídicos convencionales, tales como los descritos más adelante. En otras realizaciones, se pueden emplear fragmentos o porciones de los polipéptidos para producir los polipéptidos de longitud completa correspondientes mediante síntesis de péptidos; por consiguiente, los fragmentos pueden emplearse como intermediarios para producir los polipéptidos de longitud completa.

Los sistemas de traducción in vitro también pueden emplearse para producir uno de los polipéptidos de acuerdo con la invención o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 o 150 o más aminoácidos consecutivos de los mismos utilizando ARNm transcritos a partir de un constructo de ADN que comprende un promotor unido operativamente a un ácido nucleico que codifica el polipéptido o fragmento del mismo. En algunas realizaciones, el constructo de ADN puede linearizarse antes de llevar a cabo una reacción de transcripción in vitro. A continuación el ARNm transcrito se incuba con un extracto de traducción in vitro apropiado, tal como un extracto de reticulocitos de conejo, para producir el polipéptido o fragmento del mismo deseado.

#### Amplificación de ácidos nucleicos

Al poner en práctica la invención, se pueden producir ácidos nucleicos de acuerdo con la invención y ácidos nucleicos que codifican enzimas aldolasas, tales como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención o ácidos nucleicos modificados de acuerdo con la invención, mediante amplificación, tal como PCR. La amplificación además puede utilizarse para clonar o modificar los ácidos nucleicos de acuerdo con la invención. Por consiguiente, la invención proporciona pares de secuencias cebadoras de amplificación para amplificar ácidos nucleicos de acuerdo con la invención. Un experto en la técnica puede diseñar pares de secuencias cebadoras de amplificación para cualquier parte o para la longitud completa de estas secuencias.

En algunas realizaciones, la invención proporciona ácidos nucleicos amplificados mediante pares de cebadores de amplificación de acuerdo con la invención, tales como pares de cebadores tales como fueron establecidos en aproximadamente los primeros (5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25 o más residuos de ácidos nucleicos de acuerdo con la invención y en aproximadamente los primeros (5') 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25 o más residuos de las cadenas complementarias. En algunas realizaciones, la invención proporciona pares de secuencias cebadoras de amplificación para amplificar un ácido nucleico que codifica un polipéptido que tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, en donde el par de cebadores es capaz de amplificar un ácido nucleico que comprende una secuencia de acuerdo con la invención o fragmentos o subsecuencias de la misma. Uno o cada miembro del par de secuencias cebadoras de amplificación puede comprender un oligonucleótido que comprende al menos aproximadamente 10 a 50 o más bases consecutivas de la secuencia o aproximadamente 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25 o más bases consecutivas de la secuencia. En algunas realizaciones, la invención proporciona pares de cebadores de amplificación, en donde el par de cebadores comprende un primer miembro que tiene una secuencia tal como fue establecida en aproximadamente los primeros (5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25 o más residuos de ácidos nucleicos de acuerdo con la invención y un segundo miembro que tiene una secuencia tal como fue establecida en aproximadamente los primeros (5') 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25 o más residuos de la cadena complementaria del primer miembro.

La invención proporciona enzimas aldolasas, tales como piruvato aldolasa, tal como HMG y/o KHG aldolasa generadas mediante amplificación, tal como reacción en cadena de la polimerasa (PCR), utilizando un par de cebadores de amplificación de acuerdo con la invención. En algunas realizaciones, la invención proporciona métodos para hacer una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa mediante amplificación, tal como PCR, utilizando un par de cebadores de amplificación de acuerdo con la invención. En algunas realizaciones, el par de cebadores de amplificación amplifica un ácido nucleico de una biblioteca, tal como una biblioteca, tal como una biblioteca ambiental.

Las reacciones de amplificación también pueden utilizarse para cuantificar la cantidad de ácido nucleico en una muestra (tal como la cantidad de mensaje en una muestra celular), etiquetar el ácido nucleico (tal como aplicarlo en un arreglo o una membrana de filtro), detectar el ácido nucleico o cuantificar la cantidad de un ácido nucleico específico en una muestra. En algunas realizaciones de la invención, se amplifica el mensaje aislado a partir de una célula o una biblioteca de ADNc.

El experto en la técnica puede seleccionar y diseñar cebadores de amplificación de oligonucleótidos adecuados. Los métodos de amplificación también se conocen en la técnica e incluyen reacción en cadena de la polimerasa, PCR (ver PCR PROTOCOLS, A GUIDE TO METHODS AND APPLICATIONS, ed. Innis, Academic Press, N.Y. (1990) y PCR STRATEGIES (1995), ed. Innis, Academic Press, Inc., N.Y., reacción en cadena de la ligasa (LCR) (ver Wu (1989) Genomics 4:560; Landegren (1988) Science 241:1077; Barringer (1990) Gene 89:117); amplificación de transcripción (ver Kwoh (1989) Proc. Natl. Acad. Sci. Estados Unidos 86:1173); y replicación de secuencia

automantenida (ver Guatelli (1990) Proc. Natl. Acad. Sci. Estados Unidos 87:1874); amplificación de replicasa Q Beta (ver Smith (1997) J. Clin. Microbiol. 35:1477-1491), ensayo de amplificación de replicasa Q Beta automatizado (ver Burg (1996) Mol. Cell. Probes 10:257-271) y otras técnicas mediadas por ARN polimerasa (tales como NASBA, Cangene, Mississauga, Ontario); ver también Berger (1987) Methods Enzymol. 152:307-316; Ausubel et al. Current Protocols in Molecular Biology, John Wiley & Sons, Inc. 1997 y Sambrook et al., Molecular Cloning: A Laboratory Manual 2a Ed., Cold Spring Harbor Laboratory Press (1989). Patentes de los Estados Unidos Nos. 4.683.195 y 4.683.202; Sookninan (1995) Biotechnology 13:563-564.

#### Determinación de identidad secuencial en ácidos nucleicos y polipéptidos

La invención proporciona ácidos nucleicos que comprenden secuencias que tienen al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más o total (100%) identidad secuencial (homología) con un ácido nucleico de acuerdo con la invención (ver Listado de secuencias) en la longitud total de un ADNc, transcrito (ARNm) o gen. En algunas realizaciones, la invención proporciona polipéptidos que comprenden secuencias que tienen al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más o total (100%) identidad secuencial con un polipéptido de la SEQ ID NO: 276. El grado de identidad secuencial (homología) puede determinarse utilizando cualquier programa informático y parámetros asociados, incluidos los descritos en la presente, tales como BLAST 2.2.2. o FASTA versión 3.0t78, con los parámetros predeterminados.

Las secuencias homólogas y fragmentos de secuencias de ácido nucleico de acuerdo con la invención pueden referirse a secuencias que tienen al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más identidad secuencial (homología) con estas secuencias. La homología (identidad secuencial) puede determinarse utilizando cualquiera de los programas informáticos y parámetros descritos en la presente, incluido FASTA versión 3.0t78 con los parámetros predeterminados. Las secuencias homólogas también incluyen secuencias de ARN en las cuales las uridinas sustituyen a las timinas en las secuencias de ácido nucleico de acuerdo con la invención. Las secuencias homólogas pueden obtenerse utilizando cualquiera de los procedimientos descritos en la presente o pueden resultar de la corrección de un error de secuenciación. Sea apreciará que las secuencias de ácido nucleico de acuerdo con la invención pueden representarse con el formato tradicional de un solo carácter (Ver el interior de la contratapa de Stryer, Lubert. Biochemistry, 3a Ed., W. H Freeman & Co., Nueva York.) o en cualquier otro formato que registra la identidad de los nucleótidos en una secuencia.

En algunas realizaciones, los programas de comparación de secuencias identificados en la presente se usan en este aspecto de acuerdo con la invención, es decir, para determinar si una secuencia de ácidos nucleicos o polipéptido está dentro del alcance de acuerdo con la invención. Sin embargo, las identidades secuenciales de proteínas y/o ácidos nucleicos (homologías) pueden evaluarse utilizando cualquier algoritmo o programa de comparación secuencial conocido en la técnica. Dichos algoritmos y programas incluyen, pero de ninguna forma se limitan a, TBLASTN, BLASTP, FASTA, TFASTA y CLUSTALW (ver Pearson and Lipman, Proc. Natl. Acad. Sci. Estados Unidos 85(8):2444-2448, 1988; Altschul et al., J. Mol. Biol. 215(3):403-410, 1990; Thompson 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).

En algunas realizaciones, la homología o identidad se mide utilizando un programa informático de análisis de secuencias (tal como Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705). Dicho programa informático empareja secuencias similares asignando grados de homología a varias eliminaciones, sustituciones y otras modificaciones. En algunas realizaciones, los términos "homología" e "identidad" en el contexto de dos o más secuencias de ácidos nucleico o polipéptidos, se refiere a dos o más secuencias o subsecuencias que son iguales o tienen un porcentaje especificado de residuos de aminoácidos o nucleótidos que son iguales cuando se comparan y alinean para determinar la correspondencia máxima en una ventana de comparación o región designada tal como se mide utilizando cualquier cantidad de algoritmos de comparación secuencial o mediante alineación manual e inspección visual. En algunas realizaciones, para la comparación secuencial, una secuencia actúa como una secuencia de referencia con la cual se comparan las secuencias de prueba. Cuando se utiliza un algoritmo de comparación secuencial, las secuencias de prueba y referencia se ingresan en un ordenador, se designan las coordenadas de subsecuencia, si es necesario y se designan los parámetros del programa de algoritmo de secuencias. Se pueden utilizar los parámetros predeterminados del programa o se pueden designar parámetros alternativos. El algoritmo de comparación secuencial a continuación calcula las identidades secuenciales porcentuales para las secuencias de prueba con respecto a la secuencia de referencia, en base a los parámetros del programa.

Una "ventana de comparación", tal como se usa en la presente, incluye la referencia a un segmento de cualquiera de la cantidad de posiciones contiguas seleccionadas del grupo que consiste en de 20 a 600, generalmente aproximadamente 50 a aproximadamente 200, más generalmente aproximadamente 100 a aproximadamente 150 en el cual una secuencia pueden compararse con la secuencia de referencia con la misma cantidad de posiciones contiguas después de alinear óptimamente las dos secuencias. Los métodos de alineación de secuencias para

- comparación se conocen en la técnica. La alineación óptima 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 alineación de homología de Needleman & Wunsch, J. Mol. Biol. 48:443, 1970, mediante el método de búsqueda de similitud de person & Lipman, Proc. Nat'l. Acad. Sci. Estados Unidos 85:2444, 1988, mediante implementaciones informáticas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.) o mediante alineación manual e inspección visual. Otros algoritmos para determinar la homología o identidad incluyen, por ejemplo, además de un programa BLAST (Herramienta de Búsqueda de Alineación Local Básica en el National Center for Biological Information), ALIGN, AMAS (Análisis de Múltiples Secuencias Alineadas), AMPS (Alineación de Múltiples Secuencias Proteicas), ASSET (Herramienta de Evaluación Estadística de Segmento Alineado), BANDS, BESTSCOR, BIOSCAN (Nodo de Análisis Comparativo de Secuencias Biológicas), BLIMPS (Buscador Mejorado por Bloques), FASTA, Intervals & Points, BMB, CLUSTAL V, CLUSTAL W, CONSENSUS, LCONSENSUS, WCONSENSUS, algoritmo de Smith-Waterman, DARWIN, algoritmo de Las Vegas, FNAT (Herramienta de Alineación de Nucleótidos Forzada), Framealign, Framesearch, DYNAMIC, FILTER, FSAP (Paquete de Análisis Secuencial Fristensky), GAP (Programa de Alineación Global), GENAL, GIBBS, GenQuest, ISSC (Comparación Secuencial Sensible), LALIGN (Alineación Secuencial Local), LCP (Programa de Contenido Local), MACAW (Estación de Trabajo para Construcción y Análisis de Alineaciones Múltiples), MAP (Programa de Alineaciones Múltiples), MBLKP, MBLKN, PIMA (Alineación Multisequencia Inducida por Patrón), SAGA (Alineación Secuencial Mediante Algoritmo Genético) y WHAT-IF. Dichos programas de alineación también pueden utilizarse para someter a detección bases de datos de genoma para identificar secuencias de polinucleótidos que tienen secuencias sustancialmente idénticas. Existen varias bases de datos de genoma disponibles, por ejemplo, una porción sustancial del genoma humano está disponible como parte del Human Genome Sequencing Project (Gibbs, 1995). Al menos veintiún genomas distintos ya han sido secuenciados, incluidos por ejemplo de *M. genitalium* (Fraser et al., *Science* 270:397-403 (1995)), *M. jannaschii* (Bult et al., *Science* 23:1058-73 (1996)), *H. influenzae* (Fleischmann et al., *Science* 269:496-512 (1995)), *E. coli* (Blattner et al., *Science* 277:1453-74 (1997)) y levadura (*S. cerevisiae*) (Mewes et al., *Nature* 387:7-65 (1997)) y *D. melanogaster* (Adams et al., *Science* 287:2185-95 (2000)). También se ha avanzado considerablemente en la secuenciación de genomas de organismos modelo, tales como de ratón, *C. elegans* and *Arabidopsis* sp. Diferentes organizaciones mantienen varias bases de datos que contienen información genómica comentada con algunas informaciones funcionales y se puede acceder a las mismas a través de Internet.
- En algunas realizaciones, se utilizan los algoritmos BLAST y BLAST 2.0, los cuales se describen en Altschul et al., Nuc. Acids Res. 25:3389-3402, 1997 y Altschul et al., J. Mol. Biol. 215:403-410, 1990, respectivamente. El programa informático para llevar a cabo los análisis BLAST está disponible públicamente a través del National Center for Biotechnology Information. Este algoritmo implica primero identificar pares de secuencias con puntuación alta (HSPs) identificando palabras cortas de longitud W en la secuencia de consulta, que se emparejan o alcanzan algún umbral de puntaje T con un valor positivo cuando se alinean con una palabra de la misma longitud en una secuencia de la base de datos. T se refiere al umbral de puntaje de palabra cercana (Altschul et al., supra). Estos aciertos de palabra cercana iniciales actúan como semillas para iniciar búsquedas para encontrar HSPs más largas que los contengan. Los aciertos de palabras se extienden en ambos sentidos a lo largo de cada secuencia hasta el punto que pueda aumentarse el puntaje acumulado de alineación. Los puntajes acumulados se calculan utilizando, para secuencias de nucleótidos, los parámetros M (puntaje premio para un par de residuos que se emparejan; siempre >0). Para secuencias de aminoácidos, se utiliza una matriz de puntaje para calcular el puntaje acumulado. La extensión de los aciertos de palabras en cada sentido puede detenerse cuando: el puntaje acumulado de alineación decae a la cantidad X desde su valor máximo alcanzado; el puntaje acumulado cae hasta cero o por debajo, debido a la acumulación de una o más alineaciones residuales de puntaje negativo; o se alcanza el extremo de cualquiera de las secuencias. Los parámetros W, T y X del algoritmo BLAST determinan la sensibilidad y la velocidad de la alineación. El programa BLASTN (para secuencias de nucleótidos) utiliza como valores predeterminados 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 valores predeterminados una longitud de palabra de 3 y expectativas (E) de 10 y la matriz de puntaje BLOSUM62 (ver Henikoff & Henikoff, Proc. Natl. Acad. Sci. Estados Unidos 89:10915, 1992) alineaciones (B) de 50, expectativa (E) de 10, M=5, N=-4 y una comparación de ambas cadenas.
- El algoritmo BLAST también lleva a cabo un análisis estadístico de la similitud entre dos secuencias (ver Karlin & Altschul, Proc. Natl. Acad. Sci. Estados Unidos 90:5873, 1993). Una medida de similitud que proporciona el algoritmo BLAST es la probabilidad de suma más pequeña (P(N)), la cual proporciona un indicio de la probabilidad de que ocurra un emparejamiento al azar entre dos nucleótidos o aminoácidos. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de suma más pequeña en una comparación del ácido nucleico de prueba con el ácido nucleico de referencia es menor que aproximadamente 0,2, en algunas realizaciones menor que al menos aproximadamente 0,01 y en otras realizaciones menor que aproximadamente 0,001.
- En algunas realizaciones, las homologías secuenciales de proteínas y ácido nucleico se evalúan utilizando la Herramienta de Búsqueda de Alineación Local Básica ("BLAST"). Específicamente, se utilizan cinco programas BLAST específicos para llevar a cabo la siguiente tarea:

(1) BLASTP y BLAST3 comparan la secuencia de aminoácidos de consulta con una base de datos de secuencias de proteínas;

(2) BLASTN compara un secuencias de nucleótidos de consulta con una base de datos de secuencias de nucleótidos;

5 (3) BLASTX compara los productos de traducción conceptual de seis marcos de una secuencia de nucleótidos de consulta (ambas cadenas) con una base de datos de secuencias de proteínas;

(4) TBLASTN compara una secuencia de proteína de consulta con una base de datos de secuencias de nucleótidos traducida en los seis marcos de lectura (ambas cadenas); y

10 (5) TBLASTX compara las traducciones en los seis marcos de una secuencia de nucleótidos de consulta con las traducciones en los seis marcos de una base de datos de secuencias de nucleótidos.

Los programas BLAST identifican secuencias homólogas identificando segmentos similares, que se denominan en la presente "pares de segmento de puntaje alto", entre una secuencia de aminoácidos o ácidos nucleicos de consulta y una secuencia de prueba que, en algunas realizaciones, se obtiene de una base de datos de secuencias de proteínas o ácidos nucleicos. Los pares de segmentos de puntaje alto, en algunas realizaciones, se identifican (es decir, alinean) mediante una matriz de puntaje, de la cual se conocen muchas en la técnica. En algunas realizaciones, la matriz de puntaje utilizada es la matriz BLOSUM62 (Gonnet (1992) Science 256:1443-1445; Henikoff and Henikoff (1993) Proteins 17:49-61). En algunas realizaciones, también se pueden utilizar las matrices PAM o PAM250 (ver Schwartz y Dayhoff, eds., 1978, *Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure*, Washington: National Biomedical Research Foundation). Se puede acceder a los programas BLAST a través de la U.S. National Library of Medicine.

Los parámetros utilizados con los algoritmos mencionados anteriormente pueden adaptarse dependiendo de la longitud de secuencia y el grado de homología estudiados. En algunas realizaciones, los parámetros pueden ser los parámetros predeterminados utilizados por los algoritmos a falta de instrucciones del usuario.

#### Sistemas informáticos y productos de programas informáticos

25 La divulgación describe adicionalmente ordenadores, sistemas informáticos, medios legibles por ordenador, productos de programas informáticos y similares en los cuales están grabadas o almacenadas las secuencias de ácidos nucleicos y polipéptidos de acuerdo con la invención. Adicionalmente, al poner en práctica los métodos, tal como para determinar e identificar identidades secuenciales (para determinar si un ácido nucleico está dentro del alcance de acuerdo con la invención), homologías estructurales, motivos y similares in silico, se puede almacenar, grabar y manejar una secuencia de ácidos nucleicos o polipéptido de acuerdo con la invención en cualquier medio que pueda leerse o al cual se pueda acceder mediante un ordenador.

30 Tal como se usan en la presente, los términos "grabado" y "almacenado" se refieren al proceso de almacenar información en un medio informático. Un experto puede adoptar fácilmente cualquier método conocido para grabar información en un medio legible por ordenador para generar productos que comprenden una o más secuencias de ácido nucleico y/o polipéptido de acuerdo con la invención. Tal como se usan en la presente las expresiones y términos "ordenador", "programa informático" y "procesador" se usan en sus contextos generales más amplios e incorporan todos los dispositivos que se describen en detalle más adelante. Una "secuencia codificadora de" o una "secuencia que codifica" un polipéptido o proteína específico, es una secuencia de ácidos nucleicos que puede transcribirse y traducirse a un polipéptido o proteína cuando se coloca bajo el control de secuencias reguladoras apropiadas.

Los polipéptidos de acuerdo con la invención incluyen secuencias de acuerdo con la invención y secuencias sustancialmente idénticas a las mismas y subsecuencias y fragmentos enzimáticamente activos de cualquiera de las secuencias precedentes. En algunas realizaciones, las secuencias de polipéptidos sustancialmente idénticas u homólogas se refieren a secuencia de polipéptidos que tiene al menos 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más, o total (100%) identidad secuencial (homología) con la secuencia de acuerdo con la invención.

La homología (identidad secuencial) puede determinarse utilizando cualquiera de los programas informáticos y parámetros descritos en la presente. Una secuencia de ácidos nucleicos o polipéptido de acuerdo con la invención puede almacenarse, grabarse y manejarse en cualquier medio que pueda leerse o al cual se pueda acceder mediante un ordenador. Tal como se usan en la presente, los términos "grabado" y "almacenado" se refieren al proceso de almacenar información en un medio informático. Un experto puede adoptar fácilmente cualquiera de los métodos conocidos actualmente para grabar información en un medio legible por ordenador para generar productos que comprenden una o más de las secuencias de ácidos nucleicos de acuerdo con la invención, una o más de las secuencias de polipéptidos de acuerdo con la invención. Otra realización de la invención es un medio legible por ordenador que tiene grabadas al menos 2, 5, 10, 15 o 20 o más secuencias de ácidos nucleicos o polipéptidos de acuerdo con la invención.

La presente divulgación además se refiere a un medio legible por ordenador que tiene grabada una o más secuencias de ácidos nucleicos de acuerdo con la invención, a un medio legible por ordenador que tiene grabada una o más secuencias de polipéptidos de acuerdo con la invención, medio legible por ordenador que tiene grabadas al menos 2, 5, 10, 15 o 20 o más de las secuencias de ácidos nucleicos o polipéptidos tal como se establecieron anteriormente.

Los medios legibles por ordenador incluyen medios legibles magnéticamente, medios legibles ópticamente, medios legibles electrónicamente y medios magnéticos/ópticos. Por ejemplo, los medios legibles por ordenador pueden ser un disco duro, un disco flexible, una cinta magnética, CD-ROM, disco versátil digital (DVD), memoria de acceso aleatorio (RAM) o memoria de solo lectura (ROM) así como también otros tipos de otros medios conocidos para los expertos en la técnica.

La presente divulgación además se refiere a sistemas (tales como sistemas en base a Internet), tales como sistemas informáticos que almacenan y manejan la información secuencial descrita en la presente. Un ejemplo de un sistema informático 100 se ilustra en formato de diagrama de bloques en la FIG. 9. Tal como se usa en la presente, "un sistema informático" se refiere a los componentes de equipo, a los componentes de programación y a los componentes de almacenamiento de datos utilizados para analizar una secuencia de nucleótidos de una secuencia de ácidos nucleicos de acuerdo con la invención o una secuencia de polipéptidos de acuerdo con la invención. En algunos casos, el sistema informático 100 incluye un procesador para procesar, acceder y manejar datos de secuencias. El procesador 105 cualquier tipo conocido de unidad de central procesamiento, tal como, por ejemplo, el Pentium III de Intel Corporation o procesadores similares de Sun, Motorola, Compaq, AMD o International Business Machines.

En algunos casos, el sistema informático 100 es un sistema de uso general que comprende el procesador 105 y uno o más componentes de almacenamiento de datos internos 110 para almacenar datos y uno o más dispositivos de recuperación de datos para recuperar los datos almacenados en los componentes de almacenamiento de datos. Un experto puede apreciar fácilmente que cualquiera de los sistemas informáticos disponible actualmente es adecuado.

En un caso, el sistema informático 100 incluye un procesador 105 conectado a un bus que está conectado a una memoria principal 115 (en una realización implementada como RAM) y uno o más dispositivos de almacenamiento de datos internos 110, tales como un disco duro y/u otro medio legible por ordenador que tenga datos grabados en el mismo. En algunos casos, el sistema informático 100 además incluye uno o más dispositivos de recuperación de datos 118 para leer los datos almacenados en los dispositivos de almacenamiento de datos internos 110.

El dispositivo de recuperación de datos 118 puede representar, por ejemplo, una unidad de disco flexible, una unidad de disco compacto, una unidad de cinta magnética o un módem capaz de conectarse con un sistema de almacenamiento de datos remoto (tal como a través de Internet) etc. En algunos casos, el dispositivo para almacenamiento de datos interno 110 es un medio legible por ordenador removible tal como un disco flexible, un disco compacto, una cinta magnética, etc. que contiene lógica de control y/o datos grabados en el mismo. El sistema informático 100 puede incluir o estar programado mediante un programa informático apropiado para leer la lógica de control y/o los datos del componente de almacenamiento de datos una vez que se inserta en el dispositivo de recuperación de datos.

El sistema informático 100 incluye una pantalla 120 que se utiliza para exhibir los datos de salida para el usuario del ordenador. Cabe mencionar además que el sistema informático 100 puede estar conectado a otros sistemas informáticos 125 a-c en una red o red de área extensa para proporcionar acceso centralizado al sistema informático 100.

El programa informático para acceder y procesar las secuencias de nucleótidos de una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma o una secuencia de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma (tales como herramientas de búsqueda, herramientas de comparación y herramientas de modelado, etc.) pueden residir en la memoria principal 115 durante la ejecución.

En algunos casos, el sistema informático 100 puede comprender adicionalmente un algoritmo de comparación de secuencias para comparar una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma o una secuencia de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma, almacenadas en un medio legible por ordenador con una secuencia(s) de nucleótidos o polipéptidos de referencia almacenada(s) en un medio legible por ordenador. Un "algoritmo de comparación de secuencias" se refiere a uno o más programas que se implementan (localmente o remotamente) en un sistema informático 100 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 secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma o una secuencia de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma, almacenadas en un medio legible por



ordenador con secuencias de referencia almacenadas en un medio legible por ordenador para identificar homologías o motivos estructurales.

La Figura 10 es un diagrama de flujo que ilustra una realización de un proceso 200 para comparar una secuencia nueva de nucleótidos o proteína con una base de datos de secuencias para determinar los niveles de homología entre la secuencia nueva y las secuencias en la base de datos. La base de datos de secuencias puede ser una base de datos privada almacenada dentro de un sistema informático 100 o una base de datos pública tal como GENBANK que está disponible a través de Internet.

El proceso 200 comienza en un estado inicial 201 y a continuación pasa a un estado 202 en donde la nueva secuencia que va a compararse se almacena en una memoria en un sistema informático 100. Tal como se trató anteriormente, la memoria podría ser cualquier tipo de memoria, incluida una RAM o un dispositivo de almacenamiento interno.

A continuación el proceso 200 pasa a un estado 204 en donde se abre una base de datos de secuencias para llevar a cabo el análisis y comparación. Luego el proceso 200 pasa a un estado 206 en donde la primera secuencia almacenada en la base de datos se lee en una memoria en el ordenador. A continuación se lleva a cabo una comparación en un estado 210 para determinar si la primera secuencia es igual a la segunda secuencia. Cabe destacar que esta etapa no se limita a llevar a cabo una comparación exacta entre la secuencia nueva y la primera secuencia en la base de datos. Los expertos en la técnica conocen métodos para comparar dos secuencias de nucleótidos o proteínas, incluso si no son idénticas. Por ejemplo, se pueden introducir huecos en una secuencia para elevar el nivel de homología entre dos secuencias evaluadas. Los parámetros que controlan si se introducen huecos u otras características a una secuencia durante la comparación generalmente son ingresados por el usuario del sistema informático.

Una vez llevada a cabo la comparación de las dos secuencias en el estado 210, en el estado de decisión 210 se determina si las dos secuencias son iguales. Por supuesto, el término "igual" no está limitado a secuencias que son absolutamente idénticas. Las secuencias que están dentro de los parámetros de homología ingresados por el usuario se marcarán como "iguales" en el proceso 200.

Si se determina que las dos secuencias son iguales, el proceso 200 pasa a un estado 214 en donde se exhibe al usuario el nombre de la secuencia de la base de datos. Este estado notifica al usuario que la secuencia con el nombre que se exhibe cumple con las restricciones de homología que se ingresaron. Una vez exhibido el nombre de la secuencia almacenada al usuario, el proceso 200 pasa a un estado de decisión 218 en donde se determina si existen más secuencias en la base de datos. Si no existen más secuencias en la base de datos, a continuación el proceso 200 finaliza en un estado final 220. Sin embargo, si existen más secuencias en la base de datos, a continuación el proceso 200 pasa a un estado 224 en donde se desplaza un señalador a la siguiente secuencia en la base de datos de forma que pueda compararse con la secuencia nueva. De esta forma, la secuencia nueva se alinea y compara con cada secuencia en la base de datos.

Cabe destacar que si en el estado de decisión 212 se determinó que las secuencias no son homólogas, a continuación el proceso 200 pasará inmediatamente al estado de decisión 218 para determinar si hay otras secuencias disponibles en la base de datos para hacer la comparación.

La presente divulgación además se refiere a un sistema informático que comprende un procesador, un dispositivo de almacenamiento de datos que tiene almacenado en el mismo una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma o una secuencia de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma, un dispositivo de almacenamiento de datos que tiene almacenado en el mismo de forma recuperable secuencias de nucleótidos o secuencias de polipéptidos de referencia para compararlas con una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma o una secuencia de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma y un comparador de secuencias para llevar a cabo la comparación. El comparador de secuencias puede indicar un nivel de homología entre las secuencias comparadas o identificar motivos estructurales en el código de ácido nucleico descrito anteriormente, una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma o una secuencia de polipéptidos de acuerdo con la invención y secuencias sustancialmente idéntica a la misma o puede identificar motivos estructurales en secuencias que se comparan con estos códigos de ácido nucleico y códigos de polipéptidos. En algunos casos, el dispositivo de almacenamiento de datos puede tener almacenadas las secuencias de al menos 2, 5, 10, 15, 20, 25, 30 o 40 o más de las secuencias de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a las mismas o las secuencias de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a las mismas.

La presente divulgación además se refiere a un método para determinar el nivel de homología entre una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma o una secuencia de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma y una secuencia de nucleótidos de referencia. El método incluye leer el código de ácidos nucleicos o el código de polipéptidos y la

5 secuencia de nucleótidos o polipéptidos de referencia a través del uso de un programa informático que determina los niveles de homología y determinar 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 informático. El programa informático puede ser cualquiera de varios programas informáticos para determinar niveles de homología, incluidos los indicados específicamente en la presente (tales como BLAST2N con los parámetros predeterminados o con cualesquiera parámetros modificados). El método puede implementarse utilizando los sistemas informáticos descritos anteriormente. El método también puede llevarse a cabo leyendo al menos 2, 5, 10, 15, 20, 25, 30 o 40 o más de las secuencias de ácido nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a las mismas o las secuencias de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a las mismas a través del uso del programa informático y determinando la homología entre los códigos de ácidos nucleicos o códigos de polipéptidos y las secuencias de nucleótidos o secuencias polipéptidos de referencia.

10 La FIG. 11 es un diagrama de flujo que ilustra una realización de un proceso 250 en un ordenador para determinar si dos secuencias son homólogas. El proceso 250 comienza en un estado inicial 252 y a continuación pasa a un estado 254 en donde una primera secuencia que va a compararse se almacena en una memoria. A continuación se almacena la segunda secuencia que va a compararse en una memoria en un estado 256. Luego el proceso 250 pasa a un estado 260 en donde se lee el primer carácter en la primera secuencia y a continuación se pasa a un estado 262 en donde se lee el primer carácter de la segunda secuencia. Se entenderá que si la secuencia es una secuencia de nucleótidos, entonces el carácter será normalmente A, T, C, G o U. Si la secuencia es una secuencia de proteínas, entonces está en el código de aminoácidos de una sola letra de forma que la primera y la segunda secuencia pueden compararse fácilmente.

15 A continuación se determina en un estado de decisión 264 si los dos caracteres son iguales. Si son iguales, entonces el proceso 250 pasa a un estado 268 en donde se leen los siguientes caracteres en la primera y segunda secuencia. Luego se determina si los siguientes caracteres son iguales. Si lo son, entonces el proceso 250 continúa este ciclo hasta que dos caracteres no sean iguales. Si se determina que los dos caracteres siguientes no son iguales, el proceso 250 pasa a un estado de decisión 274 para determinar si hay más caracteres para leer en las secuencias.

20 Si no hay más caracteres para leer, entonces el proceso 250 pasa a un estado 276 en donde se exhibe al usuario el nivel de homología entre la primera y segunda secuencia. El nivel de homología se determina calculando la proporción de caracteres iguales entre las secuencias con respecto a la cantidad total de secuencias en la primera secuencia. Por consiguiente, si cada carácter en una primera secuencia de 100 nucleótidos se alinea con cada carácter en una segunda secuencia, el nivel de homología sería 100%.

25 Alternativamente, el programa informático puede ser programa informático que compara las secuencias de nucleótidos de una secuencia de ácidos nucleico tal como se establece en la invención, con una o más secuencias de nucleótidos de referencia para determinar si el código de ácidos nucleicos de acuerdo con la invención y las secuencias sustancialmente idénticas a los mismos, difieren de una secuencia de ácidos nucleicos de referencia en una o más posiciones. Opcionalmente dicho programa registra la longitud e identidad de los nucleótidos insertados, eliminados o sustituidos con respecto a la secuencia de los polinucleótidos de referencia o una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma. En algunos casos, el programa informático puede ser un programa que determina si una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma, contiene un polimorfismo en un solo nucleótido (SNP) con respecto a una secuencia de nucleótidos de referencia.

30 Por lo tanto, la presente divulgación se refiere a un método para determinar si una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma, difieren en uno o más nucleótidos de 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 informático que identifica diferencias entre secuencias de ácidos nucleicos e identificar diferencias entre el código de ácidos nucleicos y la secuencia de nucleótidos de referencia con el programa informático. En algunos casos, el programa informático es un programa que identifica polimorfismos en un solo nucleótido. El método puede implementarse mediante los sistemas informáticos descritos anteriormente y el método ilustrado en la Figura 11. El método también puede llevarse a cabo leyendo al menos 2, 5, 10, 15, 20, 25, 30 o 40 o más de las secuencias de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a las mismas y las secuencias de nucleótidos de referencia a través del uso del programa informático e identificando diferencias entre los códigos de ácidos nucleicos y las secuencias de nucleótidos de referencia con el programa informático.

35 En otros casos, el sistema en base a un ordenador puede comprender adicionalmente un identificador para identificar características en una secuencia de ácidos nucleicos de acuerdo con la invención o una secuencia de polipéptidos de acuerdo con 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 acuerdo con la invención o una secuencia de polipéptidos de acuerdo con la invención. En algunos casos, el identificador puede comprender un programa que identifica un marco de lectura abierto en una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma.

La Figura 12 es un diagrama de flujo que ilustra un caso de un proceso identificador 300 para detectar la presencia de una característica en una secuencia. El proceso 300 comienza en un estado inicial 302 y a continuación pasa a un estado 304 en donde una primera secuencia que va a verificarse para determinar sus características se almacena en una memoria 115 en el sistema informático 100. El proceso 300 a continuación pasa a un estado 306 en donde se abre una base de datos de características de secuencias. Dicha base de datos incluirá una lista cada atributo de las características junto con el nombre de la característica. Por ejemplo, un nombre de característica podría ser "Codón de iniciación" y el atributo sería "ATG". Otro ejemplo sería el nombre de característica "secuencia TAATAA" y el atributo de la característica sería "TAATAA". Un ejemplo de dicha base de datos es la producida por el University of Wisconsin Genetics Computer Group. Alternativamente, las características pueden ser motivos polipeptídicos estructurales tales como hélices alfa, láminas beta o motivos polipeptídicos funcionales tales como sitios activos enzimáticos, motivos hélice-giro-hélice u otros motivos conocidos para los expertos en la técnica.

Una vez abierta la base de datos de características en el estado 306, el proceso 300 pasa a un estado 308 en donde se lee la primera característica en la base de datos. A continuación se hace una comparación del atributo de la primera característica con la primera secuencia en el estado 310. Luego se determina en un estado de decisión 316 si el atributo de la característica se encontró en la primera secuencia. Si se encontró el atributo, entonces el proceso 300 pasa a un estado 318 en donde se exhibe al usuario el nombre de la característica encontrada.

A continuación el proceso 300 pasa a un estado decisión 320 en donde se determina si existen más características en la base de datos. Si no existen más características, a continuación el proceso 300 finaliza en un estado final 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 vuelve al estado 310 en donde el atributo de la siguiente característica se compara con la primera secuencia. Cabe señalar que si el atributo de característica no se encuentra en la primera secuencia en el estado de decisión 316, el proceso 300 pasa directamente al estado de decisión 320 para determinar si existen más características en la base de datos.

Por consiguiente, la divulgación además proporciona a un método para identificar una característica dentro de una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma o una secuencia de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma, que comprende leer el código(s) de ácidos nucleicos o código(s) de polipéptidos a través del uso de un programa informático que identifica características en el/los mismo(s) e identificar características dentro del/los código(s) de ácidos nucleicos con el programa informático. En algunos casos, el programa informático comprende un programa que identifica marcos de lectura abiertos. El método puede llevarse a cabo leyendo una sola secuencia o al menos 2, 5, 10, 15, 20, 25, 30 o 40 o más de las secuencias de ácido nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a las mismas o las secuencias de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a las mismas, a través del uso del programa informático e identificando características dentro de los códigos de ácidos nucleicos o códigos de polipéptidos con el programa informático.

Una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma o una secuencia de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma, puede almacenarse y manejarse en varios programas de procesamiento de datos en una variedad de formatos. Por ejemplo, una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma o una secuencia de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma, puede almacenarse como un texto en un archivo de procesamiento de textos, tal como Microsoft WORD™ o WORDPERFECT™ o como un archivo ASCII en una variedad de programas de bases de datos conocidos para los expertos en la técnica, tales como DB2™, SYBASE™ u ORACLE™. Además, se pueden usar muchos programas informáticos y bases de datos como algoritmos de comparación de secuencias, identificadores o fuentes de secuencias de nucleótidos o secuencias de polipéptidos de referencia para compararlas con una secuencia de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma o una secuencia de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a la misma. La siguiente lista pretende proporcionar orientación sobre programas y bases de datos que son útiles con las secuencias de ácidos nucleicos de acuerdo con la invención y secuencias sustancialmente idénticas a las mismas o las secuencias de polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a las mismas.

Los programas y bases de datos que pueden utilizarse incluyen, a modo no taxativo: MACPATTERN™ (EMBL), DISCOVERYBASE™ (Molecular Applications Group), GENEMINE™ (Molecular Applications Group), LOOK™ (Molecular Applications Group), MACLOOK™ (Molecular Applications Group), BLAST y BLAST2 (NCBI), BLASTN y BLASTX (Altschul et al., J. Mol. Biol. 215: 403, 1990), FASTA (Pearson and Lipman, Proc. Natl. Acad. Sci. Estados Unidos, 85: 2444, 1988), FASTDB (Brutlag et al. Comp. App. Biosci. 6:237-245, 1990), CATALYST™ (Molecular Simulations Inc.), Catalyst/SHAPE™ (Molecular Simulations Inc.), Cerius<sup>2</sup>.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 World Drug Index de Derwents, la base de datos BioByteMasterFile, la base de datos Genbank y la base de datos Genseqn. Muchos otros programas y bases de datos serán evidentes para un experto en la técnica, dada la presente divulgación.

Los motivos que pueden detectarse utilizando los programas mencionados anteriormente incluyen secuencias que codifican cierres de leucina, motivos hélice-giro-hélice, sitios de glicosilación, sitios de ubiquitinación, hélices alfa y láminas beta, secuencias señal que codifican péptidos señal que dirigen la secreción de las proteínas codificadas, secuencias implicadas en la regulación de la transcripción tales como homeosecuencias, extensiones ácidas, sitios activos enzimáticos, sitios de unión a sustrato y sitios de escisión enzimática.

#### Hibridación de ácidos nucleicos

La invención proporciona ácidos nucleicos aislados, sintéticos o recombinantes que se hibridan en condiciones rigurosas con una secuencia de SEQ ID NO:275. Las condiciones rigurosas pueden ser condiciones de extrema rigurosidad, condiciones de rigurosidad media y/o condiciones de rigurosidad baja, incluidas las condiciones de rigurosidad alta y reducida descritas en la presente. En algunas realizaciones, es la rigurosidad de las condiciones de lavado la que establece las condiciones que determinan si un ácido nucleico está dentro del alcance de acuerdo con la invención, tal como se trata más adelante.

“Hibridación” se refiere al proceso por el cual una cadena de ácido nucleico se une a una cadena complementaria a través de apareamiento de bases. Las reacciones de hibridación pueden ser sensibles y selectivas de forma que una secuencia específica de interés puede identificarse incluso en muestras en las que está en una concentración baja. Las condiciones adecuadamente rigurosas pueden definirse, por ejemplo, por las concentraciones de sal o formamida en las soluciones de prehibridación e hibridación o por la temperatura de hibridación y se conocen en la técnica. En otras realizaciones, la rigurosidad puede aumentarse reduciendo la concentración de sal, aumentando la concentración de formamida o elevando la temperatura de hibridación. En otras realizaciones, los ácidos nucleicos de acuerdo con la invención se definen por su capacidad para hibridarse en varias condiciones de rigurosidad (tales como alta, media y baja), tal como se establece en la presente.

En algunas realizaciones, la hibridación en condiciones de rigurosidad alta comprende aproximadamente 50% de formamida a aproximadamente 37°C a 42°C. En algunas realizaciones, las condiciones de hibridación comprenden condiciones de rigurosidad baja en aproximadamente 35% a 25% de formamida a aproximadamente 30°C a 35°C. En algunas realizaciones, las condiciones de hibridación comprenden condiciones de rigurosidad alta, tales como a 42°C en 50% de formamida, 5×SSPE, 0,3% de SDS y 200 µg/ml de ADN de esperma de salmón cizallado y desnaturalizado. En algunas realizaciones, las condiciones de hibridación comprenden estas condiciones de rigurosidad reducida, pero en 35% de formamida a una temperatura reducida de 35°C. El rango de temperatura correspondiente a un nivel de rigurosidad específico puede estrecharse adicionalmente calculando la relación entre purina y pirimidina del ácido nucleico de interés y ajustando la temperatura en consecuencia. Las variaciones en los rangos y condiciones mencionados anteriormente se conocen en la técnica.

En otras realizaciones, los ácidos nucleicos de acuerdo con la invención tal como se definen por su capacidad para hibridarse en condiciones rigurosas pueden ser los ácidos nucleicos la longitud completa de acuerdo con la invención.

En algunas realizaciones, los ácidos nucleicos de acuerdo con la invención se definen por su capacidad para hibridarse en condiciones de rigurosidad que comprenden condiciones de aproximadamente 50% de formamida a aproximadamente 37°C a 42°C. En algunas realizaciones, los ácidos nucleicos de acuerdo con la invención se definen por su capacidad para hibridarse en condiciones de rigurosidad que comprenden condiciones en aproximadamente 35% a 25% de formamida a aproximadamente 30°C a 35°C.

Alternativamente, los ácidos nucleicos de acuerdo con la invención se definen por su capacidad para hibridarse en alta rigurosidad comprendiendo condiciones a 42°C en 50% de formamida, 5×SSPE, 0,3% de SDS y un ácido nucleico de bloqueo de secuencia repetitiva, tal como ADN cot-1 o de esperma de salmón (tal como 200 µg/ml de ADN de esperma de salmón cizallado y desnaturalizado). En algunas realizaciones, los ácidos nucleicos de acuerdo con la invención se definen por su capacidad para hibridarse en condiciones de rigurosidad reducida que comprenden 35% a 40% de formamida a una temperatura reducida de 35°C a 42°C.

En reacciones de hibridación de ácidos nucleicos, las condiciones utilizadas para alcanzar un nivel específico de rigurosidad variarán dependiendo de la naturaleza de los ácidos nucleicos que se van a hibridar. Por ejemplo, se pueden tener en cuenta la longitud, el grado de complementariedad, la composición de la secuencia de nucleótidos (tal como contenido de GC con respecto a AT) y tipo de ácido nucleico (tal como ARN o ADN) de las regiones de hibridación de los ácidos nucleicos al seleccionar las condiciones de hibridación. Una consideración adicional es si uno de los ácidos nucleicos está inmovilizado, por ejemplo, en un filtro.

La hibridación puede llevarse a cabo en condiciones de rigurosidad baja, rigurosidad moderada o rigurosidad alta. Como ejemplo de hibridación de ácido nucleico, primero se prehibrida una membrana polimérica que contiene ácidos nucleicos desnaturalizados inmovilizados durante 30 minutos a 45°C en una solución que consiste en NaCl 0,9 M, 50 mM de NaH<sub>2</sub>PO<sub>4</sub>, pH 7,0, 5,0 mM de Na<sub>2</sub>EDTA, 0,5% de SDS, 10× de solución de Denhardt y 0,5 mg/ml de ácido polirriboadenílico. A continuación se agregan a la solución aproximadamente 2×10<sup>7</sup> cpm (actividad específica 4-9×10<sup>8</sup> cpm/μg) de sonda de nucleótidos etiquetada en el extremo con <sup>32</sup>P. Después de 12-16 horas de incubación, se lava la membrana durante 30 minutos a temperatura ambiente en 1×SET (150 mM de NaCl, 20 mM de clorhidrato de Tris, pH 7,8, 1 mM de Na<sub>2</sub>EDTA) que contiene 0,5% de SDS, con un lavado posterior de 30 minutos en 1×SET nueva a T<sub>r</sub>-10°C para la sonda de oligonucleótidos. A continuación la membrana se expone a una película autorradiográfica para detectar señales de hibridación. Todas las hibridaciones precedentes se considerarían hibridaciones en condiciones de rigurosidad alta.

Tras la hibridación, el filtro puede lavarse para eliminar cualquier sonda no unida específicamente detectable. También se puede variar la rigurosidad utilizada para lavar los filtros dependiendo de la naturaleza de los ácidos nucleicos que se van a hibridar, la longitud de los ácidos nucleicos que se van a hibridar, el grado de complementariedad, la composición de la secuencia de nucleótidos (tal como contenido de GC con respecto a AT) y el tipo de ácido nucleico (tal como ARN o ADN). A continuación figuran ejemplos de lavados en condiciones de rigurosidad progresivamente más altas: 2×SSC, 0,1% de SDS a temperatura ambiente durante 15 minutos (rigurosidad baja); 0,1×SSC, 0,5% de SDS a temperatura ambiente durante 30 minutos a 1 hora (rigurosidad moderada); 0,1×SSC, 0,5% de SDS durante 15 a 30 minutos entre la temperatura de hibridación y 68°C (rigurosidad alta); y NaCl 0,15M durante 15 minutos a 72°C (rigurosidad muy alta). Se puede llevar a cabo un lavado final de baja rigurosidad en 0,1×SSC a temperatura ambiente. Los ejemplos anteriores son meramente ilustrativos de un conjunto de condiciones que pueden utilizarse para lavar los filtros. Un experto en la técnica sabrá que existen numerosas recetas para lavados de rigurosidad diferentes. Más adelante se proporcionan otros ejemplos.

En algunas realizaciones, las condiciones de hibridación comprenden una etapa de lavado que comprende un lavado durante 30 minutos a temperatura ambiente en una solución que comprende 1×150 mM de NaCl, 20 mM de clorhidrato de Tris, pH 7,8, 1 mM de Na<sub>2</sub>EDTA, 0,5% de SDS, con un lavado posterior de 30 minutos en solución nueva.

Los ácidos nucleicos que se hibridaron con la sonda se identifican mediante autorradiografía u otras técnicas convencionales.

La divulgación además se refiere a identificar ácidos nucleicos que tienen niveles reducidos de identidad secuencial (homología) con la secuencia de la sonda. Por ejemplo, para obtener ácidos nucleicos de identidad secuencial (homología) reducida con respecto a la sonda detectable, se pueden usar condiciones menos rigurosas. Por ejemplo, la temperatura de hibridación puede reducirse en incrementos de 5°C de 68°C a 42°C en una solución amortiguadora de hibridación que tiene una concentración de Na<sup>+</sup> de aproximadamente 1M. Tras la hibridación, el filtro puede lavarse con 2×SSC, 0,5% de SDS a la temperatura de hibridación. Estas condiciones se consideran que son 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 se lleva a cabo a 55°C. Un ejemplo específico de condiciones de hibridación "de rigurosidad baja" es cuando la hibridación anterior se lleva a cabo a 45°C.

Ejemplos adicionales incluyen soluciones amortiguadoras, tales como 6×SSC que contiene formamida a una temperatura de 42°C. En este caso, la concentración de formamida en la solución amortiguadora de hibridación puede reducirse en incrementos de 5% de 50% a 0% para identificar clones que tienen niveles reducidos de homología con la sonda. Tras la hibridación, el filtro puede lavarse con 6×SSC, 0,5% de SDS a 50°C. Estas condiciones se consideran que son 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 se lleva a cabo a 30% de formamida. Un ejemplo específico de condiciones de hibridación "de rigurosidad baja" es cuando la hibridación anterior se lleva a cabo a 10% de formamida.

Sin embargo, la selección de un formato de hibridación puede no ser esencial - la rigurosidad de las condiciones de lavado es la que establece las condiciones que determinan si un ácido nucleico está dentro del alcance de acuerdo con la invención. Las condiciones de lavado utilizadas para identificar ácidos nucleicos dentro del alcance de acuerdo con la invención incluyen, por ejemplo: una concentración de sal de aproximadamente 0,02 molar a pH 7 y una temperatura de al menos aproximadamente 50°C o aproximadamente 55°C hasta aproximadamente 60°C; o una concentración de sal de aproximadamente NaCl 0,15 M a 72°C durante aproximadamente 15 minutos; o una concentración de sal de aproximadamente 0,2×SSC a una temperatura de al menos aproximadamente 50°C o aproximadamente 55°C hasta aproximadamente 60°C durante aproximadamente 15 a aproximadamente 20 minutos; o el complejo de hibridación se lava dos veces con una solución con una concentración de sal de aproximadamente 2×SSC que contiene 0,1% SDS a temperatura ambiente durante 15 minutos y a continuación se lava dos veces con 0,1×SSC que contiene 0,1% de SDS a 68°C durante 15 minutos; o condiciones equivalentes. Ver Sambrook ed., MOLECULAR CLONING: A LABORATORY MANUAL (2a ed.), tomos 1-3, Cold Spring Harbor Laboratory (1989), LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY: HYBRIDIZATION WITH NUCLEIC ACID PROBES, Parte I. Theory and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y. (1993) y Ausubel,

ed. John Wiley & Sons, Inc., Nueva York (1997) para obtener una descripción de la solución amortiguadora SSC y condiciones equivalentes.

Estos métodos pueden utilizarse para aislar o identificar ácidos nucleicos de acuerdo con la invención y las secuencias sustancialmente idénticas a los mismos. Por ejemplo, los métodos precedentes pueden utilizarse para  
 5 aislar o identificar ácidos nucleicos que tienen una secuencia con al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más identidad secuencial (homología) con una secuencia de ácidos nucleicos seleccionada del grupo que consiste en una de las  
 10 secuencias de acuerdo con la invención y secuencias sustancialmente idénticas a las mismas o fragmentos que comprenden 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 las secuencias complementarias a las mismas. La identidad secuencial (homología) puede medirse utilizando el algoritmo de alineación. Por ejemplo, los polinucleótidos homólogos pueden tener una secuencia codificadora que es una variante alélica de origen natural de una de las secuencias codificadoras descritas en la presente. Dichas variantes alélicas pueden tener una sustitución, eliminación o adición de uno o más nucleótidos cuando se comparan con los ácidos nucleicos de acuerdo con la invención. Adicionalmente, los  
 15 procedimientos mencionados anteriormente pueden utilizarse para aislar ácidos nucleicos que codifican polipéptidos que tienen al menos aproximadamente 99%, 95%, al menos 90%, al menos 85%, al menos 80% de identidad secuencial (homología) con un polipéptido de acuerdo con 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 tal como se determina utilizando un algoritmo de alineación de secuencias (tal como el algoritmo FASTA versión 3.0t78 con los parámetros predeterminados).

#### Sondas de oligonucleótidos y métodos para utilizarlas

La invención también proporciona sondas de ácidos nucleicos que pueden utilizarse, por ejemplo, para identificar, amplificar o aislar ácidos nucleicos que codifican un polipéptido que tiene actividad enzimática de aldolasa, tal como  
 25 piruvato aldolasa, tal como HMG y/o KHG o fragmentos del mismo o para identificar genes de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. En algunos ejemplos, la sonda comprende al menos aproximadamente 10 bases consecutivas de un ácido nucleico de acuerdo con la invención. Alternativamente, una sonda de acuerdo con la invención 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 10 a  
 30 50, aproximadamente 20 a 60, aproximadamente 30 a 70 bases consecutivas de una secuencia tal como se establece en un ácido nucleico de acuerdo con la invención. Las sondas identifican un ácido nucleico mediante unión y/o hibridación. Las sondas pueden utilizarse en arreglos de acuerdo con la invención, ver la descripción más adelante, incluidos arreglos capilares. Las sondas de acuerdo con la invención además pueden utilizarse para aislar otros ácidos nucleicos o polipéptidos.

Los ácidos nucleicos aislados, sintéticos o recombinantes de acuerdo con la invención, las secuencias complementarias a los mismos o un fragmento que comprende al menos aproximadamente 10, 15, 20, 25, 30, 35,  
 35 40, 50, 75, 100, 150, 200, 300, 400 o 500 bases consecutivas de una de las secuencias de acuerdo con la invención o las secuencias complementarias a las mismas también pueden utilizarse como sondas para determinar si una muestra biológica, tal como una muestra de aceite, contiene un organismo que tiene una secuencia de ácidos nucleicos de acuerdo con la invención o un organismo del cual se obtuvo el ácido nucleico. En dichos procedimientos, se obtiene una muestra biológica que potencialmente está albergando el organismo del cual se aisló el ácido nucleico y se obtienen ácidos nucleicos de la muestra. Los ácidos nucleicos se ponen en contacto con la sonda en condiciones que permiten que la sonda se hibride específicamente con cualquier secuencia complementaria de las que están presentes en la misma.

Cuando sea necesario, se pueden determinar las condiciones que permiten que la sonda se hibride específicamente  
 45 con secuencias complementarias poniendo en contacto la sonda con secuencias complementarias de muestras que se sabe que contienen la secuencia complementaria así como también con secuencias testigo que no tienen la secuencia complementaria. Las condiciones de hibridación, tales como la concentración de sal de la solución amortiguadora de hibridación, la concentración de formamida de la solución amortiguadora de hibridación o la temperatura de hibridación pueden variarse para identificar condiciones que permiten que la sonda se hibride específicamente con ácidos nucleicos complementarios.

Si la muestra contiene el organismo del cual se aisló el ácido nucleico, a continuación se detecta la hibridación específica de la sonda. La hibridación puede detectarse etiquetando la sonda con un agente detectable tal como un isótopo radioactivo, un tinte fluorescente o una enzima capaz de catalizar la formación de un producto detectable.

Los expertos en la técnica conocen muchos métodos para utilizar las sondas etiquetadas para detectar la presencia  
 55 de ácidos nucleicos complementarios en una muestra. Estos incluyen Transferencias Southern, Transferencias Northern, procedimientos de hibridación de colonias y transferencias en mancha. Se proporcionan protocolos para cada uno de estos procedimientos en Ausubel et al. Current Protocols in Molecular Biology, John Wiley & Sons, Inc. 1997 y Sambrook et al., Molecular Cloning: A Laboratory Manual 2a Ed., Cold Spring Harbor Laboratory Press (1989).

Alternativamente, se puede utilizar más de una sonda (al menos una que es capaz de hibridarse específicamente con cualquier secuencia complementaria que está presente en la muestra de ácido nucleico), en una reacción de amplificación para determinar si la muestra contiene un organismo que contiene una secuencia de ácidos nucleicos de acuerdo con la invención (tal como un organismo del cual se aisló el ácido nucleico). En algunas realizaciones, las sondas comprenden oligonucleótidos. En algunas realizaciones, la reacción de amplificación puede comprender una reacción PCR. Los protocolos de PCR se describen en Ausubel y Sambrook, supra. Alternativamente, la amplificación puede comprender una reacción en cadena de ligasa, 3SR o una reacción de desplazamiento de cadena. (Ver 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 dichos procedimientos, los ácidos nucleicos en la muestra se ponen en contacto con las sondas, se lleva a cabo la reacción de amplificación y se detecta cualquier producto de amplificación resultante. El producto de amplificación puede detectarse llevando a cabo electroforesis en gel en los productos de reacción y tiñendo el gel con un intercalador tal como bromuro de etidio. Alternativamente, una o más de las sondas pueden etiquetarse con un isótopo radioactivo y la presencia de un producto de amplificación radioactivo puede detectarse mediante autorradiografía después de la electroforesis en gel.

Las sondas derivadas de secuencias próximas a los extremos de la secuencia de acuerdo con la invención también pueden utilizarse en procedimientos de desplazamiento sobre el cromosoma para identificar clones que contienen secuencias genómicas adyacentes a las secuencias de acuerdo con la invención. Dichos métodos permiten el aislamiento de genes que codifican proteínas adicionales del organismo huésped.

En algunas realizaciones, los ácidos nucleicos aislados, sintéticos o recombinantes de acuerdo con la invención, 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 o más bases consecutivas de una de las secuencias de acuerdo con la invención o las secuencias complementarias a las mismas se utilizan como sondas para identificar y aislar ácidos nucleicos relacionados. En algunas realizaciones, los ácidos nucleicos relacionados pueden ser ADNc o ADN genómicos de organismos distintos al organismo de cual se aisló el ácido nucleico. Por ejemplo, los otros organismos pueden ser organismos relacionados. En dichos procedimientos, una muestra de ácido nucleico se pone en contacto con la sonda en condiciones que permiten que la sonda se hibride específicamente con secuencias relacionadas. La hibridación de la sonda con los ácidos nucleicos del organismo relacionado a continuación se detecta utilizando cualquiera de los métodos descritos anteriormente.

Al variar la rigurosidad de las condiciones de hibridación utilizadas para identificar ácidos nucleicos, tales como ADNc o ADN genómicos, los cuales se hibridan con las sondas detectables, se pueden identificar y aislar los ácidos nucleicos que tienen diferentes niveles de homología con la sonda. La rigurosidad puede variarse llevando a cabo la hibridación a temperaturas variables por debajo de las temperaturas de fusión de las sondas. La temperatura de fusión,  $T_f$ , es la temperatura (a fuerza iónica y pH definidos) en la cual el 50% de la secuencia objetivo se hibrida con una sonda perfectamente complementaria. Las condiciones muy rigurosas se seleccionan a aproximadamente 5°C menos o igual que la  $T_f$  para una sonda específica. La temperatura de fusión de la sonda puede calcularse utilizando las siguientes fórmulas:

Para sondas de entre 14 y 70 nucleótidos de longitud, la temperatura de fusión ( $T_f$ ) se calcula utilizando la fórmula:  $T_f = 81,5 + 16,6(\log [Na^+]) + 0,41(\text{fracción } G+C) - (600/N)$  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 calcularse utilizando la ecuación:  $T_f = 81,5 + 16,6(\log [Na^+]) + 0,41(\text{fracción } G+C) - (0,63\% \text{ formamida}) - (600/N)$  donde N es la longitud de la sonda.

La prehibridación puede llevarse a cabo en 6×SSC, 5× de reactivo de Denhardt, 0,5% de SDS, 100 µg/ml de ADN de esperma de salmón desnaturalizado fragmentado o 6×SSC, 5× de reactivo de Denhardt, 0,5% de SDS, 100 µg/ml de ADN de esperma de salmón desnaturalizado fragmentado, 50% de formamida. Las fórmulas de las soluciones SSC y de Denhardt se indican en Sambrook et al., supra.

En algunas realizaciones, la hibridación se lleva a cabo agregando la sonda detectable en las soluciones de prehibridación indicadas anteriormente. Cuando la sonda comprende ADN de doble hebra, el mismo se desnaturaliza antes de la adición de la solución de hibridación. En algunas realizaciones, el filtro se pone en contacto con la solución de hibridación durante un período de tiempo suficiente para permitir que la sonda se hibride con los ADNc o ADN genómicos que contienen secuencias complementarias a la misma o homólogas a la misma. Para sondas con más de 200 nucleótidos de longitud, la hibridación puede llevarse a cabo a 15-25°C por debajo de la  $T_f$ . Para sondas más cortas, tales sondas de oligonucleótidos, la hibridación puede llevarse a cabo a 5-10°C por debajo de la  $T_f$ . En algunas realizaciones, para hibridaciones en 6×SSC, la hibridación se lleva a cabo a aproximadamente 68°C. Generalmente, para hibridaciones en soluciones que contienen 50% de formamida, la hibridación se lleva a cabo a aproximadamente 42°C.

Inhibición de la expresión de enzimas aldolasas

La invención proporciona ácidos nucleicos complementarios (tales como secuencias antisentido) a los ácidos nucleicos de acuerdo con la invención, tales como ácidos nucleicos que codifican enzimas aldolasas, tales como ácidos nucleicos que comprenden antisentido, ARNsi, miARN y ribozimas. Los ácidos nucleicos de acuerdo con la invención que comprenden secuencias antisentido pueden ser capaces de inhibir el transporte, empalme o transcripción de genes que codifican enzimas aldolasas. La inhibición puede llevarse a cabo a través del direccionamiento del ADN genómico o ARN mensajero. La transcripción o función del ácido nucleico objetivo puede inhibirse, por ejemplo, mediante hibridación y/o escisión. En un conjunto ejemplar de inhibidores proporcionados por la presente invención incluye oligonucleótidos que son capaces de unirse al gen o mensaje de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, evitando o inhibiendo en ambos casos la producción o función de una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa. La asociación puede ser a través de hibridación específica de secuencia. Otro tipo de inhibidores útiles incluye oligonucleótidos que provocan la desactivación o escisión del mensaje de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. El oligonucleótido puede tener actividad enzimática la cual provoca dicha escisión, tal como ribozimas. El oligonucleótido puede estar químicamente modificado o conjugado a una enzima o composición capaz de escindir el ácido nucleico complementario. Un acervo de varios de dichos oligonucleótidos diferentes puede someterse a detección para detectar aquellos con la actividad deseada. Por consiguiente, la invención proporciona varias composiciones para la inhibición de la expresión de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa a nivel de ácido nucleico y/o proteína, tal como antisentido, ARNsi, miARN y ribozimas que comprenden secuencias de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención y los anticuerpos anti-aldolasa, tales como anti-piruvato aldolasa, tales como anti-HMG y/o anti-KHG aldolasa de acuerdo con la invención.

La inhibición de la expresión de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa puede tener una variedad de aplicaciones industriales. Por ejemplo, la inhibición de la expresión de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa puede retardar o evitar el deterioro de los alimentos. En algunas realizaciones, el uso de composiciones de acuerdo con la invención que inhiben la expresión y/o actividad de las enzimas aldolasas, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, tales como anticuerpos, oligonucleótidos antisentido, ribozimas, ARNsi y miARN se utilizan para retardar o evitar el deterioro de los alimentos. Por consiguiente, en algunas realizaciones, la invención proporciona métodos y composiciones que comprenden la aplicación sobre una planta o producto vegetal (tal como un cereal, un grano, fruto, semilla, raíz, hoja, etc.) de anticuerpos, oligonucleótidos antisentido, ribozimas, ARNsi y miARN de acuerdo con la invención para retardar o evitar el deterioro de los alimentos. Estas composiciones también pueden expresarse mediante la planta (tal como una planta transgénica) u otro organismo (tal como una bacteria u otro microorganismo transformado con un gen de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención).

Estas composiciones de acuerdo con la invención para la inhibición de la expresión de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa (tales como ARNi, ribozimas antisentido, anticuerpos) pueden utilizarse como composiciones farmacéuticas, tales como agentes antipatógenos o en otras terapias, tales como agentes antimicrobianos para, por ejemplo, la *Salmonella*.

Oligonucleótidos antisentido

La divulgación proporciona adicionalmente oligonucleótidos antisentido capaces de unirse a un mensaje de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, los cuales, en algunas realizaciones, pueden inhibir la actividad enzimática de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa dirigiéndose al ARNm. En la literatura científica y de patentes se describen estrategias para diseñar oligonucleótidos y el experto en la técnica puede diseñar dichos oligonucleótidos de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa utilizando los nuevos reactivos de acuerdo con la invención. Por ejemplo, los protocolos de desplazamiento sobre el gen/mapeo de ARN para detectar los oligonucleótidos antisentido efectivos se conocen en la técnica, ver Ho (2000) *Methods Enzymol.* 314:168-183, que describe un ensayo de mapeo de ARN, el cual está basado en técnicas moleculares estándares para proporciona un método fácil y confiable para la selección convincente de secuencias antisentido. Ver también Smith (2000) *Eur. J. Pharm. Sci.* 11:191-198.

Los ácidos nucleicos de origen natural se utilizan como oligonucleótidos antisentido. Estos oligonucleótidos antisentido pueden tener cualquier longitud; por ejemplo, en otras realizaciones, los oligonucleótidos antisentido son de aproximadamente 5 a aproximadamente 100, aproximadamente 10 a aproximadamente 80, aproximadamente 15 a aproximadamente 60, aproximadamente 18 a aproximadamente 40. La longitud óptima puede determinarse mediante detección habitual. Los oligonucleótidos antisentido pueden estar presentes en cualquier concentración. La concentración óptima puede determinarse mediante detección habitual. Se conoce una amplia variedad de nucleótidos sintéticos, de origen no natural y análogos de ácidos nucleicos que pueden tratar este problema potencial. Por ejemplo, se pueden utilizar ácidos nucleicos peptídicos (PNA) que contienen una cadena principal no iónica, tales como unidades N-(2-aminoetil)glicina. También se pueden utilizar oligonucleótidos antisentido que tienen enlaces fosforotioato, como se describen 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 cadena principal de ADN sintética proporcionados en la invención también pueden incluir ácidos nucleicos de fósforo-ditioato, metilfosfonato, fosforamidoato, alquil fosfotriéster, sulfamato, 3'-tioacetil, metileno(metilimino), 3'-N-carbamato y morfolino carbamato, tales como los descritos anteriormente.

Se puede utilizar metodología de química combinatoria para crear grandes cantidades de oligonucleótidos que pueden someterse a detección sistemática rápidamente para detectar oligonucleótidos específicos que tienen las afinidades y especificidades de unión apropiadas con cualquier objetivo, tales como las secuencias de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa sentido y antisentido de acuerdo con la invención (ver Gold (1995) J. of Biol. Chem. 270:13581-13584).

#### Ribozimas inhibitoras

La divulgación proporciona además ribozimas capaces de unir el mensaje de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. Estas ribozimas pueden inhibir la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, por ejemplo, apuntando al ARNm. Las estrategias para diseñar ribozimas y seleccionar la secuencia antisentido específica de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa como objetivo se describen bien en la bibliografía científica y de patentes, y el experto en la técnica puede diseñar dichas ribozimas usando los reactivos novedosos de acuerdo con la invención. Las ribozimas actúan uniéndose a un ARN objetivo a través de la porción de unión a ARN de una ribozima que se mantiene en proximidad cercana a una porción enzimática del ARN que escinde el ARN objetivo. Por lo tanto, la ribozima reconoce y se une a un ARN objetivo a través del apareamiento de bases complementario, y una vez unida al sitio correcto, actúa enzimáticamente para escindir e inactivar el ARN objetivo. La escisión de un ARN objetivo de este modo destruirá su capacidad de dirigir la síntesis de una proteína codificada si la escisión ocurre en la secuencia de codificación. Luego de que una ribozima se haya unido a su objetivo de ARN y lo haya escindido, puede liberarse de dicho ARN para unirse y escindir nuevos objetivos de forma repetida.

En algunas circunstancias, la naturaleza enzimática de una ribozima puede ser ventajosa sobre otras tecnologías, tales como tecnología antisentido (donde una molécula de ácido nucleico simplemente se une a un objetivo de ácido nucleico para bloquear su transcripción, traducción o asociación con otra molécula) como la concentración efectiva de la ribozima necesaria para efectuar un tratamiento terapéutico puede ser menor que aquel de un oligonucleótido antisentido. Esta ventaja potencial refleja la capacidad de la ribozima de actuar enzimáticamente. Por lo tanto, una ribozima única es capaz de escindir varias moléculas de ARN objetivo. En algunas realizaciones, una ribozima es un inhibidor altamente específico, con la especificidad de inhibición que no depende solamente del mecanismo de apareamiento de bases de unión, sino también del mecanismo por el cual la molécula inhibe la expresión del ARN al cual se une. Es decir, la inhibición está causada por la escisión del objetivo de ARN y, por lo tanto, la especificidad se define como la relación de la tasa de escisión del ARN dirigido sobre la tasa de escisión del ARN no dirigido. Este mecanismo de escisión depende de factores adicionales a aquellos que participan en el apareamiento de bases. Por lo tanto, la especificidad de la acción de una ribozima puede ser mayor que la del oligonucleótido antisentido que se une al mismo sitio de ARN.

Una molécula de ARN de ribozima enzimática puede formarse en un motivo de cabeza de martillo, motivo de horquilla, un motivo de intrón de 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; los motivos de horquilla por Hampel (1989) Biochemistry 28:4929, y Hampel (1990) Nuc. Acids Res. 18:299; el motivo del virus de la hepatitis delta por Perrotta (1992) Biochemistry 31:16; el motivo de RNasaP por Guerrier-Takada (1983) Cell 35:849; y el intrón de grupo I por Cech, Patente de los Estados Unidos No. 4.987.071. Los expertos en la técnica reconocerán que una ribozima de acuerdo con la invención, tal como una molécula de ARN enzimática de la presente invención, puede tener un sitio de unión del sustrato específico complementario a una o más de las regiones de ARN del gen objetivo. Una ribozima de acuerdo con la invención puede tener una secuencia de nucleótidos dentro o alrededor del sitio de unión del sustrato que imparte una actividad de escisión de ARN a la molécula.

#### **Interferencia de ARN (ARNi)**

La divulgación proporciona moléculas inhibitoras de ARN, las denominadas moléculas de "ARNi", que comprenden secuencias de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de acuerdo con la invención. La molécula de ARNi puede comprender una molécula de ARN de doble hebra (ARNds), tal como moléculas de ARNsi, miARN y/o ARN de horquilla corta (ARNsh). La molécula de ARNi, tal como ARNsi (ARN inhibidor pequeño) y/o miARN (micro ARN), pueden inhibir la expresión de un gen de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa. En algunos casos la molécula de ARNi, tal como ARNsi y/o miARN es de 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 ingresar en una célula y causar la degradación de un ARN de una sola hebra (ARNss) de secuencias similares o idénticas, incluyendo ARNm endógenos. Cuando una célula se expone a un ARN de doble hebra (ARNds), el ARNm del gen homólogo se degrada selectivamente mediante un proceso denominado interferencia de ARN (ARNi). Un

mecanismo básico posible detrás del ARNi es la rotura de un ARN de doble hebra (ARNds) que coincide con una secuencia de gen específica en piezas cortas denominadas ARN pequeño de interferencia, que desencadena la degradación de ARNm que coincide con su secuencia. En algunos casos los ARNi de acuerdo con la invención se usan en tratamientos de silenciamiento de gen, ver Shuey (2002) Drug Discov. Today 7:1040-1046. En algunas realizaciones, la invención proporciona métodos para degradar selectivamente ARN usando las moléculas de ARNi tales como ARNsi y/o miARN de acuerdo con la invención. El proceso puede realizarse in vitro, ex vivo o in vivo. En algunos casos, las moléculas de ARNi pueden usarse para generar una mutación de pérdida de función en una célula, órgano o un animal.

La introducción intracelular del ARNi puede ser por internalización de un ligando específico de célula objetivo unido a una proteína de unión de ARN que comprende un ARNi (tal como micro ARN) que se adsorbe. El ligando es específico a un antígeno de superficie de célula objetivo única. El ligando puede internalizarse espontáneamente luego de unirse al antígeno de superficie celular. Si el antígeno de superficie celular único no se internaliza naturalmente luego de unirse a su ligando, la internalización puede promoverse mediante la incorporación de un péptido rico en arginina u otro péptido permeable a la membrana en la estructura del ligando o la proteína de unión de ARN o unión de dicho péptido al ligando o la proteína de unión de ARN. Vea la Solicitud de Patente de los Estados Unidos No. 20060030003; 20060025361; 20060019286; 20060019258. La divulgación proporciona formulaciones basadas en lípidos para administrar, tal como introducir ácidos nucleicos de la invención como partículas de lípidos-ácidos nucleicos que comprenden una molécula de ARNi a una célula, ver por ejemplo la Publicación de la Solicitud de Patente de los Estados Unidos No. 20060008910.

#### Modificación de ácidos nucleicos – preparación de enzimas variantes de la invención

La invención proporciona métodos para generar variantes de los ácidos nucleicos de acuerdo con la invención, tales como aquellos que codifican una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa. Estos métodos pueden repetirse o usarse en varias combinaciones para generar enzimas aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa que tienen una actividad alterada o diferente o una estabilidad alterada o diferente de una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa codificada por el ácido nucleico plantilla. Estos métodos también pueden repetirse o usarse en varias combinaciones tal como para generar variaciones en la expresión del gen/mensaje, traducción del mensaje o estabilidad del mensaje. En otras realizaciones, la composición genética de una célula se altera mediante por ejemplo la modificación de un gen homólogo ex vivo, seguido por su reinserción en la célula.

Un ácido nucleico de acuerdo con la invención puede alterarse de varios modos. Por ejemplo, métodos aleatorios o estocásticos, o no estocásticos, o métodos de "evolución dirigida", ver la Patente de los Estados Unidos No. 6.361.974. Los métodos para la mutación aleatoria de genes es bien conocida en la técnica, ver la Patente de los Estados Unidos No. 5.830.696. Por ejemplo, los mutágenos pueden usarse para mutar aleatoriamente un gen. Los mutágenos incluyen por ejemplo luz ultravioleta o irradiación gamma, o un mutágeno químico, tal como mitomicina, ácido nitroso, psoralenos fotoactivados, solos o en combinación, para inducir las roturas de ADN a que sean reparadas por recombinación. Otros mutágenos químicos incluyen, por ejemplo, bisulfito de sodio, ácido nitroso, hidroxilamina, hidrazina o ácido fórmico. Otros mutágenos son análogos de precursores de nucleótidos, tales como nitrosoguanidina, 5-bromouracilo, 2-aminopurina o acridina. Estos agentes pueden agregarse a una reacción PCR en lugar del precursor de nucleótido de ese modo mutando la secuencia. Los agentes de intercalado tales como proflavina, acriflavina, quinacrina y similares también pueden usarse.

Puede usarse cualquier técnica de biología molecular, tal como mutagénesis aleatoria por PCR, ver Rice (1992) Proc. Natl. Acad. Sci. Estados Unidos 89:5467-5471; o mutagénesis de cassette múltiple combinatorio, ver Crameri (1995) Biotechniques 18:194-196. Alternativamente, los ácidos nucleicos, tales como genes, pueden reunirse luego de la fragmentación aleatoria o "estocástica", ver la Patente de los Estados Unidos No. 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 otras realizaciones, las modificaciones, adiciones o eliminaciones se introducen por medio de PCR propensa a errores, transposición, mutagénesis de cassette, mutagénesis de ensamblaje recursivo, mutagénesis de ensamblaje exponencial, mutagénesis específica de sitio, reensamblaje de genes (tal como Reensamblaje de genes, ver Patente de los Estados Unidos No. 6.537.776), mutagénesis de saturación de sitios de genes (GSSM), reensamblaje de ligación sintética (SLR), recombinación, recombinación de secuencia recursiva, mutagénesis de ADN modificada con fosfotoato, mutagénesis de plantilla que contiene uracilo, mutagénesis dúplex con brechas, mutagénesis de reparación de coincidencias puntuales, mutagénesis de cepa huésped deficiente, mutagénesis química, mutagénesis radiogénica, mutagénesis de eliminación, mutagénesis de restricción-selección, mutagénesis de restricción purificación, síntesis de gen artificial, mutagénesis de ensamblaje, creación de multímeros de ácido nucleico quimérico, mutagénesis de saturación cromosómica (CSM) y/o combinación de estos y otros métodos.

Las siguientes publicaciones describen una variedad de procedimientos de recombinación recursiva y/o métodos que pueden incorporarse a los métodos de acuerdo con 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; Cramer (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature* 391:288-291; Cramer (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. Estados Unidos* 94:4504-4509; Patten et al. (1997) "Applications of DNA Shuffling to Pharmaceuticals and Vaccines" *Current Opinion in Biotechnology* 8:724-733; Cramer 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" en: *The Encyclopedia of Molecular Biology*. VCH Publishers, Nueva York. pp. 447-457; Cramer y 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; y Stemmer (1994) "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution." *Proc. Natl. Acad. Sci. Estados Unidos* 91:10747-10751.

Los 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; y Kunkel (1987) "The efficiency of oligonucleotide directed mutagenesis" en *Nucleic Acids & Molecular Biology* (Eckstein, F. y Lilley, D. M. J. eds., Springer Verlag, Berlín); mutagénesis que usa plantillas que contienen uracilo (Kunkel (1985) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Proc. Natl. Acad. Sci. Estados Unidos* 82:488-492; Kunkel et al. (1987) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Methods in Enzymol.* 154, 367-382; y Bass et al. (1988) "Mutant Trp repressors with new DNA-binding specificities" *Science* 242:240-245); mutagénesis dirigida a oligonucleótidos (*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; y Zoller (1987) "Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template" *Methods in Enzymol.* 154:329-350); mutagénesis de ADN modificada con fosforotioato (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; y 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); mutagénesis que usa ADN dúplex con brechas (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 usarse para poner en práctica la invención incluyen reparación de no coincidencias (Kramer (1984) "Point Mismatch Repair" *Cell* 38:879-887), mutagénesis que usa cepas huésped deficientes de reparación (Carter et al. (1985) "Improved oligonucleotide site-directed mutagenesis using M13 vectors" *Nucl. Acids Res.* 13: 4431-4443; y Carter (1987) "Improved oligonucleotide-directed mutagenesis using M13 vectors" *Methods in Enzymol.* 154: 382-403), mutagénesis de eliminación (Eghedarzadeh (1986) "Use of oligonucleotides to generate large deletions" *Nucl. Acids Res.* 14: 5115), restricción-selección y restricción-purificación (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), mutagénesis por síntesis de gen total (Nambiar et al. (1984) "Total synthesis and cloning of a gene coding for the ribonuclease S protein" *Science* 223: 1299-1301; Sakamar y Khorana (1988) "Total synthesis and expression of a gene for the  $\alpha$ -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; y Grundstrom et al. (1985) "Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis" *Nucl. Acids Res.* 13: 3305-3316), reparación de rotura de doble hebra (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. Estados Unidos*, 83:7177-

7181). Los detalles adicionales sobre cualquiera de los métodos anteriores pueden encontrarse en *Methods in Enzymology* Volumen 154, que también describe controles útiles para solucionar problemas con varios métodos de mutagénesis.

5 Los protocolos que pueden usarse para poner en práctica la invención se describen, por ejemplo en la Patente de los Estados Unidos No. 5.605.793 de Stemmer (25 feb. 1997), "Métodos para la recombinación in vitro"; Patente de los Estados Unidos No. 5.811.238 de Stemmer et al. (22 set. 1998) "Métodos para generar polinucleótidos que tienen características deseadas por selección y recombinación iterativas"; Patente de los Estados Unidos No. 5.830.721 de Stemmer et al. (3 nov. 1998), "Mutagénesis de ADN por fragmentación y reensamblaje aleatorios"; Patente de los Estados Unidos No. 5.834.252 de Stemmer, et al. (10 nov. 1998) "Reacción de polimerasa complementaria final"; Patente de los Estados Unidos No. 5.837.458 de Minshull, et al. (17 nov. 1998), "Métodos y composiciones para la manipulación celular y metabólica"; WO 95/22625, Stemmer y Cramer, "Mutagénesis por fragmentación y reensamblaje aleatorios"; WO 96/33207 de Stemmer y Lipschutz "Reacción en cadena de polimerasa complementaria final"; WO 97/20078 de Stemmer y Cramer "Métodos para generar polinucleótidos que tienen características deseadas por selección y recombinación iterativa"; WO 97/35966 de Minshull y Stemmer, 10 "Métodos y composiciones para la manipulación celular y metabólica"; WO 99/41402 de Punnonen et al. "Direccionamiento de vectores de vacuna genética"; WO 99/41383 de Punnonen et al. "Inmunización de biblioteca de antígenos"; WO 99/41369 de Punnonen et al. "Manipulación de vector de vacuna genética"; WO 99/41368 de Punnonen et al. "Optimización de propiedades inmunomoduladoras de vacunas genéticas"; EP 752008 de Stemmer y Cramer, "Mutagénesis de ADN por fragmentación y reensamblaje aleatorios"; EP 0932670 de Stemmer "Evolución de la captación de ADN celular por recombinación de secuencia recursiva"; WO 99/23107 de Stemmer et al., 15 "Modificación del tropismo de virus y rango huésped por transposición de genoma viral"; WO 99/21979 de Apt et al., "Vectores de virus del papiloma humano"; WO 98/31837 de del Cardayre et al. "Evolución de células y organismos enteros por recombinación de secuencia recursiva"; WO 98/27230 de Patten y Stemmer, "Métodos y composiciones para la manipulación de polipéptidos"; WO 98/27230 de Stemmer et al., "Métodos para la optimización de la terapia de gen por transposición y selección de secuencia recursiva," WO 00/00632, "Métodos para generar bibliotecas altamente diversas"; WO 00/09679, "Métodos para obtener bancos y secuencias resultantes de polinucleótidos recombinados in vitro"; WO 98/42832 de Arnold et al., "Recombinación de secuencias de polinucleótidos usando cebadores aleatorios o definidos"; WO 99/29902 de Arnold et al., "Método para crear secuencias de polinucleótidos y polipéptidos"; WO 98/41653 de Vind, "Un método in vitro para la construcción de una biblioteca de ADN"; WO 20 98/41622 por Borchert et al., "Método para construir una biblioteca usando transposición de ADN," y WO 98/42727 de Pati y Zarling, "Alteraciones de secuencia usando recombinación homóloga".

35 Protocolos que pueden usarse para poner en práctica la invención (proporcionando detalles con respecto a varios métodos de generación de diversidad) se describen, por ejemplo, en la solicitud de Patente de los Estados Unidos No. de serie 09/407.800, "TRANSPOSICIÓN DE GENES ALTERADOS POR CODONES" de Patten et al. presentada el 28 de set. de 1999; "EVOLUCIÓN DE CÉLULAS Y ORGANISMOS ENTEROS POR RECOMBINACIÓN DE SECUENCIA RECURSIVA" de del Cardayre et al., Patente de los Estados Unidos No. 6.379.964; "RECOMBINACIÓN DE ÁCIDO NUCLEICO MEDIADO POR OLIGONUCLEÓTIDOS" de Cramer et al., Patente de los Estados Unidos No. 6.319.714; 6.368.861; 6.376.246; 6.423.542; 6.426.224 y PCT/US00/01203; "USO DE SÍNTESIS DE OLIGONUCLEÓTIDOS VARIADA DE CODONES PARA LA TRANSPOSICIÓN SINTÉTICA" de Welch et al., Patente de los Estados Unidos No. 6.436.675; "MÉTODOS PARA REALIZAR CADENAS DE CARACTERES, POLINUCLEÓTIDOS Y POLIPÉPTIDOS CON CARACTERÍSTICAS DESEADAS" de Selifonov et al., presentada el 18 de ene. de 2000, (PCT/US00/01202) y, por ejemplo, "MÉTODOS PARA REALIZAR CADENAS DE CARACTERES, POLINUCLEÓTIDOS Y POLIPÉPTIDOS CON CARACTERÍSTICAS DESEADAS" de Selifonov et al., presentada en 18 de jul. de 2000 (No. de serie de Estados Unidos 09/618.579); "MÉTODOS PARA POBLAR ESTRUCTURAS DE DATOS PARA EL USO EN SIMULACIONES EVOLUTIVAS" de Selifonov y Stemmer, presentada el 18 de ene. de 2000 (PCT/US00/01138); y "RECOMBINACIÓN MEDIADA POR PLANTILLA DE ÁCIDO NUCLEICO DE UNA SOLA HEBRA Y AISLAMIENTO DE FRAGMENTOS DE ÁCIDO NUCLEICO" de Affholter, presentada el 6 de set. de 2000 (No. de serie de Estados Unidos 09/656.549); y Patentes de los Estados Unidos No. 6.177.263; 6.153.410.

50 Los métodos no estocásticos o de "evolución dirigida" incluyen, por ejemplo, mutagénesis de saturación, tal como mutagénesis de saturación de sitios de genes (GSSM), reensamblaje de ligación sintética (SLR) o una combinación de los mismos se usan para modificar los ácidos nucleicos de acuerdo con la invención para generar enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa con propiedades nuevas o alteradas (tal como actividad bajo condiciones altamente ácidas o alcalinas, temperaturas altas o bajas y similares). Los polipéptidos 55 codificados por los ácidos nucleicos modificados pueden someterse a prueba para detectar una actividad antes de evaluarse para la formación de escisión u otra actividad. Puede usarse cualquier modalidad de evaluación o protocolo tal como usar una plataforma de arreglo de capilares. Ver la Patente de los Estados Unidos No. 6.361.974; 6.280.926; 5.939.250.

Mutagénesis de saturación de sitios de genes o GSSM

La invención también proporciona métodos para realizar enzima usando mutagénesis de saturación de sitios de genes o GSSM, como se describe en la presente, y también en la Patente de los Estados Unidos No. 6.171.820 y 6.579.258. En algunas realizaciones, los cebadores de codón que contienen una secuencia N,N,G/T degenerada se usan para introducir mutaciones puntuales en un polinucleótido tal como una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa o un anticuerpo de acuerdo con la invención, de modo de generar un juego de polipéptidos de progenie en los cuales se representa un rango completo de sustituciones de aminoácidos únicos en cada posición de aminoácidos, tal como un residuo de aminoácidos en un sitio activo de enzima o sitio de unión de ligando dirigido para modificarse. Estos oligonucleótidos pueden comprender una primera secuencia homóloga contigua, una secuencia N,N,G/T degenerada y opcionalmente una segunda secuencia homóloga. Los productos translacionales de progenie de corriente abajo del uso de dichos oligonucleótidos incluyen todos los cambios de aminoácidos posibles en cada sitio de aminoácidos a lo largo del polipéptido porque la degeneración de la secuencia N,N,G/T incluye codones para todos los 20 aminoácidos. En algunas realizaciones, un oligonucleótido degenerado (compuesto por un cassette N,N,G/T degenerado) se usa para someter cada codón original en una plantilla de polinucleótido parental a un rango completo de sustituciones de codón. En otras realizaciones, se usan al menos dos cassettes degenerados –ya sea en el mismo oligonucleótido o no, para someter al menos dos codones originales en una plantilla de polinucleótidos parental a un rango completo de sustituciones de codón. Por ejemplo, más de una secuencia N,N,G/T puede estar contenida en un oligonucleótido para introducir mutaciones de 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 otras realizaciones, los oligonucleótidos útiles para introducir adiciones y eliminaciones pueden usarse solos o en combinación con los codones que contienen una secuencia N,N,G/T para introducir cualquier combinación o permutación de adiciones, eliminaciones y/o sustituciones de aminoácidos.

En algunas realizaciones, la mutagénesis simultánea de dos o más posiciones de aminoácidos contiguas se realiza usando un oligonucleótido que contiene tripletes contiguos N,N,G/T, es decir una secuencia (N,N,G/T)<sub>n</sub> degenerada. En otras realizaciones, se usan cassettes degenerados que tienen menos degeneración que la secuencia N,N,G/T. Por ejemplo, puede desearse en algunos casos que se use (tal como en un oligonucleótido) una secuencia triplete degenerada compuesta de sólo una N, en donde dicha N puede estar en la primera, segunda o tercera posición del triplete. Cualquier otra base que incluye cualquier combinación y permutación de la misma puede usarse en las dos posiciones restantes del triplete. Alternativamente, puede desearse en algunos casos usar (tal como en un oligo) una secuencia de triplete N,N,N degenerada.

En algunas realizaciones, el uso de tripletes degenerados (tales como tripletes N,N,G/T) permite una generación sistemática y fácil de un rango completo de aminoácidos naturales (para un total de 20 aminoácidos) en todas y cada una de las posiciones de aminoácidos en un polipéptido (en otras realizaciones, los métodos también incluyen la generación de menos de todas las sustituciones posibles por residuo de aminoácido, o codón, posición). Por ejemplo, para un polipéptido de aminoácido 100, se pueden generar 2000 especies distintas (es decir 20 aminoácidos posibles por posición X 100 posiciones de aminoácidos). A través del uso de un oligonucleótido o juego de oligonucleótidos que contienen un triplete N,N,G/T degenerado, 32 secuencias individuales pueden codificar 20 posibles aminoácidos naturales. Por lo tanto, en un recipiente de reacción en el cual una secuencia de polinucleótido parental está sometida a mutagénesis de saturación usando al menos uno de dicho oligonucleótido, se generan 32 polinucleótidos de progenie distintos que codifican 20 polipéptidos distintos. Por contraste, el uso de un oligonucleótido no degenerado en mutagénesis dirigida al sitio conduce sólo a un producto de polipéptido de progenie por recipiente de reacción. Los oligonucleótidos no degenerados pueden usarse opcionalmente en combinación con cebadores degenerados divulgados; por ejemplo, los oligonucleótidos no degenerados pueden usarse para generar mutaciones puntuales específicas en un polinucleótido de trabajo. Esto proporciona un medio para generar mutaciones puntuales silenciosas específicas, mutaciones puntuales que conducen a cambios de aminoácidos correspondientes y mutaciones puntuales que causan la generación de codones de detención y la expresión correspondiente de fragmentos de polipéptidos.

En algunas realizaciones, cada recipiente de reacción de mutagénesis de saturación contiene polinucleótidos que codifican al menos 20 moléculas de polipéptidos de progenie (tales como una enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa) de tal modo que todos los 20 aminoácidos naturales se representan en una posición de aminoácidos específica que corresponde a la posición de codones mutagenizado en el polinucleótido parental (otras realizaciones usan menos de 20 combinaciones naturales). Los polipéptidos de progenie degenerados 32 veces generados a partir de cada recipiente de reacción de mutagénesis de saturación pueden estar sujetos a una amplificación clonal (tal como clonados en un huésped adecuado, tal como huésped *E. Coli* usando por ejemplo un vector de expresión) y someterse a una detección de expresión. Cuando un polipéptido de progenie individual se identifica mediante la detección para mostrar un cambio favorable en la propiedad (cuando se lo compara con el polipéptido parental, tal como formación carbono-carbono aumentada o actividad de escisión bajo condiciones alcalinas o ácidas), puede secuenciarse para identificar la sustitución de aminoácidos favorable contenida en él.

- En algunas realizaciones, tras mutagenizar todas y cada una de las posiciones de aminoácidos en un polipéptido parental usando mutagénesis de saturación como se divulga en la presente, pueden identificarse cambios de aminoácidos favorables en más de una posición de aminoácidos. Una o más nuevas moléculas de progenie pueden generarse que contienen una combinación de todas o partes de estas sustituciones de aminoácidos favorables. Por ejemplo, si 2 cambios de aminoácidos favorables específicos se identifican en cada una de las 3 posiciones de aminoácidos en un polipéptido, las permutaciones incluyen 3 posibilidades en cada posición (sin cambio del aminoácido original, y cada uno de dos cambios favorables) y 3 posiciones. Por lo tanto, hay  $3 \times 3 \times 3$  o 27 posibilidades totales, incluyendo 7 que se examinaron previamente – 6 mutaciones puntuales únicas (es decir 2 en cada una de las tres posiciones) y sin cambios en cualquier posición.
- Aun en otra realización, la mutagénesis de saturación del sitio puede usarse junto con la transposición, quimerización, recombinación y otros procesos de mutagénesis, junto con la detección. Esta invención proporciona el uso de cualquier proceso de mutagénesis, incluyendo mutagénesis de saturación, de un modo iterativo. En un ejemplo, el uso iterativo de cualquier proceso de mutagénesis se usa en combinación con la detección.
- La invención también proporciona el uso de cebadores de codón propios (que contienen una secuencia N,N,N degenerada) para introducir mutaciones puntuales en un polinucleótido, de modo de generar un juego de polipéptidos de progenie en los cuales se representa un rango completo de sustituciones de aminoácidos únicas en cada posición de aminoácidos (mutagénesis de saturación de sitios de genes (GSSM)). Los oligos usados están compuestos contiguamente por una primera secuencia homóloga, una secuencia N,N,N degenerada y en algunas realizaciones pero no necesariamente una segunda secuencia homóloga. Los productos translacionales de progenie de corriente abajo del uso de dichos oligonucleótidos incluyen todos los cambios de aminoácidos posibles en cada sitio de aminoácidos a lo largo del polipéptido porque la degeneración de la secuencia N,N,N incluye codones para los 20 aminoácidos.
- En algunas realizaciones, un oligo degenerado (compuesto por un cassette N,N,N degenerado) se usa para someter cada codón original en una plantilla de polinucleótido parental a un rango completo de sustituciones de codón. En otras realizaciones, se usan al menos dos cassettes N,N,N degenerados –ya sea en el mismo oligonucleótido o no, para someter al menos dos codones originales en una plantilla de polinucleótidos parental a un rango completo de sustituciones de codón. Por lo tanto, más de una secuencia N,N,N puede estar contenida en un oligo para introducir mutaciones de aminoácidos en más de un sitio. Esta pluralidad de secuencias N,N,N puede ser directamente contigua o separada por una o más secuencias de nucleótidos adicionales. Esta pluralidad de secuencias N,N,N puede ser directamente contigua o estar separada por una o más secuencias de nucleótidos adicionales. En otras realizaciones, los oligos útiles para introducir adiciones y eliminaciones pueden usarse 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 algunas realizaciones es posible mutagenizar simultáneamente dos o más posiciones de aminoácidos contiguas usando un oligonucleótido que contiene tripletes contiguos N,N,N, es decir una secuencia (N,N,N)<sub>n</sub> degenerada. En otras realizaciones, la presente invención proporciona el uso de cassettes degenerados que tienen menos degeneración que la secuencia N,N,N. Por ejemplo, puede desearse en algunos casos que se use (tal como en un oligo) una secuencia triplete degenerada compuesta de sólo una N, en donde dicha N puede estar en la primera, segunda o tercera posición del triplete. Cualquier otra base que incluye cualquier combinación y permutación de la misma puede usarse en las dos posiciones restantes del triplete. Alternativamente, puede desearse en algunos casos usar (tal como en un oligo) una secuencia de triplete N,N,N degenerada, una secuencia de triplete N,N,G/T o N,N,G/C.
- En algunas realizaciones, el uso de un triplete degenerado (tal como una secuencia de triplete N,N,G/T o N,N,G/C) es ventajosa por varias razones. En algunas realizaciones, la presente invención proporciona medios para generar sistemáticamente y relativamente fácilmente la sustitución del rango completo de aminoácidos posibles (para un total de 20 aminoácidos) en todas y cada una de las posiciones de aminoácidos en un polipéptido. Por lo tanto, para un polipéptido de aminoácido 100, la invención proporciona medios para generar sistemáticamente y relativamente fácilmente 2000 especies distintas (es decir, 20 aminoácidos posibles por 100 posiciones de aminoácidos). Se aprecia que se proporciona, a través del uso de un oligo que contiene una secuencia de triplete N,N,G/T o N,N,G/C degenerada, 32 secuencias individuales que codifican 20 aminoácidos posibles. Por lo tanto, en un recipiente de reacción en el cual una secuencia de polinucleótido parental está sometida a mutagénesis de saturación usando al menos uno de dicho oligo, se generan 32 polinucleótidos de progenie distintos que codifican 20 polipéptidos distintos. Por contraste, el uso de un oligo no degenerado en mutagénesis dirigida al sitio conduce sólo a un producto de polipéptido de progenie por recipiente de reacción.
- La invención también proporciona el uso de oligos no degenerados, que pueden opcionalmente usarse en combinación con cebadores degenerados divulgados. Se aprecia que en algunas situaciones, es ventajoso usar oligos no degenerados para generar mutaciones puntuales específicas en un polinucleótido de trabajo. Esto propociona un medio para generar mutaciones puntuales silenciosas específicas, mutaciones puntuales que conducen a cambios de aminoácidos correspondientes y mutaciones puntuales que causan la generación de codones de detención y la expresión correspondiente de fragmentos de polipéptidos.

- Por lo tanto, en algunas realizaciones de la presente invención, cada recipiente de reacción de mutagénesis de saturación contiene polinucleótidos que codifican al menos 20 moléculas de polipéptido de progenie de modo que todos los 20 aminoácidos se representen en una posición de aminoácidos específica que corresponde a la posición de codón mutagenizado en el polinucleótido parental. Los polipéptidos de progenie degenerados 32 veces generados a partir de cada recipiente de reacción de mutagénesis de saturación pueden estar sujetos a una amplificación clonal (tal como clonados en un huésped adecuado de *E. Coli* usando un vector de expresión) y someterse a una detección de expresión. Cuando un polipéptido de progenie individual se identifica detectando para mostrar un cambio favorable en la propiedad (cuando se compara con el polipéptido parental) puede secuenciarse para identificar la sustitución de aminoácido favorable correspondiente contenida en el mismo.
- En algunas realizaciones, tras mutagenizar todas y cada una de las posiciones de aminoácidos en un polipéptido parental usando mutagénesis de saturación como se divulga en la presente, pueden identificarse cambios de aminoácidos favorables en más de una posición de aminoácidos. Una o más nuevas moléculas de progenie pueden generarse que contienen una combinación de todas o partes de estas sustituciones de aminoácidos favorables. Por ejemplo, si 2 cambios de aminoácidos favorables específicos se identifican en cada una de las 3 posiciones de aminoácidos en un polipéptido, las permutaciones incluyen 3 posibilidades en cada posición (sin cambio del aminoácido original, y cada uno de dos cambios favorables) y 3 posiciones. Por lo tanto, hay  $3 \times 3 \times 3$  o 27 posibilidades totales, incluyendo 7 que se examinaron previamente – 6 mutaciones puntuales únicas (es decir 2 en cada una de las tres posiciones) y sin cambios en cualquier posición.
- La invención proporciona el uso de mutagénesis de saturación en combinación con procesos de mutagenización adicionales, tales como un proceso en donde dos o más polinucleótidos relacionados se introducen en una célula huésped adecuada de tal modo que se genera un polinucleótido híbrido mediante recombinación y reordenamiento reductor.
- Además de realizar mutagénesis a lo largo de la secuencia entera de un gen, la presente invención proporciona que la mutagénesis puede usarse para reemplazar cada uno de cualquier número de bases en una secuencia de polinucleótido, en donde el número de bases a ser mutagenizado es en algunas realizaciones cualquier entero de 15 a 100000. Por lo tanto, en vez de mutagenizar cada posición a lo largo de una molécula, se puede someter cada o un número discreto de bases (en algunas realizaciones un subconjunto que tiene en su totalidad de 15 a 100000) para la mutagénesis. En algunas realizaciones, se usa un nucleótido separado para mutagenizar cada posición o grupo de posiciones a lo largo de una secuencia de polinucleótidos. Un grupo de 3 posiciones a ser mutagenizadas puede ser un codón. Las mutaciones pueden introducirse usando un cebador mutagénico, que contiene un cassette heterólogo, también denominado cassette mutagénico. Ejemplos de cassettes pueden tener de 1 a 500 bases. Cada posición de nucleótidos en dichos cassettes heterólogos son 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 (E puede referirse como un oligo de diseñador).
- En algunas realizaciones, la mutagénesis de saturación está compuesta por la mutagénesis de un juego completo de cassettes mutagénicos (en donde cada cassette es, en algunas realizaciones, de aproximadamente 1-500 bases de longitud) en secuencia de polinucleótido definida para ser mutagenizados (en donde la secuencia a ser mutagenizada es, en algunas realizaciones, de aproximadamente 15 a 100000 bases de longitud). Por lo tanto, un grupo de mutaciones (que están en el rango de 1 a 100 mutaciones) se introduce en cada cassette para mutagenizarse. Un agrupamiento de mutaciones a ser introducidas en un cassette puede ser diferente o el mismo que un segundo agrupamiento de mutaciones a introducirse en un segundo cassette durante la aplicación de una ronda de mutagénesis de saturación. Dichos agrupamientos se ejemplifican mediante eliminaciones, adiciones, agrupamientos de codones particulares y agrupamientos de cassettes de nucleótidos particulares.
- En algunas realizaciones, las secuencias definidas a mutagenizarse incluyen un gen entero, vía, ADNc, un marco de lectura abierto (ORF) y un promotor entero, potenciador, represor/transactivador, origen de replicación, intrón, operador o cualquier grupo funcional de polinucleótido. Generalmente, una "secuencia definida" para este objetivo puede ser cualquier polinucleótido que tiene una secuencia de polinucleótidos de 15 bases y secuencias de polinucleótidos de longitudes de entre 15 bases y 15000 bases (esta invención específicamente nombra cada entero entre medio). Las consideraciones para elegir grupos de codones incluyen tipos de aminoácidos codificados por un cassette mutagénico degenerado.
- En algunas realizaciones, un agrupamiento de mutaciones que pueden introducirse en un cassette mutagénico, esta invención proporciona específicamente sustituciones de codones degenerados (usando oligos degenerados) que codifican 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 de ese modo.
- Reensamblaje de ligación sintética (SLR)**
- La invención proporciona un sistema de modificación génica no estocástico denominado "reensamblaje de ligación sintética" o simplemente "SLR", un "proceso de evolución dirigida" para generar polipéptidos, tales como enzimas o

anticuerpos de aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención, con propiedades nuevas o alteradas.

5 El SLR es un método de ligar fragmentos de oligonucleótidos no estocásticamente. Este método difiere de la transposición de oligonucleótidos estocástica en que los bloques de construcción de ácido nucleico no se transponen, concatenan o quimerizan aleatoriamente, sino que se ensamblan no estocásticamente. Ver la Patente de los Estados Unidos No. 6.773.900; 6.740.506; 6.713.282; 6.635.449; 6.605.449; 6.537.776. En algunas realizaciones, el SLR comprende las siguientes etapas: (a) proporcionar un polinucleótido de plantilla, en donde el polinucleótido de plantilla comprende secuencia que codifica un gen homólogo; (b) proporcionar una pluralidad de polinucleótidos de bloque de construcción, en donde los polinucleótidos de bloque de construcción se diseñan para reensamblarse en cruce con el polinucleótido de plantilla en una secuencia predeterminada, y un 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 modo que el cruce de polinucleótido de bloque de construcción se reensamble con el polinucleótido de plantilla para generar polinucleótidos que comprenden variaciones de secuencia de gen homólogo.

10 El SLR no depende de la presencia de altos niveles de homología entre polinucleótidos para rediseñarse. Por lo tanto, este método puede usarse para generar bibliotecas (o conjuntos) no estocásticamente de moléculas de progenie compuestas por más de  $10^{100}$  quimeras diferentes. El SLR puede usarse para generar bibliotecas compuestas por más de  $10^{1000}$  quimeras diferentes. Por lo tanto, realizaciones de la presente invención incluyen métodos no estocásticos para producir un juego de moléculas de ácido nucleico quimérico finalizado con un orden de ensamblaje general que se elige por diseño. Este método incluye las etapas de generar por diseño una pluralidad de bloques de construcción de ácido nucleico específicos que tienen extremos ligables compatibles mutuamente y ensamblar estos bloques de construcción de ácido nucleico de tal modo que se logre un orden de ensamblaje general diseñado.

25 Los extremos ligables compatibles mutuamente de los bloques de construcción de ácido nucleico a ensamblarse se consideran como "útiles" para este tipo de ensamblaje ordenado si permiten que los bloques de construcción se acoplen en órdenes predeterminados. Por lo tanto, el orden de ensamblaje general en el cual los bloques de construcción de ácido nucleico pueden acoplarse se especifica por el diseño de los extremos ligables. Si se usa más de una etapa de ensamblaje, entonces el orden de ensamblaje general en el cual los bloques de construcción de ácido nucleico pueden acoplarse también se especifica por el orden secuencial de las etapas de ensamblaje. En algunas realizaciones, las piezas de construcción apareadas se tratan con una enzima, tal como una ligasa (tal como ligasa ADN T4) para lograr la unión covalente de las piezas de construcción.

30 Los extremos ligables compatibles mutuamente de los bloques de construcción de ácido nucleico a ensamblarse se consideran como "útiles" para este tipo de ensamblaje ordenado si permiten que los bloques de construcción se acoplen en órdenes predeterminados. Estas plantillas de oligonucleótido parental sirven por lo tanto como una fuente de información de secuencia que ayuda en el diseño de los bloques de construcción de ácido nucleico a mutagenizarse, tal como quimerizarse o transponerse. En algunas realizaciones de este método, las secuencias de una pluralidad de plantillas de ácido nucleico parental se alinean de modo de seleccionar uno o más puntos de demarcación. Los puntos de demarcación pueden estar ubicados en un área de homología y están compuestos por uno o más nucleótidos. Estos puntos de demarcación en algunas realizaciones son compartidos por al menos dos de las plantillas progenitoras. Los puntos de demarcación pueden usarse de ese modo para delinear los límites de los bloques de construcción de oligonucleótidos para generarse de modo de rediseñar los polinucleótidos parentales. Los puntos de demarcación identificados y seleccionados en las moléculas progenitoras sirven como puntos de quimerización potencial en el ensamblaje de las moléculas de progenie quiméricas finales. Un punto de demarcación puede ser un área de homología (compuesta de al menos una base de nucleótido homóloga) compartida por al menos dos secuencias de polinucleótido parentales. Alternativamente, un punto de demarcación puede ser un área de homología que está compartida por al menos la mitad de las secuencias de polinucleótidos parentales, o puede ser un área de homología que está compartida por al menos dos tercios de las secuencias de polinucleótido parental. Aun más, en algunas realizaciones, los puntos de demarcación útiles son un área de homología que está compartida por al menos tres cuartos de las secuencias de polinucleótidos parentales o puede estar compartida por casi todas las secuencias de polinucleótidos parentales. En algunas realizaciones, un punto de demarcación es un área de homología que está compartida por todas las secuencias de polinucleótidos parentales.

35 40 45 50 55 60 En algunas realizaciones, se realiza un proceso de reensamblaje de ligación exhaustivamente para generar una biblioteca exhaustiva de polinucleótidos quiméricos de progenie. En otras palabras, todas las combinaciones ordenadas posibles de los bloques de construcción de ácido nucleico se representan en el juego de moléculas de ácido nucleico quimérico finalizado. Al mismo tiempo, en otras realizaciones, el orden de ensamblaje (es decir el orden de ensamblaje de cada bloque de construcción en la secuencia 5' o 3' de cada ácido nucleico quimerizado finalizado) en cada combinación es por diseño (o no estocástico) como se describió anteriormente. Dada la naturaleza no estocástica de la presente invención, la posibilidad de productos derivados indeseados se reduce en gran medida.



En otras realizaciones, el método de reensamblaje de ligación se realiza sistemáticamente. Por ejemplo, el método se realiza de modo de generar una biblioteca compartimentada sistemáticamente de moléculas de progenie, con compartimentos que pueden analizarse sistemáticamente, tal como uno por uno. En otras palabras la presente invención indica que a través del uso selectivo y con criterio de bloques de construcción de ácido nucleico específicos, acoplados con el uso selectivo y con criterio de reacciones de ensamblaje escalonadas secuencialmente, puede lograrse un diseño donde se hagan juegos específicos de productos de progenie en cada uno de los varios recipientes de reacción. Esto permite que se realice un examen sistemático y un procedimiento de detección. Por lo tanto, estos métodos permiten que se examine sistemáticamente un gran número posible de moléculas de progenie en grupos más pequeños. Dada su capacidad de realizar quimerizaciones de un modo que es altamente flexible pero exhaustivo y también sistemático, en particular cuando hay un bajo nivel de homología entre las moléculas progenitoras, estos métodos proporcionan la generación de una biblioteca (o juego) compuesto de un gran número de moléculas de progenie. Dada la naturaleza no estocástica de la presente invención de reensamblaje de ligación, las moléculas de progenie generadas en algunas realizaciones comprenden una biblioteca de moléculas de ácido nucleico quimerizado finalizadas que tienen un orden de ensamblaje general que se elige por diseño. Los métodos de mutagénesis de saturación y evolución dirigida optimizada también pueden usarse para generar diferentes especies moleculares de progenie. Se aprecia que la invención proporciona libertad de elección y control con respecto a la selección de los puntos de demarcación, el tamaño y el número de los bloques de construcción de ácido nucleico y el tamaño y el diseño de los acoples. Asimismo se aprecia que el requisito para la homología intermolecular es altamente relajado para la operabilidad de la presente invención. De hecho, los puntos de demarcación pueden incluso elegirse en áreas de pequeña homología o no homología intermolecular. Por ejemplo, debido a un abultamiento de codón, es decir la degeneración de codones, las sustituciones de nucleótidos pueden introducirse en bloques de construcción de ácido nucleico sin alterar el aminoácido codificado originalmente en la plantilla progenitora correspondiente. Alternativamente, un codón puede alterarse de tal modo que se altere la codificación para un aminoácido original. La presente invención indica que dichas sustituciones puedan introducirse en el bloque de construcción de ácido nucleico para mejorar la incidencia de puntos de demarcación homólogos intermoleculares y por lo tanto permitir que se logre un número aumentado de acoples entre los bloques de construcción, que a su vez permite que se genere un número mayor de moléculas quiméricas de progenie.

#### Reensamblaje de gen sintético

En algunas realizaciones, la presente invención proporciona un método no estocástico denominado reensamblaje de gen sintético, que de algún modo se relaciona con la transposición estocástica, salvo que los bloques de construcción de ácido nucleico no se transponen o concatenan o quimerizan aleatoriamente, sino que se ensamblan no estocásticamente. Ver la Patente de los Estados Unidos No. 6.537.776.

El método de reensamblaje de gen sintético no depende de la presencia de un alto nivel de homología entre polinucleótidos para transponerse. La invención puede usarse para generar bibliotecas (o conjuntos) no estocásticamente de moléculas de progenie compuestas de más de  $10^{100}$  quimeras diferentes. Posiblemente, el reensamblaje de gen sintético puede usarse aun para generar bibliotecas compuestas por más de  $10^{1000}$  quimeras de progenie diferentes.

Por lo tanto, en algunas realizaciones, la invención proporciona un método no estocástico de producir un juego de moléculas de ácido nucleico quimérico finalizado que tiene un orden de ensamblaje general que se elige por diseño, cuyo método está compuesto por las etapas de generar por diseño una pluralidad de bloques de construcción de ácido nucleico específicos que tienen extremos ligables compatibles mutuamente útiles y ensamblar estos bloques de construcción de ácido nucleico de tal modo que se logre un orden de ensamblaje general diseñado.

Los extremos ligables compatibles mutuamente de los bloques de construcción de ácido nucleico a ensamblarse se consideran como "útiles" para este tipo de ensamblaje ordenado si permiten que los bloques de construcción se acoplen en órdenes predeterminados. Por lo tanto, en algunas realizaciones, el orden de ensamblaje general en el cual los bloques de construcción de ácido nucleico pueden acoplarse está especificado por el diseño de los extremos ligables y, si se usa más de una etapa de ensamblaje, entonces el orden de ensamblaje general en el cual los bloques de construcción de ácido nucleico pueden acoplarse también está especificado por el orden secuencial de las etapas de ensamblaje. En una realización, las piezas de construcción apareadas se tratan con una enzima, tal como una ligasa (tal como ligasa ADN T4) para lograr la unión covalente de las piezas de construcción.

En otra realización, el diseño de bloques de construcción de ácido nucleico se obtiene tras el análisis de las secuencias de un juego de plantillas de ácido nucleico progenitor que sirve como base para producir un juego de progenie de moléculas de ácido nucleico quimérico finalizado. Estas plantillas de ácido nucleico progenitor sirven por lo tanto como una fuente de información de secuencia que ayuda en el diseño de los bloques de construcción de ácido nucleico a mutagenizarse, es decir como quimerizarse o transponerse.

En un ejemplo, la invención proporciona la quimerización de una familia de genes relacionados y su familia codificada de productos relacionados. En un ejemplo particular, los productos codificados son enzimas. Las enzimas aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de la presente invención pueden mutagenizarse de acuerdo con los métodos descritos en la presente.

Por lo tanto, de acuerdo con una realización de la invención, las secuencias de una pluralidad de plantillas de ácido nucleico progenitoras (tales como polinucleótidos de acuerdo con la invención) se alinean en orden para seleccionar uno o más puntos de demarcación, los cuales pueden ubicarse en un área de homología. Los puntos de demarcación pueden usarse para delinear los límites de los bloques de construcción de ácido nucleico para que se generen. Por lo tanto, los puntos de demarcación identificados y seleccionados en las moléculas progenitoras sirven como puntos de quimerización potenciales en el ensamblaje de las moléculas de proge-

En algunas realizaciones, un punto de demarcación útil es un área de homología (compuesta de 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 está 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 en algunas realizaciones casi todas las plantillas progenitoras. Aun más, en algunas realizaciones aun un punto de demarcación útil es un área de homología que está compartida por todas las plantillas progenitoras.

En una realización, el proceso de reensamblaje de gen se realiza exhaustivamente de modo de generar una biblioteca exhaustiva. En otras palabras, todas las combinaciones ordenadas posibles de los bloques de construcción de ácido nucleico se representan en el juego de moléculas de ácido nucleico quimérico finalizado. Al mismo tiempo, el orden de ensamblaje (es decir el orden de ensamblaje de cada bloque de construcción en la secuencia 5' o 3 de cada ácido nucleico quimerizado finalizado) en cada combinación es por diseño (o no estocástico). Dada la naturaleza no estocástica de este método, la posibilidad de productos derivados indeseados se reduce en gran medida.

En otras realizaciones, el método indica que el proceso de reensamblaje de gen se realiza sistemáticamente, por ejemplo para generar una biblioteca compartimentada sistemáticamente, con compartimentos que pueden analizarse sistemáticamente, tal como uno por uno. En otras palabras la invención indica que a través del uso selectivo y con criterio de bloques de construcción de ácido nucleico específicos, acoplados con el uso selectivo y con criterio de reacciones de ensamblaje escalonadas secuencialmente, puede lograrse un diseño donde se hagan juegos específicos de productos de proge- en cada uno de los varios recipientes de reacción. Esto permite que se realice un examen sistemático y un procedimiento de detección. Por lo tanto, permite que se examine sistemáticamente un gran número posible de moléculas de proge- en grupos más pequeños.

Dada su capacidad de realizar quimerizaciones de un modo que es altamente flexible pero exhaustivo y también sistemático, en particular cuando hay un bajo nivel de homología entre las moléculas progenitoras, la presente invención proporciona la generación de una biblioteca (o juego) compuesto de un gran número de moléculas de proge- Dada la naturaleza no estocástica de la presente invención de reensamblaje de genes, las moléculas de proge- generadas en algunas realizaciones comprenden una biblioteca de moléculas de ácido nucleico quimerizado finalizadas que tienen un orden de ensamblaje general que se elige por diseño. En algunas realizaciones, dicha biblioteca generada está compuesta por más de  $10^3$  a más de  $10^{1000}$  especies moleculares de proge- diferentes.

En algunas realizaciones, un juego de moléculas de ácido nucleico quimérico finalizadas producidas como se describe está compuesto de un polinucleótido que codifica un polipéptido. En una realización, este polinucleótido es un gen, que puede ser un gen fabricado por el hombre. En otra realización, este polinucleótido es una vía de gen, que puede ser una vía de gen fabricada por el hombre. En algunas realizaciones, la invención indica que uno o más genes fabricados por el hombre generados por la invención pueden incorporarse en una vía de gen fabricada por el hombre, tal como una vía operable en un organismo eucariota (incluyendo una planta).

En otra ejemplificación, la naturaleza sintética de esta etapa en la cual los bloques de construcción se generan permite el diseño y la introducción de nucleótidos (tales como uno o más nucleótidos, que pueden ser por ejemplo codones o intrones o secuencias reguladoras) que pueden eliminarse posteriormente en un proceso in vitro (tal como mutagénesis) o un proceso in vivo (tal como utilizando la capacidad de empalme de gen de un organismo huésped). Se aprecia que en muchos casos la introducción de estos nucleótidos también puede ser deseable por muchas otras razones además del beneficio posible de crear un punto de demarcación útil.

Por lo tanto, en algunas realizaciones, la invención indica que un bloque de construcción de ácido nucleico puede usarse para introducir un intrón. Por lo tanto, la invención indica que los intrones funcionales pueden introducirse en un gen fabricado por el hombre de acuerdo con la invención. En algunas realizaciones, la invención también indica que los intrones funcionales pueden introducirse en una vía de gen fabricada por el hombre de acuerdo con la invención. Por consiguiente, la invención proporciona la generación de un polinucleótido quimérico que es un gen fabricado por el hombre que contiene uno (o más) intrones introducidos artificialmente.

La invención también proporciona la generación de un polinucleótido quimérico que es una vía de gen fabricada por el hombre que contiene uno (o más) intrones introducidos artificialmente. En algunas realizaciones, los intrones introducidos artificialmente son funcionales en una o más células huésped para empalmar genes del modo que los intrones naturales sirven funcionalmente en el empalme de genes. En algunas realizaciones, la invención

proporciona procesos para producir polinucleótidos que contienen intrones fabricados por el hombre para introducirse en organismos huésped para la recombinación y/o empalme.

5 Un gen fabricado por el hombre producido usando la invención también puede servir como un sustrato para la recombinación con otro ácido nucleico. De una forma similar, una vía de gen fabricada por el hombre producida usando la invención también puede servir como un sustrato para la recombinación con otro ácido nucleico. En algunas realizaciones, la recombinación se facilita por u ocurre en áreas de homología entre el gen fabricado por el hombre que contiene intrones y un ácido nucleico, que sirve como un socio de recombinación. En algunas realizaciones, el socio de recombinación también puede ser un ácido nucleico generado por la invención, incluyendo un gen fabricado por el hombre o una vía de gen fabricada por el hombre. La recombinación puede estar facilitada por o puede ocurrir en áreas de homología que existen en uno (o más) intrones introducidos artificialmente en el gen fabricado por el hombre.

10 En algunas realizaciones, el método de reensamblaje de gen sintético de acuerdo con la invención utiliza una pluralidad de bloques de construcción de ácido nucleico, cada uno de los cuales, en algunas realizaciones, tiene dos extremos ligables. Los dos extremos ligables en cada bloque de construcción de ácido nucleico puede ser de dos extremos romos (es decir cada uno teniendo una saliente de cero nucleótidos) o en algunas realizaciones un extremo romo y una saliente, o más en algunas realizaciones aún dos salientes. En algunas realizaciones, una saliente útil para este objetivo puede ser una saliente 3' o una saliente 5'. Por lo tanto, un bloque de construcción de ácido nucleico puede tener una saliente 3' o alternativamente una saliente 5' o alternativamente dos salientes 3' o alternativamente dos salientes 5'. El orden general en el cual los bloques de construcción de ácido nucleico se ensamblan para formar una molécula de ácido nucleico quimérico finalizado se determina mediante el diseño experimental por diseño y no es aleatorio.

15 En algunas realizaciones, un bloque de construcción de ácido nucleico se genera mediante la síntesis química de dos ácidos nucleicos de una sola hebra (también denominados oligos de una sola hebra) y ponerlos en contacto para permitir que se apareen para formar un bloque de construcción de ácido nucleico de doble hebra. Un bloque de construcción de ácido nucleico de doble hebra puede ser de tamaño variable. Los tamaños de estos bloques de construcción pueden ser pequeños o grandes. Los tamaños ejemplares para el bloque de construcción están en el rango de 1 par de base (no incluyendo salientes) a 100.000 pares de base (no incluyendo salientes). También se proporcionan otros rangos de tamaño ejemplares, que tienen límites inferiores de 1 pb a 10.000 pb (incluyendo cada valor entero entremedio) y límites superiores de 2 pb a 100.000 pb (incluyendo cada valor entero entremedio).

20 Existen muchos métodos por los cuales pueden generarse bloques de construcción de ácido nucleico de doble hebra que son útiles para la invención; y estos se conocen en la técnica y el experto puede realizarlos fácilmente. En algunas realizaciones, un bloque de construcción de ácido nucleico de doble hebra se genera primero generando dos ácidos nucleicos de una sola hebra y permitiéndoles aparearse para formar un bloque de construcción de ácido nucleico de doble hebra. Las dos hebras de un bloque de construcción de ácido nucleico de doble hebra pueden ser complementarias en cada nucleótido aparte de cualquiera que forma una saliente; por lo tanto no contienen no coincidencias, aparte de cualquier saliente. En otra realización, las dos hebras de un bloque de construcción de ácido nucleico de doble hebra son complementarias en menos nucleótidos aparte de cualquiera que forma una saliente. Por lo tanto, de acuerdo con esta realización, un bloque de construcción de ácido nucleico de doble hebra puede usarse para introducir degeneración de codones. En algunas realizaciones la degeneración de codones se introduce usando la mutagénesis de saturación de sitio descrita en la presente, usando uno o más cassettes N,N,G/T o alternativamente usando uno o más cassettes N,N,N.

25 El método de la recombinación in vivo de acuerdo con la invención puede realizarse ciegamente en una piletta de híbridos desconocidos o alelos de un polinucleótido o secuencia específicos. Sin embargo, no es necesario saber la secuencia de ADN o ARN del polinucleótido específico. El abordaje de usar la recombinación dentro de una población mezclada de genes puede ser útil para la generación de cualquier proteína útil, por ejemplo, una aldolasa de acuerdo con la invención o una variante de la misma. Este abordaje puede usarse para generar proteínas que tienen especificidad o actividad alterada. Este abordaje también puede ser útil para la generación de secuencias de ácido nucleico híbridas, por ejemplo, regiones de promotor, intrones, exones, secuencias de potenciador, 31 regiones sin traducir o 51 regiones sin traducir de genes. Por lo tanto este abordaje puede usarse para generar genes que tienen tasas aumentadas de expresión. Este abordaje también puede ser útil en el estudio de secuencias de ADN repetitivas. Finalmente, este abordaje puede ser útil para realizar ribozimas o aptámeros de acuerdo con la invención.

30 En algunas realizaciones la invención descrita en la presente se dirige al uso de ciclos repetidos de reordenamiento reductor, recombinación y selección que permiten la evolución molecular dirigida de secuencias lineales altamente complejas, tales como ADN, ARN o proteínas a través de la recombinación.

#### Sistema de evolución dirigida optimizada

La invención proporciona un sistema de modificación génica no estocástico denominado "sistema de evolución dirigida optimizada" para generar polipéptidos, tales como enzimas o anticuerpos de aldolasa, tal como piruvato

aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención, con propiedades nuevas o alteradas. En algunas realizaciones, la evolución dirigida optimizada se dirige al uso de ciclos repetidos de reordenamiento reductor, recombinación y selección que permiten la evolución molecular dirigida de ácidos nucleicos a través de la recombinación.

5 La evolución dirigida optimizada permite la generación de una gran población de secuencias quiméricas evolucionadas, en donde la población generada se enriquece significativamente para las secuencias que tienen un número predeterminado de eventos de cruce. Un evento de cruce es un punto en una secuencia quimérica donde un cambio en la secuencia ocurre de una variante parental a otra variante parental. Dicho punto está normalmente en la unión de donde los oligonucleótidos de dos padres se ligan para formar una sola secuencia. Este método permite el  
10 cálculo de las concentraciones correctas de secuencias de oligonucleótidos de modo que la población quimérica final de secuencias se enriquezca para el número elegido de eventos de cruce. Esto proporciona un mayor control sobre la elección de variantes quiméricas que tienen un número predeterminado de eventos de cruce.

Además, este método proporciona medios convenientes para explorar una gran cantidad del espacio variante de proteína posible en comparación con otros sistemas. Previamente, si se generaban por ejemplo  $10^{13}$  moléculas quiméricas durante una reacción, sería extremadamente difícil analizar dicho alto número de variantes quiméricas para una actividad en particular. Además, una porción significativa de la población de progenie tendría un número muy alto de eventos de cruce que resultaban en proteínas que tenían menos posibilidad de tener niveles aumentados de una actividad particular. Usando estos métodos, la población de moléculas quiméricas puede enriquecerse para estas variantes que tienen un número particular de eventos de cruce. Por lo tanto, si bien se  
15 pueden generar aún  $10^{13}$  moléculas quiméricas durante una reacción, cada una de las moléculas elegidas para el análisis adicional tiene más probablemente, por ejemplo, sólo tres eventos de cruce. Dado que la población de progenie resultante puede desviarse para tener un número predeterminado de eventos de cruce, los límites en la variedad funcional entre las moléculas quiméricas son reducidos. Esto proporciona un número más manejable de variables cuando se calcula qué oligonucleótido de los polinucleótidos parentales originales puede ser responsable de afectar un rasgo particular.  
20  
25

Un método para crear una secuencia de polinucleótido de progenie quimérica es crear oligonucleótidos que corresponden a fragmentos o porciones de cada secuencia parental. Cada oligonucleótido en algunas realizaciones incluye una región única de superposición de modo que mezclar los oligonucleótidos resulta en una nueva variante que tiene cada fragmento de oligonucleótido ensamblado en el orden correcto. Los protocolos alternativos para  
30 realizar estos métodos de acuerdo con la invención pueden encontrarse en la Patente de los Estados Unidos No. 6.773.900; 6.740.506; 6.713.282; 6.635.449; 6.605.449; 6.537.776; 6.361.974.

El número de oligonucleótidos generados para cada variante parental tiene una relación con el número total de cruces resultantes en la molécula quimérica que se crea al final. Por ejemplo, tres variantes de secuencia de nucleótidos parentales pueden proporcionarse para sufrir una reacción de ligación de modo de encontrar una variante quimérica que tenga por ejemplo mayor actividad a una alta temperatura. Como ejemplo, un juego de 50  
35 secuencias de oligonucleótidos puede generarse correspondiendo con cada porción de cada variante parental. Por consiguiente, durante el proceso de reensamblaje de ligación podría haber hasta 50 eventos de cruce dentro de cada una de las secuencias quiméricas. La probabilidad de que cada polinucleótido quimérico generado contenga oligonucleótidos de cada variante parental en orden alterno es muy baja. Si cada fragmento de oligonucleótido está presente en la reacción de ligación en la misma cantidad molar, es probable que en algunas posiciones los oligonucleótidos del mismo polinucleótido parental se ligan próximos entre sí y, por lo tanto, no resulten en un evento de cruce. Si la concentración de cada oligonucleótido de cada padre se mantiene constante durante cualquier etapa de ligación en este ejemplo, hay una posibilidad de  $\frac{1}{3}$  (asumiendo 3 padres) de que un oligonucleótido de la misma variante parental se ligue dentro de la secuencia quimérica y no produzca un cruce.  
40

Por consiguiente, puede determinarse una función de densidad de probabilidad (PDF) para predecir la población de eventos de cruce que son probables que ocurran durante cada etapa en una reacción de ligación dado un número fijo de variantes parentales, un número de oligonucleótidos que corresponden a cada variante y las concentraciones de cada variante durante cada etapa en la reacción de ligación. Las estadísticas y la matemática detrás de la determinación del PDF se describen a continuación. Utilizando estos métodos, se puede calcular dicha función de densidad de probabilidad, y por lo tanto enriquecer la población de progenie quimérica para un número predeterminado de eventos de cruce que resultan de una reacción de ligación particular. Además, puede predeterminarse un número objetivo de eventos de cruce y el sistema entonces puede programarse para calcular las cantidades de inicio de cada oligonucleótido parental durante cada etapa en la reacción de ligación para resultar en una función de densidad de probabilidad que se centra en el número predeterminado de eventos de cruce. Estos  
45 métodos se dirigen al uso de ciclos repetidos de reordenamiento reductor, recombinación y selección que permiten la evolución molecular dirigida de un ácido nucleico codificando un polipéptido a través de la recombinación. El sistema permite la generación de una gran población de secuencias quiméricas evolucionadas, en donde la población generada se enriquece significativamente para las secuencias que tienen un número predeterminado de eventos de cruce. Un evento de cruce es un punto en una secuencia quimérica donde un cambio en la secuencia ocurre de una variante parental a otra variante parental. Dicho punto está normalmente en la unión de donde los  
50  
55  
60

oligonucleótidos de dos padres se ligan para formar una sola secuencia. Este método permite el cálculo de las concentraciones correctas de secuencias de oligonucleótidos de modo que la población quimérica final de secuencias se enriquezca para el número elegido de eventos de cruce. Esto proporciona un mayor control sobre la elección de variantes quiméricas que tienen un número predeterminado de eventos de cruce.

- 5 Además, estos métodos proporcionan un medio conveniente para explorar una gran cantidad del espacio variante de proteína posible en comparación con otros sistemas. Usando los métodos descritos en la presente, la población de moléculas quiméricas puede enriquecerse para estas variantes que tienen un número particular de eventos de cruce. Por lo tanto, si bien se pueden generar aún  $10^{13}$  moléculas quiméricas durante una reacción, cada una de las moléculas elegidas para el análisis adicional tiene más probablemente, por ejemplo, sólo tres eventos de cruce.
- 10 Dado que la población de progenie resultante puede desviarse para tener un número predeterminado de eventos de cruce, los límites en la variedad funcional entre las moléculas quiméricas son reducidos. Esto proporciona un número más manejable de variables cuando se calcula qué oligonucleótido de los polinucleótidos parentales originales puede ser responsable de afectar un rasgo particular.

- 15 En algunas realizaciones, el método crea una secuencia de polinucleótido de progenie quimérica creando oligonucleótidos que corresponden a fragmentos o porciones de cada secuencia parental. Cada oligonucleótido en algunas realizaciones incluye una región única de superposición de modo que mezclar los oligonucleótidos resulta en una nueva variante que tiene cada fragmento de oligonucleótido ensamblado en el orden correcto. Ver la Patente de los Estados Unidos No. 6.773.900; 6.740.506; 6.713.282; 6.635.449; 6.605.449; 6.537.776; 6.361.974.

#### Determinación de eventos de cruce

- 20 Las realizaciones de la invención incluyen un sistema y software que reciben una función de densidad de probabilidad (PDF) de cruce deseada, el número de genes parentales a reensamblarse y el número de fragmentos en el reensamblaje como entradas. La salida de este programa es un "PDF fragmento" que puede usarse para determinar una receta para producir genes reensamblados y la PDF de cruce estimada de esos genes. El procesamiento descrito en la presente se realiza en algunas realizaciones en MATLAB™ (The Mathworks, Natick, Mass.) un lenguaje de programación y entorno de desarrollo para la computación técnica.
- 25

#### Procesos iterativos

- 30 Cualquier proceso de acuerdo con la invención puede repetirse, tal como un ácido nucleico que codifica un fenotipo nuevo o alterado de una enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención, puede identificarse, reaislarse, modificarse nuevamente, reanalizarse para la actividad. Este proceso puede repetirse hasta que se diseñe un fenotipo deseado. Por ejemplo, puede diseñarse una vía anabólica o catabólica bioquímica entera en una célula, incluyendo, por ejemplo, una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa.

- 35 De forma similar, si se determina que un oligonucleótido particular no afecta de ningún modo el rasgo deseado (tal como un nuevo fenotipo de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa) puede eliminarse como una variable sintetizando oligonucleótidos parentales más grandes que incluyen la secuencia a eliminarse. Dado que incorporar la secuencia dentro de una secuencia mayor evita cualquier evento de cruce, no habrá más variación alguna de esta secuencia en los polinucleótidos de progenie. Esta práctica iterativa de determinar qué oligonucleótidos se relacionan más con el rasgo deseado, y cuáles no están relacionados, permite una exploración más eficiente de todas las variantes de proteína posibles que pueden proporcionar un rasgo o actividad particular.
- 40

#### Transposición in vivo

- 45 En varias realizaciones, la transposición in vivo de moléculas se usa en métodos de acuerdo con la invención para proporcionar variantes de polipéptidos de acuerdo con la invención, tal como anticuerpos de acuerdo con la invención o enzimas aldolasas de acuerdo con la invención, tal como piruvato aldolasa, HMG y/o KHG aldolasa y similares. La transposición in vivo puede realizarse utilizando la propiedad natural de las células para recombinar multímeros. Mientras que la recombinación in vivo proporcionó la principal ruta natural para la diversidad molecular, la recombinación genética sigue siendo un proceso relativamente complejo que implica 1) el reconocimiento de homologías; 2) escisión de hebras, invasión de hebras y etapas metabólicas que conducen 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.
- 50

- 55 En otras realizaciones, la invención incluye un método para producir un polinucleótido híbrido a partir de al menos un primer polinucleótido y un segundo polinucleótido. En algunas realizaciones, la invención puede usarse para producir un polinucleótido híbrido introduciendo al menos un primer polinucleótido y un segundo polinucleótido (tal como uno, o ambos, siendo una secuencia de codificación de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención) que comparten al menos una región de homología de secuencia parcial en una célula huésped adecuada. Las regiones de homología de secuencia parcial promueven procesos que

5 resultan en la reorganización de secuencia produciendo un polinucleótido híbrido. El término “polinucleótido híbrido” como se usa en la presente es cualquier secuencia de nucleótidos que resulta del método de la presente invención y contiene secuencia de al menos dos secuencias originales de polinucleótidos. Dichos polinucleótidos híbridos pueden resultar de eventos de recombinación intermolecular que promueven la integración de secuencia entre moléculas de ADN. Además, dichos 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.

10 En algunas realizaciones, el reordenamiento vivo se enfoca en procesos “intermoleculares” colectivamente denominados “recombinación”; que en las bacterias se ve generalmente como un fenómeno “dependiente de RecA”. En algunas realizaciones, la invención puede depender de procesos de recombinación de una célula huésped para recombinar y reordenar secuencias, o la capacidad de las células para mediar procesos reductores para disminuir la complejidad de secuencias casi repetidas en la célula por eliminación. Este proceso de “reordenamiento reductor” ocurre por un proceso independiente de RecA “intramolecular”.

15 En otras realizaciones de la invención, pueden generarse nuevos polinucleótidos mediante el proceso de reordenamiento reductor. El método implica la generación de constructos que contienen secuencias consecutivas (secuencias de codificación originales), su inserción en un vector apropiado y su introducción posterior en una célula huésped apropiada. El reordenamiento de las identidades moleculares individuales ocurre por procesos combinatorios entre las secuencias consecutivas en el constructo teniendo regiones de homología o entre unidades casi repetidas. El proceso de reordenamiento recombina y/o reduce la complejidad de extensión de las secuencias repetidas y resulta en la producción de especies moleculares nuevas. Pueden aplicarse varios tratamientos para mejorar la tasa de reordenamiento. Estos podrían incluir el tratamiento con luz ultravioleta, o químicos que dañan el ADN y/o el uso de líneas celulares huésped que muestran niveles mejorados de “inestabilidad genética”. Por lo tanto el proceso de reordenamiento puede implicar la recombinación homóloga o la propiedad natural de secuencias casi repetidas para dirigir su propia evolución.

25 Las secuencias repetidas o “casi repetidas” tienen un papel en la inestabilidad genética. En algunas realizaciones, las “casi repeticiones” son repeticiones que no se limitan a su estructura de unidad original. Las unidades casi repetidas pueden presentarse como un arreglo de secuencias en un constructo; unidades consecutivas de secuencias similares. Una vez ligadas, las uniones entre las secuencias consecutivas se vuelven esencialmente invisibles y la naturaleza casi repetitiva del constructo resultante ahora es continuo en el nivel molecular. El proceso de eliminación que la célula realiza para reducir la complejidad del constructo resultante opera entre las secuencias casi repetidas. Las unidades casi repetidas proporcionan un repertorio prácticamente ilimitado de plantillas sobre el cual pueden ocurrir eventos de deslizamiento. En algunas realizaciones, los constructos que contienen las casi repeticiones por lo tanto proporcionan efectivamente una elasticidad molecular suficiente tal que los eventos de eliminación (y potencialmente la inserción) pueden ocurrir prácticamente en cualquier lugar dentro de las unidades casi repetitivas.

40 Cuando las secuencias casi repetidas se ligan todas en la misma orientación, por ejemplo cabeza con cola o viceversa, la célula no puede distinguir unidades individuales. Por consiguiente, el proceso reductor puede ocurrir en todas las secuencias. En contraste, cuando por ejemplo las unidades se presentan cabeza con cabeza, en vez de cabeza con cola, la inversión delinea los criterios de valoración de la unidad adyacente de modo que la formación de la eliminación favorecerá la pérdida de unidades discretas. Por lo tanto, se prefiere con el presente método que las secuencias estén en la misma orientación. La orientación aleatoria de las secuencias casi repetidas resultará en la pérdida de la eficiencia del reordenamiento, mientras que la orientación consistente de las secuencias ofrecerá la mayor eficiencia. Sin embargo, mientras que tener menos secuencias contiguas en la misma orientación reduce la eficiencia, aún puede proporcionar suficiente elasticidad para la recuperación efectiva de nuevas moléculas. Los constructos pueden realizarse con las secuencias casi repetidas en la misma orientación para permitir una mayor eficiencia.

Las secuencias se pueden ensamblar en una orientación de cabeza a cola usando cualquiera de una variedad de métodos, que incluyen los siguientes:

- 50 • a) Se pueden utilizar cebadores que incluyen una cabeza poli-A y cola poli-T que cuando se hacen de una sola hebra proporcionarían orientación. Esto se logra teniendo las primeras nuevas bases de los cebadores realizados a partir de ARN y por lo tanto ARNasaH fácilmente eliminada.
- b) Pueden utilizarse cebadores que incluyen sitios de escisión de restricción única. Se requerirían múltiples sitios, una batería de secuencias únicas y etapas de síntesis y ligación repetidas.
- 55 • c) Las pocas bases internas del cebador podrían tiorarse y se podría usar una exonucleasa para producir moléculas con cola de forma apropiada.

En algunas realizaciones, la recuperación de las secuencias reordenadas depende de la identificación de vectores de clonación con un índice repetitivo reducido (RI). Las secuencias de codificación reordenadas pueden recuperarse

luego mediante amplificación. Los productos se reclonan y se expresan. La recuperación de vectores de clonación con RI reducido puede estar afectada por:

- 1) El uso de vectores sólo mantenidos establemente cuando el constructo se reduce en complejidad
  - 2) La recuperación física de vectores acortados mediante procedimientos físicos. En este caso, el vector de clonación se recuperaría usando procedimientos de aislamiento de plásmido estándar y se fraccionaría en tamaño en gel de agarosa o columna con un bajo corte de peso molecular utilizando procedimientos estándar.
  - 3) La recuperación de vectores que contienen genes interrumpidos que pueden seleccionarse cuando el tamaño de inserción se reduce.
  - 4) El uso de técnicas de selección directa con un vector de expresión y la selección apropiada.
- Las secuencias de codificación (por ejemplo genes) de organismos relacionados pueden demostrar un alto grado de homología y codificar productos de proteína bastante diversos. Estos tipos de secuencias son particularmente útiles en la presente invención como casi repeticiones. Sin embargo, si bien los ejemplos ilustrados a continuación demuestran el reordenamiento de secuencias de codificación originales casi idénticas (casi repeticiones), este proceso no se limita a dichas repeticiones casi idénticas.

- El siguiente ejemplo demuestra un método ejemplar de acuerdo con la invención. Se describen las secuencias de ácido nucleico de codificación (casi repeticiones) derivadas de tres (3) especies únicas. Cada secuencia codifica una proteína con un juego distinto de propiedades. Cada una de las secuencias difiere por un único o por pocos pares de base en una posición única en la secuencia. Las secuencias casi repetidas se amplifican separada o conjuntamente y se ligan en ensamblajes aleatorios de tal modo que las permutaciones y combinaciones posibles estén disponibles en la población de moléculas ligadas. El número de unidades casi repetidas puede controlarse mediante las condiciones de ensamblaje. El número promedio de unidades casi repetidas en un constructo se define como el índice repetitivo (RI).

- Una vez formados, los constructos pueden o no fraccionarse en tamaño en un gel de agarosa de acuerdo con los protocolos publicados, insertarse en un vector de clonación y transfectarse en una célula huésped apropiada. Luego se propagan las células y se efectúa un "reordenamiento reductor". La tasa del proceso de reordenamiento reductor puede estimularse mediante la introducción de daño al ADN si se desea. No es relevante si la reducción en RI está mediada por la formación de eliminación entre secuencias repetidas mediante un mecanismo "intramolecular", o mediada mediante eventos similares a la recombinación a través de mecanismos "intermoleculares". El resultado final es un reordenamiento de las moléculas en todas las combinaciones posibles.

- Opcionalmente, el método comprende la etapa adicional de analizar los miembros de biblioteca de la piletta transpuesta para identificar miembros de biblioteca transpuestos individuales que tienen la capacidad de unirse o de otro modo interactuar, o catalizar una reacción particular (tal como un dominio catalítico de una enzima) con una macromolécula predeterminada, tal como por ejemplo un receptor proteínico, un oligosacárido, virión u otro compuesto o estructura predeterminada.

- Los polipéptidos que se identifican de dichas bibliotecas pueden usarse con objetivos terapéuticos, diagnósticos, de investigación y temas relacionados (tales como catalizadores, solutos para aumentar la osmolaridad de una solución acuosa y similares) y/o pueden someterse a uno o más ciclos adicionales de transposición y/o selección.

- En otras realizaciones, se prevé que antes de o durante la recombinación u reordenamiento, los polinucleótidos generados por el método de acuerdo con la invención se sometan a agentes o procesos que promueven la introducción de mutaciones en los polinucleótidos originales. La introducción de dichas mutaciones aumentaría la diversidad de los polinucleótidos y polipéptidos híbridos codificados desde allí. Los agentes o procesos que promueven la mutagénesis pueden incluir pero no se limitan a: (+)-CC-1065, o un análogo sintético tal como (+)-CC-1065-(N3-Adenina (ver Sun y Hurley, (1992); un aducto 4'-fluoro-4-aminobifenil N-acetilado o deacetilado capaz de inhibir la síntesis de ADN (ver, por ejemplo, van de Poll et al. (1992)); o un aducto 4-aminobifenil N-acetilado o deacetilado capaz de inhibir la síntesis de ADN (ver también, van de Poll et al. (1992), pp. 751-758); cromo trivalente, una sal de cromo trivalente, un aducto de ADN de hidrocarburo policíclico aromático (PAH) capaz de inhibir la replicación de ADN, tal como 7-bromometil-benz[a]antraceno ("BMA"), tris(2,3-dibromopropil)fosfato ("Tris-BP"), 1,2-dibromo-3-cloropropano ("DBCP"), 2-bromoacroleína (2BA), benzo[a]pireno-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"). Ejemplos de medios para ralentizar o detener la amplificación PCR consiste de luz UV (+)-CC-1065 y (+)-CC-1065-(N-3-Adenina). Medios que se abarcan en particular son aductos de ADN o polinucleótidos que comprenden los aductos de ADN de los polinucleótidos o la piletta de polinucleótidos, que pueden liberarse o eliminarse mediante un proceso que incluye calentar la solución que comprende los polinucleótidos antes del procesamiento adicional.

- En otras realizaciones la invención se dirige a un método para producir proteínas recombinantes que tienen actividad biológica tratando una muestra que comprende polinucleótidos de plantilla de doble hebra que codifican una proteína

de tipo silvestre bajo condiciones de acuerdo con la invención que proporcionan la producción de polinucleótidos híbridos o reordenados.

#### Producción de variantes de secuencia

5 La invención también proporciona métodos adicionales para realizar variantes de secuencias de las secuencias de ácido nucleico (tal como una enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa) de acuerdo con la invención. En algunas realizaciones, la invención también proporciona métodos adicionales para aislar enzimas aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa usando los ácidos nucleicos y polipéptidos de acuerdo con la invención. En algunas realizaciones, la invención proporciona variantes de una aldolasa, tal como secuencia de codificación de enzima piruvato aldolasa, HMG y/o KHG aldolasa (tal como un gen, ADNc o mensaje) de acuerdo con la invención, que puede alterarse por cualquier medio, incluyendo por ejemplo métodos aleatorios o estocásticos o no estocásticos o métodos de "evolución dirigida" como se describió anteriormente.

15 Las variantes aisladas pueden ser naturales. La variante también puede crearse in vitro. Las variantes pueden crearse usando técnicas de manipulación genética tales como mutagénesis dirigida al sitio, mutagénesis química aleatoria, procedimientos de eliminación de exonucleasa III y técnicas de clonación estándar. Alternativamente, dichas variantes, fragmentos, análogos o derivados pueden crearse usando procedimientos de síntesis química o modificación. Otros métodos para realizar variantes también son conocidos por los expertos en la técnica. Esto incluye procedimientos en los cuales las secuencias de ácido nucleico obtenidas de aislados naturales se modifican para generar ácidos nucleicos que codifican polipéptidos que tienen características que mejoran su valor en aplicaciones industriales o de laboratorio. En dichos procedimientos, se genera y caracteriza un gran número de secuencias variantes que tienen una o más diferencias de nucleótidos con respecto a la secuencia obtenida del aislado natural. Estas diferencias de nucleótidos pueden resultar en cambios de aminoácidos con respecto a los polipéptidos codificados por los ácidos nucleicos de los aislados naturales.

25 Por ejemplo, pueden crearse variantes usando PCR propensas a error. En algunas realizaciones de PCR propensa a error, se realiza la PCR bajo condiciones donde la fidelidad de copia de la polimerasa ADN es baja, de tal modo que se obtenga una alta tasa de mutaciones puntuales a lo largo de la longitud entera del producto PCR. Se describe la PCR propensa a error tal como en Leung, D. W. et al., (1989) *Technique* 1:11-15; y Caldwell, R.C. & Joyce, G.F., (1992) *PCR Methods Applic.* 2:28-33. Brevemente, en dichos procedimientos, los ácidos nucleicos a mutagenizarse se mezclan con cebadores de PCR, solución amortiguadora de reacción, MgCl<sub>2</sub>, MnCl<sub>2</sub>, Taq polimerasa y una concentración apropiada de dNTP para lograr una alta tasa de mutación puntual a lo largo de la longitud entera del producto PCR. Por ejemplo, la reacción puede realizarse usando 20 fmoles de ácido nucleico a mutagenizarse, 30 pmoles de cada cebador de PCR, una solución amortiguadora de reacción que comprende 50 mM KCl, 10 mM Tris HCl (pH 8,3) y 0,01% gelatina, 7 mM MgCl<sub>2</sub>, 0,5 mM MnCl<sub>2</sub>, 5 unidades de Taq polimerasa, 0,2 mM dGTP, 0,2 mM dATP, 1 mM dCTP y 1 mM dTTP. La PCR puede realizarse por 30 ciclos de 94°C durante 1 minuto, 45°C durante 1 minuto y 72°C durante 1 minuto. Sin embargo, se apreciará que estos parámetros pueden variar según sea apropiado. Los ácidos nucleicos mutagenizados se clonan en un vector apropiado y se evalúan las actividades de los polipéptidos codificados por los ácidos nucleicos mutagenizados.

40 En algunas realizaciones, se crean las variantes usando mutagénesis dirigida de oligonucleótidos para generar mutaciones específicas del sitio en cualquier ADN clonado de interés. La mutagénesis de oligonucleótidos se describe por ejemplo en Reidhaar-Olson (1988) *Science* 241:53-57. Brevemente, en dichos procedimientos una pluralidad de oligonucleótidos de doble hebra que tienen una o más mutaciones a introducirse en el ADN clonado se sintetiza y se inserta en el ADN clonado para mutagenizarse. En algunas realizaciones, los clones que contienen el ADN mutagenizado se recuperan, se expresan y las actividades del polipéptido codificado allí se evalúan.

45 Otro método para generar variantes es la PCR de ensamblaje. La PCR de ensamblaje implica el ensamblaje de un producto PCR de una mezcla de fragmentos de ADN pequeños. Un gran número de reacciones PCR diferentes ocurre en paralelo en el mismo vial, con los productos de una reacción que ceban los productos de otra reacción. La PCR de ensamblaje se describe en la técnica, tal como en la Patente de Estados Unidos 5.965.408.

50 En algunas realizaciones, la mutagénesis PCR sexual es un ejemplo de método para generar variantes de acuerdo con la invención. En algunas realizaciones de mutagénesis PCR sexual la recombinación homóloga forzada ocurre entre moléculas de ADN de una secuencia de ADN diferente pero altamente relacionada in vitro, como resultado de la fragmentación aleatoria de la molécula de ADN basada en la homología de secuencia, seguida por la fijación del cruce mediante la extensión del cebador en una reacción PCR. La mutagénesis PCR sexual se describe por ejemplo en Stemmer (1994) *Proc. Natl. Acad. Sci. Estados Unidos* 91:10747-10751. Brevemente, en dichos procedimientos una pluralidad de ácidos nucleicos a recombinarse se digieren con Dnasa para generar fragmentos que tienen un tamaño promedio de 50 -200 nucleótidos. Los fragmentos del tamaño promedio deseado se purifican y se resuspenden en una mezcla PCR. La PCR se realiza bajo condiciones que facilitan la recombinación entre los fragmentos de ácido nucleico. Por ejemplo, la PCR puede realizarse resuspendiendo los fragmentos purificados en una concentración de 10-30 ng/μl en una solución de 0,2 mM de cada dNTP, 2,2 mM MgCl<sub>2</sub>, 50 mM KCL, 10 mM Tris HCl, pH 9,0, y 0,1% Triton X-100. Se agregan 2,5 unidades de Taq polimerasa por 100 μl de mezcla de reacción



- y se realiza la PCR usando 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, se apreciará que estos parámetros pueden variar según sea apropiado. En algunas realizaciones, los oligonucleótidos pueden estar incluidos en las reacciones PCR. En otras realizaciones, el fragmento Klenow de polimerasa ADN I puede usarse en un primer juego de reacciones PCR y la Taq polimerasa puede usarse en un juego posterior de reacciones PCR. Las secuencias recombinantes se aíslan y las actividades de los polipéptidos que codifican se evalúan.
- En algunas realizaciones, las variantes se crean mediante mutagénesis in vivo. En algunas realizaciones, las mutaciones aleatorias en una secuencia de interés se generan propagando la secuencia de interés en una cepa bacteriana, tal como la cepa *E. Coli* que lleva mutaciones en una o más de las vías de reparación de ADN. Dichas cepas "mutadoras" tienen una tasa de mutación aleatoria más alta que la de un padre de tipo silvestre. Propagar el ADN en una de estas cepas finalmente generará mutaciones aleatorias dentro del ADN. Las cepas mutadoras adecuadas para el uso para la mutagénesis in vivo se describen en la Publicación PCT No. WO 91/16427, publicada el 31 de oct. de 1991 titulada "Métodos para la creación de fenotipos a partir de poblaciones de genes múltiples".
- Las variantes también pueden generarse usando mutagénesis de cassette. En la mutagénesis de cassette una pequeña región de una molécula de ADN de doble hebra se reemplaza por un "cassette" de oligonucleótido sintético que difiere de la secuencia nativa. El oligonucleótido contiene con frecuencia una secuencia nativa completamente y/o parcialmente aleatorizada.
- También puede usarse la mutagénesis de ensamblaje recursivo para generar variantes. La mutagénesis de ensamblaje recursivo es un algoritmo para la manipulación de proteínas (mutagénesis de proteínas) desarrollada para producir poblaciones diversas de mutantes fenotípicamente relacionadas cuyos miembros difieren en la secuencia de aminoácido. Este método usa un mecanismo de retroalimentación para controlar rondas sucesivas de mutagénesis de cassette combinatorias. La mutagénesis de ensamblaje recursivo se describe por ejemplo en Arkin (1992) Proc. Natl. Acad. Sci. Estados Unidos 89:7811-7815.
- En algunas realizaciones, las variantes se crean usando mutagénesis de ensamblaje exponencial. La mutagénesis de ensamblaje exponencial es un proceso para generar bibliotecas combinatorias con un alto porcentaje de mutantes únicos y funcionales, en donde pequeños grupos de residuos se aleatorizan en paralelo para identificar, en cada posición alterada, aminoácidos que conducen a proteínas funcionales. La mutagénesis de ensamblaje exponencial se describe por ejemplo en Delegrave (1993) Biotechnology Res. 11:1548-1552. La mutagénesis aleatoria y dirigida al sitio se describe, por ejemplo, en Arnold (1993) Current Opinion in Biotechnology 4:450-455.
- En algunas realizaciones, las variantes se crean usando procedimientos de transposición en donde porciones de una pluralidad de ácidos nucleicos que codifican distintos polipéptidos se fusionan para crear secuencias de ácido nucleico quimérico que codifican polipéptidos quiméricos como se describe en la Patente de Estados Unidos 5.965.408, presentada el 9 de jul. de 1996, titulada "Método de reensamblaje de ADN mediante interrupción de la síntesis" y la Patente de los Estados Unidos No. 5.939.250, presentada el 22 de mayo de 1996, titulada "Producción de enzimas que tienen actividades deseadas por mutagénesis".
- Las variantes de los polipéptidos de acuerdo con la invención pueden ser variantes en las cuales uno o más de los residuos de aminoácidos de los polipéptidos de las secuencias de acuerdo con la invención se sustituyen con un residuo de aminoácido conservado o no conservado (en algunas realizaciones, un residuo de aminoácido conservado) y dicho residuo de aminoácido sustituido puede o no codificarse por el código genético.
- En algunas realizaciones, las sustituciones conservadoras son aquellas que sustituyen un aminoácido dado en un polipéptido por otro aminoácido de características similares. En algunas realizaciones, las sustituciones conservadoras de acuerdo con la invención comprenden los siguientes reemplazos: reemplazos de un aminoácido alifático tal como Alanina, Valina, Leucina e Isoleucina con otro aminoácido alifático; reemplazo de una Serina con una Treonina o viceversa; reemplazo de un residuo ácido tal como ácido Aspártico y ácido Glutámico con otro residuo ácido; reemplazo de un residuo que tiene un grupo amida, tal como Asparagina y Glutamina, con otro residuo que tiene un grupo amida; intercambio de un residuo básico tal como Lisina y Arginina con otro residuo básico; y reemplazo 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 aminoácido de un polipéptido de acuerdo con la invención incluyen un grupo sustituyente. En algunas realizaciones, otras variantes son aquellas en las cuales el polipéptido se asocia con otro compuesto, tal como un compuesto para aumentar la vida media del polipéptido (por ejemplo, polietilenglicol). Las variantes adicionales son aquellas en las cuales se fusionan aminoácidos adicionales al polipéptido, tal como una secuencia líder, una secuencia secretora, una secuencia de proproteína o una secuencia que facilita la purificación, enriquecimiento o estabilización del polipéptido.
- En algunas realizaciones, los fragmentos, derivados y análogos retienen la misma función o actividad biológica que los polipéptidos de acuerdo con la invención y secuencias sustancialmente idénticas a las mismas. En otras

realizaciones, el fragmento, derivado o análogo incluye una proproteína, de tal modo que el fragmento, derivado o análogo pueda activarse mediante la escisión de la porción de proproteína para producir un polipéptido activo.

Optimización de codones para lograr altos niveles de expresión de proteína en células huésped

5 La invención proporciona métodos para modificar ácidos nucleicos que codifican una enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, para modificar (tal como optimizar) el uso de codones. En algunas realizaciones, la invención proporciona métodos para modificar codones en un ácido nucleico que codifica una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa para aumentar o reducir su expresión en una célula huésped. En algunas realizaciones, la invención también proporciona ácidos nucleicos que codifican una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa modifica para aumentar su expresión en una célula huésped, enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa así modificada, y métodos para preparar las enzimas aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa modificadas. El método comprende identificar un codón "no preferido" o "menos preferido" en ácido nucleico que codifica enzimas aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa y reemplazar uno o más de estos codones no preferidos o menos preferidos por un "codón preferido" que codifica el mismo aminoácido que el codón reemplazado y al menos un codón no preferido o menos preferido en el ácido nucleico ha sido reemplazado por un codón preferido que codifica el mismo aminoácido. Un codón preferido es un codón sobrerrepresentado en secuencias codificadoras en genes en la célula huésped y un codón no preferido o menos preferido es un codón infrarrepresentado en secuencias codificadoras en genes en la célula huésped.

20 Las células huésped para expresar los ácidos nucleicos, cassettes de expresión y vectores de acuerdo con la invención incluyen bacterias, levaduras, hongos, células vegetales, células de insectos y células de mamíferos (ver la discusión anteriormente). Por lo tanto, la invención proporciona métodos para optimizar el uso de codones en todas estas células, ácidos nucleicos alterados por codones y polipéptidos realizados por los ácidos nucleicos alterados por codón. Ejemplos de células huésped incluyen bacterias gram negativas tales como *Escherichia coli*; bacterias gram positivas, tales como *Streptomyces* sp., *Lactobacillus gasseri*, *Lactococcus lactis*, *Lactococcus cremoris*, *Bacillus subtilis*, *Bacillus cereus*. Ejemplos de células huésped también incluyen organismos eucariotas, tales como varias 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. Por lo tanto, la invención también incluye ácidos nucleicos y polipéptidos optimizados para la expresión en estos organismos y especies.

30 Por ejemplo, los codones de un ácido nucleico que codifican una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa aisladas de una célula bacteriana se modifican de tal modo que el ácido nucleico se expresa óptimamente en una célula bacteriana diferente de las bacterias de las cuales la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa se derivan, un levadura, un hongo, una célula vegetal, una célula de insecto o una célula mamífera. Los métodos para optimizar codones son bien conocidos en la técnica, ver 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. También ver Narum (2001) Infect. Immun. 69:7250-7253, que describe codones de optimización en sistemas de ratón; Outchkourov (2002) Protein Expr. Purif. 24:18-24, que describe codones de optimización en levaduras; Feng (2000) Biochemistry 39:15399-15409, que describe codones de optimización en *E. coli*; Humphreys (2000) Protein Expr. Purif. 20:252-264, que describe el uso de codones de optimización que afecta la secreción en *E. coli*.

#### Animales no humanos transgénicos

45 La invención proporciona animales no humanos transgénicos que comprenden un ácido nucleico, un polipéptido (tal como una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa), un cassette o vector de expresión o una célula transfectada o transformada de acuerdo con la invención. En algunas realizaciones, la invención también proporciona métodos para realizar y usar estos animales no humanos transgénicos.

50 Los animales no humanos transgénicos pueden ser, por ejemplo, perros, cabras, conejos, ovejas, caballos, peces, cerdos (incluyendo todos los porcinos, puercos y animales relacionados), vacas, ratas y ratones, que comprenden los ácidos nucleicos de acuerdo con la invención. Estos animales pueden usarse por ejemplo en modelos in vivo para estudiar la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, la actividad o como modelos para analizar agentes que cambian la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa actividad in vivo. Las secuencias de codificación para los polipéptidos para expresarse en animales no humanos transgénicos pueden diseñarse para ser constitutivas o bajo el control de factores reguladores transcripcionales específicos de tejido, específicos de desarrollo o inducibles.

55 Los animales no humanos transgénicos pueden diseñarse y generarse usando cualquier método conocido en la técnica; ver las Patentes de Estados Unidos No. 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 la realización y el uso de células y huevos transformados y ratones, ratas, conejos, ovejas, cerdos, pollos, cabras, peces y vacas transgénicos. Ver también por ejemplo Pollock (1999) J. Immunol. Methods 231:147-157, que

describe la producción de proteínas recombinantes en la leche de los animales transgénicos que producen 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, describe la realización y el uso de mamíferos no humanos transgénicos que expresan en sus cerebros un constructo de ácido nucleico que comprende una secuencia de ADN. La Patente de Estados Unidos 5.387.742, describe la inyección de secuencias de ADN clonadas recombinantes o sintéticas en huevos de ratón fertilizados, el implante de huevos inyectados en hembras pseudo-embarazadas, y el crecimiento de crecimiento a término de ratones transgénicos. La Patente de los Estados Unidos No. 6.187.992, describe la realización y el uso de un ratón transgénico.

Los "animales knockout" también pueden usarse para realizar los métodos de acuerdo con la invención. Por ejemplo, en algunas realizaciones, los animales transgénicos o modificados de acuerdo con la invención comprenden un "animal knockout", tal como un "ratón knockout", manipulado para no expresar un gen endógeno, que se reemplaza por un gen que expresa una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de acuerdo con la invención, o una proteína de fusión que comprende una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de acuerdo con la invención

### 15 Plantas y semillas transgénicas

La invención proporciona plantas y semillas transgénicas que comprenden un ácido nucleico, un polipéptido (tal como una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa), un cassette o vector de expresión o una célula transfectada o transformada de acuerdo con la invención. La invención también proporciona productos o derivados vegetales, tales como frutas, aceites, semillas, hojas, extractos y similares, incluyendo cualquier parte de planta, que comprende un ácido nucleico y/o polipéptido (tal como una xilanasasa) de la invención, tal como en donde el ácido nucleico o polipéptido de la invención es heterólogo para la planta, parte de planta, semilla, etc. La planta transgénica (que incluye partes de planta, frutas, semillas, etc.) puede ser dicotiledónea o monocotiledónea. En algunas realizaciones, la invención también proporciona métodos para realizar y usar estas plantas y semillas transgénicas. La planta o célula vegetal transgénicas que expresan un polipéptido de la presente invención pueden construirse de acuerdo con cualquier método conocido en la técnica. Ver, por ejemplo, la Patente de los Estados Unidos No. 6.309.872.

Los ácidos nucleicos y constructos de expresión de acuerdo con la invención pueden introducirse en una célula vegetal por cualquier medio. Por ejemplo, los ácidos nucleicos o constructos de expresión pueden introducirse en el genoma de un huésped 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 enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa del huésped está regulada por elementos de control transcripcionales o translacionales endógenos. En algunas realizaciones, la invención también proporciona "plantas knockout" en donde la inserción de la secuencia génica mediante por ejemplo recombinación homóloga ha alterado la expresión del gen endógeno. Los medios para generar plantas "knockout" se conocen bien en la técnica, ver (1998) Proc Natl. Acad. Sci. Estados Unidos 95:4368-4373; Miao (1995) Plant J 7:359-365. Ver la discusión sobre plantas transgénicas a continuación.

Los ácidos nucleicos de acuerdo con la invención pueden usarse para conferir rasgos deseados en esencialmente cualquier planta, tal como plantas productoras de almidón, tal como patata, tomate, soja, remolachas, maíz, trigo, arroz, cebada y similares. Los ácidos nucleicos de acuerdo con la invención pueden usarse para manipular vías metabólicas de una planta para optimizar o alterar la expresión del huésped de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. Los ácidos nucleicos de acuerdo con la invención pueden cambiar los niveles de expresión o actividad o alterar las características de los compuestos o enzimas producidos naturalmente en una planta. Alternativamente, una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de acuerdo con la invención puede usarse en la producción de una planta transgénica para producir un compuesto no producido naturalmente por dicha planta. Esto puede reducir los costos de producción o crear un nuevo producto.

En algunas realizaciones, la primera etapa en la producción de una planta transgénica implica realizar un constructo de expresión para la expresión en una célula vegetal. Estas técnicas son bien conocidas en la técnica. Pueden incluir seleccionar y clonar un promotor, una secuencia de codificación para facilitar la unión eficiente de ribosomas para ARNm y seleccionar las secuencias de terminador de gen apropiadas. Un ejemplo de promotor constitutivo es CaMV35S, del virus mosaico del coliflor, que generalmente resulta en un alto grado de expresión en plantas. Otros promotores son más específicos y responden a señales en el ambiente interno o externo de la planta. Un ejemplo de promotor inducible por luz es el promotor del gen cab, que codifica la proteína de unión a/b de clorofila principal.

En algunas realizaciones, el ácido nucleico se modifica para lograr una expresión mayor en una célula vegetal. Por ejemplo, es probable que una secuencia de acuerdo con la invención tenga un mayor porcentaje de pares de nucleótidos A-T comparados con el que se vio en la planta, algunos de los cuales prefieren pares de nucleótidos G-C. Por lo tanto, los nucleótidos A-T en la secuencia de codificación pueden sustituirse con nucleótidos G-C sin cambiar significativamente la secuencia de aminoácidos para mejorar la producción del producto génico en las células vegetales.

El gen marcador seleccionable puede agregarse al constructo génico para identificar células o tejidos vegetales que han integrado exitosamente el transgén. Esto puede ser necesario porque lograr la incorporación y expresión de los genes en células vegetales es un evento raro, que ocurre sólo en un porcentaje bajo de tejidos o células dirigidos. Los genes marcadores seleccionables codifican proteínas que proporcionan resistencia a agentes que son normalmente tóxicos para las plantas, tales como antibióticos o herbicidas. Sólo las células vegetales que han integrado el gen marcador seleccionable sobrevivirán cuando crecen en un medio que contiene el antibiótico o herbicida apropiados. Para otros genes insertados, los genes marcadores también requieren secuencias promotoras y de terminación para el funcionamiento adecuado.

En algunas realizaciones, realizar plantas o semillas transgénicas comprende incorporar secuencias de acuerdo con la invención y opcionalmente genes marcadores en un constructo de expresión objetivo (tal como un plásmido) junto con el posicionamiento del promotor y las secuencias de terminador. Esto puede implicar la transferencia del gen modificado en la planta a través de un método adecuado. Por ejemplo, un constructo puede introducirse directamente en el ADN genómico de la célula vegetal usando técnicas tales como electroporación y microinyección de protoplastos de célula vegetal, o los constructos pueden introducirse directamente al tejido vegetal usando métodos balísticos, tales como bombardeo de partículas de ADN. Por ejemplo, ver 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 describe el uso del bombardeo de partículas para introducir transgenes en trigo; y Adam (1997) *supra*, para el uso del bombardeo de partículas para introducir YAC en células vegetales. Por ejemplo, Rinehart (1997) *supra*, usó el bombardeo de partículas para generar plantas de algodón transgénicas. Los aparatos para acelerar partículas se describen en la Patente de los Estados Unidos No. 5.015.580 y el instrumento de de aceleración de partículas disponible Bio-Rad (Biologicals) PDS-2000 (Bio-Rad, Hercules, CA); ver también John, Patente de los Estados Unidos No. 5.608.148 y Ellis, Patente de los Estados Unidos No. 5.681.730, que describe la transformación de gimnospermas medidas por partículas.

En algunas realizaciones, los protoplastos pueden inmovilizarse e inyectarse con ácidos nucleicos, tales como un constructo de expresión. Si bien la regeneración vegetal a partir de protoplastos no es fácil con los cereales, la regeneración de vegetal es posible en legumbres usando embriogénesis somática a partir de callos derivados de protoplastos. Los tejidos organizados pueden transformarse con ADN desnudo usando la técnica de pistola de genes, donde el ADN se recubre en microproyectiles de tungsteno, se disparan 1/100 del tamaño de las células que llevan el ADN en profundidad en las células y organelos. El tejido transformado luego se induce para regenerarse habitualmente mediante embriogénesis somática. Esta técnica fue exitosa en varias especies de cereal que incluyen maíz y arroz.

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

Alternativamente, los ácidos nucleicos, tales como un constructo de expresión, pueden combinarse con regiones de flaqueo de ADN adecuadas e introducirse en un vector huésped de *Agrobacterium tumefaciens* convencional. Las funciones de virulencia del huésped de *Agrobacterium tumefaciens* se dirigirá a la inserción del constructo y marcador adyacente en el ADN de la célula vegetal cuando la célula se infecte por las bacterias. Las técnicas de transformación mediadas por *Agrobacterium tumefaciens*, incluyendo el desarme y el uso de vectores binarios, se describe bien en la bibliografía científica. Ver Horsch (1984) *Science* 233:496-498; Fraley (1983) *Proc. Natl. Acad. Sci. Estados Unidos* 80:4803 (1983); *Gene Transfer to Plants*, Potrykus, ed. (Springer-Verlag, Berlín 1995). El ADN en una célula *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 alargamiento de ADN denominado ADN-T (de ~20 kb de longitud) que se transfiere a la célula vegetal en el proceso de infección y una serie de genes vir (virulencia) que dirigen el proceso de infección. *A. tumefaciens* solo puede infectar una planta a través de heridas: cuando una raíz o tallo vegetal está herido da ciertas señales químicas en respuesta a las cuales los genes vir de *A. tumefaciens* se vuelven activados y dirigen una serie de eventos necesarios para la transferencia del ADN-T del plásmido Ti al cromosoma de la planta. El ADN-T luego ingresa a la célula vegetal a través de la herida. Una especulación es que el ADN-T espera hasta que el ADN vegetal se replique o transcriba, luego se inserta en el ADN vegetal expuesto. De modo de usar *A. tumefaciens* como un vector de transgén, la sección inductora de tumor de ADN-T tiene que eliminarse mientras que retiene las regiones de límite de ADN-T y los genes vir. El transgén luego se inserta entre las regiones de límite de ADN-T, donde se transfiere a la célula vegetal y se integra a los cromosomas vegetales.

La invención proporciona la transformación de plantas monocotiledóneas usando los ácidos nucleicos de acuerdo con la invención, incluyendo cereales importantes, ver Hiei (1997) *Plant Mol. Biol.* 35:205-218. Ver también, Horsch, *Science* (1984) 233:496; Fraley (1983) *Proc. Natl. Acad. Sci. Estados Unidos* 80:4803; Thykjaer (1997) *supra*; Park (1996) *Plant Mol. Biol.* 32:1135-1148, que describen la integración de ADN-T al ADN genómico. Ver 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 algunas realizaciones, la tercera etapa implica la selección y regeneración de plantas enteras capaces de transmitir el gen objetivo incorporado a la próxima generación. Dichas técnicas de regeneración pueden usar la manipulación de ciertas fitohormonas en un medio de crecimiento de cultivo de tejidos. En algunas realizaciones, el método usa un marcador de biocida y/o herbicida que fue introducido junto con las secuencias de nucleótidos deseadas. La regeneración vegetal de protoplastos cultivados se describe en Evans et al., *Protoplasts Isolation and Culture, Handbook of Plant Cell Culture*, pp. 124-176, MacMillan Publishing Company, Nueva York, 1983; y Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. La regeneración también puede obtenerse de callos vegetales, explantes, órganos o partes de los mismos. Dichas técnicas de regeneración se describen generalmente en Klee (1987) Ann. Rev. of Plant Phys. 38:467-486. Para obtener plantas enteras de tejidos transgénicos tales como embriones inmaduros, pueden cultivarse 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 enteras se generan y producen semillas, comienza la evaluación de la progenie.

En algunas realizaciones, luego que el cassette de expresión se incorpore establemente en plantas transgénicas, puede introducirse en otras plantas por cruce sexual. Se puede usar cualquiera de un número de técnicas de reproducción, dependiendo de la especie a cruzarse. Dado que la expresión transgénica de los ácidos nucleicos de acuerdo con la invención conduce a cambios fenotípicos, las plantas que comprenden los ácidos nucleicos de acuerdo con la invención pueden cruzarse sexualmente con una segunda planta para obtener un producto final. Por lo tanto, la semilla de acuerdo con la invención puede derivar de un cruce entre dos plantas transgénicas de acuerdo con la invención, o un cruce entre una planta de acuerdo con la invención y otra planta. Los efectos deseados (tales como la expresión de los polipéptidos de acuerdo con la invención para producir una planta en la cual se altera el comportamiento de florecimiento) puede mejorarse cuando ambas plantas parentales expresan los polipéptidos (tal como una enzima aldolasa, piruvato aldolasa, HMG y/o HHG aldolasa) de acuerdo con la invención. Los efectos deseados pueden pasarse a generaciones vegetales futuros por medios de propagación estándar.

En algunas realizaciones, los ácidos nucleicos y polipéptidos de acuerdo con la invención se expresan o se insertan en cualquier planta o semilla. Las plantas transgénicas de acuerdo con la invención pueden ser dicotiledóneas o monocotiledóneas. Ejemplos de plantas transgénicas monocotiledóneas de acuerdo con la invención son pastos, tales como pasto de pradera (pasto azul, *Poa*), pasto de forraje tal como festuca, *lolium*, pasto templado, tal como *Agrostis* y cereales, tal como trigo, avena, centeno, cebada, arroz, sorgo y maíz. Ejemplos de plantas dicotiledóneas transgénicas de acuerdo con la invención son tabaco, legumbres, tales como lupinas, patata, remolacha azucarera, guisantes, frijoles y soja, y plantas crucíferas (familia *Brassicaceae*) tal como coliflor, colza y el organismo modelo próximo relacionado *Arabidopsis thaliana*. Por lo tanto, las plantas y semillas transgénicas de acuerdo con la invención incluyen un amplio rango de plantas, incluyendo pero no limitándose a especies del género *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaceis*, *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*.

En realizaciones alternativas, los ácidos nucleicos de acuerdo con la invención se expresan en plantas que contienen células de fibra, incluyendo, por ejemplo, algodón, árbol de algodón de seda (Kapok, Ceiba pentandra), sauce del desierto, arbusto de la creosata, krascheninnikovia, balsa, ramio, kenaf, cáñamo, rosa de Jamaica, yute, sisal, abacá y lino. En realizaciones alternativas, las plantas transgénicas de acuerdo con la invención pueden ser miembros del género *Gossypium*, incluyendo miembros de cualquier especie *Gossypium*, tal como *G. arboreum*; *G. herbaceum*, *G. barbadense* y *G. hirsutum*.

La invención también proporciona plantas transgénicas que han de usarse para producir grandes cantidades de los polipéptidos (tales como una enzima o anticuerpo de aldolasa, piruvato aldolasa, HMG y/o KHG aldolasa) de acuerdo con la invención. Por ejemplo, ver Palmgren (1997) Trends Genet. 13:348; Chong (1997) Transgenic Res. 6:289-296 (producción de beta-caseína de proteína de la leche humana en plantas de patata transgénicas usando un promotor de sintasa de manopina bidireccional inducible por auxina (mas1',2') con métodos de transformación de disco de hoja mediado por *Agrobacterium tumefaciens*).

Usando procedimientos conocidos, un experto puede analizar plantas de acuerdo con la invención detectando el aumento o reducción de ARN, de transgén o proteínas en plantas transgénicas. Los medios para detectar y cuantificar ARNm, o proteínas son bien conocidos en la técnica.

#### Polipéptidos y péptidos

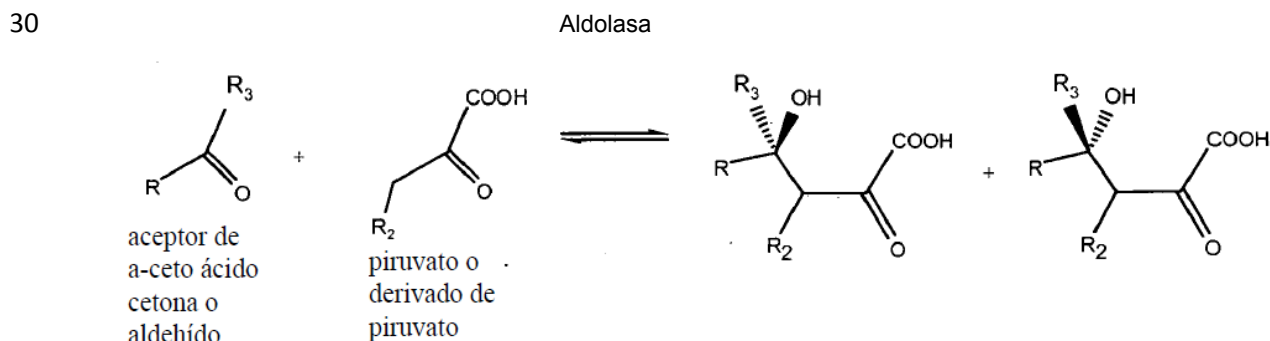
En algunas realizaciones, la invención proporciona polipéptidos aislados, sintéticos o recombinantes que tienen una identidad de secuencia de al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más o una identidad de secuencia u homología completa (100%) con una secuencia de acuerdo con la invención, tal como proteínas que tienen una secuencia tal como se establece en la SEQ ID NO:276 y fragmentos enzimáticamente activos de los mismos. La identidad de secuencia porcentual puede ser en el largo total del polipéptido.

Los polipéptidos de acuerdo con algunas realizaciones de la invención también pueden ser más cortos que el largo completo de los polipéptidos. En otras realizaciones, la divulgación proporciona polipéptidos (péptidos, fragmentos) en un rango de tamaño entre aproximadamente 5 y el largo completo de un polipéptido, tal como una enzima, tal como una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa; los tamaño ejemplares siendo de aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700 o más residuos, tales como residuos contiguos de una enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención. Los péptidos (tal como una subsecuencia de un polipéptido de acuerdo con la invención) pueden ser útiles como, por ejemplo, sondas de etiquetado, sitios activos de antígenos (inmunógenos), tolerágenos, motivos, aldolasa, tal como enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa (tal como "dominios catalíticos"), secuencias señal y/o dominios prepro.

En otras realizaciones, los polipéptidos de acuerdo con la invención que tienen actividad de aldolasa, tal como actividad de piruvato aldolasa, tal como HMG y/o KHG aldolasa son miembros de un género de polipéptidos que comparten elementos estructurales específicos, tales como residuos de aminoácidos, que se correlacionan con la actividad de aldolasa, incluyendo actividad de piruvato tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa. Estos elementos estructurados compartidos pueden usarse para la generación de rutina de aldolasa, tales como variantes de piruvato aldolasa, tal como HMG y/o KHG aldolasa. Estos elementos estructurales compartidos de la aldolasa, tal como enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa, de acuerdo con la invención, pueden utilizarse como guía para la generación de rutina de aldolasa, tales como variantes de piruvato aldolasa, tal como HMG y/o KHG aldolasa dentro del alcance del género de polipéptidos de acuerdo con la invención.

Tal como se utilizan en la presente, la expresión "aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa" comprende cualquier polipéptido o enzimas capaces de catalizar la reacción de adición de aldol o la reacción de retro-aldol (tal como polipéptidos de acuerdo con la invención, ver también la Tabla 1 y los Ejemplos 4, 5 y 6, más adelante) o cualquier modificación de un enlace carbono-carbono que contiene material, tal como en la producción de ácido R-2-hidroxi 2-(indol-3-ilmetil)-4-cetoglutarico (R-MP) y ciertos estereoisómeros de monatina, tales como monatina R,R y S,R, y sales de los mismos.

Los polipéptidos de acuerdo con algunas realizaciones de la invención catalizan la formación enlaces carbono-carbono en una reacción de aldol y tienen la capacidad de utilizar piruvato o fosfoenolpiruvato como el componente nucleofílico en la síntesis de un marco 4-hidroxi-2-cetobutirato como se muestra en el esquema general más adelante.

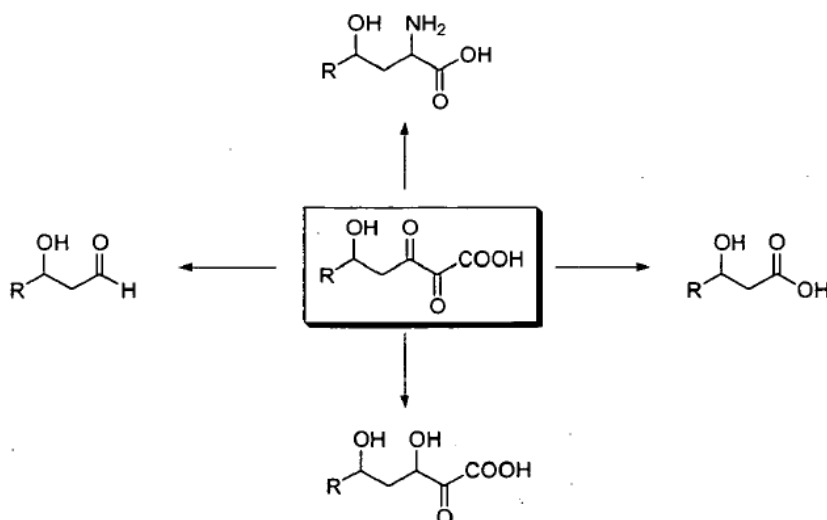


35

R=H, alquilo, alquilo sustituido, arilo, arilo sustituido, bencilo, bencilo sustituido  
R<sub>2</sub>=H, alquilo, alquilo sustituido, arilo, arilo sustituido, bencilo, bencilo sustituido  
R<sub>3</sub>=H, alquilo, alquilo sustituido, arilo, arilo sustituido, bencilo, bencilo sustituido, ácido carboxílico.

40

Sin ceñirse a ninguna teoría, se cree que el fragmento de cuatro carbonos conservado preparado en todas las condensaciones catalizadas por piruvato aldolasa es densamente y diferencialmente funcionalizado. Además, en cada aducto, cuatro estados de oxidación diferentes están contenidos en cuatro carbonos contiguos. El marco preparado por piruvato aldolasa permite entonces la preparación de ácidos α-amino-γ-hidroxicarboxílicos, ácidos β-hidroxicarboxílico, ácidos α,γ-dihidroxicarboxílicos, y azúcares de 2-desoxialdosa como se muestra en el esquema más adelante.



Por lo tanto, las piruvato aldolasas de acuerdo con algunas realizaciones de la invención pueden ser sintéticamente versátiles y pueden utilizarse en la preparación de una amplia gama de productos para usar en alimentos para animales, procesos industriales y agentes farmacéuticos (ver, por ejemplo, Gijzen, H. J. M. et al., Recent Advances in the Chemoenzymatic Synthesis of Carbohydrates and Carbohydrate Mimetics, Chem. Rev. 1996, 96, 443-473; Henderson, D. P. et al. J. Org. Chem., Stereospecific Preparation of the N-Terminal Amino Acid Moiety of Nikkomycins KX y KZ via a Multiple Enzyme Synthesis, 1997, 62, 7910-7911; Wymer, N. & Toone, E. J. Enzyme-catalyzed Synthesis of Carbohydrates. Current Opin. Chemical Biology, 2000, 4, 110-119).

Los polipéptidos de acuerdo con algunas realizaciones de la invención pueden tener más de un tipo de actividad enzimática, específicamente actividad de aldolasa y una actividad adicional, por ejemplo, tal como se establece en la Tabla 1, más adelante. Por ejemplo, un polipéptido de acuerdo con la invención puede tener actividad de aldolasa, piruvato aldolasa, actividad de HMG y/o KHG aldolasa. Adicionalmente, el polipéptido puede tener o puede pensarse que tiene actividad enzimática adicional en base a su clasificación CE. La Tabla 1 incluye la columna "Número de CE previsto". Un número de CE es el número asignado a un tipo de enzima de acuerdo con un esquema de nomenclatura de enzimas estandarizada desarrollado por la Comisión de Enzimas del Comité de Nomenclatura de la Unión Internacional de Bioquímica y Biología Molecular (IUBMB). Los resultados en la columna "Número de CE previsto" se determinan mediante un búsqueda de BLAST con respecto a la base de datos de Kegg (Enciclopedia de Kyoto de Genes y Genomas). Si la coincidencia BLAST máxima (también denominada un "acierto") tiene un valor E igual o menos que  $e^{-6}$ , el número de CE asignado a la coincidencia máxima se ingresa en la tabla. El número de CE del acierto máxima se usa como una guía de cómo puede ser el número CE de la secuencia de la invención. En casos en los que se proporciona sólo un número de CE parcial, sólo podría asignarse una clasificación amplia en base al acierto máximo. Por ejemplo, en la primera fila, para la SEQ ID NO:2, codificada por la SEQ ID NO:1, el "Número de CE Previsto se incluye como "2 . . .". Por lo tanto, la clasificación asignada es ampliamente una transferasa. Para la SEQ ID NO:26, codificada por la SEQ ID NO:25, la clasificación más específica que podría asignarse en base al acierto máximo es una aldehído-liasa.

Tabla 1

SEQ ID NO:	Actividad	Subclase de aldolasa	Número de CE previsto	SignalP Señal (AA = Aminoácido)	Fuente
1, 2	Aldolasa	HMG	2 . . .		Bacteria
3, 4	Aldolasa	HMG	2 . . .		Se desconoce
5, 6	Aldolasa	HMG	2 . . .		Se desconoce
7, 8	Aldolasa	HMG	2 . . .		Se desconoce
9, 10	Aldolasa	HMG	2 . . .		Se desconoce
11, 12	Aldolasa	HMG	2 . . .		Se desconoce

SEQ ID NO:	Actividad	Subclase de aldolasa	Número de CE previsto	SignalP Señal (AA = Aminoácido)	Fuente
13, 14	Aldolasa	HMG	2 . . .		Se desconoce
15, 16	Aldolasa	HMG	2 . . .		Se desconoce
17, 18	Aldolasa	HMG	2 . . .		Se desconoce
19, 20	Aldolasa	HMG	2 . . .		Se desconoce
21, 22	Aldolasa	HMG			Se desconoce
23, 24	Aldolasa	HMG	2 . . .		Se desconoce
25, 26	Aldolasa	HMG	4,1,2.		Se desconoce
27, 28	Aldolasa	HMG	2 . . .		Se desconoce
29, 30	Aldolasa	HMG	2 . . .		Se desconoce
31, 32	Aldolasa	HMG	2 . . .		Se desconoce
33, 34	Aldolasa	HMG	2 . . .		Se desconoce
35, 36	Aldolasa	HMG	2 . . .		Se desconoce
37, 38	Aldolasa	HMG	2 . . .		Se desconoce
39, 40	Aldolasa	HMG	2 . . .		Se desconoce
41, 42	Aldolasa	HMG	2 . . .		Se desconoce
43, 44	Aldolasa	HMG	2 . . .		Se desconoce
45, 46	Aldolasa	HMG	2 . . .		Se desconoce
47, 48	Aldolasa	HMG	2 . . .		Se desconoce
49, 50	Aldolasa	HMG	2 . . .		Se desconoce
51, 52	Aldolasa	HMG	2 . . .		Se desconoce
53, 54	Aldolasa	HMG	2 . . .		Se desconoce
55, 56	Aldolasa	HMG	2 . . .		Se desconoce
57, 58	Aldolasa	HMG	2 . . .		Se desconoce
59, 60	Aldolasa	HMG	2 . . .		Se desconoce
61, 62	Aldolasa	HMG	2 . . .		Se desconoce
63, 64	Aldolasa	HMG	2 . . .		Se desconoce
65, 66	Aldolasa	HMG	2 . . .	AA1-27	Se desconoce
67, 68	Aldolasa	HMG	2 . . .		Se desconoce
69, 70	Aldolasa	HMG	2 . . .		Se desconoce



SEQ ID NO:	Actividad	Subclase de aldolasa	Número de CE previsto	SignalP Señal (AA = Aminoácido)	Fuente
71, 72	Aldolasa	HMG	2 . . .		Se desconoce
73, 74	Aldolasa	HMG	2 . . .		Se desconoce
75, 76	Aldolasa	HMG	2 . . .		Se desconoce
77, 78	Aldolasa	HMG	2 . . .		Se desconoce
79, 80	Aldolasa	HMG	2 . . .		Se desconoce
81, 82	Aldolasa	HMG	2 . . .		Se desconoce
83, 84	Aldolasa	HMG	2 . . .		Se desconoce
85, 86	Aldolasa	HMG	2 . . .		Se desconoce
87, 88	Aldolasa	HMG	2 . . .		Se desconoce
89, 90	Aldolasa	HMG	2 . . .		Se desconoce
91, 92	Aldolasa	HMG	2 . . .		Se desconoce
93, 94	Aldolasa	HMG	2 . . .		Se desconoce
95, 96	Aldolasa	HMG	2 . . .		Se desconoce
97, 98	Aldolasa	HMG	2 . . .		Se desconoce
99, 100	Aldolasa	HMG	2 . . .		Se desconoce
101, 102	Aldolasa	HMG	2 . . .		Se desconoce
103, 104	Aldolasa	HMG	2 . . .		Se desconoce
105, 106	Aldolasa	HMG	2 . . .		Se desconoce
107, 108	Aldolasa	HMG	2 . . .		Se desconoce
109, 110	Aldolasa	HMG	2 . . .		Se desconoce
111, 112	Aldolasa	HMG	2 . . .		Se desconoce
113, 114	Aldolasa	HMG	2 . . .		Se desconoce
115, 116	Aldolasa	HMG	2 . . .		Se desconoce
117, 118	Aldolasa	HMG	2 . . .		Se desconoce
119, 120	Aldolasa	HMG	2 . . .		Se desconoce
121, 122	Aldolasa	HMG	2 . . .		Se desconoce
123, 124	Aldolasa	HMG	2 . . .		Se desconoce
125, 126	Aldolasa	HMG	2 . . .		Se desconoce
127, 128	Aldolasa	HMG	2 . . .		Se desconoce

SEQ ID NO:	Actividad	Subclase de aldolasa	Número de CE previsto	SignalP Señal (AA = Aminoácido)	Fuente
129, 130	Aldolasa	HMG	2 . . .		Se desconoce
131, 132	Aldolasa	HMG	2 . . .		Se desconoce
133, 134	Aldolasa	HMG	2 . . .		Se desconoce
135, 136	Aldolasa	HMG	2 . . .		Se desconoce
137, 138	Aldolasa	HMG	2 . . .		Se desconoce
139, 140	Aldolasa	HMG	2 . . .		Se desconoce
141, 142	Aldolasa	HMG	2 . . .		Se desconoce
143, 144	Aldolasa	HMG	2 . . .		Se desconoce
145, 146	Aldolasa	HMG	2 . . .		Se desconoce
147, 148	Aldolasa	HMG	2 . . .		Se desconoce
149, 150	Aldolasa	HMG	2 . . .		Se desconoce
151, 152	Aldolasa	HMG	2 . . .		Se desconoce
153, 154	Aldolasa	HMG	2 . . .		Se desconoce
155, 156	Aldolasa	HMG	2 . . .		Se desconoce
157, 158	Aldolasa	HMG	2 . . .		Se desconoce
159, 160	Aldolasa	HMG	2 . . .		Se desconoce
161, 162	Aldolasa	HMG	2 . . .		Se desconoce
163, 164	Aldolasa	HMG	2 . . .		Se desconoce
165, 166	Aldolasa	HMG	2 . . .		Se desconoce
167, 168	Aldolasa	HMG	2 . . .		Se desconoce
169, 170	Aldolasa	HMG	2 . . .		Se desconoce
171, 172	Aldolasa	HMG	2 . . .		Se desconoce
173, 174	Aldolasa	HMG	2 . . .		Se desconoce
175, 176	Aldolasa	HMG	2 . . .		Se desconoce
177, 178	Aldolasa	HMG	2 . . .		Se desconoce
179, 180	Aldolasa	HMG	2 . . .		Se desconoce
181, 182	Aldolasa	HMG	2 . . .	AA1-31	Se desconoce
183, 184	Aldolasa	HMG	2 . . .		Se desconoce
185, 186	Aldolasa	HMG	2 . . .		Se desconoce

<b>SEQ ID NO:</b>	<b>Actividad</b>	<b>Subclase de aldolasa</b>	<b>Número de CE previsto</b>	<b>SignalP Señal (AA = Aminoácido)</b>	<b>Fuente</b>
187, 188	Aldolasa	HMG	2 . . .		Se desconoce
189, 190	Aldolasa	HMG	2 . . .		Se desconoce
191, 192	Aldolasa	HMG	2 . . .		Se desconoce
193, 194	Aldolasa	HMG	2 . . .		Se desconoce
195, 196	Aldolasa	HMG	2 . . .		Se desconoce
197, 198	Aldolasa	HMG	2 . . .		Se desconoce
199, 200	Aldolasa	HMG	2 . . .		Se desconoce
201, 202	Aldolasa	HMG	2 . . .		Se desconoce
203, 204	Aldolasa	HMG	2 . . .		Se desconoce
205, 206	Aldolasa	HMG	2 . . .		Se desconoce
207, 208	Aldolasa	HMG	2 . . .		Se desconoce
209, 210	Aldolasa	HMG	2 . . .		Se desconoce
211, 212	Aldolasa	HMG	2 . . .		Se desconoce
213, 214	Aldolasa	HMG	2 . . .		Se desconoce
215, 216	Aldolasa	HMG	2 . . .		Se desconoce
217, 218	Aldolasa	HMG	2 . . .		Se desconoce
219, 220	Aldolasa	HMG	2 . . .		Se desconoce
221, 222	Aldolasa	HMG	2 . . .		Se desconoce
223, 224	Aldolasa	HMG	2 . . .		Se desconoce
225, 226	Aldolasa	HMG	2 . . .		Se desconoce
227, 228	Aldolasa	HMG	2 . . .		Se desconoce
229, 230	Aldolasa	HMG	2 . . .		Se desconoce
231, 232	Aldolasa	HMG	2 . . .		Se desconoce
233, 234	Aldolasa	HMG	2 . . .		Se desconoce
235, 236	Aldolasa	HMG	2 . . .		Se desconoce
237, 238	Aldolasa	HMG	2 . . .		Se desconoce
239, 240	Aldolasa	HMG	2 . . .		Se desconoce
241, 242	Aldolasa	HMG	2 . . .		Se desconoce
243, 244	Aldolasa	HMG	2 . . .		Se desconoce

SEQ ID NO:	Actividad	Subclase de aldolasa	Número de CE previsto	SignalP Señal (AA = Aminoácido)	Fuente
245, 246	Aldolasa	HMG	2 . . .		Se desconoce
247, 248	Aldolasa	HMG	2 . . .		Se desconoce
249, 250	Aldolasa	HMG	2 . . .		Se desconoce
251, 252	Aldolasa	HMG	2 . . .		Se desconoce
253, 254	Aldolasa	HMG	2 . . .		Se desconoce
255, 256	Aldolasa	HMG	2 . . .		Se desconoce
257, 258	Aldolasa	HMG	2 . . .		Se desconoce
259, 260	Aldolasa	HMG	2 . . .	AA1-18	Se desconoce
261, 262	Aldolasa	HMG	2 . . .		Se desconoce
263, 264	Aldolasa	HMG	2 . . .		Se desconoce
265, 266	Aldolasa	HMG	2 . . .		Se desconoce
267, 268	Aldolasa	HMG	2 . . .		Se desconoce
269, 270	Aldolasa	HMG	2 . . .		Se desconoce
271, 272	Aldolasa	HMG	2 . . .		Se desconoce
273, 274	Aldolasa	HMG	2 . . .		Se desconoce
275, 276	Aldolasa	HMG	2 . . .		Se desconoce
277, 278	Aldolasa	HMG	2 . . .		Se desconoce
279, 280	Aldolasa	HMG	2 . . .		Se desconoce
281, 282	Aldolasa	HMG	2 . . .		Se desconoce
283, 284	Aldolasa	HMG	2 . . .		Se desconoce
285, 286	Aldolasa	HMG	2 . . .		Se desconoce
287, 288	Aldolasa	HMG	2 . . .		Se desconoce
289, 290	Aldolasa	HMG	2 . . .		Se desconoce
291, 292	Aldolasa	HMG	2 . . .		Se desconoce
293, 294	Aldolasa	HMG	2 . . .		Se desconoce
295, 296	Aldolasa	HMG	2 . . .		Se desconoce
297, 298	Aldolasa	HMG	2 . . .		Se desconoce
299, 300	Aldolasa	HMG	2,1 . . .		Se desconoce
301, 302	Aldolasa	HMG	2,1 . . .		Se desconoce

SEQ ID NO:	Actividad	Subclase de aldolasa	Número de CE previsto	SignalP Señal (AA = Aminoácido)	Fuente
303, 304	Aldolasa	HMG	2,1 . . .		Se desconoce
305, 306	Aldolasa	KHG	4,1,2,14		Se desconoce
307, 308	Aldolasa	KHG	4,1,2,14		Se desconoce
309, 310	Aldolasa	KHG	4,1,2,14		Se desconoce
311, 312	Aldolasa	KHG	4,1,2,14		Se desconoce
313, 314	Aldolasa	KHG	4,1,2,14		Se desconoce
315, 316	Aldolasa	KHG	4,1,2,14		Se desconoce
317, 318	Aldolasa	KHG	4,1,2,14		Se desconoce
319, 320	Aldolasa	KHG	4,1,3,16		Se desconoce
321, 322	Aldolasa	KHG	4,1,2,14		Se desconoce
323, 324	Aldolasa	KHG	4,1,2,14		Se desconoce
325, 326	Aldolasa	KHG	4,1,2,14		Se desconoce
327, 328	Aldolasa	KHG	4,1,2,14		Se desconoce
329, 330	Aldolasa	KHG	4,1,3,16		Se desconoce
331, 332	Aldolasa	KHG	4,1,2,14		Se desconoce
333, 334	Aldolasa	KHG	4,1,2,14		Se desconoce

5 Polipéptidos y péptidos de acuerdo con la invención pueden aislarse de fuentes naturales, pueden ser polipéptidos sintéticos o generarse recombinantemente. Péptidos y proteínas pueden expresarse recombinantemente in vitro o in vivo. Los péptidos y polipéptidos de acuerdo con la invención pueden realizarse y aislarse usando cualquier método conocido en la técnica. También pueden usarse polipéptidos y péptidos de acuerdo con la invención, en total o en parte, utilizando métodos químicos bien conocidos en la técnica. Ver 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, las síntesis de péptidos pueden realizarse utilizando varias técnicas de fase sólida (ver por ejemplo Roberge (1995) Science 269:202; Merrifield (1997) Methods Enzymol. 289:3-13) y la síntesis automatizada puede lograrse, por ejemplo utilizando el Sintetizador de Péptidos 431A (Perkin Elmer) de acuerdo con las instrucciones proporcionadas por el fabricante.

15 Los péptidos y polipéptidos de acuerdo con la invención también pueden ser glicosilados. La glicosilación puede incorporarse post-translacionalmente ya sea químicamente o por medio de mecanismos biosintéticos celulares, en donde los últimos incorporan el uso de motivos de glicosilación conocidos, que pueden ser nativos a la secuencia o que pueden incorporarse como un péptido o agregarse a la secuencia de ácidos nucleicos de codificación. La glicosilación puede estar O-enlazada o N-enlazada.

20 En algunas realizaciones, cuando se indica, los péptidos y polipéptidos de acuerdo con la invención pueden incluir todas las formas "miméticas" y "peptidomiméticas". Los términos "miméticas" y "peptidomiméticas" se refieren a un compuesto químico sintético que tiene básicamente las mismas características estructurales y/o funcionales de los polipéptidos de acuerdo con la invención. Las formas miméticas pueden estar totalmente compuestas por análogos de aminoácidos sintéticos, no naturales o ser una molécula quimérica de aminoácidos peptídicos parcialmente naturales y análogos de aminoácidos parcialmente no naturales. Las formas miméticas pueden incorporar también cualquier cantidad de sustituciones de aminoácidos naturales conservadoras en la medida en que dichas

sustituciones no alteren también básicamente la estructura y/o actividad mimética. Al igual que con los polipéptidos de acuerdo con la invención que son variantes conservadoras o miembros de un género de polipéptidos de acuerdo con la invención (que tienen por ejemplo aproximadamente 50% o más identidad de secuencia con una secuencia de acuerdo con la invención), los experimentos de rutina determinarán si un mimético está dentro del alcance de acuerdo con la invención, es decir, si su estructura y/o función básicamente no se altera. De esta forma, en algunas realizaciones, una composición mimética está dentro del alcance de acuerdo con la invención si tiene una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa.

Composiciones miméticas de polipéptidos de acuerdo con la invención pueden contener cualquier combinación de componentes estructurales no naturales. En una realización alternativa, las composiciones miméticas de acuerdo con la invención incluyen uno o todos los siguientes tres grupos estructurales: a) grupos de enlace de residuos que no sean el enlace de amida natural ("enlace peptídico"); b) residuos no naturales en lugar de residuos aminoácidos naturales; o c) residuos que inducen un mimetismo estructural secundario, es decir, para inducir o estabilizar una estructura secundaria, tal como una conformación de giro beta, giro gama, hoja beta, hélice alfa y similares. Por ejemplo, un polipéptido de acuerdo con la invención puede caracterizarse como un mimético cuando todos o algunos de sus residuos están unidos por medios químicos que no sean enlaces peptídicos naturales. Los residuos peptidomiméticos individuales pueden estar unidos por enlaces peptídicos, otros enlaces químicos o medios de acoplamiento, tal como glutaraldehído, ésteres de N-hidroxisuccinimida, maleimidias bifuncionales, N,N'-diclohexilcarbodiimida (DCC) o N,N'-diisopropilcarbodiimida (DIC). Los grupos ligantes que pueden ser una alternativa al enlace amida tradicional ("enlace peptídico") incluyen, por ejemplo cetometileno (tal como  $\text{—C(=O)—CH}_2\text{—}$  por  $\text{—C(=O)—NH—}$ ), aminometileno ( $\text{CH}_2\text{—NH—}$ ), etileno, olefina ( $\text{CH=CH—}$ ), éter ( $\text{CH}_2\text{—O—}$ ), tioéter ( $\text{CH}_2\text{—S—}$ ), tetrazol ( $\text{CN}_4\text{—}$ ), tiazol, retroamida, tioamida o éster (ver Spatola (1983) en Química y Bioquímica de los Aminoácidos, Péptidos y Proteínas, Vol. 7, páginas 267-357, "Peptide Backbone Modifications," Marcell Dekker, NY).

Un polipéptido de acuerdo con la invención también puede caracterizarse como un mimético por contener todos o algunos de los residuos en lugar de residuos aminoácidos naturales. Residuos no naturales se describen en la biblioteca científica y de patentes; algunas composiciones no naturales ejemplares útiles como miméticos de residuos de aminoácidos naturales y lineamientos se describen más adelante. Los miméticos de aminoácidos aromáticos pueden generarse reemplazando por, por ejemplo, D- o L-nafilalanina; D- o L-fenilglicina; D- o L-2-tienilalanina; D- o L-1, -2, 3- o 4-pirenilalanina; D- o L-3-tienilalanina; 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)-fenilglicina; 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, donde el alquilo puede ser metilo, etilo, propilo, hexilo, butilo, pentilo, isopropilo, iso-butilo, sec-isotilo, iso-pentilo sustituido o insustituido o aminoácidos no ácidos. Los anillos aromáticos de un aminoácido no natural incluyen, por ejemplo, anillos aromáticos de tiazolilo, tiofenilo, porazolilo, bencimidazolilo, naftilo, furanilo, pirrolilo y piridilo.

Los miméticos de aminoácidos ácidos pueden generarse por sustitución por, por ejemplo, aminoácidos no carboxilato, manteniendo igual una carga negativa; (fosfona)alanina; treonina sulfatada. Grupos laterales carboxilo (tales como aspartilo o glutamilo) también pueden modificarse selectivamente por reacción con carbodiimidias ( $\text{R'—N—C—N—R'}$ ) tal como 1-ciclohexil-3(2-morfolinil-(4-etil) carbodiimida o 1-etil-3(4-azonia-4,4-dimetolpentil) carbodiimida. También pueden convertirse el aspartilo o glutamilo en residuos de asparaginilo y glutaminilo por reacción con iones de amonio. Miméticos de aminoácidos básicos pueden generarse por sustitución con, por ejemplo (además de lisina y arginina) los aminoácidos ornitina, citrulina o ácido (guanidino)-acético, o ácido (guanidino)alquil-acético, donde el alquilo se define anteriormente. El derivado de nitrilo (que contiene por ejemplo el resto CN en lugar de COOH) puede sustituirse por asparagina o glutamina. Los residuos de asparaginilo o glutamilo pueden desaminarse a los residuos correspondientes de aspartilo o glutamilo. Los miméticos de residuos de arginina pueden generarse haciendo reaccionar el arginil con por ejemplo uno o más reactivos convencionales, incluidos, por ejemplo fenilgloxal, 2,3-butanodiona, 1,2-ciclo-hexanodiona o ninhidrina, en algunas realizaciones en condiciones alcalinas. Los miméticos de residuos de tirosina pueden generarse haciendo reaccionar tirosilo con, por ejemplo compuestos de diazonio aromáticos o tetranitrometano, N-acetilimidazol y tetranitrometano pueden utilizarse para formar especies de O-acetil tirosilo y derivados de 3-nitro, respectivamente. Miméticos de residuos de cisteína pueden generarse haciendo reaccionar residuos de cistenilo con, por ejemplo alfa-haloacetatos tal como ácido 2-cloroacético o cloroacetamida y aminoras correspondientes; para proporcionar derivados de carboximetilo o carboxiamidometilo. Los miméticos de residuos de cisteína también pueden generarse haciendo reaccionar residuos de cisteinilo con, por ejemplo bromo-trifluoroacetona, ácido alfa-bromo-beta-(5-imidozoi) propiónico; cloroacetil fosfato, N-alquilmaleimidias, 3-nitro-2-piridil disulfuro; metil 2-piridil disulfuro; p-cloromercuribenzoato; 2-cloromercuri-4 nitrofenol; o, cloro-7-nitrobenzo-oxa-1,3-diazol. Los miméticos de lisina pueden generarse (y los residuos de amino terminales pueden alterarse) haciendo reaccionar lisinilo con, por ejemplo anhídridos succínicos y otros de ácido carboxílico. Lisina y otros miméticos de residuos que contienen alfa-amino también pueden generarse por medio de la reacción con imidoésteres, tales como metil picolinimidato piridoxal fosfato, piridoxal, cloroborohidruro, ácido trinitro-bencenosulfónico, O-metilisourea, 2,4, pentanodiona, y reacciones catalizadas por transamidasa con glioxilato. Miméticos de metionina pueden generarse por medio de reacción con, por ejemplo sulfóxido de metionina. Miméticos de prolina incluyen, por ejemplo, ácido pipercolico, ácido carboxílico de tiazolidina, 3- o 4-hidroxi prolina, deshidroprolina, 3- o 4-metilprolina o 3,3,-dimetilprolina. Los miméticos de residuos de histidina pueden generarse

5 haciendo reaccionar histidilo con, por ejemplo, dietilprocarbonato o bromuro de para-bromofenacilo. Otros miméticos incluyen, por ejemplo aquellos generados por la hidroxilación de prolina y lisina; fosforilación de los grupos hidroxilo de residuos de serilo o treonilo; metilación de los grupos alfa-amino de lisina, arginina e histidina; acetilación de la amina N-terminal; metilación de los residuos amida de cadena principal o sustitución con N-metilaminoácidos; o amidación de los grupos de carboxilo C-terminal.

10 En algunas realizaciones, un residuo, tal como un aminoácido, de un polipéptido de acuerdo con la invención puede reemplazarse por un aminoácido (o residuo peptidomimético) de quiralidad opuesta. En algunas realizaciones, cualquier aminoácido natural en la configuración L (a la que también puede hacerse referencia como R o S, dependiendo de la estructura de la entidad química) puede reemplazarse con el aminoácido del mismo tipo estructural químico o un peptidomimético, pero de quiralidad opuesta, denominado como el D-aminoácido, pero también puede denominarse como la forma R o S.

15 La invención también proporciona métodos para modificar los polipéptidos de acuerdo con la invención mediante procesos naturales, tal como procesamiento post-translacional (tal como fosforilación, acilación, etc.) o por técnicas de modificación química, y los péptidos modificados resultantes. Las modificaciones pueden ocurrir en cualquier parte en el polipéptido, incluida la estructura central del péptido, las cadenas de aminoácidos y las terminales amino o carboxilo. Se apreciará que el mismo tipo de modificación puede estar presente en el mismo grado o en grados variables en varios sitios en un polipéptido dado. También un polipéptido dado puede tener muchos tipos de modificaciones. En algunas realizaciones, las modificaciones incluyen acetilación, acilación, ADP-ribosilación, amidación, unión covalente de flavina, unión covalente de un resto hemo, unión covalente de un nucleótido o derivado de nucleótido, unión covalente de un lípido o derivado de lípido, unión covalente de un fosfatidilinositol, ciclización reticulante, formación de enlaces disulfuro, desmetilación, formación de enlaces intermoleculares covalentes, formación de cisteína, formación de piroglutamato, formilación, carboxilación gama, glicosilación, formación de anclajes GPI, hidroxilación, yodación, metilación, miristoilación, oxidación, pegilación, procesamiento proteolítico, fosforilación, prenilación, racemización, selenoilación, sulfatación y adición mediada por ARN de transferencia de aminoácidos a proteína tal como arginilación. Ver Creighton, T.E., *Proteins - Structure and Molecular Properties* 2da Ed., W.H. Freeman and Company, Nueva York (1993); *Posttranslational Covalent Modification of Proteins*, B. C. Johnson, Ed., Academic Press, Nueva York, págs. 1-12 (1983).

30 Los métodos de síntesis peptídica química de fase sólida también pueden usarse para sintetizar el polipéptido o fragmentos de acuerdo con la invención. Dicho método se ha conocido en la técnica desde principios de los años sesenta (Merrifield, R. B., *J. Am. Chem. Soc.*, 85:2149-2154, 1963) (Ver también Stewart, J. M. y Young, J. D., *Solid Phase Peptide Synthesis*, 2da Ed., Pierce Chemical Co., Rockford, Ill., pp. 11-12)) y se empleó recientemente en kits de diseño y síntesis de laboratorio disponibles en el comercio (Cambridge Research Biochemicals). Dichos kits de laboratorio disponibles comercialmente han utilizado generalmente los descubrimientos de H. M. Geysen *et al.*, *Proc. Natl. Acad. Sci., Estados Unidos*, 81:3998 (1984) y proporcionan péptidos de sintetización sobre las puntas de múltiples "varillas" o "clavijas", todas las cuales están conectadas a una única placa. Cuando se utiliza un sistema tal, una placa de varillas o clavijas se invierte y se inserta en una segunda placa de pocillos o depósitos correspondientes, que contienen soluciones para unir o anclar un aminoácido apropiado a las puntas de las varillas o clavijas. Repitiendo dicha etapa del proceso, es decir, invirtiendo e insertando las puntas de las varillas o clavijas en soluciones apropiados, los aminoácidos se incorporan en péptidos deseados. Además, están disponibles varios sistemas de síntesis peptídica de Fmoc disponible. Por ejemplo, el ensamblaje de un polipéptido o fragmento puede llevarse a cabo sobre un soporte sólido utilizando un sintetizador automatizado de Applied Biosystems, Inc. Modelo 431A™. Dicho equipo proporciona fácil acceso a los péptidos de acuerdo con la invención, ya sea por síntesis directa o por síntesis de varios fragmentos que pueden acoplarse utilizando otras técnicas conocidas.

45 Los polipéptidos de acuerdo con la invención incluyen enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa en forma activa o inactiva. Por ejemplo, los polipéptidos de acuerdo con la invención incluyen proproteínas antes de su "maduración" o procesamiento de secuencias prepro, tal como por medio de una enzima de procesamiento de proproteína, tal como una proproteína convertasa para generar una proteína madura "activa". Los polipéptidos de acuerdo con la invención incluyen enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa inactiva por otras razones, tal como antes de la "activación" por medio de un acontecimiento de procesamiento post-translacional, tal como una acción de endo- o exo-peptidasa o proteinasa, un acontecimiento de fosforilación, una amidación, una glicosilación o una sulfatación, un acontecimiento de dimerización y similares. Los polipéptidos de acuerdo con la invención incluyen todas las formas activas, incluidas las subsecuencias activas, tales como dominios catalíticos o sitios activos, de la enzima.

55 La invención incluye enzima aldolasa inmovilizada, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, anti-aldolasa, tal como anti-piruvato aldolasa, tal como anticuerpos anti-HMG y/o anti-KHG aldolasa y fragmentos de los mismos. En algunas realizaciones, la invención proporciona métodos para inhibir la actividad de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, por ejemplo usando mutantes negativos dominantes o anti-aldolasa, tal como anti-piruvato aldolasa, tal como anticuerpos anti-HMG y/o anti-KHG aldolasa de acuerdo con la invención. En algunas realizaciones, la invención incluye heterocomplejos, tales como proteínas de

fusión, heterodímeros, etc., que comprenden la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención.

5 En algunas realizaciones, los polipéptidos de acuerdo con la invención pueden tener una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, en diferentes condiciones, tal como en pH y/o temperatura extremos o, en algunas realizaciones, en presencia de agentes oxidantes. En algunas realizaciones, la invención proporciona métodos que producen la enzima aldolasa alternativa, tal como piruvato aldolasa, tal como preparaciones de HMG y/o KHG aldolasa con diferentes eficiencias y estabildades catalíticas, tal como hacia temperatura, agentes oxidantes y condiciones de lavado cambiantes. En algunas realizaciones, variantes de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa pueden producirse utilizando técnicas de mutagénesis dirigida y/o mutagénesis aleatoria. En algunas realizaciones, la evolución dirigida puede utilizarse para producir una gran variedad de enzima aldolasa, tales como variantes de la enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa con especificidades y estabilidad alternativas.

15 Las proteínas de acuerdo con la invención también son útiles como reactivos de investigación para identificar moduladores de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, tales como activadores o inhibidores de la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. Brevemente, las muestras de prueba (compuestos, caldos, extractos y similares) se incorporan a los ensayos de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa para determinar su capacidad para inhibir la escisión del sustrato. Los inhibidores identificados de esta forma pueden utilizarse en la industria y en la investigación para reducir o evitar la proteólisis indeseada. Tal como con aldolasa, tales como enzimas piruvato aldolasa, tal como HMG y/o KHG aldolasa, los inhibidores pueden combinarse para aumentar el espectro de actividad.

20 Las enzimas de acuerdo con la invención también son útiles como reactivos de investigación para digerir proteínas o para la secuenciación de proteínas. Por ejemplo, la aldolasa, tal como enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa pueden usarse para romper polipéptidos en fragmentos más pequeños para secuenciación usando, por ejemplo, un secuenciador automatizado.

25 La invención también proporciona métodos para descubrir una nueva aldolasa, tal como enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa utilizando los ácidos nucleicos, polipéptidos y anticuerpos de acuerdo con la invención. En algunas realizaciones, las bibliotecas de fagémidos se evalúan en busca del descubrimiento en base a expresión de aldolasa, tal como enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa. En otras realizaciones, se evalúan las bibliotecas de fagos lambda en busca del descubrimiento en base a expresión de la aldolasa, tal como enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa. La evaluación de las bibliotecas de fagos o fagémidos puede permitir la detección de clones tóxicos; mejor acceso al sustrato; necesidad reducida de manipular un huésped, evitar la posibilidad de cualquier sesgo que resulte de la supresión masiva de la biblioteca; y, un crecimiento más rápido en densidades de clones bajas. La evaluación de las bibliotecas de fagos o fagémidos puede ser en fase líquida o en fase sólida. En algunas realizaciones, la invención proporciona la evaluación en fase líquida. Esto proporciona una mayor flexibilidad en condiciones de ensayo; flexibilidad adicional del sustrato; sensibilidad más alta para clones débiles; y facilidad de automatización durante la evaluación de fase sólida.

30 La invención proporciona métodos de evaluación que usan las proteínas y ácidos nucleicos de acuerdo con la invención y la automatización robótica para permitir la ejecución de miles de reacciones biocatalíticas y ensayos de detección en un corto periodo de tiempo, tal como por día, así como asegurar un alto nivel de precisión y reproducibilidad (ver la descripción de los arreglos, más adelante). Como resultado, puede producirse una biblioteca de compuestos derivados en cuestión de semanas. Para más información sobre la modificación de las moléculas, incluidas molécula pequeñas, ver PCT/US94/09174; la Patente de los Estados No. 6.245.547.

35 En algunas realizaciones, los polipéptidos o fragmentos de acuerdo con la invención se obtienen a través de los procedimientos de enriquecimiento bioquímico o purificación. La secuencia de polipéptidos o fragmentos homólogos pueden determinarse por ensayos de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa (ver los Ejemplos 3, 4 y 5, más adelante), electroforesis de gel y/o microsecuenciación. La secuencia del posible polipéptido o fragmento de acuerdo con la invención puede compararse con un polipéptido de acuerdo con la invención o un fragmento, tal como que comprende al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 50 100 o 150 o más aminoácidos consecutivos de los mismos usando cualquiera de los programas descritos anteriormente.

Otra realización de la invención es un ensayo para identificar fragmentos o variantes de acuerdo con la invención, que retienen la función enzimática de los polipéptidos de acuerdo con la invención. Por ejemplo los fragmentos o variantes de dichos polipéptidos, pueden usarse para catalizar reacciones bioquímicas, que indican que el fragmento o variante retiene la actividad enzimática de un polipéptido de acuerdo con la invención. Un ejemplo de un ensayo para determinar si los fragmentos de las variantes retienen la actividad enzimática de los polipéptidos de acuerdo con la invención incluye las etapas de: poner en contacto el fragmento de polipéptido o variante con una molécula de sustrato en condiciones que permiten que el fragmento del polipéptido o variante funcione y detectar una



disminución en el nivel de sustrato o un aumento en el nivel del producto de reacción específico de la reacción entre el polipéptido y sustrato.

5 La presente invención explota las propiedades catalíticas únicas de las enzimas. Mientras que el uso de biocatalizadores (es decir, enzimas purificadas o brutas, células vivas o no vivas) en transformaciones químicas normalmente requiere la identificación de un biocatalizador particular que reacciona con un compuesto de partida específico, la presente invención usa biocatalizadores seleccionados y condiciones de reacción que son específicas para grupos funcionales que están presentes en muchos compuestos de partida, tales como molécula 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.

10 En algunas realizaciones, las reacciones biocatalíticas producen una población de derivados de un compuesto de partida único. Estos derivados pueden someterse a otra ronda de reacciones biocatalíticas para producir una segunda población de compuestos derivados. Miles de variaciones de la molécula pequeña original o compuesto pueden producirse con cada iteración de derivatización biocatalítica.

15 Las enzimas reaccionan en sitios específicos de un compuesto de partida sin afectar el resto de la molécula, un proceso que es difícil de lograr utilizando métodos químicos tradicionales. Este alto grado de especificidad biocatalítica proporciona los medios para identificar un único compuesto activo dentro de la biblioteca. La biblioteca se caracteriza por la serie de reacciones biocatalíticas utilizadas para producirlo, un denominado "historial biosintético". La evaluación de la biblioteca en busca de actividades biológicas y la localización del historial biosintético identifica la secuencia de reacción específica que produce el compuesto activo. La secuencia de reacción se repite y la estructura del compuesto sintetizado determinado. Este modo de identificación, a diferencia de otras síntesis y abordajes de detección, no requiere las tecnologías de inmovilización y los compuestos pueden sintetizarse y evaluarse libres en la solución utilizando virtualmente cualquier tipo de ensayo de detección. Cabe señalar que el alto grado de especificidad de las reacciones enzimáticas en grupos funcionales permite la "localización" de reacciones enzimáticas específicas que forman la biblioteca producida biocatalíticamente.

25 En algunas realizaciones, las etapas del procedimiento se realizan utilizando automatización robótica que permite la ejecución de miles de reacciones biocatalíticas y/o ensayos de detección por día así como asegurar un alto nivel de precisión y reproducibilidad. La automatización robótica también puede usarse para evaluar la actividad de aldolasa para determinar si un polipéptido está dentro del alcance acuerdo con la invención. Como resultado, en algunas realizaciones, puede producirse una biblioteca de compuestos derivados en cuestión de semanas que tomaría años producir utilizando métodos de detección química o enzimática "tradicionales".

30 En una realización, la invención proporciona métodos para modificar molécula pequeñas, que comprenden poner en contacto un polipéptido codificado por un polinucleótido descrito en la presente o fragmentos enzimáticamente activos de los mismos con una molécula pequeña para producir una molécula pequeña modificada. Se evalúa una biblioteca de moléculas pequeñas modificadas para determinar si una molécula pequeña modificada se encuentra presente en la biblioteca, que exhibe una actividad deseada. Una reacción biocatalítica específica que produce la molécula pequeña modificada de actividad deseada eliminando sistemáticamente cada una de las reacciones biocatalíticas utilizadas para producir una porción de la biblioteca y evaluar luego la molécula pequeña producida en la porción de la biblioteca para determinar 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 opcionalmente se repite. Las reacciones biocatalíticas se llevan a cabo con un grupo de biocatalizadores que reaccionan con restos estructurales diferentes que se encuentran dentro de la estructura de una molécula pequeña, cada uno de los biocatalizadores es específico para un resto estructural o un grupo de restos estructurales relacionados; y cada uno de los biocatalizadores reaccionan con muchas moléculas pequeñas diferentes que contienen el resto estructural distinto.

45 *Secuencias señal de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, dominios prepro y catalíticos*

50 La invención proporciona secuencias señal de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa (tales como péptidos señal (SP)), dominios prepro y dominios catalíticos (CD). Los SP, los dominios prepro y/o CD de acuerdo con la invención pueden ser péptidos aislados, sintéticos o recombinantes o pueden ser parte de una proteína de fusión, tal como un dominio heterólogo en una proteína quimérica. En algunas realizaciones, la invención proporciona ácidos nucleicos que codifican estos dominios catalíticos (CD), dominios prepro y secuencias señal (SP, tal como un péptido que tiene una secuencia que comprende/que consiste en residuos amino terminales de un polipéptido de acuerdo con la invención).

55 La invención proporciona secuencias señal aisladas, sintéticas o recombinantes (tales como péptidos señal) que consisten en o que comprenden una secuencia como se establece en los residuos 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, 1 a 44 de un polipéptido de acuerdo con la invención, tal como polipéptidos de acuerdo con la invención, ver también la Tabla 1, los Ejemplos 4, 5 y 6, más

adelante, y el Listado de Secuencias. Por ejemplo, la Tabla 1, anteriormente, establece secuencias señal (líderes) ejemplares de acuerdo con la invención, tal como en el polipéptido que tiene una secuencia como se establece en la SEQ ID NO:66, codificada, por ejemplo, por SEQ ID NO:65, tiene una secuencia señal que comprende (o que consiste en) los 27 residuos amino terminales o MSIVVTKIERAGAAVAALRTSGVATV (SEQ ID NO:407) que corresponde a los primeros 27 aminoácidos de la SEQ ID NO:66.

En algunas realizaciones, la invención proporciona secuencias señal que comprenden los primeros 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44 residuos amino terminales de un polipéptido de acuerdo con la invención.

La invención incluye polipéptidos con o sin una secuencia señal y/o una secuencia prepro. En algunas realizaciones, la invención incluye polipéptidos con secuencias señal y/o secuencias prepro heterólogas. La secuencia prepro (incluida una secuencia de acuerdo con la invención usada como un dominio prepro heterólogo) puede localizarse en el extremo terminal amino o carboxi de la proteína. En algunas realizaciones, la invención también incluye señales de secuencia aisladas, sintéticas o recombinantes, secuencias prepro y dominios catalíticos (tales como "sitios activos") que comprenden secuencias de acuerdo con la invención. El polipéptido que comprende una secuencia señal de acuerdo con la invención puede ser una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de acuerdo con la invención u otra enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa u otra enzima u otro polipéptido. Los métodos para identificar las secuencias dominio "prepro" y las secuencias señal son bien conocidos en la técnica, ver Van de Ven (1993) Crit. Rev. Oncog. 4(2):115-136. Por ejemplo, para identificar una secuencia prepro, la proteína se purifica del espacio extracelular y la secuencia de la proteína N-terminal se determina y se compara con la forma sin procesar.

Las secuencias señal (SP) y/o secuencias prepro de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención pueden ser péptidos aislados, sintéticos o recombinantes, o secuencias unidas a otra enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa o una no aldolasa, tal como no piruvato aldolasa, por ejemplo, no HMG y/o no KHG aldolasa polipéptido, tal como una proteína de fusión (quimérica). En algunas realizaciones, la invención proporciona polipéptidos que comprenden secuencias señal de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención. En algunas realizaciones, polipéptidos que comprenden secuencias señal SP y/o prepro de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención comprenden secuencias heterólogas a una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de acuerdo con la invención (tal como una proteína de fusión que comprende una SP y/o prepro de acuerdo con la invención y secuencias de otra enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa o una no aldolasa, tal como no piruvato aldolasa, por ejemplo, proteína no HMG y/o no KHG aldolasa). En algunas realizaciones, la invención proporciona enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención con SP y/o secuencias prepro heterólogas, tales como secuencias con una secuencia señal de levadura señal, una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de acuerdo con la invención pueden comprender una SP y/o prepro heteróloga en un vector, tal como un vector de la serie pPIC (Invitrogen, Carlsbad, CA).

En algunas realizaciones, las secuencias SP y/o prepro de acuerdo con la invención son identificadas siguiendo una identificación de péptidos nuevos de aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. Las vías por las cuales las proteínas se clasifican y se transportan a su ubicación celular adecuada a menudo se denominan vías que se dirigen a proteínas. Uno de los elementos más importantes en todos estos sistemas de direccionamiento es una secuencia de aminoácidos corta en el extremo amino de un polipéptido recientemente sintetizado denominado la secuencia señal. Esta secuencia señal dirige una proteína a su ubicación apropiada en la célula y se elimina durante el transporte o cuando la proteína alcanza su destino final. La mayoría de las proteínas lisosomales, de membrana o secretadas tienen una secuencia señal amino-terminal que les marca la traslocación hacia el lumen en el retículo endoplasmático. El largo de las secuencias señal puede variar de aproximadamente 10 a 65, o más, residuos aminoácidos. Varios métodos para el reconocimiento de secuencias señal son conocidos por los expertos en la técnica. Por ejemplo, en algunas realizaciones, las señales de péptidos nuevos de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa se identifican por medio de un método denominado SignalP. SignalP usa una red neural combinada que reconoce ambas señales de péptidos y sus sitios de escisión. (Nielsen (1997) "Identification of prokaryotic and eukaryotic peptides signals and prediction of their cleavage sites". Protein Engineering 10:1-6.

En algunas realizaciones, las enzimas aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención no tienen secuencias SP y/o "dominios" prepro. En algunas realizaciones, la invención proporciona la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención que carece de todo o parte de una SP y/o dominio prepro. En algunas realizaciones, la invención proporciona secuencias de ácido nucleico que codifican una secuencia señal (SP) y/o prepro de una enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa operativamente ligada a una secuencia de ácidos nucleicos de una aldolasa diferente, tal como enzima piruvato aldolasa, tal como HMG y/o KHG aldolasa u

opcionalmente puede desearse una secuencia señal (SP) y/o dominio prepro de una no aldolasa, tal como no piruvato aldolasa, por ejemplo, una proteína no HMG y/o no KHG aldolasa.

La invención también proporciona polipéptidos aislados, sintéticos o recombinantes que comprenden secuencias señal (SP), dominio prepro y/o dominios catalíticos (CD) de acuerdo con la invención y secuencias heterólogas. Las secuencias heterólogas son secuencias que no están naturalmente asociadas (tal como a una enzima) con una SP, dominio prepro y/o CD. La secuencia a la cual la SP, el dominio prepro y/o CD no se asocian naturalmente puede ser en las SP, dominio prepro y/o extremo amino terminal de CD, extremo carboxi terminal y/o ambos extremos en el SP y/o CD. En algunas realizaciones, la invención proporciona polipéptidos aislados, sintéticos o recombinantes que comprenden (o que consisten en) un polipéptido que comprende una secuencia señal (SP), un dominio prepro y/o un dominio catalítico (CD) de acuerdo con la invención con la condición de que no se asocien con ninguna secuencia a la cual no estén naturalmente asociadas (tal como una enzima aldolasa, tal como una secuencia de piruvato aldolasa, HMG y/o KHG aldolasa). De forma similar, en algunas realizaciones, la invención proporciona ácidos nucleicos aislados, sintéticos o recombinantes que codifican estos polipéptidos. Así, en algunas realizaciones, el ácido nucleico aislado, sintético o recombinante de acuerdo con la invención comprende una secuencia de codificación para una secuencia señal (SP), un dominio prepro y/o un dominio catalítico (CD) de acuerdo con la invención y una secuencia heteróloga (es decir, una secuencia que no esté naturalmente asociada con una secuencia señal (SP), un dominio prepro y/o dominio catalítico (CD) de acuerdo con la invención). La secuencia heteróloga puede estar en el extremo terminal 3', extremo terminal 5' y/o en ambos extremos de la SP, el dominio prepro y/o la secuencia de codificación CD.

*Enzima aldolasa híbrida (quimérica), tal como piruvato aldolasa, tal como HMG y/o KHG aldolasas y bibliotecas de péptidos*

En algunas realizaciones, la invención proporciona una enzima aldolasa híbrida, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa y proteínas de fusión, incluidas bibliotecas de péptidos, que comprenden secuencias de acuerdo con la invención. Las bibliotecas de péptidos de acuerdo con la invención pueden utilizarse para aislar modulares de péptidos (tales como activadores o inhibidores) de blancos, tales como sustratos, receptores, enzimas de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. Las bibliotecas de péptidos de acuerdo con la invención pueden utilizarse para identificar partes de unión formales, tales como ligandos, tales como citocinas, hormonas y similares. En algunas realizaciones, la invención proporciona proteínas quiméricas que comprenden una secuencia señal (SP), un dominio prepro y/o un dominio catalítico (CD) de acuerdo con la invención o una combinación de los mismos y una secuencia heteróloga (ver anteriormente).

En algunas realizaciones, las proteínas de fusión de acuerdo con la invención (tal como el resto de péptido) se estabilizan conformacionalmente (con respecto a péptidos lineales) para permitir una afinidad de unión más alta para los objetivos. En algunas realizaciones, la invención proporciona fusiones de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención y otros péptidos, incluidos péptidos conocidos y péptidos aleatorios. Pueden fusionarse de manera tal que la estructura de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasas no se perturbe considerablemente y que la conformación del péptido se establezca metabólicamente o estructuralmente. Esto permite la creación de una biblioteca de péptidos que se monitorea fácilmente para determinar su presencia en las células y su cantidad.

Variantes de secuencias de aminoácidos de acuerdo con la invención pueden caracterizarse por una naturaleza predeterminada de la variación, una característica que los separa de una forma natural, tal como una variación alélica o entre especies de una enzima aldolasa, tal como una secuencia de piruvato aldolasa, HMG y/o KHG aldolasa. En algunas realizaciones, las variantes de acuerdo con la invención exhiben la misma actividad biológica cualitativa que el análogo natural. De forma alternativa, las variantes pueden seleccionarse por tener características modificadas. En algunas realizaciones, mientras que el sitio o región para introducir una variación de la secuencia de aminoácidos es predeterminado, la mutación en sí misma no necesita ser determinada. Por ejemplo, para optimizar el rendimiento de una mutación en un sitio dado, la mutagénesis aleatoria puede llevarse a cabo en el codón o región objetivo y las variantes de enzima aldolasa expresadas, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa evaluadas para la combinación óptima de la actividad deseada. Las técnicas para realizar dichas mutaciones de sustitución en sitios predeterminados en el ADN que tiene una secuencia conocida son bien conocidas, como se describe por ejemplo en la presente, mutagénesis de cebador M13 y mutagénesis por PCR. La evaluación de los mutantes puede realizarse usando, por ejemplo ensayos de formación de enlaces carbono-carbono o escisión. En otras realizaciones, las sustituciones de aminoácidos pueden ser residuos únicos; las inserciones pueden realizarse en el orden de aproximadamente 1 a 20 aminoácidos, a pesar de que pueden realizarse más inserciones. Las eliminaciones pueden estar en un rango de aproximadamente 1 a aproximadamente 20, 30, 40, 50, 60, 70 residuos o más. Para obtener un derivado final con las propiedades óptimas, pueden usarse sustituciones, eliminaciones, inserciones o cualquier combinación de las mismas. Generalmente, estos cambios se realizan en unos pocos aminoácidos para minimizar la alteración de la molécula. Sin embargo, pueden tolerarse mayores cambios en ciertas circunstancias.

La invención proporciona enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa donde la estructura central del polipéptido, la estructura secundaria o la estructura terciaria, tal como una estructura de hélice

5 alfa u hoja beta, ha sido modificada. En algunas realizaciones, la carga o hidrofobicidad ha sido modificada. En algunas realizaciones, el volumen de una cadena lateral ha sido modificado. Los cambios importantes en la función o identidad inmunológica se obtienen mediante la selección de sustituciones que son menos conservadoras. Por ejemplo, pueden realizarse sustituciones que afectan más considerablemente: la estructura de la estructura central del polipéptido en el área de la alteración, por ejemplo una estructura de hélice alfa u hoja beta; una carga un sitio hidrófobo de la molécula, que puede estar en un sitio activo; o una cadena lateral. En algunas realizaciones, la invención proporciona sustituciones en polipéptidos de acuerdo con la invención donde (a) los residuos hidrófilos, tales como serilo o treonilo, se sustituyen por un residuo hidrófobo, tal como leucilo, isoleucilo, fenilalanilo, valilo o alanilo; (b) una cisteína o prolina se sustituye por cualquier otro residuo; (c) un residuo que tiene una cadena lateral electropositiva, tal como lisilo, arginilo o histidilo, se sustituye por un residuo electronegativo, tal como glutamilo o aspartilo; o (d) un residuo que tiene una cadena lateral voluminosa, tal como fenilalanina, se sustituye por uno que no tiene una cadena lateral, tal como glicina. Las variantes pueden exhibir la misma actividad biológica cualitativa (es decir, una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa) a pesar de que las variantes pueden seleccionarse para modificar las características de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, según sea necesario.

10 En algunas realizaciones, la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención comprende epítomos o etiquetas de purificación, secuencias señal u otras secuencias de fusión, etc. En algunas realizaciones, la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención puede fusionarse con un péptido aleatorio para formar un polipéptido de fusión. "Fusionado" u "operativamente ligado" en la presente significa que el péptido aleatorio y la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa están ligados, de forma tal que minimizan la alteración de la estabilidad de la enzima aldolasa, tal como piruvato aldolasa, tal como una estructura de HMG y/o KHG aldolasa, que por ejemplo retiene la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. El polipéptido de fusión (o polinucleótido de fusión que codifica el polipéptido de fusión) puede comprender también otros componentes, incluidos múltiples péptidos en múltiples aros.

15 En algunas realizaciones, los péptidos y ácidos nucleicos que los codifican son aleatorizados, ya sea completamente aleatorizados o sesgados en su aleatorización, tal como en frecuencia de nucleótidos/residuos generalmente o por posición. "Aleatorizado" significa que cada ácido nucleico y péptido consiste en nucleótidos esencialmente aleatorios y aminoácidos, respectivamente. En algunas realizaciones, los ácidos nucleicos que dan origen a los péptidos pueden sintetizarse químicamente, y por lo tanto pueden incorporar cualquier nucleótido en cualquier posición. Así, cuando los ácidos nucleicos se expresan para formar péptidos, cualquier residuo aminoácido puede incorporarse en cualquier posición. El proceso sintético puede diseñarse para generar ácidos nucleicos aleatorizados, para permitir la formación de todas o la mayoría de las combinaciones en el largo del ácido nucleico, formando así una biblioteca de ácidos nucleicos aleatorizados. La biblioteca puede proporcionar una población con una estructura lo suficientemente diversa de productos de expresión aleatorizados para influir en un rango que probablemente sea suficiente de respuestas celulares para proporcionar una o más células que exhiben una respuesta deseada. De esta forma, la invención proporciona bibliotecas de interacción lo suficiente grandes de forma tal que al menos uno de los miembros tendrán una estructura que le da afinidad por algunas moléculas, proteínas u otros factores.

20 En algunas realizaciones, una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de acuerdo con la invención es una enzima de múltiples dominios que comprende una señal peptídica, un módulo de unión a carbohidratos, una enzima aldolasa, tal como un dominio catalítico de piruvato aldolasa, HMG y/o KHG aldolasa, un ligante y/o otro dominio catalítico.

25 La invención proporciona métodos y secuencias para generar polipéptidos quiméricos que pueden codificar polipéptidos híbridos biológicamente activos (tal como una enzima aldolasa híbrida, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa). En algunas realizaciones, los polinucleótidos originales (tal como un ácido nucleico de acuerdo con la invención) codifican polipéptidos biológicamente activos. En algunas realizaciones, un método de acuerdo con la invención produce nuevos polipéptidos híbridos utilizando procesos celulares que integran la secuencia de los polinucleótidos originales de forma tal que el polinucleótido híbrido resultante codifique un polipéptido que demuestra actividades derivadas de, pero diferentes a, los polipéptidos biológicamente activos originales (tal como aldolasa o anticuerpo de acuerdo con la invención). Por ejemplo, los polinucleótidos originales pueden codificar una enzima particular (tal como aldolasa) de o que se encuentra en microorganismos diferentes, una enzima que codifica un primer polinucleótidos de un organismo o variante puede, por ejemplo, funcionar de manera efectiva en una condición ambiental particular, tal como alta salinidad, una enzima que codifica un segundo polinucleótido de un organismo diferente o variante puede funcionar de manera efectiva en una condición ambiental diferente, tal como temperaturas extremadamente altas. Un polinucleótido híbrido que contiene secuencias del primer y el segundo polinucleótido original puede codificar una enzima que exhibe características de ambas enzimas codificadas por los polinucleótidos originales. Así, la enzima codificada por el polinucleótido híbrido de acuerdo con la invención puede funcionar de manera efectiva en condiciones ambientales compartidas por cada una de las enzimas codificadas por el primer y el segundo polinucleótido, tal como alta salinidad y temperaturas extremas.

- En algunas realizaciones, un polipéptido híbrido generado por un método de acuerdo con la invención puede exhibir una enzima especializada no exhibida en las enzimas originales. Por ejemplo, siguiendo la recombinación y/o reordenamiento reductivo de polinucleótidos que codifican enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasas, el polipéptido híbrido resultante codificado por un polinucleótido híbrido puede evaluarse en busca de actividad de enzima no aldolasa especializada, tal como no piruvato aldolasa, tal como enzima no HMG y/o no KHG-aldolasa, tal como actividades de hidrolasa, peptidasa, fosforilasa, etc., obtenidas de cada una de las enzimas originales. En algunas realizaciones, el polipéptido híbrido se evalúa para determinar dichas funcionalidades químicas que distinguen al polipéptido híbrido de los polipéptidos base originales, tal como la temperatura, pH o concentración de sal en la cual el polipéptido híbrido funciona.
- 5
- 10 En algunas realizaciones, la invención se refiere a un método para producir un polipéptido híbrido biológicamente activo y evaluar dicho polipéptido en busca de actividad mejorada:
- 1) introduciendo al menos un primer polinucleótido en enlace operativo y un segundo polinucleótido en enlace operativo, compartiendo al menos el primer y el segundo polinucleótidos al menos una región de una homología parcial, en una célula huésped adecuada;
- 15
- 2) cultivando la célula huésped en condiciones que promueven la reorganización de secuencias que resulta en un polinucleótido híbrido en enlace operativo;
  - 3) expresando un polipéptido híbrido codificado por el polinucleótido híbrido;
  - 4) evaluando el polipéptido híbrido en condiciones que promueven la identificación de la actividad biológica mejorada; y
- 20
- 5) aislando un polinucleótido que codifica el polipéptido híbrido.

#### Aislar y descubrir enzimas aldolasa

La divulgación proporciona métodos para aislar y descubrir aldolasa, tal como enzimas piruvato aldolasa, HMG y/o KHG aldolasa y los ácidos nucleicos que las codifican. Pueden aislarse polinucleótidos o enzimas de organismos individuales (“aislados”), conjuntos de organismos que se han cultivado en medios definidos (“cultivos de enriquecimiento”) u organismos sin cultivar (“muestras ambientales”). Los organismos pueden aislarse, por ejemplo, mediante ciclos de selección por afinidad (biopanning) in vivo (ver detalles más abajo). El uso de un abordaje independiente del cultivo para derivar polinucleótidos que codifican bioactividades novedosas a partir de muestras ambientales es más preferible porque permite acceder a recursos de biodiversidad sin explotar. Los polinucleótidos o enzimas también pueden aislarse a partir de uno cualquiera de numerosos organismos, tales como bacterias. Además de células enteras, los polinucleótidos o enzimas también pueden aislarse a partir de preparaciones enzimáticas brutas derivadas de cultivos de estos organismos, tales como bacterias.

25

30

“Bibliotecas ambientales” se generan a partir de muestras ambientales y representan los genomas colectivos de organismos naturales obtenidos en la clonación de vectores que pueden propagarse en huéspedes procariotas adecuados. Dado que el ADN clonado se extrae inicialmente directamente de muestras ambientales, las bibliotecas no se limitan a la pequeña fracción de procariotas que pueden cultivarse en un cultivo puro. Además, una normalización del ADN ambiental presente en estas muestras podría permitir una representación más equitativa del ADN de todas las especies presentes en la muestra original. Esto puede aumentar drásticamente la eficacia de la búsqueda de genes interesantes de constituyentes menores de la muestra que pueden estar representados en órdenes de magnitud muy inferiores en comparación con la especie dominante.

35

En algunos casos se detectan sistemáticamente bibliotecas génicas generadas a partir de uno o más microorganismos sin cultivar en busca de una actividad de interés. Las posibles vías que codifican moléculas bioactivas de interés se capturan en primer lugar en células procariotas en forma de bibliotecas de expresión génica. En algunos casos, los polinucleótidos que codifican actividades de interés se aíslan de dichas bibliotecas y se introducen en una célula huésped. La célula huésped se cultiva en condiciones que promueven la recombinación y/o reordenamiento reductor, creando biomoléculas potencialmente activas con actividades novedosas o mejoradas.

40

45

La selección por afinidad in vivo puede realizarse utilizando máquinas basadas en FACS y no ópticas (tales como magnéticas). En algunos casos se construyen bibliotecas de genes complejas con vectores que contienen elementos que estabilizan el ARN transcrito. Por ejemplo, la inclusión de secuencias que resultan en estructuras secundarias tales como horquillas, que son diseñadas para flanquear las regiones transcritas del ARN, serviría para mejorar su estabilidad, incrementando así su semivida dentro de la célula. Las moléculas de sonda utilizadas en el proceso de ciclos de selección por afinidad consisten en oligonucleótidos etiquetados con moléculas reporteras que solo fluorescen tras la unión de la sonda con una molécula objetivo. Estas sondas se introducen en las células recombinantes de la biblioteca usando uno de varios métodos de transformación. Las moléculas de las sondas se unen al ARNm objetivo transcrito, lo que resulta en moléculas heterodúplex de ADN/ARN. La unión de la sonda a un

50

objetivo arrojará una señal fluorescente que es detectada y agrupada por la máquina de FACS durante el proceso de selección sistemática.

5 En algunos casos se lleva a cabo una subclonación para aislar aun más secuencias de interés. En la subclonación, una porción de ADN es amplificada y digerida, generalmente por enzimas de restricción, para cortar la secuencia deseada, la cual se liga a un vector receptor y se amplifica. En cada etapa de la subclonación se examina la porción en busca de la actividad de interés con el fin de asegurar que el ADN que codifica la proteína estructural no haya quedado excluida. El inserto puede purificarse en cualquier etapa de la subclonación, por ejemplo, mediante electroforesis en gel antes de la ligadura en el vector o en donde las células que contienen el vector receptor y las células que no contienen el vector receptor se colocan en medio selectivo que contiene, por ejemplo, un antibiótico que matará a las células que no contienen el vector receptor. Los métodos específicos de subclonación de insertos de ADNc en vectores se conocen bien en la técnica (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2a Ed., Cold Spring Harbor Laboratory Press (1989)). En otros casos, las enzimas de acuerdo con la invención son subclones. Estos subclones pueden diferir del clon base, por ejemplo, en longitud, una mutación o una etiqueta.

15 Los microorganismos de los cuales puede descubrirse, aislarse o prepararse el polinucleótido incluyen microorganismos procariotas, tales como *Eubacteria* y *Archaeobacteria* y microorganismos eucariotas inferiores tales como hongos, algunas algas y protozoos. Los polinucleótidos pueden descubrirse, aislarse o prepararse a partir de muestras ambientales, en cuyo caso el ácido nucleico puede recuperarse sin cultivar un organismo o recuperarse de uno o más organismos cultivados. En algunos casos, dichos microorganismos pueden ser extremófilos, tales como hipertermófilos, psicrófilos, psicrótrofos, halófilos, barófilos y acidófilos. Pueden usarse polinucleótidos que codifican enzimas aisladas de microorganismos extremófilos. Las enzimas de la presente invención pueden funcionar a temperaturas por encima de los 100°C, tales como las que se encuentran en manantiales calientes terrestres y respiraderos termales submarinos, o a temperaturas por debajo de 0°C, tales como las que se encuentran en las aguas árticas, en un ambiente salino saturado, tales como las que se encuentran en el Mar Muerto, a valores de pH cercanos a 0, tales como las que se encuentran en depósitos de carbón y manantiales geotérmicos ricos en azufre, o a valores de pH superiores a 11, tales como las que se encuentran en lodos de aguas servidas. En algunos casos, las enzimas de acuerdo con la invención tienen una alta actividad en una amplia gama de temperaturas y valores de pH.

20 Los polinucleótidos seleccionados y aislados como se describió anteriormente se introducen en una célula huésped adecuada. Una célula huésped adecuada es cualquier célula que es capaz de promover la recombinación y/o reordenamiento reductor. En algunas realizaciones, los polinucleótidos seleccionados ya están en un vector que incluye secuencias control adecuadas. La célula huésped puede ser una célula eucariota superior, tal como una célula de mamífero, o una célula eucariota inferior, tal como una célula de levadura o, en algunas realizaciones, la célula huésped puede ser una célula procariota, tal como una célula bacteriana. La introducción del constructo en la célula huésped puede realizarse mediante transfección con fosfato de calcio, transfección mediada por DEAE-Dextrano o electroporación.

30 Ejemplos de huéspedes incluyen 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; ver detalles más arriba. Se considera que la selección de un huésped apropiado se encuentra dentro del alcance de los expertos en la técnica a partir de las enseñanzas de la presente.

35 Pueden emplearse varios sistemas de cultivo de células de mamífero para expresar proteína recombinante; ejemplos de sistemas de expresión de 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 de mamíferos pueden comprender un origen de replicación, un promotor y potenciador adecuado y también los sitios de unión a ribosomas, sitios de poliadenilación, sitios donantes y aceptores de empalmes, secuencias de terminación transcripcional y secuencias no transcritas flanqueadoras 5' que se necesiten. Pueden usarse secuencias de ADN derivadas del empalme SV40 y los sitios de poliadenilación para proporcionar los elementos genéticos no transcritos necesarios.

45 En otras realizaciones se usan ácidos nucleicos, polipéptidos y métodos de acuerdo con la invención en vías bioquímicas o para generar novedosos polinucleótidos que codifiquen vías bioquímicas de uno o más operones o agrupamientos de genes de los mismos. Por ejemplo, las bacterias y muchos eucariotas tienen un mecanismo coordinado para regular genes cuyos productos participan en procesos relacionados. Los genes se agrupan en estructuras denominadas "agrupamientos de genes" en un solo cromosoma y se transcriben juntos bajo el control de una sola secuencia reguladora, incluido un solo promotor que inicia la transcripción de todo el agrupamiento. De esta forma, un agrupamiento de genes es un grupo de genes adyacentes que son idénticos o están relacionados, generalmente en cuanto a su función (un ejemplo de vía bioquímica codificada por un agrupamiento de genes son los policétidos).

En algunas realizaciones, el ADN del agrupamiento de genes se aísla de diferentes organismos y se liga en vectores, tales como vectores que contienen secuencias reguladoras de la expresión que pueden controlar y regular la producción de una actividad de proteína o arreglo relacionado de proteína detectable de los agrupamientos de genes ligados. El uso de vectores que tienen una capacidad excepcionalmente grande para la introducción de ADN exógeno puede ser apropiado para su uso con dichos agrupamientos de genes y se describen a modo de ejemplo en la presente de forma de incluir el factor f (o factor de fertilidad) de *E. coli*. Este factor f de *E. coli* es un plásmido que afecta la transferencia de alta frecuencia de sí mismo durante la conjugación y es ideal para obtener y propagar establemente grandes fragmentos de ADN, tales como agrupamientos de genes de muestras microbianas mezcladas. En una realización se usan vectores de clonación, denominados "fósmidos" o vectores de cromosomas artificiales bacterianos (BAC). Estos derivan del factor f de *E. coli*, que es capaz de integrar establemente grandes segmentos de ADN genómico. Cuando se integran con el ADN de una muestra ambiental mezclada es posible obtener grandes fragmentos genómicos en forma de una "biblioteca de ADN ambiental" estable. Otro tipo de vector para su uso en la presente invención es un vector cósmido. Los vectores cósmidos se diseñaron originalmente para clonar y propagar grandes segmentos de ADN genómico. La clonación en vectores cósmidos se describe detalladamente en Sambrook et al., Molecular Cloning: A Laboratory Manual, 2a Ed., Cold Spring Harbor Laboratory Press (1989). Una vez ligados a un vector apropiado, pueden introducirse en una célula huésped adecuada dos o más vectores que contengan diferentes agrupamientos de genes de sintasa de policétidos. Las regiones de homología de secuencia parcial compartidas por los agrupamientos de genes promoverán procesos que resultan en la reorganización de la secuencia, lo que resulta en un agrupamiento de genes híbrido. Posteriormente, el agrupamiento de genes híbrido novedoso puede analizarse sistemáticamente para detectar actividades mejoradas que no se encontraban en los agrupamientos de genes originales.

Los métodos para detectar sistemáticamente las distintas actividades enzimáticas son conocidos para los expertos en la técnica y se describen en la presente memoria descriptiva (ver los Ejemplos 1, 2 y 3 más adelante. Estos métodos pueden emplearse cuando se aíslan los polipéptidos y polinucleótidos de acuerdo con la invención.

En algunas realizaciones, la invención proporciona métodos para descubrir y aislar aldolasas, tales como piruvato aldolasa, tales como HMG y/o KHG aldolasa, o compuestos para modificar la actividad de estas enzimas, usando un abordaje de células enteras (ver descripción más abajo). Pueden detectarse sistemáticamente clones putativos que codifican aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de bibliotecas de ADN genómico.

#### Metodologías de detección automática y dispositivos de monitoreo "en línea"

En la puesta en práctica de los métodos de acuerdo con la invención pueden usarse distintos aparatos y metodologías conjuntamente con los polipéptidos y ácidos nucleicos de acuerdo con la invención, por ejemplo para detectar sistemáticamente polipéptidos en busca de actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, para detectar sistemáticamente compuestos como posibles moduladores, tales como activadores o inhibidores, de una actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, para anticuerpos que se unen a un polipéptido de acuerdo con la invención, para ácidos nucleicos que se hibridan con un ácido nucleico de acuerdo con la invención, para detectar sistemáticamente células que expresan un polipéptido de acuerdo con la invención y similares. Además de los distintos formatos descritos detalladamente más adelante para detectar muestras sistemáticamente, también pueden utilizarse formatos alternativos para poner en práctica los métodos de acuerdo con la invención. Estos formatos incluyen, por ejemplo, espectrómetros de masas, cromatógrafos, tales como HPLC de alto rendimiento y otras formas de cromatografía líquida, y formatos más pequeños, tales como placas de 1536 pocillos, placas de 384 pocillos, etc. Pueden adaptarse y utilizarse aparatos de detección sistemática de alto rendimiento para poner en práctica los métodos de acuerdo con la invención. Ver las Solicitudes de Patente de los Estados Unidos Nos. 20020001809 y 20050272044.

#### Arreglos de capilares

Los ácidos nucleicos o polipéptidos de acuerdo con la invención pueden inmovilizarse o aplicarse a un arreglo. Los arreglos pueden usarse para detectar sistemáticamente o para monitorear bibliotecas de composiciones (tales como pequeñas moléculas, anticuerpos, ácidos nucleicos, etc.) en busca de su capacidad de unirse o modular la actividad de un ácido nucleico o polipéptido de acuerdo con la invención. Los Arreglos de capilares, tales como el GIGAMATRIX™, Diversa Corporation, San Diego, Calif.; y los arreglos descritos, por ejemplo, en la Solicitud de Patente de los Estados Unidos No. 20020080350 A1; WO 0231203 A; WO 0244336 A, proporcionan un aparato alternativo para sostener y detectar sistemáticamente muestras. En algunas realizaciones, el Arreglo de capilares incluye una pluralidad de capilares que forman 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 tener cualquier otra forma geométrica, siempre que las paredes formen un lumen para la retención de un líquido o muestra. Los capilares del arreglo de capilares pueden mantenerse juntos a una distancia muy próxima para formar una estructura plana. Los capilares pueden unirse, por ejemplo fusionarse (tal como cuando los capilares están hechos de vidrio), pegarse, unirse o fijarse lado a lado. Además, el arreglo de capilares puede incluir material intersticial colocado entre capilares adyacentes en el arreglo para formar un dispositivo plano sólido que contenga una pluralidad de agujeros pasantes.

5 Un Arreglo de capilares puede formarse con cualquier cantidad de capilares individuales, por ejemplo, en un rango de 100 a 4.000.000 de capilares. Más aun, a un Arreglo de capilares que tenga aproximadamente 100.000 capilares individuales o más puede dársele el tamaño y la forma estándar de una placa Microtiter® para que pueda ajustarse a equipos de laboratorio estándar. Los lúmenes se llenan manualmente o automáticamente usando acción capilar o microinyección usando una aguja delgada. Las muestras de interés pueden retirarse posteriormente de los capilares individuales para analizar o caracterizar adicionalmente. Por ejemplo, se coloca una sonda delgada similar a una aguja en comunicación fluida con un capilar seleccionado para agregar o retirar material del lumen.

10 En un ensayo de detección sistemática de un solo paso, los componentes del ensayo se mezclan para dar una solución de interés antes de la inserción en el arreglo de capilares. El lumen se llena mediante acción capilar cuando al menos una porción del arreglo se sumerge en una solución de interés. Las reacciones químicas o biológicas y/o la actividad en cada capilar se monitorean en busca de eventos detectables. A menudo, un evento detectable se denomina "acierto" que normalmente puede distinguirse de capilares que producen un "no acierto" mediante detección óptica. De esta forma, los Arreglos de capilares permiten una enorme detección en paralelo de "aciertos".

15 En un ensayo de detección sistemática de múltiples pasos puede introducirse un polipéptido o ácido nucleico, tal como un ligando, en un primer componente, el cual se introduce en al menos una porción de un capilar o Arreglo de capilares. Luego puede introducirse una burbuja de aire en el capilar detrás del primer componente. Posteriormente puede introducirse un segundo componente en el capilar, en donde el segundo componente se separa del primer componente mediante la burbuja de aire. El primer y el segundo componente pueden mezclarse luego mediante la aplicación de presión hidrostática en ambos lados del arreglo de capilar para hacer colapsar la burbuja. El arreglo de capilares luego puede monitorearse en busca de un evento detectable que resulte de la reacción o no reacción de los dos componentes.

20 En un ensayo de detección sistemática de unión puede introducirse una muestra de interés como un primer líquido etiquetado con una partícula detectable en un capilar o arreglo de capilares, en donde el lumen del capilar está recubierto con un material de unión para unir la partícula detectable al lumen. Luego, el primer líquido puede retirarse del tubo capilar, con lo que la partícula detectable unida se mantiene dentro del capilar, y posteriormente puede introducirse un segundo líquido en el tubo capilar. El capilar luego puede monitorearse en busca de un evento detectable que resulte de la reacción o no reacción de la partícula con el segundo líquido.

Arreglos o "biochips"

30 Los ácidos nucleicos o polipéptidos de acuerdo con la invención pueden inmovilizarse o aplicarse a un arreglo. Los arreglos pueden usarse para detectar sistemáticamente o para monitorear bibliotecas de composiciones (tales como pequeñas moléculas, anticuerpos, ácidos nucleicos, etc.) en busca de su capacidad de unirse o modular la actividad de un ácido nucleico o polipéptido de acuerdo con la invención. Por ejemplo, en algunas realizaciones de la invención, un parámetro monitoreado es expresión de transcritos de un gen de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa. Uno o más o todos los transcritos de una célula pueden medirse mediante hibridación de una muestra que comprende transcritos de la célula o ácidos nucleicos representativos de una célula o complementarios a transcritos de una célula, mediante hibridación con ácidos nucleicos inmovilizados en un arreglo o "biochip." Mediante el uso de un "arreglo" de ácidos nucleicos en un microchip es posible cuantificar simultáneamente algunos o la totalidad de los transcritos. Alternativamente, los arreglos que comprenden ácido nucleico genómico también pueden usarse para determinar el genotipo de una cepa recientemente manipulada obtenida mediante los métodos de acuerdo con la invención. También pueden usarse "arreglos de polipéptidos" para cuantificar simultáneamente una pluralidad de proteínas. La presente invención también puede ponerse en práctica con cualquier "arreglo" conocido, también denominado "microarreglo" o "arreglo de ácidos nucleicos" o "arreglo de polipéptidos" o "arreglo de anticuerpos" o "biochip", o una variación del mismo. Los arreglos son genéricamente una pluralidad de "lugares" o "elementos objetivo", comprendiendo cada elemento objetivo una cantidad definida de una o más moléculas biológicas, tales como oligonucleótidos, inmovilizadas sobre un área definida de una superficie de sustrato para la unión específica a una molécula de muestra, tal como transcritos de ARNm.

45 Los términos "arreglo" o "microarreglo" o "biochip" o "chip", tal como se usan en la presente, son una pluralidad de elementos objetivo, comprendiendo cada elemento objetivo una cantidad definida de uno o más polipéptidos (incluidos anticuerpos) o ácidos nucleicos inmovilizados sobre un área definida de una superficie de sustrato, tal como detalla más adelante.

50 En la puesta en práctica de los métodos de acuerdo con la invención puede incorporarse cualquier arreglo y/o método para fabricar y usar arreglos en todo o en parte, o variaciones de los mismos, tal 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; ver también otros documentos, tales como WO 99/51773; WO 99/09217; WO 97/46313; WO 96/17958; ver también otras publicaciones, tales como 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. Ver también las Solicitudes de Patente de los Estados Unidos Publicadas Nos. 20010018642; 20010019827; 20010016322; 20010014449; 20010014448; 20010012537; 20010008765.

#### Anticuerpos y métodos de detección sistemática basados en anticuerpos

5 La invención proporciona anticuerpos aislados, sintéticos o recombinantes que se unen específicamente a una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de acuerdo con la invención. Pueden usarse estos anticuerpos para aislar, identificar o cuantificar las enzimas aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención o polipéptidos relacionados. Estos anticuerpos pueden usarse para aislar otros polipéptidos dentro del alcance de la invención u otras enzimas aldolasa relacionadas, tales como piruvato aldolasa, tal como HMG y/o KHG aldolasa. Los anticuerpos pueden diseñarse para unirse a un sitio activo de una enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. De esta forma, la invención proporciona métodos para inhibir enzimas aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa usando los anticuerpos de acuerdo con la invención (ver descripción más arriba con respecto a aplicaciones para composiciones de enzimas anti-aldolasa, tal como anti-piruvato aldolasa, tal como anti-HMG y/o anti-KHG aldolasa de acuerdo con la invención).

15 El término "anticuerpo" incluye un péptido o polipéptido derivado o modelado de, o básicamente codificado por, un gen de inmunoglobulina o genes de inmunoglobulina, o fragmentos de los mismos, capaz de unirse específicamente a un antígeno o epítipo, ver 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 que se unen a antígenos, es decir, "sitios de unión a antígenos" (tales como fragmentos, subsecuencias, regiones determinantes de la complementariedad (CDR)) que retienen la capacidad de unirse al antígeno, incluidos (i) un fragmento Fab, un fragmento monovalente que consiste en los dominios VL, VH, CL y CH1; (ii) un fragmento F(ab')<sub>2</sub>, un fragmento bivalente que comprende dos fragmentos Fab enlazados por un puente de disulfuro en la región bisagra; (iii) un fragmento Fd que consiste en los dominios VH y CH1; (iv) un fragmento Fv que consiste en los dominios VL y VH de un solo brazo de un anticuerpo, (v) un fragmento dAb (Ward et al., (1989) Nature 341:544-546), que consiste en un dominio VH; y (vi) una región determinante de la complementariedad (CDR) aislada. También se incluyen como referencia anticuerpos de cadena simple en el término "anticuerpo".

20 La invención proporciona fragmentos de las enzimas de acuerdo con la invención (tales como péptidos), incluidos fragmentos inmunogénicos (tales como subsecuencias) de un polipéptido de acuerdo con la invención. En algunas realizaciones, la invención proporciona composiciones que comprenden un polipéptido o péptido de acuerdo con la invención y adyuvantes o portadores y similares.

25 Los anticuerpos pueden usarse en inmunoprecipitación, tinción, columnas de inmovilización y similares. Si se desea, las secuencias de ácidos nucleicos que codifican antígenos específicos pueden generarse mediante inmunización con posterior aislamiento de polipéptido o ácido nucleico, amplificación o clonación e inmovilización de polipéptido sobre un arreglo de acuerdo con la invención. Alternativamente, los métodos de acuerdo con la invención pueden usarse para modificar la estructura de un anticuerpo producido por una célula que ha de ser modificada, tal como aumentar o disminuir la afinidad de un anticuerpo. Más aun, la capacidad de hacer o modificar anticuerpos puede ser un fenotipo manipulado e introducido en una célula mediante los métodos de acuerdo con la invención.

30 Los métodos de inmunización, producción y aislamiento de anticuerpos (policlonales y monoclonales) son conocidos para los expertos en la técnica y se describen en la bibliografía científica y de patentes; ver Coligan, CURRENT PROTOCOLS IN IMMUNOLOGY, Wiley/Greene, NY (1991); Stites (eds.) BASIC AND CLINICAL IMMUNOLOGY (7th ed.) Lange Medical Publications, Los Altos, Calif. ("Stites"); Goding, MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE (2a ed.) Academic Press, New York, N.Y. (1986); Kohler (1975) Nature 256:495; Harlow (1988) ANTIBODIES, A LABORATORY MANUAL, Cold Spring Harbor Publications, New York. También pueden generarse anticuerpos in vitro, tal como mediante el uso de sitios de unión a anticuerpos recombinantes que expresan bibliotecas de exhibición de fagos, además de los métodos in vivo tradicionales que usan animales. Ver Hoogenboom (1997) Trends Biotechnol. 15:62-70; Katz (1997) Annu. Rev. Biophys. Biomol. Struct. 26:27-45.

35 Los polipéptidos de acuerdo con 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 también pueden usarse para generar anticuerpos que se unan específicamente a los polipéptidos o fragmentos. Los anticuerpos resultantes pueden usarse en procedimientos de cromatografía de inmovilización para aislar o purificar el polipéptido o para determinar si el polipéptido está presente en una muestra biológica. En estos procedimientos, una preparación de proteína, tal como un extracto, o una muestra biológica se pone en contacto con un anticuerpo capaz de unirse específicamente a uno de los péptidos de acuerdo con 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.

55 En procedimientos de inmovilización, el anticuerpo se fija a un soporte sólido, tal como una perla u otra matriz en columna. La preparación de proteína se coloca en contacto con el anticuerpo en condiciones en las que el anticuerpo se une específicamente a uno de los polipéptidos de acuerdo con la invención o un fragmento del mismo.

Después de un lavado para retirar las proteínas no específicamente unidas, los polipéptidos unidos específicamente se eluyen.

5 La capacidad de las proteínas en una muestra biológica de unirse al anticuerpo puede determinarse usando distintos procedimientos conocidos para los expertos en la técnica. Por ejemplo, la unión puede determinarse etiquetando el anticuerpo con una etiqueta detectable, tal como un agente fluorescente, una etiqueta enzimática o un radioisótopo. Alternativamente, la unión del anticuerpo a la muestra puede detectarse usando un anticuerpo secundario que tenga dicha etiqueta detectable en el mismo. Ensayos particulares incluyen ensayos de ELISA, ensayos tipo sándwich, radioinmunoensayos y Western Blots.

10 Los anticuerpos policlonales generados contra los polipéptidos de acuerdo con 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 obtenerse mediante inyección directa de los polipéptidos en un animal o administrando los polipéptidos a un animal, por ejemplo, un animal no humano. El anticuerpo así obtenido puede unirse al polipéptido. De esta manera puede usarse incluso una secuencia que codifique solamente un fragmento del polipéptido para generar anticuerpos que puedan unirse a la totalidad del polipéptido nativo. Dichos anticuerpos pueden usarse para aislar el polipéptido de las células que expresan ese polipéptido.

15 Para la preparación de anticuerpos monoclonales puede usarse cualquier técnica que proporcione 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 humano (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).

20 Pueden adaptarse técnicas descritas para la producción de anticuerpos monocatenarios (Pat. de los Estados Unidos No. 4.946.778) para producir anticuerpos monocatenarios para los polipéptidos de acuerdo con 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. Alternativamente, pueden usarse ratones transgénicos para expresar anticuerpos humanizados para estos polipéptidos o fragmentos de los mismos.

25 Los anticuerpos generados contra los polipéptidos de acuerdo con 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 usarse para detectar sistemáticamente polipéptidos de otros organismos y muestras. En estas técnicas, los polipéptidos del organismo se ponen en contacto con el anticuerpo y se detectan aquellos polipéptidos que se unen específicamente al anticuerpo. Puede usarse cualquiera de los procedimientos descritos anteriormente para detectar la unión a anticuerpos. Uno de dichos ensayos de detección sistemática se describe en Shulman H, Eberhard A, Eberhard C, Ulitzur S, Keinan E, Bioorg Med Chem. Lett. 2000 Oct. 16; 10(20):2353-6, Highly sensitive and rapid detection of antibody catalysis by luminescent bacteria.

#### Kits

35 La divulgación proporciona, además, kits que comprenden las composiciones, tales como ácidos nucleicos, cassettes de expresión, vectores, células, semillas o plantas o partes de plantas transgénicas, polipéptidos (tales como una enzima aldolasa) y/o anticuerpos de acuerdo con la invención. Los kits también contienen material instructivo que enseña las metodologías y los usos industriales, médicos y dietéticos de acuerdo con la invención, tal como se describe en la presente.

#### 40 Manipulación de células enteras y parámetros metabólicos de medición

Los métodos de acuerdo con la invención proporcionan la evolución de células enteras, o la manipulación de células enteras, de una célula para desarrollar una nueva cepa que tenga un nuevo fenotipo, tal como una actividad nueva o modificada de aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, mediante modificación de la composición genética de la célula. Ver la solicitud de Patente de los Estados Unidos No. 20040033975.

45 La composición genética puede modificarse mediante la adición a la célula de un ácido nucleico de acuerdo con la invención, tal como una secuencia codificadora para una enzima de acuerdo con la invención. Ver WO0229032 y WO0196551.

50 Para detectar un nuevo fenotipo, al menos un parámetro metabólico de una célula modificada se monitorea en la célula en "tiempo real" o "en línea". En algunas realizaciones se monitorea en "tiempo real" o "en línea" una pluralidad de células, tal como un cultivo celular. En algunas realizaciones se monitorea en "tiempo real" o "en línea" una pluralidad de parámetros metabólicos. Los parámetros metabólicos pueden monitorearse usando las enzimas aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, de acuerdo con la invención.

El análisis de flujos metabólicos (MFA) se basa en un marco de trabajo bioquímico conocido. Se construye una matriz metabólica linealmente independiente en base a la ley de la conservación de la masa y en la hipótesis del

estado pseudoestacionario (PSSH) en los metabolitos intracelulares. En la puesta en práctica de los métodos de acuerdo con la invención se establecen redes metabólicas, incluyendo:

- identidad de todos los sustratos, productos y metabolitos intermedios de las vías
- 5 identidad de todas las reacciones químicas que interconvierten los metabolitos de las vías, la estequiometría de las reacciones de las vías,
- la identidad de todas las enzimas que catalizan las reacciones, la cinética de las reacciones enzimáticas,
- las interacciones reguladoras entre componentes de las vías, tales como interacciones alostéricas, interacciones enzima-enzima, etc.,
- 10 la compartimentalización 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 de su movimiento.

Una vez construida la red metabólica para una cepa en particular puede introducirse la presentación matemática mediante noción de matriz para estimar los flujos metabólicos intracelulares si los datos metabólicos en línea están disponibles. El fenotipo metabólico depende de los cambios de la totalidad de la red metabólica dentro de una célula. El fenotipo metabólico depende de cambios en la utilización de las vías con respecto a condiciones ambientales, regulación genética, estado de desarrollo del genotipo, etc. En algunas realizaciones de los métodos de acuerdo con la invención, luego del cálculo del MFA en línea, el comportamiento dinámico de las células, su fenotipo y otras propiedades se analizan investigando la utilización de las vías. Por ejemplo, si el suministro de glucosa aumenta y el oxígeno disminuye durante la fermentación de la levadura, la utilización de las vías respiratorias se reducirá y/o detendrá y dominará la utilización de las vías de fermentación. El control del estado fisiológico de los cultivos celulares serán posibles después del análisis de las vías. Los métodos de acuerdo con la invención pueden ayudar a determinar cómo manipular la fermentación mediante la determinación de cómo cambiar el suministro, la temperatura, el uso de inductores, etc. del sustrato para controlar el estado fisiológico de las células para moverse en la dirección deseable. En la puesta en práctica de los métodos de acuerdo con la invención, los resultados del MFA también pueden compararse con los datos de transcriptoma y proteoma para diseñar experimentos y protocolos para la manipulación metabólica o la transposición de genes, etc.

En la puesta en práctica de los métodos de acuerdo con la invención puede conferirse y detectarse cualquier fenotipo modificado o nuevo, incluidas características nuevas o mejoradas en la célula. Puede monitorearse cualquier aspecto del metabolismo o crecimiento.

Monitoreo de la expresión de un transcrito de ARNm

En algunas realizaciones de la invención, el fenotipo manipulado comprende aumentar o reducir la expresión de un transcrito de ARNm (tal como un mensaje de enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa) o generar nuevos transcritos (tal como enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa) en una célula. Esta expresión aumentada o reducida puede rastrearse mediante pruebas en busca de la presencia de una enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención o mediante ensayos de la actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. También pueden detectarse y cuantificarse transcritos de ARNm, o mensajes, mediante cualquier método conocido en la técnica, tales como Northern blot, reacciones de amplificación cuantitativa, hibridación con arreglos y similares. Las reacciones de amplificación cuantitativa incluyen, por ejemplo, PCR cuantitativa, incluida, por ejemplo, la reacción en cadena de polimerasa con transcripción inversa cuantitativa o RT-PCR; RT-PCR cuantitativa en tiempo real o "RT-PCR cinética en tiempo real" (ver Kreuzer (2001) Br. J. Haematol. 114:313-318; Xia (2001) Transplantation 72:907-914).

En algunas realizaciones de la invención, el fenotipo manipulado es generado desactivando la expresión de un gen homólogo. La secuencia codificadora o uno o más elementos de control transcripcional del gen, tal como promotores o potenciadores, pueden desactivarse. De esta forma, la expresión de un transcrito puede eliminarse completamente o solo reducirse.

En algunas realizaciones de la invención, el fenotipo manipulado comprende aumentar la expresión de un gen homólogo. Esto puede llevarse a cabo desactivando un elemento de control negativo, incluido 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 medirse mediante hibridación de una muestra que comprende transcritos de la célula o ácidos nucleicos que representan o complementan transcritos de una célula, mediante hibridación de ácidos nucleicos inmovilizados en un arreglo.

Monitoreo de la expresión de polipéptidos, péptidos y aminoácidos

En algunas realizaciones de la invención, el fenotipo manipulado comprende aumentar o reducir la expresión de un polipéptido (tal como una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa) o generar nuevos polipéptidos en una célula. Esta expresión aumentada o reducida puede rastrearse determinando la cantidad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa presente o mediante ensayos de actividad de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa. También pueden detectarse y cuantificarse polipéptidos, péptidos y aminoácidos mediante cualquier método conocido en la técnica, incluidos, por ejemplo, resonancia magnética nuclear (NMR), espectrofotometría, radiografía (radiomarcado de proteínas), electroforesis, electroforesis capilar, cromatografía líquida de alto rendimiento (HPLC), cromatografía en capa fina (TLC), cromatografía por hiperdifusión, distintos métodos inmunológicos, tales como inmunoprecipitación, inmunodifusión, inmuno-electroforesis, radioinmunoensayos (RIA), ensayos inmunoabsorbentes ligados a enzimas (ELISA), ensayos inmuno-fluorescentes, electroforesis en gel (tal como SDS-PAGE), tinción con anticuerpos, clasificador celular activado por fluorescencia (FACS), espectrometría de masas por pirólisis, espectrometría infrarroja por transformada de Fourier, espectrometría Raman, GC-MS y LC-electropulverización y espectrometrías de masas por electropulverización cap-LC en tándem y similares. También pueden detectarse sistemáticamente bioactividades novedosas usando métodos, o variaciones de los mismos, descritos en la Patente de los Estados Unidos No. 6.057.103. Más aun, como se detalla más adelante, pueden medirse uno, más de uno o todos los polipéptidos de una célula usando un arreglo de proteínas.

Aplicaciones industriales, farmacéuticas y otras

Los polipéptidos de acuerdo con la invención (tales como los que tienen aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa) pueden catalizar la formación o escisión de enlaces carbono-carbono. Las enzimas de acuerdo con la invención pueden ser catalizadores altamente selectivos. En algunas realizaciones, la invención proporciona procesos industriales que usan enzimas de acuerdo con la invención, tales como en la industria farmacéutica o de los complementos nutricionales (de la dieta) y en la industria de alimentos y piensos, tales como en métodos para elaborar productos alimenticios y de piensos o aditivos para alimentos y piensos. En algunas realizaciones, la invención proporciona procesos que usan enzimas de acuerdo con la invención en la industria médica, tal como para la fabricación de productos farmacéuticos o auxiliares o complementos dietéticos o complementos y aditivos alimenticios.

Conversión de biomasa y producción de biocombustibles limpios

La invención proporciona enzimas, tales como aldolasas, incluidas piruvato aldolasas tales como, a modo no taxativo, HMG y/o KHG aldolasas (incluidas mezclas o "cócteles" de enzimas) y métodos para la conversión de una biomasa o cualquier material lignocelulósico (por ejemplo, cualquier composición que comprenda celulosa, hemicelulosa y lignina) en combustibles (por ejemplo, bioetanol, biobutanol, biopropanol, biometanol, biodiésel), usando las enzimas de la invención, además de alimentos para animales y humanos y productos químicos. De esta forma, las composiciones y los métodos de la invención proporcionan alternativas o adyuvantes sustentables para usar como productos a base de petróleo, por ejemplo, una mezcla de bioetanol y gasolina. La invención proporciona organismos que expresan enzimas de la invención para la participación en ciclos químicos que implican la conversión natural de biomasa. En una realización, las enzimas y los métodos para la conversión se usan en conjuntos de enzimas para la despolimerización eficaz de polímeros celulósicos y hemicelulósicos y obtención de restos de carbono metabolizables. La invención proporciona métodos para descubrir e implementar las enzimas más eficaces para permitir estos importantes y nuevos procesos industriales de "conversión de biomasa" y energías alternativas.

Los métodos de la invención también incluyen tomar el material lignocelulósico convertido (procesado por enzimas de la invención) y convertirlo en combustible (por ejemplo, bioetanol, biobutanol, biopropanol, biometanol, biodiesel) mediante fermentación y/o síntesis química. En una realización, los azúcares producidos se fermentan y/o los productos no fermentables se gasifican.

Las enzimas de la invención (incluidos, por ejemplo, organismos, tales como microorganismos, por ejemplo, hongos, levadura o bacterias, que generan y en algunas realizaciones secretan enzimas recombinantes de la invención) pueden usarse o incluirse/integrarse en cualquier etapa de cualquier proceso de conversión de biomasa, por ejemplo, en cualquier paso, en varios pasos o incluirse en todos los pasos o en todos los siguientes métodos de procesos de conversión de biomasa o todas estas alternativas de biocombustible:

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

Pirólisis: es la degradación térmica de biomasa mediante calor en ausencia de oxígeno. En una realización, la biomasa se calienta hasta una temperatura entre aproximadamente 800 y 1400 grados Fahrenheit pero no se introduce oxígeno para apoyar la combustión que resulta en la creación de gas, fuel oil y carbón.

Gasificación: puede usarse biomasa para producir metano a través de calentamiento o digestión anaeróbica. El gas de síntesis, una mezcla de monóxido de carbono e hidrógeno, puede derivar de la biomasa.

5 Gas de vertedero: es generado por la descomposición (digestión anaeróbica) de basura enterrada en vertederos. Cuando el desecho orgánico se descompone genera gas que consiste en aproximadamente 50% de metano, el principal componente del gas natural.

Digestión anaeróbica: convierte materia orgánica en una mezcla de metano, el principal componente del gas natural, y dióxido de carbono. En una realización, la biomasa tal como aguas residuales (aguas servidas), estiércol o desechos del procesamiento de alimentos se mezcla con agua y se alimenta a un tanque digestor sin aire.

#### Fermentación

10 Fermentación de alcohol: el alcohol de combustible se produce convirtiendo almidón en azúcar, fermentando el azúcar para obtener alcohol y luego separando la mezcla de alcohol y agua por destilación. Materias primas tales como trigo, cebada, patata y desperdicios de papel, aserrín y paja, que contienen azúcar, almidón o celulosa, pueden convertirse en alcohol mediante fermentación con almidón.

15 Transesterificación: Una reacción ejemplar para convertir aceite en biodiésel se denomina transesterificación. En el proceso de transesterificación se hace reaccionar un alcohol (como metanol) con aceites de triglicéridos que se encuentran en aceites vegetales, grasas animales o grasas recicladas, formando ésteres de alquilo de ácidos grasos (biodiésel) y glicerina. La reacción requiere de calor y un catalizador de base fuerte, tal como hidróxido de sodio o hidróxido de potasio.

20 Biodiésel: el biodiésel es una mezcla de ésteres de alquilo de ácidos grasos hecha a partir de aceites vegetales, grasas animal o grasas recicladas. El biodiésel puede usarse como combustible para vehículos en su forma pura, pero normalmente se usa como un aditivo para diésel de petróleo para reducir los niveles de particulados, monóxido de carbono, hidrocarburos y productos tóxicos del aire de vehículos con motores diésel.

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

25 Cogeneración: es la producción simultánea de más de una forma de energía usando un solo combustible y las mismas instalaciones. En una realización, la cogeneración de biomasa tiene más crecimiento potencial que la generación de biomasa sola porque la cogeneración produce calor y electricidad.

30 En una realización, los polipéptidos de la invención tienen una actividad de aldolasa, incluida actividad de piruvato aldolasa, tal como, a modo no taxativo, actividad de HMG y/o KHG aldolasa, u otra actividad enzimática para generar biodiésel, bioetanol, biobutanol, biopropanol o biometanol a partir de un material orgánico, por ejemplo, una biomasa, tal como composiciones derivadas de plantas y animales, incluido cualquier cultivo agrícola u otra materia prima renovable, un residuo agrícola o un desecho animal, o los componentes orgánicos de desechos municipales e industriales, o microorganismos tales como algas o levadura. En una realización, los polipéptidos de la invención se usan en procesos para convertir biomasa lignocelulósica en etanol, butanol, propanol, metanol o se usan de otra forma en procesos para hidrolizar o digerir biomateriales de forma tal que pueden usarse como un biocombustible (incluidos bioetanol, biobutanol, biopropanol, biometanol o biodiésel), o para simplificar el procesamiento de la biomasa para obtener un combustible. En una realización alternativa, los polipéptidos de la invención se usan en procesos para un proceso de transesterificación haciendo reaccionar un alcohol (como metanol) con un aceite de triglicéridos en un aceite vegetal, grasa animal o grasas recicladas, formando ésteres de alquilo de ácidos grasos (biodiésel) y glicerina. En una realización, el biodiésel se obtiene a partir de aceite de soja o aceites de cocina reciclados. También pueden usarse grasas animales, otros aceites vegetales y otros aceites reciclados para producir biodiésel, dependiendo de sus costos y disponibilidad. En otras realizaciones se usan mezclas de todos los tipos de grasas y aceites para producir combustible biodiésel de la invención.

45 También pueden usarse enzimas de la invención en la refinación de glicerina. El producto derivado de la glicerina contiene catalizador sin reaccionar y jabones que son neutralizados con un ácido. Se retiran el agua y el alcohol para producir 50% a 80% de glicerina bruta. Los restantes contaminantes incluyen grasas y aceites sin reaccionar, que pueden procesarse utilizando los polipéptidos de la invención. En plantas de biodiésel grandes de la invención, la glicerina puede purificarse adicionalmente, por ejemplo hasta 99% o más de pureza, para las industrias farmacéutica y cosmética.

50 El bioetanol, biobutanol, biopropanol, biometanol y/o biodiésel obtenidos utilizando los polipéptidos de la invención pueden usarse con oxigenadores de combustible para mejorar las características de la combustión. Agregar oxígeno resulta en una combustión más completa, lo que reduce las emisiones de monóxido de carbono. Este es otro beneficio ambiental de reemplazar los combustibles de petróleo por biocombustibles (por ejemplo, un combustible de la invención). Un bioetanol, biobutanol, biopropanol, biometanol y/o biodiésel elaborado utilizando las composiciones y/o métodos de esta invención puede mezclarse con gasolina para formar una mezcla E10 (aproximadamente 5% a 10% de etanol y aproximadamente 90% a 95% de gasolina), pero puede usarse en concentraciones más altas, tales

como E85 o en su forma pura. Un bioetanol, biobutanol, biopropanol, biometanol y/o biodiésel obtenido utilizando las composiciones y/o métodos de esta invención pueden mezclarse con diésel de petróleo para formar una mezcla B20 (20% de biodiésel y 80% de diésel de petróleo), aunque pueden usarse otros niveles de mezcla hasta B100 (biodiésel puro).

5 La invención también proporciona procesos para elaborar etanol ("bioetanol"), butanol ("biobutanol"), propanol ("biopropanol"), metanol ("biometanol") y/o diésel ("biodiésel") a partir de composiciones que comprenden biomasa lignocelulósica. El material de biomasa de lignocelulosa puede obtenerse a partir de cultivos agrícolas, como un  
 10 producto derivado de la producción de alimentos para humanos y animales o como productos de desechos lignocelulósicos, tales como residuos vegetales y papel de desecho. Ejemplos de fuentes vegetales o residuos vegetales adecuados para el tratamiento con polipéptidos de la invención incluyen kelp, algas, granos, semillas,  
 15 tallos, hojas, vainas, cáscaras, mazorcas de maíz, residuos de maíz, paja, pastos (por ejemplo, pasto indio, tal como *Sorghastrum nutans*; o pasto varilla, por ejemplo, la especie *Panicum*, tal como *Panicum virgatum*) y similares, así como madera, astillas de madera, pulpa de madera y aserrín. Ejemplos de desechos de papel adecuados para el tratamiento con polipéptidos de la invención incluyen papel de fotocopias desechado, papel de impresoras de ordenadores, papel de cuadernos, papel de libretas, papel de máquinas de escribir y similares, así como periódicos, revistas, cartón y materiales de embalaje en base a papel.

En una realización, las enzimas y métodos de la invención pueden utilizarse junto con medios más "tradicionales" para obtener etanol, metanol, butanol, propanol y/o diésel a partir de biomasa, por ejemplo, como métodos que comprenden hidrolizar materiales lignocelulósicos sometiendo material lignocelulósico seco en un reactor a un catalizador compuesto por una disolución diluida de un ácido fuerte y una sal de metal; esto puede reducir la energía de activación, o la temperatura, de la hidrólisis de la celulosa para obtener rendimientos de azúcar más altos; ver, por ejemplo, las Patentes de los Estados Unidos Nos. 6.660.506; y 6.423.145.

Otra realización que incorpora el uso de enzimas de la invención comprende hidrolizar material lignocelulósico que contiene hemicelulosa, celulosa y lignina sometiendo el material a un paso de hidrólisis de una primera etapa en un medio acuoso a una temperatura y una presión seleccionadas de forma de producir principalmente la despolimerización de la hemicelulosa sin una despolimerización importante en la conversión de celulosa a glucosa. Esta etapa resulta en una suspensión en la que la fase acuosa líquida contiene monosacáridos disueltos que resultan de la despolimerización de hemicelulosa y una fase sólida que contiene celulosa y lignina. Un paso de hidrólisis de una segunda etapa puede comprender condiciones tales que al menos una porción importante de la celulosa se despolimerice, resultando dicho paso en una fase acuosa líquida que contiene los productos de la despolimerización de celulosa disueltos/solubles. Ver, por ejemplo, la Patente de los Estados Unidos No. 5.536.325. Las enzimas de la invención pueden agregarse en cualquier etapa de este proceso ejemplar.

Otra realización que incorpora el uso de enzimas de la invención comprende procesar un material de biomasa que contiene lignocelulosa mediante una o más etapas de hidrólisis de ácido diluido con aproximadamente 0,4% a 2% de ácido fuerte; y tratar un componente lignocelulósico sólido sin reaccionar del material de biomasa hidrolizado con ácido mediante deslignificación alcalina para producir precursores para termoplásticos y derivados biodegradables. Ver, por ejemplo, la Patente de los Estados Unidos No. 6.409.841. Las enzimas de la invención pueden agregarse en cualquier etapa de este proceso ejemplar.

Otra realización que incorpora el uso de enzimas de la invención comprende prehidrolizar material lignocelulósico en un reactor de prehidrólisis; agregar un líquido ácido al material lignocelulósico sólido para preparar una mezcla; calentar la mezcla hasta la temperatura de reacción; mantener la temperatura de reacción durante un tiempo suficiente como para fraccionar el material lignocelulósico y obtener una porción solubilizada que contenga al menos aproximadamente 20% de la lignina del material lignocelulósico y una fracción sólida que contenga celulosa; retirar una porción solubilizada de la fracción sólida mientras se encuentra a la temperatura de reacción o una temperatura cercana en donde la celulosa en la fracción sólida se vuelve más adecuada para la digestión enzimática; y recuperar una porción solubilizada. Ver, por ejemplo, la Patente de los Estados Unidos No. 5.705.369. Las enzimas de la invención pueden agregarse en cualquier etapa de este proceso ejemplar.

La invención proporciona métodos para obtener composiciones de combustibles para motor (por ejemplo, para motores de encendido por chispa) basadas en hidrocarburos líquidos mezclados con un alcohol de grado combustible obtenido utilizando una enzima o un método de la invención. En una realización, los combustibles obtenidos mediante el uso de una enzima de la invención comprenden, por ejemplo, mezclas de líquido de gas de carbón o líquido de gas natural-etanol, metanol, butanol, propanol y/o diésel. En una realización, un co-disolvente es 2-metiltetrahidrofurano (MTHF) derivado de biomasa. Ver, por ejemplo, la Patente de los Estados Unidos No. 6.712.866.

En una realización, los métodos de la invención para la degradación enzimática de la lignocelulosa, por ejemplo, para la producción de etanol a partir de material lignocelulósico, también pueden comprender el uso de tratamiento ultrasónico del material de biomasa; ver, por ejemplo, la Patente de los Estados Unidos No. 6.333.181.

5 En otra realización, los métodos de la invención para producir bioetanol, biobutanol, biopropanol, biometanol y/o biodiésel a partir de un sustrato celulósico comprenden proporcionar una mezcla de reacción en forma de una suspensión que comprende un sustrato celulósico, una enzima de esta invención y un agente de fermentación (por ejemplo, dentro de un recipiente de reacción, tal como un biorreactor de alimentación semi-continua de sólidos), y la mezcla de reacción se hace reaccionar en condiciones suficientes como para iniciar y mantener una reacción de fermentación (como se describe, por ejemplo, en la Solicitud de Patente de los Estados Unidos No. 20060014260). En una realización, los cálculos experimentales o teóricos pueden determinar una frecuencia de alimentación óptima. En una realización se proporcionan cantidades adicionales del sustrato celulósico y la enzima al recipiente de reacción a intervalos de acuerdo con la frecuencia de alimentación optimizada.

10 Un proceso ejemplar para la fabricación de biocombustibles (tal como bioetanol, biobutanol, biopropanol, biometanol, y/o biodiésel) de la invención se describe en las Solicitudes de Patente de los Estados Unidos Publicadas Nos. 20050069998; 20020164730; y en una realización comprende las etapas de moler la biomasa lignocelulósica (por ejemplo, hasta un tamaño de 15-30 mm), someter el producto obtenido a un pre-tratamiento de explosión de vapor (por ejemplo, a una temperatura de 190-230°C) durante entre 1 y 10 minutos en un reactor; recoger el material pre-tratado en un ciclón o producto de manufactura relacionado; y separar las fracciones líquida y sólida por filtración en un filtro prensa, introducir la fracción sólida en un depósito de fermentación y agregar una o más enzimas de la invención, por ejemplo, una celulasa y/o beta-glucosidasa (por ejemplo, disuelta en solución amortiguadora de citrato de pH 4,8).

20 Otro proceso ejemplar para fabricar biocombustibles (tales como bioetanol, biobutanol, biopropanol, biometanol y/o biodiésel) de la invención que comprende utilizar enzimas de la invención implica pre-tratar un material de inicio que comprende una materia prima lignocelulósica que comprende al menos hemicelulosa y celulosa. En una realización, el material de inicio comprende patatas, soja (colza), cebada, centeno, maíz, avenas, trigo, remolacha o remolacha azucarera o un componente o desecho o producto derivado de la producción de alimentos o piensos. El material de inicio ("materia prima") se hace reaccionar en condiciones que alteran la estructura de la fibra de la planta para producir al menos una hidrólisis parcial de la hemicelulosa y la celulosa. Las condiciones de alteración pueden comprender, por ejemplo, someter el material de inicio a una temperatura promedio de 180°C a 270°C a un pH de 0,5 a 2,5 durante un periodo de aproximadamente 5 segundos a 60 minutos; o una temperatura de 220°C a 270°C a un pH de 0,5 a 2,5 durante un periodo de 5 segundos a 120 segundos o equivalente. Esto genera una materia prima con mayor accesibilidad para su digestión por parte de una enzima, por ejemplo, una enzima celulasa de la invención. Patente de los Estados Unidos No. 6.090.595.

25 Condiciones ejemplares para la hidrólisis de material lignocelulósico incluyen reacciones a temperaturas entre aproximadamente 30°C y 48°C y/o un pH entre aproximadamente 4,0 y 6,0. Otras condiciones ejemplares incluyen una temperatura entre aproximadamente 30°C y 60°C y un pH entre aproximadamente 4,0 y 8,0.

30 Las enzimas de acuerdo con la invención pueden catalizar reacciones con gran estereo-, regio- y quimio-selectividad. Las enzimas aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención pueden manipularse para funcionar en distintos disolventes, operar con pHs extremos (por ejemplo, pHs altos y pHs bajos), temperaturas extremas (por ejemplo, altas temperaturas y bajas temperaturas), niveles de salinidad extremos (por ejemplo, alta salinidad y baja salinidad) y catalizar reacciones con compuestos que no están estructuralmente relacionados con sus sustratos fisiológicos naturales.

#### 40 Piensos y alimentos o aditivos para piensos y alimentos

Además de proporcionar auxiliares o complementos dietéticos, o complementos y aditivos alimenticios, la invención también proporciona composiciones y métodos para tratar piensos y alimentos humanos y animales y aditivos para alimentos y piensos usando un polipéptido de acuerdo con la invención, tal como una proteína que tiene actividad de enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención, y/o los anticuerpos de acuerdo con la invención. En algunas realizaciones, la invención proporciona piensos para animales, alimentos y aditivos que comprenden enzimas aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención y/o anticuerpos de acuerdo con la invención. El animal puede ser cualquier animal de granja o cualquier animal.

50 El aditivo para pienso animal de acuerdo con la invención puede ser un producto de la enzima granulado que puede mezclarse fácilmente con los componentes del pienso. De forma alternativa, los aditivos para pienso de acuerdo con la invención pueden formar un componente de una pre-mezcla. El producto de la enzima granulado de acuerdo con la invención puede estar recubierto o sin recubrir. El tamaño de las partículas de los granulados de la enzima puede ser compatible con el de los componentes del pienso y la pre-mezcla. Esto proporciona un medio seguro y conveniente para incorporar enzimas en los piensos. Alternativamente, el aditivo para pienso animal de acuerdo con la invención puede ser una composición líquida estabilizada. Esta puede ser una suspensión de base acuosa u oleosa. Ver la Patente de los Estados Unidos No. 6.245.546.

55 Las enzimas aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de la presente invención, en la modificación de pienso o un alimento, pueden procesar el alimento o pienso in vitro (modificando los componentes

del pienso o alimento) o in vivo. Los polipéptidos de acuerdo con la invención pueden agregarse a composiciones de pienso o alimento.

5 En algunas realizaciones se agrega una enzima de acuerdo con la invención en combinación con otra enzima, tal como beta-galactosidasas, catalasas, lacasas, celulasas, otras aldolasas, endoglicosidasas, endo-beta-1,4-lacasas, amiloglicosidasas, glucosidasas, glucosa isomerasas, glicosiltransferasas, lipasas, fosfolipasas, lipooxigenasas, beta-lacasas, endo-beta-1,3(4)-lacasas, cutinasas, peroxidadas, amilasas, fitasas, glucoamilasas, pectinasas, reductasas, oxidasas, descarboxilasas, fenoloxidasas, ligninasas, pululanasas, arabinanasas, hemicelulasas, mananasas, xilolacasas, xilanasas, pectín acetil esterases, rhamnogalacturonan acetil esterases, proteasas, peptidasas, proteinasas, poligalacturonasas, rhamnogalacturonasas, galactanasas, pectín liasas, transglutaminasas, pectín metilesterasas, celobiohidrolasas y/o transglutaminasas. Estos productos de digestión enzimáticos son más digeribles por el animal. De esta forma, las enzimas aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención pueden contribuir con la energía disponible del pienso o alimento o con la digestibilidad del alimento o pienso por medio de la descomposición de la celulosa.

15 En otras realizaciones, la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención puede suministrarse expresando las enzimas directamente en cosechas de pienso transgénicas (tales como plantas y semillas transgénicas y similares), tales como granos, cereales, maíz, soja, colza, lupino y similares. Como se estableció anteriormente, la invención proporciona plantas, partes de plantas y células de plantas transgénicas que comprenden una secuencia de ácidos nucleicos que codifican un polipéptido de acuerdo con la invención. En algunas realizaciones, el ácido nucleico se expresa de forma tal que la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa de acuerdo con la invención se produce en cantidades recuperables. La enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa puede recuperarse de cualquier planta o parte de planta. De forma alternativa, la planta o parte de planta que contiene el polipéptido recombinante puede usarse como tal para mejorar la calidad de un alimento o pienso, como por ejemplo mejorar el valor nutricional, la palatabilidad, etc.

25 En algunas realizaciones, la matriz de administración de enzimas de acuerdo con la invención se encuentra en forma de partículas plurales discretas, granulados o gránulos. "Gránulos" significa partículas que están comprimidas o compactadas, por ejemplo mediante granulación, extrusión o compactación similar para retirar agua de la matriz. Dicha compresión o compactación de las partículas también promueve la cohesión interpartícula de las partículas. Por ejemplo, los gránulos pueden prepararse mediante granulación del sustrato en base a granos en un molino para granulados. Los granulados preparados de esta forma se muelen o desmenuzan hasta un tamaño de gránulo adecuado para su uso como adyuvante en pienso animal. Dado que está aprobada para su uso en pienso animal, la matriz puede usarse como diluyente para la administración de enzimas en pienso animal.

35 En algunas realizaciones, la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa contenida en la matriz de administración de enzimas y en los métodos de la invención es una enzima aldolasa termoestable, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa, tal como se describe en la presente, de forma de resistir la inactivación de la enzima aldolasa, tal como piruvato aldolasa, tal como HMG y/o KHG aldolasa durante la fabricación, en la que pueden emplearse temperaturas elevadas y/o vapor para preparar la matriz de administración de enzimas granulada. Durante la digestión de pienso que contiene la matriz de administración de enzimas de la invención, los fluidos digestivos acuosos provocarán la liberación de la enzima activa. También pueden incorporarse otros tipos de enzimas termoestables y complementos nutricionales que sean termoestables en la matriz de administración que han de ser liberados en cualquier tipo de condiciones acuosas.

45 En algunas realizaciones se aplica un recubrimiento a las partículas de la matriz de la enzima con muchos fines diferentes, tales como agregar un sabor o complemento nutricional al pienso animal, retrasar la liberación de complementos y enzimas en el pienso animal en condiciones gástricas y similares. En algunas realizaciones, el recubrimiento se aplica para alcanzar un objetivo funcional, por ejemplo, cuando es deseable ralentizar la liberación de la enzima desde las partículas de la matriz o para controlar las condiciones en las que se liberará la enzima. La composición del material del recubrimiento puede ser tal que el mismo sea degradado selectivamente por el agente al cual es susceptible (tal como calor, ácido o base, enzimas u otro producto químico). De forma alternativa pueden aplicarse consecutivamente a las partículas de la matriz dos o más recubrimientos susceptibles a dichos diferentes agentes de degradación.

55 La invención también se dirige a un proceso para preparar una matriz de liberación de enzimas. De acuerdo con la invención, el proceso comprende proporcionar partículas plurales discretas de un sustrato en base a granos en un tamaño de partícula adecuado para su uso como una matriz liberadora de enzimas, en donde las partículas comprenden una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa codificada por una secuencia de aminoácidos de acuerdo con la invención. En algunas realizaciones, el proceso incluye compactar o comprimir las partículas de la matriz liberadora de enzimas para formar gránulos, lo cual en algunas realizaciones se logra mediante granulación. El agente inhibidor de moho y de cohesión, en caso de usarse, puede agregarse en cualquier momento adecuado y, en algunas realizaciones, mezclarse con el sustrato en base a granos en las proporciones deseadas antes de la granulación del sustrato en base a granos. El contenido de humedad en el molino de granulado en algunas realizaciones se encuentra en los rangos establecidos anteriormente con respecto al



contenido de humedad en el producto terminado y, en algunas realizaciones, es de aproximadamente 14-15%. En algunas realizaciones se agrega humedad a la materia prima en forma de una preparación acuosa de la enzima para ajustar la materia prima a este contenido de humedad. La temperatura en el molino de granulado en algunas realizaciones se lleva hasta aproximadamente 82°C con vapor. El molino de granulado puede operarse en cualquier condición que imparta trabajo suficiente a la materia prima para proporcionar granulados. El proceso de granulado en sí mismo es un proceso económico para eliminar agua de la composición que contiene las enzimas.

Las composiciones y los métodos de acuerdo con la invención pueden ponerse en práctica conjuntamente con la administración de prebióticos, que son azúcares de alto peso molecular, tales como fructo-oligosacáridos (FOS); galacto-oligosacáridos (GOS), material GRAS (generalmente reconocido como seguro). Estos prebióticos pueden ser metabolizados por algunas bacterias ácido lácticas (LAB) probióticas. No son digeribles por la mayoría de los microbios intestinales.

#### Tratamiento y procesamiento de alimentos

La invención proporciona alimentos y piensos que comprenden enzimas de acuerdo con la invención y métodos para usar enzimas de acuerdo con la invención en el procesamiento de alimentos y piensos. Las enzimas aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de acuerdo con la invención tienen numerosas aplicaciones en la industria del procesamiento de alimentos. En algunas realizaciones, la invención proporciona métodos para hidrolizar composiciones que comprenden celulosa incluyendo, por ejemplo, una célula de planta, una célula bacteriana, una célula de levadura, una célula de insecto o un animal o cualquier planta o parte de planta o cualquier alimento o pienso o un producto de desecho y similares.

Por ejemplo, la invención proporciona alimentos o piensos que comprenden una enzima aldolasa, tal como piruvato aldolasa, HMG y/o KHG aldolasa de la invención, tal como en un pienso, un líquido, tal como una bebida (tal como un jugo de fruta o una cerveza), un pan o una masa o un producto panificado o un precursor de una bebida (tal como una cerveza o mosto).

Los procesos de tratamiento de alimentos de acuerdo con la invención también pueden incluir el uso de cualquier combinación de otras enzimas tales como triptofanasas o tirosina descarboxilasas, lacasas, catalasas, lacasas, otras aldolasas, celulasas, endoglicosidasas, endo-beta-1,4-lacasas, amiloglicosidasas, glucosidasas, glucosa isomerasas, glicosiltransferasas, lipasas, fosfolipasas, lipooxigenasas, beta-lacasas, endo-beta-1,3(4)-lacasas, cutinasas, peroxidasas, amilasas, fitasas, glucoamilasas, pectinasas, reductasas, oxidasas, descarboxilasas, fenoloxidasas, ligninasas, pululanasas, arabinanasas, hemicelulasas, mananasas, xilolacasas, xilanasas, pectín acetil estererasas, rhamnogalacturonan acetil estererasas, proteasas, peptidasas, proteinasas, poligalacturonasas, rhamnogalacturonasas, galactanasas, pectín liasas, transglutaminasas, pectín metilesterasas, celobiohidrolasas y/o transglutaminasas

#### Composiciones farmacéuticas y complementos dietéticos

La divulgación se refiere, además, a composiciones farmacéuticas y complementos dietéticos (tales como auxiliares dietéticos) que comprenden una aldolasa de acuerdo con la invención. La actividad de aldolasa comprende actividad de piruvato aldolasa, HMG y/o KHG aldolasa. En algunos casos, las composiciones farmacéuticas y los complementos dietéticos (tales como auxiliares dietéticos) se formulan para la ingestión oral.

Los compuestos de tratamiento periodontal pueden comprender una enzima de acuerdo con la invención, tal como se describe en la Patente de los Estados Unidos No. 6.776.979. Las composiciones y métodos para el tratamiento o la profilaxis del síndrome del intestino ácido pueden comprender una enzima de acuerdo con la invención, tal como se describe en la Patente de los Estados Unidos No. 6.468.964.

En otros casos, apósitos para heridas, implantes y similares comprenden enzimas antimicrobianas (tales como antibióticas), incluida una enzima de acuerdo con la invención (incluidas, por ejemplo, secuencias de acuerdo con la invención). Las enzimas de acuerdo con la invención también pueden usarse en apósitos de alginato, apósitos de barrera antimicrobiana, apósitos para quemaduras, vendas vendas de compresión, herramientas de diagnóstico, apósitos de gel, apósitos hidrocélulares (de espuma), apósitos hidrocoloides, apósitos I.V., campos de incisión, apósitos de baja adherencia, apósitos absorbentes de olores, vendas de pasta, apósitos post-operatorios, manejo de cicatrices, cuidado de la piel, apósitos de película transparente y/o cicatrización de heridas. Las enzimas pueden usarse en la limpieza de heridas, en la preparación del lecho de heridas, para tratar úlceras por presión, úlceras en las piernas, quemaduras, úlceras de pie diabético, cicatrices, fijación I.V., heridas quirúrgicas y heridas menores. Las enzimas pueden usarse en composiciones de desbridamiento enzimático estériles, tales como ungüentos. En varios casos, la aldolasa se formula como un comprimido, gel, píldora, implante, líquido, aerosol, película, micela, polvo, alimento, granulado de pienso o como una formulación encapsulada.

Las composiciones farmacéuticas y los complementos dietéticos también pueden incluir el uso de cualquier combinación de otras enzimas, tales como beta-galactosidasas, catalasas, lacasas, celulasas, otras aldolasas, endoglicosidasas, endo-beta-1,4-lacasas, amiloglicosidasas, glucosidasas, glucosa isomerasas,

5 glicosiltransferasas, lipasas, fosfolipasas, lipooxigenasas, beta-lacasas, endo-beta-1,3(4)-lacasas, cutinasas, peroxidasas, amilasas, fitasas, glucoamilasas, pectinasas, reductasas, oxidasas, descarboxilasas, fenoloxidasas, ligninasas, pululaninas, arabinanasas, hemicelulasas, mananasas, xilolacasas, xilanasas, pectín acetil esterasas, rhamnogalacturonan acetil esterasas, proteasas, peptidasas, proteínasas, poligalacturonasas, rhamnogalacturonasas, galactanasas, pectín liasas, transglutaminasas, pectín metilesterasas, celobiohidrolasas y/o transglutaminasas.

Vías biosintéticas para producir R,R y otros estereoisómeros de monatina

10 Tal como se describe, entre otros, en el documento WO 03/091396 A2 (ver las Figuras 1-3 y 11-13), la monatina puede producirse a partir de triptófano a través de una vía de múltiples etapas que implica conversiones biológicas (es decir, que facilita la reacción de un sustrato para formar un producto con un polipéptido). Una vía descrita implica convertir biológicamente el triptófano en indol-3-piruvato, convertir biológicamente indol-3-piruvato en ácido 2-hidroxi 2-(indol-3-ilmetil)-4-ceto glutárico ("MP") y convertir biológicamente el MP en monatina. En algunas realizaciones, los polipéptidos de la invención pueden usarse para facilitar la reacción de indol-3-piruvato para formar MP. En algunas realizaciones, los polipéptidos de la invención pueden usarse para facilitar preferentemente la producción de R-MP.

15 En algunas realizaciones, un polipéptido seleccionado de polipéptidos aislados o recombinantes de la SEQ ID NO:276 o fragmentos o subsecuencias de los mismos que tienen actividad de aldolasa pueden ser útiles para facilitar una reacción dentro de una vía de múltiples etapas para producir un producto seleccionado de monatina, derivados de monatina, sales de los mismos y combinaciones de los mismos. En una realización, los polipéptidos con actividad de aldolasa pueden ser útiles para facilitar una reacción en la cual indol-3-piruvato se convierte en MP como una etapa dentro de una vía de múltiples etapas para producir un producto seleccionado de monatina, derivados de monatina, sales de los mismos y combinaciones de los mismos.

20 En otra realización, un polipéptido seleccionado de polipéptidos aislados o recombinantes con actividad de HMG aldolasa de la SEQ ID NO:276 o fragmentos o subsecuencias de los mismos que tienen actividad de aldolasa pueden ser útiles para facilitar una reacción entre indol-3-piruvato y una fuente de carbono C3 como una etapa dentro de una vía de múltiples etapas para producir un producto seleccionado de monatina, derivados de monatina, sales de los mismos y combinaciones de los mismos. En una realización, los polipéptidos con actividad de HMG aldolasa pueden ser útiles para facilitar una reacción en la cual indol-3-piruvato se convierte en MP como una etapa dentro de una vía de múltiples etapas para producir un producto seleccionado de monatina, derivados de monatina, sales de los mismos y combinaciones de los mismos.

25 Adicionalmente, un polipéptido codificado por una secuencia de ácidos nucleicos que tiene al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más o identidad de secuencia completa (100%) con un ácido nucleico de la SEQ ID NO:275 en todo el largo completo de un ADNc, transcripto (ARNm) o gen, o más residuos pueden ser útiles para facilitar una reacción entre indol-3-piruvato y una fuente de carbono C3 como una etapa dentro de una vía de múltiples etapas para producir un producto seleccionado de monatina, derivados de monatina, sales de los mismos o combinaciones de los mismos. En una realización, el o los polipéptidos, o fragmentos o subsecuencias de los mismos con actividad de aldolasa pueden ser útiles para facilitar una reacción en la cual indol-3-piruvato se convierte en MP como una etapa dentro de una vía de múltiples etapas para producir un producto seleccionado de monatina, derivados de monatina, sales de los mismos y combinaciones de los mismos.

30 En otra realización de la invención, uno o más polipéptidos con actividad de aldolasa de HMG codificada por una secuencia de ácidos nucleicos que tiene al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más o identidad de secuencia más completa (100%) con un ácido nucleico de SEQ ID NO. 275 en la longitud completa de ADNc, transcripto (ARNm) o gen pueden ser útiles para facilitar una reacción entre indol-3-piruvato y una fuente de carbono C3 como una etapa dentro de una vía de múltiples etapas para producir un producto seleccionado de monatina, derivados de monatina, sales de la misma o combinaciones de la misma. En una realización, el o los polipéptidos con actividad de aldolasa HMG pueden ser útiles para facilitar una reacción en la cual indol-3-piruvato se convierte en MP como una etapa dentro de una vía de múltiples etapas para producir un producto seleccionado de monatina, derivados de monatina, sales de los mismos y combinaciones de los mismos.

35 Asimismo, un polipéptido con actividad de aldolasa codificada por una secuencia de ácidos nucleicos que hibrida en condiciones rigurosas con un ácido nucleico de la SEQ ID NO:275 pueden ser útiles para facilitar una reacción entre indol-3-piruvato y una fuente de carbono C3 como una etapa dentro de una vía de múltiples etapas para producir un producto seleccionado de monatina, derivados de monatina, sales de los mismos y combinaciones de los mismos. En una realización, el o los polipéptidos con actividad de aldolasa pueden ser útiles para facilitar una reacción en la cual indol-3-piruvato se convierte en MP como una etapa dentro de una vía de múltiples etapas para producir un producto seleccionado de monatina, derivados de monatina, sales de los mismos y combinaciones de los mismos.

40 En otra realización de la invención, un polipéptido con actividad de HMG aldolasa codificado por una secuencia de ácidos nucleicos que se hibrida en condiciones rigurosas con un ácido nucleico de la SEQ ID NO:275, puede ser útil

para facilitar una reacción entre indol-3-piruvato y una fuente de carbono C3 como una etapa dentro de una vía de múltiples etapas para producir un producto seleccionado de monatina, derivados de monatina, sales de los mismos y combinaciones de los mismos. En una realización, el polipéptido con actividad de HMG aldolasa puede ser útil para facilitar una reacción en la cual indol-3-piruvato se convierte en MP como una etapa dentro de una vía de múltiples etapas para producir un producto seleccionado de monatina, derivados de monatina, sales de los mismos y combinaciones de los mismos.

Los polipéptidos con actividad de aldolasa descritos en la presente en la presente pueden ser útiles para facilitar una reacción entre indol-3-piruvato y una fuente de carbono C3. La fuente de carbono C3 puede ser, a modo no taxativo, oxaloacetato, piruvato o un derivado de piruvato, tal como fosfoenolpiruvato. En una realización, la fuente de carbono C3 es piruvato.

Ejemplos de enzimas útiles para la conversión del producto de reacción entre indol-3-piruvato y la fuente de carbono C3 a monatina incluyen miembros de las clases de enzimas: triptófano aminotransferasas (2.6.1.27), triptófano deshidrogenasas (1.4.1.19), D-aminoácido deshidrogenasas (1.4.99.1), glutamato deshidrogenasas (1.4.1.2-4), fenilalanina deshidrogenasa (EC 1.4.1.20), triptófano-fenilpiruvato transaminasas (2.6.1.28), o más generalmente miembros de la familia de la aminotransferasa (2.6.1.-) tal como aspartato aminotransferasa (EC 2.6.1.1), tirosina aminotransferasa (aromática) (2.6.1.5), D-triptófano aminotransferasa o D-alanina aminotransferasa (2.6.1.21) (ver la Figura 2 del documento WO 03/091396 A2). Esta reacción también puede llevarse a cabo usando reacciones químicas. La aminación del cetoácido (MP) se lleva a cabo por aminación reductiva usando amoniaco y cianoborohidruro de sodio. Las Figuras 11-13 del documento WO 03/091396 A2 muestran polipéptidos adicionales que pueden usarse para convertir MP a monatina, así como para proporcionar mayores rendimientos de monatina a partir de indol-3-piruvato o triptófano. En una realización, estas enzimas se utilizan para catalizar la conversión de MP, el producto de reacción entre indol-3-piruvato y piruvato, a monatina.

El perfil de sabor de una composición de monatina puede alterarse controlando la cantidad relativa de los varios estereoisómeros de monatina en la composición. La presente invención proporciona vías y sustancias para producir composiciones de monatina con un porcentaje deseado de Monatina R,R y/o monatina S,R.

La quiralidad de los compuestos de monatina producidos por las vías divulgadas puede verse afectada tanto por pH como por los polipéptidos usados para las conversiones biológicas. Los polipéptidos con actividad de aldolasa descritos en la presente, pueden utilizarse para controlar la quiralidad del carbono-2 de monatina (ver la Fórmula I anterior) en la reacción en la cual indol-3-piruvato se convierte en MP.

Una vez que el producto de reacción de la reacción entre indol-3-piruvato y la fuente de carbono C3 se produce, puede agregarse el grupo amino estereoespecíficamente. La configuración R o S del carbono-4 (ver la Fórmula I anterior) puede generarse dependiendo de si se usa la ácido aromático D o L aminotransferasa. Muchas aminotransferasas son específicas para el isómero L; sin embargo, las D-triptófano aminotransferasas existen en ciertas plantas (Kohiba y Mito, Procedimientos del 8vo Simposio Internacional sobre Vitamina B<sub>6</sub> y catálisis de carbonilo, Osaka, Japón 1990). Además, se han identificado las D-alanina aminotransferasas (2.6.1.21), D-metionina-piruvato aminotransferasas (2.6.1.41) y (R)-3-amino-2-metilpropanoato aminotransferasas (2.6.1.61), (S)-3-amino-2-metilpropanoato aminotransferasa (2.6.1.22) y D-fenilglicina aminotransferasa. Ciertas aminotransferasas pueden aceptar sólo el sustrato por esta reacción con una configuración particular en el carbono C2. Por lo tanto, incluso si la conversión al producto de reacción entre indol-3-piruvato y la fuente de carbono 3 no es estereoespecífica, la estereoquímica del producto final puede controlarse a través de la selección apropiada de aminotransferasa. Dado que las reacciones son reversibles, el producto de reacción sin reaccionar (isómero no deseado) puede reciclarse nuevamente en forma de sus constituyentes y una mezcla racémica del producto de reacción puede volver a formarse.

Un ejemplo de un donante amino adecuado para la adición de un grupo amino a un producto de la reacción entre el indol-3-piruvato y la fuente de carbono C3 incluye, a modo no taxativo, un aminoácido, tal como alanina, aspartato, lisina, glutamato, glicina y triptófano.

Con referencia ahora a las figuras, cabe señalar lo siguiente. Las gráficas de flujo identifican vías para producir monatina, pero no se limitan a un método particular para poner en práctica las vías. Por ejemplo, las vías pueden ponerse en práctica in vivo, in vitro o una combinación de los mismos.

Asimismo, poner en práctica las vías no requiere que cada uno de los componentes identificados (tales como reactivos y enzimas) sean proporcionados explícitamente por el profesional de la salud, en la medida en que se proporcionen suficientes componentes, o fuentes de compuestos, y condiciones de reacción de forma tal que la vía pueda proceder potencialmente. En otras palabras, por ejemplo, si una figura representa un proceso para producir una composición de monatina, que incluye producir indol-3-piruvato a partir de L-triptófano, producir ácido 2-hidroxi 2-(indol-3-ilmetil)-4-ceto glutárico ("precursor de monatina" o "MP") a partir de indol-3-piruvato, y producir monatina a partir de MP, en donde cada reacción se facilita por medio de una enzima apropiada, se contempla que poner en práctica la vía incluye combinar L-triptófano con  $\alpha$ -cetoglutarato y enzimas contempladas para facilitar las reacciones identificadas, y en condiciones adecuadas para que cada una de las reacciones ocurra sin proporcionar

explícitamente también indol-3-piruvato o MP. En ese caso, el L-triptófano podría reaccionar con  $\alpha$ -cetoglutarato para producir indol-3-piruvato. Debido a las condiciones fijadas y la enzima proporcionada, el indol-3-piruvato producido a partir de la reacción de L-triptófano podría reaccionar para formar MP, y luego debido a las condiciones establecidas y la enzima proporcionada, el MP producido a partir de la reacción de indol-3-piruvato podría reaccionar para formar monatina.

También cabe señalar que las vías representadas no requieren que el profesional de la salud proporcione explícitamente los materiales de partida o enzimas identificados. En otras palabras, se contempla que la práctica de cualquiera de las vías que identifican L-triptófano como un material de partida incluirían proporcionar un compuesto que puede producir L-triptófano, en condiciones adecuadas para que ocurra la producción de L-triptófano y combinar dicho compuesto con enzimas capaces de facilitar la serie de reacciones establecidas en condiciones que serían adecuadas para que ocurran dichas reacciones. Como otro ejemplo, se contempla también que poner en práctica la vía identificada incluiría proporcionar un microorganismo genéticamente manipulado para producir monatina de acuerdo con la vía descrita, y proporcionar condiciones apropiadas para que ocurra el proceso de fermentación. Por ejemplo, un microorganismo, que naturalmente produce grandes cantidades de L-triptófano, podría manipularse genéticamente para producir o sobreproducir una o más de las enzimas usadas para facilitar las reacciones en la vía a monatina, y se podrían proporcionar condiciones apropiadas para que así el microorganismo produzca monatina.

La Figura 1 identifica la realización particular en la que una aldolasa R-específica facilita la reacción de indol-3-piruvato y piruvato para formar R-MP. La gráfica de flujo de la Figura 1 representa esquemáticamente un proceso de acuerdo con la invención para realizar una composición de monatina que incluye monatina R,R. Como se muestra en la Figura 1, la vía general implica una reacción de triptófano para formar indol-3-piruvato para producir MP, y una reacción de MP para producir monatina, incluida monatina R,R.

La Figura 1 ilustra además permutaciones específicas de esta vía general, diseñadas para aumentar la producción de la forma R,R de monatina a costa de las formas S,S, R,S y S,R de monatina. En particular, la Figura 1 ilustra la realización en donde: la enzima aminotransferasa utilizada en la reacción de L-triptófano tiene mayor actividad y/o especificidad para dicha reacción con respecto a las reacciones de MP y monatina 4S o la oxidasa tiene mayor actividad y/o especificidad para el L-triptófano que para la monatina 4R; la enzima que facilita la reacción de indol-3-piruvato es un polipéptido con actividad de aldolasa divulgado en la presente y la enzima que facilita la reacción de MP es una enzima D de especificidad amplia, preferiblemente evolucionada para funcionar de manera más eficiente con el isómero R de MP.

La Figura 1 también ilustra permutaciones particulares diseñadas para hacer más económica la producción de monatina R,R. Por ejemplo, en la Figura 1, el L-triptófano— a diferencia del D-triptófano o combinaciones de L- y D-triptófano—se identifica como el material de partida. Mientras que la elección de la forma específica de triptófano no afecta la quiralidad de los compuestos de monatina finales en la composición de monatina (porque la reacción de triptófano forma indol-3-piruvato, que no tiene quiralidad), algunos pueden preferir utilizar L-triptófano como material de partida al menos porque el L-triptófano es actualmente menos costoso y puede obtenerse más fácilmente que el D-triptófano.

Enfocándonos ahora en la primera reacción que se muestra en la Figura 1, cuando el triptófano se convierte en indol-3-piruvato cualquiera de uno o más de  $\alpha$ -cetoglutarato, oxaloacetato y piruvato reacciona para formar un aminoácido (glutamato, aspartato y alanina, respectivamente). La Figura 1 representa la realización en donde el material de partida de triptófano es L-triptófano, y el  $\alpha$ -cetoglutarato, oxaloacetato y/o piruvato producen la forma isomérica L del aminoácido (tal como L-glutamato, L-aspartato y/o L-alanina, respectivamente).

Como se muestra en la Figura 1, un abordaje para mejorar la producción de derivados de monatina R,R implica facilitar la reacción de L-triptófano con una enzima que tiene mayor especificidad, mayor actividad o ambas para triptófano a diferencia de MP o monatina, y facilita la reacción de MP con una enzima D. Como se divulga en el documento WO 03/091396 A2, ciertas enzimas pueden facilitar la reacción de triptófano para producir indol-3-piruvato, así como la reacción de aminación de MP para producir monatina. El uso de una L-aminotransferasa en la etapa de aminación crea un centro quiral S en la posición C-4 de monatina, mientras que el uso de una enzima D crea un centro quiral en la posición C-4 de monatina. Por lo tanto, en el caso en el que L-aminotransferasa, que facilita la reacción de triptófano, es también activo en la reacción de MP, puede formarse monatina R,S y S,S, dependiendo de la forma de MP presente. Además, ciertas otras enzimas—las L-aminoácido oxidasas—pueden no sólo facilitar la reacción de triptófano para formar indol-3-piruvato, pero puede tener una actividad secundaria para la degradación de monatina R,R. De acuerdo con algunas realizaciones, esta actividad secundaria de 4R es minimizada o eliminada. Una actividad secundaria de oxidasa en las formas 4S de monatina las reduciría o minimizaría del producto final y podría ser deseable dependiendo en la composición final deseada. Por consiguiente, cuanto mayor sea la especificidad y/o actividad de la enzima L seleccionada para triptófano con respecto al MP o monatina, mayor será la cantidad de R,R y S,R producida con respecto a la monatina S,S y R,S.

Enzimas adecuadas para la reacción de triptófano, de acuerdo con la realización ilustrada en la Figura 1, incluyen: L-aminotransferasas capaces de facilitar una reacción de L-triptófano para formar indol-3-piruvato, y que tienen mayor especificidad para dicha reacción con respecto a la reacción de R-MP para formar isómeros 4S de monatina; y, L-

aminoácido oxidasas capaces de facilitar una reacción de L-triptófano para formar indol-3-piruvato, y que tienen mayor especificidad y/o actividad para dicha reacción con respecto a la reacción de isómeros 4R de monatina para formar MP, y equivalentes funcionales de cualquiera de los anteriores. Más específicamente, ejemplos no taxativos de enzimas adecuadas pueden seleccionarse de L-triptófano aminotransferasas (E.C. 2.6.1.27) y tirosina aminotransferasas (aromáticas) (EC 2.6.1.5) y L-aminoácido oxidasas (EC 1.4.3.2), y mutantes derivados de enzimas que tienen actividad de aspartato aminotransferasa.

El Ejemplo 16 identifica una enzima específica, un polipéptido HEXaspC mutante que incluye una sustitución de Pro 9 por Tyr y una sustitución de Arg 122 por Gly útil para facilitar las reacciones de L-triptófano y  $\alpha$ -KG, oxaloacetato, piruvato o combinaciones de los mismos para formar indol-3-piruvato y L-glutamato, L-aspartato y L-alanina, respectivamente. Otra enzima específica que tiene actividad "limitada" es TatA, la triptófano aminotransferasa de *S. meliloti*. Otras enzimas adecuadas para la reacción de triptófano de acuerdo con realizaciones preferidas de la vía que se muestra en la Figura 1 incluyen aquellas con las siguientes características: una enzima que transamina MP a 1/10 la tasa o menos que la tasa de L-triptófano como en el Ejemplo 16 o una enzima cuando se usa con una racemasa, como en Ejemplo 18, que produce más de 90% de los isómeros 4R de monatina.

Ejemplos de enzimas que no tienen mayor especificidad para la conversión de L-triptófano a indol-3-piruvato en comparación con la conversión de MP a monatina incluyen: HEXaspC (Ejemplo 16), aminotransferasa de especificidad amplia de *Leishmania major* (WO 03/091396 A2), la aminotransferasa de porcinos (WO 03/091396 A2) y TatA de *Rhodobacter sphaeroides* (Ejemplo 18). Sin embargo, estas enzimas pueden evolucionar por ejemplo a través de mutagénesis para tener actividad limitada para R-MP y/o monatina R,R con respecto al triptófano.

Si nos centramos ahora en la segunda reacción identificada en la Figura 1, la elección de la enzima para facilitar la reacción de indol-3-piruvato para formar MP determina la cantidad relativa de monatina R,R con respecto a la monatina S,R producida. En general, cuanto mayor sea la cantidad relativa de R-MP con respecto al S-MP producido, mayor será la cantidad relativa de monatina R,R con respecto a la monatina S,R producida (cuando una D-enzima facilita la reacción de MP para formar monatina). Cuando se desea una composición de monatina que tiene la forma R,R de monatina como su único componente de monatina, debería usarse una enzima que produce selectivamente R-MP a diferencia de S-MP (una "enzima R-específica"). Los polipéptidos con actividad de aldolasa descrita en la presente son útiles para producir selectivamente R-MP, a diferencia de S-MP. Varios ejemplos de enzimas aldolasa R-específicas se demuestran en la Tabla 1, anterior, en los Ejemplos 4, 5 y 6, más adelante, y en el Listado de Secuencias.

Si nos centramos ahora en la última etapa de la vía identificada en la Figura 1, se muestra que la reacción de R-MP para formar monatina R,R es facilitada por una D-aminotransferasa de amplia especificidad, por ejemplo D-alanina aminotransferasa (E.C. 2.6.1.21, que también se conoce como D-aminoácido aminotransferasa o D-aspartato aminotransferasa) o una D-aminoácido deshidrogenasa. Como se describe anteriormente, la conversión de MP a monatina es una reacción de aminación, que crea un centro quiral en el carbono C-4 de monatina. Donde la forma R-quiral se desea en la posición C-4, deberían usarse enzimas que producen centros quirales "R" en aminoácidos.

De acuerdo con algunos casos, la D-aminotransferasa tiene mayor especificidad, mayor actividad o ambas para la R-MP que para indol-3-piruvato. De acuerdo con algunas realizaciones, la D-aminotransferasa tiene actividad limitada para el indol-3-piruvato. Las enzimas con dichas características pueden evolucionar o mutar de enzimas existentes, por ejemplo, como se muestra en el Ejemplo 16.

Los Ejemplos 9 a 12 ilustran la producción de monatina R,R a partir de D-triptófano.

La Figura 2 ilustra un método de producción de monatina R,R y monatina S,R. Mientras que en la realización de la Figura 1, la aldolasa usada en la reacción de indol-3-piruvato para formar R-MP afecta la relación entre R,R:S,R formada, en la realización de la Figura 2, la enzima D que facilita la conversión de MP a monatina determina la relación entre R,R:S,R formada. De acuerdo con la vía de la Figura 2, si una enzima no estereoespecífica se usa para facilitar la conversión de indol-3-piruvato a MP, entonces tanto S-MP y R-MP pueden formarse. Si una aldolasa no estereoselectiva se utiliza para convertir indol-3-piruvato a MP, es necesaria una transaminasa estereoselectiva para convertir el MP a monatina R,R o monatina S,R. Como se muestra en la Figura 2, el uso de una D-aminotransferasa o D-aminoácido deshidrogenasa que sea estereoespecífica para R-MP resulta en la producción de monatina R,R.

La Figura 3 ilustra otra vía alternativa para apuntar a una producción de monatina R,R. La vía de la Figura 3 es una modificación de la vía de la Figura 1, en donde indol-3-piruvato se produce indirectamente, en vez de directamente, a partir de L-triptófano. Más específicamente, el L-triptófano se convierte en D-triptófano, y el D-triptófano se convierte luego a indol-3-piruvato.

La conversión de L-triptófano a D-triptófano puede facilitarse por medio de una triptófano racemasa o equivalente funcional de la misma. El Ejemplo 15 proporciona fuentes potenciales de triptófano racemasas y métodos de detección para identificar dichas enzimas. También se contempla que una triptófano racemasa puede evolucionar

(tal como por medio de mutagénesis o manipulación recombinante) para un mejor rendimiento a partir de una aminoácido racemasa existente.

5 Ejemplos no taxativos de triptófano racemasas incluyen homólogos o mutantes de aminoácido racemasas (EC 5.1.1.-), por ejemplo serina racemasa, en donde los homólogos o mutantes son capaces de convertir L-triptófano a D-triptófano. Ejemplos no taxativos de fuentes de las cuales la aminoácido racemasa puede derivar incluyen: microorganismos tales como *Salmonella typhimurium*, *Escherichia coli*, *Bacillus subtilis*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Schizosaccharomyces pombe*, *Bacillus cereus*, *Enterococcus gallinarum*, *Pediococcus pentosaceus*, *Bacillus pumilus*, *Lactobacillus fermenti*, *Lactobacillus brevis*, *Aquifex pyrophilus*, *Lactobacilli*, *Streptococcus*, *Anabaena* sp., *Pseudomonas striata*, *Lentinus edodes*, *Scapharca brouhtonii* *Desulfurococcus* sp., *Thermococcus* sp. y *Pseudomonas striata*. Ejemplos no taxativos adicionales de fuentes de las cuales la aminoácido racemasa puede derivar incluyen gusano de seda, cerebro de rata o cerebro de ratón.

10 Ejemplos no taxativos de fuentes potenciales de las cuales las triptófano racemasas adecuadas pueden derivar incluyen: microorganismos tales como *Pseudomonas*, por ejemplo, *Pseudomonas chlororaphis* (*Pseudomonas aureofaciens*) (ATCC 15926) y *Burkholderia pyrrocina* (ATCC15958). Ejemplos no taxativos adicionales de fuentes potenciales de las cuales las triptófano racemasas adecuadas pueden derivar incluyen plantas, por ejemplo plantas de tabaco, tales como *Nicotiana tabacum*, plantas de trigo, tales como, *Triticum aestivum*, remolachas, tomates y *Sclerochiton ilicifolius*.

15 La vía que se muestra en la Figura 3 tiene ciertos beneficios, incluyendo que incluso cuando la monatina R,R es el producto deseado, puede usarse la misma enzima para la reacción que produce indol-3-piruvato que para la reacción que produce monatina. Es decir, en la vía ilustrada en la Figura 1, una L-aminotransferasa (o L-enzima adecuada) facilita la reacción que produce indol-3-piruvato, pero una D-aminotransferasa facilita la reacción que produce monatina. Por contraste, en la vía de la Figura 3, cierta D-aminotransferasa que facilita la reacción que produce indol-3-piruvato, también puede facilitar la reacción que produce monatina. Por consiguiente, en las vías de acuerdo con la Figura 3, pueden preferirse D-aminotransferasas de especificidad amplia cuando se desea usar la misma enzima para la reacción que forma indol-3-piruvato que para la reacción que forma monatina. Por contraste, en las vías de acuerdo con las Figuras 1, 2, 4, 6, 7 y 8, la producción de monatina puede proceder de manera más eficiente cuando se selecciona una D-aminotransferasa que tiene actividad limitada y/o especificidad para indol-3-piruvato en comparación con R-MP.

20 Otro beneficio de la vía representada esquemáticamente en la Figura 3 es que el producto aminoácido de la reacción acoplada a la reacción que produce indol-3-piruvato puede usarse ahora como un material de partida en la reacción acoplada a la reacción que produce monatina. Es decir, en la vía ilustrada en la Figura 1, el L-triptófano reacciona para producir indol-3-piruvato y al mismo tiempo oxaloacetato, alfa-cetoglutarato y/o piruvato reaccionan para producir un L-aminoácido. Dado que la reacción de R-MP para formar monatina se acopla con una reacción que utiliza un D-aminoácido como un sustrato, el L-aminoácido de la reacción que forma indol-3-piruvato no se recicla, en las condiciones que se muestran, para usar en la reacción acoplada a la reacción de R-MP. Por contraste, en la vía ilustrada en la Figura 3, la reacción de D-triptófano para formar indol-3-piruvato se acopla a una reacción que forma un producto D-aminoácido. Dicho D-aminoácido puede reciclarse para usar en la reacción acoplada a la reacción de R-MP. Esto permite el uso de cantidades no-estequiométricas de un aceptor amino en la etapa uno. En algunos casos de la invención, el D-aminoácido es D-alanina.

25 Las Figuras 4 y 5 ilustran modificaciones adicionales de la vía que se muestra en la Figura 1, cuyas modificaciones están dirigidas a reciclar el producto aminoácido formado por la reacción acoplada con la reacción de L-triptófano con el reactivo aminoácido de la reacción acoplada a la reacción de MP con monatina.

30 Con referencia a la Figura 4, el reciclaje se logra proporcionando una enzima que puede facilitar la conversión de un L-aminoácido a un D-aminoácido y viceversa. Más específicamente, donde como se muestra en la Figura 4,  $\alpha$ -KG reacciona para formar L-glutamato cuando L-triptófano reacciona para formar indol-3-piruvato, se puede proporcionar una glutamato racemasa (EC 5.1.1.3) o equivalente funcional que puede facilitar la conversión del L-glutamato a D-glutamato y viceversa. En dicho caso, el L-glutamato formado junto con la producción de indol-3-piruvato se elimina en por su conversión a D-glutamato, y el D-glutamato formado a partir de la conversión de L-glutamato está disponible entonces como un sustrato para la reacción acoplada con el MP a la reacción de monatina. De manera similar, el  $\alpha$ -KG formado en la reacción de D-glutamato está disponible como un sustrato para la reacción acoplada a la reacción de L-triptófano con indol-3-piruvato.

35 Ejemplos no taxativos de fuentes potenciales de las cuales la glutamato racemasa puede derivar incluyen *Pediococcus pentosaceus*, *Bacillus pumilus*, *Lactobacillus fermenti*, *Lactobacillus brevis*, *E. coli*, *Aquifex pyrophilus* y *Bacillus subtilis*. Más específicamente (también a modo no taxativo), la glutamato racemasa puede expresarse a partir de un ácido nucleico tal como el gen de *pediococcus pentaosaceus murl* (No. de Acceso de Genbank L22789), o glutamato racemasa de *Lactobacillus brevis*.

40 En el caso en el que el oxaloacetato reacciona para formar L-aspartato cuando el L-triptófano reacciona para formar indol-3-piruvato, puede proporcionarse una aspartato racemasa (EC 5.1.1.13) o equivalente funcional para convertir

L-aspartato en D-aspartato. En dicho caso, el L-aspartato junto con la producción de indol-3-piruvato se elimina en por su conversión a D-aspartato, y el D-aspartato formado a partir de la conversión de L-aspartato está disponible entonces como un sustrato para la reacción acoplada a la reacción de MP con monatina. De manera similar, el oxaloacetato formado en la reacción de D-aspartato está disponible para actuar como un sustrato para la reacción acoplada a la reacción de L-triptófano con indol-3-piruvato.

Ejemplos no taxativos de enzimas adecuadas que tienen actividad de aspartato racemasa incluyen ASPR-101 (BioCatalytics, Inc., Pasadena, CA) y homólogos o mutantes de una aminoácido racemasa (EC 5.1.1.-) que es capaz de facilitar la conversión de L-aspartato a D-aspartato.

Ejemplos no taxativos de fuentes potenciales de las cuales las racemasas de aspartato pueden derivar incluyen: *Desulfurococcus*, *Thermococcus*, molusco bivalvo *Scapharca brouhtonii*, *Acinetobacter*, *Agrobacterium*, *Archaeoglobus*, *Bacillus*, *Bordetella*, *Bradyrhizobium*, *Brevibacterium*, *Burkholderia*, *Campylobacter*, *Candida*, *Caulobacter*, *Clostridium*, *Desulfitobacterium*, *Desulfotalea*, *Enterococcus*, *Erwinia*, *Escherichia*, *Ferroplasma*, *Helicobacter*, *Klebsiella*, *Lactobacillus*, *Mannheimia*, *Medicago*, *Mesorhizobium*, *Methanococcus*, *Methanosarcina*, *Oceanobacillus*, *Oenococcus*, *Pediococcus*, *Polaribacter*, *Pseudomonas*, *Pyrococcus*, *Ralsonia*, *Shigella*, *Sinorhizobium*, *Salmonella*, *Sphingomonas*, *Streptococcus*, *Thermoanaerobacter*, *Vibrio*, *Wolinella*, *Xanthomonas*, *Xanthobacter*, *Yersinia* y *Zymomonas*.

En los casos en los que el piruvato reacciona para formar L-alanina cuando el L-triptófano reacciona para formar indol-3-piruvato, se puede proporcionar una alanina racemasa o equivalente funcional para convertir L-alanina a D-alanina. En dicho caso, la L-alanina formada junto con la producción de indol-3-piruvato se elimina por su conversión a D-alanina, y la D-alanina formada a partir de la conversión de L-alanina está disponible entonces para actuar como un sustrato para la reacción acoplada a la reacción de MP con monatina. De manera similar, el piruvato formado en la reacción de D-alanina está disponible para actuar como un sustrato para la reacción acoplada con el L-triptófano a la reacción de indol-3-piruvato.

Ejemplos no taxativos de alanina racemasas adecuadas incluyen A8936 (Sigma-Aldrich, St. Louis, MO).

Ejemplos no taxativos de fuentes potenciales de las cuales la alanina racemasa puede derivar incluyen: *Brucella abortus*, *Streptococcus faecalis*, *Salmonella typhimurium*, *Escherichia coli*, *Bacillus subtilis*, *Bacillus stearothermophilus*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Schizosaccharomyces pombe*, *Bacillus cereus* y *Lentinus edodes*.

Los Ejemplos 18 y 21 ilustran el uso de las racemasas anteriores, su impacto en el aumento de la relación del producto de monatina deseado, y proporcionan fuentes potenciales para las enzimas racemasa.

Con respecto a la Figura 5, se usa una aminotransferasa de estereoinversión para facilitar la reacción de R-MP para formar monatina. A pesar de que típicamente la reacción de R-MP (o S-MP) para formar monatina R,R (o monatina S,R) se acopla con la reacción de un D-aminoácido, una aminotransferasa de estereoinversión puede facilitar que las reacciones acopladas de R-MP (o S-MP) formen monatina R,R (o monatina S,R) usando un L-aminoácido. De esta forma, el producto de L-aminoácido de la reacción de L-triptófano aminotransferasa puede usarse como un sustrato para la transaminación de MP a monatina, y el producto (es decir, oxaloacetato, piruvato y/o  $\alpha$ -KG) de la reacción acoplada a la reacción de MP con monatina puede usarse como un material de partida para la reacción acoplada a la reacción de L-triptófano con indol-3-piruvato. Ejemplos no taxativos de aminotransferasas de estereoinversión que pueden usarse incluyen D-fenilglicina aminotransferasa (EC 2.6.1.72, también conocida como D-4-hidroxifenilglicina aminotransferasa) y D-metionina aminotransferasa (EC 2.6.1.41, también conocida como D-met-aminotransferasa y D-metionina-piruvato aminotransferasa). Ejemplos no taxativos de fuentes potenciales de las cuales la D-fenilglicina aminotransferasa puede derivar incluyen *Pseudomonas*, tales como *Pseudomonas putida* LW-4 y *Pseudomonas stutzeri* ST-201. Ejemplos no taxativos de fuentes potenciales de las cuales la D-metionina aminotransferasa puede derivar incluyen coliflor y maní.

Los ejemplos 19 y 20 juntos proporcionan fuentes potenciales de enzimas de estereoinversión, y métodos de realización de dichas enzimas. Los ejemplos también proporcionan métodos de detección para identificar dichas enzimas. También se contempla que dichas enzimas pueden evolucionar de enzimas de estereoinversión que se conocen o se encuentran en la naturaleza. Como un ejemplo no taxativo, la aminotransferasa de estereoinversión puede ser un homólogo o mutante de una D-aminoácido aminotransferasa o un homólogo o mutante de una aminoácido racemasa (EC 5.1.1.-).

Las Figuras 6-8 también ilustran modificaciones a la vía de la Figura 1. Las vías ilustradas en las Figuras 6-8 proporcionan métodos para impulsar reacciones de equilibrio eliminando el subproducto de la reacción de triptófano y en algunos casos proporcionando un sustrato para la reacción de MP.

Con referencia a la Figura 6, la vía que se muestra elimina el producto L-aminoácido de la reacción acoplada a la reacción de triptófano convirtiéndolo en un L-aminoácido diferente, y luego proporciona un sustrato para la reacción acoplada a la reacción de MP convirtiendo el L-aminoácido recién formado en un D-aminoácido. Específicamente, se

muestra que el L-triptófano reacciona junto con el oxaloacetato para formar indol-3-piruvato y L-aspartato. Un aspartato 4-d Descarboxilasa (EC 4.1.1.12) o equivalente funcional se usa para facilitar la conversión de L-aspartato a L-alanina y dióxido de carbono, y una enzima con actividad de alanina racemasa se usa para facilitar la conversión de L-alanina a D-alanina. Dicha D-alanina puede servir como un donante amino para la conversión de R-MP a monatina.

5 Con respecto a la Figura 7, la vía que se muestra ilustra métodos adicionales para eliminar el producto de L-aminoácido de la reacción acoplada a la reacción de triptófano. Las realizaciones como se presentan en la figura producen un subproducto que no está disponible para reaccionar en la dirección inversa, por ejemplo debido a volatilidad (tal como dióxido de carbono) o por conversión espontánea en un producto final sin reaccionar. Un ejemplo de dicho abordaje incluye los casos en los que  $\alpha$ -KG reacciona junto con L-triptófano para producir L-glutamato, puede proporcionarse una glutamato descarboxilasa (EC 4.1.1.15) o equivalente funcional que puede facilitar la conversión de de L-glutamato a 4-aminobutanoato (con dióxido de carbono como subproducto). Ejemplos no taxativos de fuentes potenciales de las cuales la L-glutamato descarboxilasa puede derivar incluyen: *Clostridium perfringens*, *C. welchii* o *E. coli*.

10 Otro ejemplo de dicho abordaje para continuar con la reacción de triptófano incluye donde oxaloacetato reacciona junto con L-triptófano, puede proporcionarse una aspartato descarboxilasa (EC 4.1.1.1) o equivalente funcional para facilitar la conversión de L-aspartato a  $\beta$ -alanina (con dióxido de carbono como un subproducto).

Con referencia a la Figura 8, la vía que se muestra ilustra otros métodos adicionales para eliminar el producto L-aminoácido de la reacción acoplada a la reacción de triptófano y que proporciona un sustrato para la reacción acoplada a la reacción de MP. Específicamente, en los casos en los que  $\alpha$ -KG reacciona junto con L-triptófano para formar L-glutamato, puede proporcionarse una enzima con actividad de L-alanina aminotransferasa y piruvato, en donde la enzima L-alanina aminotransferasa facilita la reacción de piruvato y L-glutamato para formar L-alanina. También se puede proporcionar una alanina racemasa o equivalente funcional para facilitar la conversión de la L-alanina a D-alanina. Dicha D-alanina puede usarse como un sustrato junto con MP para formar monatina y piruvato. Ver los Ejemplos 18 y 21.

#### Vías biosintéticas para producir R,R y otros estereoisómeros de derivados de monatina

Los métodos de la invención descrita que usan polipéptidos con actividad de aldolasa descritos en la presente pueden usarse para facilitar la reacción entre un indol-3-piruvato sustituido y una fuente de carbono C3.

30 Las enzimas útiles para facilitar una reacción entre un indol-3-piruvato sustituido y una fuente de carbono C3 incluyen un polipéptido con actividad de aldolasa de la SEQ ID NO:276 o fragmentos o subsecuencias del mismo que tienen actividad de aldolasa.

En una realización, uno o más polipéptidos con actividad de HMG aldolasa de la SEQ ID NO:276 o fragmentos o subsecuencias de los mismos que tienen actividad de aldolasa pueden ser útiles para facilitar una reacción entre un indol-3 piruvato y una fuente de carbono C3.

35 Alternativamente, uno o más polipéptidos con actividad de aldolasa codificada por una secuencia de ácidos nucleicos que tiene al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más, o una identidad de secuencia completa (100%) con un ácido nucleico de la SEQ ID NO:275 en la longitud completa de un ADNc, transcripto (ARNm) o gen, puede ser útil para facilitar una reacción entre un indol-3-piruvato y una fuente de carbono C3.

40 En una realización de la invención, un polipéptido con actividad de HMG aldolasa codificada por una secuencia de ácidos nucleicos que tiene al menos aproximadamente 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más, o una identidad de secuencia completa (100%) con un ácido nucleico de la SEQ ID NO:275 en todo el largo completo de un ADNc, transcripto (ARNm) o gen puede ser útil para facilitar una reacción entre un indol-3-piruvato y una fuente de carbono C3.

45 Un polipéptido con actividad de aldolasa codificada por una secuencia de ácidos nucleicos que se hibrida en condiciones rigurosas con un ácido nucleico de la SEQ ID NO:275 puede ser útil para facilitar una reacción entre un indol-3-piruvato sustituido y una fuente de carbono C3.

50 En una realización de la invención, un polipéptido con actividad de HMG aldolasa codificada por una secuencia de ácidos nucleicos que se hibrida en condiciones rigurosas con un ácido nucleico de la SEQ ID NO:275 puede ser útil para facilitar una reacción entre un indol-3-piruvato sustituido y una fuente de carbono C3.

En una realización, el grupo sustituyente del indol-3-piruvato sustituido es un átomo de halógeno unido a cualquier átomo de carbono del anillo indol. En otra realización, el grupo sustituyente es un átomo de cloro unido a cualquier carbono del anillo indol. En otra realización adicional, el derivado de monatina es ácido 4-hidroxi-4-(6-metilindol-3-ilmetil)glutámico.



Los polipéptidos que tienen actividad de aldolasa, y de acuerdo con algunas realizaciones de la invención, pueden usarse en una vía en la cual una o más etapas son una reacción de síntesis química. Por ejemplo, en algunas realizaciones, uno o más polipéptidos que tienen actividad de aldolasa pueden facilitar una reacción entre piruvato e indol-3-piruvato para proporcionar precursor de monatina. El precursor de monatina puede purificarse luego.

5 Entonces puede utilizarse una reacción de aminación reductiva del precursor de monatina para proporcionar monatina.

Los polipéptidos que tienen actividad de aldolasa, y de acuerdo con algunas realizaciones de la invención así como otras enzimas usadas en el proceso para producir monatina y derivados de monatina pueden usarse en forma pura, bruta, aislada o de suspensión de sulfato de amonio.

10 Los polipéptidos que tienen actividad de aldolasa y de acuerdo con algunas realizaciones de la invención pueden optimizarse usando los agentes de estabilización, incluyendo ditioneitol ("DTT") y  $\beta$ -mercaptoetanol.

La monatina o derivado de monatina que se produce que utiliza uno o más de los polipéptidos divulgados en la presente, es generalmente al menos aproximadamente 50 a aproximadamente 99% monatina R,R o derivado de monatina R,R, en peso de la monatina o derivado de monatina total producida. En otras realizaciones, la monatina o derivado de monatina producido que utiliza uno o más de los polipéptidos divulgados en la presente, es mayor que 60% monatina R,R o derivado de monatina R,R, en peso de la monatina total producida; por ejemplo, la monatina R,R o derivado de monatina R,R es 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99% de la monatina o derivado de monatina total producida. Alternativamente, pueden combinarse varias cantidades de dos o más preparaciones de monatina o derivados de monatina de manera que resulte en una preparación que sea un porcentaje deseado de monatina R,R o derivado de monatina R,R. Por ejemplo, una preparación de monatina que es 60% monatina R,R puede combinarse con una preparación de monatina que es 90% monatina R,R; si se combinan cantidades iguales de 60% y 90% de monatina R,R, la preparación de monatina resultante sería 75% monatina R,R.

La monatina o derivado de monatina, o un intermediario (incluido un precursor de monatina), producida utilizando uno o más polipéptidos divulgados en la presente, puede purificarse a partir de los componentes de la reacción. En una realización, la monatina, derivado de monatina o intermediario, tal como precursor de monatina, puede purificarse simplemente eliminando la sustancia que debe purificarse de la preparación de enzimas en la cual se sintetiza.

En otras realizaciones, el intermediario, precursor de monatina, monatina o derivado de monatina se purifica a partir de la preparación en la cual se sintetizó de forma tal que la composición o preparación "purificada" resultante sea al menos aproximadamente 5-60% monatina en peso del total de los compuestos orgánicos. En otra realización, la monatina, derivado de monatina o intermediario, tal como un precursor de monatina, puede purificarse hasta un grado de pureza de al menos aproximadamente 70%, 80%, 90%, 95% o 99% en peso de los compuestos orgánicos totales. La monatina, derivado de monatina o el intermediario (incluido el precursor de monatina), producida utilizando uno o más de los polipéptidos divulgados en la presente, puede purificarse a partir de los componentes de la reacción por medio de cualquier método conocido por un experto en la técnica. Óptimamente, la monatina purificada o intermediario puede recristalizarse repetidamente hasta que se logre el grado deseado de pureza.

Los siguientes ejemplos se ofrecen para ilustrar, a modo no taxativo, la invención reivindicada.

## EJEMPLOS

### 40 Ejemplo 1

Detección de monatina, precursor de monatina, triptófano, alanina, aspartato y glutamato

El presente ejemplo describe los procedimientos utilizados para detectar la presencia de monatina, precursor de monatina (PM), triptófano, aspartato, alanina y glutamato. Asimismo, el presente ejemplo describe un procedimiento para separar y detectar los cuatro estereoisómeros de monatina.

### 45 Análisis del monitoreo de reacción múltiple (MRM) LC/MS/MS de monatina y triptófano

Se realizaron análisis de mezclas para monatina y triptófano derivados de reacciones bioquímicas *in vitro* o *in vivo* mediante el uso de un instrumento de espectrometría de masas en tándem con cromatografía líquida (LC/MS/MS) Micromass de Waters que incluye un cromatógrafo para líquidos 2795 de Waters con un monitor de absorbancia 996 de fotodiodo (PDA) de Waters colocado en serie entre el cromatógrafo y un espectrómetro de masa de triple cuadrúpulo Quattro Ultima de Micromass. Se realizaron separaciones por LC utilizando una columna de cromatografía de fase inversa Xterra MS C8 de 2,1 mm x 250 mm a 40°C. La fase móvil de LC consistió en A) agua que contenía (i) 0,05% (v/v) de ácido trifluoroacético o (ii) 0,3% de ácido fórmico y 10 mM de formiato de amonio y B) metanol con (i) 0,05% (v/v) de ácido trifluoroacético o (ii) 0,3% de ácido fórmico y 10 mM de formiato de amonio.

Si la fase móvil de LC consistió en A) agua con un 0,05% (v/v) de ácido trifluoroacético y B) metanol con 0,05% (v/v) de ácido trifluoroacético, la elución en gradiente fue lineal desde el 5% B hasta el 35%B, durante 0-4 minutos, lineal desde el 35% B hasta el 60% B, durante 4-6,5 minutos, lineal desde el 60% B hasta el 90% B, durante 6,5-7 minutos, isocrática a 90% B, durante 7-11 minutos, lineal desde el 90% B hasta el 95% B, durante 11-12 minutos, lineal desde el 95% B hasta el 5% B, durante 12-13 minutos, con un período de reequilibrado de 2 minutos entre las series. La tasa de flujo fue de 0,25 mL/min y se monitoreó una absorbancia de PDA de 200 nm a 400 nm. Todos los parámetros de ESI-MS se optimizaron y seleccionaron en base a la generación de iones moleculares protonados ([M+H]<sup>+</sup>) de los analitos de interés, y la producción de iones de fragmentos característicos. Los siguientes parámetros instrumentales se utilizaron para el análisis del monitoreo de reacción múltiple (MRM) 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 la Fuente: 100°C; temperatura de desolvatación; 350°C; Gas de desolvatación: 500 L/h; Cono de gas: 50 L/h; Baja resolución de masas (Q1): 12,0; Alta resolución de masas (Q1): 12,0; Energía iónica: 0,2; Entrada: -5 V; Energía de colisión: 8; Salida: IV; Baja resolución de masas (Q2): 15; Alta resolución de masas (Q2): 15; Energía iónica (Q2): 3,5; Multiplicador: 650. Se utilizan cinco transiciones de MRM padre/hija monatina específicas para detectar específicamente las reacciones de monatina in vitro e in vivo. Las transiciones monitoreadas son 293,1 a 158,3, 293,1 a 168,2, 293,1 a 211,2, 293,1 a 230,2 y 293,1 a 257,2. El triptófano se monitorea con la transición MRM 204,7 a 146,4. Para la cuantificación estándar interna de monatina y triptófano, se analizaron cuatro patrones de calibración con cuatro proporciones diferentes de cada analito respecto de d5-triptófano y d5-monatina. Estos datos están sujetos a un análisis lineal de mínimos cuadrados para formar una curva de calibración para la monatina y el triptófano. A cada muestra se le agrega una cantidad fija de d5-triptófano y d5-monatina (la d5-monatina se sintetizó de d5-triptófano conforme a los procedimientos de la patente WO03/091396A2) y las proporciones de respuesta (monatina/d5-monatina; triptófano/d5-triptófano) utilizadas junto con las curvas de calibración descritas anteriormente para calcular la cantidad de cada analito en las mezclas.

Si la fase de LC fue A) agua con un 0,3% de ácido fórmico y 10 mM de formiato de amonio y B) metanol con un 0,3% de ácido fórmico y 10 mM de formiato de amonio, la elución en gradiente fue lineal de un 5% B a un 45% B, 0-8,5 minutos, lineal desde un 45% B a un 90% B, 8,5-0 minutos, isocrática desde un 90% B hasta un 90% B, 9-12,5 minutos, lineal desde un 95% B hasta un 5% B, 12,5-13 minutos, con un período de reequilibrado de 4 minutos entre las series. La tasa de flujo fue de 0,27mL/min, y la absorbancia de PDA se monitoreó desde 210 nm hasta 400 nm. Todos los parámetros de ESI-MS se optimizaron y seleccionaron en base a la generación de iones moleculares protonados ([M+H]<sup>+</sup>) de los analitos de interés y la producción de iones de fragmentos característicos. Los parámetros instrumentales utilizados para esta fase móvil secundaria son los mismos que los parámetros anteriores. Se utilizan cuatro transiciones MRM padre-hija monatina-específicas y una transición padre/hija triptófano específica para detectar específicamente las reacciones de monatina y triptófano in vivo e in vitro. Las transiciones monitoreadas son 293,1 a 158,0, 293,1 a 168,0, 293,1 a 211,5 y 293,1 a 257,0. El triptófano se monitorea con la transición MRM 205,2 a 146,1. Para la cuantificación estándar interna de monatina y triptófano, se analizaron cuatro patrones de calibración que contienen cuatro proporciones diferentes de cada analito respecto de d5-triptófano y d5-monatina. Estos datos están sujetos a un análisis lineal de mínimos cuadrados para formar una curva de calibración para monatina y triptófano. A cada muestra se le agrega una cantidad fija de d5-triptófano y d5-monatina (la d5-monatina se sintetizó de d5-triptófano conforme a los procedimientos de la patente WO03/091396 A2), y se utilizan las proporciones de respuesta (monatina/d5-monatina; triptófano/d5-triptófano) junto con las curvas de calibración descritas anteriormente para calcular la cantidad de cada analito en las mezclas. Las transiciones de masa padre-hija monitoreadas para d5-triptófano y d5-monatina son 210,2 a 151,1, y 298,1 a 172,0, respectivamente.

#### Medición exacta de masa de monatina

Se realizó un análisis MS de alta resolución mediante el uso de un espectrómetro de masas de cuadrupolo híbrido/tiempo de vuelo Q-Star de Applied Biosystems-Perkin Elmer. La masa medida para la monatina protonada utilizó triptófano como patrón interno de calibración de masa. La masa calculada de monatina protonada, en base a la composición básica C<sub>14</sub>H<sub>17</sub>N<sub>2</sub>O<sub>5</sub> es 293,1137. La monatina producida utilizando el proceso biocatalítico descrito en los Ejemplos 2 y 3 mostró una masa medida de 293,1144. Este es un error de medición de la masa menor a 2 partes por millón (ppm), que proporciona prueba concluyente de la composición básica de la monatina producida por vía enzimática.

#### Medición quiral LC/MS/MS (MRM) de monatina.

La determinación de la distribución de estereoisómeros de monatina en reacciones in vitro e in vivo se logró por derivatización con 1-fluoro-2-4-dinitrofenil-5-L-alanina amida (FDAA), seguida de la medición MRM LC/MS/MS de fase inversa.

#### 55 Derivatización de monatina con FDAA

En una muestra o patrón de 50 µL y en un patrón interno de 10 µL se agregaron 100 µL o 200 µL de un 1% de solución de FDAA en acetona. Se agregaron veinte o cuarenta µL de bicarbonato de sodio 1,0 M, respectivamente y la mezcla se incubó por 1 hora a 40°C mezclándose ocasionalmente. La muestra se removió y enfrió y neutralizó

con 20 µL de HCl 2,0 M (se puede necesitar más HCl para realizar la neutralización de una mezcla biológica tamponada). Después de completar la desgasificación, las muestras estaban listas para analizar por LC/MS/MS.

Monitoreo de reacción múltiple LC/MS/MS para la determinación de la distribución de estereoisómeros de monatina en reacciones in vitro e in vivo

- 5 Se realizaron análisis utilizando la instrumentación de LC/MS/MS anteriormente descrita. Las separaciones por LC capaces de separar los cuatro estereoisómeros de monatina (específicamente FDAA-monatina) se realizaron en una columna de cromatografía de fase inversa de 2,0 x 250 mm (3 µm) de Phenomenex Luna C18 (2) a 40°C. La fase móvil de LC consistió en A) agua con un 0,05% (masa/volumen) de acetato de amonio y B) acetonitrilo. La elución fue isocrática al 13% B, 0-2 minutos, lineal desde el 13% B hasta el 30% B, 2-15 minutos, lineal desde el 30% B hasta el 80% B, 15-16 minutos, isocrática al 80% B, 16-21 minutos, y lineal desde el 80% B hasta el 13% B, 21-22 minutos con un período de reequilibrado de 8 minutos entre las series. La tasa de flujo fue de 0,23 mL/min, y la absorbancia de PDA se monitoreó desde 200 nm hasta 400 nm. Todos los parámetros de la ESI-MS se optimizaron y seleccionaron en base a una generación de iones moleculares desprotonados ([M-H]<sup>-</sup>) de FDAA-monatina, y la producción de iones de fragmento característico.
- 10
- 15 Los siguientes parámetros instrumentales se utilizaron para el análisis de LC/MS de la monatina en el modo ESI/MS de ión negativo: Capilar: 2,0 kV; Cono: 25 V; Hex 1: 10 V; Apertura: 0 V; Hex 2: 0 V; Temperatura de la fuente: 100°C; temperatura de desolvatación: 350°C; Gas de desolvatación: 500 L/h; Cono de gas: 50 L/h; Baja resolución de masas (Q1): 12,0; Alta resolución de masas (Q1): 12,0; Energía iónica: 0,2; Entrada: -5V; Energía de colisión: 20; Salida: 1V; Baja resolución de masas (Q2): 12; Alta resolución de masas (Q2): 12; Energía iónica (Q2): 3,0; Multiplicador: 650. Se utilizan tres transiciones padre-hija monatina-FDAA específicas para detectar específicamente las reacciones de FDAA-monatina in vitro e in vivo. Las transiciones monitoreadas para monatina son 543,2 a 268,1, 543,2 a 499,3, y 543,2 a 525,3. La transición de masa derivada del patrón interno de monatina fue 548,2 a 530,3. La identificación de los estereoisómeros de FDAA-monatina se basa en el tiempo de retención cromatográfica comparada con los estereoisómeros de monatina sintéticos purificados y los datos espectrales de la masa. Se utiliza un estándar interno para monitorear el progreso de la reacción y para confirmar el tiempo de retención del estereoisómero S,S.
- 20
- 25

Detección por fluorescencia post columna con cromatografía líquida de aminoácidos que incluyen glutamato y alanina.

- 30 La cromatografía líquida con la detección por fluorescencia post columna (LC/OPA) para la determinación de glutamato y alanina en reacciones in vivo e in vitro se realizó en un sistema de LC 2690 de Waters o un equivalente combinado con un detector de fluorescencia mediante escaneo 474 y un módulo de reacción post columna de Waters. También se realizaron análisis semi cuantitativos de monatina y triptófano utilizando este procedimiento. Las separaciones de LC se realizaron en una columna de intercambio de iones cargados con interacción de sodio a 60°C. La fase móvil A consistió en un tampón Na 328 de Pickering (Pickering Laboratories, Inc; Mountain View, CA)
- 35 La fase móvil B consistió en un tampón Na 740 de Pickering. La elución en gradiente fue desde un 0% hasta 100% B, 0-20 minutos, isocrática al 100% B, 20-36 minutos y lineal desde un 100% B hasta un 0% B, 36-37 minutos, con al menos un período de reequilibrado de 5 minutos entre segmentos, dependiendo de la matriz de la muestra. La tasa de flujo para la fase móvil fue de 0,5 mL/min. La tasa de flujo para la solución de derivatización post columna de OPA fue de 0,5 mL/min. Las configuraciones del detector de fluorescencia fueron EX 338-340 nm y Em 420-425 nm.
- 40 Se empleó norleucina como un patrón interno para el análisis. La identificación de aminoácidos se basó en los datos del tiempo de retención cromatográfica para patrones purificados.

Detección de L- y D-aminoácidos por LC/MS/MS

- 45 Las muestras que contienen una mezcla de L- y D-aminoácidos como lisina, alanina, metionina, tirosina, leucina, fenilalanina, triptófano, glutamato, y aspartato obtenidas de experimentos de reacción bioquímica se trataron por primera vez con ácido fórmico para desnaturalizar la proteína. Posteriormente la muestra se centrifugó y filtró a través de un filtro de jeringa de nylon de 0,45 µm antes del análisis de LC/MS/MS. La identificación de L- y D-aminoácidos se basó en el tiempo de retención y la detección selectiva de masa. La separación por LC se logró mediante el uso de un sistema de cromatografía líquida 2690 de Waters y una columna de cromatografía de 2,1 mm x 250mm Chirobiotic TAG de ASTEC con una temperatura de columna de 45°C. Las fases móviles A y B de la LC consistieron de un 0,25% de ácido acético y un 0,25% de ácido acético en metanol, respectivamente. Se utilizó una elución isocrática para todos los procedimientos para separar los isómeros L y D. La lisina se eluyó utilizando un 80% de fase móvil A y un 20% de fase móvil B. El glutamato, la alanina y la metionina se separaron con la elución del 60% de la fase móvil A y el 40% de la fase móvil B y una tasa de flujo de 0,25 mL/min. El aspartato, el triptófano, la tirosina, la leucina, y la fenilalanina se separaron isoméricamente con 30% de la fase móvil A y con un 70% de la fase móvil B con una tasa de flujo de 0,3 mL/min para todos menos la fenilalanina, que se utilizó a una tasa de flujo de 0,25 mL/min.
- 50
- 55

El sistema de detección para el análisis de los L- y D-aminoácidos incluyó un detector de fotodiodo (PDA) 996 de Waters y un espectrómetro de masa de triple cuadrúpolo Quattro Ultima de Micromass. El PDA, que escanea de 195

a 350 nm, se colocó en serie entre el sistema de cromatografía y el espectrómetro de masa. Los parámetros para el espectrómetro de masa de triple cuadrúpulo Quattro Ultima de Micromass que funciona en modo de ionización positiva por electroasperción (+ESI) fueron los siguientes: Capilar: 3,0 kV; Cono: 20 V; Hex 1: 15 V; Apertura: 1 V; Hex 2: 0 V; Temperatura de fuente: 100°C; temperatura de desolvatación: 350°C; gas de desolvatación: 530 L/h; cono de gas: 30 L/h; Baja resolución de masas Q1: 12,5; Alta resolución de masas Q1: 12,5; Energía iónica 1: 0,2; Entrada: -5; Colisión: 8; Salida 1: 10; Baja resolución de masas Q2: 12,5; Alta resolución de masas Q2: 12,5; Energía iónica 2: 0,5; Multiplicador: 650 V. Los experimentos de MS/MS con un modo de Monitoreo de reacción múltiple (MRM) se establecieron para monitorear selectivamente las transiciones de reacción de 147,8 a 84,2 y de 147,8 a 102,1 de glutamato, de 134,00 a 74,30 y de 134,00 a 88,2 para aspartato, de 147,3 a 85,0 para lisina, de 150,3 a 104,8 para metionina, de 182,3 a 137,0 para tirosina, de 132,3 a 87,0 para leucina y de 166,3 a 121,0 para fenilalanina. Cuando se enumeraron dos transiciones, las últimas se utilizaron para su cuantificación. Para el triptófano, se determinaron los experimentos de MS/MS con el modo de Monitoreo de reacción múltiple (MRM) para monitorear selectivamente las transiciones de reacción de 205,2 a 118,2, de 205,2 a 146,1 y de 205,2 a 188,2 y la transición desde 212,1 a 151,1 para el d8-DL-triptófano. La cuantificación de triptófano se logró determinando la proporción de la respuesta de analito de la transición de 205,2 a 146,1 en el patrón interno, d8-D, L-triptófano. En forma alternativa, la cuantificación de triptófano, glutamato y ácidos aspárticos se basaron en las respuestas de señal de  $m/z=146,5$ ,  $m/z=102,1$  y  $m/z=88,2$ , respectivamente.

Producción de monatina y precursor de monatina (MP) para patrones y ensayos

Producción de monatina

Una mezcla racémica de monatina R,R y monatina S,S se produjo en forma sintética como se describe en la Patente de los Estados Unidos No. 5.128.482.

La monatina R,R y la monatina S,S se separaron mediante derivatización e hidrólisis. En síntesis, la mezcla racémica de monatina se esterificó, el grupo libre de amino se bloqueó con Cbz, se formó una lactona, y la lactona S,S se hidrolizó selectivamente utilizando una enzima de proteasa inmovilizada. La monatina también se puede separar como se describe en Bassoli, A. et al., Eur. J. Org. Chem., 8:1652-1658 (2005).

Producción de MP

Se produjo R-MP mediante la transaminación de monatina R,R utilizando una D-aminotransferasa de amplia gama AT-103 (BioCatalytics, Pasadena, CA) en un tampón de fosfato de potasio 0,1 M, mediante el uso de piruvato de sodio como el aceptor de amino. Se produjo S-MP mediante la transaminación de monatina S,S utilizando una L-aminotransferasa AT-102 (BioCatalytics, Pasadena, CA) en un tampón de fosfato de potasio 0,1 M mediante el uso de piruvato de sodio como el aceptor de amino. Ambas reacciones se realizaron a 30°C y con un pH de aproximadamente 8,0-8,3 por aproximadamente 20 horas. Ambos compuestos se purificaron utilizando una escala preparativa HPLC con una resina hidrófoba (XADTM1600) de Rohm & Haas (Filadelfia, PA) que eluye en agua. Se obtuvieron las muestras que contienen más del 90% de precursor de monatina puro y se congelaron a seco.

## 35 Ejemplo 2

Detección del precursor de monatina

El presente ejemplo describe los procedimientos utilizados para separar y detectar los dos enantiómeros del precursor de monatina.

Método no quiral para la detección del precursor de monatina

Las muestras de reacción en placas de 96 pocillos se inyectaron en una columna de 3,0 x 150mm, 3,5  $\mu\text{m}$  de Agilent Zorbax RX-C18 utilizando un muestreador automático CTCPal (LEAP Technologies, Carrboro, N.C.). Los productos se separaron utilizando un gradiente de H<sub>2</sub>O/ACN (0,1% de ácido fórmico):

Tiempo: 0,00 min 5% B

Tiempo: 4,00 min 100% B

Tiempo: 5,00 min 100% B

Tiempo: 5,10 min 5% B

Tiempo: 6,50 min 5% B

Las bombas LC-10ADvp (Shimadzu, Kioto, Japón) proporcionaron el gradiente a 0,8 mL/min. Los productos se detectaron utilizando un espectrómetro de masas de triple cuadrúpulo API4000 de Turbolon (Applied Biosystems,

Foster City, CA). La vaporización de iones y el monitoreo múltiple de iones se realizó para los analitos de interés en el modo de ión negativo y cada análisis duró 6,5 minutos.

Piruvato = 87,1[M-H+] –

Indol-3-piruvato = 202,1[M-H+] –

5 Producto = 290,0 [M-H+] –

Análisis quiral por CE de precursores de monatina R y S

Se utilizó un instrumento de electroforesis capilar MDQ de P/ACE™ (Beckman Coulter, Fullerton, CA). Se utilizó el Kit de Desarrollo Quiral, que incluye pequeñas cantidades de varios selectores quirales, tampones necesarios y 2 capilares (Beckman, Fullerton, CA). En forma alternativa, para el ensayo de MP únicamente, se pueden obtener en forma separada los siguientes reactivos y otros suministros de Beckman Coulter (Fullerton, CA) o en otra parte:

10

Capilar revestido N-CHO; 50 µm de ID, 65 cm de longitud total o un capilar de sílice fundida.

25 mM de tampón de fosfato, pH 5.

25 mg de hidroxipropil-β-ciclodextrina

Una solución de acondicionamiento capilar de 10 ml (en forma alternativa se puede utilizar un 0,5% de solución de óxido de polietileno, M<sub>w</sub>, 600.000 o 300.000 daltons)

15

Análisis por electroforesis capilar (CE)

Se utilizó un capilar neutral revestido, 50 µm ID, de 60 cm (50 cm para la detección) o 30 (20) cm junto con una detección de DAD (o UV simple) a 214 nm. El capilar de separación se calentó a 15°C, y las muestras a 4°C. El tampón de separación consistió en 20 mM de hidroxipropil-β-ciclodextrina, 25 mM de fosfato y un pH de 5. La inyección de la muestra generalmente fue de 0,5 psi, 5 s. La separación se realizó a 500V/cm, polaridad inversa (15 kV para un capilar de 30 cm, 30 kV para 60 cm). La corriente típica utilizada durante la separación fue de -28 µA. Los tiempos típicos de migración para los picos de MP fueron 3,5 minutos (20 cm de longitud efectiva) u 8 minutos (50 cm).

20

Un paso opcional de limpieza/lavado/acondicionamiento capilar previo a las series de muestra utilizó H<sub>2</sub>O durante 4 minutos, HCl 0,1 M durante 1 minuto, H<sub>2</sub>O durante 1,5 minutos, solución de acondicionamiento capilar durante 4 minutos, H<sub>2</sub>O durante 1 minuto y tampón de separación durante 4 minutos.

25

Un resumen del procedimiento de serie fue el siguiente: enjuague del tampón de separación por 1-2 minutos, inyección de muestra 5 s a 0,5 psi, separación durante 5-10 minutos a una polaridad de voltaje inversa de 15 o 30 kV dependiendo de la longitud capilar.

### 30 Ejemplo 3

Ensayo general de aldolasas de piruvato

Un procedimiento ejemplar para medir la actividad de diferentes aldolasas de piruvato utiliza un sustrato general, 4-carboxi-4-hidroxi-2-oxoadipato (CHA). El ensayo de CHA se adaptó de los ensayos de la literatura (por ejemplo, véase E.E. Dekker & R.P. Kitson, *J. Biol. Chem.* 267, 10507-10514, 1992). Un ensayo típico incluyó 50 mM de fosfato de sodio con un pH de 7,5, 1 mM de MgCl<sub>2</sub>, 1 mM de CHA, 10 µg/ml de D-lactato deshidrogenasa (LDH) de *Lactobacillus leichmanii* (Sigma-Aldrich, St. Louis, MO), 0,5 mM de NADH. El ensayo comenzó mediante la incorporación de enzima (generalmente entre 1 y 5 µL). La liberación de piruvato, junto con la formación de NAD<sup>+</sup> se monitoreó permanentemente en un espectrómetro a 340 nm.

35

El CHA se sintetizó de conformidad con el procedimiento descrito en Tack, B.F. Chapman, P.J., y S. Dagley. *J. Biol. Chem.* 247 6438-6443 (1972).

40

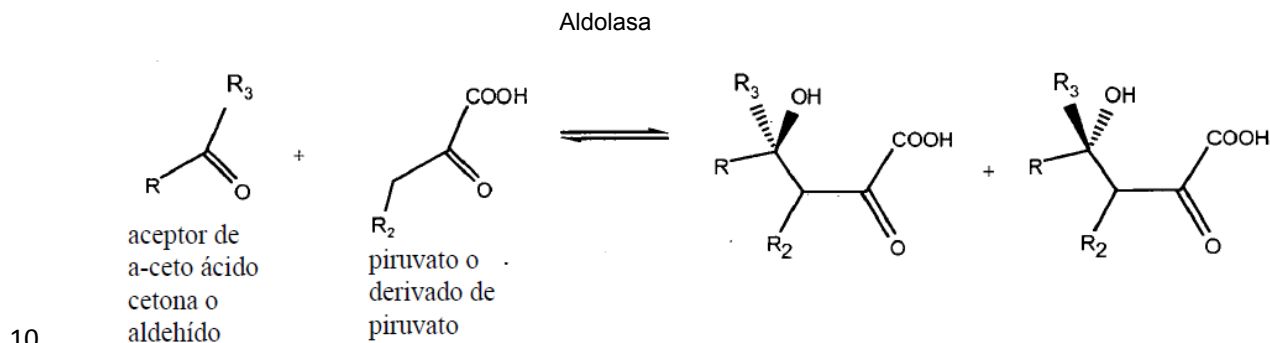
Una unidad de actividad enzimática, como aldolasa de piruvato, como enzima de aldolasa de HMG y/o KHG, se define como la cantidad que libera piruvato suficiente para reducir la absorbancia a 340 nm por una OD de 1 por minuto.

### Ejemplo 4

45 Descubrimiento de nuevas ceto-hidroxi-glutarato (KHG) e hidroxi-metil-ceto-glutarato (HMG) aldolasas de las bibliotecas ambientales de Diversa

Más de 150 aldolasas únicas de HMG y 15 aldolasas de KHG se descubrieron controlando las bibliotecas de ADN de Diversa. Estos genes de aldolasa se secuenciaron y subclonaron en un vector de expresión adecuado. Este vector se transformó posteriormente en un huésped de expresión adecuado para la producción de cantidades suficientes de la aldolasa para la caracterización de la enzima. Un conjunto seleccionado de aldolasas se testearon para la actividad en CHA y también para la formación de un precursor de monatina (MP). Todas las enzimas descubiertas y descritas en esta patente tienen potencial para su uso en otras reacciones que forman enlaces carbono-carbono entre el aceptor de alfa-cetoácido y piruvato o el donante derivado de piruvato como se ejemplifica en el esquema de reacción general que aparece a continuación.

5



R = H, alquilo, alquilo sustituido, arilo, arilo sustituido, bencilo, bencilo sustituido

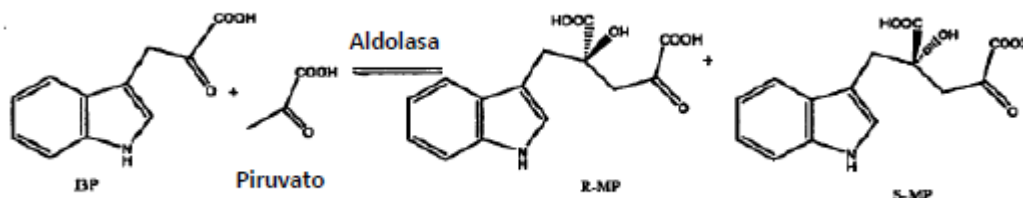
R<sub>2</sub> = H, alquilo, alquilo sustituido, arilo, arilo sustituido, bencilo, bencilo sustituido

R<sub>3</sub> = H, alquilo, alquilo sustituido, arilo, arilo sustituido, bencilo, bencilo sustituido, ácido carboxílico

### Ejemplo de referencia 5

15 Caracterización de las aldolasas seleccionadas

Las aldolasas seleccionadas se caracterizaron respecto de su capacidad para catalizar la conversión de indol-3-piruvato y piruvato en el precursor de monatina (MP) como se muestra en el siguiente esquema:



20 Las figuras 13 y 14 muestran las actividades de 58 aldolasas diferentes en la formación del MP medido por LC/MS/MS.

25 Las reacciones de aldol se realizaron con 20 mM de indol-3-piruvato ("I3P"), 50 mM de piruvato, 100 mM de fosfato de sodio con un pH de 7, 1 mM de MgCl<sub>2</sub>, 100 µg/mL de aldolasa. Las reacciones se incubaron a temperatura ambiente en la oscuridad. Las alícuotas (30 µL) se removieron varias veces y las reacciones se detuvieron mediante el almacenamiento de las muestras en hielo. Una porción de cada alícuota se sometió a un análisis de CE mientras que la porción restante se diluyó 1:1000 en un 50% de acetonitrilo y se sometió al análisis de LC/MS.

**Tabla 2: Enantioselectividad de diferentes aldolasas para la formación de MP a partir de I3P y piruvato determinado por CE quiral**

Aldolasa	% de R-MP (plazo de 23 horas)	CHA (U/mg) (en base a la proteína total en lisado)	Expresión relativa
SEQ ID NO: 28 (codificada por la SEQ ID NO: 27)	98+	31,8	###

Aldolasa	% de R-MP (plazo de 23 horas)	CHA (U/mg) (en base a la proteína total en lisado)	Expresión relativa
SEQ ID NO: 116 (codificada por la SEQ ID NO: 115)	95	304	##
SEQ ID NO: 76 (codificada por la SEQ ID NO: 75)	50	424	###
SEQ ID NO: 298 (codificada por la SEQ ID NO: 297)	98+	388	##
SEQ ID NO: 44 (codificada por la SEQ ID NO: 43)	70	332	##
SEQ ID NO: 54 (codificada por la SEQ ID NO: 53)	98+	95	#
SEQ ID NO: 148 (codificada por la SEQ ID NO: 147)	90	200	##
SEQ ID NO: 46 (codificada por la SEQ ID NO: 45)	70	174	##
SEQ ID NO: 134 (codificada por la SEQ ID NO: 133)	90	576	####
SEQ ID NO: 142 (codificada por la SEQ ID NO: 141)	98+	55	#
SEQ ID NO: 122 (codificada por la SEQ ID NO: 121)	98+	38	##
SEQ ID NO: 74 (codificada por la SEQ ID NO: 73)	80	484	####
SEQ ID NO: 64 (codificada por la SEQ ID NO: 63)	95	38	#
SEQ ID NO: 108 (codificada por la SEQ ID NO: 107)	98+	40	#
SEQ ID NO: 796(codificada por la SEQ ID NO: 95)	98+	No detectado	#
SEQ ID NO: 126 (codificada por la SEQ ID NO: 125)	95	124	##
SEQ ID NO: 80 (codificada por la SEQ ID NO: 79)	98+	No detectado	#
SEQ ID NO: 36 (codificada por la SEQ ID NO: 35)	98+	80	##
SEQ ID NO: 62 (codificada por la SEQ ID NO: 61)	98+	No detectado	#
SEQ ID NO: 112(codificada por la SEQ ID NO: 111)	98+	No detectado	#
SEQ ID NO: 130 (codificada por la SEQ ID NO: 129)	98+	38	##

Aldolasa	% de R-MP (plazo de 23 horas)	CHA (U/mg) (en base a la proteína total en lisado)	Expresión relativa
SEQ ID NO: 94 (codificada por la SEQ ID NO: 93)	98+	47	##
SEQ ID NO: 102 (codificada por la SEQ ID NO: 101)	No detectado	No detectado	#
SEQ ID NO: 58 (codificada por la SEQ ID NO: 57)	98+	59	##
SEQ ID NO: 88 (codificada por la SEQ ID NO: 87)	50	510	###
SEQ ID NO: 50 (codificada por la SEQ ID NO: 49)	98+	144	##
SEQ ID NO: 106 (codificada por la SEQ ID NO: 105)	98+	No detectado	##
SEQ ID NO: 40 (codificada por la SEQ ID NO: 39)	40	406	###
SEQ ID NO: 42 (codificada por la SEQ ID NO: 41)	98+	92	##
SEQ ID NO: 278(codificada por la SEQ ID NO: 277)	95	2,0	#
SEQ ID NO: 162 (codificada por la SEQ ID NO: 161)	95	11,8	#
SEQ ID NO: 276 (codificada por la SEQ ID NO: 275)	98+	100,4	#####
SEQ ID NO: 178 (codificada por la SEQ ID NO: 177)	95	38,8	#
SEQ ID NO: 202 (codificada por la SEQ ID NO: 201)	No detectado	No detectado	#
SEQ ID NO: 166 (codificada por la SEQ ID NO: 165)	98+	85,5	##
SEQ ID NO: 218 (codificada por la SEQ ID NO: 217)	95	49,1	##
SEQ ID NO: 224 (codificada por la SEQ ID NO: 223)	98+	23,2	#
SEQ ID NO: 226 (codificada por la SEQ ID NO: 225)	98+	128,3	#
SEQ ID NO: 244 (codificada por la SEQ ID NO: 243)	98+	40,4	#
SEQ ID NO: 250 (codificada por la SEQ ID NO: 249)	95	6,0	#
SEQ ID NO: 252 (codificada por la SEQ ID NO: 251)	95	20,2	##



Aldolasa	% de R-MP (plazo de 23 horas)	CHA (U/mg) (en base a la proteína total en lisado)	Expresión relativa
SEQ ID NO: 264 (codificada por la SEQ ID NO: 263)	95	9,9	##
SEQ ID NO: 268 (codificada por la SEQ ID NO: 267)	95	2,0	#
SEQ ID NO: 272 (codificada por la SEQ ID NO: 271)	95	6,7	#
SEQ ID NO: 184 (codificada por la SEQ ID NO: 183)	95	No detectado	#
SEQ ID NO: 282 (codificada por la SEQ ID NO: 281)	05	36,7	###
SEQ ID NO: 186 (codificada por la SEQ ID NO: 185)	95	4,2	#
SEQ ID NO: 192 (codificada por la SEQ ID NO: 191)	95	11,9	#
SEQ ID NO: 200 (codificada por la SEQ ID NO: 199)	95	17,9	##
SEQ ID NO: 280 (codificada por la SEQ ID NO: 279)	50	No detectado	#
SEQ ID NO: 284 (codificada por la SEQ ID NO: 283)	90	2,2	#
SEQ ID NO: 172 (codificada por la SEQ ID NO: 171)	95	8,4	#
SEQ ID NO: 180 (codificada por la SEQ ID NO: 179)	98+	61,0	###
SEQ ID NO: 168 (codificada por la SEQ ID NO: 167)	98+	9,3	#
SEQ ID NO: 228 (codificada por la SEQ ID NO: 227)	98+	38,7	###
SEQ ID NO: 236(codificada por la SEQ ID NO: 235)	95	13,1	#
SEQ ID NO: 238 (codificada por la SEQ ID NO: 237)	98+	22,3	##
SEQ ID NO: 240 (codificada por la SEQ ID NO: 239)	95	No detectado	#
SEQ ID NO: 270 (codificada por la SEQ ID NO: 269)	40	4,6	#
SEQ ID NO: 156(codificada por la SEQ ID NO: 155)	98+	133,0	###

Sírvase notar que las selectividades para R-MP de 98+% indican que no se detectó S-MP alguno. Dada la sensibilidad del ensayo de CE, los resultados indican que al menos el 98+% de MP formado es el enantiómero R.

Por lo tanto, las enzimas que se indican con 98+% son al menos 98% selectivas hacia R-MP y pueden ser hasta 100% selectivas.

5 La tabla 2 también muestra la actividad de las enzimas en un sustrato de aldolasa general, CHA, así como también la expresión relativa de cada enzima, como determina SDS-PAGE. Sírvase notar que varias enzimas no mostraron una actividad detectable en CHA, pero exhibieron actividad al generar un MP.

En resumen, las aldolasas muestran una amplia gama de actividades, expresiones y selecciones. Asimismo, hay muchas aldolasas que muestran selectividades altamente finas (98% o más) para R-MP.

### Ejemplo de referencia 6

Descubrimiento de aldolasas de piruvato en plantas

10 Se diseñaron y utilizaron cebadores degenerados de PCR (ver a continuación) para extraer genes de aldolasa de ADNc preparado de *Sclerochiton ilicifolius*. Los extremos 5' y 3' de los genes se recuperaron y los genes de longitud total se amplificaron por PCR.

<u>SEQ ID NO:</u>	<u>Nombre del cebador</u>	<u>Cebador</u>
335	F1	AARGTBTWGARGACAATG (SEQ ID NO: 335)
336	F2	GCDGAGAWCAAYGGRTGG (SEQ ID NO: 336)
337	R1	CCATCRSYATCDGCRADAGCCA (SEQ ID NO: 337)
338	R2	GCRTADAGCCAYTCNCCRTC (SEQ ID NO: 338)

### Ejemplo 7

15 Clonación de D-aminoácido aminotransferasa de *Bacillus sphaericus*

La D-aminoácido aminotransferasa de *Bacillus sphaericus* (EC 2.6.1.21 también conocida como D-alanina aminotransferasa o D-aspartato aminotransferasa) se produjo recombinantemente para su uso en ensayos combinados con varias aldolasas. Esta enzima es homóloga a las D-amino transferasas anteriormente descritas para la producción de monatina (publicación de los Estados Unidos No. 20040063175 y publicación de los Estados Unidos No. 20050282260).

*Cepas*

Se cultivó *B. sphaericus* (ATCC número 10208) en nutriente agar a 30°C durante 24 horas. Se colocaron grupos de colonias en 100 µL de agua estéril y se calentaron durante 5 minutos a 95°C para alterar las células. Se utilizaron tres µL en amplificaciones posteriores de Reacciones en Cadena de la Polimerasa (PCR).

25 *Protocolo de Reacción en cadena de la polimerasa*

Se diseñaron cebadores para clonar en vectores pET 28b y pET 30a (Novagen, Madison, WI), utilizando los sitios NcoI y BamHI. La construcción pET 30 contiene una terminal N marcada con Histidina y Serina mientras que la construcción pET28 no está marcada.

Cebadores de *Bacillus sphaericus*:

30 Terminal N: 5'-GATATACCATGGCATACTCATTATGGAATG-3' (SEQ ID NO: 383) y terminal C: 5'-GTTATCGGATCCTTAGGCATTAATTGAAATTG-3' (SEQ ID NO: 384)

35 La región de codificación se amplificó utilizando el siguiente protocolo de PCR. En una reacción de 50 µl se utilizaron 3 µl de plantilla, 1,6 µl de cada cebador, 0,25 mM de cada dNTP, 3,5 U de Polimerasa de Alta Fidelidad (Roche, Indianápolis, IN) y un tampón IX de Expand™ con Mg. El programa del termociclador utilizado incluyó un arranque en caliente a 94°C durante 3 minutos, seguido de 8 repeticiones de las siguientes etapas: 94°C durante 30 segundos, 52°C durante 30 segundos, y 72°C durante 2 minutos. Se realizaron veintidós ciclos con una temperatura homogénea de 58°C. Después de 30 ciclos, la muestra se mantuvo a 72°C durante 7 minutos y posteriormente se almacenó a 4°C. Se obtuvieron los productos libres de PCR del tamaño correcto (aproximadamente 850 bp para dicho gen DAT).

40

## Clonación

Los productos PCR se purificaron utilizando un kit de purificación de PCR QIAquick de Qiagen (Qiagen, Valencia, CA) y se digirieron con BamHI y NcoI en un tampón BamHI (New England Biolabs, Ipswich, MA).

- 5 El vector y los insertos digeridos se purificaron utilizando el Kit de Extracción de Gel QIAquick de Qiagen (Qiagen, Valencia, CA). Las ligaduras se realizaron utilizando el Kit de Ligadura Rápida de Roche (Roche, Indianápolis, IN) y se purificaron utilizando el kit de purificación de PCR QIAquick. Las ligaduras se transformaron en *Escherichia coli* de la cepa DH10B utilizando una cubeta de 0,2 cm y un sistema Gene Pulser II de Bio Rad como describe el manual de electroporación de Bio Rad (Bio Rad, Hercules, CA). Las células se recuperaron en 900 µL de SOC por 30 minutos a 37°C a 225 rpm. Las células se ubicaron en placas de LB agar que contenían kanamicina (25µg/mL).
- 10 El ADN plasmídico se purificó utilizando un kit spin miniprep de Qiagen (Qiagen, Valencia, CA) y se monitoreó respecto de los insertos correctos para la digestión de restricción con BamHI y NcoI. Las secuencias de plásmidos que parecían tener el inserto correcto se verificaron a través de la secuencia de terminación de la cadena didesoxi del ADN en Agencourt BioScience Corporation (Beverly, MA). La secuencia verificó la secuencia de codificación encontrada en la Región número de orden AF081278 de NCBI: 134.985 (gi: 3513754), que produce una proteína con una secuencia de aminoácido como enumera el número de orden AAC33964 (gi: 3513755).
- 15

## Expresión génica y ensayos

- 20 El ADN plasmídico se subclonó en el huésped de expresión de *E.coli* BL21 (DE3) (Novagen, Madison, WI). Los cultivos se desarrollaron y los plásmidos se aislaron utilizando el kit miniprep de Qiagen (Qiagen, Valencia, CA) y se analizaron mediante una digestión de restricción para confirmar la identidad. La inducción se produjo en un medio LB que contenía kanamicina (50 µg/mL). Las células se desarrollaron a una OD 600 entre 0,4 y 0,8, inducidas con 0,1 mM de IPTG (isopropil tiogalactósido) y se tomaron muestras a las 4 horas después de la inducción. Se prepararon extractos celulares conforme al protocolo que acompaña el reactivo de Novagen BugBuster™ (Novagen, Madison, WI) (con benzonucleasa y una mezcla de inhibidor de proteasa completo de Roche (Roche, Indianápolis, IN)). Se obtuvieron niveles muy altos de proteína soluble con el peso molecular previsto, como indica SDS-PAGE.
- 25 Para algunas reacciones, el producto génico pET 30 se purificó utilizando cartuchos His-Bind conforme a los protocolos del fabricante (Novagen, Madison, WI). Las fracciones de eluyente se desalaron en columnas PD-10 (GE Healthcare, Piscataway, NJ) y se eluyeron en 25-100 mM de tampón de fosfato de potasio, con un pH de 7,5.

- 30 Los extractos celulares se analizaron para la actividad de D-aminotransferasa mediante la producción de alanina a partir de piruvato y D-triptófano utilizando el siguiente protocolo. Las reacciones de 1 mL se desarrollaron generalmente en 100mM de tampón de fosfato de potasio (pH 7,5), 50 µM de piridoxal fosfato, 25 mM de piruvato de sodio y 50 mM de D-triptófano. Las reacciones se iniciaron mediante la incorporación de extractos libres de células o enzimas purificadas y se incubaron durante entre 15 minutos y 24 horas a 30°C con agitación suave. Se agregó ácido fórmico a la concentración definitiva de dos por ciento para detener la reacción y la proteína precipitada se eliminó mediante centrifugación. También se realizaron reacciones de control sin proteína agregada. También se utilizaron puntos de tiempo cero como controles negativos. Se detectó alanina utilizando una derivatización OPA como se describe en el Ejemplo 1.
- 35

## Ejemplo de referencia 8

Comparación de la Producción total de monatina y la distribución isomérica para los polipéptidos con actividad de aldolasa de SEQ ID NO: 8; SEQ ID NO: 4, SEQ ID NO: 12 y SEQ ID NO: 28 y ProA de *C. testosteroni*.

- 40 Se adquirió transaminasa AT-103 (una D-aminotransferasa de amplia especificidad) de BioCatalytics (Pasadena, CA) y esta enzima o la enzima recombinante producida en el Ejemplo 7 se utilizó en reacciones combinadas con aldolasas de HMG para producir monatina a partir de D-triptófano y piruvato como se describe en la solicitud de patente de los Estados Unidos publicada No. 20050282260. La aldolasa ProA de *C. testosteroni* se utilizó como un aldolasa de referencia a los fines comparativos y se preparó como se describe en las solicitudes de patente no. 20040063175 y WO 03091396 A2 publicadas. Las aldolasas probadas se aislaron y transformaron como se describe en el Ejemplo 4 anterior.
- 45

- 50 Para producir cantidades de muestra de cada aldolasa, se desarrollaron cultivos de 50 mL en un medio LB que contenía ampicilina (100 µg/mL) a una OD 600 de aproximadamente 0,5. Las cepas con SEQ ID NO: 7, SEQ ID NO: 3, y SEQ ID NO: 11 se indujeron con 100µM de IPTG. La cepa con SEQ ID NO:27 se indujo con 200 µg/L de anhidrotetraciclina. Las células se cultivaron durante 5 horas después de la inducción y los extractos celulares se prepararon conforme a los protocolos del fabricante (Novagen, Madison, WI, reactivo Bugbuster). Se agregaron inhibidores de benzonucleasa y proteasa. Las proteínas solubles en los extractos celulares se separaron en una Estación de Electroforesis Automatizada-Experion de Bio-Rad Laboratories y se analizaron para verificar su concentración y porcentaje utilizando la versión de Software de Experion 1.1.98.0.

- 55 Se incorporaron por cada 1 mL de mezcla de reacción: aproximadamente 50 µg de aldolasa (proporcionada en extractos celulares salvo que se indique lo contrario), 4 mM de MgCl<sub>2</sub>, 50 mM de D-triptófano, 0,5 mg de D-

aminotransferasa de *B. sphaericus* purificada, 200 mM de piruvato de sodio, 100mM de tampón de fosfato de potasio con un pH de 7,5 y 0,05 mM de PLP. Los experimentos se desarrollaron por duplicado con controles negativos a los cuales no se les incorporó aldolasa. Las muestras se incubaron durante 1 hora, 2 horas y 24 horas (entre 17 y 20 horas) a 30°C con una leve agitación. Se producen pequeñas cantidades de monatina (<0,5 ppm) sin aldolasa en reacciones de 24 horas, dadas las reacciones no enzimáticas catalizadas por magnesio y fosfato. Estos valores se restaron de las cifras reflejadas a continuación y se muestran los resultados promedio. Los únicos estereoisómeros detectados cuando se produce monatina utilizando estos procedimientos son los estereoisómeros R,R y S,R. El porcentaje de R,R se enumera a continuación y se determinó a través del área de pico de LC de fase inversa.

10 **Tabla 3: Total de monatina producida de D-triptófano y % de R,R**

Aldolasa (punto de tiempo)	Total de monatina (ppm)	% de monatina R,R
SEQ ID NO: 8 (1 hora)	15,65	89,7
SEQ ID NO: 8 (18 horas)	129,22	79,0
SEQ ID NO: 4 (1 hora)	3,22	94,8
SEQ ID NO: 4 (18 horas)	12,14	93,8
SEQ ID NO: 12 (1 hora)	2,35	100
SEQ ID NO: 12 (18 horas)	11,89	98,65
SEQ ID NO: 28 (1 hora)	14,70	100
SEQ ID NO: 28 (18 horas)	95,14	97,35
Enzima purificada (ProA) de <i>C. testosteroni</i> ProA (1 hora)	16,63	86,45
Enzima purificada (ProA) de <i>C. testosteroni</i> (18 horas)	86,86	63,1

La muestra de 18 horas de la SEQ ID NO: 28 también se analizó para la distribución estereoisomérica mediante el procedimiento de derivatización de FDAA enumerado en el Ejemplo 1, que dio un resultado de 94,9% de monatina R,R y 5,1% de monatina S,R.

15 Se realizaron los mismos experimentos lado a lado utilizando L-triptófano como el sustrato inicial y combinando las aldolasas con L-aminotransferasa de especificidad amplia HexAspC producida como se describe en la solicitud de patente publicada No. 20050282260 y purificada. Las reacciones deberían principalmente generar monatina S,S y monatina R,S. Las reacciones también se complementaron con 10mM de alfa-cetoglutarato como el aceptor de amino para la transaminación de L-triptófano. Nuevamente, los resultados duplicados se promedian a continuación para la monatina total (restando los niveles básicos sin la presencia de aldolasa) y el porcentaje de monatina S,S se refleja en base al área de pico de LC de fase inversa. En algunos casos, dado que las aldolasas son R-específicas y producen poca monatina en total, las estimaciones de fase inversa de la distribución estereoisomérica son menos precisas por la asimetría del pico de triptófano que se puede coeluir con el pico de monatina S,S/R,R. Las tendencias aún son informativas al comparar la especificidad R de las aldolasas. Los resultados de mayores análisis utilizando el procedimiento de derivatización de FDAA se reflejan entre paréntesis para las distintas muestras y son más precisos. Las cifras totales de monatina por encima de aproximadamente 400 ppm son mayores que la gama lineal de la escala de los parámetros utilizados para cuantificar los resultados, por lo que son resultados cualitativos. La aldolasa ProA de *C. testosteroni* generalmente produce entre 95 y 100% de monatina S,S como se refleja en la solicitud de Patente de los Estados Unidos No. 2005028266 publicada.

30 **Tabla 4: Monatina total producida de L-triptófano y % de S,S**

Aldolasa (punto de tiempo)	Monatina total (ppm)	% de monatina S,S
SEQ ID NO:8 (1 hora)	138,65	78,9
SEQ ID NO:8 (18 horas)	600,3	78,15

SEQ ID NO:4 (1 hora)	por debajo del control negativo	95,65
SEQ ID NO:4 ( 18 horas)	28,5	87,6
SEQ ID NO:12 (1 hora)	por debajo del control negativo	93,55
SEQ ID NO:12 (18 horas)	24,9	75 (59,35)
SEQ ID NO:28 (1 hora)	17,85	55,05 (18,9)
SEQ ID NO:28 (18 horas)	135,5	27,25 (19,1)
Enzima purificada ProA de <i>C. testosteroni</i> (1 hora)	440,35	92,5
Enzima purificada ProA de <i>C. testosteroni</i> (18 horas)	958,3	92,15

5 Se puede ver que la especificidad R del polipéptido con actividad de aldolasa de SEQ ID NO: 28 es relativamente alto en relación con la enzima ProA de referencia. Esto también se refleja en el bajo % de monatina S,S producido a pesar del alto grado de especificidad de la aminotransferasa HexAspC para S-MPen estas reacciones. Las cifras totales de monatina, cuando se compara la producción de monatina S,S respecto de la producción de monatina R,R no son indicativas de la actividad de aldolasa. La D-aminotransferasa es menos activa que HexAspC, especialmente en las concentraciones de MP que están presentes en estas reacciones.

10 Para mayor comparación del polipéptido con actividad de aldolasa de SEQ ID NO: 28 con la enzima ProA de *C. testosteroni*, se utilizaron varias proporciones de D-aminotransferasa respecto de aldolasa en reacciones que comenzaron con D-triptófano (no se realizaron muestras duplicadas para estos experimentos). Las reacciones se realizaron como se explica anteriormente. Para las reacciones en las cuales la concentración de aldolasa se mantuvo constante, se utilizó aproximadamente 50 µg. Para reacciones en las cuales la D-aminotransferasa se mantuvo constante, se utilizó 0,5 mg. Para la concentración de 2 y 10 mg/mL de D-aminotransferasa, se utilizó una enzima liofilizada. Para las dos concentraciones más altas de D-aminotransferasa se realizaron duplicados.

15

**Tabla 5: Efecto de la concentración de D-aminotransferasa en la producción de monatina R,R**

Aldolasa	Concentración de D-aminotransferasa	Tiempo	Monatina total (ppm aproximado)	% de monatina R,R
SEQ ID NO: 28	0,25 mg/ml	1 hora	2	100
SEQ ID NO: 28	0,25 mg/ml	24 horas	141	97,1
SEQ ID NO: 28	0,5 mg/ml	1 hora	8	100
SEQ ID NO: 28	0,5 mg/ml	24 horas	273	96,5
SEQ ID NO: 28	1 mg/ml	1 hora	34	100
SEQ ID NO: 28	1 mg/ml	24 horas	638	96,5
SEQ ID NO: 28	2 mg/ml	1 hora	979	100
SEQ ID NO: 28	2 mg/ml	24 horas	1910	97,3
SEQ ID NO: 28	10 mg/ml	1 hora	2930	99,1
SEQ ID NO: 28	10 mg/ml	24 horas	2950	96,5
ProA (purificada) de <i>C. testosteroni</i>	0,25 mg/ml	1 hora	4	78,7

Aldolasa	Concentración de D-aminotransferasa	Tiempo	Monatina total (ppm aproximado)	% de monatina R,R
ProA (purificada) de <i>C. testosteroni</i>	0,25 mg/ml	24 horas	257	61,1
ProA (purificada) de <i>C. testosteroni</i>	0,5 mg/ml	1 hora	25	79,0
ProA (purificada) de <i>C. testosteroni</i>	0,5 mg/ml	24 horas	480	62,5
ProA (purificada) de <i>C. testosteroni</i>	1 mg/ml	1 hora	74	73,8
ProA (purificada) de <i>C. testosteroni</i>	1 mg/ml	24 horas	810	68,1
ProA (purificada) de <i>C. testosteroni</i>	2 mg/ml	1 hora	325	73,1
ProA (purificada) de <i>C. testosteroni</i>	2 mg/ml	24 horas	2220	71,9
ProA (purificada) de <i>C. testosteroni</i>	10 mg/ml	1 hora	2910	59,7
ProA (purificada) de <i>C. testosteroni</i>	10 mg/ml	24 horas	2450	67,5
SEQ ID NO: 8	0,25 mg/ml	1 hora	4	92,3
SEQ ID NO: 8	0,25 mg/ml	24 horas	219	69,8
SEQ ID NO: 8	0,5 mg/ml	1 hora	14	84,9
SEQ ID NO: 8	0,5 mg/ml	24 horas	426	67,5
SEQ ID NO: 8	1 mg/ml	1 hora	62	84,2
SEQ ID NO: 8	1 mg/ml	24 horas	877	68,7

5 Para los niveles de monatina superiores a 400 ppm, los resultados no se encuentran en la gama lineal de la curva estándar y son valores aproximados únicamente. La cantidad máxima de monatina R,R producida, cuando se diluyó correctamente fue de 1100 ppm. El análisis estereoisómero de FDAA se realizó para el polipéptido con actividad de aldolasa de SEQ ID NO: 28 con muestras de D-aminotransferasa de 10 mg/mL. A las dos horas, la muestra contenía un 98,5% de monatina R,R. A las 17 horas, la muestra contenía un 95,9% de monatina R,R. El polipéptido con actividad de aldolasa de SEQ ID NO: 28 produjo porcentajes altos de monatina R,R, aún después de plazos largos de incubación y utilizando grandes cantidades de aminotransferasa. Si se proporciona D-aminotransferasa adecuada, el polipéptido con actividad de aldolasa de SEQ ID NO: 28 produce tanta monatina como la aldolasa ProA de *C. testosteroni* indicando una actividad específica similar.

10

**Tabla 6: Efecto de la concentración de aldolasa en la producción de monatina R,R**

Aldolasa	Concentración de aldolasa	Tiempo	Monatina total (ppm)	% de R,R monatina
SEQ ID NO: 28	25 µg/ml	1 hora	7,0	100

SEQ ID NO: 28	25 µg/ml	24 horas	275	97,4
SEQ ID NO: 28	25 µg/ml	1 hora	9,0	97,3
SEQ ID NO: 28	25 µg/ml	24 horas	334	95,7
SEQ ID NO: 28	100 µg/ml	24 horas	297	93,3
ProA (purificada) de <i>C. testosteroni</i>	25 µg/ml	1 hora	16	78,2
ProA (purificada) de <i>C. testosteroni</i>	25 µg/ml	24 horas	491	73,2
ProA (purificada) de <i>C. testosteroni</i>	50 µg/ml	1 hora	18	64,1
ProA (purificada) de <i>C. testosteroni</i>	50 µg/ml	24 horas	437	63,0
ProA (purificada) de <i>C. testosteroni</i>	100 µg/ml	1 hora	26	62,5
ProA (purificada) de <i>C. testosteroni</i>	100 µg/ml	24 horas	513	61,5
SEQ ID NO: 8	25 µg/ml	1 hora	11,0	88,1
SEQ ID NO: 8	25 µg/ml	24 horas	337,0	74,7
SEQ ID NO: 8	50 µg/ml	1 hora	14,0	78,2
SEQ ID NO: 8	50 µg/ml	24 horas	406,0	67,8
SEQ ID NO: 8	100 µg/ml	1 hora	24,0	70,1
SEQ ID NO: 8	100 µg/ml	24 horas	329,0	63,9

Cuando se varía la concentración de aldolasa, no hay mucho aumento en la monatina total. El porcentaje de R,R disminuye con el tiempo y también con la concentración de aldolasa, especialmente cuando no hay suficiente D-aminotransferasa.

- 5 Para poder examinar la especificidad R de la aldolasa probada, los experimentos se realizaron partiendo de L-triptófano y aminotransferasa HexAspC, que se produjo y purificó como se describe en la solicitud de patente de los Estados Unidos publicada No. 20050282260. La HexAspc muestra una selectividad fuerte para la transaminación de S-MP versus R-MP, por lo tanto, los porcentajes por encima del 50% de monatina R,S indican una aldolasa altamente estereoespecífica. Se proporcionaron diez mM de alfa-cetoglutarato como un aceptor de amino. Sin embargo, en concentraciones altas, la L-aminotransferasa también utiliza piruvato. En estas reacciones, generalmente solo se producen monatina S,S y monatina R,S dentro de los límites de detección del protocolo de derivatización de FDAA.
- 10

Tabla 7: Efecto de la concentración de L-aminotransferasa en la producción de monatina S,S

Aldolasa	Concentración de D-aminotransferasa	Tiempo	Monatina total (ppm aproximado)	% de Monatina S,S
SEQ ID NO: 28	0,25 mg/ml	1 hora	13	33,8
SEQ ID NO: 28	0,25 mg/ml	24 horas	127	34,2
SEQ ID NO: 28	0,5 mg/ml	1 hora	31	30,9
SEQ ID NO: 28	0,5 mg/ml	24 horas	272	26,8
SEQ ID NO: 28	1 mg/ml	1 hora	34	20,3
SEQ ID NO: 28	1 mg/ml	24 horas	385	23,5
ProA (purificada) de <i>C. testosteroni</i>	0,25 mg/ml	1 hora	523	94,2
ProA (purificada) de <i>C. testosteroni</i>	0,25 mg/ml	24 horas	1817	93,7
ProA (purificada) de <i>C. testosteroni</i>	0,5 mg/ml	1 hora	602	91,8
ProA (purificada) de <i>C. testosteroni</i>	0,5 mg/ml	24 horas	2122	89,9
ProA (purificada) de <i>C. testosteroni</i>	1 mg/ml	1 hora	873	90,2
ProA (purificada) de <i>C. testosteroni</i>	1 mg/ml	24 horas	1237	82,6
SEQ ID NO: 8	0,25 mg/ml	1 hora	339	86,3
SEQ ID NO: 8	0,25 mg/ml	24 horas	1499	88,0
SEQ ID NO: 8	0,5 mg/ml	1 hora	211	80,3
SEQ ID NO: 8	0,5 mg/ml	24 horas	1328	83,1
SEQ ID NO: 8	1 mg/ml	1 hora	400	74,6
SEQ ID NO: 8	1 mg/ml	24 horas	1370	79,0



Tabla 8: Efecto de la concentración de aldolasa en la producción de monatina S,S

Aldolasa	Concentración de aldolasa	Tiempo	Monatina total (ppm)	% de Monatina S,S
SEQ ID NO: 28	25 µg/ml	1 hora	11	25,1
SEQ ID NO: 28	25 µg/ml	24 horas	112	20,0
SEQ ID NO: 28	50 µg/ml	1 hora	18	31,8
SEQ ID NO: 28	50 µg/ml	24 horas	160	27,0
SEQ ID NO: 28	100 µg/ml	1 hora	33	33,2
SEQ ID NO: 28	100 µg/ml	24 horas	238	41,4
ProA (purificada) de <i>C. testosteroni</i>	25 µg/ml	1 hora	305	86,4
ProA (purificada) de <i>C. testosteroni</i>	25 µg/ml	24 horas	1094	87,5
ProA (purificada) de <i>C. testosteroni</i>	50 µg/ml	1 hora	575	90,9
ProA (purificada) de <i>C. testosteroni</i>	50 µg/ml	24 horas	1449	89,9
ProA (purificada) de <i>C. testosteroni</i>	100 µg/ml	1 hora	817	93,6
ProA (purificada) de <i>C. testosteroni</i>	100 µg/ml	24 horas	1360	89,7
SEQ ID NO: 8	25 µg/ml	1 hora	134	70,7
SEQ ID NO: 8	25 µg/ml	24 horas	728	76,3
SEQ ID NO: 8	50 µg/ml	1 hora	197	80,0
SEQ ID NO: 8	50 µg/ml	24 horas	928	81,4
SEQ ID NO: 8	100 µg/ml	1 hora	279	86,7
SEQ ID NO: 8	100 µg/ml	24 horas	1383	86,8

5 Para aldolasas que son altamente R-específicas, como el polipéptido con actividad de aldolasa de SEQ ID NO: 28, se produce menos monatina total y la mayor cantidad de aldolasa aumenta la monatina total (así como también el % de S,S). Estas aldolasas producen menos sustrato de S-MP, el sustrato preferido para la L-aminotransferasa utilizado. Para enzimas que son menos específicas para R, como ProA, la mayor cantidad de aldolasa no mejora significativamente la producción total de monatina o el % de monatina S,S. Aumentar la cantidad de L-aminotransferasa agregada disminuye el porcentaje de monatina S,S producida. En base al análisis anterior, el polipéptido con actividad de aldolasa de SEQ ID NO: 8 se encuentra entre ProA y el polipéptido con actividad de aldolasa de SEQ ID NO: 28 en cuanto a la R-especificidad, que coincide con la información anterior donde el % de R-MP se mide para la etapa de aldol únicamente.

10

Subclonación de SEQ ID NO: 27

Los siguientes cebadores se utilizaron para ampliar en PCR el gen de aldolasa: 5'-gaggagctcgagtcagacgtatttcagtccttttc-3' (SEQ ID NO: 385) y 5'-agaagacatgatgattatcagccggggac-3' (SEQ ID NO: 386). La SEQ ID NO: 27 del gen de aldolasa codifica el polipéptido con actividad de aldolasa de SEQ ID NO: 28. El producto resultante de PCR se digirió con *XhoI* y *NdeI* para cortar en los sitios que habían sido integrados en los cebadores. El fragmento se purificó con gel (Kit de extracción de Gel QIAquick (Qiagen, Valencia, CA)) y se ligó (utilizando T4 ADN ligasa) con pET28b que había sido digerido con *XhoI* y *NdeI* y purificado con gel. La ligadura se transformó en células químicamente competentes TOP10F'. Las colonias que crecen en las placas se monitorearon para verificar insertos y varias aisladas con insertos se sometieron al análisis de secuencia del ADN (Agencourt, Beverly, MA).

Purificación del Polipéptido con actividad de aldolasa de SEQ ID NO: 28.

Los clones confirmados de aldolasa se transformaron en BL21 DE3 o BL21DE3 pLyss. Los cultivos de 24 horas desarrollados con el antibiótico correcto se diluyeron en un medio limpio (generalmente 1:100) y crecieron hasta alcanzar una  $DO_{600}$  de ~0,6 con una aireación a 37°C. Los cultivos se indujeron con 1mM de IPTG y se cambiaron a 30°C (con aireación) y se continuó con la incubación durante la noche. Las células se cultivaron mediante centrifugación. El sedimento celular se sometió a un ciclo de congelación-descongelación para ayudar a la lisis celular. El sedimento celular se lisó en BugBuster y Benzona (Novagen, Madison, WI) (conforme al protocolo del fabricante). El residuo celular se eliminó mediante centrifugación. El extracto proteico crudo se aplicó a una columna HisBind (Novagen, Madison, WI) que había sido preparado conforme al protocolo del fabricante. La columna se lavó y la proteína se eluyó conforme al protocolo del fabricante. La proteína purificada se desaló con columnas PD-10 (GE Healthcare, Piscataway, N.J.). El tampón utilizado para el intercambio consistió en 50 mM de fosfato de potasio con pH de 7,5, 100 mM de NaCl<sub>2</sub>, 4 mM de MgCl<sub>2</sub>. La proteína purificada se concentró con concentradores centrífugos Amicon (Millipore, Billerica, MA).

**Ejemplo de referencia 9**

Comparación de la producción total de monatina y la distribución isomérica para polipéptidos con actividad de aldolasa de SEQ ID NO: 40, SEQ ID NO: 298, SEQ ID NO: 36, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 96, SEQ ID NO: 54, SEQ ID NO: 122, SEQ ID NO: 142, SEQ ID NO: 130, SEQ ID NO: 112, SEQ ID NO: 94, SEQ ID NO: 80 y SEQ ID NO: 28.

La transaminasa AT-103 (una D-aminotransferasa de especificidad amplia) se adquirió de BioCatalytics (Pasadena, CA) y esta enzima o la enzima recombinante producida en el Ejemplo 7 se utilizó en reacciones combinadas con aldolasas de HMG para producir monatina a partir de D-triptófano y piruvato como se describe en la solicitud de patente de los Estados Unidos publicada No. 20050282260. El polipéptido con actividad de aldolasa de SEQ ID NO: 28 (marcada con histidina) se utilizó como una aldolasa de referencia a los fines comparativos y se produjo y purificó como se describe al final del Ejemplo 8. Las otras aldolasas probadas se aislaron y transformaron como se describe en el Ejemplo 4 anterior.

Para producir cantidades de muestra de cada aldolasa, se desarrollaron cultivos de 25mL en un medio LB que contenía ampicilina (100 µg/mL), a una  $OD_{600}$  de aproximadamente 0,4. Las cepas se indujeron con 100 µM de IPTG. Las células se desarrollaron por 4 horas con posterioridad a la inducción, y los extractos celulares se prepararon conforme a los protocolos del fabricante (Novagen, Madison, WI, reactivo Bugbuster) con benzonucleasa. Las proteínas solubles en los extractos celulares se separaron en una Estación de Electroforesis Automática-Experion de Bio-Rad Laboratories (Bio-Rad, Hércules, CA) y se analizaron para verificar la concentración y el porcentaje utilizando la versión de Software de Experion 1.1.98.0.

Se incorporaron por cada 1 mL de mezcla de reacción: aproximadamente 50 µg de aldolasa (proporcionada en extractos celulares salvo que se indique lo contrario), 4 mM de MgCl<sub>2</sub>, 50 mg/mL de D-triptófano, 2 mg de AT-103 (BioCatalytics, Pasadena, CA), 200 mM de piruvato de sodio, 100mM de tampón de fosfato de potasio con un pH de 7,5 y 0,05 mM de PLP. El D-triptófano no es soluble a esta concentración alta, pero se utilizó para garantizar que las reacciones se mantuvieran en cantidades de saturación de D-triptófano. Los experimentos se desarrollaron por duplicado con controles negativos donde no se incorporó aldolasa. Las muestras se incubaron durante 2 y 24 horas (entre 17 y 20 horas) a 30°C con una leve agitación. Las pequeñas cantidades de monatina se producen en 24 horas sin aldolasa (~0,5 ppm) dadas las reacciones no enzimáticas catalizadas por magnesio y fosfato. Los valores típicos para el % de monatina R,R son 50% para estas muestras. Los valores de control negativo se restaron de los números reflejados a continuación y se muestran los resultados medios. Los únicos estereoisómeros detectados cuando se produce monatina utilizando estos procedimientos son los estereoisómeros R,R y S,R. El porcentaje de R,R se enumera a continuación y se determinó mediante el área de pico de fase inversa. Se desarrolló el mismo experimento después del almacenamiento de extractos celulares y del polipéptido purificado con actividad de aldolasa de SEQ ID NO: 28 para 2 meses a -20° C. En este caso se utilizó 50 mM de D-triptófano como en el Ejemplo 8. Se agregó el doble de cantidad de aldolasa con la excepción del polipéptido con actividad de aldolasa de

SEQ ID NO: 28 para el cual se utilizó 50 µg nuevamente. Estos resultados se reflejan a la derecha de la Tabla 9. Los resultados de la derivatización de FDAA para la distribución isomérica se reflejan entre paréntesis.

**Tabla 9: monatina total producida a partir de D-triptófano y % de R,R**

Aldolasa (punto de tiempo)	Monatina total (ppm)	% de R,R monatina	Monatina total (ppm)	% de monatina R,R
SEQ ID NO: 40 (2 horas)	336	82,7	238	66,8
SEQ ID NO: 40 (18 horas)	707,55	76,2	748,5	62,4
SEQ ID NO: 298 (2 horas)	394,3	98,0	183	91,9 (91,6)
SEQ ID NO: 298 (18 horas)	819,5	96,1	648,5	80,0
SEQ ID NO: 36 (2 horas)	56	98,2	52,5	94,0
SEQ ID NO: 36 (18 horas)	123,25	96,9	296	88,5
SEQ ID NO: 62 (2 horas)	1,15	78,4	0	n/d
SEQ ID NO: 62 (18 horas)	0,95	73,8	16	89,2
SEQ ID NO: 64 (2 horas)	16,7	98,8	24	96,9
SEQ ID NO: 64 (18 horas)	43,7	97,6	161	98,5
SEQ ID NO: 96 (2 horas)	30,4	99,2	29	96,0
SEQ ID NO: 96 (18 horas)	80,8	98,3	200	97,3
SEQ ID NO: 54 (2 horas)	183,1	99,4	135,5	98,2 (98,8)
SEQ ID NO: 54 (18 horas)	457,7	98,9	488,5	96,9
SEQ ID NO: 122 (2 horas)	129,3	97,9	126	97,8 (99,1)
SEQ ID NO: 122 (18 horas)	289,85	95,8	471,5	94,4
SEQ ID NO: 142 (2 horas)	40,4	98,3	58,5	95,9
SEQ ID NO: 142 (18 horas)	82,3	97,3	388	96,8
SEQ ID NO: 42 (2 horas)	335,9	98,2	206,5	93,3
SEQ ID NO: 42 (18 horas)	612,45	96,6	630,5	82,9
SEQ ID NO: 130 (2 horas)	77,5	99,3	60,5	98,5 (99,6)
SEQ ID NO: 130 (18 horas)	177,45	99,1	368,5	98,4
SEQ ID NO: 112 (2 horas)	20,4	99,0	27	98,6
SEQ ID NO: 112 (18 horas)	57,75	98,3	186,5	99,3
SEQ ID NO: 108 (2 horas)	44,4	98,7	41	97,0
SEQ ID NO: 108 (18 horas)	111,7	98,0	265,5	93,3 (96,4)

Aldolasa (punto de tiempo)	Monatina total (ppm)	% de R,R monatina	Monatina total (ppm)	% de monatina R,R
SEQ ID NO: 94 (2 horas)	69,4	98,2	56	94,4
SEQ ID NO: 94 (18 horas)	181,95	96,9	341	84,8
SEQ ID NO: 80 (2 horas)	27,9	98,9	29,5	98,8
SEQ ID NO: 80 (18 horas)	74	97,9	219	96,6
SEQ ID NO: 28 - purificado (2 horas)	131,3	99,5	53	96,7 (99,6)
SEQ ID NO: 28 - purificado (18 horas)	407,4	99,2	257	99,2

5 Se puede ver que algunas enzimas son más estables al almacenamiento que otras aldolasas, en base a las proporciones de actividad. Un producto secundario, más probablemente 4-hidroxi-4-metil glutamato también puede formarse durante estas reacciones. Las enzimas mencionadas se clasificaron según su especificidad hacia la producción de monatina comparando sus áreas de pico en relación con el subproducto. Los resultados fueron el polipéptido con la actividad de aldolasa de SEQ ID NO: 122, SEQ ID NO: 80, SEQ ID NO: 108, SEQ ID NO: 96, SEQ ID NO: 122, SEQ ID NO: 130, SEQ ID NO: 36, SEQ ID NO: 94, SEQ ID NO: 298, SEQ ID NO: 40, SEQ ID NO: 142, SEQ ID NO: 54, SEQ ID NO: 64, SEQ ID NO: 28, SEQ ID NO: 62.

10 En base a los experimentos iniciales, los polipéptidos con actividad de aldolasa de SEQ ID NO: 298, SEQ ID NO: 54, y SEQ ID NO: 42 parecían ser los más prometedores en términos del nivel de actividad y el % de monatina R,R producida. Las enzimas se subclonaron en vectores de expresión pET con y sin marcas de histidina.

Clonación de SEQ ID NO: 297, SEQ ID NO: 53 y SEQ ID NO: 41.

Cebadores utilizados para la clonación:

Tabla 10

SEQ ID NO:	cebador 5'	cebador 3'
297	5'-agaagacatatgggtgctgctgccaac-3' (SEQ ID NO: 387)	5'-ataataggatccttagacatattgaggccc-3' (SEQ ID NO: 388)
53	5'-ataatacatatgaagccgggtggtgc-3' (SEQ ID NO: 389)	5'-agaagaggatccttagacataggtgagcccc-3' (SEQ ID NO: 390)
41	5'-ataataccatgggtgctggtccag-3' (SEQ ID NO: 391)	5'-agaagaggatccttagacatattcagcccc-3' (SEQ ID NO: 392)

15 La SEQ ID NO: 297, la SEQ ID NO: 53 y la SEQ ID NO: 41 se amplificaron por PCR y digirieron con las enzimas adecuadas (*NdeI* y *BamHI* para los productos en PCR que contienen SEQ ID NO: 297 y SEQ ID NO: 53, *NcoI* y *BamHI* para los productos en PCR que contienen SEQ ID NO: 41) y purificados con gel (Kit de extracción con Gel QIAquick (Qiagen, Valencia, CA)). La SEQ ID NO: 297 y la SEQ ID NO: 53 se ligaron individualmente en pET28 que había sido digerido con *NdeI* y *BamHI* y purificado con gel. La SEQ ID NO: 41 se ligó a pET30 que había sido digerido con *NcoI* y *BamHI* y purificado con gel. La ligadura se transformó en TOP10. Las colonias se monitorearon

20

para verificar insertos. Las colonias aisladas con un inserto se sometieron al análisis de secuencia de ADN (Agencourt, Beverly, MA).

#### Purificación de Aldolasas

5 Los clones confirmados de aldolasas se transformaron en BL21 DE3 o BL21 DE3 pLysS. Los cultivos de 24 horas desarrollados con el antibiótico adecuado se diluyeron en un medio limpio (generalmente 1:100) y alcanzaron una OD<sub>600</sub> de ~0,6 con una aireación a 37°C. Los cultivos posteriormente se indujeron con 1 mM de IPTG y se cambiaron a 30°C (con aireación) y la incubación se continuó durante 24 horas. Las células se cultivaron mediante centrifugación. El sedimento celular básicamente se sometió a un ciclo de congelación-descongelación para ayudar a la lisis celular. El sedimento celular se lisó en BugBuster y Benzonasa (Novagen, Madison, WI) (conforme al protocolo del fabricante). El residuo celular se eliminó mediante centrifugación. El extracto proteico crudo se aplicó a una columna HisBind (Novagen, Madison, WI) que había sido preparada conforme al protocolo del fabricante. La columna se lavó y la proteína se eluyó conforme al protocolo del fabricante. La proteína purificada se desaló con columnas PD 10 (GE Healthcare, Piscataway, N.J.). El tampón utilizado para el intercambio consistió en 50 mM de fosfato de potasio con un pH de 7,5, 100 mM de NaCl, y 4 mM MgCl<sub>2</sub>. La proteína purificada se concentró con concentradores centrífugos Amicon (Millipore, Billerica, MA).

#### Prueba de aldolasas purificadas

20 Las aldolasas purificadas se valoraron respecto de su capacidad para producir monatina R,R a partir de D- triptófano. Se incorporaron por cada 1 mL de mezcla de reacción: aproximadamente 50 µg de aldolasa purificada, 4 mM de MgCl<sub>2</sub>, 50 mM de D-triptófano, 0,5 mg de D-aminotransferasa de *B. sphaericus* purificada, 200 mM de piruvato de sodio, 100mM de tampón de fosfato de potasio con un pH de 7,5 y 0,05 mM de PLP. Las muestras se tomaron a las 2 horas y a las 24 horas. Los resultados se reflejan en la Tabla 11 a continuación.

**Tabla 11: monatina total producida a partir de D-triptófano y % de R,R**

Aldolasa (punto de tiempo)	Monatina total (ppm)	% de monatina R,R (área de pico de LC de fase inversa)	% de monatina R,R (derivatización de FDAA)
SEQ ID NO: 298 (2 horas)	16,95	88,5	n/d
SEQ ID NO: 298 (durante la noche)	212	77,6	71
SEQ ID NO: 54 (2 horas)	12,05	96,7	n/d
SEQ ID NO: 54 (durante la noche)	161,85	93,0	91,1
SEQ ID NO: 42 (2 horas)	20,95	80,3	n/d
SEQ ID NO: 42 (durante la noche)	223,2	69,1	62,1
SEQ ID NO: 28 (2 horas)	14,25	95,8	n/d
SEQ ID NO: 28 (durante la noche)	176,6	93,4	92,3

25 Los mismos experimentos se desarrollaron utilizando L-triptófano como el sustrato inicial y combinando las aldolasas con la L- aminotransferasa HexAspC de especificidad amplia producida y purificada como describe la solicitud de Patente de los Estados Unidos No. publicada 20050282260 (0,5 mg de proteína purificada). Los resultados se demuestran a continuación en la Tabla 12 para la producción total de monatina (restando los niveles básicos sin presencia de aldolasa) y el porcentaje de monatina S,S se demuestra en base al área de pico de LC de fase inversa. Las cifras superiores a 400 ppm están fuera de la gama lineal de la curva estándar, y son aproximados.

30

Tabla 12: monatina total producida a partir de L-triptófano y % de S,S

Aldolasa (punto de tiempo)	Monatina total (ppm)	% de monatina S,S (área de pico de LC de fase inversa)	% de monatina S,S (derivatización de FDAA)
SEQ ID NO: 298 (1 hora)	186,6	64,0	n/d
SEQ ID NO: 298 (durante la noche)	197,5	64,3	67,6
SEQ ID NO: 54 (1 hora)	70,4	36,5	n/d
SEQ ID NO: 54 (durante la noche ras)	87,8	41,7	42,1
SEQ ID NO: 42 (1 hora)	401,1	80,9	n/d
SEQ ID NO: 42 (durante la noche)	507,5	82,9	85,8
SEQ ID NO: 28 (1 hora)	56,2	30,1	n/d
SEQ ID NO: 28 (durante la noche)	88,8	32,2	33,8

Estos datos y los datos de la monatina R,R ilustran que para la especificidad R- MP, los polipéptidos con actividad de aldolasa tienen el siguiente orden: SEQ ID NO: 28, SEQ ID NO: 54, SEQ ID NO: 298, SEQ ID NO: 42.

## 5 Ejemplo de referencia 10

Comparación de la producción total de monatina y la distribución isomérica para los polipéptidos con actividad de aldolasa de SEQ ID NO: 116, SEQ ID NO: 76, SEQ ID NO: 44, SEQ ID NO: 148, SEQ ID NO: 46, SEQ ID NO: 134, SEQ ID NO: 74, SEQ ID NO: 126, SEQ ID NO: 102, SEQ ID NO: 58, SEQ ID NO: 88, SEQ ID NO: 50, SEQ ID NO: 106, SEQ ID NO: 304, SEQ ID NO: 300 y SEQ ID NO: 28.

10 La enzima recombinante producida en el Ejemplo 7 se utilizó en reacciones combinadas con aldolasas de HMG para producir monatina a partir de D-triptófano y piruvato como describe la solicitud de patente de los Estados Unidos publicada No. 20050282260. El polipéptido con actividad de aldolasa de SEQ ID NO: 28 se utilizó como referencia en estos ensayos y había sido purificado como se describe en el Ejemplo 8.

15 Para producir cantidades de muestra de cada aldolasa, se desarrollaron cultivos de 25 mL en un medio LB que contenía ampicilina (100 µg/mL), hasta alcanzar una  $DO_{600}$  de aproximadamente 0,5. Los cultivos se indujeron con 1mM de IPTG. Las células se cambiaron a 30°C y se desarrollaron en 24 horas. Los extractos celulares se prepararon conforme a los protocolos del fabricante (Novagen, Madison, WI, reactivo Bugbuster) También se agregó benzonucleasa. Las proteínas solubles en los extractos celulares se separaron en una Estación de Electroforesis Automática Experion de Bio Rad Laboratories (Bio Rad, Hércules, CA) y se analizaron para la concentración y el porcentaje utilizando la versión de Software Experion 1.1.98.0.

20 Se incorporaron por cada 1 mL de mezcla de reacción: aproximadamente 50 µg de aldolasa (proporcionada en extractos celulares salvo que se indique lo contrario), 4 mM de  $MgCl_2$ , 50 mM de D-triptófano, 0,5 mg de D-aminotransferasa de *B. sphaericus* purificada, 200 mM de piruvato de sodio, 100mM de tampón de fosfato de potasio con un pH de 7,5 y 0,05 mM de PLP. Se agregó ditioteitrol (DTT) (concentración definitiva de 2mM) a las muestras descritas a continuación. Los experimentos se desarrollaron por duplicado. Las muestras se incubaron durante 2 horas y durante la noche (20 horas) a 30°C con una leve agitación. Los resultados medios se muestran a continuación. Los únicos estereoisómeros detectados cuando se produjo monatina utilizando estos procedimientos son los estereoisómeros R,R y S,R. El porcentaje de R,R se enumera a continuación y se determinó por el área de pico de LC de fase inversa.

30

Tabla 13: monatina total producida a partir de D-triptófano y % de R,R

Aldolasa (punto de tiempo)	Monatina total (ppm)	% de monatina R,R
SEQ ID NO: 116 (2 horas)	34,5	97
SEQ ID NO: 116 (18 horas )	99	95
SEQ ID NO: 76 (2 horas)	40	76
SEQ ID NO: 76 (18 horas)	112	67
SEQ ID NO: 44 (2 horas)	32,5	97
SEQ ID NO: 44 (18 horas)	93,5	94
SEQ ID NO: 148 (2 horas)	31,5	94
SEQ ID NO: 148 (18 horas)	98	89
SEQ ID NO: 46 (2 horas)	42,5	84
SEQ ID NO: 46 (18 horas)	169	72
SEQ ID NO: 134 (2 horas)	43,5	92
SEQ ID NO: 134 (18 horas)	113	86
SEQ ID NO: 74 (2 horas)	23,5	96
SEQ ID NO: 74 (18 horas)	78,5	92
SEQ ID NO: 126 (2 horas)	18	94
SEQ ID NO: 126 (18 horas)	72	92
SEQ ID NO: 102 (2 horas)	1	0
SEQ ID NO: 102 (18 horas)	4,5	91
SEQ ID NO: 58 (2 horas)	23	92
SEQ ID NO: 58 (18 horas)	122	88
SEQ ID NO: 88 (2 horas)	57,5	74
SEQ ID NO: 88 (18 horas)	200,5	64
SEQ ID NO: 50 (2 horas)	32,5	99
SEQ ID NO: 50 (18 horas)	131,5	97
SEQ ID NO: 106 (2 horas)	4,5	78
SEQ ID NO: 106 (18 horas)	20	95
SEQ ID NO: 304 (2 horas)	0	0
SEQ ID NO: 304 (18 horas)	0,45	55
SEQ ID NO: 304 DTT (2 horas)	0	0
SEQ ID NO: 304 DTT (18 horas)	0,55	53

<b>Aldolasa (punto de tiempo)</b>	<b>Monatina total (ppm)</b>	<b>% de monatina R,R</b>
SEQ ID NO: 300 (2 horas)	0,85	34
SEQ ID NO: 300 (18 horas)	5,5	36
SEQ ID NO: 300 DTT (2 horas)	1,5	55
SEQ ID NO: 300 DTT (18 horas)	9	40
SEQ ID NO: 28 (2 horas)	25	99
SEQ ID NO: 28 (18 horas)	69	97

Los números de la producción total de monatina oscilaron entre 1 ppm y más de 200 ppm y el % de R,R osciló entre 0% y 99% . Dado que la aminotransferasa fue la misma para todas las aldolasas, el cambio de aldolasa puede afectar significativamente tanto la cantidad de monatina producida como la distribución estereoisomérica de la monatina producida. El DTT (como en las muestras a continuación) pareció aumentar la cantidad de monatina total producida.

Se realizaron los mismos experimentos que anteriormente utilizando L- triptófano como el sustrato de inicio y combinando las aldolasas (proporcionadas como extracto celular) con L- aminotransferasa HexAspC de amplia especificidad de parcialmente purificada (0,5 mg de HexAspC). Los resultados medios (de los duplicados) se reflejan a continuación en la Tabla 14 para la producción total de monatina (restando los niveles básicos sin la presencia de aldolasa) y el % de monatina S,S se refleja en base al área de pico de LC de fase inversa. Las cifras superiores a 400 ppm están fuera de la gama lineal de la curva estándar y son aproximados. El polipéptido purificado con actividad de aldolasa de SEQ ID NO: 28 se utilizó como referencia. Los polipéptidos con actividad de aldolasa de SEQ ID NO: 34 y la SEQ ID NO: 300 (derivado de planta) se utilizaron con y sin 2 mM de DTT. Shannon y Marcus (The Journal of Biological Chemistry 237; 3342-3347, 1962) utilizaron mercaptoetanol como un agente reductor en la purificación original de aldolasa de HMG de maní.

**Tabla 14: Monatina total producida a partir de L-triptófano y % de S,S**

<b>Aldolasa (punto de tiempo)</b>	<b>Monatina total (ppm)</b>	<b>% de monatina S,S</b>
SEQ ID NO: 116 (2 horas)	129	47,9
SEQ ID NO: 116 (21 horas)	207	56,4
SEQ ID NO: 76 (2 horas)	949	90,6
SEQ ID NO: 76 (21 horas)	1181	89,0
SEQ ID NO: 44 (2 horas)	128	55,0
SEQ ID NO: 44 (21 horas)	237	61,7
SEQ ID NO: 148 (2 horas)	199	71,5
SEQ ID NO: 148 (21 horas)	358	74,4
SEQ ID NO: 46 (2 horas)	346	79,3
SEQ ID NO: 46 (21 horas)	757	83,3
SEQ ID NO: 134 (2 horas)	215	69,2
SEQ ID NO: 134 (21 horas)	370	74,1
SEQ ID NO: 74 (2 horas)	75	51,4
SEQ ID NO: 74 (21 horas)	137	58,8



<b>Aldolasa (punto de tiempo)</b>	<b>Monatina total (ppm)</b>	<b>% de monatina S,S</b>
SEQ ID NO: 126 (2 horas)	47	56,7
SEQ ID NO: 126 (21 horas)	113	56,7
SEQ ID NO: 102 (2 horas)	igual que control	n/d
SEQ ID NO: 102 (21 horas)	11,5	61,1
SEQ ID NO: 58 (2 horas)	113	71,9
SEQ ID NO: 58 (21 horas)	351	75,5
SEQ ID NO: 88 (2 horas)	852	90,1
SEQ ID NO: 88 (21 horas)	1352	88,8
SEQ ID NO: 50 (2 horas)	62	30,8
SEQ ID NO: 50 (21 horas)	145	38,6
SEQ ID NO: 106 (2 horas)	3,5	31,0
SEQ ID NO: 106 (21 horas)	45	34,4
SEQ ID NO: 304 (2 horas)	igual que control	n/d
SEQ ID NO: 304+ DTT (2 horas)	1	n/d
SEQ ID NO: 304 (21 horas)	igual que control	n/d
SEQ ID NO: 304 + DTT (21 horas)	10	n/d
SEQ ID NO: 300 (2 horas)	73	75,2
SEQ ID NO:300 + DTT (2 horas)	121	83
SEQ ID NO: 300 (21 horas)	91	63,6
SEQ ID NO: 300 + DTT (21 horas)	197	71,6
SEQ ID NO: 28 (2 horas)	55	35,1
SEQ ID NO: 28 (21 horas)	87	40,4

### Ejemplo de referencia 11

Efecto del ditiotreitól (DTT) en la producción de monatina

- 5 Varias de las enzimas en el Ejemplo 10 se eligieron para mayor estudio. Las aldolasas derivadas de planta mostraron una mejora con la adición de DTT como agente reductor. Se notó que las aldolasas derivadas de microbios a partir de muestras ambientales también contienen un porcentaje alto de residuos de cisteína. Por lo tanto, se realizaron otros experimentos para ver si el DTT también aumentó la producción de monatina para aldolasas no de planta.
- 10 Se incorporaron por cada 1 mL de mezcla de reacción: aproximadamente 50 µg de aldolasa (proporcionada en extractos celulares salvo que se indique lo contrario), 4 mM de MgCl<sub>2</sub>, 50 mM de D-triptófano, 2 mg de AT-103, 200 mM de piruvato de sodio, 100 mM de tampón de fosfato de potasio con un pH de 7,5 y 0,05 mM de PLP. Se agregó ditiotreitól (concentración definitiva de 2mM) a las muestras indicadas a continuación. Los experimentos se desarrollaron por duplicado. Las muestras se incubaron durante 2 horas a 30°C con una leve agitación. Los resultados medios se muestran a continuación respecto de la monatina total como se determina por LC/MS/MS, con la producción básica de monatina (sin control de aldolasa) restada.
- 15

**Tabla 15: monatina total producida a partir de D-triptófano**

<b>Aldolasa</b>	<b>Monatina total (ppm)</b>
SEQ ID NO: 116	126
SEQ ID NO: 116 + DTT	102
SEQ ID NO: 44	107
SEQ ID NO: 44 + DTT	103
SEQ ID NO: 46	88
SEQ ID NO: 46 + DTT	161
SEQ ID NO: 58	118
SEQ ID NO: 58 + DTT	141
SEQ ID NO: 50	243
SEQ ID NO: 50 + DTT	170
SEQ ID NO: 28 proteína purificada	174
SEQ ID NO: 28 + DTT proteína purificada	196

El control no aldolasa produjo 10 ppm de monatina total con y sin DTT, indicando que el DTT no afecta la reacción general mediante la reducción de subproductos y no afecta la actividad de D-aminotransferasa. Los polipéptidos con actividad de aldolasa de SEQ ID NO: 46, SEQ ID NO: 58, SEQ ID NO: 28 demostraron un beneficio a partir de la incorporación de DTT. El polipéptido con actividad de aldolasa de SEQ ID NO: 46 mostró el mayor beneficio, aproximadamente una actividad 1,8 veces más alta con 2 mM de DTT. Dos polipéptidos con actividad de aldolasa parecen haber sido inhibidos por el DTT (SEQ ID NO: 116 y SEQ ID NO: 50) a la vez que no se vio efecto alguno dentro del error experimental para el polipéptido con actividad de aldolasa de SEQ ID NO: 44. Sin embargo, es posible que para cada aldolasa utilizada exista una concentración óptima de DTT para detectar un beneficio de proporcionar el agente reductor.

El polipéptido con actividad de aldolasa de SEQ ID NO: 88 fue elegido para estudiar el efecto de la concentración de DTT en la producción de monatina. Se realizaron reacciones combinadas como anteriormente. Los resultados aparecen en la Figura 15. La concentración óptima de DTT en este ensayo osciló entre 2,5 y 5 mM para la cantidad de aldolasa incorporada. Lo interesante es que si no incluyó DTT, la cantidad de monatina producida fue casi tan alta como los 2,5 mM de DTT, pero incorporar cantidades subóptimas de DTT (0,5-1mM) realmente parece generar un efecto inhibitor así como también la incorporación de una alta cantidad de DTT (20mM).

### **Ejemplo 12**

Comparación de la producción total de monatina y la distribución isomérica para los polipéptidos con actividad de aldolasa de SEQ ID NO: 278, SEQ ID NO: 162, SEQ ID NO: 276, SEQ ID NO: 178, SEQ ID NO: 202, SEQ ID NO: 166, SEQ ID NO: 218, SEQ ID NO: 224, SEQ ID NO: 226, SEQ ID NO: 244, SEQ ID NO: 250, SEQ ID NO: 252, SEQ ID NO: 264, SEQ ID NO: 268, SEQ ID NO: 272, SEQ ID NO: 184, SEQ ID NO: 282, SEQ ID NO: 186, SEQ ID NO: 192, SEQ ID NO: 200, SEQ ID NO: 280, SEQ ID NO: 284, SEQ ID NO: 172, SEQ ID NO: 180, SEQ ID NO: 168, SEQ ID NO: 228, SEQ ID NO: 236, SEQ ID NO: 238, SEQ ID NO: 240, SEQ ID NO: 270, SEQ ID NO: 156 y SEQ ID NO: 28.

La enzima recombinante producida en el Ejemplo 7 se utilizó en reacciones combinadas con aldolasas de HMG para producir monatina a partir de D-triptófano y piruvato como se describe en la solicitud de patente de los Estados Unidos publicada No. 20050282260. El polipéptido con actividad de aldolasa de SEQ ID NO: 28 se utilizó como referencia en estos ensayos y había sido purificada como se describe en el Ejemplo 8.

Para producir cantidades de muestra de cada aldolasa, se desarrollaron cultivos de 25 mL en un medio LB que contenía ampicilina (100 µg/mL), a una DO<sub>600</sub> de aproximadamente 0,5. Los cultivos se indujeron con 1 mM de IPTG. Las células se cambiaron a 30°C y se desarrollaron durante la noche. Los extractos celulares se prepararon utilizando un reactivo Bugbuster conforme a los protocolos del fabricante (Novagen, Madison, WI). También se

incorporó benzonucleasa. Las proteínas solubles en los extractos celulares se separaron en una Estación de Electroforesis Automática - Experion de Bio-Rad Laboratories (Bio-Rad, Hércules, CA) y se analizaron para verificar la concentración y el porcentaje utilizando la versión de software de Experion 1.1.98.0.

- 5 Se incorporaron por cada 1 mL de mezcla de reacción: aproximadamente 200 µg de los polipéptidos con actividad de aldolasa de SEQ ID NO: 278, SEQ ID NO: 162, SEQ ID NO: 276, SEQ ID NO: 178, SEQ ID NO: 202, SEQ ID NO: SEQ ID NO: 166, SEQ ID NO: 218, SEQ ID NO: 244, SEQ ID NO: 250, SEQ ID NO: 252, SEQ ID NO: 264, SEQ ID NO: 268, SEQ ID NO: 272, SEQ ID NO: 184, SEQ ID NO: 282, SEQ ID NO: 186, SEQ ID NO: 192, y SEQ ID NO: 200 o 50 µg del polipéptido con actividad de aldolasa de SEQ ID NO: 280, SEQ ID NO: 284, SEQ ID NO: 172, SEQ ID NO: 180, SEQ ID NO: 168, SEQ ID NO: 228, SEQ ID NO: 236, SEQ ID NO: 238, SEQ ID NO: 240, SEQ ID NO: 10 270, y SEQ ID NO: 156 (proporcionado en extractos celulares salvo que se indique lo contrario), 4 mM de MgCl<sub>2</sub>, 50 mM de D-triptófano, 0,5 mg de D-aminotransferasa de *B. sphaericus* purificada, 200 mM de piruvato de sodio, 100 mM de tampón de fosfato de potasio con un pH de 7,5 y 0,05 mM de PLP. Los experimentos se desarrollaron por duplicado. Las muestras se incubaron durante 2 horas y durante la noche (20 horas) a 30°C con una leve agitación. Los resultados medios aparecen a continuación. Los únicos estereoisómeros detectados cuando se produce 15 monatina utilizando estos procedimientos son los estereoisómeros R,R y S,R. El porcentaje de R,R se enumera a continuación y se determinó mediante el área de pico de LC de fase inversa.

**Tabla 16: monatina total producida a partir de D-triptófano y % de R,R**

Aldolasa (punto de tiempo)	Monatina total (ppm)	% de monatina R,R
SEQ ID NO: 278 (1 hora)	11,35	100
SEQ ID NO: 278 (18 horas)	282,15	96
SEQ ID NO: 162 (1 hora)	19,35	100
SEQ ID NO: 162 (18 horas)	277,9	98
SEQ ID NO: 276 (1 hora)	27,2	100
SEQ ID NO: 276 (18 horas)	421	98
SEQ ID NO: 178 (1 hora)	24,8	98
SEQ ID NO: 178 (18 horas)	394,25	94
SEQ ID NO: 202 (1 hora)	0	0
SEQ ID NO: 202 (18 horas)	19,2	91
SEQ ID NO: 166 (1 hora)	42,8	89
SEQ ID NO: 166 (18 horas)	601,25	71
SEQ ID NO: 218 (1 hora)	15,6	99
SEQ ID NO: 218 (18 horas)	456,05	96
SEQ ID NO: 224 (1 hora)	19,7	98
SEQ ID NO: 224 (18 horas)	406,55	93
SEQ ID NO: 226 (1 hora)	41,3	95
SEQ ID NO: 226 (18 horas)	460,15	84
SEQ ID NO: 244 (1 hora)	11,6	99
SEQ ID NO: 244 (18 horas)	168,3	98
SEQ ID NO: 250 (1 hora)	20,25	95

ES 2 383 767 T3

<b>Aldolasa (punto de tiempo)</b>	<b>Monatina total (ppm)</b>	<b>% de monatina R,R</b>
SEQ ID NO: 250 (18 horas)	289,25	89
SEQ ID NO: 252 (1 hora)	48,4	81
SEQ ID NO: 252 (18 horas)	335,8	73
SEQ ID NO: 264 (1 hora)	31,65	82
SEQ ID NO: 264 (18 horas)	252,35	77
SEQ ID NO: 268 (1 hora)	12,95	98
SEQ ID NO: 268 (18 horas)	252,55	95
SEQ ID NO: 272 (1 hora)	13,8	98
SEQ ID NO: 272 (18 horas)	165,8	98
SEQ ID NO: 184 (1 hora)	19,55	96
SEQ ID NO: 184 (18 horas)	221,85	94
SEQ ID NO: 282 ( 1 hora)	29,75	95
SEQ ID NO: 282 (18 horas)	399,05	91
SEQ ID NO: 186 (1 hora)	14,4	94
SEQ ID NO: 186 (18 horas)	116,15	93
SEQ ID NO: 192 (1 hora)	17,1	97
SEQ ID NO: 192 (18 horas)	131,25	97
SEQ ID NO: 200 (1 hora)	32,1	97
SEQ ID NO: 200 (18 horas)	331,05	94
SEQ ID NO: 28 (1 hora) (200 µg)	32,1	100
SEQ ID NO: 28 (18 horas) (200µg)	111,45	99
SEQ ID NO: 280 (1 hora)	0	n/d
SEQ ID NO: 280 (18 horas)	3,25	61
SEQ ID NO: 284 (1 hora)	2,3	100
SEQ ID NO: 284 (18 horas=	55,35	98
SEQ ID NO: 172 (1 hora)	12,75	99
SEQ ID NO: 172 (18 horas)	205,9	96
SEQ ID NO: 180 (1 hora)	38,7	93
SEQ ID NO: 180 (18 horas)	310,9	75
SEQ ID NO: 168 (1 hora)	28	98
SEQ ID NO: 168 (18 horas)	301,1	90

<b>Aldolasa (punto de tiempo)</b>	<b>Monatina total (ppm)</b>	<b>% de monatina R,R</b>
SEQ ID NO: 228 (1 hora)	39,2	99
SEQ ID NO: 228 (18 horas)	367	95
SEQ ID NO: 236 (1 hora)	14,85	96
SEQ ID NO: 236 (18 horas)	250,05	90
SEQ ID NO: 238 (1 hora)	30,05	97
SEQ ID NO: 238 (18 horas)	466,15	90
SEQ ID NO: 240 (1 hora)	2,65	100
SEQ ID NO: 240 (18 horas)	51,55	96
SEQ ID NO: 270 (1 hora)	12,2	91
SEQ ID NO: 270 (18 horas)	214,3	83
SEQ ID NO: 156 (1 hora)	62,5	88
SEQ ID NO: 156 (18 horas)	623,9	71
SEQ ID NO: 28 (1 hora) (50 µg)	31,3	98
SEQ ID NO: 28 (18 horas) (50 µg)	444,25	97

Los números de la producción total de monatina fueron desde indetectable a más de 600 ppm y el % de R,R oscilaron entre el 61% y el 100%. Dado que la aminotransferasa fue la misma para todas las aldolasas, el cambio de aldolasa puede tener un impacto significativo tanto en la cantidad de monatina producida como en la distribución estereoisomérica de la monatina producida.

Se realizaron los mismos experimentos que anteriormente utilizando L-triptófano como el sustrato inicial y combinando las aldolasas (proporcionadas como extractos celulares) con L-aminotransferasa HexAspC de especificidad amplia producida y purificada como se describe en la solicitud de patente de los Estados Unidos publicada No. 20050282260 (0,5 mg de proteína purificada). Los resultados se reflejan en la Tabla 17 para la producción total de monatina (restando los niveles básicos sin presencia de aldolasa) y el porcentaje de monatina S,S se refleja en base al área de pico de LC de fase inversa. Las cifras superiores a 400 ppm están fuera de la gama lineal de la curva estándar y son aproximadas. La Tabla 12 muestra los resultados para la enzima R-específica de referencia, el polipéptido con actividad de aldolasa de SEQ ID NO: 28 que se ensayó al mismo tiempo.

**Tabla 17: Monatina total producida a partir de L-triptófano y % de S,S**

<b>Aldolasa (punto de tiempo)</b>	<b>Monatina total (ppm)</b>	<b>% de S,S monatina (área de pico de LC de fase inversa)</b>	<b>% de monatina S,S (derivatización de FDAA)</b>
SEQ ID NO: 278 (1 hora)	14,6	21,0	n/d
SEQ ID NO: 278 (durante la noche)	7905	24,6	n/d
SEQ ID NO: 162 (1 hora)	14	15,4	n/d
SEQ ID NO: 162(durante la noche)	105,6	17,3	n/d
SEQ ID NO: 276 (1 hora)	35,8	10,2	n/d

Aldolasa (punto de tiempo)	Monatina total (ppm)	% de S,S monatina (área de pico de LC de fase inversa)	% de monatina S,S (derivatización FDAA)
SEQ ID NO: 276 (durante la noche)	67,8	9	15,1
SEQ ID NO: 218 (1 hora)	11,7	20,3	n/d
SEQ ID NO: 218(durante la noche)	49,9	17,7	22,0
SEQ ID NO: 244 (1 hora)	6,2	18,0	n/d
SEQ ID NO: 244 (durante la noche)	24,4	14,7	19,6
SEQ ID NO: 268 (1 hora)	6,3	24,1	n/d
SEQ ID NO: 268 (durante la noche)	61,2	23,3	29,2
SEQ ID NO: 272 (1 hora)	5,7	20,5	n/d
SEQ ID NO: 272 (durante la noche)	43,6	19,9	22,9
SEQ ID NO: 192 (1 hora)	6,8	19,0	n/d
SEQ ID NO: 192(durante la noche)	56,4	20,0	24,4
SEQ ID NO: 172 (1 hora)	29,9	35,6	n/d
SEQ ID NO: 172(durante la noche)	184,2	42,4	45,6
SEQ ID NO: 228 (1 hora)	59,6	23,9	n/d
SEQ ID NO: 228 (durante la noche)	182	35,6	38,0

### Ejemplo 13

Producción de monatina a partir de indol-3-piruvato utilizando una D-aminotransferasa

5 La transaminasa AT-103 fue parte de una biblioteca de transaminasas adquirida de BioCatalytics (Pasadena, CA) y la enzima se probó para la producción de monatina en reacciones combinadas utilizando la ProA aldolasa de *C. testosteroni*. La aldolasa se preparó como se describe en la patente No. WO 03/091396 A2. La AT-103 es una D-transaminasa de amplia especificidad (E.C. 2.6.1.21) de la especie *Bacillus* que requiere un D-aminoácido (como un D-glutamato, un D-aspartato o D-alanina) como el donante de aminoácido. Las enzimas y los componentes/sustratos adicionales se incorporaron directamente al tampón de reacción proporcionado en el kit que contenía 100 mM de tampón de fosfato de potasio con un pH de 7,5, 100 mM de donante amino y 0,1 mM de PLP. A un mL de tampón de reacción se agregaron: 4 mg de indol-3-piruvato, 20 mg de piruvato, aproximadamente 50 µg de ProA proporcionada en un extracto celular, 1 µL 2 M de MgCl<sub>2</sub>, y 2 mg de enzima aminotransferasa. Las reacciones se realizaron por duplicado. Las reacciones se incubaron durante la noche a 30°C con una leve agitación (100 rpm). Las muestras se filtraron y sometieron a un análisis de LC/MS/MS de fase inversa como se describe en el Ejemplo 1. Los resultados indicaron que aproximadamente 370 µg/mL de monatina se produjeron utilizando la enzima AT-103. Los resultados fueron posteriormente analizados para determinar las proporciones de S,R/R,S respecto de la R,R/monatina S,S en base a las áreas de pico de los dos grupos de estereoisómeros que se resuelven durante la separación cromatográfica. De la monatina total producida por AT-103, el 69% era R,R/monatina S,S en relación con los isómeros mezclados. Esta enzima es homóloga a la enzima DAT de *Bacillus subtilis* descrita en la patente No. WO 03/091396 A2 que se conoce por tener una amplia especificidad para D-aminoácidos. El análisis quiral se realizó utilizando la metodología FDAA descrita en el Ejemplo 1, que verificó que la D-aminotransferasa estaba produciendo principalmente monatina R,R y alguna monatina S,R como se esperaba. Otros experimentos de transaminación con

monatina S,S o monatina R,R y  $\alpha$ -cetoglutarato como sustratos verificaron que la enzima de BioCatalytics era altamente selectiva para la configuración-D del carbono 4 como se esperaba. En estos experimentos, no se detectó glutamato en la reacción con monatina S,S y  $\alpha$ -cetoglutarato como sustratos.

- 5 Para disminuir la cantidad de monatina S,S o monatina R,S producida como subproductos en reacciones combinadas con AT-103 (la D-transaminasa de amplia gama) y la ProA aldolasa, la aldolasa se purificó utilizando cartuchos His-Bind conforme a los protocolos del fabricante (Novagen, Madison, WI.). La enzima purificada preferentemente no debería contener actividades de L-aminotransferasa salvajes que pueden estar presentes en extractos celulares (como las actividades TyrB o AspC de *E. coli* nativas). Este eluyente His-Bind se desaló para eliminar imidazol utilizando columnas PD-10 (G25 Sephadex, GE Healthcare, Piscataway, NJ) y se eluyó en 50 mM de Tris-Cl con un pH de 7. Los experimentos se realizaron por duplicado en un volumen de 1 mL y contenían 100 mM de tampón Tris-Cl, pH de 7,8, 50  $\mu$ g de ProA aldolasa, 4 mg de indol-3-piruvato, 1 o 2 mg de D-aminotransferasa, 200 mM de piruvato de sodio, 2 mM de  $MgCl_2$ , 3 mM de fosfato de potasio, 0,1 mM de PLP y 14,7 mg de D-glutamato. Los tubos se incubaron a 30°C con una leve agitación. Se tomaron puntos de tiempo de dos horas y se congelaron inmediatamente a -20°C. El pH se ajustó a las dos horas de 5 hasta entre 7 y 8 utilizando NaOH y los ensayos se incubaron durante la noche. Las muestras se filtraron y analizaron para verificar la monatina como se describe en el Ejemplo 1. Las muestras de dos horas no mostraron cantidades detectables de monatina, probablemente a raíz del pH bajo. Las muestras de 24 horas contenían cerca de 190 ng/mL de monatina cuando se utilizó un 1 mg de D-aminotransferasa y cerca del 84% era monatina R,R y el 16% era monatina S,R. Cuando se utilizaron 2 mg de D-aminotransferasa, se produjeron 540 ng/mL de monatina, cerca del 71% era monatina R,R.
- 10
- 15
- 20 Se realizaron experimentos similares utilizando un tampón de aminotransferasa de BioCatalytics (BioCatalytics, Pasadena, CA), que contenía 100 mM de fosfato de potasio con un pH de 7,5, 0,1 mM de PLP y 100 mM de D-glutamato. Se agregaron indol-3-piruvato sólido y D-aminotransferasa como se hizo anteriormente. Se agregó ProA aldolasa (50  $\mu$ g),  $MgCl_2$ , y 50 mM de piruvato de las soluciones concentradas. Los ensayos se trataron como se hizo anteriormente, a pesar de que no se necesitó un ajuste de pH en este caso. Se realizó un control negativo con la enzima y el tampón proporcionados por BioCatalytics, que no contenía monatina. Los resultados experimentales se reflejan en la Tabla 18.
- 25

**Tabla 18: Producción de monatina a partir de indol-3-piruvato en tampón de fosfato**

Mg D-aminotransferasa	Tiempo (horas)	Monatina total (ng/mL)	% de R,R
0	2	0	n/d
1	2	6780	No determinado
2	2	13170	55%
0	16	0	n/d
1	16	15000	No determinado
2	16	28930	51%

La producción de monatina en un tampón de fosfato es claramente más elevada que en sistemas tamponados Tris.

- 30 Para comparar las actividades del *B. subtilis* DAT clonado de la patente WO 03/091396 A2 con la enzima de BioCatalytics (AT-103) (BioCatalytics, Pasadena, CA), se realizaron ensayos adicionales. El gen dat de *B. subtilis* también se subclonó en pET30 para eliminar la marca de Histidina 6. Las enzimas marcadas y no marcadas se produjeron en BL21 (DE3) como se describe en la patente WO 03/091396 A2. Se realizaron extractos celulares y los ensayos de proteína total se desarrollaron para estimar la concentración de proteína como se describió anteriormente. Se realizaron reacciones duplicadas de un mL que contenían: 500  $\mu$ g de D-aminotransferasa, 50  $\mu$ g de ProA aldolasa, 100 mM de fosfato de potasio con un pH de 7,5, 3 mM de  $MgCl_2$ , 4 mg de indol-3-piruvato, 200 mM de piruvato de sodio, 7,35 mg (50 mM) de D-glutamato y 0,1 mM de PLP. Las muestras se incubaron a 30°C durante 1, 2 y 24 horas y se filtraron para el análisis por LC/MS/MS. Las muestras solo contenían estereoisómeros S,R y R,R de monatina como determina el protocolo de derivatización FDAA descrito en el Ejemplo 1. Los resultados están resumidos en la Tabla 19 a continuación. El % de R,R se determinó por áreas de pico que se separaron mediante cromatografía de fase inversa.
- 35
- 40

Tabla 19: comparación de enzimas D-aminotransferasa

Enzima	Tiempo (horas)	Monatina (ppb)	% de monatina R,R
B. sub DAT-HIS	1	512	no determinado
B. sub DAT no marcado	1	1056	no determinado
AT-103 de BioCatalytics	1	2353	no determinado
B. sub DAT-HIS	2	894	80-90%
B. sub DAT no marcado	2	1913	□80%
AT-103 de BioCatalytics	2	6887	92,5%
B. sub DAT-HIS	16	3014	31
B. sub DAT no marcado	16	5612	33
AT-103 de BioCatalytics	16	16131	66

La eliminación de la marca de Histidina 6 parece haber mejorado la actividad de la D-aminotransferasa de *B. subtilis*. Sin embargo, el homólogo de la D-aminotransferasa de BioCatalytics claramente tuvo la actividad más alta. El homólogo de la D-aminotransferasa de BioCatalytics también demostró una mayor especificidad de sustrato para el precursor de R-monatina. Los mayores plazos de incubación parecen reducir el exceso enantiomérico de la R,R, monatina que se produce.

Dado que las enzimas de *Bacillus* D-aminotransferasa tienen una preferencia por el piruvato como un aceptor amino y por D-alanina como un donante amino, se esperaba que la D-alanina pudiera utilizarse como el donante amino para la conversión de MP en monatina con resultados mejores o similares. Se realizaron reacciones duplicadas de un mL que contenían: 500 µg de D-aminotransferasa, 50 µg de ProA aldolasa purificada, 100 mM de fosfato de potasio con un pH de 7,5, 3 mM de MgCl<sub>2</sub>, 4 mg de indol-3-piruvato, 100 mM de piruvato de sodio, 25 mM de D-glutamato o D-alanina y 0,1 mM de PLP. Las muestras se incubaron durante 2 horas y se trataron como se hizo anteriormente previo al análisis. Cuando se utilizó D-alanina como el donante amino, se produjeron niveles de monatina apenas más altos de lo esperado (23 frente a 21 ppm). Asimismo, se espera que las concentraciones altas de piruvato puedan inhibir la etapa de transaminación, por lo cual administrar dosis de piruvato en cantidades más pequeñas con el tiempo puede mejorar la proporción general de la producción de monatina. Se puede ver a partir de los datos anteriores que a pesar de que se utilizó la mitad del piruvato en este caso, en relación con la tabla anterior, se produjo una cantidad significativamente mayor de monatina. AT-103 es un ejemplo de una enzima con actividad limitada para S-MP. A pesar de que la ProA aldolasa utilizada en este estudio genera más que 90-95% de S-MP, AT-103 genera hasta 92% de monatina R,R.

#### Ejemplo 14

Producción de monatina R,R a partir de D-triptófano

Se incorporaron por cada 1 mL de mezcla de reacción: aproximadamente 60 µg de ProA aldolasa de *C. testosteroni* (proporcionada en extractos celulares como se describe en la patente WO 03/091396), 4 mM de MgCl<sub>2</sub>, 50 mM de D-triptófano, 0,5 mg de D-aminotransferasa (AT-103) de BioCatalytics (BioCatalytics, Pasadena, CA), 100 mM de piruvato de sodio, 100mM de tampón de fosfato de potasio con un pH de 7,5 o 100 mM de tampón de acetato de sodio con un pH de 8, 0,05 mM de PLP, 3 mM de fosfato de potasio (solamente para las reacciones de acetato) y 10 mM de α-cetoglutarato. Los experimentos se desarrollaron por duplicado con controles negativos en donde no se agregó aldolasa. Las muestras se incubaron durante la noche (20 horas) a 30°C con una leve agitación. El pH real de las muestras de acetato de sodio fue de aproximadamente 5, mientras que el pH definitivo para las muestras tamponadas de fosfato fue de aproximadamente 7. Ninguna de las aldolasas parecía tener una actividad significativa con un pH de 5; la muestra que contenía ProA aldolasa estaba apenas por encima del control negativo pero probablemente no estaba por encima del error experimental. En el fosfato de potasio, la ProA aldolasa produjo 73,4 ppm de monatina con una proporción R,R:S,R de 1,7:1 (~63% de R,R a partir de D-triptófano).

Dado que las enzimas de *Bacillus* D-aminotransferasa tienen una preferencia por el piruvato como un aceptor amino y por D-alanina como un donante amino, se esperaba que la incorporación de alfa-cetoglutarato no fuera necesaria cuando se producía R,R o monatina S,R a partir de D-triptófano. El experimento anterior se repitió (en 100 mM de tampón de fosfato de potasio) utilizando ProA aldolasa purificada (50-60 µg) y un tiempo de incubación de 2,5 horas.



Se desarrollaron experimentos por duplicado con y sin alfa-cetoglutarato. Con la incorporación de 10 mM de alfa-cetoglutarato, se formaron 56,1 ppm de monatina a partir de D-triptófano (79,5% de R,R, 20,5% de S,R). Sin alfa-cetoglutarato, se formaron 102,5 ppm de monatina (79% de R,R, 21% de S,R).

### Ejemplo 15

#### 5 Triptófano racemasa

La monatina R,R ha sido producida utilizando D-aminotransferasa y una aldolasa cuando se utilizó D-triptófano como el material inicial (Ejemplo 14). Sin perjuicio de lo expuesto, el L-triptófano puede ser un material inicial preferido por varias razones. Por ejemplo, el L-triptófano puede ser más económico y estar más disponible que el D-triptófano. La presente solicitud describe varios procedimientos para obtener un triptófano racemasa activo. Los rendimientos de monatina R,R se mejoran utilizando una aldolasa R-específica, es decir, una aldolasa que en forma preferente o selectiva produce R-MP. La Figura 3 ilustra un procedimiento para producir monatina R,R enriquecida con estereoisómeros a partir de L-triptófano utilizando triptófano racemasa, una D-aminotransferasa y una aldolasa R-específica.

Una selección para un triptófano racemasa se crea construyendo una cepa que requeriría una racemasa activa para el crecimiento. Un triptófano auxótrofo necesita una fuente de L-triptófano cuando crece en un medio mínimo. Complementar el medio con D-triptófano es una forma de seleccionar para una racemasa que convierte D-triptófano en L-triptófano. Los triptófanos auxótrofos se probaron para el crecimiento en un medio mínimo complementado con D-triptófano. Las cepas, CAG18455 y CAG18579 del centro Coli Genetic Stock Center y NRRL12264 (también lipA-,  $\lambda$ DE3 lisogenizado y curado de su plásmido) no crecieron cuando se complementaron con D-triptófano pero crecieron cuando se las complementó con L-triptófano. La *E. coli* puede utilizarse como un organismo huésped pero también pueden utilizarse otros organismos huésped como levadura, otra bacteria u otros organismos eucariotas. Un triptófano auxótrofo (especialmente NRRL12264 (también llamado lipA,  $\lambda$ DE3 lisogenizado y curado de su plásmido)) crecerá en D-triptófano cuando se haya transformado con una D-aminotransferasa. Esto confirma la capacidad de la *E. coli* de transportar D-triptófano en la célula.

Saicher y Lingens describieron la presencia de un triptófano racemasa en *Pseudomonas aeruginosa* (ATCC 15926). La triptófano racemasa también ha sido descrito en varias plantas, que incluyen tabaco, remolacha, tomate, y trigo y la enzima parece estar inducida por condiciones de estrés o sequía osmótica. La triptófano racemasa puede desempeñar una función en *Sclerochiton illicifolius* en la vía de producción de la monatina nativa. Para aislar esta actividad de racemasa, se construye una biblioteca de expresión a partir de ATCC15926 (u otro organismo con actividad de triptófano racemasa) y la biblioteca se transforma en el triptófano auxótrofo. Una cepa se selecciona para su cultivo utilizando D-triptófano como la fuente de triptófano. Un procedimiento similar también se utiliza para monitorear muchas cepas con racemasas conocidas para buscar una racemasa con actividad en D-triptófano. Los ejemplos incluyen: alanina, serina y glutamato racemasas (T. Yoshimura y N. Esaki, "Amino Acid Racemases: Functions and Mechanisms". Journal of Bioscience and Bioengineering, Vol. 96, No. 2, 103-109, 2003). La alanina racemasa es PLP dependiente y ha sido clonada de la *Salmonella typhimurium* (gen dadB). Otras fuentes incluyen *Escherichia coli*, *Bacillus subtilis*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Schizosaccharomyces pombe* y *Bacillus cereus*. Un hongo basidiomiceto, *Lentinus edodes*, también contiene una alanina racemasa de amplia actividad. La serina racemasa también es PLP dependiente y se encuentra en eucariotas (como el ADNc del gusano de seda, cerebro de rata, cerebro de ratón) así como en bacterias (*Enterococcus gallinarum*). La glutamato racemasa es PLP independiente y ha sido clonada de *Pediococcus pentosaceus*, *Bacillus pumilus*, *Lactobacillus fermenti*, *Lactobacillus brevis*, *E. coli*, *Aquifex pyrophilus*, y *Bacillus subtilis*. La glutamato racemasa es muy específica y por lo tanto, aún los aminoácidos aspartato, asparagina, y glutamina de estructura similar no son sustratos para la enzima. También existen aspartato racemasas y son independientes de PLP. Estas enzimas se encuentran en cepas de *Lactobacilli*, *Streptococcus*, y algunas archaeas como las cepas de *Desulfurococcus* y *Thermococcus*. El molusco bivalvo *Scapharca brouhtonii* también contiene una aspartato racemasa. Otras racemasas encontradas en la biblioteca incluyen aminoácido racemasa (EC 5.1.1.10) de *Anabaena* sp. y *Pseudomonas striata*, prolina racemasa, fenilalanina racemasa multifuncional. Las epimerasas o racemasas relacionadas también se prueban. Las racemasas potenciales se prueban para garantizar que no sean D-triptófano aminotransferasas. Este monitoreo se realiza mediante análisis de secuencia y/o un ensayo de enzimas.

Las enzimas que pasan la prueba como una racemasa se monitorean para verificar la actividad en la monatina como describe el Ejemplo 17. Idealmente, se obtiene una enzima que es muy específica para triptófano y tiene actividad de racemasa escasa o nula en la monatina.

Una triptófano racemasa también puede evolucionar y/o mejorar (mediante mutagénesis o ingeniería recombinante) a partir de una racemasa, una transaminasa o una epimerasa existente. Asimismo, dado que se conocen las estructuras de cristal para las alanina aminotransferasas, estas se pueden utilizar como base para una mutagénesis racional basada en la estructura. El proceso descrito anteriormente se utiliza como una selección inicial para la actividad de triptófano racemasa y como una pantalla para una mejor actividad.

Bibliotecas de triptófano racemasa

Construcción de bibliotecas:

- Se obtuvieron *Burkholderia pyrrocina* (ATCC15958) y *Pseudomonas chlororaphis* (ATCC15926) de la American Type Culture Collection. Se desarrollaron como sugiere la ATCC y se preparó el ADN genómico conforme al procedimiento de Mekalanos JJ. (Duplicación y amplificación de genes tóxicos en *Vibrio cholerae*. Cell. 1983. 35:253-63). El ADN genómico fue parcialmente digerido con la enzima de restricción *Sau3AI*. Los fragmentos 1-3 Kbp se purificaron con gel utilizando un Kit de Extracción de Gel QIAquick de Qiagen (Qiagen, Valencia, CA). El ADN purificado se ligó en pTrc99a (Amersham, Piscataway, NJ) que había sido digerido con *Bam*HI y purificado como se indicó anteriormente. La ligadura se realizó a temperatura ambiente con una incubación de 24 horas utilizando una proporción molar inserto-vector de 3:1. La biblioteca ligada se transformó en células químicamente competentes TOP10F (Invitrogen, Carlsbad, CA) y se colocó en una placa en un medio LB con 100 µg/ml de ampicilina. Después de la incubación durante 24 horas de las placas de transformación, las colonias se sacaron de las placas, se lavaron con el medio LB líquido y el sedimento celular de tamaño adecuado se preparó utilizando un kit miniprep QIAquick de Qiagen (Qiagen, Valencia, CA). Se reunieron y prepararon cerca de 30.000 colonias.
- El plásmido reunido se transformó en CAG18455 (*trpC83:Tn10, rph-1*) o CAG18579 (*trpC:Tn10kan, rph-1*). Ambas cepas son triptófano auxótrofos por lo que no crecerán en un medio mínimo M9 (Difco) salvo que el medio se vea complementado con triptófano. Los transformantes se colocaron en placas en un medio mínimo M9 complementado con D-triptófano. Esto selecciona para una cepa que puede convertir D-triptófano en L-triptófano.

- Antes de la transformación de la biblioteca, las cepas se probaron para verificar su crecimiento en un medio mínimo con L-triptófano o D-triptófano. Las cepas se probaron para verificar su crecimiento en un medio mínimo complementado con D-triptófano y no se observó crecimiento alguno. Ambas cepas crecieron en medios idénticos complementados con L-triptófano en lugar de D-triptófano. Asimismo, un derivado de NRRL12264 (la cepa utilizada había sido curada del plásmido de operón triptófano, lisogenizado con λDE3, y eliminado para *lipA*, junto con otras mutaciones cromosómicamente codificadas (*serB*,  $\Delta$ *trpED*, *tnaA2*, *aroP*) (esta cepa no pudo crecer en un medio mínimo complementado con D-triptófano pero creció en un medio idéntico complementado con L-triptófano en lugar de D-triptófano) se transformó con una aminotransferasa D-específica de *Bacillus subtilis* (WO 03/091396). La expresión de la D-aminotransferasa fue impulsada por el promotor T7. La cepa transformada pudo crecer en un medio mínimo M9 complementado con D-triptófano.

- Se monitorean las colonias que crecen en el medio D-triptófano. El plásmido se aísla y vuelve a transformarse en una cepa madre (CAG18455 o CAG 18579) para confirmar que el crecimiento en un medio D-triptófano depende del plásmido y no de una mutación del huésped. Se analizan las secuencias nucleotídicas de los plásmidos que complementan el triptófano auxótrofo. Los clones que se considera que tienen un gen de triptófano racemasa son analizados con posterioridad.

- La triptófano racemasa de otras fuentes de tejido se aísla en una forma similar. Existen informes de colección de la actividad de la triptófano racemasa tanto en células en cultivo de tejido de tabaco (*Nicotiana tabacum* L. var. Wisconsin 38) (Miura, G. A. and Mills, S. E. The conversion of D-tryptophan to L-tryptophan in cell cultures of tobacco. Plant Physiol. 1971 47:483-487) como en extractos proteicos crudos de trigo (*Triticum aestivum*) (Rekoslavskaya, N. I., Yur'eva, O. V., Shibanova, L. A., y Salyaev, R. K. Synthesis and physiological function of D-tryptophan during wheat germination. Russian Journal of Plant Physiology 1997. 44:196-203) Una biblioteca de expresión de ADNc se crea a partir de tejido, como se describe en la colección y la biblioteca de expresión se utiliza para transformar un triptófano auxótrofo como se describió anteriormente.

Ensayo de triptófano racemasa

- Los clones que se identifican como aquellos que potencialmente tienen triptófano racemasa se transforman en una cepa de *E. coli* comúnmente utilizada para la expresión de proteínas recombinantes, como BL21. Las células se desarrollan en el caldo LB a una densidad óptica 600 nm de 0,4-0,6. El promotor que impulsa la expresión de la racemasa es inducido con IPTG (0,1 mM de concentración definitiva). Después de la inducción, las células pueden expresar la proteína por 1 a 3 horas a 37°C (con aireación). Las células se cultivan y alisan mediante filtro de émbolo, sonicación o a través de medios químicos (como BugBuster (Novagen, Madison, WI)). Las células lisadas se centrifugan para eliminar los restos celulares. El extracto clarificado se utilizó directamente en los ensayos.
- Se agregan las cantidades variables de extracto en una solución de forma tal que la concentración definitiva es 50 mM de fosfato de potasio (ph 7,0) y 2 mM de L-triptófano. Se agrega piridoxal-5'-fosfato a una concentración definitiva de 10 µM. Las muestras se incuban y posteriormente se analizan por LC/MS. La presencia de un pico de D-triptófano cuando solo se utiliza L-triptófano como un sustrato indica un resultado positivo. La concentración de D-triptófano debería aumentar con el aumento del tiempo hasta que se alcanza un equilibrio y la tasa debería aumentar también con la concentración de proteína hasta que la concentración de enzima es suficientemente alta que ya no se satura con el sustrato. El D-triptófano también puede convertirse en L-triptófano como se indicó anteriormente.

Un gen complementario puede codificarse para una D-aminotransferasa. (Un "gen complementario" es un gen que, cuando se expresa, anula una mutación en un organismo. Por ejemplo, si un organismo tiene mutación nula en uno de los genes necesarios para la síntesis del triptófano por la célula, un gen complementario podría ser uno que, cuando se expresa, permite que la cepa crezca en un medio mínimo (es decir, sin triptófano). Esta reacción requiere un alfa-cetoácido como un  $\alpha$ -cetoglutarato, oxalacetato, o piruvato como un aceptor amino. Estos compuestos probablemente estén presentes en un extracto celular (en cantidades pequeñas). Estos compuestos pueden eliminarse utilizando una columna desaladora PD-10 (GE Healthcare, Piscataway, NJ) y el ensayo aún puede realizarse en extracto crudo. La actividad de triptófano racemasa se purifica utilizando una cromatografía de columna convencional. Finalmente, el marco abierto de lectura identificado como un triptófano racemasa potencial se clona en un vector de expresión con una marca de afinidad. La triptófano racemasa potencial posteriormente se purifica mediante cromatografía por afinidad. En ambos casos, la proteína purificada se utiliza en ensayos de enzima básicamente como se describió anteriormente.

#### Ingeniería genética inversa de triptófano racemasa

La triptófano racemasa puede purificarse de fuentes vegetales o microbianas mediante técnicas de purificación proteica tradicionales que incluyen el fraccionamiento de sulfato de amonio y la cromatografía de columna convencional. Una vez que la proteína se ha purificado de forma tal que un muestreo puede aislarse en un gel 2-D, se utilizan técnicas de microsecuencia de péptidos o de secuencias de aminoácido convencionales del tipo Edman (ver [golgi.harvard.edu/microchem/](http://golgi.harvard.edu/microchem/) para las descripciones de protocolos y el equipo utilizado para este tipo de trabajo). Sin embargo, en algunos casos, la secuencia de genoma del organismo no puede utilizarse como una fuente de la proteína para la purificación de proteínas porque dicha secuencia no se ha determinado aún. En dicha situación, el primer grupo de cebadores degenerados puede diseñarse en base a una secuencia disponible del relativo conocido más cercano de la fuente de proteína. El PCR y el genoma degenerados se desarrollan posteriormente conforme a los protocolos establecidos para aislar la secuencia de codificación de la triptófano racemasa.

#### Producción de monatina triptófano racemasa

Se incorporaron por cada 1 mL de mezcla de reacción: aproximadamente 60  $\mu$ g de ProA aldolasa de *C. testosteroni* (proporcionada en extractos celulares como se indica en la patente WO 03/091396 A2), 100  $\mu$ L/mL de extracto celular de triptófano racemasa o 1 mg/mL de triptófano racemasa purificada, 4 mM de  $MgCl_2$ , 50 mM de L-triptófano, 0,5 mg de D-aminotransferasa de BioCatalytics (AT-103) (BioCatalytics, Pasadena, CA), 100 mM de piruvato de sodio, 100mM de tampón de fosfato de potasio con un pH de 7,5, 0,05 mM de PLP y 10 mM de  $\alpha$ -cetoglutarato. Dado que el piruvato es un aceptor amino aceptable para la D-aminotransferasa de especificidad amplia, el  $\alpha$ -cetoglutarato es opcional. Los experimentos se desarrollan por duplicado con controles negativos en donde no se incorpora aldolasa o triptófano racemasa. Las muestras se incuban por 1 hora o durante la noche (20 horas) a 30°C con una leve agitación.

La triptófano racemasa se prueba para verificar la actividad en la monatina. Un ensayo similar al del Ejemplo 17 se utilizó con monatina como sustrato y se comparó con la actividad de la enzima en el triptófano. La enzima ideal tiene actividad en el triptófano pero tiene actividad escasa o nula en otros aminoácidos, especialmente, glutamato y monatina. Si la enzima tiene una actividad significativa en la monatina, la enzima se puede mutagenizar para disminuir la actividad en la monatina y/o en el glutamato a la vez que mantiene la actividad de triptófano intacta o a un nivel lo suficientemente alto para que la enzima sea útil en la producción de monatina. Las técnicas que se pueden utilizar para la mutagénesis incluyen, a modo no taxativo, PCR propenso a cometer errores, una mutagénesis de sitio dirigido, la modelación para identificar objetivos de la mutagénesis dirigida a sitio (sitios que pueden incluirse en la unión de sustrato), el pasaje a través de cepas mutagénicas y la transposición de ADN.

Las racemasas mutagenizadas pueden monitorearse para verificar la actividad de triptófano utilizando un ensayo sobre placa como se describió anteriormente. Los clones que retienen actividad de triptófano se monitorean para verificar una pérdida de actividad en la monatina.

### **Ejemplo 16**

#### Mutagénesis dirigida a sitio de HEXAspC

##### Resumen experimental

Se encontró que el hexamutante de *E. coli* AspC (HEXaspC) tenía una mejor actividad en comparación con AspC para la producción de monatina S,S como se describe en el ejemplo 6 de la WO 03/091396 A2. El HEX (número de acceso: /AHFA gi:127190) contiene las siguientes mutaciones de AspC (numeración para *E. coli*): V35L, K37Y, T43I, N64L, T104S y N285S. Con base en el análisis estructural y los informes de la bibliografía (S. Rothman y J. Kirsch, *J. Mol. Biol.* (2003), 327, 593-608; S. Rothman et al., *Protein Science* (2004), 13: 763-772), se crearon 5 mutantes adicionales que se esperaba que aumentaran la actividad cinética de los sustratos utilizados en la vía de producción de la monatina: L-triptófano, S-MP o ambos. Dos de los mutantes aumentaron las velocidades de transaminación

para el triptófano y la monatina S,S. Dos de los mutantes mostraron una mayor estereoselectividad para la formación de monatina S,S mientras que uno era menos estereoselectivo. Con base en esto, se espera que la D-aminotransferasa de amplia especificidad de *Bacillus* sp. que tiene mutaciones similares sea útil como la D-aminotransferasa en las vías de monatina R,R mostradas en la figura 3 y descritas en el ejemplo 15. Uno de los mutantes (HEXaspCP9T/R122G) presentó una actividad superior para la transaminación del L-triptófano, pero la actividad en la producción de monatina S,S o en la transaminación de monatina S,S disminuyó significativamente. Por consiguiente, se espera que esta enzima sea útil en la primera etapa de las vías de producción de la monatina R,R mostradas en las figuras 1, 2, 4, 5, 6, 7 y 8 y descritas en los ejemplos 18 y 19. En general, la aminotransferasa que presenta una actividad similar a la de AspC sobre el L-triptófano y una actividad limitada sobre R-MP y S-MP sería útil para los procesos descritos en las figuras 1, 2, 4, 5, 6, 7 y 8.

Procedimientos y materiales

El gen HEX clonado en pUC19 fue proporcionado por el profesor J. F. Kirsch (Departamento de Biología Molecular y Celular, Universidad de California, Berkeley, Berkeley, California 94720-3206) y se utilizó como la plantilla para la clonación del gen en pET23a. Véase James J. Onuffer y Jack F. Kirsch, *Redesign of the substrate specificity of Escherichia coli aspartate aminotransferase to that of Escherichia coli tyrosine aminotransferase by homology modeling and site-directed mutagenesis*, *Protein Science*, 4: 1750-1757 (1995). Véase también NCBI número de acceso 1AHF\_A GI:1127190 (secuencia de aminoácido HEX). Se diseñaron los siguientes cebadores para clonar el gen HEX en el vector pET23a (Novagen, Madison, WI.):

Cebadores HEXaspC:

Terminal N: 5'-GCGGAACATATGTTTGAGAACATTACCGCC-3'; (SEQ ID NO: 393)

Terminal C: 5'-ATAACCGGATCCTTACAGCACTGCCACAATCG-3' (SEQ ID NO: 394)

Se utilizó el siguiente protocolo de PCR para la amplificación de genes: En una reacción de 100 µL, se agregó 50 ng de plantilla de ADN, 1,0 µM de cada cebador, 0,2 mM de cada dNTP, 1 U *Pfu* Turbo Polimerasa (Stratagene, La Jolla, CA), y plantilla *Pfu* clonada una vez. El programa del termociclador utilizó un comienzo en caliente a 94°C durante 5 minutos; seguido de 25 ciclos de una etapa desnaturalización a 94°C (30 segundos), una etapa de recocido a 55°C (1 minuto) y una etapa de extensión a 72°C (2 minutos) y finalmente una etapa de acabado a 72°C (7 minutos). Se utilizaron las técnicas de biología molecular estándar para clonar el producto de la PCR en pET23a mediante el uso de sitios de restricción *Nde*I y *Bam*HI.

Se piensa que el residuo de triptófano en la posición 130 es importante para juntar las interacciones con el anillo de piridoxilo, pero además parece ser una fuente de impedimento estérico con el sustrato del precursor de S-monatina (MP), con base en las observaciones de modelado de proteína. Por consiguiente, se utilizó un aminoácido con una cadena lateral hidrófoba más pequeña (fenilalanina) para reemplazar al triptófano. El resto de las mutaciones se basó en los datos de cinética en la bibliografía, a pesar de que se crearon nuevas combinaciones de mutaciones deseables. Todas las mutaciones a HEXaspC, con la excepción de W130F, se realizaron mediante el uso del kit Multi-Change de Stratagene (Stratagene, La Jolla, CA) de acuerdo con las instrucciones del fabricante. La mutación de W130F se realizó mediante el uso del kit QuikChange de Stratagene (Stratagene, La Jolla, CA) de conformidad con las instrucciones del fabricante con la única excepción de que la temperatura de extensión para la reacción de PCR disminuyó a 66°C. Los cebadores para el kit multi-change (multicambio) se diseñaron mediante el uso de la herramienta de diseño de cebadores multi-kit QuikChange en <www.stratagene.com>, excepto para los cebadores de mutación simple W130F.

Las secuencias de cebadores se enumeran a continuación en la tabla 20.

**Tabla 20**

Cebador	Secuencia (5' a 3')
aspCW130F_hacia atrás	CGCTCTTATGGTTCGGTTTGCTTGGGTTGCTCACCC (SEQ ID NO: 395)
aspCW130F_hacia adelante	GGGTGAGCAACCCAAGCTTTCGAACCATAAGAGCG (SEQ ID NO: 396)
R122G-1 <sup>a</sup>	CAAAAAATACCAGCGTTAAGGGAGTGTGGGTGAGCAACC (SEQ ID NO: 397)

P9T_4 <sup>a</sup>	CATTACCGCCGCTACTGCCGACCCGATTC (SEQ ID NO: 398)
I68V-1 <sup>a</sup>	CACCAAAAATTACCTCGGCGTAGACGGCATCCCTGAATT (SEQ ID NO: 399)
T156A <sup>a</sup>	TGATGCGGAAAAATCACGCTCTTGACTTCGATGCAC (SEQ ID NO: 400)
<sup>a</sup> Denota un cebador que se modificó mediante fosforilación 5'	

#### Expresión de los genes mutantes HEXaspC y análisis de la actividad enzimática

Se inocularon los cultivos líquidos (5 mL) del sistema de autoinducción 2 Novagen Overnight Express™ (Catálogo #71366-3; soluciones 1-6) (Novagen, Madison, WI) de las placas frescas o de las reservas de glicerol congeladas de las siguientes cepas:

*E. coli* BL21(DE3)::HEXaspCpET23a

*E. coli* BL21(DE3)::HEXaspCW130FpET23a

*E. coli* BL21(DE3)::HEXaspCT156ApET23a

*E. coli* BL21(DE3)::HEXaspCP9T/T156ApET23a

10 *E. coli* BL21(DE3)::HEXaspCP9T/R122GpET23a

*E. coli* BL21(DE3)::HEXaspCR122G/T156ApET23a.

Los cultivos se incubaron a 37°C a 230 rpm durante 6 a 8 horas. Se determinó la OD<sub>600</sub> de cada cultivo y se calculó el volumen de cultivo necesario para obtener una OD<sub>600</sub> de 0,03 a 0,05 en 25 mL. Los volúmenes calculados de cada cultivo líquido se transfirieron a matraces que contenían 25 mL del mismo medio. El sistema de autoinducción 2 Overnight Express™ es un medio completo, químicamente definido para la expresión de alto nivel con sistemas de expresión inducidos por IPTG que utilizan la lactosa como agente de inducción y no requieren el control del crecimiento celular. Los cultivos de Overnight Express se incubaron a 30°C con agitación a 230 rpm durante 18 horas. Las células se cosecharon mediante centrifugación y se lavaron una vez con MOPS frío 50 mM, pH 7,0. A continuación, las células se lisaron mediante el uso de Bugbuster™ (libre de amina primaria) Reactivo de extracción (Novagen Catálogo #70923-3) (Novagen, Madison, WI) que contenía 1 µL/mL de benzonasa nucleasa (Novagen Catálogo #70746-3) (Novagen, Madison, WI), 5 µL/mL de Cóctel de inhibidores de proteasa grupo II (Novagen Catálogo #539132) (Novagen, Madison, WI) y 0,33 µL/10 mL r-Lisozima (Novagen Catálogo #71110-3) (Novagen, Madison, WI) de acuerdo con el protocolo recomendado de Novagen. Después de la incubación a 25°C durante 15 minutos con agitación suave, los restos celulares de cada suspensión se sedimentaron mediante centrifugación a 21.000 g durante 15 minutos a 4°C. El sobrenadante se decantó cuidadosamente y se analizó como el extracto libre de células. Las fracciones de cuerpo de inclusión se aislaron mediante la suspensión de las fracciones de restos celulares en un 30% de reactivo de extracción Bugbuster™ (libre de amina primaria), la centrifugación a 21.000×g durante 10 minutos; la suspensión de los gránulos centrifugados en un 10% de reactivo de extracción Bugbuster™ (libre de amina primaria), y la centrifugación nuevamente para aislar los gránulos lavados. Los extractos libres de células y las fracciones de cuerpo de inclusión se analizaron para verificar la expresión de proteínas por SDS-PAGE sobre geles con gradiente del 4 al 15% (BioRad # 161-1104, Hercules, CA). Para las muestras de extracto celular, se cargaron veinte microgramos de proteína soluble en cada ruta de gel (mezcladas previamente con tampón de carga de proteínas 1× y calentadas a 95°C durante 5 minutos). Las fracciones de cuerpo de inclusión se disolvieron en tampón de carga de proteínas 1× (0,2 mL), se calentaron durante 10 minutos a 95°C y se cargaron 5 µL de cada solución por ruta de gel. Se calculó la cantidad de cada mutante HEX en comparación con la proteína soluble total cargada en cada ruta mediante análisis de intensidad de banda con el uso de la herramienta de gel 1D de Labworks Bioluming (UVP, Inc. Upland, CA) y se informa a continuación:

Tabla 21

Muestra	proteína HEXaspC/ total proteína soluble
<i>E. coli</i> BL21(DE3)::HEXaspCP9T/T156ApET23a CFE	0,310
<i>E. coli</i> BL21(DE3)::HEXaspCP9T/R122ApET23a CFE	0,145
<i>E. coli</i> BL21(DE3)::HEXaspCpET23a CFE	0,172
<i>E. coli</i> BL21(DE3)::HEXaspCR122A/T156ApET23a CFE	0,174
<i>E. coli</i> BL21(DE3)::HEXaspCW130FpET23a CFE	0,114
<i>E. coli</i> BL21(DE3)::HEXaspCT156ApET23a CFE	0,120

El análisis de los geles mostró que el mutante HEXaspCR122A/T156A fue la única proteína que se encontró en cantidades sustanciales como cuerpos de inclusión. La proteína HEXaspCP9T/T156A proporcionó el nivel de expresión más alto, aproximadamente un 90% mejor que la proteína HEXaspC. Por el contrario, las proteínas W130F, T156A y P9T/R122G se expresaron en concentraciones inferiores a la HEXaspC.

La actividad de las proteínas mutantes HEXaspC para la producción de S,S-monatina se midió mediante el uso de las siguientes condiciones de reacción: Cada 1 mL de reacción contenía 50 mM de TAPS, pH 8,2, 4 mM de MgCl<sub>2</sub>, 3 mM de fosfato de sodio, pH 8,0, 200 mM de piruvato de sodio (pH ajustado a 8), 5 mM de α-cetoglutarato (pH ajustado a 8), 50 mM de triptófano, 0,05 mM de piridoxal 3-fosfato, 50 µg/mL de ProA aldolasa (agregada como un extracto libre de células) y concentraciones variables (aproximadamente 50 y 500 µg/mL) de aminotransferasa (agregada como un extracto libre de células). Se utilizó agua desaireada para preparar las soluciones madre y para ajustar el volumen de las mezclas de reacción a 1,0 mL. Se agregó fosfato piroxidial justo antes de la adición de las enzimas. Los tubos de reacción se incubaron a 30°C con leve agitación durante 4 horas. Las muestras (0,01 mL) se retiraron a las 1, 2 y 4 horas después de la adición de las enzimas, se filtraron y se analizaron mediante LC/MS/MS, como se describe en el ejemplo 1. La producción de monatina se normalizó con base en la cantidad de aminotransferasa presente en las reacciones. En las condiciones de estos ensayos, la HEXaspC y la HEXaspCT156A produjeron la mayor cantidad de monatina por mg de aminotransferasa mientras que la proteína P9T/R122G produjo la menor, seguida por la HEXaspCW130F. Las enzimas HEXaspCW130F y P9T/R122G mostraron la mayor estereoselectividad para S-MP (mayor que un 98% de S,S-monatina), incluso cuando se utilizaron altas concentraciones de enzimas (mayores que 300 µg/mL). El porcentaje de producto de S,S-monatina disminuyó a menos de un 90% en las reacciones enzimáticas que contenían la enzima P9T/T156A en una concentración elevada. Los otros mutantes mostraron una estereoselectividad de producto muy similar al mutante HEXaspC original (aproximadamente 95% de S,S-monatina). El análisis del producto de la reacción que contenía la enzima HEXaspC mediante el uso del reactivo de derivatización de la FDAA descrito en el ejemplo 1 mostró que el segundo estereoisómero formado fue la R,S-monatina.

#### Ensayo de la actividad de la aminotransferasa de triptófano y monatina

Los mutantes se probaron para verificar la actividad de transaminación mediante el uso de monatina S,S y L-triptófano como sustratos. La actividad de la aminotransferasa se midió mediante el seguimiento de la formación del coproducto de la reacción, glutamato, por HPLC con derivatización post columna con OPA como se describe en el ejemplo 1. La mezcla de reacción contenía, en 1,0 mL, 100 mM de tampón HEPPS, pH 8,0, 20 mM de alfa-cetoglutarato, 0,08 mM de piridoxal fosfato, 25 mM de triptófano o monatina S,S y enzima (proporcionado como 2,5 mg de proteína de extractos celulares). Todos los componentes excepto la enzima se mezclaron juntos, la enzima se agregó para comenzar la reacción y la solución de reacción se incubó a 30°C (agitación suave) durante 90 minutos. Las reacciones se realizaron por duplicado, con controles negativos donde no se agregaron enzimas. La reacción se detuvo mediante la adición de un 10% de ácido fórmico (concentración final), la mezcla se centrifugó a 21.000 rpm y el sobrenadante se extrajo cuidadosamente y se filtró. Los datos se corrigieron para los niveles de fondo de glutamato y para la dilución desde la adición de ácido para precipitar las proteínas, después se normalizaron por la cantidad de aminotransferasa mutante agregada. Cuando se utilizó triptófano como sustrato, el HEXaspC produjo 13,0 mM de glutamato por mg de aminotransferasa por hora. La actividad relativa, expresada como porcentaje de los mutantes es la siguiente: HEXaspCW130F (156%), HEXaspCT156A (151%), HEXaspCP9T/T156A (63,7%), HEXaspCP9T/R122G (116%) y HEXaspCR122G/T156A (107%). Cuando se utilizó monatina S,S como sustrato, el HEXaspC produjo 7,43 mM de glutamato por mg de aminotransferasa por hora. La actividad relativa, expresada como porcentaje de los mutantes es la siguiente: HEXaspCW130F (113%), HEXaspCT156A (87,7%), HEXaspCP9T/T156A (67,3%), HEXaspCP9T/R122G (11,2%) y HEXaspCR122G/T156A (114%).

El mutante HEXaspCP9T/R122G presentó una actividad superior para la conversión de triptófano a indol-3-piruvato, pero una actividad inferior para la transaminación de monatina S,S. La proporción de triptófano para la actividad de monatina fue de 18,2 en comparación con 1,75 para HEXaspC, lo que lo hizo un candidato deseable para la producción de monatina R,R con el uso de vías que requieren una L-aminotransferasa como las descritas en los ejemplos 18 y 19. Como tal, el HEXaspCP9T/R122G es un ejemplo de una aminotransferasa con actividad limitada sobre la monatina S,S (y MP).

La mayoría de las mutaciones mejoró la actividad del L-triptófano, pero únicamente dos mutantes aumentaron la actividad del L-triptófano y la monatina S,S (HEXaspCW130F y HEXaspCR122G/T156A). Debido a que 25 mM del sustrato se utilizó en estos ensayos, es más probable que las enzimas estuvieran saturadas y la actividad es una reflexión de  $k_{cat}$  de las enzimas. No obstante, en las condiciones en las que se realizaron los ensayos para la producción de la monatina S,S (anteriormente) es improbable que la concentración de S-MP sea suficiente para saturar la enzima, por consiguiente, no se refleja el aumento en  $k_{cat}$ . Se ha informado que, para sustratos similares, que algunas de las mutaciones hicieron aumentar la  $k_{cat}$  pero también aumentaron la  $K_m$  aparente para el sustrato de aminoácido. Al utilizar el aumento de las concentraciones de los sustratos, se espera que estos dos mutantes proporcionen un beneficio en las tasas de producción de monatina S,S en comparación con HEXaspC. La mutación de HEXaspCT156A parece tener tasas de transaminación de triptófano superiores sin tener un efecto significativo sobre la tasa de transaminación de MP en las condiciones anteriores para la producción de monatina S,S.

Mediante la comparación de las estructuras de HEXaspC y una de las enzimas de D-aminotransferasa de *Bacillus* sp. (ver, por ejemplo, S. Sugio, G. A. Petsko, J. M. Manning, K. Soda, y D. Ringe, *Biochemistry* 34 (1995) pp. 9661-9669), pueden trazarse las mutaciones W130F, R122G, T156A y HEX de AspC para los residuos correspondientes en la estructura de D-aminotransferasa. Se espera que mutaciones similares en la D-aminotransferasa de amplia especificidad mejoren la producción global de monatina R,R como se describe en el ejemplo 14. Por ejemplo, la funcionalidad proporcionada por el triptófano 130 en AspC es reemplazada por D-aminotransferasas en *Bacillus* mediante enlace de hidrógeno entre las cadenas laterales de las serinas 179 a 181 y glutamato 166 (esquema de numeración YM-1). Para disminuir el impedimento estérico, puede mutarse el glutamato a un residuo de aspartato. Algunas D-aminotransferasas presentan un residuo de treonina en la posición 179, lo que aumentaría el impedimento estérico y debería evitarse. La enzima de *B. sphearicus* tiene una alanina en lugar de una serina en la posición 181, que también puede reducir el impedimento estérico.

La información adicional de los estudios de la aspartato aminotransferasa también puede aplicarse a la D-aminotransferasa. Mientras que la enzima AspC tiene una arginina en el sitio activo que interactúa con la cadena lateral de los sustratos de dicarboxilato, la D-aminotransferasa tiene un asa de Ser240 a Ser243. Las cadenas laterales de Ser240, Tr242 y Ser243 siguen la misma dirección y forman un bolsillo con el grupo hidroxilo de Ser180 que proporciona una superficie para que los sustratos no polares y polares puedan interactuar. La Ser180 se encuentra implicada en la unión de PLP; no obstante, para mejorar la actividad de una D-aminotransferasa con R-MP, se puede modificar los residuos de Ser240, Tr242 o Ser243 para que acepten sustratos mayores o para favorecer a los sustratos cargados negativamente. Por ejemplo, la Tr242 puede mutarse a Ser para reducir la longitud de la cadena lateral. Uno de los residuos puede mutarse a lisina o arginina, como Ser243. Los residuos (numeración YM-1) Val30-Val36 se ubican en una hebra beta a lo largo del sitio activo de la D-aminotransferasa y también son importantes para la actividad. Se piensa que Tir31, Val33, Glu32 y Lis35 están de cara al sitio activo. Tir31, Glu32 y Val33 son invariables en todos los homólogos de *Bacillus*. Ro et al. (FEBS Lett 398 (1996) pp. 141-145) mutagenizaron la Val33 a Ala y encontraron un leve aumento en la eficiencia catalítica para la transaminación de alfa-cetoglutarato y una eficiencia catalítica significativamente superior para los sustratos más voluminosos (impedimento estérico menor). En algunos homólogos Lis35 se reemplaza con Arg, pero el impedimento estérico es una preocupación, se prefiere el residuo de Lis. Las valinas 34 y 36 también son preferidas sobre las sustituciones conservadoras como la isoleucina, nuevamente debido al menor impedimento estérico para moléculas mayores como MP.

### Ejemplo 17

#### Clonación, expresión y prueba de racemasas de glutamato y aspartato

Este ejemplo describe los procedimientos utilizados para clonar y probar las enzimas de racemasas de aminoácidos, que pueden utilizarse para interconvertirse entre L-glutamato y D-glutamato (o L- y D-aspartato o L- y D-alanina). Las racemasas de glutamato, aspartato o alanina son útiles en una vía biosintética para producir monatina R,R cuando una etapa en dicha vía produce un L-aminoácido (como L-glutamato, L-aspartato o L-alanina) y otra etapa en la vía consume un D-aminoácido (como D-glutamato, D-aspartato o D-alanina). La figura 4 ilustra una vía biosintética para producir monatina R,R a partir de L-triptófano mediante el uso de una aminotransferasa específica para L-triptófano, una aldolasa R específica, una D-aminotransferasa y una glutamato racemasa (aspartato o alanina).

Los genes se clonaron en los vectores pET28 y pET30 para generar las proteínas sin marcar y las proteínas de fusión con terminal N escindible HIS<sub>6</sub>-marcador/T7-marcadores. Las proteínas resultantes se purificaron mediante el uso de cromatografía de afinidad por iones metálicos inmovilizados.

Resumen experimental

Los genes que codifican las racemasas de glutamato (EC 5.1.1.3) de *Lactobacillus brevis* (Genbank No. de acceso D29627, secuencia de ácidos nucleicos) y *Pediococcus pentosaceus* (gen murl) (Genbank No. de acceso L22789) se clonaron y se expresaron en *E. coli*. Los extractos se probaron para verificar la actividad en la conversión de L-glutamato a D-glutamato y D-glutamato a L-glutamato. La enzima aspartato racemasa de BioCatalytics (EC 5.1.1.13) (BioCatalytics, Pasadena, CA) también se probó para la interconversión entre L- y D-aspartato.

Aislamiento de ADN Genómico para la clonación

El ADN genómico de *L. brevis* (ATCC 8287D) se obtuvo de la colección de cultivos "American Type Culture Collection". Se cultivó *P. pentosaceus* (ATCC 25745) a 37°C en caldo MRS de *lactobacilli* y se utilizaron 2 ml para el aislamiento del ADN genómico mediante el uso del procedimiento de Mekalanos J J. de duplicación y ampliación de genes de toxina en células de *Vibrio cholerae*. 1983. 35:253-63.

Protocolo de la reacción en cadena de la polimerasa

Los cebadores se diseñaron con sitios de restricción 5' y salientes para clonar a los vectores pET 28 y pET30 (Novagen, Madison, WI).

15 Cebadores de glutamato racemasa de *L. brevis*:

Terminal N: 5'-GCGGCGCCATGGAAAATGATCCGATTGGTCTAATG-3', (SEQ ID NO: 401), y

Terminal C: 5'-GCGGCGGTGCGACGCAATTACAATTGTGTTTGTGTC-3'. (SEQ ID NO: 402).

Cebadores de glutamato racemasa de *P. pentosaceus*:

Terminal N: 5'-GCGGCGCCATGGATGTATGTATAATTTTATTTAG-3', (SEQ ID NO: 403), y

20 Terminal C: 5'-GCGGCGGTGCGACAAATTTTCATTATTCTAATTT-3' (SEQ ID NO: 404).

El gen derivado de *L. brevis* se amplificó mediante el uso del siguiente protocolo de PCR. En una reacción de 50 µL se utilizó 0,150 µg de plantilla, 1,6 µM de cada cebador, 0,4 mM de cada dNTP, 2,8 U de polimerasa Expand High Fidelity™ (Roche, Indianápolis, Ind.), 0,5 U de polimerasa Pfu (Stratagene, La Jolla, CA) y tampón 1× Expand™ con Mg. El programa termociclador utilizado incluyó un comienzo en caliente a 96°C durante 3 minutos, 8 repeticiones de las siguientes etapas: 94°C durante 30 segundos, 52°C durante 45 segundos y 72°C durante 2 minutos, seguidas por 22 repeticiones de las siguientes etapas: 94°C durante 30 segundos, 60°C durante 45 segundos y 72°C durante 2 minutos. Después de 22 repeticiones la muestra se mantuvo a 72°C durante 7 minutos y a continuación se almacenó a 4°C. Este protocolo de PCR produjo un producto de aproximadamente 830 bp, a juzgar por la comparación con los marcadores de tamaño de ADN.

30 El gen derivado de *P. pentosaceus* se amplificó mediante el uso del siguiente protocolo de PCR. En una reacción de 50 µL se agregó 0,15 µg de plantilla, 1,6 µM de cada cebador, 0,4 mM de cada dNTP, 2,8 U de polimerasa Expand High Fidelity™, 0,5 U de polimerasa y tampón 1× Expand™ con Mg. El programa termociclador utilizado incluyó un comienzo en caliente a 96°C durante 3 minutos, seguido de 8 repeticiones de las siguientes etapas: 94°C durante 30 segundos, 37°C durante 45 segundos y 72°C durante 2 minutos, seguidas por 8 repeticiones de las siguientes etapas: 94°C durante 30 segundos, 45°C durante 45 segundos y 72°C durante 2 minutos, seguidas por 14 repeticiones de las siguientes etapas: 94°C durante 30 segundos, 55°C durante 45 segundos y 72°C durante 2 minutos. Después de 14 repeticiones la muestra se mantuvo a 72°C durante 7 minutos y a continuación se almacenó a 4°C. Este protocolo de PCR produjo un producto de aproximadamente 840 bp, a juzgar por la comparación con los marcadores de tamaño de ADN.

Clonación

Los productos de la PCR se purificaron en geles del 0,8% TAE-agarosa mediante el uso del kit de extracción en gel de Qiagen (Qiagen, Valencia, CA). Los productos de la PCR se cuantificaron mediante el uso de un espectrofotómetro SmartSpec 3000™. Los productos se digirieron con enzimas de restricción NcoI y Sall de acuerdo con los protocolos recomendados del fabricante (Biolaboratorios de Nueva Inglaterra, Beverly, MA) y se purificaron en geles del 0,8% TAE-agarosa mediante el uso del kit de extracción en gel de Qiagen (Qiagen, Valencia, CA). Los vectores pET28 y pET30 se prepararon mediante digestión con las enzimas de restricción NcoI y Sall seguidos del tratamiento con fosfatasa alcalina sérica y purificación en geles del 0,8% TAE-agarosa mediante el uso del kit de extracción en gel de Qiagen (Qiagen, Valencia, CA).

50 Los vectores e inserciones digeridos se ligaron mediante el uso del kit de ligadura rápida de ADN (Roche, Indianápolis, IN). Aproximadamente 50 ng de la inserción tratada, 100 ng del vector tratado (proporción de inserción a vector 3 a 1), 5 U de T4 ADN ligasa y 1× de tampón de ligación se incubaron durante 5 minutos a temperatura



ambiente. Las reacciones de ligación se purificaron mediante el uso del kit de purificación del producto de la PCR altamente puro (Roche, Indianápolis, IN) y se utilizaron para transformar las células electrocompetentes DH10B de *E. coli* (Invitrogen, Carlsbad, CA). Se agregaron 10 µl de cada reacción de ligación a 40 µl de células DH10B y se transformaron mediante electroporación mediante el uso del impulsor de genes Bio-Rad II (Bio-Rad, Hercules, CA) en las siguientes condiciones: 2,5 kV, 25° F, 200 ohm en una cubeta de 0,2 cm. Las células se dejaron recuperar en 1 mL de medio SOC a temperatura ambiente durante 1 hora a 37°C con agitación a 225 rpm. Las células se colocaron en placas LB que contenían kanamicina (50 µg/mL).

El ADN plásmido se purificó a partir de los transformantes resultantes mediante el uso del kit spin miniprep de Qiagen (Qiagen, Valencia, CA) y se analizó para detectar las inserciones correctas mediante enzimas de restricción con NcoI y Sall. Las secuencias de plásmidos que parecen tener una inserción correcta se verificaron mediante el secuenciado de ADN de terminación de la cadena didesoxi.

#### Expresión y análisis génicos

El ADN plásmido, verificado por análisis de secuencias, se subclonó en hospedadores de expresión de *E. Coli* BL21(DE3) (Novagen, Madison, WI). Los cultivos se desarrollaron. Los plásmidos se aislaron mediante el uso del kit miniprep de Qiagen (Qiagen, Valencia, CA) y se analizaron mediante enzimas de restricción para confirmar la identidad.

La inducción en BL21(DE3) se realizó inicialmente con glutamato racemasas de *L. brevis* y *P. pentosaceus* en los vectores pET28 (sin marcar) y pET 30 (marcados con histidina). Se realizó un estudio temporal con cultivos desarrollados en 250 mL de kanamicina que contenía LB (50 mg/L) a una OD<sub>600</sub> entre 0,5 y 0,6, inducidos con 100 mM de IPTG (isopropil tiogalactósido) y muestreados a las 0 y 3 horas después de la inducción. Las células procedentes de 600 µL (hora 0) y 275 µL (a las 3 horas) se volvieron a colocar en suspensión en tampón de dodecil sulfato de sodio que contenía 2-mercaptoetanol calentado a 95°C durante 10 minutos y se enfriaron. Se analizaron por SDS-PAGE alícuotas de estas muestras de la proteína celular completa mediante el uso de un gel en gradiente del 4 al 15%.

Se prepararon también extractos celulares procedentes de los cultivos de 3 horas mediante la colocación en suspensión de los sedimentos celulares procedentes de 5 ml de cultivo en 0,625 ml de reactivo BugBuster™ de Novagen (Novagen, Madison, WI) que contenía 0,625 µl de benzonasa nucleasa y 3 µL de mezcla de inhibidor de proteasa grupo #3 (Calbiochem-Novabiochem Corp., San Diego, CA) a temperatura ambiente durante 20 minutos con leve agitación y centrifugación a 16.000 x g para eliminar el residuo celular. Los sobrenadantes (extractos celulares) se cargaron en geles con gradiente del 4 al 15% para el análisis de las proteínas celulares solubles.

Las muestras de las 3 horas de glutamato racemasas de *L. brevis* y glutamato racemasas de *P. pentosaceus* clonadas mostraron una proteína total y soluble que correspondió al tamaño correcto (aproximadamente 31 kDa). El producto génico en pET30 (marcado con histidina) de *L. brevis* se sobreexpresó a un nivel superior y también fue más soluble (>20% de proteína soluble) que el producto génico en pET 28 (sin marcar) de *L. brevis*, así como los productos génicos de *P. pentosaceus* en ambos vectores. El producto génico de *P. pentosaceus* mostró la misma sobreexpresión y solubilidad en los vectores pET28 y pET30, que fue significativamente inferior a la observada para el producto génico en pET30 de *L. brevis*.

Las células de los cultivos inducidos (250 mL) se centrifugaron y se lavaron una vez con NaCl al 0,85%. Los sedimentos celulares se volvieron a suspender en 5 mL/g de peso de células húmedas de reactivo BugBuster™ (Novagen, Madison, WI) que contenía 5 µl/ml de la mezcla de inhibidor de proteasa del conjunto n° 3 (Calbiochem-Novabiochem Corp., San Diego, CA) y 1 µl/ml de benzonasa nucleasa. Las muestras se incubaron a temperatura ambiente durante 20 minutos en un agitador orbital. El residuo celular insoluble se eliminó por centrifugación a 16.000 x g durante 20 minutos a 4°C.

Los extractos celulares se analizaron para verificar la actividad de la glutamato racemasa mediante el uso del siguiente protocolo. Las reacciones de 400-µL se llevaron a cabo en 10 mM de fosfato de potasio (pH 8,0), 0,2 mM de DTT y 10 mM de L-glutamato o D-glutamato. Las reacciones se iniciaron mediante la adición de 20-100 µL de extractos libres de células y se incubaron a temperatura ambiente. Se tomaron alícuotas de las muestras durante un periodo de tiempo de 1 minuto, 5 minutos, 10 minutos, 20 minutos y 1 hora (las muestras del minuto cero sirvieron como reacciones de control). Se agregó 2 M de ácido fórmico (25 µL) a cada alícuota de muestra de 40 µL para detener la reacción y la proteína precipitada se eliminó mediante centrifugación. Los sobrenadantes se extrajeron y congelaron a -80°C hasta que se analizaron mediante LC/MS/MS.

Los resultados de los ensayos de los extractos celulares de la inducción de pET30 con 100 mM de IPTG (3 horas) demuestran que las enzimas de *L. brevis* (Genbank acceso no. BAA06106.1 GI:468450) y *P. pentosaceus* (Genbank acceso no. AAA16761.1 GI:349029) tienen niveles significativos de actividad de racemasas en ambos isómeros de glutamato. La racemasa de *P. pentosaceus* (20 µL de extractos celulares) alcanzó el equilibrio entre L- y D-glutamato a los 10 a 20 minutos del comienzo con cualquiera de los sustratos. La enzima de *L. brevis* (20 µL de extractos celulares) también alcanzó el equilibrio en aproximadamente 20 minutos.

Se analizó la enzima aspartato racemasa parcialmente purificada (catálogo #ASPR-101) comprada a BioCatalytics, Inc. (Pasadena, CA) para verificar la actividad en el L-aspartato y D-aspartato mediante el uso de un protocolo similar al que antecede. La enzima comercial mostró actividad de racemasa en ambos isómeros. Mediante el uso de 0,5-1 mg de enzima, se logró el equilibrio a los 20 a 60 minutos.

- 5 También se analizaron las tres racemasas (glutamato racemasa de *L. brevis*, glutamato racemasa de *P. pentosaceus* y aspartato racemasa de BioCatalytics) para verificar la actividad en la monatina S,S mediante el uso del siguiente protocolo. Las reacciones de 400- $\mu$ L se llevaron a cabo en 10 mM de fosfato de potasio (pH 8,0), 0,2 mM de DTT y 10 mM de monatina S,S. Las reacciones se iniciaron mediante la adición de extractos libres de células (*L. brevis* y *P. pentosaceus*) o enzima purificada (aspartato racemasa de BioCatalytics) y se incubaron a temperatura ambiente. Se tomaron alícuotas de las muestras durante un periodo de tiempo de 1 minuto, 5 minutos, 10 minutos, 20 minutos y 1 hora (las muestras del minuto cero sirvieron también como muestras sin enzima). Se agregó 2 M de ácido fórmico (25  $\mu$ L) a cada alícuota de muestra de 40  $\mu$ L para detener la reacción y la proteína precipitada se eliminó mediante centrifugación. Los sobrenadantes se extrajeron y congelaron a  $-80^{\circ}\text{C}$  hasta que se analizaron mediante LC/MS/MS (ejemplo 1). No se notó ninguna disminución en la concentración de monatina S,S a lo largo del tiempo ni hubo ningún aumento en la monatina S,R (presente inicialmente como subproducto contaminante <5%, incluso en el control sin enzimas). Por consiguiente, ninguna de las racemasas analizadas mostró actividad con respecto a la monatina.

### Ejemplo 18

Producción de monatina R,R a partir de L-triptófano mediante el uso de glutamato o aspartato racemasas

- 20 Este ejemplo describe los procedimientos para producir monatina R,R enriquecida estereoisoméricamente a partir de L-triptófano mediante el uso de una L-triptófano (L-tirosina o aromática) aminotransferasa, ProA aldolasa, glutamato o aspartato racemasa y D-aminoácido aminotransferasa de amplia especificidad. La figura 5 es un diagrama que ilustra la vía. Este enfoque de producción de monatina R,R enriquecida estereoisoméricamente requiere una enzima para la etapa 1 que tenga una actividad reducida en la producción de monatina a partir del precursor de monatina (MP). Con base en resultados anteriores, se utilizaron los productos de genes *tatA* de *Sinorhizobium meliloti* y *Rhodobacter sphaeroides* descritos en el ejemplo 1 de WO 03/091396 A2.

Materiales y procedimientos

- Las glutamato racemasas de *L. brevis* y *P. pentosaceus* se produjeron en *E. coli* como se describe en el ejemplo 17. En algunos casos la versión marcada con His<sub>6</sub> de estas enzimas se purificó mediante el uso de cartuchos His-Bind 900 de acuerdo con los protocolos del fabricante (Novagen, Madison, WI) y se desalaron para extraer el imidazol mediante el uso de columnas PD-10 (G25 Sephadex, GE Healthcare, Piscataway, NJ). Las enzimas se eluyeron en 25 mM de fosfato de potasio, pH 8,0. La aspartato racemasa (ASPR-101) y la D-amino transferasa (AT-103) se compraron a BioCatalytics, Inc. (Pasadena, CA). Las tirosina (aromática) aminotransferasas de *S. meliloti* y *R. sphaeroides* se prepararon como se describe en el ejemplo 1 de WO 03/091396 A2. La ProA aldolasa de *Comamonas testosteroni* se preparó como se describe en el Ejemplo 4 de WO 03/091396 A2. La totalidad de los ensayos de proteínas se realizó mediante el uso de un ensayo de proteínas Bio-Rad de acuerdo con los protocolos del fabricante (Hercules, CA).

Reducción en cantidad de monatina S,S producida mediante el uso de racemasas

- Las mezclas de reacción (1 mL de volumen, realizadas por duplicado) contenían 100 mM de tampón de fosfato de potasio (pH 8), 2 mM de MgCl<sub>2</sub>, 0,05 mM de piridoxal 5'-fosfato (PLP), 200 mM de piruvato de sodio, 5 mM de  $\alpha$ -cetoglutarato u oxaloacetato de sodio, aproximadamente 280  $\mu$ g/mL de TatA de *S. meliloti* proporcionado en un extracto celular, 1 mg/mL de D-amino transferasa (AT-103) de BioCatalytics (BioCatalytics, Pasadena, CA), 100  $\mu$ L/mL de extracto celular de glutamato racemasa o 1 mg/mL de aspartato racemasa y aproximadamente 100  $\mu$ g/mL de ProA aldolasa proporcionada como extracto celular. Se agregó triptófano sólido a una concentración de 10,2 mg/ml. Los controles negativos no contenían racemasa. Las mezclas se incubaron a 30°C (agitación a 250 rpm) durante 1 hora, 2 horas o durante la noche. Las muestras se centrifugaron para extraer el precipitado, se filtraron por jeringa y se almacenaron a  $-80^{\circ}\text{C}$  antes del análisis de la monatina mediante el uso del procedimiento de LC/MS/MS descrito en el ejemplo 1. La mayoría de las muestras contenía >95% de monatina S,S, debido a las cantidades de L-amino transferasa nativa presente en los extractos celulares. No obstante, las muestras que contenían racemasa tenían una cantidad reducida de monatina total como resultado de las enzimas de racemasa, lo que provocó que el L-glutamato estuviera menos disponible para la transaminación de MP. Sin racemasa, se produjo monatina de 1545-2355 ppm (predominantemente S,S) durante el lapso de tiempo. Con las racemasas presentes, se produjo únicamente monatina (aspartato racemasa) de 340-879 ppm (enzima de *L. brevis*), 444-531 ppm (enzima de *P. pentosaceus*) y 506-1460 ppm. Estos datos indican que las racemasas son activas en las condiciones de reacción requeridas para producir monatina. Para minimizar la formación de monatina S,S a partir de enzimas de extracto celular como las aspartato aminotransferasas, se realizaron experimentos adicionales con enzimas purificadas y una mayor proporción de enzimas de D-amino transferasa a L-amino transferasa.

Conversión de L-triptófano en isómeros de monatina que contienen 4-R

Los experimentos que anteceden se repitieron mediante el uso de aproximadamente 54 µg de L-aminotransferasa purificada (ya sea TatA de *S. meliloti* o *R. sphaeroides*), 1 mg de aspartato aminotransferasa (BioCatalytics, Pasadena, CA), 1 mg de D-aminotransferasa, oxaloacetato como el aceptante amino y 75 µg de aldolasa purificada. Las reacciones se realizaron por duplicado con un tiempo de muestreo de 2 horas y un tiempo de incubación durante la noche. Los controles negativos se realizaron con L-aminotransferasa de *S. meliloti* pero no racemasa. Además de la cuantificación del pico de R,R/S,S y S,R/monatina R,S basada en la cromatografía de fase inversa, se determinó el porcentaje de cada estereoisómero mediante el uso de la técnica de derivatización de la FDAA descrita en el ejemplo 1. Los resultados son los siguientes:

5

Tabla 22

L-Aminotransferasa	Tiempo de incubación	Monatina total (ppm)	% S,S	% R,R	% R,S	% S,R
TatA de <i>S. meliloti</i>	2 horas	17,1	10,2	58,1	0,8	31,0
TatA de <i>S. meliloti</i>	2 horas	15,8	13,3	55,3	1,0	30,4
TatA de <i>S. meliloti</i>	durante la noche	77,7	25,8	40,0	1,3	32,9
TatA de <i>S. meliloti</i>	durante la noche	67,9	29,4	37,3	1,5	31,8
TatA de <i>R. sphaeroides</i>	2 horas	241,2	96,3	2,3	0,8	0,6
TatA de <i>R. sphaeroides</i>	2 horas	223,2	95,7	2,7	1,0	0,6
TatA de <i>R. sphaeroides</i>	durante la noche	600,4	96,6	1,8	0,5	1,1
TatA de <i>R. sphaeroides</i>	durante la noche	618,5	96,1	2,1	0,5	1,3
sin control de racemasa	2 horas	7,1	92,0	1,4	6,6	0,0
sin control de racemasa	2 horas	5,7	94,0	1,2	4,8	0,0
sin control de racemasa	durante la noche	44,6	93,5	1,3	4,7	0,5
sin control de racemasa	durante la noche	37,5	95,4	0,9	3,7	0,0

15

La presencia de racemasa aumentó claramente la cantidad total de monatina producida cuando se utilizó TatA de *S. meliloti* como la enzima para la transaminación de L-triptófano. Los niveles de monatina aumentaron de un promedio de 6,4 a 16,5 ppm en el ensayo de 2 horas y de 41 a 73 ppm en el ensayo durante la noche. Además, el porcentaje de R,R formado aumentó de aproximadamente un 1% a tanto como un 58% mediante el uso de la enzima racemasa. El S,R estereoisómero de monatina, otro poderoso edulcorante, fue el otro componente principal, lo que aumentó de casi 0 en los controles negativos a un 31%. El TatA de *R. sphaeroides* presentó más actividad en S-MP que la L-transaminasa de *S. meliloti*, lo que demostró la importancia de tener una enzima que presenta una alta especificidad de sustrato para el L-triptófano en comparación con MP cuando los isómeros 4-R de monatina son los productos deseados. Como aproximadamente un 10% de la monatina total fue 4S en el intervalo de tiempo de dos horas, pudo considerarse que el TatA de *S. meliloti* tenía una actividad limitada en el MP.

20

Los experimentos se repitieron con el TatA de *S. meliloti* purificado (54 µg) y la glutamato racemasa de *L. brevis*. Cuando se utilizó glutamato racemasa purificada, se utilizaron aproximadamente 64 µg por 1 mL de reacción. También se probaron los extractos celulares que contenían glutamato racemasa y se utilizaron 1,4 mg de proteína

soluble. Se utilizó nuevamente un control negativo sin racemasa y todas las muestras se realizaron por duplicado. Los resultados son los siguientes:

Tabla 23

Glutamato racemasa	Tiempo de incubación	Monatina total (ppm)	% S,S	% R,R	% R,S	% S,R
<i>L. brevis</i> (purificada)	2 horas	3,3	49,1	34,2	3,7	13,0
<i>L. brevis</i> (purificada)	2 horas	3,6	47,9	35,2	3,5	13,4
<i>L. brevis</i> (purificada)	durante la noche	29,3	58,9	24,7	3,2	13,2
<i>L. brevis</i> (purificada)	durante la noche	40,2	55,1	25,0	4,7	15,3
<i>L. brevis</i> (extracto celular)	2 horas	10,5	45,8	35,9	1,1	17,2
<i>L. brevis</i> (extracto celular)	2 horas	10,5	47,4	33,9	1,1	17,6
<i>L. brevis</i> (extracto celular)	durante la noche	79,4	70,3	17,9	1,3	10,5
<i>L. brevis</i> (extracto celular)	durante la noche	80,1	69,1	19,1	1,1	10,7
ninguno	2 horas	2,7	84,1	7,1	6,3	2,4
ninguno	2 horas	3,2	84,9	6,0	6,8	2,2
ninguno	durante la noche	36,5	92,3	2,3	4,2	1,2
ninguno	durante la noche	30,5	92,7	2,0	4,1	1,3

- 5 Nuevamente, es claro que la adición de la racemasa aumenta la monatina total producida a partir de L-triptófano, así como aumenta las cantidades relativas de isómeros de monatina que contienen 4R en comparación con la monatina S,S. El uso de aldolasa, racemasa y L-aminotransferasa purificada aumenta enormemente la capacidad de control de la formación de estereoisómeros deseada. La proporción de L a D aminotransferasa es también una forma de manipular la estereoquímica del producto final.
- 10 Al comparar los resultados mostrados en las tablas 18 y 19 en el ejemplo 13 con los resultados con condiciones de reacción similares a las condiciones que anteceden, uno puede ver que se formaron aproximadamente 7 a 29 ppm de monatina a partir de indol-3-piruvato y los porcentajes de monatina R,R formados fueron de aproximadamente un 51 a un 90%. El uso de la aspartato racemasa aumentó la cantidad total de monatina producida de 16 a 78 ppm de monatina con un % de R,R de aproximadamente un 40 a un 58%. Además, se utilizó una materia prima más estable y menos costosa, L-triptófano. En el ejemplo 14, se produjeron aproximadamente 73 ppm de monatina a partir de D-triptófano en una proporción de R,R:S,R de aproximadamente 1,7:1. La cantidad total de isómeros 4R fue > al 80% de la monatina total. Debido a que la R,R-monatina y S,R-monatina son poderosos edulcorantes (1000 veces más dulce que la sacarosa) la capacidad de enriquecer estos isómeros sin la necesidad de sustratos de D-aminoácidos costosos es crítica.
- 20 Como se describe en los ejemplos 13 y 14, la D-alanina puede servir como donante amino para la transaminación de MP a monatina. Muchas L-aminotransferasas tienen la capacidad de utilizar piruvato como aceptante amino en cierta medida y producir L-alanina. Debido a que las reacciones mencionadas anteriormente utilizan altas concentraciones de piruvato, es probable que cierta cantidad de piruvato se convierta en L-alanina. Por ejemplo, durante la transaminación de L-triptófano, se encontró que la enzima HexAsp descrita en el ejemplo 16 convierte de un 10 a un 18% del piruvato (concentraciones iniciales de 50 a 200 mM) a L-alanina en 2 horas si el cetoglutarato
- 25

alfa se encuentra ausente. La enzima mostró una preferencia de 10 veces para el cetoglutarato alfa cuando ambos aceptores amino se encontraban presentes en concentraciones mayores (>50 mM). La AspC (descrita en WO 03/091396 A2) también produjo cierta cantidad de L-alanina a partir de piruvato. Por consiguiente, se espera que se pueda omitir la adición de cetoglutarato alfa u oxaloacetato en las reacciones anteriores y utilizar alanina racemasa (EC 5.1.1.1) en lugar de glutamato o aspartato racemasa. Las enzimas alanina racemasa se identificaron primero en *Brucella abortus* y *Streptococcus faecalis* (Marr, A. G. y Wilson, P. W. Arch. Biochem. Biophys., 49 (1954) 424-433 y Wood, W. A. y Gunsalus, I. C. J. Biol. Chem., 190 (1951) 403-416). El gen *dadB* en *Salmonella typhimurium* se identificó como la fuente de actividad de alanina racemasa y pueden encontrarse varios cientos de homólogos en las bases de datos genómicas. Otras fuentes conocidas de actividad de alanina racemasa son *Escherichia coli*, *Bacillus subtilis*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Schizosaccharomyces pombe* y *Bacillus cereus*. El hongo basidiomiceto, *Lentinus edodes*, también contiene una alanina racemasa de amplia actividad. Un homólogo termoestable de *Bacillus stearothermophilus* se encuentra disponible para su compra a Sigma-Aldrich (catálogo #A8936) (Sigma-Aldrich, St. Louis, MO) y ha sido inmovilizado para aplicaciones comerciales (Inagaki, K., Biochemistry, 25: 3268 1986). La alanina racemasa se convierte a una glutamato o aspartato racemasa mediante procedimientos aleatorios como PCR mutagénica, pasaje a través de cepas mutagénicas u otros procedimientos conocidos por los expertos en la técnica. Una evolución más centrada de la alanina racemasa se centra en los residuos del sitio activo, que incluyen Lis129, Met134 y residuos que incluyen entre Gli283 y Trp288 (numeración de *Bacillus stearothermophilus*).

### Ejemplo 19

20 D-fenilglicina aminotransferasa (D-4-hidroxifenilglicina aminotransferasa)

Como se muestra en la figura 3, la D-fenilglicina aminotransferasa es útil en una vía biosintética para la producción de monatina. Por ejemplo, la D-fenilglicina aminotransferasa produce monatina R,R a partir de R-MP con L-glutamato como el donante amino.

Síntesis de PCR de 4 D-hidroxifenilglicina aminotransferasa de *P. stutzeri* a partir de cebadores de oligonucleótidos

25 Este ejemplo describe los procedimientos que se utilizaron para sintetizar la 4 D-hidroxifenilglicina aminotransferasa, una enzima estereoinversora que puede utilizarse para convertir el precursor de monatina R en monatina R,R mediante el uso de L-glutamato como el donante amino.

Diseño de cebadores

30 La secuencia publicada (Genbank acceso No. AY319935, secuencia de ácidos nucleicos; Genbank acceso No. AAQ8290, secuencia de proteínas) para 4 D-hidroxifenilglicina aminotransferasa de *Pseudomonas stutzeri* (4 D-HPG AT) se utilizó como plantilla para el diseño de cebadores de PCR. De forma alternativa, la 4-D-hidroxifenilglicina aminotransferasa de *Pseudomonas putida*, (CAD42450 (proteína), AX467211 (nucleótido)) se utiliza como una plantilla de secuencia. Se diseñó un total de 34 cebadores directos y 35 cebadores inversos; los cebadores directos e inversos fueron 40 cebadores que compartían 20 pares de base superpuesta. Además, se diseñaron 2 cebadores externos con sitios de restricción 5' y salientes para clonar a los vectores pET 28 y pET30 (Novagen, Madison, WI).

Cebadores externos 4 D-HPG AT de *P. stutzeri*: Terminal N (con sitio NdeI):

5'-GGCCGGCATATGTCGATCCTTAACGACTACAAACGT-3' (SEQ ID NO: 405) y terminal C (con sitio XhoI):

5'-GGAAGGCTCGAGTCATGATTGGTTTCCAGACAAATT-3' (SEQ ID NO: 406).

Protocolo de la reacción en cadena de la polimerasa

40 La secuencia de genes de *P. stutzeri* se amplificó mediante el uso de los siguientes protocolos. La reacción de PCR primaria de 100 µL incluyó 0,05 µM de cada uno de los 69 cebadores internos, 0,4 mM de cada dNTP, 10 U de rTth Polimerasa XL (Roche, Indianápolis, IN), 0,625 U de Pfu polimerasa (Stratagene, La Jolla, CA), 1×XL de tampón y 1 mM de Mg(OAc)<sub>2</sub>. El programa termociclador utilizado incluyó un comienzo en caliente a 94°C durante 3 minutos, 15 repeticiones de las siguientes etapas: 94°C durante 30 segundos, 42°C durante 30 segundos y 68°C durante 15 segundos, seguidas por 10 repeticiones de las siguientes etapas: 94°C durante 30 segundos, 52°C durante 30 segundos y 68°C durante 30 segundos, seguidas por 10 repeticiones de las siguientes etapas: 94°C durante 30 segundos, 60°C durante 30 segundos y 68°C durante 1 minuto y 15 segundos. Después de los 10 ciclos finales la muestra se mantuvo a 68°C durante 7 minutos y a continuación se almacenó a 4°C. Este protocolo de PCR produjo un frotis de producto a aproximadamente 0,5 kb sobre un gel de TAE-agarosa al 0,8%.

50 La reacción de PCR secundaria se configuró mediante el uso de la reacción de PCR primaria como plantilla. La reacción de PCR secundaria de 100 µL incluyó 2,5 µL de la reacción de PCR primaria, 0,5 µM de cada uno de los 2 cebadores externos (con sitios de restricción NdeI y XhoI), 0,4 mM de cada dNTP, 10 U de rTth Polimerasa XL, 0,625 U de Pfu polimerasa, 1×XL tampón y 1 mM de Mg(OAc)<sub>2</sub>. El programa termociclador utilizado incluyó un comienzo en caliente a 94°C durante 3 minutos, 10 repeticiones de las siguientes etapas: 94°C durante 30

segundos, 52°C durante 30 segundos y 68°C durante 1 minuto y 30 segundos, seguidas por 15 repeticiones de las siguientes etapas: 94°C durante 30 segundos, 60°C durante 30 segundos y 68°C durante 1 minuto y 30 segundos. Después de las repeticiones la muestra se mantuvo a 68°C durante 7 minutos y a continuación se almacenó a 4°C. Este protocolo de PCR produjo una banda de producto particular a aproximadamente 1,4 kb sobre un gel de TAE-agarosa al 0,8%.

Los productos de la PCR se purificaron en geles TAE-agarosa al 0,8% mediante el uso del kit de extracción en gel de Qiagen (Qiagen, Valencia, CA). El producto se clonó en TOPO y se transformó en células TOP 10 de acuerdo con el protocolo del fabricante (Invitrogen, Carlsbad, CA). El ADN plásmido se purificó a partir de los transformantes resultantes mediante el uso del kit spin miniprep de Qiagen (Qiagen, Valencia, CA) y se analizó para detectar las inserciones correctas mediante enzimas de restricción con NdeI y XhoI. Las secuencias de plásmidos que parecen tener una inserción correcta se verificaron mediante el secuenciado de ADN de terminación de la cadena didesoxi con los cebadores universales M13 directos y M13 inversos. De los 10 clones secuenciados, todos tenían al menos una mutación de la secuencia deseada. El mejor clon tenía una única mutación de par de base que resultó en un cambio de aminoácido. La secuencia de este clon se corrigió mediante el uso del protocolo de mutagénesis QuickChange de acuerdo con las recomendaciones del fabricante (Stratagene, La Jolla, CA).

El clon TOPO corregido se digirió con enzimas de restricción NdeI y XhoI de acuerdo con los protocolos recomendados del fabricante (Biolaboratorios de Nueva Inglaterra, Beverly, MA) y se purificaron en geles de TAE-agarosa AL 0,8% mediante el uso del kit de extracción en gel de Qiagen (Qiagen, Valencia, CA). Los vectores pET28 y pET30 se prepararon mediante la digestión con enzimas de restricción NdeI y XhoI seguidos del tratamiento con fosfatasa alcalina sérica y purificación en geles de TAE-agarosa al 0,8% mediante el uso del kit de extracción en gel de Qiagen (Qiagen, Valencia, CA).

Los vectores e inserciones digeridos se ligaron mediante el uso del kit de ligadura rápida NEB (Beverly, MA). Aproximadamente 50 ng de la inserción tratada, 100 ng del vector tratado (proporción de inserción a vector 3 a 1), 5 U de T4 ADN ligasa y 1× de tampón de ligación se incubaron durante 5 minutos a temperatura ambiente. La mezcla de ligación se transformó en células químicamente competentes TOP10F' (Invitrogen, Carlsbad, CA). Las células se dejaron recuperar en 0,25 mL de medio SOC a temperatura ambiente durante 1 hora a 37°C con agitación a 225 rpm. Las células se colocaron en placas LB que contenían kanamicina (50 µg/mL). El ADN plásmido se purificó a partir de los transformantes resultantes mediante el uso del kit spin miniprep de Qiagen (Qiagen, Valencia, CA) y se analizó para detectar las inserciones correctas mediante enzimas de restricción con NdeI y XhoI.

### 30 Expresión y análisis génicos

El ADN plásmido se transformó en hospedadores de expresión de *E. Coli* BL21(DE3) (Novagen, Madison, WI). Los cultivos se desarrollaron y los plásmidos se aislaron mediante el uso del kit miniprep de Qiagen (Qiagen, Valencia, CA) y se analizaron mediante enzimas de restricción para confirmar la identidad.

La inducción en BL21(DE3) se realizó inicialmente con 4 D-HPG de *P. stutzeri* en los vectores pET 28 (marcado con histidina) y pET 30 (sin marcar). Se realizó un estudio temporal con cultivos desarrollados en 250 mL de kanamicina que contenía LB (50 mg/L) a una OD<sub>600</sub> de entre 0,5 y 0,6, inducidos con 100 mM de IPTG (isopropil tiogalactósido) y muestreados a las 0 y 3 horas después de la inducción. Un volumen adecuado de células procedentes de la hora 0 y la hora 3 se volvió a colocar en suspensión en 40 µL de tampón de dodecil sulfato de sodio que contenía 2-mercaptoetanol y se calentó a 95°C durante 10 minutos y se enfrió. Se analizaron por SDS-PAGE alícuotas de estas muestras de la proteína celular completa mediante el uso de un gel en gradiente del 4 al 15%.

Se prepararon también extractos celulares procedentes de los cultivos de 3 horas mediante la colocación en suspensión de los sedimentos celulares procedentes de 5 ml de cultivo en 0,625 ml de reactivo BugBuster™ de Novagen (Novagen, Madison, WI) que contenía 0,625 µl de benzonasa nucleasa y 3 µL de mezcla de inhibidor de proteasa conjunto #3 (Calbiochem-Novabiochem Corp., San Diego, CA) a temperatura ambiente durante 20 minutos con leve agitación y centrifugación a 16.000 x g para eliminar el residuo celular. Los sobrenadantes (extractos celulares) se cargaron en geles con gradiente del 4 al 15% para el análisis de las proteínas celulares solubles. Cuando fue requerido, la proteína se purificó mediante el uso de cartuchos His-Bind 900 de acuerdo con los protocolos del fabricante (Novagen, Madison, WI) y se desaló para extraer el imidazol mediante el uso de columnas PD-10 (G25 Sephadex, GE Healthcare, Piscataway, NJ).

### 50 Organismos con D-fenilglicina aminotransferasa (DPGAT)

Los organismos del género *Pseudomonas* y géneros similares, con una D-fenilglicina aminotransferasa estereoinversora (también llamada D-4-hidroxifenilglicina aminotransferasa) se aíslan de la siguiente forma. Las muestras de suelo se incuban en placas petri con el siguiente medio: (por litro) 15 g de Agar, 3,4 g de KH<sub>2</sub>PO<sub>4</sub>, 3,55 g de Na<sub>2</sub>HPO<sub>4</sub>, 0,2 g de MgSO<sub>4</sub>·7H<sub>2</sub>O, 8 mg de CaCl<sub>2</sub>·2H<sub>2</sub>O, 10 mg de extracto de levadura, 1 ml de solución de elementos de traza 1000×, 1 g de D-fenilglicina (D-4-hidroxifenilglicina).

Los aislados se probaron mediante PCR para detectar la presencia de una aminotransferasa estereoinversora (cebadores diseñados a partir de las D-fenilglicina aminotransferasas conocidas) o se enriquecieron adicionalmente para detectar la presencia de una aminotransferasa estereoinversora de la siguiente forma: los aislados de las placas se cultivaron en medio líquido como antecede menos el agar a 30°C con agitación a una OD<sub>600</sub> de aproximadamente 1,0. Las células se cosecharon mediante centrifugación y se lavaron dos veces con NaCl al 0,85%. Una muestra de 10 mg (peso húmedo) se suspendió en 1 ml de tampón de fosfato de potasio (pH 7,0) y 5 mM de D-fenilglicina (o D-4-hidroxifenilglicina). Se agregaron 15 mM de ácido acético (aminooxi) neutralizado para duplicar las muestras preparadas en la forma descrita anteriormente. El consumo de D-fenilglicina (o D-4-hidroxiglicina) se midió mediante HPLC. Se seleccionaron los aislados capaces de degradar la D-fenilglicina (o D-4-hidroxifenilglicina), pero que lo hacían en una proporción menor en la presencia de ácido acético (aminooxi) para un mayor análisis. Los aislados se probaron mediante PCR para verificar la presencia de aminotransferasa estereoinversora (cebadores designados a partir de las D-fenilglicina aminotransferasas conocidas).

La presencia de aminotransferasa estereoinversora se confirmó mediante el desarrollo de un cultivo en un medio líquido como se describe anteriormente, la cosecha de las células y la preparación de un extracto bruto libre de células (CFE) y la prueba para detectar la actividad enzimática de D-fenilglicina aminotransferasa (o D-4-hidroxifenilglicina aminotransferasa). Se agregó CFE a la mezcla de reacción con las siguientes concentraciones finales: CAPS 0,1 M (pH 9,5), 60 mM de L-glutamato (sal de sodio), 5 mM de benzoilformiato (o 4-hidroxibenzoato) y 50 µM de PLP. La reacción inversa se midió mediante la adición de CFE a la mezcla de reacción con las siguientes concentraciones: 50 mM de fosfato de potasio (pH 7,0), 60 mM de D-fenilglicina (o D-4-hidroxifenilglicina), 5 mM de α-cetoglutarato, 50 µM de PLP. Los ensayos se incubaron a 35°C y se tomaron alícuotas en intervalos de tiempo y se detuvieron mediante ebullición durante 2 minutos. El producto se cuantificará mediante el procedimiento de HLPC de Gil-Av, E., Tishbee, A., Hare, P. E., *Resolution of underivatized amino acids by reversed phase chromatography. J. Am. Chem. Soc., 102: 5115-5117 (1980)* o mediante los procedimientos descritos en el ejemplo 1 (medición de formación de glutamato).

Como una alternativa para los procedimientos basados en la PCR, la D-fenilglicina aminotransferasa estereoinversora se purifica a partir de la bacteria aislada mediante técnicas de purificación de proteínas convencionales que incluyen el fraccionamiento de sulfato de amonio y cromatografía de columna convencional. Una vez que la proteína ha sido purificada en un grado razonable, se utilizan técnicas de microsecuenciado de péptidos o secuenciado de aminoácidos tipo Edman convencional (véase [golgi.harvard.edu/microchem/](http://golgi.harvard.edu/microchem/) para descripciones de los protocolos y equipamientos utilizados para este tipo de trabajo). Los cebadores degenerados se diseñan con base en la secuencia disponible del pariente más cercano conocido de la fuente de proteínas. A continuación se lleva a cabo la PCR degenerada y el paseo genómico de acuerdo con los protocolos establecidos para aislar la secuencia que codifica la D-fenilglicina aminotransferasa estereoinversora.

#### Producción de monatina DPGAT

Las D-hidroxifenilglicina aminotransferasas, como se describe en (1) y (2) anteriormente, se utilizan en extractos libres de proteínas brutos o purificados como se describe en (1) anteriormente. Las tirosina (aromática) aminotransferasas de *S. melloti* y *R. sphaeroides* se prepararon como se describe en el ejemplo 1 de WO 03/091396 A2. La ProA aldolasa de *Comamonas testosteroni* se preparó como se describe en el ejemplo 4 de WO 03/091396 A2. La totalidad de los ensayos de proteínas se realizó mediante el uso de un ensayo de proteínas Bio-Rad de acuerdo con los protocolos del fabricante (Hercules, CA).

Las mezclas de reacción (1 mL de volumen, realizadas por duplicado) contenían 100 mM de tampón de fosfato de potasio (pH 8), 2 mM de MgCl<sub>2</sub>, 0,05 mM de piridoxal 5'-fosfato (PLP), 200 mM de piruvato de sodio, 5 mM de α-cetoglutarato de sodio, aproximadamente 280 µg/mL de TatA de *S. melloti* proporcionado en un extracto celular, 100 µL/mL de extracto celular de D-hidroxifenilglicina aminotransferasa o 1 mg/mL de D-hidroxifenilglicina aminotransferasa purificada y aproximadamente 100 µg/mL de ProA aldolasa proporcionada como extracto celular. Se agregó triptófano sólido a una concentración de 10,2 mg/ml. Los controles negativos se configuraron sin D-hidroxifenilglicina aminotransferasa. Las muestras se incubaron a 30°C con leve agitación durante aproximadamente 1 hora o durante la noche. Las muestras se centrifugaron para extraer el precipitado, se filtraron por jeringa y se almacenaron a -80°C antes del análisis de la monatina mediante el uso del procedimiento de LC/MS/MS descrito en el ejemplo 1.

Las D-hidroxifenilglicina aminotransferasas con una mayor actividad para la producción de monatina se preparan mediante técnicas de mutagénesis conocidas por los expertos en la técnica, que incluyen: PCR mutagénica, pasaje a través de cepas mutagénicas o mutagénesis dirigida al sitio. Las D-hidroxifenilglicina aminotransferasas mejoradas se seleccionan por el crecimiento en un medio mínimo con R,R-monatina como la fuente de nitrógeno. Inicialmente la selección se basó en el crecimiento pero como se seleccionaron las aminotransferasas mejoradas el análisis se basó en la velocidad de crecimiento. Es decir, se cultivaron las células con versiones mutadas del gen y el gen se expresó en un medio mínimo con R,R-monatina como la fuente de nitrógeno. Las velocidades de crecimiento celular con las versiones mutadas del gen se compararon con la versión sin mutar. Se seleccionaron las células con una velocidad de crecimiento más rápida y se analizó la aminotransferasa de forma adicional. La D-hidroxifenilglicina aminotransferasa se mutagenizó mediante técnicas disponibles como la PCR propensa a errores y el pasaje a través

de las cepas mutagénicas o mediante procedimientos para los cuales se ha obtenido una licencia como reestructuración de ADN y otras tecnologías de evolución dirigida.

#### Ensayo de DPGAT

5 Se cultivaron las células con D-hidroxifenilglicina aminotransferasa recombinante, la proteína se expresó y las células se lisaron como se describe en el ejemplo 17 o mediante protocolos estándar. La proteína se expresó en su receptor nativo mediante el uso de los procedimientos descritos (*Wiyakrutta, W., Meevootisom, V. A stereoinverting D-phenylglycine aminotransferase from Pseudomonas stutzeri ST-201: purification, characterization, and application for D-phenylglycine synthesis. J. Biotechnol., 55: 193-203 (1997)*).

10 Se agregó extracto libre de células a la mezcla de reacción con las siguientes concentraciones finales: 100 mM de fosfato de potasio (pH 7,0 a 8,5), 60 mM de D-fenilglicina (o D-4-hidroxifenilglicina), 5 mM de  $\alpha$ -cetoglutarato, 50  $\mu$ M de PLP. Los ensayos se incubaron a temperatura ambiente, se tomaron alícuotas en intervalos de tiempo y se detuvieron mediante la adición de un volumen igual de ácido fórmico. El producto (L-glutamato) se cuantificó mediante los procedimientos descritos en el ejemplo 1.

#### **Ejemplo de referencia 20**

15 Descubrimiento de un gen de D-metionina Aminotransferasa

##### Antecedentes

20 Se considera que la D-metionina-piruvato aminotransferasa (EC 2.6.1.41) es otro ejemplo, aunque raro, de una transaminasa estereoinversora. Esta enzima cataliza la conversión reversible de la D-metionina y piruvato a L-alanina y 4-metiltilio-2-oxobutanoato. El oxaloacetato, fenilpiruvato, 2-oxobutirato, 2-oxovalerato, 2-oxoheptanoato, glioxilato y oxoglutarato también puede servir como aceptantes amino.

25 Se piensa que la transaminación de D o L metionina es parte de una vía para la producción de etileno en plantas superiores (coliflor, tomatera, manzano, guisante, banano, maní) así como microorganismos en el suelo (*Escherichia coli*, *Pseudomonas pisi*, *Pseudomonas aeruginosa*, *Bacillus mycoides*, *Acinetobacter calcoaceticus*, *Aeromonas hydrophila* B12E, *Rhizobium trifolii* N2P7, *Penicillium digitatum*, *Saccharomyces cerevisiae*, *Corynebacterium* D7F). *D.C. Billington, B. T. Golding, y S. B. Primrose Biochem J. (1979) 182, 827-836*. En las bacterias, la L-metionina se utiliza típicamente como el sustrato en los estudios de producción de etileno y se piensa que las enzimas de amplia especificidad como TyrB o AspC de *E. coli* son las responsables de la transaminación. No obstante, *S. B. Primrose J. Gen. Microbiol. (1976), 95(1), 159-65* y *S. B. Primrose J. Gen. Microbiol. (1977), 98, 519-528* mostraron que la cepa SPA O de *E. coli* (colección de cultivos de la Universidad de Warwick) produjo casi tanto etileno a partir de D-metionina como a partir de L-metionina en cultivos en lote. Debido a que no se identificó ninguna D-aminotransferasa de amplia especificidad en *E. coli*, una posible explicación puede ser que la D-aminoácido deshidrogenasa de *E. coli* (codificada por el gen *dadA*) convierte a la D-metionina en 4-metiltilio-2-oxobutanoato. También es posible que exista una metionina racemasa en *E. coli*; no obstante, ninguna enzima de dicho tipo ha sido descrita en la bibliografía.

35 Contrariamente a *E. coli*, en las cabezuelas de coliflor (preparaciones de extracto mitocondrial) y en las semillas germinativas de maní, la producción de etileno fue mayor cuando se proporcionó D-metionina y piruvato al extracto enzimático en comparación con la L-metionina y piruvato (*L. W. Mapson, J. F. March, and D. A. Wardale Biochem J (1969) 115, 653-661; J. I. Durham, P. W. Morgan, J. M. Prescott y d C. M. Lyman Phytochemistry (1973) 12, 2123-2126*). Por consiguiente, la posibilidad de una combinación de metionina racemasa y una L-aminotransferasa no se encuentra respaldada por los datos. La actividad de la deshidrogenasa se descartó por diálisis de los extractos celulares de coliflor, no se encontró ningún NAD en las mezclas de ensayo. Se descartó la actividad de oxidasa ya que no se detectó consumo de oxígeno y no hubo ningún requisito de FAD. La D-metionina aminotransferasa de los tejidos de maní se purificó, mostró depender de PLP y mostró ser independiente de la actividad de la L-metionina aminotransferasa. Existe una posibilidad de que estas D-metionina—piruvato aminotransferasas produzcan efectivamente D-alanina como un subproducto (similar a las enzimas de *Bacillus* descritas en los ejemplos 13 y 14) y que las células contengan alanina racemasa para reciclar la D-alanina nuevamente a L-alanina (o un donante amino análogo). En cualquiera de los casos, el descubrimiento de la D-aminotransferasa de amplia especificidad en plantas más altas es ventajoso para el desarrollo de procesos que producen R,R monatina o monatina S,R.

##### Resumen experimental

50 La D-metionina aminotransferasa se purifica parcialmente a partir de las cabezuelas de coliflor y de los embriones de maní germinativos mediante el uso de los protocolos de cromatografía estándar y del sistema Pharmacia AKTA Explorer. Las secuencias de proteínas de las proteínas homólogas se determinan mediante técnicas de huella genética LC/MS/MS y la búsqueda en la base de datos realizada por el centro Harvard Microchemistry. Las regiones codificadoras de los genes de la planta se clonan a partir de la biblioteca de ADNc mediante el uso de los protocolos de PCR estándar o mediante la síntesis del gen como se describe en el ejemplo 19.



De forma alternativa, las bibliotecas de expresión de ADNc se construyen (Stratagene, La Jolla, CA) a partir de tejido de coliflor o semillas de maní desarrolladas en la presencia de D-metionina (y con producción de etileno). Las bibliotecas se transforman en auxótrofos de metionina de *E. coli* a partir del centro de linaje genético (Yale) de *E. coli* como las cepas RC519 o AB1931. Los plásmidos de las cepas capaces de crecer en un medio mínimo que contiene D-metionina contienen la región codificadora de interés (véase el ejemplo 15, una técnica de análisis análoga).

Una vez que las regiones codificadoras de interés se obtienen y se expresan en una cepa de laboratorio estándar de *E. coli*, los productos génicos resultantes pueden utilizarse en ensayos para producir monatina R,R como se describe en el ejemplo 19 en lugar de la D-hidroxifenilglicina aminotransferasa, con la excepción del pH que es de 7,5 (el pH óptimo para la aminotransferasa). Si la D-metionina aminotransferasa tiene un requisito estricto de sustratos donantes de D-aminoácido, la enzima puede utilizarse para preparar monatina R,R como se describe en los ejemplos 13 y 14. Los genes pueden mutagenizarse y analizarse para detectar el aumento en la actividad como se describe en el ejemplo 19.

#### Procedimientos

##### 15 Aislamiento de coliflor

Se extrajeron 400 gramos de cabezuelas de coliflor recién recogidas con 400 mL de una solución de sacarosa/tampón a 4°C (0,4 M de sacarosa y 0,1 M de tampón de fosfato de sodio, pH 7,4) mediante la alternación de remojo y mezclado con el uso de una mezcladora. El residuo celular se extrajo por filtración con estopilla y la solución resultante se centrifugó a 40,000×g durante 30 minutos a 4°C. El material sólido (que contenía organelos mitocondriales) se volvió a suspender en 20 mL de tampón de fosfato de sodio 10 mM, pH 7,4 y las enzimas se extrajeron con 200 mL de acetona fría (-30°C). La suspensión se volvió a centrifugar y el precipitado se secó mediante el uso de Savant Speed Vac. El material sólido se disolvió en 10 mM de tampón de fosfato de sodio, pH 7,4 y la acetona residual se extrajo mediante el uso de una columna PD-10 (GE Healthcare, Piscataway, NJ).

La actividad de la aminotransferasa se analizó mediante la incubación de la preparación de enzimas con 5 mM de D-metionina, 1 mM de piruvato, 0,05 mM de PLP y 2 mM de EDTA en 0,1 M de tampón de fosfato de sodio, pH 7,4. Los ensayos se realizaron a 25°C durante 16 horas. El 4-Metilíto-2-oxobutanoato se midió a través de la formación del derivado de 2,4-dinitrofenilhidrazona, mediante el uso de LC/MS (m/z de 328) y metodologías similares descritas en el ejemplo 1. Se preparó una solución (p/v) al 0,4% de 2,4-dinitrofenilhidrazina en 2M de ácido sulfúrico y se agregó la mitad del volumen a la mezcla de ensayo después de la incubación. La mezcla se mezcló con leve agitación a 30°C durante 30 minutos y el precipitado se recolectó mediante centrifugación y se analizó mediante LC/MS. Las fracciones de proteínas separadas mediante técnicas cromatográficas estándar se analizaron para verificar la actividad en una forma similar, pero la alanina coproducto se midió por la técnica de derivatización post columna con OPA descrita en el ejemplo 1.

##### Aislamiento de maní (*Arachia hypogea* L. cv. Starr)

La enzima D-metionina aminotransferasa del homogenado de embriones de maní germinativo (menos las cotiledóneas) se purifica de conformidad con el procedimiento de *J. I. Durham, P. W. Morgan, J. M. Prescott y C. M. Lyman Phytochemistry (1973) 12, 2123-2126*. Los agentes reductores se utilizan durante la preparación de extractos brutos para estabilizar las enzimas y el residuo celular se elimina mediante centrifugación a 33,000×g. La fracción de un 35 a un 50% de sulfato de amonio se purifica de forma adicional mediante incubación a baja temperatura y mediante la eliminación de las proteínas en el precipitado. El sobrenadante se fracciona de forma adicional mediante el uso de acetona. Los grupos activos a continuación se purifican de forma adicional mediante cromatografía de filtración en gel (Sephadex 200, GE Healthcare, Piscataway, NJ).

A medida que las fracciones de proteínas se enriquecen con la proteína transaminasa, se utiliza la electroforesis en gel 2D para separar la enzima de interés para el microsecuenciado. Después de la elucidación de los homólogos, las regiones codificadoras en las secuencias de la planta depositadas en NCBI, la proteína D-aminotransferasa se produce recombinantemente en *Escherichia coli* mediante el uso de técnicas de biología molecular estándar. Se espera que los extractos celulares de las cabezuelas de coliflor o de las semillas de maní o de las enzimas homólogas producidas recombinantemente puedan utilizarse en la producción de monatina R,R como se describe en el ejemplo 19 (si es una transaminasa estereoinversora) o en los ejemplos 13 y 14 (si es una D-aminotransferasa de amplia especificidad).

#### **Ejemplo de referencia 21**

##### L-alanina aminotransferasa/ alanina racemasa/ D-alanina aminotransferasa

Las figuras 4, 6 y 8 ilustran las vías biosintéticas para producir monatina R,R enriquecida estereoisoméricamente a partir de L-triptófano mediante el uso de una L-aminoácido aminotransferasa (como L-alanina-aminotransferasa y/o L-triptófano-aminotransferasa), una aldolasa R-específica, una alanina racemasa y una D-alanina aminotransferasa.

En el ejemplo 16 se describe una aminotransferasa triptófano-específica. De forma alternativa, las tirosina (aromática) aminotransferasas de *S. meliloti* y *R. sphaeroides* se prepararon como se describe en el ejemplo 1 de WO 03/091396 A2. La ProA aldolasa de *Comamonas testosteroni* se preparó como se describe en el Ejemplo 4 de WO 03/091396 A2. La totalidad de los ensayos de proteínas se realizó mediante el uso de un ensayo de proteínas Bio-Rad de acuerdo con los protocolos del fabricante (Bio-Rad, Hercules, CA). La alanina racemasa se compró a Sigma-Aldrich (St. Louis, MO) (catálogo número A8936). La D-alanina aminotransferasa se compró a BioCatalytics (catálogo número AT-103) (BioCatalytics, Pasadena, CA).

Las L-alanina aminotransferasas se distribuyen ampliamente en eucariotas, bacterias y archaea. Se identificó que los siguientes organismos (con base en la homología de secuencia) contenían una L-alanina aminotransferasa (E.C. 2.6.1.2): *Arabidopsis thaliana*, *Ashbya gossypii*, *Azotobacter vinelandii*, *Bifidobacterium longum*, *Caenorhabditis elegans*, *Candida albicans*, *Candida glabrata*, *Chlamydomonas reinhardtii*, *Cryptococcus neoformans*, *Debaryomyces hansenii*, *Homo sapiens*, *Hordeum vulgare*, *Kluyveromyces lactis*, *Magnaporthe grisea*, *Medicago truncatula*, *Mus musculus*, *Neurospora crassa*, *Oryza sativa*, *Phanerochaete chrysosporium*, *Pinus taeda*, *Pseudomonas putida*, *Pyrococcus abyssi*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Rattus norvegicus*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Takifugu rubripes*, *Trypanosoma cruzi*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Vibrio vulnificus*, *Yarrowia lipolytica* y *Zea mays*. Además, muchas aminotransferasas presentan un bajo nivel de actividad de alanina aminotransferasa y dado los altos niveles de L-glutamato y piruvato pueden convertirse en L-alanina y  $\alpha$ -cetoglutarato. La enzima con baja actividad se mejora con técnicas de mutagénesis estándar como la PCR propensa a errores y el pasaje a través de las cepas mutagénicas o mediante técnicas de evolución dirigida. El gen para una L-alanina aminotransferasa se clona mediante el uso de cebadores de diseño disponibles públicamente y mediante el uso de técnicas estándar para amplificar, clonar, expresar y purificar el gen/enzima.

Las mezclas de reacción (1 mL de volumen, realizadas por duplicado) contenían 100 mM de tampón de fosfato de potasio (pH 8), 2 mM de  $MgCl_2$ , 0,05 mM de piridoxal 5'-fosfato (PLP), 200 mM de piruvato de sodio, 5 mM de  $\alpha$ -cetoglutarato de sodio, aproximadamente 280  $\mu g/mL$  de TatA de *S. meliloti* proporcionado en un extracto celular (u otra aminotransferasa L-triptófano específica), 100  $\mu L/mL$  de extracto celular de alanina racemasa o 1 mg/mL de alanina racemasa purificada, aproximadamente 280  $\mu g/mL$  de D-alanina aminotransferasa proporcionada en un extracto celular y aproximadamente 100  $\mu g/mL$  de ProA aldolasa proporcionada como extracto celular. Se agregó triptófano sólido a una concentración de 10,2 mg/ml. Los controles negativos se configuraron sin alanina racemasa. Las muestras se incubaron a 30°C con leve agitación durante aproximadamente 1 hora o durante la noche. Las muestras se centrifugaron para extraer el precipitado, se filtraron por jeringa y se almacenaron a -80°C antes del análisis de la monatina mediante el uso del procedimiento de LC/MS/MS descrito en el ejemplo 1. En estas mezclas de reacción, si la L-triptófano aminotransferasa puede utilizar  $\alpha$ -cetoglutarato pero no piruvato, como aceptante amino, puede agregarse una L-alanina aminotransferasa que convierte el L-glutamato y piruvato a L-alanina y  $\alpha$ -cetoglutarato, como se muestra en la figura 6.

## Ejemplo 22

Subclonación de genes que codifican las aldolasas de la SEQ ID NO: 276, 244, 218, 228, 162, 50, 74, 44 y 116

Los genes que codifican las aldolasas que tienen las secuencias de aminoácidos de la SEQ ID NO: 276, 244, 218, 228, 162, 50, 74, 44 y 116 se subclonaron en el vector de expresión pET28b (Novagen, Madison, WI) con marcadores His con terminales N para permitir la purificación. La SEQ ID NO:275 es la secuencia del gen que codifica la aldolasa que tiene la secuencia de aminoácidos de la SEQ ID NO:276. La SEQ ID NO:243 es la secuencia del gen que codifica la aldolasa que tiene la secuencia de aminoácidos de la SEQ ID NO:244. La SEQ ID NO:217 es la secuencia del gen que codifica la aldolasa que tiene la secuencia de aminoácidos de la SEQ ID NO:218. La SEQ ID NO:227 es la secuencia del gen que codifica la aldolasa que tiene la secuencia de aminoácidos de la SEQ ID NO:228. La SEQ ID NO:161 es la secuencia del gen que codifica la aldolasa que tiene la secuencia de aminoácidos de la SEQ ID NO:162. La SEQ ID NO:49 es la secuencia del gen que codifica la aldolasa que tiene la secuencia de aminoácidos de la SEQ ID NO:50. La SEQ ID NO:73 es la secuencia del gen que codifica la aldolasa que tiene la secuencia de aminoácidos de la SEQ ID NO:74. La SEQ ID NO:43 es la secuencia del gen que codifica la aldolasa que tiene la secuencia de aminoácidos de la SEQ ID NO:44. La SEQ ID NO:115 es la secuencia del gen que codifica la aldolasa que tiene la secuencia de aminoácidos de la SEQ ID NO:116.

Los cebadores utilizados para la clonación se muestran en la tabla 24.

Tabla 24

ADN de aldolasa SEQ ID NO.:	cebador 5'	cebador 3'
275	5'- ATAAGACATATGCCTATCGTTGTTACGAAG- 3' (SEQ ID NO: 339)	5'- ATAAGAGGATCCTTATTCTCGGGCAGCCGCT C-3' (SEQ ID NO: 340)
243	5'- ATAAGACATATGAACAGACCTGTGGTTGTC -3' (SEQ ID NO: 341)	5'- ATAAGAGGATCCTTACAGGTAAGGACCGA G-3' (SEQ ID NO: 342)
217	5'- ATAAGACATATGAGCGTGGTCATCCGAAAC -3' (SEQ ID NO: 343)	5'- ATAAGAGGATCCTTACTTCGCTTTGTTATAGG C-3' (SEQ ID NO: 344)
227	5'-ATAAGACATATGAACAAG (SEQ ID NO: 345)	5'-ATAAGAGGATCCTTACAA (SEQ ID NO: 346)
161	5'- ATAAGACATATGAGCGTGGTCGTACCCGG- 3' (SEQ ID NO: 347)	5'- ATAAGAGGATCCTTAGCCGTTTTTCCCGTCGG TG-3' (SEQ ID NO: 348)
49	5'- AGAAGACATATGATGAGCATCGTCGTCCA GAAC-3' (SEQ ID NO: 349)	5'- AGAAGAGGATCCTCAGACATATTCAGGCCCT TG-3' (SEQ ID NO: 350)
73	5'- AGAAGACATATGATGAGCGTGGTCATCAC C-3' (SEQ ID NO: 351)	5'- ACAACAGGATCCCTATTTCTTCTCCGGCGTTT C-3' (SEQ ID NO: 352)
43	5'-ATAATACATATGAGCGTC (SEQ ID NO: 353)	5'-ATAATAGGATCCTTAGAC (SEQ ID NO: 354)
115	5'-AGAAGACATATGATGTCG (SEQ ID NO: 355)	5'-AGAAGAGGATCCTCAGAT (SEQ ID NO: 356)

Los genes que codifican las aldolasas anteriores se amplificaron mediante PCR y se digirieron con enzimas adecuadas (Nde I y BamH I) y se purificaron en gel (kit de extracción en gel QiAquick® (Qiagen, Valencia, CA)). Los digestos se ligaron individualmente a pET28 que había sido digerido con Nde I y BamH I y se purificaron en gel. La ligación se transformó en células TOP10 (Invitrogen, Carlsbad, CA). El ADN miniprep de las colonias individuales se analizó para verificar la presencia de inserciones mediante el análisis de tamaño con el uso de electroforesis en gel de agarosa. Los aislados con una inserción se sometieron a análisis de secuencias de ADN (Agencourt, Beverly, MA).

#### 10 Purificación de aldolasas

Los clones de aldolasa confirmados se transformaron en BL21(DE3) o BL21(DE3) pLysS. La inducción se llevó a cabo durante la noche en caldo Terrific a 30°C. Los cultivos desarrollados durante la noche con el antibiótico adecuado se diluyeron en un medio fresco (típicamente 1:100) y se cultivaron a una OD<sub>600</sub> de aproximadamente 0,6 con aireación a 37°C. Los cultivos después se indujeron con 1 mM de IPTG y se cambiaron a 30°C (con aireación) y la incubación continuó durante la noche. Las células se cosecharon mediante centrifugación. El sedimento celular se sometió típicamente a un ciclo de congelado y descongelado para ayudar a la lisis celular. El sedimento celular se lisó en BugBuster y Benzonasa Nucleasa (Novagen, Madison, WI) (de acuerdo con el protocolo del fabricante). El residuo celular se eliminó mediante centrifugación. El extracto de proteína bruto se aplicó a una columna HIS-Bind de 10 mg de capacidad (Novagen, Madison, WI) que había sido preparada de acuerdo con el protocolo del fabricante. La columna se lavó y la proteína se eluyó de acuerdo con el protocolo del fabricante. La proteína purificada se desaló con columnas PD-10 (GE Healthcare, Piscataway, NJ) y se eluyó en 50 mM de tampón de fosfato de potasio, pH 7,5, que contenía 4 mM de MgCl<sub>2</sub>, 200 mM de NaCl. La proteína se concentró con concentradores de centrifugación Amicon (5000 MW de corte) (Millipore, Billerica, MA). Después de la concentración se notó que las aldolasas de la SEQ ID NO: 44, SEQ ID NO: 28 y SEQ ID NO: 276 mostraron un cierto nivel de

precipitación, a pesar de que la actividad aún era bastante alta. Las proteínas se almacenaron a  $-80^{\circ}\text{C}$  hasta su ensayo.

5 Los ensayos de proteínas se realizaron mediante el uso del kit Pierce BCA (Pierce, Rockford, IL) y el protocolo de placa de microvaloración con albúmina de suero bovino ("BSA") como la proteína estándar. Se utilizó el sistema de electroforesis Experion Pro260 (Bio-Rad, Hercules, CA) o SDS-PAGE para estimar el porcentaje de aldolasa en la muestra purificada y para evaluar los niveles de expresión en el extracto celular soluble y en la proteína total.

#### Prueba de las aldolasas purificadas

10 Las aldolasas purificadas se probaron para verificar su capacidad de producir monatina R,R a partir de D-triptófano y se compararon con las aldolasas de la SEQ ID NO:28 y la SEQ ID NO:54 preparadas de la misma forma. Los ensayos se realizaron en tubos de microcentrifugadora (por duplicado) con proteína purificada, mediante el uso de la misma concentración de enzima por ensayo (50  $\mu\text{g}/\text{mL}$ ) con la excepción de la SEQ ID NO:244, para la cual se utilizó 25  $\mu\text{g}/\text{mL}$ . La SEQ ID NO:243 no se expresó bien y proporcionó menores cantidades de proteína purificada. Se utilizaron 2 mg/mL de AT-103 de BioCatalytics (BioCatalytics, Pasadena, CA) como la D-aminotransferasa. Cada 1 mL de mezcla de reacción se agregó lo siguiente: aldolasa, 4 mM de  $\text{MgCl}_2$ , 50 mM de D-triptófano, D-aminotransferasa, 200 mM de piruvato de sodio, 100 mM de tampón de fosfato de potasio, pH 7,5 y 0,05 mM de PLP. Las muestras se incubaron a  $30^{\circ}\text{C}$ . Se tomaron muestras a los 30 minutos, a la hora, a las 3 horas y durante la noche (19 horas). La tabla 25 muestra los resultados promediados de la monatina total producida en cada intervalo de tiempo y el % de monatina R,R producido, como se determina por las áreas de los picos de la fase inversa. En la columna 3 de la tabla 25, se realizó un análisis de derivatización adicional de la FDAA por LC/MS/MS como se describe en el ejemplo 1, para algunas de las reacciones y dichos resultados se muestran en el paréntesis.

15

20

**Tabla 25: Monatina total producida a partir de D-triptófano y % de R,R**

Aldolasa (horas)	Monatina total (ppm)	% de monatina R,R
SEQ ID NO: 28 (0,5)	16	99,1
SEQ ID NO: 28 (1)	53,2	99,2 (99,0)
SEQ ID NO: 28 (3)	207,8	98,6 (98,1)
SEQ ID NO: 28 (19)	544,9	95,3 (93,2)
SEQ ID NO: 44 (0,5)	46,2	88,0
SEQ ID NO: 44 (1)	92,5	86,8
SEQ ID NO: 44 (3)	319,7	76,4
SEQ ID NO: 44 (19)	762,9	67,1
SEQ ID NO: 54 (0,5)	35,3	96,2
SEQ ID NO: 54 (1)	82,7	96,1
SEQ ID NO: 54 (3)	280,1	92,9
SEQ ID NO: 54 (19)	715,3	77,1
SEQ ID NO: 74 (0,5)	51,1	92,6
SEQ ID NO: 74 (1)	89,3	94,3
SEQ ID NO: 74 (3)	269,5	89,9
SEQ ID NO: 74(19)	701,9	76,2
SEQ ID NO: 50 (0,5)	55,9	96,7
SEQ ID NO: 50 (1)	96,5	96,2
SEQ ID NO: 50 (3)	272,2	95,6
SEQ ID NO: 50 (19)	645,8	88,5

ES 2 383 767 T3

SEQ ID NO: 162 (0,5)	37,3	95,7
SEQ ID NO: 162 (1)	75,0	97,1
SEQ ID NO: 162 (3)	261,0	95,9
SEQ ID NO: 162 (19)	633,1	87,0
SEQ ID NO: 276 (0,5)	37,8	98,8
SEQ ID NO: 276 (1)	71,2	99,3 (99,5)
SEQ ID NO: 276 (3)	245,2	99,0 (99,0)
SEQ ID NO: 276 (19)	585,4	96,7 (96,1)
SEQ ID NO: 244 (0,5)	30,2	97,7
SEQ ID NO: 244 (1)	46,4	98,3 (99,2)
SEQ ID NO: 244 (3)	165	98,4 (98,7)
SEQ ID NO: 244 (19)	572,5	95,6 (93,7)
SEQ ID NO: 228 (0,5)	52	95,0
SEQ ID NO: 228 (1)	81,7	96,5
SEQ ID NO: 228 (3)	251	95,9
SEQ ID NO: 228 (19)	723	87,1
sin aldolasa (0,5)	0	
sin aldolasa (1)	0	
sin aldolasa (3)	0,6	58,3
sin aldolasa (19)	6,5	61,5

La aldolasa de la SEQ ID NO:276 mantuvo un alto nivel de actividad, así como la mayor estereoespecificidad para la producción de monatina R,R. El almacenamiento de esta enzima en un tampón con la omisión de cloruro de sodio parece reducir el nivel de precipitación notado anteriormente. La concentración de magnesio en el tampón de almacenamiento no parece tener un impacto sobre el nivel de precipitación.

5

Las aldolasas de la SEQ ID NO:276, SEQ ID NO: 28 y SEQ ID NO: 244, todas demostraron una alta actividad y una alta estereoselectividad para la producción de monatina R,R. Estas tres enzimas se prepararon como se describe en el ejemplo 23 y se ensayaron anaerobiamente como se describe en el ejemplo 24, mediante el uso de recipientes de suero de 10 mL. Se realizaron ensayos de 7 mL mediante el uso de 0,05 mg/mL de cada aldolasa (purificada) y 2 mg/mL de D-aminotransferasa purificada de *B. sphaericus* preparada como se describe en el ejemplo 24. La actividad de cada aldolasa en la producción de monatina a partir de D-triptófano se comparó con la aldolasa HMG de *S. meliloti* preparada como se describe en el ejemplo 27. Se estimó la monatina total mediante el uso del procedimiento LC-OPA descrito en el ejemplo 1. El porcentaje de monatina R,R formada se determinó mediante el uso del procedimiento de derivatización de la FDAA descrito en el ejemplo 1. Los resultados se muestran a continuación en la tabla 26.

10

15

Tabla 26

Aldolasa	Tiempo (horas)	Monatina (g/L)	% R,R formada
<i>S. meliloti</i>	23	3,9	82,0
SEQ ID NO: 28	23	4,0	84,6
SEQ ID NO: 276	23	4,0	95,7

SEQ ID NO: 244	23	3,7	88,8
<i>S. meliloti</i>	47	4,5	76,2
SEQ ID NO: 28	47	4,3	84,6
SEQ ID NO: 276	47	4,3	93,2
SEQ ID NO: 244	47	4,5	85,2

Estos resultados demuestran que la aldolasa de la SEQ ID NO: 276 también produce niveles altos de monatina R,R en reacciones anaerobias de mayor volumen.

### Ejemplo 23

#### 5 Expresión y purificación de la aldolasa de la SEQ ID NO:276

La clonación del receptor BL21 (DE3)pLysS de *E. coli* que tiene el gen de la aldolasa enumerado como la SEQ ID NO:275 en el plásmido pET28b se describe anteriormente en el ejemplo 22.

10 La aldolasa de la SEQ ID NO:276 con un marcador de purificación HIS<sub>6</sub> amino-terminal se produjo mediante el uso del sistema Overnight Express II de Biosciences EMD (Novagen, Madison, WI) (soluciones 1 a 6) que contenía 50 µg/mL de kanamicina en matraces de agitación. Este sistema de expresión induce la expresión de los sistemas inducidos por IPTG sin la necesidad de controlar el crecimiento celular. Después de la inoculación de alícuotas de 200 mL del medio (en matraces de 1 L) de los cultivos líquidos o placas de la construcción de aldolasa, los cultivos se incubaron a 30°C durante la noche con agitación a 225 rpm. Cuando la OD<sub>600,n</sub>, alcanzó un mínimo de 6, las células se cosecharon mediante centrifugación y se lavaron una vez con tampón.

15 Para preparar el extracto libre de células que contenía la aldolasa, las células se suspendieron en 3 a 4 volúmenes de 100 mM de fosfato de potasio, pH 7,8 y después se mezclaron mediante el uso de un homogeneizador Microfluidics (Microfluidics, Newton, MA) (3 pasadas a 18.000 psi), con mantenimiento de la temperatura de la suspensión a menos de 15°C. De forma alternativa, el extracto libre de células se preparó mediante el uso del reactivo de extracción BugBuster® de EMD Biosciences (libre de amina primaria) (Novagen, Madison, WI) que  
20 contenía 1 µL/mL de Benzonasa® Nucleasa, 5 µL/mL de mezcla de inhibidor de proteasa conjunto II y 0,033 µL/mL de rLisozima™ de acuerdo con el protocolo del fabricante. Todas las etapas de purificación subsiguientes se llevaron a cabo a 4°C. La suspensión de células se centrifugó durante 20 a 30 minutos a 15.000-20.000×g para extraer los residuos celulares. Se aplicó una alícuota de 20 a 25 mL del extracto libre de células a una columna de 45 mL de la resina de flujo rápido GE Healthcare Chelating Sepharose™ (forma de níquel (II)) (GE Healthcare,  
25 Piscataway, NJ) que había sido equilibrada previamente con 100 mM de fosfato de potasio que contenía 200 mM de cloruro de sodio. Para generar níquel a partir de la resina, la resina se lavó con 150 mL de hexahidrato de sulfato de níquel (II) 200 mM y después con 150 mL de agua destilada. Después de cargar la muestra, la columna se lavó/eluyó con 150 mL del tampón de equilibración que contenía 25 mM de imidazol, 150 mL del tampón de equilibración que contenía 50 mM de imidazol y 150 mL del tampón de equilibración que contenía 500 mM de  
30 imidazol. La proteína marcada con HIS<sub>6</sub> se eluyó en el último lavado. El lavado con 500 mM de imidazol se concentró con dispositivos de centrifugación con filtro Millipore/Amicon Centricon Plus-70 (MWCO 10 kDa) (Millipore, Billerica, MA) a 15-20 mL de acuerdo con las instrucciones del fabricante. El imidazol y el cloruro de sodio se extrajeron mediante el pasaje a través de las columnas desechables de GE Healthcare PDIO (GE Healthcare, Piscataway, NJ) (2,5 mL de muestra por columna) equilibrados previamente con 100 mM de fosfato de potasio, pH  
35 7,8. La aldolasa purificada se eluyó con 3,5 mL por columna del mismo tampón.

Se determinó la concentración de proteínas de cada fracción mediante el uso del kit de ensayo Pierce BCA (Pierce, Rockford, IL) mediante el uso de BSA como el estándar de proteína. Se determinó la pureza de cada fracción y el nivel de expresión en el extracto libre de células mediante el uso de un sistema de chip microcapilar Bio-Rad Experion (Bio-Rad, Hercules, CA) o mediante el uso de geles con gradiente de SDS-poliacrilamida del 4 al 15% de  
40 Bio-Rad en un aparato celular Mini PROTEAN® 3 (Bio-Rad, Hercules, CA). La proteína se visualizó en los geles de poliactilamida mediante el uso de tinción de Coomassie G-250 de Bio-Rad Bio-Safe (Bio-Rad, Hercules, CA) y se destiñó con agua. Típicamente, este procedimiento produce aproximadamente 50 mg de enzima a partir de 400 mL de cultivo durante la noche que tiene de un 85 a un 95% de pureza, a juzgar por el software Experion. Se almacenaron alícuotas (1 a 5 mL) de la enzima purificada a -80°C hasta su uso. La preparación de la enzima en  
45 esta forma redujo el nivel de precipitación de la enzima señalada anteriormente. La presencia de magnesio en el tampón de almacenamiento no tuvo ningún efecto sobre el nivel de precipitación.

**Ejemplo 24**

Producción de monatina R,R mediante el uso de aldolasa de la SEQ ID NO:276: Optimización de las condiciones de reacción

5 La D-alanina aminotransferasa de *Bacillus sphaericus* (ATCC cepa 10208) clonada en el ejemplo 7 se purificó como la proteína marcada con HIS<sub>6</sub> como se describe a continuación mediante el uso del sistema EMD Biosciences Overnight Express System II (Novagen, Madison, WI) para su cultivo e inducción. El sistema EMD Biosciences Overnight Express System II (soluciones 1 a 6) (Novagen, Madison, WI) contenía 50 µg/mL de kanamicina en matraces de agitación. Este sistema de expresión induce la expresión de los sistemas inducidos por IPTG sin la necesidad de controlar el crecimiento celular. Después de la inoculación de alícuotas de 200 mL del medio (en matraces de 1 L) de los cultivos líquidos o placas de la construcción de aminotransferasa, los cultivos se incubaron a 10 30°C durante la noche con agitación a 225 rpm. Cuando la OD<sub>600</sub> nm alcanzó un mínimo de 6, las células se cosecharon mediante centrifugación y se lavaron una vez con tampón.

15 Para preparar el extracto libre de células que contenía la D-alanina aminotransferasa, las células se suspendieron en 3 a 4 volúmenes de 100 mM de fosfato de potasio, pH 7,8, que contenía 50 µM de piridoxal fosfato (PLP) y después se mezclaron mediante el uso de un homogeneizador Microfluidics (Microfluidics, Newton, MA) (3 pasadas a 18.000 psi), con mantenimiento de la temperatura de la suspensión a menos de 15°C. De forma alternativa, el extracto libre de células se preparó mediante el uso del reactivo de extracción BugBuster® de EMD Biosciences (libre de amina primaria) (Novagen, Madison, WI) que contenía 1 µL/mL de Benzonasa® Nucleasa, 5 µL/mL de mezcla de inhibidor de proteasa conjunto II y 0,033 µL/mL de rLisozima™ de acuerdo con el protocolo del fabricante. Todas las etapas de purificación subsiguientes se llevaron a cabo a 4°C. El extracto celular se centrifugó durante 20 a 30 minutos a 20 15.000×g para extraer los residuos celulares. Se aplicó una alícuota de 20 a 25 mL del extracto libre de células a una columna de 45 mL de la resina de flujo rápido GE Healthcare Chelating Sepharose™ (forma de níquel (II)) (GE Healthcare, Piscataway, NJ) que había sido equilibrada previamente con 100 mM de fosfato de potasio que contenía 200 mM de cloruro de sodio y 50 µM de PLP. Para generar níquel a partir de la resina, la resina se lavó con 150 mL de hexahidrato de sulfato de níquel (II) 200 mM y después con 150 mL de agua destilada. Después de cargar la muestra, la columna se lavó/eluyó con 150 mL del tampón de equilibración que contenía 25 mM de imidazol, 150 mL del tampón de equilibración que contenía 50 mM de imidazol y 150 mL del tampón de equilibración que contenía 500 mM de imidazol. La proteína marcada con HIS<sub>6</sub> se eluyó en el último lavado. El lavado con 500 mM de imidazol se concentró con dispositivos de centrifugación con filtro Millipore/Amicon Centricon Plus-70 (MWCO 10 kDa) (Millipore, Billerica, MA) a 15 a 20 mL de acuerdo con las instrucciones del fabricante. El imidazol y el cloruro de sodio se extrajeron mediante el pasaje a través de las columnas desechables de GE Healthcare PDIO (GE Healthcare, Piscataway, NJ) (2,5 mL de muestra por columna) equilibrados previamente con 100 mM de fosfato de potasio, pH 7,8, que contenía 0,5 µM de PLP. La aminotransferasa purificada se eluyó con 3,5 mL por columna del mismo tampón.

35 Se determinó la concentración de cada fracción mediante el uso del kit de ensayo Pierce BCA (Pierce, Rockford, IL) mediante el uso de BSA como el estándar de proteína. Se determinó la pureza de cada fracción y el nivel de expresión en el extracto libre de células mediante el uso de un sistema de chip microcapilar Bio-Rad Experion (Bio-Rad, Hercules, CA) o mediante el uso de geles con gradiente de SDS-poliacrilamida del 4 al 15% de Bio-Rad (Bio-Rad, Hercules, CA) en un aparato celular Mini PROTEAN® 3. La proteína se visualizó en geles de poliacrilamida mediante el uso de tinción de Coomassie G-250 de Bio-Rad Bio-Safe (Bio-Rad, Hercules, CA) y se destiñó con agua. Típicamente, este procedimiento produce aproximadamente 50 mg de enzima a partir de 200 mL de cultivo durante la noche que tiene de un 85 a un 90% de pureza, a juzgar por el software Experion y a partir del análisis de los geles SDS-PAGE. Se almacenaron alícuotas (1 a 5 mL) de la enzima purificada a -80°C hasta su uso.

45 La aldolasa de la SEQ ID NO:276 (clonada en el ejemplo 22) se purificó como la proteína marcada con HIS<sub>6</sub> como se describe en el ejemplo 23.

50 El cofactor metálico preferido para la aldolasa de la SEQ ID NO:276 se determinó mediante el análisis de una variedad de metales divalentes. Las reacciones se llevaron a cabo anaerobiamente en frascos de suero de 10 mL con 7 mL de los volúmenes finales. La solución a granel que consistía en 100 mM de fosfato de potasio (pH 7,8), 200 mM de piruvato de sodio, 0,05 mM de PLP y Tween 80 (v/v) al 0,01%, se preparó a un volumen final de 48,8 mL y se roció con nitrógeno durante 30 minutos. El D-triptófano (143 mg; concentración final de 100 mM) se colocó en siete recipientes de suero de 10 mL. A cada uno de los recipientes se agregaron 0,014 mL de una solución a granel de 2 M de un catión metálico divalente, preparado a partir de sal de cloruro (concentración final de 4 mM). Para el control negativo, se agregaron 0,014 mL de dH<sub>2</sub>O. Los recipientes de suero se taparon con septos de goma y se rociaron con nitrógeno mediante 16 a 18 agujas estériles. En condiciones anaerobias, se agregaron 5,625 mL de la solución a granel anaerobia a cada frasco de suero anaerobio. Subsiguientemente, la D-alanina aminotransferasa de *B. sphaericus* y la aldolasa de la SEQ ID NO:276 se agregaron anaerobiamente a cada frasco de suero a una concentración final de 2 mg/mL y 0,05 mg/mL, respectivamente. Las soluciones se incubaron a temperatura ambiente con mezclado suave durante 18 horas. La monatina final se analizó de conformidad con los procedimientos descritos en el ejemplo 1 mediante el uso del procedimiento de detección fluorescente de aminoácidos de cromatografía líquida post-columna. (Tabla 27).

Tabla 27

Cofactor metálico	Monatina final (mM) a las 18 horas
Ninguno (control negativo)	1,7
Magnesio	10,6
Manganeso	10,0
Cobalto	6,7
Zinc	4,9
Níquel	1,5
Calcio	0,7

Las condiciones de reacción para la aldolasa de la SEQ ID NO:276 se investigaron de forma adicional con un experimento factorial fraccional de dos niveles diseñado mediante el uso del software estadístico Design Expert 7.0.0 (Stat-Ease, Inc.; Minneapolis, MN). El diseño de análisis consistió en un único bloque de cinco factores a dos niveles con cuatro puntos centrales (20 ejecuciones en total). Los cinco factores para optimizarse fueron la concentración del cofactor metálico, el pH de reacción, la concentración de Tween® 80, la proporción de piruvato a triptófano y la concentración de aldolasa (Tabla 28).

Los tubos de polipropileno cónicos (14 mL) que contenían 143 mg de D-triptófano se desoxigenaron en una cámara anaerobia (Coy Laboratory Products, Inc; Grass Lake, MI) durante la noche. Las soluciones a granel de 2 M de MgCl<sub>2</sub>; 1 M de fosfato de potasio a un pH 7,0, 7,75 y 8,5; Tween® 80 (v/v) al 10%; 2 M de piruvato de sodio y 10 mM de PLP (piridoxal 5'-fosfato) se prepararon en agua desgasificada y se equilibraron en la cámara anaerobia durante la noche. Las soluciones a granel de D-alanina aminotransferasa de *B. sphaericus* y la aldolasa de la SEQ ID NO:276 se descongelaron inmediatamente sobre hielo en una cámara anaerobia. Las soluciones a granel se agregaron a los tubos cónicos de 14 mL que contenían el D-triptófano para obtener las concentraciones determinadas por el diseño estadístico (Tabla 28). El agua desgasificada se agregó a cada tubo para llevar el volumen final, junto con las adiciones de enzimas, a 7,0 mL. Los tubos se incubaron a temperatura ambiente en una cámara anaerobia con mezclado suave durante 24 horas. Se analizó la concentración de monatina y la pureza isomérica de acuerdo con los procedimientos descritos en el ejemplo 1, mediante el uso del procedimiento de detección fluorescente de aminoácidos de cromatografía líquida post-columna y del control de reacciones múltiples por LC/MS/MS para la determinación de la distribución de estereoisómeros de monatina en el procedimiento de reacciones in vitro e in vivo (procedimiento de derivatización de la FDAA).

Tabla 28

Ejecución #	estándar #	Bloque	Mg (mM)	pH	Tween® (%)	Pir:Trp	Aldolasa de la SEQ ID NO: 276 (mg/mL)
20	1	Bloque 1	5,00	7,75	0,01	2,00	0,05
8	2	Bloque 1	9,00	8,50	0,02	1,00	0,01
3	3	Bloque 1	1,00	8,50	0,00	1,00	0,01
16	4	Bloque 1	9,00	8,50	0,02	3,00	0,09
7	5	Bloque 1	1,00	8,50	0,02	1,00	0,09
12	6	Bloque 1	9,00	8,50	0,00	3,00	0,01
6	7	Bloque 1	9,00	7,00	0,02	1,00	0,09
2	8	Bloque 1	9,00	7,00	0,00	1,00	0,01
15	9	Bloque 1	1,00	8,50	0,02	3,00	0,01



Ejecución #	estándar #	Bloque	Mg (mM)	pH	Tween® (%)	Pir:Trp	Aldolasa de la SEQ ID NO: 276 (mg/mL)
4	10	Bloque 1	9,00	8,50	0,00	1,00	0,09
5	11	Bloque 1	1,00	7,00	0,02	1,00	0,01
1	12	Bloque 1	1,00	7,00	0,00	1,00	0,09
13	13	Bloque 1	1,00	7,00	0,02	3,00	0,09
14	14	Bloque 1	9,00	7,00	0,02	3,00	0,01
17	15	Bloque 1	5,00	7,75	0,01	2,00	0,05
11	16	Bloque 1	1,00	8,50	0,00	3,00	0,09
18	17	Bloque 1	5,00	7,75	0,01	2,00	0,05
9	18	Bloque 1	1,00	7,00	0,00	3,00	0,01
19	19	Bloque 1	5,00	7,75	0,01	2,00	0,05
10	20	Bloque 1	9,00	7,00	0,00	3,00	0,09

- El análisis estadístico de los datos indicó que el pH de reacción, la proporción piruvato:triptófano y la concentración de aldolasa fueron factores significativos que afectan la valoración de monatina, la pureza isomérica y la producción de carbono. Se generó una gráfica de conveniencia mediante el uso del software Design Expert en el cual se variaron los factores para maximizar los objetivos de una mayor valoración de monatina y una mayor pureza isomérica en las condiciones de exceso de piruvato. Las condiciones de reacción indicadas como óptimas fueron 1 mM de MgCl<sub>2</sub>, pH>8,0, Tween® 80 al 0,01% (v/v) y 0,01 mg/mL de aldolasa de la SEQ ID NO:276. Esta es una reducción de 5 veces en la cantidad típica de aldolasa utilizada, así como una reducción de 4 veces en la cantidad de metal divalente típicamente utilizado.
- 5
- 10 Se realizaron experimentos adicionales para determinar el intervalo de pH óptimo para el proceso de reacción. Las soluciones a granel de tampón EPPS 1M se prepararon en incrementos de unidades de pH de 0,2 entre un pH de 7,0 y 9,0. Estas soluciones se desgasificaron y se equilibraron en la cámara anaerobia durante la noche. Los tubos de polipropileno (14 mL) que contenían 143 mg de D-triptófano se desoxigenaron en una cámara anaerobia durante la noche. Las soluciones a granel de 2 M de MgCl<sub>2</sub>, Tween® 80 al 10% (v/v), 2 M de piruvato de sodio y 10 mM de PLP se prepararon en agua desgasificada y se equilibraron en la cámara anaerobia. Las preparaciones de D-alanina aminotransferasa de *B. sphaericus* y de aldolasa de la SEQ ID NO:276 se descongelaron sobre hielo y se utilizaron inmediatamente en la cámara anaerobia. Las soluciones a granel se agregaron a los tubos cónicos de 14 mL para proporcionar una concentración final de 100 mM de EPPS, 200 mM de piruvato, 100 mM de triptófano, 1 mM de MgCl<sub>2</sub>, Tween® 80 al 0,01% (v/v), 0,05 mM de PLP, 2 mg/mL de D-alanina aminotransferasa de *B. sphaericus* y 0,01 mg/mL de la aldolasa de la SEQ ID NO:276 en un volumen total de 7 mL por tubo. Las reacciones se incubaron a temperatura ambiente en una cámara anaerobia con leve agitación durante 22 horas. Se extrajeron las muestras y éstas se analizaron para verificar la monatina como se describe en el ejemplo 1 mediante el uso del procedimiento de control de reacciones múltiples por LC/MS/MS (Tabla 29).
- 15
- 20

Tabla 29

pH de reacción	Monatina (mM) a las 22 horas
7,0	5,8
7,2	9,9
7,8	14,0
8,0	14,2
8,2	14,3

pH de reacción	Monatina (mM) a las 22 horas
8,4	12,6
8,6	12,3
8,8	10,8
9,0	11,1

Los resultados indicaron que la formación de monatina aumentó con el incremento del pH entre 7,0 y 8,0. La formación de monatina alcanzó un máximo en el intervalo de pH de 8,0 y 8,2 y disminuyó en un pH mayor a 8,4. Además, la pureza isomérica de monatina disminuyó en un pH mayor a 8,4.

- 5 Se realizó una mayor optimización de reacción mediante el uso de la aldolasa de la SEQ ID NO: 276 (0,01 mg/mL), 2 mg/mL de T243N 4978 DAT (sin marcar, del ejemplo 26), 1 mM de MgCl<sub>2</sub>, 200 mM de piruvato de sodio, 0,05 mM de PLP, Tween-80 al 0,01% y 100 mM de D-triptófano a pH 8,5. Las células utilizadas para producir la aldolasa y la D-aminotransferasa se abrieron en 50 mM de EPPS, pH 8,4, mediante el uso de un homogeneizador Microfluidics (Microfluidics, Newton, MA) (3 pasadas a 20.000 psi). Los residuos celulares se eliminaron mediante centrifugación (20.000×g durante 30 minutos) y los extractos celulares clarificados se utilizaron en las reacciones enzimáticas. No se utilizó ningún tampón adicional, pero las mezclas de reacción se ajustaron al pH 8,5 mediante el uso de hidróxido de sodio y se enjuagaron con nitrógeno antes de la adición de la enzima. Se llevaron a cabo reacciones de 250 mL en 0,7 L, seis para fermentadores agitados (Infors AG, Bottmingen, Suiza) a 30°C mediante el uso de nitrógeno en el espacio de aire. Se agregó fosfato de potasio a una concentración final de 0, 5, 10, 20 y 50 mM. La adición de 5 a 10 mM de fosfato resultó ser óptima, la cual produjo 3,5 g/L de monatina (cuantificada por LC/MS/MS como se describe en ejemplo 1).

#### Ejemplo de referencia 25

Clonación de una nueva D-aminoácido aminotransferasa de *Bacillus*

- 20 Se produjo D-aminoácido aminotransferasa de *Bacillus* ("DAAT" o "DAT") (EC 2.6.1.21, también conocida como D-alanina aminotransferasa o D-aspartato aminotransferasa) recombinantemente. Esta aminotransferasa se utilizó en ensayos acoplados con aldolasas para la producción de monatina R,R. Esta enzima aminotransferasa es homóloga a las D-aminotransferasas descritas previamente para la producción de monatina (solicitud de los Estados Unidos publicada No. 20040063175 y solicitud publicada de los Estados Unidos No. 20050282260). El organismo ATCC 4978—*Bacillus sphaericus* depositado originalmente como *Bacillus rotans*—se solicitó a ATCC y se utilizó para preparar AND genómico. Se diseñaron cebadores degenerados en las regiones de conservación de secuencias de proteínas de D-aminotransferasas conocidas de *Bacillus* y se utilizaron para la amplificación de la reacción en cadena de la polimerasa ("PCR") de la secuencia de genes de la DAAT interna de la cepa ATCC mencionada anteriormente. El paseo genómico se llevó a cabo mediante el uso del kit universal BD GenomeWalker™ (Clontech, Mountain View, CA). Los análisis de secuencias (Agencourt BioScience Corporation, Beverly, MA) verificaron una secuencia codificadora de longitud total para el gen de DAAT de ATCC 4978. La secuencia de ADN del gen de DAAT de ATCC 4978 es la SEQ ID NO:357 y se muestra a continuación. El gen de la SEQ ID NO:357 puede amplificarse a través de los protocolos de PCR estándar y clonarse mediante el uso de técnicas de ADN recombinante. El gen de la SEQ ID NO:357 puede además reconstruirse mediante cualquier procedimiento conocido por el entendido en la técnica, como los procedimientos de ensamblaje de PCR conocidos por el entendido en la técnica.

La secuencia de ADN de ATCC 4978 de DAAT es:

atgagttata gcttatggaa tgaccaaatt gtgaatgatg aagaagtagt agttgataag  
 gaggaccgtg gctatcaatt tggcgatggt gtttatgaag ttgtaaaagt atataacggt gaattattta  
 cagcggagga gcatgtcgat cgtttttacg cgagtgtga aaaaattcgc gttacgatcc  
 ctatacaaa agacaaattg catcaattat tgcacagtt agttgaaatg aataaagttc aaacaggaca  
 tattatttc caaattacgc gtggtgcagg ccctcgtaat catattttcc ctggtgatga agtgaagcca  
 gtatfaacag gtaataccaa ggaaaatcca cgtcccgtag caaacttga aaaaggtgtg  
 aaagcaacat ttgtagaaga cattcgttgg ttacgctgtg acattaaatc attaaattta ctggtgcgg  
 tacttgctaa acaagaagca catgaaaaag gatgctatga agcggttta catcgtgatg  
 aaatcgtaac agaaggctct tcttcaata tttatggaat taaagatggc gtattataca cacatccagc  
 gaataacttc atcttaaattg gtattacag tcaagtaatc attaaatgtg ctgctgaaat tggcttacca  
 gtgaaggaag aagcaatgac aaaaactcag ctcttgcaa tggatgaagt gattgtttca  
 tcaacgactt cagaagtaac gccaaattac gacatagatg gaacagtaat tgggtcgggt  
 aaaccgggtg actggacacg taaattacaa gcacaattg atacgaaaat cccaaaaggt  
 attcgcgat aa

(SEQ ID NO: 357)

5 La secuencia de aminoácidos del gen de DAAT de ATCC 4978 como se codifica por la secuencia de ADN anterior es la SEQ ID NO:358 y se muestra a continuación:

Met Ser Tyr Ser Leu Trp Asn Asp Gln Ile Val Asn Asp Glu Glu Val Val  
 Val Asp Lys Glu Asp Arg Gly Tyr Gln Phe Gly Asp Gly Val Tyr Glu Val  
 Val Lys Val Tyr Asn Gly Glu Leu Phe Thr Ala Glu Glu His Val Asp Arg  
 Phe Tyr Ala Ser Ala Glu Lys Ile Arg Val Thr Ile Pro Tyr Thr Lys Asp Lys  
 Leu His Gln Leu Leu His Gln Leu Val Glu Met Asn Lys Val Gln Thr Gly  
 His Ile Tyr Phe Gln Ile Thr Arg Gly Ala Gly Pro Arg Asn His Ile Phe Pro  
 Gly Asp Glu Val Lys Pro Val Leu Thr Gly Asn Thr Lys Glu Asn Pro Arg  
 Pro Val Ala Asn Phe Glu Lys Gly Val Lys Ala Thr Phe Val Glu Asp Ile  
 Arg Trp Leu Arg Cys Asp Ile Lys Ser Leu Asn Leu Leu Gly Ala Val Leu  
 Ala Lys Gln Glu Ala His Glu Lys Gly Cys Tyr Glu Ala Val Leu His Arg  
 Asp Glu Ile Val Thr Glu Gly Ser Ser Ser Asn Ile Tyr Gly Ile Lys Asp Gly  
 Val Leu Tyr Thr His Pro Ala Asn Asn Phe Ile Leu Asn Gly Ile Thr Arg  
 Gln Val Ile Ile Lys Cys Ala Ala Glu Ile Gly Leu Pro Val Lys Glu Glu Ala  
 Met Thr Lys Thr Gln Leu Leu Ala Met Asp Glu Val Ile Val Ser Ser Thr  
 Thr Ser Glu Val Thr Pro Ile Ile Asp Ile Asp Gly Thr Val Ile Gly Ala Gly  
 Lys Pro Gly Asp Trp Thr Arg Lys Leu Gln Ala Gln Phe Asp Thr Lys Ile  
 Pro Lys Gly Ile Arg Ala

(SEQ ID NO: 358)

10 La nueva D-aminotransferasa obtenida a partir de la cepa ATCC 4978 tiene una secuencia de proteínas que presenta diferentes cambios de residuos de aminoácidos al compararla con la secuencia de D-aminotransferasa publicada de ATCC 10208 de *B. sphaericus*. La DAAT de ATCC 4978 tiene únicamente un 72% de identidad con la DAAT de *B. sphaericus* (ATCC 10208). Mientras que esta cepa se enumera como *B. sphaericus* en la ATCC, se

depositó como *B. rotans*. Con base en las alineaciones de las secuencias y las diferencias destacadas entre esta nueva DAAT y la DAAT de *B. sphaericus*, se identificaron varios residuos candidatos que pueden evaluarse por su rol (individualmente o en combinación) en la mayor actividad de DAAT para la biosíntesis de monatina R,R, en estas, así como otras secuencias de DAAT.

5 **Ejemplo de referencia 26**

Caracterización de mutantes de D-Aminotransferasa de ATCC 4978

Resumen experimental

10 El nuevo gen de la D-aminotransferasa (descrito en el ejemplo 25) de la cepa ATCC 4978 de *Bacillus* se mutagenizó mediante el uso de procedimientos dirigidos al sitio. Los genes mutantes se expresaron y ensayaron para verificar las actividades de interés para las vías de producción de monatina, especialmente cuando se acopla a una o más aldolasas.

15 Además de las ideas enumeradas en el ejemplo 16 para los objetivos de la mutagénesis dirigida al sitio, se desarrollaron otras ideas para el anclaje real de R-MP en el sitio activo de la estructura cristalina de YM-1, las alineaciones de secuencias de aminoácidos primarios se utilizaron para determinar si era probable que la proteína aminotransferasa 4978 tuviera características estructurales en dicha región. Se esperaba que las siguientes mutaciones adicionales fueran beneficiosas (mediante el uso de la numeración de aminoácidos de la aminotransferasa 4978). Se pensó que la mutagénesis de alanina 153 a arginina estabilizaría al segundo grupo carboxilo del sustrato (R-MP). Es probable que este cambio aumente el impedimento estérico, de forma tal de compensar esto, los residuos de serina en las posiciones 181 y 182 se cambiaron a alanina o glicina. También se 20 manejó la hipótesis de que se podría introducir una arginina en la posición 180, 181 o 182 y convertir uno o más de los otros residuos de serina a alanina o glicina para hacer lugar para la cadena lateral de arginina más voluminosa. La fenilalanina en el aminoácido 200 se encuentra espacialmente cerca del lugar donde se predice que el R-MP se anclará en el sitio activo y existe una amplia variabilidad en este residuo entre las D-aminotransferasas que catalizan la transaminación de monatina bastante bien. Se pensó que las modificaciones de los aminoácidos en esta posición 25 podrían ser útiles. Se predijo que la mutación de la leucina 151 a fenilalanina mejoraría potencialmente las interacciones con el anillo de indol del sustrato.

30 Con base en la bibliografía, se partió de la hipótesis que la mutación de treonina 243 a aspargina puede mejorar la selectividad de R-MP para las reacciones de transaminación. De igual modo, se pensó que la mutagénesis de aspargina 100 a alanina puede mejorar la actividad específica de la enzima para las reacciones de transaminación de monatina (Ro, et al., *FEBS Lett* 398:141-145, (1996); Sugio, S, et al., *Biochemistry* 34:9661-9669, (1995); EP1580268).

35 Lee et al. caracterizaron los mutantes de la región 141 a 144 (asa) y encontraron que las D-aminotransferasas con EYcY en lugar de LRcD (que es nativo para nuestra proteína) tienden a tener una Km inferior para los sustratos de ácido dicarboxílico. (Lee S G, Hong S P, Song J J, Kim S J, Kwak M S, Sung M H. *Functional and structural characterization of the most stable D-amino acid aminotransferases from Geobacillus spp. Appl Environ Microbiol.* 2006 Febrero; 72(2):1588-94). Debido a que el MP es un sustrato de ácido dicarboxílico, similar al alfa-ceto glutarato y las concentraciones de MP son bastante bajas en una mezcla de reacción de producción de monatina, una Km inferior podría ayudar potencialmente a la actividad de una DAT mutante para la producción de monatina.

40 A continuación se describen los procedimientos para crear los mutantes de D-aminotransferasa 4978, así como analizar los resultados con el uso de estos mutantes.

Mutagénesis

45 Los cebadores para mutagénesis se diseñaron de acuerdo con las sugerencias enumeradas en el kit Multi-Change de Stratagene (Stratagene, La Jolla, CA). Los cebadores se 5'-fosforilaron. La mutagénesis se realizó mediante el uso del kit Multi-Change de Stratagene (Stratagene, La Jolla, CA) de acuerdo con las instrucciones del fabricante. Las plantillas utilizadas para la mutagénesis fueron las construcciones de DAT 4978 de pET30 (sin marcar) o pET28 (marcado) descritas en el ejemplo 25. Las secuencias de cebadores se enumeran a continuación en la tabla 30:

**Tabla 30**

Nombre del mutante	Cambio de aminoácido	Cebador
4978-22	T243N	GTGATTGTTTCATCAACGAATTCAGAAGTAACGCC (SEQ ID NO: 359)

Nombre del mutante	Cambio de aminoácido	Cebador
10	T243R	GTGATTGTTTCATCAACGCGTTCAGAAGTAACGCC (SEQ ID NO: 360)
7	T243S	GTGATTGTTTCATCAACGAGTTCAGAAGTAACGCC (SEQ ID NO: 361)
19	T243A	GTGATTGTTTCATCAACGGCTTCAGAAGTAACGCC (SEQ ID NO: 362)
15	N100A	GTGCAGGCCCTCGTGCTCATATTTCCCTGG (SEQ ID NO: 363)
B	T243Q	GAAGTGATTGTTTCATCAACGCAGTCAGAAGTAACCGCCAATTAT C (SEQ ID NO: 364)
2	T243N/N100	los cebadores anteriores se utilizan juntos

Las células XL10-Gold de *E.coli* (Stratagene, La Jolla, CA) se transformaron y las preparaciones de plásmido purificadas resultantes se secuenciaron para verificar que se incorporaran las mutaciones correctas.

#### Expresión y ensayo

5 Las preparaciones de ADN plásmido que contenían los mutantes correctos o la DAT 4978 de tipo salvaje se transformaron en el huésped de expresión BL21(DE3) de *E. coli* (Novagen, Madison, WI). Los cultivos se desarrollaron mediante el uso de los protocolos descritos anteriormente, los plásmidos se aislaron mediante el uso del kit miniprep de Qiagen (Qiagen, Valencia, CA) y se analizaron mediante digestión restrictiva, como se describe anteriormente, para confirmar la presencia de una inserción.

10 La inducción del gen de DAAT se realizó típicamente en un medio LB que contenía kanamicina (50 µg/mL). Las células se cultivaron a una OD<sub>600 nm</sub> de 0,4 a 0,8 a 37°C., se indujeron con 0,1 mM de IPTG (isopropil tiogalactósido) y se muestrearon a las 3 a 4 horas después de la inducción.

15 Los extractos celulares se prepararon con el reactivo BugBuster y benzonasa nucleasa (Novagen, Madison, WI). Se realizaron ensayos de 1 ml a 30°C con leve agitación que contenían 10,2 mg de D-triptófano, 0,05 mM de PLP, 4 mM de MgCl<sub>2</sub>, 100 mM de tampón de fosfato de potasio, pH 7,5, aproximadamente 50 µg de aldolasa, 200 mM de piruvato y 0,150 a 0,5 mg/mL de D-aminotransferasa proporcionados como extractos celulares. La totalidad de los ensayos de proteínas se realizó mediante el uso del kit de proteínas totales Bio-Rad (Coomassie) (Bio-Rad, Hercules, CA) o el kit BCA de Pierce (Pierce, Rockford, IL) y el porcentaje de expresión de la D-aminotransferasa se estimó mediante SDS-PAGE o el sistema de electroforesis automatizada Experion de Bio-Rad (Bio-Rad, Hercules, CA). Se tomaron muestras a las 3 horas y durante la noche.

20 La aldolasa R-específica de la SEQ ID NO: 28 se utilizó en ensayos con aproximadamente 0,150 mg/mL de D-aminotransferasa.

25 Los primeros ensayos mostraron que los siguientes mutantes (sin marcar) presentaron actividad de transaminación (en orden de mayor a menor): T243N, T243S, T243N/N100A, N100A. Se notó que el T243N parecía aumentar la estereo-pureza de la monatina R,R producida. Los ensayos se repitieron mediante el uso de ProA aldolasa purificada de *Comamonas testosteroni* (100 µg/ml) y 0,50 mg/ml de mutantes de D-aminotransferasa (sin marcar, proporcionados como extracto celular). Se tomaron muestras a las 2 horas y durante la noche. Los resultados para las proteínas activas se muestran a continuación, se promediaron los resultados duplicados. El % de monatina R,R se determinó por área de pico en HPLC de fase inversa y a continuación se midió mediante el uso del procedimiento de derivatización de la FDAA descrito en el ejemplo 1. En la columna 3 de la tabla 31, se realizó un análisis adicional de derivatización de la FDAA por LC/MS/MS como se describe en el ejemplo 1, para algunas de las reacciones y dichos resultados se muestran en el paréntesis. Únicamente la R,R y monatina S,R se producen a partir de D-triptófano. El mutante T243R no pareció producir monatina en las condiciones probadas y el mutante T243A produjo niveles muy bajos de monatina.

35

Tabla 31

Enzima sin marcar (tiempo – hora o durante la noche)	Monatina total (ppm)	% R,R
tipo salvaje 4978 (2 horas)	4,7	41,6
tipo salvaje 4978 (durante la noche)	43,2	35,1 (30,9)
T243S (2 horas)	55,0	37,4 (21,7)
T243S (durante la noche)	97,7	35,5 (29,8)
T243N (2 horas)	73,2	86,7 (88,3)
T243N (durante la noche)	120,9	86,3 (86,1)
N100A (2 horas)	12,0	40,8
N100A (durante la noche)	22,3	41
T243A (2 horas)	0,8	~100

5 A pesar de que los ensayos se realizaron con la estimación del porcentaje de D-aminotransferasa mediante el uso del software Experion de Bio-Rad es claro que los mutantes T243S y T243N presentan una actividad mayor en comparación con la enzima de tipo salvaje. El mutante T243N proporcionó además un beneficio adicional del aumento dramático del % de monatina R,R formada. Esta enzima tiene una mayor preferencia para el R-MP en comparación con el S-MP en reacciones de transaminación. El mutante N100A no aumentó la actividad por sí solo o en combinación con T243N, contrariamente a lo que se sugirió en la bibliografía. El mutante V34A dirigido al sitio de la DAT 4978 sin marcar también se creó mediante el uso de procedimientos similares como se describe anteriormente. Se encontró que el mutante V34A dirigido al sitio tenía una actividad significativamente menor que la enzima de tipo salvaje en las condiciones probadas.

15 Otro punto de interés en los ensayos iniciales fue que la enzima de tipo salvaje pareció tener una mayor actividad cuando se produjo con un marcador His de terminal N. La mutagénesis subsiguiente se realizó en la versión marcada del gen. Además, los mutantes más prometedores anteriores se subclonaron en pET28b que tenía un marcador His con terminal N. Estos se purificaron con el uso de columnas HIS-bind de Novagen y del protocolo del fabricante con los tampones recomendados (Novagen, Madison, WI). El tampón de las fracciones de eluyente se intercambió, mediante el uso de columnas PD10 de GE Healthcare (GE Healthcare, Piscataway, NJ), con el tampón utilizado en los ensayos.

20 Se realizaron ensayos de 1 ml con D-aminotransferasa purificada (0,5 mg/ml) y aldolasa R-específica purificada de la SEQ ID NO:276 (50 µg/ml) a 30°C con leve agitación y que contenían 10,2 mg de D-triptófano, 0,05 mM de PLP, 200 mM de piruvato, 4 mM de MgCl<sub>2</sub> y 100 mM de tampón de fosfato de potasio, pH 7,5 Las muestras duplicadas se incubaron durante 2 horas y durante la noche. Como control positivo, se utilizó la DAT de *Bacillus sphaericus* (clonada en el ejemplo 7) en los mismos ensayos. Los resultados se presentan en la tabla 32:

Tabla 32

Enzima - marcada (tiempo: hora o durante la noche)	Monatina total (ppm)	% R,R
tipo salvaje 4978 (2 horas)	43	98,4
tipo salvaje 4978 (durante la noche)	96,7	98,3 (95,9)
T243N (2 horas)	197,5	100
T243N (durante la noche)	301,2	99,9 (99,6)
DAT de <i>B. sphaericus</i> (2 horas)	58,2	99,7

Enzima - marcada (tiempo: hora o durante la noche)	Monatina total (ppm)	% R,R
DAT de <i>B. sphaericus</i> (durante la noche)	221,7	98,7 (96,6)
T243Q (2 horas)	7,1	100
T243Q (durante la noche)	12,4	98,8

- 5 Los datos anteriores mostraron que el mutante T243N produce claramente la mayor cantidad de monatina a las 2 horas. A medida que pasa el tiempo, la proporción del control positivo del mutante T243N con respecto a la DAT de *B. sphaericus* se reduce. Este resultado sugiere que el mutante T243N no es tan estable durante la reacción de monatina como la DAT de *B. sphaericus*. Cuando se analizó en condiciones similares, la enzima T243S (marcada, purificada) presentó niveles de actividad similares a los del mutante T243N; no obstante, el porcentaje de monatina R,R producido fue menor (97,2% en ambas 2 horas y durante la noche). El mutante T243N/N100A tiene una actividad menor a la del mutante T243N. No obstante, ambos T243S y T243N/N100A presentaron una actividad mayor a la de la DAT 4978 de tipo salvaje.
- 10 Los ensayos de transaminación se realizaron para determinar qué proporciones de reacción se mejoraron mediante el uso del mutante T243N en lugar de la DAT de *B. sphaericus*. Se realizaron ensayos de 0,5 ml a 30°C y se tomaron como intervalos de tiempo 1 hora, 2 horas y 5 horas. Los ensayos contenían 25 mM de monatina o D-triptófano, 25 mM de piruvato, 100 mM de fosfato de potasio, pH 7,5, 50 µM de PLP y 0,1 mg de D-aminotransferasa (marcada, purificada). En el caso en el que se utilizaron menos de 100 µg de DAT, la cantidad de alanina se normalizó a 100 µg de D-aminotransferasa. Las muestras se trataron con ácido fórmico y se analizaron mediante LC-OPA para verificar la presencia del coproducto, alanina. Los resultados se muestran en las tablas 33 y 34.
- 15

**Tabla 33: Actividad de transaminación con monatina R,R como sustrato**

Enzima	D-alanina (mM)
DAT 4978 tipo salvaje (2 horas)	0,54
DAT 4978 tipo salvaje (5 horas)	1,11
T243N/N100A (2 horas)	1,32
T243N/N100A (5 horas)	2,78
T243S (2 horas)	1,5
T243S (5 horas)	2,61
T243N (2 horas)	1,26
T243N (5 horas)	2,65
DAT de <i>B. sphaericus</i> (2 horas)	0,97
DAT de <i>B. sphaericus</i> (5 horas)	2,2

20 **Tabla 34: Actividad de transaminación con D-triptófano como sustrato**

Enzima	D-alanina (mM)
DAT 4978 de tipo salvaje (1 hora)	4,55
DAT 4978 de tipo salvaje (2 horas)	8,47
T243N/N100A (1 hora)	8,52

Enzima	D-alanina (mM)
T243N/N100A (2 horas)	12,67
T243S (1 hora)	4,89
T243S (2 horas)	8,1
T243N (1 hora)	7,19
T243N (2 horas)	10,83
DAT de <i>B. sphaericus</i> (1 hora)	8,7
DAT de <i>B. sphaericus</i> (2 horas)	12,54

Para las reacciones de D-triptófano, los resultados muestran que algunas de las enzimas alcanzaron el equilibrio a las 2 horas. Las reacciones de monatina R,R limitan claramente la velocidad y las mejoras a esta actividad tienen un impacto mayor sobre las velocidades de producción de monatina de D-triptófano.

- 5 Se realizaron ensayos adicionales para examinar la estabilidad del mutante T243N de la DAT 4978. La enzima de tipo salvaje también pierde actividad con el paso del tiempo. El ejemplo 27 describe procedimientos para mejorar la estabilidad del mutante T243N de la D-aminotransferasa. Cuando las versiones recién preparadas marcadas y sin marcar del mutante T243N se prepararon y compararon con respecto a su actividad, se encontró que la versión sin marcar tenía una mejor estabilidad temporal, lo que la hace una mejor versión de la enzima para utilizarla en reacciones de producción de monatina.

Se prepararon otros mutantes de la DAT 4978 mediante procedimientos comúnmente conocidos por los expertos en la técnica. No obstante, todas estas mutaciones resultaron en una proteína que era insoluble en las condiciones en que fueron preparadas y por consiguiente, no podían analizarse para verificar su actividad. Las mutaciones que resultaron en proteínas insolubles fueron:

- 15 S180 A/S181A/S182R;  
L151F;  
V34G;  
S181R;  
A153R/S181A/S182A;  
20 A153R/S182A;  
A153R/S182G;  
S180R/S181A/S182G;  
S180R/S181A/S182A;  
S180R/S181G/S182G;  
25 S180G/S181R/S182G; y  
S180A/S181R/S182A.

Mutagénesis adicional

- 30 Para crear el mutante F200M de DAT 4978, el marco de lectura abierto de la DAT 4978 de tipo salvaje del ejemplo 25 (marcada) se amplificó con los cebadores 73 y 80 (a continuación) y la polimerasa de ADN PfuTurbo (Stratagene, La Jolla, CA) y se clonó en pCRII-Blunt (Invitrogen, Carlsbad, CA). Se verificó su secuencia (Agencourt, Beverly, MA). Las regiones 5' y 3' se amplificaron mediante el uso de los cebadores 80 y 96 y 99 y 103, respectivamente. El ADN amplificado a continuación se purificó en gel mediante el uso del kit de extracción en gel QIAquick de Qiagen (Qiagen, Valencia, CA). El ADN amplificado se sometió nuevamente a PCR mediante el uso de los cebadores 80 y 99. El ADN amplificado se purificó en gel como se describe anteriormente y se clonó en pCRII Blunt y se verificó su secuencia. El marco de lectura abierto de la DAT se subclonó como un fragmento de enzima de restricción NdeI/XhoI en pET28b.



Tabla 35

Número de cebador	Secuencia
73	CATATGAGTTATAGCTTATGGAATGACCAAATTGTGAATG (SEQ ID NO: 365)
80	CTCGAGTGCGGCCGCAAGCTTGTGCGACGGAGCTC (SEQ ID NO: 366)
96	AATATTTATGGAATTAAGATGGCGTATTATACACACATCCAGCGAATAACATGATCTTAA ATGGTATTACACGTCAAGTAATCATTAAATGTGC (SEQ ID NO: 367)
99	GGCCAGTGAATTGTAATACGACTCACTATAGGGC (SEQ ID NO: 368)
103	CGCCATCTTTAATTCCATAAATATTTGAAGAAGAGCCTTCTG (SEQ ID NO: 369)

5 Los siguientes cebadores se diseñaron para mutagénesis dirigida al sitio adicional mediante el kit de mutagénesis dirigida al sitio QuikChange® (Stratagene, La Jolla, CA). La mutagénesis se realizó mediante el uso del kit Multi-Change de Stratagene (Stratagene, La Jolla, CA) de acuerdo con las instrucciones del fabricante. La plantilla utilizada para la mutagénesis fue la construcción de DAT 4978 de pET28 (marcado) descrita en el ejemplo 25. También se creó un mutante doble mediante el uso del mutante F200Y como plantilla y se realizó una ronda adicional de mutagénesis con el cebador T243N (enumerado anteriormente).

Tabla 36

Mutante	Oligo
141-LRcD-144->EYcY	GCAACATTTGTAGAAGACATTCGTTGGGAATACTGTTACATTAAATCATTAAATTTACTTG GTGCG (SEQ ID NO: 370)
F200Y	GTATTATACACACATCCAGCGAATAACTACATCTTAAATGGTATTACACGTCAAG (SEQ ID NO: 371)
S244K	GCAATGGATGAAGTGATTGTTTCATCAACGACTAAAGAAGTAACGCCAATTATCGACATA GATG (SEQ ID NO: 372)
243-TS-244->NK	GCAATGGATGAAGTGATTGTTTCATCAACGAATAAAGAAGTAACGCCAATTATCGACATA GATG (SEQ ID NO: 373)
243-TS-244->NR	GCAATGGATGAAGTGATTGTTTCATCAACGAATCGTGAAGTAACGCCAATTATCGACATA GATG (SEQ ID NO: 374)

10 Las regiones codificadoras mutantes se verificaron por secuenciado de ADN (Agencourt). Los plásmidos a los que se les verificó la secuencia se transformaron en células BL21(DE3) (Novagen, Madison, WI).

#### Expresión y ensayo

15 Los cultivos que contenían 100 ml de LB con 50 µg/ml de kanamicina en un matraz deflector de 500 ml inoculado con un ml de un cultivo durante la noche y cultivados a 37°C a una densidad óptica (a 600 nm) de aproximadamente 0,6. La producción de proteína se indujo mediante IPTG a una concentración final de 1 mM. Las células se incubaron a 30°C durante 4,5 horas después de la adición de IPTG. Las células se centrifugaron y congelaron a -80°C. Las células (preparadas mediante el uso del reactivo BugBuster de Novagen (Novagen, Madison, WI) que contenía 1 µL/mL de benzonasa nucleasa, 5 µL/mL de mezcla de inhibidor de proteasa II y 0,033 µL/mL de rLisozima de acuerdo con el protocolo recomendado de Novagen) se mezclaron y se analizaron por SDS-PAGE. Los mutantes (141-LRcD-144->EYcY) y (243-TS-244->NR) resultaron en proteínas insolubles en las condiciones en las cuales se prepararon. El mutante 243-TS-244->NK no presentó actividad cuantificable en las condiciones probadas y es probablemente una enzima de actividad débil en comparación con el tipo salvaje como es el mutante S244K.

25 Las proteínas marcadas con His se purificaron de la siguiente forma. Las columnas HIS-bind (Novagen, Madison, WI) se equilibraron con 10 mL de fosfato de potasio 100 mM, pH 7,8, que contenía 200 mM de NaCl y 50 µM de PLP. Los extractos libres de células se cargaron en la columna. Las columnas se lavaron con 10mL de tampón de equilibración, 10 mL de tampón de equilibración que contenía 25 mM de imidazol y 10 mL de tampón de equilibración que contenía 50 mM de imidazol. Las proteínas se eluyeron con 5 ml de tampón de equilibración que contenía 500 mM de imidazol. Las proteínas se desalaron mediante el uso de columnas PD10 que se equilibraron en

100 mM de fosfato de potasio, pH 7,8 que contenía 50  $\mu$ M de PLP. Las proteínas purificadas se concentraron y cuantificaron mediante el uso del ensayo Bradford (Bio-Rad, Hercules, CA).

5 Los mutantes de D-aminotransferasa se analizaron mediante el uso de 500  $\mu$ g/ml de D-aminotransferasa, 50  $\mu$ g/ml de aldolasa de la SEQ ID NO:276, 4 mM de  $MgCl_2$ , 50 mM de fosfato de potasio, pH 8, 200 mM de piruvato de sodio, 0,05 mM de PLP y 20,4 mg/ml de D-triptófano para las condiciones de ensayo. El volumen final fue de 1,25 ml. Las muestras (200  $\mu$ l) se tomaron después de 0,5, 1, 2 y 14 horas y se congelaron hasta que el experimento se completó. Las muestras se filtraron, se diluyeron de 1 a 10 y se analizaron mediante LC/MS/MS como se describe en el ejemplo 1.

10 La D-aminotransferasa 4978 de tipo salvaje del ejemplo 25 se utilizó como una referencia para el porcentaje de actividad relativa. La tabla 37 muestra la actividad relativa de cada mutante en cada intervalo de tiempo.

**Tabla 37**

D-aminotransferasa	tiempo (horas)	% actividad
4978 de tipo salvaje	0,5	100
T243N	0,5	270
F200M	0,5	50
F200Y	0,5	70
F200M/T243N	0,5	183
S244K	0,5	4
4978 de tipo salvaje	1	100
T243N	1	289
F200M	1	55
F200Y	1	81
F200M/T243N	1	203
S244K	1	6
4978 de tipo salvaje	2	100
T243N	2	266
F200Y	2	79
F200M/T243N	2	185
S244K	2	6
4978 de tipo salvaje	14	100
T243N	14	254
F200M	14	56
F200Y	14	80
F200M/T243N	14	168
S244K	14	8

El T243N fue el mejor mutante de todos los probados para verificar la actividad en la producción de monatina R,R.

**Ejemplo de referencia 27**

Estabilización del mutante T243N de la D-aminotransferasa de la cepa ATCC 4978

Como se muestra en el ejemplo 25, la actividad inicial de la DAT mutante T243N es significativamente mayor que para la DAT de *B. sphaericus*, pero la actividad disminuye más rápidamente. Experimentos adicionales, con el uso del protocolo anaerobio descrito a continuación, indicaron que la actividad inicial de la DAT mutante T243N, fue hasta 8 veces mayor que la DAT de *B. sphaericus*, no obstante, la actividad disminuyó rápidamente aún en las condiciones anaerobias. Los siguientes estudios se realizaron para tratar de mantener la mayor actividad durante un periodo extendido de tiempo.

El mutante T243N de la D-aminotransferasa de la cepa 4978 (descrito en el ejemplo 25) y la aldolasa HMG de *S. meliloti* se purificaron como las proteínas marcadas con HIS<sub>6</sub> como se describe a continuación. La aldolasa de la SEQ ID NO:276 se purificó como se describe en el ejemplo 23.

Purificación de DAT a partir de ATCC 4978 (mutante T243N)

El mutante T243N de la D-aminotransferasa de la cepa de ATCC 4978 con un marcador de purificación HIS<sub>6</sub> amino terminal (descrito en el ejemplo 26) se produjo mediante el uso del sistema Overnight Express II de Biosciences EMD (soluciones 1 a 6) (Novagen, Madison, WI) que contenía 50 µg/mL de kanamicina en matraces de agitación. Este sistema de expresión induce la expresión de los sistemas inducidos por IPTG sin la necesidad de controlar el crecimiento celular. Después de la inoculación de alícuotas de 200 mL del medio (en matraces de 1 L) de los cultivos líquidos o placas del receptor BL21(DE3) de *E. coli* que tiene el gen para la D-aminotransferasa mutante T243N de la cepa 4978 de ATCC en el plásmido pET28b, los cultivos se incubaron a 30°C durante la noche con agitación a 225 rpm. Cuando la OD<sub>600</sub> alcanzó un mínimo de 6, las células se cosecharon mediante centrifugación y se lavaron una vez con tampón.

El extracto libre de células se preparó mediante el uso del reactivo de extracción BugBuster® de EMD Biosciences (libre de amina primaria) (Novagen, Madison, WI) que contenía 1 µL/mL de Benzonasa® Nucleasa, 5 µL/mL de mezcla de inhibidor de proteasa conjunto II (Calbiochem—Novabiochem Corp., San Diego, CA) y 0,033 µL/mL de rLisozima™ (Novagen, Madison, WI) de acuerdo con el protocolo del fabricante. Todas las etapas de purificación subsiguientes se llevaron a cabo a 4°C. El extracto celular se centrifugó durante 20 a 30 minutos a 15.000×g para extraer los residuos celulares. Se aplicó una alícuota de 20 a 25 mL del extracto libre de células a una columna de 45 mL de la resina de flujo rápido GE Healthcare Chelating Sepharose™ (forma de níquel (II)) (GE Healthcare, Piscataway, NJ) que había sido equilibrada previamente con 100 mM de fosfato de potasio que contenía 200 mM de cloruro de sodio y 50 µM de PLP. Para generar níquel a partir de la resina, la resina se lavó con 150 mL de hexahidrato de sulfato de níquel (II) 200 mM y después con 150 mL de agua destilada. Después de cargar la muestra, la columna se lavó/eluyó con 150 mL del tampón de equilibración que contenía 25 mM de imidazol, 150 mL del tampón de equilibración que contenía 50 mM de imidazol y 150 mL del tampón de equilibración que contenía 500 mM de imidazol. La proteína marcada con HIS<sub>6</sub> se eluyó en el último lavado. El lavado con 500 mM de imidazol se concentró con dispositivos de centrifugación con filtro Millipore/Amicon Centricon Plus-70 (MWCO 10 kDa) (Millipore, Billerica, MA) a 15 a 20 mL de acuerdo con las instrucciones del fabricante. El imidazol y el cloruro de sodio se extrajeron mediante el pasaje a través de las columnas desechables de GE Healthcare PDIO (GE Healthcare, Piscataway, NJ) (2,5 mL de muestra por columna) equilibradas previamente con 100 mM de fosfato de potasio, pH 7,8, que contenía 0,5 µM de PLP. La aminotransferasa purificada se eluyó con 3,5 mL por columna del mismo tampón. Se determinó la concentración de cada fracción mediante el uso del kit de ensayo Pierce BCA (Pierce, Rockford, IL) mediante el uso de BSA como la el estándar de proteína.

Se determinó la pureza de cada fracción y el nivel de expresión en el extracto libre de células mediante el uso de un sistema de chip microcapilar Bio-Rad Experion (Bio-Rad, Hercules, CA) o mediante el uso de geles con gradiente de SDS-poliacrilamida del 4 al 15% de Bio-Rad (Bio-Rad, Hercules, CA) en un aparato celular Mini PROTEAN® 3. La proteína se visualizó en geles de poliacrilamida mediante el uso de tinción de Coomassie G-250 de Bio-Rad Bio-Safe (Bio-Rad, Hercules, CA) y se destiñó con agua. Típicamente, este procedimiento produce aproximadamente 20 mg de enzima a partir de 200 mL de cultivo durante la noche que tiene de un 85 a un 90% de pureza, a juzgar por el software Experion y a partir del análisis de los geles SDS-PAGE. Se almacenaron alícuotas (1 a 5 mL) de la enzima purificada a -80°C hasta su uso.

La purificación de la aldolasa de la SEQ ID NO:276 es como se describe en el ejemplo 23.

Purificación de la aldolasa HMG de *S. meliloti*

La aldolasa HMG de *S. meliloti* con un marcador de purificación HIS<sub>6</sub> amino terminal (clonación descrita en la solicitud de los Estados Unidos publicada No. 20040063175 y WO 03091396 A2) se produjo mediante la inducción de los cultivos desarrollados en caldo de Luria-Bertani que contenía 50 mg/L de kanamicina con 0,2 mM de IPTG. Después de la inoculación de alícuotas de 800 mL del medio de cualquiera de los cultivos líquidos o placas del receptor BL21(DE3) de *E. coli* que tiene el gen para la aldolasa HMG de *S. meliloti* en pET30(Xa/LIC), los cultivos se

incubaron a 37°C con agitación a 225 rpm. Cuando la densidad óptica alcanzó una OD<sub>600nm</sub> de 0,5 a 0,75, se agregó IPTG y los cultivos se incubaron a 30°C con agitación a 225 rpm durante 4 horas. Las células se cosecharon mediante centrifugación y se lavaron una vez con tampón.

5 Para preparar el extracto libre de células que contenía la aldolasa de *S. melliloti*, las células se suspendieron en 3 a 4 volúmenes de 50 mM de EPPS(N-(2-hidroxietil)piperazina-N'-(ácido 3-propanosulfónico), pH 8,2 que contenía 100 mM de NaCl y después se mezclaron mediante el uso de un homogeneizador Microfluidics (Microfluidics, Newton, MA) (3 pasadas a 18.000 psi), con mantenimiento de la temperatura de la suspensión a menos de 15°C. La suspensión de células se centrifugó durante 20 a 30 minutos a 15.000-20.000×g para extraer los residuos celulares. La proteína marcada con HIS<sub>6</sub> se purificó mediante el uso de columnas HIS-Bind® de EMD Biosciences (Novagen, Madison, WI) de acuerdo con el protocolo recomendado del fabricante con una excepción: las columnas se lavaron con una solución 1:1 de tampón de unión:tampón de lavado en lugar de tampón de lavado únicamente. La fracción de elución se concentró con dispositivos de centrifugación con filtro Millipore/Amicon Centricon Plus15 (MWCO 5 kDa) (Millipore, Billerica, MA) a 7 a 10 mL de acuerdo con las instrucciones del fabricante. El imidazol y el cloruro de sodio se extrajeron mediante el pasaje a través de las columnas desechables de GE Healthcare PD10 (GE Healthcare, Piscataway, NJ) (2,5 mL de muestra por columna) equilibrado previamente con 50 mM de EPPS, pH 8,2, que contenía 100 mM de NaCl. La aldolasa purificada se eluyó con 3,5 mL por columna del mismo tampón. Se determinó la concentración de proteínas de cada fracción mediante el uso del kit de ensayo Pierce BCA (Pierce, Rockford, IL) mediante el uso de BSA como el estándar de proteína.

20 Se determinó la pureza de cada fracción y el nivel de expresión en el extracto libre de células mediante el uso de geles con gradiente de SDS-poliacrilamida del 4 al 15% de Bio-Rad (Bio-Rad, Hercules, CA) en un aparato celular Mini PROTEAN® 3. La proteína se visualizó en geles de poliacrilamida mediante el uso de tinción de Coomassie G-250 de Bio-Rad Bio-Safe (Bio-Rad, Hercules, CA) y se destiñó con agua. Típicamente, este procedimiento produce aproximadamente 15 a 20 mg de la enzima a partir de 800 mL de cultivo y es de un 85 a un 95% pura. Se almacenaron alícuotas (1 a 5 mL) de la enzima purificada a -80°C hasta el uso.

#### 25 Ensayos de producción de monatina

Los tubos cónicos de polipropileno (14 mL) que contenían 143 mg de D-triptófano se desoxigenaron en una cámara anaerobia durante la noche. Las soluciones a granel de 1 M de tampón EPPS (pH 8,2), 2 M de MgCl<sub>2</sub>, 2 M de piruvato de sodio y 10 mM de PLP se prepararon en agua desgasificada y se equilibraron en la cámara anaerobia. Las soluciones a granel de Tween 80 al 10% (v/v), Tween 20 al 1% (v/v), Triton X-100 al 1% (v/v), acetona al 100%, etanol al 100% y glicerol (p/v) al 50% se equilibraron en la cámara anaerobia junto con 0,7 g de trehalosa, inositol, sorbitol y eritritol en 2 mL de tubos de microcentrifugadora. Las preparaciones de las enzimas purificadas se descongelaron sobre hielo y se utilizaron inmediatamente en la cámara anaerobia. Las soluciones a granel se agregaron a los tubos cónicos de 14 mL para proporcionar una concentración final de 100 mM de EPPS, 200 mM de piruvato, 100 mM de triptófano, 1 mM de MgCl<sub>2</sub>, 0,05 mM de PLP, 0,5 mg/mL de D-aminotransferasa y 0,01 mg/mL de aldolasa de la SEQ ID NO:276 o 0,05 mg/mL de aldolasa HMG de *S. melliloti*. Los componentes estabilizadores de enzima propuestos se agregaron en varias concentraciones finales (Tablas 38 y 39) para llevar el volumen final de reacción a 7 mL por tubo. Las reacciones se incubaron a temperatura ambiente en una cámara anaerobia con leve agitación durante 24 horas. Se extrajeron las muestras periódicamente y éstas se analizaron para verificar la monatina como se describe en el ejemplo 1 mediante el uso del procedimiento de control de reacciones múltiples por LC/MS/MS. Se calcularon las velocidades iniciales a partir de las muestras extraídas entre las 0 y 3 horas después de la adición de la enzima.

**Tabla 38**

Aditivo	Múltiplo de la velocidad inicial de formación de monatina	Múltiplo de la valoración final de monatina (20 horas)
Ninguno	1,0	1,0
Tween 80 al 0,01% (v/v)	1,3	1,4
Tween 80 al 0,1% (v/v)	1,3	1,5
Tween 20 al 0,01% (v/v)	1,1	1,5
Tween X-100 al 0,01% (v/v)	1,1	1,2
Acetona al 5% (v/v)	0,4	0,3
Etanol al 5% (v/v)	0,7	0,5
Glicerol al 1% (p/v)	1,9	1,1

Glicerol al 5% (p/v)	1,4	1,4
□ Glicerol al 10% (p/v)	1,1	1,7
Trehalosa al 10% (p/v)	1,0	1,3
Inositol al 10% (p/v)	1,3	1,5
Sorbitol al 10% (p/v)	1,1	1,3
Eritritol al 10% (p/v)	0,8	1,0

**Tabla 39 (SEQ ID NO: 276 aldolasa)**

Aditivo	Múltiplo de la velocidad inicial de formación de monatina	Múltiplo de la valoración final de monatina a las 22 horas
Tween 80 al 0,01% (v/v)	1,0	1,0
Glicerol al 1% (p/v)	1,2	0,9
Glicerol al 5% (p/v)	1,5	1,5
Glicerol al 10% (p/v)	1,7	2,1

- 5 La adición de detergente al 0,01%-0,1% (v/v), como Triton X-100, Tween 20 o Tween 80, o poliol al 1% a 10% (p/v), como glicerol, trehalosa, inositol o sorbitol, mejoró la estabilidad de la D-aminotransferasa T243N durante la duración del experimento.

#### Ejemplo de referencia 28

A: Clonación de racemasa de aminoácido de amplia especificidad (BAR) KT2440 de *Pseudomonas putida*

- 10 Se identificó una BAR en KT2440 de *P. putida* mediante el uso de la información en la bibliografía (Roise, D., Soda, K., Yagi, T., Walsch, C. T., Biochemistry 23, 5195-5201, (1984)). El sitio activo de una enzima de BAR de *P. striata* se secuenció y se informó —LTAVLKADAYGXGIGL (SEQ ID NO:375). Esta secuencia se utilizó para BLAST, la secuencia genómica de KT2440 de *P. putida* disponible en NCBI. Se identificó una proteína con una secuencia de consenso casi idéntica. Los cebadores se diseñaron para clonar el gen del ADN genómico obtenido de la colección de cultivos "American Type Culture Collection" (ATCC, Manassas, VA).

- 15 (5'-AGAAGACATATGCCCTTTCGCCGTAGGG-3' (SEQ ID NO: 376 y  
5'-AGAAGAGGATCCTCAGTCGACGAGTATCTTCG-3') (SEQ ID NO: 377)).

- 20 La PCR se llevó a cabo en condiciones estándar y el producto de la PCR se purificó (kit de purificación de PCR QIAquick, Qiagen, Valencia, CA). El producto de la PCR purificado se digirió con Nde I y BamH I. El producto de la PCR digerido se purificó en gel (kit de extracción en gel QIAquick, Qiagen, Valencia, CA) y se ligó a pET30 y pET28 que habían sido digeridos y purificados en gel en una forma similar. Los clones con inserciones se secuenciaron (Agencourt, Beverly, MA) y se identificaron los aislados con la secuencia correcta (pET30 KT2440 BAR y pET28 KT2440BAR) y se utilizaron en estudios posteriores.

La secuencia de ADN de KT2440 BAR es:

atgccctttcgccgtacccttctggctgcatccctggcacttctgatcaccggacaggccccctgtatgcg  
 gcaccaccgtgtgcatgagacaacggcaccacaacaccctgaccgtgcaaaaagcaatgcctgggtcga  
 agtcagcggcagcggcctgcagcacaacatccgcacgctgcaggccgagctggccggcaagtccaag  
 ctgtgcccgtgctcaaggccgatgcctatggccacggatcggcctggtaatgcatcgatcatcgccc  
 aaggcgtgccctgctggcgggtggccagcaacgaggaggccccgctggcgcgccaagtggcttcac  
 cgggcaactgggtgagggtacgctggccagcctcagcgagctggaagatggcttgagtagacatgg  
 aagagctgggtggcagcgcggaatttggccgccaaggccgatgcatcgccgcgccatggcaagac  
 cttgcgattcacatggcgtcaactccagcggcatgagccgcaacgggggtggagatggccacctggtc  
 cggccgtgggaagcgtgcagatcaccgaccagaagcacctcaagctggcgcgctgatgaccact  
 tcgccgtggaagacaaggacgatgtacgcaagggcctggcggcattcaacgagcagaccgactggtg  
 atcaagcacgccaggctggaccgagcaagctcacctgcacgcccaactcgttcgctacgctgga  
 agtcccgaagcgcgctggacatggtacgaacgggtggcgcgctgttcggcgacaccgtgccggcg  
 cgcaccgagtacaacgtgcatgagttcaaatcgacgctggcggcggtgcacagctatccggccgg  
 caacaccgtgggctatgaccgcaccttcaccctggcccgtgattcgcggctggccaacattacggctgg  
 gtactccgatggctaccgcccgggtattcaccaacaaggccatgtgctgatcaacggccaccgtgtgcc  
 ggtcgtgggcaagggtgctgatgaacacgctgatggcgtgatgcaccgacttcctgatgtgaaggggggt  
 aacgaagtgggtgctgttcggcaagcaggccggggcgaaatcaccaggccgagatggaagaaatca  
 acggcgcggttctcgcgattgtacaccgatggggcaattccaaccgaagatactcgtcgactga

(SEQ ID NO: 378)

La secuencia de aminoácidos de KT2440 BAR es:

Mpfrtllaaslallitgqaplyaapplsmdngntnlvtqnsnawvevsasalqhnirtlqaelagksklc  
 avlkadayghgiglvmpsiaaqgvpcvavasneearvvrasgftgqlvrirlaslseledglqydme  
 lvgsaefarqadaiaarhgktrihmalnssgmsrngvematwsgrgealqitdqkhlklvalmthfa  
 vedkddvrkglaafneqtdwlikharldrsklilhaansfatlevpearldmvrvtggalfgdtvpartey  
 kramqfkshvaavhsypagntvgdydrftlardsrlanitvgysdyrrvftnkghvlinghrvpvvg  
 5 kvsmntlmvdtvdfpdvkgnevvlfgkqaggeitqaemecingalladlytwgnsnpkilvd

(SEQ ID NO: 379)

B) Purificación de KT2440 BAR de *P. putida*.

El plásmido de pET30 KT2440 BAR descrito anteriormente se transformó en DE3 pLysS (Invitrogen, Carlsbad, CA).  
 La cepa resultante se cultivó en LB o caldo Terrific a 37°C con aireación a una OD<sub>600nm</sub> de 0,4 a 0,6 y se indujo con 1  
 10 mM de IPTG. La incubación continuó durante 3 a 4 horas a 37°C con aireación. Las células se cosecharon mediante  
 centrifugación y el sedimento celular se almacenó a -80°C hasta su uso. El sedimento celular se descongeló sobre  
 hielo. Las células se lisaron con BugBuster (Novagen, Madison, WI) y benzonasa (Novagen, Madison, WI). El  
 residuo celular se eliminó mediante centrifugación y el extracto libre de células se utilizó inmediatamente o se  
 almacenó a -80°C. El gen de KT2440 BAR también se clonó en los sitios Ndel-BamHI de pET28 y se transformó en  
 15 BL21 DE3 pLysS. Esta construcción no pareció expresar proteínas solubles muy eficientemente, por lo que la  
 versión sin marcar (pET30 KT2440 BAR) se estudió en estudios futuros.

5 El extracto se aplicó a una columna UnoQ (Bio-Rad, Hercules, CA) que había sido equilibrada con al menos 5 volúmenes de columna del tampón A (25 mM de fosfato de potasio, pH 8,0, 10 µM de piridoxal-5'-fosfato (PLP)). La columna se lavó con 2 volúmenes de columna del tampón A. La proteína se eluyó con un gradiente lineal del tampón B (tampón A+NaCl 1M) del tampón B del 0 al 100% sobre 20 volúmenes de columna y se recolectaron fracciones de 5ml del momento en que el gradiente comenzó. Las fracciones se analizaron mediante el uso del procedimiento Amplex Red descrito en el ejemplo 4 parte 7. En resumen, se combinó 100 µg de D-aminácido oxidasa (Sigma-Aldrich, St. Louis, MO), 0,05 mM de FAD, 25 mM de L-trp y un pequeño volumen de la fracción para analizarse en 50 µL de H<sub>2</sub>O y se agregó al tampón de reacción Amplex Red preparado como se indica en el protocolo del fabricante. Las fracciones con actividad se desalaron con una columna PD-10 (GE Healthcare, Piscataway, NJ) y se concentraron con concentradores de centrifugación Amicon (Millipore, Billerica, MA). La proteína purificada se almacenó a -80°C.

10 C) Producción y ensayo del mutante Y354A de alanina racemasa de *Geobacillus stearothermophilus* con actividad de triptófano racemasa.

15 La alanina racemasa de tipo salvaje de *Geobacillus stearothermophilus* (SEQ ID NO:380, mostrada a continuación, se clonó en pET30 y se utilizó como plantilla para la mutagénesis dirigida al sitio para realizar el cambio Y354A. El gen de la SEQ ID NO:380 puede amplificarse a través de los protocolos de PCR estándar y clonarse mediante el uso de técnicas de ADN recombinante estándar. El gen de la SEQ ID NO:380 puede además reconstruirse mediante cualquier procedimiento conocido por el entendido en la técnica, como los procedimientos de ensamblaje de PCR conocidos por el entendido en la técnica.

20 Las secuencias de ADN y aminoácidos de la alanina racemasa de tipo salvaje de *Geobacillus stearothermophilus* se muestran a continuación como la SEQ ID NO:380:

```
atggacgagt  ttcaccgca  tacgtggcg  gaagtggatt  tggacgcat  ttacgacaal
gtggagaatt  tgcgccgtt  gctgccggac  gacacgcaca  ttatggcgg  cgtgaaggcg
aacgcctatg  gacatgggga  tgtcaggtg  gcaaggacag  cgctcgaagc  gggggcctcc
cgcctggcgg  ttccctttt  ggatgaggcg  ctcgctttaa  gggaaaaagg  aatcgaagcg
ccgattctag  ttctcggggc  ttcccgtcca  gctgatcgcg  cgctggccgc  ccagcagcgc
attgcctga  ccgtgtccg  ctccgactgg  ttggaagaag  cgctccccct  ttacagcggc
cctttccta  ttcattcca  ttgaaaatg  gacaccggca  tgggacggct  tggagtgaaa
gacgaggaag  agacgaaacg  aatcgtagcg  ctgattgagc  gccatccgca  ttttgtctt
gaaggggtgt  acacgcattt  tgcgactcgc  gatgaggtga  acaccgatta  ttttctat

cagtataccc  gtttttgca  catgctcga  tggctgccgt  cgcgcccgcc  gctcgtccat
tgcgccaaca  ggcgagcgtc  gctccgttc  cctgaccgga  cgttcaatat  ggtccgcttc
ggcattgcca  tgatgggct  tgccccgtcg  cccggcatca  agccgctgct  gccgtatcca
ttaaagaag  catttcgct  ccatagccgc  ctctacacg  tcaaaaaact  gcaaccaggc
gaaaagggtga  gctatgggtc  gacgtacact  gcgcagacgg  aggagtggat  cgggacgatt
ccgatcggct  atgcggacgg  ctggctccgc  cgctgcagc  actttcatgt  ccttgtgac
ggacaaaagg  cgccgattgt  cggccgcat  tgcattggacc  agtgcattgat  ccgctgcct
ggtccgctgc  cggtcggcac  gaaggtgaca  ctgattggtc  gccaaagggga  cgaggttaatt
tccattgatg  atgtcgtcgc  ccatttgga  acgatcaact  acgaagtgcc  ttgcacgac
agtattcgag  tgccccgat  tttttccgc  cataagcgt  taatggaagt  gagaaacgcc
gttggccgcg  ga
```

(SEQ ID NO: 380)

25 La secuencia de aminoácidos codificada de la alanina racemasa (*Geobacillus stearothermophilus*) se muestra a continuación como la SEQ ID NO:381:

Met Asp Glu Phe His Arg Asp Thr Trp Ala Glu Val Asp Leu Asp Ala Ile Tyr Asp  
 Asn Val Glu Asn Leu Arg Arg Leu Leu Pro Asp Asp Thr His Ile Met Ala Val Val  
 Lys Ala Asn Ala Tyr Gly His Gly Asp Val Gln Val Ala Arg Thr Ala Leu Glu Ala  
 Gly Ala Ser Arg Leu Ala Val Ala Phe Leu Asp Glu Ala Leu Ala Leu Arg Glu Lys  
 Gly Ile Glu Ala Pro Ile Leu Val Leu Gly Ala Ser Arg Pro Ala Asp Ala Ala Leu  
 Ala Ala Gln Gln Arg Ile Ala Leu Thr Val Phe Arg Ser Asp Trp Leu Glu Glu Ala  
 Ser Ala Leu Tyr Ser Gly Pro Phe Pro Ile His Phe His Leu Lys Met Asp Thr Gly  
 Met Gly Arg Leu Gly Val Lys Asp Glu Glu Glu Thr Lys Arg Ile Val Ala Leu Ile  
 Glu Arg His Pro His Phe Val Leu Glu Gly Val Tyr Thr His Phe Ala Thr Ala Asp  
 Glu Val Asn Thr Asp Tyr Phe Ser Tyr Gln Tyr Thr Arg Phe Leu His Met Leu Glu  
 Trp Leu Pro Ser Arg Pro Pro Leu Val His Cys Ala Asn Ser Ala Ala Ser Leu Arg  
 Phe Pro Asp Arg Thr Phe Asn Met Val Arg Phe Gly Ile Ala Met Tyr Gly Leu Ala  
 Pro Ser Pro Gly Ile Lys Pro Leu Leu Pro Tyr Pro Leu Lys Glu Ala Phe Ser Leu  
 His Ser Arg Leu Val His Val Lys Lys Leu Gln Pro Gly Glu Lys Val Ser Tyr Gly  
 Ala Thr Tyr Thr Ala Gln Thr Glu Glu Trp Ile Gly Thr Ile Pro Ile Gly Tyr Ala Asp  
 Gly Trp Leu Arg Arg Leu Gln His Phe His Val Leu Val Asp Gly Gln Lys Ala Pro  
 Ile Val Gly Arg Ile Cys Met Asp Gln Cys Met Ile Arg Leu Pro Gly Pro Leu Pro  
 Val Gly Thr Lys Val Thr Leu Ile Gly Arg Gln Gly Asp Glu Val Ile Ser Ile Asp  
 Asp Val Ala Arg His Leu Glu Thr Ile Asn Tyr Glu Val Pro Cys Thr Ile Ser Tyr  
 Arg Val Pro Arg Ile Phe Phe Arg His Lys Arg Ile Met Glu Val Arg Asn Ala Val  
 Gly Arg Gly

(SEQ ID NO: 381)

- 5 La mutagénesis se realizó mediante el uso del kit de mutagénesis dirigida a múltiples sitios QuickChange (Stratagene, La Jolla, CA). El siguiente cebador mutagénico se utilizó para realizar el cambio Y354A, 5'-gccatttgaaacgatcaacgcggaagtgccttgacgatcag-3'

(SEQ ID NO:382).

- 10 La mutagénesis dirigida al sitio se realizó como se describe en el protocolo del fabricante. Se secuenciaron varios aislados (Agencourt, Beverly, MA) y se seleccionó un aislado con la secuencia correcta y se utilizó para el análisis adicional.

El mutante simple pET30Y354A se transformó en BL21(DE3)pLysS de *E. coli*. La proteína purificada se preparó de la siguiente forma: La cepa resultante se cultivó en LB o caldo Terrific (a 37°C con aireación) a una OD<sub>600nm</sub> de 0,4 a 0,6 y se indujo con 1 mM de IPTG. La incubación continuó a 37°C con aireación durante aproximadamente 3 horas. Las células se cosecharon mediante centrifugación y el sedimento celular se almacenó a -80°C.

- 15 El sedimento celular se descongeló sobre hielo y se volvió a suspender en un volumen adecuado de BugBuster (Novagen, Madison, WI) más Benzonasa nucleasa (Novagen, Madison, WI). El residuo celular se eliminó mediante centrifugación y el extracto libre de células se aplicó a una columna HIS-Bind (Novagen, Madison, WI) que había sido equilibrada con tampón de unión. La columna se lavó con tampón de unión y tampón de lavado y la proteína se eluyó con tampón de elusión (como lo indica el protocolo del fabricante). La proteína purificada se desaló mediante el uso de una columna PD-10 (GE Healthcare, Piscataway, NJ). La proteína se desaló en 50 mM de fosfato de potasio, pH 8,0 y 10 µM de piridoxal-5'-fosfato de acuerdo con el protocolo del fabricante. La proteína se concentró mediante el uso del concentrador de centrifugación Amicon (Millipore, Billerica, MA). La proteína purificada y concentrada se dividió en alícuotas pequeñas y se almacenó a -80°C hasta su uso.

- 25 El Y354A purificado se compare con la alanina racemasa de tipo salvaje (preparada en la forma descrita anteriormente) en los ensayos de alanina y triptófano. Los ensayos se realizaron a 37°C in 50 mM de tampón de fosfato de potasio, pH 8 y 10 µM de PLP mediante el uso de 400 µL de proteína purificada concentrada (>1 mg/ml) y



50 mM de sustrato. La detección de D-alanina y D-triptófano se realizó mediante el uso de la metodología de aminoácido quiral descrita en el ejemplo 1.

**Tabla 40**

Enzima	Sustrato	Tiempo	D isómero producido (ppm)
Tipo salvaje	L-triptófano	0	nd*
		10	nd
		30	nd
		60	nd
		1080	nd
Y354A		0	nd
		10	198
		30	568
		60	1386
		1080	10080
Tipo salvaje	L-alanina	0	5140
		10	5960
		30	6280
		60	6500
		1080	5040
Y354A		0	4760
		10	4980
		30	4980
		60	4200
		1080	5000

\*nd = no detectado.

- 5 Estos datos se analizaron sin el uso de un estándar interno; por consiguiente estos datos son semi-cuantitativos y deberían utilizarse a los efectos comparativos. Sin embargo, estos resultados muestran que la mutación simple del Y354A es suficiente para ampliar la especificidad de la alanina racemasa de forma tal que pueda catalizar la racemización de aminoácidos mediante el uso de sustratos alternativos.

D) Ensayo de la enzima BAR.

- 10 La enzima Bar se analizó de la siguiente forma. Los ensayos Amplex Red se prepararon en la forma descrita en este ejemplo. Se utilizó KT2440 BAR de *P. putida* a 200 µg (purificado como se describe en este ejemplo). La alanina racemasa de tipo salvaje de *G. stearothermophilus* y el Y354A se purificaron como se describe en este ejemplo y se utilizó a 200 µg/mL o 1000 µg/mL. El CE es un extracto libre de células que se preparó como se describe en este ejemplo. Los resultados para el intervalo de tiempo de 60 minutos se muestran en la tabla 41 a continuación.

15

**Tabla 41**

Enzima	Lectura de fluorómetro (a los 60 minutos)
BAR (200)	56943
Y354A (200)	7860
Y354A (1000)	13587
WT alanina racemasa (200)	3646
WT alanina racemasa (1000)	3639
BAR CE (5 µl)	16228
BAR CE (10 µl)	26662
BAR CE (50 µl)	>58000
Sin enzimas	1510

5 La proteína purificada también se analizó para verificar la actividad de triptófano racemasa en 50 mM de fosfato de potasio, pH 8, 10 µM de PLP y 30 mM de L-triptófano como se describe anteriormente. Se utilizaron 200 µg/mL o 1000 µg/mL de enzima purificada en los ensayos (indicado en el paréntesis). Se analizó el D-triptófano mediante el uso del procedimiento de aminoácido quiral descrito en el ejemplo 1 para la detección.

**Tabla 42**

Enzima	Tiempo	D-triptófano (ppm)
BAR (200)	0	0
	5	172
	10	410
	20	844
	30	1318
	60	2362
	120	2594
	240	2762
	1080	2294
	Y354A (200)	0
5		0
10		0
20		0
30		12
60		22
120		44
240		56
1080		368

Y354A (1000)	0	0
	5	0
	10	12
	20	18
	30	40
	60	80
	120	146
	240	218
	1080	1164

El ensayo indicó que la enzima KT2440 BAR de *P. putida* es mucho más activa sobre el triptófano que la enzima derivada de *G. stearothermophilus* y su mutante Y354A.

E) Producción de monatina con KT2440 BAR de *P. putida*.

- 5 Se realizó un ensayo de producción de monatina con la KT2440 BAR purificada de *P. putida* (como se purifica anteriormente) (100 µg) o Y354A purificado (como se purifica anteriormente) (500 µg), D-aminotransferasa (BioCatalytics AT-103 (Pasadena, CA)) (500 µg) y la aldolasa de la SEQ ID NO:276) (como se purifica en el ejemplo 23) (50 µg). El experimento de producción de monatina que comienza con L-triptófano se realizó de la siguiente forma. Además de las enzimas anteriores, se agregó por 1 mL de mezcla de reacción lo siguiente: 4 mM de MgCl<sub>2</sub>, 50 mM de L-triptófano, 100 mM de piruvato de sodio, 100 mM de tampón de fosfato de potasio, pH 7,5 y 0,05 mM de PLP.
- 10

Como control, el experimento se realizó en la forma descrita anteriormente sin racemasa y comenzó con D-triptófano en lugar de L-triptófano. En la tabla 43 a continuación se presenta un resumen de los resultados.

**Tabla 43**

Sustrato	Racemasa	Tiempo	Monatina total	% R,R	% S,S	% R,S	% S,R
L-Trp	Y354A	2 horas	Ninguno detectado				
		18 horas	Ninguno detectado				
L-Trp	BAR	2 horas	Ninguno detectado				
		18 horas	38,6 ppm	92,1	5		2,9
L-Trp	Ninguno	2 horas	Ninguno detectado				
		18 horas	Ninguno detectado				
D-trp	Ninguno	2 horas	19,9 ppm	Sin probar	Sin probar	Sin probar	Sin probar
		18 horas	221,25 ppm	97,8	0,2		2

- 15 No se detectó monatina mediante el uso de Y354A en este experimento. Esta racemasa fue utilizada en el pasado (datos no mostrados) para producir monatina, pero se utilizaba un nivel mucho mayor de enzima (al menos 2 mg y

hasta 10 mg para ver mayores niveles de monatina). La KT2440 BAR de *P. putida* se utilizó para producir monatina a partir de L-triptófano. Los 100 µg utilizados en este experimento no fueron suficientes para ver producción de monatina después de dos horas pero fue suficiente para ver producción de monatina después de 18 horas. La distribución de estereoisómeros indicó que la mayor parte de la monatina producida es el isómero R,R. No se produjo ningún isómero R,S. Este resultado indica que la KT2440 BAR no es capaz de racemizar de forma detectable el isómero R,R de monatina (la racemización del isómero R,R produciría el isómero R,S). Se produjo una cantidad significativa del isómero S,S en este experimento. Esto es probablemente debido al hecho de que el AT-103 utilizado en este experimento no es altamente purificado y puede contener L-aminotransferasas del extracto celular y hay una gran cantidad de L-triptófano presente para servir como donante amino para la transaminación de S-MP.

### Ejemplo de referencia 29

Clonación y expresión de arginina racemasa de *Pseudomonas taetrolens*

Resumen experimental

La arginina racemasa de *Pseudomonas taetrolens* (también conocida como *P. graveolens*) (Genbank No. de acceso AB096176, secuencia de ácidos nucleicos) y su mutante 1384M se clonaron, se expresaron y se analizaron para verificar su actividad en la conversión de L-triptófano a D-triptófano. Este gene s un 72% idéntico al gen de BAR de *P. putida* de KT2440 y un 73% idéntico al gen de BAR de *P. putida* de NBRC 12996 descrito anteriormente. La secuencia de aminoácidos es un 72% idéntica a ambas proteínas de BAR de *P. putida*.

Protocolo de la reacción en cadena de la polimerasa

Se cultivó *Pseudomonas taetrolens* (ATCC 4683) en caldo nutriente a 28°C con agitación a 225 rpm. La reacción en cadena de la polimerasa se realizó en todas las células mediante el uso de cebadores diseñados con sitios de restricción 5' y salientes para clonar a los vectores pET 28 y pET30 (Novagen, Madison, WI).

Las secuencias de cebadores fueron las siguientes:

Terminal N: 5'-ATAATACATATGCCCTTCTCCCGTACCC-3' (SEQ ID NO: 408) y

Terminal C: 5'-GCGGCGGGATCCTTACTGATCTTTCAGGATT-3' (SEQ ID NO: 409).

El gen derivado de *P. taetrolens* se amplificó mediante el uso del siguiente protocolo de PCR. Veinticinco µL de las células cultivadas se lisaron a 96°C durante 10 minutos. El residuo celular se eliminó mediante centrifugación y el sobrenadante se utilizó como plantilla para la PCR. La reacción de 100 µL contenía 5 µL de plantilla (sobrenadante de células lisadas), 1,6 µM de cada cebador, 0,3 mM de cada dNTP, 10 U de rT<sup>th</sup> Polimerasa XL (Applied Biosystems, Foster City, CA), 1×XL de tampón y 1 mM de Mg(OAc)<sub>2</sub>. El programa termociclador utilizado incluyó un comienzo en caliente a 94°C durante 3 minutos, 8 repeticiones de las siguientes etapas: 94°C durante 30 segundos, 52°C durante 30 segundos y 68°C durante 2 minutos, seguidas por 22 repeticiones de las siguientes etapas: 94°C durante 30 segundos, 58°C durante 30 segundos y 68°C durante 2 minutos. Después de las repeticiones la muestra se mantuvo a 68°C durante 7 minutos y a continuación se almacenó a 4°C. Este protocolo de PCR produjo un producto de 1230 bp.

Clonación

Los productos de la PCR se purificaron en geles TAE-agarosa al 0,8% mediante el uso del kit de extracción en gel de Qiagen (Qiagen, Valencia, CA). El producto se clonó en TOPO y se transformó en células TOP 10 de acuerdo con el protocolo del fabricante (Invitrogen, Carlsbad, CA). El ADN plásmido se purificó a partir de los transformantes resultantes mediante el uso del kit spin miniprep de Qiagen (Qiagen, Valencia, CA) y se analizó para detectar las inserciones correctas mediante enzimas de restricción con Nde I y BamH I. Las secuencias de plásmidos que parecen tener una inserción correcta se verificaron mediante el secuenciado de ADN de terminación de la cadena didesoxi con los cebadores universales M13 directos y M13 inversos.

El clon TOPO correcto se digirió con enzimas de restricción Nde I y BamH I de acuerdo con los protocolos recomendados del fabricante (Biolaboratorios de Nueva Inglaterra, Beverly, MA) y se purificaron en geles de TAE-agarosa al 0,8% mediante el uso del kit de extracción en gel de Qiagen (Qiagen, Valencia, CA). Los vectores pET28 y pET30 se prepararon mediante la digestión con enzimas de restricción Nde I y BamH I seguidos del tratamiento con fosfatasa alcalina sérica (Roche, Indianápolis, IN) y purificación en geles de TAE-agarosa al 0,8% mediante el uso del kit de extracción en gel de Qiagen (Qiagen, Valencia, CA). Los vectores e inserciones digeridos se ligaron mediante el uso del kit de ligadura rápida de ADN (Roche, Indianápolis, IN). Aproximadamente 50 ng de la inserción tratada, 100 ng del vector tratado (proporción de inserción a vector 3 a 1), 5 U de T4 ADN ligasa y 1× de tampón de ligación se incubaron durante 5 minutos a temperatura ambiente. La reacción de ligación se desaló mediante el uso del kit de purificación del producto de la PCR altamente puro (Roche, Indianápolis, IN) y se utilizó para transformar las células electrocompetentes DH10B de *E. coli* (Invitrogen, Carlsbad, CA). Se agregaron 10 µl de cada reacción de

ligación a 40 µl de células DH10B y se transformaron mediante electroporación mediante el uso del impulsor de genes Bio-Rad II en las siguientes condiciones: 2,5 kV, 25° µF, 200 ohm en una cubeta de 0,2 cm. Las células se dejaron recuperar en 1 mL de medio SOC a temperatura ambiente durante 1 hora a 37°C con agitación a 225 rpm. Las células se colocaron en placas LB que contenían kanamicina (50 µg/mL). El ADN plásmido se purificó a partir de los transformantes resultantes mediante el uso del kit spin miniprep de Qiagen (Qiagen, Valencia, CA) y se analizó para detectar las inserciones correctas mediante enzimas de restricción con Nde I y BamH I.

#### Expresión y análisis génicos

El ADN plásmido se transformó en el huésped de expresión de *E. Coli* BL21(DE3) pLysS (Novagen, Madison, WI). Los cultivos se desarrollaron y los plásmidos se aislaron mediante el uso del kit miniprep de Qiagen (Qiagen, Valencia, CA) y se analizaron mediante enzimas de restricción para confirmar la identidad.

La inducción en BL21DE3 pLysS se realizó inicialmente en ambos vectores pET 28 (marcado con histidina) y pET 30 (sin marcar). Se realizó un estudio temporal con cultivos desarrollados a 37°C en 100 mL de kanamicina que contenía LB (50 mg/L) a una OD<sub>600</sub> de 0,5, inducidos con 100 mM de IPTG (isopropil tiogalactósido) y muestreados a las 0 y 3 horas después de la inducción. Las células procedentes de la hora 0 y la hora 3 se volvieron a colocar en suspensión en tampón de dodecil sulfato de sodio 1× que contenía 2-mercaptoetanol y se calentaron a 95°C durante 10 minutos y se enfriaron. Se analizaron por SDS-PAGE alícuotas de estas muestras de la proteína celular completa mediante el uso de un gel en gradiente del 4 al 15%.

Se prepararon también extractos celulares procedentes de los cultivos de 3 horas mediante la colocación en suspensión de los sedimentos celulares procedentes de 5 ml de cultivo en reactivo BugBuster™ de Novagen que contenía benzonasa nucleasa y mezcla de inhibidor de proteasa grupo #3 (Calbiochem-Novabiochem Corp., San Diego, CA) a temperatura ambiente durante 20 minutos con leve agitación y centrifugación a 16.000 x g para eliminar el residuo celular. Los sobrenadantes (extractos celulares) se cargaron en geles con gradiente del 4 al 15% para el análisis de las proteínas celulares solubles.

La muestra de las 3 horas de la arginina racemasa clonada de *P. taetrolens* mostró una banda de proteínas total que correspondió al tamaño correcto (aproximadamente 45 kDa) en el vector pET 30 (sin marcar). El producto de genes de *P. taetrolens* en pET 30 se sobreexpresó a un nivel mayor que el del producto de genes de *P. taetrolens* en pET 28 (marcado con histidina), pero ninguno de los vectores proporcionó una banda de proteínas solubles visible.

Las células de los cultivos inducidos (100 mL) se centrifugaron y se lavaron una vez con NaCl al 0,85%. Los sedimentos celulares se volvieron a suspender en 5 mL/g de peso de células húmedas de reactivo BugBuster™ (Novagen, Madison, WI) que contenía 5 µl/ml de la mezcla de inhibidor de proteasa del conjunto nº 3 (Calbiochem-Novabiochem Corp., San Diego, CA) y 1 µl/ml de benzonasa nucleasa. Las muestras se incubaron a temperatura ambiente durante 20 minutos en un agitador orbital. El residuo celular insoluble se eliminó por centrifugación a 16.000 x g durante 20 minutos a 4°C.

Los extractos celulares se analizaron para verificar la actividad de la triptófano racemasa mediante el uso del siguiente protocolo. Se llevaron a cabo reacciones de 1 mL en 50 mM de fosfato de potasio (pH 8,0), 0,05 mM de PLP y 30 mM L-triptófano. Las reacciones se iniciaron con la adición de extractos libres de células y se incubaron a 30°C durante la noche. Se tomaron alícuotas de las muestras después de la incubación durante la noche (las muestras del minuto cero sirvieron como reacción de control). Se agregó ácido fórmico concentrado (5 µL) a cada alícuota de muestra de 250 µL para detener la reacción y la proteína precipitada se eliminó mediante centrifugación. Los sobrenadantes se extrajeron y congelaron a -80°C hasta que se analizaron para detectar D-triptófano mediante el procedimiento de aminoácido quiral descrito en el ejemplo 1.

Los resultados de los ensayos de los extractos celulares de la inducción de pET28 y pET30 con 100 mM de IPTG (3 horas) demostraron que los clones de *P. taetrolens* muestran actividad de racemasa en L-triptófano. Nuevamente, la versión marcada de BAR parece ser menos activa y puede precipitarse o ser menos soluble que la versión sin marcar (pET28). La tabla 44 a continuación muestra los resultados iniciales, aunque no cuantitativos ya que se obtuvo una proteína muy poco soluble.

**Tabla 44**

Tratamiento	Intervalo de tiempo	Sustrato	Extracto de racemasa (200 µg)	D-trp conc (µg/mL)
pET28/ <i>P. taetrolens</i>	0	L-trp	500 µL	nd
pET30/ <i>P. taetrolens</i>	0	L-trp	500 µL	nd

Tratamiento	Intervalo de tiempo	Sustrato	Extracto de racemasa (200 µg)	D-trp conc (µg/mL)
pET28/ <i>P. taetrolens</i>	durante la noche	L-trp	500 µL	140
pET30/ <i>P. taetrolens</i>	durante la noche	L-trp	500 µL	226

Se repitió la inducción de la construcción del pET30 (sin marcar) mediante el uso de las mismas condiciones como se menciona anteriormente y se observó una banda de proteínas solubles visible en SDS-PAGE. El ensayo se repitió mediante el uso de las mismas condiciones descritas anteriormente y se obtuvieron los resultados como se muestra en la tabla 45 a continuación.

5

Tabla 45

Tratamiento	Intervalo de tiempo	Sustrato	Extracto de racemasa (µL)	D-trp conc (µg/mL)
<i>P. taetrolens</i> -pET30	0	L-trp	300	Nd
<i>P. taetrolens</i> -pET30	0	L-trp	150	Nd
<i>P. taetrolens</i> -pET30	2 horas	L-trp	300	319
<i>P. taetrolens</i> -pET30	2 horas	L-trp	150	308
<i>P. taetrolens</i> -pET30	Durante la noche	L-trp	300	1586
<i>P. taetrolens</i> -pET30	Durante la noche	L-trp	150	1658

Nuevamente, se observó que la duplicación de los volúmenes no escala a mayor actividad. Para futuros trabajos, se determinó la extracción de la proteína de Bugbuster lo más rápido posible después de la preparación de los extractos celulares y para almacenar la proteína en 50 mM de tampón de fosfato, pH 8 que contiene 0,01 mM de PLP. El detergente en Bugbuster puede inhibir la reacción o puede causar una pérdida de actividad tras el almacenamiento.

10

La inducción de la construcción pET30 se llevó a cabo nuevamente y el extracto celular se procesó con cromatografía de intercambio de aniones (como en el ejemplo 28) para proporcionar un extracto puro. El ensayo se repitió con esta preparación parcialmente purificada. Los números en el paréntesis en la columna de extracto de racemasa de la tabla 46 a continuación indican la cantidad aproximada de enzima de racemasa parcialmente purificada utilizada en el ensayo. Los resultados del ensayo se muestran en la tabla 46 a continuación.

15

Tabla 46

Fuente de enzimas	Intervalo de tiempo	Sustrato	Extracto de racemasa	D-trp conc (µg/mL)
KT2440	0	L-trp	75 µL (90 µg)	nd
NBRC 12996	0	L-trp	42 µL (200 µg)	nd
NBRC 12996	0	L-trp	21 µL (100 µg)	nd
<i>P. taetrolens</i>	0	L-trp	108 µL (200 µg)	nd
<i>P. taetrolens</i>	0	L-trp	54 µL (100 µg)	nd

Fuente de enzimas	Intervalo de tiempo	Sustrato	Extracto de racemasa	D-trp conc (µg/mL)
KT2440	2 horas	L-trp	75 µL (90 µg)	661
NBRC 12996	2 horas	L-trp	42 µL (200 µg)	408
NBRC 12996	2 horas	L-trp	21 µL (100 µg)	208
<i>P. taetrolens</i>	2 horas	L-trp	108 µL (200 µg)	862
<i>P. taetrolens</i>	2 horas	L-trp	54 µL (100 µg)	547
KT2440	durante la noche	L-trp	75 µL (90 µg)	2386
NBRC 12996	durante la noche	L-trp	42 µL (200 µg)	2382
NBRC 12996	durante la noche	L-trp	21 µL (100 µg)	1706
<i>P. taetrolens</i>	durante la noche	L-trp	108 µL (200 µg)	2029
<i>P. taetrolens</i>	durante la noche	L-trp	54 µL (100 µg)	2099

La no linealidad de la muestra durante la noche en este caso es probablemente debido al hecho de que las reacciones están alcanzando el equilibrio. Claramente, la BAR de *P. taetrolens* tiene una actividad significativa para la racemización de triptófano, como la tienen la 12996 BAR y KT2440 BAR. Parece que la KT2440 BAR y la BAR de *P. taetrolens* tienen una actividad similar, que es levemente mayor a la de la 12996 BAR.

5

La secuencia de ADN de la arginina racemasa de *P. taetrolens* se muestra a continuación como la SEQ ID NO:410. La secuencia de PCR proporcionó dos cambios en comparación con la secuencia de NCBI publicada. Específicamente, la secuencia de PCR contenía una adenosina en lugar de una guanina en la posición 902 y una citosina en lugar de una guanina en la posición 921. Estos cambios de ADN resultaron en una mutación silenciosa así como en un cambio de glicina a aspartato en la posición de aminoácido 301.

10

```

ATGCCCTTCTCCCGTACCCTGCTCGCCCTTCCCTTGGCAT
GGCATTGCTGCAAAACCCGGCCTTGTGTGCGCCACCCCTG
TCGATGACCGACGGCGTAGCTCAAGTGAATACCCAGGAC
AGCAATGCCTGGGTGCAAATCAATAAAGCCGCGTTTCGAG
CACAACATACGGACTCTGCAAACCGCCCTCGCCGGCAAG
TCGCAGATCTGCGCCGTAAGGCGGATGCCTATGGC
CACGGTATCGGCTTGTGATGCCCTCGGTGATCGCCATGG
GTGTTCCCTGTGTCGGTGTGCGCCAGCAACGAAGAAGCCC
GCGTTCGTGCGCGAGAGCGGTTTCAAGGGTCAACTGATAC
GCGTGCACCCGCTGCCCTGAGCGAACTGGAAGCTGCAC
TGCCGTACAACATGGAAGAGCTGGTGGGCAACCTGGACT
TCGCGGTCAAGGCCAGCCTGATTGCCGAGGATCACGGTC
GCCCCGTGGTGGTGCACCTGGGTCTGAATTCCAGCGGCA
TGAGCCGTAACGGAGTGGACATGACCACCGCTCAGGGCC

```

GTCGTGATGCGGTAGCTATCACCAAGGTGCCAAACCTGG  
 AAGTGC GGCGATCATGACCCACTTCGCGGT CGAAGATG  
 CTGCCGACGTGCGTGCCGGGCTCAAGGCCTCAATCAGC  
 AAGCCCAATGGCTGATGAACGTGGCCCAGCTTGATCGCA  
 GCAAGATCACCTGCACGCGGCCAACTCGTTCGCCACAC  
 TGGAGGTGCCC GAATCGCATCTGGACATGGTCCGCCCCG  
 GCGGCGCGCTGTTCCGGCGACACCGTACCGTCCCACACCG  
 AGTACAAGCGGGT CATGCAGTTCAAGTCCCACGTGGCGT  
 CGGTCAACAGCTACCCCAAGGGCAACACCGTCCGTTATG  
 ACCGCACGTACACCTGGGCCGCGACTCGCGGCTGGCCA  
 ACATCACCGTCCGCTACTCTGACGGCTACCGCCGCGCGTT  
 TACCAATAAAGGGATTGTGCTGATCAACGGCCATCGCGT  
 GCCAGTGGTGGGCAAAGTCTCGATGAACACCCTGATGGT  
 GGACGTC ACTGACGCGCCGGATGTGAAAAGCGGGCGATGA  
 AGTGGTGCTGTTCCGGGCACCAGGGCAAGGCCGAGATTAC  
 CCAGGCTGAGATCGAAGACATCAACGGTGC ACTGCTTGC  
 GGATCTGTATACCGTGTGGGGCAATCCAACCCTAAAAT  
 CCTGAAAGATCAGTAA

(SEQ ID NO: 410)

La proteína codificada por el gen de la SEQ ID NO:410 se analizó mediante el programa de predicción de péptidos señal Signal P 3.0 ([www.cbs.dtu.dk/services/SignalP/](http://www.cbs.dtu.dk/services/SignalP/)) y se predijo una secuencia líder de 23 aminoácidos.

#### 5 1384M Mutagénesis de BAR de *P. taetrolens*

La mutagénesis se llevó a cabo mediante el uso del kit de mutagénesis dirigida a múltiples sitios QuickChange (Stratagene, La Jolla, CA), mediante el uso del gen de BAR de *P. taetrolens* en el pET30 que resulta en una proteína sin marcar. Se utilizó el siguiente cebador mutagénico para realizar el cambio 1384M: 5'-TACCCAGGCTGAGATGGAAGACATCAACG-3' (SEQ ID NO:411).

10 La mutagénesis dirigida al sitio se realizó como se describe en el protocolo del fabricante. Se secuenciaron varios aislados (Agencourt, Beverly, MA) y se seleccionó un aislado con la secuencia correcta y se utilizó para el análisis adicional.

15 El plásmido se transformó en células BL21(DE3) (Novagen, Madison, WI). La proteína recombinante se produjo en un medio Overnight Express II (Novagen, Madison, WI) que contenía 50 µg/mL de kanamicina de acuerdo con los protocolos del fabricante. Los extractos libres de células se prepararon mediante el uso de BugBuster (Novagen, Madison, WI) de acuerdo con los protocolos del fabricante, se desalaron y se analizaron para verificar el porcentaje de expresión de la proteína diana mediante el uso del procedimiento Experion descrito anteriormente.

20 La totalidad de los ensayos de proteínas se realizó mediante el uso del kit BCA de Pierce (Pierce, Rockford, IL). Se realizaron ensayos de triptófano racemasa con la enzima mutante mediante el uso de la enzima de tipo salvaje preparada de la misma forma que un control positivo. Los ensayos contenían por mL: 30 mM de L-triptófano, 50 mM de fosfato de potasio, pH 8, 10 µM de PLP y aproximadamente 100 µg de proteína racemasa en un extracto libre de células. En el caso donde no se utilizaron 100 µg (con base en el porcentaje de expresión de Experion y los números de proteínas totales de Pierce), los resultados se normalizaron. Se tomaron muestras en el minuto 0, a los 30 minutos, a las 2 horas y durante la noche, se trataron con ácido fórmico al 2%, se filtraron y se diluyeron 1:10 para su análisis mediante el uso del procedimiento de aminoácido quiral descrito en el ejemplo 1.

25 La enzima de tipo salvaje pareció producir 49,1 ppm de D-triptófano en 30 minutos, mientras que el mutante 1384M produjo 108 ppm. El intervalo de tiempo de 2 horas fue similar, la enzima de tipo salvaje produjo 229,4 ppm D-triptófano contra 541,7 para el mutante 1384M. La mutación 1384M parece haber duplicado aproximadamente la actividad de BAR de *P. taetrolens*. El intervalo de tiempo durante la noche para el mutante también es mayor, pero a medida que las reacciones se aproximan al equilibrio la diferencia entre las actividades se reduce. Cuando se realizaron ensayos para la producción de monatina como en el ejemplo 28, el 1384M no pareció proporcionar ningún beneficio con respecto a la enzima de tipo salvaje de *P. taetrolens*.



**Ejemplo de referencia 30**

Ensayo de extracto de *A. caviae*

El ATCC 14486 de *Aeromonas caviae* se cultivó en un caldo de nutrientes a 37°C. Las células del cultivo (200 mL) se centrifugaron y se lavaron una vez con NaCl al 0,85%. Los sedimentos celulares se volvieron a suspender en 5 mL/g de peso de células húmedas de reactivo BugBuster™ (Novagen, Madison, WI) que contenía 5 µl/ml de la mezcla de inhibidor de proteasa del conjunto nº 3 (Calbiochem-Novabiochem Corp., San Diego, CA) y 1 µl/ml de benzonasa nucleasa. Las muestras se incubaron a temperatura ambiente durante 20 minutos en un agitador orbital. El residuo celular insoluble se eliminó por centrifugación a 16.000 x g durante 20 minutos a 4°C. El extracto libre de células se desaló sobre una columna PD-10 (GE Healthcare, Piscataway, NJ).

Los extractos libres de células se analizaron para verificar la actividad de la triptófano racemasa mediante el uso del siguiente protocolo. Se llevaron a cabo reacciones de 1 mL en 50 mM de fosfato de potasio (pH 8,0), 0,05 mM de PLP y 30 mM L-triptófano. Las reacciones se iniciaron con la adición de extractos libres de células (de 100 µL o 500 µL) y se incubaron a 30°C durante la noche. Se tomaron alícuotas de las muestras a las 2 horas y después de la incubación durante la noche (las muestras del minuto cero sirvieron como reacciones de control). Se agregó ácido fórmico concentrado (5 µL) a cada alícuota de muestra de 250 µL para detener la reacción y la proteína precipitada se eliminó mediante centrifugación. Los sobrenadantes se extrajeron y congelaron a -80°C hasta que se analizaron para detectar D-triptófano mediante el procedimiento de aminoácido quirál descrito en el ejemplo 1.

Los resultados del ensayo de los extractos celulares de *A. caviae* demostraron actividad de racemasa sobre L-triptófano como se muestra en la tabla 47.

**Tabla 47**

Tratamiento	Intervalo de tiempo	sustrato	Extracto racemasa de	D-trp conc (µg/mL)
<i>A. caviae</i>	0	L-trp	100 µL	nd
<i>A. caviae</i>	0	L-trp	500 µL	nd
<i>A. caviae</i>	2 horas	L-trp	100 µL	2
<i>A. caviae</i>	2 horas	L-trp	500 µL	19
<i>A. caviae</i>	durante la noche	L-trp	100 µL	45
<i>A. caviae</i>	durante la noche	L-trp	500 µL	130

Después de encontrar actividad en los extractos celulares de *A. caviae*, se diseñaron los cebadores degenerados (con base en las regiones conservadas de homólogos de BAR conocidos) para obtener el gen de BAR de estas especies. Las secuencias de cebador degenerado se muestran a continuación:

Aer deg F2: 5'-GCCAGCAACGARGARGCMCGCGT-3' (SEQ ID NO: 412); y

Aer deg R1: 5'-TGGCCSTKGATCAGCACA-3' (SEQ ID NO:413)

donde K indica G o T, R indica A o G, S indica C o G y M indica A o C.

Los cebadores que anteceden se utilizaron para amplificar la PCR a fragmentos de ADN de 715 bp de ADN genómico de *A. caviae* (ATCC 14486). Se utilizó el siguiente protocolo de PCR: La reacción de 50 µL contenía 0,5 µL de plantilla (aproximadamente 100 ng de ADN genómico de *A. caviae*), 1,6 µM de cada cebador, 0,3 mM de cada dNTP, 10 U de rT<sup>th</sup> Polimerasa XL (Applied Biosystems, Foster City, CA), 1×XL de tampón, 1 mM de Mg(OAc)<sub>2</sub> 2,5 µL de dimetil sulfóxido. El programa termociclador utilizado incluyó un comienzo en caliente a 94°C durante 3 minutos y 30 repeticiones de las siguientes etapas: 94°C durante 30 segundos, 53°C durante 30 segundos y 68°C durante 2 minutos. Después de las 30 repeticiones la muestra se mantuvo a 68°C durante 7 minutos y a continuación se almacenó a 4°C. Este protocolo PCR generó el producto 715 bp.

Clonación

Los productos de la PCR se purificaron en geles TAE-agarosa al 0,8% mediante el uso del kit de extracción en gel de Qiagen (Qiagen, Valencia, CA). El producto se clonó en TOPO y se transformó en células TOP 10 de acuerdo con el protocolo del fabricante (Invitrogen, Carlsbad, CA). El ADN plásmido se purificó a partir de los transformantes resultantes mediante el uso del kit spin miniprep de Qiagen (Qiagen, Valencia, CA) y se analizó para detectar las

inserciones correctas mediante enzimas de restricción con EcoR 1. Las secuencias de plásmidos que parecen tener una inserción correcta se verificaron mediante el secuenciado de ADN de terminación de la cadena didesoxi con los cebadores universales M13 directos.

- 5 La secuencia de ADN del producto de la PCR de *A. caviae* se muestra a continuación como la SEQ ID NO:414), donde las regiones de secuencia de cebador degenerado se encuentran subrayadas:

GCCAGCAACGARGARGCMCGCGTTGCCCGCGAGAAGGGCTTCC  
 AAGGTCGCCTGATGCGGGTACGTGCCGCCACCCCGGATGAAGT  
 GGAGCAGGCCCTGCCCTACAAGCTGGAGGAGCTCATCGGCAGC  
 CTGGAGAGCGCCAAGGGGATCGCCGACATCGCCAGCGCCATC  
 ACACCAACATCCCGGTGCACATCGGCCTGAACTCCGCCGGCAT  
 GAGCCGCAACGGCATCGATCTGCGCCAGGACGATGCCAAGGCC  
 GATGCCCTGGCCATGCTCAAGCTCAAGGGGATCACCCCGGTGCG  
 GCATCATGACCCACTTCCCGGTGGAGGAGAAAGAGGACGTCAA  
 GCTGGGGCTGGCCCAGTTCAAGCTGGACTACCAGTGGCTCATC  
 GACGCCGCAAGCTGGATCGCAGCAAGCTCACCATCCACGCCG  
 CCAACTCCTTCGCCACCCTGGAAGTACCGGAAGCCTACTTTGAC  
 ATGGTGCGCCCGGGCGGCATCATCTATGGCGACACCATTCCCTC  
 CTACACCGAGTACAAGAAGGTGATGGCGTTCAAGACCCAGGTC  
 GCCTCCGTCAACCACTACCCGGCGGGCAACACCGTCCGGCTATG  
 ACCGCACCTTACCCTCAAGCGCGACTCCCTGCTGGCCAACCTG  
 CCGATGGGCTACTCCGACGGCTACCGCCGCGCCATGAGCAACA  
 AGGCCTATGTGCTGATCMASGGCCA

(SEQ ID NO: 414), donde R indica A o G, S indica C o G y M indica A o C.

- 10 La secuencia de aminoácidos de la enzima parcial BAR de *A. caviae* se muestra como la SEQ ID NO:415 a continuación.

ASNEEARVAREKGFEGRLMRVRAATPDEVEQALPYKLEELI  
 GSLESAKGIADIAQRHHTNIPVHIGLNSAGMSRNGIDLQDD  
 AKADALAMLKLGITPVGIMTHFPVEEKEDVKLGLAQFKL  
 DYQWLIDAGKLDKSLTIHAANSFATLEVPEAYFDMVRPG  
 GIIYGDTPSYTEYKKVMAFKTQVASVNHYPAGNTVGYDRT  
 FTLKRDSLLANLPMGYSDGYRRAMSNKAYVLIXG

donde X es H, Q, N o K (SEQ ID NO:415).

- 15 El fragmento de proteína de consenso de la SEQ ID NO:415 es más de un 89% homólogo en el nivel de aminoácido con respecto a la secuencia TIGR publicada para *A. hydrophila*. Se esperaba que debido a que la proteína altamente relacionada de *Aeromonas hydrophila* mostraba una actividad de racemasa de amplia especificidad, así como los extractos celulares de *A. caviae*, la región codificadora de longitud completa para *A. caviae*, una vez obtenida, produciría una racemasa que también tendría una amplia especificidad con actividad sobre triptófano. Se utilizaron los procedimientos de paseo genómico para obtener la secuencia de genes de longitud completa del gen de BAR de *A. caviae* mostrado a continuación como la SEQ ID NO:416.

atgcacaaga aaactgct cgcgaccctg atctttggcc tgctggccgg ccaggcagtc  
 gccgccccct atctgccgt cgccgacgac caccgcaacg gtcaggaaca gaccgccgcc  
 aacgcctggc tggaaagtga tctcggcgcc ttcgagcaca acatccagac cctgaagaat  
 cgctcgggtg acaagggccc gcagatctgc gccatcatga aggcggacgc ctacggtcac  
 ggcatcgacc tgctggtccc ttccgtggtc aaggcaggca tcccctgcat cggcatcgcc  
 agcaacgaag aagcacgtgt tccccgcgag aagggtctcg aaggtcgctt gatgctggga  
 cgtgccgcca ccccgatga agtggagcag gccctgccct acaagctgga ggagctcacc  
 ggagcctgg agagcgcca ggggatcgcc gacatcgccc agcgccatca caccaacatc  
 ccggtgcaca tggcctgaa ctccgcccgc atgagccgca acggcatcga tctgcccag  
 gacgatgcca aggcgatgc cctggccatg ctcaagctca aggggatcac cccggtcggc  
 atcatgacc acttcccgtt ggaggagaaa gaggacgtca agctggggct ggcccagttc  
 aagctggact accagtggct catcgacgcc ggcaagctgg atcgagcaa gctcaccatc  
 cacgccgcca actccttccg caccctggaa gtaccggaag cctactttga catggtgcgc  
 ccggcgccga tcatctatgg cgacaccatt cctctctaca ccgagtaca gaaggtgatg  
 gcggtcaaga cccaggtcgc ctccgtcaac cactaccggc cgggcaacac cgtcggctat  
 gaccgcacct tcacctcaa gcgcgactcc ctgctggcca acctgccgat gggtactcc  
 gacggctacc gccgcgcat gagcaacaag gcctatgtgc tgatccatgg ccagaaggcc  
 cccgtcgtgg gcaagacttc catgaacacc accatggtgg acgtcaccga catcaagggg  
 atcaaacccg gtagcaggtt ggtcctgttc ggacgccagg gtagtccga ggtgaaacaa  
 tctgatctgg aggagtacaa cgggtgccctc ttggcggaca tgtacaccgt ctggggctat  
 accaaccca agaagatcaa gcgctaa

(SEQ ID NO: 416).

La secuencia de aminoácidos para la BAR nativa de *A. caviae* se muestra a continuación como la SEQ ID NO:417.

5 1 mhkktllatl ifgllagqav aapylpladd hrngqeetaa  
 41 nawlevdlga fehniqtln rlgdkgpqic aimkadaygh  
 81 gidllvpsv kagipcgia sneearvare kgfegrmlrv  
 121 raatpdeveq alpykleeli gslesakgia diaqrhhtni  
 161 pvhiglnsag msrngidlrq ddakadalam lkkgitpvg  
 10 201 imthfpveek edvklglaqf kldyqwlida gklrsklti  
 241 haansfatle vpeayfdmvr pggiiygdti psyteykkvm  
 281 afktqvasvn hypagntvgy drtflkrds llanlpmgys  
 321 dgyrramsnk ayvlihgqka pvvgktsmnt tmvdvtdikg  
 361 ikpgdevlfl grqgdaevkq sdleeyngal ladmytwggy  
 15 401 tnpkkikr.

(SEQ ID NO: 417)

Se predice que los primeros 21 residuos de aminoácidos con terminal N de la SEQ ID NO:417 son un péptido señal mediante el uso del programa Signal P 3.0 ([www.cbs.dtu.dk/services/SignalP/](http://www.cbs.dtu.dk/services/SignalP/)). Las pruebas experimentales confirmaron que el producto de expresión se secretó en el periplasma de *E. coli* y el péptido señal se escindió como se predijo. Cuando se clonó y expresó el gen de longitud completa mediante el uso de los procedimientos descritos anteriormente, se encontró que tenía una actividad comparable, pero mayor que la de la BAR de *P. taetrolens*.

### Ejemplo 31

Producción de la aldolasa de la SEQ ID NO: 276 en un huésped de expresión alternativo

El gen de la SEQ ID NO: 275 se subclonó mediante el uso de procedimientos de biología molecular estándar en un derivado del vector pET23d (Novagen, Madison, WI) que contenía el gen *metE* de *E. coli* y el promotor insertado en el sitio de restricción NgoMIV y un segundo sitio de restricción *pis* que se agregó para una fácil extracción del gen beta lactamasa (*baa*). La construcción de este vector que contenía una inserción para un gen de oxigenasa mio-inositol se describe en PCT WO2006066072 en los ejemplos 2 y 20. La inserción de aldolasa se confirmó mediante el secuenciado de ADN (Agencourt Bioscience Corporation; Beverly, MA) y el plásmido con la secuencia de inserción correcta se transformó en el receptor de expresión BW30384(DE3) $\Delta$ ompT $\Delta$ metE de *E. coli*. La construcción de este huésped de expresión y el protocolo de transformación también se describen en la PCT WO2006066072 (ejemplos 21 y 22). El gen de la aldolasa se expresó mediante el uso del sistema de autoinducción Overnight Express™ II de Novagen (Novagen, Madison, WI) que contenía 100 mg/L de ampicilina. Este sistema se describió en el ejemplo 24 para la expresión de la D-alanina aminotransferasa de *Bacillus sphaericus* (cepa ATCC 10208). Los extractos libres de células que contenían la aldolasa se produjeron mediante el uso del reactivo de extracción BugBuster™ de Novagen (libre de amina primaria) (Novagen, Madison, WI) que contenía 1  $\mu$ L/mL de benzonasa nucleasa, 0,033  $\mu$ L/mL de r-Lisozima y 5  $\mu$ L/mL de mezcla de inhibidor de proteasa conjunto II de acuerdo con el protocolo recomendado del fabricante para la lisis celular. Las proteínas solubles en los extractos libres de células se separaron en una estación de electroforesis automatizada Experion de Bio-Rad (Bio-Rad, Hercules, CA) y se analizaron para verificar el porcentaje de expresión de proteínas solubles mediante el uso del software Experion versión 1.1.98.0 como se describe en el ejemplo 12.

Para intentar mejorar la expresión de aldolasa en *E. coli*, los codones dos a siete de la secuencia codificadora de ADN se mutaron para los cambios sugeridos en el análisis de la secuencia de tipo salvaje mediante el uso del algoritmo HY de *E. coli* ProteoExpert RTS de Roche. Estos cambios se realizaron mediante el uso del kit de mutagénesis dirigida a múltiples sitios QuikChange® o el kit de mutagénesis dirigida al sitio QuikChange® II XL (Stratagene, La Jolla, CA) de acuerdo con el protocolo sugerido del fabricante con 0,75  $\mu$ L de solución Quik cada 25  $\mu$ L de mezcla de reacción y la transformación a células ultracompetentes XL10-Gold®. Las mutaciones se generaron mediante el uso de cebadores (cada 45 nucleótidos de longitud) diseñados con el programa en la web de diseño de cebadores QuikChange® de Stratagene disponible en línea en [www.stratagene.com.qcprimerdesign](http://www.stratagene.com.qcprimerdesign).

Tabla 48

	Codón 1	Codón 2	Codón 3	Codón 4	Codón 5	Codón 6	Codón 7
Tipo salvaje	ATG	CCT	ATC	GTT	GTT	ACG	AAG
ProteoExpert #1	ATG	CCA	ATT	GTT	GTA	ACT	AAA
ProteoExpert #2	ATG	CCA	ATT	GTT	GTT	ACT	AAA
ProteoExpert #6	ATG	CCA	ATT	GTT	GTA	ACC	AAA
ProteoExpert #10	ATG	CCA	ATT	GTT	GTT	ACC	AAA

Las secuencias del gen de aldolasa con los cambios de codones anteriores se transformaron en el huésped de expresión BW30384(DE3) $\Delta$ ompT $\Delta$ metE de *E. coli* y después se expresaron mediante el uso del sistema de autoinducción 2 Novagen Overnight Express™ (Novagen, Madison, WI). Ninguna de estas mutaciones resultó en niveles más altos de expresión de genes al compararlas con la secuencia de tipo salvaje. Los resultados típicos de los extractos libres de células se analizaron mediante el software 1.1.98.0 Experion Pro260 de Bio-Rad mostrado en la tabla 49 a continuación, donde la columna titulada expresión de proteínas muestra los valores para el % total de proteínas solubles.

Los extractos libres de células (0,025 mg/mL de proteínas solubles por ensayo) se analizaron para verificar su capacidad de producir monatina R,R a partir de 200 mM de piruvato de sodio y 100 mM de D-triptófano mediante el uso del protocolo descrito en el ejemplo 7. Las reacciones (volumen total de 7 mL) se llevaron a cabo en tubos de polipropileno de 14 mL en una cámara anaerobia mediante el uso de D-aminotransferasa purificada de *B. sphaericus* (ATCC 10208) a una concentración final de 2 mg/mL. La concentración de monatina producida a la 1, 4 y 20 horas después de la adición de las enzimas se muestra en la tabla 49 a continuación. La tabla 49 muestra que el extracto libre de células generado a partir de la construcción que contenía la secuencia de aldolasa de tipo salvaje produjo una concentración de monatina a las 20 horas levemente mayor a la de los extractos libres de células a partir de las construcciones que tenían las mutaciones ProteoExpert. Las concentraciones de monatina se determinaron mediante el procedimiento de LC/MS/MS descrito en el ejemplo 1.

Tabla 49

	Expresión de proteínas (%)	[Monatina] mM-1 hora	[Monatina]; mM-4 horas	[Monatina] mM-20 horas
Tipo salvaje	22	2,1	4,2	13,4
ProteoExpert #1	18	1,3	3,1	12,4
ProteoExpert #2	18	2,1	4,7	11,6
ProteoExpert #6	18	1,4	1,9	12,4
ProteoExpert #10	20	1,8	1,8	11,7

Estos datos demuestran que la aldolasa de la SEQ ID NO: 276 puede producirse en un huésped de expresión alternativo sin inducción de IPTG.

El gen *bla* se extrajo del vector y las fermentaciones subsiguientes para producir aldolasa se realizaron sin el uso de antibiótico como se describe en la PCT WO2006066072 para otra enzima. Los niveles de expresión en las fermentaciones por lote alimentado a 30°C alcanzaron un máximo a las 6 a 8 horas después de la inducción, lo que produjo la aldolasa de la SEQ ID NO: 276 al 25 al 30% de proteína soluble, de acuerdo con los datos de Experion. Los estudios de estabilidad no mostraron ninguna pérdida aparente de actividad de aldolasa cuando el producto de fermentación se dejó durante 6 horas a 15 o 30°C (en condiciones de oxígeno limitado así como en condiciones aireadas) antes de la concentración y disrupción celular. Se encontró que la aldolasa era igualmente estable almacenada como un extracto libre de células o como un sedimento celular cuando se almacenó durante 5 días a -80°C. No se requirió el lavado del sedimento celular antes del almacenamiento a -80°C y en realidad causó una pequeña disminución en la actividad. Se encontró que las células que se volvieron a suspender en agua destilada, sobrenadante de fermentación o 100 mM de tampón de fosfato de potasio (pH 7,8) no presentaron pérdida en la actividad o en la concentración de proteínas (a juzgar por SDS-PAGE) cuando se almacenaron durante 11 días a temperatura ambiente o a 4°C. El extracto libre de células producido en tampón de fosfato de potasio no mostró ninguna pérdida de actividad de aldolasa al almacenarse a 4°C o a temperatura ambiente durante 5 días. Las células pueden abrirse en un tampón de fosfato, hasta un 25% de sobrenadante o agua con recuperación comparable de actividad de aldolasa; no obstante, se encontró que la adición de 1 mM de MgCl<sub>2</sub> a agua mejora ligeramente la actividad de la aldolasa. Estos datos muestran que la proteína aldolasa es lo suficientemente estable para ser comercialmente útil.

Se describieron varias realizaciones de la invención. Las realizaciones de la invención incluyen uno o más de los aspectos descritos anteriormente. Otras realizaciones se encuentran dentro del alcance de las siguientes reivindicaciones.

LISTA DE SECUENCIAS

5	<210> 275 <211> 690 <212> ADN <213> Desconocido	
10	<220> <223> ADN obtenido de una muestra del entorno	
	<400> 275	
	atgcctatcg ttgttacgaa gatcgaccga cccagcgcgg cggacgtcga aaggatcgcc	60
	gcctatggtg tcgcgacctt gcatgaagcg caaggacgaa ccgggttgat ggcgtccaat	120
	atgcgccc aa tctatcgccc tgcgcacatt gccgggcccc cggtgacctg ccttggtggc	180
	cctggcgaca attggatgat ccatgtcgcc gtcgaacagt gccagccggg agatgtcctg	240
	gtcgtggtac cgaccagccc ctgcgaagac ggctatttcg gcgatctgct ggcgacctcg	300
	ctgcggtcgc gcgggggtcaa aggtctgatc atcgaggccg gcgtacgca tctcgcgaca	360
	ttgaccgaga tgaattccc ggtctggtcc aaggcgggtg tcgcgcaagg aacgggtcaag	420
	gagaccatcg ccagcgtcaa tgtgccccctc gtctgcgcgg gcgcccgcac cgtgcccggc	480
	gatctgatcg ttgccgacga cgacggggtc gtcgtgatcc caagacgttc cgttccggcg	540
	gtcctttcca gcgccgaggc ccgcgaagag aaggaagccc gcaaccgcgc ccgcttcgaa	600
	gctggcgagc tgggcctcga cgtctacaac atgcgccagc gcctggccga caagggttg	660
	cgctatgtcg agcggctgcc cgaggaatag	690
15	<210> 276 <211> 229 <212> PRT <213> Desconocido	
20	<220> <223> Proteína obtenida de una muestra del entorno	
25	<220> <221> DOMAIN <222> (20) ... (172) <223> Dimetilmenaquinona metiltransferasa	
	<400> 276	

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

**REIVINDICACIONES**

1.- Un ácido nucleico aislado, sintético o recombinante que comprende

- 5 (a) un ácido nucleico que codifica al menos un polipéptido, en donde el ácido nucleico comprende una secuencia que tiene al menos 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más o 100% de identidad de secuencia con respecto a la secuencia de SEQ ID NO:275 en la longitud total de un ADNc, transcripto (ARNm) o gen,

10 en donde el ácido nucleico codifica al menos un polipéptido que tiene una actividad de aldolasa o codifica un polipéptido o péptido que actúa como un epítipo o inmunógeno y es capaz de generar un anticuerpo específico de aldolasa;

- 15 (b) un ácido nucleico que codifica al menos un polipéptido, en donde el ácido nucleico comprende una secuencia que se hidrida en condiciones rigurosas con el complemento de un ácido nucleico que comprende la secuencia de SEQ ID NO:275, en donde el ácido nucleico codifica al menos un polipéptido que tiene una actividad de aldolasa o codifica un polipéptido o péptido que actúa como un epítipo o inmunógeno y es capaz de generar un anticuerpo específico de aldolasa;

(c) un ácido nucleico que codifica al menos un polipéptido que tiene una actividad de aldolasa, en donde el polipéptido comprende la secuencia de SEQ ID NO:276 o fragmentos enzimáticamente activos de la misma;

(d) el ácido nucleico de (a), (b) o (c) que codifica un polipéptido que tiene una actividad de aldolasa pero que carece de una secuencia señal,

20 (e) el ácido nucleico de (a), (b), (c) o (d) que codifica un polipéptido que tiene una actividad de aldolasa que comprende además una secuencia heteróloga; o

(f) una secuencia de ácidos nucleicos complementaria del ácido nucleico de (a), (b), (c), (d) o (e).

2.- Un par de cebadores de amplificación, en donde el par de cebadores de amplificación

25 (a) es capaz de amplificar un ácido nucleico que comprende una secuencia de la reivindicación 1 o una subsecuencia de la misma, o

(b) es el par de cebadores de (a), en donde un miembro del par de cebadores comprende un oligonucleótido que comprende al menos 10 a 50 bases consecutivas de la secuencia de la reivindicación 1 o 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 o

30 (c) es el par de cebadores de (b), que comprende un primer miembro que tiene una secuencia de los primeros (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 la reivindicación 1, y un segundo miembro que tiene una secuencia de los primeros (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 hebra complementaria del primer miembro.

35 3.- Un cassette de expresión, vector o vehículo de clonación que comprende un ácido nucleico de la reivindicación 1.

4.- Una célula transformada que comprende un ácido nucleico de la reivindicación 1 o que comprende un cassette de expresión, vector o vehículo de clonación de la reivindicación 3.

40 5.- Un animal no humano, planta, parte de planta o semilla transgénicos que comprende un ácido nucleico de la reivindicación 1 o que comprende un cassette de expresión, vector o vehículo de clonación de la reivindicación 3 o una célula transformada como se establece en la reivindicación 4.

6.- Un polipéptido o péptido aislado, sintético o recombinante que tiene una actividad de aldolasa que comprende

45 (a) una secuencia de aminoácidos que tiene al menos 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más o 100% de identidad de secuencia con respecto a SEQ ID NO:276 en la longitud total del polipéptido o enzima, y/o fragmentos enzimáticamente activos de la misma;

(b) una secuencia de aminoácidos codificada mediante un ácido nucleico de la reivindicación 1, en donde el polipéptido tiene

(i) una actividad de aldolasa o



- (ii) tiene actividad inmunogénica ya que es capaz de generar un anticuerpo que se une específicamente a un polipéptido que tiene una secuencia de (a), y/o fragmentos enzimáticamente activos de la misma;
- (c) el polipéptido de (a) o (b) que tiene una actividad de aldolasa pero que carece de una secuencia señal;
- 5 (d) el polipéptido de (a), (b) o (c) que tiene una actividad de aldolasa y que comprende además una secuencia heteróloga; o
- (e) el polipéptido de (a), (b), (c) o (d), en donde el polipéptido comprende al menos un sitio de glicosilación.
- 7.- Un anticuerpo aislado, sintético o recombinante que se une específicamente a un polipéptido de la reivindicación 6.
- 8.- Un método para producir un polipéptido recombinante que comprende las etapas de:
- 10 (a) proporcionar un ácido nucleico de la reivindicación 1; y
- (b) expresar el ácido nucleico de la etapa (a) en condiciones que permitan la expresión del polipéptido, produciendo así un polipéptido recombinante.
- 9.- Un método para generar una variante de un ácido nucleico que codifica un polipéptido con una actividad de aldolasa que comprende las etapas de:
- 15 (a) proporcionar un ácido nucleico plantilla que comprende un ácido nucleico de la reivindicación 1; y
- (b) modificar, eliminar o agregar uno o más nucleótidos en la secuencia plantilla o una combinación de los mismos, para generar una variante del ácido nucleico plantilla.
- 10.- El método de la reivindicación 9, que es un método para modificar codones en un ácido nucleico que codifica un polipéptido de aldolasa, comprendiendo el método las siguientes etapas:
- 20 (a) proporcionar un ácido nucleico de la reivindicación 1; y
- (b) identificar un codón en el ácido nucleico de la etapa (a) y reemplazarlo por un codón diferente que codifica el mismo aminoácido que el codón reemplazado, modificando así codones en un ácido nucleico que codifica una aldolasa.
- 11.- Una secuencia señal o líder aislada, sintética o recombinante que consiste en una secuencia de aminoácidos como se establece en los residuos amino terminales 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, 1a 25, 1 a 26, 1 a 27, 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 una secuencia de aminoácidos de la reivindicación 6.
- 12.- Un método para escindir un enlace carbono-carbono en una composición que comprende las siguientes etapas:
- 30 (a) proporcionar un polipéptido de la reivindicación 6 o un polipéptido codificado por un ácido nucleico de la reivindicación 1;
- (b) proporcionar una composición que comprende un enlace carbono-carbono; y
- (c) poner en contacto el polipéptido de la etapa (a) con la composición de la etapa (b) en condiciones en las que la aldolasa escinde el enlace carbono-carbono en la composición.
- 13.- Un método para formar un enlace carbono-carbono que comprende las siguientes etapas:
- 35 (a) proporcionar un polipéptido de la reivindicación 6 o un polipéptido codificado por un ácido nucleico de la reivindicación 1;
- (b) proporcionar un donante y un compuesto aceptor; y
- (c) poner en contacto el polipéptido de la etapa (a) con los compuestos de la etapa (b) en condiciones en las que la aldolasa forma el enlace carbono-carbono.
- 40 14.- El método de la reivindicación 13 que es un método para obtener un ácido D-glutámico 4-sustituido que comprende las siguientes etapas:
- (a) proporcionar un polipéptido de la reivindicación 6 o un polipéptido codificado por un ácido nucleico de la reivindicación 1;
- (b) proporcionar un aceptor de  $\alpha$ -cetoácido y un piruvato o un donante de piruvato; y

(c) poner en contacto el polipéptido de la etapa (a) con los compuestos de la etapa (b) en condiciones en las que la aldolasa cataliza la síntesis de un ácido D-glutámico 4-sustituido.

15.- El método de la reivindicación 13 que es un método para obtener un 2-ceto-glutarato 3,4-sustituido que comprende las siguientes etapas:

5 (a) proporcionar un polipéptido de la reivindicación 6 o un polipéptido codificado por un ácido nucleico de la reivindicación 1;

(b) proporcionar un donante y un compuesto aceptor, en donde opcionalmente el donante y el aceptor son un piruvato o un donante de piruvato y un aceptor de  $\alpha$ -cetoácido, una cetona y/o un aldehído; y

10 (c) poner en contacto el polipéptido de la etapa (a) con los compuestos de la etapa (b) en condiciones en las que la aldolasa cataliza la síntesis de un 2-ceto-glutarato 3,4-sustituido.

16.- Una composición que comprende un polipéptido de la reivindicación 6 o un polipéptido codificado por un ácido nucleico de la reivindicación 1.

17.- Un método para obtener un alimento, pienso o bebida que comprende poner en contacto una composición que comprende un compuesto que contiene un enlace carbono-carbono con un polipéptido de la reivindicación 6 o un polipéptido codificado por un ácido nucleico de la reivindicación 1.

18.- Un método para la conversión de una biomasa o cualquier material lignocelulósico en un combustible que comprende poner en contacto la biomasa o material lignocelulósico con un polipéptido de la reivindicación 6 o un polipéptido codificado por un ácido nucleico de la reivindicación 1 o un fragmento enzimáticamente activo del mismo.

20

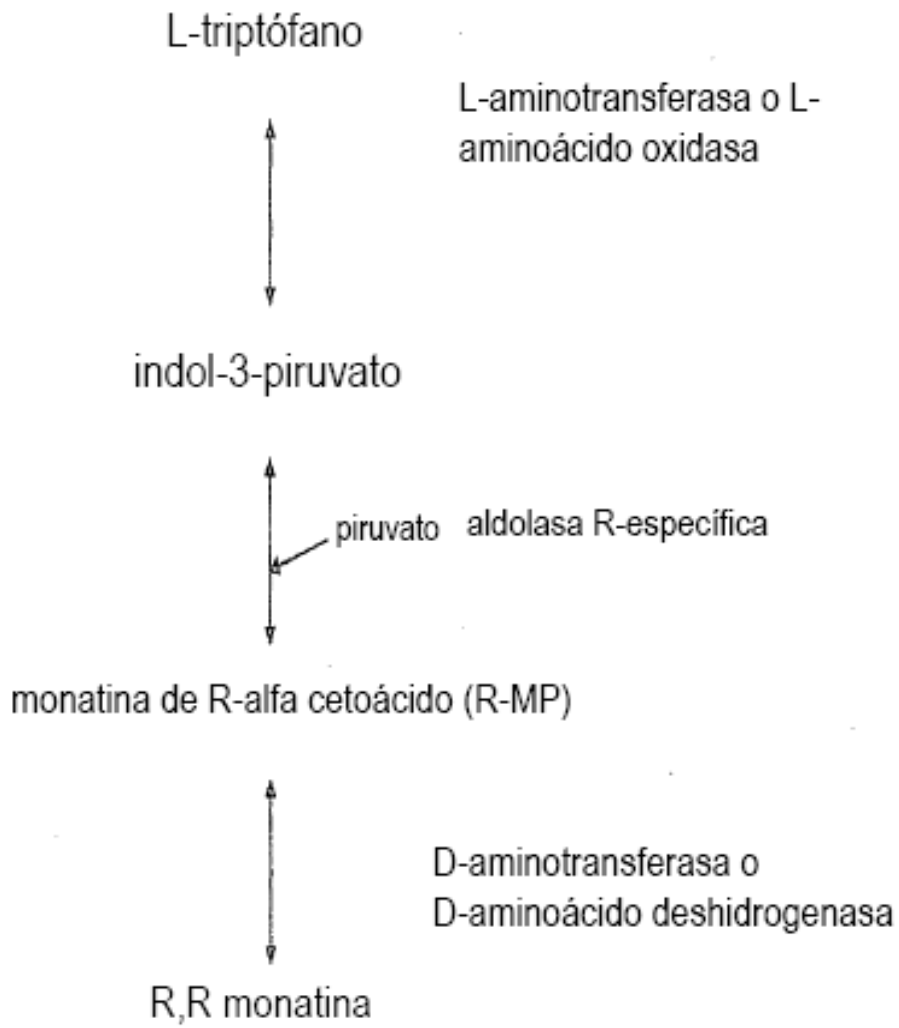


FIGURA 1

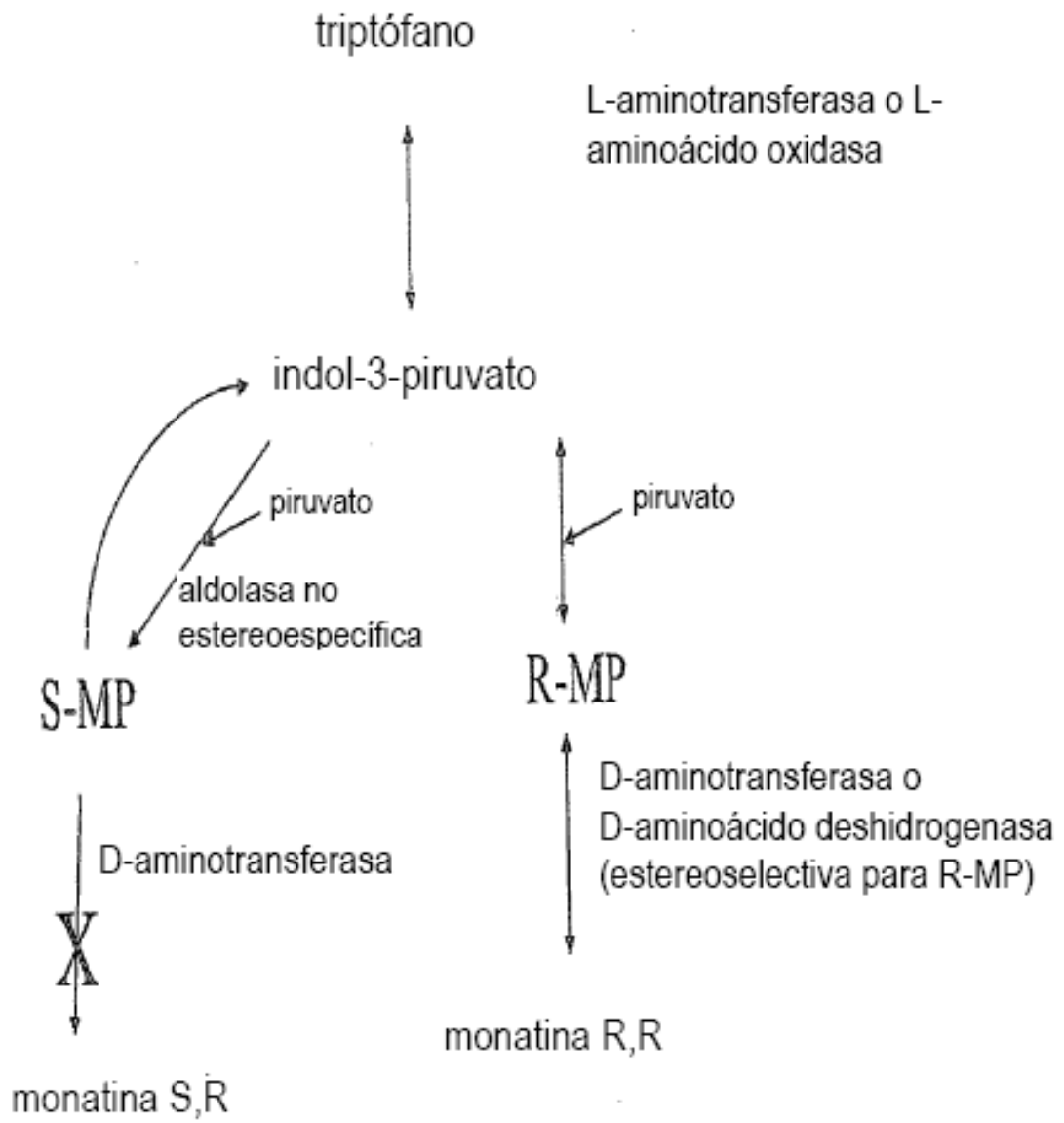


FIGURA 2

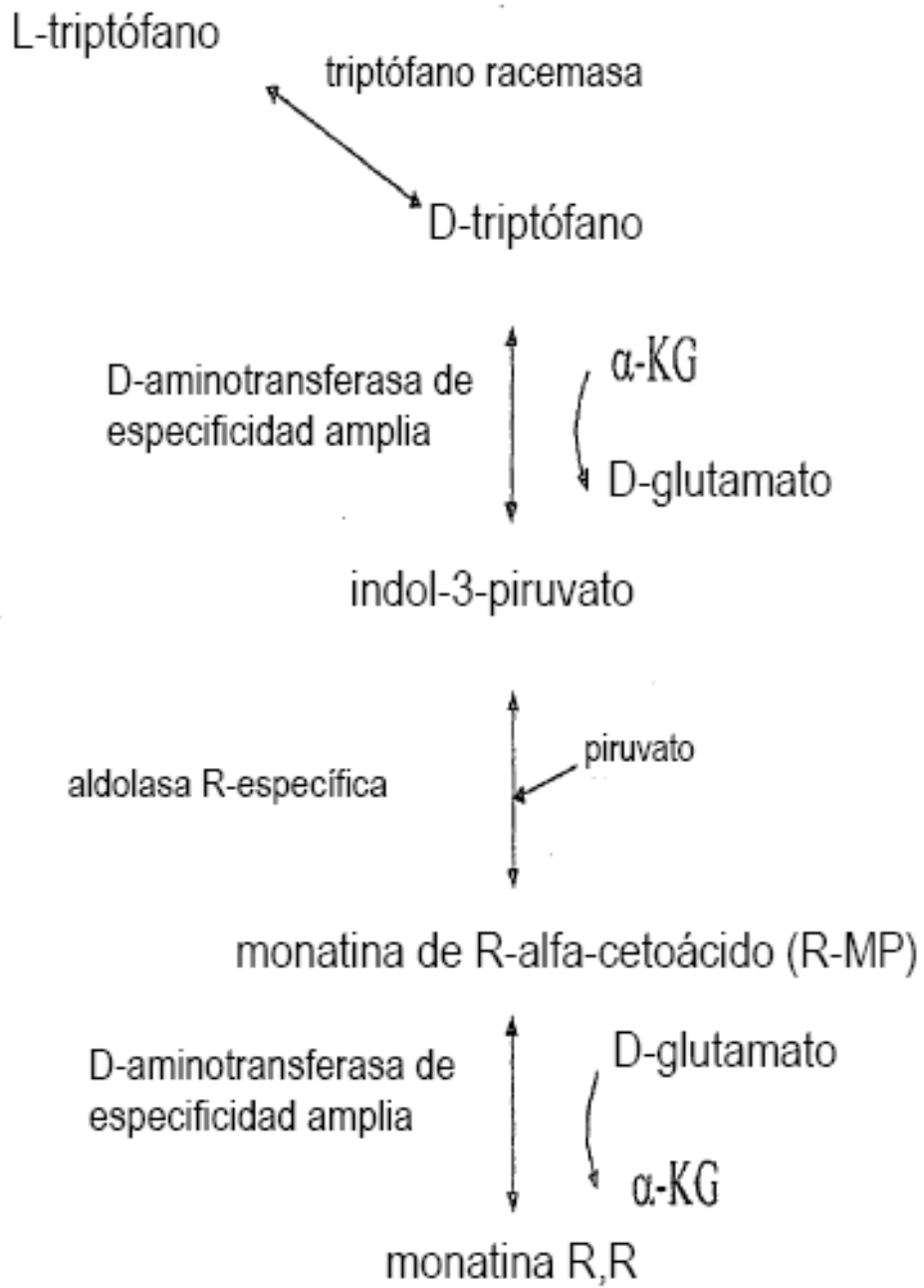


FIGURA 3

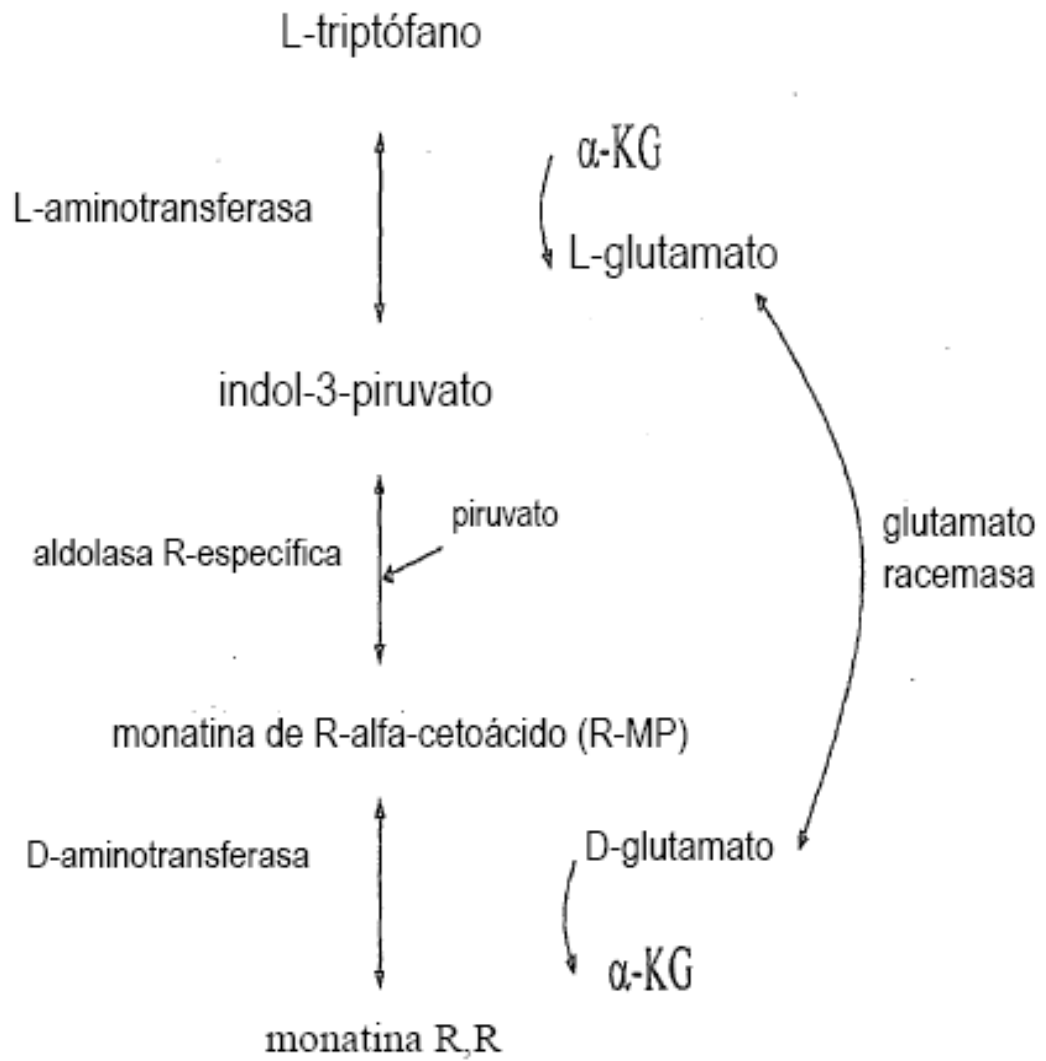


FIGURA 4

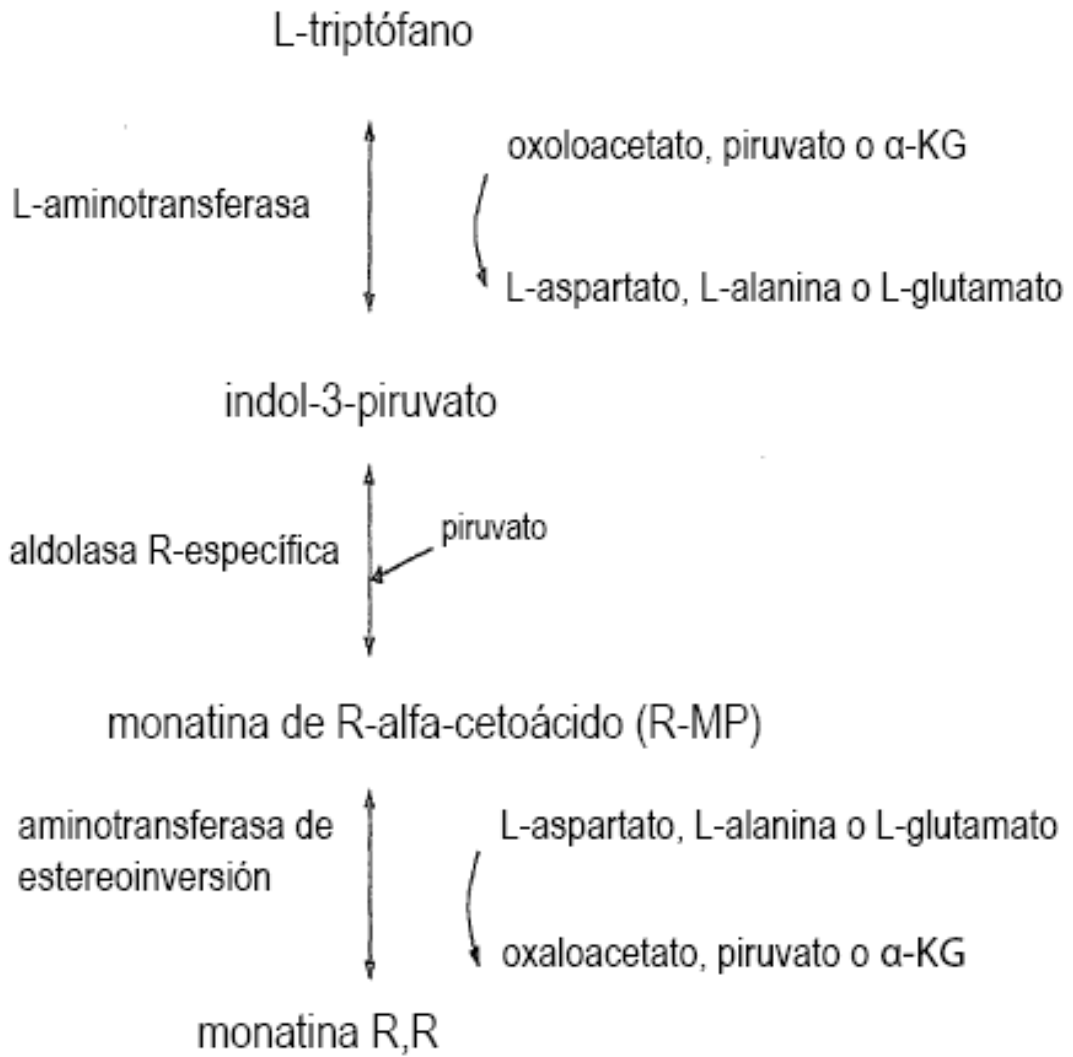


FIGURA 5

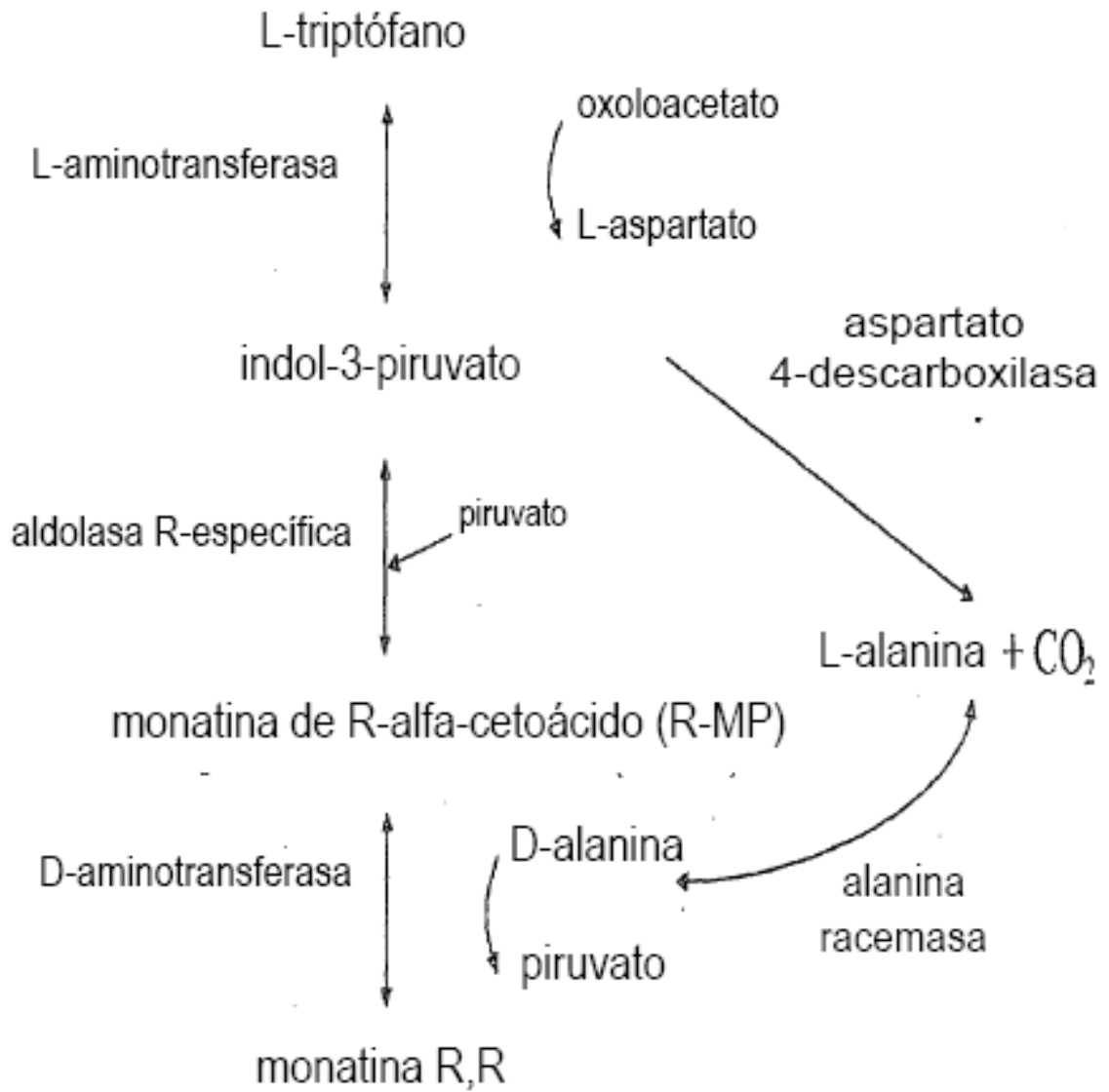


FIGURA 6



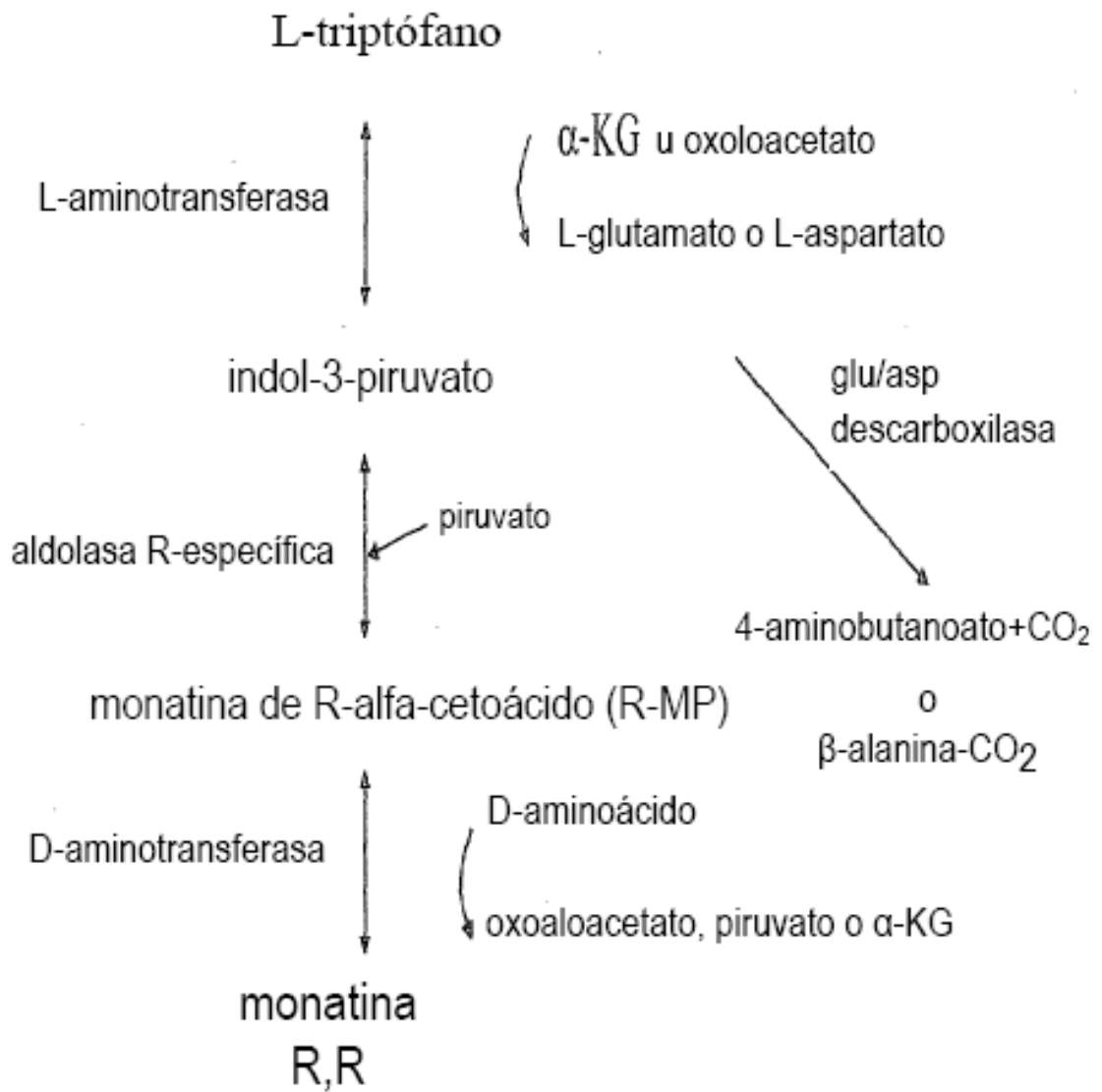


FIGURA 7

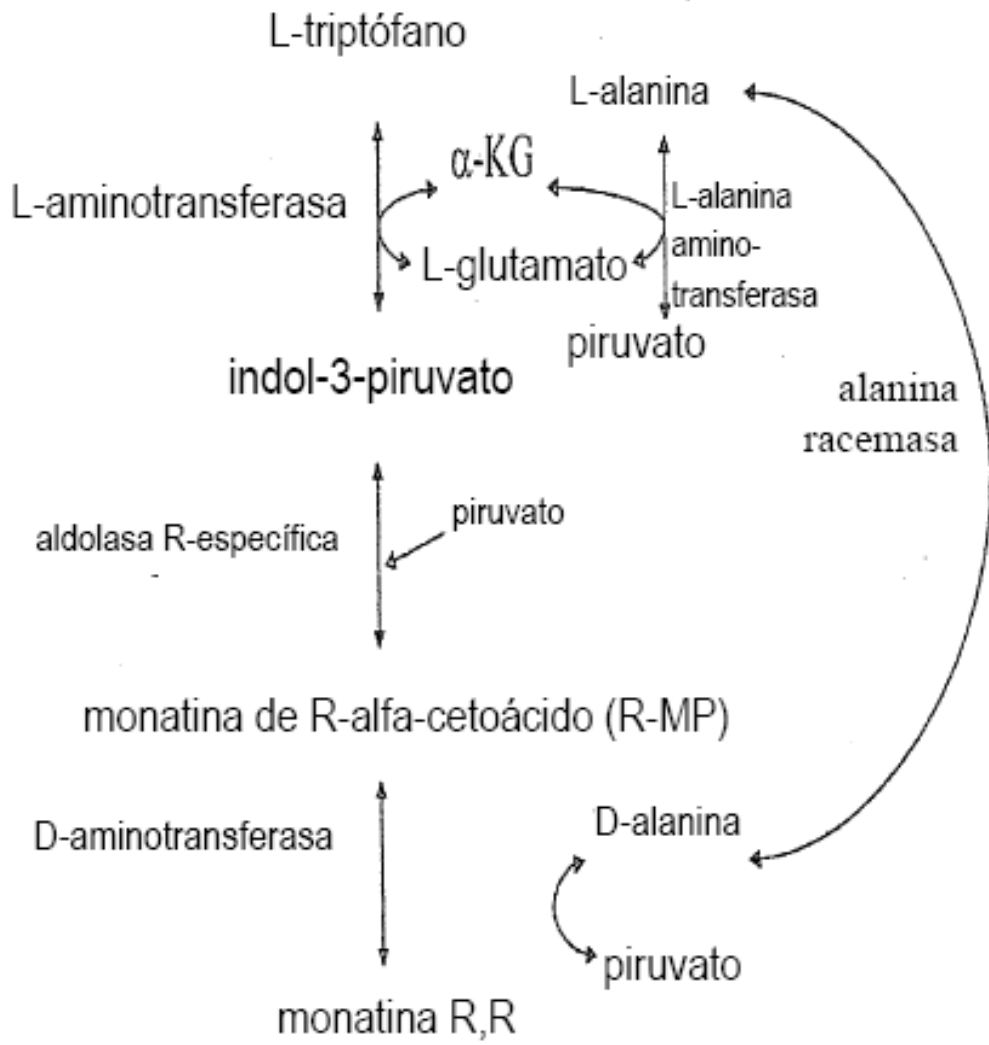


FIGURA 8

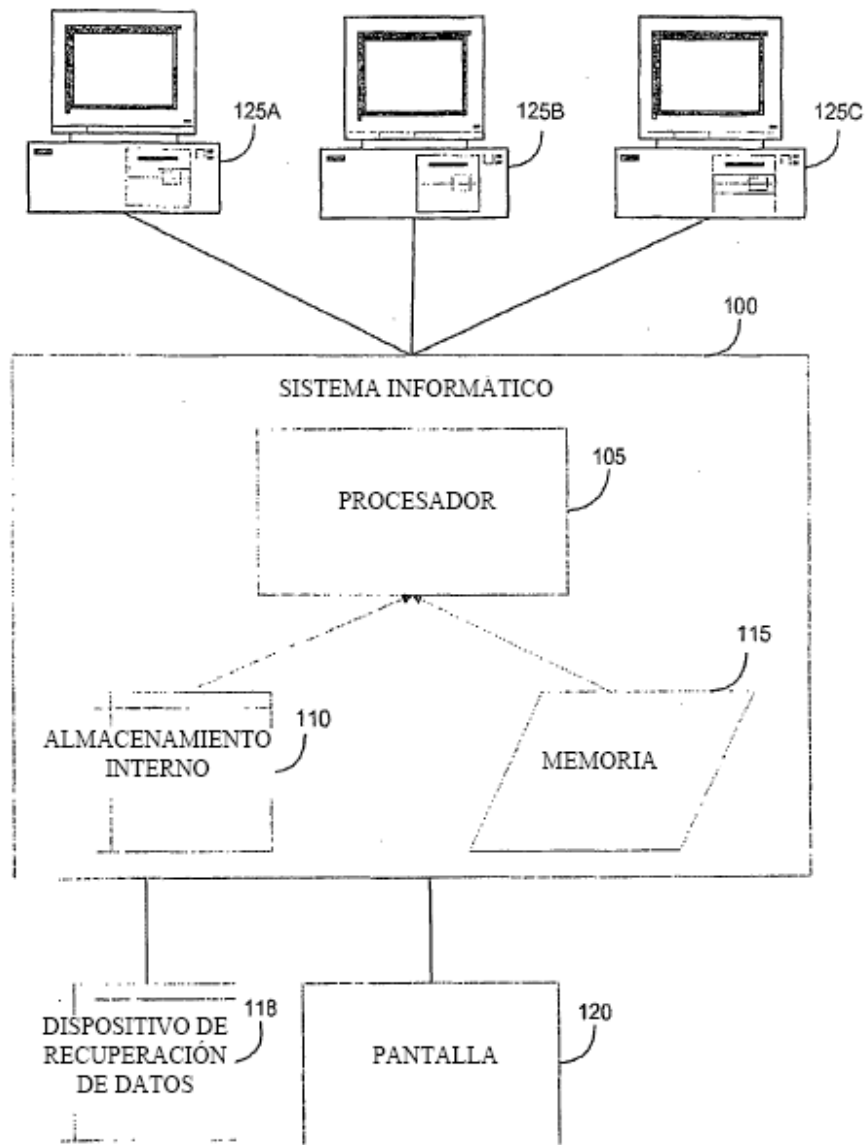


FIGURA 9

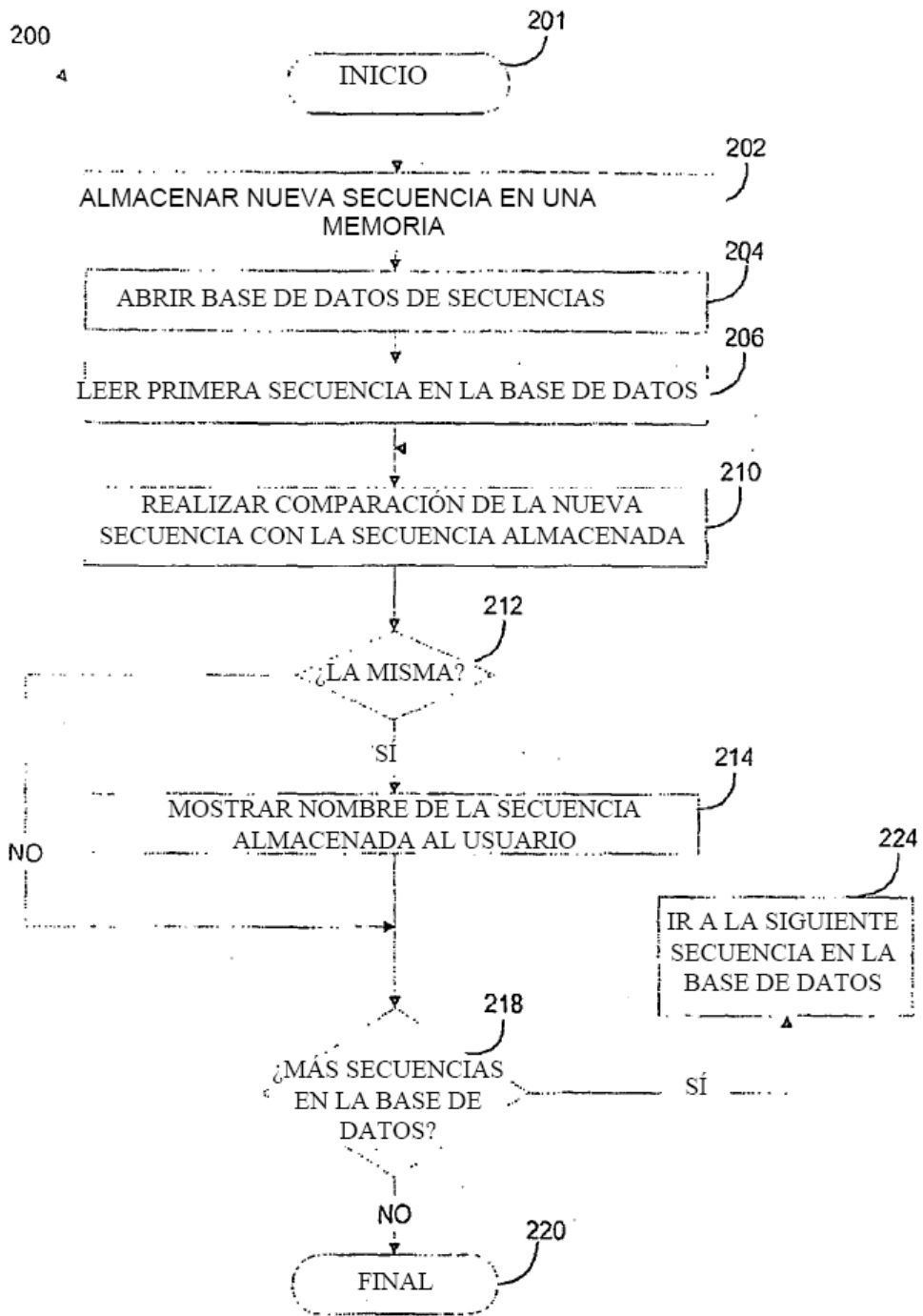


FIGURA 10

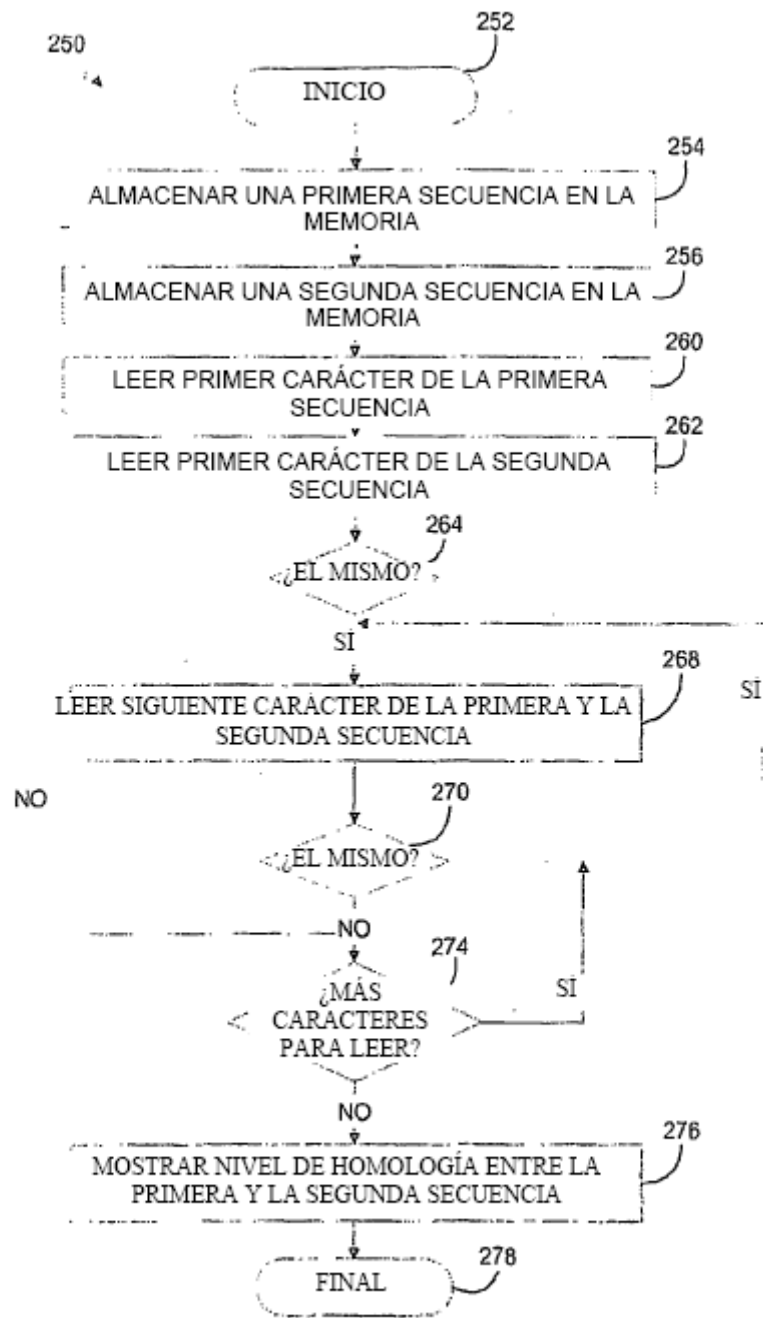


FIGURA 11

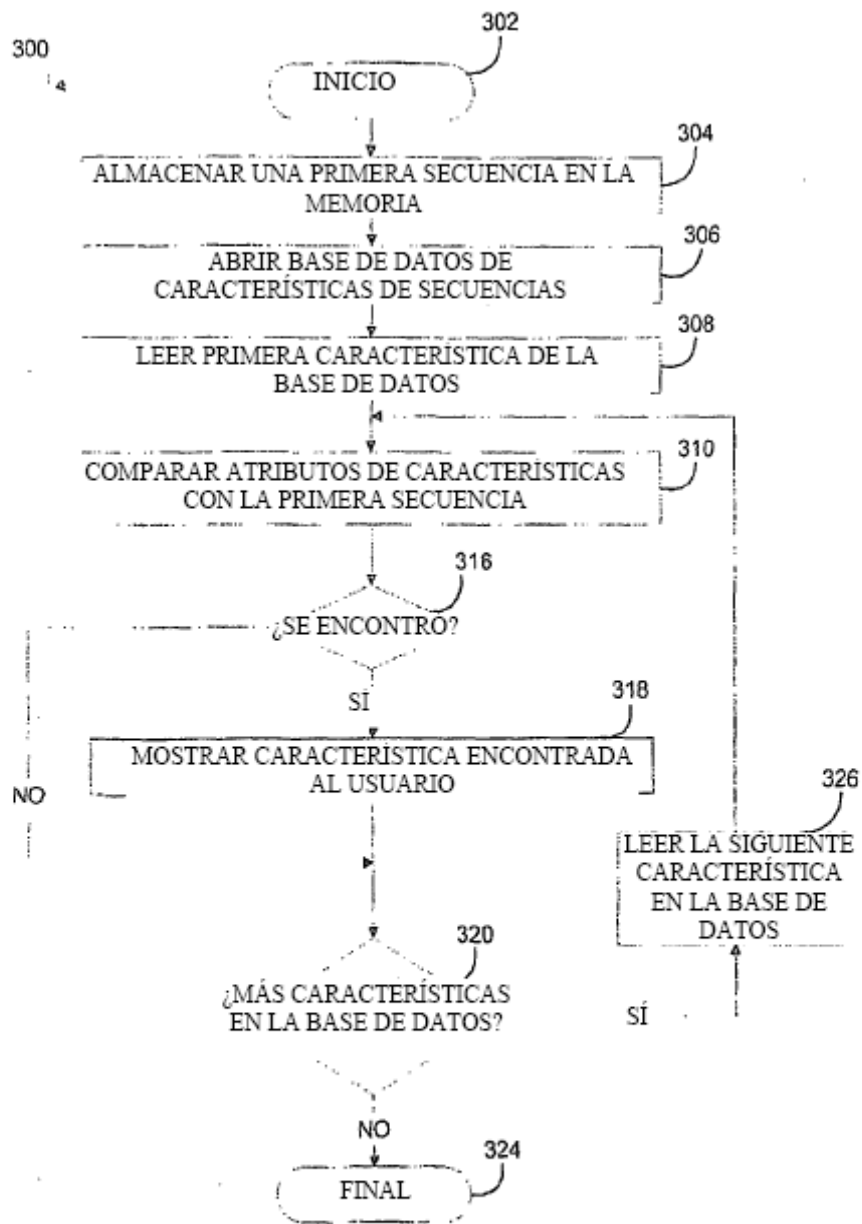


FIGURA 12

FIGURA  
13

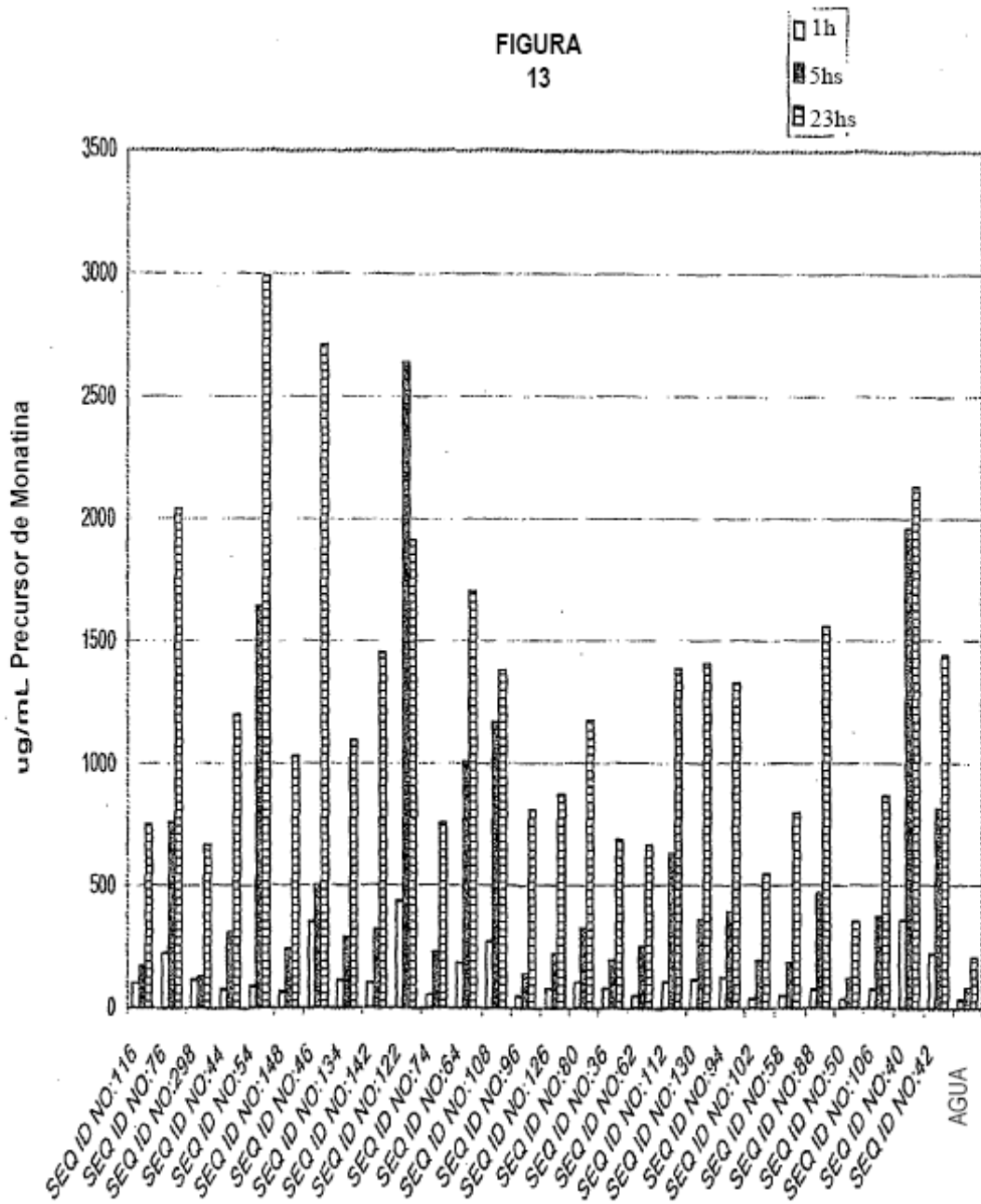
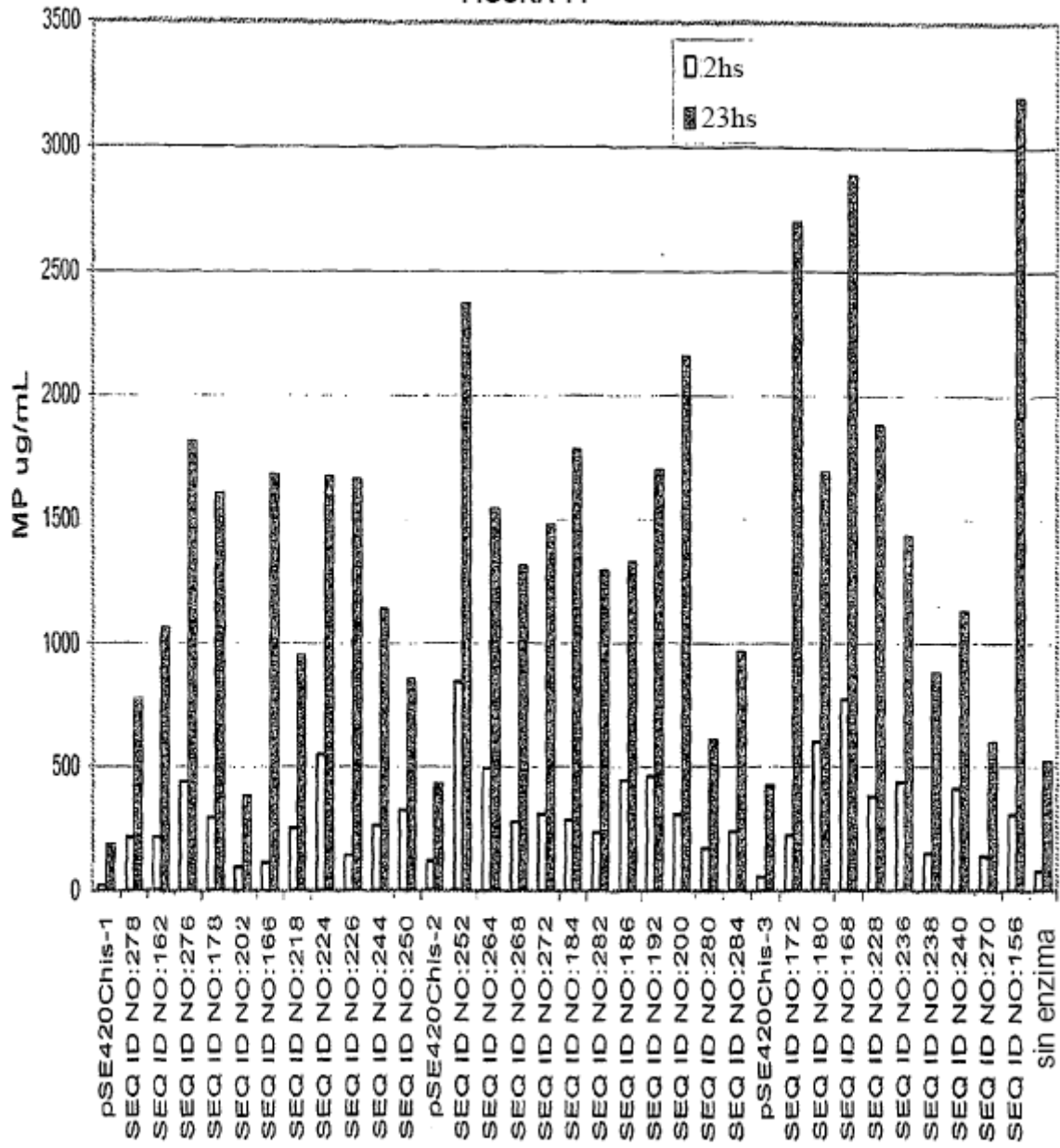


FIGURA 14





Efecto de DTT en la SEQ ID NO:88

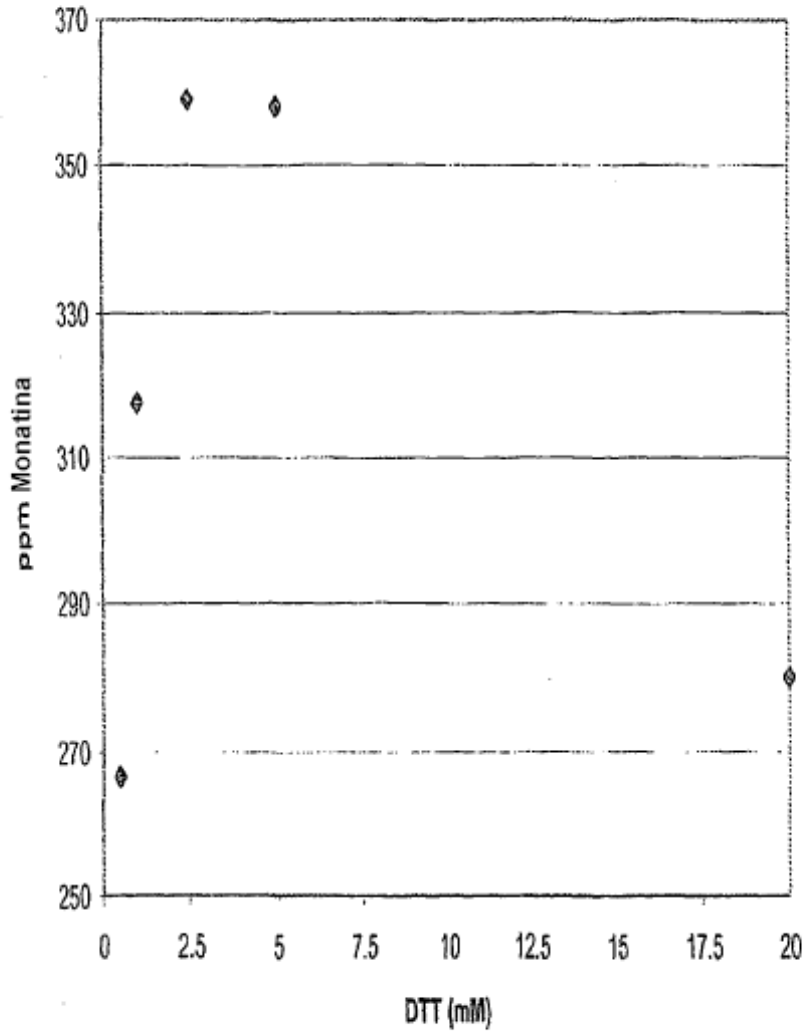


FIGURA 15