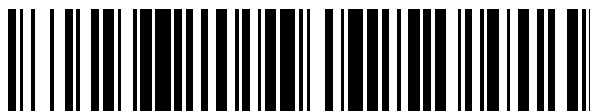


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 599 644**

51 Int. Cl.:

C12N 9/10 (2006.01)

C12N 9/88 (2006.01)

G06F 19/00 (2011.01)

C12P 13/00 (2006.01)

G06F 19/22 (2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **13.08.2010 E 14190794 (9)**

97 Fecha y número de publicación de la concesión europea: **19.10.2016 EP 2857505**

54 Título: **Un proceso para la identificación y preparación de una omega-transaminasa (R)-específica**

30 Prioridad:

02.09.2009 EP 09011271
19.05.2010 EP 10005203

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

02.02.2017

73 Titular/es:

LONZA AG (100.0%)
Lonzastrasse 38
3930 Visp, CH

72 Inventor/es:

HOEHNE, MATTHIAS;
BORNSCHEUER, UWE;
ROBINS, KAREN y
SCHÄTZLE, SEBASTIAN

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 599 644 T3

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

DESCRIPCIÓN

Un proceso para la identificación y preparación de una omega-transaminasa (R)-específica

La presente invención se refiere a procesos para la preparación de una amina quiral ópticamente activa.

Las aminas quirales desempeñan un papel importante en las industrias farmacéutica, agroquímica y química. Se usan frecuentemente como intermedios o sintonas para la preparación de diversas sustancias activas fisiológicamente, por ejemplo farmacéuticamente activas, tales como derivados de cefalosporina o pirrolidina. En un gran número de las diversas aplicaciones de las aminas quirales, solo una forma particular ópticamente activa, ya sea el enantiómero (R) o el enantiómero (S), es fisiológicamente activa. Por consiguiente, existe una necesidad de proporcionar procesos para la preparación de aminas quirales en una forma ópticamente activa.

Dicha necesidad queda parcialmente cubierta por la preparación de aminas quirales mediante cristalización de sales diastereoméricas por adición de ácidos carboxílicos quirales (Breuer et al., *Angewandte Chemie* (2004) 116, 806-843). Otros métodos químicos usan síntesis enantioselectiva para reducir precursores proquirales con dobles enlaces C=N.

Entre los diferentes métodos enzimáticos que se han empleado para la síntesis de aminoácidos y aminas ópticamente activos, las ω -transaminasas (ω -TAs) han recibido recientemente una atención creciente, debido a su potencial para la síntesis asimétrica de aminas ópticamente activas, que se emplean frecuentemente como unidades de construcción en la preparación de numerosos productos farmacéuticos.

Las ω -transaminasas son enzimas dependientes de PLP (piridoxal fosfato) que catalizan reacciones de transferencia de grupos amino. Cuando se emplean ω -transaminasas, se pueden producir principalmente aminas denominadas enantioenriquecidas y/o ópticamente activas puras a través de dos estrategias de reacción, (i) la síntesis asimétrica que parte de cetonas, y (ii) la resolución cinética que parte de aminas racémicas. Aunque las ω -transaminasas exhiben una buena enantioselectividad en general, no se han usado con frecuencia en la síntesis asimétrica, aunque en este caso teóricamente es posible obtener un rendimiento del 100%. Un requisito específico en una síntesis asimétrica que emplee ω -transaminasas es el desplazamiento del equilibrio hacia el lado del producto, especialmente cuando se usa un aminoácido como la alanina como donante de amino, ya que en este caso el equilibrio está del lado de los sustratos (cetona, alanina) y no del lado de los productos (amina, piruvato); otro requisito es que la estereoselectividad de la enzima debe ser perfecta, que no es siempre el caso de las ω -transaminasas. Por lo tanto, las ω -transaminasas se usan principalmente para la resolución cinética de aminas racémicas, donde la enantioselectividad no tiene por qué ser necesariamente perfecta. Por tanto, aunque se han investigado resoluciones cinéticas de aminas racémicas, la limitación a un rendimiento máximo del 50 % dificulta considerablemente su aplicación. Por otro lado, la síntesis asimétrica requiere métodos para desplazar el equilibrio desfavorable hacia la síntesis de enantiómeros de aminas ópticamente puras para las que se han desarrollado varios métodos, lo cual es un prerrequisito clave de procesos eficientes que permitan el uso de transaminasas a escala industrial. Dichos métodos se describen en el documento WO 2007/093372.

El documento WP 0 987 332 A1 describe un proceso para producir compuestos amino ópticamente activos, a saber compuestos (R)-amino, mediante una enzima microbiana, en particular una ω -transaminasa (R)-selectiva derivada de *Arthrobacter* sp. El documento EP 1 038 953 A1 describe una ω -transaminasa (R)-selectiva adicional, que, no obstante, deriva de *Mycobacterium aurum*.

Iwasaki et al. (*Biotechnol. Let.* 2003 (25), 1843-1846), Koszelewski et al. (*Adv. Synth. Catal.* 2008 (350), 2761-2766) e Iwasaki et al. (*Appl. Microbiol. Biotechnol.* 2006 (69), 499-505) describen una transaminasa R-específica procedente de *Arthrobacter* sp. La identificación de microorganismos que proporcionen transaminasas (S)- ó (R)-selectivas se realiza habitualmente seleccionando el microorganismo obtenido, por ejemplo, de muestras de suelo y enriqueciéndolas en un cultivo (Jun et al., *App. Microbiol. Biotechnol.* 2004 (70), 2529-2534, y Shin et al. (*Biosc. Biotechnol. Biochem.* 2001 (65), 1782-1788)). Especialmente, todas las ω -transaminasas R-selectivas descritas fueron obtenidas mediante cultivo de enriquecimiento. Dichos métodos requieren mucho tiempo, ya que para lograr un proceso eficiente a menudo se prefiere sobre-expresar la enzima en un organismo heterólogo, tal como *Escherichia coli*. Esto requiere el aislamiento de la secuencia génica de los organismos naturales o salvajes ("wild type" en inglés), pero la clonación de la secuencia de DNA no siempre se logra con éxito y es un proceso que requiere mucho tiempo.

En particular, durante el desarrollo del proceso o para la identificación de una nueva ω -TA, la purificación de la enzima y la caracterización de sus propiedades enzimáticas son de gran interés. Sin embargo, dado que la actividad de ω -TA se determina habitualmente con métodos de bajo rendimiento como HPLC o electroforesis capilar (EC), la determinación de las propiedades de la enzima bastante a menudo supone una etapa limitante.

En general, en comparación con las ω -transaminasas (S)-selectivas el número de ω -transaminasas (R)-selectivas disponibles es mucho más limitado. Esto se contrapone a la elevada demanda que existe de ω -transaminasas (R)-selectivas, que son altamente deseables para síntesis asimétrica de (R)-enantiómeros de diversas aminas quirales. Por tanto, sigue existiendo la necesidad de proporcionar transaminasas (R)-selectivas adicionales y los medios y

métodos para obtenerlas a fin de producir R-enantiómeros, en particular aminas ópticamente activas, por ejemplo para aplicaciones farmacéuticas o agroquímicas, que hasta la fecha o bien no se encuentran disponibles en absoluto o bien no pueden obtenerse mediante un proceso económicamente viable, preferiblemente a escala industrial.

5 Por lo tanto, la presente invención se basa en el problema técnico de proporcionar medios y métodos simples, rápidos y fiables para identificar, caracterizar y obtener ω -transaminasas (R)-selectivas para la producción de (R)-enantiómeros deseados, preferiblemente en forma ópticamente pura, que sean adecuados preferiblemente para la identificación, caracterización y preparación a escala industrial.

La presente invención resuelve este problema técnico mediante la provisión de las enseñanzas de las reivindicaciones independientes.

10 Por lo tanto, el presente documento proporciona un proceso para la preparación de una amina quiral ópticamente activa, que comprende hacer reaccionar al menos un aceptor amino y al menos un donante amino con una ω -transaminasa (R)-selectiva (también denominada ω -TA) preparada mediante un método que comprende las siguientes etapas:

15 a) proporcionar al menos una secuencia de biomolécula de consulta de al menos una ω -transaminasa (R)-selectiva, preferiblemente una transaminasa que pertenezca al tipo de plegamiento IV de enzimas dependientes de PLP, y al menos un banco de biomoléculas,

20 b) buscar en el banco de biomoléculas con la secuencia de biomolécula de consulta para identificar un grupo de primeras secuencias de biomolécula diana, en donde las primeras secuencias de biomolécula diana tienen un grado de identidad de secuencia de al menos un 20%, preferiblemente un 32%, respecto a la secuencia de biomolécula de consulta, calculado a nivel de aminoácidos,

c) seleccionar en el grupo de primeras secuencias de biomolécula diana un grupo de segundas secuencias de biomolécula diana, que no comprendan, a nivel de aminoácidos, al menos una de las estructuras de secuencia de aminoácidos c1) a c3), siendo

25 c1) en la posición 95 a 97 una secuencia de aminoácidos Tyr Xa1 Xa2, siendo Xa1 un aminoácido Ile, Val, Leu, Met, Phe, y siendo Xa2 un aminoácido Arg o Lys o

c2) en la posición 97 a 99 una secuencia de aminoácidos Tyr Xaa Gln, siendo Xaa un aminoácido, preferiblemente un aminoácido convencional y en la región de las posiciones 105 a 111, preferiblemente en la posición 107 a 109, una secuencia de aminoácidos Arg Xaa Xa3, siendo Xa3 un aminoácido, que preferiblemente es His o

30 c3) en la posición 38 Thr, en la posición 97 Lys y en la posición 107 a 109 una secuencia de aminoácidos Arg Xa4 Xa5, siendo Xa4 un aminoácido, preferiblemente Gly, y siendo Xa5 un aminoácido, que preferiblemente es Tyr,

y que comprende

35 c4) en la posición 95 un aminoácido distinto a Tyr, Arg, Lys, o en la posición 95 Tyr, pero en la posición 97 no Arg o Lys y

c5) en la posición 40 no Lys o Arg y

c6) en la región de las posiciones 161 a 165, preferiblemente en la posición 163, al menos una Lys

para identificar un grupo de segundas secuencias de biomolécula diana y

40 d) proporcionar, preferiblemente preparar, una biomolécula que tenga la segunda secuencia de biomolécula diana identificada en la etapa c) y que sea, o que codifique al menos parcialmente, una proteína con la actividad de una ω -transaminasa (R)-selectiva.

45 La terminología "una biomolécula que sea o que codifique al menos parcialmente una proteína" significa preferiblemente que la biomolécula puede ser una proteína con la actividad de una ω -transaminasa (R)-selectiva o, en caso de que la biomolécula sea una secuencia de nucleótidos, al menos parte de dicha secuencia de nucleótidos codifique dicha proteína, preferiblemente una proteína de longitud completa. Por consiguiente, en caso de que la biomolécula proporcionada sea una molécula de secuencia de nucleótidos, al menos una parte de dicha molécula de secuencia de nucleótidos codifica una proteína con la actividad de una ω -transaminasa (R)-selectiva, por lo que posiblemente una parte adicional de dicha molécula de nucleótidos presenta alguna otra función, por ejemplo una función reguladora o de replicación. Por lo tanto, la terminología "que proporciona una biomolécula que tiene la segunda secuencia de biomolécula diana identificada en la etapa c) y que es o codifica al menos parcialmente una proteína con la actividad de una ω -transaminasa (R)-selectiva" preferiblemente es equivalente a la terminología "que proporciona una biomolécula que tiene la segunda secuencia de biomolécula diana identificada en la etapa c) que

puede ser una proteína con la actividad de una ω -transaminasa (R)-selectiva o que puede ser una molécula de secuencia de nucleótidos que incluye una secuencia que codifica dicha proteína siempre que la proteína presente la actividad de una ω -transaminasa (R)-selectiva”.

5 En una realización de la presente invención, en la etapa c2) la posición de la estructura Arg Xaa Xa3 puede variar en 1-2 aminoácidos respecto a las posiciones 107 a 109, es decir, puede estar en las posiciones 105 a 111.

En una realización de la presente invención, en la etapa c6) la posición de Lys puede variar en 1-2 aminoácidos respecto a la posición 163, es decir, puede estar en las posiciones 161 a 165.

10 En el contexto de la presente invención cada secuencia de aminoácidos dada, en particular la estructura de secuencia de aminoácidos, o la secuencia de DNA o la estructura de secuencia de DNA, se proporciona en la dirección desde el extremo amino al extremo carboxi o dirección 5' a 3', lo que sea aplicable, y a menos que se especifique lo contrario. Preferiblemente, las secuencias dadas en la dirección dada se proporcionan como una tira continua sin ningún nucleótido o aminoácido interviniente, lo que sea aplicable.

15 El proceso para la preparación de una ω -transaminasa (R)-selectiva se puede llevar a cabo de un modo rápido, eficiente, simple, fiable y fácil. Por lo tanto, la presente invención permite identificar y preparar una ω -transaminasa (R)-selectiva, que no ha sido conocida o accesible previamente, lo que abre numerosas rutas para proporcionar aminas quirales, preferiblemente (R)-enantiómeros, en particular en rutas de síntesis asimétrica muy eficientes. Se podría demostrar con éxito que incluso proteínas con un bajo grado de identidad de secuencia, por ejemplo de tan solo un 35%, respecto a una ω -transaminasa (R)-selectiva conocida, en contra de lo esperado, son de hecho ω -transaminasas (R)-selectivas. De un modo particularmente ventajoso e inesperado, la presente invención permite proporcionar preferiblemente ω -transaminasas (R)-selectivas procedentes de *Aspergillus terreus* y *Mesorhizobium loti*, siendo la ω -transaminasa de *Aspergillus* la primera ω -transaminasa eucariótica que convierte sustratos con (R)-selectividad.

25 Básicamente, la presente invención se basa en las enseñanzas técnicas de que determinadas proteínas o sus secuencias de nucleótidos codificadoras, incluso con un grado bajo de identidad de secuencia con respecto a ω -transaminasas (R)-selectivas conocidas, pueden usarse como fuente potencial para identificar y preparar ω -transaminasas (R)-selectivas, en donde las ω -transaminasas (R)-selectivas putativas son sometidas a escrutinio y se preparan discriminando las no deseadas y de este modo seleccionando de forma positiva proteínas que presenten estructuras de secuencia particulares en su secuencia de aminoácidos, que se atribuyen a la actividad enzimática deseada. Por lo tanto, la presente invención usa específicamente características estructurales particulares, en particular la ausencia y la presencia de aminoácidos particulares en ω -TA (R)-selectivas putativas, para identificar y preparar ω -transaminasas (R)-selectivas.

30 Por consiguiente, en una primera etapa a) del presente proceso, una secuencia de biomolécula de consulta de al menos una ω -transaminasa (R)-selectiva, por ejemplo una ω -transaminasa (R)-selectiva conocida, o al menos una parte de secuencia característica de dicha ω -transaminasa (R)-selectiva, que preferiblemente es capaz en concreto de identificar una enzima dependiente de PLP de plegamiento de tipo IV, se proporciona junto con al menos un banco de biomoléculas. Preferiblemente, la secuencia de biomolécula de consulta es un ORF (marco de lectura abierto, del inglés “open reading frame”) de una ω -TA (R)-selectiva, una secuencia de nucleótidos codificadora, o una parte característica de los mismos. En otra realización, la secuencia de biomolécula de consulta es ella misma una secuencia de aminoácido de ORF o una parte característica de la misma.

35 En el contexto de la presente invención, un parte características de la secuencia de biomolécula de consulta de al menos una ω -transaminasa o liasa (R)-selectiva, en particular de la ω -transaminasa (R)-selectiva conocida, es una secuencia de biomolécula que en su forma de molécula de secuencia de DNA se hibrida en las siguientes condiciones con la secuencia de biomolécula de consulta de longitud completa, en particular a la secuencia de DNA del ORF.

40 Los métodos para la hibridación de ácidos nucleicos tales como el DNA son bien conocidos y se describen, por ejemplo, en “Molecular Cloning”, Tercera edición (2001); “Methods for General and Molecular Bacteriology”, ASM Press (1994); “Immunology methods manual”, Academic Press (Molecular), y muchos otros libros de texto estándar.

45 A continuación se incluye un ejemplo de hibridación en condiciones severas. Se incuban un filtro con un ácido nucleico inmovilizado sobre dicho filtro y el ácido nucleico usado como sonda en una disolución que comprende un 50% de formamida, 5 \times SSC (cloruro sódico 750 mM y citrato sódico 75 mM), fosfato sódico 50 mM (pH 7,6), 5 \times disolución de Denhardt, 10% de sulfato de dextrano y 20 μ g/L de DNA de esperma de salmón desnaturalizado a 42 °C durante una noche. Después de la incubación, el filtro es lavado en 0,2 \times disolución SSC (aprox. 65°C). Estas condiciones de hibridación severas pueden modificarse ajustando la concentración de formamida (las condiciones se vuelven menos severas al reducir la concentración de formamida) y cambiando las concentraciones de sales y las condiciones de temperatura.

55 Una hibridación en condiciones menos severas se lleva a cabo, por ejemplo, incubando un filtro con un ácido nucleico inmovilizado sobre el mismo y un ácido nucleico usado como sonda en una disolución que comprende

6×SSCE: cloruro sódico 3 mol/L, dihidrógeno fosfato sódico 0,2 mol/L y EDTA 0,02 mol/L, pH 7,4), 0,5% de SDS, 30% de formamida y 100 µg/L de DNA de esperma de salmón desnaturalizado a 37 °C durante una noche, y lavando el filtro con 1×disolución SSC que contiene un 0,1% de SDS (50 °C).

5 En una realización particularmente preferida, la presente invención entiende con la expresión “al menos una secuencia de biomolécula de consulta de al menos una ω -transaminasa (R)-selectiva” una secuencia de biomolécula que es adecuada para seleccionar secuencias de biomolécula según la presencia o ausencia de las estructuras de secuencia c1) a c6), tal y como se han identificado en la presente memoria. Por consiguiente, dicha secuencia de biomolécula de consulta es una secuencia que sirve para hacer un escrutinio y para identificar aquellas secuencias de biomolécula que no comprendan a nivel de aminoácidos al menos una de las estructuras de secuencia de aminoácidos c1) a c3), y que comprendan a nivel de aminoácidos las estructuras de secuencia c4), c5) y c6). Dicha secuencia de biomolécula de consulta puede venir presentada como información de secuencia de aminoácidos o como una molécula de secuencia de DNA o como información de secuencia de DNA.

15 En una realización preferida, la parte característica de la secuencia de biomolécula de consulta usada en la etapa a) abarca, preferiblemente consiste en, la región que va desde la posición 30 a la 170, lo más preferiblemente de la posición 30 a la 120 del ORF de una ω -transaminasa (R)-selectiva.

El banco de biomoléculas se utiliza para la búsqueda en una etapa b) posterior con la secuencia de biomolécula de consulta para identificar un grupo de primeras biomoléculas diana, que muestren al menos un grado mínimo de identidad de secuencia de al menos el 20%, preferiblemente el 25%, preferiblemente el 32%, preferiblemente al menos el 33%, lo más preferiblemente el 34%, al menos el 35%, al menos el 36%, al menos el 40%, al menos el 50%, al menos el 60%, al menos el 70% y al menos el 80%, al menos el 90% o al menos el 95%, con respecto a la secuencia de biomolécula de consulta, en base al nivel de aminoácidos, y en donde dicho grupo de primeras biomoléculas diana representa una primera selección del banco de biomoléculas usado en la etapa a). Dicho grado de identidad de secuencia es preferiblemente una identidad de secuencia entre al menos la parte característica del ORF, preferiblemente esencialmente el ORF completo, en particular el ORF completo de la secuencia de biomolécula de consulta y al menos la parte característica del ORF, preferiblemente esencialmente el ORF completo, en particular el ORF completo de las secuencias de biomolécula escrutadas. Posteriormente a dicha etapa b), de este grupo de primeras secuencias de biomolécula diana se seleccionan en la etapa c) aquellas que no comprenden como estructura de secuencia c1) en la posición 95 a 97 una secuencia de aminoácidos Tyr Xa1 Xa2, siendo Xa1 un aminoácido Ile, Val, Leu, Met, Phe, y siendo Xa2 un aminoácido Arg o Lys, o que no comprendan como estructura de secuencia c2) en la posición 97 a 99 una secuencia de aminoácidos Tyr Xaa Gln, siendo Xaa un aminoácido, preferiblemente un aminoácido habitual, y en la región entre las posiciones 105 y 111, preferiblemente en las posiciones 107 a 109, una secuencia de aminoácidos Arg Xaa Xa3, siendo Xa3 un aminoácido, preferiblemente un aminoácido habitual, siendo preferiblemente His, o que no comprenda la estructura de secuencia c3) en la posición 38 Thr, en la posición 97 Lys y en la posición 107 a 109 una secuencia de aminoácidos Arg Xa4 Xa5, siendo Xa4 un aminoácido, preferiblemente un aminoácido habitual, que preferiblemente es Gly, y siendo Xa5 un aminoácido, preferiblemente un aminoácido habitual, que preferiblemente es Tyr, mediante lo cual se discrimina y se descargan dichas secuencias, que no muestran al menos una de las estructuras de secuencia identificadas anteriormente c1), c2) ó c3). En una selección posterior o simultánea, se seleccionan las secuencias de biomolécula que de acuerdo a la estructura de secuencia c4) muestran en la posición 95 un aminoácido distinto de Tyr, Arg, Lys, o en la posición 95 Tyr, pero en la posición 97 no Arg y no Lys, y que de acuerdo a la estructura de secuencia c5) no presentan en la posición 40 Lys ni Arg, y que de acuerdo a la estructura de secuencia c6) tienen en la región entre las posiciones 161 y 165 al menos una Lys, preferiblemente una Lys, preferiblemente en la posición 163 Lys, de tal modo que se selecciona y se identifica un grupo de segundas secuencias de biomolécula diana. El grupo de segundas secuencias de biomolécula diana identificadas y sometidas a escrutinio para las estructuras de secuencia identificadas anteriormente representa secuencias de biomolécula que son o que codifican una proteína con la actividad de una ω -transaminasa (R)-selectiva, que es proporcionada de este modo.

En el contexto de la presente invención, una transaminasa es una enzima dependiente de piridox-alfosfato que cataliza la transferencia de grupos amino, estando clasificada preferiblemente en el tipo IV de plegamiento. Las transaminasas se clasifican como E.C. 2.6.1.X. De acuerdo a la presente invención, la transaminasa es una ω -transaminasa (R)-selectiva. En el contexto de la presente invención una proteína con la actividad de una ω -transaminasa (R)-selectiva es una proteína que es capaz, en las condiciones de reacción apropiadas, de catalizar una transferencia de grupos nitrogenados tales como grupos amino desde un donante a un aceptor tal como es capaz de hacerlo una ω (omega)-transaminasa (R)-selectiva (beta-alanina-piruvato transaminasa). En el contexto de la presente invención una ω -transaminasa (R)-selectiva preferiblemente es una enzima con el código de clasificación E.C.2.6.1.18.

En el contexto de la presente invención, el término “amina quiral ópticamente activa” se refiere a lo mismo que el término “amina quiral enantioméricamente activa”. Estos términos en particular se refieren a una preparación que está esencialmente libre del enantiómero no deseado. Por consiguiente, una amina quiral ópticamente activa comprende esencialmente un exceso de un enantiómero o incluso consiste únicamente en un enantiómero.

En particular, en el contexto de la presente invención, una amina quiral ópticamente activa tiene una pureza óptica de al menos 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99,5, 99,6, 99,7, 99,8 y en particular de al menos 99,9%.

En la presente invención la pureza óptica se expresa en % de exceso de un enantiómero respecto al otro enantiómero. Por tanto, la pureza óptica en % es el cociente de la diferencia entre las concentraciones del enantiómero (R) y del enantiómero (S) y la suma de las concentraciones de ambos enantiómeros (pureza óptica de A en % = $([A] - [S]) : ([A] + [S]) \times 100$, en donde A y S representan las concentraciones de los enantiómeros (R) y (S) o viceversa).

En el contexto de la presente invención, un banco de biomoléculas es una fuente de biomoléculas por sí mismo o es una colección de datos de secuencias de biomoléculas, en particular datos de secuencia de polinucleótidos o polipéptidos.

En el contexto de la presente invención, una biomolécula preferiblemente es una molécula de polinucleótido que porta información genética, en particular una molécula de DNA. En una realización preferida adicional de la presente invención una biomolécula es un polipéptido, en particular una proteína, que comprende una serie de aminoácidos. Por tanto, un banco de biomoléculas puede ser una fuente física de biomoléculas, en particular puede ser un banco génico, en particular una biblioteca genómica o de DNAC, o es una colección de información sobre dichas biomoléculas, en particular una colección de datos de secuencia, en particular secuencias de aminoácidos o secuencias de polinucleótidos, en particular secuencias de DNA.

La presente invención también se refiere a secuencias de biomoléculas, secuencias de aminoácidos, secuencias de polinucleótidos o secuencias de nucleótidos, en particular secuencias de DNA, en donde dicha nomenclatura designa por un lado la sustancia física *per se*, que significa una proteína, un polipéptido o una molécula de DNA, y por otro lado el contenido de información de dicha sustancia química, que, en el caso de un polinucleótido, es el tipo, orden y número de sus nucleótidos y, en el caso de un polipéptido es el tipo, orden y número de los aminoácidos individuales que conforman el polipéptido.

En el contexto de la presente invención los términos “biomolécula”, “molécula de DNA”, “molécula de polinucleótido”, “polipéptido” o “proteína” se refiere a sustancias químicas *per se*. En el caso de que la presente invención se refiera específicamente a una información de secuencia de aminoácidos, a una información de secuencia de DNA, a una información de secuencia de polinucleótidos o una información de secuencia de biomoléculas, no se refiere a la forma física de una biomolécula, sino más bien a la información contenida en ella, es decir, el tipo, orden y número de sus constituyentes, a saber, aminoácidos y nucleótidos.

La presente invención, en una realización A) preferida, en sus etapas a) a c) es aplicable a un banco de biomoléculas que es una colección de datos de secuencia de, en una realización, secuencias de polinucleótidos, en particular secuencias de DNA o, en otra realización, secuencias de aminoácidos, las cuales se usan ambas para realizar una búsqueda en la etapa b) con herramientas de alineamiento de secuencia, preferiblemente tales como BLAST, que es una herramienta de búsqueda de alineamiento local básica, para identificar un grupo de primeras secuencias de biomolécula diana (Altschul, S.F. et al. 1990. J. Mol. Biol. 215: 403; Altschul, S.F. et al. 1997. Nucleic Acid Res. 25:3389-3402). Otros programas adecuados incluyen GAP, BESTFIT y FASTA en el Paquete de Software Wisconsin Genetics (Genetics Computer Group (GCG), Madison, WI, EE.UU.). En una etapa posterior c) el grupo de primeras secuencias de biomoléculas identificadas es sometido a etapas de selección adicionales durante las cuales se usan las estructuras de secuencia identificadas para identificar y seleccionar, negativamente con las estructuras de secuencia c1) a c3) y positivamente con las estructuras de secuencia c4) a c6), la secuencia de biomolécula deseada. Una vez identificado el segundo grupo de secuencias de biomoléculas, la presente invención utiliza dicha información de secuencias para preparar las correspondientes moléculas de oligo- o poli-nucleótidos, por ejemplo cebadores de hibridación, para realizar un escrutinio y seleccionar, en la forma física de bancos de biomoléculas, moléculas de DNA que codifican las enzimas identificadas en el segundo grupo. Por tanto, en una realización preferida, se puede usar la información de secuencia obtenida en la etapa c) para clonar un gen correspondiente a partir de una biblioteca de DNA genómico o DNAC y para la expresión de la proteína, por ejemplo, en *E. coli*, *Saccharomyces cerevisiae*, *Pichia pastoris* u otras. En otra realización preferida de la presente invención, la información de secuencia obtenida en la etapa c) se usa para sintetizar desde cero la ω -transaminasa (R)-selectiva. En dicha realización preferida es posible usar la información de secuencia obtenida en la etapa c) para la síntesis de un gen, por ejemplo, mediante la utilización de un codón optimizado y estabilidad de ARNm, y clonando y expresando dicho gen. De acuerdo a dicha realización preferida, es posible usar como el al menos un banco de biomoléculas en la etapa a) un banco de biomoléculas que contenga información de secuencia de DNA o información de secuencia de aminoácidos. En el caso de que el banco de biomoléculas contenga información de secuencia de DNA, dicha información tiene que ser traducida a información de secuencia de aminoácidos a fin de ser procesada según las etapas b) a c), en las que se usan secuencias de aminoácidos como secuencia de biomolécula de consulta, y como estructuras de secuencia de aminoácidos en las etapas b) a c), que se llevan a cabo en el banco de secuencias de DNA usando las secuencias de biomolécula de consulta y las estructuras de secuencia c1) a c6) traducidas de nuevo a información de secuencia de DNA.

En otra realización preferida de la presente invención, las presentes enseñanzas se aplican en una realización B) en la etapa a) a un banco de biomoléculas que está presente en la forma física de un banco genético, en particular una biblioteca genética tal como una biblioteca de DNAC-, metagenoma o genómica, usando moléculas de DNA que codifican toda una ω -transaminasa (R)-selectiva, o al menos la parte característica, para buscar, identificar y seleccionar un grupo de primeras secuencias de biomolécula, en particular moléculas de secuencia de DNA que, calculadas a nivel de aminoácidos, presenten un grado de identidad de secuencia de al menos el 20%, preferiblemente el 32%, con respecto a la molécula de secuencia de DNA de consulta. La posterior etapa c) se lleva a cabo preferiblemente en dicho grupo de primeras moléculas de DNA usando cebadores de molécula de nucleótido que identifican y seleccionan, positiva o negativamente, las estructuras de secuencia deseadas c1) a c6).

En el contexto de la presente invención, el término "estructura de secuencia" se refiere a las características selectivas específicas de la secuencia de aminoácidos de la ω -transaminasa (R)-selectiva putativa, identificadas específicamente en la etapa c) del proceso reivindicado. En particular, la primera estructura de secuencia c1) es la estructura de secuencia identificada en la posición 95 a 97, que en dicho orden es una secuencia de aminoácidos Tyr Xa1 Xa2, siendo Xa1 un aminoácido Ile, Val, Leu, Met, Phe, y siendo Xa2 un aminoácido Arg o Lys, y que, según la presente invención, no deberían estar presentes en la secuencia de biomolécula seleccionada finalmente. La segunda estructura de secuencia c2) es la estructura de secuencia identificada en la posición 97 a 99, que en dicho orden es una secuencia de aminoácidos Tyr Xaa Gln, siendo Xaa un aminoácido, y en la región de la posición 105 a 111, preferiblemente en la posición 107 a 109, una secuencia de aminoácidos, que en dicho orden es Arg Xaa Xa3, siendo Xaa un aminoácido y siendo Xa3 un aminoácido, que preferiblemente es His, y que tampoco debería estar presente en la secuencia de biomolécula seleccionada finalmente. La tercera estructura de secuencia c3) es la estructura de secuencia que se caracteriza por la presencia de Thr en la posición 38, en la posición 97 Lys y en la posición 107 a 109 una secuencia de aminoácidos Arg Xa4 Xa5, siendo Xa4 un aminoácido, que preferiblemente es Gly, y siendo Xa5 un aminoácido, que preferiblemente es Tyr, estructura c3) que no debería estar presente en la secuencia de biomolécula seleccionada finalmente.

La cuarta estructura de secuencia c4) requiere que en la posición 95 haya un aminoácido que no sea Tyr, Arg, Lys, o en la posición 95 que sea Tyr, pero en la posición 97 que no haya Arg o Lys, lo que la convierte en una ω -transaminasa (R)-selectiva putativa preparada según la presente invención. La quinta estructura de secuencia c5) requiere que en la posición 40 no haya Lys ni Arg en la ω -transaminasa (R)-selectiva putativa identificada. La sexta estructura de secuencia c6) requiere que en la región de la posición 161 a la 165 esté presente al menos una Lys, preferiblemente una Lys, preferiblemente una Lys en la posición 163, en la ω -TA (R)-selectiva finalmente identificada y preparada.

En el contexto de la presente invención, la identificación y la localización de posiciones de aminoácido se determina como se indica a continuación. Las primeras secuencias de biomolécula diana son alineadas, preferiblemente se alinean múltiplemente unas con otras y con la secuencia de biomolécula de consulta. El alineamiento se puede realizar con un software de alineamiento convencional, tal como STRAP, en particular ClustalW, preferiblemente ClustalW3D. En otra realización también es posible alinear las secuencias por pares, es decir cada una de las primeras secuencias de biomolécula diana con la secuencia de biomolécula de consulta. En una realización preferida, las primeras secuencias de biomolécula diana se alinean con la BCAT de *E. coli*. En otra realización preferida, las primeras secuencias de biomoléculas diana se alinean con otra secuencia de biomolécula de consulta usada en la presente invención, por ejemplo la ω -transaminasa (R)-selectiva.

La anotación de las posiciones de aminoácidos tal cual se proporcionan en la presente invención está determinada por la posición de la correspondiente estructura de secuencia en la secuencia de biomolécula de consulta usada como patrón de posiciones. Un alineamiento como el descrito anteriormente alinea las correspondientes posiciones de aminoácidos de las primeras secuencias de biomoléculas diana con las estructuras de secuencia presentes en la secuencia de biomolécula de consulta.

Como ejemplo, el uso de BCAT de *E. coli* como patrón para la anotación de la posición, se usa la secuencia de aminoácidos BCAT de *E. coli* conocida de las posiciones 92 a 100, a saber TSAYIRPLI (SEQ ID NO: 9) para identificar las posiciones 95 a 99 en la ω -transaminasa putativa, que se analiza para determinar la ausencia de las estructuras de secuencia c1), c2), c3) y la presencia de c4). La secuencia de aminoácidos conocida de la posición 35 a la 42, a saber VFEGIRCY (SEQ ID NO: 10) de la BCAT de *E. coli* marca la posición de la G en la posición G para la estructura de secuencia c3) y c5). La secuencia de aminoácidos DVGMGVNP de la secuencia de aminoácidos BCAT de *E. coli* (SEQ ID NO: 11) marca la posición 104 a 111 para la estructura de secuencia c3) en las posiciones 105 a 111. La secuencia de aminoácidos PTAAGAGN de las posiciones 159 a 167 de la BCAT de *E. coli* conocida (SEQ ID NO: 12) marca la posición 163 como una K para la estructura de secuencia c6).

Por tanto, en una realización preferida de la presente invención la biomolécula es una proteína y la secuencia de biomolécula es una secuencia de aminoácidos. Por consiguiente, en una realización preferida el banco de biomoléculas es un banco, en particular una base de datos, que reúne información de diversas proteínas, en particular información de secuencias de aminoácidos. Los bancos de bases de datos preferidos son el banco de datos de proteína NCBI, el banco de datos UniProtKB/SwissProt y el banco de datos UniProtKB/TrEMBL.

En una realización preferida también es posible que la biomolécula sea una molécula de DNA y la secuencia de biomolécula sea una secuencia de DNA. Por consiguiente, en una realización preferida, el banco de biomoléculas es un banco, en particular una base de datos, que reúne información de diversas secuencias de polinucleótidos, en particular secuencias de DNA.

5 En una realización preferida, la presente invención usa un proceso acorde a lo anterior, en donde el banco de biomoléculas es una base de datos de biomoléculas y la base de datos de biomoléculas se usa para realizar la búsqueda de la etapa b) con un programa de alineamiento de secuencia de biomolécula, en particular BLAST. En una realización preferida se prevé que si el banco de biomoléculas es una base de datos de biomoléculas, la base de datos de biomoléculas se usa para realizar la búsqueda de la etapa b) con una secuencia de aminoácidos, si la base de datos de biomoléculas es una base de datos de secuencias de aminoácidos, o con una secuencia de DNA, si la base de datos de biomoléculas es una base de datos de secuencias de DNA.

10 La invención prevé en la etapa d) proporcionar finalmente una biomolécula, preferiblemente una molécula de DNA o una molécula de secuencia de aminoácido, es decir, una proteína, tanto mediante síntesis desde cero de la transaminasa como mediante aislamiento a partir de los polinucleótidos de un banco genético que codifican la transaminasa deseada con la ayuda de cebadores, estando definidos en base a las segundas secuencias de biomoléculas. Los polinucleótidos obtenidos, en particular moléculas de secuencia de DNA, se usan para ser sobre-expresadas en las condiciones apropiadas en un medio de cultivo apropiado para expresar la transaminasa deseada.

20 Se puede proporcionar un proceso para realizar el escrutinio y la identificación de una ω -transaminasa (R)-selectiva, que comprende las etapas a) a c) identificadas anteriormente, en particular proporcionando al menos una secuencia de biomolécula de consulta de al menos una ω -transaminasa (R)-selectiva y al menos un banco de biomoléculas, realizar una búsqueda en el banco de biomoléculas con la secuencia de biomolécula de consulta para identificar un grupo de primeras secuencias de biomolécula diana, en donde las primeras secuencias de biomoléculas tienen un grado de identidad de secuencia de al menos un 20%, preferiblemente un 25%, preferiblemente un 32% con respecto a la secuencia de biomolécula de consulta, calculado a nivel de aminoácidos, seleccionar, en el grupo de primeras secuencias de biomolécula, secuencias de biomoléculas que no comprendan, a nivel de aminoácido, ninguna de las estructuras de secuencia c1) a c3), siendo c1) en la posición 95 a 97 una secuencia de aminoácidos Tyr Xa1 Xa2, siendo Xa1 un aminoácido Ile, Val, Leu, Met, Phe, y siendo Xa2 un aminoácido Arg o Lys, o c2) siendo la posición 97 a 99 una secuencia de aminoácidos Tyr Xaa Gln, siendo Xaa un aminoácido y en la región de la posición 105 a 111, preferiblemente 107 a 109, una secuencia de aminoácidos Arg Xaa Xa3, siendo Xa3 preferiblemente un aminoácido, preferiblemente His, o c3) siendo la posición 38 Thr, en la posición 97 Lys y en la posición 107 a 109 una secuencia de aminoácidos Arg Xa4 Xa5, siendo Xa4 un aminoácido, que preferiblemente es Gly, y siendo Xa5 un aminoácido, que preferiblemente es Tyr, y que no comprende las estructuras de secuencia c4), c5) y c6), siendo c4) a nivel de aminoácidos en la posición 95 diferente de Tyr, Arg, Lys, o en la posición 95 Tyr, pero en la posición 97 no Arg o Lys, y siendo c5) en la posición 40 Lys o Arg, y siendo c6) en la región 161 a 165, preferiblemente en la posición 163, Lys para identificar un grupo de segundas secuencias de biomolécula, que son secuencias de biomolécula de una ω -TA (R)-selectiva.

40 En una realización preferida adicional de la presente invención, las presentes enseñanzas proporcionan ω -transaminasas (R)-selectivas que pueden obtenerse de acuerdo a los procesos de preparación de la presente invención. En particular, la presente invención proporciona proteínas y secuencias de DNA, en particular moléculas de DNA, que codifican dicha proteína a partir de *Mesorhizobium loti* y *Aspergillus terreus*, tal como se identifica en la SEQ ID NO: 1 y 2 correspondientes a *Mesorhizobium loti*, y 3 y 4 correspondientes a *Aspergillus terreus*. La presente invención proporciona en particular y en la realización más preferida la enseñanza de que las ω -transaminasas (R)-selectivas obtenidas y preparadas según la presente invención, por ejemplo las identificadas en la SEQ ID NO: 1 y 3, pueden usarse en una reacción de transaminación, en particular pueden usarse en un proceso para la preparación de una amina quiral ópticamente activa, preferiblemente un (R)-enantiómero de una amina quiral, que comprende hacer reaccionar al menos un aceptor de amino y al menos un donante de amino con la ω -transaminasa (R)-selectiva según la presente invención, en particular de SEQ ID NO: 1 ó 3, y obtener la amina quiral ópticamente activa. En una realización preferida, el proceso para la preparación de una amina quiral ópticamente activa es una síntesis asimétrica que usa un compuesto que contiene un grupo ceto, preferiblemente una cetona y un donante de amino. En otra realización preferida, el método de preparación usa una reacción de resolución cinética. Teniendo como eductos el donante de amino, que preferiblemente es una mezcla de aminas racémicas, y un aceptor de amino, preferiblemente cetonas.

55 En la realización preferida de la presente invención según la cual se sintetiza una amina quiral ópticamente activa usando una ω -transaminasa (R)-selectiva de la presente invención en una síntesis asimétrica que parte de cetonas, el grado de conversión preferido a la amina quiral ópticamente activa deseada, es decir al (R)-enantiómero, es de al menos 70%, 80%, 90%, 95%, 98%, 99%, 99,5% y lo más preferiblemente 100%.

60 En otra realización preferida según la cual se usa la ω -transaminasa (R)-selectiva de la presente invención en una reacción de resolución cinética que parte de aminas racémicas, el grado de conversión preferido a la amina quiral ópticamente activa, preferiblemente el (S)-enantiómero, es de al menos 30, 40, 45, 46, 47, 48, 49, en particular 50%.

Las concentraciones para analizar la pureza óptica y la conversión pueden determinarse por ejemplo mediante HPLC, electroforesis capilar (EC), cromatografía de gases (CG) o métodos fotométricos o fluorimétricos.

5 Por tanto, en una realización preferida, la presente invención se refiere a un proceso para la preparación de una amina quiral ópticamente activa, comprendiendo dicho proceso la reacción de un compuesto aceptor de amino que comprende un grupo ceto y una mezcla racémica de una amina en presencia de una transaminasa, en particular una ω -transaminasa (R)-selectiva, preferiblemente según la presente invención, para obtener un (S)-enantiómero de la amina quiral.

10 En otra realización preferida de la presente invención se proporciona un proceso para la preparación de una amina quiral ópticamente activa, en particular un (R)-enantiómero de dicha amina, proceso que comprende hacer reaccionar un compuesto aceptor de amino que comprende un grupo ceto y un donante de amino en presencia de una ω -transaminasa (R)-selectiva, en particular una obtenible según la presente invención, para obtener un (R)-enantiómero de la amina.

15 En el contexto de la presente invención un aceptor de amino es una molécula capaz de aceptar un grupo amino transferido desde un donante de amino por acción de una transaminasa, en particular una ω -transaminasa. En una realización particularmente preferida de la presente invención el aceptor de amino contiene una funcionalidad de cetona.

20 En una realización particularmente preferida de la presente invención el aceptor de amino se selecciona del grupo que consiste en ácido fenilpirúvico, una sal del mismo, ácido pirúvico, una sal del mismo, acetofenona, 2-cetoglutarato, 3-oxobutirato, 2-butanona, 3-oxopirrolidona (3-OP), 3-piridilmetilcetona (3-PMK), éster de etilo de ácido 3-oxobutírico (3-OBEE), éster de metilo de ácido 3-oxopentanoico (3-OPME), N-1-boc-oxopiperidinona, N-1-boc-3-oxopirrolidina (B3OP), 3-oxo-piperidina, N-1-boc-3-oxopiperidina (B3OPi), 1-Cbz-3-oxopiperidina (C3OPi), 1-Cbz-3-oxopirrolidina (C3OP), alquil-3-oxo-butanoatos, metoxiacetona y 1-oxotetralona.

25 En el contexto de la presente invención un donante de amino es una molécula capaz de proporcionar un grupo amino a un aceptor de amino usando una transaminasa, en particular una ω -transaminasa. En una realización preferida el donante de amino es una amina o un aminoácido.

30 En una realización particularmente preferida el donante de amino se selecciona del grupo que consiste en β -alanina, alanina, α -metilbencilamina (α -MBA), glutamato, fenilalanina, glicina, 3-aminobutirato, isopropilamina, 2-aminobutano y γ -aminobutirato o una sal, por ejemplo un cloruro, de cualquiera de ellos. En una realización particularmente preferida, el producto de cetona obtenido puede ser ácido fenilpirúvico, un sal del mismo, ácido pirúvico, una sal del mismo, ácido glioxílico, una sal del mismo, acetofenona, 2-cetoglutarato, acetona, 3-oxobutirato, 2-butanona, 3-oxopirrolidina (3-OP), 3-piridilmetilcetona (3-PMK), éster de etilo de ácido 3-oxobutírico (3-OBEE), éster de metilo de ácido 3-oxopentanoico (3-OPME), N-1-boc-oxopiperidinona y N-1-boc-3-oxopirrolidina (B3OP) o una sal, por ejemplo un cloruro, de cualquiera de ellos.

35 En una realización preferida adicional, la presente invención se refiere a un proceso para la preparación de una amina quiral ópticamente activa que es seleccionada del grupo de aminas que tienen un grupo amino ópticamente activo, en particular aminas con grupos alquílicos, grupos alquílicos ramificados o grupos arilalquílicos. En particular, dichas aminas, en particular aminas mono- o bi-cíclicas, son en particular aminas cíclicas de 5 a 6 miembros o hidrocarburos heterocíclicos S-, O- ó N-sustituídos o aminas aromáticas, en particular aminas aromáticas sustituidas con alquilo o alcoxi. En una realización preferida, las aminas quirales obtenidas se seleccionan del grupo que
40
45
consiste en fenilalalina, alanina, 3-aminopiperidina, alquil-3-amino-butanoatos, 3-aminopirrolidina (3-AP), 3-piridil-1-etilamina (3-PEA), N-1-boc-3-aminopirrolidina (B3AP), N-1-boc-3-aminopiperidina (B3APi), 1-Cbz-3-amino-piperidina (C3APi), 1-Cbz-3-aminopirrolidina (C3AP), éster de etilo de ácido 3-aminobutírico (3-ABEE), éster de metilo de ácido 3-aminopentanoico (3-APME), α -metilbencilamina (α -MBA), 1-aminotetralina, 3,4-dimetoxi fenil acetona, α -metil-4-(3-piridil)-butanamina, γ -aminobutirato, glutamato, isopropilamina, β -aminobutirato, sec-butilamina, metoxiisopropilamina, derivados de 3-aminopirrolidina, 1-N-Boc-3-aminopiperidina, cefalosporina y derivados de cefalosporina.

En una realización particularmente preferida, la presente invención prevé por tanto hacer reaccionar 3OP con una ω -transaminasa (R)-selectiva y un donante amino para obtener (R)-3AP ópticamente activa.

50 En una realización preferida adicional, la presente invención prevé hacer reaccionar 3-PMK con una ω -transaminasa (R)-selectiva y un donante amino para obtener (R) 3-PEA ópticamente activa.

En una realización preferida adicional, la presente invención prevé hacer reaccionar 3-OPME con una ω -transaminasa (R)-selectiva y un donante amino para obtener (R) 3-APME ópticamente activo.

En una realización preferida adicional, la presente invención prevé hacer reaccionar B3OPi con una ω -transaminasa (R)-selectiva y un donante amino para obtener (R)-B3APi ópticamente activa.

En una realización preferida adicional, la presente invención prevé hacer reaccionar C3OPi con una ω -transaminasa (R)-selectiva y un donante amino para obtener (R)-C3APi ópticamente activa.

En una realización preferida adicional, la invención prevé hacer reaccionar C3OP con una ω -transaminasa (R)-selectiva y un donante amino para obtener (R)-C3AP ópticamente activa.

5 En una realización preferida adicional, la invención prevé hacer reaccionar acetofenona con una ω -transaminasa (R)-selectiva y un donante amino para obtener (R)- α -MBA ópticamente activa.

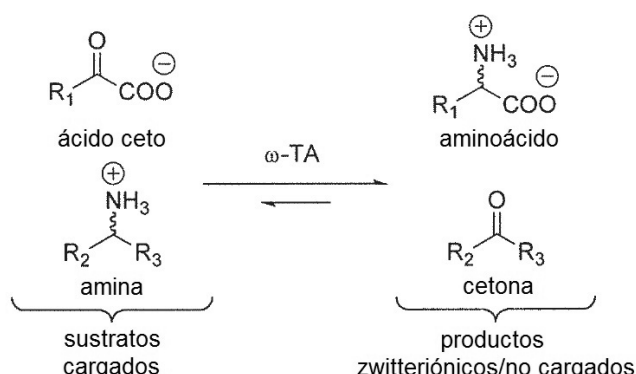
10 En una realización preferida adicional, la invención prevé hacer reaccionar un aceptor de amino, en particular hidrocarburos o aromáticos mono- o bicíclicos, cíclicos de 5 a 6 miembros que contengan un grupo oxo o heterocíclicos S-, O- o N-sustituidos, en particular aromáticos sustituidos con alquilo o alcoxi, con un donante de amino y una ω -transaminasa (R)-selectiva para obtener aminas, en particular aminas mono- o bi-cíclicas, en particular aminas cíclicas de 5 a 6 miembros, o aminas de hidrocarburos o aromáticas heterocíclicas S-, O- o N-sustituidas, en particular aminas aromáticas sustituidas con alquilo o alcoxi, en particular en forma (R).

15 En una realización particularmente preferida de la presente invención, el aceptor de amino y el donante de amino se hacen reaccionar con la transaminasa en medio acuoso, por ejemplo en tampón fisiológico. En una realización particularmente preferida la reacción de transaminación se lleva a cabo a un pH en el rango de 5 a 9, en particular de 7 a 8,5. En una realización preferida particular, la reacción se lleva a cabo en un rango de temperaturas entre 10 y 65°C, preferiblemente entre de 20 a 50°C, en particular de 18 a 25°C, preferiblemente a temperatura ambiente o de 34°C a 39°C, en particular 37°C. En una realización preferida adicional de la presente invención el aceptor de amino y el donante de amino se proporcionan en una relación molar de 1:1 a 1:5, en particular de 1:1 a 1:2. En una
20 realización preferida de la presente invención la actividad enzimática puede ser de 1 a 20.000 μ mol/min.

Se puede llevar a cabo un proceso para el análisis de una transaminasa, en particular para la caracterización de propiedades de una transaminasa, que comprende las siguientes etapas:

- i. proporcionar un aceptor de amino cargado, un donante de amino cargado y una transaminasa, preferiblemente una ω -transaminasa obtenible según la presente invención
- 25 ii. hacer reaccionar el aceptor de amino y el donante de amino con la transaminasa en un medio de reacción, y con ello
- iii. determinar la conductividad del medio de reacción en un primer conjunto de condiciones de reacción y
- iv. después de la etapa iii) determinar la conductividad del medio de reacción en un segundo conjunto de condiciones de reacción, para obtener al menos dos valores de conductividad que reflejen las propiedades
30 de la transaminasa.

En el curso de una reacción catalizada por una transaminasa, preferiblemente catalizada por una ω -TA, de los sustratos cargados, el donante de amino, preferiblemente una amina, y el aceptor de amino, preferiblemente un ácido ceto, por ejemplo piruvato, la conductividad disminuye ya que se forma una cetona no cargada y el componente amino zwitter-iónico, preferiblemente un aminoácido, por ejemplo alanina.



35 El proceso para el análisis permite una medición sencilla del progreso de la reacción. Preferiblemente, lo más adecuado es un tampón de baja conductividad, particularmente el tampón CHES (ácido N-ciclohexil-2-aminoetanosulfónico) de baja conductividad, a fin de evitar una conductividad inicial demasiado elevada. Preferiblemente, se puede realizar una calibración del proceso de conductividad mediante simulación de diferentes conversiones. A modo de ejemplo, para el par de sustrato estándar α -metilbencilamina y piruvato, una conversión de
40 1 mM corresponde a un cambio de 44 μ S. Una validación del presente proceso comparando las velocidades de

- reacción medidas con electroforesis capilar dio lugar a una conformidad excelente. Los extractos celulares no interfieren de forma significativa con el presente proceso. Puesto que el piruvato es el aceptor de amino común de virtualmente todas las ω -TA, el presente proceso puede usarse para investigaciones de la actividad de transaminasa con diferentes donantes de amino. Además, también puede obtenerse información sobre la enantioselectividad de la enzima.
- 5 Se puede llevar a cabo un proceso para analizar una transaminasa, en particular para la caracterización de las propiedades de una transaminasa, que permita analizar la actividad de la transaminasa en dependencia de, por ejemplo, el valor de pH o la temperatura del medio de reacción, o que permita analizar la estabilidad de la reacción, el efecto de aditivos o de composiciones tamponantes.
- 10 Según el proceso de análisis se puede llevar a cabo una primera medición de la conductividad del medio de reacción en un primer conjunto de condiciones de reacción, y después se puede llevar a cabo al menos una segunda medición de la conductividad a fin de poder comparar ambos valores de conductividad obtenidos y extraer conclusiones de las actividades y propiedades de la transaminasa evaluada. Un descenso de la conductividad muestra en una reacción de transaminasa según la presente invención la actividad de dicha transaminasa. El reconocimiento de un descenso reducido, un descenso acelerado o de ningún descenso de la conductividad permite extraer conclusiones sobre las propiedades de la transaminasa.
- 15 Un conjunto de condiciones de reacción puede ser preferiblemente un conjunto de condiciones de reacción cuyas condiciones se seleccionan preferiblemente del grupo que consiste en temperatura, valor de pH y composición del medio de reacción, preferiblemente las condiciones de reacción identificadas anteriormente, excepto la concentración de eductos y productos. En una realización de la presente invención, el conjunto de condiciones de reacción se mantiene constante a lo largo del tiempo. En otra realización de la invención, el conjunto de condiciones de reacción puede variar con el tiempo de reacción.
- 20 Se puede llevar a cabo un proceso para el análisis de una transaminasa, en donde el medio de reacción es un tampón de baja conductividad. En una realización particularmente preferida del proceso, el aceptor de amino cargado es piruvato.
- 25 El presente proceso para el análisis de una transaminasa puede usarse después del proceso de preparación de una ω -transaminasa (R)-selectiva de la presente invención, y extiende las enseñanzas de la presente invención no solo a proporcionar nuevas y ventajosas ω -transaminasas (R)-selectivas sino también para determinar sus características.
- Otras realizaciones preferidas de la presente invención son materia objeto de las sub-reivindicaciones.
- 30 La presente invención se ilustra en más detalle en los siguientes ejemplos y en el listado de secuencias acompañante.
- SEQ ID NO: 1: muestra la secuencia de aminoácidos completa (ORF) de la ω -TA (R)-selectiva procedente de *Mesorhizobium loti*.
- SEQ ID NO: 2: muestra la secuencia de DNA que codifica la SEQ ID NO: 1.
- 35 SEQ ID NO: 3: muestra la secuencia de aminoácidos completa (ORF) de la ω -TA (R)-selectiva procedente de *Aspergillus terreus*.
- SEQ ID NO: 4: muestra la secuencia de DNA que codifica la SEQ ID NO: 3.
- SEQ ID NO: 5: muestra la secuencia de aminoácidos completa (ORF) de la ω -TA (R)-selectiva procedente de *Mycobacterium aurum*.
- 40 SEQ ID NO: 6: muestra la secuencia de DNA que codifica la SEQ ID NO: 5.
- SEQ ID NO: 7: muestra la secuencia de aminoácidos completa (ORF) de la ω -TA (R)-selectiva procedente de *Arthrobacter* sp.
- SEQ ID NO: 8: muestra la secuencia de DNA que codifica la SEQ ID NO: 7.
- 45 SEQ ID NO: 9: muestra la estructura de secuencia de BCAT de *E. coli* usada para la determinación de las posiciones 95 a 99.
- SEQ ID NO: 10: muestra la estructura de secuencia de BCAT de *E. coli* usada para la determinación de la posición 38.
- SEQ ID NO: 11: muestra la estructura de secuencia de BCAT de *E. coli* usada para la determinación de la posición 107.

- SEQ ID NO: 12: muestra la estructura de secuencia de BCAT de *E. coli* usada para la determinación de la posición 163.
- SEQ ID NO: 13: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Penicillium chrysogenum*.
- 5 SEQ ID NO: 14: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 13.
- SEQ ID NO: 15: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Aspergillus niger*.
- SEQ ID NO: 16: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 15.
- SEQ ID NO: 17: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Aspergillus oryzae*.
- SEQ ID NO: 18: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 17.
- 10 SEQ ID NO: 19: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Aspergillus fumigatus*.
- SEQ ID NO: 20: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 19.
- SEQ ID NO: 21: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Neosartorya fischeri*.
- SEQ ID NO: 22: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 21.
- 15 SEQ ID NO: 23: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Gibberella zeae*.
- SEQ ID NO: 24: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 23.
- SEQ ID NO: 25: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Hyphomonas neptunium*.
- SEQ ID NO: 26: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 25.
- 20 SEQ ID NO: 27: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Mesorhizobium loti* MAFF303099.
- SEQ ID NO: 28: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 27.
- SEQ ID NO: 29: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Roseobacter* sp.
- SEQ ID NO: 30: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 29.
- 25 SEQ ID NO: 31: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Marinomonas* sp.
- SEQ ID NO: 32: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 31.
- SEQ ID NO: 33: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Rhizobium etli*.
- SEQ ID NO: 34: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 33.
- 30 SEQ ID NO: 35: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Rhodiferax ferriducens*.
- SEQ ID NO: 36: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 35.
- SEQ ID NO: 37: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Jannaschia* sp.
- SEQ ID NO: 38: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 37.
- 35 SEQ ID NO: 39: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Labrenzia alexandrii*.
- SEQ ID NO: 40: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 39.
- SEQ ID NO: 41: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Burkholderia* sp.
- SEQ ID NO: 42: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 41.
- 40 SEQ ID NO: 43: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Burkholderia cenocepacia*.

SEQ ID NO: 44: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 43.

SEQ ID NO: 45: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de protobacteria alfa.

SEQ ID NO: 46: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 45.

5 SEQ ID NO: 47: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de protobacteria gamma.

SEQ ID NO: 48: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 47.

SEQ ID NO: 49: muestra la secuencia de DNA que codifica la ω -TA (R)-selectiva procedente de *Mycobacterium vanbaalenii*.

SEQ ID NO: 50: muestra la secuencia de aminoácidos completa (ORF) de la SEQ ID NO: 49.

10 Figura 1: muestra un cromatograma de B3APi obtenida mediante síntesis estándar.

Figura 2: muestra un cromatograma de B3APi obtenida mediante asimétrica según la invención.

Figura 3: muestra un cromatograma de C3AP obtenida mediante asimétrica según la invención.

Figura 4: muestra un cromatograma de MPPA obtenida mediante asimétrica según la invención.

Figura 5: muestra un cromatograma de B3AP obtenida mediante asimétrica según la invención.

15 **Ejemplos:**

Ejemplo 1 – Identificación de ω -transaminasas (R)-selectivas

20 La secuencia de aminoácidos de la ω -TA (R)-selectiva de *Mycobacterium aurum*, tal cual se presenta en el documento EP 1 038 953 A1 (SEQ ID NO: 5) y la secuencia de aminoácidos de la ω -TA (R)-selectiva de *Arthrobacter* sp., tal cual se presenta en el documento EP 0 987 332 A1 (SEQ ID NO: 7), se usa como secuencias de biomolécula de consulta. El banco de biomoléculas usado en este ejemplo es el banco de datos de proteínas pubmed del NCBI (<http://www.ncbi.nlm.nih.gov/pubmed>) (13 julio 2009).

25 Usando una búsqueda BLAST con la secuencia de aminoácidos de ω -TA de *M. aurum* o *Arthrobacter* sp. (SEQ ID NO: 5 ó 7) como consulta con los parámetros estándar (matriz de puntuación BLOSUM62, tamaño de palabra: 3, costes de hueco: existencia – 11, extensión – 1) se ha identificado un primer grupo de 100 secuencias de aminoácido diversas de diferentes organismos, que tienen todas un grado mínimo del 30 % de identidad de secuencia con respecto a la secuencia de consulta.

30 Para el BLAST se ha usado, el 13 de julio de 2009, “secuencias de proteína no redundantes (nr)” (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome). En este primer grupo de 100 aminoácidos diversos que representan las primeras secuencias de biomolécula diana, dichas secuencias han sido buscadas e identificadas por una estructura de secuencia que no presenta las estructuras de secuencia c1), c2) ó c3) y que no muestra las estructuras de secuencia c4), c5) y c6). Se pudo identificar 21 ORF y se enumeran a continuación en la Tabla 1.

Tabla 1

Nº	Organismos fuente	% de identidad con		Nº ident. gen/ Nº ident. proteína (base de datos NCBI)
		AS- ω -TA SEQ ID NO: 7	MA- ω -TA SEQ ID NO: 5	
1	<i>Aspergillus terreus</i> NIH2624	44	40	115385557
2	<i>Penicillium chrysogenum</i> Wisconsin 54-1255	44	42	211591081
3	<i>Aspergillus niger</i>	40	36	145258936
4	<i>Aspergillus oryzae</i> RIB40	41	40	169768191
5	<i>Aspergillus fumigatus</i> Af293	41	38	70986662
6	<i>Neosartorya fischeri</i> NRRL 181	41	38	119483224
7	<i>Gibberella zeae</i> PH-1	40	39	46109768
8	<i>Hyphomonas neptunium</i> ATCC 15444	44	40	114797240

Nº	Organismos fuente	% de identidad con		Nº ident. gen/ Nº ident. proteína (base de datos NCBI)
		AS- ω -TA SEQ ID NO: 7	MA- ω -TA SEQ ID NO: 5	
9	<i>Mycobacterium vanbaalenii</i> PYR-1	50	91	120405468
10	<i>Mesorhizobium loti</i> MAFF303099	38	37	13471580
11	<i>Mesorhizobium loti</i>	35	37	20804076
12	<i>Roseobacter</i> sp. MED193	37	37	86137542
13	<i>Marinomonas</i> sp. MED121	36	34	87122653
14	<i>Rhizobium etli</i> CIAT 652	33	32	190895112
15	<i>Rhodferax ferrireducens</i> T118	38	36	89899273
16	<i>Jannaschia</i> sp. CCS1	37	32	89053613
17	<i>Labrenzia alexandrii</i> DFL-11	42	36	EEE43073
18	<i>Burkholderia</i> sp. 383	32	32	78059900
19	<i>Burkholderia cenocepacia</i> HI2424	36	33	ABK12047
20	<i>Alpha proteobacterium</i> HTCC2255	36	34	ZP_01448442
21	<i>Gamma proteobacterium</i>	27	26	
	Proteobacteria alfa	49	100	SEQ ID 5
	Proteobacteria gamma	100	49	SEQ ID 7

AS: *Aspergillus* sp., MA: *Mycobacterium aurum*

Ejemplo 2 – Preparación y análisis de las transaminasas de *Aspergillus terreus*, *Mycobacterium vanbaalenii* y *Mesorhizobium loti*

5 2.1. La ω -TA de *Mycobacterium vanbaalenii*, entrada 9 de la Tabla 1 (SEQ ID NO: 49 y 50), de *Aspergillus terreus*, entrada 1 de la Tabla 1 (SEQ ID NO: 3 y 4), y de *Mesorhizobium loti*, entrada 11 de la Tabla 1 (SEQ ID NO: 1 y 2), han sido obtenidas y usadas en forma adaptada a utilización de codón (para *E. coli*) para expresar las enzimas en *Escherichia coli*. La transaminasa de *Mycobacterium vanbaalenii* se denomina en adelante Mva-TA, la transaminasa de *Aspergillus terreus* Ate-TA y la de *Mesorhizobium loti* Mlo-TA.

2.2. Acetofenoensayo

10 La Mva-TA y la Ate-TA convirtieron (R)- α -MBA al menos 100 veces más rápido que el (S)-enantiómero en un acetofenoensayo con amina 2,5 mM y piruvato 2,5 mM a pH 7,5 y 30°C.

Ensayo: se monitorizó el aumento de absorción de la acetofenona formada durante la reacción a 245 nm.

La Mlo-TA no convirtió ni la (R)- ni la (S)- α -MBA.

15 A continuación, se han evaluado aminas adicionales en presencia de piruvato o α -cetoglutarato como aceptores de amino (amina 10 mM, piruvato 10 mM, PLP 0,1 mM, tampón de fosfato pH 7,5, incubación 24 h a 30°C, análisis mediante cromatografía de capa fina).

Se pudo observar conversión para 2-aminoheptano, 2-aminopentano, 1,3-dimetilbutilamina y 4-fenil-2-aminobutano. Asimismo, se pudo detectar una conversión mínima de isopropilamina.

20 No se pudo detectar conversión con otros donantes de amino tales como D-alanina, L-valina, γ -aminobutirato, etilamina, bencilamina, putrescina, 2-amino-3-metilbutano y 3,3-dimetil-2-aminobutano.

De este modo, se demostró que las tres proteínas son ω -TA.

En particular, usando como sustratos (R)- y (S)-2-aminoheptano, solo el enantiómero (R) fue convertido significativamente. Por tanto, también la Mlo-TA es una ω -TA (R)-selectiva. No se observó actividad DATA (D-aminoácido transaminasa) o BCAT (aminotransferasa de cadena ramificada) en ninguna de las tres proteínas.

25 2.3. Ensayo de conductividad.

Asimismo, se usaron 1-N-boc-3-aminopirrolidina (B3AP), 1-N-boc-3-aminopiperidina (B3APi) y 1-Cbz-3-aminopiperidina (C3APi) como sustratos para determinar las actividades relativas de dichas sustancias contra el sustrato modelo α -MBA.

Durante la reacción de la amina y el piruvato (ambos sustratos se cargan a pH 7,5) en alanina y cetona (la cetona no tiene carga, la alanina es un compuesto zwitteriónico y no contribuye a la conductividad), la monitorización de la cinética de la conductividad permite concluir las velocidades de conversión.

5 Antes de comenzar la reacción, se llevó a cabo una calibración determinando diferentes conversiones en dependencia de diversas concentraciones de alanina, piruvato, cetona y amina.

La reducción de la conductividad fue por conversión de mM de α -MBA, B3AP, B3APi y C3APi 44 μ S, 50 μ S, 48,5 μ S y 49,3 μ S.

10 Además de las tres transaminasas expresadas recombinantemente, se evaluó una ATA-117 (R)-selectiva disponible comercialmente en Codexis usando las siguientes condiciones de reacción: tampón CHES 50 mM, pH 7,5, ajuste de pH con BIS-TRIS (=Bis(2-hidroxietil)amino-tris(hidroximetil)metano), PLP 0,1 mM, amina 5-6 mM y piruvato, reacción a 25°C.

Se pudo demostrar una conversión de los sustratos para Ate-TA. Las actividades relativas fueron 2 % para (R)-B3AP y (R,S)-C3APi y de 1 % para (R)-B3APi, en comparación con la (R)- α -MBA.

15 2.4. Determinación de la enantioselectividad mediante síntesis asimétrica de ω -TA (R)-selectiva de *Aspergillus terreus* y *Mesorhizobium loti*

Mediante un método independiente se demuestra que la Mlo y la Ate-TA son (R)-selectivas y convierten los sustratos en los productos deseados con una enantioselectividad excelente.

Como prueba definitiva de la (R)-enantioselectividad de ambas transaminasas, se llevó a cabo una síntesis asimétrica de diferentes aminas y se determinó la pureza óptica de las mismas.

20 Como donante de amino se usó un exceso de 100 a 1 de alanina. La conversión no se determinó exactamente, sino que se estimó de forma aproximada. En experimentos adicionales se podrían desarrollar métodos para incrementar la conversión (PDC). Se obtuvieron enantioselectividades entre altas y excelentes con ambas transaminasas, excepto con C3APi, en donde solo se obtuvo una enantioselectividad muy baja.

Tabla 2

Amina	Ate-TA		Mlo-TA	
	%ee	%c	%ee	%c
B3AP	99,8	40 \pm 20	-	-
C3AP	99,6	40 \pm 20	-	-
B3APi	>99,9	30 \pm 20	-	-
C3APi	49	30 \pm 20		
MPPA	-	-	98	15 \pm 10

25

Métodos

Las transaminasas fueron expresadas en *E. coli* BL21.

30 El medio de cultivo (50 mL de medio LD-Amp) fue inducido con un 0,2% de rhamnosa, cuando se obtuvo una densidad óptica (DO) de 0,5, y fue cultivado durante 12 horas a 25°C. Posteriormente, las células fueron lavadas con un tampón de fosfato sódico, se centrifugaron, se suspendieron en 1 mL de tampón de fosfato sódico (pH 7,5) y se congelaron en alícuotas de 200 μ L.

35 Para la biocatálisis se añadieron 5 μ L de una disolución de cetona 500 mM en DMSO, 22 mg de D-alanina y 10 μ L de PLP 10 mM a dichas alícuotas. La mezcla de reacción se completó hasta un volumen final de 500 μ L con un tampón de fosfato sódico 50 mM (pH 7,5). Las mezclas de reacción fueron incubadas durante una noche a 25 °C y 500 rpm. Para la determinación del exceso enantiomérico mediante EC se añadieron 200 μ L de NaOH 1 N a 400 μ L de muestra, se extrajo con 400 μ L de diclorometano y se separó la fase orgánica. A continuación la fase orgánica fue extraída con 100 μ L de un tampón de fosfato de trietilamonio 5 mM (pH 3). Posteriormente, la fase acuosa obtenida se inyectó en la EC.

Programa de separación:

40 EC-Capilar 30 cm, 50 μ m de diámetro interno, temperatura 15 °C

- lavar con tampón de fosfato de trietilamonio pH 3, 1 minuto, 2,07 bar (30 psi)

- lavar con un 5% de ciclo dextrina altamente sulfatada (HSaCD o HSyCD), 1 minuto, 0,69 bar (10 psi)
 - inyección: 5-10 s, 0,069 bar (1 psi)
 - sumergir en agua
 - separación: voltaje 10 ó 15 kV
- 5 - detección: MPPA, C3AP y C3APi a 200 nm; B3AP, B3APi a 190 nm

Condiciones de separación:

Tabla 3

	Segmento de separación [cm]	Selector quiral	Tiempo de migración [min]		Voltaje durante la separación
			S-Amin	R-Amin	
MPPA	10	HSyCD	2,4	2,8	15 kV
B3AP	10	HSyCD	4,55	6,1	15 kV
B3APi	20	HSyCD	12,0	12,4	15 kV
C3AP	20	HSaCD	10,2	11,3	10 kV
C3APi	20	HSaCD	6,8	7,2	15 kV

Ejemplo 4 – Preparación y análisis de todas las transaminasas identificadas en la Tabla 1.

10 4.1. Expresión y purificación de las transaminasas

15 Los marcos de lectura abiertos optimizados para codón que codifican las proteínas con los números de entrada 1, 2, 3, 4, 8, 9, 11, 13, 14, 15, 17, 18 y 21 de la Tabla 1 fueron insertados en pGASTON entre los sitios de restricción NdeI y BamHI usando una estrategia de clonación independiente de ligación. Los ORF optimizados para codón que codifican todas las demás proteínas fueron ordenados ya subclonados en pET-22b. Se cultivaron cepas de *E. coli* BL21 (DE3) transformadas en 400 mL de medio LB suplementado con ampicilina (100 µg/mL). Las células fueron incubadas inicialmente a 37°C en un agitador giratorio hasta que la DO₆₀₀ alcanzó 0,7. A continuación las células fueron inducidas mediante la adición de rhamnosa 0,2% (pGASTON) o IPTG 0,1 mM (pET-22b), respectivamente, y al mismo tiempo la temperatura de incubación se redujo a 20°C. Tras la inducción, se continuó con la incubación durante otras 20 h. Para seguir la expresión se extrajeron alícuotas a diversos tiempos tras la inducción.

20 La partícula de células (~3 g de peso húmedo) se lavó dos veces con tampón de fosfato (pH 7,5, 50 mM), que contenía PLP 0,1 mM a 4°C. Tras disrupción (prensa francesa) la suspensión celular fue centrifugada (4000 × g, 30 min) y el sobrenadante resultante se hizo pasar a través de un filtro de 0,5 µm antes del análisis cromatográfico. La cromatografía se llevó a cabo usando un Purificador ÁKTA (GE Healthcare). El extracto celular filtrado se aplicó a una columna de 5 mL de IMAC Sepharose™ 6 Fast Flow (GE Healthcare). La columna se lavó con un caudal de 5 mL min⁻¹ con un volumen de 10 columnas de tampón de fosfato 50 mM, pH 7,5, que contenía NaCl 300 mM, PLP 0,1 mM e imidazol 30 mM (para evitar unión inespecífica) y la actividad ATA se eluyó con 10 volúmenes de columna de tampón de fosfato (pH 7,5, 50 mM), que contenía NaCl 300 mM, PLP 0,1 mM e imidazol 300 mM (caudal de 5 mL min⁻¹). Las fracciones que contenían actividad se agruparon y se desalinizaron mediante cromatografía en gel con un tampón de Tricina 20 mM, pH 7,5, que contenía PLP 0,01 mM. Las enzimas purificadas fueron almacenadas a 30 4°C.

La cantidad de cada proteína purificada a partir de aproximadamente 3 g de células (peso húmedo) se muestra a continuación en la Tabla 4.

Tabla 4

Entrada	Fuente de enzimas	Rendimiento de proteína tras purificación [mg]
1	<i>Aspergillus terreus</i>	8,6
2	<i>Penicillium chrysogenum</i>	26,2
3	<i>Aspergillus niger</i>	-
4	<i>Aspergillus oryzae</i>	20,6
5	<i>Aspergillus fumigatus</i>	14,8
6	<i>Neosartorya fischeri</i>	23,3

Entrada	Fuente de enzimas	Rendimiento de proteína tras purificación [mg]
7	<i>Gibberella zeae</i>	4,8
8	<i>Hyphomonas neptunium</i>	6,5
9	<i>Mycobacterium vanbaalenii</i>	8,9
10	<i>Mesorhizobium loti</i> MAFF303099	6,9
11	<i>Mesorhizobium loti</i>	5,3
12	<i>Roseobacter sp.</i>	27,5
13	<i>Marinomonas sp.</i>	23,7
14	<i>Rhizobium etli</i>	6,5
15	<i>Rhodoferrax ferrireducens</i>	7,5
16	<i>Jannaschia sp.</i>	24,8
17	<i>Labrenzia alexandrii</i>	12,5
18	<i>Burkholderia sp.</i>	41,6
19	<i>Burkholderia cenocepacia</i>	-
20	Proteobacteria alfa	-
21	Proteobacteria gamma	2,6

4.2. Caracterización de la especificidad de sustrato de ω -TA (R)-selectiva

- 5 Para determinar la actividad hacia α -metilbencil amina (α -MBA) en el escrutinio inicial de las proteínas expresadas, se usó un ensayo basado en acetofenona: se hizo reaccionar una disolución de (R)- ó (S)- α -MBA 2,5 mM y piruvato en presencia de la enzima purificada y se correlacionó el incremento de la absorbancia a 245 nm con la formación de acetofenona. Las conversiones de las aminas 2-aminohexano, 4-fenil-2-aminobutano y 1-N-Boc-3-aminopirrolidina fueron monitorizadas mediante un ensayo de conductividad: se hizo reaccionar una disolución que contenía la amina 10 mM y piruvato en presencia de la amina transaminasa purificada y se relacionó el descenso de la conductividad con la conversión del sustrato.
- 10 Para investigar la actividad DATA y BCAT se midió el descenso de NADH espectrofotométricamente a 340 nm usando ensayos acoplados de deshidrogenasa: se hizo reaccionar una disolución de ácido α -cetoglutarico 5 mM y D-alanina en presencia de la transaminasa purificada, 1 U/mL de lactato deshidrogenasa y NADH 0,5 mM para medir la actividad DATA. Se usó una disolución que contenía ácido 3-metil-2-oxobutírico 5 mM y L-glutamato, cloruro amónico 10 mM, 1 U/mL de glutamato deshidrogenasa y NADH 0,5 mM para medir la actividad BCAT.
- 15 Todas las reacciones se llevaron a cabo en tampón de Tricina 20 mM, pH 7,5, que contenía PLP 0,01 mM. Se ajustó el pH del tampón con 1,8-diazabicyclo[5.4.0]undec-7-eno.

Los resultados se muestran a continuación en la Tabla 5.

Tabla 5: Actividades específicas para varias sustancias.

Sustratos	piruvato 1		piruvato 2		piruvato 3		piruvato 4		2KG D-Ala	MOB L-Glu
	R	S	R	S	R	S	R	S		
1	15,2	<0,001	2,91	<0,001	9,7	<0,001	0,031	<0,001	<0,001	0,003
2	1,3	<0,001	1,1	0,044	5,6	<0,001	0,264	<0,001	<0,001	<0,001
3	- ^{a)}	-	-	-	-	-	-	-	-	-
4	3,7	0,001	1,4	0,023	5,2	0,002	0,051	0,002	<0,001	<0,001
5	4,1	<0,001	2,4	<0,001	4,5	<0,001	0,009	<0,001	<0,001	0,005
6	4,5	<0,001	7,4	<0,001	6,0	<0,001	0,013	<0,001	<0,001	0,005
7	18,6	<0,001	19,6	<0,001	8,2	<0,001	<0,001	<0,001	<0,001	0,016
8	3,6	<0,001	3,2	0,225	20,7	<0,001	0,163	<0,001	<0,001	0,012
9	4,7	<0,001	5,6	<0,001	2,6	<0,001	<0,001	<0,001	<0,001	0,003
10	0,011	<0,001	0,003	<0,001	0,010	<0,001	0,001	<0,001	0,004	0,004
11	0,013	<0,001	0,124	<0,001	0,002	<0,001	<0,001	<0,001	<0,001	0,005
12	0,003	<0,001	0,001	<0,001	0,001	<0,001	0,001	<0,001	0,003	0,002

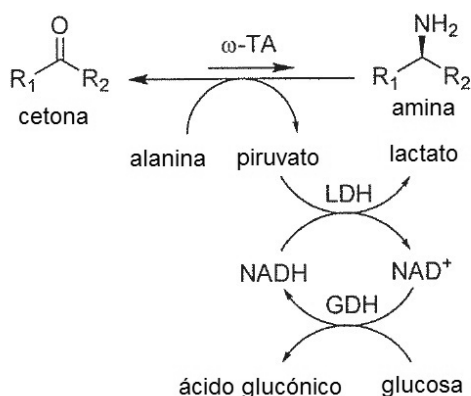
Sustratos	piruvato 1		piruvato 2		piruvato 3		piruvato 4		2KG D-Ala	MOB L-Glu
	R	S	R	S	R	S	R	S		
Entrada										
13	0,002	<0,001	0,020	<0,001	0,003	<0,001	<0,001	<0,001	<0,001	0,003
14	0,867	<0,001	0,012	<0,001	0,260	<0,001	<0,001	<0,001	0,020	0,016
15	0,056	<0,001	0,001	<0,001	0,307	<0,001	<0,001	<0,001	0,010	0,098
16	0,059	0,007	0,071	0,002	0,370	0,068	0,022	<0,001	0,062	0,020
17	0,060	0,003	0,073	0,001	0,120	0,027	0,205	0,002	0,063	0,023
18	0,017	<0,001	0,002	<0,001	1,1	0,007	<0,001	<0,001	<0,001	0,001
19	- ^{a)}	-	-	-	-	-	-	-	-	-
20	-	-	-	-	-	-	-	-	-	-
21	0,028	<0,001	0,610	0,004	0,034	<0,001	<0,001	<0,001	<0,001	0,031

1 – aminohexano, 2 – α-MBA, 3 – 4-fenil-2-aminobutano, 4 – 1-N-Boc-3-aminopirrolidina, 2KG – 2-cetoglutarato, D-Ala – D-alanina, L-Glu – L-glutamato, MOB – ácido 3-metil-2-oxobutírico. El número de entrada corresponde a la Tabla 1. Todas las medidas se realizaron al menos por duplicado. La desviación de las medidas individuales respecto al valor medio fue < 10 %.

^{a)} La medida no fue posible debido a que el rendimiento de proteína durante la expresión fue muy bajo/la proteína era inestable durante la purificación.

4.3. Síntesis asimétrica de (R)-aminas 1-4 (ver leyenda de la Tabla 5 anterior) con ω-Tas de *Aspergillus Terreus*, *Mesorhizobium loti* y *Mycobacterium vanbaalenii*

10 Las síntesis asimétricas se realizaron a 30°C durante 24 horas en tampón de fosfato sódico (100 mM, pH 7) que contenía monohidrato de piridoxal-5'-fosfato PLP (1 mM) y NAD⁺ (1 mM) en tubos Eppendorf de 1,5 mL.



15 La mezcla de reacción contenía cetona 50 mM, L-alanina (5 equivalentes, 250 mM), lactato deshidrogenasa de corazón bovino (90 U), glucosa (150 mM) y glucosa deshidrogenasa (15 U). Se expresó ω-TA de *Aspergillus terreus*, *Mesorhizobium loti* y *Mycobacterium vanbaalenii* (entrada 1, 11 y 9 de la Tabla 1) en BL21 de *E. coli* tal como se ha descrito anteriormente, se congelaron en alícuotas y se aplicaron directamente como biocatalizadores de célula completa sin purificación adicional. La conversión se midió mediante detección de las aminas formadas (1, cromatografía de gases (CG); 2-4 electroforesis capilar (EC)). El análisis quiral de 2-4 se llevó a cabo usando EC como se ha descrito anteriormente. El valor de exceso enantiomérico (% ee) para 1 se analizó mediante CG. Tras la extracción de la amina con acetato de etilo, se realizó una derivatización a trifluoroacetamida añadiendo un exceso de 20 a 1 de anhídrido de ácido trifluoroacético. Tras purgar con nitrógeno para eliminar el exceso de anhídrido y el ácido trifluoroacético residual, el compuesto derivatizado se disolvió en acetato de etilo (50 µL) y la línea base se separó usando un Shimadzu GC14A que fue equipado con una columna de Heptakis-(2,3-di-O-acetil-6-O-*tert*-butildimetilsilil)-β-ciclodextrina (25 m x 0,25 mm). Los tiempos de retención fueron 16,0 minutos ((S)-1) y 16,2 minutos ((S)-2) con un gradiente de temperatura en el horno de 80 °C/10 min // 20 °C // 180 °C/10 min.

Los resultados se presentan a continuación en la Tabla 6.

Tabla 6

Aminas formadas	Ω -TA	Conversión [%] ^b	Exceso enantiomérico [%eeP] ^c
1	Ate	32	>99
1	Mlo	41	>99
1	Mva	35	>99
2	Ate	15	>99
2	Mlo	1	95,0
2	Mva	2	>99
3	Ate	14	>99
4	Ate	11	>99

^a Condiciones de reacción: cetona 50 mM, D-alanina 250 mM, tampón de fosfato sódico 100 mM pH 7,0, PLP 1 mM, NADH 1 mM. El piruvato co-producto de la reacción se eliminó con lactato deshidrogenasa (LDH). Para el reciclado del cofactor se usó glucosa deshidrogenasa (GDH). ^b Las conversiones no han sido optimizadas. La desviación de una medida individual respecto al valor medio no excedió el 10 %. El compuesto 4 sólo fue convertido mediante Ate-TA. ^c (R)-enantiómeros.

5

LISTA DE SECUENCIAS

<110> Lonza AG

<120> Un proceso para la identificación y preparación de una omega-transaminasa (R)-específica

<130> 202663

<160> 50

<170> PatentIn versión 3.5

<210> 1

<211> 322

<212> PRT

<213> Mesorhizobium loti

<400> 1

```

Met Thr Leu Ala Thr Thr Asp Ala Thr Val Gly Val Pro Glu Val Glu
1          5          10          15

Thr Thr His Lys Asp Thr Arg Arg Tyr Pro His Gly Val Ala Phe Met
          20          25          30

Asp Gly Gln Tyr Leu Pro Met Ser Glu Ala Lys Ile Ser Val Leu Asp
          35          40          45

Trp Gly Phe Leu His Ser Asp Ala Thr Tyr Asp Thr Val His Val Trp
          50          55          60

Glu Gly Arg Phe Phe Arg Leu Asp Leu His Leu Asp Arg Phe Phe Arg
65          70          75          80

Gly Met Asp Arg Leu Arg Met Lys Leu Pro Tyr His Arg Arg Glu Val
          85          90          95

Glu Arg Val Leu Ser Asn Cys Val Ala Leu Ser Gly His Lys Ser Ala
          100          105          110

Tyr Val Glu Met Ile Cys Thr Arg Gly Gly Ser Pro Thr Phe Ser Arg
          115          120          125

Asp Pro Arg Glu Ala Glu Asn Arg Phe Ile Ala Phe Ala Val Pro Phe
          130          135          140

Gly Ser Val Ala Asn Lys Glu Gln Leu Glu Arg Gly Leu His Val Gly
145          150          155          160

Val Ser Glu Thr Val Arg Ile Pro Pro Lys Ser Val Asp Pro Thr Ile
          165          170          175

```

ES 2 599 644 T3

Lys Asn Tyr His Trp Leu Asp Leu Val Arg Gly Leu Tyr Asp Ala Tyr
 180 185 190

Asp Val Gly Ala Glu Thr Ala Leu Ile Met Asp Thr Asn Gly Asn Ile
 195 200 205

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

Lys Thr Pro Ala Phe Gly Val Leu Pro Gly Ile Thr Arg Gln Ser Val
 225 230 235 240

Phe Asp Leu Cys Gly Glu Val Gly Leu Ala Val Thr Ala Ala Asp Leu
 245 250 255

Pro Arg Leu Glu Leu Gly Glu Ala Asp Glu Val Phe Ile Thr Ser Thr
 260 265 270

Ala Gly Gly Ile Met Pro Val Thr Arg Val Asp Gly Ser Ser Ile Gly
 275 280 285

Ser Gly Lys Val Gly Val Val Thr Arg Gln Leu Met Asp Leu Tyr Trp
 290 295 300

Gln Lys His Ser Asp Asp Ala Trp Ser Thr Pro Val Lys Tyr Ala Ser
 305 310 315 320

Gly Ser

<210> 2

<211> 969

<212> DNA

<213> Mesorhizobium loti

<400> 2

```

atgaccctgg caaccaccga tgcaaccggt ggtggtccgg aagttgaaac cacccataaa      60
gatacccgtc gttatccgca tgggtgttgca tttatggatg gtcagtatct gccgatgagc      120
gaagcaaaaa ttagcgttct ggattggggg tttctgcatt ctgatgccac ctatgatacc      180
gttcatgttt ggggaaggtcg tttttttcgt ctggatctgc atctggatcg cttttttcgt      240
ggtatggatc gtctgcgcat gaaactgccg tatcatcgtc gtgaagttga acgtgttctg      300
agcaattgtg ttgcactgag cggtcataaa agcgcctatg tggaaatgat ttgtaccctg      360
ggtggtagcc cgaccttag ccgtgatccg cgtgaagcag aaaatcgctt tattgcattt      420
gcagttccgt ttggttctgt ggcaaataaa gaacagctgg aacgtggtct gcatgttgg      480
    
```

ES 2 599 644 T3

gtagcgaaa ccgttcgtat tcctcogaaa agcgttgatc cgaccattaa aaattatcat 540
 tggctggatc tggttcgtgg tctgtatgat gcctatgatg ttggtgcaga aaccgcactg 600
 attatggata ccaatggcaa tattgcagaa ggtccgggtt ttaacgtggt tacctgtaaa 660
 aatcgtcagc tgaaaacacc ggcatttggg gttctgcctg gtattacacg tcagagcgtt 720
 tttgatctgt gtggtgaagt tggctctggca gttaccgcag cagatctgcc tcgtctggaa 780
 ctgggtgaag cagatgaagt ttttattacc agcaccgcag gggcattat gccggttacc 840
 cgtgttgatg gtagcagcat tggtagcggg aaagttgggt ttggtaccgc tcagctgatg 900
 gatctgtatt ggcagaaaca ttctgatgat gcatgggtcta caccgggtaa atatgcctca 960
 ggatcctga 969

<210> 3
 <211> 328
 <212> PRT
 <213> Aspergillus terreus

<400> 3

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

ES 2 599 644 T3

145					150					155				160	
Ala	Val	Val	Ala	Arg	Thr	Val	Arg	Arg	Val	Pro	Pro	Gly	Ala	Ile	Asp
				165					170					175	
Pro	Thr	Val	Lys	Asn	Leu	Gln	Trp	Gly	Asp	Leu	Val	Arg	Gly	Met	Phe
			180					185					190		
Glu	Ala	Ala	Asp	Arg	Gly	Ala	Thr	Tyr	Pro	Phe	Leu	Thr	Asp	Gly	Asp
		195					200					205			
Ala	His	Leu	Thr	Glu	Gly	Ser	Gly	Phe	Asn	Ile	Val	Leu	Val	Lys	Asp
	210					215					220				
Gly	Val	Leu	Tyr	Thr	Pro	Asp	Arg	Gly	Val	Leu	Gln	Gly	Val	Thr	Arg
225					230					235					240
Lys	Ser	Val	Ile	Asn	Ala	Ala	Glu	Ala	Phe	Gly	Ile	Glu	Val	Arg	Val
				245					250					255	
Glu	Phe	Val	Pro	Val	Glu	Leu	Ala	Tyr	Arg	Cys	Asp	Glu	Ile	Phe	Met
			260					265					270		
Cys	Thr	Thr	Ala	Gly	Gly	Ile	Met	Pro	Ile	Thr	Thr	Leu	Asp	Gly	Met
		275					280					285			
Pro	Val	Asn	Gly	Gly	Gln	Ile	Gly	Pro	Ile	Thr	Lys	Lys	Ile	Trp	Asp
	290					295					300				
Gly	Tyr	Trp	Ala	Met	His	Tyr	Asp	Ala	Ala	Tyr	Ser	Phe	Glu	Ile	Asp
305					310					315					320
Tyr	Asn	Glu	Arg	Asn	Ser	Gly	Ser								
				325											

<210>

<211> 987

<212>DNA

<213> Aspergillus terreus

<400> 4

atggcaagca tggataaagt ttttgccggt tatgcagcac gtcaggcaat tctggaaagc	60
accgaaaacca ccaatccgtt tgcaaaaggt attgcatggg ttgaaggatga actggttcog	120
ctggcagaag cacgtattcc gctgctgat cagggtttta tgcatagcga tctgacctat	180
gatgttcoga gcgtttggga tggctgttt ttttgtctgg atgatcatat taccogtctg	240
gaagccagct gtaccaaact gcgtctgctg ctgccgctgc ctcgtgatca ggtaaacia	300

ES 2 599 644 T3

attctggttg aaatggttgc caaaagcggg attcgtgatg catttgtgga actgattggt 360
 acccgtggtc tgaaggtgt tctggcacc cgtccggaag atacgtgaa taatctgtat 420
 atgtttgtgc agccgtatgt ttgggttatg gaaccggata tgcagcgtgt tgggtgtagc 480
 gcagttggtg cacgtaccgt tctcgtggt cgcctgggtg caattgatcc gaccgttaa 540
 aatctgcagt ggggtgatct ggttcgtggt atgtttgaag cagcagatcg tgggtcaacc 600
 tatccgtttc tgaccgatgg tcatgcacat ctgaccgaag gtagcggttt taacattgtg 660
 ctggtgaaag atggtgttct gtatacaccg gatcgtgggtg ttctgcaggg tgttacacgt 720
 aaaagcgtga ttaatgcagc agaagccttt ggtattgaag tgcgtggtga atttgttccg 780
 gttgaactgg catatcgtct tcatgaaatt tttatgtgta ccaccgcagg cggattatg 840
 ccgattacca cctgggatgg tatgccggtt aatggtggtc agattggtcc gattaccaa 900
 aaaatttggg atggctattg ggcaatgcat tatgatgcag cctatagctt tgaattgat 960
 tataatgaac gcaattcagg atcctga 987

<210> 5
 <211> 339
 <212> PRT
 <213> Mycobacterium aurum

<400> 5

Met Thr Ala Leu Ser Asp Leu Gly Thr Ser Asn Leu Val Ala Val Glu
 1 5 10 15

Pro Gly Ala Ile Arg Glu Asp Thr Pro Ala Gly Ser Val Ile Gln Tyr
 20 25 30

Ser Asp Tyr Glu Leu Asp Thr Ser Ser Pro Phe Ala Gly Gly Val Ala
 35 40 45

Trp Ile Glu Gly Glu Tyr Leu Pro Ala Glu Glu Ala Lys Ile Ser Ile
 50 55 60

Phe Asp Thr Gly Phe Gly His Ser Asp Leu Thr Tyr Thr Val Ala His
 65 70 75 80

Val Trp His Gly Asn Ile Phe Arg Leu Gly Asp His Leu Asp Arg Leu
 85 90 95

Leu Asp Gly Ala Ser Lys Leu Arg Leu Asp Ala Gly Tyr Ser Lys Asp
 100 105 110

Glu Leu Ala Glu Ile Thr Lys Lys Cys Val Ser Met Ser Gln Leu Arg
 115 120 125

ES 2 599 644 T3

Glu Ser Phe Val Asn Leu Thr Val Thr Arg Gly Tyr Gly Lys Arg Lys
 130 135 140

Gly Glu Lys Asp Leu Ser Lys Leu Thr His Gln Val Tyr Ile Tyr Ala
 145 150 155 160

Ile Pro Tyr Leu Trp Ala Phe Pro Pro Ala Glu Gln Ile Phe Gly Thr
 165 170 175

Thr Ala Ile Val Pro Arg His Val Arg Arg Ala Gly Arg Asn Thr Val
 180 185 190

Asp Pro Thr Ile Lys Asn Tyr Gln Trp Gly Asp Leu Thr Ala Ala Ser
 195 200 205

Phe Glu Ala Lys Asp Arg Gly Ala Arg Thr Ala Ile Leu Leu Asp Ser
 210 215 220

Asp Asn Cys Val Ala Glu Gly Pro Gly Phe Asn Val Cys Ile Val Lys
 225 230 235 240

Asp Gly Lys Leu Ala Ser Pro Ser Arg Asn Ala Leu Pro Gly Ile Thr
 245 250 255

Arg Lys Thr Val Phe Glu Leu Ala Asp Gln Met Gly Ile Glu Ala Thr
 260 265 270

Leu Arg Asp Val Thr Ser Arg Glu Leu Tyr Asp Ala Asp Glu Leu Met
 275 280 285

Ala Val Thr Thr Ala Gly Gly Val Thr Pro Ile Asn Ser Leu Asp Gly
 290 295 300

Glu Ala Val Gly Asn Gly Glu Pro Gly Pro Leu Thr Val Ala Ile Arg
 305 310 315 320

Asp Arg Phe Trp Ala Leu Met Asp Glu Pro Gly Pro Leu Ile Glu Thr
 325 330 335

Ile Glu Tyr

- <210> 6
- <211> 1020
- <212> DNA
- <213> Mycobacterium aurum
- <400> 6

ES 2 599 644 T3

```

atgactgctc tttcagacct cggcacctcc aacctggtgg ccgtcgagcc cggcgccatc      60
cgcgaggaca ccccggcccg ctcggtgatc cagtacagcg actacgaact ggacacctcc      120
agcccgttcg ccggcggcgt cgcctggatc gagggcgaat acctgcocggc cgaagaagcg      180
aagatctcca tcttcgacac cggattcggc cattccgatc tgacctacac cgtcgcgcat      240
gtatggcaag gcaacatctt ccggctcggc gaccacctgg accggttgct cgacggggcg      300
tccaagctgc gcctggagcg cgggtacagc aaggacgaac tggccgagat caccaagaag      360
tgcggtgcca tgtcgcagct gcgcgaatcg ttcgtgaatc tgaccgtcac ccggggatac      420
ggaaagcgca agggcgagaa ggacctgtcc aagctcacc atcaggtgta catctacgcc      480
atcccgtacc tgtgggcctt cccgccgcc gagcagatct tcggcaccac cgcgatcgtg      540
ccgcgccatg tccgccgcgc cggccgcaac accgtcgacc cgaccatcaa gaactaccag      600
tggggtgatc tcaccgcagc cagtttcgaa gccaaaggacc gtggtgcgcg caccgcgatc      660
ctgctcgact cggacaactg cgtggccgaa ggtccgggct tcaacgtgtg catcgtcaag      720
gacggcaagc tggcctcccc gtcccggaac gcggtgcccg gcatcaccgg taagacggtg      780
ttcgaactgg ccgaccagat gggcatcgaa gccaccctgc gcgacgtcac cagccgtgaa      840
ctctacgagc ccgacgagtt gatggcggtc accaccgccc gcggggtcac accgatcaac      900
tcgctggatg gcgaggccgt gggcaacggc gagcccggtc cactgacggt ggccatccgg      960
gaccggttct gggcgctgat ggacgagccg ggcccgtgta tcgaaacgat cgaatactga     1020

```

```

<210> 7
<211> 330
<212> PRT
<213> Arthrobacter

```

```
<400> 7
```

```
Met Ala Phe Ser Ala Asp Thr Ser Glu Ile Val Tyr Thr His Asp Thr
1           5           10           15
```

```
Gly Leu Asp Tyr Ile Thr Tyr Ser Asp Tyr Glu Leu Asp Pro Ala Asn
20           25           30
```

```
Pro Leu Ala Gly Gly Ala Ala Trp Ile Glu Gly Ala Phe Val Pro Pro
35           40           45
```

```
Ser Glu Ala Arg Ile Ser Ile Phe Asp Gln Gly Tyr Leu His Ser Asp
50           55           60
```

```
Val Thr Tyr Thr Val Phe His Val Trp Asn Gly Asn Ala Phe Arg Leu
65           70           75           80
```

```
Asp Asp His Ile Glu Arg Leu Phe Ser Asn Ala Glu Ser Met Arg Ile
```

ES 2 599 644 T3

				85						90					95			
Ile	Pro	Pro	Leu	Thr	Gln	Asp	Glu	Val	Lys	Glu	Ile	Ala	Leu	Glu	Leu			
			100					105					110					
Val	Ala	Lys	Thr	Glu	Leu	Arg	Glu	Ala	Phe	Val	Ser	Val	Ser	Ile	Thr			
		115					120					125						
Arg	Gly	Tyr	Ser	Ser	Thr	Pro	Gly	Glu	Arg	Asp	Ile	Thr	Lys	His	Arg			
	130					135					140							
Pro	Gln	Val	Tyr	Met	Tyr	Ala	Val	Pro	Tyr	Gln	Trp	Ile	Val	Pro	Phe			
145					150					155					160			
Asp	Arg	Ile	Arg	Asp	Gly	Val	His	Ala	Met	Val	Ala	Gln	Ser	Val	Arg			
				165					170					175				
Arg	Thr	Pro	Arg	Ser	Ser	Ile	Asp	Pro	Gln	Val	Lys	Asn	Phe	Gln	Trp			
			180					185					190					
Gly	Asp	Leu	Ile	Arg	Ala	Val	Gln	Glu	Thr	His	Asp	Arg	Gly	Phe	Glu			
		195					200					205						
Ala	Pro	Leu	Leu	Leu	Asp	Gly	Asp	Gly	Leu	Leu	Ala	Glu	Gly	Ser	Gly			
	210					215					220							
Phe	Asn	Val	Val	Val	Ile	Lys	Asp	Gly	Val	Val	Arg	Ser	Pro	Gly	Arg			
225					230					235					240			
Ala	Ala	Leu	Pro	Gly	Ile	Thr	Arg	Lys	Thr	Val	Leu	Glu	Ile	Ala	Glu			
				245					250					255				
Ser	Leu	Gly	His	Glu	Ala	Ile	Leu	Ala	Asp	Ile	Thr	Leu	Ala	Glu	Leu			
			260					265						270				
Leu	Asp	Ala	Asp	Glu	Val	Leu	Gly	Cys	Thr	Thr	Ala	Gly	Gly	Val	Trp			
	275						280					285						
Pro	Phe	Val	Ser	Val	Asp	Gly	Asn	Pro	Ile	Ser	Asp	Gly	Val	Pro	Gly			
	290					295					300							
Pro	Ile	Thr	Gln	Ser	Ile	Ile	Arg	Arg	Tyr	Trp	Glu	Leu	Asn	Val	Glu			
305					310					315					320			
Ser	Ser	Ser	Leu	Leu	Thr	Pro	Val	Gln	Tyr									
			325						330									

ES 2 599 644 T3

<210> 8
 <211> 993
 <212> DNA
 <213> *Arthrobacter* sp.

<400> 8
 atggcattca ggcgagatac ctccgagatc gtctacacgc acgacaccgg cctcgactac 60
 atcaacttata ggcactacga actcgatcct gctaaccocg tcgaggaggg tgcggcatgg 120
 atcgaggggtg cattcgtgcc gccgtcggag gcgaggatct cgatcttoga tcagggttac 180
 ctccactcgg acgtcaccta cacggtcttc cacgtctgga acggaaatgc attccgcctc 240
 gacgaccaca tcgaacgcct cttctccaac gcggagtcga tgcgcatcat cctcccgctc 300
 acacaggacg aagtgaagga gattgcgctc gaactcgtcg cgaagaccga attgctgtgag 360
 gccttcgtgt ccgtgtcgat taccocgggt tacagctcga ctccgggcca gcgcgacatc 420
 acgaagcacc gccgcaggt gtacatgtat gccgtcccat atcagtggat cgtgccggtt 480
 gaccgaattc ggcagcgggt gcacgccatg gtgcacaga gcgtgcgccc aaccccgcgc 540
 agctcgatcg accctcaggt caagaacttc cagtgggggg atctgatccg tgcggttcaa 600
 gagagcgcag accgcgggtt cgaggctccc cttctgctcg acggcgatgg actgcttgcc 660
 gagggctcgg ggttcaacgt cgtcgtgatc aaggacggcg tcgtgcgcag cccgggtcga 720
 gcggcgctcc ccggcattac gcggaagacc gtgctcgaga tcgccgaatc gctcggacac 780
 gaggcgattc togccgacat cacgctcgtc gaactgctcg acgccgacga agtgctcggc 840
 tgcaagactg cgggcggagt gtggccattc gtcagcgtgg acggcaacc cactctggac 900
 ggggttcccg gcccacacac ccagtcgatc atccgtcgtt actgggagct gaatgtcga 960
 agctcgtcgt tgcttacgcc tgtgcagtac tga 993

<210> 9
 <211> 9
 <212> PRT
 <213> *Escherichia coli*

<400> 9
 Thr Ser Ala Tyr Ile Arg Pro Leu Ile
 1 5

<210> 10
 <211> 8
 <212> PRT
 <213> *Escherichia coli*

<400> 10
 Val Phe Glu Gly Ile Arg Cys Tyr
 1 5

ES 2 599 644 T3

<210> 11
 <211> 8
 <212> PRT
 <213> Escherichia coli

<400> 11

Asp Val Gly Met Gly Val Asn Pro
 1 5

<210> 12
 <211> 9
 <212> PRT
 <213> Escherichia coli

<400> 12

Pro Thr Ala Ala Lys Ala Gly Gly Asn
 1 5

<210> 13
 <211> 960
 <212> DNA
 <213> Penicillium chrysogenum

<400> 13
 atggctacaa tggaaaaaat cttcgccgcc taccacgagc gccaaaagct tcttgcagcg 60
 aacaccacc ccttcgcaaa aggtgtcgtc tgggtggagg gagaacttac tcctctccat 120
 gaagcccgtc tccaatcct agaccaaggc ttcattgcaca gcgacttgac atacgatggt 180
 ccctctgtct gggatggagc cttttcccg ctcgatgacc acatcaccgc gttggaagcc 240
 agctgcacca agctacgcat gaaactcccc ctcccacgcg acgaggtgaa gcagattctg 300
 gtcgatatgg ttgcaaagag tggcatccgc gacgcgtttg tcgaaatcat cgtgacgcgt 360
 ggattgaaag gtgtgcgagg ctctcgccct gaggatatcg tcaaccgat ctatatgttt 420
 attcaaccct acgtctggtg tatggaacct gaggtgcagc ctgtgggtgg aagcgcaatt 480
 atcgcaagga ctgtccgccc cgtcccgcct ggctgcatcg acccactgt caagaatctg 540
 caatggggtg atctggttcg cggcctttc gaggcttctg atcgtggcgc cgaatatccc 600
 ttctgaccg atggtgacac caacctcacc gaaggttccg gcttcaacat tgttctcgtg 660
 aaggacaata ttctgtacac tccagctcgc ggagtacttg aaggtgtgac acgcaagagt 720
 gtgattgatg tcgctcgagc cagcggcttt gacattaagg tcgagttggt acctgtccaa 780
 atggcttatg atgcgatga aatttttatg tgtaccactg ctggaggtat catgcccatc 840
 accagtcttg atggcaagcc cgtgaacgac ggaaagggtg ggtctgttac caagaagatc 900
 tgggatgggt actgggctat ccaactatgat cctgcctaca gcttcgagat tgcctattag 960

<210> 14
 <211> 319

ES 2 599 644 T3

<212> PRT

<213> Penicillium chrysogenum

<400> 14

Met Ala Thr Met Glu Lys Ile Phe Ala Ala Tyr His Glu Arg Gln Lys
 1 5 10 15

Leu Leu Ala Ala Asn Thr His Pro Phe Ala Lys Gly Val Ala Trp Val
 20 25 30

Glu Gly Glu Leu Thr Pro Leu His Glu Ala Arg Ile Pro Ile Leu Asp
 35 40 45

Gln Gly Phe Met His Ser Asp Leu Thr Tyr Asp Val Pro Ser Val Trp
 50 55 60

Asp Gly Arg Phe Phe Arg Leu Asp Asp His Ile Thr Arg Leu Glu Ala
 65 70 75 80

Ser Cys Thr Lys Leu Arg Met Lys Leu Pro Leu Pro Arg Asp Glu Val
 85 90 95

Lys Gln Ile Leu Val Asp Met Val Ala Lys Ser Gly Ile Arg Asp Ala
 100 105 110

Phe Val Glu Ile Ile Val Thr Arg Gly Leu Lys Gly Val Arg Gly Ser
 115 120 125

Arg Pro Glu Asp Ile Val Asn Arg Ile Tyr Met Phe Ile Gln Pro Tyr
 130 135 140

Val Trp Cys Met Glu Pro Glu Val Gln Pro Val Gly Gly Ser Ala Ile
 145 150 155 160

Ile Ala Arg Thr Val Arg Arg Val Pro Pro Gly Cys Ile Asp Pro Thr
 165 170 175

Val Lys Asn Leu Gln Trp Gly Asp Leu Val Arg Gly Leu Phe Glu Ala
 180 185 190

Ser Asp Arg Gly Ala Glu Tyr Pro Phe Leu Thr Asp Gly Asp Thr Asn
 195 200 205

Leu Thr Glu Gly Ser Gly Phe Asn Ile Val Leu Val Lys Asp Asn Ile
 210 215 220

Leu Tyr Thr Pro Ala Arg Gly Val Leu Glu Gly Val Thr Arg Lys Ser
 225 230 235 240

ES 2 599 644 T3

Val Ile Asp Val Ala Arg Ala Ser Gly Phe Asp Ile Lys Val Glu Leu
 245 250 255

Val Pro Val Gln Met Ala Tyr Asp Ala Asp Glu Ile Phe Met Cys Thr
 260 265 270

Thr Ala Gly Gly Ile Met Pro Ile Thr Ser Leu Asp Gly Lys Pro Val
 275 280 285

Asn Asp Gly Lys Val Gly Ser Val Thr Lys Lys Ile Trp Asp Gly Tyr
 290 295 300

Trp Ala Ile His Tyr Asp Pro Ala Tyr Ser Phe Glu Ile Ala Tyr
 305 310 315

<210> 15
 <211> 984
 <212> DNA
 <213> *Aspergillus niger*

<400> 15
 atggcatcca tgaaccaagt tcttactgaa tatgccactc gccgcgcgac actagaagcc 60
 agtaaaaacc cctacgcca ggaatcgcc tgggttgaag ggcaactcgt cccctcagg 120
 gaggcccgca tccccctaat tgatcaaggc tttttacgca gtgatttaac ctacgatgtc 180
 atctcogtct gggatggctg gttctttcgc ctagatgacc acctttcccg acttgaattg 240
 gcctgcgcga aatcgcgtct caagttgcc atttcccgcg atgaagtga acaatccctg 300
 gttaggatgg tcgctcaaag tggattcga gatgcatatg tggctttgat tgtgacgcgg 360
 ggattgcaga gtgttcgagg tgccaagccg gaggacttgg tgaacaacct gtacatgttt 420
 gtacaacct atgtatgggt aatggaacca gaggtccaac gggtcggtgg aagtgtgtt 480
 gttactcgaa ctgttcgtcg ggtgccccca ggagctatct atcctactgt aaagaacct 540
 caatggggtg acctgaccgg aggtatgctc gaggccgccc atcgaggctc catgtacccg 600
 ttctgacgg atggagatgg ccatctcacc gaggggtccg gatacaatat cgttctaate 660
 aaggccggtg ccatttatac gcctgatcgc ggogtgcgac atggtgtcac caggacaagt 720
 gtcattgatg ttgcaagagc ttgtggtatc caagttcacc tcgaagctgt gccggtggag 780
 ttggtatc agtgtgatga gatattcatg tgcacaacag caggtggaat catgcccac 840
 actgagctgg atggcaagcc tgtaaatggg gggcggattg gtccgatcac gaagaagac 900
 tgggaocgggt attgggtat gcattatgat ccagcctaca gcttcgcagt tagttatgat 960
 gacgatcaa aagcaaagct ctga 984

ES 2 599 644 T3

<210> 16
 <211> 327
 <212> PRT
 <213> *Aspergillus niger*

<400> 16

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

ES 2 599 644 T3

Ile Tyr Thr Pro Asp Arg Gly Val Leu His Gly Val Thr Arg Thr Ser
225 230 235 240

Val Ile Asp Val Ala Arg Ala Cys Gly Ile Gln Val His Leu Glu Ala
245 250 255

Val Pro Val Glu Leu Val Tyr Gln Cys Asp Glu Ile Phe Met Cys Thr
260 265 270

Thr Ala Gly Gly Ile Met Pro Ile Thr Glu Leu Asp Gly Lys Pro Val
275 280 285

Asn Gly Gly Arg Ile Gly Pro Ile Thr Lys Lys Ile Trp Asp Gly Tyr
290 295 300

Trp Gly Met His Tyr Asp Pro Ala Tyr Ser Phe Ala Val Ser Tyr Asp
305 310 315 320

Asp Gly Ser Lys Ala Lys Leu
325

<210> 17
<211> 981
<212> DNA
<213> *Aspergillus oryzae*

<400> 17
atgacatcta tgaacaaagt attttccggt tactacgagc gcaaggctcg tctagataac 60
agtgacaacc gctttgcgaa aggaattgcc tacgtccagg gatctttcgt cccactcgcc 120
gacgcaagag tcccactcct cgacgagggt ttcattgcata gcgacctcac gtacgatgtg 180
ccatcggtct gggatgggag ctttttccgc cttgatgatc atctcagtcg attggaagat 240
agttgtgaaa agatgogact gaagatccca ctgtccaggg acgaagtcaa gcaaacccta 300
agggagatgg ttgctaagag tggaatcgaa gatgcctttg tggagctgat cgtcactcgt 360
ggcctgaaag gggctccgtg caataagcca gaggatcttt tcgacaatca tctctatctg 420
atcgtcatgc cgtatgtctg ggtgatggag ccgcccattc aacataaccg aggtactgag 480
atcattgccc gtacagtaag gcgcaactcc ccgggtgctt togatcctac catcaagaat 540
ctccagtggg gggacttgac acggggtcta tttgaagcgg ctgaccgtgg cgcggattac 600
ccatttctct cagatggaga taccaatctc acagaaggat ccggtttcaa tatagtgttg 660
gttaaagatg gtattatcta cagccccgac cgtgggtgtc tggaaggcat tacacgtaag 720
agtgtttttg atattgccc ggtcaagaac atcgaggctc gcgttcaggt ggtgccactc 780
gaacatgcct atcaagccga tgagatattc atgtgtacta ctgctgggtg cattatgcct 840
atcacgaaac tcgatgggaa accgatccgg aatggagaag tcggtcccct tactacaaag 900

ES 2 599 644 T3

atatgggatg agtactgggc gatgcactat gaccgaaat atagctctgc tatcgattac 960

aggggccatg aggtaactg a 981

<210> 18
 <211> 326
 <212> PRT
 <213> *Aspergillus oryzae*
 <400> 18

Met Thr Ser Met Asn Lys Val Phe Ser Gly Tyr Tyr Glu Arg Lys Ala
 1 5 10 15

Arg Leu Asp Asn Ser Asp Asn Arg Phe Ala Lys Gly Ile Ala Tyr Val
 20 25 30

Gln Gly Ser Phe Val Pro Leu Ala Asp Ala Arg Val Pro Leu Leu Asp
 35 40 45

Glu Gly Phe Met His Ser Asp Leu Thr Tyr Asp Val Pro Ser Val Trp
 50 55 60

Asp Gly Arg Phe Phe Arg Leu Asp Asp His Leu Ser Arg Leu Glu Asp
 65 70 75 80

Ser Cys Glu Lys Met Arg Leu Lys Ile Pro Leu Ser Arg Asp Glu Val
 85 90 95

Lys Gln Thr Leu Arg Glu Met Val Ala Lys Ser Gly Ile Glu Asp Ala
 100 105 110

Phe Val Glu Leu Ile Val Thr Arg Gly Leu Lys Gly Val Arg Gly Asn
 115 120 125

Lys Pro Glu Asp Leu Phe Asp Asn His Leu Tyr Leu Ile Val Met Pro
 130 135 140

Tyr Val Trp Val Met Glu Pro Ala Ile Gln His Thr Gly Gly Thr Ala
 145 150 155 160

Ile Ile Ala Arg Thr Val Arg Arg Thr Pro Pro Gly Ala Phe Asp Pro
 165 170 175

Thr Ile Lys Asn Leu Gln Trp Gly Asp Leu Thr Arg Gly Leu Phe Glu
 180 185 190

Ala Ala Asp Arg Gly Ala Asp Tyr Pro Phe Leu Ser Asp Gly Asp Thr
 195 200 205

ES 2 599 644 T3

Asn Leu Thr Glu Gly Ser Gly Phe Asn Ile Val Leu Val Lys Asp Gly
 210 215 220

Ile Ile Tyr Thr Pro Asp Arg Gly Val Leu Glu Gly Ile Thr Arg Lys
 225 230 235 240

Ser Val Phe Asp Ile Ala Gln Val Lys Asn Ile Glu Val Arg Val Gln
 245 250 255

Val Val Pro Leu Glu His Ala Tyr His Ala Asp Glu Ile Phe Met Cys
 260 265 270

Thr Thr Ala Gly Gly Ile Met Pro Ile Thr Lys Leu Asp Gly Lys Pro
 275 280 285

Ile Arg Asn Gly Glu Val Gly Pro Leu Thr Thr Lys Ile Trp Asp Glu
 290 295 300

Tyr Trp Ala Met His Tyr Asp Pro Lys Tyr Ser Ser Ala Ile Asp Tyr
 305 310 315 320

Arg Gly His Glu Gly Asn
 325

<210> 19
 <211> 972
 <212> DNA
 <213> *Aspergillus fumigatus*

<400> 19
 atggcctcta tggacaaagt cttttcggga tattatgocg gccagaagct gcttgaacgg 60
 agcgacaatc ctttctctaa gggcattgct tatgtggaag gaaagctcgt cttacctagt 120
 gatgctagaa taccgctact cgacgaaggt ttcatgcaca gtgacctaac ctatgatggt 180
 atatcggttt gggatggtcg cttctttcga ttggacgacg atttgcaacg gattttggaa 240
 agctgcgata agatgcccgt caagttccca cttgcaactga gctcagtgaa aaatattctg 300
 gctgagatgg tcgccaagag tggatccgg gatgocgttg tggaagttat tgtgacacgt 360
 ggtctgacag gtgtacgtgg ttcgaagcct gaggatctgt ataatacaa catatacctg 420
 cttgttcttc catacatttg ggttatggcg cctgagaacc agctccatgg tggcgaggct 480
 atcattacaa ggacagtgcg acgaacaccc ccaggtgcat ttgatcctac tatcaaaaat 540
 ctacagtggg gtgatttaac aaagggactt tttgaggcaa tggaccgtgg cgccacatac 600
 ccatttctca ctgatggaga caccaacctt actgaaggat ctggtttcaa cattgttttg 660
 gtgaagaacg gtattatcta taccctgat cgaggtgtct tgcgagggat cacacgtaaa 720

ES 2 599 644 T3

agtgtgattg acgttgcccg agccaacagc atcgacatcc gccttgaggt cgtaccagtg 780
gagcaggctt atcactctga tgagatcttc atgtgcacaa ctgccggcgg cattatgcct 840
ataacattgc ttgatggtca acctgttaat gacggccagc ttggcccaat cacaaagaag 900
atatgggatg gctattggga gatgcactac aatccggcgt atagttttcc tgttgactat 960
ggcagtggct aa 972

<210> 20
<211> 323
<212> PRT
<213> *Aspergillus fumigatus*

<400> 20

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

ES 2 599 644 T3

Thr Ile Lys Asn Leu Gln Trp Gly Asp Leu Thr Lys Gly Leu Phe Glu
 180 185 190

Ala Met Asp Arg Gly Ala Thr Tyr Pro Phe Leu Thr Asp Gly Asp Thr
 195 200 205

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

Ile Ile Tyr Thr Pro Asp Arg Gly Val Leu Arg Gly Ile Thr Arg Lys
 225 230 235 240

Ser Val Ile Asp Val Ala Arg Ala Asn Ser Ile Asp Ile Arg Leu Glu
 245 250 255

Val Val Pro Val Glu Gln Ala Tyr His Ser Asp Glu Ile Phe Met Cys
 260 265 270

Thr Thr Ala Gly Gly Ile Met Pro Ile Thr Leu Leu Asp Gly Gln Pro
 275 280 285

Val Asn Asp Gly Gln Val Gly Pro Ile Thr Lys Lys Ile Trp Asp Gly
 290 295 300

Tyr Trp Glu Met His Tyr Asn Pro Ala Tyr Ser Phe Pro Val Asp Tyr
 305 310 315 320

Gly Ser Gly

<210> 21
 <211> 972
 <212> DNA
 <213> *Neosartorya fischeri*

<400> 21
 atggcctcta tggacaaagt cttttcggga tatcatgogc gccagaagct gcttgaacgg 60
 agcgacaatc ctttctctaa gggcattgcc tatgtggaag gaaagctcgt cttaccacgc 120
 gacgccagaa taccgctact tgacgaaggc ttcattgcaag gtgacctaac ttatgatggt 180
 acaacggttt gggatggacg cttcttttoga ttggatgata atatgcaacg gatcctggaa 240
 agctgcgata aatgctggct caagttccca cttgcaccga gcacggtgaa aatatacctg 300
 gctgagatgg tcgccaagag tggattcgg gatgcgtttg tggagttat cgtgacacgt 360
 ggtctgacag gtgtacgtgg ttogaagccc gaggatctgt ataataacaa catatacctg 420
 cttgttctcc catacgtttg gggtatggcg cctgagaacc agctccttgg tggcagtgct 480
 atcattacaa ggacagtgcg acgaacaccc cgggtgcat ttgatcctac gatcaaaaat 540

ES 2 599 644 T3

ctacagtggg gtgacttaac aaagggactt tttgaggcaa tggaccgtgg cgcaacgtac 600
 ccattttctca ctgacggaga caccaacctt accgaaggat ctggatttaa cattgtcttg 660
 gtgaagaacg gtattatcta taccctgat cgaggtgtct tgcgagggat cacacgtaaa 720
 agtgtgattg acgttgcccg agccaacaac atcgacatcc gccttgaggt cgtaccagtg 780
 gagcaggttt atcaactccga tgaatcttc atgtgcacaa cagccggtgg cattatgcct 840
 ataacgttgc ttgatgtca accagttaat gacggccagg ttggcccgat cacaaagaag 900
 atatgggatg gttactggga gatgcactac aatccggcgt atagttttcc ggtcgactat 960
 ggcagtggct aa 972

<210> 22
 <211> 323
 <212> PRT
 <213> *Neosartorya fischeri*

<400> 22

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

ES 2 599 644 T3

Ile Ile Thr Arg Thr Val Arg Arg Thr Pro Pro Gly Ala Phe Asp Pro
165 170 175

Thr Ile Lys Asn Leu Gln Trp Gly Asp Leu Thr Lys Gly Leu Phe Glu
180 185 190

Ala Met Asp Arg Gly Ala Thr Tyr Pro Phe Leu Thr Asp Gly Asp Thr
195 200 205

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

Ile Ile Tyr Thr Pro Asp Arg Gly Val Leu Arg Gly Ile Thr Arg Lys
225 230 235 240

Ser Val Ile Asp Val Ala Arg Ala Asn Asn Ile Asp Ile Arg Leu Glu
245 250 255

Val Val Pro Val Glu Gln Val Tyr His Ser Asp Glu Ile Phe Met Cys
260 265 270

Thr Thr Ala Gly Gly Ile Met Pro Ile Thr Leu Leu Asp Gly Gln Pro
275 280 285

Val Asn Asp Gly Gln Val Gly Pro Ile Thr Lys Lys Ile Trp Asp Gly
290 295 300

Tyr Trp Glu Met His Tyr Asn Pro Ala Tyr Ser Phe Pro Val Asp Tyr
305 310 315 320

Gly Ser Gly

<210> 23

<211> 978

<212> DNA

<213> Gibberella zeae

<400> 23

atgtcgacca tggacaagat cttcgccggc cagcccagc gccaaagccac cctcgtcgca 60

agcgacaaca tcttcgcaa cggcattgcc tggatccaag gcgagctcgt cccctcaat 120

gaagcccgca tccccctcat ggaccaaggt ttcattgcac gcgacttgac ctacgatgac 180

cctgcagtct gggatggtcg tttcttcogt cttgatgacc atctcgaccg tctcgaggca 240

agcgtcaaga agatgogaat gcaattcccc attccccogc atgagatcag aatgactott 300

ctcgacatgc tcgccaagag tggaatcaag gatgcttttg ttgagctcat tgtcactcgt 360

ES 2 599 644 T3

ggcttgaagc ctgttcgtga ggccaagcct ggtgaggtct tgaacaacca cctctacttg 420
 atcgtccaac cctacgtctg ggtcatgagc cccgaagctc agtacgtcgg cggtaatgcc 480
 gttatcgcac gaactgttcg tcgaatccct cctggatcca tggatcccac catcaagaac 540
 ctccaatgga gtgatttcac ccgcggcatg ttogaagcat acgatcgtgg agcacaatac 600
 cccttcctca ccgacggcga cacaaacatc accgaaggat ctggtttcaa cgttgtcttt 660
 gtcaagaaca acgttattta caccocgaac cgaggagttt tgcagggaat taccagaaag 720
 agtgtgatcg acgctgccaa gtggtgtggt catgaagttc gagtggagta tgtccctggt 780
 gagatggcct atgaagctga tgagatcttc atgtgtacta ctgctggagg aatcatgcct 840
 atcaccacca tggatggaaa gccagtcaag gacggaaagg tcgggcctgt cacaaggcc 900
 atctgggatc ggtactgggc gatgcactgg gaggatgagt tcagtttcaa gattgactac 960
 cagaaactga agctgtag 978

<210> 24
 <211> 325
 <212> PRT
 <213> *Gibberella zeae*

<400> 24

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

ES 2 599 644 T3

Lys Pro Gly Glu Val Leu Asn Asn His Leu Tyr Leu Ile Val Gln Pro
 130 135 140

Tyr Val Trp Val Met Ser Pro Glu Ala Gln Tyr Val Gly Gly Asn Ala
 145 150 155 160

Val Ile Ala Arg Thr Val Arg Arg Ile Pro Pro Gly Ser Met Asp Pro
 165 170 175

Thr Ile Lys Asn Leu Gln Trp Ser Asp Phe Thr Arg Gly Met Phe Glu
 180 185 190

Ala Tyr Asp Arg Gly Ala Gln Tyr Pro Phe Leu Thr Asp Gly Asp Thr
 195 200 205

Asn Ile Thr Glu Gly Ser Gly Phe Asn Val Val Phe Val Lys Asn Asn
 210 215 220

Val Ile Tyr Thr Pro Asn Arg Gly Val Leu Gln Gly Ile Thr Arg Lys
 225 230 235 240

Ser Val Ile Asp Ala Ala Lys Trp Cys Gly His Glu Val Arg Val Glu
 245 250 255

Tyr Val Pro Val Glu Met Ala Tyr Glu Ala Asp Glu Ile Phe Met Cys
 260 265 270

Thr Thr Ala Gly Gly Ile Met Pro Ile Thr Thr Met Asp Gly Lys Pro
 275 280 285

Val Lys Asp Gly Lys Val Gly Pro Val Thr Lys Ala Ile Trp Asp Arg
 290 295 300

Tyr Trp Ala Met His Trp Glu Asp Glu Phe Ser Phe Lys Ile Asp Tyr
 305 310 315 320

Gln Lys Leu Lys Leu
 325

<210> 25

<211> 966

<212> DNA

<213> Hyphomonas neptunium

<400> 25

atgctgacat ttcaaaaagt gctgaccogga tttcagacgc gggcagacgc togcgcagag 60

cgaacagacg cgttcgcca cggtattgcc tggatcgaaa acgaatttgt goccacogga 120

aaggcgcgga taccgatact tgatcagggc ttctcgcatt cggatctgac ctatgacgtc 180

ES 2 599 644 T3

ccgcagctct ggaatggccg gatcttccgt ctogatgata acctggaccg tctggaagtt 240
 agctgcgcca aaatgctggc cccgctgccc atcgccaggc cagagctccg caggctggtg 300
 atggagctgg tctccaggag tgggctgccc gacgcatatg ttgagatcat cgtgaccgcc 360
 gggttgaagt tctgctgagg cgcacaagcc gaagacatca tcccgaacct gtatctcatg 420
 gcggtgcctt acgtctggat tctcccgtt gaataccaga accatggggc gccggccgtc 480
 gtaacgcgaa cggctgcccg cacacccccg ggggccctgg acccgacaat caagaacctt 540
 cagtggggcg atctctgccc agggcttatg gaagcaggtg accgggactc tttcttcccc 600
 attcttcccc acggcgacgg gaacgccaca gaaggcgtg gctataatat cgtccttgtc 660
 agaatgggg aacttcacac acccctgccc ggggtgctgg aaggcatcac acgcaggaca 720
 gtactggaaa tagcggcccg gcgcccctg aaaacacatg ttaccgagat accatccag 780
 gcactctatg aatgcatga attgttcatg tgttcgacgg caggcggcat aatgccactc 840
 gtccttctgg atgggaacat tgtgggtgat ggaacagtcg gtccggttac acggatgatt 900
 tgggaggctt actgggatct tcatgatgat cctcagctca gcgagcctgt gacttacgcg 960
 ccctga 966

<210> 26
 <211> 321
 <212> PRT
 <213> Hyphonmonas neptunium
 <400> 26

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

ES 2 599 644 T3

Tyr Val Glu Ile Ile Val Thr Arg Gly Leu Lys Phe Leu Arg Gly Ala
115 120 125

Gln Ala Glu Asp Ile Ile Pro Asn Leu Tyr Leu Met Ala Val Pro Tyr
130 135 140

Val Trp Ile Leu Pro Leu Glu Tyr Gln Asn His Gly Ala Pro Ala Val
145 150 155 160

Val Thr Arg Thr Val Arg Arg Thr Pro Pro Gly Ala Leu Asp Pro Thr
165 170 175

Ile Lys Asn Leu Gln Trp Gly Asp Leu Val Arg Gly Leu Met Glu Ala
180 185 190

Gly Asp Arg Asp Ser Phe Phe Pro Ile Leu Pro Asp Gly Asp Gly Asn
195 200 205

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

Leu His Thr Pro Arg Arg Gly Val Leu Glu Gly Ile Thr Arg Arg Thr
225 230 235 240

Val Leu Glu Ile Ala Ala Ala Arg Gly Leu Lys Thr His Val Thr Glu
245 250 255

Ile Pro Ile Gln Ala Leu Tyr Glu Cys Asp Glu Leu Phe Met Cys Ser
260 265 270

Thr Ala Gly Gly Ile Met Pro Leu Val Leu Leu Asp Gly Asn Ile Val
275 280 285

Gly Asp Gly Thr Val Gly Pro Val Thr Arg Met Ile Trp Glu Ala Tyr
290 295 300

Trp Asp Leu His Asp Asp Pro Gln Leu Ser Glu Pro Val Thr Tyr Ala
305 310 315 320

Pro

<210> 27
<211> 963
<212> DNA
<213> Mesorhizobium loti MAFF303099

ES 2 599 644 T3

```

<400> 27
atggaccaga cgacggcaac acaggcttca aagccgctgc ccacagtcgg cgaccgccac      60
gtcgaccgcg attcctaccc cgacggcatc gccttcctcg acggccagta tctgccgatg      120
tcgcaagcca aggtgtcggg gctggactgg ggcttcctgc attccgacgc cacctacgac      180
acggtgcatg tctggaacgg ccgcttcttc cgcctcgacc tgcattctga ccgcttcttc      240
ggcggactgg aaaagctgcg catgaccatc cccttcgaca gggatggcgt ggcogagatc      300
ctgcacaatt gtgtgcacct ttcgggccat cgcgcgcgct atgtcgaaat gctgtgcaag      360
cgcggcgcag cgcggacctt cagccgcgat ccgcccagc cgatcaaccg cttcatggcc      420
ttcgcggtac ccttcggctc ggtcgccaat gccgagcagt tgcagcgcgg cctgcgcgtc      480
gccatcagcg acaaggtgcg catcccgcgg gcttcggctg atccgtgat caagaactac      540
cattggctcg atctgggtcg cggcctctac gacgcctatg acagcgggtc ggagaccgcg      600
ctcattctcg acttcaacgg caatgtgcc gagggcccgg gcttcaacgt cttctgcgtc      660
aaggacggca aactgtcgac gccggccatc ggcgtgctgc ccggcatcac gcgccgcaca      720
gtcttcgatc tctgcgccga agaaggtctt gccgccgccg ccgccgatgt cagcgtcgcc      780
ggcctcaagg cggccgacga ggtcttcata acctcgaccg ccgggggcat catgccggtg      840
accgagatcg acggcgcggc gatcgcggac ggcaaggtcg ggccggttac cagccggcta      900
atggcgctct actggcaaaa gcacgacgat ccggcctggt cgtctcaggt gaagtatccc      960
tga
    
```

```

<210> 28
<211> 320
<212> PRT
<213> Mesorhizobium loti MAFF303099
    
```

```

<400> 28
Met Asp Gln Thr Thr Ala Thr Gln Ala Ser Lys Pro Leu Pro Thr Val
1          5          10          15

Gly Asp Arg His Val Asp Pro His Ser Tyr Pro Asp Gly Ile Ala Phe
          20          25          30

Leu Asp Gly Gln Tyr Leu Pro Met Ser Gln Ala Lys Val Ser Val Leu
          35          40          45

Asp Trp Gly Phe Leu His Ser Asp Ala Thr Tyr Asp Thr Val His Val
          50          55          60

Trp Asn Gly Arg Phe Phe Arg Leu Asp Leu His Leu Asp Arg Phe Phe
65          70          75          80
    
```

ES 2 599 644 T3

Gly Gly Leu Glu Lys Leu Arg Met Thr Ile Pro Phe Asp Arg Asp Gly
85 90 95

Val Ala Glu Ile Leu His Asn Cys Val Ala Leu Ser Gly His Arg Ala
100 105 110

Ala Tyr Val Glu Met Leu Cys Thr Arg Gly Ala Ser Pro Thr Phe Ser
115 120 125

Arg Asp Pro Arg Gln Ala Ile Asn Arg Phe Met Ala Phe Ala Val Pro
130 135 140

Phe Gly Ser Val Ala Asn Ala Glu Gln Leu Gln Arg Gly Leu Arg Val
145 150 155 160

Ala Ile Ser Asp Lys Val Arg Ile Pro Pro Ala Ser Val Asp Pro Ser
165 170 175

Ile Lys Asn Tyr His Trp Leu Asp Leu Val Arg Gly Leu Tyr Asp Ala
180 185 190

Tyr Asp Ser Gly Ala Glu Thr Ala Leu Ile Leu Asp Phe Asn Gly Asn
195 200 205

Val Ala Glu Gly Pro Gly Phe Asn Val Phe Cys Val Lys Asp Gly Lys
210 215 220

Leu Ser Thr Pro Ala Ile Gly Val Leu Pro Gly Ile Thr Arg Arg Thr
225 230 235 240

Val Phe Asp Leu Cys Ala Glu Glu Gly Leu Ala Ala Ala Ala Ala Asp
245 250 255

Val Ser Val Ala Ala Leu Lys Ala Ala Asp Glu Val Phe Ile Thr Ser
260 265 270

Thr Ala Gly Gly Ile Met Pro Val Thr Glu Ile Asp Gly Ala Ala Ile
275 280 285

Ala Asp Gly Lys Val Gly Pro Val Thr Ser Arg Leu Met Ala Leu Tyr
290 295 300

Trp Gln Lys His Asp Asp Pro Ala Trp Ser Ser Gln Val Lys Tyr Pro
305 310 315 320

<210> 29
<211> 936
<212> DNA

ES 2 599 644 T3

<213> Roseobacter sp.

<400> 29
tcaaggatat cggatcgctt tccgaaaacg atcatcttca tgcacctccc aatacgttcc 60
agtgatctcc ttggttatcg ggccgaccaa tccgcttcca attgaagttt catcgatctt 120
ggtgacgggc atgacgccgc cagcgggtga ggtcacgaac acttcatccg ccgccctcaa 180
ttcacttggt gccagatcat ccgtcgcgca ggtgatctgt aattcgtcac aaagatcaaa 240
gattgtttgg cgtgttattc ccatcagcac cccaaacttg ggcgaggaaa ttttgccgtt 300
tttcacggcg aagacgttga aaccgggccc ttccgctatg tttccggtgg cgtctaaaag 360
aatggcgggt tccgctccct gcgcatacgc cgcatacaaa cctttgacca ggtccagcca 420
gtgatagttc tttacggttg gatcaacgga gctgggtgga atgcccgaaa ggtcgggtgac 480
ggctgcgtgt aagccacgtc tcatctgttc tgggtttgca accgacccaa atggaattgc 540
gaaggcgatg aagcggttga cggcatcgcg agggtcgagg ctaaacttag gagaagtgcc 600
gcgtgtgcag atgaactcga catacgcctc ctggagacct gatagagcaa cgcaattgtg 660
caaaatctca gtcactctcg cagcaccaaa cgggatcgcg atgtgcaact tttccatccc 720
tgogaagaac cgtcaagat gatcatcaag cctgaaaaat gctcccttcc acacatgcgc 780
aacatcgtat gtcgcatctg agtgaggaa gccattatcc aacaccgaaa tctttgcctc 840
gctgatcgga agatattggc catccatgaa ggcaatgccg ggagggtacg ttgccgggtc 900
agaatgacga tcagacagcg caggcaattt atacat 936

<210> 30

<211> 311

<212> PRT

<213> Roseobacter sp.

<400> 30

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

ES 2 599 644 T3

Ile Pro Phe Gly Arg Ala Glu Met Thr Glu Ile Leu His Asn Cys Val
85 90 95

Ala Leu Ser Gly Leu Gln Asp Ala Tyr Val Glu Phe Ile Cys Thr Arg
100 105 110

Gly Thr Ser Pro Thr Phe Ser Arg Asp Pro Arg Asp Ala Val Asn Arg
115 120 125

Phe Ile Ala Phe Ala Ile Pro Phe Gly Ser Val Ala Asn Pro Glu Gln
130 135 140

Met Arg Arg Gly Leu His Ala Ala Val Thr Asp Leu Val Arg Ile Pro
145 150 155 160

Pro Ser Ser Val Asp Pro Thr Val Lys Asn Tyr His Trp Leu Asp Leu
165 170 175

Val Lys Gly Leu Tyr Ala Ala Tyr Ala Gln Gly Ala Glu Thr Ala Ile
180 185 190

Leu Leu Asp Ala Thr Gly Asn Ile Ala Glu Gly Pro Gly Phe Asn Val
195 200 205

Phe Ala Val Lys Asn Gly Lys Ile Ser Ser Pro Lys Phe Gly Val Leu
210 215 220

Met Gly Ile Thr Arg Gln Thr Ile Phe Asp Leu Cys Asp Glu Leu Gln
225 230 235 240

Ile Thr Cys Ala Thr Asp Asp Leu Ala Pro Ser Glu Leu Arg Ala Ala
245 250 255

Asp Glu Val Phe Val Thr Ser Thr Ala Gly Gly Val Met Pro Val Thr
260 265 270

Lys Ile Asp Glu Thr Ser Ile Gly Ser Gly Leu Val Gly Pro Ile Thr
275 280 285

Lys Glu Ile Thr Glu Ala Tyr Trp Arg Met His Glu Asp Asp Arg Phe
290 295 300

Arg Lys Ala Ile Arg Tyr Pro
305 310

<210> 31
<211> 945
<212> DNA

ES 2 599 644 T3

<213> *Marinomonas* sp.

<400> 31
ctatatatttt gccgggggat aatgaatggg taggcaccaa tcagagtttt tgtgcttttc 60
ccaatatgca gattttaaca attggaagat aggtccgggt gtgccacttc caatcacatt 120
atggtcgatt ttggtaatgg gcataatgcc acccgccggt gacgtaatga atacttcgtc 180
agcattcctt aatcggtag gcgttacctc agttgcacag cagcttatgt gtaattcgtc 240
acaaagatca aagacgggtc ggcgagttat gcctggcaaa accccttttt ctggtggtgt 300
aattatcttt cccttaacac aaaaaacatt aaagccaggg ccttcggcga tattaccctc 360
tgaatccaac aggatggcgg tctcaccacc tttttcatag gcacgtata acccggttac 420
caaatcaagc caatggtaat ttttgacctt tgaatcaacc gagtttggcg gaatacggac 480
aatagagcta ataatggcat gcagtcocatt tttcatttga tcttgatttg cgaccgagcc 540
aaagggaaacg gcgaaagcca taaaacgatt gattgagtct cttgggtccc ggctgaaatc 600
cggcgaattc cctcttgtgc atatcatttc gacataggcg tttttatgac ccgataaggc 660
cacgcaatta tgcaaaattt cagcgacctc ttctttggag taaggcattg tcatatgaat 720
tttctctaac cctgagaaaa aacgttcaag gtacaaatcc aacctgaaaa atgcaccttg 780
ccagacatgg acaacatcat aggttgcgtc agaatgcaaa aagccataat caagaatgga 840
tagttttgct tttgacatat ctaaattattg accatccatg taggcgacgc cttttggata 900
attatgtggg tccttatatt cattggacaa ttgcagtaaa gccat 945

<210> 32

<211> 314

<212> PRT

<213> *Marinomonas* sp.

<400> 32

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

ES 2 599 644 T3

Pro Tyr Ser Lys Glu Glu Val Ala Glu Ile Leu His Asn Cys Val Ala
85 90 95

Leu Ser Gly His Lys Asn Ala Tyr Val Glu Met Ile Cys Thr Arg Gly
100 105 110

Asn Ser Pro Asp Phe Ser Arg Asp Pro Arg Asp Ser Ile Asn Arg Phe
115 120 125

Met Ala Phe Ala Val Pro Phe Gly Ser Val Ala Asn Gln Asp Gln Met
130 135 140

Lys Asn Gly Leu His Ala Ile Ile Ser Ser Ile Val Arg Ile Pro Pro
145 150 155 160

Asn Ser Val Asp Ser Lys Val Lys Asn Tyr His Trp Leu Asp Leu Val
165 170 175

Thr Gly Leu Tyr Asp Ala Tyr Glu Lys Gly Gly Glu Thr Ala Ile Leu
180 185 190

Leu Asp Ser Glu Gly Asn Ile Ala Glu Gly Pro Gly Phe Asn Val Phe
195 200 205

Cys Val Lys Gly Lys Ile Ile Thr Thr Pro Glu Lys Gly Val Leu Pro
210 215 220

Gly Ile Thr Arg Arg Thr Val Phe Asp Leu Cys Asp Glu Leu His Ile
225 230 235 240

Ser Cys Cys Ala Thr Glu Val Thr Pro Thr Asp Leu Arg Asn Ala Asp
245 250 255

Glu Val Phe Ile Thr Ser Thr Ala Gly Gly Ile Met Pro Ile Thr Lys
260 265 270

Ile Asp His Asn Val Ile Gly Ser Gly Thr Thr Gly Pro Ile Phe Gln
275 280 285

Leu Leu Lys Ser Ala Tyr Trp Glu Lys His Lys Asn Ser Asp Trp Cys
290 295 300

Leu Pro Ile His Tyr Pro Pro Ala Lys Ile
305 310

<210> 33
<211> 939
<212> DNA

ES 2 599 644 T3

<213> Rhizobium etli

<400> 33

```

gtgctcgatc gggcgagga atccgccttt cagaaaggct cgcctatgt cgatggaaaa      60
attataccaa tcgaagaggc gaaggtacca cttttagact ggggctttct gcgttcggac      120
gcgtgtcagg acacagtgtc ggtgtgggat ggcagctttt tccgtttgac tgaccacctg      180
gatcgatttg agcgatctgt ccagcgcctt cggatggata cagctcctgt tacaagaagc      240
gatatccacc ggatagtcca caaactcgtt gccgtgtgcg gtttccgtga tgcttacgtt      300
caaatcatca tgaccgcggy ggggcccccg atcggtagtc gggatcttcg tttatgctcg      360
aacagcttcc aggcattctg cgtgccgtac atgtggatcg cgaacccoga aaagcaggaa      420
attggcatgg cagtgcacgt cagccggcgc gtcogaatcc cacctcagtc cgtagatccg      480
ttggtgaagc attaccattg gctcgacttt gagatgggct tgttcgaagc ttacgagaac      540
ggtgccgata cgggtggttct cactgatctc gatggcaata tcacagaggg gcccggttc      600
aacgttttctg cgggtgatcga cggagtctctg aggaccctt ctttcggcat gctggatgga      660
atgaccagga ggacagttat ggaactgtgc aacgagctga atctcgatgt cacgcaagag      720
accatttctt tggagcggct tctaatcgcc tcggaaatct tccttacgac gaccgcaggc      780
gggattatcc cggctctcttc ggtgaacgga accggaatcg gtttcggatc agtgggtgaa      840
caaactcgcc gcattcaccg gtcatactgg gataagcgaa gcagcggatg gtatggcgag      900
cctgtcgcct acgccccaaa agctctcttg gagccctag      939

```

<210> 34

<211> 312

<212> PRT

<213> Rhizobium etli

<400> 34

```

Met Leu Asp Arg Gly Glu Glu Ser Ala Phe Gln Lys Gly Ser Ala Tyr
1           5           10          15

Val Asp Gly Lys Ile Ile Pro Ile Glu Glu Ala Lys Val Pro Leu Leu
20          25          30

Asp Trp Gly Phe Leu Arg Ser Asp Ala Cys Gln Asp Thr Val Ser Val
35          40          45

Trp Asp Gly Ser Phe Phe Arg Leu Thr Asp His Leu Asp Arg Phe Glu
50          55          60

Arg Ser Val Gln Arg Leu Arg Met Asp Thr Ala Pro Val Thr Arg Ser
65          70          75          80

```

ES 2 599 644 T3

Asp Ile His Arg Ile Val His Lys Leu Val Ala Val Cys Gly Phe Arg
85 90 95

Asp Ala Tyr Val Gln Ile Ile Met Thr Arg Gly Arg Pro Pro Ile Gly
100 105 110

Ser Arg Asp Leu Arg Leu Cys Ser Asn Ser Phe Gln Ala Phe Cys Val
115 120 125

Pro Tyr Met Trp Ile Ala Asn Pro Glu Lys Gln Glu Ile Gly Met Ala
130 135 140

Val His Val Ser Arg Arg Val Arg Ile Pro Pro Gln Ser Val Asp Pro
145 150 155 160

Leu Val Lys His Tyr His Trp Leu Asp Phe Glu Met Gly Leu Phe Glu
165 170 175

Ala Tyr Glu Asn Gly Ala Asp Thr Val Val Leu Thr Asp Leu Asp Gly
180 185 190

Asn Ile Thr Glu Gly Pro Gly Phe Asn Val Phe Ala Val Ile Asp Gly
195 200 205

Val Leu Arg Thr Pro Ser Phe Gly Met Leu Asp Gly Met Thr Arg Arg
210 215 220

Thr Val Met Glu Leu Cys Asn Glu Leu Asn Leu Asp Val Thr Gln Glu
225 230 235 240

Thr Ile Ser Leu Glu Arg Leu Leu Ile Ala Ser Glu Ile Phe Leu Thr
245 250 255

Thr Thr Ala Gly Gly Ile Ile Pro Val Ser Ser Val Asn Gly Thr Gly
260 265 270

Ile Gly Phe Gly Ser Val Gly Glu Gln Thr Arg Arg Ile His Arg Ser
275 280 285

Tyr Trp Asp Lys Arg Ser Ser Gly Trp Tyr Gly Glu Pro Val Ala Tyr
290 295 300

Ala Gln Lys Ala Leu Leu Glu Pro
305 310

<210> 35
<211> 906
<212> DNA

ES 2 599 644 T3

<213> Rhodoferrax ferrireducens

<400> 35
atgccagcac ccgatctttc caaaggcgtc gcctttgtac gtgggcagta cgtccccatt 60
ggtgaagcaa gcattccgat aactgactgg ggtttcctgc gctctgacgc cacctatgac 120
gtggtgacgg tgtgggatgg ttcccttttt cgccctggacg cccatctgga gcgcttcag 180
cgcagctgcc aacgctggcg gcttgaccgc gggctgacgc ccgggcaaat caccggcgtg 240
ttgtcgcaat gcgtgcgcct gagcgggctt cgcgcctctt atgtcgagat gatctgcacc 300
cggggccagc cgccttgggg atcgcgcgac ccacgcctgg ccgtcaatca gttctatgcc 360
tttgcggtgc cctatgtctg gcttgccaat gccagcagc gtgaagcggg actgcacctg 420
atgatcagcg acgtgcaacg tattcctgcg acctcctgg acccttcggc caagaactac 480
cactggaacg acttgacgat gggactgctg ggggcgctgg acgccggggc cgacagtgtt 540
gtgctggtcg actcagtcgg gaacgtggtc gagggaccgc ggttcaactg gttttgtgtc 600
agccacggcg cgcttgtgac gccagcagag ggcattgctt aaggggtgtc gcgtgcacc 660
gtcatcgaga tggcgcgcgc tctggggctg gagacacaac tgcgtgcctt gcccgccgat 720
gagttgcgtg gagccgagga ggtgttcac tccacctcgg gcggtggtgt actgcccgtg 780
agccgggtgg acaaacgtcc ggttggtgat ggcctccgg gccgatcac gcagcgcctg 840
gtacagacct actgggcctg gcatgccgat ccggtgtaca gccagccgat agattattcg 900
ctttga 906

<210> 36

<211> 301

<212> PRT

<213> Rhodoferrax ferrireducens

<400> 36

Met Pro Ala Pro Asp Leu Ser Lys Gly Val Ala Phe Val Arg Gly Gln
1 5 10 15

Tyr Val Pro Ile Gly Glu Ala Ser Ile Pro Ile Thr Asp Trp Gly Phe
20 25 30

Leu Arg Ser Asp Ala Thr Tyr Asp Val Val Thr Val Trp Asp Gly Ser
35 40 45

Phe Phe Arg Leu Asp Ala His Leu Glu Arg Phe Met Arg Ser Cys Gln
50 55 60

Arg Trp Arg Leu Asp Pro Gly Leu Thr Pro Gly Gln Ile Thr Gly Val
65 70 75 80

ES 2 599 644 T3

Leu Ser Gln Cys Val Arg Leu Ser Gly Leu Arg Ala Ser Tyr Val Glu
85 90 95

Met Ile Cys Thr Arg Gly Gln Pro Pro Trp Gly Ser Arg Asp Pro Arg
100 105 110

Leu Ala Val Asn Gln Phe Tyr Ala Phe Ala Val Pro Tyr Val Trp Leu
115 120 125

Ala Asn Ala Gln Gln Arg Glu Ala Gly Leu His Leu Met Ile Ser Asp
130 135 140

Val Gln Arg Ile Pro Ala Thr Ser Val Asp Pro Ser Ala Lys Asn Tyr
145 150 155 160

His Trp Asn Asp Leu Thr Met Gly Leu Leu Gly Ala Leu Asp Ala Gly
165 170 175

Ala Asp Ser Val Val Leu Val Asp Ser Val Gly Asn Val Val Glu Gly
180 185 190

Pro Gly Phe Asn Val Phe Cys Val Ser His Gly Ala Leu Val Thr Pro
195 200 205

Ser Glu Gly Met Leu Glu Gly Val Ser Arg Arg Thr Val Ile Glu Met
210 215 220

Ala Arg Ala Leu Gly Leu Glu Thr Gln Leu Arg Ala Leu Pro Ala Asp
225 230 235 240

Glu Leu Arg Gly Ala Glu Glu Val Phe Ile Ser Thr Ser Gly Gly Gly
245 250 255

Val Leu Pro Val Ser Arg Val Asp Lys Arg Pro Val Gly Asp Gly Arg
260 265 270

Pro Gly Pro Ile Thr Gln Arg Leu Val Gln Thr Tyr Trp Ala Trp His
275 280 285

Ala Asp Pro Val Tyr Ser Gln Pro Ile Asp Tyr Ser Leu
290 295 300

<210> 37

<211> 900

<212> DNA

<213> Jannaschia sp.

<400> 37

tcaataggcg atgtcggttc ggtgctcggc ccgctgcatc cagtcgaaat accgccgccg

60

ES 2 599 644 T3

cagggtccagc gccacgggcc ctgccgcctc gttggagaag acgcggttat cgacccgcgc 120
 caccgggatc acgccgccgc ccgaggacga caggaagacc tcatcggcct ccaggaattc 180
 gtccagggggc agcgggctg tttcgaccgt caggcccgcc tctgcgccca tttccagcac 240
 cgtgcgggcg gtgatgcat gcaacacacc gtggtccgac gtcacgatcc ggtcaccgaa 300
 cagggcgaag gcgttgaagc cgggcccttc ggtcacatgg cccgcgtggt ccagcagcaa 360
 gaccgtctca aacccttgt ctttcgcctc gaacaaccgc ccggtgaaat cgccccaatg 420
 gtagtttttc acggtgggat caacgctgtc ggcggggatg cggcggtgtc gtttggcgat 480
 ccagacggag gttccacgct cggcgatttc gggtttcacg atgtgcacgt acggcacgca 540
 ccaggcgtag aagtgttg cacaatcgcg cgggtcccgc gatccgggca ccgggttgcg 600
 cccccgcgcg gcgaccatgg cgacatagct gtcgcgcaat ccgcttgcg ccaccatccg 660
 ggtcagggcc gttttcacgc cgtccccgtc gcgccgata tccatccgca gcgcgccgac 720
 ggaggccatg aaccgcgcca cgtaatcctc cagccggaag aagccgcccc gccagaccgg 780
 caccacgtca taggcgatgt cggaatgggt caggcccaaa tcggtgacgg ggcacgcggc 840
 ctctgcgatg gggatgatgc ggctcccat ccaggccgcg ccggttgata aatcattcat 900

<210> 38
 <211> 299
 <212> PRT
 <213> Jannaschia sp.

<400> 38

Met Asn Asp Leu Ser Asn Gly Ala Ala Trp Met Gly Ser Arg Ile Ile
 1 5 10 15

Pro Ile Ala Glu Ala Ala Ile Pro Val Thr Asp Trp Gly Leu Thr His
 20 25 30

Ser Asp Ile Ala Tyr Asp Val Val Pro Val Trp Arg Gly Gly Phe Phe
 35 40 45

Arg Leu Asp Asp Tyr Val Ala Arg Phe Met Ala Ser Val Gly Ala Leu
 50 55 60

Arg Met Asp Ile Gly Arg Asp Gly Asp Gly Val Lys Thr Ala Leu Thr
 65 70 75 80

Arg Met Val Ala Ala Ser Gly Leu Arg Asp Ser Tyr Val Ala Met Val
 85 90 95

Ala Ala Arg Gly Arg Asn Pro Val Pro Gly Ser Arg Asp Pro Arg Asp
 100 105 110

ES 2 599 644 T3

Cys Ala Asn His Phe Tyr Ala Trp Cys Val Pro Tyr Val His Ile Val
 115 120 125

Lys Pro Glu Ile Ala Glu Arg Gly Thr Ser Val Trp Ile Ala Lys Arg
 130 135 140

Thr Arg Arg Ile Pro Ala Asp Ser Val Asp Pro Thr Val Lys Asn Tyr
 145 150 155 160

His Trp Gly Asp Phe Thr Gly Gly Leu Phe Glu Ala Lys Asp Lys Gly
 165 170 175

Phe Glu Thr Val Leu Leu Leu Asp His Ala Gly His Val Thr Glu Gly
 180 185 190

Pro Gly Phe Asn Ala Phe Ala Leu Phe Gly Asp Arg Ile Val Thr Ser
 195 200 205

Asp His Gly Val Leu His Gly Ile Thr Arg Arg Thr Val Leu Glu Met
 210 215 220

Ala Ala Glu Ala Gly Leu Thr Val Glu Thr Arg Pro Leu Pro Leu Asp
 225 230 235 240

Glu Phe Leu Glu Ala Asp Glu Val Phe Leu Ser Ser Ser Gly Gly Gly
 245 250 255

Val Ile Pro Val Ala Arg Val Asp Asn Arg Val Phe Ser Asn Asp Ala
 260 265 270

Ala Gly Pro Val Ala Leu Asp Leu Arg Arg Arg Tyr Phe Asp Trp Ile
 275 280 285

Thr Arg Ala Glu His Arg Thr Asp Ile Ala Tyr
 290 295

<210> 39
 <211> 936
 <212> DNA
 <213> *Labrenzia alexandrii*

<400> 39
 ctaatcccga taagaaatct ccgtccggtt atccggacgg gcggccatt caaataggc 60
 cttgtgaagt gctgtgtaa tcggaccgg cgcatttt gaaaagatcc ggtcatccac 120
 ccgggtgacc ggtgcaacgc cggcaccgga cgtggtgatg aaaacttctg cgctctcaaa 180
 gaactcacta agcgccaagg cccgtatctc gatctcaagc cctgttctg ccgcatatc 240

ES 2 599 644 T3

cagcacogtc tgacggctga tgccttccaa aacaccgctt tcagcogtca ccagcgttcg 300
 gccctttaca gcaaacacat tgaagcccgg cccttcggtc acgttgccat catcatcgag 360
 caagatgaca gtctccgcgc cgtggtcctt ggccctcaaac agacccttgg tgaaatcacc 420
 ccagtggtag tttttgacct tcggattgac actgcccgggc gaaatccggg gcacggtctt 480
 ggcgatatga acatggggcgc ctctttccgc gatctccggc cggatgaogt gaacataggg 540
 cacgcaccat gcaaagaatt ggttcttgca gtcccgggga tcgcgcgaac ccggcacctg 600
 gttcacaccc cgggaagtga ccattgccac gtaggccttt cgcagccccg tcgccgccac 660
 gagcgattga agggcgtctt ccatctcctc atcgctcaac ccagggtcca accgcagttc 720
 gtccatcgag cggcggaaac gggccagata atggggcaaa cgaagaatg caccatccaa 780
 gaccgggacc acatcatagg taatgtccga atggatcagc cccaatctg taacaggcaa 840
 tgccgcctcg gaaatcgca tgagcctgcc gcccatccac gcggcacogt tggacaaatc 900
 tgagatggta acgcgaacat cgacgggtgga ggacat 936

<210> 40
 <211> 311
 <212> PRT
 <213> *Labrenzia alexandrii*

<400> 40

Met Ser Ser Thr Val Asp Val Arg Val Thr Ile Ser Asp Leu Ser Asn
 1 5 10 15

Gly Ala Ala Trp Met Gly Gly Arg Leu Met Pro Ile Ser Glu Ala Ala
 20 25 30

Leu Pro Val Thr Asp Trp Gly Leu Ile His Ser Asp Ile Thr Tyr Asp
 35 40 45

Val Val Pro Val Leu Asp Gly Ala Phe Phe Arg Leu Pro His Tyr Leu
 50 55 60

Ala Arg Phe Arg Arg Ser Met Asp Glu Leu Arg Leu Asp Pro Gly Leu
 65 70 75 80

Ser Asp Glu Glu Met Glu Asp Ala Leu Gln Ser Leu Val Ala Ala Thr
 85 90 95

Gly Leu Arg Lys Ala Tyr Val Ala Met Val Thr Ser Arg Gly Val Asn
 100 105 110

Gln Val Pro Gly Ser Arg Asp Pro Arg Asp Cys Lys Asn Gln Phe Phe
 115 120 125

ES 2 599 644 T3

Ala Trp Cys Val Pro Tyr Val His Val Ile Arg Pro Glu Ile Ala Glu
 130 135 140

Arg Gly Ala His Val His Ile Ala Lys Thr Val His Arg Ile Ser Pro
 145 150 155 160

Gly Ser Val Asn Pro Lys Val Lys Asn Tyr His Trp Gly Asp Phe Thr
 165 170 175

Lys Gly Leu Phe Glu Ala Lys Asp His Gly Ala Glu Thr Val Ile Leu
 180 185 190

Leu Asp Asp Asp Gly Asn Val Thr Glu Gly Pro Gly Phe Asn Val Phe
 195 200 205

Ala Val Lys Gly Arg Thr Leu Val Thr Ala Glu Ser Gly Val Leu Glu
 210 215 220

Gly Ile Ser Arg Gln Thr Val Leu Asp Ile Ala Ala Glu Gln Gly Leu
 225 230 235 240

Glu Ile Glu Ile Arg Ala Leu Ala Leu Ser Glu Phe Phe Glu Ser Asp
 245 250 255

Glu Val Phe Ile Thr Thr Ser Gly Gly Gly Val Ala Pro Val Thr Arg
 260 265 270

Val Asp Asp Arg Ile Phe Ser Asn Asp Ala Pro Gly Pro Ile Thr Thr
 275 280 285

Ala Leu His Lys Ala Tyr Phe Glu Trp Ala Ala Arg Pro Asp Asn Arg
 290 295 300

Thr Glu Ile Ser Tyr Arg Asp
 305 310

<210> 41
 <211> 981
 <212> DNA
 <213> Burkholderia sp.

<400> 41
 ttatgccggg accgacgaga agtattcagc cggctgcgcg tgccatcctg cccagcgctt 60
 ctcccagtag aggttgtgaa tcttctogga gatcggggcgg gggccggttgc ggttgccgag 120
 cggctgatcg togattgcgc tgacggggcat gatcccgct gccgacgacg aggtgaatgc 180
 ttcgtcagcc tcgcgagct gggtcgggt gtacttgccg atgttcacct tgatgccgag 240

ES 2 599 644 T3

ctcggcggcg atatcgaaga cgtctcggcg cgtgatgcc agcaggcagc ctcggcggg 300
 cgtatagagc tccccgttct tggcgaaaaa cacgttcgca ccggctgctt cggtgaggtt 360
 gtcgctttcg tcgacgagaa ccgccagtc cttctcctgc gtcattggctt cgaacagcgc 420
 gagcttcatg tccatccagt ggaagttctt ggccgtggga tcgaccgcct tcgcggaaat 480
 gcggttgta agcttgctga tcagcagggt cgagcctcgc gtgcggacct cgtcgtcagc 540
 ctggaagaag aacggcacgg caaacgcgaa catcgcgttc ttcattggcgc cgcggtcgcg 600
 gcgatcgacg ctcagcgggc cgcgggtgac gcaccaccag acgtacgcgt ccttcaggcc 660
 ggcgttgcc acgaggttgt ggagaatctc cttgacctgg tcgcgattga acggattctc 720
 aaggaagaac ttgcgccagc actcttccat gcgcgtgagg tgatcgtcga gcgggaagaa 780
 gttgccccgc gagacgggta cgacatcgtg gccgcgtcgc gcatgcagga atcctgcgtc 840
 gacgagcggg acggtcgtt cgggtgatggg gacgtagttg ccgtcacaga aggcgctgcc 900
 atcttcgtag cggggttcat gcggagcgcg agcatgcagc gggttctcgt gcatgatctg 960
 ctggacttga atgattgcc a t 981

<210> 42
 <211> 326
 <212> PRT
 <213> Burkholderia sp.

<400> 42

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

ES 2 599 644 T3

Tyr Val Trp Trp Cys Val Thr Arg Gly Pro Leu Ser Val Asp Arg Arg
 115 120 125
 Asp Arg Gly Ala Met Lys Asn Ala Met Phe Ala Phe Ala Val Pro Phe
 130 135 140
 Phe Phe Gln Ala Asp Asp Glu Val Arg Thr Arg Gly Ser Asn Leu Leu
 145 150 155 160
 Ile Ser Lys Leu Tyr Asn Arg Ile Ser Ala Lys Ala Val Asp Pro Thr
 165 170 175
 Ala Lys Asn Phe His Trp Met Asp Met Lys Leu Ala Leu Phe Glu Ala
 180 185 190
 Met Thr Gln Glu Lys Asp Trp Ala Val Leu Val Asp Glu Ser Asp Asn
 195 200 205
 Leu Thr Glu Ala Ala Gly Ala Asn Val Phe Phe Ala Lys Asn Gly Glu
 210 215 220
 Leu Tyr Thr Pro Ala Glu Gly Cys Leu Leu Gly Ile Thr Arg Gln Ser
 225 230 235 240
 Val Phe Asp Ile Ala Ala Glu Leu Gly Ile Lys Val Asn Ile Gly Lys
 245 250 255
 Tyr Thr Ala Thr Gln Leu Arg Glu Ala Asp Glu Ala Phe Thr Ser Ser
 260 265 270
 Ser Ala Gly Gly Ile Met Pro Val Ser Ala Ile Asp Asp Gln Pro Leu
 275 280 285
 Gly Asn Arg Asn Gly Pro Gly Pro Ile Ser Glu Lys Ile His Asn Leu
 290 295 300
 Tyr Trp Glu Lys Arg Trp Ala Gly Trp His Ala Gln Pro Ala Glu Tyr
 305 310 315 320
 Phe Ser Ser Val Pro Ala
 325

<210> 43
 <211> 954
 <212> DNA
 <213> Burkholderia cenocepacia

<400> 43
 tcagccgtcc gcgtagtoga ccgccaggct ccacgccgga tcgccgtgct tcgccagta

60

ES 2 599 644 T3

```

cgcgctogaac agccggcgcg tcatcggccc cggacggcgg tcgccgatcg tcgcgctcgtt      120
cagcccgctg accggcatga tgccgcgggc cgtcgacgtg atgaacactt cgtccgctgc      180
gcgcaactgt gcatcgtcga tgcgcgcagc ctgggctcgc atgcccacgc ccgtcgccag      240
ctcgaacacc gtctgcccgc tgatgccatg cagcacgccg cgctcggggc tcgagcagccg      300
gccgtcgcgc acgacgaaca cgttgaacct gggcccttcg gcgatgctgc cgtccgtgca      360
cttcagcagg accgactccg cgcccggctc atagcccttc agcagcccgg cgacgaggtc      420
cagccagtga tagttcttga tctgctgctc gaccgattcg ggcggaatgc gccgcacgtc      480
gtcgatcaag tggaggtgca gcccttcgcg caactgccgc tcgttcgcca ccgatccgta      540
cggcaccgcg aacgcgatga actggttcac cgcacgcgcg gggtcgcccg tgaaggtcgg      600
cgacacgccg cgcgtacaca gcactcgcac gtaggcgtgg cgcagccccg agcggcgcac      660
gcattcgacg aggatgtcgc gcagcgcacg gtcgggtcaac ggcacgttca gccgcaaccg      720
cgcaagcgag cgcctgaatc gtcgatatg cttgtcgagg cggagaagc ggcggtcca      780
aacatgcacg gtgtcgtacg tgacgtcgga gtgcagaaag ccccagtoga gcaccgatac      840
gcgcgcgtcg gcgatcggaa tgaagcggcc gttcatgtag gcggcgcctt gcggaacgc      900
gggcgcttcg accgatggaa tgccatgatg ggccgaacgt gtctcccggc gcac      954

```

```

<210> 44
<211> 317
<212> PRT
<213> Burkholderia cenocepacia

```

```

<400> 44

```

```

Met Pro Arg Glu Thr Arg Ser Ala His His Gly Ile Pro Ser Val Glu
1           5           10           15

```

```

Ala Pro Ala Phe Pro Gln Gly Ala Ala Tyr Met Asn Gly Arg Phe Ile
          20           25           30

```

```

Pro Ile Ala Asp Ala Arg Val Ser Val Leu Asp Trp Gly Phe Leu His
          35           40           45

```

```

Ser Asp Val Thr Tyr Asp Thr Val His Val Trp Asn Gly Arg Phe Phe
          50           55           60

```

```

Arg Leu Asp Lys His Ile Glu Arg Phe Arg Arg Ser Leu Ala Arg Leu
65           70           75           80

```

```

Arg Leu Asn Val Pro Leu Thr Asp Asp Ala Leu Arg Asp Ile Leu Val
          85           90           95

```

ES 2 599 644 T3

Glu Cys Val Arg Arg Ser Gly Leu Arg His Ala Tyr Val Glu Met Leu
 100 105 110

Cys Thr Arg Gly Val Ser Pro Thr Phe Ser Arg Asp Pro Arg Asp Ala
 115 120 125

Val Asn Gln Phe Ile Ala Phe Ala Val Pro Tyr Gly Ser Val Ala Asn
 130 135 140

Glu Arg Gln Leu Arg Glu Gly Leu His Leu His Val Ile Asp Asp Val
 145 150 155 160

Arg Arg Ile Pro Pro Glu Ser Val Asp Pro Gln Ile Lys Asn Tyr His
 165 170 175

Trp Leu Asp Leu Val Ala Gly Leu Leu Lys Gly Tyr Asp Ala Gly Ala
 180 185 190

Glu Ser Val Leu Leu Lys Cys Thr Asp Gly Ser Ile Ala Glu Gly Pro
 195 200 205

Gly Phe Asn Val Phe Val Val Arg Asp Gly Arg Leu Arg Thr Pro Glu
 210 215 220

Arg Gly Val Leu His Gly Ile Thr Arg Gln Thr Val Phe Glu Leu Ala
 225 230 235 240

Thr Ala Met Gly Ile Asp Ala Gln Ala Ala Arg Ile Asp Asp Ala Gln
 245 250 255

Leu Arg Asp Ala Asp Glu Val Phe Ile Thr Ser Thr Ala Gly Gly Ile
 260 265 270

Met Pro Val Thr Arg Leu Asn Asp Ala Thr Ile Gly Asp Gly Arg Pro
 275 280 285

Gly Pro Met Thr Arg Arg Leu Phe Asp Ala Tyr Trp Ala Lys His Gly
 290 295 300

Asp Pro Ala Trp Ser Leu Ala Val Asp Tyr Ala Asp Gly
 305 310 315

- <210> 45
- <211> 864
- <212> DNA
- <213> alpha proteobacterium

<400> 45
 ttaatatcca actggatcac gatgtttacc ttcattgtgc agctgccaat aggcattcat

ES 2 599 644 T3

tatacgtttc gacaaaaggcc cgatattgcc ggaagcaatt ggtcgttcgt caattcttgt 120
aacaggcatc acgcccccg cagtcgaggt tatgaagacc tcgtcagcgg tttaagctt 180
gtcgacagta atgtcttgcg catgacatct aactccaat tcttcgcaaa tgcataaaat 240
ggtttgccgc gtaattccca acaagacact tacgtcgggc gttgaaacca caccattggt 300
gatcgagaaa atattgaatc ctggcccttc ggacacgttg ccattaaggt caatcagagt 360
ggctgtgtct gctccacgtt cgtaggcagc ataaagccct ttgaccatat ccaaccaatg 420
gtaattcttg acggtaggat caacagagct tgggtggaatg cgaacgacat ccgtgatcgc 480
cagatgcaaa ccacactcca actgttcttt gctggccact gaaccgaacg gaatggcaaa 540
agctatgaag cggttttctg catctcttgg atctcgactg aagtgggtg acgttctct 600
tgtgcagatg aactcaacat atgcattttt gagcttgac aacgccacgc aattatgaag 660
aatttctatg atttgcgctt tgggtgaagg aattgacata tgcaaagcgt ccattccacg 720
ggaaaatcga tctaaatgat cacctaactt gaaaaatgct ccgcccata catgggcaac 780
gtcataagtc gcgtctgaat gaagaaacc ataatacaaa acggacaatt tcgcctcatg 840
gatagtagg tactgcccac ccat 864

<210> 46
<211> 287
<212> PRT
<213> alpha proteobacterium
<400> 46

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

ES 2 599 644 T3

Phe Gly Ser Val Ala Ser Lys Glu Gln Leu Glu Cys Gly Leu His Leu
 115 120 125

Ala Ile Thr Asp Val Val Arg Ile Pro Pro Ser Ser Val Asp Pro Thr
 130 135 140

Val Lys Asn Tyr His Trp Leu Asp Met Val Lys Gly Leu Tyr Ala Ala
 145 150 155 160

Tyr Glu Arg Gly Ala Asp Thr Ala Thr Leu Ile Asp Leu Asn Gly Asn
 165 170 175

Val Ser Glu Gly Pro Gly Phe Asn Ile Phe Ser Ile Thr Asn Gly Val
 180 185 190

Val Ser Thr Pro Asp Val Ser Val Leu Leu Gly Ile Thr Arg Gln Thr
 195 200 205

Ile Phe Asp Ile Cys Glu Glu Leu Glu Val Arg Cys His Ala Gln Asp
 210 215 220

Ile Thr Val Asp Lys Leu Lys Thr Ala Asp Glu Val Phe Ile Thr Ser
 225 230 235 240

Thr Ala Gly Gly Val Met Pro Val Thr Arg Ile Asp Glu Arg Pro Ile
 245 250 255

Ala Ser Gly Asn Ile Gly Pro Leu Ser Lys Arg Ile Met His Ala Tyr
 260 265 270

Trp Gln Leu His Asn Glu Gly Lys His Arg Asp Pro Val Glu Tyr
 275 280 285

<210> 47

<211> 882

<212> DNA

<213> gamma proteobacterium

<400> 47

atgagtgacg agcccattat ctatatcaac ggtgactatc ttccactgtc tcagggcgcg 60

gtatccccgg tggatcaggg attcctgctc ggtgatggcg tattcgatgt ggtctcgcgc 120

tggaagggca acatcttcaa gctcgacgcc catctcgatc ggttctttga ctcgattcag 180

gccgcgcgac tcaatcaoga catgagtoga gacgcgtgga aggaagcgat catogaaacc 240

acgcgtcgca atggactoga cgatgcctcg attogcttta tcgtgaccog cggcgcagccc 300

aaaggggtgg ttgctgatcc cggggatttt aaaccgacgt gcatcgtctg ggtggcgcct 360

ES 2 599 644 T3

tatatcttcc tcgcgatga ggagaaacgc cgcaatggta ttgcctgat gattagcgcg 420
 acgcgggggtt tcctgctga caccctggac cctcgttaca aatgcctcga ccgcctgcat 480
 tcacagctga ttcggttga ggcctggag gcgggttatg acgatgcgct ttggctcgat 540
 cattccggtc acgtgtccga gtcagcagcg agcaacctgt ttatcgtcaa gaatggtgtg 600
 ttgtacaccc cttcagcagg aattctgctc ggcatcacac gggacaccat tctcgagctc 660
 gcgaccgagc tggacatccc ctggaagag cgacagctca gtgcgttcga tgtctatata 720
 gccgatgagg tcttcacctg cagcacagcg ggtggcgcgc ttccggtcag ggaggtcgca 780
 ggtcgaacga ttcgggcac aacccccggc ccgattaccc aggcaatcga caacgcgtat 840
 tggcgatgc gtgaaacaga ccggtacgcg acgcccgttt aa 882

<210> 48
 <211> 293
 <212> PRT
 <213> gamma proteobacterium

<400> 48

Met Ser Asp Glu Pro Ile Ile Tyr Ile Asn Gly Asp Tyr Leu Pro Leu
 1 5 10 15

Ser Gln Ala Arg Val Ser Pro Val Asp Gln Gly Phe Leu Leu Gly Asp
 20 25 30

Gly Val Phe Asp Val Val Ser Ala Trp Lys Gly Asn Ile Phe Lys Leu
 35 40 45

Asp Ala His Leu Asp Arg Phe Phe Asp Ser Ile Gln Ala Ala Arg Leu
 50 55 60

Asn His Asp Met Ser Arg Asp Ala Trp Lys Glu Ala Ile Ile Glu Thr
 65 70 75 80

Thr Arg Arg Asn Gly Leu Asp Asp Ala Ser Ile Arg Phe Ile Val Thr
 85 90 95

Arg Gly Glu Pro Lys Gly Val Val Ala Asp Pro Arg Asp Phe Lys Pro
 100 105 110

Thr Cys Ile Val Trp Val Ala Pro Tyr Ile Phe Leu Ala Asp Glu Glu
 115 120 125

Lys Arg Arg Asn Gly Ile Arg Leu Met Ile Ser Ala Thr Arg Gly Phe
 130 135 140

ES 2 599 644 T3

Pro Ala Asp Thr Leu Asp Pro Arg Tyr Lys Cys Leu Asp Arg Leu His
 145 150 155 160

Ser Gln Leu Ile Arg Leu Glu Ala Leu Glu Ala Gly Tyr Asp Asp Ala
 165 170 175

Leu Trp Leu Asp His Ser Gly His Val Ser Glu Ser Ala Ala Ser Asn
 180 185 190

Leu Phe Ile Val Lys Asn Gly Val Leu Tyr Thr Pro Ser Ala Gly Ile
 195 200 205

Leu Arg Gly Ile Thr Arg Asp Thr Ile Leu Glu Leu Ala Thr Glu Leu
 210 215 220

Asp Ile Pro Trp Lys Glu Arg Gln Leu Ser Ala Phe Asp Val Tyr Ile
 225 230 235 240

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

Arg Glu Val Ala Gly Arg Thr Ile Arg Gly Thr Thr Pro Gly Pro Ile
 260 265 270

Thr Gln Ala Ile Asp Asn Ala Tyr Trp Ala Met Arg Glu Thr Asp Arg
 275 280 285

Tyr Ala Thr Pro Leu
 290

<210> 49
 <211> 1014
 <212> DNA
 <213> Mycobacterium vanbaalenii

<400> 49
 atgggcattg ataccggcac cagcaatctg gttgcagttg aaccgggtgc aattcgtgaa 60
 gatacaccgg caggcagcgt tattcagtat agcgattatg aaattgatta tagcagcccc 120
 tttgccggtg gtgttgcattg gattgaaggt gaatatctgc ctgcagaaga tgcaaaaatt 180
 agcatttttg ataccggttt tggatcatagc gatctgacct ataccgttgc acatgtttgg 240
 catggcaata tttttcgtct gggcgatcat ctggatcgtc tgctggatgg tgcacgtaaa 300
 ctgcgtctgg atagcgggta taccaaagat gaactggccg atattaccaa aaatgtgtg 360
 agcctgagcc agctgctgga aagctttggt aatctgacca ttaccctggg ttatggtaaa 420
 cgtaaaggcg aaaaagatct gagcaaacgt acccatcagg tgtatattta tgccattccg 480
 tatctgtggg catttcctcc ggcagagcag atttttggca ccaccgcagt tgttccgcgt 540

ES 2 599 644 T3

catgttcgtc gtgcoggtcg taatacogtt gatccgacca ttaaaaatta tcagtggggt 600
gatctgaccg cagcaagctt tgaagcaaaa gatcgtggtg cacgtaccgc aattctgatg 660
gatgccgata attgtgttgc agaaggtccg ggttttaacg tgtgcattgt gaaagatggt 720
aaactggcaa gcccgagccg taatgcaactg ccgggtatta cacgtaaac cgtgtttgaa 780
attgccgggtg caatgggtat tgaagcagca ctgcgtgatg ttaccagcca tgaactgtat 840
gatgccgatg aaattatggc agttaccacc gcaggcgggtg ttaccccgat taataccctg 900
gatggtgttc cgattggtga tggngaaccg ggtccggtta ccgttgccat togtgatcgt 960
ttttgggcac tgatggatga acctggtccg ctgattgaag caattcagta ttaa 1014

<210> 50
<211> 337
<212> PRT
<213> *Mycobacterium vanbaalenii*

<400> 50

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

ES 2 599 644 T3

Tyr Leu Trp Ala Phe Pro Pro Ala Glu Gln Ile Phe Gly Thr Thr Ala
 165 170 175
 Val Val Pro Arg His Val Arg Arg Ala Gly Arg Asn Thr Val Asp Pro
 180 185 190
 Thr Ile Lys Asn Tyr Gln Trp Gly Asp Leu Thr Ala Ala Ser Phe Glu
 195 200 205
 Ala Lys Asp Arg Gly Ala Arg Thr Ala Ile Leu Met Asp Ala Asp Asn
 210 215 220
 Cys Val Ala Glu Gly Pro Gly Phe Asn Val Cys Ile Val Lys Asp Gly
 225 230 235 240
 Lys Leu Ala Ser Pro Ser Arg Asn Ala Leu Pro Gly Ile Thr Arg Lys
 245 250 255
 Thr Val Phe Glu Ile Ala Gly Ala Met Gly Ile Glu Ala Ala Leu Arg
 260 265 270
 Asp Val Thr Ser His Glu Leu Tyr Asp Ala Asp Glu Ile Met Ala Val
 275 280 285
 Thr Thr Ala Gly Gly Val Thr Pro Ile Asn Thr Leu Asp Gly Val Pro
 290 295 300
 Ile Gly Asp Gly Glu Pro Gly Pro Val Thr Val Ala Ile Arg Asp Arg
 305 310 315 320
 Phe Trp Ala Leu Met Asp Glu Pro Gly Pro Leu Ile Glu Ala Ile Gln
 325 330 335
 Tyr

REIVINDICACIONES

- 5 1. Un proceso para la preparación de una amina quiral ópticamente activa, que comprende hacer reaccionar al menos un receptor de amino con al menos un donante de amino con una ω -transaminasa (R)-selectiva preparada mediante un método que comprende las siguientes etapas:
- a) proporcionar al menos una secuencia de biomolécula de consulta de al menos una ω -transaminasa (R)-selectiva y al menos un banco de biomoléculas,
- 10 b) realizar la búsqueda en el banco de biomoléculas con la secuencia de biomolécula de consulta para identificar un grupo de primeras secuencias de biomoléculas diana, en donde las primeras secuencias de biomoléculas diana tienen un grado de identidad de secuencia de al menos el 20% con respecto a la secuencia de biomolécula de consulta, calculado a nivel de aminoácidos,
- c) seleccionar en el grupo de primeras secuencias de biomoléculas diana un grupo de segundas secuencias de biomoléculas diana, que no comprenda, a nivel de procesos aminoácidos, al menos una de las siguientes estructuras de aminoácido c1) a c3) con
- 15 c1) en la posición 95 a 97 una secuencia de aminoácidos Tyr Xa1 Xa2, siendo Xa1 un aminoácido Ile, Val, Leu, Met, Phe, y siendo Xa2 un aminoácido Arg o Lys, o
- c2) en la posición 97 a 99 una secuencia de aminoácidos Tyr Xaa Gln, siendo Xaa un aminoácido y en la región de la posición 105 a 111 una secuencia de aminoácidos Arg Xaa Xa3, siendo Xa3 un aminoácido, que preferiblemente es His, o
- 20 c3) en la posición 38 Thr, en la posición 97 Lys y en la posición 107 a 109 una secuencia de aminoácidos Arg Xa4 Xa5, siendo Xa4 un aminoácido, que preferiblemente es Gly, y siendo Xa5 un aminoácido, que preferiblemente es Tyr
- y que comprende
- 25 c4) en la posición 95 un aminoácido distinto de Tyr, Arg, Lys, o en la posición 95 Tyr, pero en la posición 97 no Arg o Lys, y
- c5) en la posición 40 no Lys o Arg, y
- c6) en la región de la posición 161 a 165 al menos Lys
- para identificar un grupo de segundas secuencias de biomoléculas diana, y
- 30 d) proporcionar una biomolécula que tenga la segunda secuencia de biomolécula diana identificada en la etapa c) y que sea, o que codifique al menos parcialmente, una proteína con la actividad de una ω -transaminasa (R)-selectiva,
- y obtener la amina quiral ópticamente activa.
2. El proceso según la reivindicación 1, en el que la biomolécula es una proteína y la secuencia de biomolécula es una secuencia de aminoácidos.
- 35 3. El proceso según la reivindicación 1, en donde la biomolécula es una molécula de DNA y la secuencia de biomolécula es una secuencia de DNA.
4. El proceso según una cualquiera de las reivindicaciones precedentes, en donde el banco de biomoléculas es una base de datos de biomoléculas y la base de datos de biomoléculas se utiliza para la búsqueda de la etapa b) con una herramienta de alineamiento de secuencias de biomoléculas, en particular BLAST.
- 40 5. El proceso según una cualquiera de las reivindicaciones precedentes, en donde la proteína o la molécula de DNA proporcionadas en la etapa d) se proporciona mediante síntesis partiendo de cero.
6. El proceso según una cualquiera de las reivindicaciones 1 y 3, en donde el banco de biomoléculas es un banco de genes y el banco de genes se utiliza para la búsqueda de la etapa b) con una molécula de secuencia de DNA de consulta.
- 45 7. El proceso según la reivindicación 6, en donde en la etapa c) se usan cebadores de secuencia de DNA para seleccionar el grupo de segundas secuencias de biomoléculas.
8. El proceso según una cualquiera de las reivindicaciones precedentes, en donde la secuencia de biomolécula de consulta es una secuencia que representa toda una ω -transaminasa (R)-selectiva funcional, o una parte de la misma.

9. El proceso según una cualquiera de las reivindicaciones precedentes, en donde la ω -transaminasa (R)-selectiva usada para hacer reaccionar el al menos un aceptor de amino con el al menos un donante de amino se selecciona del grupo de secuencias de biomoléculas que consisten en las SEQ ID NO: 1 y 3.

Figura 1 - patrón de B3APi

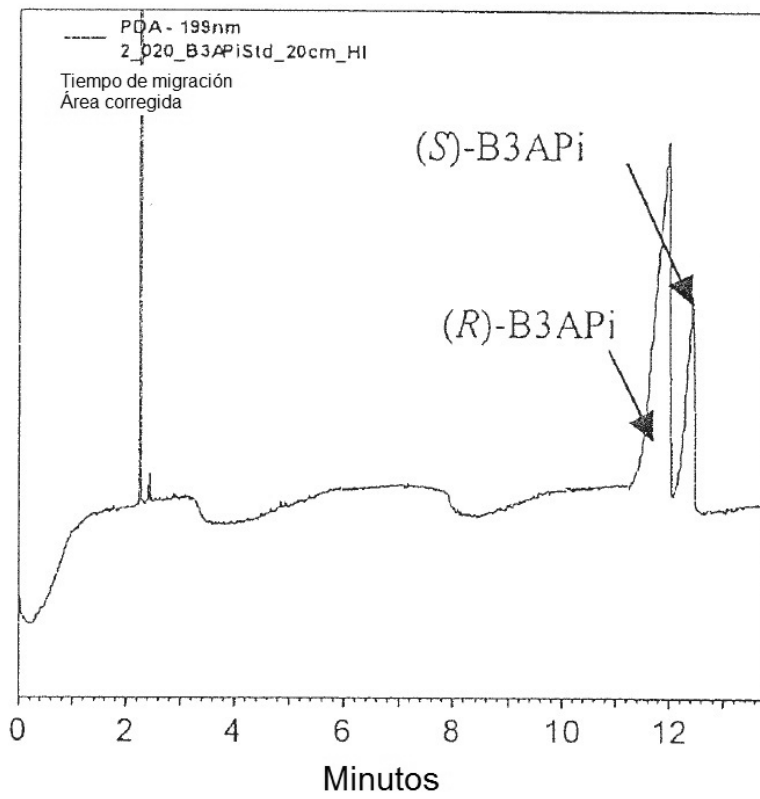


Figura 2 - Síntesis asimétrica de B3APi

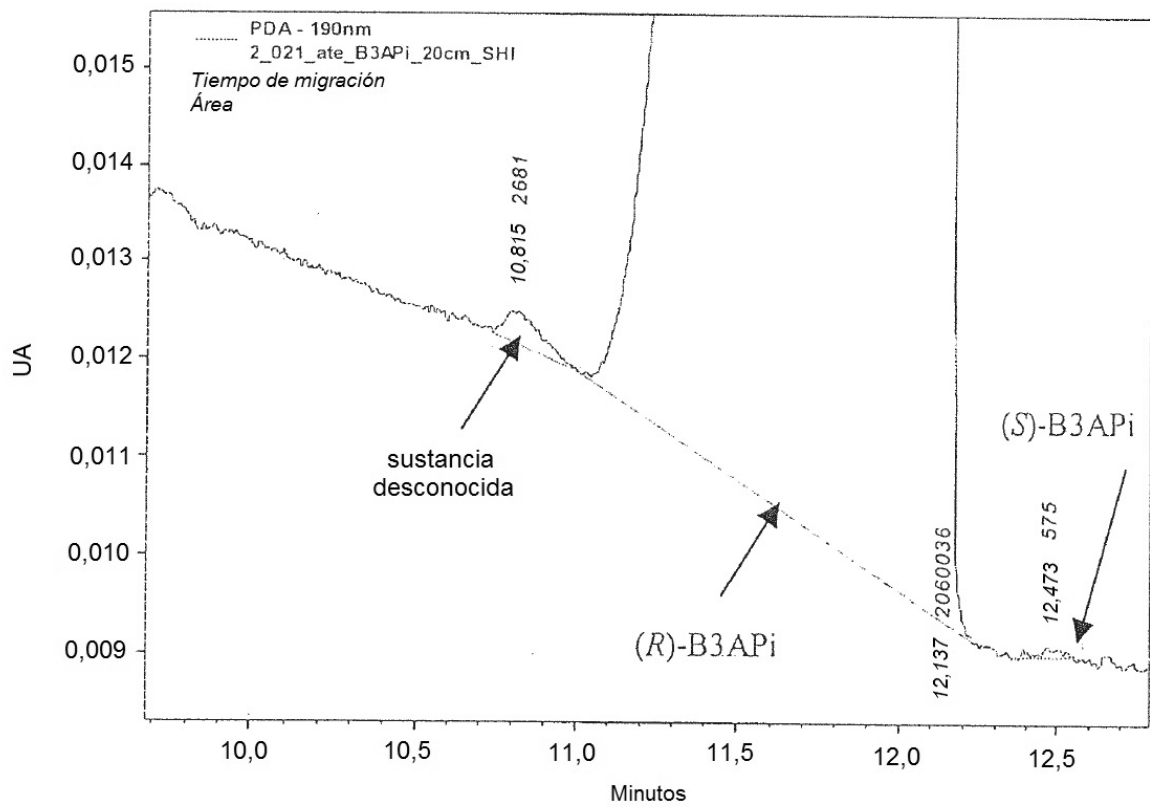


Figura 3 - síntesis asimétrica de C3AP

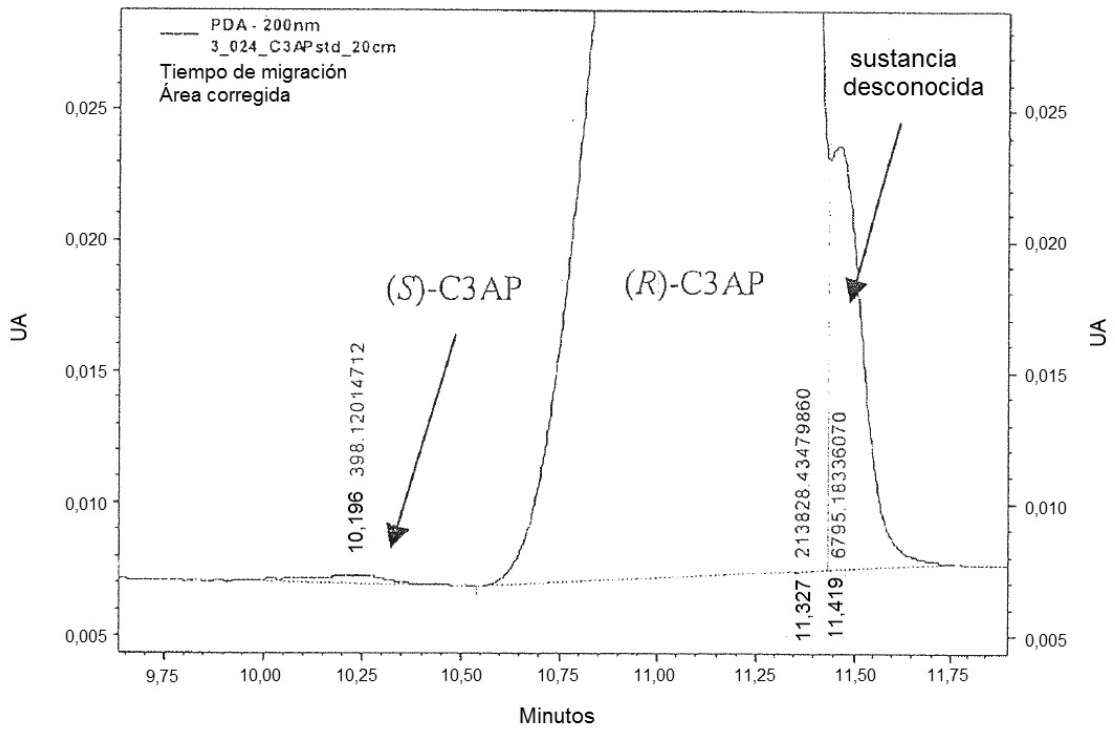


Figura 4 - síntesis asimétrica de MPPA

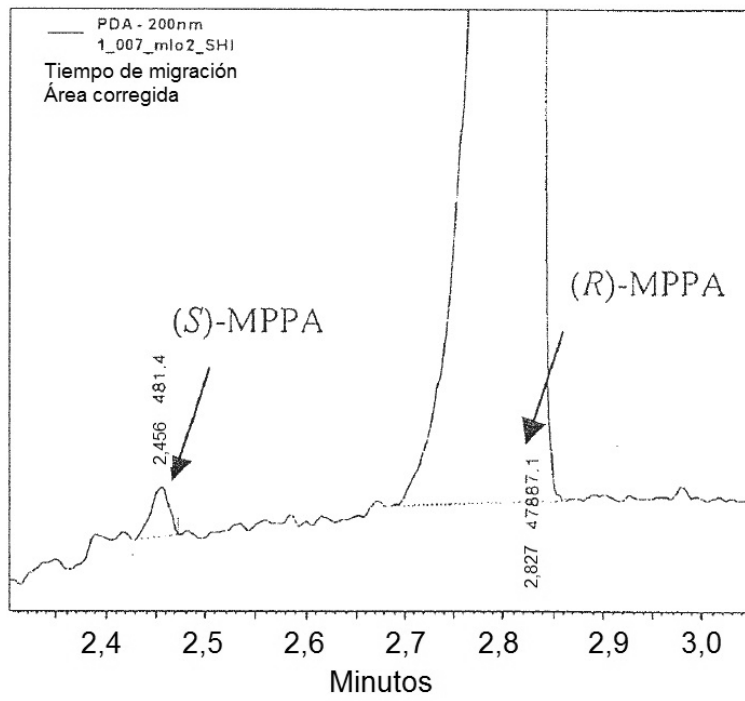


Figura 5 - síntesis asimétrica de B3AP

