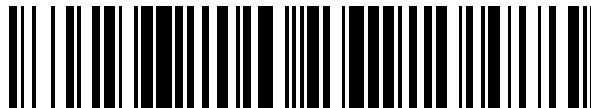


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 562 458**

51 Int. Cl.:

**C12N 9/10** (2006.01)

**C12N 9/88** (2006.01)

**G06F 19/00** (2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **13.08.2010 E 10743046 (4)**

97 Fecha y número de publicación de la concesión europea: **02.12.2015 EP 2473601**

54 Título: **Proceso de identificación y preparación de una transaminasa omega específica de (R)**

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:

**04.03.2016**

73 Titular/es:

**LONZA AG (100.0%)**

**Lonzastrasse  
3930 Visp, CH**

72 Inventor/es:

**HOEHNE, MATTHIAS;  
BORNSCHEUER, UWE;  
ROBINS, KAREN y  
SCHAETZLE, SEBASTIAN**

74 Agente/Representante:

**VALLEJO LÓPEZ, Juan Pedro**

**ES 2 562 458 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

Proceso de identificación y preparación de una transaminasa omega específica de (R)

5 La presente invención se refiere a los procesos de cribado y preparación de transaminasas  $\omega$  específicas de (R).

Las aminas quirales tienen un importante papel en la industria farmacéutica, agroquímica y química. Se utilizan frecuentemente como intermediarios o sintonas para la preparación de varias sustancias activas fisiológicamente, por ejemplo farmacéuticamente, tales como derivados de cefalosporinas o pirrolidina. En un gran número de las  
10 distintas aplicaciones de las aminas quirales, solamente una forma particular ópticamente activa, o bien el enantiómero (R) o el (S) es fisiológicamente activa. En consecuencia, existe la necesidad de proporcionar procesos para la preparación de aminas quirales en una forma ópticamente activa.

Estas necesidades se cumplen parcialmente preparando aminas quirales por cristalización de sales diastereoméricas por medio de la adición de ácidos carboxílicos quirales (Breuer et al., *Angewandte Chemie* (2004) 116, 806-843). Otros métodos químicos utilizan la síntesis enantioselectiva reduciendo precursores proquirales con  
15 dobles enlaces C=N.

Entre los distintos métodos enzimáticos que se han empleado para la síntesis de aminoácidos y aminas ópticamente activas, las transaminasas  $\omega$  ( $\omega$ -TAs) han recibido recientemente una mayor atención debido a su potencial para la  
20 síntesis asimétrica de aminas ópticamente activas, que se utilizan frecuentemente como bloques de construcción para la preparación de numerosos productos farmacéuticos.

Las transaminasas  $\omega$  son enzimas dependientes de PLP (fosfato de piridoxal) que catalizan reacciones de  
25 transferencia de grupos amino. Cuando se emplean las transaminasas  $\omega$ , pueden producirse las aminas denominadas enantioenriquecidas y/o puras ópticamente activas mediante dos estrategias de reacción (i) la síntesis asimétrica partiendo de cetonas y (ii) la resolución cinética partiendo de aminas racémicas. Aunque las transaminasas  $\omega$  muestran una buena enantioselectividad en general, no se han utilizado frecuentemente en síntesis asimétrica, aunque en este caso un rendimiento del 100 % es teóricamente. Un requerimiento específico en  
30 una síntesis asimétrica que emplea transaminasas  $\omega$  es desplazar el equilibrio hacia el lado del producto, especialmente cuando se utiliza un aminoácido como la alanina como donante de amino, ya que en este caso el equilibrio está en el lado de los sustratos (cetona, alanina) y no en el lado de los productos (amina, piruvato); otro requerimiento es que la estereoselectividad de la enzima tiene que ser perfecta, lo que no es siempre el caso de las transaminasas  $\omega$ . Por lo tanto, las transaminasas  $\omega$  se utilizan principalmente para la resolución cinética de aminas  
35 racémicas, en las que no tiene que ser necesariamente perfecta. Por lo tanto, aunque se han investigado las resoluciones cinéticas de aminas racémicas, la limitación a un rendimiento máximo del 50 % dificulta su aplicación considerablemente. Por otro lado, la síntesis asimétrica necesita métodos para desplazar el equilibrio no favorable hacia la síntesis de enantiómeros individuales de aminas ópticamente puras para lo que se han desarrollado varios métodos, los cuales son un pre-requisito clave para los procesos eficaces que hacen posible el uso de  
40 transaminasas a escala industrial. Tales métodos se desvelan en el documento WO 2007/093372.

El documento EP 0 987 332 A1 desvela un proceso de producción de compuestos de amina ópticamente activos, a saber compuestos de (R)-amina, mediante una enzima microbiana, en particular una transaminasa  $\omega$  selectiva de (R) procedente de *Arthrobacter sp.* El documento EP 1 038 953 A1 desvela una transaminasa  $\omega$  selectiva de (R)  
45 más, la cual, sin embargo, procede de *Mycobacterium aurum*.

Iwasaki et al. (*Biotechnol. Let.* 2003 (25), 1843-1846), Koszelewski et al. (*Adv. Synth. Catal.* 2008 (350), 2761-2766) y Iwasaki et al. (*Appl. Microbiol. Biotechnol.* 2006 (69), 499-505) desvelan una transaminasa específica de (R) de *Arthrobacter sp.* La identificación de microorganismos que proporcionan transaminasas selectivas de (S) o (R) útiles  
50 se efectúa habitualmente seleccionando microorganismos que se obtienen, por ejemplo, de muestras de suelo y enriqueciéndolos en cultivos (Jun et al., *App. Microbiol. Biotechnol.* 2004 (70), 2529-2534 y Shin et al. (*Biosc. Biotechnol. Biochem.* 2001 (65), 1782-1788)). Especialmente todas las transaminasas  $\omega$  selectivas de R descritas se obtuvieron por enriquecimiento en cultivo. Tales métodos consumen tiempo, aunque para un proceso eficaz se prefiere a menudo que se sobre-exprese la enzima en un organismo heterólogo como *Escherichia coli*. Esto necesita el aislamiento de la secuencia genética de organismos de tipo silvestre pero la clonación de la secuencia de ADN  
55 no siempre es satisfactoria y es un proceso que consume mucho tiempo.

En particular, durante el proceso de desarrollo o para la identificación de nuevas  $\omega$ -TA, es de gran interés la purificación de la enzima y la caracterización de sus propiedades enzimáticas. Aunque, sin embargo, la actividad de la  $\omega$ -TA habitualmente se determina con métodos de bajo rendimiento como la HPLC o electroforesis de capilaridad (CE), la determinación de las propiedades de la enzima es bastante a menudo la etapa limitante.  
60

En general, en comparación con las transaminasas  $\omega$  selectivas de (S) el número de transaminasas  $\omega$  selectivas de (R) es mucho más limitado. Esto se encuentra en claro contraste con la alta demanda de transaminasas  $\omega$  selectivas de (R) que son altamente deseables para la síntesis asimétrica de enantiómeros (R) de varias aminas quirales. Por lo tanto, sigue existiendo la necesidad de proporcionar más transaminasas  $\omega$  selectivas de (R) y medios y métodos  
65

para obtenerlas con el fin de producir enantiómeros (R), en particular aminas ópticamente activas, por ejemplo, para aplicaciones farmacéuticas o agroquímicas, que hasta ahora no están en absoluto disponibles o no lo están en un proceso económicamente viable, preferentemente a escala industrial.

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

10 La presente invención resuelve este problema técnico proporcionando las enseñanzas de las reivindicaciones independientes.

Por lo tanto la presente enseñanza proporciona un proceso para la preparación de una transaminasa  $\omega$  selectiva de (R) (también denominada  $\omega$ -TA), que comprende las siguientes etapas:

15 a) proporcionar al menos una secuencia de biomolécula incógnita de al menos una transaminasa  $\omega$  selectiva de (R), preferentemente una transaminasa que pertenezca a las enzimas dependientes de PLP de plegamiento tipo IV, y al menos un banco de biomoléculas,

20 b) cribar el banco de biomoléculas con la secuencia de biomolécula incógnita para identificar un grupo de primeras secuencias de biomolécula diana, en el que las primeras secuencias de biomolécula diana tienen un grado de identidad de secuencia de al menos un 20 %, preferentemente de un 32 %, con la secuencia de biomolécula incógnita, calculada a nivel de aminoácidos,

25 c) seleccionar en el grupo de las primeras secuencias de biomolécula diana un grupo de segundas secuencias biomolécula diana, que no comprende, a nivel de aminoácidos, al menos uno de los siguientes motivos de secuencia de aminoácido c1) a c3) con

30 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

35 c2) en la posición 97 a 99 una secuencia de aminoácidos Tyr Xaa Gln, siendo Xaa un aminoácido, preferentemente un aminoácido convencional y en la región desde la posición 105 a 111, preferentemente en la posición 107 a 109, una secuencia de aminoácidos Arg Xaa Xa3, siendo Xa3 un aminoácido, siendo preferentemente His o

40 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, siendo preferentemente Gly, y siendo Xa5 un aminoácido, siendo preferentemente Tyr, y que comprenden

c4) en la posición 95 otro aminoácido distinto de Tyr, Arg, Lys, o en la posición 95 Tyr, pero en la posición 97 ni Arg ni Lys y

45 c5) en la posición 40 ni Lys ni Arg y

c6) en la región desde la posición 161 a 165, preferentemente en la posición 163, al menos una Lys

50 para identificar un grupo de segundas secuencias de biomolécula diana y

d) proporcionar, preferentemente preparar, una biomolécula que tiene la segunda secuencia de biomolécula diana identificada en la etapa c) y que es, o que codifica al menos parcialmente, una proteína con la actividad de una transaminasa  $\omega$  selectiva de (R).

55 La expresión “una biomolécula que es o que codifica al menos parcialmente una proteína” significa preferentemente que la biomolécula puede ser o una proteína con la actividad de una transaminasa  $\omega$  selectiva de (R) o, en caso de que la biomolécula sea una secuencia de nucleótidos, que al menos una parte de dicha secuencia de nucleótidos codifica dicha proteína, preferentemente una proteína de longitud completa. En consecuencia, en caso de que la biomolécula que se proporciona 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 la transaminasa  $\omega$  selectiva de (R), en que posiblemente una parte adicional de dicha molécula de nucleótido tiene alguna otra función, por ejemplo una función reguladora o replicadora. Por lo tanto, la expresión “proporcionar una molécula 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 la transaminasa  $\omega$  selectiva de (R)” es preferentemente equivalente a la expresión “proporcionar una biomolécula que tiene la segunda secuencia de biomolécula diana identificada en la etapa c) que puede ser una proteína con la actividad de la transaminasa  $\omega$  selectiva de (R) o que puede ser una molécula de secuencia de

nucleótido que incluye una secuencia que codifica dicha proteína, siempre que la proteína tenga la actividad de la transaminasa  $\omega$  selectiva de (R)".

5 En una realización de la presente invención en la etapa c2) la posición del motivo Arg Xaa Xa3 puede variar en 1-2 aminoácidos a partir de las posiciones 107-109, 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 en 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, en particular un motivo de secuencia de aminoácido, o secuencia de ADN, o motivo de secuencia de ADN determinado se proporciona en la dirección de extremo amino a extremo carboxilo o de 5' a 3', cualquiera que se aplique, y a menos de que especifique otra cosa. Preferentemente, las secuencias determinadas en la dirección dada se proporcionan como una serie continua sin nucleótidos o aminoácidos intermedios, cualquiera que se aplique.

15 La presente invención proporciona un proceso para la preparación de una transaminasa  $\omega$  selectiva de (R) de una manera rápida, eficaz, simple, fiable y fácil. La presente invención permite por lo tanto identificar y preparar transaminasas- $\omega$  selectivas de (R), que bien no se conocían o no habían estado accesibles previamente, dando lugar a numerosas maneras de proporcionar aminos quirales, preferentemente enantiómeros (R), particularmente en  
20 rutas de síntesis asimétricas muy eficaces. Se podría demostrar satisfactoriamente que incluso proteínas con un bajo grado de identidad de secuencia, por ejemplo de solo el 35 %, con una transaminasa  $\omega$  selectiva de (R) en contraste con las expectativas, son de hecho transaminasas  $\omega$  selectivas de (R). De una manera particularmente ventajosa e inesperada, la presente invención permite proporcionar preferentemente transaminasas  $\omega$  selectivas de (R) de *Aspergillus terreus* y *Mesorhizobium loti*, de manera que la transaminasa de *Aspergillus* es la primera  
25 transaminasa  $\omega$  de eucariota y convierte los sustratos con selectividad (R).

Básicamente, la presente invención se basa en la enseñanza técnica de que las proteínas o sus secuencias de nucleótidos codificantes, incluso con un bajo grado de identidad de secuencia con transaminasas  $\omega$  selectivas de (R) conocidas pueden utilizarse como una fuente potencial para identificar y preparar transaminasas  $\omega$  selectivas de (R),  
30 de forma que las supuestas transaminasas  $\omega$  selectivas de (R) se criban y preparan discriminando proteínas no deseadas y por tanto seleccionando positivamente aquellas que presenten motivos de secuencia particulares en la secuencia de aminoácidos, a la que se atribuye la actividad enzimática deseada. Por lo tanto, la presente invención utiliza específicamente características estructurales particulares, en particular la ausencia y presencia de aminoácidos particulares en la supuesta  $\omega$ -TA selectiva de (R), para identificar y preparar transaminasas  $\omega$   
35 selectivas de (R).

En consecuencia, en una primera etapa a) del presente proceso, se proporciona una secuencia de biomolécula incógnita de al menos una transaminasa  $\omega$  selectiva de (R), por ejemplo una transaminasa  $\omega$  selectiva de (R) conocida, o al menos una parte de la secuencia característica para dicha transaminasa  $\omega$  selectiva de (R), que preferentemente es capaz en particular de identificar una enzima dependiente de PLP de plegamiento tipo IV, junto con al menos un banco de biomoléculas. Preferentemente la secuencia de biomolécula incógnita es una ORF (fase de lectura abierta) de una  $\omega$ -TA selectiva de (R), secuencia de nucleótidos codificante, o una parte característica de la misma. En otra realización, la secuencia de biomolécula incógnita es la ORF-secuencia de aminoácidos en sí o una parte característica de la misma.

45 En el contexto de la presente invención una parte característica de la secuencia de biomolécula incógnita de al menos una transaminasa  $\omega$  selectiva de (R), en particular de la transaminasa  $\omega$  selectiva de (R) conocida, es una secuencia de biomolécula que en forma de su secuencia de ADN se hibrida en las siguientes condiciones a la secuencia de biomolécula incógnita de longitud completa, en particular a la secuencia de ADN de la ORF.

50 Los métodos para la hibridación de ácidos nucleicos, tales como ADN, se conocen bien 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 convencionales.

55 Un ejemplo de una hibridación en condiciones rigurosas es el siguiente. Se incuban un filtro con un ácido nucleico inmovilizado sobre este y el ácido nucleico usado como sonda en una solución que comprende un 50 % de formamida, 5 x SSC (750 mM de cloruro sódico y 75 mM de citrato sódico), 50 mM de fosfato sódico (pH 7,6), 5 x de solución de Denhardt, un 10 % de sulfato de dextrano y 20  $\mu$ g/l de ADN de esperma de salmón desnaturalizado a 42 °C durante una noche. Tras la incubación, el filtro se lava con solución 0,2 x SSC (a 65 °C). Estas condiciones rigurosas de hibridación se pueden modificar ajustando la concentración de formamida (las condiciones serán menos rigurosas a medida que disminuye la concentración de formamida) y cambiando la concentración de sales y las condiciones de temperatura.

65 La hibridación en condiciones menos rigurosas se lleva a cabo, por ejemplo, incubando un filtro con un ácido nucleico inmovilizado sobre este y un ácido nucleico que se utiliza como sonda en una solución que comprende 6 x SSCE (20 x SSCE ; 3 mol/l de cloruro sódico, 0,2 mol/l de dihidrogenofosfato de sodio y 0,02 mol/l de EDTA, pH

7,4), SDS al 0,5 %, un 30 % de formamida y 100 µg/l de ADN de esperma de salmón desnaturalizado a 37 °C durante una noche, y lavando el filtro con 1 x de solución SSC que contiene 0,1 % SDS (50 °C).

5 En una realización particularmente preferida, se entiende en la presente invención con la expresión “al menos una secuencia de biomolécula incógnita de al menos una transaminasa ω selectiva de (R)” una secuencia de biomolécula que es adecuada para seleccionar secuencias de biomoléculas de acuerdo con la presencia o ausencia de los motivos de secuencia c1) a c6), tal como se identifican en el presente documento. En consecuencia, dicha secuencia de biomolécula incógnita es una secuencia que criba e identifica aquellas secuencias de biomolécula que no comprenden a nivel de aminoácidos al menos uno de los motivos de secuencias de aminoácidos c1) a c3) y que comprenden a nivel de aminoácidos los motivos de secuencia c4), c5) y c6). Dicha secuencia de biomolécula puede plasmarse como información de una secuencia de aminoácidos o como una molécula de secuencia de ADN o información de secuencia de ADN.

15 En una realización preferida la parte característica de la secuencia de biomolécula incógnita que se utiliza en la etapa a) engloba, preferentemente consiste en, la región desde las posiciones 30 a 170, más preferentemente 30 a 120 de la ORF de una transaminasa ω selectiva de (R).

20 Se hace una búsqueda en la etapa b) posterior en el banco de biomoléculas con la secuencia de biomolécula incógnita para identificar un grupo de primeras secuencias de biomolécula diana, que muestran al menos un grado mínimo de identidad de secuencia de al menos un 20 %, preferentemente un 25 %, preferentemente un 32 %, preferentemente al menos un 33 %, más preferentemente un 34 %, al menos un 35 %, al menos un 36 %, al menos un 40 %, al menos un 50 %, al menos un 60 %, al menos un 70 % y al menos un 80 %, al menos un 90 % o al menos un 95 % con la secuencia de biomolécula incógnita, basándose en el nivel de aminoácidos, y en el que dicho grupo de primeras biomoléculas diana representa una primera selección del banco de moléculas que se utiliza en la etapa a). Dicho grado de identidad de secuencia es preferentemente una identidad de secuencia entre al menos la parte característica, preferentemente esencialmente toda, en particular toda, la ORF de la secuencia de biomolécula incógnita y al menos la parte característica, preferentemente esencialmente toda, en particular toda la ORF de las secuencias de biomoléculas cribadas. Posteriormente a dicha etapa b) en este grupo de primeras secuencias de biomolécula diana, esas secuencias se seleccionan en una etapa c), que no comprende como motivo 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 comprende como motivo 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 de la posición 105 a 111, preferentemente en la posición 107 a 109, una secuencia de aminoácidos Arg Xaa Xa3, siendo Xa3 un aminoácido, preferentemente un aminoácido habitual, siendo preferentemente His o el cual no comprende un motivo 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, preferentemente un aminoácido habitual, siendo preferentemente Gly, y siendo Xa5 un aminoácido, preferentemente un aminoácido habitual, siendo preferentemente Tyr, discriminando y descartando de esta manera las secuencias, que presentan al menos uno de los motivos de secuencia c1), c2) o c3) que se han identificado anteriormente. En una selección adicional o simultánea se seleccionan las secuencias de biomolécula que de acuerdo con el motivo 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 ni Arg ni Lys y que de acuerdo con el motivo de secuencia c5) en la posición 40 no tiene ni Lys ni Arg y que de acuerdo con el motivo de secuencia c6) en la región desde la posición 161 a 165 tiene al menos una Lys, preferentemente una Lys, preferentemente en la posición 163 Lys, para seleccionar e identificar de esta manera un grupo de segundas secuencias de biomoléculas diana. El grupo de segundas secuencias de biomolécula diana que se identifica y selecciona respecto de los motivos de secuencia identificados anteriormente representan secuencias de biomolécula que son o codifican una proteína con la actividad de una transaminasa ω selectiva de (R), que se proporciona de esta manera.

50 En el contexto de la presente invención, una transaminasa es una enzima dependiente de fosfato de piridoxal que cataliza la transferencia de grupos amino, que se clasifican preferentemente en el tipo IV de plegamiento. Las transaminasas se clasifican en E.C.2.6.1.X. De acuerdo con la presente invención, la transaminasa es una transaminasa ω selectiva de (R). En el contexto de la presente invención una proteína con la actividad de una transaminasa ω selectiva de (R) es una proteína que es capaz de catalizar en condiciones de reacción adecuadas una transferencia de grupos nitrogenados tal como los grupos amino de un donante a un receptor tal como es capaz de hacerlo una transaminasa ω (omega) selectiva de (R), (beta-alanina-piruvato transaminasa). En el contexto de la presente invención una transaminasa ω selectiva de (R) es preferentemente una enzima con el código de clasificación E.C.2. 6.1.18.

60 En el contexto de la presente invención el término amina quiral ópticamente activa se refiere a la misma materia objeto que la expresión amina quiral enantioméricamente activa. Estas expresiones se refieren en particular a una preparación que está esencialmente libre, en una realización incluso más preferida libre, del enantiómero no deseado. En consecuencia, una amina quiral ópticamente activa comprende esencialmente un exceso de un enantiómero o incluso consiste solamente en un enantiómero.

65

En particular, en el contexto de la presente invención, una amina quiral ópticamente activa tiene una pureza óptica de al menos un 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 un 99,9 %.

5 En la presente invención la pureza óptica se proporciona en % de exceso de un enantiómero respecto del otro enantiómero. Por lo tanto, la pureza óptica en % es el cociente de la diferencia entre las concentraciones de enantiómero (R) y (S) y la suma de las concentraciones de ambos enantiómeros (pureza óptica de A en % =  $\frac{[A]-[B]}{[A]+[B]} \times 100$ , en que A y B representan las concentraciones de enantiómeros (R) y (S) o viceversa).

10 En el contexto de la presente invención un banco de biomoléculas es una fuente de biomoléculas por si misma o una colección de datos de secuencias de biomolécula, en particular datos de secuencias de polinucleótidos o polipéptidos.

15 En el contexto de la presente invención una biomolécula es preferentemente una molécula de polinucleótido que alberga información genética, en particular una molécula de ADN. En una realización preferida adicional de la presente invención una biomolécula es un polipéptido, en particular una proteína, que comprende un número de aminoácidos. Por lo tanto, un banco de biomoléculas puede ser una fuente física de biomoléculas, en particular puede ser un banco de genes, en particular una biblioteca de ADNc o genómica o es una colección de información  
20 acerca de dichas biomoléculas, en particular una colección de datos de secuencia, en particular secuencias de aminoácidos o secuencias de polinucleótido, en particular secuencias de ADN.

La presente invención también se refiere a secuencias de biomoléculas, a secuencias de aminoácido, a secuencias de polinucleótido o a secuencias de nucleótido, en particular a secuencias de ADN, de forma que dicha expresión  
25 designa por una parte la sustancia física en sí, lo que significa una proteína, un polipéptido o una molécula de ADN, y por otra parte el contenido de información de dicha sustancia química, que, en caso de hacer referencia a un polinucleótido, es el tipo, orden y número de sus nucleótidos y, en caso de hacer referencia a un polipéptido se refiere al tipo, orden y número de los aminoácidos únicos que forman el polipéptido.

30 En el contexto de la presente invención los términos "biomolécula", "molécula de ADN", "molécula de polinucleótido", "polipéptido" o "proteína" se refieren a sustancias químicas per se. En caso de que la presente invención se refiera específicamente a una información de secuencia de aminoácidos, a una información de secuencia de ADN, a una información de secuencia de polinucleótido o a una información de secuencia de biomolécula, esta no se refiere a la forma física de una biomolécula, sino más bien a la información contenida en ella, es decir, al tipo, orden y número  
35 de sus constituyentes, a saber aminoácidos o nucleótidos.

La presente invención puede aplicarse, en una realización preferida A), en sus etapas a) a c) a un banco de biomoléculas que es una colección de datos de secuencia de, en una realización, secuencias de polinucleótidos, en particular secuencias de ADN o, en otra realización, de secuencias de aminoácidos, ambas de las cuales se criban  
40 en una etapa b) con herramientas de alineamiento de secuencia, preferentemente tales como BLAST, que es un herramienta de búsqueda y 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 programas informáticos Wisconsin Genetics (Genetics Computer Group (GCG), Madison, WI, USA). En una etapa c) adicional,  
45 el grupo de primeras secuencias de biomolécula identificadas se somete a etapas de selección adicionales en el transcurso de las cuales los motivos de secuencia identificadas se utilizan para identificar y seleccionar negativamente con los motivos de secuencia c1) a c3) y seleccionar positivamente con los motivos de secuencia c4) a c6) la secuencia de biomolécula deseada. Una vez que se identifica el segundo grupo de secuencias de biomoléculas, la presente invención utiliza dicha información de secuencia para preparar las moléculas de oligo o polinucleótido correspondientes, por ejemplo cebadores de hibridación, para cribar y seleccionar en formas físicas de bancos de biomoléculas, moléculas de secuencia de ADN que codifiquen las enzimas identificadas en el segundo grupo. Por lo tanto, en una realización preferida, la información de secuencia que se obtiene en la etapa c) puede utilizarse para la clonación de un gen correspondiente a partir de un ADN genómico o una biblioteca de ADNc y la expresión de la proteína en, por ejemplo, *E. coli*, *Saccharomyces cerevisiae*, *Pichia pastoris* u otras. En otra  
50 realización preferida de la presente invención, la información de secuencia que se obtiene en la etapa c) se utiliza para sintetizar *de novo* la transaminasa  $\omega$  selectiva de (R) deseada. En dicha realización preferida es posible utilizar la información de secuencia obtenida en la etapa c) para la síntesis de un gen, con por ejemplo, un uso de codones optimizado y estabilidad de ARNm, y clonar y expresar dicho gen. De acuerdo con dicha realización preferida, es posible utilizar como el al menos un banco de biomoléculas de la etapa a) un banco de biomoléculas que contiene información de secuencias de ADN o información de secuencias de aminoácidos. En el caso de que el banco de biomoléculas contenga información de secuencia de ADN, dicha información tiene que traducirse en información de secuencia de aminoácidos con el fin de procesarse de acuerdo con las etapas b) a c), en las que se usan secuencias de aminoácidos como secuencias de biomolécula incógnita y como motivos de secuencia de aminoácidos para su procesamiento de acuerdo con las etapas b) a c), que se llevan a cabo en el banco de  
60 secuencias de ADN utilizando secuencias de biomolécula incógnita y motivos de secuencia c1) a c6) traducidos inversamente en información de secuencia de ADN.

En otra realización preferida de la presente invención, la presente enseñanza se aplica en una realización B) en la etapa a) a un banco de biomoléculas que está presente en forma física de un banco genético, particularmente en una biblioteca genética tal como una biblioteca de ADNc, metagenómica o genómica, utilizando moléculas de ADN que codifican todas o al menos la parte característica de al menos una transaminasa  $\omega$  selectiva de (R), para cribar, identificar y seleccionar un grupo de primeras secuencias de biomolécula, en particular moléculas de secuencia de ADN, que, calculadas a nivel de aminoácidos, tienen un grado de identidad de secuencia de al menos un 20 %, preferentemente de un 32 % con la molécula de secuencia de ADN incógnita. La etapa c) posterior se lleva a cabo preferentemente en dicho grupo de primeras moléculas de ADN utilizando cebadores de moléculas de nucleótidos que identifican y seleccionan positiva y negativamente los motivos de secuencia c1) a c6) deseados.

En el contexto de la presente invención la expresión "motivo de secuencia" se refiere a las características selectivas específicas de la secuencia de aminoácidos de la supuesta transaminasa  $\omega$  selectiva de (R) según se identifica específicamente en la etapa c) del proceso reivindicado. En particular, el primer motivo de secuencia c1) es el motivo de secuencia identificado en la posición 95 a 97 siendo en este orden 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, de acuerdo con la presente invención, no debería estar presente en la secuencia de biomolécula seleccionada finalmente. El segundo motivo de secuencia c2) es el motivo de secuencia que se identifica en la posición 97 a 99 siendo en este orden una secuencia de aminoácidos Tyr Xaa Gln, siendo Xaa un aminoácido y en la región desde la posición 105 a 111, preferentemente en la posición 107 a 109 una secuencia de aminoácidos Arg Xaa Xa3, siendo Xaa un aminoácido y siendo Xa3 un aminoácido, siendo preferentemente His, y que tampoco debería estar presente en la secuencia de biomolécula que se selecciona finalmente. El tercer motivo de secuencia c3) es el motivo de secuencia caracterizado por la presencia de Thr en la posición 38, Lys en la posición 97 y en la posición 107 a 107 una secuencia de aminoácido Arg Xa4 Xa5, siendo Xa4 un aminoácido, siendo preferentemente Gly y siendo Xa5 un aminoácido, siendo preferentemente Tyr, cuyo motivo c3) no debería estar presente en la secuencia de biomolécula que se selecciona finalmente.

El cuarto motivo de secuencia c4) necesita que en la posición 95 esté presente un aminoácido distinto de Tyr, Arg, Lys, o en la posición 95 Tyr, pero que en la posición 97 no estén presentes ni Arg ni Lys, lo que reúne los requisitos para una supuesta transaminasa  $\omega$  selectiva de (R) preparada de acuerdo con la presente invención. El quinto motivo de secuencia c5) necesita que en la posición 40 no esté presente ni Lys ni Arg en la supuesta transaminasa  $\omega$  selectiva de (R) que se prepara e identifica finalmente. El sexto motivo de secuencia c6) necesita que en la región desde la posición 161 a 165 al menos esté presente una Lys, preferentemente una Lys, preferentemente en la posición 163 Lys en la  $\omega$ -TA selectiva de (R) que se prepara o identifica finalmente.

En el contexto de la presente invención, la identificación y localización de las posiciones de aminoácidos se determinan de la siguiente manera. Las primeras secuencias de biomolécula diana se alinean, preferentemente se alinean de manera múltiple entre sí y la secuencia de biomolécula incógnita. El alineamiento se puede hacer con programas informáticos de alineamiento convencionales, tales como STRAP, en particular ClustalW, preferentemente ClustalW3D. En otra realización también es posible alinear las secuencias por pares, es decir, cada primera secuencia de biomolécula diana con la secuencia de biomolécula incógnita. En una realización preferida las primeras secuencias de biomolécula diana se alinean a la BACT de *E. coli*. En otra realización preferida, las primeras secuencias de biomolécula se alinean a otra secuencia de biomolécula incógnita que se utiliza en la presente invención, por ejemplo la transaminasa  $\omega$  selectiva de (R).

La notación de las posiciones de aminoácidos, tal como se proporcionan en la presente invención, se determina por la posición del motivo de secuencia correspondiente en la secuencia de biomolécula incógnita que se utiliza como la posición convencional. El alineamiento tal como se describe anteriormente alinea las posiciones de aminoácidos correspondientes de las primeras secuencias de biomolécula diana a los motivos de secuencia presentes en la secuencia de biomolécula incógnita.

Como ejemplo, utilizando la BCAT de *E. coli* como referencia para la notación de la posición se utiliza la secuencia de aminoácidos de BCAT de *E. coli* conocida desde la posición 92 a 100, a saber TSAYIRPLI (SEC ID N°: 9) para identificar las posiciones 95 a 99 en la supuesta transaminasa  $\omega$ , que se analiza respecto de la ausencia de los motivos de secuencia c1), c2), c3) y de la presencia de c4). La secuencia de aminoácidos conocida desde la posición 35 a 42, a saber VFEGIRCY (SEC ID N°: 10) de la BCAT de *E. coli* marca la posición de la G en la posición 38 para los motivos de secuencia c3) y c5). La secuencia de aminoácidos DVGMGVNP en la secuencia de aminoácidos BCAT de *E. coli* (SEC ID N°: 11) marca las posiciones 104 a 111 para el motivo de secuencia c3) en las posiciones 105 a 111. La secuencia de aminoácidos PTAAKAGGN desde las posiciones 159 a 167 de la BCAT de *E. coli* conocida (SEC ID N°: 12) marca la posición 163 como una K en el motivo de secuencia c6).

Por lo tanto, en una realización preferida de la presente invención la biomolécula es una proteína y la secuencia de biomolécula es una secuencia de aminoácidos. En consecuencia, en una realización preferida, el banco de biomoléculas es un banco, en particular una base de datos, cuyo banco es un banco con información recogida acerca de varias proteínas, en particular información de secuencia de aminoácidos. Los bancos de bases de datos preferidos son el banco de datos de proteínas del NCBI, el UniProtKB/SwissProt y el banco de datos UniProtKB/TrEMBL.

En una realización preferida también es posible que la biomolécula sea una molécula de ADN y que la secuencia de biomolécula sea una secuencia de ADN. En consecuencia, en una realización preferida, el banco de biomoléculas en un banco, en particular una base de datos, con información recogida acerca de varias secuencias de polinucleótidos, en particular de secuencias de ADN.

5 En una realización preferida, la presente invención utiliza un proceso de acuerdo con lo anterior, en el que el banco de biomoléculas es una base de datos de biomoléculas y la base de datos de biomoléculas se criba en la etapa b) con un programa de alineamiento de secuencias de biomoléculas, 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 criba en la etapa b) con una secuencia de aminoácidos, si la base de datos es una base de datos de secuencias de aminoácidos, o con una secuencia de ADN, si la base de datos de biomoléculas es una base de datos de secuencias de ADN.

15 La invención prevé en la etapa d) proporcionar finalmente una biomolécula, preferentemente una molécula de ADN o una molécula de secuencia de aminoácidos, es decir, una proteína, o mediante la síntesis *de novo* de la transaminasa o mediante aislamiento de un banco genético físico, polinucleótidos que codifican la transaminasa deseada con la ayuda de cebadores, que se definen basándose en las segundas secuencias de biomoléculas. Los polinucleótidos que se obtienen, en particular moléculas de secuencias de ADN, se utilizan para su expresión en condiciones adecuadas en un medio de cultivo adecuado para expresar la transaminasa deseada.

20 La presente invención también se refiere a un proceso para el cribado y la identificación de una transaminasa  $\omega$  selectiva de (R), que comprende las etapas a) a c) identificadas anteriormente, en particular proporcionando al menos una secuencia de biomolécula incógnita de al menos una transaminasa  $\omega$  selectiva de (R) y al menos un banco de biomoléculas, cribando el banco de biomoléculas con la secuencia de biomolécula incógnita para identificar un grupo de primeras secuencias de biomolécula diana, en que las primeras secuencias de biomolécula tienen un grado de identidad de secuencia de al menos un 20 %, preferentemente un 25 %, preferentemente un 32 % respecto de la secuencia de biomolécula incógnita, que se calcula a nivel de aminoácidos, seleccionando en el grupo de primeras secuencias de biomolécula diana secuencias de biomoléculas que no comprenden, a nivel de aminoácidos, ninguno de los motivos de secuencia c1) a c3) siendo c1) en la posición 95 a 97 la secuencia de aminoácidos Tyr Xa1 Xa2, siendo Xa1 un aminoácido Ile, Val, Leu, Met, Phe, y siendo Xa2 un aminoácido Arg o Lys o siendo c2) en la posición 97 a 99 una secuencia de aminoácidos Tyr Xaa Gln, siendo Xaa un aminoácido y en la región desde la posición 105 a 111, preferentemente 107 a 109, una secuencia de aminoácidos Arg Xaa Xa3, siendo Xa3 preferentemente un aminoácido, preferentemente His o siendo 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, siendo preferentemente Gly, y siendo Xa5 un aminoácido, siendo preferentemente Tyr, y que no comprende los motivos de secuencia c4), c5) y c6) siendo c4) a nivel de aminoácidos en la posición 95 un aminoácido distinto de Tyr, Arg, Lys o en la posición 95 Tyr, pero en la posición 97 ni Arg ni Lys y siendo c5) en la posición 40 ni Lys ni Arg y siendo c6) en la región desde la posición 161 a 165, preferentemente en la posición 163, Lys para identificar un grupo de segundas secuencias de biomoléculas, que son secuencias de biomolécula de una  $\omega$ -TA selectiva de (R).

40 En una realización preferida más de la presente invención la presente enseñanza proporciona transaminasas  $\omega$  selectivas de (R) que se pueden obtener de acuerdo con los procesos de preparación de la presente invención. En particular, la presente invención proporciona proteínas y secuencias de ADN, en particular moléculas de ADN, que codifican dicha proteína de *Mesorhizobium loti* y *Aspergillus terreus* como se identifican en las SEC ID N°: 1 y 2 para *Mesorhizobium loti* y 3 y 4 para *Aspergillus terreus*. Las transaminasas  $\omega$  selectiva de (R) que pueden obtenerse o prepararse de acuerdo con la presente invención, por ejemplo las que se identifican en las SEC ID N°: 1 y 3, se pueden utilizar en una reacción de transaminación, en particular se puede utilizar en un proceso para la preparación de una amina quiral ópticamente activa, preferentemente un enantiómero (R) de una amina quiral, que comprende hacer reaccionar al menos un aceptor de amino y al menos un donante de amino con la transaminasa  $\omega$  selectiva de (R) de acuerdo con la presente invención, en particular de SEC ID N°: 1 o 3, y obtener la amina quiral ópticamente activa. El proceso para la preparación de una amina quiral ópticamente activa puede ser una síntesis asimétrica que utiliza un compuesto que contiene un grupo ceto preferentemente una cetona y un donante de amino. El método de preparación puede ser utilizando una reacción de resolución cinética. Siendo el donante de amino preferentemente una mezcla racémica de aminas y un aceptor de amino, preferentemente cetonas, en forma de aductos.

55 Una amina quiral ópticamente activa se puede sintetizar utilizando una transaminasa  $\omega$  selectiva de (R) de la presente invención en una síntesis asimétrica partiendo de cetonas el grado preferido de conversión en la amina quiral ópticamente activa deseada, es decir, el enantiómero (R) es al menos un 70 %, 80 %, 90 %, 95 %, 98 %, 99 %, 99.5 % y más preferentemente el 100 %.

60 La transaminasa  $\omega$  selectiva de (R) de la presente invención se puede utilizar en una reacción de resolución cinética partiendo de aminas racémicas al grado preferido de conversión en la amina quiral ópticamente activa, preferentemente el enantiómero (S), es al menos un 30, 40, 45, 46, 47, 48, 49, en particular el 50 %.

65 Las concentraciones para el análisis de la pureza óptica y la conversión se pueden determinar por ejemplo utilizando HPLC, electroforesis de capilaridad (CE), cromatografía de gas (GC) o métodos foto o fluorimétricos.



Puede llevarse a cabo un proceso para la preparación de una amina quiral ópticamente activa, comprendiendo dicho método hacer reaccionar un compuesto aceptor de amina que comprende un grupo ceto y una mezcla racémica de una amina en presencia de una transaminasa, en particular una transaminasa  $\omega$  selectiva de (R), preferentemente de acuerdo con la presente invención, para obtener un enantiómero (S) de la amina quiral.

5 Puede llevarse a cabo un proceso para la preparación de una amina quiral ópticamente activa, en particular un enantiómero (R) de dicha amina, comprendiendo ese proceso hacer reaccionar un compuesto aceptor de amino que comprende un grupo ceto y un donante de amino en presencia de una transaminasa  $\omega$  selectiva de (R), en particular que se obtiene de acuerdo con la presente invención, para obtener un enantiómero (R) de la amina.

10 Un aceptor de amina es una molécula capaz de aceptar un grupo amino transferido desde un donante de amino por una transaminasa, en particular una transaminasa  $\omega$ . El aceptor de amino contiene preferentemente una funcionalidad cetona.

15 El aceptor de amino se puede seleccionar de entre el grupo que consiste en ácido fenilpirúvico, una sal del mismo, ácido pirúvico, una sal del mismo, acetofenona, 2-cetoglutarato, 3-oxobutirato, 2-butanona, 3-oxopirrolidina (3-OP), 3-piridilmetilcetona (3-PMK), éster etílico del ácido 3-oxobutírico (3-OBEE), éster metílico del ácido 3-oxopentanoico (3-OPME), N-1-boc-oxo-piperidinona, 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.

Un donante de amino es una molécula capaz de proporcionar un grupo amino a un aceptor de amino utilizando una transaminasa, en particular una transaminasa  $\omega$ . El donante de amino puede ser preferentemente una amina o un aminoácido.

25 El donante de amino se puede seleccionar de entre el grupo que consiste en  $\beta$ -alanina, alanina,  $\alpha$ -metilbencilamina ( $\alpha$ -MBA), glutamato, fenilalanina, glicina, 3-aminobutirato, isopropilamina, 2-aminobutano y  $\gamma$ -aminobutirato o una sal, por ejemplo un cloruro, de cualquiera de los mismos. En una realización particularmente preferida, el producto cetónico que se obtiene puede ser ácido fenilpirúvico, una sal del mismo, ácido pirúvico una sal del mismo, ácido glicoxílico, una sal del mismo, acetofenona, 2-cetoglutarato, acetona, 3-oxobutirato, 2-butanona, 2-oxopirrolidina (3-OP), 3-piridilmetilcetona (3-PMK), éster etílico del ácido 3-oxobutírico (3-OBEE), éster metílico del ácido 3-oxopentanoico (3-OPME), N-1-boc oxopiperidinona y N-1-boc-3-oxopirrolidina (B3OP) o una sal, por ejemplo un cloruro, de uno cualquiera de los mismos.

35 Se puede llevar a cabo un proceso para la preparación de una amina quiral ópticamente activa que se selecciona de entre el grupo de aminas que tienen un grupo amino ópticamente activo, en particular aminas con grupos alquilo, grupos alquilo ramificados o grupos arilalquilo. En particular, estas aminas, en particular aminas mono o bicíclicas, son en particular aminas cíclicas de 5 a 6 miembros o hidrocarburos heterocíclicos O-, S- o N-sustituídos o aminas aromáticas, en particular aminas aromáticas sustituidas con alquilo o alcoxi. En una realización preferida, las aminas quirales obtenidas se seleccionan entre el grupo que consiste en fenilalanina, alanina, 3-aminopiperidina, alquil-3-amino-butanoatos, 3-aminopirrolidina (3-AP), 3-piridil-1-etil-amina (3-PEA), N-1-boc-3-aminopirrolidina (B3AP), N-1-boc-3-aminopiperidina (B3APi), 1-Cbz-3-aminopiperidina (C3APi), 1-Cbz-3-aminopirrolidina (CEAP), éster etílico del ácido 3-aminobutírico (3-ABEE), éster metílico del ácido 3-aminopentanoico (3-APME),  $\alpha$ -metilbencilamina ( $\alpha$ -MBA), 1-aminotetralina, 3,4-dimetoxi fenil acetona,  $\alpha$ -metil-4-(3-piridil)-butanamina,  $\gamma$ -aminobutirato, glutamato, isopropilamina,  $\beta$ -aminobutirato, sec-butilamina, metoxiisopropilamina, derivados de 3-aminopirrolidina, 1-N-Boc-3-aminopiperidina, cefalosporina y derivados de la cefalosporina.

Puede hacerse reaccionar 3OP con una transaminasa  $\omega$  selectiva de (R) y un donante de amino para obtener (R) 3AP ópticamente activa.

50 Puede hacerse reaccionar 3-PMK con una transaminasa  $\omega$  selectiva de (R) y un donante de amino para obtener (R) 3-PEA ópticamente activa.

Puede hacerse reaccionar 3-OBEE con una transaminasa  $\omega$  selectiva de (R) y un donante de amino para obtener (R) 3-ABEE ópticamente activa.

55 Puede hacerse reaccionar 3-OPME con una transaminasa  $\omega$  selectiva de (R) y un donante de amino para obtener (R) 3-APME ópticamente activa.

60 Puede hacerse reaccionar B3OP con una transaminasa  $\omega$  selectiva de (R) y un donante de amino para obtener (R) B3AP ópticamente activa.

Puede hacerse reaccionar B3OPi con una transaminasa  $\omega$  selectiva de (R) y un donante de amino para obtener (R) B3APi ópticamente activa.

65

Puede hacerse reaccionar C3OPi con una transaminasa ω selectiva de (R) y un donante de amino para obtener (R) C3APi ópticamente activa.

5 Puede hacerse reaccionar C3OP con una transaminasa ω selectiva de (R) y un donante de amino para obtener (R) C3AP ópticamente activa.

Puede hacerse reaccionar acetofenona con una transaminasa ω selectiva de (R) y un donante de amino para obtener (R) α-MBA ópticamente activa.

10 Puede hacerse reaccionar un aceptor de amino, en particular grupos cíclicos, aromáticos o hidrocarburo heterocíclicos S-, O-, o N-sustituidos de 5 a 6 miembros que contienen un grupo oxo mono o bicíclico en particular, grupos aromáticos sustituidos con alquilo o alcoxi, con un donante de amino y una transaminasa ω selectiva de (R) para obtener aminas, en particular aminas mono o bicíclicas, en particular aminas cíclicas de 5 a 6 miembros o hidrocarburos heterocíclicos S-, O-, o N-sustituidos o aminas aromáticas, en particular aminas aromáticas sustituidas con alquilo o alcoxi, en particular en forma (R).

15 Pueden hacerse reaccionar el aceptor de amino y el donante de amino con la transaminasa en medio acuoso, por ejemplo, tampón fisiológico. En una realización particularmente preferida la reacción de transaminación se lleva a cabo a un pH en el intervalo de 5 a 9, en particular de 7 a 8,5. En una realización particular preferida, la reacción se lleva a cabo a un intervalo de temperaturas de 10 a 65 °C, preferentemente de 20 a 50 °C, en particular de 18 a 25 °C, preferentemente a temperatura ambiente o a 34 °C a 39 °C, en particular a 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 realización preferida de la presente invención la actividad enzimática puede ser de 1 a 20.000 μmol/min.

20 Puede llevarse a cabo un proceso para el análisis de una transaminasa, en particular para la caracterización de las propiedades de una transaminasa, que comprende las siguientes etapas:

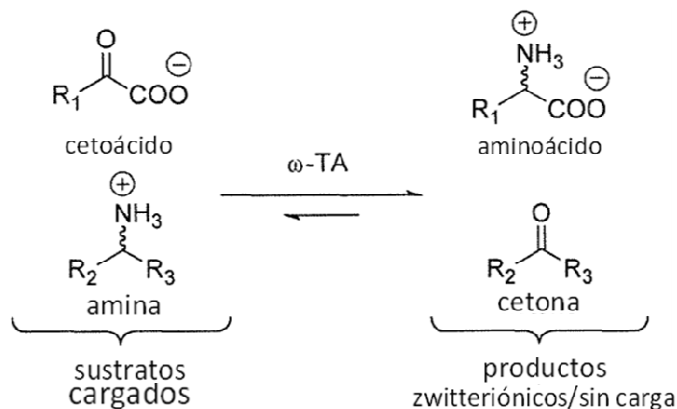
30 i. proporcionar un aceptor de amino cargado, un donante de amino cargado y una transaminasa, preferentemente una transaminasa ω, más preferentemente una transaminasa ω selectiva de (R) que se obtiene de acuerdo con la presente invención

35 ii. hacer reaccionar el aceptor de amino y el donante de amino con la transaminasa en un medio de reacción, y de esta manera

iii. determinar la conductividad del medio de reacción en un primer grupo de condiciones de reacción y

40 iv. posteriormente a la etapa iii) determinar la conductividad del medio de reacción en un segundo grupo de condiciones de reacción, para obtener al menos dos valores de conductividad que reflejen las propiedades de la transaminasa.

45 En el curso de una reacción catalizada por una transaminasa, preferentemente una ω-TA, de los sustratos de donante de amino cargado, preferentemente una amina, y el aceptor de amino, preferentemente un cetoácido, por ejemplo un piruvato, la conductividad disminuye ya que se forman una cetona no cargada y el componente amino zwitteriónico, preferentemente aminoácido, por ejemplo alanina.



50 El proceso de análisis permite una medición simple del progreso de la reacción. Preferentemente, un tampón de baja conductividad, particularmente el tampón CHES (ácido N-ciclohexil-2-aminoetanosulfónico) de baja conductividad es el más adecuado para evitar una conductividad inicial demasiado alta. Preferentemente, se puede hacer una calibración del proceso de conductividad mediante la simulación de diferentes conversiones. Como ejemplo, para la

pareja de sustrato convencional  $\alpha$ -metilbencilamina y piruvato, una conversión de 1 mM corresponde a una carga de 44  $\mu$ S. Una validación del presente proceso comparando las tasas de reacción medidas mediante electroforesis de capilaridad dieron como resultado una conformidad excelente. Los extractos celulares no interfieren significativamente en el presente proceso. Ya que el piruvato es el aceptor de amino común de virtualmente todas las  $\omega$ -TA, el presente proceso puede utilizarse para investigaciones de la actividad de transaminasa hacia diferentes donantes de amino. Además, también puede obtenerse información acerca de la enantioselectividad de la enzima.

Puede llevarse a cabo un proceso para analizar una transaminasa, en particular para la caracterización de las propiedades de una transaminasa, que permite analizar la actividad de la transaminasa dependiendo de, por ejemplo, el valor del pH o la temperatura del medio de reacción o permite analizar la estabilidad de la reacción, el efecto de los aditivos o las composiciones tampón.

De acuerdo con el proceso de análisis se puede llevar a cabo una primera medición de la conductividad del medio de reacción en un primer grupo de condiciones de reacción y posteriormente se puede llevar a cabo al menos una segunda medición de la conductividad del medio de reacción con el fin de ser capaces de comparar ambos valores de conductividad obtenidos y extraer conclusiones acerca de las actividades y propiedades de la transaminasa ensayada. Una disminución de la conductividad muestra en una reacción de transaminasa de acuerdo con la presente invención la actividad de dicha transaminasa. El reconocimiento de una disminución reducida, de una disminución acelerada o de ninguna disminución de la conductividad permite extraer conclusiones acerca de las propiedades de la transaminasa.

Un grupo de condiciones de reacción puede ser preferentemente un grupo de condiciones cuyas condiciones se seleccionan preferentemente entre el grupo que consiste en temperatura, valor de pH y composición del medio de reacción, preferentemente condiciones de reacción tal como se identificaron anteriormente, a excepción de la concentración de aductos y productos. En una realización de la presente invención se mantiene constante el grupo de condiciones de reacción durante el tiempo de reacción. En otra realización de la invención, el grupo de condiciones de reacción puede ser diferente durante el tiempo de reacción.

Puede llevarse a cabo un proceso para el análisis de una transaminasa, en el que el medio de reacción es un tampón de baja conductividad. En una realización particularmente preferida del proceso, el aceptor de amino cargado es piruvato.

El proceso para el análisis de una transaminasa puede usarse posteriormente al proceso para la preparación de una transaminasa  $\omega$  selectiva de (R) de la presente invención y extiende la enseñanza de la presente invención no solo a proporcionar nuevas y ventajosas transaminasa  $\omega$  selectiva de (R) sino también a permitir la determinación de sus características.

Las realizaciones preferidas adicionales de la presente invención son la materia objeto de las reivindicaciones dependientes.

La presente invención se ilustra en más detalle en los siguientes ejemplos y en el listado de secuencias adjunto.

La SEC ID N°: 1 muestra la secuencia de aminoácidos completa (ORF) de la  $\omega$ -TA selectiva de (R) de *Mesorhizobium loti*,

La SEC ID N°: 2 muestra la secuencia de ADN que codifica la SEC ID N°: 1,

La SEC ID N°: 3 muestra la secuencia de aminoácidos completa (ORF) de la  $\omega$ -TA selectiva de (R) de *Aspergillus terreus*,

La SEC ID N°: 4 muestra la secuencia de ADN que codifica la SEC ID N°: 3,

La SEC ID N°: 5 muestra la secuencia de aminoácidos completa (ORF) de la transaminasa  $\omega$  selectiva de (R) de *Mycobacterium aurum*,

La SEC ID N°: 6 muestra la secuencia de ADN que codifica la SEC ID N°: 5,

La SEC ID N°: 7 muestra la secuencia de aminoácidos completa (ORF) de la transaminasa  $\omega$  selectiva de (R) de *Arthrobacter sp.*,

La SEC ID N°: 8 muestra la secuencia de ADN que codifica la SEC ID N°: 7,

La SEC ID N°: 9 muestra el motivo de secuencia del BCAT de *E. coli* que se utiliza para las determinaciones de las posiciones 95 a 99,

## ES 2 562 458 T3

- La SEC ID Nº: 10 muestra el motivo de secuencia del BCAT de *E. coli* que se utiliza para las determinaciones de la posición 38,
- 5 La SEC ID Nº: 11 muestra el motivo de secuencia del BCAT de *E. coli* que se utiliza para las determinaciones de la posición 107,
- La SEC ID Nº: 12 muestra el motivo de secuencia del BCAT de *E. coli* que se utiliza para las determinaciones de la posición 163,
- 10 La SEC ID Nº: 13 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Penicillium chrysogenum*,
- La SEC ID Nº: 14 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID Nº: 13,
- 15 La SEC ID Nº: 15 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Aspergillus niger*,
- La SEC ID Nº: 16 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID Nº: 15,
- La SEC ID Nº: 17 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Aspergillus oryzae*,
- 20 La SEC ID Nº: 18 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID Nº: 17,
- La SEC ID Nº: 19 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Aspergillus fumigatus*,
- 25 La SEC ID Nº: 20 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID Nº: 19,
- La SEC ID Nº: 21 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Neosartorya fischeri*,
- La SEC ID Nº: 22 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID Nº: 21,
- 30 La SEC ID Nº: 23 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Gibberella zeae*,
- La SEC ID Nº: 24 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID Nº: 23,
- 35 La SEC ID Nº: 25 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Hyphomonas neptunium*,
- La SEC ID Nº: 26 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID Nº: 25,
- 40 La SEC ID Nº: 27 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de de *Mesorhizobium loti* MAFF303099,
- La SEC ID Nº: 28 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID Nº: 27,
- 45 La SEC ID Nº: 29 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Roseobacter sp.*,
- La SEC ID Nº: 30 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID Nº: 29,
- La SEC ID Nº: 31 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Marinomonas sp.*,
- 50 La SEC ID Nº: 32 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID Nº: 31,
- La SEC ID Nº: 33 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Rhizobium etli*,
- 55 La SEC ID Nº: 34 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID Nº: 33,
- La SEC ID Nº: 35 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Rhodoferrax ferrireducens*,
- 60 La SEC ID Nº: 36 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID Nº: 35,
- La SEC ID Nº: 37 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Jannaschia sp.*,
- La SEC ID Nº: 38 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID Nº: 37,
- 65 La SEC ID Nº: 39 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Labrenzia alexandrii*,

La SEC ID N°: 40 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID N°: 39,

La SEC ID N°: 41 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Burkholderia sp.*,

5 La SEC ID N°: 42 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID N°: 41,

La SEC ID N°: 43 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Burkholderia cenocepacia*,

10 La SEC ID N°: 44 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de 43,

La SEC ID N°: 45 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de proteobacteria alfa,

15 La SEC ID N°: 46 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID N°: 45,

La SEC ID N°: 47 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de proteobacteria gamma,

La SEC ID N°: 48 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID N°: 47,

20 La SEC ID N°: 49 muestra la secuencia de ADN que codifica la  $\omega$ -TA selectiva de (R) de *Mycobacterium vanbaalenii* y

La SEC ID N°: 50 muestra la secuencia de aminoácidos completa (ORF) de la SEC ID N°: 49.

25 La Figura 1 muestra un cromatograma de B3APi obtenida por síntesis convencional.

La Figura 2 muestra un cromatograma de B3APi obtenida mediante la síntesis asimétrica de acuerdo con la invención.

30 La Figura 3 muestra un cromatograma de C3AP obtenida mediante la síntesis asimétrica de acuerdo con la invención.

La Figura 4 muestra un cromatograma de MPPA obtenida mediante la síntesis asimétrica de acuerdo con la invención.

35 La Figura 5 muestra un cromatograma de B3AP obtenida mediante la síntesis asimétrica de acuerdo con la invención.

## Ejemplos

40 Ejemplo 1 - Identificación de transaminasas  $\omega$  selectivas de (R)

La secuencia de aminoácidos de la  $\omega$ -TA selectiva de (R) de *Mycobacterium aurum* que se encuentra en el documento EP 1 038 953 A1 (SEC ID N°: 5) y la secuencia de aminoácidos de la  $\omega$ -TA selectiva de (R) de *Arthrobacter sp.* tal como se encuentra en el documento EP 0 987 332 A1 (SEC ID N°: 7), se utilizan como secuencias de biomolécula incógnita. El banco de biomoléculas que se utiliza en este ejemplo es el banco de datos de proteína de pubmed del NCBI ((<http://www.ncbi.nlm.nih.gov/pubmed> (13 Julio 2009)).

50 Utilizando un cribado por BLAST con la secuencia de aminoácidos de la  $\omega$ -TA de *M. aurum* o *Arthrobacter sp.* (SEC ID N°: 5 o 7) como incógnita utilizando parámetros convencionales (matriz de puntuación BLOSUM62, tamaño de palabra: 3, costes de hueco: existencia - 11, extensión -1) se han identificado un primer grupo de 100 secuencias de aminoácidos distintas de diferentes organismos, que tienen todas un grado mínimo de identidad de secuencia del 30 % con la secuencia incógnita.

55 Para el programa BLAST se han utilizado las "secuencias de proteína no redundantes (nr)" el 13 de julio de 2009 ([http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST\\_PROGRAMS=blastp&PAGE\\_TYPE=BlastSearch&SHOW\\_DEFAULTS=on&LINK\\_LOC=blasthome](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 distintos que representan las primeras secuencias de biomolécula diana, las secuencias se han cribado e identificado buscando un motivo de secuencia que no muestre los motivos de secuencia c1), c2) o c3) y que muestre los motivos de secuencia c4), c5) y c6). Se pudieron identificar 21 ORF y se enumeran en la Tabla 1 posterior.

60

Tabla 1

Nº	Organismos fuente	% Identidad con		Nº de ident. genética/ Nº ident. Proteico (base de datos NCBI)
		AS- $\omega$ TA SEC ID 7	MA- $\omega$ -TA SEC ID 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
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>Rhodofera 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>Proteobacteria alfa</i> HTCC2255	36	34	ZP_01448442
21	<i>Proteobacteria gamma</i>	27	26	
	<i>Mycobacterium aurum</i>	49	100	SEC ID 5
	<i>Arthrobacter</i> sp.	100	49	SEC 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 Las  $\omega$ -TA de *Mycobacterium vanbaalenii*, entrada 9 en la Tabla 1 (SEC ID Nº: 49 y 50), de *Aspergillus terreus*, entrada 1 en la Tabla 1 (SEC ID Nº: 3 y 4), y *Mesorhizobium loti*, entrada 11 de la Tabla 1 (SEC ID Nº: 1 y 2), se han obtenido y utilizado en forma adaptada al codón de uso (para *E. coli*) para expresar las enzimas en *Escherichia coli*. La transaminasa de *Mycobacterium vanbaalenii* se denomina a continuación Va-TA, la transaminasa de *Aspergillus terreus* Ate-TA y la de *Mesorhizobium loti* Malo-TA.

10

## 2.2 Ensayo de acetofenona

La Mva-TA y Ate-TA convertían la (R)- $\alpha$ -MBA al menos 100 veces más rápido que al enantiómero (S) en un ensayo de acetofenona con 2,5 mM de amina y 2,5 mM de piruvato a pH 7,5 y 30 °C.

15

Ensayo: El aumento de absorción de la acetofenona que se forma durante la reacción se controló a 245 nm.

20

La Mlo-TA no convertía ni la (R)- ni la (S)-  $\alpha$ -MBA.

A continuación, se han ensayado aminas adicionales en presencia de piruvato o de  $\alpha$ -cetoglutarato como receptores de amino (10 mM de amina, 10 mM de piruvato, 0,1 mM de PLP, tampón fosfato pH 7,5, incubación durante 24 h a 30 °C, se analizó por cromatografía en capa fina).

25

Pudo observarse una conversión para 2-aminoheptano, 2-aminopentano, 1,3-dimetilbutilamina y 4-fenil-2-aminobutano. También se podía detectar una mínima conversión de isopropilamina.

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

Por lo tanto, se demostró que las tres proteínas eran  $\omega$ -TA.

5 En particular, utilizando (R)- y (S)-2-aminohexano como sustrato, solo se convertía significativamente el enantiómero (R). Por lo tanto, también la Mlo-TA es una  $\omega$ -TA selectiva de (R). No se apreció actividad DATA (D-aminoácido transaminasa) ni BCAT (aminotransferasa de cadena ramificada) para las tres proteínas.

### 10 2.3 Ensayo de conductividad

También se utilizaron como sustrato la 1-N-boc-3-aminopirrolidina (B3AP), 1-N-boc-3-aminopiperidina (B3APi) y 1-Cbz-3-aminopiperidina (C3APi) para determinar las actividades relativas de estas sustancias contra el sustrato modelo  $\alpha$ -MBA.

15 Durante la reacción de la amina y el piruvato (a pH 7,5 ambos sustratos están cargados) a alanina y cetona (la cetona no tiene carga, la alanina es un compuesto zwitteriónico y no contribuye a la conductividad) el control de las cinéticas de conductividad permite deducir las tasas de conversión.

20 Antes del inicio de la reacción, se llevó a cabo una calibración determinando diferentes conversiones dependiendo de diversas concentraciones de alanina, piruvato, cetona y amina.

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

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

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

### 35 2.4 Determinación de la enantioselectividad mediante síntesis asimétrica de aminas con las $\omega$ -TA selectivas de (R) tanto de *Aspergillus terreus* como de *Mesorhizobium loti*

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

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

45 Como donante de amino se utilizó alanina con cien veces de exceso. La conversión no se determinó exactamente, sino que se estimó aproximadamente. En experimentos adicionales se podrían desarrollar diferentes métodos para aumentar la conversión (PDC). Se obtuvieron enantioselectividades de altas a excelentes con ambas transaminasas, excepto con C3APi en la que solamente 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

### 50 Métodos:

Las transaminasas se expresaron en *E. coli* BL21.

El medio de cultivo (50 ml de medio LD-Amp) se indujo con un 0,2 % de ramnosa, cuando se obtuvo una densidad óptica (DO) de 0,5, y se cultivó durante 12 horas a 25 °C. Posteriormente, se lavaron las células 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 µl.

5 Para la biocatálisis se añadieron 5 µl de una solución de cetona 500 mM en DMSO, 22 mg de D-alanina y 10 µl de PLP a estas alícuotas. La mezcla de reacción se enrasó a un volumen total de 500 µl con tampón de fosfato sódico 50 mM (pH 7,5). Las mezclas de reacción se incubaron durante una noche a 25 °C y 500 rpm. Para la determinación del exceso enantiomérico por medio de CE se añadieron 200 µl de NaOH 1 N a 400 µl de la muestra, se extrajo con 10  
400 µl de diclorometano y se separó la fase orgánica. La fase orgánica se extrajo entonces con 100 µl de un tampón de fosfato de trietilamonio 5 mM (pH 3). Posteriormente, la fase acuosa obtenida se inyectó en la CE.

El programa de separación:

- 15 CE-Capilar 30 cm, 50 µm de diámetro interior, temperatura 15 °C
- aclarado con tampón de fosfato de trietilamonio pH 3, 1 min, 206,8 kPa
  - aclarado con ciclodextrina al 5 % altamente sulfatada (HsaCD o HSyCD), 1 min, 68,9 kPa
  - 20 - inyección: 5 - 10 s, 6,9 kPa
  - inmersión en agua
  - 25 - separación: voltaje 10 o 15 kV
  - detección: MPPA, C3AP y C3APi a 200 nm; B3AP, B3APi a 190 nm

Condiciones de separación:

30

Tabla 3

	Segmento de separación [cm]	Selector quirál	Tiempo de migración [min]		Voltaje durante la separación
			S-Amina	R-Amina	
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

#### 35 4.1 Expresión y purificación de las transaminasas

La fases de lectura abiertas optimizadas por codones codifican las proteínas con números de entrada 1, 2, 3, 4, 8, 9, 11, 13, 14, 15, 17, 18 y 21 de la Tabla 1 se insertaron en pGASTON entre los sitios de restricción de NdeI y BamHI utilizando una estrategia de clonación independiente de ligadura. Las ORF optimizada por codones que codifican el resto de proteínas se ordenaron ya subclonadas 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 se incubaron inicialmente a 37 °C con un agitador giratorio hasta que la DO<sub>600</sub> llegó a 0,7. Las células se indujeron entonces mediante la adición de 0,2 % de ramnosa (pGASTON) o 0,1 mM de IPTG (pET-22b), respectivamente, y al mismo tiempo se bajó la temperatura de incubación a 20 °C. Tras la inducción se continuó con la incubación durante otras 45 20 h. Se retiraron alícuotas en varios puntos de tiempo tras la inducción para seguir la expresión.

El aglomerado celular (~3 g en peso húmedo) se lavó dos veces con tampón fosfato (pH 7,5, 50 mM), que contenía 0,1 mM de PLP a 4 °C. Tras la destrucción (en prensa francesa) la suspensión celular se centrifugó (5000 x g, 30 min) y el sobrenadante resultante se pasó a través de un filtro de 0,5 µm antes de la cromatografía. Se llevó a cabo la cromatografía utilizando un Purificador ÄKTA (GE Healthcare). El extracto celular filtrado se aplicó a una columna de 5 ml de IMAC Sepharose™ 6 Fast Flow (GE Healthcare). Se lavó la columna con un caudal de 5 ml min<sup>-1</sup> con un volumen de 10 columnas de tampón fosfato 50 mM, pH 7,5, que contenía 300 mM de NaCl, 0,1 mM de PLP, y 30 mM de imidazol (para evitar la unión no específica) y la actividad ATA se eluyó con volúmenes de 10 columnas de tampón fosfato (pH 7,5, 50 mM), que contenía 300 mM de NaCl, 0,1 mM de PLP y 300 mM de imidazol (caudal de 5



ml min<sup>-1</sup>). Las fracciones que contenían la actividad se agruparon y desalaron por medio de cromatografía en gel con un tampón Tricina 20 mM pH 7,5 que contenía 0,01 mM de PLP. Las enzimas purificadas se almacenaron a 4 °C.

- 5 La cantidad de cada proteína purificada a partir de aproximadamente 3 g de células (peso húmedo) se dan en la Tabla 4 posterior.

Tabla 4

Entrada	Fuente enzimática	Rendimiento en proteína tras la 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
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>Marimonas sp.</i>	23,7
14	<i>Rhizobium etli</i>	6,5
15	<i>Rhodofera 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

- 10 4.2 Caracterización de la especificidad de sustrato de la  $\omega$ -TA selectiva de (R)

Para determinar la actividad hacia la  $\alpha$ -metilbencil amina ( $\alpha$ -MBA) en el cribado inicial de las proteínas expresadas, se utilizó un ensayo basado en acetofenona: se hizo reaccionar una solución de (R)- o (S)-  $\alpha$ -MBA 2,5 mM y piruvato en presencia de la enzima purificada y se correlacionó el aumento de absorbancia a 245 nm con la formación de acetofenona. Se controlaron las conversiones de las aminas 2-aminohexano, 4-fenil-2-aminobutano y 1-N-Boc-3-aminopirrolidina utilizando un ensayo de conductividad: Se hizo reaccionar una solución que contenía amina 10 mM y piruvato en presencia de la amina transaminasa purificada y se relacionó la disminución de la conductividad con la conversión del sustrato.

- 20 Para investigar la actividad DATA y BCAT se midió el descenso de NADH espectrofotométricamente a 340 nm utilizando ensayos de acoplamiento a deshidrogenasa: se hizo reaccionar una solución de ácido  $\alpha$ -cetoglutarico 5 mM y D-alanina en presencia de la transaminasa purificada, 1 U/ml de lactato deshidrogenasa y 0,5 de NADH para medir la actividad de DATA. Se utilizó una solución que contenía ácido 3-metil-2-oxobutírico 5 mM y L-glutamato, 10 mM de cloruro amónico, 1 U/ml de glutamato deshidrogenasa y 0,5 mM de NADH para medir la actividad de BCAT.

- 25 Todas las reacciones tuvieron lugar en tampón de Tricina 20 mM pH 7,5 que contenía 0,01 mM de PLP. El pH del tampón se ajustaba con 1,8-Diazabiciclo [5.4.0] undecil-7-eno.

Los resultados se dan en la Tabla 5 a continuación.

Tabla 5: Actividades específicas para varios sustratos.

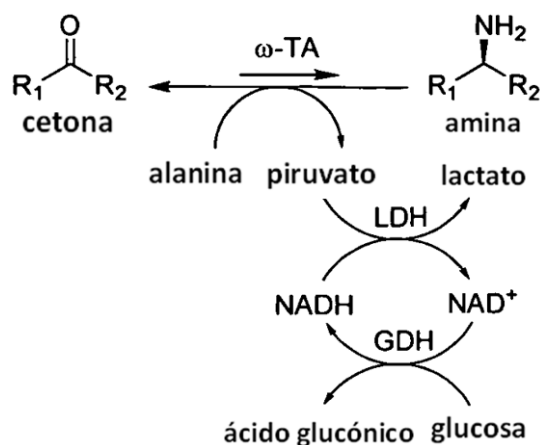
Sustratos	piruvato 1		piruvato 2		piruvato 3		piruvato 4		2KG D-Ala	MOB L-Glu
	Entrada	R	S	R	S	R	S	R		
1	15,2	<0,001	2,91	<0,001	9,7	<0,001	0,031	<0,001	<0,001	0,003
2	1,3	<0,001	1,1	0,044	5,6	<0,001	0,264	<0,001	<0,001	<0,001
3	- a)	-	-	-	-	-	-	-	-	-
4	3,7	0,001	1,4	0,023	5,2	0,002	0,051	0,002	<0,001	<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
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** -  $\alpha$ -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 con la Tabla 1. Todas las mediciones se hicieron al menos por duplicado. La desviación de mediciones únicas del valor medio era < 10 %.

**a)** No fue posible la medición ya que el rendimiento de proteína durante la expresión era muy baja/ la proteína era inestable durante la purificación.

- 5 4.3 Síntesis asimétrica de aminas (R) 1-4 (véase la leyenda en la Tabla 5 anterior) con  $\omega$ -TA de *Aspergillus terreus*, *Mesorhizobium loti* y *Mycobacterium vanbaalenii*

Se llevaron a cabo las síntesis asimétricas a 30 °C durante 24 horas en tampón de fosfato sódico (100 mM, pH 7) que contenía fosfato-5' de piridoxal PLP monohidrato (1 mM) y NAD<sup>+</sup> (1 mM) en tubos Eppendorf de 1,5 ml.



La mezcla de reacción contenía 50 mM de cetona, L-alanina (5 equiv., 250 mM), lactato deshidrogenasa de corazón bovino (90 U), glucosa (150 mM) y glucosa deshidrogenasa (15 U). Se expresaron  $\omega$ -TA de *Aspergillus terreus*, *Mesorhizobium loti* y *Mycobacterium vanbaalenii* (entradas 1, 11 y 9 de la Tabla 1) en *E. coli* BL21 como se ha descrito anteriormente, se congelaron en alícuotas y se aplicaron directamente a la biocatálisis celular sin más purificación. La conversión se midió por detección de las aminas formadas (1, cromatografía de gas (GC); 2-4 electroforesis de capilaridad (CE)). Se llevó a cabo el análisis quiral de 2-4 utilizando la CE como se ha descrito anteriormente. El valor del exceso enantiomérico (%ee) para 1 se analizó por GC. Tras la expresión de la amina con acetato de etilo, se llevó a cabo la derivación con la trifluoroacetamida añadiendo un exceso de 20 veces de anhídrido de ácido trifluoroacético. Tras purgar con nitrógeno para retirar el exceso de anhídrido y el ácido trifluoroacético residual, el compuesto derivado se disolvió en acetato de etilo (50  $\mu$ l) y se separó la línea base utilizando un Shimadzu GC14A que estaba equipado con una columna heptaquis-(2,3-di-O-acetil-6-O-*tert*-butildimetilsilil)- $\beta$ -ciclodextrina (25 m x 0,25 mm). Los tiempos de retención fueron 16,0 min ((S)-1) y 16,2 min ((S)-2) a un gradiente de temperatura del horno de 80 °C/10 min// 20 °C// 180 °C/10 min.

Los resultados se dan en la Tabla 6 siguiente.

Tabla 6

aminas formadas	$\omega$ -TA	Conversión [%] <sup>b</sup>	exceso enantiomérico [%eeP] <sup>c</sup>
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

<sup>a</sup> Condiciones de reacción: 50 mM cetona, 250 mM D-alanina, 100 mM tampón de fosfato sódico pH 7.0, 1 mM PLP, 1 mM NADH. El co-producto de piruvato de la reacción se retiró con lactato deshidrogenasa (LDH). Para el reciclado, se utilizó glucosa deshidrogenasa (GDH). <sup>b</sup> Las conversiones no estaban optimizadas. La desviación de una medición única del valor medio no excedía el 10 %. El compuesto 4 se convertía solo por Ate-TA. <sup>c</sup>(R)-enantiómeros.

20

**Listado de secuencias**

<110> Lonza AG

25

<120> Proceso de identificación y preparación de una transaminasa omega específica de (R)

<130> 202663

<160> 50

ES 2 562 458 T3

<170> PatentIn versión 3.5

<210> 1

<211> 322

5 <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  
 Lys Asn Tyr His Trp Leu Asp Leu Val Arg Gly Leu Tyr Asp Ala Tyr  
 180 185 190

10

ES 2 562 458 T3

Asp Val Gly Ala Glu Thr Ala Leu Ile Met<sup>-</sup> 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> ADN  
 <213> *Mesorhizobium loti*

5

<400>2

atgaccctgg caaccaccga tgcaaccggt ggtgttccgg aagttgaaac cacccataaa 60  
 gatacccgtc gttatccgca tgggtttgca tttatggatg gtcagtatct gccgatgagc 120  
 gaagcaaaaa ttagcgttct ggattgggggt tttctgcatt ctgatgccac ctatgatacc 180  
 gttcatgttt gggaaggctc ttttttctgt ctggatctgc atctggatcg ctttttctgt 240  
 ggtatggatc gtctgcgat gaaactgccg tatcatcgtc gtgaagttga acgtgttctg 300  
 agcaattgtg ttgcactgag cggtcataaa agcgcctatg tggaaatgat ttgtaccctg 360  
 ggtggtagcc cgaccttag ccgtgatccg cgtgaagcag aaaatcgctt tattgcattt 420  
 gcagttccgt ttggttctgt ggcaaataaa gaacagctgg aacgtggtct gcatgttggt 480  
 gttagcgaaa ccgttcgtat tcctccgaaa agcgttgatc cgaccattaa aaattatcat 540  
 tggctggatc tggttcgtgg tctgtatgat gcctatgatg ttggtgcaga aaccgcactg 600  
 attatggata ccaatggcaa tattgcagaa ggtccgggtt ttaacgtgtt taccgtgaaa 660  
 aatcgtcagc tgaaaacacc ggcatttgggt gttctgcctg gtattacacg tcagagcgtt 720  
 tttgatctgt gtggtgaagt tgggtctggca gttaccgcag cagatctgcc tcgtctggaa 780

10

ES 2 562 458 T3

ctgggtgaag cagatgaagt ttttattacc agcaccgcag gcggcattat gccggttacc 840  
 cgtgttgatg gtagcagcat tggtagcggg aaagttgggt ttgttaccg tcagctgatg 900  
 gatctgtatt ggcagaaaca ttctgatgat gcatggctca caccggtaa atatgcctca 960  
 ggatcctga 969

5 <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  
 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

10

ES 2 562 458 T3

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  
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>4  
<211> 987  
<212> ADN  
<213> *Aspergillus terreus*  
  
<400> 4

5

ES 2 562 458 T3

```

atggcaagca tggataaagt tttgcccgt tatgcagcac gtcaggcaat tctggaagc      60
accgaaacca ccaatccgtt tgcaaaaggt attgcatggg ttgaaggtga actggttccg      120
ctggcagaag cacgtattcc gctgctggat cagggtttta tgcatagcga tctgacctat      180
gatgttccga gcgtttggga tggctgtttt tttcgtctgg atgatcatat taccctctg      240
gaagccagct gtaccaaact gcgtctgcgt ctgccgctgc ctcgtgatca ggttaaacia      300
attctggttg aaatggttgc caaaagcggg attcgtgatg catttggtga actgattggt      360
accctggttc tgaaggtgtt tcgtggcacc cgtccggaag atatcgtgaa taatctgtat      420
atgtttgtgc agccgtatgt ttgggttatg gaaccggata tgcagcgtgt tgggtgtagc      480
gcagttgttg cacgtaccgt tcgtcgtggt ccgcctgggt caattgatcc gaccgttaa      540
aatctgcagt ggggtgatct ggttcgtggt atgtttgaag cagcagatcg tggtgcaacc      600
tatccgtttc tgaccgatgg tgatgcacat ctgaccgaag gtagcggttt taacattgtg      660
ctggtgaaag atggtgttct gtatacaccg gatcgtgggt ttctgcaggg tgttacacgt      720
aaaagcgtga ttaatgcagc agaagccttt ggtattgaag tgcgtgttga atttgttccg      780
gttgaactgg catatcgtg tgatgaaatt tttatgtgta ccaccgcagg cggattatg      840
ccgattacca ccctggatgg tatgccggtt aatggtggtc agattggtcc gattaccaa      900

aaaatttggg atggctattg ggcaatgcat tatgatgcag cctatagctt tgaaattgat      960
tataatgaac gcaattcagg atcctga                                          987

```

5

```

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

<400>5

```



ES 2 562 458 T3

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

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

ES 2 562 458 T3

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> ADN  
 <213> *Mycobacterium aurum*  
 <400>6

5

atgactgctc tttcagacct cggcacctcc aacctggtgg ccgtcgagcc cggcgccatc 60  
 cgcgaggaca cccccgccgg ctcggtgatac cagtacagcg actacgaact ggacacctcc 120  
 agcccgttcg cggcgccggt cgacctggatc gagggcgcaat acctgccggc cgaagaagcg 180  
 aagatctcca tcttcgacac cggattcggc cattccgatac tgacctacac cgtcgagcat 240  
 gtatggcacg gcaacatctt ccggctcggc gaccacctgg accggttgct cgacggggcg 300  
 tccaagctgc gcctggacgc cgggtacagc aaggacgaac tggccgagat caccaagaag 360  
 tgcgtgtcga tgtcgcagct gcgcgaatcg ttcgtgaatc tgaccgtcac ccggggatac 420  
 ggaaagcgca agggcgagaa ggacctgtcc aagctcacc atcaggtgta catctacgcc 480  
 atcccgtacc tgtgggcctt cccgcccgcc gagcagatct tcggcaccac cgcgatcgtg 540  
 ccgcccattg tccgccgcgc cggccgcaac accgtcgacc cgaccatcaa gaactaccag 600  
 tggggtgatc tcaccgcagc cagtttcgaa gccaaaggacc gtggtgcgcg caccgcgatc 660  
 ctgctcgact cggacaactg cgtggccgaa ggtccgggct tcaacgtgtg catcgtcaag 720  
 gacggcaagc tggcctcccc gtcccggaac gcgttgccgg gcatcaccgc taagacggtg 780  
 ttcgaactgg ccgaccagat gggcatcgaa gccaccctgc gcgacgtcac cagccgtgaa 840  
 ctctacgacg ccgacgagtt gatggcggtc accaccgcgg gcgggggtcac accgatcaac 900  
 tcgctggatg gcgaggccgt gggcaacggc gagcccggtc cactgacggt ggccatccgg 960  
 gaccggttct gggcgctgat ggacgagccg ggcccgctga tcgaaacgat cgaatactga 1020

10

ES 2 562 458 T3

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

5

<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  
 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

ES 2 562 458 T3

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

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

5

<400>8

atggcattca gcgccgatac ctccgagatc gtctacacgc acgacaccgg cctcgactac 60  
 atcacttata gcgactacga actcgatcct gctaaccgc tcgcgggagg tgcggcatgg 120  
 atcgagggtg cattcgtgcc gccgtcggag gcgcggatct cgatcttcga tcagggttac 180  
 ctccactcgg acgtcaccta cacggtcttc cacgtctgga acggaaatgc attccgcctc 240  
 gacgaccaca tcgaacgcct cttctccaac gcggagtcga tgcgcatcat cctccgcctc 300  
 acacaggacg aagtgaagga gattgcgctc gaactcgtcg cgaagaccga attgcgtgag 360  
 gccttcgtgt ccgtgtcgat taccgcgggt tacagctcga ctccgggcga gcgcgacatc 420  
 acgaagcacc gcccgaggt gtacatgtat gccgtccat atcagtggat cgtgccgttt 480  
 gaccgaattc gcgacggcgt gcacgccatg gtcgcacaga gcgtgcgccg aacccgcgc 540  
 agctcgatcg accctcaggt caagaacttc cagtgggggg atctgatccg tgcggttaa 600  
 gagacgcacg accgcgggtt cgaggctccc cttctgctcg acggcgatgg actgcttgcc 660  
 gagggctcgg ggttcaacgt cgctcgtgat aaggacggcg tcgtgcgcag cccgggtcga 720  
 gcggcgctcc ccggcattac gcggaagacc gtgctcgaga tcgccgaatc gctcggacac 780  
 gaggcgattc tcgccgacat cacgctcgtc gaactgctcg acgccgacga agtgcctggc 840  
 tgcacgactg cgggcccaggt gtggccattc gtcagcgtgg acggcaacc catctcggac 900  
 ggggttccc gccccatcac ccagtcgatc atccgctgtt actgggagct gaatgtcgag 960  
 agctcgtcgt tgcttacgcc tgtgcagtac tga 993

10

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

ES 2 562 458 T3

<400>9

Thr Ser Ala Tyr Ile Arg Pro Leu Ile  
1 5

5 <210> 10  
<211> 8  
<212> PRT  
<213> *Escherichia coli*

10 <400> 10

Val Phe Glu Gly Ile Arg Cys Tyr  
1 5

15 <210> 11  
<211> 8  
<212> PRT  
<213> *Escherichia coli*

20 <400> 11

Asp Val Gly Met Gly Val Asn Pro  
1 5

25 <210> 12  
<211> 9  
<212> PRT  
<213> *Escherichia coli*

<400> 12

Pro Thr Ala Ala Lys Ala Gly Gly Asn  
1 5

30 <210> 13  
<211> 960  
<212> ADN  
<213> *Penicillium chrysogenum*

35 <400> 13

ES 2 562 458 T3

atggctacaa	tggaaaaaat	cttcgccgcc	taccacgagc	gccaaaagct	tcttgcagcg	60
aacacccacc	ccttcgcaaa	aggtgtcgct	tgggtggagg	gagaacttac	tcctctccat	120
gaagcccgta	tccaatcct	agaccaaggc	ttcatgcaca	gcgacttgac	atacgatgtt	180
ccctctgtct	gggatggacg	ctttttccgg	ctcgatgacc	acatcacccg	gttggaaagcc	240
agctgcacca	agctacgcat	gaaactcccc	ctcccacgcg	acgaggtgaa	gcagattctg	300
gtcgatatgg	ttgcaaagag	tggcatccgc	gacgcgtttg	tcgaaatcat	cgtagcgcgt	360
ggattgaaag	gtgtgcgagg	ctctcgccct	gaggatatcg	tcaaccgtat	ctatatgttt	420
attcaaccct	acgtctggtg	tatggaacct	gaggtgcagc	ctgtgggtgg	aagcgcaatt	480
atcgcaagga	ctgtccgccg	cgccccgcct	ggctgcatcg	acccccactgt	caagaatctg	540
caatgggggtg	atctggttcg	cggccttttc	gaggettctg	atcgtggcgc	cgaatatccc	600
ttcctgaccg	atggtgacac	caacctcacc	gaaggttccg	gcttcaacat	tgttctcgtg	660
aaggacaata	ttctgtacac	tccagctcgc	ggagtacttg	aaggtgtgac	acgcaagagt	720
gtgattgatg	tcgctcgagc	cagcggcttt	gacattaagg	tcgagttggt	acctgtccaa	780
atggcttatg	atgcggatga	aatTTTTatg	tgtaccactg	ctggagggat	catgcccatac	840
accagtcttg	atggcaagcc	cgtgaacgac	ggaaaggttg	ggtctgttac	caagaagatc	900
tgggatgggt	actgggctat	ccactatgat	cctgcctaca	gcttcgagat	tgcttattag	960

<210> 14  
 <211> 319  
 <212> PRT  
 <213> *Penicillium chrysogenum*  
  
 <400> 14

5

ES 2 562 458 T3

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

ES 2 562 458 T3

				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
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> ADN  
 <213> *Aspergillus niger*  
 <400> 15

5



ES 2 562 458 T3

atggcatcca tgaaccaagt tcttactgaa tatgccactc gccgcgcgac actagaagcc	60
agtaaaaacc cctacgcaa gggaatcgcc tgggttgaag ggcaactcgt ccccctcagg	120
gaggcccga tccccctaat tgatcaaggc tttttacgca gtgatttaac ctacgatgtc	180
atctccgtct gggatggctg gttctttcgc ctagatgacc acctttcccg acttgaattg	240
gcctgcgcga aatcgcgtct caagttgcc atttcccgcg atgaagtga acaatccctg	300
gttaggatgg tcgctcaaag tggatttcga gatgcatatg tggctttgat tgtgacgcgg	360
ggattgcaga gtgttcgagg tgccaagccg gaggacttgg tgaacaacct gtacatgttt	420
gtacaaccct atgtatgggt aatggaacca gaggtccaac gggtcggtgg aagtgctgtt	480
gttactcgaa ctgttcgctg ggtgccccca ggagctattt atcctactgt aaagaaccta	540
caatggggtg acctgacctg aggtatgctc gaggccgccg atcgaggctc catgtacctg	600
ttcctgacgg atggagatgg ccatctcacc gaggggtccg gatacaatat cgttctaate	660
aaggccgggt ccatttatac gcctgatcgc ggcgtgctgc atgggtgcac caggacaagt	720
gtcattgatg ttgcacgagc ttgtggtatc caagttcacc tcgaagctgt gccggtggag	780
ttggtatata agtgtgatga gatattcatg tgcacaacag cagggtggaat catgccatc	840
actgagctgg atggcaagcc tgtaaattggg gggcggattg gtccgatcac gaagaagatc	900
tgggacgggt attggggtat gcattatgat ccagcctaca gcttcgcagt tagttatgat	960
gacggatcaa aagcaaagct ctga	984

<210> 16  
 <211> 327  
 <212> PRT  
 <213> *Aspergillus niger*  
 <400> 16

5

ES 2 562 458 T3

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

ES 2 562 458 T3

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

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

5

<400> 17

atgacatcta tgaacaaagt attttccggt tactacgagc gcaaggctcg tctagataac	60
agtgacaacc gctttgcgaa aggaattgcc tacgtccagg gatctttcgt cccactcgcc	120
gacgcacgag tcccactcct cgacgagggt ttcatgcata gcgacctcac gtacgatgtg	180
ccatcggtct gggatgggcg cttttccgc cttgatgatc atctcagtcg attggaagat	240
agttgtgaaa agatgcgact gaagatccca ctgtccaggg acgaagtcaa gcaaacccta	300
agggagatgg ttgctaagag tggaaatcgaa gatgcctttg tggagctgat cgtcactcgt	360
ggcctgaaaag ggggtccgtgg caataagcca gaggatcttt tcgacaatca tctctatctg	420
atcgtcatgc cgtatgtctg ggtgatggag cccgccatcc aacataccgg aggtactgcg	480
atcattgccc gtacagtacg gcgactccc cccggtgctt tcgatcctac catcaagaat	540
ctccagtggg gggacttgac acggggtcta tttgaagcgg ctgaccgtgg cgcggattac	600
ccatttctct cagatggaga taccaatctc acagaaggat ccggtttcaa tatagtgttg	660
gttaaagatg gtattatcta cacgcccgac cgtgggtgtc tgggaaggcat tacacgtaag	720
agtggttttg atattgccc ggtcaagaac atcgagggtcc gcgttcaggt ggtgccactc	780
gaacatgcct atcacgccga tgagatattc atgtgtacta ctgctggtgg cattatgcct	840
atcacgaaac tcgatgggaa accgatccgg aatggagaag tcggtcccct tactacaaag	900
atatgggatg agtactgggc gatgcactat gacccgaaat atagctctgc tatcgattac	960
aggggcatg agggtaactg a	981

10

ES 2 562 458 T3

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

5

<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

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

ES 2 562 458 T3

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> ADN  
<213> *Aspergillus fumigatus*

<400> 19

atggcctcta tggacaaagt cttttcggga tattatgcdc gccagaagct gcttgaacgg 60  
agcgacaatc cttttctctaa gggcattgct tatgtggaag gaaagctcgt cttacctagt 120  
gatgctagaa taccgctact cgacgaaggt ttcatgcaca gtgacctaac ctatgatgtt 180  
atatcggttt gggatggtcg cttctttcga ttggacgatc atttgcaacg gattttggaa 240  
agctgcgata agatgctggct caagttccca cttgcaactga gctcagttaa aaatattctg 300  
gctgagatgg tcgccaagag tggatcccg gatgctggtt tggaagttat tgtgacacgt 360  
ggctctgacag gtgtacgtgg ttcgaagcct gaggatctgt ataataacaa catataacctg 420  
cttgttcttc catacatttg ggttatggcg cctgagaacc agctccatgg tggcgaggct 480  
atcattacaa ggacagtgcg acgaacaccc ccaggtgcat ttgatcctac tatcaaaaat 540  
ctacagtggg gtgatttaac aaagggactt ttgaggcaa tggaccgtgg cgccacatac 600  
ccattttctca ctgatggaga caccaacctt actgaaggat ctggtttcaa cattgttttg 660  
gtgaagaacg gtattatcta taccctgat cgagggtgtc tgcgagggat cacacgtaaa 720  
agtgtgattg acgttgcccg agccaacagc atcgacatcc gccttgaggt cgtaccagtg 780  
gagcaggctt atcactctga tgagatcttc atgtgcacaa ctgccggcgg cattatgcct 840  
ataacattgc ttgatggtca acctgttaat gacggccagg ttggccaat cacaagaag 900  
atatgggatg gctattggga gatgcactac aatccggcgt atagttttcc tgttgactat 960  
ggcagtggct aa 972

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

ES 2 562 458 T3

<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

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

ES 2 562 458 T3

				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

5 <210>21  
 <211> 972  
 <212> ADN  
 <213> *Neosartorya fischeri*

<400>21

```

atggcctcta tggacaaagt ctttctggga tatcatgcgc gccagaagct gcttgaacgg      60
agcgacaatc ctttctctaa gggcattgcc tatgtggaag gaaagctcgt cttaccacgc      120
gacgccagaa taccgctact tgacgaaggc ttcatgcacg gtgacctaac ttatgatggt      180
acaacggttt gggatggacg cttctttcga ttggatgadc atatgcaacg gatcctggaa      240
agctgcgata aaatgcggct caagtccca cttgcaccga gcacggtgaa aaatatcctg      300
gctgagatgg tcgccaagag tggattcgg gatgctttg tggaagtat cgtgacacgt      360
ggcttgacag gtgtacgtgg ttcgaagccc gaggatctgt ataataaaa catatacctg      420
cttgttctcc catacgtttg ggttatggcg cctgagaacc agctccttgg tggcagtgct      480
atcattacaa ggacagtgcg acgaacacc cgggtgcat ttgatcctac gatcaaaaaat      540
ctacagtggg gtgacttaac aaagggactt tttgaggcaa tggaccgtgg cgcaacgtac      600
ccatttctca ctgacggaga caccaacctt accgaaggat ctggatttaa cattgtcttg      660
gtgaagaacg gtattatcta taccctgat cgaggtgtct tgcgaggat cacacgtaaa      720
agtgtgattg acgttgcccg agccaacaac atcgacatcc gccttgaggc cgtaccagtg      780
gagcaggttt atcactccga tgaatcttc atgtgcacaa cagccggtgg cattatgcct      840
ataacgttgc ttgatggca accagttaat gacggccagg ttggcccgat cacaaagaag      900
atatgggatg gttactggga gatgcactac aatccggcgt atagttttcc ggtcgactat      960
ggcagtggtc aa                                               972

```

10

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

<400> 22

ES 2 562 458 T3

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

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



# ES 2 562 458 T3

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

5  
<210> 23  
<211> 978  
<212> ADN  
<213> *Gibberella zeae*  
  
<400> 23

```
atgtcgacca tggacaagat ctctcgcggc cacgcccagc gccaaagccac cctcgtcgca      60  
agcgacaaca tcttcgcaa cggcattgcc tggatccaag gcgagctcgt cccctcaat      120  
gaagcccgca tccccctcat ggaccaaggt ttcattgcac gcgacttgac ctacgatgac      180  
cctgcagtct gggatggtcg tttcttccgt cttgatgacc atctcgaccg tctcgaggca      240  
agcgtcaaga agatgcgaat gcaattcccc attccccgcg atgagatcag aatgactctt      300  
ctcgacatgc tcgccaagag tggaaatcaag gatgcttttg ttgagctcat tgtcactcgt      360  
ggcttgaagc ctgttcgtga ggccaagcct ggtgaggctt tgaacaacca cctctacttg      420  
atcgtccaac cctacgtctg ggtcatgagc cccgaagctc agtacgctgg cggtaatgca      480  
gttatcgcac gaactgttcg tcgaaatcct cctggatcca tggatccac catcaagaac      540  
ctccaatgga gtgatttcac ccgcggcatg ttcgaagcat acgatcgtgg agcacaatac      600  
cccttctca cgcacggcga cacaacatc accgaaggat ctggtttcaa cgttgtcttt      660  
gtcaagaaca acgttattta caccccgaac cgaggagttt tgcagggaat taccagaaag      720  
agtgtgatcg acgctgcaa gtggtggtgt catgaagttc gagtggagta tgtccctggt      780  
gagatggcct atgaagctga tgagatcttc atgtgtacta ctgctggagg aatcatgcct      840  
atcaccacca tggatggaaa gccagtcaag gacggaaaagg tctgggctgt cacaaaggcc      900  
atctgggatc ggtactgggc gatgcactgg gaggatgagt tcagtttcaa gattgactac      960  
cagaaactga agctgtag                        978
```

10  
  
15  
<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

ES 2 562 458 T3

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				
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	

ES 2 562 458 T3

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

5 <210> 25  
 <211> 966  
 <212> ADN  
 <213> *Hyphomonas neptunium*  
 <400> 25

atgctgacat ttcaaaaagt gctgaccgga tttcagacgc gggcagacgc tcgcgagag 60  
 cgaacagacg cgttcgccga cggtattgcc tggatcgaaa acgaatttgt gcccatcgga 120  
 aaggcgcgga taccgatact tgatcagggc ttcctgcatt cggatctgac ctatgacgtc 180  
 cccgcagtct ggaatggccg gatcttccgt ctcgatgatc acctggaccg tctggaagtt 240  
 agctgcgcca aaatgctgct cccgctgccg atcgcaggc cagagctccg caggctggtg 300  
 atggagctgg tctccaggag tgggctgctg gacgcatatg ttgagatcat cgtgaccgc 360  
 gggttgaagt tcctgagagg cgcacaagcc gaagacatca tcccgaacct gtatctcatg 420  
 gcggtgcctt acgtctggat tctcccgtt gaataccaga accatggcgc gccggccgtc 480  
 gtaacgcgaa cggtgcgccg cacacccccg ggggcccctg acccgacaat caagaacctt 540  
 cagtggggcg atctcgtccg agggcttatg gaagcaggtg accgggactc tttcttcccc 600  
 attcttcccc acggcgacgg gaacgccaca gaaggcgtg gctataatat cgtccttgtc 660  
 aggaatgggg aacttcacac acccgcgcgc ggggtgctgg aaggcatcac acgcaggaca 720  
 gtactggaaa tagcggccgc gcgcggcctg aaaacacatg ttaccgagat acccatccag 780  
 gcactctatg aatgcgatga attgttcatg tgttcgacgg caggcggcat aatgccactc 840  
 gtccttctgg atgggaacat tgtgggtgat ggaacagtcg gtccggttac acggatgatt 900  
 tggggaggctt actgggatct tcatgatgat cctcagctca gcgagcctgt gacttacgcg 960  
 10 ccctga 966

15 <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

ES 2 562 458 T3

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		
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

ES 2 562 458 T3

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

5 <212> ADN

<213> *Mesorhizobium loti* MAFF303099

<400> 27

atggaccaga cgacggcaac acaggettca aagccgctgc ccacagtcgg cgaccgccac 60  
 gtcgaccgcg attcctacce cgacggcatc gccttcctcg acggccagta tctgccgatg 120  
 tcgcaagcca aggtgtcggg gctggactgg ggcttcctgc attccgacgc cacctacgac 180  
 acggtgcatg tctggaacgg ccgcttcttc cgcttcgacc tgcattctga ccgcttcttc 240  
 ggcggactgg aaaagctgcg catgaccatc cccttcgaca gggatggcgt ggccgagatc 300  
 ctgcacaatt gtgtcgccct ttcgggccat cgcgcccgtc atgtcgaaat gctgtgcacg 360  
 cgcggcgcat cgccgacctt cagcccgcat ccgcccagg cgatcaaccg cttcatggcc 420  
 ttcgcccgtc ccttcggctc ggtcgccaat gccgagcagt tgcagcgcgg cctgcccgtc 480  
 gccatcagcg acaaggtgcg catcccgccg gcttcggctc atccgctgat caagaactac 540  
 cattggctcg atctgggtcg cggcctctac gacgcctatg acagcgggtc ggagaccgcg 600  
 ctcattctcg acttcaacgg caatgtcgcc gagggcccgg gcttcaacgt cttctgctc 660  
 aaggacggca aactgtcgac gccggccatc ggcgtgctgc ccggcatcac gcgccgcaca 720  
 gtcttcgatc tctgcccga agaaggtctt gcccccggc ccgccgatgt cagcgtcgcc 780  
 gcgctcaagg cggccgacga ggtcttcac acctcgaccg ccggcggcat catgccgggtg 840  
 accgagatcg acggcggcgc gatcggcgc ggcaaggctc ggccggttac cagccggcta 900  
 atggcgctct actggcaaaa gcacgacgat ccggcctggt cgtctcaggt gaagtatccc 960  
 tga 963

10

<210> 28

<211> 320

15 <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

ES 2 562 458 T3

	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
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

ES 2 562 458 T3

5  
 <210> 29  
 <211> 936  
 <212> ADN  
 <213> *Roseobacter* sp.  
 <400> 29

tcaaggatat cggatcgctt tccgaaaacg atcatcttca tgcattctcc aatacgcctc 60  
 agtgatctcc ttggttatcg ggccgaccaa tccgcttcca attgaagttt catcgatctt 120  
 ggtgacgggc atgacgccgc cagcgggtga ggtcacgaac acttcatccg ccgcctcaa 180  
 ttcacttggg gccagatcat ccgtcgcgca ggtgatctgt aattcgtcac aaagatcaaa 240  
 gattgtttgg cgtgttattc ccatcagcac cccaaacttg ggcgaggaaa ttttgccggt 300  
 tttcacggcg aagacgttga aaccgggccc ttccgctatg tttccggtgg cgtctaaaag 360  
 aatggcgggt tcggctccct gcgcatacgc cgcatacaaa ctttgacca ggtccagcca 420  
 gtgatagttc tttacggtg gatcaacgga gctgggtgga atgaggacaa ggtcggtgac 480  
 ggctgctgtg aagccacgtc tcatctgttc tgggtttgca accgaccaa atggaattgc 540  
 gaagggcgtg aagcgggtga cggcatcgcg agggctcgcg ctaaactgtag gagaagtgcc 600  
 gcgtgtgcag atgaactcga catacgcata ctggagacct gatagagcaa cgcaattgtg 660  
 caaaatctca gtcattctcg cagcaccaaa cgggatcgac atgtgcaact tttccatccc 720  
 tgcgaagaac cggtaagat gatcatcaag cctgaaaaat gctcccttcc acacatgccc 780  
 aacatcgtat gtcgcatctg agtggaggaa gccattatcc aacaccgaaa tctttgcctc 840  
 gctgatcgga agatattggc catccatgaa ggcaatgccg ggagggtagc ttgccgggtc 900  
 10 agaatgacga tcagacagcg caggcaatth atacat 936

15  
 <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 2 562 458 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> ADN  
<213> *Marinomonas* sp.

5

<400> 31



ctatatTTTT gccgggggat aatgaatggg taggcaccaa tcagagTTTT tgtgctTTTT 60  
 ccaatatgca gattttaaca attggaagat aggtccggtt gtgccacttc caatcacatt 120  
 atggtcgatt ttggaatgg gcataatgcc acccgccgtt gacgtaatga atacctcgtc 180  
 agcattcctt aaatcggtag gcgttacctc agttgcacag cagcttatgt gtaattcgtc 240  
 acaaagatca aagacggttc ggcgagttat gcctggcaaa acccctTTTT ctgggtgTTT 300  
 aattatcttt cccttaacac aaaaaacatt aaagccaggg ccttcggcga tattaccctc 360  
 tgaatccaac aggatggcgg tctcaccacc tttttcatag gcatcgtata acccggttac 420  
 caaatcaagc caatggtaat ttttgacctt tgaatcaacc gagtttggcg gaatacggac 480  
 aatagagcta ataatggcat gcagtcatt tttcatttga tcttgatttg cgaccgagcc 540  
 aaaggaagc gcgaaagcca taaaacgatt gattgagtct cttgggtccc ggctgaaatc 600  
 cggcgaattc cctcttgtgc atatcatttc gacataggcg tttttatgac ccgataaggg 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

5

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  
 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

10

ES 2 562 458 T3

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> ADN  
<213> *Rhizobium etli*

5

<400> 33

ES 2 562 458 T3

gtgctcgatc gcggcgagga atccgccttt cagaaaggct ccgcctatgt 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 tgcttacggt	300
caaatcatca tgaccgcgg gcggcccccg atcggtagtc gggatcttcg tttatgctcg	360
aacagcttcc aggcattctg cgtgccgtac atgtggatcg cgaaccccgaa aaagcaggaa	420
attggcatgg cagtgcacgt cagccggcgc gtccgaatcc cacctcagtc cgtagatccg	480
ttggtgaagc attaccattg gctcgacttt gagatgggct tgttcgaagc ttacgagaac	540
ggtgccgata cggtggttct cactgatctc gatggcaata tcacagaggg gcccggcttc	600
aacgttttcg cggatgatcga cggagttctg aggacccctt ctttcggcat gctggatgga	660
atgaccagga ggacagttat ggaactgtgc aacgagctga atctcgatgt cacgcaagag	720
accatttctt tggagcggct tctaatacgc tcggaatct tccttacgac gaccgcaggc	780
gggattatcc cggctctctc ggtgaacgga accggaatcg gtttcggatc agtgggtgaa	840
caaactcgcc gcattcaccg gtcatactgg gataagcgaa gcagcggatg gtatggcgag	900
cctgtcgctt acgcccacaaa agctctcttg gagccctag	939

5 <210> 34  
 <211> 312  
 <212> PRT  
 <213> *Rhizobium etli*  
 <400> 34

ES 2 562 458 T3

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  
 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

ES 2 562 458 T3

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> ADN  
 <213> *Rhodofera ferrireducens*

5

<400> 35

atgccagcac ccgatctttc caaaggcgtc gcctttgtac gtgggcagta cgccccatt 60  
 ggtgaagcaa gcattccgat aactgactgg ggtttcctgc gctctgacgc cacctatgac 120  
 gtggtgacgg tgtgggatgg ttcctttttt cgcctggacg cccatctgga gcgcttcacg 180  
 cgcagctgcc aacgctggcg gcttgacccg gggctgacgc ccgggcaaat caccggcgtg 240  
 ttgtcgcaat gcgtgcgctt gagcgggctt cgcgcctctt atgtcgagat gatctgcacc 300  
 cggggccagc cgcttggggg atcgcgcgac ccacgcctgg ccgtcaatca gttctatgcc 360  
 tttgcggtgc cctatgtctg gcttgccaat gcccagcagc gtgaagcggg actgcacctg 420  
 atgatcagcg acgtgcaacg tattcctgcg acctccgtgg acccttcggc caagaactac 480  
 cactggaacg acttgacgat gggactgctg ggggcgctgg acgccggggc cgacagtgtt 540  
 gtgctggtcg actcagtcgg gaacgtggtc gagggacccg ggttcaacgt gttttgtgtc 600  
 agccacggcg cgcttgtgac gcccagcgag ggcattgctt aaggggtgtc gcgtcgcacc 660

10

ES 2 562 458 T3

gtcatcgaga tggcgcgcgc tctggggctg gagacacaac tgcgtgcctt gcccgccgat 720  
 gagttgcgtg gagccgagga ggtgttcac tccacctcgg gcggtggtgt actgcccgtg 780  
 agccgggtgg acaaacgtcc ggttggtgat ggccgtccgg ggccgatcac gcagcgctg 840  
 gtacagacct actgggcctg gcatgccgat ccggtgtaca gccagccgat agattattcg 900  
 ctttga 906

5 <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  
 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

ES 2 562 458 T3

	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> ADN  
 <213> *Jannaschia* sp.

5

<400> 37

tcaatagggc	atgtcggttc	ggtgctcggc	ccgcgtgata	cagtcgaaat	accgccgccg	60
caggctcagc	gccacggggc	ctgccgcata	gttgagagaag	acgcggttat	cgacccgcgc	120
caccgggata	acgccgccgc	ccgaggacga	caggaagacc	tcatacggcct	ccaggaattc	180
gtccaggggc	agcgggcgtg	tttcgaccgt	caggccccgc	tctcgggcca	ttccagcac	240
cgtagcggcg	gtgatgccat	gcaacacacc	gtggtccgac	gtcacgatcc	ggtcaccgaa	300
cagggcgaag	gcgttgaagc	ccggcccttc	ggtcacatgg	cccgcgtggt	ccagcagcaa	360
gaccgtctca	aaccccttgt	ctttcgcttc	gaacaacccg	ccggtgaaat	cgccccaatg	420
gtagtttttc	acggtgggat	caacgctgtc	ggcggggatg	cggcgggtgc	gtttggcgat	480
ccagacggag	gttccacgct	cggcgatttc	gggtttcacg	atgtgcacgt	acggcacgca	540
ccaggcgtag	aagtggttg	cacaatcgcg	cgggtcccgc	gatccgggca	ccgggttgcg	600
ccccgcgcgc	gcgaccatgg	cgacatagct	gtcgcgcaat	ccgcttgcg	ccaccatccg	660
ggtcagggcc	gttttcacgc	cgccccgctc	gcgcccgata	tccatccgca	gcgcgccgac	720
ggaggccatg	aaccgcgcca	cgtaatcctc	cagccggaag	aagccgcccc	gccagaccgg	780
caccacgtca	taggcgatgt	cggaatgggt	caggccccaa	tcggtgacgg	ggatcgcggc	840
ctctgcgatg	gggatgatgc	ggctccccat	ccaggccgcg	ccgcttgata	aatcattcat	900

10

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

15

<400> 38

ES 2 562 458 T3

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



ES 2 562 458 T3

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> ADN  
 <213> *Labrenzia alexandrii*

5

<400> 39

ctaatcccga taagaaatct ccgtccggtt atccggacgg gcggcccatt caaaataggc 60  
 cttgtgaagt gctgtcgtaa tcggaccggt cgcacatctt gaaaagatcc ggtcatccac 120  
 ccgggtgacc ggtgcaacgc cgccaccgga cgtggtgatg aaaacttcgt cgctctcaaa 180  
 gaactcacta agcgcgaagg cccgtatctc gatctcaagc ccctgttctg ccgcatatc 240  
 cagcaccgtc tgacggctga tgccttccaa aacaccgctt tcagccgtca ccagcgttcg 300  
 gccctttaca gcaaacacat tgaagcccgg cccttcggtc acgttgccat catcatcgag 360  
 caagatgaca gtctccgctc cgtggtcctt ggctctcaaac agacccttgg tgaatcacc 420  
 ccagtggtag tttttgacct tcggattgac actgcccggc gaaatccggt gcacggtctt 480  
 ggcgatatga acatgggctc ctctttccgc gatctccggc cggatgacgt gaacataggg 540  
 cacgcacat gcaaagaatt ggttcttgca gtcccgggga tcgctcgaac ccggcacctg 600  
 gttcacacc ccggaagtga ccattgccac gtaggcctt cgcagccccg tcgcccacc 660  
 gagcgattga agggcgtctt ccatctctc atcgtcaac ccagggtcca accgcagttc 720  
 gtccatcgag cggcggaaac gggccagata atggggcaaa cgaaagaatg caccatccaa 780  
 gaccgggacc acatcatagg taatgtccga atggatcagc cccaatctg taacaggcaa 840  
 tgcggcctcg gaaatcggca tgagcctgcc gccatccac gcggcaccgt tggacaaatc 900  
 tgagatggta acgcaacat cgacgggtgga ggacat 936

10

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

15

<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

ES 2 562 458 T3

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  
 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

ES 2 562 458 T3

<210>41  
 <211> 981  
 <212> ADN  
 <213> *Burkholderia sp.*

5 <400> 41

```

ttatgccggg accgacgaga agtattcagc cggctgcgcg tgccatcctg cccagcgctt      60
ctcccagtag aggttgtgaa tcttctcggg gatcggggccg gggccgttgc ggttgccgag      120
cggctgatcg tcgattgcmc tgacgggcat gatcccgcct gccgacgacg aggtgaatgc      180
ttcgtcagcc tcgcmcagct gggctcgggt gtacttgccc atgttcacct tgatgccgag      240
ctcggcggcg atatcgaaga cgctctggcg cgtgatgccc agcaggcagc cttcggcggg      300
cgtatagagc tccccgttct tggcgaaaaa cacgttcgca cgggctgctt cggtgaggtt      360
gtcgccttcg tcgacgagaa ccgcccagtc cttctcctgc gtcattggctt cgaacagcmc      420
gagcttcatg tccatccagt ggaagttctt ggcctgggga tcgaccgcct tcgcggaaat      480
gcggttgtag agcttgctga tcagcaggtt cgagcctcmc gtgcmcacct cgtcmcagc      540
ctggaagaag aacggcacgg caaaccmcga catcmcgttc ttcattggcmc cgcggtcmc      600
gcgatcmcag ctcagcmggc cgcgggtgac gcaccaccag acgtaccmcgt cttcagggc      660
ggcgttgcmc acgaggttgt ggagaatctc cttgacctgg tcgcmattga acggattctc      720
aaggaagaac ttcgcmcagc actcttccat gcgcmtgagg tgatcmcga gccggaagaa      780
gttgccccgc gagacggtga cgacatcmga ggcgcmcgc gcatcmcagga atcctcmcgc      840
gacgagcmgg acggtcmcct cggtgatggg gacgtagttg ccmgtcacaga aggcgctcmc      900
atcttcmtag cmgggttcat gcmgagcmc agcatcmcag cmgttctcmgt gcatcmctg      960
ctggacttga atgattcmca t                                     981
    
```

10 <210> 42  
 <211> 326  
 <212> PRT  
 <213> *Burkholderia sp.*

15 <400> 42

```

Met Ala Ile Ile Gln Val Gln Gln Ile Met His Glu Asn Pro Leu His
1           5           10
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
    
```

ES 2 562 458 T3

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

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> ADN  
<213> *Burkholderia cenocepacia*

ES 2 562 458 T3

<400> 43

```

tcagccgtcc gcgtagtcga ccgccaggct ccacgccgga tcgccgtgct tcgcccagta      60
cgcgtcgaac agccggcgcg tcatcggccc cggacggccc tcgccgatcg tcgctcgtt      120
cagccgcgtg accggcatga tgccgccggc cgtcgcgctg atgaacaact cgtccgcgtc      180
gcgcaactgt gcatcgtcga tgcgcgcagc ctgggcgtcg atgcccacgc ccgtcgcag      240
ctcgaacacc gtctgccgcg tgatgccatg cagcacgccg cgctcgggcg tccgcagccg      300
gccgtcgcgc acgacgaaca cgttgaaccc gggcccttcg gcgatgctgc cgtccgtgca      360
cttcagcagg accgactccg cgcccgcgtc atagcccttc agcagcccgg cgacgaggtc      420
cagccagtga tagttcttga tctgcggatc gaccgattcg ggcggaatgc gccgcacgtc      480
gtcgatcacg tggaggtgca gcccttcgcg caactgccgc tcgttcgcga ccgatccgta      540
cggcaccgcg aacgcgatga actggttcac cgcacgcgcg gggtcgcggc tgaaggtcgg      600
cgacacgccg cgcgtacaca gcatctcgac gtaggcgtgg cgcagccccg agcggcgcac      660
gcattcgacg aggatgtcgc gcagcgcacg gtcggtcaac ggcacgttca gccgcaaccg      720
cgcaagcgag cgcctgaatc gctcgatatg cttgtcgcgg cggagaagaag ggccgttcca      780
aacatgcacg gtgtcgtacg tgacgtcggg gtgcagaaag ccccagtcga gcaccgatac      840
gcgcgcgctc gcgatcggaa tgaagcggcc gttcatgtag gcggcgccct gcggaaacgc      900
gggcgcttcg acggatggaa tgccatgatg ggccgaacgt gtctcccgcg gcac          954
    
```

5 <210> 44  
 <211> 317  
 <212> PRT  
 <213> *Burkholderia cenocepacia*

10 <400> 44

```

Met Pro Arg Glu Thr Arg Ser Ala His His Gly Ile Pro Ser Val Glu
 1           5           10
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
Glu Cys Val Arg Arg Ser Gly Leu Arg His Ala Tyr Val Glu Met Leu
    
```

ES 2 562 458 T3

			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> ADN  
 <213> Proteobacteria alfa  
 <400> 45

5

ES 2 562 458 T3

```

ttaatattca actggatcac gatgtttacc ttcattgtgc agctgccaat aggcatgcat      60
tatacgtttc gacaaaggcc cgatattgcc ggaagcaatt ggtcgttcgt caattcttgt      120
aacaggcatc acgcccccg cagtcgaggt tatgaagacc tcgtcagcgg ttttaagctt      180
gtcgacagta atgtcttgcg catgacatct aacttccaat tcttcgcaa tgcaaaaaat      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 tggtggaatg cgaacgacat ccgtgatcgc      480
cagatgcaaa ccacactcca actgttcttt gctggccact gaaccgaacg gaatggcaaa      540
agctatgaag cggttttctg catctcttgg atctcgactg aagttgggtg acgttcctct      600
tgtgcagatg aactcaacat atgcattttt gagcttggac aacgccacgc aattatgaag      660
aatttctatg atttgcgctt tgggtgaagg aattgacata tgcaaagcgt ccattccacg      720
ggaaaatcga tctaaatgat cacctaactt gaaaaatgct ccgccccata catgggcaac      780
gtcataagtc gcgtctgaat gaagaaaccc ataatccaaa acggacaatt tcgcctcatg      840
gataggtagg tactgccccat ccat      864

```

<210> 46  
 <211> 287  
 <212> PRT  
 <213> Proteobacteria alfa  
  
 <400> 46

5

ES 2 562 458 T3

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  
 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





ES 2 562 458 T3

```

atgagtgacg agcccattat ctatatcaac ggtgactatc ttccactgtc tcaggcgcgg      60
gtatccccgg tggatcaggg attcctgctc ggtgatggcg tattcgatgt ggtctccgcc      120
tggaaggcga acatcttcaa gctcgacgcc catctcgatc ggttctttga ctcgattcag      180
gccgcgcgac tcaatcacga catgagtcga gacgcgtgga aggaagcgat catcgaaacc      240
acgcgtcgca atggactcga cgatgcctcg attcgcctta tcgtgacccg cggcgagccc      300
aaaggggtgg ttgctgatcc cggggatttt aaaccgacgt gcatcgtctg ggtggcgctt      360
tatatcttcc tcgcggatga ggagaaacgc cgcaatggta ttcgcctgat gattagcgcg      420
acgcgggggt tccctgctga caccctggac cctcgttaca aatgcctcga ccgcctgcat      480
tcacagctga ttcggcttga ggccctggag gcgggttatg acgatgcgct ttggctcgat      540
cattccggtc acgtgtccga gtcagcagcg agcaacctgt ttatcgtcaa gaatggtgtg      600
ttgtacacc cttcagcagg aattctgctc gccattacac gggacaccat tctcgagctc      660
gcgaccgagc tggacatccc ctgaaagag cgacagctca gtgcgttcga tgtctatata      720
gccgatgagg tcttcacctg cagcacagcg ggtggcgcgc ttccggtcag ggaggtcgca      780
ggtcgaacga ttcgcggcac aacccccggc ccgattaccc aggcaatcga caacgcgtat      840

tgggcgatgc gtgaaacaga ccggtacgcg acgccgcttt aa                          882

```

```

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

<400> 48

```

5

ES 2 562 458 T3

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

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

ES 2 562 458 T3

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> ADN  
<213> *Mycobacterium vanbaalenii*

5

<400> 49

atgggcattg ataccggcac cagcaatctg gttgcagttg aaccggtgc aattcgtgaa 60  
gatacaccgg caggcagcgt tattcagtat agcgattatg aaattgatta tagcagcccg 120  
tttgccggtg gtgttgcacg gattgaaggt gaatatctgc ctgcagaaga tgcaaaaatt 180  
agcatttttg ataccggttt tggatcatagc gatctgacct ataccgttgc acatgttttg 240  
catggcaata tttttcgtct gggcgatcat ctggatcgtc tgctggatgg tgcacgtaaa 300  
ctgcgtctgg atagcggtta taccaaagat gaactggccg atattaccaa aaaatgtgtg 360  
agcctgagcc agctgctga aagctttgtt aatctgacca ttaccctggt ttatggtaaa 420  
cgtaaaggcg aaaaagatct gagcaaaact acccatcagg tgtatatatta tgccattccg 480  
tatctgtggg catttctctc ggcagagcag atttttggca ccaccgcagt tgttccgcgt 540  
catgttcgtc gtgccggtcg taataccgtt gatccgacca ttaaaaatta tcagtggggt 600  
gatctgaccg cagcaagctt tgaagcaaaa gatcgtggtg cacgtaccgc aattctgatg 660  
gatgccgata attgtggttc agaaggtccg ggttttaacg tgtgcattgt gaaagatggt 720  
aaactggcaa gcccgagccg taatgcactg ccgggtatta cacgtaaac cgtgtttgaa 780  
attgccggtg caatgggtat tgaagcagca ctgctgatg ttaccagcca tgaactgtat 840  
gatgccgatg aaattatggc agttaccacc gcaggcgggt ttaccctgat taataccctg 900  
gatggtgttc cgattggtga tgggtgaaccg ggtccggtta ccgttgcat tcgtgatcgt 960  
ttttgggcac tgatggatga acctggtccg ctgattgaag caattcagta ttaa 1014

10

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

15

<400> 50

Met Gly Ile Asp Thr Gly Thr Ser Asn Leu Val Ala Val Glu Pro Gly

ES 2 562 458 T3

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
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

ES 2 562 458 T3

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**

1. Un proceso para la preparación de una transaminasa  $\omega$  selectiva de (R), que comprende las siguientes etapas:
  - 5 a) proporcionar al menos una secuencia de biomolécula incógnita de al menos una transaminasa  $\omega$  selectiva de (R) y al menos un banco de biomoléculas,
  - b) cribar el banco de biomoléculas con la secuencia de biomolécula incógnita 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 % con la secuencia de biomolécula incógnita, que se calcula a nivel de aminoácidos,
  - 10 c) seleccionar en el grupo de las primeras secuencias de biomolécula diana un grupo de segundas secuencias de biomolécula diana, que no comprende, a nivel de aminoácidos, al menos uno de los siguientes motivos de secuencia de aminoácidos c1) a c3) siendo
    - 15 c1) en las posiciones 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 las posiciones 97 a 99 una secuencia de aminoácidos Tyr Xaa Gln, siendo Xaa un aminoácido y en la región desde la posición 105 a 111 una secuencia de aminoácidos Arg Xaa Xa3, siendo Xa3 un aminoácido, siendo preferentemente His o
    - 20 c3) en la posición 38 Thr, en la posición 97 Lys y en las posiciones 107 a 109 una secuencia de aminoácidos Arg Xa4 Xa5, siendo Xa4 un aminoácido, siendo preferentemente Gly, y siendo Xa5 un aminoácido, siendo preferentemente Tyr y que comprende
    - 25 c4) en la posición 95 otro aminoácido distinto de Tyr, Arg, Lys, o en la posición 95 Tyr, pero en la posición 97 ni Arg ni Lys y
    - c5) en la posición 40 ni Lys ni Arg y
    - c6) en la región desde la posición 161 a 165 al menos una Lys para identificar un grupo de segundas secuencias de biomolécula diana y
  - 30 d) proporcionar una biomolécula que tiene una segunda secuencia de biomolécula diana identificada en la etapa c) y siendo, o codificando al menos parcialmente, una proteína con la actividad de una transaminasa  $\omega$  selectiva de (R).
- 35 2. Un proceso para cribar una transaminasa  $\omega$  selectiva de (R) que comprende las etapas a) a c) de la reivindicación 1.
3. El proceso de acuerdo con las reivindicaciones 1 o 2, en el que la biomolécula es una proteína y la secuencia de biomolécula es una secuencia de aminoácidos.
- 40 4. El proceso de acuerdo con las reivindicaciones 1 o 2, en el que la biomolécula es una molécula de ADN y la secuencia de biomolécula es una secuencia de ADN.
5. El proceso de acuerdo con una cualquiera de las reivindicaciones anteriores, en el que el banco de biomoléculas es una base de datos de biomoléculas y la base de datos de biomoléculas se criba en la etapa b) con una herramienta de alineamiento de secuencias de biomoléculas, en particular BLAST.
- 45 6. El proceso de acuerdo con una cualquiera de las reivindicaciones 1, 3, 4 y 5, en el que la proteína o la molécula de ADN proporcionadas en la etapa d) se proporcionan por síntesis *de novo*.
- 50 7. El proceso de acuerdo con una cualquiera de las reivindicaciones 1, 2 y 4, en el que el banco de biomoléculas es un banco genético y el banco genético se criba en la etapa b) con una molécula de secuencia de ADN incógnita.
8. El proceso de acuerdo con la reivindicación 7, en el que en la etapa c) se utilizan cebadores de secuencia de ADN para seleccionar el grupo de segundas secuencias de biomolécula.
- 55 9. El proceso de acuerdo con una cualquiera de las reivindicaciones anteriores, en el que la secuencia de biomolécula incógnita es una secuencia que representa la totalidad o parte de una transaminasa  $\omega$  selectiva de (R) funcional.

Figura 1 - B3Api convencional

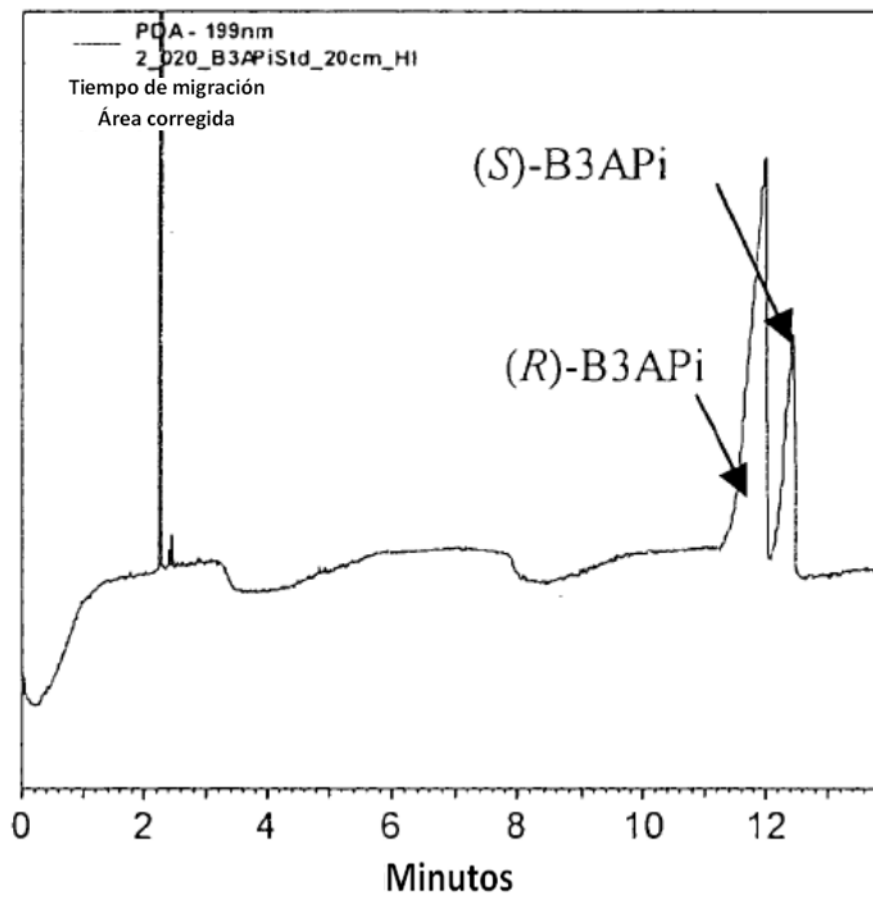




Figura 2 - Síntesis asimétrica de B3APi

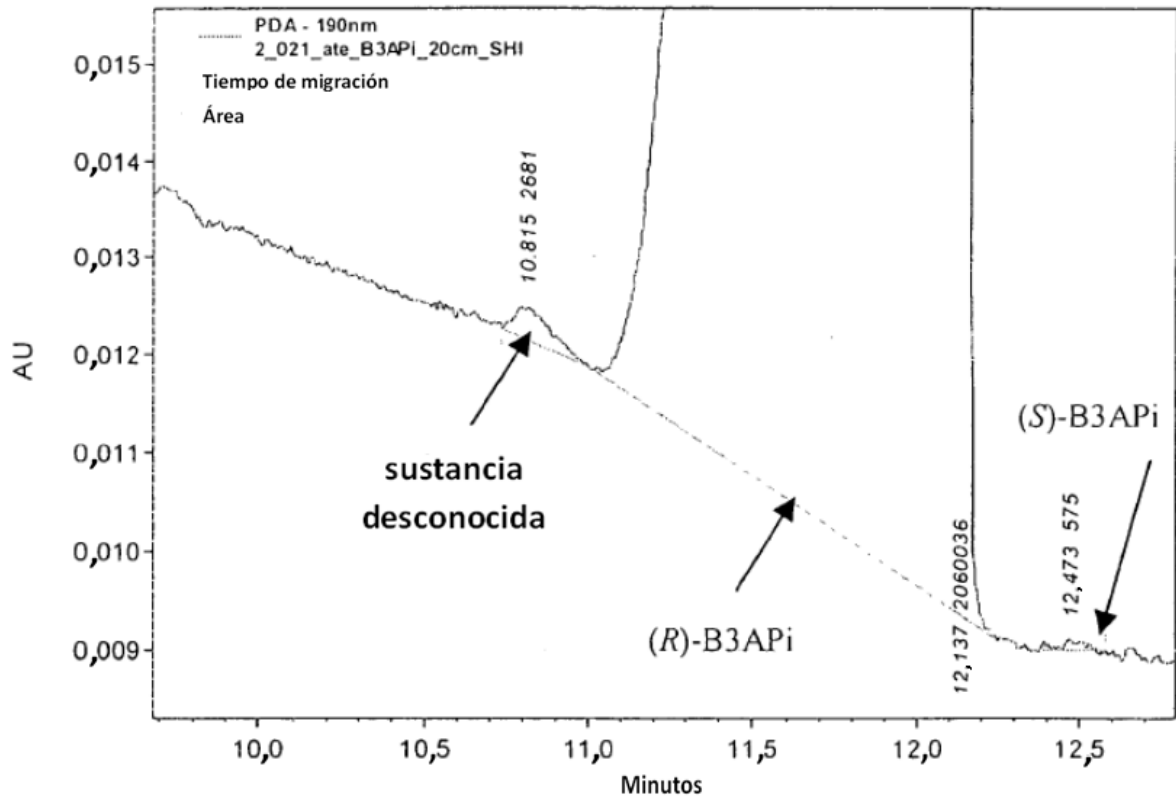
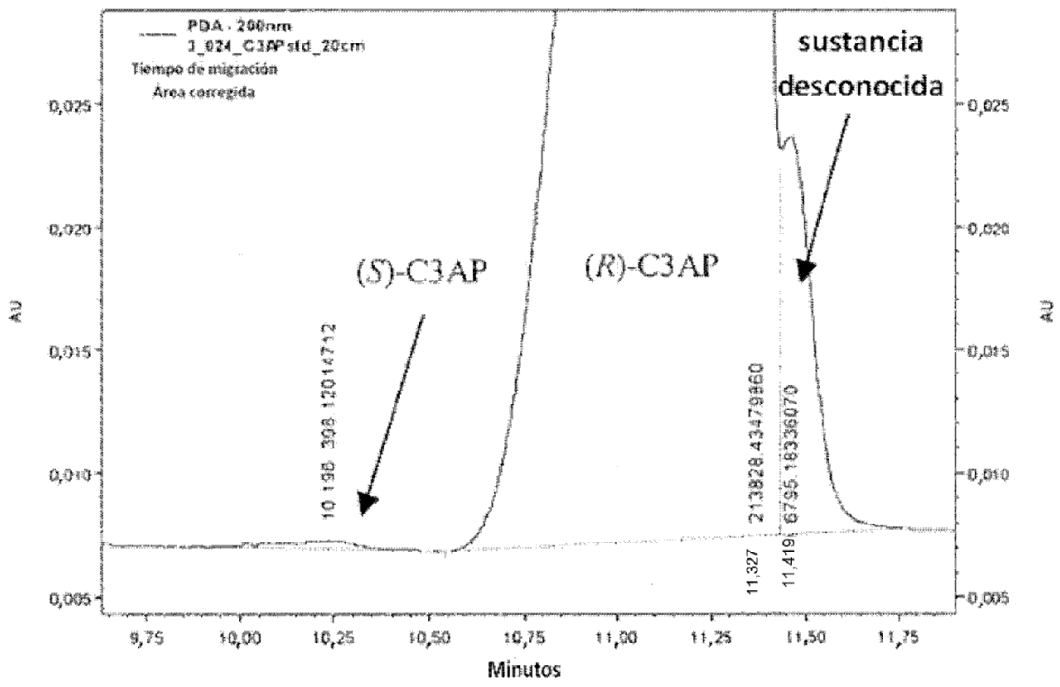
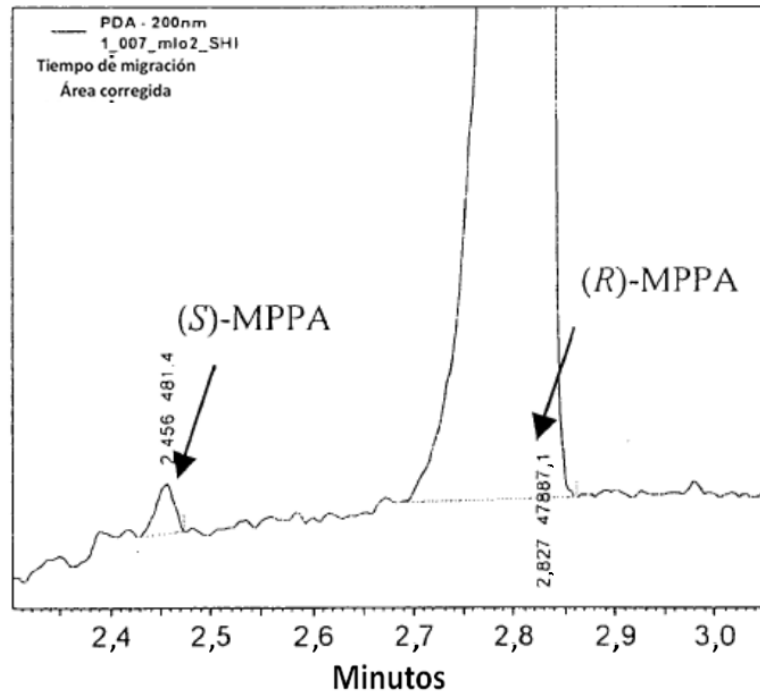


Figura 3 - Síntesis asimétrica de C3AP



**Figura 4 - Síntesis asimétrica de MPPA**



**Figura 5 - Síntesis asimétrica de B3AP**

