

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 809 727**

51 Int. Cl.:

C11D 3/38 (2006.01)
C02F 3/34 (2006.01)
C12R 1/01 (2006.01)
C07H 15/04 (2006.01)
C11D 1/66 (2006.01)
C11D 1/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **06.04.2017** **PCT/EP2017/058296**

87 Fecha y número de publicación internacional: **12.10.2017** **WO17174748**

96 Fecha de presentación y número de la solicitud europea: **06.04.2017** **E 17719188 (9)**

97 Fecha y número de publicación de la concesión europea: **03.06.2020** **EP 3440179**

54 Título: **Biotensioactivos glicolipopeptídicos**

30 Prioridad:

06.04.2016 GB 201605875

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.03.2021

73 Titular/es:

CRODA INTERNATIONAL PLC (100.0%)
Cowick Hall Snaith
Goole, Yorkshire DN14 9AA, GB

72 Inventor/es:

KERR, RUSSELL GREIG;
HALTLI, BRADLEY ARNOLD;
MARCHBANK, DOUGLAS HUBERT y
BERRUÉ, FABRICE

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 809 727 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

Biotensioactivos glicolipopeptídicos

Campo de la invención

- 5 Esta invención se refiere generalmente a los campos de la química de tensioactivos, bioquímica, y microbiología. Más específicamente, la invención se refiere a biotensioactivos que tienen un oligómero lipídico hidrófobo unido covalentemente a una cadena de péptido o semejante a péptido (p. ej., aminoácido no proteinogénico o aminoácido único) y un resto de carbohidrato, varias secuencias de aminoácidos y ácidos nucleicos que codifican componentes de rutas biosintéticas para estos biotensioactivos, y métodos para preparar y usar estos biotensioactivos.

Antecedentes

- 10 Los tensioactivos son compuestos químicos anfifílicos que poseen restos tanto hidrófobos como hidrófilos que les permiten interaccionar con sistemas polares y no polares. Los tensioactivos ejercen su actividad en interfaces entre diferentes fases (gas, líquido, sólido) y, como resultado, presentan un intervalo de funciones que incluyen, pero no se limitan a, la capacidad de actuar como detergentes, emulsionantes, agentes humectantes y agentes espumantes. La mayor parte de los tensioactivos químicos son sulfatos o sulfonatos de alquilo derivados de fuentes petro u
- 15 oleoquímicas. El uso de estos productos ha crecido de forma constante con un consumo mundial estimado de 13 millones de toneladas en 2008 y un valor de mercado estimado de 27 billones de dólares (USD) en 2012. En respuesta a preocupaciones medioambientales y de sostenibilidad, muchas empresas que utilizan tensioactivos químicos en sus productos han explorado alternativas responsables con el medioambiente como reemplazos parciales o totales de los tensioactivos químicos. Una alternativa a los tensioactivos químicos son los biotensioactivos, que son moléculas
- 20 activas en superficie que se originan a partir de microorganismos. Estos tensioactivos ofrecen ventajas sobre los tensioactivos químicos tales como la producción a partir de productos base producidos de forma sostenible, biodegradabilidad y menor toxicidad.

US 2013/0331466 describe soforolípidos modificados como agentes solubilizantes de aceite.

US 2013/0085067 describe soforolípidos modificados para la inhibición de patógenos de plantas.

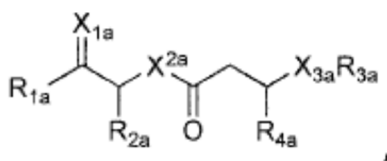
Sumario

- 25 Se descubrió que la bacteria *Variovorax paradoxus* RKNM-096, depositada el 10 de abril, 2015 como número de acceso NRRL B-67038 bajo los términos del Tratado de Budapest con la Agricultural Research Service Culture Collection (NRRL, 1818 North University Street, Peoria, Illinois, 61064) produce una clase previamente desconocida de biotensioactivos denominados "glicolipopéptidos". A diferencia de los biotensioactivos conocidos, los
- 30 glicolipopéptidos típicamente contienen un oligómero lipídico hidrófobo unido covalentemente a una cadena peptídica y un resto de carbohidrato.

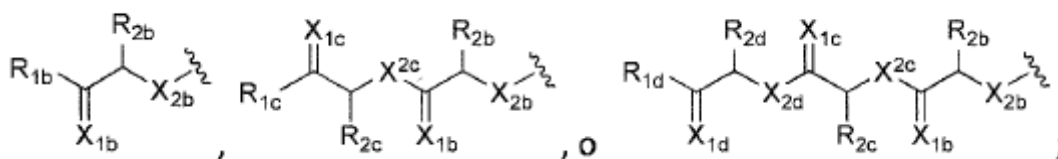
- El depósito de NRRL B-67038 en apoyo de esta solicitud lo hizo Nautilus Bioscience Canada Inc., 550 Univ. Ave., Charlottetown, PE, Canadá, C1A4P3. Nautilus Bioscience Canada Inc. autoriza al solicitante a hacer referencia al material biológico depositado en esta solicitud y proporciona su consentimiento sin reservas e irrevocable de que los
- 35 materiales estén disponibles para el público según las leyes nacionales apropiadas que gobiernan el depósito de estos materiales, tales como la Norma 31 y 33 EPC. La solución de expertos bajo la Norma 32 EPC también se solicita por la presente.

- En la presente memoria se describen biotensioactivos purificados que incluyen un componente lipídico hidrófobo que incluye un extremo carboxilo y un extremo hidroxilo, en donde el componente lipídico está unido covalentemente a (i)
- 40 una cadena de péptido o semejante a péptido en el extremo carboxilo del componente lipídico y (ii) un resto de carbohidrato en el extremo hidroxilo del componente lipídico a través de una unión glicosídica. La cadena de péptido o semejante a péptido puede incluir un dipéptido serina-leucinol, el componente lipídico puede incluir tres restos de ácido β -hidroxialcanoico (p. ej., en donde la longitud de cada cadena acilo del componente lipídico es C₆, C₈, C₁₀, o C₁₂), y el resto de carbohidrato puede incluir un resto de ramnosa unido al componente lipídico a través de una unión
- 45 glicosídica. En determinadas realizaciones, el resto de carbohidrato puede incluir dos restos de ramnosa y/o un grupo acetilo. Los análogos y derivados de estos glicolipopéptidos pueden prepararse por métodos convencionales.

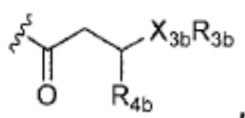
Los glicolipopéptidos pueden tener la estructura:



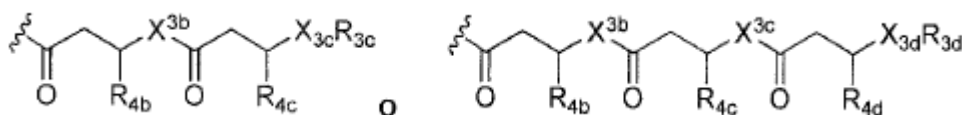
en donde R_{1a} es H, OH, OCH₃, SH, S(CH₃), NH₂, NH(CH₃), N(CH₃)₂, o un péptido o estructura semejante a péptido que tiene la estructura:



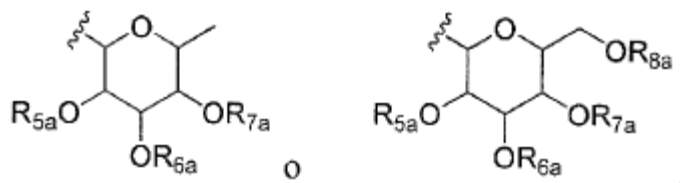
5 en donde R_{1b}, R_{1c}, y R_{1d}, son H, OH, OCH₃, SH, S(CH₃), NH₂, NH(CH₃), o N(CH₃)₂; R_{2a}, R_{2b}, R_{2c}, y R_{2d} son cada uno independientemente una cadena lateral de aminoácido; X_{1a}, X_{1b}, X_{1c}, y X_{1d} son cada uno independientemente un átomo de oxígeno o dos átomos de hidrógeno; X_{2a}, X_{2b}, X_{2c}, y X_{2d} son cada uno independientemente NH, N(CH₃), u O; R_{3a} es una porción de carbohidrato o un monómero lipídico que tiene la estructura:



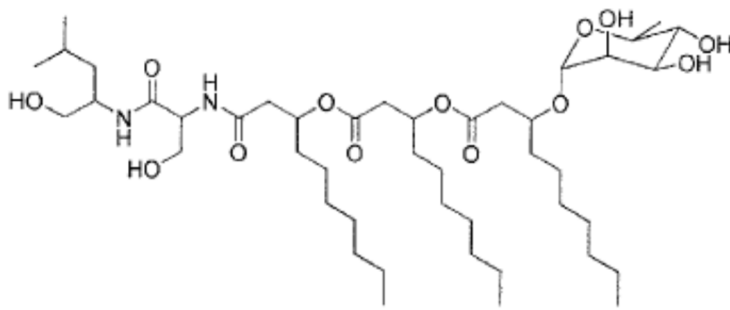
o un oligómero lipídico que tiene la estructura de:

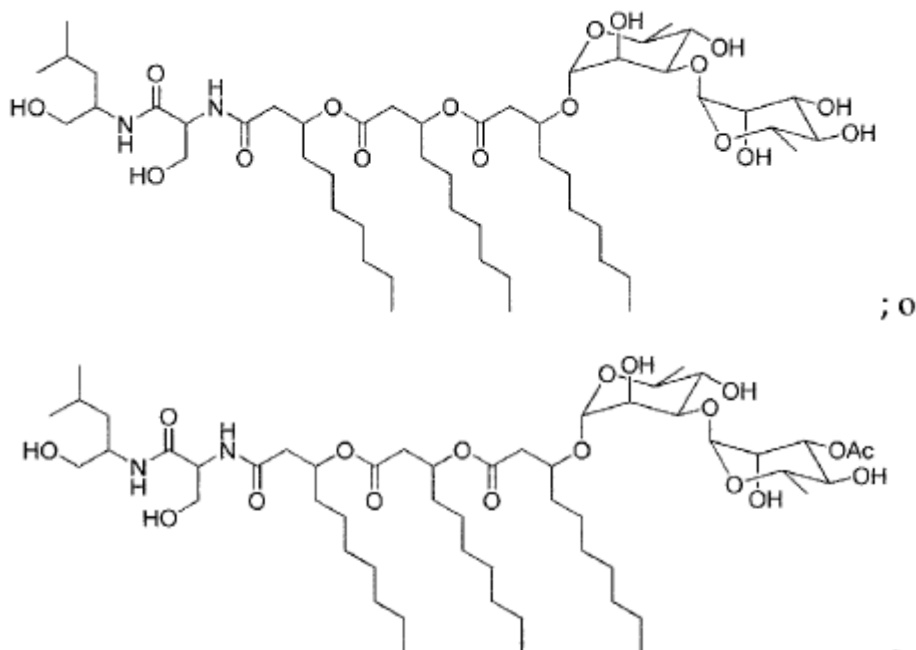


en donde X_{3a}, X_{3b}, X_{3c}, y X_{3d} son cada uno independientemente NH, N(CH₃), u O; R_{3a}, R_{3b}, R_{3c}, y R_{3d} incluye una porción de carbohidrato que incluye un monómero que tiene la estructura:



15 en donde R_{5a}, R_{6a}, R_{7a}, y R_{8a} son cada uno independientemente un átomo de hidrógeno, metilo, acetilo, o un carbohidrato; y R_{4a}, R_{4b}, R_{4c}, y R_{4d} son cada uno independientemente un átomo de hidrógeno, metilo, o un grupo hidrocarburo C₂ a C₁₉ saturado o insaturado de cadena lineal, ramificada, cíclico, o aromático. Los glicolipopéptidos naturales incluyen aquellos que tienen las siguientes estructuras:





En la presente memoria también se describen composiciones emulsionadas (p. ej., emulsiones de aceite en agua o agua en aceite) que incluyen: un componente polar, un componente no polar, y uno o más de los biotensioactivos descritos anteriormente; y un método para preparar una emulsión de agua en aceite o aceite en agua mezclando conjuntamente un componente polar, un componente no polar, y uno o más de los biotensioactivos descritos anteriormente. En la presente memoria se describe además un método para preparar uno de los biotensioactivos descritos anteriormente

(a) aislando un microorganismo que incluye el biotensioactivo,

(b) poniendo el microorganismo en un cultivo en condiciones que promuevan a síntesis del biotensioactivo, y

(c) aislando el biotensioactivo del cultivo; y un microorganismo aislado preparado por ingeniería para producir uno de los biotensioactivos descritos anteriormente, en donde un conjunto de genes heterólogos implicados en la biosíntesis del biotensioactivo se ha introducido en el microorganismo.

A no ser que se defina otra cosa, todos los términos técnicos usados en la presente memoria tienen el mismo significado que el entendido comúnmente por un experto en la técnica a la que pertenece esta invención. Las definiciones entendidas comúnmente de términos químicos y biológicos pueden encontrarse en Rieger et al., Glossary of Genetics: Classical and Molecular, 5ª edición, Springer- Verlag: Nueva York, 1991; y A Dictionary of Chemistry, Ed. J. Daintith, 7ª Ed., Oxford University Press, 2016.

Tal y como se usa en la presente memoria, cuando se hace referencia a un compuesto químico o a una molécula, el término "purificado" significa separado de componentes con los que aparece en la naturaleza o en una mezcla producida artificialmente. Típicamente, una molécula está purificada cuando carece de al menos aproximadamente el 10% (p. ej., al menos el 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, 99,9%, y 100%), en peso (excluyendo el disolvente), de los componentes con los que aparece en la naturaleza o en una mezcla producida artificialmente. La pureza puede medirse por cualquier método apropiado, p. ej., cromatografía en columna, electroforesis en gel de poliacrilamida, o análisis por HPLC.

Por "identidad de secuencia" se quiere decir la relación entre dos secuencias de aminoácidos o entre dos secuencias de nucleótidos. En la presente memoria, el grado de identidad entre dos secuencias de aminoácidos o dos secuencias de desoxirribonucleótidos se determina usando el algoritmo de Needleman-Wunsch (Needleman y Wunsch, 1970, J. Mol. Biol. 48: 443-453) como se implementa en el programa Needle del paquete EMBOSS (EMBOSS: The European Molecular Biology Open Software Suite, Rice et al., 2000, Trends in Genetics 16: 276-277; <http://emboss.org>), preferiblemente la versión 3.0.0 o posterior. Los parámetros opcionales usados son penalización por apertura de hueco de 10, penalización por extensión de hueco de 0,5, y la matriz de sustitución EBLOSUM62 (EMBOSS versión de BLOSUM62) para secuencias de aminoácidos o la matriz de sustitución EDNAFULL (EMBOSS versión de NCBI NUC4.4) para secuencias de nucleótidos. El resultado de Needle marcado "identidad más larga" (obtenido usando la opción -nobrief) se usa como el porcentaje de identidad y se calcula como sigue:

$$\frac{(\text{Residuos de Aminoácidos o Nucleótidos Idénticos} \times 100)}{(\text{Longitud del Alineamiento} - \text{Número Total de Huecos en el Alineamiento})}$$

Aunque pueden usarse métodos y materiales similares o equivalentes a los descritos en la presente memoria en la práctica o ensayo de la presente invención, a continuación, se describen métodos y materiales adecuados. En el caso de conflicto, prevalecerá la presente memoria descriptiva, incluyendo las definiciones.

Además, las realizaciones particulares que se discuten a continuación son solo ilustrativas y no se pretende que sean limitantes.

Breve descripción de los dibujos

La FIG. 1 es una ilustración de correlaciones HMBC ($^1\text{H} \rightarrow ^{13}\text{C}$) y COSY seleccionadas (enlaces en negrita) de NB-RLP1006 y iones de fragmentos asignados de disociación inducida por colisión MS/MS de los glicolipopéptidos.

La FIG. 2 es un gel de poliacrilamida desnaturalizante que muestra R1pE etiquetado con His purificado (A) y el análisis por UPLC-HRMS de reacciones enzimáticas en las que la enzima se incubó con NB-RLP860 y dTDP-L-ramnosa.

La FIG. 3 es una comparación esquemática de la agrupación génica del glicolipopéptido de *V. paradoxus* RKNM-096 con agrupaciones génicas homólogas identificadas en *I. limosus* DSM 16000 y *J. agaricidamnorum* DSM9628. Los genes que codifican proteínas homólogas a proteínas de la agrupación génica de *V. paradoxus* se indican con patrones de flechas rellenas. La identidad y similitud con las proteínas de *V. paradoxus* se indican bajo las flechas (% de identidad/% de similitud). La organización del dominio NRPS se indica bajo flechas que representan genes que codifican sintetasas peptídicas no ribosomales (NRPS). Dominios: C - condensación, A - adenilación, T - tiolación/proteína portadora de peptidilo, R - reductasa. La notación con subíndice indica un sustrato de dominio A posible. Los marcadores por encima de las flechas en las agrupaciones génicas de *I. limosus* y *J. agaricidamnorum* indican las ID de las proteínas.

La FIG.4 es la secuencia de ácido nucleico SEQ ID NO:1.

La FIG.5 es la secuencia de ácido nucleico SEQ ID NO:2.

La FIG.6 es la secuencia de ácido nucleico SEQ ID NO:3.

La FIG.7 es la secuencia de ácido nucleico SEQ ID NO:5.

La FIG.8 es la secuencia de ácido nucleico SEQ ID NO:7.

La FIG.9 es la secuencia de ácido nucleico SEQ ID NO:9.

La FIG.10 es la secuencia de ácido nucleico SEQ ID NO: 11.

La FIG. 11 es la secuencia de ácido nucleico SEQ ID NO: 13.

La FIG.12 es la secuencia de ácido nucleico SEQ ID NO: 15.

La FIG.13 es la secuencia de ácido nucleico SEQ ID NO: 17.

La FIG.14 es la secuencia de ácido nucleico SEQ ID NO: 19.

La FIG.15 es la secuencia de ácido nucleico SEQ ID NO:21.

La FIG.16 es la secuencia de aminoácidos SEQ ID NO:4.

La FIG.17 es la secuencia de aminoácidos SEQ ID NO:6.

La FIG.18 es la secuencia de aminoácidos SEQ ID NO:8.

La FIG.19 es la secuencia de aminoácidos SEQ ID NO:10.

La FIG.20 es la secuencia de aminoácidos SEQ ID NO:12.

La FIG.21 es la secuencia de aminoácidos SEQ ID NO:14.

La FIG.22 es la secuencia de aminoácidos SEQ ID NO:16.

La FIG.23 es la secuencia de aminoácidos SEQ ID NO:18.

La FIG.24 es la secuencia de aminoácidos SEQ ID NO:20.

La FIG.25 es la secuencia de aminoácidos SEQ ID NO:22.

La FIG.26 es la secuencia de aminoácidos SEQ ID NO:23.

La FIG.27 es la secuencia de aminoácidos SEQ ID NO:24.

Descripción detallada

La invención es como se define en las reivindicaciones adjuntas. En la presente memoria se describen composiciones de tensioactivos glicolipopeptídicos, métodos para preparar y usar dichos biotensioactivos, y bacterias y cultivos bacterianos que producen glicolipopéptidos. Las realizaciones preferidas descritas más adelante ilustran la adaptación de estas composiciones y métodos. Sin embargo, a partir de la descripción de estas realizaciones, pueden realizarse y/o llevarse a la práctica otros aspectos de la invención sobre la base de la descripción proporcionada más adelante.

Metodología general

En la presente memoria se describen métodos que implican química orgánica, bioquímica, microbiología, y biología molecular convencionales. Dichos métodos se describen, p. ej., en Clayden et al., Organic Chemistry, Oxford University Press, 1ª edición (2000); Molecular Cloning: A Laboratory Manual, 2ª ed., vol. 1-3, Sambrook et al., ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001; Current Protocols in Molecular Biology, Ausubel et al., ed., Greene Publishing and Wiley-Interscience, Nueva York; y en los distintos volúmenes de Methods in Microbiology y Methods in Biochemistry and Molecular Biology ambos publicados por Elsevier.

Glicolipopéptidos

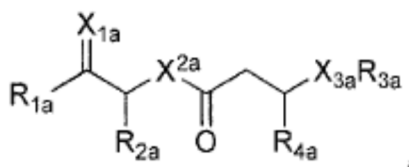
Los glicolipopéptidos naturales y los análogos y derivados sintéticos de los mismos incluyen típicamente un componente lipídico hidrófobo que incluye un extremo carboxilo y un extremo hidroxilo, en donde el componente lipídico está unido covalentemente a (i) una cadena de péptido o semejante a péptido en el extremo carboxilo del componente lipídico y (ii) un resto de carbohidrato en el extremo hidroxilo del componente lipídico a través de una unión glicosídica.

La cadena peptídica puede comprender en el intervalo de entre 2 y 10 aminoácidos, preferiblemente 2 a 8, más preferiblemente 2 a 4 aminoácidos. La cadena peptídica puede comprender lo más preferiblemente 2 aminoácidos. La cadena de péptido o semejante a péptido puede comprender y/o consistir en un dipéptido serina-leucinol.

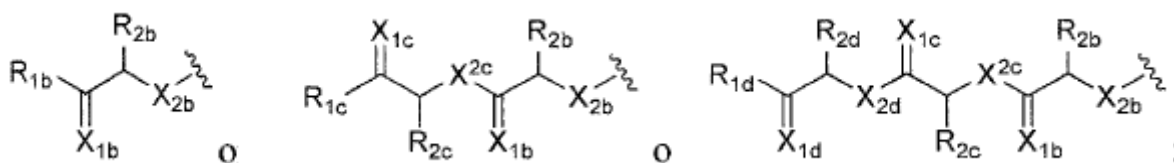
El componente lipídico puede comprender en el intervalo de entre 1 y 6 restos de ácido alcanoico, preferiblemente 2 a 4, y más preferiblemente 3. Lo más preferiblemente el componente lipídico puede incluir tres restos de ácido β-hidroxi alcanoico. La longitud de cada cadena acilo del componente lipídico puede estar en el intervalo de entre C₄ a C₂₀, preferiblemente C₆ a C₁₆, más preferiblemente C₈ a C₁₄. Lo más preferiblemente, la longitud de cada cadena acilo puede seleccionarse de C₈, C₁₀, o C₁₂.

El resto de carbohidrato puede seleccionarse de sacáridos incluyendo glucosa, fructosa, galactosa, manosa, ribosa, o variantes desoxisacárido incluyendo desoxirribosa, fucosa, o ramnosa. Preferiblemente, el resto de carbohidrato es ramnosa. En particular, un resto de ramnosa unido al componente lipídico a través de una unión glicosídica. En determinadas realizaciones, el resto de carbohidrato puede incluir uno, dos, o tres restos de ramnosa y/o un grupo acetilo. Preferiblemente, el resto de carbohidrato incluye dos.

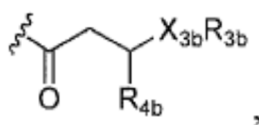
Los glicolipopéptidos pueden incluir la estructura:



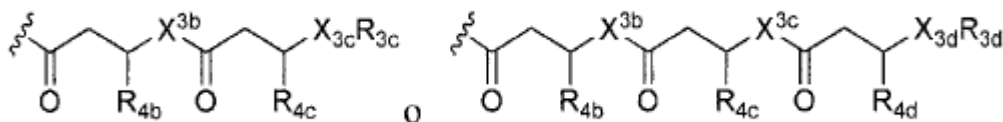
en donde R_{1a} es H, OH, OCH₃, SH, S(CH₃), NH₂, NH(CH₃), N(CH₃)₂, o un péptido o estructura semejante a péptido que tiene la estructura:



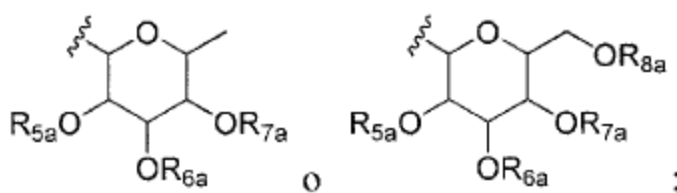
en donde R_{1b}, R_{1c}, y R_{1d}, son H, OH, OCH₃, SH, S(CH₃), NH₂, NH(CH₃), o N(CH₃)₂; R_{2a}, R_{2b}, R_{2c}, y R_{2d} son cada uno independientemente una cadena lateral de aminoácido; X_{1a}, X_{1b}, X_{1c}, y X_{1d} son cada uno independientemente un átomo de oxígeno o dos átomos de hidrógeno; X_{2a}, X_{2b}, X_{2c}, y X_{2d} son cada uno independientemente NH, N(CH₃), u O; R_{3a} es una porción de carbohidrato o un monómero lipídico que tiene la estructura:



o un oligómero lipídico que tiene la estructura de:

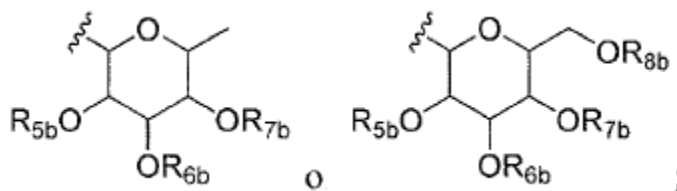


- 5 en donde X_{3a}, X_{3b}, X_{3c}, y X_{3d} son cada uno independientemente NH, N(CH₃), u O; R_{3a}, R_{3b}, R_{3c}, y R_{3d} incluye una porción de carbohidrato que incluye un monómero que tiene la estructura:



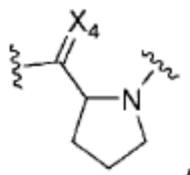
en donde R_{5a}, R_{6a}, R_{7a}, y R_{8a} son cada uno independientemente un átomo de hidrógeno, metilo, acetilo, o un carbohidrato; y R_{4a}, R_{4b}, R_{4c}, y R_{4d} son cada uno independientemente un átomo de hidrógeno, metilo, o un grupo hidrocarburo C₂ a C₁₉ saturado o insaturado, de cadena lineal o ramificada, cíclico, o aromático.

- 10 En lo anterior, al menos uno de R_{6a}, R_{7a}, y R_{8a} puede incluir un monómero de carbohidrato que tiene la estructura:

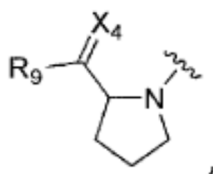


en donde R_{5b}, R_{6b}, R_{7b}, y R_{8b} son cada uno independientemente un átomo de hidrógeno, metilo, acetilo, o un carbohidrato.

- 15 En determinadas realizaciones, el péptido o porción semejante a péptido incluye al menos una prolina o monómero semejante a prolina que tiene la estructura:

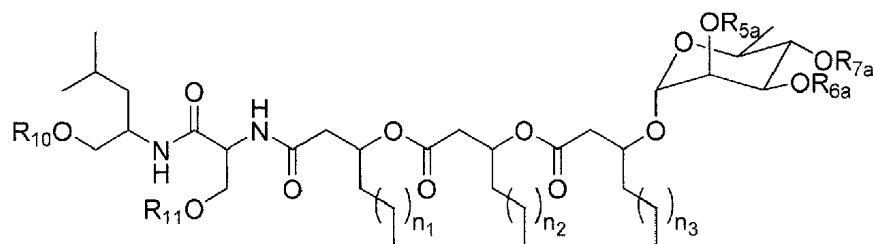


en donde X₄ es un átomo de oxígeno o dos átomos de hidrógeno, o una única prolina o monómero semejante a prolina o una prolina o monómero semejante a prolina terminal que tiene la estructura:

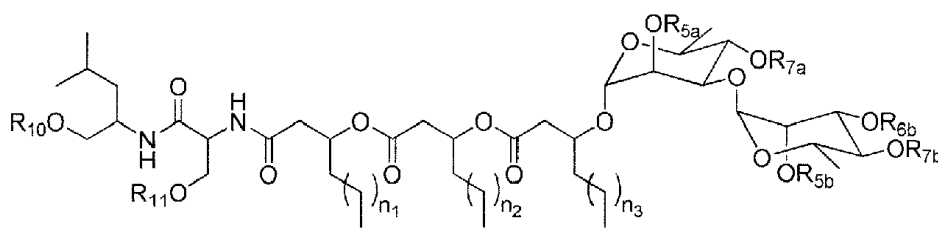


- 20 en donde R₉ es H, OH, OCH₃, SH, S(CH₃), NH₂, NH(CH₃), o N(CH₃)₂; y X₄ es un átomo de oxígeno o dos átomos de hidrógeno.

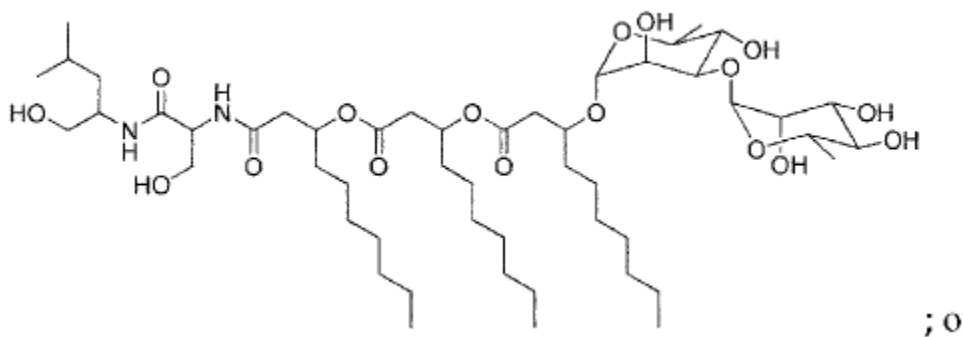
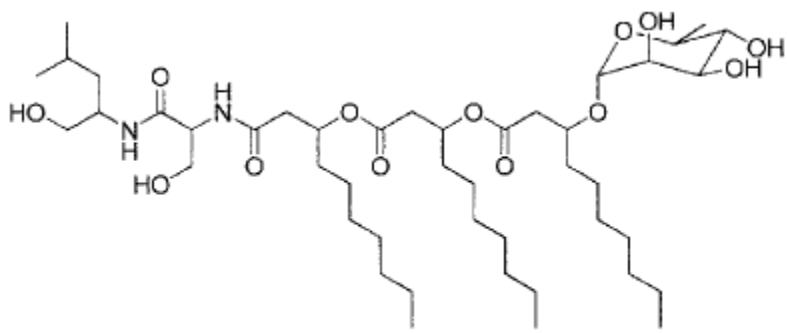
Los glicolipopéptidos pueden tener las siguientes estructuras:

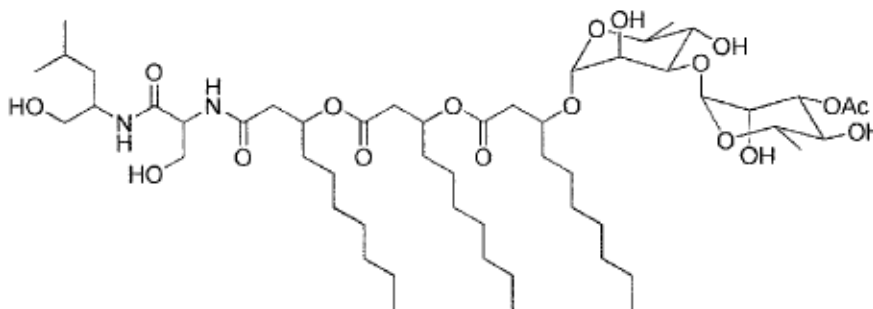


5 en donde R_{5a} , R_{6a} , R_{7a} , R_{10} , y R_{11} son cada uno independientemente un átomo de hidrógeno o acetilo; y n_1 , n_2 , y n_3 son números enteros que varían cada uno independientemente de 1 a 7;



en donde R_{5a} , R_{5b} , R_{6b} , R_{7a} , R_{7b} , R_{10} , y R_{11} son cada uno independientemente un átomo de hidrógeno o acetilo; y n_1 , n_2 , y n_3 son números enteros que varían cada uno independientemente de 1 a 7;





Un experto en la técnica puede preparar derivados, análogos, y otras variantes de los glicolipopeptidos anteriores. Por ejemplo, la composición de aminoácidos y la longitud de la cadena peptídica podrían modificarse de una manera combinatoria, introduciendo aminoácidos proteinogénicos o no naturales para modular la solubilidad, equilibrio hidrófilo-lipófilo (HLB), y otras características de tensioactivo de los glicolipopeptidos. La porción peptídica también puede contener aminoácidos con grupos funcionales cargados, que pueden dar lugar a tensioactivos catiónicos, aniónicos, o zwitteriónicos con aplicaciones de tensioactivo únicas. La funcionalidad ácido carboxílico en la posición C-terminal del péptido también puede reducirse a un grupo hidroxilo primario. De forma similar, la porción lipídica puede contener diversos números (p. ej., 1, 2, 3, 4 o más) de unidades β -hidroxialcanoato, que en sí mismas pueden estar comprendidas por grupos hidrocarburo C_2 a C_{19} saturados o insaturados, de cadena lineal o ramificada, cíclicos, o aromáticos. Los restos de ramnosa podrían estar unidos entre sí a través de uniones glicosídicas 1,2, 1,3, o 1,4, que pueden poseer bien la configuración α o β . Además de la ramnosa, la porción de carbohidrato también puede estar compuesta por glucosa u otras unidades de monosacárido.

Podrían prepararse variantes de los biotensioactivos glicolipopeptídicos de *Variovorax paradoxus* RKNM-096 que tienen propiedades alteradas. Las propiedades alteradas de dichas variantes pueden incluir, pero no se limitan a, alteraciones en las propiedades de emulsificación, espumación y reducción de la tensión superficial presentadas bajo diferentes condiciones físico-químicas tales como, pero no limitadas a, temperatura, pH, y salinidad.

Las variovaricinas descritas en la presente memoria pueden estar al menos un 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, 99, 99,5, 99,9, o 99,99 por ciento purificadas (en peso). Pueden estar en forma cristalina o no cristalina (amorfa), y en algunos casos también pueden obtenerse como sales derivadas de ácidos orgánicos e inorgánicos tales como: acético, trifluoroacético, láctico, cítrico, tartárico, formato, succínico, maleico, malónico, glucónico, clorhídrico, bromhídrico, fosfórico, nítrico, sulfúrico, metanosulfónico y ácidos conocidos similares. Las sales pueden prepararse adaptando procedimientos conocidos comúnmente.

En algunas realizaciones, la composición incluye compuestos adicionales tales como vehículos, otros tensioactivos (p. ej., tensioactivos no glicolipopeptídicos), o compuestos biológicamente activos (tensioactivos no glicolipopeptídicos, tales como agentes farmacéuticos u otros agentes antimicrobianos no glicolipopeptídicos). La adición de los agentes mencionados anteriormente a tensioactivos glicolipopeptídicos puede seleccionarse por un experto en la técnica sobre la base de la aplicación elegida.

La composición puede incluir un vehículo, tal como vehículos farmacéuticamente aceptables convencionales como se describe en *Remington: The Science and Practice of Pharmacy*, The University of the Sciences in Philadelphia, Editores, Lippincott, Williams, & Wilkins, Filadelfia, Pa., 21ª Edición (2005).

Los vehículos farmacéuticamente aceptables varían dependiendo del modo de administración. Las formulaciones fluidas usadas para inyección parenteral pueden incluir fluidos tales como agua, disolución salina fisiológica, dextrosa acuosa o glicerol. Las formulaciones sólidas pueden incluir vehículos sólidos altamente purificados tales como estearato de magnesio, almidón, o lactosa. Las composiciones farmacéuticas también pueden contener cantidades menores de sustancias auxiliares no tóxicas, tales como tampones y conservantes.

En algunas realizaciones, las composiciones incluyen un tensioactivo no glicolipopeptídico. Los ejemplos incluyen tensioactivos no iónicos, catiónicos, aniónicos y anfóteros. Los ejemplos representativos de tensioactivos aniónicos incluyen carboxilatos, sulfonatos, sulfonatos de petróleo, sulfonatos de alquilbenceno, sulfonato de naftaleno, sulfonatos de olefina, sulfatos de alquilo, sulfatos, aceites y grasas naturales sulfatadas, ésteres sulfatados, alcanolamidas sulfatadas, alquilfenoles, alquilfenoles etoxilados y sulfatados y ramnolípidos. Los ejemplos de tensioactivos catiónicos incluyen sales de amonio cuaternario, etilendiaminas sustituidas con N, N, N', N' tetraquis y 2-alkuil-1-hidroxetil-2-imidazolin. Los ejemplos de tensioactivos no iónicos incluyen alcoholes alifáticos etoxilados, tensioactivos de polioxietileno, ésteres carboxílicos, ésteres de polietilén glicol, éster de anhidrosorbitol y derivados etoxilados, glicol ésteres de ácidos grasos, amidas carboxílicas, condensados de monoalcanolamina y amidas de polioxietileno ácido graso. Los ejemplos de tensioactivos anfóteros incluyen sales de sodio del ácido N-coco-3-aminopropiónico, N-tallow-3-iminodipropionato y N-cocoamidetil-N-hidroxietilglicina, así como hidróxido de N-carboximetil-N-dimetil-N-(9-octadecenil) amonio. En realizaciones adicionales, la composición incluye uno o más

alimentos o aditivos alimentarios, agentes cosméticos o farmacéuticos o agentes antimicrobianos (tales como un agente antibacteriano o antifúngico).

Métodos para preparar glicolipopéptidos

Los glicolipopéptidos descritos en la presente memoria pueden prepararse por aislamiento o purificación a partir de cepas bacterianas que los producen, tales como *Variovorax paradoxus* RKNM-096. Como se describe en la sección de Ejemplos más adelante, las bacterias que producen uno o más glicolipopéptidos pueden aislarse de hábitats naturales u obtenerse a partir de fuentes accesibles públicamente. Puede determinarse que las bacterias producen glicolipopéptidos por los métodos descritos en los Ejemplos. La bacteria productora de glicolipopéptido puede ponerse en un biorreactor (recipiente) que contiene medio de cultivo adecuado, e incubarse entonces en condiciones que promueven la replicación bacteriana y la producción de uno o más glicolipopéptidos. El o los glicolipopéptidos producidos pueden purificarse o aislarse de la mezcla de cultivo por técnicas convencionales tales como extracción seguida de separación cromatográfica (p. ej., usando cromatografía líquida de ultra alta resolución). Pueden realizarse análisis químicos (determinación del peso molecular, punto de fusión, RMN, espectroscopía IS, etc.) para confirmar la estructura y pureza del o de los glicolipopéptidos aislados. Alternativamente, los glicolipopéptidos descritos en la presente memoria pueden prepararse por síntesis total o semisíntesis, p. ej., como se describe en la presente memoria.

Agrupaciones génicas de glicolipopéptidos y métodos de uso

Como se describe en el Ejemplo 7 más adelante, se caracterizaron las agrupaciones génicas biosintéticas de glicolipopéptido y de ramnosa de *V. paradoxus* RKNM-096. Los polipéptidos codificados en la agrupación génica funcionan de una manera coordinada para sintetizar la serie NB-RLP de biotensioactivos. La secuencia de nucleótidos que codifica estos genes y las secuencias de aminoácidos de los polipéptidos correspondientes se muestran en el listado de secuencias. Otras secuencias de aminoácidos y las secuencias de ácido nucleico que comparten al menos un 70% (p. ej., al menos un 70, 80, 90, 95, 97, 98, o 99%) de identidad de secuencia con las mostradas en el listado de secuencias también podrían usarse en los métodos y composiciones descritas en la presente memoria, particularmente cuando dichas otras secuencias presentan (o codifican una molécula que presenta) al menos un 50% (p. ej., al menos un 50, 60, 70, 80, 90, o 100%) de la actividad enzimática del polipéptido nativo correspondiente. Las secuencias de ácido nucleico que codifican los mismos polipéptidos descritas en la presente memoria, pero que no se incluyen en el listado de secuencias, también podrían usarse.

Los polinucleótidos anteriores podrían usarse en un método para producir enzimas biosintéticas recombinantes. Como un ejemplo, dicho método podría incluir cultivar una célula huésped (p. ej., *E. coli* u otra célula huésped procariota o eucariota adecuada) que contiene un vector de expresión que tiene una secuencia de ácido nucleico de una o más de SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19 y SEQ ID NO:21 en un medio de cultivo en condiciones adecuadas para la expresión de la proteína recombinante en la célula huésped, y b) aislar la o las proteínas recombinantes de la célula huésped o del medio de cultivo.

También se contempla un método para producir un glicolipopéptido en una célula huésped heteróloga mediante la expresión de la agrupación génica biosintética completa o parcial. Este método podría incluir las etapas de a) cultivar una célula huésped que contiene un vector de expresión que tiene secuencias de ácido nucleico que comprenden SEQ ID NO: 1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19 y SEQ ID NO:21 en un medio de cultivo en condiciones adecuadas para la expresión de las proteínas recombinantes en la célula huésped, y b) aislar los glicolipopéptidos producidos del medio de cultivo.

Se contemplan además métodos para usar una molécula de ácido nucleico que hibrida con o incluye una parte de SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19 o SEQ ID NO:21 como una sonda o cebador de PCR para identificar otros organismos capaces de producir glicolipopéptidos o biotensioactivos estructuralmente similares.

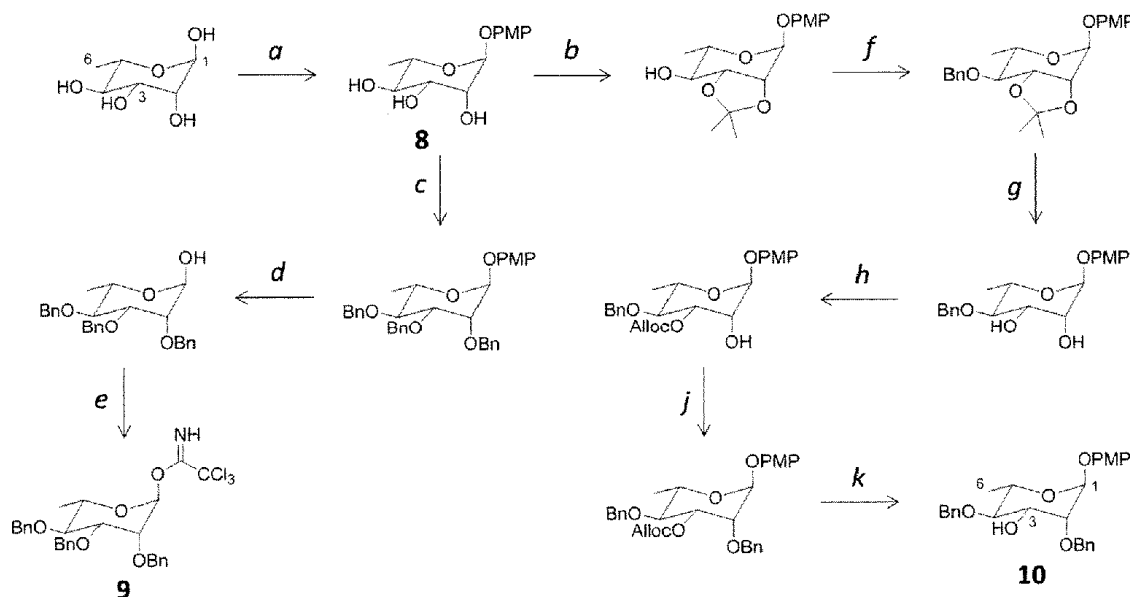
Síntesis

Los compuestos de la presente invención pueden conseguirse usando métodos químicos como se indica en la presente memoria.

La síntesis total de los glicolipopéptidos puede conseguirse usando metodología sintética establecida para ensamblar bloques de construcción disponibles comercialmente. Un análisis retrosintético de NB-RLP1006 (1) demuestra la idoneidad de la síntesis total. Como un ejemplo, un experto en la técnica de la síntesis orgánica puede acoplar el sustituyente dipéptido (4) al ácido tridecanoico (5) y realizar una glicosilación química del intermedio lipopéptido (2) usando el donante de glicosilo 3 como se muestra más adelante. El resto de dipéptido puede prepararse usando métodos de acoplamiento de amida estándar, mientras el ácido tridecanoico puede generarse a partir de *trans*-2-decenoato de etilo disponible comercialmente. Mientras, el sustituyente diramnosa unido en α -1,3 (3) puede ensamblarse usando el donante de glicosilo 6 y el aceptor de glicosilo 7. Se entiende que esta estrategia general, u otras estrategias similares en las que se ensamblan materiales de partida disponibles comercialmente, podría permitir la síntesis de análogos de glicolipopéptido. Por ejemplo, pueden incorporarse diferentes aminoácidos en la porción de

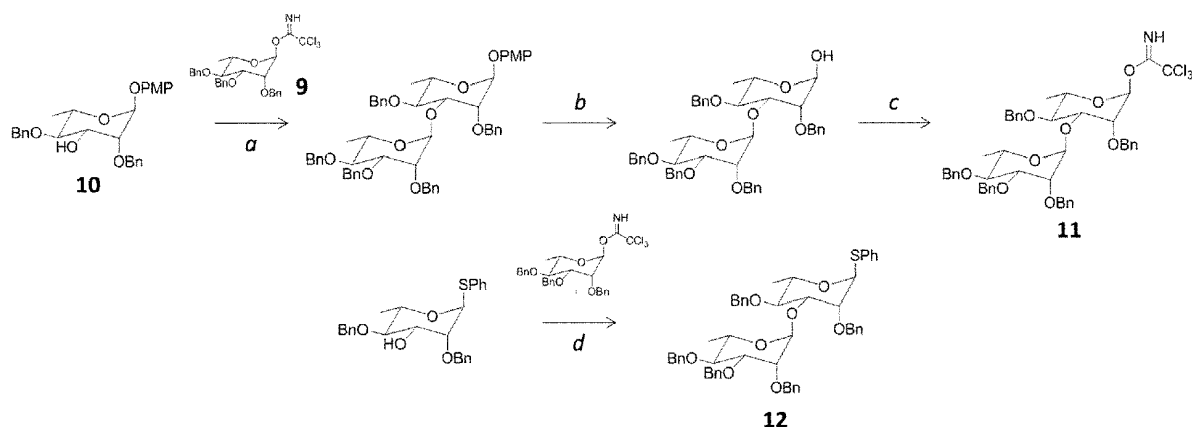
péptido o semejante a péptido, mientras la longitud de la cadena peptídica puede incrementarse o disminuirse. De forma similar, podrían hacerse modificaciones estructurales a las porciones de lípido y carbohidrato de los glicolipopéptidos para producir análogos con características de biotensioactivo potencialmente útiles.

Para generar el sustituyente carbohidrato, diramnosa unida en α -1,3, deben realizarse varias manipulaciones de grupo protector para permitir la glicosilación regioselectiva del azúcar ramnosa en la posición 3-OH (Esquema 2). El *p*-metoxifenil α -L-ramnopiranosido (**8**), que sirve como un precursor sintético para ambos restos de ramnosa, puede sintetizarse a partir de L-ramnosa disponible comercialmente en tres etapas. El azúcar ramnosa terminal puede prepararse entonces por perbencilación de **8** y eliminación del sustituyente *p*-metoxifenilo para permitir la síntesis del tricloroacetimidato de ramosilo (**9**). Mientras, una secuencia de seis etapas de manipulaciones de grupo protector puede proporcionar *p*-metoxifenil 2,4-di-*O*-benciloxiramnopiranosido (**10**) (Cai, X.; et al. Carbohydr. Res. 2010, 345, 1230), que puede glicosilarse en la posición 3-OH para conseguir la unión glicosídica α -1,3 entre los dos sustituyentes ramnosa. Se espera que el efecto anomérico dirija la formación de una unión α -glicosídica con alta estereoselectividad en esta glicosilación química (Takahashi, O.; et al. Carbohydr. Res. 2007, 342, 1202).



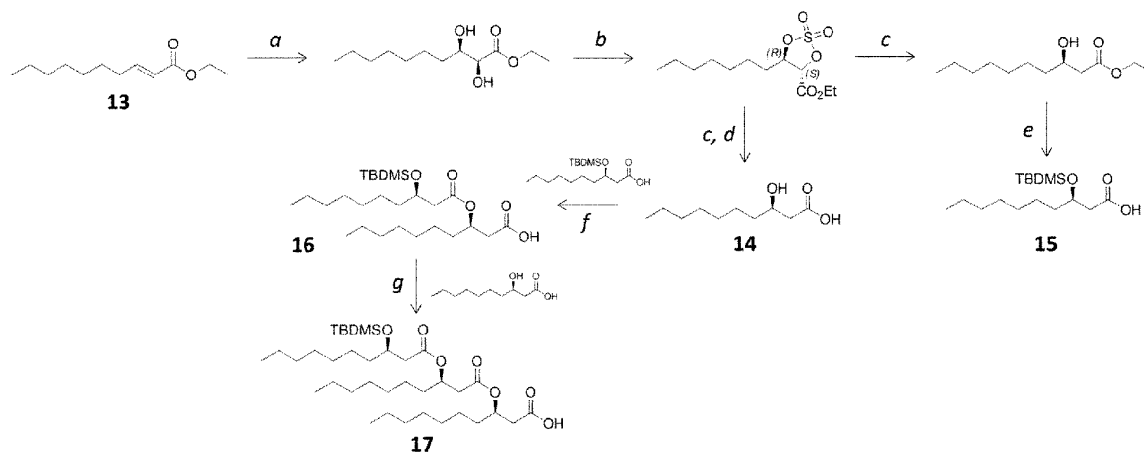
Esquema 2. Síntesis del donante de ramosilo **9** y del aceptor de glicosilo **10** como bloques de construcción para el ensamblaje del sustituyente diramnosa. Reactivos y condiciones: (a) (i) Ac_2O , piridina, 70°C , 16 h. (ii) *p*-metoxifenol, $\text{BF}_3\cdot\text{Et}_2\text{O}$, CH_2Cl_2 , $0^\circ\text{C} \rightarrow 25^\circ\text{C}$, 3 h. (iii) NaOCH_3 , CH_3OH . (b) $\text{CH}_3\text{C}(\text{OCH}_3)_2\text{CH}_3$, DMF, $\text{TsOH}\cdot\text{H}_2\text{O}$. (c) BnBr , Bu_4NI , NaH , DMF, $0^\circ\text{C} \rightarrow 25^\circ\text{C}$, 3 h. (d) CH_3CN al 80%, CAN, 35°C , 30 min. (e) CCl_3CN , DBU, CH_2Cl_2 , 25°C , 30 min. (f) BnBr , Bu_4NI , NaH , DMF, $0^\circ\text{C} \rightarrow 25^\circ\text{C}$, 3 h. (g) AcOH al 70%, 70°C , 3 h. (h) AllocCl , piridina, DMF, CH_2Cl_2 , $-15^\circ\text{C} \rightarrow 0^\circ\text{C}$, 3 h. (j) BnBr , Bu_4NI , NaH , DMF, $0^\circ\text{C} \rightarrow 25^\circ\text{C}$, 3 h.; (k) NaBH_4 , $\text{Pd}[\text{P}(\text{C}_6\text{H}_5)_3]_4$, $\text{CH}_3\text{COONH}_4$, $\text{CH}_3\text{OH}:\text{THF}$ (1:1), -5°C , < 30 min.

Para ensamblar al sustituyente diramnosa, el donante de glicosilo **9** puede unirse al aceptor de glicosilo **10** a través de la activación del tricloroacetimidato anomérico usando $\text{BF}_3\cdot\text{Et}_2\text{O}$ o TMSOTf (Esquema 3). El grupo protector *p*-metoxifenilo anomérico puede reemplazarse entonces por un buen grupo saliente, tal como un tricloroacetimidato, para permitir la glicosilación del resto de ácido decanoico. Alternativamente, puede instalarse un grupo tiofenilo en lugar del grupo *p*-metoxifenilo durante la reacción A del Esquema 2. Esta estrategia permitiría seguir una glicosilación ortogonal dado el papel dual del grupo tiofenilo anomérico como un grupo protector y saliente (Gampe, CM.; et al. Tetrahedron 2011, 67, 9771; Wu, C.-Y.; Wong, C.-H. Top. Curr. Chem. 2011, 301, 223). Se sabe que la síntesis total de NB-RLP860 podría conseguirse usando el donante de ramosilo **9** u otros donantes de ramosilo adecuados. Además, se reconoce que un químico experto podría modificar el resto de carbohidrato de los glicolipopéptidos mediante el uso de donantes de glicosilo distintos de **9**, **11**, o **12**.



Esquema 3. Síntesis de **11** y **12** como donantes de glicosilo para la glicosilación del resto de ácido tridecanoico. Reactivos y condiciones: (a) $\text{BF}_3 \cdot \text{Et}_2\text{O}$ (cat.), **9**, CH_2Cl_2 anhidro, tamices moleculares 3Å, -78°C , 2 h. (b) CH_3CN al 80%, CAN, 35°C , 30 min. (c) CCl_3CN , DBU, CH_2Cl_2 , 25°C , 30 min. (d) $\text{BF}_3 \cdot \text{Et}_2\text{O}$ (cat.), **9**, CH_2Cl_2 anhidro, tamices moleculares 3Å, -78°C , 2 h.

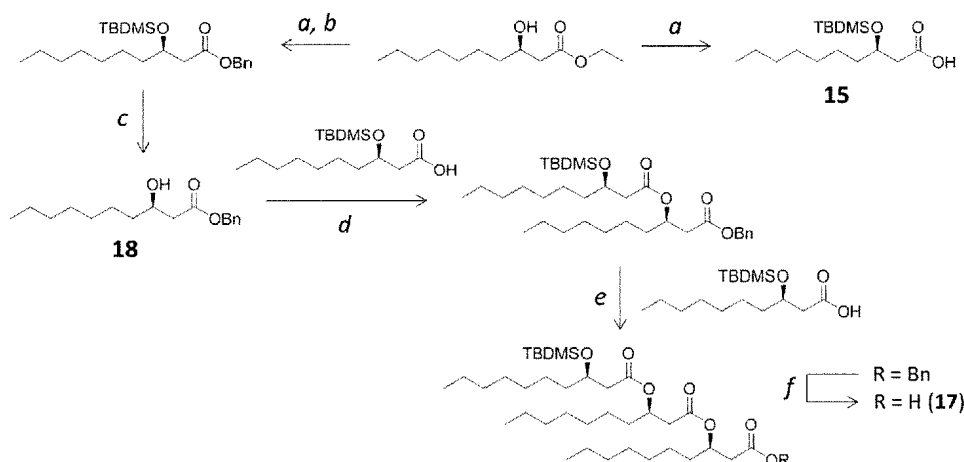
Para ensamblar el resto de ácido tridecanoico, puede servir el *trans*-2-decenoato de etilo (**13**) disponible comercialmente como un precursor para la síntesis de ácido 3-hydroxidecanoico (**14**) a través de una β -oxidación de cinco etapas establecida (Schneekloth, J.S.; et al. Bioorg. Med. Chem. Lett. 2006, 16, 3855; Pandey, S.K.; Kumar, P. Eur. J. Org. Chem. 2007, 369). Los rendimientos globales reportados en la bibliografía para esta síntesis varían entre el 50-85% (Esquema 4). Aunque el ácido ($\pm$)-3-hydroxidecanoico también está disponible comercialmente, este precursor racémico tiene un coste prohibitivo y no representa probablemente una ruta económicamente viable. Después de generar el ácido 3-(*tert*-butildimetilsilil) decanoico (**15**), este bloque de construcción puede unirse dos veces a **14** a través de una esterificación de Steglich para generar el ácido di y tridecanoico sililado (**16** y **17**, respectivamente) de NB-RLP1006 y otros glicolipopéptidos. En esta estrategia, el ácido carboxílico **15** se activa como un éster de *N*-hidroxisuccinimida antes de la adición de **14**, obviando un grupo protector adicional para la funcionalidad ácido carboxílico de **14**.



Esquema 4. Síntesis del resto de ácido tridecanoico (**17**) de NB-RLP1006. Reactivos y condiciones: (a) $(\text{DHQD})_2\text{PHAL}$ (1% en moles), OsO_4 0,1 M (0,5% en moles), K_2CO_3 , $\text{K}_3\text{Fe}(\text{CN})_6$, $\text{CH}_3\text{SO}_2\text{NH}_2$, *t*-BuOH: H_2O (1:1), 0°C , 24 h. (b) (i) SOCl_2 , Et_3N , CH_2Cl_2 , 0°C , 30 min. (ii) RuCl_3 , NaIO_4 , CCl_4 : CH_3CN : H_2O (2:2:3), 0°C , 1 h. (c) NaBH_4 , DMAC, 25°C , 30 min. (d) NaOH , Acetona: H_2O (1:1), 25°C , 16-24 h. (e) (i) TBDMSCl, imidazol, DMF, 25°C . (ii) NaOH , Acetona: H_2O (1:1), 25°C , 16-24 h. (f) (i) ácido 3-benciloxidecanoico, DIC, NHS, $0^\circ\text{C} \rightarrow 25^\circ\text{C}$, 3 h. (ii) ácido 3-hydroxidecanoico, Et_3N , DMAP, 25°C , 3 h. (g) (i) ácido 3-(3-(benciloxi)decanoiloxi) decanoico, DIC, NHS, $0^\circ\text{C} \rightarrow 25^\circ\text{C}$, 3 h. (ii) ácido 3-hydroxidecanoico, Et_3N , DMAP, 25°C , 3 h.

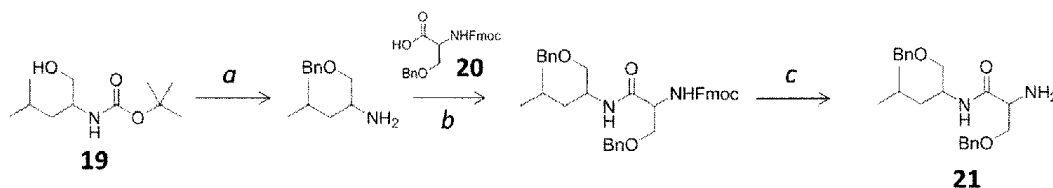
También está disponible una estrategia alternativa en la que el grupo ácido carboxílico de **14** se protege como un éster de bencilo antes de la esterificación (Esquema 5). En esta estrategia, los bloques de construcción **15** y **18** se unen entre sí en una síntesis que requiere etapas adicionales para instalar y eliminar los grupos protectores éter de sililo y éster de bencilo. Se sabe que un químico experto en la técnica de la síntesis orgánica podría utilizar cualquiera de las estrategias para introducir restos de hidrocarburos C_2 a C_{19} saturados o insaturados, de cadena lineal o ramificada,

cíclicos, o aromáticos con el fin de modificar la porción lipídica de los glicolipopéptidos. Se anticipa que los análogos generados a través de esta estrategia también pueden presentar propiedades de tensioactivo.



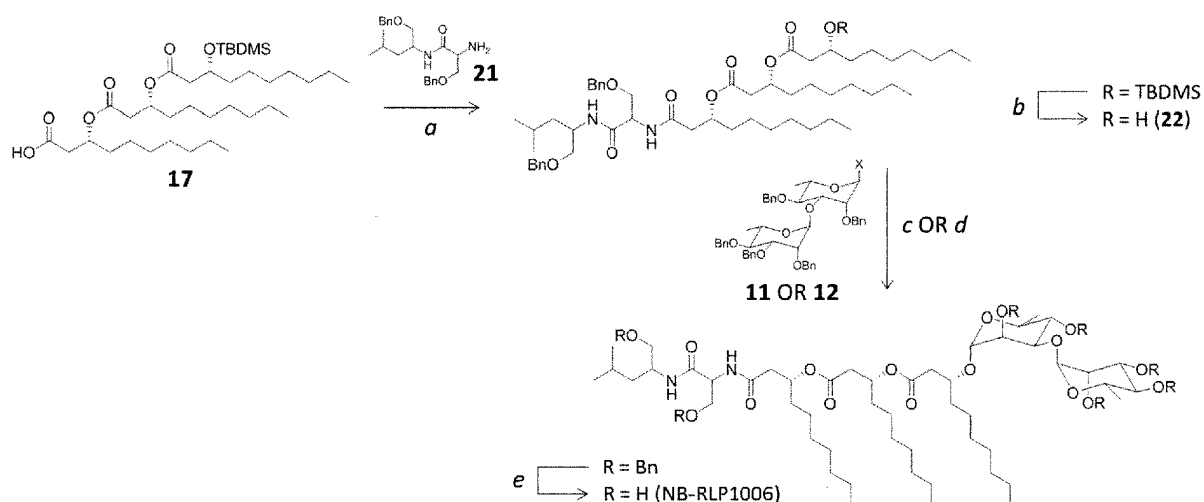
Esquema 5. Ruta alternativa al resto de ácido tridecanoico (**17**) de NB-RLP1006. Reactivos y condiciones: (a) (i) TBDMSCl, imidazol, DMF, 25 °C. (ii) NaOH, Acetona:H₂O (1:1), 25 °C, 16-24 h. (b) (i) DIC, NHS, 0 °C → 25 °C, 3 h. (ii) BnOH, Et₃N, DMAP, 25 °C, 3 h. (c) Bu₄NF, THF, 25 °C, 3 h. (d) (i) ácido 3-(*terc*-butildimetilsililo)decanoico (**15**), DIC, NHS, 0 °C → 25 °C, 3 h. (ii) 3-hidroxi-decanoato de bencilo (**18**), Et₃N, DMAP, 25 °C, 3 h. (e) (i) Bu₄NF, THF, 25 °C, 3 h. (ii) ácido 3-(*terc*-butildimetilsililo)decanoico (**15**), DIC, NHS, 0 °C → 25 °C, 3 h. (iii) 3-(3-hidroxi-decanoilo)decanoato de bencilo, Et₃N, DMAP, 25 °C, 3 h. (f) Pd/C al 10% (20% en peso), H₂, CH₂Cl₂, 25 °C, 16 h.

El dipéptido leucinol-serina puede ensamblarse a partir de Boc-leucinol (**19**) y Fmoc-Ser(Bzl)-OH (**20**) disponibles comercialmente usando química de acoplamiento de amida bien establecida (Esquema 6) (Valeur, E.; Bradley, M. Chem. Soc. Rev. 2009, 38, 606). La secuencia de reacción de cinco etapas implica proteger el grupo hidroxilo primario de **19** como un éter de bencilo y acoplar los dos aminoácidos antes de eliminar el grupo protector 9-fluorenilmetiloxycarbonilo (Fmoc) para generar el dipéptido leucinol-serina (**21**) que es estable para acoplamiento de amida al ácido decanoico. Es concebible que pudieran ensamblarse otros aminoácidos disponibles comercialmente, incluyendo, pero no limitado a, D-aminoácidos y β-aminoácidos, de una manera similar para introducir modificaciones estructurales en la porción de péptido del glicolipopéptido. El extremo C del péptido o porción semejante a péptido podría existir como una funcionalidad ácido carboxílico o reducirse a un grupo hidroxilo primario. Otras modificaciones en la posición C-terminal incluyen, pero no se limitan a, alquilación, acilación, glicosilación, fosforilación, y sulfatación. La longitud de la cadena del péptido o porción semejante a péptido podría incrementarse mediante el acoplamiento de monómeros de aminoácidos adicionales al intermedio dipéptido. Alternativamente, podría acoplarse un único monómero de aminoácido al intermedio ácido tridecanoico (**17**) para disminuir la longitud de la cadena.



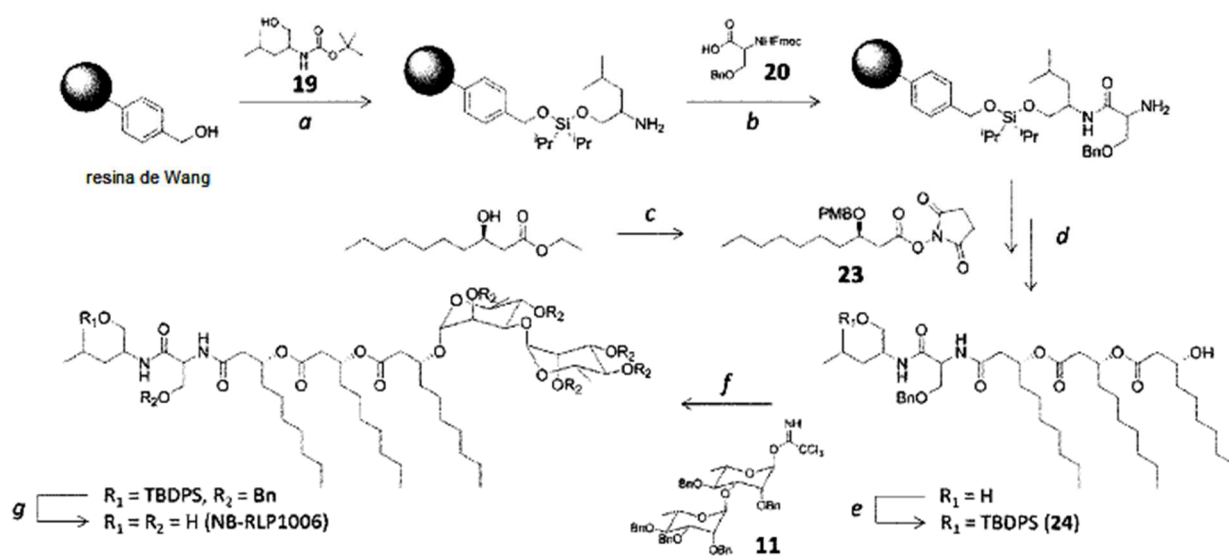
Esquema 6. Preparación del residuo de dipéptido leucinol-serina de NB-RLP1006. Reactivos y condiciones: (a) (i) BnBr, Bu₄NI, NaH, DMF, 0 °C → 25 °C, 3 h. (ii) HCl, CH₃OH, 25 °C, 1 h. (b) (i) Fmoc-Serina(Bzl)-OH, DIC, NHS, 0 °C → 25 °C, 3 h. (ii) Leucinol(Bzl), Et₃N, DMAP, 25 °C, 3 h. (c) Piperidina (20% en vol), DMF, 25 °C, 3 h.

El ácido tridecanoico (**17**) puede experimentar fácilmente acoplamiento de amida al dipéptido bencilado (**21**), después de lo cual el grupo protector *terc*-butildimetil sililo puede eliminarse usando fluoruro de tetrabutilamonio para proporcionar el aceptor de glicosilo **22** (Esquema 7). La glicosilación de **22** bien con **11** o **12**, seguida de desprotección global a través de hidrogenólisis de los éteres de bencilo, suministra el glicolipopéptido NB-RLP1006 desprotegido.



Esquema 7. Etapas finales en el ensamblaje del glicolipopéptido NB-RLP1006. Reactivos y condiciones: (a) (i) DIC, NHS, 0 °C → 25 °C, 3 h. (ii) Leucinol(Bzl)-Ser(Bzl)NH₂ (**21**), Et₃N, DMAP, 25 °C, 3 h. (b) Bu₄NF, THF, 25 °C, 3 h. (c) Si el donante de glicosilo **11** (X = CNHCl₃), BF₃·Et₂O (cat.), tamices moleculares 3 Å, CH₂Cl₂ anhidro, -78 °C, 2 h. (d) Si donante de glicosilo **12** (X = SPh), NIS, TfOH (10% en moles), 1,2-DCE, tamices moleculares 4 Å, (e) Pd/C al 10% (20% en peso), H₂, CH₂Cl₂, 25 °C, 16 h.

Aunque la síntesis total de NB-RLP1006 puede requerir entre 14-18 etapas (secuencia lineal más larga, 34-40 etapas en total), la síntesis podría acelerarse utilizando técnicas sintéticas en fase sólida, en las que el residuo de leucinol terminal se inmoviliza en un soporte sólido. Por ejemplo, el Leucinol(Bzl) (**19**) puede unirse a un grupo *p*-alcoxibencil hidroxilo unido a poliestireno (resina de Wang) a través de una unión éter de sililo (Esquema 8) (Scott, P.J.H. Linker Strategies in Solid-Phase Synthesis, John Wiley & Sons Ltd: Chichester, Reino Unido, 2009; p 50-51). Siguiendo las metodologías de acoplamiento de amida y esterificación de Steglich descritas previamente (Coin, I.; et al. Nat. Protoc. 2007, 2, 3247.), los residuos remanentes de serina y ácido decanoico pueden unirse entonces en una estrategia por etapas usando Fmoc-Ser(Bzl)-OH (**20**) y ácido 3-(*tert*-butildimetilsilil)decanoico (**23**). Después de liberar el intermedio lipopéptido, el grupo hidroxilo primario puede protegerse selectivamente como un éter de *tert*-butildifenilsililo para proporcionar el aceptor de glicosilo **24**. El glicolipopéptido NB-RLP1006 puede sintetizarse entonces por glicosilación química y eliminación de los grupos protectores éter de sililo y bencilo. También pueden producirse análogos de los glicolipopéptidos usando técnicas sintéticas en fase sólida como se describe en la síntesis en fase sólida anterior de NB-RLP1006.



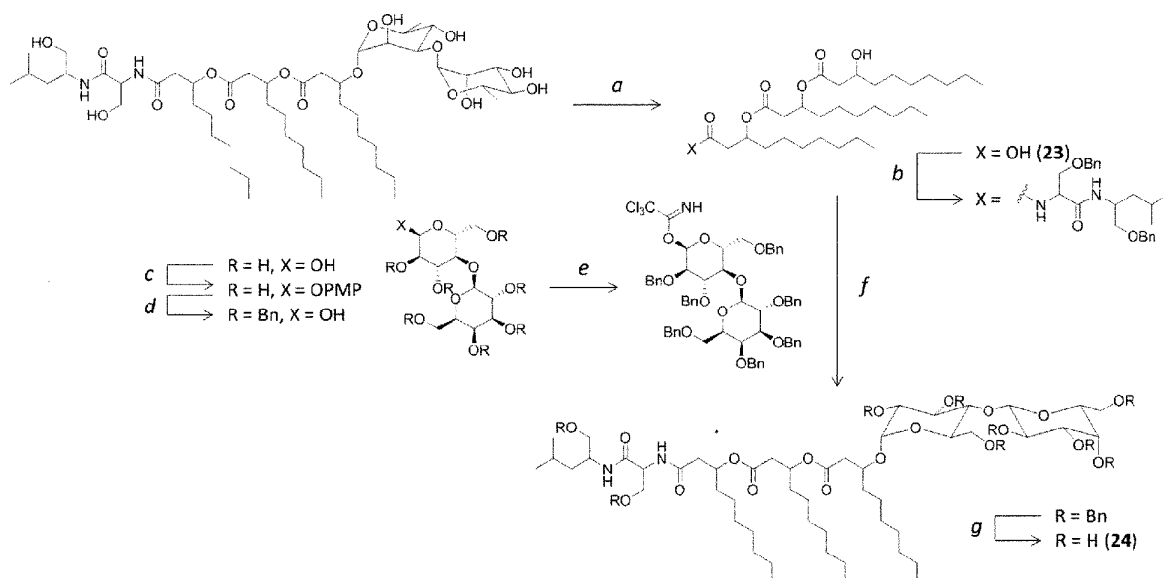
Esquema 8. Síntesis total en fase sólida de NB-RLP1006. Reactivos y condiciones: (a) (i) Leucinol(Bzl) (**19**), ⁱPr₂SiCl₂, imidazol, DMF, 25 °C, 1 h. (ii) HCl, CH₃OH, 25 °C, 1 h. (b) (i) Fmoc-Ser(Bzl)-OH (**20**), DIC, NHS, 0 °C → 25 °C, 3 h. (ii) Et₃N, DMAP, 25 °C, 3 h. (c) (i) PMBCl, NaH, DMF, 0 °C → 25 °C, 3 h. (ii) NaOH, Acetona:H₂O (3:1), 25 °C, 16-24 h. (iii) DIC, NHS, 0 °C → 25 °C, 3 h. (d) (i) **23**, Et₃N, DMAP, 25 °C, 3 h. (ii) CH₃CN al 80%, CAN, 35 °C, 30 min. (iii) **23**, Et₃N, DMAP, 25 °C, 3 h. (iv) CH₃CN al 80%, CAN, 35 °C, 30 min. (v) **23**, Et₃N, DMAP, 25 °C, 3 h. (vi) CH₃CN al

80%, CAN, 35 °C, 30 min. (vii) Bu₄NF, THF, 25 °C, 3 h. (e) TBDPSCI, imidazol, DMAP, DMF, 25 °C, 2 h. (f) BF₃•Et₂O (cat.), CH₂Cl₂ anhidro, -78 °C, 2 h. (g) (i) Bu₄NF, THF, 25 °C, 3 h. (ii) Pd/C al 10% (20% en peso), H₂, CH₂Cl₂, 25 °C, 16 h.

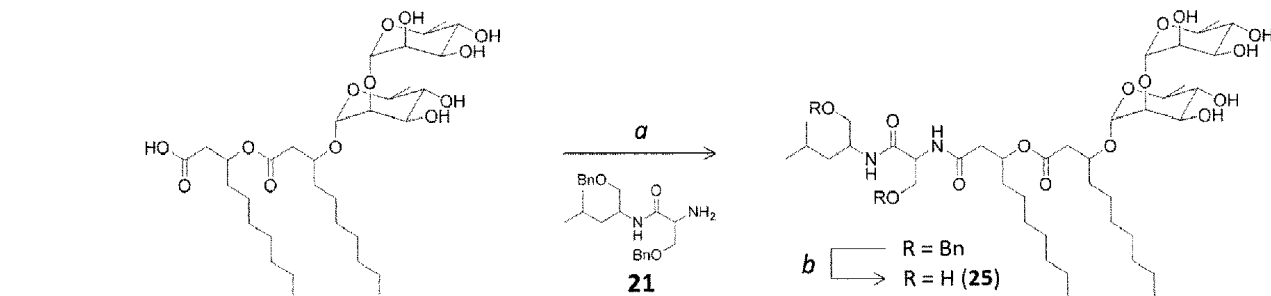
Semisíntesis de los glicolipopéptidos

- 5 Los análogos sintéticos de NB-RLP1006 y otros glicolipopéptidos también pueden ser interesantes para evaluar las relaciones estructura-actividad de esta clase de biotensioactivos. A diferencia de la síntesis total, una semisíntesis podría representar una estrategia rápida para desarrollar varios análogos de glicolipopéptido. Por ejemplo, las estrategias pueden implicar una semisíntesis del ácido tridecanoico (**23**) por hidrólisis ácida de la mezcla de glicolipopéptidos (Esquema 9). Véase, Miao, S.; et al. J. Agric. Food Chem. 2015, 63, 3367. El ácido tridecanoico (**23**)
- 10 podría acoplarse entonces a la porción de péptido y glicosilarse con disacáridos disponibles comercialmente, tales como lactosa o maltosa, para generar nuevos análogos de glicolipopéptido (p. ej., **24**). La aglicona de glicolipopéptidos también puede producirse por *V. paradoxus* RKNM-096 y experimentar una glicosilación química para producir análogos similares. También se sabe que los ramnolípidos podrían utilizarse como un precursor avanzado y unirse a varios dipéptidos (p. ej., **21**) a través del grupo funcional ácido carboxílico para producir glicolipopéptidos similares a NB-RLP1006 (Esquema 10). Dada la disponibilidad comercial de los ramnolípidos, concebiblemente un experto en la técnica de la síntesis orgánica también reconocería que las cadenas peptídicas distintas de leucinol-serina podrían introducirse para expandir la diversidad estructural de los análogos de glicolipopéptido accesibles a través de esta estrategia semisintética.

20



- Esquema 9. Ruta semisintética alternativa a análogos de glicolipopéptido a partir del ácido tridecanoico (**23**) y disacáridos disponibles comercialmente. Reactivos y condiciones: (a) HCl 1M, CH₃CN:H₂O (1:1), 95 °C, 16 h. (b) (i) DIC, NHS, 0 °C → 25 °C, 3 h. (ii) Leucinol(Bzl)-Ser(Bzl)NH₂ (**21**), Et₃N, DMAP, 25 °C, 3 h. (c) (i) Ac₂O, piridina, 70 °C, 16 h. (ii) *p*-metoxifenol, BF₃•Et₂O, CH₂Cl₂, 0 °C → 25 °C, 3 h. (iii) NaOCH₃, CH₃OH. (d) (i) BnBr, Bu₄NI, NaH, DMF, 0 °C → 25 °C, 3 h. (ii) CH₃CN al 80%, CAN, 35 °C, 30 min. (e) CCl₃CN, DBU, CH₂Cl₂, 25 °C, 30 min. (f) BF₃•Et₂O (cat.), CH₂Cl₂ anhidro, -78 °C, 2 h. (g) Pd/C al 10% (20% en peso), H₂, CH₂Cl₂, 25 °C, 16 h.



Esquema 10. Semisíntesis de un análogo de glicolipopéptido (**25**) a partir de ramnolípidos disponibles comercialmente. Reactivos y condiciones: (a) (i) PyBOP, DMF, 0 °C → 25 °C, 2 h. (ii) Leucinol(Bzl)-Ser(Bzl)NH₂ (**21**), Et₃N, 0 °C → 25 °C, 2 h. (b) Pd/C al 10% (20% en peso), H₂, CH₂Cl₂, 25 °C, 16 h.

Concebiblemente, un experto en la técnica de la síntesis orgánica podría aislar glicolipopéptidos naturales a partir de una fermentación microbiana y sintetizar derivados, análogos, y otras variantes estructurales. Por ejemplo, las modificaciones que podrían ocurrir en R_{5a}, R_{5b}, R_{6a}, R_{6b}, R_{7a}, R_{7b}, R₁₀, y R₁₁ incluyen, pero no se limitan a, alquilación, acilación, glicosilación, fosforilación, y sulfatación. Los glicolipopéptidos también podrían experimentar una hidrólisis básica para producir compuestos semejantes a ramnolípidos con propiedades de tensioactivo potencialmente útiles. También se sabe que una reacción de hidrólisis básica proporcionaría NB-RLP374.

Métodos de uso

Los glicolipopéptidos descritos en la presente memoria podrían usarse de forma similar a otros tensioactivos. Pueden usarse, por ejemplo, como detergentes, emulsionantes, dispersantes, agentes humectantes, agentes espumantes, o inhibidores/disruptores de biopelículas. Un uso típico sería para la preparación de emulsiones para formulaciones cosméticas o farmacéuticas (p. ej., emulsiones de agua en aceite o aceite en agua), donde uno o más glicolipopéptidos o derivados o análogos de los mismos se mezclan con un componente polar y un componente no polar.

Las propiedades de los tensioactivos de esta invención también les hacen adecuados como emulsionantes, particularmente en emulsiones de aceite en agua o agua en aceite, p. ej., en aplicaciones de cuidado personal. Los productos en emulsión para el cuidado personal pueden tomar la forma de cremas y leches que incluyen de forma deseable y típica emulsionante para ayudar a la formación y estabilidad de la emulsión. Típicamente, los productos en emulsión para el cuidado personal usan emulsionantes (incluyendo estabilizantes de la emulsión) en cantidades de aproximadamente el 3 a aproximadamente el 5% en peso de la emulsión.

La fase de aceite de dichas emulsiones es típicamente aceites emolientes del tipo usado en productos para el cuidado personal o cosméticos, que los materiales grasos que son líquidos a temperatura ambiente o sólidos a temperatura ambiente, siendo a granel habitualmente un sólido ceroso, siempre que sea líquido a una temperatura elevada, típicamente hasta 100 °C, más habitualmente aproximadamente 80 °C, de manera que dichos emolientes sólidos tienen de forma deseable temperaturas de fusión menores de 100 °C, y habitualmente menores de 70 °C, a las cuales pueden incluirse en y emulsionarse en la composición.

La concentración de la fase de aceite puede variar ampliamente y la cantidad de aceite es típicamente del 1 al 90%, habitualmente del 3 al 60%, más habitualmente del 5 al 40%, particularmente del 8 al 20%, y especialmente del 10 al 15% en peso de la emulsión total. La cantidad de agua (o poliol, p. ej., glicerina) presente en la emulsión es típicamente mayor del 5%, habitualmente del 30 al 90%, más habitualmente del 50 al 90%, particularmente del 70 al 85%, y especialmente del 75 al 80% en peso de la composición total. La cantidad de tensioactivo usada en dichas emulsiones puede estar en el intervalo del 0,001 al 10% en peso de la emulsión, preferiblemente del 0,01 al 6% en peso, más preferiblemente del 0,1 al 5% en peso, más preferiblemente del 1 al 3% en peso. La cantidad de tensioactivo usada en dichas emulsiones es típicamente del 2 al 5,5%, en peso de la emulsión.

Las formulaciones de uso final de dichas emulsiones incluyen hidratantes, cremas solares, productos postsolares, mantecas corporales, cremas en gel, productos que contienen mucho perfume, cremas perfumadas, productos para el cuidado del bebé, acondicionadores del pelo, productos tonificadores de la piel y para el blanqueamiento de la piel, productos sin agua, productos antiperspirantes y desodorantes, productos bronceadores, limpiadores, emulsiones espumantes 2 en 1, emulsiones múltiples, productos sin conservantes, productos sin emulsionante, formulaciones suaves, formulaciones exfoliantes, p. ej., que contienen lechos sólidos, formulaciones de silicona en agua, productos que contienen pigmentos, emulsiones pulverizables, cosméticos con color, acondicionares, productos de ducha, emulsiones espumantes, desmaquillante, desmaquillante de ojos, y toallitas. Un tipo de formulación preferida es un protector solar que contiene uno más protectores solares orgánicos y/o protectores solares inorgánicos, tales como óxidos metálicos, pero que incluye de forma deseable al menos un dióxido de titanio y/u óxido de cinc particulado.

Todas las características descritas en la presente memoria pueden combinarse con cualquiera de los aspectos anteriores, en cualquier combinación. Debe entenderse que la invención no está limitada a los detalles de las realizaciones anteriores, que se describen solo como ejemplo. Son posibles muchas variaciones.

Con el fin de que la presente invención pueda entenderse más fácilmente, se hará ahora referencia, como ejemplo, a la siguiente descripción.

Ejemplos

Ejemplo 1: aislamiento de *Variovorax paradoxus* RKNM-096.

La cepa bacteriana RKNM-096 se aisló a partir de tierra recogida del área Battle Bluffs al oeste de Kamloops, Columbia Británica. RKNM-096 se aisló como una colonia mucoide con pigmentación amarilla, y se purificó por subcultivo seriado. La bacteria se identificó por análisis genético del ARNr 16S, que indicó que RKNM-096 era una cepa de *V. paradoxus*.

Ejemplo 2: identificación de *Variovorax paradoxus* RKNM-096 como un productor de biotensioactivo.

V. paradoxus RKNM-096 se identificó como un productor de biotensioactivo en un cribado dirigido a identificar productores bacterianos de biotensioactivos con propiedades emulsionantes. El ensayo utilizado para identificar productores bacterianos de biotensioactivos fue el ensayo de actividad de emulsificación. En este ensayo, se crecieron cultivos en 10 mL de medio líquido en tubos de vidrio de 25 mm x 150 mm a 30 °C con agitación a 200 rpm durante 5 días. Después de 5 días, las células se retiraron por centrifugación y se mezclaron 3,5 mL del caldo de cultivo sin células con 3,5 mL de queroseno en un tubo de ensayo de 13 mm x 100 mm con tapa de rosca. Los tubos se agitaron con vórtex durante dos minutos y entonces se dejaron permanecer toda la noche a temperatura ambiente, después de lo cual se midieron la altura de la emulsión (h_{emuls}) y la altura total (h_{total}) del líquido en el tubo. Se calculó el índice de emulsificación (E_{24}) usando la ecuación $E_{24} = h_{\text{emuls}}/h_{\text{total}} \times 100\%$. Los caldos de fermentación de *V. paradoxus* RKNM-096 cultivado en caldo ISP2 (maltosa al 0,4%, extracto de levadura al 0,4%, dextrosa al 1,0%, pH 7,0) presentaron un valor de E_{24} del 50,7%.

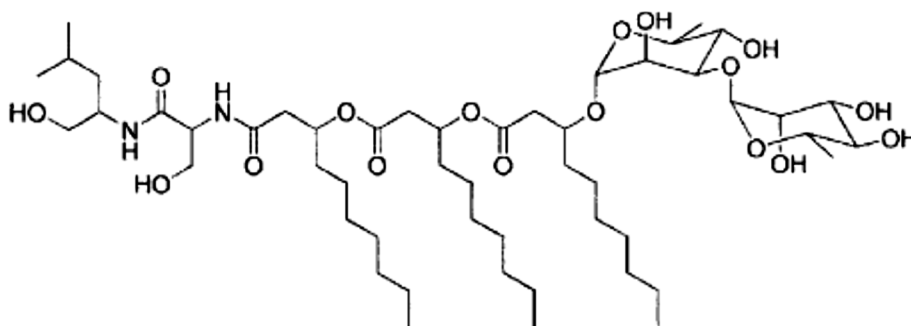
Ejemplo 3: identificación de biotensioactivos glicolipopeptídicos producidos por *Variovorax paradoxus* RKNM-096.

Para determinar si una molécula pequeña era responsable de la actividad de emulsificación observada, *V. paradoxus* RKNM-096 se fermentó en caldo ISP2 como se ha descrito anteriormente y el caldo se extrajo dos veces con 10 mL de acetato de etilo (EtOAc). El extracto de EtOAc se lavó entonces dos veces con 10 mL de agua para eliminar cualesquiera componentes del medio polares remanentes del extracto de EtOAc. Para propósitos de comparación, se extrajo un blanco de medio ISP2 de una manera idéntica. Los extractos de EtOAc se evaporaron en vacío y se reconstituyeron en CH₃OH a una concentración de 0,5 mg/mL.

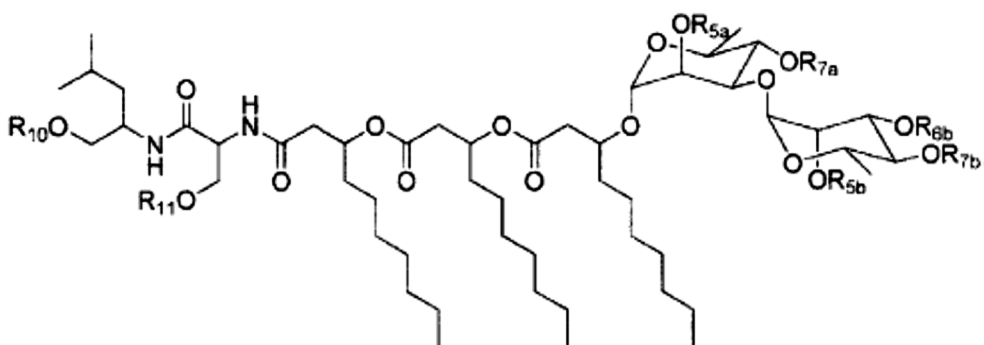
Los extractos se separaron por cromatografía líquida de ultra alto rendimiento (UPLC; Accela™, Thermo Fisher Scientific Mississauga, ON, Canadá) y los eluatos se analizaron con un detector de matriz de fotodiodos (200-600 nm) (PDA; Accela™, Thermo Fisher Scientific Mississauga, ON, Canadá), un detector de dispersión de luz evaporativo (ELSD; Sedex, Sedere, Alfortville, Francia) y un espectrómetro de masa de alta resolución que utiliza ionización por electropulverización (HRESIMS) (Orbitrap Exactive; Thermo Fisher Scientific, Mississauga, ON, Canadá) (modo positivo, monitorización m/z 200-2.000). La separación cromatográfica se consiguió con una columna Kinetex 1,7 μm . C₁₈ 100 Å 50 x 2,1 mm (Phenomenex, Torrance, CA, EE. UU.) y un gradiente lineal de H₂O al 95%/ácido fórmico al 0,1% (FA) (disolvente A) y acetonitrilo al 5% (CH₃CN)/FA al 0,1% (disolvente B) a disolvente B al 100% durante 5 min seguido de mantener el disolvente B al 100% durante 3 min con una tasa de flujo de 400 $\mu\text{L}/\text{min}$. El examen del cromatograma ELSD del extracto de *V. paradoxus* RKNM-096 reveló cinco picos prominentes. El primer pico eluyó a los 0,50 min y estaba presente en el blanco de medio, lo que indica que este pico estaba compuesto por componentes del medio. Los siguientes cuatro picos (1-4) eluyeron a los 3,0 min, 5,04 min, 5,29 min y 5,39 min en el cromatograma ELSD, respectivamente. Estos picos no se observaron en los extractos del blanco de medio, lo que indica que eran productos metabólicos de *V. paradoxus* RKNM-096. El pico 1 eluyó a los 3,00 min y el examen del espectro de masa del pico correspondiente en el cromatograma iónico total (3,04 min) reveló la presencia de dos iones con relaciones de masa a carga (m/z) de 375,2855 y 397,2673, lo que es consistente con la $[\text{M} + \text{H}]^+$ y $[\text{M} + \text{Na}]^+$ anticipada para un compuesto con una fórmula molecular de C₁₉H₃₈N₂O₅ y masa de 374,2781. Los espectros de masa de los picos 2 - 4 se examinaron de una manera idéntica y los iones $[\text{M} + \text{H}]^+$ se identificaron como m/z 1007,6628, 1049,6778, y 1049,6734, respectivamente. La diferencia en masa entre los iones $[\text{M} + \text{H}]^+$ asociados con los picos 3 y 4 fue 4,2 ppm, lo que sugiere que estos dos compuestos tenían probablemente una fórmula molecular idéntica, sin embargo, la ligera diferencia en el tiempo de retención indicó que eran probablemente análogos estructurales muy relacionados.

Los compuestos también se elucidaron usando RMN. Los datos de RMN indicaron la presencia de cuatro grupos carbonilo además de dos residuos de azúcar con desplazamientos químicos de carbono anomérico característicos a δ_{C} 101,4 y δ_{C} 103,9. Las correlaciones clave COSY y HMBC permitieron la caracterización química del leucinol derivado de aminoácido, un residuo de serina, y tres ácidos 3-hidroxicanoicos (Fig. 1). La conectividad entre los diferentes restos se confirmó adicionalmente por espectrometría de masa en tándem. Los dos residuos de desoxihexosa se identificaron por interpretación de las correlaciones COSY ¹H-¹H y análisis de la constante de acoplamiento. El pequeño acoplamiento J presentado por el protón anomérico H-1' (δ_{H} 4,79, d, $J = 1,4$ Hz) y el protón metino H-2' (δ_{H} 3,86, dd, $J = 3,2, 1,4$ Hz) situó a los protones H-1' y H-2' en la posición ecuatorial, mientras el mayor acoplamiento J para H-4' (δ_{H} 3,53 - 3,48, app. t) indicó la relación axial con H-3' y H-5', y por lo tanto sugirió un residuo α -ramnopiranosilo. El pico cruzado de HMBC entre el protón anomérico H-1' y C-3C (δ_{C} 76,5) demostró la unión de este azúcar al resto de ácido 3-hidroxicanoico. El segundo residuo de azúcar también se identificó como un α -ramnopiranosilo sobre la base de los valores de la constante de acoplamiento. El pequeño acoplamiento J para H-1" (δ_{H} 5,01, d, $J = 1,5$ Hz) y H-2" (δ_{H} 3,98, dd, $J = 3,3, 1,5$ Hz) indicó la orientación ecuatorial de estos protones, mientras la mayor constante de acoplamiento para H-4" (δ_{H} 3,40, app. t, $J = 9,5$ Hz) demostró su relación axial con H-3" y H-5". Una correlación HMBC clave entre H-3' y C-1" estableció una unión glicosídica 1,3- α entre los dos restos de ramnopiranosilo.

La estructura de NB-RLP1006 es:



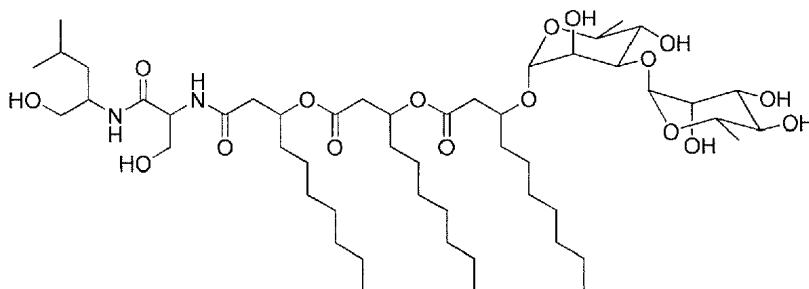
- 5 Los extractos orgánicos de *V. paradoxus* RKNM-096 también se fraccionaron por cromatografía en fase normal automatizada, seguido de HPLC en fase inversa, lo que proporcionó NB-RLP1048A y NB-RLP1048B. Sobre la base del análisis de HRESIMS (NB-RLP1048A: HRESIMS m/z 1049,6778 $[M + H]^+$; NB-RLP1048B: HRESIMS 1049,6734 $[M + H]^+$, calcd para $C_{53}H_{97}N_2O_{18}$, 1049,6731), se determinó que estos compuestos eran análogos monoacetilados de NB-RLP1006. La fórmula molecular aparente de estos compuestos es $C_{53}H_{96}N_2O_{18}$. Sobre la base del análisis de RMN, NB-RLP1048A consistía en una mezcla inseparable de glicolipopéptidos acetilados con la estructura:



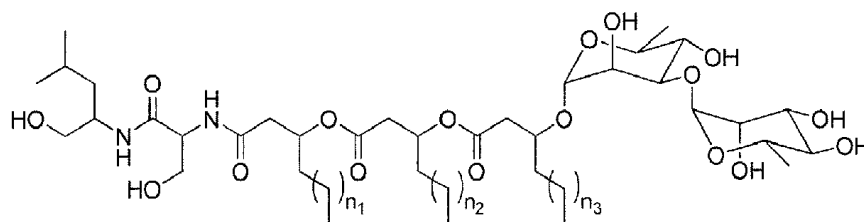
donde cualquier grupo R único es un grupo acetilo, mientras todos los demás grupos R son átomos de hidrógeno.

- 10 La estructura química de NB-RLP1048B se determinó por técnicas espectroscópicas RMN 1D y 2D, confirmando la localización del grupo acetilo en la posición C-3".

La estructura química de NB-RLP1048B es:



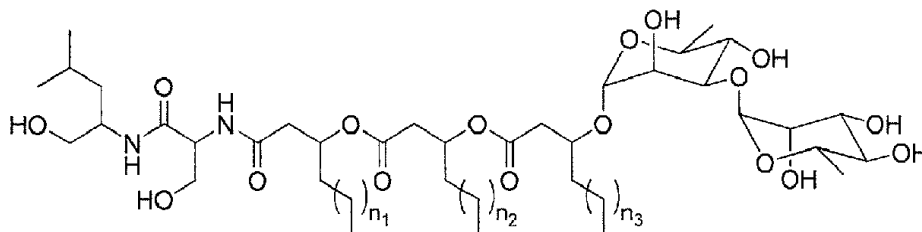
- 15 El esquema de fraccionamiento descrito también rindió varios otros análogos de glicolipopéptido producidos por *V. paradoxus* RKNM-096 en cantidades menores, incluyendo 10,9 mg de NB-RLP978. Los datos de RMN de 1H y ^{13}C fueron casi idénticos a los de NB-RLP1006. La fórmula molecular aparente de NB-RLP978 es $C_{49}H_{90}N_2O_{17}$ (HRESIMS m/z 979,6307 $[M + H]^+$, calcd para $C_{49}H_{91}N_2O_{17}$, 979,6312). Sobre la base de la espectrometría de masa en tandem, se determinó que NB-RLP978 era una mezcla inseparable de tres análogos muy relacionados, NB-RLP978A, NB-RLP978B, y NB-RLP978C, que contienen una cadena acilo C_8 en una de las posiciones de ácido 3-hidroxicanoico.
- 20 La estructura química de NB-RLP978A-C es:



cualquier cadena acilo única es C_8 (es decir, n_1 , n_2 , o $n_3 = 3$) mientras las cadenas acilo remanentes son C_{10} (es decir, $n = 5$). NB-RLP978A: $n_1 = n_2 = 5$, $n_3 = 3$; NB-RLP978B: $n_1 = n_3 = 5$, $n_2 = 3$; NB-RLP978C: $n_1 = 3$, $n_2 = n_3 = 5$.

La purificación por HPLC de fase inversa de NB-RLP1006 y NB-RLP978A-C, también rindió NB-RLP950, una mezcla inseparable de compuestos con una fórmula molecular aparente de $C_{47}H_{86}N_2O_{18}$ (HRESIMS m/z 951,5982 $[M + H]^+$, calcd para $C_{47}H_{86}N_2O_{18}$, 951,5999), lo que es consistente con un análogo de NB-RLP1006 que carece de cuatro grupos metileno. El espectro de RMN de 1H de NB-RLP950 fue casi idéntico al de NB-RLP1006 y NB-RLP978. No se obtuvo un espectro de ^{13}C debido a material insuficiente. Sobre la base de la espectrometría de masa en tándem, se determinó que NB-RLP950 era una mezcla de seis análogos muy relacionados: NB-RLP950A, NB-RLP950B, NB-RLP950C, NB-RLP950D, NB-RLP950E, y NB-RLP950F. Estos análogos de glicolipopéptido bien contienen dos cadenas de acilo C_8 o una cadena de acilo C_6 en las posiciones de ácido 3-hidroxicanoico.

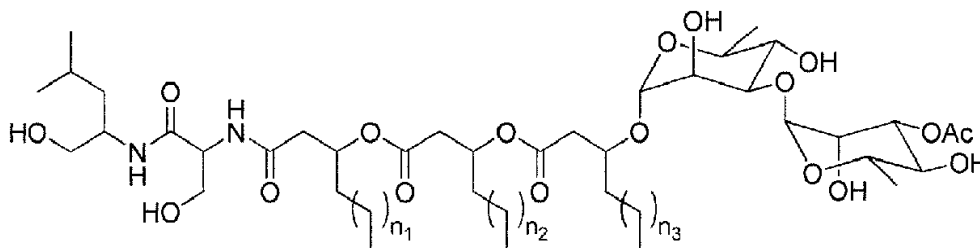
La estructura química de NB-RLP950A-F es:



donde cualesquiera dos cadenas acilo son C_8 (p. ej., $n_1 = n_2 = 3$ y $n_3 = 5$) mientras la cadena acilo remanente es C_6 (es decir, $n = 1$). NB-RLP950A: $n_1 = 5$, $n_2 = n_3 = 3$; NB-RLP950B: $n_2 = 5$, $n_1 = n_3 = 3$; NB-RLP950C: $n_3 = 5$, $n_1 = n_2 = 3$; NB-RLP950D: $n_1 = n_2 = 5$, $n_3 = 1$; NB-RLP950E: $n_1 = n_3 = 5$, $n_2 = 1$; NB-RLP950F: $n_2 = n_3 = 5$, $n_1 = 1$.

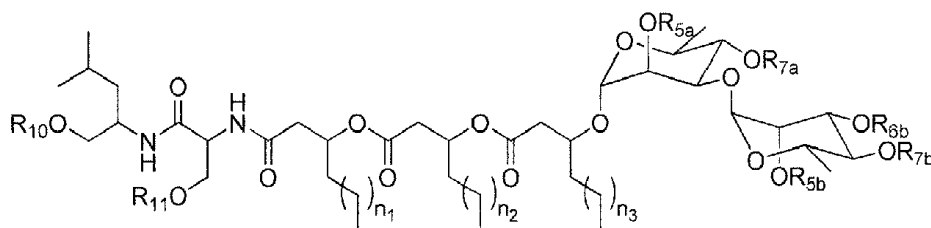
La purificación por HPLC de fase inversa de NB-RLP 1048B también rindió NB-RLP1020. Los datos de RMN de 1H y ^{13}C de NB-RLP1020 fueron casi idénticos a los de NB-RLP1048B. La fórmula molecular aparente de NB-RLP1020 es $C_{51}H_{92}N_2O_{18}$ (HRESIMS m/z 1021,6415 $[M + H]^+$, calcd para $C_{51}H_{93}N_2O_{18}$, 1021,6418). Sobre la base de la espectrometría de masa en tándem, se determinó que NB-RLP1020 era una mezcla inseparable de tres análogos muy relacionados, NB-RLP1020A, NB-RLP1020B, y NB-RLP1020C, que comprenden una cadena acilo C_8 en una de las posiciones del ácido 3-hidroxicanoico.

La estructura química de NB-RLP1020A-C es:



donde cualquier cadena acilo única es C_8 (es decir, n_1 , n_2 , o $n_3 = 3$) mientras las cadenas acilo remanentes son C_{10} (es decir, $n = 5$). NB-RLP1020A: $n_1 = n_2 = 5$, $n_3 = 3$; NB-RLP1020B: $n_1 = n_3 = 5$, $n_2 = 3$; NB-RLP1020C: $n_1 = 3$, $n_2 = n_3 = 5$.

El fraccionamiento por HPLC en fase inversa también rindió una mezcla inseparable de compuestos con una fórmula molecular aparente de $C_{51}H_{92}N_2O_{18}$ (HRESIMS m/z 1021,6477 $[M + H]^+$, calcd para $C_{51}H_{93}N_2O_{18}$, 1021,6418). De manera similar a NB-RLP1020A-C, la estructura de estos compuestos es:

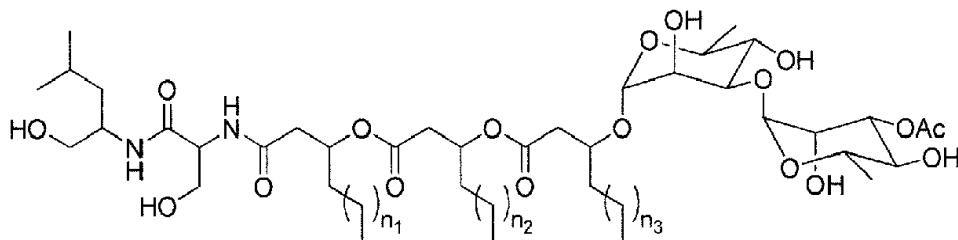


donde cualquier grupo R único es un grupo acetilo, mientras todos los demás grupos R son átomos de hidrógeno y donde cualquier cadena acilo única es C₈ (es decir, n₁, n₂, o n₃ = 3) mientras las cadenas acilo remanentes son C₁₀ (es decir, n = 5).

- 5 El fraccionamiento por HPLC en fase inversa también rindió NB-RLP1076. Los datos de RMN de ¹H y ¹³C fueron casi idénticos a los de NB-RLP1020A-C y NB-RLP1048B. La fórmula molecular aparente de NB-RLP1076 es C₅₅H₁₀₀N₂O₁₈ (HRESIMS *m/z* 1077,7046 [M + H]⁺, calcd para C₅₅H₁₀₁N₂O₁₈, 1077,7044). Sobre la base de la espectrometría de masa en tandem, se determinó que NB-RLP1076 era una mezcla inseparable de tres análogos muy relacionados, NB-RLP1076A, NB-RLP1076B, y NB-RLP1076C, que comprenden una cadena acilo C₁₂ en una de las posiciones del

10

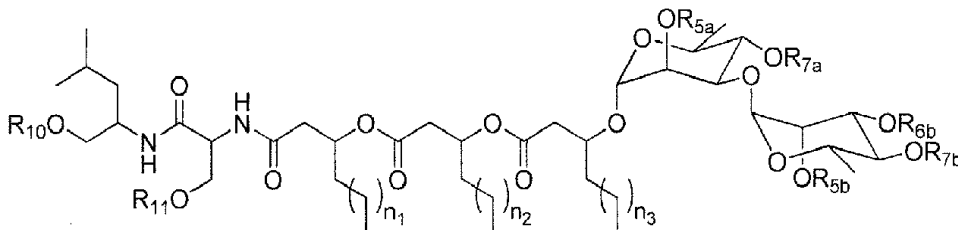
La estructura química de NB-RLP1076 A-C es:



donde cualquier cadena acilo única es C₁₂ (es decir, n₁, n₂, o n₃ = 7) mientras las cadenas acilo remanentes son C₁₀ (es decir, n = 5). NB-RLP1076A: n₁ = n₂ = 5, n₃ = 7; NB-RLP1076B: n₁ = n₃ = 5, n₂ = 7; NB-RLP1076C: n₁ = 7, n₂ = n₃ = 5.

15

El fraccionamiento por HPLC en fase inversa también rindió una mezcla inseparable de compuestos con una fórmula molecular aparente de C₅₅H₁₀₀N₂O₁₈ (HRESIMS *m/z* 1077,7098 [M + H]⁺, calcd para C₅₅H₁₀₁N₂O₁₈, 1077,7044). De manera similar a NB-RLP1076 A-C, la estructura de estos compuestos es:



20

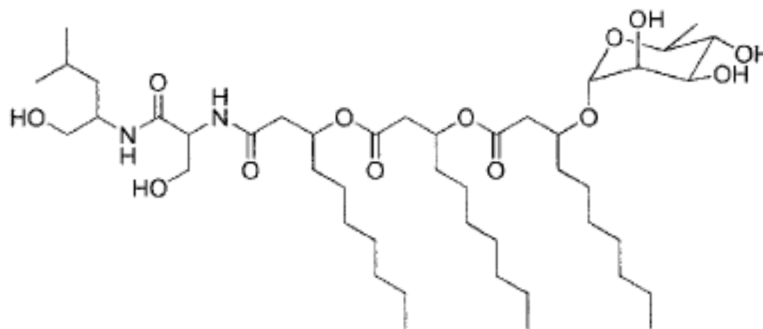
donde cualquier grupo R único es un grupo acetilo, mientras todos los demás grupos R son átomos de hidrógeno y donde cualquier cadena acilo única es C₁₂ (es decir, n₁, n₂, o n₃ = 7) mientras las cadenas acilo remanentes son C₁₀ (es decir, n = 5).

25

Usando porciones de la agrupación génica biosintética del biotensioactivo glicolipopeptídico de *V. paradoxus* RKNM-096 como en sondas *in silico* frente a genomas bacterianos publicados (descrito más adelante), identificamos *Janthinobacterium agaricidamnosum* DSM 9628 como un productor potencial de biotensioactivos glicolipopeptídicos similares a los aislados a partir de *V. paradoxus* RKNM-096. *J. agaricidamnosum* se cultivó y se extrajo como se ha descrito anteriormente para *V. paradoxus* RKNM-096 y el extracto orgánico resultante (110,4 mg) se sometió a cromatografía en fase inversa automatizada con una columna RediSep C₁₈ usando un gradiente de H₂O/CH₃OH. Las fracciones que contenían el glicolipopeptido (77,6 mg) se combinaron y una porción de este material se sometió a separación adicional por HPLC en fase inversa, lo que rindió 17,1 mg de NB-RLP860 y 6,4 mg de NB-RLP832. El análisis de NB-RLP860 por HRESIMS (HRESIMS *m/z* 861,6033 [M + H]⁺, calcd para C₄₅H₈₅N₂O₁₃, 861,6046) indicó una fórmula molecular aparente de C₄₅H₈₄N₂O₁₃ y cinco grados de insaturación. Los datos de RMN de ¹H y ¹³C de NB-RLP860 fueron similares a NB-RLP1006, excepto que el espectro de RMN carecía de resonancias pertenecientes al segundo resto de α-ramnopiranos.

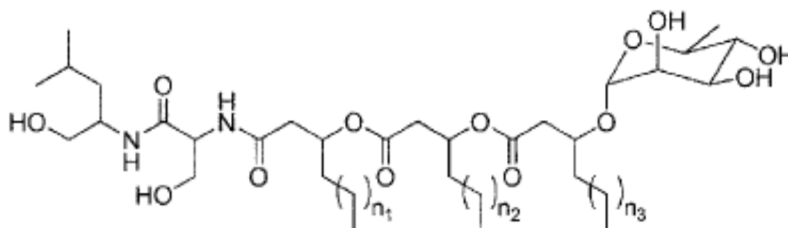
30

La estructura química de NB-RLP860 se determinó por espectroscopía RMN 1D y 2D. La estructura de NB-RLP860 es:



- 5 El análisis de NB-RLP832 por HRESIMS (HRESIMS m/z 833,5734 $[M + H]^+$, calcd para $C_{45}H_{81}N_2O_{13}$, 833,5733) indicó una fórmula molecular aparente de $C_{45}H_{80}N_2O_{13}$. Los datos de RMN de 1H y ^{13}C fueron casi idénticos a los de NB-RLP860. Sobre la base de la espectrometría de masa en tandem, se determinó que NB-RLP832 era una mezcla inseparable de tres análogos muy relacionados, NB-RLP832A, NB-RLP832B, y NB-RLP832C, que comprenden una cadena acilo C_8 en una de las posiciones del ácido 3-hidroxicanoico.

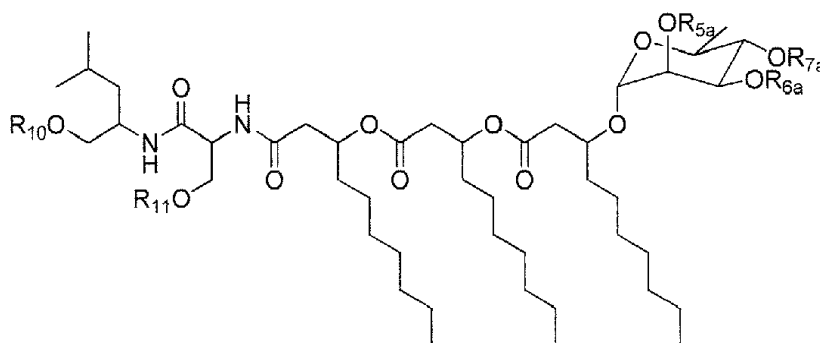
La estructura química de NB-RLP832A-C es



- 10 donde cualquier cadena acilo única es C_8 (es decir, n_1 , n_2 , o $n_3 = 3$) mientras las cadenas acilo remanentes son C_{10} (es decir, $n = 5$). NB-RLP832A: $n_1 = n_2 = 5$, $n_3 = 3$; NB-RLP832B: $n_1 = n_3 = 5$, $n_2 = n_3$; NB-RLP832C: $n_1 = 3$, $n_2 = n_3 = 5$.

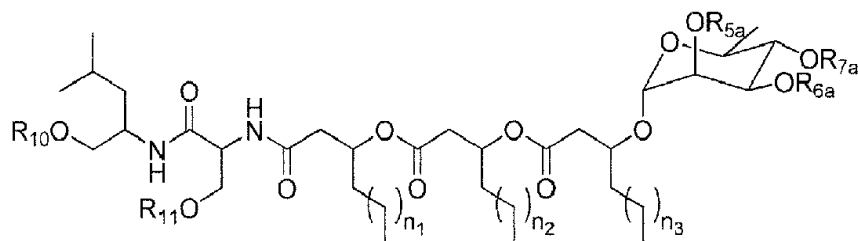
- 15 Los glicolipopéptidos NB-RLP860 y NB-RLP832A-C también se detectaron en pequeñas cantidades en extractos orgánicos de *V. paradoxus* RKNM-096 por análisis de LC-MS. El análisis de los cromatogramas de HRESIMS reveló iones $[M + H]^+$ de m/z 861,6073 y m/z 833,5749, que son consistentes con la m/z predicha de los iones $[M + H]^+$ para NB-RLP860 (calcd para $C_{45}H_{85}N_2O_{13}$, m/z 861,6046 $[M + H]^+$) y NB-RLP832A-C (calcd para $C_{45}H_{81}N_2O_{13}$, m/z 833,5733 $[M + H]^+$).

- 20 El análisis de extractos orgánicos de *V. paradoxus* RKNM-096 también reveló tres picos en el cromatograma HRESIMS que presentan iones $[M + H]^+$ de m/z 903,6213, lo que es consistente con los iones predichos $[M + H]^+$ para un análogo acetilado de NB-RLP860 (m/z 903,6152 $[M + H]^+$). Como estos compuestos se produjeron en pequeñas cantidades, los intentos de determinar inequívocamente sus estructuras por espectroscopía RMN fueron prohibitivos. Estos compuestos no se detectaron en extractos orgánicos de *J. agaricidamnosum* DSM 9628. Dados los iones fragmentos observados de m/z 715,5480 (b) y 598,4310 (bf), estos compuestos se identificaron como glicolipopéptidos acetilados
- 25 NB-RLP902 con la estructura:



donde cualquier grupo R único es un grupo acetilo, mientras todos los demás grupos R son átomos de hidrógeno.

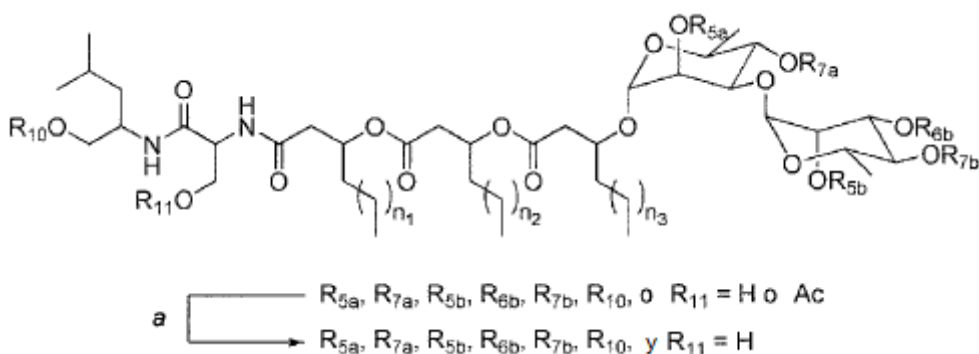
Las fracciones generadas por cromatografía en fase inversa automatizada de extractos orgánicos de *V. paradoxus* RKNM-096 se enriquecieron con NB-RLP902. También se detectó en los cromatogramas HRESIMS de estas fracciones un pequeño pico que presentó un ion $[M + H]^+$ de m/z 875,5888, que es consistente con un análogo de NB-RLP902 que carece de dos grupos metileno. Este ion $[M + H]^+$ no se observó en los extractos orgánicos de *J. agaricidamnorum* DSM 9628. El ion fragmento observado de 687,5164 (b) indica que este compuesto también es un glicolipopéptido. De manera similar a NB-RLP978A-C, NB-RLP1020A-C, y NB-RLP832A-C, se propone que este pico está comprendido por tres compuestos NB-RLP874A-C con la estructura:



donde cualquier grupo R único es un grupo acetilo, mientras todos los demás grupos R son átomos de hidrógeno y donde cualquier cadena acilo única es C_8 (es decir, $n_1, n_2, o n_3 = 3$) mientras las cadenas acilo remanentes son C_{10} (es decir, $n = 5$).

Ejemplo 4: Desacetilación de NB-RLP1048A y otros biotensioactivos glicolipopeptídicos acetilados producidos por *Variovorax paradoxus* RKNM-096.

Se sabe que la cantidad relativa de NB-RLP1006 y glicolipopéptidos acetilados (p. ej., NB-RLP1048A) producida por *V. paradoxus* RKNM-096 puede variar entre lotes usando diferentes medios de cultivo y condiciones de fermentación. Como resultado, las propiedades de tensioactivo del producto de glicolipopéptido extraído también pueden variar. Como la consistencia del producto es importante para ser competitivo en la industria de los biotensioactivos, se desarrolló un método para retirar selectivamente el acetato de $R_{5a}, R_{5b}, R_{6a}, R_{6b}, R_{7a}, R_{7b}, R_{10}$, y R_{11} para generar un producto de glicolipopéptido consistente comprendido por NB-RLP1006 con una pureza $>95\%$ en peso (Esquema 1). El método utiliza NaOH en un intervalo estrecho de concentración para retirar selectivamente los restos de acetato sin inducir una hidrólisis adicional de las uniones amida, éster, o glicosídicas del glicolipopéptido. La concentración de NaOH y el disolvente de la reacción tienen ambos un papel demostrado en el control del grado de hidrólisis y para conseguir selectividad. Las concentraciones óptimas de NaOH son directamente proporcionales a la concentración y composición de los glicolipopéptidos en el medio de reacción. Los disolventes de la reacción con una mayor composición de agua, tal como H_2O :acetona (9:1), mostraron una mejor selectividad y minimizaron la hidrólisis de las uniones éster entre los restos de ácido β -hidroxialcanoico.



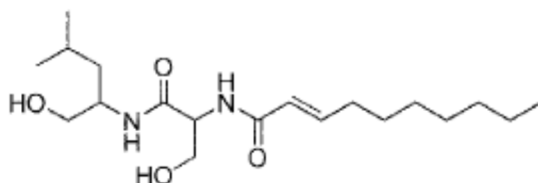
Esquema 1. Hidrólisis selectiva de restos acetato en una mezcla de biotensioactivos glicolipopeptídicos que contienen análogos acetilados de NB-RLP1006 (p. ej., NB-RLP1048A). Reactivos y condiciones: (a) glicolipopéptido (1,0 mg/mL), NaOH (1,5 mM), H_2O :acetona (9:1), 25 °C, 40 h.

Se sabe que la desacetilación de la mezcla de glicolipopéptidos puede conseguirse con variaciones del método descrito en la presente memoria. Es posible que puedan utilizarse bases inorgánicas distintas de NaOH, incluyendo, pero no limitado a, LiOH, KOH, Na_2CO_3 , NH_3 , y NH_4OH , o bases orgánicas, incluyendo, pero no limitado a, hidróxido de tetrabutilamonio o alquilaminas. La desacetilación selectiva también puede conseguirse enzimáticamente usando esterazas, incluyendo, pero no limitado a, acetilesterazas y lipasas.

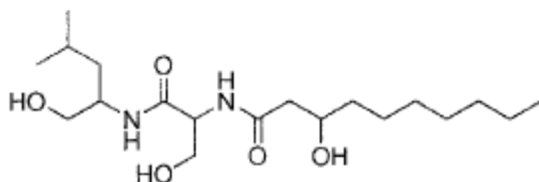
Se sabe que la hidrólisis de la mezcla de glicolipopéptidos produce varios productos, incluyendo, pero no limitado a, los lipopéptidos NB-RLP356 (HRESIMS m/z 357,2745 $[M + H]^+$, calcd para $C_{19}H_{37}N_2O_4$, 357,2748; m/z 379,2567 $[M + Na]^+$, calcd para $C_{19}H_{36}N_2O_4Na$, 379,2565), NB-RLP374 (HRESIMS m/z 375,2851 $[M + H]^+$, calcd para $C_{19}H_{39}N_2O_5$, 375,2854), y NB-RLP526 (HRESIMS m/z 527,4054 $[M + H]^+$, calcd para $C_{29}H_{55}N_2O_6$, 527,4055), y los glicolípidos NB-RLP480 (HRESIMS m/z 481,2599 $[M + H]^+$, calcd para $C_{22}H_{41}O_{11}$, 481,2643; m/z 503,2465 $[M + Na]^+$, calcd para

C₂₂H₄₀O₁₁Na, 503,2463) y NB-RLP650 (HRESIMS m/z 651,3962 [M + H]⁺, calcd para C₃₂H₅₉O₁₃, 651,3950). Dadas sus estructuras anfifílicas, también se espera que estos compuestos se comporten como agentes tensioactivos y puedan presentar propiedades de tensioactivo que pueden ser únicas o complementarias a los glicolipopéptidos. Se sabe que estos compuestos se forman durante el proceso de desacetilación descrito en la presente memoria y están así presentes en el producto final de glicolipopéptido. Aunque normalmente están presentes en pequeñas cantidades (< 5% en peso), estos compuestos pueden contribuir a las características de tensioactivo del producto de glicolipopéptido. La hidrólisis de los glicolipopéptidos también puede ocurrir espontáneamente, por ejemplo, durante la extracción y purificación, para generar estos compuestos. Por ejemplo, el lipopéptido NB-RLP374 se detecta en el extracto orgánico de *V. paradoxus* R NM-096 antes de que el material de glicolipopéptido se someta a cualquier modificación aguas abajo.

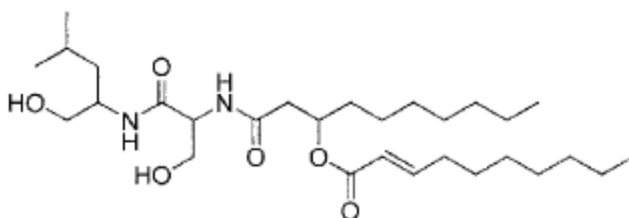
La estructura química de NB-RLP356 es:



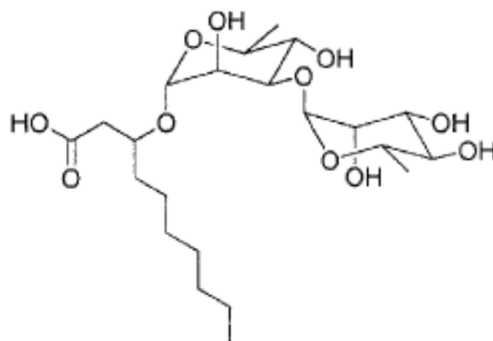
La estructura química de NB-RLP374 es:



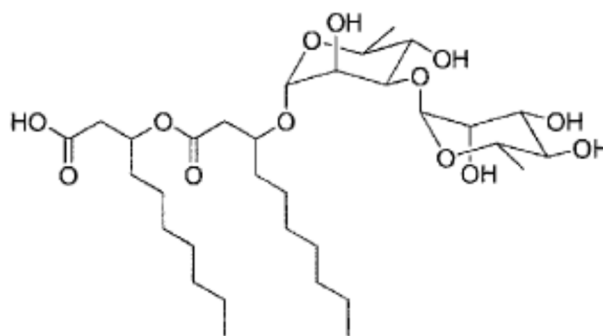
La estructura química de NB-RLP526 es:



La estructura química de NB-RLP480 es:



La estructura química de NB-RLP650 es:



Ejemplo 5: actividad en superficie.

Como se resume en la Tabla 1, las concentraciones micelares críticas (CMC) de NB-RLP1006, NB-RLP978, NB-RLP860, y NB-RLP1048B se determinaron por el método Du Nouy utilizando un microtensiómetro multicanal Kibron Delta-8 (Kibron Inc., Helsinki, Finlandia). Todas las muestras se prepararon en agua desionizada desgasificada (Millipore, Etobicoke, ON, CA) a concentraciones que variaron de 0 a 2,0 mM. Todas las mediciones se registraron entre 24 y 25 °C y se realizaron en duplicado. La concentración micelar crítica tanto de NB-RLP1006 como de NB-RLP978 fue 0,20 mM (0,02% en peso). Las mediciones de tensión superficial indicaron que NB-RLP1006 y NB-RLP978 eran capaces de reducir la tensión superficial del agua de 72 a 35,5 mN/m a sus CMC. Mientras, NB-RLP860 y NB-RLP1048B presentaron valores de CMC de 0,85 mM (0,07 y 0,09% en peso, respectivamente), reduciendo la tensión superficial del agua hasta 36,2 y 36,9 mN/m, respectivamente. La actividad en superficie de NB-RLP1006 se comparó con los ramnolípidos A y B, que se purificaron a partir de una mezcla de ramnolípidos disponible comercialmente (R90; AGAE Technologies, Corvallis, OR, EE. UU.) por HPLC en fase inversa. Los ramnolípidos A y B presentaron ambos una CMC de 0,06 mM (0,003 y 0,004% en peso, respectivamente) en la que la tensión superficial del agua se redujo hasta 28,2 y 39,0 mN/m, respectivamente. Los valores de CMC más altos para NB-RLP860 y NB-RLP1048B pueden deberse a su baja solubilidad acuosa.

Table 1. Propiedades de tensioactivo de glicolipopéptidos aislados en comparación con ramnolípidos. Se muestran la concentración micelar crítica (CMC) y la reducción de la tensión superficial del agua.

Compuesto	CMC (mM)	Tensión Superficial Mínima (mN/m)
NB-RLP1006	0,20	35,5
NB-RLP1048B	0,85	36,9
NB-RLP860	0,85	36,2
NB-RLP978	0,20	35,5
Ramnolípidos A	0,06	28,2
Ramnolípidos B	0,06	39,0

La curvatura característica (C_c) de NB-RLP1006 se determinó usando el modelo de la diferencia hidrófila-lipófila-curvatura neta promedio (HLD-NAC) para calcular el desplazamiento en el potencial químico cuando NB-RLP1006 se transfiere desde la fase de aceite a la acuosa como una función de la salinidad por la siguiente ecuación general:

$$HLD = F(S) - k \times EACN + F(A) - \alpha \times \Delta T + C_c$$

donde $F(S)$ es una función de la salinidad, k es un coeficiente igual a 0,17, EACN (número de carbonos alcano efectivo) es el número de carbonos en la fase de aceite de alcano, α es un coeficiente dependiente del tipo de tensioactivo (iónico, etoxilatos, etc), y ΔT es el efecto de la temperatura. Se prepararon cuatro mezclas de NB-RLP1006 y dihexil sulfosuccinato de sodio (SDHS) con una concentración total de tensioactivo de 1,8 mg/mL usando las siguientes relaciones NB-RLP1006/SDHS: 0, 12, 24, y 40% en peso de NB-RLP1006. Se realizó un barrido de electrolitos para cada mezcla variando la concentración de NaCl del 0 al 6,0% (p/v). Cada mezcla se añadió a un volumen igual de tolueno, que constituyó la fase de aceite, y se agitó vigorosamente. La salinidad óptima (S^*) se identificó como la concentración de NaCl en la cual se formó una microemulsión de Winsor Tipo III, en donde la fase media separada estaba compuesta por un volumen igual de aceite y agua. Se generó un gráfico de las relaciones molares NB-RLP1006/SDHS frente a S^* y se calculó C_c a partir de la línea con mejor ajuste. Se determinó que el valor de C_c para NB-RLP1006 era +5,2, un valor que refleja la naturaleza hidrófoba de este biotensioactivo.

Las propiedades emulsionantes de NB-RLP1006 se determinaron usando el índice de emulsificación como se ha descrito anteriormente. El NB-RLP1006 puro presentó una actividad de emulsificación fuerte con un valor de E_{24} del 53% a 1 mg/mL en agua desionizada. La emulsificación de NB-RLP1006 depende del pH con valores de E_{24} del 8, 38, y 31% a pH 3, 6, y 8, respectivamente. El tipo de emulsión formada por NB-RLP1006 (p. ej., aceite en agua o agua en aceite) se determinó usando el ensayo de la dilución de gota. Se formó una emulsión mezclando vigorosamente una disolución de 1 mg/mL de RLP1006 en agua desionizada con un volumen igual de queroseno durante 1 min. Una porción (20 μ L) de la emulsión se transfirió a 0,5 mL de agua desionizada y 0,5 mL de queroseno y se monitorizó la dilución de la emulsión en cada líquido. La emulsión formada por NB-RLP1006 se dispersó fácilmente en la fase acuosa, indicando que la fase continua de la emulsión era agua y que se formó una emulsión de aceite en agua (o/w) por NB-RLP1006 en estas condiciones.

Estos resultados establecieron que NB-RLP1006 es un biotensioactivo potente capaz de disminuir la tensión superficial del agua hasta 35,5 mN/m con una CMC comparable con la de otros biotensioactivos bien caracterizados, los ramnolípidos A y B. NB-RLP1006 también presenta una actividad de emulsificación fuerte formando emulsiones de o/w en las condiciones descritas en la presente memoria.

Ejemplo 6: ensayo de citotoxicidad de los glicolipopéptidos.

Para evaluar el perfil de seguridad de los glicolipopéptidos, se realizaron ensayos de citotoxicidad frente a dos líneas celulares humanas normales, células de fibroblastos BJ ATCC CRL-2522 y queratinocitos epidérmicos de adulto (HEKa; Life Technologies, Carlsbad, CA, EE. UU.). Los fibroblastos BJ se crecieron y se mantuvieron en 15 mL de medio esencial mínimo de Eagle suplementado con suero fetal bovino (10% v/v), penicilina (100 μ U) y estreptomycin (100 μ g/mL). Las células HEKa se crecieron y se mantuvieron en 15 mL de medio EPI life (Life Technologies, Carlsbad, CA, EE. UU.) suplementado con suplemento de crecimiento HKGS (10% v/v; Life Technologies, Carlsbad, CA, EE. UU.) y 50 μ g/mL de gentamicina (Sigma-Aldrich, St. Louis, MO, EE. UU.). Las células se cultivaron en matraces de cultivo celular T75cm² y se incubaron a 37 °C en una atmósfera humidificada de CO₂ al 5%. Para los fibroblastos BJ, el medio de cultivo se repuso cada dos a tres días y no se dejó que las células superaran el 80% de confluencia. Para las células HEKa, el medio de crecimiento se repuso cada 2 d hasta que las células alcanzaron el 50% de confluencia y después el medio se repuso cada 24 h hasta que se obtuvo el 80% de confluencia. Al 80% de confluencia, las células se contaron, se diluyeron hasta 10.000 células/pocillo en medio de crecimiento sin antibióticos y se transfirieron 90 μ L de suspensión celular a los pocillos de placas de cultivo celular tratadas de 96 pocillos. Las placas se incubaron como antes para permitir que las células se adhirieran a las placas durante 24 h antes del tratamiento. Se usó DMSO como el vehículo a una concentración final del 1%. Todos los compuestos ensayados se volvieron a solubilizar en DMSO y se preparó una serie de dilución para cada línea celular usando el medio de crecimiento de cultivo celular respectivo, de las cuales se añadieron 10 μ L a los pocillos de ensayo rindiendo ocho concentraciones finales que variaron de 512 μ g/mL a 8 μ g/mL por pocillo (volumen final del pocillo de 100 μ L). Los fibroblastos y las células HEKa se incubaron como se ha descrito previamente durante 24 h. Todas las muestras se ensayaron en triplicado. Cada placa contenía cuatro blancos de medio sin inocular (medio + DMSO al 1%), cuatro controles de crecimiento no tratados (medio + DMSO al 1% + células), y una columna que contenía un control positivo de piritiona de cinc diluido de forma seriada. Se añadió AlamarBlue (Life Technologies, Carlsbad, CA, EE. UU.) a cada pocillo 24 h después del tratamiento (10% v/v). Se monitorizó la fluorescencia (560/12 excitación, 590 nm emisión) usando un lector de placas Varioskan Flash Multimode tanto al tiempo cero como 4 h después de la adición de AlamarBlue. Después de restar la fluorescencia a tiempo cero de las lecturas a 4 h, se calculó el porcentaje de viabilidad celular respecto a los pocillos de control de vehículo. Se mostró una baja actividad citotóxica frente a las líneas celulares HEKa y fibroblastos BJ. Los valores observados de CI_{50} y MIC_{90} para los glicolipopéptidos fueron significativamente mayores que el control positivo piritiona de cinc, que sirvió como una referencia de la industria para agentes antimicrobianos tópicos (Tabla 2). Estos resultados indican que los glicolipopéptidos presentan una baja citotoxicidad frente a las células cutáneas humanas y pueden ser así seguros para su uso en aplicaciones que implican el contacto dérmico, tal como productos cosméticos.

Tabla 2. Resultados de los ensayos de citotoxicidad para los glicolipopéptidos. Los valores indican las concentraciones inhibitoras mitad de la máxima (CI_{50}) y la concentración inhibitora mínima que da lugar al 90% de inhibición del crecimiento (MIC_{90}) en μ g/mL. El error se reporta como desviación estándar.

Células Eucariotas				
Compuesto	HEKa (CI_{50})	HEKa (MIC_{90})	BJ (CI_{50})	BJ (MIC_{90})
NB-RLP1006	15,5 $\pm$ 1,7	64 - 128	19,5 $\pm$ 2,4	32
NB-RLP1048B	19,3 $\pm$ 4,0	64 - 128	18,7 $\pm$ 1,6	32
NB-RLP860	15,5 $\pm$ 1,6	128	16,3 $\pm$ 0,3	32
Piritiona de cinc	0,20 $\pm$ 0,001	1	2,2 $\pm$ 0,3	4

Ejemplo 7: Secuenciación de las agrupaciones génicas biosintéticas del glicolipopéptido y de ramnosa de *V. paradoxus* RKNM-096.

Para establecer la base genética para la biosíntesis de los nuevos biotensioactivos glicolipopeptídicos descritos aquí, se secuenció el genoma de *V. paradoxus* RKNM-096. *V. paradoxus* RKNM-096 se cultivó en caldo ISP2 y se aisló el ADN genómico usando el Kit de Aislamiento de ADN Microbiano UltraClean® según las recomendaciones del fabricante (Mo Bio, Carlsbad, CA, EE. UU.). El genoma se secuenció en McGill University y Genome Quebec Innovation Centre (Montreal, QC, CA) usando 2 células SMRT en un secuenciador PacBio RSII (Pacific Biosciences, Menlo Park, CA, EE. UU.). Se generaron un total de 140.476 sublecturas brutas con una longitud promedio de 11.269 pb y el ensamblaje del genoma se consiguió usando un proceso de trabajo HGAP (Chin *et al.* [2013] *Nature Methods* 10, 563). Brevemente, las sublecturas brutas se generaron a partir de archivos de datos raw .bas.h5 PacBio. Se extrajo un valor de corte de la longitud de la sublectura (30X) de las sublecturas y se usó en la etapa de preensamblaje (BLASR), que consiste en el alineamiento de sublecturas cortas en sublecturas largas (Chaisson y Tesler [2012] *BMC Bioinformatics* 13, 238). Como los errores en las lecturas de PacBio son aleatorios, el alineamiento de lecturas cortas múltiples en lecturas más largas permite la corrección de los errores de secuenciación en lecturas largas. Estas lecturas largas corregidas se usaron entonces como semillas en un ensamblaje posterior preparado usando el ensamblador Celera (Myers *et al.* [2000] *Science* 287, 2196), que genera cóntigos. Estos cóntigos se "pulieron" entonces alineando lecturas brutas en cóntigos (BLASR) que se procesaron entonces a través de un algoritmo de llamada de variantes (Quiver) que genera secuencias consenso de alta calidad usando realineamientos locales y puntuaciones de calidad PacBio (Chin *et al.* [2013] *Nature Methods* 10, 563). Se obtuvieron más de 161.717,463 pb de sublecturas largas corregidas y dio lugar al ensamblaje de dos cóntigos. Un cóntigo contenía 7.193.071 pb mientras el otro contenía 1.767 pb. El genoma se anotó usando el servidor RAST (Aziz *et al.* [2008] *BMC Genomics* 9, 75; Overbeek *et al.* [2014] *Nucleic Acid Res.* 42, D206; Brettin *et al.* [2015] *Sci Rep.* 5, 8265). La función de marcos de lectura abiertos (ORF) identificada por la anotación RAST se exploraron adicionalmente por BLASTP (Altschul *et al.* [1997] *Nucleic Acids Res.* 25, 3389) y el análisis del dominio conservado (Marchler-Bauer y Bryant [2004] *Nucleic Acids Res.* 32, W327) de secuencias de aminoácidos deducidas.

Sobre la base de la estructura de NB-RLP1006 se estableció la hipótesis de que su biosíntesis requeriría una NRPS para sintetizar el dipéptido, una o más aciltransferasas para acilar el péptido y generar el resto 3-(3-(3-hidroxicanoiloxi) decanoiloxi) decanoilo y una o más glicosiltransferasas. El escaneo del genoma para genes que codifican NRPS identificó dos loci. Un locus contenía un único gen que codificaba NRPS seguido de dos glicosiltransferasas, así este locus (12.721 pb) se analizó adicionalmente. Se identificaron seis ORF en este locus, que se predijo que jugaban un papel integral en la biosíntesis de glicolipopéptido (Tabla 3). Los seis genes, designados *rlpA* a *rlpF*, están orientados en la misma dirección y forman una región contigua en el genoma de *V. paradoxus* RKNM-096.

Tabla 3. Funciones deducidas de los Orf identificados en las agrupaciones génicas biosintéticas de glicolipopéptido (Seq. ID:1) y dTDP-L-ramnosa (Seq. ID:2) de *V. paradoxus* RKNM-096.

Seq. ID. (ADN)	Fuente	Nombre	Inicio	Parada	Seq. ID. (Proteína)	Tamaño (aa)	Función Propuesta		
3	Seq. ID:1	<i>rlpA</i>	121	1.035	4	304	Regulador	transcripcional	LysR
5	Seq. ID:1	<i>rlpB</i>	1.437	8.912	6	2.491	Péptido	sintetasa	no ribosomal
7	Seq. ID:1	<i>rlpC</i>	8.924	10.243	8	439	dTDP-ramnosil transferasa		
9	Seq. ID:1	<i>rlpD</i>	10.276	10.488	10	70	Proteína MbtH		
11	Seq. ID:1	<i>rlpE</i>	10.497	11.465	12	322	dTDP-ramnosil transferasa		
13	Seq. ID:1	<i>rlpF</i>	11.462	12.721	14	419	Transportador MFS		
15	Seq. ID:2	<i>rmlB</i>	299	1.378	16	359	dTDP-glucosa	4,6-	deshidratasa
17	Seq. ID:2	<i>rmlD</i>	1.375	2.265	18	296	dTDP-4-deshidroramnosa reductasa		

Seq. ID. (ADN)	Fuente	Nombre	Inicio	Parada	Seq. ID. (Proteína)	Tamaño (aa)	Función Propuesta
19	Seq. ID:2	<i>rmlA</i>	2.298	3.194	20	298	Glucosa-1-fosfato timidililtransferasa
21	Seq. ID:2	<i>rmlC</i>	3.191	3.736	22	181	dTDP-4-deshidroramnososa 3,5-epimerasa

Genes implicados en la regulación. El primer gen, *rlpA*, codifica una proteína que presenta similitud con reguladores transcripcionales que pertenecen a la familia LysR. El análisis de dominios conservados indicó que RlpA contenía un dominio de hélice-giro-hélice amino terminal y un dominio de unión a sustrato LysR carboxi terminal, lo que es consistente con la arquitectura de dominios de los reguladores transcripcionales LysR. Esta familia de reguladores puede funcionar como activadores o represores transcripcionales (Maddocks y Oyston [2009] *Microbiology* 154, 3609), por lo tanto, es probable que RlpA juegue un papel en la regulación de la biosíntesis de glicolipopéptidos.

Genes implicados en la biosíntesis de péptidos. Después de *rlpA* hay un gen grande, *rlpB*, (7.476 pb), que codifica una NRPS. El análisis de dominios (Bachmann y Ravel [2009] *Meth. Enzymol.* 458, 181) indicó que la NRPS consiste en dos módulos (M1 y M2) con la siguiente organización de dominios (C-A-PCP)_{M1}-(C-A-PCP-R)_{M2}. La estructura dimodular y la organización de dominios sugiere que RlpB genera un dipéptido, lo que es consistente con la estructura de los glicolipopéptidos de *V. paradoxus* RKNM-096. El primer dominio del primer módulo de RlpB es un dominio de condensación. La presencia de un dominio C al comienzo de un módulo de inicio de NRPS es característica de los péptidos acilados. Los dominios C amino terminales pueden catalizar la formación de enlaces amida entre el primer aminoácido de un péptido y un ácido graso. El ácido graso puede ser presentado al dominio C como un intermedio acil-ACP, como en el caso de la biosíntesis de CDA (Kopp *et al.* [2008] *J. Am. Chem. Soc.* 130, 2656), o un intermedio acil-CoA, como en el caso de la biosíntesis de surfactina (Krass *et al.* [2010] *Chem. Biol.* 17, 872). Se realizó un análisis filogenético del dominio C del módulo de inicio de RlpB (residuos 12-437) usando el programa NaPDoS (Ziemert *et al.* [2012] *PLoS One* 7, e34064). El dominio RlpB se agrupa cerca de los dominios C de los módulos de inicio que catalizan la condensación de un precursor de ácido graso con un aminoácido. El dominio C más relacionado en la base de datos de referencia NaPDoS fue el módulo de inicio de la NRPS de bacililactina (38% de identidad), que cataliza la condensación de 2,3-dihidroxibenzoil-ACP con glicina (May *et al.* [2001] *J. Biol. Chem.* 278, 7209). Esto sugiere que la biosíntesis de glicolipopéptidos empieza con la condensación de un ácido graso con el primer aminoácido del péptido (serina). Análisis similares del segundo dominio C indicaron que estaba relacionado muy de cerca con el segundo dominio C de la NRPS dimodular de bacililactina, DhbF (54% de identidad). El análisis filogenético reveló que el dominio C de M2 de RlpB se agrupaba con los dominios C que catalizan la condensación de dos L-aminoácidos (Ziemert *et al.* [2012] *PLoS One* 7, e34064), lo que es consistente con la estructura del glicolipopéptido.

Para predecir la especificidad de sustrato de los dominios A de RlpB, se extrajeron códigos de especificidad de sustrato de los sitios activos del dominio A (8 residuos entre los restos A3 y A6) y se compararon con códigos de especificidad de dominios A conocidos usando la herramienta NRPS Predictive Blast (Bachmann y Ravel [2009] *Meth. Enzymol.* 458, 181). El código de especificidad del dominio A de M1 fue el más similar a los dominios A de las NRPS de nostopeptólido, pioverdina, CDA y enterobactina que activan L-serina (75-87% de identidad, 87-100% de similitud, valor E 0,023-0,039), lo que sugiere que la L-serina se incorpora por M1. Esta observación es consistente con la estructura de los glicolipopéptidos. El código de especificidad del dominio A de M2 mostró una baja homología (50% de identidad, 100% de similitud, valor E 0,98) con un dominio A de NRPS de tirocidina (TycB), que activa L-fenilalanina o L-triptófano (Mootz y Marahiel [1997] *J. Bacteriol.* 179, 6843). Este bajo nivel de similitud impide la predicción de la especificidad de sustrato de este dominio A. Sobre la base de la estructura de los glicolipopéptidos de *V. paradoxus* RKNM-096, se esperaba que el segundo dominio A activara L-leucina. La comparación del código de especificidad del dominio A del módulo 2 de RlpB con los códigos de especificidad de leucina (Stachelhaus *et al.* [1999] *Chem. Biol.* 6, 493) también reveló una baja similitud, así el código de especificidad del dominio A de M2 de RlpB puede representar una nueva variante para leucina, aunque es necesaria una evidencia bioquímica para establecer la especificidad de sustrato de este dominio. También se analizaron los dominios PCP de RlpB y se encontró que ambos contenían el resto del dominio PCP central con una serina invariable que representa el sitio de unión de 4'-fosfopanteteína (Konz y Marahiel [1999] *Chem. Biol.* 6, R39).

El dominio final de RlpB es un dominio R. Los dominios R utilizan NAD(P)H como un cofactor para liberar reductivamente los productos finales unidos a PCP como un aldehído o alcohol (Du y Lou [2010] *Nat. Prod. Rep.* 27, 255). La presencia de un residuo de leucinol en el extremo carboxi del resto dipéptido del glicolipopéptido es consistente con la liberación de un intermedio dipéptido acilado por un dominio R. Colectivamente, la estructura y organización de los dominios de RlpB, así como la especificidad de sustrato predicha de los dominios individuales, son consistentes con la estructura de los glicolipopéptidos producidos por *V. paradoxus* RKNM-096.

Se encontró un gene pequeño (*rlpD*), en 3' de *rlpB*, que codifica una proteína de 70 aminoácidos que muestra similitud con proteínas semejantes a MbtH. Estas proteínas se encuentran frecuentemente en asociación con NRPS y se ha

demostrado que son esenciales para la producción de péptidos no ribosomal. (Baltz [2014] *J. Ind. Microbiol. Biotechnol.* 41, 357). Recientemente, se ha mostrado que estas proteínas facilitan las reacciones de adenilación a través de la interacción directa con los dominios A (Herbst *et al.* [2013] *J. Biol. Chem.* 288, 1991). Así, predecimos que RlpD interacciona con uno o ambos dominios A de RlpB para facilitar la formación de dipéptidos.

5 Genes implicados en la glicosilación. La glicosilación del dipéptido acilado generado por RlpB está catalizada probablemente por dos ORF (*rlpC* y *E*) en 3' *rlpB*. La secuencia de aminoácidos deducida de *rlpC* (439 aa) muestra similitud con la familia GT1 de glicosiltransferasas, que utilizan azúcares activados como sustratos para transferir restos de azúcar a un conjunto diverso de moléculas aceptoras (Breton *et al.* [2006] *Glycobiology* 16, 29R). La secuencia de aminoácidos deducida de *rlpE* (322 aa) muestra similitud con dTDP-ramnosiltransferasas. En la biosíntesis de ramnolípidos se utilizan dos glicosiltransferasas para transferir secuencialmente dos unidades de ramnosilo al componente lipídico del ramnolípido (Deziel *et al.* [2003] *Microbiology* 149, 2005). RhlB transfiere ramnosa de dTDP-L-ramnosa al grupo β -hidroxilo libre del ácido 3-(3'-hidroxidecanoiloxi)decanoico (HDD) para generar mono-RL, mientras di-RL se forma por la transferencia de una ramnosa adicional de dTDP-L-ramnosa a mono-RL por RhlC (Abdel-Mawgoud *et al.* [2011] en *Biosurfactants*, Springer- Verlag, Berlín Heidelberg). La relación entre RlpC y RlpE y los homólogos RhlB y RhlC de *P. aeruginosa* PAO1, *B. thialandensis* E264 y *B. pseudomallei* 1710B se investigó a través de la generación de un árbol filogenético (método de agrupamiento pareado no ponderado con método de la media aritmética). En este análisis, RlpC se agrupó con los ortólogos de RhlB, mientras RlpE se agrupó con los ortólogos de RhlC. Aunque RlpC se agrupó con los ortólogos de RhlB, no se agrupó fuertemente ya que mostró una identidad de secuencia limitada con estas enzimas (18,6-23,1%). Por el contrario, RlpE compartió entre el 39,6-40,7% de identidad con los ortólogos de RhlC. Estos datos sugieren que RlpC y RlpE realizan funciones similares a RhlB y RhlC, respectivamente. Establecimos la hipótesis de que RlpC cataliza la ramnosilación de un intermedio dipéptido acilado utilizando dTDP-L-ramnosa como el donante de carbohidrato. La limitada homología de secuencia entre RlpC y los ortólogos de RhlB puede reflejar la diferencia significativa en los sustratos de glicosilación utilizados por las enzimas. Se predice que RlpE cataliza la segunda reacción de glicosilación, transfiriendo ramnosa de dTDP-L-ramnosa al producto de la reacción de RlpC.

No se encontraron genes que codifican la biosíntesis de dTDP-L-ramnosa muy cerca de la agrupación génica del glicolipopéptido. El escaneo del genoma para homólogos de genes biosintéticos de ramnosa en *P. aeruginosa* PAO1 (*rmIBDAC*) identificó cuatro genes que presentaban una fuerte similitud de secuencia con los de *P. aeruginosa* (identidad/similitud: RmlB - 79%/89%, RmlD - 60%/71%, RmlA - 78%/89%, RmlC - 66%/80%). En el genoma de *V. paradoxus* RKNM-096 los cuatro genes están agrupados y se encuentran en el mismo orden que en *P. aeruginosa* (*rmIBDAC*) (Rahim *et al.* [2000] *Microbiology* 146, 2803). Este locus proporciona probablemente los sustratos dTDP-L-ramnosa utilizados por RlpC y RlpE. La modulación de la expresión de uno o más componentes de la ruta biosintética de la dTDP-L-ramnosa por un experto en la técnica puede ser una estrategia efectiva para incrementar los rendimientos de glicolipopéptido.

35 Genes implicados en el transporte. Directamente en 3' de *rlpE* hay un ORF (*rlpF*) que codifica una proteína, que es similar a los transportadores de la familia de facilitador mayor de una variedad de bacterias. RlpF presenta un 38% de identidad y 54% de similitud con PA1131 de *P. aeruginosa* PAO1, que está inmediatamente en 5' de *rhlC* (Dubeau [2009] *BMC Microbiol* 9, 263). RlpF está implicado probablemente en el eflujo de glicolipopéptido.

40 Genes implicados en la biosíntesis del resto lipídico. En la biosíntesis de ramnolípidos, el resto HDD se produce por RhlA, que condensa dos moléculas de β -hidroxidecanoil-ACP de la biosíntesis de ácidos grasos para rendir ácido 3-(3'-hidroxidecanoiloxi)decanoico. El escaneo del genoma de *V. paradoxus* RKNM-096 para homólogos de RhlA no identificó ninguna proteína con una similitud significativa con RhlA. Así, la generación del resto lipídico de los glicolipopéptidos de RKNM-096 está dirigida probablemente por un nuevo mecanismo, todavía no identificado.

45 Genes implicados en la acetilación de glicolipopéptido. Los análogos acetilados de NB-RLP1006 son abundantes en los caldos de fermentación de *V. paradoxus* RKNM-096. No se identificaron genes que codifican acetiltransferasas en la agrupación génica. Así, es probable que la acetilación esté catalizada por una enzima codificada en otro lugar del genoma de *V. paradoxus* RKNM-096.

Biosíntesis propuesta. La biosíntesis de glicolipopéptido empieza presumiblemente con la formación del resto 3-(3-(3'-hidroxidecanoiloxi)decanoiloxi)decanoilo a través de un mecanismo todavía no identificado. Después de la formación del resto lipídico, este se presenta probablemente al dominio C de M1 de RlpB que condensa el resto lipídico con L-serina. El M2 de RlpB incorpora entonces L-leucina para formar un intermedio dipéptido acilado unido a PCP que se libera de la enzima por el dominio R C-terminal de RlpB, dando lugar a la formación de un residuo de L-leucinol terminal. La dTDP-L-ramnosa, producida por el operón *rmIBDAC*, se utiliza entonces por las ramnosiltransferasas RlpC y RlpE para glicosilar secuencialmente la aglicona dando lugar a la producción del glicolipopéptido glicosilado final NB-RLP1006. NB-RLP1006 serviría como un sustrato para la acetilación para formar NB-RLP1048A y NB-RLP1048B.

60 Para demostrar la implicación de la agrupación génica *rlpA-rlpF* en la biosíntesis de glicolipopéptidos en *V. paradoxus* RKNM-096, se expresó *rlpE* en *E. coli* y la actividad de la enzima se demostró usando NB-RPL860 como un sustrato. El análisis por bioinformática indicó que RlpE cataliza la segunda ramnosilación en la biosíntesis de glicolipopéptidos, convirtiendo los glicolipopéptidos mono-ramnosilados (p. ej., NB-RLP832 y NB-RLP860) en glicolipopéptidos di-

ramnosilados (p. ej., NB-RLP978 y NB-RLP1006). El gen *rlpE* se clonó en pET28a (EMD Millipore, Darmstadt, AL) con una etiqueta de hexa-histidina amino terminal usando técnicas de clonación estándar y la clonación sin mutación se verificó por secuenciación. Debido al alto contenido en GC de *rlpE*, se eligió *E. coli* Rossetta DE3 *pLysS* (EMD Millipore) como el huésped de expresión ya que esta cepa expresa ARNt para codones raros ricos en GC (AGG, CCA, GGA). Se usó una única colonia para inocular 50 mL de LB Miller (EMD Millipore) suplementado con 50 µg/mL de kanamicina (Sigma- Aldrich) y 34 µg/mL de cloranfenicol (Sigma-Aldrich) y el matraz se incubó a 37 °C con agitación a 250 rpm toda la noche. Los cultivos de expresión (50 mL) se realizaron en LB Miller suplementado con kanamicina y cloranfenicol. Estos cultivos se inocularon con 0,5 mL del cultivo de toda la noche y se cultivaron a 37 °C y 250 rpm hasta que la densidad óptica (600 nm) alcanzó 0,5, después de lo cual se añadió IPTG a una concentración final de 1,0 mM para inducir la expresión de proteínas y los cultivos se incubaron a 15 °C durante 24 h. Las células se recogieron por centrifugación (6.000 x g durante 5 min) y se lavaron una vez con Tris-HCl 20 mM (pH 8,0). El sedimento celular se congeló a -80 °C hasta que pudo realizarse la purificación. Para purificar RlpE etiquetado con His, las células se descongelaron, se suspendieron en tampón de lisis (NaCl 500 mM, glicerol al 5%, Triton X-100 al 1%, Tris-HCl 25 mM, pH 8,0) y después se lisaron mediante sonicación. Los restos celulares y la proteína insoluble se eliminaron por centrifugación a 15.000 x g durante 30 min. El sobrenadante se mezcló con 0,5 mL de resina HisPur Ni-NTA (Thermo Fisher Scientific). La resina se lavó seis veces con 1,0 mL de imidazol 75 mM. El RlpE etiquetado con His se eluyó con 1,0 mL de imidazol 250 mM. Se realizaron y combinaron cuatro lotes de elución. El tampón de elución de imidazol se intercambió con tampón de enzima (Tris-HCl 25 mM, glicerol al 10%) y se concentró por filtración centrífuga usando un filtro rotativo Macrosep 3 kDa (Pall). Después de la concentración, la enzima se repartió en partes alícuotas y se almacenó a -80 °C. La pureza de la enzima se analizó por electroforesis en gel de poli(acrilamida) desnaturante (gel al 4-15% Mini-PROTEAN preparado comercialmente, 160 V, 30 min; Bio-Rad). El peso molecular calculado de RlpE etiquetado con His fue 38,2 kDa. El peso molecular aparente de la proteína purificada fue 33,05 kDa, lo que está muy de acuerdo con el peso molecular esperado (Fig. 2A).

La actividad de RlpE se estableció incubando la enzima (0,1 µM) en tampón de reacción (Tris-HCl 25 mM pH 8,0, MgCl₂ 2,5 mM) con 1 mM de TDP-L-ramnosa y NB-RLP860 0,5 mM. Las reacciones (200 µL) se incubaron a 30 °C durante

4 h. Una porción (25 µL) de la reacción se retiró a los 15 s, 1 min, 5 min, 20 min, 1 h y 4 h. La reacción se paró por la adición de dos volúmenes de metanol, seguido de congelación ultra rápida. Las reacciones paradas se separaron por UPLC (Accela™, Thermo Fisher Scientific Mississauga, ON, Canadá) y los eluatos se analizaron por HRESIMS (LTQ Orbitrap Velos; Thermo Fisher Scientific) (modo positivo, monitorización *m/z* 200-2.000). La separación cromatográfica se consiguió con una columna Hypersil Gold 1,9 µm C₁₈ 175 Å 50 x 2,1 mm (thermo Fisher Scientific) y un gradiente lineal de H₂O al 50%/FA al 0,1% (disolvente A) y acetonitrilo al 50% (CH₃CN)/FA al 0,1% (disolvente B) a disolvente B al 100% durante 5 min, seguido de mantener el disolvente B al 100% durante 3 min con una tasa de flujo de 300 µL/min. Las reacciones realizadas con enzima hervida no mostraron conversión de NB-RLP860 a NB-RLP1006. Por el contrario, las reacciones enzimáticas que contenían 6xHis-RlpE intacta dieron lugar a la conversión completa de NB-RLP860 a NB-RLP1006 después de 4 h (Fig. 2B). Estos datos indican que RlpE cataliza la segunda etapa de ramnosilación en la biosíntesis de glicolipopéptido en *V. paradoxus* RKN-096. Como los genes para la biosíntesis de productos naturales en bacterias están típicamente agrupados, este descubrimiento también proporciona una evidencia fuerte que confirma la agrupación génica propuesta como el locus responsable para la biosíntesis de glicolipopéptidos.

También exploramos la capacidad de 6His-RlpE purificado de añadir iterativamente unidades de ramnosa a NB-RLP860 mediante el escaneo de los datos de HRMS para masas consistentes con los tensioactivos glicolipopeptídicos que contienen tres residuos de ramnosa (calc'd [M+H]⁺ 1153,7204), cuatro residuos de ramnosa (calc'd [M+H]⁺ 1299,7783) y cinco residuos de ramnosa (calc'd [M+H]⁺ 1445,8367). Se obtuvieron masas consistentes con productos de reacción trirramnosilados y tetrarramnosilados y difirieron de los pesos moleculares esperados en <1,03 partes por millón (ppm) y < 0,6 ppm, respectivamente. De forma interesante, se observaron dos picos para cada masa, lo que sugiere que unidades de ramnosa adicionales están unidas en dos posiciones diferentes de la estructura de NB-RLP1006. Respecto a la producción de NB-RLP1006, los glicolipopéptidos tri-ramnosilados y tetra-ramnosilados constituyeron el 2,99% y 0,14% de los productos de la reacción. No se detectaron glicolipopéptidos penta-ramnosilados. Estos datos indican que puede usarse RlpE expresado recombinantemente para generar análogos de glicolipopéptidos con hasta cuatro residuos de ramnosa. Dichas modificaciones pueden alterar las propiedades funcionales del glicolipopéptido. Las propiedades pueden incluir, pero no se limitan a, humectación, espumación, capacidad tensioactiva y emulsificación.

La elucidación de la ruta biosintética para los biotensioactivos glicolipopeptídicos producidos por *V. paradoxus* RKNM-096 establece el estadio para la modificación racional de la ruta biosintética para generar nuevos análogos o para incrementar los rendimientos. Los análogos pueden generarse por los expertos en la técnica a través de la modificación de las enzimas responsables de la biosíntesis e incorporación de las porciones de lípido, péptido y carbohidrato de la molécula. Los rendimientos pueden incrementarse por los expertos en la técnica mediante la modificación de genes reguladores y/o promotores, mediante la sobreexpresión de enzimas que representan etapas limitantes en la ruta biosintética o mediante la inactivación de enzimas que llevan a cabo reacciones no deseables. El conocimiento de la ruta biosintética también permite la expresión en un huésped heterólogo, lo que puede permitir mejoras en el rendimiento o la generación de análogos de glicolipopéptido.

Ejemplo 8: identificación de biotensioactivos relacionados en otras bacterias.

Se realizó la secuenciación de las agrupaciones génicas biosintéticas de glicolipopéptido y ramnosa de *V. paradoxus* RKNM-096. Antes del descubrimiento de la serie de glicolipopéptidos de biotensioactivos y la agrupación génica biosintética asociada descrita en la presente memoria, no habría sido posible predecir de manera exacta la producción de biotensioactivos glicolipeptídicos relacionados sobre la base únicamente del análisis de la secuencia de ADN. La identificación de la agrupación génica biosintética de glicolipopéptidos permite ahora la interrogación dirigida de genomas microbianos para agrupaciones génicas relacionadas, lo que puede tener el potencial de producir nuevos biotensioactivos glicolipeptídicos. Como *rlpC* codifica una nueva ramnosiltransferasa, que glicosila un intermedio dipéptido acilado característico de la clase de glicolipopéptidos de biotensioactivos, usamos la secuencia de aminoácidos deducida de este gen para buscar genomas bacterianos disponibles para homólogos. Esta búsqueda identificó homólogos que presentan RlpC a partir de una amplia variedad de bacterias. Investigamos entonces regiones genómicas que flanquean los genes que codifican los homólogos de RlpC para la presencia de homólogos de los otros genes biosintéticos de glicolipopéptidos. Se presentarán dos ejemplos para demostrar la utilidad de usar secuencias de la agrupación génica del glicolipopéptido como sondas para descubrir productores de nuevos biotensioactivos potenciales.

Se identificó una agrupación génica homóloga en el genoma de *Janthinobacterium agaricidamnosum* DSM 9628 (no. de acceso GenBank NZJHG322949.1) (Fig. 3). *J. agaricidamnosum* es una beta-proteobacteria como *V. paradoxus*, pero pertenece a una familia diferente. El homólogo de RlpC en esta cepa (WP_038493268.1) presentó un 68% de identidad con RlpC. El escaneo del genoma alrededor del homólogo de RlpC identificó otros homólogos de genes presentes en la agrupación génica del glicolipopéptido. Directamente en 3' del homólogo de RlpC había una proteína semejante a MtbH (WP_038493269.1) que compartía el 69% de identidad con RlpD. En 5', se identificó una NRPS dimodular (WP038499875.1), que mostró un 68% de identidad con RlpB y contenía una organización de dominios idéntica ([C-A-PCP]_{M1}-[C-A-PCP-R]_{M2}). El análisis de los sitios activos (Bachmann y Ravel [2009] *Meth. Enzymol.* **458**, 181) indicó que la especificidad de sustrato predicha también concordaba con la de RlpB, con el código de especificidad del dominio A de M1 concordando con el de L-serina y el código de especificidad del dominio A de M2 concordando con el del dominio A de M2 de RlpB, indicando que la L-leucina se incorpora por M2 (Fig. 3). También se encontraron un dominio C y dominio R en el extremo amino y carboxi de la NRPS de *J. agaricidamnosum*, respectivamente. Esto sugiere que la biosíntesis se inicia por condensación de un intermedio acilo con serina, y se termina por la liberación reductora de un dipéptido acilado, similar a lo que se predice para la biosíntesis de glicolipopéptido en *V. paradoxus* RKNM-096. No se encontró homólogo de RlpE en la agrupación génica de *J. agaricidamnosum* DSM 9628, lo que indica que el producto de la agrupación contiene probablemente un único residuo de ramnosa. También se detectó una agrupación génica con una organización altamente similar a la de *J. agaricidamnosum* DSM 9628 en el genoma de *V. paradoxus* DSM 21786 (no. de acceso GenBank NC 022247.1). Colectivamente, estos datos sugieren que *J. agaricidamnosum* DSM9628 y *V. paradoxus* DSM 21786 poseen la capacidad de producir nuevos biotensioactivos con estructuras relacionadas a las producidas por *V. paradoxus* RKNM-096. Sobre la base del análisis de bioinformática presentado aquí, predecimos que el o los compuestos producidos por estas bacterias serían un dipéptido L-serinil-L-leucinol *N*-acilado que porta un único residuo de ramnosa.

El escaneo del genoma usando la secuencia de RlpC también identificó una agrupación génica posible de biotensioactivo en la alfa-proteobacteria relacionada de forma más distante *Inquilinus limosus* DSM 16000 (no. de acceso Genbank NZ_AUHM01000002.1) (Fig. 3). El homólogo de RlpC (WP_026869107.1) en *I. limosus* compartía un 61% de identidad con la proteína de *V. paradoxus* RKNM-096. Se identificaron genes que codifican una proteína semejante a MtbH (WP_026869104.1) y una NRPS (WP026869105.1) inmediatamente en 5' del homólogo de RlpC. La proteína semejante a MtbH compartía un 56% de identidad con RlpD. La NRPS era una enzima monomodular con la siguiente organización de dominios: C-A-T-R (Fig. 3). El análisis del sitio activo del dominio A (Bachmann y Ravel [2009] *Meth. Enzymol.* **458**, 181) indicó que la L-serina es el sustrato probable de esta enzima. La presencia de un dominio C en el extremo N y un dominio R en el extremo C sugiere que el producto de la NRPS es un serinol acilado. También se detectó un homólogo de RlpE (WP_034850803.1) en la agrupación génica de *I. limosus* (41% de identidad), lo que sugiere que el intermedio serinol acilado puede glicosilarse secuencialmente para rendir un producto que porta un resto diramnosilo similar a NB-RLP1006. El producto final puede exportarse fuera de la célula a través de la acción de un exportador MFS (WP_034850806.1) que comparte un 52% de identidad con RlpF.

Para validar nuestra estrategia *in silico* para identificar productores de biotensioactivos glicolipeptídicos, obtuvimos *J. agaricidamnosum* DSM 9628 y *V. paradoxus* DSM 21786 de la colección de cultivos Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ). Cada cepa se fermentó en una variedad de medios de cultivo para promover la producción de biotensioactivos predichos. Las fermentaciones se extrajeron dos veces con un volumen igual de EtOAc. La capa orgánica se evaporó y los extractos concentrados resultantes se analizaron por UPLC-PDA-ELSD-HRESIMS como se ha descrito anteriormente para NB-RLP1006 (Ejemplo 3). Se observaron tres picos prominentes eluyendo a los 3,07, 5,05 y 5,51 min en los cromatogramas ELSD y HRESIMS de *J. agaricidamnosum* DSM 9628. El pico a los 3,07 min (HRESIMS *m/z* 1182,6217 [M + H]⁺, calcd para C₅₆H₈₅N₁₂O₁₆, 1181,6201) podría atribuirse al compuesto conocido jagaracina reportado previamente de esta cepa (Graupner et al. [2012] *Angew Chem. Int. Ed. Engl.* **51**: 13173). La extracción de los espectros de masa para los picos que eluyen a los 5,05 y 5,51 min reveló iones [M + H]⁺ de *m/z* 833,5741 y *m/z* 861,6033, respectivamente. Los iones [M + H]⁺ observados mostraron una diferencia de masa de -1,5 y 1,0 ppm de los iones *m/z* [M + H]⁺ predichos para los glicolipopéptidos monoramnosilo NB-RLP832 (*m/z* 833,5741 [M + H]⁺) y NB-RLP860 (*m/z* 861,6033 [M + H]⁺), respectivamente, lo que indica que los

compuestos esperados habían sido producidos por *J. agaricidamnorum* DSM 9628. De manera similar a NB-RLP978A-C y NB-RLP1020A-C, la masa de NB-RLP832 concordó de cerca con la predicha para un análogo de NB-RLP860 que carece de dos grupos metileno. Estos compuestos se purificaron y sus estructuras se elucidaron usando una combinación de experimentos de RMN 1D y 2D. Este análisis confirmó inequívocamente que el biotensioactivo monoramnosilado esperado había sido producido por *J. agaricidamnorum* DSM 9628 (véase el Ejemplo 3).

Un análisis idéntico de los extractos de la fermentación de *V. paradoxus* DSM 21786 también reveló la presencia de un pico que eluía a los 5,51 min en el cromatograma HRESIMS. La inspección del espectro de masa asociado con este pico reveló la presencia de un ion $[M + H]^+$ con una m/z de 861,6104, que se diferenciaba de la masa esperada ($[M + H]^+$ m/z 861,6046) por 5,8 ppm. El tiempo de retención idéntico y la masa monoisotópica indicaron que tanto *J. agaricidamnorum* DSM 9628 como *V. paradoxus* DSM 21786 producen NB-RLP860.

Listado de secuencias

<110> Croda International Plc

<120> Biotensioactivos glicopeptídicos

<130> RIC/MP7270267

<160> 24

<170> PatentIn version 3.3

<210> 1

<211> 12721

<212> ADN

<213> *Variovorax paradoxus*

<400> 1

ES 2 809 727 T3

gtcgtgtctc cttcttttcg tggggtgttc caacggggccg actgggaggt cggctgaaaa	60
ccgctcgcca gtgtgcgtgc cgcaagggtt gccttcaata aaataatcaa gctaagtaat	120
atgaatggca tgcataatga ctcggtcgac ctcaatctgc tgcgcctgtt cgatgcggtc	180
taccgcgagc gcagcgtgag ccgcgccgcg gaggcgtggt gcctcacgca gcctgcggca	240
agccatgggc tgggacggct ggggctgctt ttgaaagacg cgctcttcac gcgtgcccc	300
ggcggcgtgg cggccacgcc ggcgccgac cggctcgcg tggcgggtgca ggcggcgtc	360
ggcacgatcg aagcggcgct gcacgagccc gatcgcttcg agcccaggt gtcgcgcaag	420
agctttcgta ttcacatgag cgacatcggc gagggcgct tcctgccgc gctgatggcg	480
cggctcgcg agctggcgcc cggcgtgcgg ctggagacct tgccgctctt gcctgcggag	540
gttgcgccc cactcgacag cggccgcatc gatttcgct tcggctttct ctcgaccgtg	600
cgcgacacgc agcgcacgca tcttctgaaa gaccgctaca tcgtgctgct gcgcaagggc	660
catcccttg tgaagcgccg gcgcaagggg caggcgctgc tcgaggcgct gcaggagctc	720
gactacgtgg cggcgcgcac gcacgccgac acgctgcgca tcttgagtt gctcaacctc	780
gaagaccgcc tgcgcctcac gaccgagcac ttcattggtgc taccggccat cgtgcgcgcc	840
accgatctcg cgttggtgat gccgcgcaac atcgcgcgag ggtttgcgga ggaggcggc	900
tacgcgatcg tcgagccgcc gtttccgctg cgcgatttca gcgtgtcgct gactggagc	960
aagcgcttcg agggcgacct ggccaaccgt tgggtgcggc aggtgatcac ggcgctgttc	1020
tccgagcgcg gctgaagttc gaccaccaa gtacgcgccg cgcgggtgaa gcgcgcgca	1080
ctgcgcgagt aacacgccga gagattcccc tacagctttc tcgccagtt gctgcatcg	1140
aacattcttt tggggtgcat gacgcgcgaa atacgatgaa agccttcgat tccgaaagcc	1200
gcgattcagg tcgcaacttc gggatgaaat ctttcgcgct caaagacgtt cgtgaaatgt	1260
tttcttccct aaaaccgtca ctgaaagtgt tgaaccact tgtacagtgg actggcaatg	1320
tgaacggatt gttaccgcgg agcaccggca tttctccttg agcggccgat gcacgacgcg	1380

ES 2 809 727 T3

tccattttcac	gcgcacatgc	atcgttgcc	atttactca	agacctggag	aagtgcata	1440
gtaccgtcga	tcagctgggc	cgcaccgccc	cccttacctc	ggggcagatg	gcgatgtggc	1500
tcggcgcaaa	gttcgcgtcg	cccgcaccca	atttcaatct	cgccgaagcc	atcgacatcg	1560
caggcgagat	cgacccccgc	atcttcctgg	cggccatgcg	acaggtggcc	gatgaagtgc	1620
aggccacgcg	cctgagcttc	atcgataccc	cgcaagggcc	acgacaggtc	gtcgcgcccc	1680
ttttcaccgg	cgagatcccc	tacctcgacc	tcagcggcga	gagcgatccg	caggccgagg	1740
ccgagcgctg	gatgcatacg	gactacaccc	gcagcatcga	cctcgcgcac	gggcagctgt	1800
ggctgtccgc	gctgatccgc	ctcgcgcccc	atcgccacat	ctgggtaccac	cgcagccatc	1860
acatcgcgct	cgacggcttc	agcggcgggc	tcatacgcacg	ccgcttcgcc	gacatctaca	1920
ccgcgatggg	cgacaacaac	gcagcggtgc	ccgaagactc	gcgccttgca	ccgatctcgc	1980
agctggccga	cgaagaacat	gcctatcgcg	agtcggggccg	cttcccgcgc	gaccgccagt	2040
actggaccga	gcgcttcgcc	gatgcacccg	atccgttgag	cctcgcctcg	caccgctcgg	2100
tcaacgtcgg	tggcctcttg	cgccagacgg	tgcacctgcc	ggcggccagc	gtgcaagccc	2160
tgcagaccat	cgcgcaagag	ctcggcacca	cgctgccgca	aatcctcatc	gccaccaccg	2220
cggcctacct	gtaccgcgca	acgggcatcg	aggacatggc	aatcggcatc	cccgtcaccg	2280
cgcgccacaa	cgaccgcatg	cgcgcggtgc	ccgcgatggg	ggccaacgcg	ctgccgctgc	2340
gcctggcgat	gcgcgcggac	ctgccgattc	cggaactgat	ccgcgaagtc	ggccggcaga	2400
tgcggcagat	cctcgcgcac	cagtcgtatc	gctacgagca	tttgcgcagc	gacctcaaca	2460
tgctgggtgaa	caaccggcag	ctcttcacca	ccgtgggtcaa	cgtcgagccc	ttcgactacg	2520
acttccgctt	tgcggggccat	gccgcgaagc	cgcgcaacct	ctcgaacggc	acggccgagg	2580
acctcggcgt	cttctgttac	gagcgcggca	acgggcagga	cctgcagatc	gacttcgacg	2640
ccaaccccg	ggtgcacacc	gcagaggaac	tggccgatca	ccagcgccgg	ctgcttgctt	2700
tcatacgacg	cgtgatccgc	ctgccgttgc	aggccgtcgg	ccagatcgac	ctgctcggtg	2760
ccgaagagcg	gcagcaattg	ctggtcgagt	ggaacgacac	ggcccacgcc	gtgcccgaca	2820
cccatctcac	cgcgttgatc	gaagcgcagc	tcgcagccga	tccgcaagcc	atcgcatctg	2880
gcttcgacgg	cgaggcgatg	aacaacgaag	aactgaaccg	ccgcgcaca	cgtctcggcc	2940
acctgctgcg	cgcacgcggc	gctggcccgg	agcgcaccgt	ggcgtcgcg	atcccgcgtt	3000
cgatggacct	gatgattgcc	ttgctcgcca	cgttgaagac	cggcgcggcc	tacctgccgg	3060
tcgatccgga	tttcccggcg	gaccgcatcg	ccttcattgt	cggcgatgcg	cagcccgtgt	3120
gcctcgtcac	gaccgaagcc	ctcgcggagt	cgtgcgggc	agccgcccc	acattgctgc	3180
tcgatgtagc	gcaaacgatt	gcggatctgg	agagttgcaa	cgacaccaac	ccgggcatcg	3240

ES 2 809 727 T3

cgatcgaccc	ttcgcatccg	gcctatgtga	tctacacctc	gggctcgacc	ggcatgcccc	3300
aggggtgcgg	cgtgtcgcac	cgcgccatcg	tcaaccgcct	gcgctggatg	caggaccgct	3360
acggccttca	ggccgacgac	cgcgtgctgc	agaagacgcc	ttccagcttc	gacgtgtcgg	3420
tgtgggagtt	cttctggccg	ctgatcgacg	gtgccacgct	ggtgcttgcg	aaaccgggcg	3480
gccacaagga	tgcggcctac	ctcgcggggc	tgatcgcgga	ggagggcatc	accacgatcc	3540
acttcgtgcc	gtcgtatgctc	gaggtcttcc	tgctcgagcc	cacggcgggc	gcatgcacca	3600
cgtcgcgcgc	cgtgatctgc	agcggcgaag	ccttgctgcc	cgcgctgcaa	tgcgagttcc	3660
agcagcacct	ctcgtgcgag	ctgcacaacc	tctacggtcc	gaccgaggcc	gcggtcgacg	3720
tcacctcgtg	ggagtgcgaa	cgcacggacg	acgcagaagc	ctcgagcggt	cccacgcgcc	3780
gcccgatctg	gaacaccag	atgcacgtgc	tcgacagcgg	cctgcagccc	gtgccggccg	3840
gcgtgactgg	cgagctgtac	atcgcggggc	tcggcctcgc	acgcggctac	ctcaagcgcc	3900
cgttgctgag	cgcgcgagct	ttcatcgcca	acccctacgg	cacaccgggc	agccgcgtgt	3960
accgcaccgg	cgacctcgcg	cgtgtgcgca	aggacggcag	ccttgacttc	ctcggccgcg	4020
ccgaccagca	ggtgaagatc	cggggcctgc	gcatcgagcc	gggagagatc	gaatccgtgc	4080
tgctgcagca	tcgcgaagtc	gcgcaggccg	ccgtggtggc	gcgcgaagac	gtaccgggcg	4140
aaaagcgtct	cgtggcctac	gtcgttgoga	cggacgctgc	cgatccgcaa	gcggccgaac	4200
tgcgcacgcg	cctcgcgcaa	tcgctgcccg	agtacatggt	gccttcggcc	ttcgtcagcc	4260
tcccgtcgct	gccgctcgga	cccagcggca	agctcgaccg	caaggcgctg	ccgccccccg	4320
aagtgcaggc	cgccacgcgc	tacgcgcgcg	cgcgcacgcc	gaccgaaaag	atcctggccg	4380
gcctctgggc	cgagacgctg	catttgccgc	gcgtcgggtg	caacgacaac	ttcttcgaac	4440
tcggcggcca	ctcgtgatg	atcgtgcagc	tcatgtcgat	gatccggcag	caattcatga	4500
tcgacctgcc	ggtcgacacg	ctgttccagg	tctccaccat	cgcgggcctt	gccgagctgc	4560
tcgaccagga	atcggtcgcc	cgtccgagcc	tgactccgat	gccgcgcccc	gcgcgcattc	4620
cgtgtcctt	cgcgcagcgc	cgcctgtggc	tgatgaacca	gctcgaaggc	gcgaaccggg	4680
cctacaacat	gccgctcgcg	ctgcgcctgt	cgggtgtgct	cgatcgcaac	gcattgcatg	4740
cggcgctcgg	cgacctggtg	cagcggccag	agagcctgcg	cacgggtctac	ccgaacgaag	4800
acgggctgcc	gtaccagcac	atcctcgacg	gcgcggatgc	gcgtccggcg	gtgatcgagg	4860
ccgacagcag	cgaagaagaa	atcgcggcgc	agcttcacgc	cgtcgcgggc	catgccttcg	4920
atctcggcag	cgcggcgccc	ttgcgcgtct	acctgttcaa	gctcgccggc	gacgaacacg	4980
tgctgctgct	gctcacgcac	cacattgccg	gcgatggcgc	ctcgtcgtcg	ccgctagcgc	5040
gcgacatcag	cgtggcctat	gccgcgcgct	gcgaaggcaa	ggcgcggggc	tgggagccgc	5100
tgccgctgca	atacgccgac	tacgcgctgt	ggcagcagga	gctgctcggc	agcgaagacg	5160

ES 2 809 727 T3

atgccgagag catggccggc cggcagcgtg agttctggcg ttccctcgctg agcgacctgc 5220
 ccgagcaact ggcgctgccc gtcgaccacg cacggccgct cgtgccgacc taccgcggcg 5280
 atgtggtccc gctgcagatt ccgtcgcatg tgcataaacg catcctgcaa ctggcgcgcg 5340
 acgggcaggg cagcgtcttc atggtgctgc aggcgcgact cgcgggcctc ctgagccgcc 5400
 tcggcgcggg cgacgacatc gtcacggca gcccggtcgc ggggcgcagc gaccatgcgc 5460
 tggacgaact catcggtgc ttcgtcaaca cgctggtgct gcgcaactgac acctcgggcc 5520
 agccgagcct gcgcgagctg gtctcgcgcg tgcgcgccac caacctcgcg gcctatgcga 5580
 accaggagtt tccgtacgac cgcctcgtgg agctgctgcg tccgggcgcg tcgcgcgcca 5640
 acctgccgct gttccaggtc atgctgggct tccagggcac gagccgcctg tcgttcagcc 5700
 tgccgggcct gtcgatcgcg ccgcagccgg tggccatcga caccgcgaag ttcgacctgt 5760
 cgttcacatc cggcgagcaa cgcggtgccg atggcctgcc gggcggcac cccggcgcca 5820
 tccagtacag caccgacctg ttcgagcgca gcacggtcga ggccatgggc gcgcggtgg 5880
 tgcgtttgct ggaagaggcc tgcgaggcgc ccgacgatgc ggtgagtggc ctcgccatcc 5940
 tgagcgcgga agaaaccgac cgcctgctgt ccgactggag cggccgcacg cgcgaccttg 6000
 cgccgctctc gttcgccgac atggtggcct cgcacgcccgc ggagcgcccg cttgcagatg 6060
 cagtgggtgct cgacgacgcg accgtcagct acgccgaact cgatgcacgc gccaacgggc 6120
 tctcgacact gctgcgtgcg caaggcatcg ggggtggcgc catcgctcgcg acagtgtgc 6180
 cgcgttcgct cgacctcatc gtggcgcaact tggccatcgt gaaggccggc ggggcctacc 6240
 tgcccatcga ccccaaccac atggccgcgc gcagcgccct cgtgttcgag gaggcgcgc 6300
 ccgccgcggt gctgacgcac gatgcgctgt tgcccgagct ggtcggcggt ccccgctgca 6360
 tcgcgctcga cagcgacagc atggttgccg cgctggccat ccagtgggat acgccgctgg 6420
 tgcacgccc caatccacag gatgccgcct acctcatcta cacctccggc tccaccggca 6480
 tgccaaggcg cgtggtggtg ccgcatgcgg gcctgggcag cctcggcacc gcgatggcgg 6540
 agcggctcgt catcgccac ggctcgcgcg tgcgtcagtt ctccctccagc ggcttcgacg 6600
 cgtcggtgat ggaccagctg atggcctttg gcgcgggtgc cgcgctggtg gtgccggggc 6660
 cggagcaact gtcggcacg gagctggccg atctgctcga gaagcaggcc gtgagccacg 6720
 cgctgattcc gcccgccgcg ctcgcgacct tgccgcacgg cgagttccc cacctgcaga 6780
 cgctggtggt gggcgcgcat gcctgcaccg ccgcgctggc ggcgaagtgg tcgcaaggcc 6840
 gccgcatgat caacgcctac gggccgaccg agatcacat ctgcgcgagc atgagcgcg 6900
 cgatgacggc cgaggagttg ccctccatcg gccagccgat ctggaacacg cggatgtatg 6960
 tgctcgacag cgccctgcaa ccggtgccgc cgggtgtcgc gggcgagctc tacatcgccg 7020

ES 2 809 727 T3

gcagcggcgt	ggcgcgcggc	tatctcaacc	ggccggcatt	gagtgcgga	cgcttcacgc	7080
ccgacccgca	tggcgcgccc	ggcagccgca	tgtaccgcag	cggcgacctc	gcacgctggc	7140
gcgcgcgacg	cacgctcgac	ttcctcggcc	gcgcgcgacca	gcaggtgaag	atccggggct	7200
tccgcatcga	gccgggcgag	atcgaatccg	tgtgtctcaa	gcacccgttg	atcacgcagg	7260
ccgcctgat	cgcccgcgag	gacgtgccc	gcgagaagcg	cctggtcgcc	tacttcgtcg	7320
ccggttccga	gccgcagccc	accgagctgc	gcgcccacat	ggcgcaggcc	ttgcccgaact	7380
acatggtgcc	ttcgcccttc	gtgcgcctgc	cgctcgtgcc	gctcacgcaa	agcggcaagc	7440
tgcacaagaa	ggcgtgcgcg	gtgcccgaac	agcagccgcg	cgcgctgtac	gtggagcccc	7500
gcacgcccgc	cgagaaactg	ctgcgcggcc	tctggtccga	gacgctgcac	ctggagcgtg	7560
tcgcatcca	cgacaacttc	ttcgagatcg	gcgggcattc	gctcatggcg	atccagctgg	7620
gcctgcgcgt	ccgccagcag	gtgcgcgcgg	acttcccgcg	cgccgaggtc	tacaaccgcc	7680
cgacgattgc	cgacctggcc	gcctggtctg	acaacgaagg	cggcacggtc	gaggcgctgg	7740
acctgtcgcg	cgagctcgac	ctgcccgcg	acatccgccc	gcaggccact	gcaccgaagc	7800
tgcacccgcg	ccgcgtgttc	ctcaccggcg	cgagcggtt	cgtcggcagt	cacctgctgg	7860
ccgcgtgtt	gcgcgacacc	gcggcctgcg	tggctctcca	cgtcgcgcgcg	cccgaacgagc	7920
agggcggcga	gcagcgccctc	aagcgcacgc	tggcccagcg	ccagctcggg	gcgatctggg	7980
acaacgcgcg	catcaaggtc	gtgaccggcg	acctcggcaa	gccgcgcctg	ggcctcgatg	8040
acgtgcccgt	gcaactggtg	cgcgacggct	gcgacgccat	ctaccactgc	gccgcgcagg	8100
tgcacttcct	gcatccctac	gcgagcctca	agcccgcgaa	cgtcgacagc	gtggtcacgc	8160
tgtctgaatg	gacggcgcag	gggcgcgcga	agagcatgca	ctacgtctcc	acgtggtctg	8220
tgatcgacca	gaacaacaag	gaagacacca	tcaccgagca	atcggcgctg	gcctcatgga	8280
gcgggctggt	cgacggctac	agccagagca	agtgggtcgg	cgatgcgctg	gccgcgcagg	8340
cgcaggcgcg	gggcatgccg	gtggcgatct	accggctggg	ggcagtcacc	ggcgaccaca	8400
cgcacgcgat	ctgcaatgcc	gacgacctga	tctggcgcg	ggcgcatctc	tatgccgacc	8460
tggaaagcgt	ttccgatatg	gacctgccgc	tcaacctcac	accggtggac	gacgtggcgc	8520
gcgccatcct	cggccttgcg	gcgcagagg	cctcgtgggg	ccaggtgttc	cacctgatga	8580
gccaggcggc	gctgcgggtg	cgcgacattc	cgacgtctt	cgagcgcatg	ggcatgcggc	8640
tggagccggt	cggtctggag	ccctggtctg	agcgcgcgca	tgcacggctg	gccgtcgcgc	8700
atgacgcgga	cctggccgcg	gtgctcgcca	tcctcgaccg	ctacgacacc	acggccacgc	8760
cgccgcagg	gagcggcgcg	gccacgcgat	cgagctcga	ggccatcggc	gcgccgatcc	8820
gcccgggtgga	ccgcgacctg	ctgcagcgct	acttcgtcga	cctgggcata	gacaccaagg	8880
cgcgccgcgc	cctggaaacc	accacttcat	aggagcacac	ggaatggcac	gctatctcat	8940

ES 2 809 727 T3

cgcagcaacc gccttgccgg gacacgtcct gccgatgctg gccatcgcgc agcatctggt 9000
 gaaccagggg cacgaggtgc gggcgcacac cgcgagccag ttcagggcgc aggccgaggc 9060
 gaccggtgcg ggcttcacgc ccttcgagcg cacgatcgac ttcgactacc gcgacctgga 9120
 caagcgcttt cccgagcgcc agcgcatcgc ctcggcgat gcgcagctgt gcttcggcct 9180
 gaagcacttc tttgccgatg cgatggccgc gcagcatgcg ggcttgcaat cgatcctcga 9240
 agacttcgag gccgatgcca tcgtggtcga cacgatgttc tgcggcactt tcccgtgct 9300
 gctaggcaag gagcggaag accgcccggc catcgtcggc atcggcatct cggcgctgcc 9360
 gctctcgagc tgcgacaccg ccttcttcgg caacgcgctg ccgcccgtct ccacgccgga 9420
 agggcggggtg cgcaacaagg cgatgaacgc caacctcaa caggcgatgt tcggcgaggt 9480
 gcaacgctac ttcgacacgc tgctcgcgcg ttccgggcctg gccgcgctgc ccgatttctt 9540
 cgtcgatgcg atggtgaagc tgcccgatct ttacctgcag ctaccgcgc cttcgttcga 9600
 ataccgcgc agcgacctgc ccgctcggt gcatttcgtc ggcccgtgc tctcgccgc 9660
 gagccgcgac ttcacgccgc ccgagtgggt gcacgagctg gacgacggcc gctcggtcgt 9720
 gctggtcacg cagggcacgc tggccaacca gaatccgtcg cagctgatcg gcccgacgct 9780
 gcaggcgctg gccggcgaca agaacatcct cgtcatcgcc accaccggcg gcccggtgcc 9840
 gcccgccctg acggtgaacc tgcccgcga cgcgcgctg gtgcccgttc tgccctacga 9900
 ccggctgctg cccaagctgc acgcgatggt caccaacggc ggctacggct cggcgaacca 9960
 tgcattgagc ctcggtgtgc cgctggtggt ggccggcacc tccgaagaga agcccagat 10020
 cgcgcgcgc gtggcctggt cgggcgcggg catcaacctc gccaccggcc agccgaccgc 10080
 gcgccaggtc ggcgacgcgg tgcgcaagg actgggcaac tcgacctatc gccagcgtgc 10140
 ggcggtgctg cgtgaggact tcgcttgcca tcgcgcgctg accggcatcg ccggcgccct 10200
 cgaggcactt ctgcaaacct tcgcatccgc ggaaatggct tgaacctgaa cccatacga 10260
 caaaggaaat cccagatgag caaccgcttc gacgacaaga acgccagctt ccagggtgctg 10320
 gtgaacgacg agggccagca ctcgctgtgg cccgccttca tcgcccgtgcc cgcgggtgg 10380
 cagggtggcg tggcgccgac cgaccgcgac gcctgcagcg cctacatcgc ggcgaaactgg 10440
 caggacatgc gcccgcttc gctggtggtg gccacggcgg ccggtgacg ccgaggatgt 10500
 ccttcccgtt cggtgccgtc gtcgtcacct atttcccgac cggcgagcaa gtggcgaacc 10560
 tccattcgct ggccgctcg tgtccgcacc tctgctggt cgacaacacg ccgaggtgg 10620
 gcgattggca tgcggcgctc gtcgatgcgg gcgtttcggg gctgcacaac ggcaaccgcg 10680
 gcggcatcgc ggggccttc aaccgcggca tcatcgacct cgaagcgcg ggcccggaac 10740
 tcttcttct gctcgaccag gattcgaagc tgccaccgg ctacttcgat gccatgtcg 10800

ES 2 809 727 T3

aggctgcat ggtggcccg gagcggaagg gcgagggcaa tggtagaggaa gacgcggcct 10860
 tcctgatcgg cccgctcgtc cacgacacga acctggacgc gctgatcccg caattcggcc 10920
 tccagggcaa acgcgtctac cagttcgacc tggggcagcc cttcaccgag ccgctgatgc 10980
 gctgcgcctt catgatttcc tcgggctccc tgatttcgcg cggcgcttg gcccgatcg 11040
 gccggttcga cgagcgctat gtgatcgacc acgtggacac cgactactgc atgcgtgccc 11100
 tgggtcgcgg cgtgccgctc tacctgaatc cgcacgtcgt gctgcggcac cagattggcg 11160
 acatccgtgc ccggtcgtg ttcggctgga agatccaact catcaactac ccggccgcgc 11220
 ggcgtacta catcgcgcg aatgccatcg atctctcgcg ggcgcatgtg cgcgccttcc 11280
 ccgcatcct gttcatcaac gtttacacgc tcaagcagat cctgccgatg ctgatgttcg 11340
 agcgcgaccg cttcaagaag accatcgcg tgatgctcgg ctgcttcgat ggcctgttcg 11400
 ggcggtcgg gggcctcgg gaggtgcac cgcgggatgg caaatacctg gcccgcgcg 11460
 attgaccgcc acccttccag cgcgcgcgt acgcgcgcgc gcgctcgct tcatctcgt 11520
 cacggtgctg atcgacttca tggcggttcg cctgatcctg ccgggcctgc cgcacctggt 11580
 ggagcggctg gccggcgga gcacggtaac ggcgcggtac tggatcgctg tgttcggcac 11640
 cgcgttcgcg gcgatccagt tcgtgagctc gccgatccag ggcgcgctgt ccgaccgctt 11700
 cggcgggcg ccggtgatcc tgctgtcgtg cttcggcctc ggcgtggatt tcgtgttcat 11760
 ggcctcggcc gacagcctgc cgtggctggt cgtcggcgg gtggtctccg gcgtgttctc 11820
 ggccagcttc accatcgcca atgcctacat cgccgatgtg acgctgccgg aggagcgcg 11880
 ccgcagctac ggcacgtg gggcccgctt cggcatgggc ctggtgttcg gcccggtgct 11940
 cggcgggcaa ctgagccaca tcgatccgcg cctgccgttc tggttcgcg ccggcttgac 12000
 gctgctcagc ttctgctacg gatggttcgt gttgccgaa tcgctgccgc ccgagcgcg 12060
 tgcccgcaag ttcgactggt cgcgatccaa tcgggttggg acgctggtgc tgcacaagcg 12120
 ctatccgag gtgttcggac tggcgcggt gatcttctc gtgaacctg ctcagtacgt 12180
 ctatccagc gtgttcgtgc tgttcgcga ctaccgtat cactggaagg aagacgcgt 12240
 gggctgggtg ctcggcgcg tggcggtgct cagcgtgctg gtcaatgcgc tgttgatcg 12300
 gccggcggtg aagcgttcg gcgagcgcg cgcctgttg ctcggcatgg gcttcggcg 12360
 gctcggttc gtcatcatcg ggtttgccga cgtgggatg atctctctg ccggggtgcc 12420
 gttcggcatt ctgctggcgt tcgccggacc ggcgcgcgag gcgctggtca cgtgcaggt 12480
 cggcaccgcc gagcagggcc gcatccagg ggcgctcacc agcctggtgt cgtggcgcg 12540
 catcgtcgg ccggcgatgt tcgccggcag cttcggttac ttcacggcg cggacgcgc 12600
 ggtgcaactt ccggcgcgcg cgttttctc cgtgcggcg ttcctctgca tcggcacgt 12660
 gatcgcggtg cgctacgcac agccgaagcc cgcgacggca gcggtgcccg agccgacctg 12720

a 12721

- 5 <210> 2
- <211> 3959
- <212> ADN

<213> Variovorax paradoxus

<400> 2

ccgctgcgcc	tcgcaacggg	tttgctcctt	cggtgcatcg	cgatccctgc	gggtgcgatg	60
gctctccaga	cggcgtttga	tgtgatgcag	tactgacccc	ctgttcgggc	cgacctgagc	120
gtttatggga	gtttgcgcct	tcggtagggc	caccgggggtg	gcccgccttc	ctgcagtggg	180
gcgattgtag	gtgggcactg	ccaatgcgcc	aaccccgga	gtttcggccc	ttgggcccgat	240
gggataatca	tccgttcatt	cgccggaggg	cgatcgttcg	acaacaacag	gggaccccat	300
gatcctggta	accggcggcg	caggcttcat	tggcgccaat	ttcgtactcg	actggctcgc	360
acagagcgat	gaaccggtcg	tgaacctaga	caagctgacc	tacgcgggca	acctcgagac	420
gctcgcacg	ctcaaggaca	accgaagca	catcttcgtg	cagggcgaca	tcggcgacag	480
cgcgctgctc	gaccgcctgc	tggccgagca	caagccgcgt	gccgtggtca	acttcgcggc	540
cgaatcgac	gtcgaccgct	cgatccacgg	ccccgaagac	ttcgtgcaga	ccaacgtgct	600
gggcaccttc	cgctgctcg	aatccgtcgc	cggtttctgg	aatgccctgc	cgcccgacca	660
gaagcccgcc	ttccgcttcc	tgcattgtgc	gaccgacgag	gtctacggct	cgctctccaa	720
gaccgacccg	gccttcaccg	aagagaacaa	gtacgagccc	aacagcccgt	actcggccag	780
caagcccgcc	agcgaccacc	tcgtgcgcgc	ctggcaccac	acctacggcc	tgccggtggg	840
caccaccaac	tgctogaaca	actacgggcc	gttccacttc	cccgagaagc	tcattcccct	900
gatgatcgct	aacgcgctgg	ggggcaagcc	gctgcccggt	tacggcgacg	gcatgcaggt	960
gcgcgactgg	ctctacgtga	aggaccactg	cagcgccatc	cgccgcgtgc	tcgaagccgg	1020
caagctcggc	gagacctaca	acgtgggcgg	ctggaacgag	aagcccaaca	tcgagatcgt	1080
caacaccgct	tgccgcgtgc	tcgacgagct	gagccccaag	gccggcggca	agccgtacaa	1140
ggaacagatc	acctatgtga	ccgaccgccc	cgccacgac	cgccgctacg	cgatcgacgc	1200
acgcaagctc	gagcgcgaa	tcggctggaa	acctgccgag	accttcgaca	gcggcatccg	1260
caagacggtc	gagtgggtacc	tcgcgaacgg	cgagtgggtg	cgcaacgtgc	aaagcggcgc	1320
gtaccgcgag	tggttcgaga	agcaatacga	cgccgcaccg	gcgaaggcca	ccgcatgaag	1380
ctgctgctgc	tgggcaaggg	cggacaggtc	ggctgggagc	tgcaacgcag	cctcgcgccc	1440
ctgggccaac	tggtggcgct	cgatttcgac	agcaccgact	tcaacgccga	cttcagtgcg	1500
cccgcgcagc	tgcccgagac	agtgcctgaag	gtgcgccccg	acgtcatcgt	caatgccgca	1560
gcgcacaccg	cggtcgacaa	ggccgagagc	gagcccgagt	tcgcgcgcaa	gctcaacgcc	1620

ES 2 809 727 T3

acctcgcccc	gcgtgggtggc	cgaagccgcg	cagcagatcg	gcgcgctgat	ggttcactac	1680
tcgaccgact	acgtcttcga	cggcagcggc	agcaagccgt	ggaagaaga	cgatgcgacc	1740
ggcccgcctca	gcgtctacgg	cagcaccaag	ctcgaaggcg	agcaactgg	ggcaaagcac	1800
tgtgcgaagc	acctgatctt	tcgcaccagc	tgggtctatg	ccgcgcgcgg	cggcaacttc	1860
gccaagacca	tgtgcgcat	cgccaaggag	cgcgacaagc	tgaccgtcat	cgacgaccag	1920
ttcggcgcgc	ccaccggcgc	ggaactgctg	gccgacatca	ccgcgcacgc	gattcgcgcg	1980
acgctgcagg	acccgtccaa	ggccgggctc	tatcacgcgg	tggccgggtg	cgtgaccacg	2040
tggcacggct	atgcgcgctt	cgtgatcgag	caggccaagg	cggcggggcg	ggaactgaag	2100
gccggccccg	aagcggctga	gcccggtgcc	accacggcat	tcccgcgcgc	ggccaggcgg	2160
ccgcacaact	cgcgcttggg	caccaccaag	ctgcaatcga	ccttcggcct	cgtgctgccc	2220
gagtggcagt	ccggcgtcgc	ccgcatgttg	cgcgaaacct	tctgatattc	gcagagcaag	2280
agagacacga	acaccccatg	accaagacga	cgcaacgcaa	aggcatcatc	ctcgccggtg	2340
gctcggggac	ccgcctgcac	cccgcgcgcg	ttgccatgag	caaacaactg	ctgccggtgt	2400
acgacaagcc	gatgatctat	taccgcgtga	gcacgctgat	gctggggcgc	atgcgcgaca	2460
tcctgatcat	cagcacgcgc	caggacacgc	cgcgtttcca	gcaactgctg	ggggatggca	2520
gccaatgggg	catcaacctg	cagtacgcgc	tgcagccgag	cccggatggt	ctggcgcaag	2580
cgttcacat	cggtgacaag	ttcgtgggca	acgacccgag	tgcgctggtg	ctgggggaca	2640
acatcttcta	tggccacgac	ttcgcccatc	tgctggccga	tgccgcgcgc	aagacctcgg	2700
gtgcgacggt	gttcgcctac	cacgtgcacg	accccgagcg	ctacggcggtg	gtggccttcg	2760
atgccaaggg	cagggcgagc	agcatcgaag	aaaagccgct	caagcccaag	agcagctatg	2820
cggtcacggg	cctctacttc	tacgacaacc	aggtcgtcga	catcgccaag	gccgtgaagc	2880
cgagcgcgcg	cggcgaaact	gagatcaccg	cggcacaacc	ggcgtatctc	gacctcgacc	2940
agctgaacgt	gcagatcatg	cagcgcggct	atgcgtggct	cgataccggt	acgcacgaca	3000
gcctgctgga	agccgggagc	ttcattgcca	cgctcgagca	ccgccagggg	ctgaagatcg	3060
catgccccga	agagatcgca	tggcgcaatg	gcttcattct	aaccgagcaa	ctcgaaaagc	3120
tcgcgcgcgc	gctggaaaag	agcggctacg	gcaagtacct	caagcacctg	ctgaacgacg	3180
aggtgcgctc	gtgaaggcca	cgccacacct	gattcctgac	gtgctcgtga	tcgagccgaa	3240
gggtgttggc	gatgcacggg	gcttcttctt	cgaaagcttc	aaccagaagg	ccttcgacga	3300
agcgatcggc	aagcatgtcg	acttcgtgca	ggacaacccat	tcgcgatcgg	ccaagggtgt	3360
gctgcggggg	ctgcattacc	aggtccagca	gccgcaaggc	aagctcgtgc	gggtggtgcg	3420
tgggtcgggtg	ttcgacgtgg	ccgtcgacat	ccgcaagtcg	tcgccgactt	ttggcaaatg	3480

ES 2 809 727 T3

gggtgggtgtc gagttgaacg aagacaacca caagcagctc tgggtgccgg caggattcgc 3540
gcacggtttc ctggtgttga gcgagaccgc ggaattcctc tacaagacca ccgactacta 3600
cgcgcccgcc cacgagcgcg cgattgtctg gaacgacccc gctgtcggta ttcgatggcc 3660
ggatgtggga ggggcaccgg tcctgtcgaa gaaggacgaa gacgggtgtc ttctgcaagc 3720
ggcagagggt ttctagtgtc ctttcgtcag atagcggggc ggcttcgcgt atcgggatcc 3780
cgcggtgagc ccgcaagagt gccctgagag ggggggcgaa aaactcacia cgccactgcc 3840
tcgagcaaac gtgcgtctcg cagctttctg aagtgtgtgc accttctttt tttttctctt 3900
acatctttga aatgattttg aaaatccgcg gcgatcgcat gcatgctgct ggaatcacc 3959

<210> 3

<211> 915

5 <212> ADN

<213> Variovorax paradoxus

<400> 3

atgaatggca tgcataatcga ctcggtcgac ctcaatctgc tgcgcctgtt cgatgcggtc 60
taccgcgagc gcagcgtgag ccgcgcccgc gaggcgtggt gcctcacgca gcctgcggca 120
agccatgggc tgggacggct gcggctgctt ttgaaagacg cgctcttcac gcgtgcccc 180
ggcggcgtgg cgcacacgcc gcgcgcccgc cggctcgcgg tggcgggtgca ggcggcgcctc 240
ggcacgatcg aagcggcgct gcacgagccc gatcgcttcg agccccaggt gtcgcgcaag 300
agctttcgta ttcacatgag cgacatcggc gagggcgct tcctgcccgc gctgatggcg 360
cggctcggcg agctggcgcc cggcgtgcgg ctggagaccc tgcgcctctt gcctgcggag 420
gttgcgccc cactcgacag cggccgcata gatttcgcct tcggctttct ctcgaccgtg 480
cgcgacacgc agcgcacgca tcttctgaaa gaccgctaca tcgtgctgct gcgcaagggc 540
catccctttg tgaagcgccg gcgcaagggg caggcgctgc tcgaggcgct gcaggagctc 600
gactacgtgg cggcgcgcac gcacgcccgc acgctgcgca tcttgagtt gctcaacctc 660
gaagaccgcc tgcgcctcac gaccgagcac ttcattggtgc taccggccat cgtgcgcgcc 720
accgatctcg cggcgggtgat gccgcgcaac atcgcgcgag ggcttgcgga ggagggcgcc 780
tacgcgatcg tcgagccgcc gtttccgctg cgcgatttca gcgtgtcgct gcactggagc 840
aagcgcttcg agggcgaccc ggccaaccgt tgggtgcggc aggtgatcac ggcgctgttc 900
tccgagcgcg gctga 915

10

<210> 4

<211> 304

<212> PRT

<213> Variovorax paradoxus

15

<400> 4

ES 2 809 727 T3

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

ES 2 809 727 T3

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

Phe Ser Val Ser Leu His Trp Ser Lys Arg Phe Glu Gly Asp Pro Ala
275 280 285

Asn Arg Trp Leu Arg Gln Val Ile Thr Ala Leu Phe Ser Glu Arg Gly
290 295 300

<210> 5

<211> 7476

5 <212> ADN

<213> Variovorax paradoxus

<400> 5

atgagtaccg	tcgatcagct	gggccgcacc	gccccctta	cctcggggca	gatggcgatg	60
tggctcggcg	caaagttcgc	gtcgcccgac	accaatttca	atctcgccga	agccatcgac	120
atcgcaggcg	agatcgaccc	cgcgatcttc	ctggcggcca	tgcgacaggt	ggccgatgaa	180
gtcgaggcca	cgcgcctgag	cttcacatgat	accccgcaag	ggccacgaca	ggtcgtcgcg	240
cccgttttca	ccggcgagat	cccctacctc	gacctcagcg	gcgagagcga	tccgcaggcc	300
gaggccgagc	gctggatgca	tgcggaactac	acccgcagca	tcgacctcgc	gcacgggcag	360
ctgtggctgt	ccgcgctgat	ccgcctcgcg	cccgatcgcc	acatctggtg	ccaccgcagc	420
catcacatcg	cgtctgacgg	cttcagcggc	ggcctcatcg	cacgccgctt	cgccgacatc	480
tacaccgcga	tggtcgacaa	caacgcagcg	gtgcccgaag	actcgcgctt	tgacccgatc	540
tcgcagctgg	ccgacgaaga	acatgcctat	cgcgagtccg	gccgcttccc	gcgcgaccgc	600
cagtactgga	ccgagcgctt	cgccgatgca	cccgatccgt	tgagcctcgc	ctcgcacccgc	660
tcggtcaacg	tcggtggcct	cttgcgccag	acggtgcacc	tgccggcggc	cagcgtgcaa	720
gccctgcaga	ccatcgcgca	agagctcggc	accacgctgc	cgcaaatacct	catcgccacc	780
accgcggcct	acctgtaccg	cgcaacgggc	atcgaggaca	tggaatacgg	catccccgtc	840
accgcgcgcc	acaacgaccg	catgcgccgc	gtgcccgcga	tggtggccaa	cgcgctgccg	900
ctgcgcctgg	cgatgcgcgc	ggacctgccg	attccggaac	tgatccgcga	agtcggccgg	960
cagatgcggc	agatcctcgc	gcaccagtgc	tatcgctacg	agcatttgcg	cagcgacctc	1020
aacatgctgg	tgaacaaccg	gcagctcttc	accacgctgg	tcaacgtcga	gcccttcgac	1080
tacgacttcc	gctttgctgg	ccatgccgcg	aagccgcgca	acctctcgaa	cggcacggcc	1140
gaggacctcg	gcattcttct	gtacgagcgc	ggcaacgggc	aggacctgca	gatcgacttc	1200
gacgcccaacc	ccgcgggtgca	caccgcagag	gaactggccg	atcaccagcg	ccggctgctt	1260
gccttcatcg	acgccgtgat	ccgcctgccg	ttgcaggccg	tcggccagat	cgacctgctc	1320

ES 2 809 727 T3

ggtgccgaag agcggcagca attgctggtc gagtggaacg acacggccca cgcctgccc 1380
 gacacccatc tcaccgcgtt gatcgaagcg cagctcgag ccgatccgca agccatcgca 1440
 ttgcgcttcg acggcgaggg gatgaacaac gaagaactga accgcccgc caaccgtctc 1500
 gccacactgc tgcgcgcacg cggcgctggc ccggagcgca ccgtggcgt cgcgatcccg 1560
 cgttcgatgg acctgatgat tgccttgctc gccacgttga agaccggcg gcctacctg 1620
 ccggtcgatc cggatttccc ggcgacgc atcgcttca tgctcgcgca tgcgcagccc 1680
 gtgtgcctcg tcacgaccga agcctcgcg gagtcgctgc cggcagccgc cccacattg 1740
 ctgctcgatg tagcgcaaac gattgcggat ctggagagtt gcaacgacac caaccgggc 1800
 atcgcgatcg acccttcgca tccggcctat gtgatctaca cctcgggctc gaccggcatg 1860
 cccaagggtg cggtcgtgtc gcaccgcgc atcgtcaacc gcctgcgctg gatgcaggac 1920
 cgctacggcc ttcaggccga cgaccgcgtg ctgcagaaga cgccttcag cttcgacgtg 1980
 tcggtgtggg agttcttctg gccgctgac gacgggtcca cgtcgggtgt tgcgaaaccg 2040
 ggcggccaca aggatgcggc ctacctcgcg gggctgatcg cggaggagg catcaccacg 2100
 atccacttcg tgccgctgat gctcgaggtc ttctgctcg agccacggc gggcgcatgc 2160
 accacgctgc gccgcgtgat ctgcagcggc gaagccttgt cgcgcgctg gcaatcgag 2220
 ttccagcagc acctctcgtg cgagctgcac aacctctacg gtccgaccga ggcgcggtc 2280
 gacgtcacct cgtgggagtg cgaacgcacg gacgacgag aagcctcgag cgttccatc 2340
 ggccgccga tctggaacac ccagatgcac gtgctcgaca gcggcctgca gccctgccc 2400
 gccggcgtga ctggcgagct gtacatcgcg ggcgtcgcc tcgcacgcgg ctacctcaag 2460
 cgcctcgttc tgagcgccga gcgtttcatc gccaaacct acggcacacc cggcagccgc 2520
 atgtaccgca ccggcgacct cgcgcgctgg cgcaaggacg gcagccttga cttcctcggc 2580
 cgcgcgacc agcaggtgaa gatccggggc ctgcgcatcg agccgggaga gatcgaatcc 2640
 gtgctgctgc agcatccgca agtcgcgag gccgcgctgg tggcgcgca agacgtaccg 2700
 ggcgaaaagc gtctcgtggc ctacgtcgtt gcgacggacg ctgccgatcc gcaagcgcc 2760
 gaactgcgca cgcgcctcgc gcaatcgctg cccgagtaca tgggtgcctt gcccttcgtc 2820
 agcctccgt cgtgcgct cggaccagc ggcaagctcg accgcaaggc gctgcgccc 2880
 cccgaagtgc aggcgcacc gccgtacgc gcgcgcgca cgcgaccga aaagatcctg 2940
 gccggcctct gggccgagac gctgcatttg ccgcgctcg gtgtcaacga caacttcttc 3000
 gaactcggcg gccactcgct gatgatcgtg cagctcatgt cgatgatccg gcagcaattc 3060
 atgatcgacc tgccggtcga cacgctgttc caggtctcca ccatcgcggg ccttgccgag 3120
 ctgctcgacc aggaatcggc cgcctcgc agcctgactc cgatgccgcg ccccgcgcg 3180
 attccgctgt ccttcgcgca gcgcgcctg tggctgatga accagctcga aggcgcgaac 3240

ES 2 809 727 T3

ccggcctaca acatgccgct cgcgctgcgc ctgtcgggtg tgctcgatcg caccgcattg 3300
catgcggcgc tcggcgacct ggtgcagcgc cagagagcc tgcgcacggt ctacccgaac 3360
gaagacgggc tgccgtacca gcacatcctc gacggcgcgg atgcgcgtcc ggcggtgatc 3420
gagggcgaca gcagcgaaga agaaatcgcg gcgcagcttc acgccgctgc gggccatgcc 3480
ttcgatctcg gcagcgcggc gcccttgcgc gtctacctgt tcaagctcgc cggcgacgaa 3540
cacgtgctgc tgctgctcac gcaccacatt gccggcgatg gcgcctcgct gctgccgcta 3600
gcgcgcgaca tcagcgtggc ctatgccgcg cgctgcgaag gcaaggcgcc gggctgggag 3660
ccgctgcgcg tgcaatacgc cgactacgcg ctgtggcagc aggagctgct cggcagcgaa 3720
gacgatgccg agagcatggc cggccgccag cgtgagttct ggcgttcctc gctgagcgac 3780
ctgcccgagc aactggcgct gcccgctgac cagcacggc cgctcgtgcc gacctaccgc 3840
ggcgatgtgg tcccgctgca gattccgctg catgtgcatg aacgcatacct gcaactggcg 3900
cgcgacgggc aggccagcgt cttcatggtg ctgcaggccg cactcgcggg cctcctgagc 3960
cgcctcggcg cgggcgacga catcgtcatc ggcagcccg tcgcggggcg cagcgacct 4020
gcgctggacg aactcatcgg ctgcttcgtc aacacgctgg tgctgcgcac tgacacctcg 4080
ggccagccga gcctgcgcga gctggtctcg cgcgtgcgcg ccaccaacct cgcggcctat 4140
gcgaaccagg agtttccgta cgaccgcctc gtggagctgc tgcgctccgg cgcctcgcgc 4200
gccaaacctg cgctgttcca ggtcatgctg ggcttcagg gcacgagccg cctgtcgttc 4260
agcctgccgg gcctgtcgat cgcgccgcag ccggtggcca tcgacaccgc gaagttcgac 4320
ctgtcgttca tcctcggcga gcaacgcggt gccgatggc tcgcggggcg catctccggc 4380
ggcatccagt acagcaccga cctgttcgag cgcagcacgg tcgaggccat gggcgcgcg 4440
ctggtgcgtt tgctggaaga ggcctgcgag gcgcccgacg atgcggtgag tggcctcgcc 4500
atcctgagcg cggaagaaac cgaccgcctg ctgtccgact ggagcggccg cagcgcgac 4560
cttgccgcgc tctcgttcgc cgacatggtg gcctcgcatg ccgcggagcg cccgcttgca 4620
gatgcagtgg tgctcgacga cgcgaccgtc agctacgccg aactcgatgc acgcgccaac 4680
cggctctcgc acctgctgcg tgcgcaaggc atcggggttg gcgccatcgt cgcgacagt 4740
ctgcccggtt cgctcgacct catcgtggcg cacttgcca tcgtgaaggc cggcgcggcc 4800
tacctgcca tcgacccaa ccacatggcc gcgcgcagcg ccttcgtgtt cgaggaggcc 4860
gcgcccgccg cgggtgctgac gcacgatgcg ctgttgcccg agctggctcg cgttccccgc 4920
tgcatcgcgc tcgacagcga cagcatggtt gccgcgctgg ccatccagtc ggatacgccg 4980
ctggtgcatg cggccaatcc acaggatgcc gcctacctca tctacacctc cggctccacc 5040
ggcatgcca agggcggtgt ggtgccgcat gcgggcctgg gcagcctcg caccgcgatg 5100

ES 2 809 727 T3

gcggagcggc tcgtcatcgg ccacggctcg cgcgtgctgc agttctcctc cagcggcttc 5160

gacgcgtcgg tgatggacca gctgatggcc ttggcgccg gtgccgcgt ggtggtgccg 5220

gggcccggagc aactgctcgg cacggagctg gccgatctgc tcgagaagca ggccgtgagc 5280

cacgcgtcga ttccgccgcg cgcgtcgcg accctgccgc acggcgagtt cccgcacctg 5340

cagacgctgg tggtcggcgg cgatgcctgc accgccgcg tggcggcgaa gtggtcgcaa 5400

ggccgccgca tgatcaacgc ctacggcccc accgagatca ccatctgcgc gagcatgagc 5460

gcgccgatga cggccgagga gttgccctcc atcgccacgc cgatctggaa cagcgggatg 5520

tatgtgctcg acagcgccct gcaaccgggtg ccgccgggtg tcgoggcgga gctctacatc 5580

gccggcagcg gcgtggcgcg cggctatctc aaccggccgg cattgagtgc ggaacgcttc 5640

atcgccgacc cgcattggcg gcccggcagc cgcattgtacc gcagcggcga cctcgcacgc 5700

tggcgcgccg acggcacgct cgacttcctc ggccgcgcg accagcaggt gaagatccgg 5760

ggcttcgcga tcgagccggg cgagatcgaa tcggtgctgc tcaagcaccg gttgatcacg 5820

caggccgcgg tgatcgcccg cgaggacgtg cccggcgaga agcgctggt cgcctacttc 5880

gtcgccgggt ccgagccgca gccaccgag ctgcgcgcg acatggcgca ggccttgccc 5940

gactacatgg tgccttcggc cttcgtgcgc ctgcctgcgc tgcgctcac gcaaagcggc 6000

aagctcgaca agaaggcgct gccggtgcc gaccagcagc ccgccgcgt gtacgtggag 6060

ccccgcacgc cgaccgagaa actgctcgcg ggccctctggt ccgagacgct gcacctggag 6120

cgtgtcggca tccacgacaa cttcttcgag atcgccgggc attcgctcat ggcgatccag 6180

ctgggcatgc gcatccgcca gcaggtgcgc gcggacttcc cgcacgccga ggtctacaac 6240

cgcgcgacga ttgccgacct ggccgcctgg ctcgacaacg aaggcggcac ggtcgaggcg 6300

ctggacctgt cgcgcgagct cgacctgcc gcgcacatcc gcccgcaggc cactgcaccg 6360

aagctcgcac cgcgcgcgt gttcctcacc ggccgcgagc gcttcgtcgg cagtcacctg 6420

ctggccgcgc tgttcgcgca caccgcggcc tgcgtggtct gccacgtgcg cgcgcccgac 6480

gagcaggccg gcgagcagcg cctcaagcgc acgctggccc agcggccagct cggtgcgatc 6540

tgggacaacg cgcgcacaa ggtcgtgacc ggcgacctcg gcaagcccg cctgggcctc 6600

gatgacgctg ccgtgcaact ggtgcgcgac ggctgcgacg ccatctacca ctgcgcccg 6660

caggtcgact tcctgcatcc ctacgcgagc ctcaagcccc cgaacgtcga cagcgtggtc 6720

acgctgctcg aatggacggc gcagggggcg gcgaagagca tgcactacgt ctccacgctg 6780

gctgtgatcg accagaacaa caaggaagac accatcaccg agcaatcggc gctggcctca 6840

tggagcgggc tggtcgacgg ctacagccag agcaagtggg tcggcgatgc gctggcccg 6900

gaggcgcagg cgcgcggcat gccggtggcg atctaccggc tgggggcagt caccggcgac 6960

cacacgcacg cgatctgcaa tgccgacgac ctgatctggc gcgtggcgca tctctatgcc 7020

ES 2 809 727 T3

gacctggaag cgattcccga tatggacctg ccgctcaacc tcacaccggt ggacgacgtg 7080
 gcgcgcgccca tcctcggcct tgcggcgag gaggcctcgt ggggccaggt gttccacctg 7140
 atgagccagg cggcgctgcg ggtgcgcgac attccgcacg tcttcgagcg catgggcatg 7200
 cggctggagc cggtcgggct ggagccctgg ctgcagcgcg cgcacgcacg gctggccgctc 7260
 gcgcacgacc gcgacctggc cgcgggtgctc gccatcctcg accgctacga caccacggcc 7320
 acgccgccgc aggtgagcgg cgcggccacg catgcgcagc tcgaggccat cggcgcgccg 7380
 atccgcccg tggaccgcga cctgctgcag cgctacttcg tcgacctggg catcgacacc 7440
 aaggcgcgcc gcgccctgga aaccaccact tcatag 7476

<210> 6

<211> 2491

5 <212> PRT

<213> Variovorax paradoxus

<400> 6

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

ES 2 809 727 T3

Tyr Thr Ala Met Val Asp Asn Asn Ala Ala Val Pro Glu Asp Ser Arg
 165 170 175
 Leu Ala Pro Ile Ser Gln Leu Ala Asp Glu Glu His Ala Tyr Arg Glu
 180 185 190
 Ser Gly Arg Phe Pro Arg Asp Arg Gln Tyr Trp Thr Glu Arg Phe Ala
 195 200 205
 Asp Ala Pro Asp Pro Leu Ser Leu Ala Ser His Arg Ser Val Asn Val
 210 215 220
 Gly Gly Leu Leu Arg Gln Thr Val His Leu Pro Ala Ala Ser Val Gln
 225 230 235 240
 Ala Leu Gln Thr Ile Ala Gln Glu Leu Gly Thr Thr Leu Pro Gln Ile
 245 250 255
 Leu Ile Ala Thr Thr Ala Ala Tyr Leu Tyr Arg Ala Thr Gly Ile Glu
 260 265 270
 Asp Met Ala Ile Gly Ile Pro Val Thr Ala Arg His Asn Asp Arg Met
 275 280 285
 Arg Arg Val Pro Ala Met Val Ala Asn Ala Leu Pro Leu Arg Leu Ala
 290 295 300
 Met Arg Ala Asp Leu Pro Ile Pro Glu Leu Ile Arg Glu Val Gly Arg
 305 310 315 320
 Gln Met Arg Gln Ile Leu Arg His Gln Ser Tyr Arg Tyr Glu His Leu
 325 330 335
 Arg Ser Asp Leu Asn Met Leu Val Asn Asn Arg Gln Leu Phe Thr Thr
 340 345 350
 Val Val Asn Val Glu Pro Phe Asp Tyr Asp Phe Arg Phe Ala Gly His
 355 360 365
 Ala Ala Lys Pro Arg Asn Leu Ser Asn Gly Thr Ala Glu Asp Leu Gly
 370 375 380
 Ile Phe Leu Tyr Glu Arg Gly Asn Gly Gln Asp Leu Gln Ile Asp Phe
 385 390 395 400
 Asp Ala Asn Pro Ala Val His Thr Ala Glu Glu Leu Ala Asp His Gln

ES 2 809 727 T3

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

ES 2 809 727 T3

Ser Phe Asp Val Ser Val Trp Glu Phe Phe Trp Pro Leu Ile Asp Gly
 660 665 670
 Ala Thr Leu Val Leu Ala Lys Pro Gly Gly His Lys Asp Ala Ala Tyr
 675 680 685
 Leu Ala Gly Leu Ile Ala Glu Glu Gly Ile Thr Thr Ile His Phe Val
 690 695 700
 Pro Ser Met Leu Glu Val Phe Leu Leu Glu Pro Thr Ala Gly Ala Cys
 705 710 715 720
 Thr Thr Leu Arg Arg Val Ile Cys Ser Gly Glu Ala Leu Ser Pro Ala
 725 730 735
 Leu Gln Ser Gln Phe Gln Gln His Leu Ser Cys Glu Leu His Asn Leu
 740 745 750
 Tyr Gly Pro Thr Glu Ala Ala Val Asp Val Thr Ser Trp Glu Cys Glu
 755 760 765
 Arg Thr Asp Asp Ala Glu Ala Ser Ser Val Pro Ile Gly Arg Pro Ile
 770 775 780
 Trp Asn Thr Gln Met His Val Leu Asp Ser Gly Leu Gln Pro Val Pro
 785 790 795 800
 Ala Gly Val Thr Gly Glu Leu Tyr Ile Ala Gly Val Gly Leu Ala Arg
 805 810 815
 Gly Tyr Leu Lys Arg Pro Leu Leu Ser Ala Glu Arg Phe Ile Ala Asn
 820 825 830
 Pro Tyr Gly Thr Pro Gly Ser Arg Met Tyr Arg Thr Gly Asp Leu Ala
 835 840 845
 Arg Trp Arg Lys Asp Gly Ser Leu Asp Phe Leu Gly Arg Ala Asp Gln
 850 855 860
 Gln Val Lys Ile Arg Gly Leu Arg Ile Glu Pro Gly Glu Ile Glu Ser
 865 870 875 880
 Val Leu Leu Gln His Pro Gln Val Ala Gln Ala Ala Val Val Ala Arg
 885 890 895
 Glu Asp Val Pro Gly Glu Lys Arg Leu Val Ala Tyr Val Val Ala Thr
 900 905 910

ES 2 809 727 T3

Asp Ala Ala Asp Pro Gln Ala Ala Glu Leu Arg Thr Arg Leu Ala Gln
 915 920 925
 Ser Leu Pro Glu Tyr Met Val Pro Ser Ala Phe Val Ser Leu Pro Ser
 930 935 940
 Leu Pro Leu Gly Pro Ser Gly Lys Leu Asp Arg Lys Ala Leu Pro Pro
 945 950 955 960
 Pro Glu Val Gln Ala Ala Thr Pro Tyr Ala Ala Pro Arg Thr Pro Thr
 965 970 975
 Glu Lys Ile Leu Ala Gly Leu Trp Ala Glu Thr Leu His Leu Pro Arg
 980 985 990
 Val Gly Val Asn Asp Asn Phe Phe Glu Leu Gly Gly His Ser Leu Met
 995 1000 1005
 Ile Val Gln Leu Met Ser Met Ile Arg Gln Gln Phe Met Ile Asp
 1010 1015 1020
 Leu Pro Val Asp Thr Leu Phe Gln Val Ser Thr Ile Ala Gly Leu
 1025 1030 1035
 Ala Glu Leu Leu Asp Gln Glu Ser Val Ala Arg Pro Ser Leu Thr
 1040 1045 1050
 Pro Met Pro Arg Pro Ala Arg Ile Pro Leu Ser Phe Ala Gln Arg
 1055 1060 1065
 Arg Leu Trp Leu Met Asn Gln Leu Glu Gly Ala Asn Pro Ala Tyr
 1070 1075 1080
 Asn Met Pro Leu Ala Leu Arg Leu Ser Gly Val Leu Asp Arg Thr
 1085 1090 1095
 Ala Leu His Ala Ala Leu Gly Asp Leu Val Gln Arg His Glu Ser
 1100 1105 1110
 Leu Arg Thr Val Tyr Pro Asn Glu Asp Gly Leu Pro Tyr Gln His
 1115 1120 1125
 Ile Leu Asp Gly Ala Asp Ala Arg Pro Ala Val Ile Glu Ala Asp
 1130 1135 1140
 Ser Ser Glu Glu Glu Ile Ala Ala Gln Leu His Ala Ala Ala Gly
 1145 1150 1155

ES 2 809 727 T3

His	Ala	Phe	Asp	Leu	Gly	Ser	Ala	Ala	Pro	Leu	Arg	Val	Tyr	Leu
1160						1165					1170			
Phe	Lys	Leu	Ala	Gly	Asp	Glu	His	Val	Leu	Leu	Leu	Leu	Thr	His
1175						1180					1185			
His	Ile	Ala	Gly	Asp	Gly	Ala	Ser	Leu	Leu	Pro	Leu	Ala	Arg	Asp
1190						1195					1200			
Ile	Ser	Val	Ala	Tyr	Ala	Ala	Arg	Cys	Glu	Gly	Lys	Ala	Pro	Gly
1205						1210					1215			
Trp	Glu	Pro	Leu	Pro	Leu	Gln	Tyr	Ala	Asp	Tyr	Ala	Leu	Trp	Gln
1220						1225					1230			
Gln	Glu	Leu	Leu	Gly	Ser	Glu	Asp	Asp	Ala	Glu	Ser	Met	Ala	Gly
1235						1240					1245			
Arg	Gln	Arg	Glu	Phe	Trp	Arg	Ser	Ser	Leu	Ser	Asp	Leu	Pro	Glu
1250						1255					1260			
Gln	Leu	Ala	Leu	Pro	Val	Asp	His	Ala	Arg	Pro	Leu	Val	Pro	Thr
1265						1270					1275			
Tyr	Arg	Gly	Asp	Val	Val	Pro	Leu	Gln	Ile	Pro	Ser	His	Val	His
1280						1285					1290			
Glu	Arg	Ile	Leu	Gln	Leu	Ala	Arg	Asp	Gly	Gln	Ala	Ser	Val	Phe
1295						1300					1305			
Met	Val	Leu	Gln	Ala	Ala	Leu	Ala	Gly	Leu	Leu	Ser	Arg	Leu	Gly
1310						1315					1320			
Ala	Gly	Asp	Asp	Ile	Val	Ile	Gly	Ser	Pro	Val	Ala	Gly	Arg	Ser
1325						1330					1335			
Asp	His	Ala	Leu	Asp	Glu	Leu	Ile	Gly	Cys	Phe	Val	Asn	Thr	Leu
1340						1345					1350			
Val	Leu	Arg	Thr	Asp	Thr	Ser	Gly	Gln	Pro	Ser	Leu	Arg	Glu	Leu
1355						1360					1365			
Val	Ser	Arg	Val	Arg	Ala	Thr	Asn	Leu	Ala	Ala	Tyr	Ala	Asn	Gln
1370						1375					1380			
Glu	Phe	Pro	Tyr	Asp	Arg	Leu	Val	Glu	Leu	Leu	Arg	Pro	Gly	Arg

ES 2 809 727 T3

1385		1390		1395
Ser Arg Ala Asn Leu Pro Leu Phe Gln Val Met Leu Gly Phe Gln 1400 1405 1410				
Gly Thr Ser Arg Leu Ser Phe Ser Leu Pro Gly Leu Ser Ile Ala 1415 1420 1425				
Pro Gln Pro Val Ala Ile Asp Thr Ala Lys Phe Asp Leu Ser Phe 1430 1435 1440				
Ile Leu Gly Glu Gln Arg Gly Ala Asp Gly Leu Pro Gly Gly Ile 1445 1450 1455				
Ser Gly Gly Ile Gln Tyr Ser Thr Asp Leu Phe Glu Arg Ser Thr 1460 1465 1470				
Val Glu Ala Met Gly Ala Arg Leu Val Arg Leu Leu Glu Glu Ala 1475 1480 1485				
Cys Glu Ala Pro Asp Asp Ala Val Ser Gly Leu Ala Ile Leu Ser 1490 1495 1500				
Ala Glu Glu Thr Asp Arg Leu Leu Ser Asp Trp Ser Gly Arg Thr 1505 1510 1515				
Arg Asp Leu Ala Pro Leu Ser Phe Ala Asp Met Val Ala Ser His 1520 1525 1530				
Ala Ala Glu Arg Pro Leu Ala Asp Ala Val Val Leu Asp Asp Ala 1535 1540 1545				
Thr Val Ser Tyr Ala Glu Leu Asp Ala Arg Ala Asn Arg Leu Ser 1550 1555 1560				
His Leu Leu Arg Ala Gln Gly Ile Gly Val Gly Ala Ile Val Ala 1565 1570 1575				
Thr Val Leu Pro Arg Ser Leu Asp Leu Ile Val Ala His Leu Ala 1580 1585 1590				
Ile Val Lys Ala Gly Ala Ala Tyr Leu Pro Ile Asp Pro Asn His 1595 1600 1605				
Met Ala Ala Arg Ser Ala Phe Val Phe Glu Glu Ala Ala Pro Ala 1610 1615 1620				

ES 2 809 727 T3

Ala Val	Leu Thr His Asp	Ala	Leu Leu Pro Glu	Leu	Val Gly Val
1625		1630		1635	
Pro Arg	Cys Ile Ala Leu	Asp	Ser Asp Ser Met	Val	Ala Ala Leu
1640		1645		1650	
Ala Ile	Gln Ser Asp Thr	Pro	Leu Val His Ala	Ala	Asn Pro Gln
1655		1660		1665	
Asp Ala	Ala Tyr Leu Ile	Tyr	Thr Ser Gly Ser	Thr	Gly Met Pro
1670		1675		1680	
Lys Gly	Val Val Val Pro	His	Ala Gly Leu Gly	Ser	Leu Gly Thr
1685		1690		1695	
Ala Met	Ala Glu Arg Leu	Val	Ile Gly His Gly	Ser	Arg Val Leu
1700		1705		1710	
Gln Phe	Ser Ser Ser Gly	Phe	Asp Ala Ser Val	Met	Asp Gln Leu
1715		1720		1725	
Met Ala	Phe Gly Ala Gly	Ala	Ala Leu Val Val	Pro	Gly Pro Glu
1730		1735		1740	
Gln Leu	Leu Gly Thr Glu	Leu	Ala Asp Leu Leu	Glu	Lys Gln Ala
1745		1750		1755	
Val Ser	His Ala Leu Ile	Pro	Pro Ala Ala Leu	Ala	Thr Leu Pro
1760		1765		1770	
His Gly	Glu Phe Pro His	Leu	Gln Thr Leu Val	Val	Gly Gly Asp
1775		1780		1785	
Ala Cys	Thr Ala Ala Leu	Ala	Ala Lys Trp Ser	Gln	Gly Arg Arg
1790		1795		1800	
Met Ile	Asn Ala Tyr Gly	Pro	Thr Glu Ile Thr	Ile	Cys Ala Ser
1805		1810		1815	
Met Ser	Ala Pro Met Thr	Ala	Glu Glu Leu Pro	Ser	Ile Gly Gln
1820		1825		1830	
Pro Ile	Trp Asn Thr Arg	Met	Tyr Val Leu Asp	Ser	Ala Leu Gln
1835		1840		1845	
Pro Val	Pro Pro Gly Val	Ala	Gly Glu Leu Tyr	Ile	Ala Gly Ser
1850		1855		1860	

ES 2 809 727 T3

Gly Val 1865	Ala Arg Gly Tyr 1870	Leu Asn Arg Pro Ala Leu 1875	Ser Ala Glu
Arg Phe 1880	Ile Ala Asp Pro His 1885	Gly Ala Pro Gly Ser 1890	Arg Met Tyr
Arg Ser 1895	Gly Asp Leu Ala Arg 1900	Trp Arg Ala Asp Gly 1905	Thr Leu Asp
Phe Leu 1910	Gly Arg Ala Asp Gln 1915	Gln Val Lys Ile Arg 1920	Gly Phe Arg
Ile Glu 1925	Pro Gly Glu Ile Glu 1930	Ser Val Leu Leu Lys 1935	His Pro Leu
Ile Thr 1940	Gln Ala Ala Val Ile 1945	Ala Arg Glu Asp Val 1950	Pro Gly Glu
Lys Arg 1955	Leu Val Ala Tyr Phe 1960	Val Ala Gly Ser Glu 1965	Pro Gln Pro
Thr Glu 1970	Leu Arg Ala His Met 1975	Ala Gln Ala Leu Pro 1980	Asp Tyr Met
Val Pro 1985	Ser Ala Phe Val Arg 1990	Leu Pro Ser Leu Pro 1995	Leu Thr Gln
Ser Gly 2000	Lys Leu Asp Lys Lys 2005	Ala Leu Pro Val Pro 2010	Asp Gln Gln
Pro Ala 2015	Ala Leu Tyr Val Glu 2020	Pro Arg Thr Pro Thr 2025	Glu Lys Leu
Leu Ala 2030	Gly Leu Trp Ser Glu 2035	Thr Leu His Leu Glu 2040	Arg Val Gly
Ile His 2045	Asp Asn Phe Phe Glu 2050	Ile Gly Gly His Ser 2055	Leu Met Ala
Ile Gln 2060	Leu Gly Met Arg Ile 2065	Arg Gln Gln Val Arg 2070	Ala Asp Phe
Pro His 2075	Ala Glu Val Tyr Asn 2080	Arg Pro Thr Ile Ala 2085	Asp Leu Ala
Ala Trp 2090	Leu Asp Asn Glu Gly 2095	Gly Thr Val Glu Ala 2100	Leu Asp Leu

ES 2 809 727 T3

Ser Arg 2105	Glu Leu Asp Leu Pro 2110	Ala His Ile Arg Pro 2115	Gln Ala Thr
Ala Pro 2120	Lys Leu Ala Pro Arg 2125	Arg Val Phe Leu Thr 2130	Gly Ala Ser
Gly Phe 2135	Val Gly Ser His Leu 2140	Leu Ala Ala Leu Leu 2145	Arg Asp Thr
Ala Ala 2150	Cys Val Val Cys His 2155	Val Arg Ala Pro Asp 2160	Glu Gln Ala
Gly Glu 2165	Gln Arg Leu Lys Arg 2170	Thr Leu Ala Gln Arg 2175	Gln Leu Gly
Ala Ile 2180	Trp Asp Asn Ala Arg 2185	Ile Lys Val Val Thr 2190	Gly Asp Leu
Gly Lys 2195	Pro Arg Leu Gly Leu 2200	Asp Asp Ala Ala Val 2205	Gln Leu Val
Arg Asp 2210	Gly Cys Asp Ala Ile 2215	Tyr His Cys Ala Ala 2220	Gln Val Asp
Phe Leu 2225	His Pro Tyr Ala Ser 2230	Leu Lys Pro Ala Asn 2235	Val Asp Ser
Val Val 2240	Thr Leu Leu Glu Trp 2245	Thr Ala Gln Gly Arg 2250	Ala Lys Ser
Met His 2255	Tyr Val Ser Thr Leu 2260	Ala Val Ile Asp Gln 2265	Asn Asn Lys
Glu Asp 2270	Thr Ile Thr Glu Gln 2275	Ser Ala Leu Ala Ser 2280	Trp Ser Gly
Leu Val 2285	Asp Gly Tyr Ser Gln 2290	Ser Lys Trp Val Gly 2295	Asp Ala Leu
Ala Arg 2300	Glu Ala Gln Ala Arg 2305	Gly Met Pro Val Ala 2310	Ile Tyr Arg
Leu Gly 2315	Ala Val Thr Gly Asp 2320	His Thr His Ala Ile 2325	Cys Asn Ala
Asp Asp	Leu Ile Trp Arg Val	Ala His Leu Tyr Ala	Asp Leu Glu

ES 2 809 727 T3

2330	2335	2340
Ala Ile Pro Asp Met Asp Leu Pro Leu Asn Leu Thr Pro Val Asp 2345 2350 2355		
Asp Val Ala Arg Ala Ile Leu Gly Leu Ala Ala Gln Glu Ala Ser 2360 2365 2370		
Trp Gly Gln Val Phe His Leu Met Ser Gln Ala Ala Leu Arg Val 2375 2380 2385		
Arg Asp Ile Pro His Val Phe Glu Arg Met Gly Met Arg Leu Glu 2390 2395 2400		
Pro Val Gly Leu Glu Pro Trp Leu Gln Arg Ala His Ala Arg Leu 2405 2410 2415		
Ala Val Ala His Asp Arg Asp Leu Ala Ala Val Leu Ala Ile Leu 2420 2425 2430		
Asp Arg Tyr Asp Thr Thr Ala Thr Pro Pro Gln Val Ser Gly Ala 2435 2440 2445		
Ala Thr His Ala Gln Leu Glu Ala Ile Gly Ala Pro Ile Arg Pro 2450 2455 2460		
Val Asp Arg Asp Leu Leu Gln Arg Tyr Phe Val Asp Leu Gly Ile 2465 2470 2475		
Asp Thr Lys Ala Arg Arg Ala Leu Glu Thr Thr Thr Ser 2480 2485 2490		

<210> 7

<211> 1320

5 <212> ADN

<213> Variovorax paradoxus

<400> 7

atggcacgct atctcatcgc agcaaccgcc ttgccgggac acgtcctgcc gatgctggcc	60
atcgcgacgc atctggtgaa ccaggggcac gaggtgcggg tgcacaccgc gagccagttc	120
agggcgacgg ccgaggcgac cgggtgcgggc ttcacgccct tcgagcgcac gatcgacttc	180
gactaccgcy acctggacaa gcgctttccc gagcgccagc gcatcgccctc ggcgcatgcy	240
cagctgtgct tcggcctgaa gcacttcttt gccgatgcga tggccgcgca gcatgcgggc	300
ctgcaatcga tcctcgaaga cttcgaggcc gatgccatcg tggctcgacac gatgttctgc	360
ggcactttcc cgctgctgct aggcaaggag cgcgaagacc gcccggccat cgtcggcatc	420

ggcatctcgg cgctgccgct ctcgagctgc gacaccgcct tcttcggcac cgcgctgccg 480
ccgtcgtcca cgcggaagg gcgggtgcgc aacaaggcga tgaacgcaa cctcaaacag 540
gcgatgttcg gcgaggtgca acgctacttc gacacgctgc tcgcgcgttc gggcctggcc 600
gcgctgcccg atttcttcgt cgatgcgatg gtgaagctgc ccgatcttta cctgcagctc 660
accgcgcctt cgttcgaata cccgcgcagc gacctgcccg cgtcggtgca tttcgtcggc 720
ccgctgctct cgcgcgcgag ccgcgacttc acgcgcgccg agtgggtggca cgagctggac 780
gacggccgct cggtcgtgct ggtcacgcag ggcacgctgg ccaaccagaa tccgtcgcag 840
ctgatcggcc cgacgctgca ggcgctggcc ggcgacaaga acatcctcgt catcgccacc 900
accggcggcc cggtgccgcc cgccctgacg gtgaacctgc ccgccaacgc ccgcgtggtg 960
ccgttcctgc cctacgaccg gctgctgccc aagctgcacg cgatggtcac caacggcggc 1020
tacggctcgg tcaaccatgc attgagctc ggtgtgccgc tgggtgggtggc cggcacctcc 1080
gaagagaagc ccgagatcgc cgcgcgcgtg gcctggctcg gcgcgggcat caacctcgcc 1140
accggccagc cgaccgcgcg ccaggtcggc gacgcggtgc gcaaggtact gggcaactcg 1200
acctatcgcc agcgtgcggc ggtgctgcgt gaggacttcg cttgccatcg cgcgctgacc 1260
ggcatcgccg gcgccctcga ggcacttctg caaaccttcg catccgcgga aatggcttga 1320

<210> 8

<211> 439

5 <212> PRT

<213> Variovorax paradoxus

<400> 8

Met Ala Arg Tyr Leu Ile Ala Ala Thr Ala Leu Pro Gly His Val Leu
1 5 10 15

Pro Met Leu Ala Ile Ala Gln His Leu Val Asn Gln Gly His Glu Val
20 25 30

Arg Val His Thr Ala Ser Gln Phe Arg Ala Gln Ala Glu Ala Thr Gly
35 40 45

Ala Gly Phe Thr Pro Phe Glu Arg Thr Ile Asp Phe Asp Tyr Arg Asp
50 55 60

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

Gln Leu Cys Phe Gly Leu Lys His Phe Phe Ala Asp Ala Met Ala Ala
85 90 95

Gln His Ala Gly Leu Gln Ser Ile Leu Glu Asp Phe Glu Ala Asp Ala

ES 2 809 727 T3

100					105					110				
Ile	Val	Val	Asp	Thr	Met	Phe	Cys	Gly	Thr	Phe	Pro	Leu	Leu	Gly
		115					120					125		
Lys	Glu	Arg	Glu	Asp	Arg	Pro	Ala	Ile	Val	Gly	Ile	Gly	Ile	Ala
	130					135					140			
Leu	Pro	Leu	Ser	Ser	Cys	Asp	Thr	Ala	Phe	Phe	Gly	Thr	Ala	Pro
145					150					155				160
Pro	Ser	Ser	Thr	Pro	Glu	Gly	Arg	Val	Arg	Asn	Lys	Ala	Met	Ala
				165					170					175
Asn	Leu	Lys	Gln	Ala	Met	Phe	Gly	Glu	Val	Gln	Arg	Tyr	Phe	Thr
			180					185					190	
Leu	Leu	Ala	Arg	Ser	Gly	Leu	Ala	Ala	Leu	Pro	Asp	Phe	Phe	Asp
		195					200					205		
Ala	Met	Val	Lys	Leu	Pro	Asp	Leu	Tyr	Leu	Gln	Leu	Thr	Ala	Ser
	210					215					220			
Phe	Glu	Tyr	Pro	Arg	Ser	Asp	Leu	Pro	Ala	Ser	Val	His	Phe	Gly
225					230					235				240
Pro	Leu	Leu	Ser	Pro	Ala	Ser	Arg	Asp	Phe	Thr	Pro	Pro	Glu	Trp
				245					250					255
His	Glu	Leu	Asp	Asp	Gly	Arg	Ser	Val	Val	Leu	Val	Thr	Gln	Thr
			260					265					270	
Leu	Ala	Asn	Gln	Asn	Pro	Ser	Gln	Leu	Ile	Gly	Pro	Thr	Leu	Ala
		275					280					285		
Leu	Ala	Gly	Asp	Lys	Asn	Ile	Leu	Val	Ile	Ala	Thr	Thr	Gly	Pro
	290					295					300			
Val	Pro	Pro	Ala	Leu	Thr	Val	Asn	Leu	Pro	Ala	Asn	Ala	Arg	Val
305					310					315				320
Pro	Phe	Leu	Pro	Tyr	Asp	Arg	Leu	Leu	Pro	Lys	Leu	His	Ala	Val
				325					330					335
Thr	Asn	Gly	Gly	Tyr	Gly	Ser	Val	Asn	His	Ala	Leu	Ser	Leu	Val
			340					345					350	

ES 2 809 727 T3

Pro Leu Val Val Ala Gly Thr Ser Glu Glu Lys Pro Glu Ile Ala Ala
355 360 365

Arg Val Ala Trp Ser Gly Ala Gly Ile Asn Leu Ala Thr Gly Gln Pro
370 375 380

Thr Ala Arg Gln Val Gly Asp Ala Val Arg Lys Val Leu Gly Asn Ser
385 390 395 400

Thr Tyr Arg Gln Arg Ala Ala Val Leu Arg Glu Asp Phe Ala Cys His
405 410 415

Arg Ala Leu Thr Gly Ile Ala Gly Ala Leu Glu Ala Leu Leu Gln Thr
420 425 430

Phe Ala Ser Ala Glu Met Ala
435

<210> 9

<211> 213

5 <212> ADN

<213> Variovorax paradoxus

<400> 9

atgagcaacc	cgttcgacga	caagaacgcc	agcttcacagg	tgctgggtgaa	cgacgagggc	60
cagcactcgc	tgtggcccg	cttcacgcgc	gtgcccgcgc	gctggcaggt	ggcgctggcg	120
ccgaccgacc	gcgacgcctg	cagcgcctac	atcgcggcga	actggcagga	catgcgccc	180
cgttcgctgg	tggtggccac	ggcgcccggc	tga			213

10

<210> 10

<211> 70

<212> PRT

<213> Variovorax paradoxus

15

<400> 10

ES 2 809 727 T3

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

Asn Asp Glu Gly Gln His Ser Leu Trp Pro Ala Phe Ile Ala Val Pro
20 25 30

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

Ala Tyr Ile Ala Ala Asn Trp Gln Asp Met Arg Pro Arg Ser Leu Val
50 55 60

Val Ala Thr Ala Ala Gly

65 70

5 <210> 11

<211> 969

<212> ADN

<213> Variovorax paradoxus

10 <400> 11

```

atgtccttcc cgttcggtgc cgtcgtcgtc acctatttcc cgaccggcga gcaagtggcg      60
aacctccatt cgctggcggc ctctgttcgc cacctctgcg tggtcgacaa cacgccgcag      120
gtgggcgatt ggcatgcggc gctcgtcgat gcgggcgttt cgggtgctgca caacggcaac      180
cgcgggcgga tcgcggggcg cttcaaccgc ggcacatcgc acctcgaagc gcggggcgcc      240
gaactcttct tcctgctcga ccaggattcg aagctgccac ccggctactt cgatgccatg      300
tgcgaggctg cgatgggtgg ccgggagcgg aaggcgagg gcaatggtga ggaagacgcg      360
gccttcctga tcggcccgcg cgtccacgac acgaacctgg acgcgctgat cccgcaattc      420
ggcctccagg gcaaacgcgt ctaccagttc gacctgcggc agcccttcac cgagccgctg      480
atgcgctgcg ccttcatgat ttctcgggc tccctgattt cgcgcggcgc ctgggcccgg      540
atcggccggt tcgacgagcg ctatgtgatc gaccacgtgg acaccgacta ctgcatgcgt      600
gccctgggtc gcggcgtgcc gctctacctg aatccgcacg tcgtgctgcg gcaccagatt      660
ggcgacatcc gtgcccggtc gctgttcggc tggaagatcc acttcatcaa ctaccgggc      720
gcgcggcgct actacatcgc gcgcaatgcc atcgatctct cgcgggcgca tgtgcgcgcc      780
tttcccgcga tcctgttcat caacgtttac acgctcaagc agatcctgcc gatgctgatg      840
ttcgagcgcg accgcttcaa gaagaccatc gcgctgatgc tcggctgctt cgatggcctg      900
ttcgggcggc tcgggggcct cggcgaggtg catccgcgga tgggcaaata cctgggcccgc      960
agcgattga                                     969

```

<210> 12

<211> 322

<212> PRT

<213> Variovorax paradoxus

5 <400> 12

Met	Ser	Phe	Pro	Phe	Gly	Ala	Val	Val	Val	Thr	Tyr	Phe	Pro	Thr	Gly
1				5					10					15	

Glu	Gln	Val	Ala	Asn	Leu	His	Ser	Leu	Ala	Ala	Ser	Cys	Pro	His	Leu
			20					25					30		

Cys	Val	Val	Asp	Asn	Thr	Pro	Gln	Val	Gly	Asp	Trp	His	Ala	Ala	Leu
		35					40					45			

ES 2 809 727 T3

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

<211> 1260

<212> ADN

<213> Variovorax paradoxus

5 <400> 13

```

ttgaccgcca cccttcacgc gccgcgcgta cgcgcgcgcg cgctcgcctt catcttcgtc      60
acggtgctga tcgaacttcat ggcgttcggc ctgatactgc ccggcctgcc gcacctggtg      120
gagcggctgg ccggcggcag caccgtaacg gcggcgctact ggatcgctgt gttcggcacc      180
gcgttcgcgg cgatccagtt cgtgagctcg ccgatccagg gcgcgctgtc cgaccgcttc      240
gggcggcggc cggtgatcct gctgtcgtgc ttcggcctcg gcgtggattt cgtgttcattg      300
gccctggccg acagcctgcc gtggctgttc gtcggccggg tggctctccg cgtgttctcg      360
gccagcttca ccatacgcaa tgcctacatc gccgatgtga cgctgccgga ggagcgcgcc      420
cgcagctacg gcatacgtgg ggccgcgttc ggcattgggc tgggtgttcgg gccggtgctc      480
ggcgggcaac tgagccacat cgatccgcgc ctgccgttct ggttcgcggc cggcttgacg      540
ctgctcagct tctgctacgg atgggtcgtg ttgccgaat cgctgccgcc cgagcggcgt      600
gcccgcaagt tcgaactggtc gcatacgcaat ccggttggga cgctggtgct gctcaagcgc      660
tatccgcagg tgttcggact ggcggcggtg atcttctctg tgaacctggc tcagtacgtc      720
tatcccagcg tgttcgtgct gttcggccgc taccggtatc actggaagga agacgccgtg      780
ggctgggtgc tcggcgcggt ggcgtgctc agcgtgctgg tcaatgcgct gttgatcggg      840
ccgggcgtga agcgttctcg cgagcgcgcg gccctgttgc tcggcatggg cttcggcgtg      900
ctcggcttcg tcatcatcgg gtttgccgac gctggatgga tcctcctggc cggggtgccg      960
ttcggcattc tgctggcggt cgccggaccg gcggcgcagg cgctggtcac gctgcaggtc     1020
ggcaccgccg agcagggcgg catccagggg gcgctcacca gcctgggtgc ggtggcgggc     1080
atcgtcgggc cggcgatggt cgccggcagc ttcggttact tcatcggcgc ggacgcgcgc     1140
gtgcacttgc cgggcgcgcc gtttttcctc gctgcggcgt tcctctgcat cggcacgctg     1200
atcgcgtggc gctacgcaca gccgaagccc gcgacggcag cggtgcccga gccgacctga     1260

```

<210> 14

<211> 419

10 <212> PRT

<213> Variovorax paradoxus

<400> 14

ES 2 809 727 T3

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

ES 2 809 727 T3

245								250				255				
Glu	Asp	Ala	Val	Gly	Trp	Val	Leu	Gly	Ala	Val	Gly	Val	Leu	Ser	Val	
			260					265					270			
Leu	Val	Asn	Ala	Leu	Leu	Ile	Gly	Pro	Gly	Val	Lys	Arg	Phe	Gly	Glu	
			275					280					285			
Arg	Arg	Ala	Leu	Leu	Leu	Gly	Met	Gly	Phe	Gly	Val	Leu	Gly	Phe	Val	
			290					295					300			
Ile	Ile	Gly	Phe	Ala	Asp	Ala	Gly	Trp	Ile	Leu	Leu	Ala	Gly	Val	Pro	
305					310					315					320	
Phe	Gly	Ile	Leu	Leu	Ala	Phe	Ala	Gly	Pro	Ala	Ala	Gln	Ala	Leu	Val	
				325					330					335		
Thr	Leu	Gln	Val	Gly	Thr	Ala	Glu	Gln	Gly	Arg	Ile	Gln	Gly	Ala	Leu	
			340					345					350			
Thr	Ser	Leu	Val	Ser	Val	Ala	Gly	Ile	Val	Gly	Pro	Ala	Met	Phe	Ala	
			355					360					365			
Gly	Ser	Phe	Gly	Tyr	Phe	Ile	Gly	Ala	Asp	Ala	Pro	Val	His	Leu	Pro	
			370					375					380			
Gly	Ala	Pro	Phe	Phe	Leu	Ala	Ala	Ala	Phe	Leu	Cys	Ile	Gly	Thr	Leu	
385					390					395					400	
Ile	Ala	Trp	Arg	Tyr	Ala	Gln	Pro	Lys	Pro	Ala	Thr	Ala	Ala	Val	Pro	
				405					410					415		

Glu Pro Thr

<210> 15

$\langle 211 \rangle$ 1080

5 <212> ADN

<213> Variovorax paradoxus

<400> 15

atgatacctgg	taaccggcgg	cgcaggcttc	attggcgcca	atttcgtact	cgactggctc	60
gcacagagcg	atgaaccggt	cgtgaaccta	gacaagctga	cctacgcggg	caacctcgag	120
acgctcgcac	cgctcaagga	caaccogaag	cacatcttcg	tgcagggcga	catcggcgac	180
agcgcgctgc	tcgaccgcct	gctggccgag	cacaagccgc	gtgccgtggg	caacttcgcg	240
gccgaatcgc	acgtcgaccg	ctcgatccac	ggccccgaag	acttcgtgca	gaccaacgtg	300

ES 2 809 727 T3

ctgggcacct tccgcctgct cgaatccgtg cgcggtttct ggaatgccct gccggccgac 360
cagaaggccg ccttccgctt cctgcatgtg tgcaccgacg aggtctacgg ctgcgtctcc 420
aagaccgacc cggccttcac cgaagagaac aagtacgagc ccaacagccc gtactcggcc 480
agcaaggccg ccagcgacca cctcgtgcgc gcctggcacc acacctacgg cctgccggtg 540
gtcaccacca actgctcgaa caactacggg ccgttccact tccccgagaa gctcattccc 600
ctgatgatcg tcaacgcgct ggcgggcaag ccgctgcccg tgtacggcga cggcatgcag 660
gtgcgcgact ggctctacgt gaaggaccac tgcagcgcca tccgccgctg gctcgaagcc 720
ggcaagctcg gcgagaccta caacgtgggc ggctggaacg agaagcccaa catcgagatc 780
gtcaacaccg tctgcgcgct gctcgacgag ctgagcccca aggccggcgg caagccgtac 840
aaggaacaga tcacctatgt gaccgaccgc cccggccacg accgccgcta cgcgatcgac 900
gcacgcaagc tcgagcgcga actcggctgg aaacctgccg agaccttcga cagcggcatc 960
cgcaagacgg tcgagtggta cctcgcgaac ggcgagtggg tgcgcaacgt gcaaagcggc 1020
gcgtaccgcg agtgggtcga gaagcaatac gacgccgcac cggcgaaggc caccgcatga 1080

<210> 16

<211> 359

5 <212> PRT

<213> Variovorax paradoxus

<400> 16

Met Ile Leu Val Thr Gly Gly Ala Gly Phe Ile Gly Ala Asn Phe Val
1 5 10 15

Leu Asp Trp Leu Ala Gln Ser Asp Glu Pro Val Val Asn Leu Asp Lys
20 25 30

Leu Thr Tyr Ala Gly Asn Leu Glu Thr Leu Ala Ser Leu Lys Asp Asn
35 40 45

Pro Lys His Ile Phe Val Gln Gly Asp Ile Gly Asp Ser Ala Leu Leu
50 55 60

Asp Arg Leu Leu Ala Glu His Lys Pro Arg Ala Val Val Asn Phe Ala
65 70 75 80

Ala Glu Ser His Val Asp Arg Ser Ile His Gly Pro Glu Asp Phe Val
85 90 95

Gln Thr Asn Val Leu Gly Thr Phe Arg Leu Leu Glu Ser Val Arg Gly
100 105 110

Phe Trp Asn Ala Leu Pro Ala Asp Gln Lys Ala Ala Phe Arg Phe Leu

ES 2 809 727 T3

115		120		125
His Val Ser Thr Asp Glu Val Tyr Gly Ser Leu Ser Lys Thr Asp Pro				
130		135		140
Ala Phe Thr Glu Glu Asn Lys Tyr Glu Pro Asn Ser Pro Tyr Ser Ala				
145		150		155
Ser Lys Ala Ala Ser Asp His Leu Val Arg Ala Trp His His Thr Tyr				
		165		170
Gly Leu Pro Val Val Thr Thr Asn Cys Ser Asn Asn Tyr Gly Pro Phe				
		180		185
His Phe Pro Glu Lys Leu Ile Pro Leu Met Ile Val Asn Ala Leu Ala				
		195		200
Gly Lys Pro Leu Pro Val Tyr Gly Asp Gly Met Gln Val Arg Asp Trp				
		210		215
Leu Tyr Val Lys Asp His Cys Ser Ala Ile Arg Arg Val Leu Glu Ala				
		225		230
Gly Lys Leu Gly Glu Thr Tyr Asn Val Gly Gly Trp Asn Glu Lys Pro				
		245		250
Asn Ile Glu Ile Val Asn Thr Val Cys Ala Leu Leu Asp Glu Leu Ser				
		260		265
Pro Lys Ala Gly Gly Lys Pro Tyr Lys Glu Gln Ile Thr Tyr Val Thr				
		275		280
Asp Arg Pro Gly His Asp Arg Arg Tyr Ala Ile Asp Ala Arg Lys Leu				
		290		295
Glu Arg Glu Leu Gly Trp Lys Pro Ala Glu Thr Phe Asp Ser Gly Ile				
		305		310
Arg Lys Thr Val Glu Trp Tyr Leu Ala Asn Gly Glu Trp Val Arg Asn				
		325		330
Val Gln Ser Gly Ala Tyr Arg Glu Trp Val Glu Lys Gln Tyr Asp Ala				
		340		345
Ala Pro Ala Lys Ala Thr Ala				
		355		

<210> 17

<211> 891

5 <212> ADN

<213> Variororax paradoxus

<400> 17

ES 2 809 727 T3

```

atgaagctgc tgctgctggg caagggcgga caggtcggct gggagctgca acgcagcctc   60
gcgccccctgg gcgaactggt ggcgctcgat ttcgacagca ccgacttcaa cgccgacttc   120
agtcgccccg agcagctggc cgagacagtg ctgaagggtgc gccccgacgt catcgtcaat   180
gccgcagcgc acaccgcggt cgacaaggcc gagagcgagc ccgagttcgc gcgcaagctc   240
aacgccacct cgcccgccgt ggtggccgaa gccgcgcagc agatcggcgc gctgatggtt   300
cactactcga ccgactacgt cttcgacggc agcggcagca agccgtggaa agaagacgat   360
gcgaccggcc cgctcagcgt ctacggcagc accaagctcg aaggcgagca actggtggca   420
aagcactgtg cgaagcacct gatctttcgc accagctggg tctatgccgc gcgcggcggc   480
aacttcgcca agaccatgct gcgcatcgcc aaggagcgcg acaagctgac cgtcatcgac   540
gaccagttcg gcgcgcccac cggcgcgga ctgctggccg acatcaccgc gcacgcgatt   600
cgcgcgacgc tgcaggacct gtccaaggcc gggctctatc acgcggtggc cgggtggcgtg   660
accacgtggc acggetatgc gcgcttcgtg atcgagcagg ccaaggcggc gggcgtggaa   720
ctgaaggccg gccccgaagc ggtcgagccc gtgcccacca cggcattccc gacgccggcc   780
aggcgccgcg acaactcgcg cctggacacc accaagctgc aatcgacctt cggcctcgtg   840
ctgcccagat ggcagtcggt cgtcgcccgc atgttgcgcg aaaccttctg a           891

```

<210> 18

<211> 296

5 <212> PRT

<213> Variororax paradoxus

<400> 18

```

Met Lys Leu Leu Leu Leu Gly Lys Gly Gly Gln Val Gly Trp Glu Leu
1           5           10          15

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

Ser Thr Asp Phe Asn Ala Asp Phe Ser Arg Pro Glu Gln Leu Ala Glu
          35          40          45

Thr Val Leu Lys Val Arg Pro Asp Val Ile Val Asn Ala Ala Ala His
          50          55          60

Thr Ala Val Asp Lys Ala Glu Ser Glu Pro Glu Phe Ala Arg Lys Leu
65          70          75          80

```

10

ES 2 809 727 T3

Asn Ala Thr Ser Pro Gly Val Val Ala Glu Ala Ala Gln Gln Ile Gly
85 90 95

Ala Leu Met Val His Tyr Ser Thr Asp Tyr Val Phe Asp Gly Ser Gly
100 105 110

Ser Lys Pro Trp Lys Glu Asp Asp Ala Thr Gly Pro Leu Ser Val Tyr
115 120 125

Gly Ser Thr Lys Leu Glu Gly Glu Gln Leu Val Ala Lys His Cys Ala
130 135 140

Lys His Leu Ile Phe Arg Thr Ser Trp Val Tyr Ala Ala Arg Gly Gly
145 150 155 160

Asn Phe Ala Lys Thr Met Leu Arg Ile Ala Lys Glu Arg Asp Lys Leu
165 170 175

Thr Val Ile Asp Asp Gln Phe Gly Ala Pro Thr Gly Ala Glu Leu Leu
180 185 190

Ala Asp Ile Thr Ala His Ala Ile Arg Ala Thr Leu Gln Asp Pro Ser
195 200 205

Lys Ala Gly Leu Tyr His Ala Val Ala Gly Gly Val Thr Thr Trp His
210 215 220

Gly Tyr Ala Arg Phe Val Ile Glu Gln Ala Lys Ala Ala Gly Val Glu
225 230 235 240

Leu Lys Ala Gly Pro Glu Ala Val Glu Pro Val Pro Thr Thr Ala Phe
245 250 255

Pro Thr Pro Ala Arg Arg Pro His Asn Ser Arg Leu Asp Thr Thr Lys
260 265 270

Leu Gln Ser Thr Phe Gly Leu Val Leu Pro Glu Trp Gln Ser Gly Val
275 280 285

Ala Arg Met Leu Arg Glu Thr Phe
290 295

<210> 19

<211> 897

5 <212> ADN

<213> Variovorax paradoxus

<400> 19

ES 2 809 727 T3

```

atgaccaaga cgacgcaacg caaaggcatc atcctcgccg gtggctcggg caccgcctg      60
caccgcgga cgcttgccat gagcaaaca ctgctgccg tgtacgaca gccgatgatc      120
tattaccgcg tgagcacgct gatgctgggc ggcagcgcg acatcctgat catcagcacg      180
ccgcaggaca cgccgcgttt ccagcaactg ctgggggatg gcagccaatg gggcatcaac      240
ctgcagtacg cggcgcagcc gagcccgat ggtctggcg aggcgttcac catcggtgac      300
aagttcgtgg gcaacgaccc gagtgcgctg gtgctggggg acaacatctt ctatggccac      360
gacttcgccc atctgctggc cgatgccgac gccaaagacct cgggtgcgac ggtgttcgcc      420
taccacgtgc acgacccga gcgctacggc gtggtggcct tcgatgcaa gggcagggcg      480
agcagcatcg aagaaaagcc gctcaagccc aagagcagct atgcggtcac gggcctctac      540
ttctacgaca accaggtcgt cgacatcgcc aaggccgtga agccgagcgc gcgcggcgaa      600
ctcgagatca ccgcgggtcaa ccaggcgtat ctgcacctcg accagctgaa cgtgcagatc      660
atgcagcgcg gctatgctg gctcgatacc ggtacgcacg acagcctgct ggaagccggg      720
cagttcattg ccacgctcga gcaccgccag gggctgaaga tcgcatgccc cgaagagatc      780
gcatggcgca atggcttcac ctcaaccgag caactcgaaa agctcgcggc gccgctggaa      840
aagagcggct acggcaagta cctcaagcac ctgctgaacg acgaggtgcg ctcgtga      897

```

<210> 20

<211> 298

5 <212> PRT

<213> Variovorax paradoxus

<400> 20

```

Met Thr Lys Thr Thr Gln Arg Lys Gly Ile Ile Leu Ala Gly Gly Ser
1          5          10          15

```

```

Gly Thr Arg Leu His Pro Ala Thr Leu Ala Met Ser Lys Gln Leu Leu
20          25          30

```

```

Pro Val Tyr Asp Lys Pro Met Ile Tyr Tyr Pro Leu Ser Thr Leu Met
35          40          45

```

```

Leu Gly Gly Met Arg Asp Ile Leu Ile Ile Ser Thr Pro Gln Asp Thr
50          55          60

```

```

Pro Arg Phe Gln Gln Leu Leu Gly Asp Gly Ser Gln Trp Gly Ile Asn
65          70          75          80

```

```

Leu Gln Tyr Ala Val Gln Pro Ser Pro Asp Gly Leu Ala Gln Ala Phe
85          90          95

```

```

Ile Ile Gly Asp Lys Phe Val Gly Asn Asp Pro Ser Ala Leu Val Leu

```

ES 2 809 727 T3

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

<210> 21

<211> 546

5 <212> ADN

<213> Variovorax paradoxus

<400> 21

gtgaaggcca cgccacctc gattcctgac gtgctcgtga tcgagccgaa ggtgtttggc	60
gatgcacggg gcttcttctt cgaaagcttc aaccagaagg ccttcgacga agcgatcggc	120
aagcatgtcg acttcgtgca ggacaacccat tcgcgatcgg ccaaggtgt gctgcggggg	180

ES 2 809 727 T3

ctgcattacc aggtccagca gccgcaaggc aagctcgtgc ggggtggtgcg tgggtgcggtg 240
 ttcgacgtgg ccgtcgacat ccgcaagtcg tcgccgactt ttggcaaagt ggtgggtgtc 300
 gagttgaacg aagacaacca caagcagctc tgggtgccgg caggattcgc gcacggtttc 360
 ctggtgttga gcgagaccgc ggaattcctc tacaagacca ccgactacta cgcgcccgcc 420
 cacgagcgcg cgattgtctg gaacgacccc gctgtcggta ttcgatggcc ggatgtggga 480
 ggggcaccgg tcctgtcgaa gaaggacgaa gacgggtgtc ttctgcaagc ggcagaggtt 540
 ttctag 546

<210> 22

<211> 181

5 <212> PRT

<213> Variovorax paradoxus

<400> 22

Met Lys Ala Thr Pro Thr Ser Ile Pro Asp Val Leu Val Ile Glu Pro
 1 5 10 15

Lys Val Phe Gly Asp Ala Arg Gly Phe Phe Phe Glu Ser Phe Asn Gln
 20 25 30

Lys Ala Phe Asp Glu Ala Ile Gly Lys His Val Asp Phe Val Gln Asp
 35 40 45

Asn His Ser Arg Ser Ala Lys Gly Val Leu Arg Gly Leu His Tyr Gln
 50 55 60

Val Gln Gln Pro Gln Gly Lys Leu Val Arg Val Val Arg Gly Ala Val
 65 70 75 80

Phe Asp Val Ala Val Asp Ile Arg Lys Ser Ser Pro Thr Phe Gly Lys
 85 90 95

Trp Val Gly Val Glu Leu Asn Glu Asp Asn His Lys Gln Leu Trp Val
 100 105 110

Pro Ala Gly Phe Ala His Gly Phe Leu Val Leu Ser Glu Thr Ala Glu
 115 120 125

Phe Leu Tyr Lys Thr Thr Asp Tyr Tyr Ala Pro Ala His Glu Arg Ala
 130 135 140

Ile Val Trp Asn Asp Pro Ala Val Gly Ile Arg Trp Pro Asp Val Gly
 145 150 155 160

Gly Ala Pro Val Leu Ser Lys Lys Asp Glu Asp Gly Cys Leu Leu Gln

ES 2 809 727 T3

165

170

175

Ala Ala Glu Val Phe
180

<210> 23

<211> 1029

5 <212> ADN

<213> Variovorax paradoxus

<400> 23

ATGGGCAGCA	GCCATCATCA	TCATCATCAC	AGCAGCGGCC	TGGTGCCGCG	CGGCAGCCAT	60
ATGTCCTTCC	CGTTCGGTGC	CGTCGTCGTC	ACCTATTTCC	CGACCGGCGA	GCAAGTGGCG	120
AACCTCCATT	CGCTGGCGGC	CTCGTGTCCG	CACCTCTGCG	TGGTCGACAA	CACGCCGCAG	180
GTGGGCGATT	GGCATGCGGC	GCTCGTCGAT	GCGGGCGTTT	CGGTGCTGCA	CAACGGCAAC	240
CGCGGCGGCA	TCGCGGGCGC	CTTCAACCGC	GGCATCATCG	ACCTCGAAGC	GCGGGGCGCC	300
GAACTCTTCT	TCCTGCTCGA	CCAGGATTCG	AAGCTGCCAC	CCGGCTACTT	CGATGCCATG	360
TGCGAGGCTG	CGATGGTGGC	CCGGGAGCGG	AAGGGCGAGG	GCAATGGTGA	GGAAGACGCG	420
GCCTTCCTGA	TCGGCCCGCT	CGTCCACGAC	ACGAACCTGG	ACGCGCTGAT	CCCGCAATTC	480
GGCCTCCAGG	GCAAACGCGT	CTACCAGTTC	GACCTGCGGC	AGCCCTTCAC	CGAGCCGCTG	540
ATGCGCTGCG	CCTTCATGAT	TTCTTCGGGC	TCCCTGATTT	CGCGCGGCGC	CTGGGCCCCG	600
ATCGGCCGGT	TCGACGAGCG	CTATGTGATC	GACCACGTGG	ACACCGACTA	CTGCATGCGT	660
GCCCTGGGTC	GCGGCGTGCC	GCTCTACCTG	AATCCGCACG	TCGTGCTGCG	GCACCAGATT	720
GGCGACATCC	GTGCCCGGTC	GCTGTTTCGGC	TGGAAGATCC	ACTTCATCAA	CTACCCGGCC	780
GCGCGGCGCT	ACTACATCGC	GCGCAATGCC	ATCGATCTCT	CGCGGGCGCA	TGTGCGCGCC	840
TTTCCCGCGA	TCCTGTTTCA	CAACGTTTAC	ACGCTCAAGC	AGATCCTGCC	GATGCTGATG	900
TTCGAGCGCG	ACCCTTTCAA	GAAGACCATC	GCGCTGATGC	TCGGCTGCTT	CGATGGCCTG	960
TTCGGGCGGC	TCGGGGGCGT	CGGCGAGGTG	CATCCGCGGA	TGGGCAAATA	CCTGGGCGCG	1020
AGCGATTGA						1029

10

<210> 24

<211> 342

<212> PRT

<213> Variovorax paradoxus

15

<400> 24

MGSSHHHHHH	SSGLVPRGSH	MSFPFGAVVV	TYFPTGEQVA	NLHSLAASCP	HLCVVDNTPQ	60
VGDWHAALVD	AGVSVLHNGN	RGGIAGAFNR	GIIDLEARGA	ELFFLLDQDS	KLPPGYFDAM	120

ES 2 809 727 T3

CEAAMVARER KEGNGEEDA AFLIGPLVHD TNLDALIPQF GLQGKRVYQF DLRQPFTEPL 180
MRCAFMISG SLISRGAWAR IGRFDERYVI DHVDTDYCMR ALGRGVPLYL NPHVVLRHQI 240
GDIRARSLFG WKIHFINYPA ARYYIARNA IDLSRAHVRA FPAILFINVY TLKQILPMLM 300
FERDRFKKTI ALMLGCFDGL FGRLGGLGEV HPRMGKYLGR SD 342

REIVINDICACIONES

1. Un biotensioactivo purificado que comprende un componente lipídico hidrófobo que comprende un extremo carboxilo y un extremo hidroxilo, en donde el componente lipídico está unido covalentemente a (i) una cadena de péptido o semejante a péptido en el extremo carboxilo del componente lipídico y (ii) un resto de carbohidrato en el extremo hidroxilo del componente lipídico a través de una unión glicosídica, en donde el resto de carbohidrato incluye uno, dos o tres residuos de ramnosa.
2. El biotensioactivo purificado según la reivindicación 1, en donde la cadena peptídica comprende en el intervalo de entre 2 y 10 aminoácidos.
3. El biotensioactivo purificado según la reivindicación 1 o reivindicación 2, en donde el componente lipídico comprende en el intervalo de entre 1 y 6 restos de ácido alcanoico.
4. El biotensioactivo purificado según cualquier reivindicación anterior, en donde el componente lipídico comprende una cadena acilo, y en donde la longitud de dicha cada cadena acilo está en el intervalo de entre C₄ a C₂₀.
5. El biotensioactivo purificado según cualquier reivindicación anterior, en donde el resto de carbohidrato puede seleccionarse de sacáridos incluyendo glucosa, fructosa, galactosa, manosa, ribosa, o variantes desoxisacáridos incluyendo desoxirribosa o fucosa.
6. El biotensioactivo purificado según cualquier reivindicación anterior, en donde el componente lipídico comprende tres restos de ácido β-hidroxi alcanoico.
7. El biotensioactivo purificado de la reivindicación 4, en donde la longitud de cada cadena acilo del componente lipídico es C₁₀.
8. El biotensioactivo purificado según cualquier reivindicación anterior, en donde el resto de carbohidrato comprende un resto de ramnosa unido al componente lipídico a través de una unión glicosídica.
9. El biotensioactivo purificado según la reivindicación 8, en donde el resto de carbohidrato comprende dos restos de ramnosa.
10. El biotensioactivo purificado según cualquier reivindicación anterior, en donde el componente lipídico comprende tres restos de ácido β-hidroxi alcanoico y la longitud de cada cadena acilo del componente lipídico es C₁₀.
11. Un método para preparar el biotensioactivo de cualquiera de las reivindicaciones 1 a 10, el método que comprende las etapas de:
 - (a) aislar un microorganismo que comprende el biotensioactivo;
 - (b) poner el microorganismo en un cultivo en condiciones que promueven la síntesis del biotensioactivo; y
 - (c) aislar el biotensioactivo del cultivo.
12. El método según la reivindicación 11, en donde el microorganismo pertenece al género y especie *Variovorax paradoxus* y es la cepa RKNM-096 según está depositada en el NRRL con el número de acceso B-67038.
13. Una composición emulsionada de aceite en agua o agua en aceite que comprende un componente polar, un componente no polar, y el biotensioactivo como se reivindica en cualquiera de las reivindicaciones 1 a 10.
14. Un microorganismo aislado preparado por ingeniería para producir el biotensioactivo de cualquiera de las reivindicaciones 1 a 10, en donde se ha introducido en el microorganismo un conjunto de genes heterólogos que presenta una similitud de al menos el 70% con las SEQ ID 3, 5, 7, 9, 11 y 13.
15. Un método para modificar tensioactivos glicolipopeptídicos naturales mediante la adición de restos de ramnosa adicionales usando RlpE expresado recombinantemente [SEQ ID 11, 12, 23 y 24].

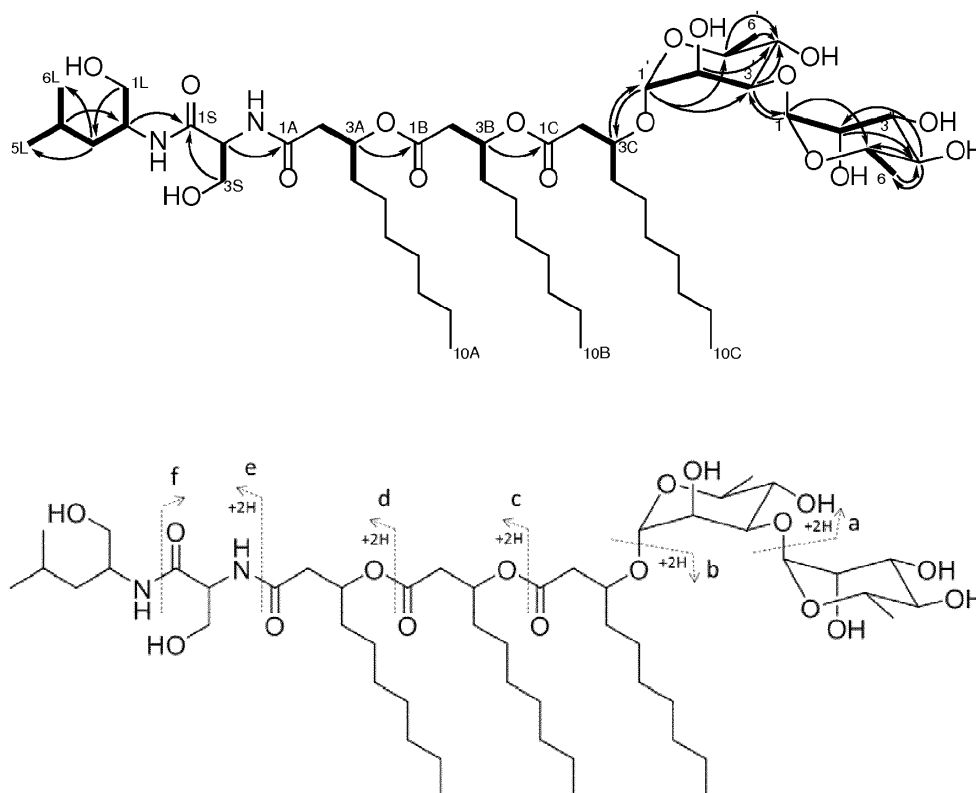


Fig. 1

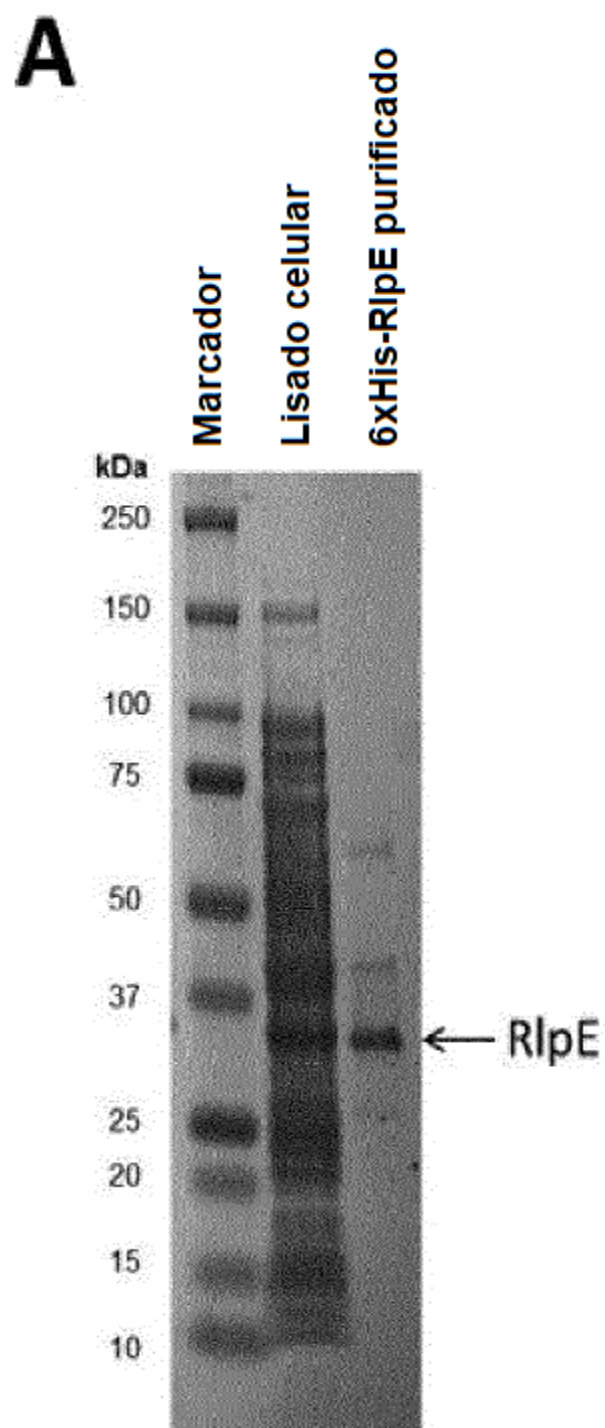


Fig. 2

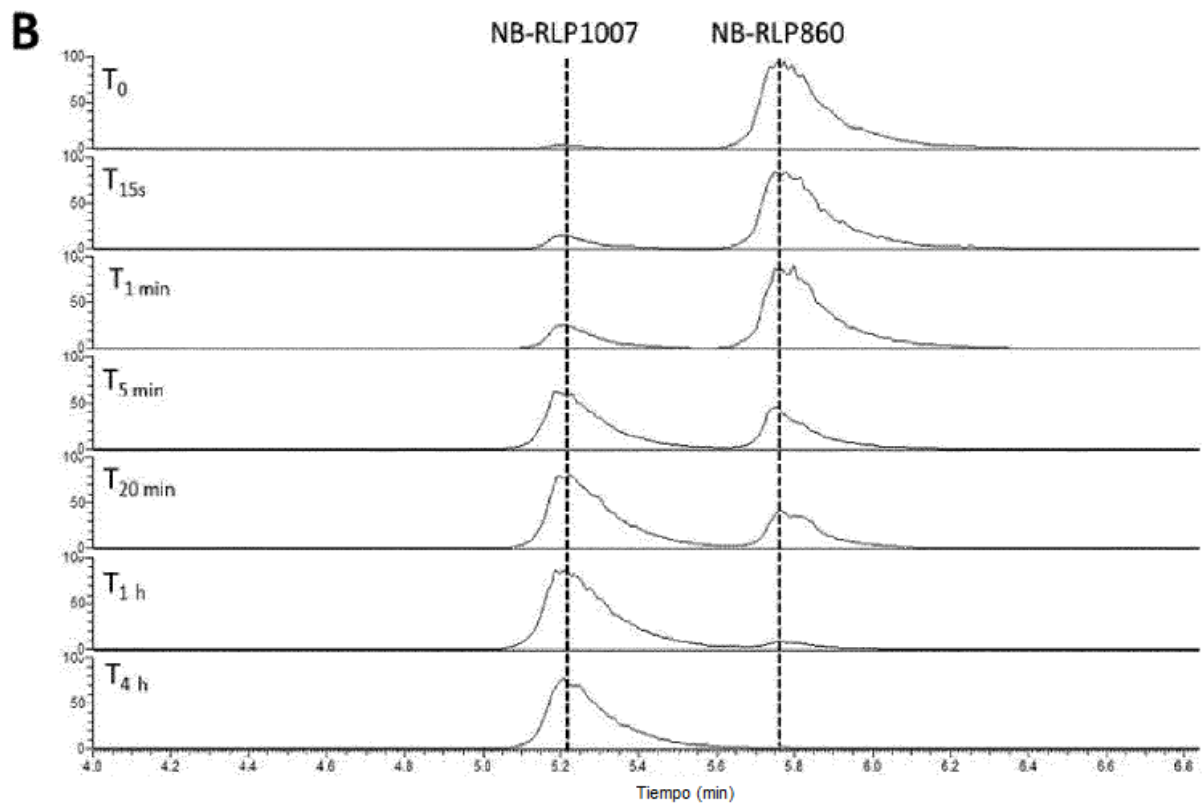
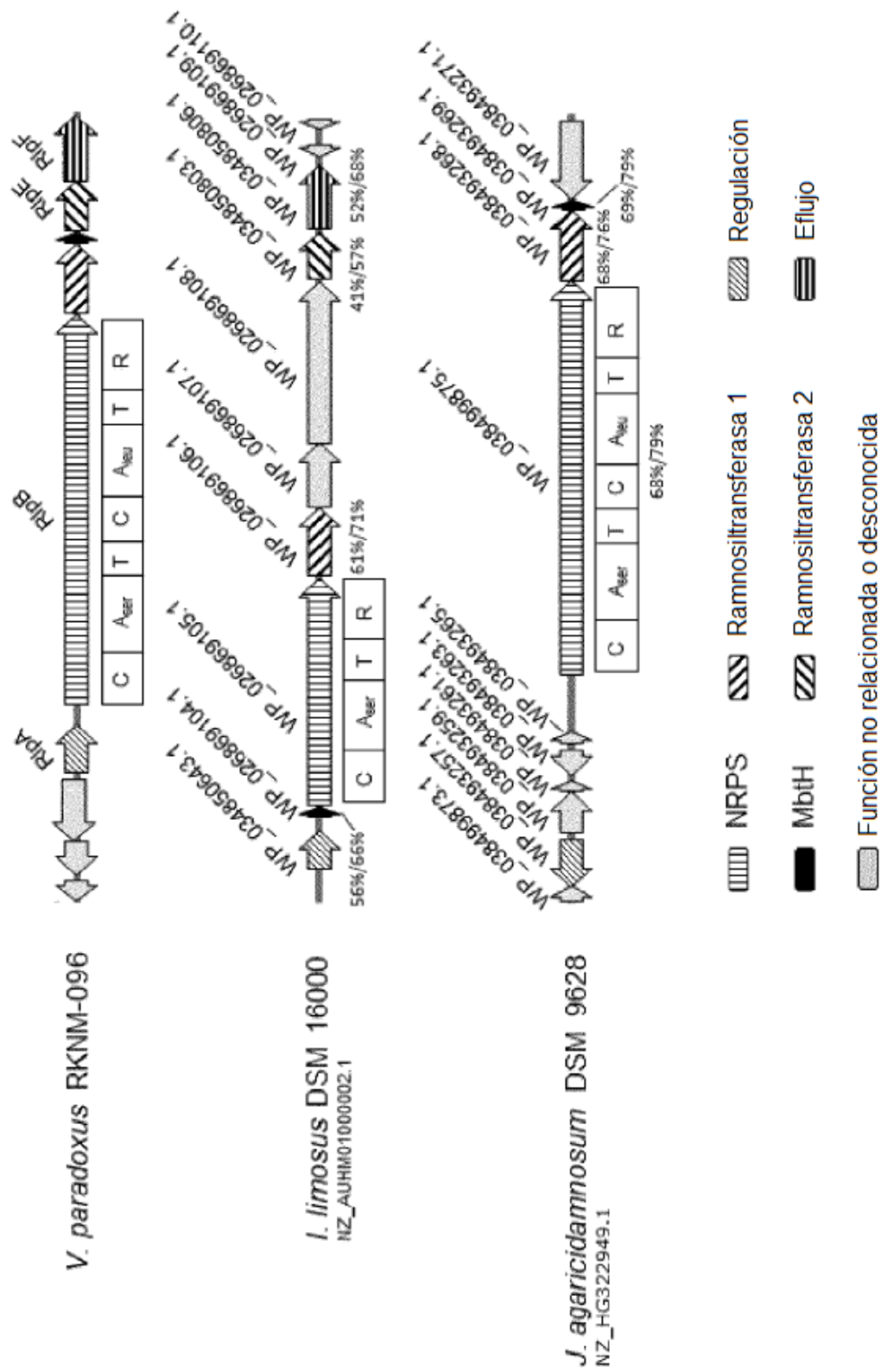


Fig. 2



[illegible]

gccagccgatctggaacacgaggatgtatgtgctcgacagcgccctgcaaccgggtgccgcccgggtgtcgccggcgag
 ctctacatcgccggcagcggtggcgcgcggtatctcaaccggccggcattgagtgccgaacgcttcatcgccga
 cccgatggcgcgcccgagccgcatgtaccgagcgggcgacctcgacgctggcgcgccgacggcacgctcgact
 tccctcgccggcgcccgaccagcaggtgaagatccggggcttccgcatcgagccggcgagatcgaatccgtgctgctc
 aagcaccggttgatcacgcaggccgctgatcgcccgaggaagctgcccgcgagaagcgctggtcgccactt
 cgtcgccggttccgagccgagccaccgagctgcgcgcccacatggcgcgaggccttgcccgaactacatggtgcctt
 cggccttcgtgcgctgcccgtcgctgcccgtcacgcaagcggaagctcgacaagaaggcgctgcccgtgcccgcac
 cagcagcccgccgctgtacgtggagccccgcacgcccagaccagagaaactgctcgccggcctctggtccgagacgct
 gcacctggagcgtgtcgcatccacgacaacttcttcgagatcgccggcgattcgctcatggcgatccagctgggca
 tgcgcatccgcccagcaggtgcccggcgacttcccgacgcccaggtctacaaccgcccagcattgcccacctggcc
 gcctggctcgacaacgaaggcgccagcgtcgaggcgctggacctgtcgcgcgagctcgacctgcccgcgacatccg
 cccgagggcactgcaccgaagctcgacccgcccgcgtgttccctaccggcgcgagcggttcgtcggcagtcacc
 tgcctggcgcgctgttgccgacacccgcccgtcgctggtctgccacgtgcgcgcccgcagcagcagggccggcgag
 cagcgctcaagcgacgctggcccagcgccagctcggtgcatctgggacaacgcgcgcatcaaggtcgtagccgg
 cgacctcggaagcgcgccctggcctcgatgacgctgcccgtgcaactggtgcgcgacggctgcgacgccatctacc
 actgcgcgcgaggtcgacttccctgcatccctacgcgagcctcaagcccggaacgtcgacagcgtggtcacgctg
 ctgcaatggacggcgagggcgcgcggaagagcatgcaactacgtctccacgctggctgtgatcgaccagaacaaca
 ggaagacaccatcaccgagcaatcgccgctggcctcatggagcggtggtcgacggctacagccagagcaagtggg
 tcggcgatgctggtggcccgagggcgagggcgcgccgcatgcccgtggcgatctaccggctggggcgagtcccgcc
 gaccacacgcacgcatctgcaatgcccagacctgatctggcgctggcgcatctctatgcccacctggaagcgat
 tcccgatattgacctgcccgtcaacctcacaccggtggacgacgtggcgcgcccatcctcgcccttgccggcgag
 aggcctcgtagggccaggtgttccacctgatgagccagggcgcgctgcccgtgcccgcacattccgcacgtcttcgag
 cgcatgggcatgcccgtggagccggtgcccgtggagccctggctgcagcgcgcgcatgcaaggctggccgtcgcgca
 tgaccgagacctggccgggtgctcgccatcctcgaccgctacgacaccacggccacgcccgcaggtgagcgccg
 cggccacgcatgcccagctcgaggccatcgccgcccgatccgcccgtggaccgagacctgctgcagcgctacttc
 gtgcacctgggcatcgacaccaaggcgccgcccgtggaaaccaccacttcataggagcacacggaatggcacgc
 tatctcatcgagcaaccgcttgcccggacacgtcctgcccgtgctggccatcgcgagcatctggtgaaccagg
 gcacgaggtgcccgtgcacaccgagccagttcagggcgagggccagggcgaccggtgcccgttcacgcccctcg
 agcgacgcatcgacttcgactaccgagacctggacaagcgcttcccgagcgccagcgcatcgcccgccgcatgcg
 cagctgtgcttcggcctgaagcacttcttggcgatgcatggccgcgagcatgcccgtgcaatcgatcctcga
 agacttcgagggcgatgccatcggtcgacacgatgttctgcccacttcccgctgctgctaggcaaggagcgcg
 aagaccgcccggccatcgctggcatcgccatctcgccgctgcccgtctcgagctgcgacaccgcttcttcggcacc
 gcgtgcccggctcgccacgcccgaaggcggtgcccgaacaaggcgatgaacgccaacctcaaacaggcgatgtt
 cggcgaggtgcaacgctacttcgacacgctgctcgccgcttcgggcctggccgctgcccgatcttcttcgtcgatg
 cgatggtgaagctgcccgatcttacctgcagctcaccgccccttcgttcgaataccgcgcgagcgacctgcccgcg
 tcggtgcatttcgtcgcccgctgctctcgcccgcgagccggaacttcacgcccggagtggtggcacgagctgga
 cgacggccgctcggtcgctggtcacgcagggcacgctggccaaccagaatccgtcgcgatgatcgcccgcagcg
 tgcaggcgctggccggcgacaagaacatcctcgatcgccaccaccggcgcccggtgcccggccgctgacgggtg
 aacctgcccggcaacgcccgcgtggtgcccgttctgcccacgaccgctgctgcccgaagctgcagcgatggtcac
 caacggcggtacggctcggtcaacctgcatgagcctcggtgtgcccgtggtggtggccggcacctccgaagaga
 agcccagatcgccgcgcgctggcctggtcgggcgccggcatcaacctcgccaccggccagccgacccgcccag
 gtggcgagcggtgcccgaaggtactgggcaactcgacctatcgccagcgctgcccgggtgctgctgaggacttcgc
 ttgccatcgcgctgacggcgatcgccggcgccctcgaggcaettctgcaaaccttcgcatccgcccgaatggctt
 gaacctgaaccccatcgacaaaggaaatcccagatgagcaaccggttcgacgacaagaacggccagcttccaggtgc
 tgggtgaacgacgagggccagcactcgctgtggcccgccttcacgcccgtgcccggcgctggcaggtggcgctggcg
 ccgaccgaccgagcgctcgagcgctacatcgccggaactggcaggacatgcccggcgcttcgctggtgggtggc
 cagggcgccggctgacggcgaggtgcttcccgctcggtgcccgtcgctcgctcacctatttcccgaccggcgagca
 agtggcgaaacctccattcgctggcgccctcggtgcccacctctgctggtcgacaacacggccgaggtggcgat
 ggcagcgccgctcgatgcccggcttcggtgctgcacaacggcaaccgcccggcgatcgccggcgcccttcaac

cgcggcatcatcgacctcgaagcgcggggcgccgaactcttcttctctgctcgaccaggattcgaagctgccaccgg
 ctacttcgatgccatgtgcgaggtgcgatgggtggcccgaggagcggaagggcgagggcaatggtgaggaagacgcgg
 ccttcctgatcgggccgctcgtccacgacacgaacctggacgcgctgatcccgcaattcggcctccagggaacgc
 gtctaccagttcgacctgcggcagcccttcaccgagccgctgatgcgctgcgccttcattgatttctcgggctccct
 gatttcgcgcggcgctggggccggatcggcgggttcgacgagcgctatgtgatcgaccagtggaacaccgactact
 gcatgcgtgccctgggtcgcggtgcgctctacctgaatccgcacgtcgtgctgcggcaccagattggcgacatc
 cgtgcccggtcgctgttcgggtggaagatccacttcattcaactaccggcgcgcgcgctactacatcgcgcgcaa
 tgccatcgatctctcgcgggcgcatgtgcgcgcctttcccgcatcctgttcattcaacgtttacacgctcaagcaga
 tectgccgatgctgatgttcgagcgcgaccgcttcaagaagaccatcgcgctgatgctcgggtgcttcgatggcctg
 ttccggcggtcgggggcctcgcgaggtgcatccgcggtgggcaaatacctgggcgcgagcgattgaccgccacc
 ctccagcgccgcgctacgcgcgcgcgctcgccttcattctcgtcacgggtgctgatcgacttcattggcggttcgg
 cctgatcctgccggcctgccgcacctgggtggagcggtggcgggcgcgagcacggtaacggcgcggtactggatcg
 ctgtgttcggcaccgcgttcgcgcgatccagttcgtgagctcgccgatccaggcgcgctgtccgaccgcttcggg
 cggcgccgggtgatcctgctgtcgtgcttcggcctcgcggtggatttcgtgttcattggcctggccgacagcctgcc
 gtggctgttcgtcgccgggtgggtctccggcggtgttcgcccagcttcaccatcgccaatgcctacatcgccgatg
 tgacgctgccggaggagcgcgcccgagctacggcatcgtggggcgcgcttcggcatgggcctgggtgttcgggccc
 gtgctcgccgggcaactgagccacatcgatccgcgcctgccgttctgggttcgcgccggcttgacgctgctcagctt
 ctgctacggatgggttcgtgttgcccgaatcgtgcgcgccgagcgcgctgcccgcaagttcgactggctcgatgcca
 atccgggtgggacgctgggtgctgctcaagcgctatccgcaggtgttcggactggcgcggtgatcttcctcgtgaac
 ctgggtcagtagctctatcccagcggttcgtgctgttcgcccactaccggatcactggaaggaagacgcggtggg
 ctgggtgctcgccgcgggtggcgctgctcagcgctgctgggtcaatgcgctgttgatcgggccggcggtgaagcgcttcg
 gcgagcgccgcgcctgttgctcgcatgggcttcggcgctgctcggttcgtcatcatcggggttgccgacgctgga
 tggatcctcctggcggggtgcccgttcggcattctgctggcggttcgcccgaaccggcgcgagcgctgggtcacgct
 gcaggtcgccaccgagcagggcgcatccagggggcgctcaccagcctgggtgctgggtggcgggcatcgtcgggc
 cggcgatgttcgcccgcagcttcgggttacttcattcgccgcggacgcgcgggtgcaattgcggggcgcgccgttttct
 ctcgctgcggcggttcctctgcatcggcacgctgatcgcggtggcgctacgcacagccgaagcccgcgacggcagcggt
 gcccgagccgacctga

Fig. 4

SEQ ID NO 2

LONGITUD: 3959

TIPO: ADN

ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 2

```

1  CCGCTGCGCC TCGCAACGGG TTGCTCCTT CGGTGCAATG CGATCCCTGC GGGTGCATG
61  GCTCTCCAGA CGGCGTTTGA TGTGATGCAG TACTGACCCC CTGTTTCGGGCG GACCTGAGC
121 GTTTATGGGA GTTTGCGCCT TCGGTAGGGC CACCGGGGTG GCCCGCTCTC CTGCAGTGGG
181 GCGATTGTAG GTGGGCACTG CCAATGCGCC AACCCTGGGA GTTTCGGCCC TTGGGCCGAT
241 GGGATAATCA TCCGTTTATT CGCCGGAGGG CGATCGTTTC ACAACAACAG GGGACCCCAT
301 GATCCTGGTA ACCGGCGGGC CAGGCTTTCAT TGGCGCCAAT TTCGTAATCG ACTGGCTCGC
361 ACAGAGCGAT GAACCGGTTC TGAACCTAGA CAAGCTGACC TACGCGGGCA ACCTCGAGAC
421 GCTCGCATCG CTCAAGGACA ACCCGAAGCA CATCTTCGTG CAGGGCGACA TCGGCGACAG
481 CGCGCTGCTC GACCGCTTGC TGGCCGAGCA CAAGCCGCGT GCCGTGGTCA ACTTCGCGGC
541 CGAATCGCAC GTCGACCGCT CGATCCACGG CCCCAGAACG TTCGTGCAGA CCAACGTGCT
601 GGGCACCTTC CGCTGCTCG AATCCGTGCG CGGTTTCTGG AATGCCCTGC CGGCCGACCA
661 GAAGGCCGCC TTCCGCTTCC TGCATGTGTC GACCGACGAG GTCTACGGCT CGCTCTCCAA
721 GACCGACCCG GCCTTCACCG AAGAGAACAA GTACGAGCCC AACAGCCCGT ACTCGGCCAG
781 CAAGGCCGCC AGCGACCACC TCGTGCAGCG CTGGCACCAC ACCTACGGCC TGCCGGTGGT
841 CACCACCAAC TGCTCGAACA ACTACGGGCC GTTCCACTTC CCCGAGAAGC TCATTCCCTC
901 GATGATCGTC AACGCGCTGG CGGGCAAGCC GCTGCCCGTG TACGGCGACG GCATGCAGGT
961 GCGCGACTGG CTCTACGTA AGGACCACTG CAGCGCCATC CGCCGCGTGC TCGAAGCCGG
1021 CAAGCTCGGC GAGACCTACA ACGTGGGCGG CTGGAACGAG AAGCCCAACA TCGAGATCGT
1081 CAACACCGTC TGCGCGCTGC TCGACGAGCT GAGCCCCAAG GCCGGCGGCA AGCCGTACAA
1141 GGAACAGATC ACCTATGTGA CCGACCGCCC CGGCCACGAC CGCCGCTACG CGATCGACGC
1201 ACGCAAGCTC GAGCGCGAAC TCGGCTGGAA ACCTGCCGAG ACCTTCGACA GCGGCATCCG
1261 CAAGACGGTC GAGTGGTACC TCGCGAACGG CGAGTGGGTG CGCAACGTGC AAAGCGGCGC
1321 GTACCGCGAG TGGGTCGAGA AGCAATACGA CGCCGCACCG GCGAAGGCCA CCGCATGAAG
1381 CTGCTGCTGC TGGGCAAGGG CGGACAGGTC GGCTGGGAGC TGCAACGCAG CCTCGCGCCC
1441 CTGGGCGAAC TGGTGGCGCT CGATTTTCGAC AGCACCAGT TCAACGCCGA CTTCACTGCG
1501 CCCGAGCAGC TGGCCGAGAC AGTGCTGAAG GTGCGCCCCG ACGTCATCGT CAATGCCGCA
1561 GCGCACACCG CGGTCGACAA GGCCGAGAGC GAGCCCGAGT TCGCGCGCAA GCTCAACGCC
1621 ACCTCGCCCG GCGTGGTGGC CGAAGCCGCG CAGCAGATCG GCGCGCTGAT GGTTCACTAC
1681 TCGACCGACT ACGTCTTCGA CGGCAGCGGC AGCAAGCCGT GGAAAGAAGA CGATGCGACC
1741 GGCCCGCTCA GCGTCTACGG CAGCACCAAG CTCGAAGGCG AGCAACTGGT GGCAAAGCAC
1801 TGTGCGAAGC ACCTGATCTT TCGCACCAGC TGGGTCTATG CCGCGCGCGG CGGCAACTTC
1861 GCCAAGACCA TGCTGCGCAT CGCCAAGGAG CGCGACAAGC TGACCGTCAT CGACGACCAG
1921 TTCGGCGCGC CCACCGGCGC GGAAGTGTG GCGGACATCA CCGCGCACGC GATTGCGCGC
1981 ACGCTGCAGG ACCCGTCCAA GGCCGGGCTC TATCACGCGG TGGCCGGTGG CGTGACCACG
2041 TGGCACGGCT ATGCGCGCTT CGTGATCGAG CAGGCCAAGG CGGCGGGCGT GGAAGTGAAG
2101 GCCGGCCCCG AAGCGGTGCA GCGCGTGCCC ACCACGGCAT TCCCAGCGCC GGCCAGGCGG
2161 CCGCACAACG CGCGCTGGA CACCACCAAG CTGCAATCGA CCTTCGGCCT CGTGCTGCCC
2221 GAGTGGCAGT CCGGCGTCGC CCGCATGTTG CGCGAAACCT TCTGATATTC GCAGAGCAAG
2281 AGAGACACGA ACACCCCATG ACCAAGACGA CGCAACGCAA AGGCATCATC CTCGCCGGTG
2341 GCTCGGGCAC CGGCTGCGAC CCGCGCACGC TTGCCATGAG CAAACAACG CTGCCGGTGT
2401 ACGACAAGCC GATGATCTAT TACCCGCTGA GCACGCTGAT GCTGGGCGGC ATGCGCGACA
2461 TCCTGATCAT CAGCACGCCG CAGGACACGC CGCGTTTCCA GCAACTGCTG GGGGATGGCA

```

```

2521 GCCAATGGGG CATCAACCTG CAGTACGCGG TGCAGCCGAG CCCGGATGGT CTGGCGCAGG
2581 CGTTCATCAT CGGTGACAAG TTCGTGGGCA ACGACCCGAG TGCGCTGGTG CTGGGGGACA
2641 ACATCTTCTA TGGCCACGAC TTCGCCCATC TGCTGGCCGA TGCCGACGCC AAGACCTCGG
2701 GTGCGACGGT GTTCGCCTAC CACGTGCACG ACCCCGAGCG CTACGGCGTG GTGGCCTTCG
2761 ATGCCAAGGG CAGGGCGAGC AGCATCGAAG AAAAGCCGCT CAAGCCCAAG AGCAGCTATG
2821 CGGTCACGGG CCTCTACTTC TACGACAACC AGGTCGTCGA CATCGCCAAG GCCGTGAAGC
2881 CGAGCGCGCG CGGCGAACTC GAGATCACCG CGGTCAACCA GGCGTATCTC GACCTCGACC
2941 AGCTGAACGT GCAGATCATG CAGCGCGGCT ATGCGTGGCT CGATACCGGT ACGCACGACA
3001 GCCTGCTGGA AGCCGGGCAG TTCATTGCCA CGCTCGAGCA CCGCCAGGGG CTGAAGATCG
3061 CATGCCCCGA AGAGATCGCA TGGCGCAATG GCTTCATCTC AACCAGAGCA CTGAAAAAGC
3121 TCGCGGCGCC GCTGGAAGAAG AGCGGCTACG GCAAGTACCT CAAGCACCTG CTGAACGACG
3181 AGGTGCGCTC GTGAAGGCCA CGCCACCTC GATTCTTGAC GTGCTCGTGA TCGAGCCGAA
3241 GGTGTTTGGC GATGCACGGG GCTTCTTCTT CGAAAGCTTC AACCAGAAGG CCTTCGACGA
3301 AGCGATCGGC AAGCATGTCT ACTTCGTGCA GGACAACCAT TCGCGATCGG CCAAGGGTGT
3361 GCTGCGGGGG CTGCAATTAC AGGTCCAGCA GCCGCAAGGC AAGCTCGTGC GGGTGGTGGC
3421 TGGTGGGGTG TTCGACGTGG CCGTCGACAT CCGCAAGTCG TCGCCGACTT TTGGCAAATG
3481 GGTGGGTGTC GAGTTGAACG AAGACAACCA CAAGCAGCTC TGGGTGCCGG CAGGATTCGC
3541 GCACGGTTTC CTGGTGTGTA GCGAGACCGC GGAATTCCTC TACAAGACCA CCGACTACTA
3601 CGCGCCCGCC CACGAGCGCG CGATTGTCTG GAACGACCCC GCTGTCGGTA TTCGATGGCC
3661 GGATGTGGGA GGGGCACCGG TCCTGTGCGA GAAGGACGAA GACGGGTGTC TTCTGCAAGC
3721 GGCAGAGGTT TTCTAGTGTC CTTTCGTCTG ATAGCGGGGC GGCTTCGCGT ATCGGGATCC
3781 CGCGTTGAGC CCGCAAGAGT GCCCTGAGAG GGGGGGCGAA AAACTCACAA CGCCACTGCC
3841 TCGAGCAAAC GTGCGTCTCG CAGCTTTCTG AAGTTGTTGC ACCTTCTTTT TTTTCTCTT
3901 ACACTTTTGA AATGATTTTG AAAATCCGCG GCGATCGCAT GCAIGCTGCT GGAATCACC

```

Fig. 5

SEQ ID NO 3
 LONGITUD: 915
 TIPO: ADN
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 3

```

1 ATGAATGGCA TGCATATCGA CTCGGTCGAC CTCAATCTGC TCGCGCTGTT CGATGCGGTC
61 TACCGCGAGC GCAGCGTGAG CCGCGCCGCG GAGTCGCTGG GCCTCACGCA GCCTGCGGCA
121 AGCCATGGGC TGGGACGGCT GCGGCTGCTT TTGAAAGACG CGCTCTTCAC GCGTGCCCCC
181 GGCGGCGTGG CGCCACGCGC GCGCGCCGAC CGGCTCGCGG TGGCGGTGCA GGCGGCGCTC
241 GGCACGATCG AAGCGGCGCT GCACGAGCCC GATCGCTTCG AGCCCCAGGT GTCGCGCAAG
301 AGCTTTCGTA TTCACATGAG CGACATCGGC GAGGGGCGCT TCCTGCCCCG GCTGATGGCG
361 CGGCTCGGCG AGCTGGCGCC CGGCGTGCGG CTGGAGACCC TGCCGCTCTT GCCTGCGGAG
421 GTTGCGCCCG CACTCGACAG CGGCCGATC GATTTTCGCT TCGGCTTTCT CTCGACCGTG
481 CGCGACACGC AGCGCACGCA TCTTCTGAAA GACCGCTACA TCGTGCTGCT GCGCAAGGGC
541 CATCCCTTTG TGAAGCGCCG GCGCAAGGGG CAGGCGCTGC TCGAGGCGCT GCAGGAGCTC
601 GACTACGTGG CGGTGCGCAC GCACGCCGAC ACGCTGCGCA TCTTGAGTT GCTCAACCTC
661 GAAGACCGCC TCGCCTCAC GACCGAGCAC TTCAITGGTG TACCGGCCAT CGTGCGCGCC
721 ACCGATCTCG CGGTGGTGAT GCCGCGCAAC ATCGCGCGAG GGTTCGCGGA GGAGGGCGGC
781 TACGCGATCG TCGAGCCGCC GTTTCGCTG CGCGATTTCA GCGTGTGCTG GCACTGGAGC
841 AAGCGCTTCG AGGGCGACCC GGCCAACCGT TGGTTGCGGC AGGTGATCAC GGCGCTGTTC
901 TCCGAGCGCG GCTGA

```

Fig. 6

SEQ ID NO 5

LONGITUD: 7476

TIPO: ADN

ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 5

```

1  ATGAGTACCG TCGATCAGCT GGGCCGCACC GCCCCCTTA CCTCGGGGCA GATGGCGATG
61  TGGCTCGGCG CAAAGTTCGC GTCGCCCCGAC ACCAATTTCA ATCTCGCCGA AGCCAICGAC
121  ATCGCAGGCG AGATCGACCC CGCGATCTTC CTGGCGGCCA TCGACAGGT GGCCGATGAA
181  GTCGAGGCCA CGCGCTGAG CTTTCATCGAT ACCCGCAAG GGCCACGACA GGTCTCGCG
241  CCCGTTTTCA CCGCGAGAT CCCCTACCTC GACCTCAGCG GCGAGAGCGA TCCGAGGCC
301  GAGGCCGAGC GCTGGATGCA TGCGGACTAC ACCCGCAGCA TCGACCTCGC GCACGGGCAG
361  CTGTGGCTGT CCGCGCTGAT CCGCCTCGCG CCCGATCGCC ACATCTGGTA CCACCGCAGC
421  CATCACATCG CGCTCGACGG CTTCAGCGGC GGCCTCATCG CACGCCGCTT CGCCGACATC
481  TACACCGCGA TGGTCGACAA CAACGCAGCG GTGCCCCAAG ACTCGCGCCT TGCACCGATC
541  TCGCAGCTGG CCGACGAAGA ACATGCCTAT CGCGAGTCCG GCGGCTTCCC GCGCGACCGC
601  CAGTACTGGA CCGAGCGCTT CGCCGATGCA CCCGATCCGT TGAGCCTCGC CTCGCACCGC
661  TCGGTCAACG TCGGTGGCCT CTTGCGCCAG ACGGTGCACC TGCCGGCGGC CAGCGTGCAA
721  GCCCTGCAGA CCAICGCGCA AGAGCTCGGC ACCACGCTGC CGAAATCCT CATCGCCACC
781  ACCGCGGCCT ACCTGTACCG CGCAACGGGC ATCGAGGACA TGGCAATCGG CATCCCCGTC
841  ACCGCGCGCC ACAACGACCG CATGCGCCGC GTGCGCGCA TGGTGGCCAA CGCGCTGCCG
901  CTGCGCCTGG CGATGCGCGC GGACCTGCCG ATTCCGGAAC TGATCCGCGA AGTCGGCCGG
961  CAGATGCGGC AGATCCTGCG GCACCAGTCG TATCGCTACG AGCATTTGCG CAGCGACCTC
1021  AACATGCTGG TGAACAACCG GCAGCTCTTC ACCACCGTGG TCAACGTCGA GCCCTTCGAC
1081  TACGACTTCC GCTTTGCGGG CCATGCCGCG AAGCCGCGCA ACCTCTCGAA CGGCACGGCC
1141  GAGGACCTCG GCATCTTCCT GTACGAGCGC GGCAACGGGC AGGACCTGCA GATCGACTTC
1201  GACGCCAACC CCGCGGTGCA CACCGCAGAG GAACTGGCCG ATCACCAGCG CCGGCTGCTT
1261  GCCTTTCATCG ACGCCGTGAT CCGCCTGCCG TTGCAGGCCG TCGGCCAGAT CGACCTGCTC
1321  GGTGCCGAAG AGCGGCAGCA ATTGCTGGTC GAGTGGAACG ACACGGCCCA CGCCGTGCCG
1381  GACACCCATC TCACCGCGTT GATCGAAGCG CAGCTCGCAG CCGATCCGCA AGCCAICGCA
1441  TTGCGCTTCG ACGGCGAGGC GATGAACAAC GAAGAACTGA ACCGCCGCGC CAACCGTCTC
1501  GCCCACCTGC TCGCGCACG CCGCGCTGGC CCGGAGCGCA CCGTGGCGCT CGCGATCCCG
1561  CGTTGATGAG ACCTGATGAT TGCCTTGCTC GCCACGTTGA AGACCGGCGC GGCTTACCTG
1621  CCGGTCGATC CGGATTTCCC GCGGACCGC ATCGCCTTCA TGCTCGGCGA TCGCGAGCCC
1681  GTGTGCCTCG TCACGACCGA AGCCTCGCG GAGTCGCTGC CGGCACCGC CCCCACATTG
1741  GTGCTCGATG TAGCGCAAAC GATTGCGGAT CTGGAGAGTT GCAACGACAC CAACCCGGGC
1801  ATCGCGATCG ACCCTTCGCA TCCGGCCTAT GTGATCTACA CCTCGGGCTC GACCGGCATG
1861  CCCAAGGGTG CGGTGCTGTC GCACCGCGCC ATCGTCAACC GCCTGCGCTG GATGCAGGAC
1921  CGCTACGGCC TTCAGGCCGA CGACCGCGTG CTGCAGAAGA CGCCTTCCAG CTTGACGCTG
1981  TCGGTGTGGG AGTTCTTCTG GCCGCTGATC GACGGTGCCA CGCTGGTGCT TGCGAAACCG
2041  GGCGGCCACA AGGATGCGGC CTACCTCGCG GGGCTGATCG CGGAGGAGGG CATCACCACG
2101  ATCCACTTCG TGCCGTCGAT GCTCGAGGTC TTCTGCTCG AGCCACGGC GGGCGCATGC
2161  ACCACGCTGC GCCGCGTGAT CTGCAGCGGC GAAGCCTTGT CGCCCGCGCT GCAATCGCAG
2221  TTCCAGCAGC ACCTCTCGTG CGAGCTGCAC AACCTCTACG GTCCGACCGA GGCCGCGGTC
2281  GACGTCACTT CGTGGGAGTG CGAACGCACG GACGACGCGA AAGCCTCGAG CGTTCCTATC
2341  GGCCGCCCGA TCTGGAACAC CCAGATGCAC GTGCTCGACA GCGGCTGCA GCGCGTGCCG
2401  GCGGCGTGTA CTGGCGAGCT GTACATCGCG GCGCTCGGCC TCGCACGCGG CTACCTCAAG
2461  CGCCCGTTGC TGAGCGCCGA GCGTTTCATC GCCAACCCCT ACGGCACACC CGGCAGCCGC
2521  ATGTACCGCA CCGGCGACCT CGCGCGCTGG CGCAAGGACG GCAGCCTTGA CTTCTCGGC
2581  CGCGCCGACC AGCAGGTGAA GATCCGGGGC CTGCGCATCG AGCCGGGAGA GATCGAATCC
2641  GTGCTGCTGC AGCATCCGCA AGTCGCGCAG GCGGCGGTGG TGGCGCGCGA AGACGTACCG
2701  GGCGAAAAGC GTCTCGTGGC CTACGTCGTT GCGACGGACG CTGCCGATCC GCAAGCGGCC
2761  GAACTGCGCA CGCGCCTCGC GCAATCGCTG CCCGAGTACA TGGTGCCTTC GGCCTTCGTC

```

```

2821 AGCCTCCCGT CGCTGCCGCT CGGACCCAGC GGCAAGCTCG ACCGCAAGGC GCTGCCGCCC
2881 CCCGAAGTGC AGGCCGCCAC GCCGTACGCC GCGCCGCGCA CGCCGACCGA AAAGATCCTG
2941 GCCGGCCTCT GGGCCGAGAC GCTGCATTTC CCGCGCGTCG GTGTCAACGA CAACTTCTTC
3001 GAACTCGGCG GCCACTCGCT GATGATCGTG CAGCTCATGT CGATGATCCG GCAGCAATTC
3061 ATGATCGACC TGCCGGTCGA CACGCTGTTC CAGGTCTCCA CCATCGCGGG CCTTGCCGAG
3121 CTGCTCGACC AGGAATCGGT CGCCCGTCCG AGCCTGACTC CGATGCCGCG CCCCGCGCGC
3181 ATTCCGCTGT CCTTCGCGCA GCGCCGCTTG TGGCTGATGA ACCAGCTCGA AGGCGCGAAC
3241 CCGGCCTACA ACATGCCGCT CGCGCTGCGC CTGTCCGGTG TGCTCGATCG CACCGCATTG
3301 CATGCGGCGC TCGGCGACCT GGTGCAGCGC CACGAGAGCC TGCGCACGGT CTACCCGAAC
3361 GAAGACGGGC TGCCGTACCA GCACATCCTC GACGGCGCGG ATGCGCGTCC GCGGGTGATC
3421 GAGGCCGACA GCAGCGAAGA AGAAATCGCG GCGCAGCTTC ACGCCGCTGC GGGCCATGCC
3481 TTCGATCTCG GCAGCGCGGC GCCCTTGCGC GTCTACCTGT TCAAGCTCGC CGGCGACGAA
3541 CACGTGCTGC TGCTGCTCAC GCACCACATT GCCGGCGATG GCGCCTCGCT GCTGCCGCTA
3601 GCGCGCGACA TCAGCGTGGC CTATGCCGCG CGCTGCGAAG GCAAGGCGCC GGGCTGGGAG
3661 CCGCTGCCGC TGCAATACGC CGACTACGCG CTGTGGCAGC AGGAGCTGCT CGGCAGCGAA
3721 GACGATGCCG AGAGCATGGC CGGCCGCCAG CGTGAGTTCT GGCGTTCTTC GCTGAGCGAC
3781 CTGCCCAGAG AACTGGCGCT GCCCGTCGAC CACGCACGGC CGCTCGTGCC GACCTACCGC
3841 GGCGATGTGG TCCCCTGCA GATTCCGTTC CATGTGCATG AACGCATCCT GCAACTGGCG
3901 CGCGACGGGC AGGCCAGCGT CTTCAATGGT CTGCAGGCCG CACTCGCGGG CCTCTGAGC
3961 CGCCTCGGCG CGGGCGACGA CATCGTCATC GGCAGCCCGG TCGCGGGGCG CAGCGACCAT
4021 GCGCTGGACG AACTCATCGG CTGCTTCGTC AACACGCTGG TGCTGCGCAC TGACACCTCG
4081 GGCCAGCCGA GCCTGCGCGA GCTGGTCTCG CCGGTGCGCG CCACCAACCT CGCGGCCAT
4141 GCGAACCAGG AGTTTCCGTA CGACCGCCTC GTGGAGCTGC TGCGTCCGGG CCGCTCGCGC
4201 GCCAACCTGC CGCTGTTCCA GGTCAITGCT GGTTCACAG GCACGAGCCG CCTGTCTTC
4261 AGCCTGCCGG GCCTGTCGAT CGCGCCGCGC CCGGTGGCCA TCGACACCGC GAAGTTCGAC
4321 CTGTCTTTCA TCCTCGGCGA GCAACCGCGT GCCGATGGCC TGCCGGGCGG CATCTCCGGC
4381 GGCATCCAGT ACAGCACCGA CCTGTTTCGAG CGCAGCACGG TCGAGGCCAT GGGCGCGCGG
4441 CTGGTGCCTT TGCTGGAAGA GGCTGCGAG GCGCCCGACG ATGCGGTGAG TGGCTCGCC
4501 ATCCTGAGCG CGGAAGAAAC CGACCGCCTG CTGTCCGACT GGAGCGGCCG CACGCGCGAC
4561 CTTGCGCCGC TCTCGTTTCG CGACATGGTG GCCTCGCATG CCGCGGAGCG CCCGCTTGCA
4621 GATGCAGTGG TGCTCGACGA CGCGACCGTC AGCTACGCCG AACTCGATGC ACGCGCCAAC
4681 CGGCTCTCGC ACCTGCTGCG TGCGCAAGGC ATCGGGGTTG GCGCCATCGT CGCGACAGTG
4741 CTGCCGCGTT CGCTCGACCT CATCGTGGCG CACTTGGCCA TCGTGAAGGC CGGCGCGGCC
4801 TACCTGCCCA TCGACCCCAA CCACATGGCC GCGCGCAGCG CCTTCGTGTT CGAGGAGGCC
4861 GCGCCCGCCG CCGTGCTGAC GCACGATGCG CTGTTGCCCC AGCTGGTCGG CGTTCCCCGC
4921 TGCATCGCGC TCGACAGCGA CAGCATGGTT GCCGCGCTGG CCATCCAGTC GGATACGCCG
4981 CTGGTGCATG CGGCCAATCC ACAGGATGCC GCCTACCTCA TCTACACCTC CGGCTCCACC
5041 GGCATGCCCA AGGGCGTGGT GGTGCCGCAT GCGGGCCTGG GCAGCCTCGG CACCGCGATG
5101 GCGGAGCGGC TCGTCATCGG CCACGGCTCG CCGGTGCTGC AGTTCTCTCT CAGCGGCTTC
5161 GACGCGTCGG TGATGGACCA GCTGATGGCC TTTGGCGCCG GTGCCGCGCT GGTGGTGCCG
5221 GGGCCGGAGC AACTGCTCGG CACGGAGCTG GCCGATCTGC TCGAGAAGCA GGCCGTGAGC
5281 CACGCGCTGA TTCCGCCCGC CGCGCTCGCG ACCCTGCCGC ACGGCGAGTT CCCGCACCTG
5341 CAGACGCTGG TGGTCGGCGG CGATGCCCTGC ACCGCCGCGC TGCGGGCGAA GTGGTCGCAA
5401 GGCCGCGCA TGATCAACGC CTACGGCCCC ACCGAGATCA CCATCTGCGC GAGCATGAGC
5461 GCGCCGATGA CGGCCGAGGA GTTGCCCTCC ATCGGCCAGC CGATCTGGAA CACGCGGATG
5521 TATGTGCTCG ACAGCGCCCT GCAACCGGTG CCGCCGGGTG TCGCGGGCGA GCTCTACATC
5581 GCCGGCAGCG GCGTGGCGCG CGGCTATCTC AACC GGCCG CATTGAGTGC GGAACGCTTC
5641 ATCGCCGACC CGCATGGCGC GCCCGGCAGC CGCATGTACC GCAGCGGCGA CCTCGCACGC
5701 TGGCGCGCCG ACGGCACGCT CGACTTCCTC GGCCGCGCCG ACCAGCAGGT GAAGATCCGG
5761 GGCTTCCGCA TCGAGCCGGG CGAGATCGAA TCCGTGCTGC TCAAGCACCC GTTGATCACG
5821 CAGGCGCCCG TGATCGCCCG CGAGGACGTG CCCGCGGAGA AGCGCCTGCT CGCCTACTTC
5881 GTCGCGGGTT CCGAGCCGCA GCCACCGAG CTGCGCGCCC ACATGGCGTA GGCCTTGCCC
5941 GACTACATGG TGCTTTCGGC CTTCTGCGC CTGCGCTCGC TGCCGCTCAC GCAAGCGGC
6001 AAGCTCGACA AGAAGGCGCT GCCGGTGCCC GACCAGCAGC CCGCCGCGCT GTACGTGGAG
6061 CCGCGCACGC CGACCGAGAA ACTGCTCGCG GGCCTCTGGT CCGAGACGCT GCACCTGGAG
6121 CGTGTGCGCA TCCACGACAA CTTCTTCGAG ATCGGCGGGC ATTGCTCAT GCGCAICGAG
6181 CTGGGCATGC GCATCCGCCA GCAGGTGCGC GCGGACTTCC CGCACGCCGA GGTCTACAAC

```

```

6241 CGCCCGACGA TTGCCGACCT GGCCGCCTGG CTCGACAACG AAGGCGGCAC GGTGAGGCG
6301 CTGGACCTGT CGCGCGAGCT CGACCTGCCC GCGCACATCC GCGCGCAGGC CACTGCACCG
6361 AAGCTCGCAC CGCGCCGCGT GTTCTCTACC GGCGCGAGCG GCTTCGTGCG CAGTCACCTG
6421 CTGGCCGCGC TGTTCGCGCA CACCGCGGCC TGGTGGTCT GCCACGTGCG CGCGCCCGAC
6481 GAGCAGGCGG GCGAGCAGCG CCTCAAGCGC ACGCTGGCCC AGCGCCAGCT CGGTGCGATC
6541 TGGGACAACG CGCGCATCAA GGTCTGTACC GGCGACCTCG GCAAGCCGCG CCTGGGCCTC
6601 GATGACGCTG CCGTGCAACT GGTGCGCGAC GGCTGCGACG CCATCTACCA CTGCGCCGCG
6661 CAGGTGCACT TCCTGCATCC CTACGCGAGC CTCAAGCCCG CGAACGTCGA CAGCGTGGTC
6721 ACGCTGCTCG AATGGACGGC GCAGGGGCGC GCGAAGAGCA TGCATACGT CTCCACGCTG
6781 GCTGTGATCG ACCAGAACAA CAAGGAAGAC ACCATCACCG AGCAATCGGC GCTGGCCTCA
6841 TGGAGCGGGC TGGTCGACGG CTACAGCCAG AGCAAGTGGG TCGGCGATGC GCTGGCCCGC
6901 GAGGCGCAGG CGCGCGGCAT GCCGGTGGCG ATCTACCGGC TGGGGGAGT CACCGGCGAC
6961 CACACGCACG CGATCTGCAA TGCCGACGAC CTGATCTGGC GCGTGGCGCA TCTCTATGCC
7021 GACCTGGAAG CGATTCCCGA TATGGACCTG CCGCTCAACC TCACACCGGT GGACGACGTG
7081 GCGCGCGCCA TCCTCGGCC TCGGGCGCAG GAGGCCTCGT GGGGCCAGGT GTTCCACCTG
7141 ATGAGCCAGG CGGCGCTGCG GGTGCGCGAC ATTCCGCACG TCTTCGAGCG CATGGGCATG
7201 CGGCTGGAGC CGGTGCGGCT GAGGCGCTGG CTGCAGCGCG GCTGACGACG GCTGGCCGTC
7261 GCGCATGACC GCGACCTGGC CGCGGTGCTC GCCATCCTCG ACCGCTACGA CACCGCGGCC
7321 ACGCCGCCGC AGGTGAGCGG CGCGGCCACG CATGCGCAGC TCGAGGCCAT CGGCGCGCCG
7381 ATCCGCCCGG TGGACCGCGA CCTGCTGCGC CGCTACTTCG TCGACCTGGG CATCGACACC
7441 AAGGCGCGCC GCGCCCTGGA AACCACCACT TCATAG

```

Fig. 7

SEQ ID NO 7
 LONGITUD: 1320
 TIPO: ADN
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 7

```

1 ATGGCAGCT ATCTCATCGC AGCAACCGCC TTGCCGGGAC ACGTCTGCG GATGCTGGCC
61 ATCGCGCAGC ATCTGGTGAA CCAGGGGCAC GAGGTGCGGG TGACACCGC GAGCCAGTTC
121 AGGGCGCAGG CCGAGGCGAC CGGTGCGGGC TTCACGCCCT TCGAGCGCAC GATCGACTTC
181 GACTACCGCG ACCTGGACAA GCGCTTTCCC GAGCGCCAGC GCATCGCCTC GGCGCATGCG
241 CAGCTGTGCT TCGGCCTGAA GCACTTCTTT GCCGATGCGA TGGCCGCGCA GCATGCGGGC
301 CTGCAATCGA TCCTCGAAGA CTTTCGAGGC GATGCCATCG TGGTCGACAC GATGTTCTGC
361 GGCACCTTCC CGCTGCTGCT AGGCAAGGAG CGCGAAGACC GCGCGGCCAT CGTCGCGATC
421 GGCATCTCGG CGCTGCCGCT CTCGAGCTGC GACACCGCCT TCTTCGGCAC CGCGCTGCCG
481 CCGTCGTCCA CGCCGGAAGG GCGGGTGC GC AACAAGGCGA TGAACGCCAA CCTCAAACAG
541 GCGATGTTTC GCGAGGTGCA ACGCTACTTC GACACGCTGC TCGCGCGTTC GGGCCTGGCC
601 GCGCTGCCCG ATTTCTTCGT CGATGCGATG GTGAAGCTGC CCGATCTTTA CCTGCAGCTC
661 ACCGCGCCTT CGTTCGAATA CCCGCGCAGC GACCTGCCCC CGTCGGTGCA TTTCGTGCGC
721 CCGCTGCTCT CGCCCGCGAG CCGCGACTTC ACGCCGCCCG AGTGGTGGCA CGAGCTGGAC
781 GACGGCCGCT CGGTCTGTGT GGTACGCGAG GGCACGCTGG CCAACCAGAA TCCGTGCGAG
841 CTGATCGGCC CGACGCTGCA GGCGCTGGCC GGCGACAAGA ACATCCTCGT CATCGCCACC
901 ACCGGCGGCC CGGTGCCGCC CGCCCTGACG GTGAACCTGC CCGCCAACGC CCGCGTGGTG
961 CCGTTCCTGC CCTACGACCG GCTGCTGCCC AAGCTGCACG CGATGGTCAC CAACGCGCGC
1021 TACGGCTCGG TCAACCATGC ATTGAGCCTC GGTGTGCCGC TGGTGGTGGC CGGCACCTCC
1081 GAAGAGAAGC CCGAGATCGC CGCGCGCGTG GCCTGGTTCGG GCGCGGGCAT CAACCTCGCC
1141 ACCGGCCAGC CGACCGCGCG CCAGGTGCGG GACGCGGTGC GCAAGGTACT GGGCAACTCG
1201 ACCTATCGCC AGCGTGC GGTGCTGCGT GAGGACTTCG CTTGCCATCG CGCGCTGACC
1261 GGCATCGCCG GCGCCCTCGA GGCATTTCTG CAAACCTTCG CATCCGCGGA AATGGCTTGA

```

Fig. 8

SEQ ID NO 9
 LONGITUD: 213
 TIPO:ADN
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 9
 1 ATGAGCAACC CGTTCGACGA CAAGAACGCC AGCTTCCAGG TGCTGGTGAA CGACGAGGGC
 61 CAGCACTCGC TGTGGCCCGC CTTTCATCGC GTGCCC GCCG GCTGGCAGGT GGCGCTGGCG
 121 CCGACCGACC GCGACGCCCTG CAGCGCCTAC ATCGCGCGCA ACTGGCAGGA CATGCCCCCG
 181 CGTTCGCTGG TGGTGGCCAC GCGCGCCGGC TGA

Fig. 9

SEQ ID NO 11
 LONGITUD: 969
 TIPO:ADN
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 11
 1 ATGTCCTTCC CGTTCGGTGC CGTCGTCGTC ACCTATTTCC CGACCGGCGA GCAAGTGGCG
 61 AACCTCCATT CGCTGGCGGC CTCGTGTCCG CACCTCTGCG TGGTCGACAA CACGCCGCGAG
 121 GTGGGCGATT GGCATGCGGC GCTCGTCGAT GCGGGCGTTT CCGGTGCTGCA CAACGGCAAC
 181 CGCGGCGGCA TCGCGGGCGC CTTCAACCGC GGCATCATCG ACCTCGAAGC GCGGGGCGCC
 241 GAACTCTTCT TCCTGCTCGA CCAGGATTCG AAGCTGCCAC CCGGCTACTT CGATGCCATG
 301 TGCAGAGGCTG CGATGGTGGC CCGGGAGCGG AAGGGCGAGG GCAATGGTGA GGAAGACGCG
 361 GCCTTCCTGA TCGGCCCGCT CGTCCACGAC ACGAACCTGG ACGCGCTGAT CCCGCAATTC
 421 GGCCTCCAGG GCAAACGCGT CTACCAGTTC GACCTGCGGC AGCCCTTCAC CGAGCCGCTG
 481 ATGCGCTGCG CCTTCATGAT TTCTCGGGC TCCCTGATTT CGCGCGGCGC CTGGGCGCGG
 541 ATCGGCCCGGT TCGACGAGCG CTATGTGATC GACCACGTGG ACACCGACTA CTGCATGCGT
 601 GCCCTGGGTC GCGGCGTGCC GCTCTACCTG AATCCGCACG TCGTGCTGCG GCACCAGATT
 661 GGCGACATCC GTGCCCCTGC GCTGTTCGGC TGGGAAGATCC ACTTCATCAA CTACCCGCGC
 721 GCGCGGCGCT ACTACATCGC GCGCAATGCC ATCGATCTCT CGCGGGCGCA TGTGCGCGCC
 781 TTTCCCGCGA TCCTGTTCAT CAACGTTTAC ACGCTCAAGC AGATCCTGCC GATGCTGATG
 841 TTCGAGCGCG ACCGCTTCAA GAAGACCATC GCGCTGATGC TCGGCTGCTT CGATGGCCTG
 901 TTCGGGCGGC TCGGGGGCCT CGCGGAGGTG CATCCGCGGA TGGGCAAATA CCTGGGCGCG
 961 AGCGATTGA

Fig. 10

SEQ ID NO 13

LONGITUD: 1260

TIPO: ADN

ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 13

```

1   TTGACCGCCA CCCTTCCAGC GCCGCGCGTA CGCCGCGCCG CGCTCGCCTT CATCTTCGTC
61  ACGGTGCTGA TCGACTTCAT GGC GTTCGGC CTGATCCTGC CCGGCCTGCC GCACCTGGTG
121 GAGCGGCTGG CCGGCGGCAG CACGGTAAAC GCGGCGTACT GGATCGCTGT GTTCGGCACC
181 GCGTTCGCGG CGATCCAGTT CGTGAGCTCG CCGATCCAGG GCGCGCTGTC CGACCGCTTC
241 GGGCGGCGGC CGGTGATCCT GCTGTCTGTC TTCGGCCTCG GCGTGGATTT CGTGTTCATG
301 GCCCTGGCCG ACAGCCTGCC GTGGCTGTTT GTGCGCCGGG TGGTCTCCGG CGTGTTCCTG
361 GCCAGCTTCA CCAICGCCAA TGCCTACATC GCCGATGTGA CGCTGCCGGA GGAGCGCGCC
421 CGCAGCTACG GCATCGTGGG GGCCGCGTTC GGCAIGGGCC TGGTGTTCGG GCCGGTGTCT
481 GGCGGGCAAC TGAGCCACAT CGATCCGCGC CTGCCGTTCT GGTTCGCGGC CCGCTTGACG
541 CTGCTCAGCT TCTGCTACGG ATGGTTCGTG TTGCCC GAAT CGCTGCCGCC CGAGCGGCGT
601 GCGCGCAAGT TCGACTGGTC GCATGCCAAT CCGGTTGGGA CGCTGGTGCT GCTCAAGCGC
661 TATCCGCAGG GTTTCGGACT GCGGCGCGTG ATCTTCCTCG TGAACCTGGC TCAGTACGTC
721 TATCCAGCGT GTTTCGTGCT GTTCGCCGAC TACCGGTATC ACTGGAAGGA AGACGCCGTG
781 GGCTGGGTGC TCGGCGCGGT GGGCGTGCTC AGCGTGCTGG TCAATGCGCT GTTGATCGGG
841 CCGGGCGTGA AGCGCTTCGG CGAGCGCCGC GCCCTGTTGC TCGGCATGGG CTTGCGCGTG
901 CTCGGCTTCG TCATCATCGG GTTTGCCGAC GCTGGATGGA TCCTCCTGGC CGGGGTGCCG
961 TTCGGCATTG TGCTGGCGTT CGCCGGACCG GCGGCGCAGG CGCTGGTCAC GCTGCAGGTC
1021 GGCACCGCCG AGCAGGGCCG CATCCAGGGG GCGCTACCA GCCTGGTGTC GGTGGCGGGC
1081 ATCGTCGGGC CGGCGATGTT CGCCGGCAGC TTCGGTACT TCAICGGCGC GGACGCGCCG
1141 GTGCACTTGC CCGGCGCGCC GTTTTCCTC GCTGCGGCGT TCCTCTGCAT CGGCACGCTG
1201 ATCGCGTGGC GCTACGCACA GCCGAAGCCC GCGACGCGAG CGGTGCCCGA GCCGACCTGA

```

Fig. 11

SEQ ID NO 15
 LONGITUD: 1080
 TIPO: ADN
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 15

1	ATGATCCTGG	TAACCGGCGG	CGCAGGCTTC	ATTGGCGCCA	ATTTCGTACT	CGACTGGCTC
61	GCACAGAGCG	ATGAACCGGT	CGTGAACCTA	GACAAGCTGA	CCTACGCGGG	CAACCTCGAG
121	ACGCTCGCAT	CGCTCAAGGA	CAACCCGAAG	CACATCTTCG	TGCAGGGCGA	CATCGGCGAC
181	AGCGCGCTGC	TCGACCGCCT	GCTGGCCGAG	CACAAGCCGC	GTGCCGTGGT	CAACTTCGCG
241	GCCGAATCGC	ACGTCGACCG	CTCGATCCAC	GGCCCCGAAG	ACTTCGTGCA	GACCAACGTG
301	CTGGGCACCT	TCCGCCTGCT	CGAATCCGTG	CGCGGTTTCT	GGAATGCCCT	GCCGGCCGAC
361	CAGAAGGCCG	CCTTCCGCTT	CTTGCATGTG	TCGACCGACG	AGGCTACGG	CTCGCTCTCC
421	AAGACCGACC	CGGCCCTTAC	CGAAGAGAAC	AAGTACGAGC	CCAACAGCCC	GTACTCGGCC
481	AGCAAGGCCG	CCAGCGACCA	CCTCGTGCGC	GCCTGGCACC	ACACCTACGG	CCTGCCGGTG
541	GTCACCACCA	ACTGCTCGAA	CAACTACGGG	CCGTTCCACT	TCCCCGAGAA	GCTCAITCCC
601	CTGATGATCG	TCAACGCGCT	GGCGGGCAAG	CCGCTGCCCC	TGTACGGCGA	CGGCAIGCAG
661	GTGCGCGACT	GGCTCTACGT	GAAGGACCAC	TGCAGCGCCA	TCCGCCGCGT	GCTCGAAGCC
721	GGCAAGCTCG	GCGAGACCTA	CAACGTGGGC	GGCTGGAACG	AGAAGCCCAA	CATCGAGATC
781	GTCAACACCG	ICTGCGCGCT	GCTCGACGAG	CTGAGCCCCA	AGGCCGGCGG	CAAGCCGTAC
841	AAGGAACAGA	TCACCTATGT	GACCGACCGC	CCCGGCCACG	ACCGCCGCTA	CGCGATCGAC
901	GCACGCAAGC	TCGAGCGCGA	ACTCGGCTGG	AAACCTGCCG	AGACCTTCGA	CAGCGGCATC
961	CGCAAGACGG	TCGAGTGGTA	CCTCGCGAAC	GGCGAGTGGG	TGCGCAACGT	GCAAAGCGGC
1021	GCGTACCGCG	AGTGGGTCTGA	GAAGCAATAC	GACGCCGCAC	CGGCGAAGGC	CACCGCATGA

Fig. 12

SEQ ID NO 17
 LONGITUD: 891
 TIPO: ADN
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 17

1	ATGAAGCTGC	TGCTGCTGGG	CAAGGGCGGA	CAGGTCGGCT	GGGAGCTGCA	ACGCAGCCTC
61	GCGCCCCCTGG	GCGAACTGGT	GGCGCTCGAT	TTCGACAGCA	CCGACTTCAA	CGCCGACTTC
121	AGTCGCCCCG	AGCAGCTGGC	CGAGACAGTG	CTGAAGGTGC	GCCCCGACGT	CATCGICAAT
181	GCCGCAGCGC	ACACCGCGGT	CGACAAGGCC	GAGAGCGAGC	CCGAGTTCGC	GCGCAAGCTC
241	AACGCCACCT	CGCCCGGCGT	GGTGGCCGAA	GCCGCGCAGC	AGATCGGCGC	GCTGAIGGTT
301	CACTACTCGA	CCGACTACGT	CTTCGACGGC	AGCGGCAGCA	AGCCGTGGAA	AGAAGACGAT
361	GCGACCGGCC	CGCTCAGCGT	CTACGCGAGC	ACCAAGCTCG	AAGGCGAGCA	ACTGGTGGCA
421	AAGCACTGTG	CGAAGCACCT	GATCTTTTCG	ACCAGCTGGG	TCTATGCCGC	GCGCGGCGGC
481	AACCTTCGCA	AGACCATGCT	GCGCATCGCC	AAGGAGCGCG	ACAAGCTGAC	CGTCATCGAC
541	GACCAGTTCG	GCGCGCCAC	CGGCGCGGAA	CTGCTGGCCG	ACAACCCGCG	GCACGCGATT
601	GCGCGGACGC	TGCAGGACCC	GTCCAAGGCC	GGGCTCTATC	ACGCGGTGGC	CGGTGGCGTG
661	ACCACGTGGC	ACGGCTATGC	GCGCTTCGTG	ATCGAGCAGG	CCAAGGCGGC	GGGCGTGGAA
721	CTGAAGGCCG	GCCCCGAAGC	GGTCGAGCCC	GTGCCCACCA	CGGCATTCCC	GACGCCGGCC
781	AGGCGGCCGC	ACAACCTCGC	CCTGGACACC	ACCAAGCTGC	AATCGACCTT	CGGCCTCGTG
841	CTGCCCAGAT	GGCAGTCCGG	CGTCGCCCCG	ATGTTGCGCG	AAACCTTCTG	A

Fig. 13

SEQ ID NO 19
 LONGITUD: 897
 TIPO: ADN
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 19

1	ATGACCAAGA	CGACGCAACG	CAAAGGCATC	ATCCTCGCCG	GTGGCTCGGG	CACCCGCCTG
61	CACCCCGCGA	CGCTTGCCAT	GAGCAAACAA	CTGCTGCCGG	TGTACGACAA	GCCGATGATC
121	TATTACCCGC	TGAGCACGCT	GATGCTGGGC	GGCATGCGCG	ACATCCTGAT	CATCAGCACG
181	CCGCAGGACA	CGCCGCGTTT	CCAGCAACTG	CTGGGGGATG	GCAGCCAATG	GGGCATCAAC
241	CTGCAGTACG	CGGTGCAGCC	GAGCCCGGAT	GGTCTGGCGC	AGGCGTTCAT	CATCGGTGAC
301	AAGTTCGTGG	GCAACGACCC	GAGTGCCTG	GTGCTGGGGG	ACAACATCTT	CTATGGCCAC
361	GACTTCGCCC	ATCTGCTGGC	CGATGCCGAC	GCCAAGACCT	CGGGTGCAGC	GGTGTTCGCC
421	TACCACGTGC	ACGACCCCGA	GCGCTACGGC	GTGGTGGCCT	TCGATGCCAA	GGGCAGGGCG
481	AGCAGCATCG	AAGAAAAGCC	GCTCAAGCCC	AAGAGCAGCT	ATGCGGTCAC	GGGCCTCTAC
541	TTCTACGACA	ACCAGGTCGT	CGACATCGCC	AAGGCCGTGA	AGCCGAGCGC	GCGCGCGGAA
601	CTCGAGATCA	CCGCGGTCAA	CCAGGCGTAT	CTCGACCTCG	ACCAGCTGAA	CGTGCAGATC
661	ATGCAGCGCG	GCTATGCGTG	GCTCGATACC	GGTACGCACG	ACAGCCTGCT	GGAAGCCGGG
721	CAGTTCATTG	CCACGCTCGA	GCACCGCCAG	GGGCTGAAGA	TCGCATGCCC	CGAAGAGATC
781	GCAIGGCGCA	ATGGCTTCAT	CTCAACCGAG	CAACTCGAAA	AGCTCGCGGC	GCCGCTGGAA
841	AAGAGCGGCT	ACGGCAAGTA	CCTCAAGCAC	CTGCTGAACG	ACGAGGTGCG	CTCGTGA

Fig. 14

SEQ ID NO 21
 LONGITUD: 546
 TIPO: ADN
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 21

1	GTGAAGGCCA	CGCCACCTC	GATTCTTGAC	GTGCTCGTGA	TCGAGCCGAA	GGTGTGTTGGC
61	GATGCACGGG	GCTTCTTCTT	CGAAAGCTTC	AACCAGAAGG	CCTTCGACGA	AGCGATCGGC
121	AAGCATGTCT	ACTTCGTGCA	GGACAACCAT	TCGCGATCGG	CCAAGGGTGT	GCTGCGGGGG
181	CTGCATTACC	AGGTCCAGCA	GCCGCAAGGC	AAGCTCGTGC	GGGTGGTGGC	TGGTGCGGTG
241	TTTCGACGTG	CCGTGACAT	CCGCAAGTCG	TCGCCGACTT	TTGGCAAATG	GGTGGGTGTC
301	GAGTTGAACG	AAGACAACCA	CAAGCAGCTC	TGGGTGCCGG	CAGGATTCGC	GCACGGTTTC
361	CTGGTGTGTA	GCGAGACCGC	GGAATTCTCT	TACAAGACCA	CCGACTACTA	CGCGCCCGCC
421	CACGAGCGCG	CGATTGTCTG	GAACGACCCC	GCTGTGCGTA	TTTCGATGGC	GGATGTGGGA
481	GGGGCACCGG	TCCTGTGCAA	GAAGGACGAA	GACGGGTGTC	TTCTGCAAGC	GGCAGAGGTT
541	TTCTAG					

Fig. 15

SEQ ID NO 4

LONGITUD: 304

TIPO: PÉPTIDO

ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 4

MNGMHIDSVDLNLRLFDVYRERSVSRAAESLGLTQPAASHGLGRLRLLLKDALFTRAPGGVAPTPRADRLAVAVQ
AALGTIEAALHEPDRFEPQVSRKSFRIHMSDIGEGRFLPALMARLGELAPGVRLETLPLLPAEVAPALDSGRIDFAF
GFLSTVRDTQORTHLLKDRYIVLLRKGHFVKRRRKQALLEALQELDYVAVRTHADTLRILQLLNLEDRLRLTTEHF
MVLPAIVRATDLAVVMPRNIARGFAEEGGYAIVEPPFPLRDFSLSLHWSKRFEGDPANRWLRQVITALFSESG

Fig. 16

SEQ ID NO 6
 LONGITUD: 2491
 TIPO: PÉPTIDO
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 6

MSTVDQLGRTAPLTSGQMAMWLGAKFASPDNTNFNLAEAIDIAGEIDPAIFLAAMRQVADEVEATRLSFIDTPQGPRQ
 VVAPVFTGEIPYLDLSGESDPQAEAERWMHADYTRSIDLHAGQLWLSALIRLAPDRHIWYHRSHHIALDGFSGGLIA
 RRFADIYTAMVDNNAAVPEDSRLAPISQLADEEHAYRESGRFPRDRQYWTERFADAPDPLSLASHRSVNVGGLLRQT
 VHLPAASVQALQTTIAQELGTTLPQILIATTAAYLYRATGIEDMAIGIPVTARHNDRMRRVPAMVANALPLRLAMRAD
 LPIPELIREVGRQMRQILRHQSYRYEHLRSDLNMLVNNRQLFTTVNVNEPFDYDFRFRAGHAAPRNLSNGTAEDLGI
 FLYERNGNGQDLQIDFDANPAVHTAEELADHQRRLLAFIDAVIRLPLQAVGQIDLLGAEERQQLLVEWNDTAHAVPDT
 HLTALIEAQLAADPQAIALRFDGEAMNNEELNRRANRLAHLRLRARGAGPERTVALAIPRSMDLMIALLATLKTGAAY
 LPVDPDFPADRIAFMLGDAQPVCLVTEALAESLPAAAPTLLLDVAQTIADLESCNDTNPGIAIDPSHPAYVIYTS
 STGMPKGAUVSHRAIVNRLRWMDRYGLQADDRVLQKTPSSFDVSVWEFFWPLIDGATLVLAKEGPKGKDAAYLAGLI
 AEEGITTITHFVPSMLEVFLLEPTAGACTTLRRVICSGEALSPALQSQFQQLHLSCELHNLYGPTAAVDVTSWECERT
 DDAEASSVPIGRPIWNTQMHVLDLGLQPVPAVGTGELYIAGVGLARGYLKRPLLSAERFIANPYGTPGSRMYRTGDL
 ARWRKDGSLDFLGRADQQVKIRGLRIEPEIESVLLQHPQVAQAQAVVAREDPGEKRLVAYVVATDAADPQAELRT
 RLAQSLPEYMPVSAFVSLPSLPLGPGSKLDRKALPPPEVQAATPYAAPRTPEKILAGLWAE TLHLPRVGVNDNFFE
 LGGHSLMIVQLMSMIRQQFMIDLPVDTLQVSTIAGLAELLDQESVARESLTPMPRPARIPLSFAQRRLWLMNQLEG
 ANPAYNMPLALRLSGVLDRTALHAALGDLVQRHESLRTVYPNEDGLPYQHILDGADARPAVIEADSSEEEIAAQLHA
 AAGHAFDLGSAAPLRVYLFKLAGDEHVLLLTTHIAGDGASLLPLARDISVAYAAARCEGKAPGWEPPLQYADYALW
 QQELLGSEDDAESMAGRQREFWRSSLSLPEQLALPDHARPLVPTYRGDVVPLQIPSHVHERILQLARDGQASVFM
 VLQAAALAGLLSRLGAGDDIVIGSPVAGRSDHALDELIGCFVNTLVLRDTDTSGQPSLRELVSVRATNLAAAYANQEPF
 YDRLVELLRPGRSRANLPLFQVMLGFQGTSLRSLFSLPGLSIAPQPVAIDTAKFDLSFILGEQRGADGLPGGISGGIQ
 YSTDLFERSTVEAMGARLVRLLEEACEAPDDAVSGLAILSAEETDRLLSDWSGRTRDLAPLSFADMVASHAAERPLA
 DAVVLDDATVSYAELDARANRLSHLLRAQGIGVGAIVATVLPRLSLDLIVAHLAIVKAGAAAYLPIDPNHMAARSAFVF
 EEAAPAAVLTHDALLPELVGVPRCIALDSDSMVAALAIQSDTPIVHAANPQDAAYLIYTSGSTGMPKGVVVPHAGLG
 SLGTAMAEERLVIGHGSRVLQFSSSGFDASVMDQLMAFGAGAAALVVPGEQLLGTTELADLLEKQAVSHALIPPAALAT
 LPHGEFPHLQTLVVGDACTAALAAKWSQGRRMINAYGPTTEITICASMSAPMTAEELPSIGQPIWNTRMYVLDLALQ
 PVPPGVAGELYIAGSGVARGYLNRPALSAERFIADPHGAPGSRMYRSGDLARWRADGTLDLFLGRADQQVKIRGFRIE
 PGEIESVLLKHPLITQAAVIAREDPGEKRLVAYFVAGSEPQPTELRAHMAQALPDYMVPVSAFVRLPSLPLTQSGKL
 DKKALPVPDQQAALYVEPRTPEKLLAGLWSETLHLERVGIDHNFFEIGGHSLMAIQLMRIRQQVRADFPFAEVY
 NRPTIADLAALWLDNEGGTVEALDLSRELDLPAHIRPQATAPKLAPRRVFLTGASGFVGSLLAALLRDTAACVCHV
 RAPDEQAGEQRLKRTLAQRQLGAIWDNARIKVVTDGLGKPRGLDAAVQLVRDGCDAIYHCAAQVDFLHPYASLKP
 ANVDSVVTLLLEWTAQGRAKSMHYVSTLAVIDQNNKEDTITEQSALASWSGLVDGYSQSKWVGDALAREAQARGMPVA
 IYRLGAVTGDHTHAICNADDLIWRVAHLYADLEAIPDMDLPLNLTPVDDVARAILGLAAQEASWGQVFHLSQAALR
 VRDIPHVFERMGMRLPEVGLPWLQRAHARLAVAHDRDLAAVLAILDRYDTTATPPQVSGAATHAQLEAIGAPIRPV
 DRDLLQRYFVDLGDITKARRALETTS

Fig. 17

SEQ ID NO 8
 LONGITUD: 439
 TIPO: PÉPTIDO
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 8
 MARYLIAATALPGHVLPMIAIAQHVLNQGHEVRVHTASQFRAQAEATGAGFTPFERTIDFDYRDLDKRFPERQRIAS
 AHAQLCFGLKHFFADAMAAQHAGLQSILEDFEADAIVVDTMFCGTFPLLLGKEREDRPAIVGIGISALPLSSCDTAF
 FGTALPPSSSTPEGRVRNKAMNANLKQAMFGEVQRYFDILLARSGLAALPDFFVDAMVKLPDLYLQLTAPSFEYPRSD
 LPASVHFVGPPLSPASRDFTPEWWHELDGGRSVVLVTQGTILANQNPSQLIGPTLQALAGDKNILVIATTGGPVPPA
 LTVNLPANARVVPFLPYDRLLPKLHAMVTNGGYGSVNHALSLGVPLVVAGTSEEKPEIAARVAWSGAGINLATGQPT
 ARQVGDAVRKVLGNSTYRQRAAVLREDFACHRALGTGIAGALEALLQTFASAEMA

Fig. 18

SEQ ID NO 10
 LONGITUD: 70
 TIPO: PÉPTIDO
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 10
 MSNPFDDKNASFQVLVNDEGQHSLWPAFIAVPAGWQVALPTDRDACSAYIAANWQDMRPRSLVVATAAG

Fig. 19

SEQ ID NO 12
 LONGITUD: 322
 TIPO: PÉPTIDO
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 12
 MSFPFGAVVVTYFPTGEQVANLHSLAASCPHLCVVDNTPQVGDWHAALVDAGVSVLHNGNRGGIAGAFNRGIIDLEA
 RGAELEFFLLDQSKLPPGYFDAMCEAMVARERKGEENGEEAAFLIGPLVHDTNLDALIPQFGLQGKRVYQFDLRQ
 PFTEPLMRCAFMISSGSLISRGAWARIGRFDERYVIDHVDTDYCMRALGRGVPLYLNPHVVLVRHQIGDIRARSLFGW
 KIHFINYPARRYIARNALDLSRAHVRAFPAILFINVYTLKQILPMLMFERDRFKKTIALMLGCFDGLFGRLLGGLG
 EVHPRMGKYLGRSD

Fig. 20

SEQ ID NO 14
 LONGITUD: 419
 TIPO: PÉPTIDO
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 14
 MTATLPAPRVRRRAALAFIFVTVLIDFMAFGLILPGLPHLVERLAGGSTVTAAYWIAVFGTAFAAIQFVSSPIQGALS
 DRFGRRPVILLSCFGLGVDFVFMALADSLPWLFVGRVVSQVFSASFTIANAYIADVTLPEERARSYGIVGAAFGMGL
 VFGPVLGGQLSHIDPRLPFWFAAGLTLLSFCYGWFLVLPESLPPERRARKFDWSHANPVGTLVLLKRYPQVFGLAAYI
 FLVNLAQYVYPSVFVLFADYRYHWKEDAVGWVLGAVGVLSVLVNALLIGPGVKRFGERRALLLGMGFVLGFVIIGF
 ADAGWILLAGVPFGILLAFAGPAAQALVTLQVGTAEQGRIQGALTSLVSVAGIVGPAMFAGSFGYFIGADAPVHLPG
 APFFLAALFLCIGTLIAWRYAQPKPATAAVPEPT

Fig. 21

SEQ ID NO 16
 LONGITUD: 359
 TIPO: PÉPTIDO
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 16
 MILVTGGAGFIGANFVLDWLAQSDEPVVNLDKLTYAGNLETLASLKDNPKHIFVQGDIGDSALLDRLLAEHKPRVV
 NFAAESHVDRSIHGPEDFVQTNVLGTFRLLSVRGFWNALPADQKAARFLHVSTDEVYGSLSKTDPAFTEENKYE
 NSPYSASKAASDHLVRAWHHTYGLPVVTTNCSNNYGPFFHPEKLIPLMIVNALAGKPLPVYGDGMQVRDWLYVKDHC
 SAIRRVLEAGKLGETYNVGGWNEKPNIEIVNTVCALLDELSPKAGGKPYKEQITYVTDPRPGHRRYDAIDARKLEREL
 GWKPAETFDGIRKTVEWYLANGWVRNVQSGAYREWVEKQYDAAPAKATA

Fig. 22

SEQ ID NO 18
 LONGITUD: 296
 TIPO: PÉPTIDO
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 18
 MKLLLLGKGGQVGWELQRLAPLGLVALDFDSTDFNADFSRPEQLAETVLKVRPDVIVNAAAHTAVDKAESEPEFA
 RKLNATSPGVVAEAAQQIGALMVHYSTDYVFDGSGSKPWKEDDATGPLSVYGSTKLEGEQLVAKHCAKHLIFRTSWV
 YAARGGNFAKTMRLIAKERDKLTVIDDQFGAPTGAELLADITAHAIRATLQDPSKAGLYHAVAGGVTTWHGYARFVI
 EQAKAAGVELKAGPEAVEPVPTTAFPTPARRPHNSRLDTTKLQSTFGLVLPEWQSGVARMLRETF

Fig. 23

SEQ ID NO 20
LONGITUD: 298
TIPO: PÉPTIDO
ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 20
MTKTTQQRKGIILAGGSGTRLHPATLAMSKQLLPVYDKPMIYYPLSTLMLGGMRDILIIISTPQDTPRFQQLLGDGSQW
GINLQYAVQPSPDGLAQAFIIGDKFVGNDFSAVLGDNIFYGHDFAHLLADADAKTSGATVFAYHVHDPERYGVVAF
DAKGRASSIEEKPLKPKSSYAVTGLYFYDNQVVDIAKAVKPSARGELEITAVNQAYLDLDQLNVQIMQRGYAWLDTG
THDSLLEAGQFIATLEHRQGLKIACPEEIAWRNGFISTEQLEKLAAPLEKSGYGKYLKHLNDEVRS

Fig. 24

SEQ ID NO 22
LONGITUD: 181
TIPO: PÉPTIDO
ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 22
MKATPTSIPDVLVIEPKVFGDARGFFFESEFNQKAFDEAIGKHVDFVQDNHRSRAGVLRGLHYQVQQPQGKLVVRVVR
GAVFDVAVDIRKSSPTFGKWVGVELNEDNHKQLWVPAGFAHGFLVLSETAEFLYKTTDYYAPAHERAIVWNDPAVGI
RWPDVGGAPVLSKKDEDGCLLQAAEVF

Fig. 25

SEQ ID NO 23
 LONGITUD: 1029
 TIPO: ADN
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 23

1	ATGGGCAGCA	GCCATCATCA	TCATCATCAC	AGCAGCGGCC	TGGTGCCGCG	CGGCAGCCAT
61	ATGTCCTTCC	CGTTCGGTGC	CGTCGTCGTC	ACCTATTTCC	CGACCGGCCGA	GCAAGTGGCG
121	AACCTCCATT	CGCTGGCGGC	CTCGTGTCGG	CACCTCTGCG	TGGTCGACAA	CACGCCGCAG
181	GTGGGCGATT	GGCATGCGGC	GCTCGTCGAT	GCGGGCGTTT	CGGTGCTGCA	CAACGGCAAC
241	CGCGGCGGCA	TCGCGGGCGC	CTTCAACCGC	GGCATCATCG	ACCTCGAAGC	GCGGGGCGCC
301	GAACCTCTTCT	TCCTGCTCGA	CCAGGATTCG	AAGCTGCCAC	CCGGCTACTT	CGATGCCATG
361	TGCGAGGCTG	CGATGGTGGC	CCGGGAGCGG	AAGGGCGAGG	GCAATGGTGA	GGAAGACGCG
421	GCCTTCCTGA	TCGGCCCCTG	CGTCCACGAC	ACGAACCTGG	ACGCGCTGAT	CCCGCAATTC
481	GGCCTCCAGG	GCAAACGCGT	CTACCAGTTC	GACCTGCGGC	AGCCCTTCAC	CGAGCCGCTG
541	ATGCGCTGCG	CCTTCAIGAT	TTCCTCGGGC	TCCCTGATTT	CGCGCGGCGC	CTGGGCCCCTG
601	ATCGGCCGGT	TCGACGAGCG	CTATGTGATC	GACCACGTGG	ACACCGACTA	CTGCATGCGT
661	GCCCTGGGTC	GCGGCGTGCC	GCTCTACCTG	AATCCGCACG	TCGTGCTGCG	GCACCAGATT
721	GGCGACATCC	GTGCCCCTGC	GCTGTTTCGG	TGGAAGATCC	ACTTCATCAA	CTACCCGGCC
781	GCGCGGCGCT	ACTACATCGC	GCGCAATGCC	ATCGATCTCT	CGCGGGCGCA	TGTGCGCGCC
841	TTTCCCGCGA	TCCTGTTCAT	CAACGTTTAC	ACGCTCAAGC	AGATCCTGCC	GATGCTGATG
901	TTCGAGCGCG	ACCGCTTCAA	GAAGACCATC	GCGCTGATGC	TCGGCTGCTT	CGATGGCCTG
961	TTCGGGCGGC	TCGGGGGCCT	CGGCGAGGTG	CATCCGCGGA	TGGGCAAATA	CCTGGGCGCG
1021	AGCGATTGA					

Fig. 26

SEQ ID NO 24
 LONGITUD: 342
 TIPO: PÉPTIDO
 ORGANISMO: *Variovorax paradoxus*

SECUENCIA: 24

MGSSHHHHHSSGLVPRGSHMSFPFGAVVVITYFPTGEQVANLHSLAASCPHLCVVDNTPQVGDWHAALVDAGVSVLH
 NGNRGGIAGAFNRGIIDLEARGAELFFLLDQDSKLPPGYFDAMCEAMVARERKGEENGEEDAAFLIGPLVHDTNLD
 ALIPQFGLQGRVYQFDLRQPFTEPLMRCAFMISGSLISRGAWARIGRFDERYVIDHVDTDYCMRALGRGVPLYL
 NPHVVLRHQIGDIRARSLFGWKIHFINYPARRYYIARNAIDLSRAHVRAFPAILFINVYTLKQILPMLMFERDRFK
 KTIALMLGCFDGLFGRLGGLGEVHPRMGKYLGRSD

Fig. 27