(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 660 227**

(51) Int. Cl.:

A61K 39/385 (2006.01)
A61K 39/02 (2006.01)
A61K 39/108 (2006.01)
C07K 1/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **16.11.2010** **PCT/US2010/002980**

(87) Fecha y número de publicación internacional: **26.05.2011** **WO11062615**

(96) Fecha de presentación y número de la solicitud europea: **16.11.2010** **E 10831896 (5)**

(97) Fecha y número de publicación de la concesión europea: **20.12.2017** **EP 2501406**

(54) Título: **Sistema biosintético que produce polisacáridos inmunogénicos en células procariotas**

(30) Prioridad:

19.11.2009 US 272931 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
21.03.2018

(73) Titular/es:

GLAXOSMITHKLINE BIOLOGICALS S.A. (100.0%)
Rue de l'Institut, 89
1330 Rixensart, BE

(72) Inventor/es:

WACKER, MICHAEL y
WAECHTER, CHARLES

(74) Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 660 227 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

Sistema biosintético que produce polisacáridos inmunogénicos en células procariotas

Referencia a solicitud relacionada

La presente solicitud reivindica el beneficio de prioridad bajo la 35 U.S.C. § 119(e) de la Solicitud Provisional de Patente de EE. UU. Nº 61/272,931, presentada el 19 de noviembre de 2009, que se incorpora en el presente documento por referencia en su totalidad.

Campo de la invención

La presente invención se refiere al uso de un sistema biosintético y proteínas para preparar una vacuna. Además, la invención se refiere a un sistema biosintético procariota recombinante que tiene una epimerasa que inicia la síntesis de un oligo o polisacárido con un monosacárido especificado en el extremo reductor. La invención se refiere adicionalmente a proteínas N-glicosiladas producidas con glicanos en un sistema de expresión y vacunas bioconjugadas producidas a partir de dichas proteínas N-glicosiladas que comprenden glicanos inmunogénicos, y proporciona procedimientos para la producción de proteínas N-glicosiladas.

Antecedentes de la invención

Las glicoproteínas son proteínas que tienen uno o más polímeros de azúcares unidos covalentemente. La glicosilación unida a N de proteínas es un proceso esencial y conservado que se produce en el retículo endoplásmico de organismos eucariotas. Es importante para el plegamiento proteico, oligomerización, estabilidad, control de calidad, clasificación y transporte de proteínas secretoras y de membrana (Helenius, A., y Aebi, M. (2004). Roles of N-linked glycans in the endoplasmic reticulum. Annu. Rev. Biochem. 73, 1019-1049).

La glicosilación tiene una profunda influencia sobre la inmunogenicidad, la estabilidad y la semivida de una proteína. Además, la glicosilación puede ayudar a la purificación de proteínas por cromatografía, por ejemplo, por cromatografía de afinidad con ligandos de lectina unidos a una fase sólida que interactúa con los restos glicosilados de la proteína. Por lo tanto es una práctica establecida producir muchas proteínas glicosiladas recombinantemente en células eucariotas para proporcionar patrones de glicosilación biológica y farmacéuticamente útiles.

El documento WO 2003/07467 (Aebi y col.) demostró que el agente patógeno *Campylobacter jejuni* transmitido por los alimentos, que es una bacteria, podía N-glicosilar sus proteínas, lo que era una característica única entre los organismos procariotas conocidos excepto ciertas especies de archaea. La maquinaria necesaria para la glicosilación está codificada por 12 genes que están agrupados en el denominado locus pgl. La destrucción de la N-glicosilación afecta a la invasión y patogénesis de *C. jejuni* pero no es letal como en la mayoría de los organismos eucariotas (Burda P. y M. Aebi, (1999). The dolichol pathway of N-linked glycosylation. Biochem Biophys Acta 1426(2):239-57). ES posible reconstituir la N-glicosilación de las proteínas de *C. jejuni* expresando recombinantemente el locus pgl y un receptor de glicoproteínas en *E. coli* al mismo tiempo (Wacker y col. (2002). N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science 298, 1790-1793).

Los N-glicanos tienen un glicano unido a una secuencia de consenso en una proteína. La secuencia de consenso de N-glicosilación conocida en una proteína permite la N-glicosilación de proteínas recombinantes diana en organismos procariotas. Dichos organismos comprenden una oligosacárido transferasa (“OT”; “OTasa”), tal como, por ejemplo, una oligosacaril transferasa de *C. jejuni*, que es una enzima que transfiere el glicano a una secuencia de consenso de la proteína.

El documento WO 2003/07467 (Aebi y col.) enseña sobre un organismo procariota en el que se introduce un ácido nucleico que codifica (i) glicosil transferasas específicas para el ensamblaje de un oligosacárido en un vehículo lipídico, (ii) una proteína recombinante diana que comprende una secuencia de consenso “N-X-S/T”, en la que X puede ser cualquier aminoácido excepto prolina, y (iii) una oligosacaril transferasa, tal como, por ejemplo, una oligosacaril transferasa de *C. jejuni* que une covalentemente dicho oligosacárido con la secuencia de consenso de la proteína diana. Dicho organismo procariota produce N-glicanos con una estructura específica que se define por el tipo de la glicosil transferasa específica.

El documento WO 2006/119987 (Aebi y col.) describe proteínas, así como medios y procedimientos para producir proteínas, con eficacia para la N-glicosilación en organismos procariotas *in vivo*. Se describe adicionalmente una introducción eficaz de N-glicanos en proteínas recombinantes para modificar la inmunogenicidad, estabilidad, actividad biológica, profiláctica y/o terapéutica de dichas proteínas, y la provisión de una célula huésped que presenta eficazmente proteínas N-glicosiladas recombinantes de la presente divulgación en su superficie. Además, describe una proteína N-glicosilada recombinante que comprende una o más de las siguientes secuencias de aminoácidos N-glicosiladas optimizadas:

D/E - X - N - Z - S/T (secuencia de consenso optimizada),

en la que X y Z puede ser cualquier aminoácido natural excepto Pro, y en la que se introduce al menos una de dichas secuencias de aminoácidos parcialmente N-glicosiladas. La introducción de secuencias de aminoácidos parciales específicas (secuencias de consenso optimizadas) en las proteínas dan lugar a proteínas que están N-glicosiladas eficazmente por una oligosacaril transferasa en estas posiciones introducidas.

la biosíntesis de diferentes polisacáridos está conservada en las células bacterianas. Los polisacáridos se ensamblan en vehículos lipídicos a partir de precursores comunes (nucleótidos de azúcares activados) en la membrana citoplasmática por diferentes glicosil transferasas con especificidad definida. Se proporcionan lipopolisacáridos ("LPS") solo en bacterias Gram-negativas, por ejemplo, *Shigella* spp., *Pseudomonas* spp. y *E. coli* (ExPEC, EHEC).

La síntesis de LPS comienza con la adición de un monosacárido a un vehículo lipídico de fosfato de undecaprenilo ("Und-P-P") en el lado citoplasmático de la membrana. El antígeno se construye mediante la adición secuencial de monosacáridos a partir de nucleótidos de azúcares activados por diferentes glicosil transferasas, y el polisacárido unido al lípido tiene un movimiento de difusión transversal a lo largo de la membrana mediante una flipasa. La unidad de repetición de antígeno se polimeriza mediante una reacción enzimática. El polisacárido se transfiere entonces al lípido A por una ligasa WaaL que forma el LPS que se exporta a la superficie, mientras que el polisacárido capsular se libera del vehículo lipídico después de la polimerización y se exporta a la superficie. La ruta biosintética de estos polisacáridos hace posible la producción de bioconjugados de LPS *in vivo*, la captura de polisacáridos en el periplasma en un vehículo proteico.

Dichos complejos de oligo o polisacáridos (es decir, restos de azúcares) y proteínas (es decir, vehículos proteicos) se pueden utilizar como vacunas conjugadas para proteger contra varias infecciones bacterianas. Se han utilizado satisfactoriamente vacunas conjugadas para proteger contra infecciones bacterianas. La conjugación de un polisacárido antigénico con un vehículo proteico es necesaria para la respuesta de memoria protectora, ya que los polisacáridos son inmunógenos independientes de linfocitos T. Los polisacáridos se han conjugado con vehículos proteicos mediante diferentes procedimientos químicos, utilizando grupos de activación reactivos en el polisacárido así como el vehículo proteico.

Las vacunas conjugadas se pueden administrar a niños para proteger contra infecciones bacterianas y también puede proporcionar una respuesta inmunitaria de larga duración a adultos. Se ha descubierto que las construcciones del documento WO 2009/104074 (Fernandez, y col.) generan una respuesta de IgG en animales. Se ha descubierto que una respuesta de IgG contra una vacuna de conjugados polisacárido-proteína específica de *Shigella* O en los seres humanos se correlaciona con una protección inmunitaria en seres humanos. (Passwell, J.H. y col., "Safety and Immunogenicity of Improved *Shigella* O-Specific Polysaccharide-Protein Conjugate Vaccines in Adults in Israel" Infection and Immunity, 69(3): 1351- 1357 (Mar. 2001)). Se cree que el polisacárido {es decir, los restos de azúcares} desencadenan una respuesta inmunitaria a corto plaza que es específica del azúcar. Además, el sistema inmunitario humano genera una fuerte respuesta contra polisacáridos específicos de las estructuras superficiales de bacterias, tales como los antígenos O y polisacáridos capsulares. Sin embargo, como la respuesta inmunitaria a los polisacáridos es dependiente de IgM, el sistema inmunitario no desarrolla memoria. El vehículo proteico que vehicula el polisacárido desencadena una respuesta de IgG que es dependiente de linfocitos T y que proporciona una protección de larga duración ya que el sistema inmunitario desarrolla memoria.

La *E. coli* 0157 es una cepa enterohemorrágica responsable de aproximadamente dos tercios de todos los casos recientes de síndrome hemolítico-urémico y plantea serios problemas en salud humana (Law, D. (2000) J. App. Microbiol., 88, 729-745; Wang, L., y Reeves, P. R. (1998) Infect. Immun. 66, 3545-3551).

La cepa 0157 de E. coli produce un antígeno O que contienen la unidad tetrasacarídica repetida (4-N-acetil perosamina →fucosa- glucosa→GalNAc) (a-D-PerNAc-cc-L-Fuc-P-D-Glc-cc-D-GalNAc) (Perry, M. B., MacLean, L. y Griffith, D. W. (1986) Biochem. Cell. Biol., 64, 21 -28). El tetrasacárido se pre-ensambla sobre pirofosfato de undecaprenilo. La envoltura celular de *E. coli* contiene una membrana plasmática interna, una capa de peptoglicano que soporta la tensión y una membrana externa asimétrica que consiste en una monocapa interna fosfolipídica y una monocapa externa compuesta por LPS bacteriano. El LPS contiene tres componentes, el ancla A lipídica, el centro que contiene ácido 2-desoxi-D-mano-oct-2-ulosónico, y la región del antígeno O (véase: Raetz, C. R. H. y Whitfield, C. (2002) Annu. Rev. Biochem., 71, 635-700; Whitfield, C. (2006) Ann. Rev. Biochem. 75, 39-68; Samuel, G. y Reeves, P. R. (2003) Carbohydrate Research, 338, 2503-2519; y las ref. en el mismo para las revisiones del ensamblaje de antígeno O de LPS bacterianos).

Los componentes del antígeno O del LPS bacteriano son polisacáridos extremadamente diversos grandes que pueden ser homopoliméricos, compuestos de un único monosacárido repetido, o heteropoliméricos, que contienen 10-30 repeticiones de 3-6 unidades de azúcar (Reeves, P. R., Hobbs, M.., Valvano, M. A., Skurnik, M., Whitfield, C, Coplin, D., ido, N., lena, J., Maskell, D., Raetz, C. R. H., y Rick, P. D. (1996) Trends Microbiol., 4, 495-503). Los antígenos O, por tanto, son la característica dominante de la superficie celular bacteriana y constituyen determinantes de virulencia y patogenicidad determinantes (Law, D. (2000) J. App. Microbiol, 88, 729-745; Spears, K. J., Roe, A. J. y Gaily, D. L. (2006) FEMS Microbiol. Lett., 255, 187-202; Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Wang, Q., Reeves, P. R. y Wang, L (2008) FEMS Microbiol. Rev. 32, 627-653; Stenutz, R., Weintraub, A. y Widmalm, G. (2006) FEMS Microbiol. Rev. 30, 382-403). Se habían identificado cepas de *E. coli*

con más de 190 serotipos O individuales, atribuidos a estructuras únicas de antígeno O (Stenutz, R., Weintraub, A. y Widmalm, G. (2006) FEMS Microbiol. Rev. 30, 382-403).

Las unidades repetidas de antígeno O se pre-ensamblan en la cara citosólica de la membrana interna unida a pirofosfato de undecaprenilo. Las unidades repetidas unidas a lípidos se difunden transversalmente (flip-flop) respecto a la superficie periplásmica de la membrana interna y se polimerizan antes del transporte a la membrana externa y la unión a LPS. La mayoría de las unidades de repetición del antígeno O heteropolimérico tienen N-acetilglucosamina ("GlcNAc") o N-acetilgalactosamina ("GalNAc") en el extremo reductor.

Se había asumido que la biosíntesis de los intermediarios lipídicos se inicia por la transferencia de GlcNAc-P o GalNAc-P desde sus derivados de nucleótidos de azúcar respectivos al undecaprenilo monofosfato ("Und-P") catalizado por WecA (Samuel, G. y Reeves, P. R. (2003) Carbohydrate Research, 338, 2503-2519; Alexander, D. C. y Valvano, M. A. (1994) J. Bacteriol, 176, 7079-7084; Zhang, L., Radziejewska-Lebrecht, J., Krajewska-Pietrasik, D., Tolvanen, P. y Skurkik, M. (1997) Mol. Microbiol. 23, 63-76; Amor, P. A. y Whitfield, C. (1997) Mol.

Microbiol. 26 (145-161); Wang, L. y Reeves, P. R. (1998) Infect. Immun. 66, 3545-3551). Aunque las propiedades y especificidad de la actividad GlcNAc-fosfotransferasa de WecA se había caracterizado (Rush, J. S., Rick, P. D. y Waechter, C. J. (1997) Glycobiology, 7, 315-322), la conclusión de que WecA cataliza la síntesis de GalNAc-P-P-Und se basó en estudios genéticos (Wang, L. y Reeves, P. R. (1998) Infect. Immun. 66, 3545-3551). Dichos estudios genéticos anteriores indicaban que la biosíntesis del intermediario tetrasacarídico unido a lípidos se inició mediante transferencia enzimática de GalNAc-P de UDP-GalNAc a Und-P catalizado por WecA (Wang, L. y Reeves, P. R. (1998) Infect. Immun. 66, 3545-3551). Sin embargo, no había pruebas enzimáticas directas que demostraran que la WecA utiliza UDP-GalNAc como donante de GalNAc-P.

Además, se propuso previamente que los genes gne y gneI de *E. coli* 055 codificaban una UDP-GlcNAc 4-epimerasa (Wang, L., Huskic, S., Cisterne, A., Rothemund, D. y Reeves, P.R. (2002) J. Bacterid. 184, 2620-2625; Guo, H., Yi, W., Li, L. y Wang, P. G. (2007) Biochem. Biophys. Res. Commun., 356, 604-609). Los informes previos identificaron dos genes de *E. coli* 055 (Wang, L., Huskic, S., Cisterne, A., Rothemund, D. y Reeves, P.R. (2002) J. Bacterid. 184, 2620-2625) y *E. coli* 086 (Guo, H., Yi, W., Li, L. y Wang, P. G. (2007) Biochem. Biophys. Res. Commun., 356, 604-609), el gne de *E. coli* 055 y gneI de *E. coli* 086, respectivamente, que son un 100 % idénticos al gen Z3206 de la misma familia genética.

En consecuencia, un experto habría dado lugar a la creencia de que el gen Z3206 también codifica una UDP-GlcNAc/UDP-GalNAc epimerasa.

Feldman y col Proc. Natl. Acad. Sci US (2005) vol. 102, nº 8 páginas 3016-3021 desvela la modificación de glicosilación proteica unida a N con estructuras lipopolisacarídicas de antígeno O diversas. Rush y col J. Biol. Chem. (2009) vol. 285, nº 3, páginas 1671-1680 desvela una nueva epimerasa que convierte el GlcNAc-P-P-undecaprenol en GalNAc-P-P-undecaprenol en la *E. coli* 0157.

Breve sumario de la invención

Se ha descubierto ahora sorprendentemente que una epimerasa codificada por el gen Z3206 en *E. coli* 0157 cataliza una reacción que sintetiza N-acetilgalactosamina ("GalNAc") en pirofosfato de undecaprenilo, que inicia la formación de un oligo o polisacárido.

1. En un aspecto, la presente invención se refiere a un organismo procariota recombinante que es una *E. coli* que comprende ácidos nucleicos que codifican:

i) una epimerasa que sintetiza N-acetilgalactosamina sobre pirofosfato de undecaprenilo, en el que dicho ácido nucleico que codifica dicha epimerasa es al menos un 95 % idéntica a la SEQ ID NO: 1;

ii) glicosil transferasas que ensamblan un oligo o polisacárido en un vehículo lipídico;

iii) una oligosacárido transferasa de *Campylobacter jejuni*; y

iv) una proteína que comprende una o más secuencias de consenso introducidas recombinantemente, D/E-X-N-Z-S/T, en la que X y Z son independientemente cualquier aminoácido natural excepto prolina

en el que el oligosacárido ensamblado por dichas glicosil transferasas se pueden transferir desde el vehículo lipídico a la proteína diana que tiene una o más secuencias de consenso optimizadas por la oligosacaril transferasa.
La divulgación incluye adicionalmente glicosil transferasas que sintetizan todo o una parte de un polisacárido que tiene GalNAc en el extremo reductor, y que además incluye glicosil transferasas que sintetiza todo o una parte de un polisacárido antigénico que tiene GalNAc en el extremo reductor.

En otro aspecto, la divulgación se refiere a una epimerasa que produce GalNAc sobre pirofosfato de undecaprenilo, y, en un aspecto adicional, la epimerasa se codifica por el gen Z3206.

En un aspecto adicional, la presente divulgación se refiere a un sistema de expresión para producir una proteína N-glicosilada que comprende: una secuencia de nucleótidos que codifica una oligosacaril transferasa; una secuencia de nucleótidos que codifica un vehículo proteico; al menos un agrupamiento genético de oligo o polisacárido, en el que el polisacárido contiene GalNAc en el extremo reductor; y una secuencia de ácido nucleico que codifica una epimerasa.

En un aspecto adicional más, la presente divulgación se dirige a un sistema biosintético en procariotas que comprende el gen Z3206 que codifica una epimerasa que convierte GlcNAc-P-P-Und en GalNAc -P-P-Und.

En otro aspecto adicional más, la presente divulgación se refiere a un sistema biosintético procariota que comprende el gen gne de *E. coli* 055 o el gen gne J de *E. coli* 086 que codifica una epimerasa que convierte GlcNAc-P-P-Und en GalNAc -P-P-Und.

En otro aspecto más, la presente divulgación se refiere a una proteína N-glicosilada que comprende al menos una secuencia de consenso D/E-X-N-Z-S/T introducida, en laque X y Z pueden ser cualquier aminoácido natural excepto prolina, y un glicano que tiene N-acetilgalactosamina en el extremo reductor.

En otro aspecto más, la presente divulgación se refiere a una vacuna bioconjugada que comprende una proteína N-glicosilada que tiene al menos una secuencia de consenso D/E-X-N-Z-S/T introducida, en la que X y Z pueden ser cualquier aminoácido natural excepto prolina; un glicano inmunogénico que tiene N-acetilgalactosamina en el extremo reductor; y un adyuvante.

En un aspecto adicional, la invención se refiere a un procedimiento de producción de una proteína glicosilada unida a N, que comprende:

a.) la introducción en un organismo huésped *E. coli* de ácidos nucleicos que codifican:

i) una epimerasa que sintetiza N-acetilgalactosamina sobre pirofosfato de undecaprenilo en el que dicho ácido nucleico que codifica dicha epimerasa es al menos un 95 % idéntica a la SEQ ID NO: 1;

ii) glicosil transferasas que ensamblan un oigo o polisacárido en un vehículo lipídico;

iii) una oligosacaril transferasa de *Campylobacter jejuni*; y

iv) una proteína que comprende una o más de una secuencia de consenso D/E-X-N-Z-S/T introducida, en la que X y Z puede ser cualquier aminoácido natural excepto prolina; y

b.) cultivar dicho organismo huésped hasta que se produzca al menos una proteína N-glicosilada.

En un aspecto adicional, la presente divulgación se refiere al uso de un sistema biosintético y las proteínas para preparar una vacuna bioconjugada.

En un aspecto adicional, la presente divulgación se refiere a procedimientos para la producción de mono, oligo, y polisacáridos, y en un aspecto adicional más la divulgación se refiere a procedimientos de producción de glicanos antigénicos y proteína N-glicosiladas.

Breve descripción de los dibujos

La FIG. 1 muestra el transcurso de tiempo de la síntesis de [^{3}H]GlcNAc/GalNAc-P-P-Und por las fracciones de membrana de *E. coli* 0157. La fracción de membrana de la cepa 0157 de *E. coli* se incubó con UDP-[^{3}H]GlcNAc durante los tiempos indicados a 37 ºC. Los productos [^{3}H]lípidos se extrajeron y se ensayó la incorporación de [^{3}H]GlcNAc en [^{3}H]GlcNAc-P-P-Und (O) y [^{3}H]GalNAc-P-P-Und (·) como se describe en el Ejemplo 2.

La FIG. 2 muestra la ruta biosintética propuesta para la formación de GalNAc-P-P-Und a partir de GlcNAc-P-P-Und.

La FIG. 3 muestra la purificación y caracterización de [^{3}H] GalNAc-P-P-Und sintetizado por fracciones de membrana de la cepa 0157 de *E. coli*. Las fracciones de membrana de *E. coli* 0157 se incubaron con UDP- [^{3}H] GlcNAc y los lípidos [^{3}H] GalNAc se purificaron como se describe en el Ejemplo 3. FIG. 3A, se muestra un cromatograma de capa fina para preparar lípidos [H]HexNAc en gel G (Quantum 1) de sílice impregnado con borato después de la purificación en DEAE-celulosa. FIG. 3B, se muestra la cromatografía de capa fina de [^{3}H] GalNAc-P-P-Und purificado en gel G de sílice impregnado con borato (Baker, Si250) después de la recuperación de la placa de preparación en el panel A. FIG. 3C, se muestra el cromatograma en papel en disminución (papel Nº 1 Whatman impregnado con borato) de la [H]-amino azúcar recuperada después de hidrólisis ácida moderada de [H] GalNAc-P-P-Und purificado en la FIG. 3B. FIG. 3D, cromatograma en papel en disminución (Whatman Nº 3 MM) del [^{3}H]HexNAc-alditol producido por reducción de la [^{3}H] amino azúcar de la FIG. 3C con NaBFU.

La FIG. 4 muestra el marcado metabólico de células de *E. coli* 21546 y células de *E. coli* 21546 después de la transformación con pMLBAD: Z3206. Se marcaron metabólicamente la *E. coli* 21546 (FIG. 4A) y *E. coli* 21546:

pMLBAD/Z3206 (FIG. 4B) con [^{3}H] GlcNAc durante 5 min a 37 ºC. Se extrajo [^{3}H] GlcNAc/ GalNAc-P-P-Und, se liberaron de contaminantes hidrosolubles y se separaron por cromatografía de capa fina sobre placas de gel de sílice impregnadas de borato (Baker Si250) como se describe en el Ejemplo 3. Los lípidos radioactivos se detectaron utilizando un cromatoescáner Bioscan. Las posiciones cromatográficas de GalNAc-P-P-Und y GlcNAc-P-P-Und se indican con flechas.

La FIG. 5 muestra una cromatografía de capa fina de [^{3}H] GlcNAc/ GalNAc-P-P-Und formado por incubación de fracciones de membrana de las cepas de *E. coli* con UDP-[^{3}H] GlcNAc. Las fracciones de membrana de las cepas K12 (FIG. 5A), 0157 (FIG. 5B), 21546 (FIG. 5C) y 21546:pMLBAD/Z3206 (FIG. 5D) de *E. coli* se incubaron con UDP-[^{3}H] GlcNAc durante 10 min a 37 ºC, y los productos [^{3}H]lípidos se extrajeron, se liberaron de contaminantes hidrosolubles por partición, y se separaron por cromatografía de capa fina sobre placas de gel de sílice impregnado de borato (Baker si250) como se describe en el Ejemplo 3. Las posiciones cromatográficas de GalNAc-P-P-Und y GlcNAc-P-P-Und se indican con flechas.

La FIG. 6 muestra la descarga de GlcNAc-P por incubación con UMP. Las fracciones de membrana de *E. coli* 21546:Z3206 se pre-incubaron con UDP-[^{3}H] GlcNAc para el marcado enzimático de GlcNAc-P-P-Und durante 10 min (FIG. 6A) a 37 ºC seguido por un segundo periodo de incubación con 1 mM de UMP incluido durante 1 min (FIG. 6B) o 2 min (FIG. 6C). Después de los periodos de incubación indicados se extrajeron los [^{3}H] GlcNAc/ GalNAc-P-P-Und y se resolvieron por cromatografía de capa fina en placas de gel de sílice impregnado con borato (Baker Si250) como se describe en el Ejemplo 3. Las posiciones cromatográficas de GalNAc-P-P-Und y GlcNAc-P-P-Und se indican con flechas.

La FIG. 7 muestra la conversión de [^{3}H] GlcNAc-P-P-Und y [^{3}H] GalNAc-P-P-Und en el producto [^{3}H] HexNAc-P-P-Und pertinente catalizado por membranas de la cepa 21546 que expresa Z3206. Las fracciones de membrana a partir de la cepa 21546 (FIG. 7B, FIG. 7E) y 21546: pMLBAD/Z3206 (FIG. 7C y FIG. 7F) de *E. coli* se incubaron con [^{3}H] GlcNAc-P-P-Und (FIG. 7A, FIG. 7B y FIG. 7C) o [^{3}H] GalNAc-P-P-Und purificadas (paneles de la FIG. 7D, FIG. 7E, y FIG. 7F) (dispersados ultrasónicamente en un 1 % de Triton-X100) durante 1 min a 37 ºC. Se extrajeron [^{3}H] GlcNAc/ GalNAc-P-P-Und, se resolvieron por cromatografía de capa fina sobre placas de gel de sílice impregnado de borato (Baker Si250) y se detectaron con un radiocromatoescáner Bioscan AR2000 como se describe en el Ejemplo 3.

La FIG. 8 muestra el análisis SDS-PAGE de proteína AcrA glicosilada y no glicosilada. Se separaron extractos periplásmicos preparados a partir de células de *E. coli* DH5a que albergan el plásmido de expresión AcrA y el operón pgl Agne complementado con pML-BAD:Z3206 (calle 1), pMLBAD:gne (calle 2) o el vector de control pMLBAD (calle 3) mediante SDS-PAGE 10 % y se transfirieron a membranas de celulosa. AcrA y sus formas glicosiladas se detectaron con antisuero anti AcrA. Se indica la posición correspondiente a las bandas de AcrA no glicosilada (AcrA) y glicosilada (gAcra).

La FIG. 9 muestra los genes que se habían identificado por Liu B y col. {Structure and genetics of *Shigella* O antigens FEMS Microbiology Review, 2008. 32: p. 27).

La FIG. 10 es un esquema que muestra la región de ADN que contiene los genes necesarios para sintetizar en antígeno O de *S. flexneri* 6.

La FIG. 11 muestra la expresión del antígeno O de *S. flexneri* 6 en *E. coli*. Se visualizó el LPS por tinción con plata o por trasferencia a membranas de nitrocelulosa y detección de anticuerpos dirigidos contra *S. flexneri* 6.

La FIG. 12 muestra la HPLC del antígeno O. Análisis LLO de células de *E. coli* (SCM3) que no contienen el Z3206 (-) de *S. flexneri*, células de *E. coli* (SCM3) que contienen Z3206 (+) o células de *E. coli* (SCM3) vacías.

La FIG. 13 muestra la trasferencia de Western de proteínas purificadas con níquel de células de *E. coli* que expresan EPA, pglB, y antígeno o de *S. flexneri* Z3206 (+/-).

Descripción detallada de la invención

La presente divulgación engloba un sistema biosintético procariota recombinante que comprende ácidos nucleicos que codifican una epimerasa que sintetiza un oligo o polisacárido que tiene una N-acetilgalactosamina en el extremo reductor, y proteínas N-glicosiladas que tienen N-acetilgalactosamina en el término reductor del glicano.

La expresión “secuencia(s) parcial de aminoácidos” también se hace referencia como “secuencia(s) de consenso optimizada” o “secuencia(s) de consenso”. La secuencia de consenso optimizada se N-glicosila mediante una oligosacaril transferasa (“OST”, “OTasa”), mucho más eficazmente que la secuencia de consenso regular “N-X-S/T”.

En general, la expresión “proteína N-glicosilada recombinante” se refiere a cualquier poli u oligopéptido producido en una célula huésped que no comprende de manera natural el ácido nucleico que codifica dicha proteína. En el contexto de la presente invención, esa expresión se refiere a una proteína producida recombinantemente en una célula huésped procariota, por ejemplo, *Escherichia* spp. , *Campylobacter* spp., *Salmonella* spp., *Shigella* spp., *Helicobacter* spp., *Pseudomonas* spp., *Bacillus* spp., y en realizaciones adicionales *Escherichia coli*, *Campylobacter*

jejuni, *Salmonella typhimurium*, etc., en el que el ácido nucleico que codifica dicha proteína se ha introducido en dicha célula huésped y en la que la proteína codificada está N-glicosilada por la OTasa, siendo dicha enzima transferasa de origen natural o siendo introducida recombinantemente en dicha célula huésped.

De acuerdo con el código de una letra aceptada internacionalmente para los aminoácidos, las abreviaturas D, E, N, S y T denotan ácido aspártico, ácido glutámico, asparagina, serina, y treonina, respectivamente.

Las proteínas de acuerdo con la divulgación comprenden una o más de las secuencias de consenso optimizadas (D/E-X-N-Z-S/T que se introducen en la proteína y se N-glicosilan. Por lo tanto, las proteínas de la presente divulgación se diferencian de las N-glicoproteínas de *C. jejuni* de origen natural que también contienen la secuencia de consenso optimizada pero no comprende ninguna secuencia de consenso optimizada adicional (introducida).

La introducción de la secuencia de consenso optimizada se puede conseguir mediante la adición, eliminación y/o sustitución de uno o más aminoácidos. La adición, eliminación y/o sustitución de uno o más aminoácidos con el fin de introducir la secuencia de consenso optimizada se puede conseguir por estrategias sintéticas químicas, en las que, en vista de la presente divulgación, serían bien conocidas por los expertos en la técnica tales como la síntesis peptídica química asistida por fase sólida. De manera alternativa, y preferido para polipéptidos más grandes, las proteínas de la presente invención se pueden preparar por técnicas recombinantes que serían técnicas convencionales en la técnica a la luz de la divulgación.

Las proteínas de la presente divulgación tienen la ventaja de que se pueden producir con alta eficacia y en cualquier huésped. En una realización de la divulgación, el huésped comprende un operón pgl funcional de Campylobacter spp., por ejemplo, de *C. jejuni*. En realizaciones adicionales, las oligosacaril transferasas de Campylobacter spp. para la práctica de la divulgación son de Campylobacter coli o Campylobacter lari. En vista de la divulgación, las oligosacaril transferasas serían evidentes para un experto en la técnica. Por ejemplo, se desvelan oligosacaril transferasas en las referencias tales como Szymanski, CM. y Wren, B.W. (2005) Protein glycosylation in bacterial mucosal pathogens, Nat. Rev. Microbiol. 3:225-237. El operón pgl funcional puede estar presente naturalmente cuando dicho huésped procariota es Campylobacter spp., o, por ejemplo *C. jejuni*. Sin embargo como se ha demostrado anteriormente en la técnica y se ha mencionado anteriormente, el operón pgl se puede transferir en las células y mantenerse funcional en dicho nuevo ambiente celular.

La expresión “operón pgl funcional de Campylobacter spp., preferentemente *C. jejuni*” significa que se refiere al agrupamiento de ácidos nucleicos que codifican la oligosacaril transferasa (OTasa) de Campylobacter spp., por ejemplo, *C. jejuni*, y una o más glicosil transferasa específicas capaces de ensamblar un oligosacárido en un vehículo lipídico, y en el que dicho oligosacárido se pueden transferir desde el vehículo lipídico a la proteína diana que tiene una o más secuencias de aminoácidos optimizadas. D/E-X-N-Z-S/T por la OTasa. Se tiene que entender que la expresión “operón pgl funcional de Campylobacter spp., preferentemente de *C. jejuni*” en el contexto de la presente invención no se refiere necesariamente a un operón como una unidad transcripcional singular. La expresión simplemente necesita la presencia de los componentes funcionales para la N-glicosilación de la proteína recombinante en una célula huésped. Estos componentes se pueden transcribir como un ARNm o más separados y se pueden regular juntos o por separado. Por ejemplo, la expresión también engloba componentes funcionales posicionados en el ADN genómico y plasmídico en una célula huésped. Con el fin de eficacia, en una realización, todos los componentes del operón pgl funcional se regulan y expresan simultáneamente.

La oligosacaril transferasa puede originar, en algunas realizaciones, de Campylobacter spp., y en otras realizaciones, a partir de *C. jejuni*. En realizaciones adicionales, la oligosacaril transferasas se puede originar de otros organismos que son conocidos por los expertos en la técnica y tienen una oligosacaril transferasa, tal como, por ejemplo, *Woloinella* spp. y organismos eucariotas.

La una o más glicosil transferasas específicas capaces de ensamblar un oligosacárido de un vehículo lipídico se puede originar de la célula huésped o se introduce recombinantemente en dicha célula huésped, siendo la única limitación funcional que el oligosacárido ensamblado por dichas glicosil transferasas se pueden transferir desde el vehículo lipídico a la proteína diana que tiene una o más secuencias de consenso optimizadas por la OTasa. Por lo tanto, la selección de la célula huésped que comprende glicosil transferasa especificas naturalmente y/o sustituyendo las glicosil transferasas presentes naturalmente en dicho huésped así como la introducción de glicosil transferasas heterólogas que hacen posible a los expertos en la técnica a variar los N-glicanos unidos al sitio de consenso de N-glicosilación optimizado en las proteínas de la presente divulgación.

Como resultado de lo anterior, la presente divulgación proporciona el diseño individual de patrones de N-glicanos en las proteínas de la presente divulgación. Las proteínas pueden, por lo tanto, individualizarse en su patrón de N-glicanos para ajustarse a las necesidades biológicas, farmacéuticas y de purificación.

En realizaciones de la presente divulgación, las proteínas pueden comprender una pero también más de una, tal como al menos dos, al menos 3 o al menos 5 de dichas secuencias de aminoácido optimizadas N-glicosiladas.

La presencia de una o más secuencias de aminoácidos optimizadas N-glicosiladas en las proteínas de la presente divulgación puede ser una ventaja por el aumento de su inmunogenicidad, aumento de su estabilizar, afectar su actividad biológica, prologar su semivida biológica y/o simplificar su purificación.

La secuencia de consenso optimizada puede incluir cualquier aminoácido excepto prolina en las posiciones X y Z. La expresión “cualquier aminoácido” quiere decir que engloba aminoácidos comunes y raros así como a derivados sintéticos de aminoácidos y análogos que seguirán permitiendo que la secuencia de consenso optimizada se N-glicosile por la OTasa. Se prefieren aminoácidos de origen natural comunes y raros para X y Z. X y Z pueden ser el mismo o diferentes.

Se señala que X y Z pueden ser diferentes en cada una de las secuencias de consenso optimizada en una proteína de acuerdo con la presente divulgación.

El N-glicano unido de la secuencia de consenso optimizada se determinará por la glicosil transferasa específica y su interacción cuando se ensambla el oligosacárido de un vehículo lipídico por transferencia mediante la OTasa. En vista de la presente divulgación, los expertos en la técnica serán capaces de diseñar un N-glicano variando el tipo y cantidad de las glicosil transferasas específicas presentes en la célula huésped deseada.

“Monosacárido” como se utiliza en el presente documento se refiere a un resto de azúcar. “Oligo y polisacárido” se refiere a dos o más restos de azúcar. El término “glicanos” como se utiliza en el presente documento se refiere a mono, oligo o polisacáridos. “N-glicanos” se define en el presente documento como mono, oligo o polisacáridos de composiciones variables que están unidas a un nitrógeno de ε-amida de un resto de asparagina en una proteína mediante una unión N-glicosídica. En una realización, los N-glicanos transferidos por una OTasa se ensamblan sobre un ancla lipídica de pirofosfato de undecaprenol (“Und-P-P”) que está presente en la membrana citoplasmática de una bacteria Gram-negativa o positiva. Están implicados en la síntesis del antígeno o, polisacárido y peptidoglicano O (Bugg, T. D., y Brandish, P. E. (1994). From peptidoglycan to glycoproteins: common features of lipid-linked oligosaccharide biosynthesis. FEMS Microbiol Lett 119, 255-262; Valvano, M. A. (2003). Export of O-specific lipopolysaccharide. Front Biosci 8, S452-471).

Se llevaron a cabo estudios para determinar si la biosíntesis de un tetrasacárido repetido unido a un lípido (4-N-acetil perosamina →fucosa→glucosa-GalNAc) se iniciaba por la formación de GalNAc-P-P-Und por WecA. Cuando las fracciones de membrana de las cepas K12, 0157, y PR4019, una cepa que sobre-expresa WecA de *E. coli*, se incubaban con UDP-[^{3}H] GalNAc, no se detectaba ninguna síntesis enzimática de [^{3}H] GlcNAc-P-P-Und ni [^{3}H] GalNAc-P-P-Und. Sin embargo, cuando se incubaban fracciones de membrana de la cepa 0157 con UDP-[^{3}H] GlcNAc, se observaban dos productos marcados enzimáticamente con las propiedades químicas y cromatográficas de [^{3}H] GlcNAc-P-P-Und y [^{3}H] GalNAc-P-P-Und, confirmando que la cepa 0157 contenía una epimerasa capaz de la interconversión de GlcNAc-P-P-Und y GalNAc-P-P-Und. La presencia de una epimerasa se confirmó también mostrando que la [H] GlcNAc-P-P-Und exógena se convertía en [^{3}H] GalNAc-P-P-Und cuando se incubaban con membranas de la cepa 0157. Cuando la cepa 0157 se marcaba metabólicamente con [^{3}H] GlcNAc, se detectaban tanto [^{3}H] GlcNAc-P-P-Und como [^{3}H] GalNAc-P-P-Und. La transformación de la cepa 21546 de *E. coli* con el gen Z3206 hacía posible que estas células sintetizaran GalNAc-P-P-Und *in vivo* e *in vitro*. La reversibilidad de la reacción de epimerasas demostró mostrando que [^{3}H] GlcNAc-P-P-Und se reformaba cuando las membranas de la cepa 0157 se incubaba [^{3}H] GalNAc-P-P-Und exógena. La incapacidad de Z3206 para complementar la pérdida del en gne en la expresión del sistema de N-glicosilación de *Campylobacter jejuni* en *E. coli* indicaba que no funciona como una epimerasa de UDP-GlcNAc/ UDP-GalNAc. Basándose en estos resultados, se confirmó que la síntesis de GalNAc-P-P-Und se sintetiza de manera reversible por un GlcNAc-P-P-Und epimerasa después de la formación de GlcNAc-P-P-Und por WecA en *E. coli* 0157.

La reacción de inicio del ensamblaje de la subunidad del antígeno O en *E. coli* 0157 se investigó para confirmar que la síntesis de GalNAc-P-P-Und se cataliza por algunos mecanismos previamente desconocidos más que por WecA. Las pruebas presentadas en el presente documento demuestran que GalNAc-P-P-Und no se sintetiza por transferencia de GalNAc-P a partir de UDP-GalNAc catalizada por WecA sino más bien por la epimerización reversible del 4-OH de GlcNAc-P-P-Und catalizada por una epimerasa codificada por el gen Z3206 en *E. coli*.

En consecuencia, la divulgación engloba una nueva ruta biosintética para el ensamblaje de un componente importante de la superficie celular bacteriana así como una nueva ruta biosintética para la síntesis de GalNAc-P-P-Und. Una realización adicional de la divulgación incluye la epimerasa bacteriana como una nueva diana para los agentes antimicrobianos.

La *E. coli* 0157 sintetiza un antígeno O con la estructura tetrasacarídica repetida (4-N-acetil-perosamina-fucosa→glucosa→GalNAc). Se demuestra en el presente documento que la biosíntesis del intermediario tetrasacárido unido a lípido no se iniciaba por la transferencia enzimática de GalNAc-P a partir de UDP-GalNAc a Und-P catalizado por WecA, contrariamente a los estudios genéticos anteriores (Wang, L. y Reeves, P. R. (1998) Infect. Immun. 66, 3545-3551). La divulgación descrita en el presente documento, obtenida por búsquedas de homología y confirmada después por experimentos genéticos, enzimológicos, y experimentos de marcado metabólico, demuestran que la WecA no utiliza el UDP-GalNAc como sustrato, sino que la WecA es necesaria para sintetizar GlcNAc-P-P-Und que entonces se convierte reversiblemente en GalNAc-P-P-Und por una epimerasa codificada por el gen Z3206 de la cepa 0157.

El gen Z3206 de la presente divulgación pertenece a una familia de genes presentes en varias cepas que producen unidades repetidas del antígeno O de superficie que contienen restos GalNAc en sus extremos reductores (Tabla 1).

La secuencia de gel Z3206 se muestra en la SEQ ID NO: 1. Los informes previos identificaban que dos genes de *E. coli* 055 (Wang, L., Huskic, S., Cisterne, A., Rothemund, D. y Reeves, P.R. (2002) J. Bacteriol. 184, 2620-2625) y *E. coli* 086 (Guo, H., Yi, W., Li, L. y Wang, P. G. (2007) Biochem. Biophys. Res. Comm., 356, 604-609), el gen gne de *E. coli* 055 y el gen gneI de *E. coli* 086, respectivamente, que son un 100 % idénticos al gen Z3206 (Tabla 1). La secuencia del gen gne de *E. coli* 055 se muestra como la SEQ ID NO: 3, y la secuencia del gen gneI de *E. coli* 086 se muestra como la SEQ ID NO: 5.

TABLA 1

Correlación del gen Z3206 en cepas bacterianas que expresan cadenas del antígeno O con GalNAc en los extremos reductores		
	% de identidad con Z3206	GalNAc en el extremo reductor de la unidad de repetición del antígeno O
gne de *E. coli* 055 (SEQ ID NO: 3)	100	Sí
gneI de E. coli 086 (SEQ ID NO: 5)	100	Sí
gne de Shigella boydii 018 (SEQ ID NO: 7)	88	Sí
gne de *Salmonella enterica* 030 (SEQ ID NO: 9)	94	Sí
gne de *C. jejuni* (SEQ ID NO: 11)	21	No
galE de *E. coli* K12 (SEQ ID NO: 13)	27	No
gne2 de *E. coli* 086 (SEQ ID NO: 15)	18	Sí

En consecuencia, los inventores concluyen que el gne de *E. coli* 055 y gen I de *E. coli* 086 también codifican epimerasas capaces de convertir el GlcNAc-P-P-Und en GalNAc-P-P-Und en las cepas 055 y 086, respectivamente, que también producen unidades de repetición del antígeno O con GalNAc en los extremos reductores.

Dos estrategias experimentales en el presente estudio indican que la proteína Z3206 no cataliza la epimerización de UDP-GlcNAc a UDP-GalNAc en la cepa 0157. Primero, cuando las membranas de la cepa 0157 se incubaron con [^{3}H]UDP-GalNAc, no se detectaban ni [^{3}H] GlcNAc-P-P-Und ni [^{3}H] GalNAc-P-P-Und (Tabla 3). Si Z3206 catalizara la conversión de [^{3}H]UDP-GalNAc a [^{3}H]UDP-GlcNAc, se esperaría que se debería observar [^{3}H] GlcNAc-P-P-Und.

Segundo, los inventores han demostrado que la Z3206 marcada con hemaglutinina era incapaz de complementar el sistema indicador de N-glicosilación en *C. jejuni* dependiente de UDP-GalNAc (FIG. 8).

El gen gne de *E. coli* 055 de la cepa 055 (Wang, L., Huskic, S., Cisterne, A., Rothemund, D. y Reeves, P.R. (2002) J. Bacteriol. 184, 2620-2625) también se ensayó en cuanto a la actividad epimerasa incubando extractos en bruto con UDP-GalNAc y ensayando indirectamente la conversión a UDP-GlcNAc midiendo un aumento de la reactividad con p-dimetil-aminobenzaldehído después de la hidrólisis ácida. En ambos estudios, la formación del producto se basaba en los cambios de la reactividad con p-dimetilaminobenzaldehído, y sin caracterización definitiva del producto final azúcar nucleotídico. También se demostró que el 90 % del gne I de *E. coli* 086 marcado con polihistidina puro tenía un bajo nivel de actividad UDP-glucosa epimerasa con respecto a gne 2 en un ensayo acoplado.

En consecuencia, una realización de la invención se refiere a un sistema biosintético procariota que contiene el gen Z3206, el gen gne de *E. coli* 055 o el gen gne I de *E. coli* 086 que convierte el GlcNAc-P-P-Und en GalNAc-P-P-Und.

Es significativo que la *E. coli* 086, que sintetiza un antígeno O que contiene restos de GalNAc, que presumiblemente necesitaría UDP-GalNAc como donante glicosílico para el extremo GalNAc no reductor adicional, también posee un gen GlcNAc 4-epimerasa, llamado gne 2, en el agrupamiento genético del antígeno O (Guo, H., Yi, W., Li, L. y Wang, P. G. (2007) Biochem. Biophys. Res. Commun., 356, 604-609). Este gen de epimerasa adicional tiene una alta homología con el gen GalE del agrupamiento genético del ácido colánico y parece que es una UDP-GlcNAc 4-epimerasa capaz de sintetizar UDP-GalNAc.

El gen Z3206 parece que está altamente conservado en los serotipos O de *E. coli* iniciados con GalNAc. En un recientes estudio, se exploraron 62 cepas de *E. coli*, con estructuras de unidades repetidas de antígenos O establecidos, en cuanto a la expresión de Z3206 mediante un procedimiento basado en una reacción en cadena de la polimerasa utilizando cebadores de nucleótido diseñados para detectar específicamente el gen Z3206 de *E. coli* 0157 (Wang, L., Huskic, S., Cisterne, A., Rothemund, D. y Reeves, P.R. (2002) J. Bacteriol. 184, 2620-2625). En este estudio se detectó el Z3206 en 16 de las 22 cepas de *E. coli* que se sabía que contenían GalNAc, y en solo 4 de las 40 cepas que carecían de GalNAc. Además, una exploración similar de 22 cepas que contenían GalNAc con cebadores diseñados para detectar una epimerasa alternativa con actividad UDP-GlcNAc 4-epimerasa (el gen GalE de *E. coli* 0113) no detectaba cepas que albergaran este gen, indicando que el Z3206 es el gen de la GlcNAc 4-

epimerasa asociado más comúnmente con la presencia de un GalNAc en el extremo reductor en las unidades repetidas del antígeno O de *E. coli*.

El análisis de la secuencia proteica de Z3206 por una variedad de algoritmos de predicción topológica basada en internet indica que la proteína Z3206 no es altamente hidrófoba. La mayoría de los algoritmos de predicción topológica indican que Z3206 es una proteína soluble de 37 kDa, aunque TMPred (Hofmann, K., y Stoffel, W. (1993) Biol. Chem. Hoppe-Seyler 374, 166 (extr.)) predecía una hélice transmembrana en el extremo N única débil. Sin embargo, la transferencia de Western después de la SDS-PAGE de fracciones celulares de las células de *E. coli* que expresaban Z3206 marcada con hemaglutinina muestra claramente que la proteína marcada se asocia con la fracción particulada después de la lisis hipotónica de las células. Los experimentos preliminares demostraban que la proteína se mantiene asociada con la fracción particulada después de la incubación de la fracción de la membrana con KCl 1 M, pero se solubiliza en una forma activa por incubación con un 0,1 % de Triton X-100.

El Z3206 de *E. coli* 157 tiene una homología de secuencia significativa con la familia de deshidrogenasas/reductasas de cadena corta de las oxido-reductasas incluyendo el motivo GXXGXXG (plegamiento de Rossman), que coincida con el bolsillo de unión al NAD(P) (Allard, S. T. M., Giraud, M. F., y Naismith, J. H. (2001) Cell. Mol. Life Sci. 58, 1650-1655) y la secuencia conservada S3/44Y^{a}, implicada en la abstracción y donación de protones (Field, R. A. y Naismith, J. H. (2003) Biochemistry 42, 7637-7647). El modelo molecular basado en las estructuras cristalinas de la UDP-Glc 4-epimerasa, otro miembro de la familia de deshidrogenasas/reductasas de cadena corta, sugiere que después de la abstracción de hidrido, el intermediario 4-ceto rota alrededor del fosfato β de UDP para presentar la cara opuesta del intermediario ceto y permitir la re-inserción del hidrido del lado opuesto, invirtiendo de esta manera la configuración del hidroxilo en el carbono 4. La presencia de estas secuencias conservadas sugiere que la Z3206 funciona probablemente mediante un mecanismo similar. Aunque la distribución en equilibrio de los productos de la epimerasa, que se ven en la FIG. 7, parece que favorecen la formación de GlcNAc-P-P-Und, la utilización de GalNAc-P-P-Und para el ensamblaje de la unidad repetida del antígeno O dirigiría la reacción de epimerización en la dirección de GalNAc-P-P-Und por acción masiva.

La epimerización de los restos glicosilo de intermediarios lipídicos poli-isoprenoides no se ha informado ampliamente en la naturaleza. En un estudio previo, la 2-epimerización del ribosil-P-decaprenol para formar arabinosil-P-decaprenol, se informó de un donante de arabinosilo en la síntesis de arabinogalactano en micobacterias (Mikusova, K., Huang, H., Yagi, T., Holsters, M., Vereecke, D., D'Haeze, W., Scherman, M. S., Brennan, P. J., McNeil, M. R., y Crick, D. C. (2005) J. Bacteriol. 187, 8020-8025). El arabinosil-P-decaprenol se forma mediante una reacción de oxidación/reducción en dos etapas que necesita dos proteínas micobacterianas Rv3790 y Rv3791. Aunque la epimerización se estimulaba modestamente por adición de NAD y NADP, ni la Rv3790 ni Rv3791 contenían el plegamiento de Rossman o el motivo S 24YXXXK, característicos de la familia de deshidrogenasas/reductasas de cadena corta (Allard, S. T. M., Giraud, M.-F. y Naismith, J. H. (2001) Cell. Mol. Life Sci. 58, 1650-1655; Field, R. A. y Naismith, J. H. (2003) Biochemistry 42, 7637-7647).

En resumen, se describe una nueva ruta biosintética para la formación de GalNAc-P-P-Und mediante la epimerización de GlcNAc-P-P-Und.

Se ha demostrado que varios antibióticos inhiben la síntesis de GlcNAc-P-P-Und, pero tiene una utilidad limitada debido a que también bloquean la síntesis de GlcNAc-P-P-dolicol, el intermediario unido a dolicol iniciador de la ruta de la N-glicosilación de las proteínas. Aunque el GlcNAc-P-P-dolicol es un equivalente en mamíferos relacionado estructuralmente al intermediario glicolípido bacteriano, el GlcNAc-P-P-Und, no existen pruebas de una reacción de epimerización similar que convierta el GlcNAc-P-P-dolicol en GalNAc-P-P-dolicol en células eucariotas. Por lo tanto, esto aumenta la posibilidad de que en las cepas en las que el antígeno O de superficie que contienen GalNAc en los extremos reductores están implicados en un proceso patológico. La síntesis del antígeno O podría bloquearse potencialmente inhibiendo las epimerasas bacterianas.

Una realización de la presente divulgación implica un epimerasa que convierte GlcNAc-P-P-Und (N-acetilglucosaminil-pirofosforil-undecaprenol) en GalNAc-P-P-Und (N-acetilgalactosaminil-pirofosforil-undecaprenol) en *E. coli* 0157. Un aspecto ejemplar adicional más de la invención implica el inicio de síntesis de tetrasacáridos de repetidos unidos a lípidos que tienen GalNAc en el extremo reductor.

La base de otro aspecto de la invención incluye el descubrimiento de que el *Campylobacter jejuni* contiene un sistema de glicosilación unida a N de proteínas general. Se ha demostrado que distintas proteínas de *C. jejuni* se modifican mediante un heptasacárido. Este heptasacárido se ensambla sobre el pirofosfato de undecaprenilo, el vehículo lipídico, en el lado citoplasmático de la membrana interna por la adición por etapas de monosacáridos nucleotídicos activados catalizados por glicosil transferasas específicas. El oligosacárido unido al lípido se difunde transversalmente (flip-flop) en el espacio periplásmico mediante una flipasa, por ejemplo, PglK. En la etapa final de la glicosilación unida a N de la proteína, la oligosacaril- transferasa (por ejemplo, PglB) cataliza la transferencia de oligosacárido desde el vehículo lipídico a restos de asparagina (Asn) con la secuencia de consenso D/E-X-N-Z-S/T, en la que X y Z puede ser cualquiera aminoácido excepto prolina. El agrupamiento de glicosilación para el heptasacárido se ha trasferido satisfactoriamente a *E. coli* y se han producido glicoproteínas unidas a N de Campylobacter.

Se ha demostrado que la PglB no tiene una especificidad estricta para el sustrato de azúcar unido a un lípido. Los polisacáridos antigénicos unidos en pirofosfato de undecaprenilo se capturan por la PglB en el periplasma y se transfieren a un vehículo proteico (Feldman, 2005; Wacker, M., y col., Sustrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. Proc Natl Acad Sci USA, 2006. 103(18): p. 7088-93). La enzima también transferirá una matriz diversa de oligosacáridos unidos a pirofosfato de undecaprenilo (UPP) si contienen una hexosamina N-acetilada en el extremo reductor. La secuencia de nucleótidos de PglB y la secuencia de aminoácidos para PglB están publicados en el documento WO2009/104074.

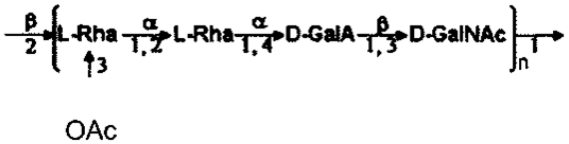
En consecuencia, una realización de la divulgación implica una proteína N-glicosilada recombinante que comprende: una o más de una secuencia de consenso introducida, D/E-X-N-Z-S/T, en la que X y Z pueden ser cualquier aminoácido natural excepto prolina; y un oligo o polisacárido que tiene N-acetilgalactosamina en el extremo reductor y está unido por N a cada una de dichas una o más secuencias de consenso introducidas por un enlace N-glicosídico.

En una realización adicional, la presente divulgación se refiere a un sistema biosintético procariota recombinante para la producción de todo o parte de un polisacárido que comprende una epimerasa que sintetiza N-acetilgalactosamina ("GalNAc") sobre pirofosfato de undecaprenilo. En una realización adicional, todo o una parte del polisacárido es antigénico.

En otra realización, la presente divulgación se refiere a un sistema biosintético procariota recombinante que comprende: una epimerasa que sintetiza GalNAc sobre pirofosfato de undecaprenilo; y glicosil transferasas que sintetizan un polisacárido que tiene GalNAc en el extremo reductor.

Una realización de la divulgación que comprende adicionalmente un sistema biosintético procariota recombinante que comprende una epimerasa que sintetiza GalNAc sobre pirofosfato de undecaprenilo y glicosil transferasas que sintetizan un polisacárido, en el que dicho polisacárido tiene la siguiente estructura: α-D-PerNAc-α-L-Fuc-β-D-Glc-α-D-GalNAc; y en el que GalNAc está en el extremo reductor de dicho polisacárido.

El sistema biosintético procariota recombinante puede producir mono, oligo o polisacáridos de distintos orígenes. Dichos oligo y polisacáridos pueden ser de origen procariota o eucariota. Los oligo o polisacáridos de origen procariota pueden ser de bacterias Gram-negativas o Gram-positivas. En una realización de la invención, el oligo o polisacárido es de *E. coli*. En un aspecto adicional de la invención, dicho oligo o polisacárido es de *E. coli* 0157. En otra realización, dicho oligo o polisacárido comprende la siguiente estructura: α-D-PerNAc-α-L-Fuc- β-D-Glc-α-D-GalNAc. En una realización adicional de la invención, el oligo o polisacárido es de *Shigella flexneri*. En una realización adicional más, el oligo o polisacárido es de *Shigella flexneri* 6. En un aspecto adicional más, dicho oligo o polisacárido comprende la siguiente estructura:



Las realizaciones de la divulgación incluyen adicionalmente proteínas de distintos orígenes. Dichas proteínas incluyen proteínas nativas a organismos procariotas y eucariotas. El vehículo proteico puede ser, por ejemplo, AcrA o un vehículo proteico que se ha modificado para que contenga la secuencia de consenso para la glicosilación proteica, es decir, D/E-X-N-Z-S/T, en la que X y Z puede ser cualquier aminoácido excepto prolina (por ejemplo, una Exotoxina de *Pseudomonas aeruginosa* ("EPA")). En una realización de la invención, la proteína es EPA de *Pseudomonas aeruginosa*.

Un aspecto adicional de la divulgación implica nuevas vacunas bioconjugadas que tienen GalNAc en el extremo reductor del n-glicano. Una realización adicional de la divulgación implica una nueva estrategia para producir dichas vacunas bioconjugadas que utiliza células bacterianas recombinantes que contienen una epimerasa que produce GalNAc sobre pirofosfato de undecaprenilo. En una realización, las vacunas bioconjugadas pueden utilizarse para tratar o prevenir enfermedades bacterianas. En realizaciones adicionales, las vacunas bioconjugadas pueden tener un potencial terapéutico y/o profiláctico para el cáncer u otras enfermedades.

Una dosificación de vacunación típica para seres humanos es aproximadamente 1 a 25 μg, preferentemente aproximadamente 1 μg a aproximadamente 10 μg, más preferentemente aproximadamente 10 μg. Opcionalmente, una vacuna, tal como una vacuna bioconjugada de la presente invención, incluye un adyuvante.

En una realización adicional, la presente divulgación se refiere a un sistema de expresión para la producción de una vacuna bioconjugada contra al menos una bacteria que comprende: una secuencia de nucleótidos que codifica una oligosacaril transferasa; una secuencia de nucleótidos que codifica un vehículo proteico; al menos un agrupamiento genético de polisacáridos de la al menos una bacteria, en la que el polisacárido contienen GalNAc en el extremo reductor; y una secuencia de ácido nucleico que codifica una epimerasa. En una realización adicional, el agrupamiento genético de polisacáridos codifica un polisacárido antigénico.

En otra realización adicional más, la presente invención se refiere a un sistema de expresión para la producción de una vacuna bioconjugada contra al menos una bacteria que comprende: una secuencia de nucleótidos que codifica una oligosacaril transferasa; una secuencia de nucleótido que codifica un vehículo proteico que comprende al menos un secuencia de consenso insertada, D/E-X-N-Z-S/T, en el que X y Z puede ser cualquier aminoácido natural excepto prolina; al menos un agrupamiento genético de polisacárido de la al menos una bacteria, en el que el polisacárido contiene GalNAc en el extremo reductor; y el gen Z3206. En una realización adicional, el agrupamiento genético de polisacárido codifica un polisacárido antigénico.

En otra realización, la presente divulgación se refiere a una vacuna bioconjugada que comprende: un vehículo proteico; al menos una cadena de polisacárido inmunogénico unida al vehículo proteico, en el que dicho polisacárido tiene GalNAc en el extremo reductor, y adicionalmente en el que dicho GalNAc está unido directamente al vehículo proteico; y un adyuvante.

En una realización adicional más, la presente divulgación se refiere a una vacuna bioconjugada que comprende. un vehículo proteico que comprende al menos una secuencia de consenso insertada, D/E-X-N-Z-S/T, en la que X y Z puede ser cualquier aminoácido natural excepto prolina; al menos un polisacárido inmunogénico de al menos una bacteria, unido al vehículo proteico, en el que al menos un polisacárido inmunogénico contiene GalNAc en el extremo reductor unido directamente al vehículo proteico; y, opcionalmente, un adyuvante.

Otra realización de la divulgación se refiere a un procedimiento de producción de una vacuna bioconjugada, comprendiendo dicho procedimiento: ensamblaje de un polisacárido que tiene GalNAc en el extremo reductor en un organismo recombinante mediante el uso de glicosil transferasas; unir dicho GalNAc a un resto de asparagina de una o más proteínas diana en dicho organismo recombinante, en el que dicha una o más proteínas diana contienen uno o más epítopos de linfocitos T.

En una realización adicional, la presente divulgación se refiere a un procedimiento de producción de una vacuna bioconjugada, comprendiendo dicho procedimiento: la introducción de información genética que codifica un aparato metabólico que lleva a cabo la N-glicosilación de una proteína diana en un organismo procariota para producir un organismo procariota modificado; en el que la información genética necesaria para la expresión de una o más proteínas recombinantes se introduce en dicho organismo procariota; en el que la información genética necesaria para la expresión de la epimerasa de la cepa 0157 de *E. coli* se introduce en dicho organismo procariota; y en el que el aparato metabólico comprende glicosil transferasas de un tipo que ensambla un polisacárido que tiene GalNAc en el extremo reductor en un vehículo lipídico, y una oligosacaril transferasa, uniendo covalentemente la oligosacaril transferasa el GalNAc del polisacárido a un resto de asparagina de la proteína diana, y conteniendo la proteína diana al menos un epítopo de linfocito T; producir un cultivo del organismo procariota modificado; y obtener las proteínas glicosiladas del medio de cultivo.

Un aspecto adicional de la presente divulgación se refiere a una composición farmacéutica. Un aspecto adicional de la divulgación implica una composición farmacéutica que comprende al menos una proteína N-glicosilada de acuerdo con la invención. A la luz de la divulgación del presente documento, la preparación de medicamentos que comprenden las proteínas se conoce bien en la técnica. un aspecto adicional más de la divulgación se refiere a una composición farmacéutica que comprende un antibiótico que inhibe una epimerasa que convierte GlcNAc-P-P-Und en GalNAc-P-P-Und. En una realización preferida, la composición farmacéutica de la invención comprende un excipiente, diluyente y/o adyuvante farmacéuticamente aceptable.

Los excipientes, diluyentes y/o adyuvantes adecuados se conocen bien en la técnica. Un excipiente o diluyente puede ser un material sólido, semi-sólido o líquido que puede servir como vehículo o medio para el principio activo. Un experto en la técnica en el campo de la preparación de composiciones puede seleccionar fácilmente la forma apropiada y modo de administración dependiendo de las características particulares del producto seleccionado, la enfermedad o afección que se va a tratar, el estadio de la enfermedad o afección, y otras circunstancias relevantes (Remington's Pharmaceutical Sciences, Mack Publishing Co. (1990)). La proporciona y naturaleza del diluyente o excipiente farmacéuticamente aceptable se determinan por la solubilidad y propiedades químicas del principio farmacéuticamente activo seleccionado, la vía de administración escogida, y la práctica farmacéutica convencional. La preparación farmacéutica puede adaptarse para su uso oral, parenteral o tópico y puede administrarse al paciente en forma de comprimidos, cápsulas, supositorios, solución, suspensiones o similares. Los compuestos farmacéuticamente activos de la presente divulgación, aunque son efectivos por sí mismos, se pueden formular y administrar en forma de sus sales farmacéuticamente aceptables, tales como sales por adición de ácido o sales por adición de bases, con fines de estabilidad, conveniencia de cristalización, aumento de solubilidad, y similares.

En ciertos casos en los que las secuencias específicas de nucleótidos o aminoácidos se señalan, se entenderá que la presente divulgación engloba secuencias homólogas que tengan la misma funcionalidad de las secuencias señaladas. En una realización de la divulgación, dichas secuencias son al menos un 90 % homólogas. En otras realizaciones adicionales más, dichas secuencias son al menos un 95 % homólogas.

La determinación del porcentaje de identidad entre dos secuencias de nucleótidos o aminoácidos se conoce por los expertos en la técnica.

Las secuencias de ácido nucleico que se describen en el presente documento, tal como las que se describen en el listado de secuencias posterior, son ejemplos solamente, y será evidente para un experto en la técnica que las secuencias se pueden combinar de diferentes maneras. Realizaciones adicionales de la divulgación incluyen variantes de ácidos nucleicos. Una variante de un ácido nucleico (por ejemplo, un ácido nucleico con codón optimizado) puede ser sustancialmente idéntica, es decir, al menos un 80 % idéntica, por ejemplo, un 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 99,5 % idéntica, a las SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 1 1, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28 o SEQ ID NO: 29. Variantes de ácidos nucleicos de una secuencia que contiene las SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 1 1, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28 o SEQ ID NO: 29 incluyen ácidos nucleicos con una sustitución ,variación, modificación, remplazo, eliminación, y/o adición de uno o más nucleótidos (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10, 1 1, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175 o 200 nucleótidos) de una secuencia que contiene las SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 1 1, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21 , SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28 o SEQ ID NO: 29, o partes de las mismas.

Por ejemplo, en una realización de la presente divulgación, dichas variantes incluyen ácidos nucleicos que codifican una epimerasa que convierte GlcNAc-P-P-Und en GalNAc-P-P-Und y que i) se expresan en una célula huésped, tal como, por ejemplo, *E. coli* y ii) son sustancialmente idénticas a SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 o SEQ ID NO: 9, o partes de las mismas.

Los ácidos nucleicos descritos en el presente documento incluyen ADN recombinante y ADN sintético (por ejemplo, sintético químicamente). Los ácidos nucleicos pueden ser de doble cadena o cadena sencilla En el caso de ácidos nucleicos de cadena sencilla, el ácido nucleico puede ser una cadena en sentido o una cadena antisentido. Los ácidos nucleicos se pueden sintetizar utilizando análogos derivados de oligonucleótidos.

Los plásmidos que incluyen un ácido nucleico descrito en el presente documento se pueden transfectar o transformar en células huésped para la expresión. Las técnicas para la transfección y transformación son conocidas por el experto en la técnica.

Se tiene que entender que el término “o”, como se utiliza en el presente documento, denota alternativas que pueden, si es apropiado, combinarse; es decir, el término “o” incluye cada uno de las alternativas enumeradas por separado así como su combinación. Como se utiliza en el presente documento, a menos que el contexto dicte claramente otra cosa, las referencias al singular, tal como las formas singulares de “un”, “una” y “el”, incluyen el plural, y las referencias al plural incluyen el singular.

La invención se define adicionalmente en referencia a los siguientes ejemplos que describen adicionalmente las composiciones y procedimientos de la presente invención, así como su utilizad. Será evidente para los expertos en la técnica que pueden practicarse modificaciones de las composiciones y procedimientos, y mantenerse en el ámbito de la invención.

Ejemplos

Cepas bacterianas y Plásmidos – Las cepas de E. coli PR4019 (Rush, J. S., Rick, P. D. y Waechter, C. J. (1997) Glycobiology, 7, 315-322) y PR21546 (Meier-Dieter, U., Starman, R., Barr, K., Mayer, H. y Rick, P. D. (1990) J. Biol. Chem., 265, 13490-13497) fueron donadas generosamente por el Dr. Paul Rick, Bethesda, MD, y *E. coli* 0157:H45 (Stephan, R., Borel, N., Zweifel, C, Blanco, M., y Blanco, J.E. (2004) BMC Microbiol 4: 10) era un regalo del Dr. Claudio Zweifel, Instituto Veterinario, Universidad de Zurich. *E. coli* H5or (Invitrogen) se utilizó como el huésped para los experimentos de clonación y para los análisis de glicosilación proteica. Los plásmidos utilizados se enumeran en la Tabla 2.

TABLA 2

Plásmidos utilizados en los Ejemplos		
Plásmido	Descripción	Ref.
pMLBAD	Vector de clonación, Tmp^R	Lefebre & Valvano (2002)
pMLBAD:Z3206 (SEQ ID NO: 23)	Z3206 en pMLBAD, Tmp^R, expresión controlada por un promotor inducible por arabinosa	Ejemplos del presente documento
pMLBAD:gne (SEQ ID NO: 24)	*gne* en pMLBAD, Tmp^R, expresión controlada por un promotor inducible por arabinosa	Ejemplos del presente documento

(continuación)

Plásmidos utilizados en los Ejemplos		
Plásmido	Descripción	Ref.
pACYCpgl	Agrupamiento de *pgl* de *C. jejuni* Cm^R	Wacker, y col. (2002)
pACYC*gne::kan*	Agrupamiento de *pgl* de *C. jejuni* que contiene un casete kan en *gne,* Cm^R, Kan^R	Linton, y col. (2005)
pWA2	AcrA soluble pariglásmica marcada con hexa-His bajo el control del promotor Tet en pBR322, Amp^R	Feldman, y col. (2005)

Materiales – Se obtuvieron [I,6-^{3}H]GlcNAc (30 Ci/mmol), UDP-[I-^{3}H]GlcNAc (20 Ci/mmol) y UDP-[6-^{3}H]GalNAc (20 Ci/mmol) en American Radiolabeled Chemicals (St. Louis, MO). Las placas de capa fina en gel G de sílice Quantum 1 son un producto de Quantum Industries (Fairfield, NJ),y placas de Gel G de sílice Si250 Baker se fabricaron en Mallinckrodt Chemical Works. El extracto de levadura y Bacto-peptona son productos de BD Biosciences. Todos los demás productos químicos se obtuvieron de fuentes comerciales convencionales. Se añadieron Trimetoprim (50 µg/ml), cloranfenicol (20 µg/ml), ampicilina (100 µg/ml), y kanamicina (50 µg/ml) si era necesario.

Construcción de Plásmidos recombinantes – La cepa DH5a de *E. coli* se utilizó para los experimentos de clonación de ADN y los plásmidos construidos se verificaron por secuenciación de ADN. El gen Z3206 se amplificó a partir de *E. coli* 0157: H45 por PCR con los oligonucleótidos Z3206-Fw y Z3206-RvHA (AAACCCGGGATGAACGATAACGTTTTGCTC (SEQ ID NO: 17) y AAATCTAGATTAAGCGTAATCTGGAACATC GTATGGGTACTCAGAAACAAACGTTATGTC (SEQ ID NO: 18); los sitios de restricción están subrayados). El fragmentos de PCR se digirió con SmaI y XbaI y se ligó en el vector pMLBAD escindido en SmaI-XbaI (Lefebre, M. D. y Valvano M. A. (2002) Appl Environ Microbiol 68: 5956-5964). Esto daba como resultado el plásmido pMLBAD:Z3206 (SEQ ID NO: 23) que codifica Z3206 con un marcador de hemaglutinina en el extremo C.

El gne se amplificó a partir del pACYCpgl (Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wrenn, B.W., Aebi, M. (2002) Science 298, 1790-1793), que codifica el agrupamiento pgl de *Campylobacter jejuni*, con los oligonucleótidos gne-Fw y gne-RV (AAACCATGGATGAAAATTCTTA TTAGCGG (SEQ ID NO: 19) y AAATCTAGATTAAGCGTAATCTGGAACATCGTATGGGTAGCACTGTTTTTCCCAATC (SEQ ID NO: 20); los sitios de restricción están subrayados). El producto de la PCR se digirió con NcoI y XbaI y se ligó en los mismos sitios del pMLBAD para generar el plásmidos pMLBAD: gne (SEQ ID NO: 24) que codifica Gne con un marcador de hemaglutinina en el extremo C (Tabla 2).

Condiciones de cultivo, Expresión proteica e inmunodetección – Las cepas de *E. coli* se cultivaron en medio Luria-Bertani (un 1 % extracto de levadura, un 2 % de Bacto-peptona, un 0,6 % de NaCl) a 37 ºC con agitado vigoroso. La expresión inducible por arabinosa se consiguió añadiendo arabinosa a una concentración final de un 0,02-0,2 % (p/v) a las células de *E. coli* hasta un A_{600} de 0,05-0,4. Se añadió la misma cantidad de arabinosa de nuevo 5 h tras la inducción, y se continuó la incubación durante 4-15 h.

Procedimientos analíticos – Las concentraciones proteicas se determinaron utilizando el ensayo proteico BCA (Pierce) después de la precipitación de las proteínas de membrana con desoxicolato y ácido tricloroacético de acuerdo con el boletín Pierce Biotechnology “Eliminate Interfering Sustances from Samples for BCA Protein Assay”. Las muestras se analizaron en cuanto a la radioactividad por espectrometría de centelleo en un espectrofotómetro de centelleo líquido Packard Tri-Carb 2100TR después del a adición de 0,5 ml de un 1 % de SDS y 4 ml de Econosafe Economical Biodegradable Counting Mixture (Research Products International, Corp., Mount Prospect, IL).

Ejemplo 1: Identificación de un gen de *E. coli* 0157 que codifica GlcNAc-P-P-Und 4-epimerasa

Los inventores describen en el presente documento el sorprendente descubrimiento de una nueva ruta biosintética en la que se forma GalNAc-P-P-Und por epimerización del 4-OH de GlcNAc-P-P-Und catalizada por la acción desconocida anteriormente de una 4-epimerasa. En esta ruta, el GlcNAc-P-P-Und se forma por la transferencia de GlcNAc-P del UDP-GlcNAc, catalizada por WecA, y después se epimeriza el GlcNAc-P-P-Und en GalNAc-P-P-Und por la GlcNAc-P-P-Und-4-epimerasa, que era una ruta desconocida previamente (FIG. 2).

El gen que codifica un candidato para la GlcNAc-P-P-Und-4-epimerasa se identificó por búsqueda de homología de ADN. Las búsquedas de homología se llevaron a cabo utilizando las bases de datos de la U.S. National Library of Medicine que se encuentra en http://blast.ncbi.nlm.nih.gov/Blast.cgi. Se exploraron las secuencias genómicas de las diferentes bacterias que codifican las unidades repetidas del antígeno O que tienen un GalNAc en el extremo reductor. Un grupo con una unidad repetida que contenía GalNAc en el extremo reductor, y un segundo grupo que carecía de un GalNAc terminal en la unidad repetida se compararon para identificar las epimerasas potenciales.

Utilizando estos criterios, se identificó el Z3206 como candidata de GlcNAc-P-P-Und-4-epimerasa (Tabla 1).

Los genes de GlcNAc 4- epimerasa presentes en las cepas de *E. coli* con unidades repetidas del antígeno O que contenían GalNAc se pueden separar en dos grupos de homología como se muestra en la Tabla 1. Se descubrió sorprendentemente que un grupo de homología (que contenía el gneI) se correlacionaba claramente con la presencia de GalNAc como el azúcar de inicio en la unidad repetida del antígeno O. Se descubrió sorprendentemente adicionalmente que el segundo grupo (que contenía gneI) presentaba un alto grado de similitud con la UDP-Glc epimerasa, GalE, y se encuentra en las cepas de *E. coli* que no inician la síntesis de unidades repetidas del antígeno O con GalNAc. Se identificó el gen Z3206 en *E. coli* 0157, un gen con un alto grado de homología con gneI, como un candidato para la GlcNAc-P-P-Und-4-epimerasa. La localización genómica del gen Z3206 es consistente con un papel en esta ruta, ya que reside entre galF del agrupamiento del antígeno O y wcaM que pertenece al agrupamiento del ácido colánico.

La investigación descrita en los Ejemplos 2-1 1 confirma además los descubrimientos anteriores, incluyendo la GlcNAc 4-epimerasa (*E. coli* 0157 Z3206) como que cataliza la formación de GalNAc-P-P-Und.

Ejemplo 2: UDP-GalNAc no es un sustrato para WecA de *E. coli* (GlcNAc-fosfotransferasa)

Para determinar si la WecA de *E. coli* utilizará el UDP-GalNAc como un donante GalNAc-P para formar GalNAc-P-P-Und, se incubaron fracciones de membrana de las cepas K12, PR4019 de *E. coli*, una cepa que sobre-expresa WecA, y 0157, que sintetiza un tetrasacárido de unidad repetida de antígeno O con GalNAc en el extremo reductor presumiblemente iniciado por la síntesis de GalNAc-P-P-Und, con UDP-[^{3}H]GalNAc.

Preparación de la membrana de *E. coli* – Las células bacterianas se recolectaron por centrifugación a 1.000 x g durante 10 min, se lavaron una vez con solución salina tampón de fosfato enfriada en hielo, una vez con agua fría, y una vez con 10 mM de Tris-HCl, pH 7,4, 0,25 M de sacarosa, 10 mM EDTA que contenía 0,2 mg/ml de lisozima, y se incubó a 30 ºC durante 30 min. Las células bacterianas se recuperaron por centrifugación a 1.000 x g durante 10 minutos, se re-suspendieron rápidamente en 40 volúmenes de Tris-HCl 10 mM enfriado en hielo, pH 7,4, y se colocaron en hielo. Después de 10 min las células se homogeneizaron con 15 golpes con un homogeneizador de Dounce ajustado y se suplementaron con 0,1 mM de fluoruro de fenilmetilsulfonilo y sacarosa a una concentración final de 0,25 M. Las células sin romper se retiraron por centrifugación a 1.000 x g durante 10 min, y los envoltorios celulares se recuperaron por centrifugación a 40.000 x g durante 20 min. La fracción de membrana se res-suspendió en 10 mM de Tris-HCl, pH 7,4, 0,25 M de sacarosa, 1 mM de EDTA y de nuevo se sedimentó a 40.000 x g y se re-suspendió en el mismo tampón a una concentración de ~ 20 mg/ml. Las fracciones de membrana se almacenaron a -20 ºC hasta que se necesitaran.

Ensayo para la biosíntesis de [^{3}H] GlcNAc-P-P-Und y [^{3}H] GalNAc-P-P-Und en membranas de *E. coli* – Las mezclas de reacción *in vitro* para la síntesis de GlcNAc-P-P-Und y GalNAc-P-P-Und contenían 50 mM de Tris-HCl, pH 8, 40 mM de $MgCl_2$, 5 mM de ditiotreitol, 5 mM de 5'AMP, fracción de membrana de *E. coli* (50-200 µg de proteína de membrana), y 5 µM de UDP-[^{3}H]GlcNAc/GalNAc (500-2500 dpm/pmol) en un volumen total de 0,05 ml. Después de la incubación a 37 ºC se terminaron las reacciones por la adición de 40 volúmenes de $CHCl_3/CH_3OH$ (2:1), y el extracto lipídico total que contiene [H]HexNAc-P-P-undecaprenoles se preparó como se había descrito previamente (Waechter, C. J., Kennedy, J. L. y Harford, J. B. (1976) Arch. Biochem. Biophys. 174, 726-737). Después de la partición, la fase orgánica se secó con una corriente de nitrógeno y se re-disolvió en 1 ml de $CHCl_3/CH_3OH$ (2:1) y se retiró una alícuota (0,2 ml), se secó en un vial de centelleo, y se analizó en cuanto a la radioactividad por espectrometría de centelleo líquida en un espectrómetro de centelleo líquido Packard Tri-Carb 2100 TR. Para determinar la tasa de síntesis de [^{3}H] GlcNAc-P-P-Und o [^{3}H] GalNAc-P-P-Und, se secó el extracto lipídico bajo una corriente de nitrógeno, se re-disolvió en una pequeño volumen de $CHCl_3/CH_3OH$ (2:1), y se esparció en una placa de gel de sílice Baker Si250 impregnada de borato de 10 x 20 cm, y se desarrolló la placa con $CHCl_3$, CH_3OH, H_2O, 0,2 M de borato sódico (65:25: 2:2). Los glicolípidos individuales se detectaron en un Bioscan AR2000 Imaging Scanner (Bioscan, Washington, D.C). Las tasas biosintéticas de cada glicolípido se calcularon multiplicando la cantidad total de radioactividad en [^{3}H] GlcNAc/ GalNAc-P-P-Und por el porcentaje de los [^{3}H]glicolípidos individuales.

Las fracciones de membrana de las diferentes cepas de *E. coli* (K12, PR4019 y 0157) se incubaron con UDP-[^{3}H]GlcNAc o UDP-[^{3}H]GalNAc y se determinó la incorporación en [^{3}H] GlcNAc/ GalNAc-P-P-Und como se ha descrito anteriormente. Como se ve en la Tabla 3, no se detectaron glicolípidos marcados después de la incubación con UDP-[^{3}H]GalNAc, solo era detectable el GlcNAc-P-P-Und cuando las fracciones de membrana se incubaron con UDP- [^{3}H]GlcNAc.

TABLA 3

Síntesis de [H]GlcNAc/GalNAc-P-P-undecaprenol en fracciones de membrana de *E. coli* utilizando UDP-[^{3}H]GlcNAc o UDP-[^{3}H]GalNAc como sustrato			
Fuente de membranas	Nucleótidos de azúcares añadidos	[^{3}H]Glicolípidos formados	
		GlcNAc-P-P-Und	GalNAc-P-P-Und
		(*pmol/mg*)	(*pmol/mg*)
K12	UDP-[^{3}H]GlcNAc	6,4	<0,01
K12	UDP-[^{3}H]GalNAc	<0,01	<0,01
PR4019	UDP-[^{3}H]GlcNAc	44	<0,01
PR4019	UDP-[^{3}H]GalNAc	<0,01	<0,01
0157	UDP-[^{3}H]GlcNAc	1,5	0,5
0157	UDP-[^{3}H]GalNAc	<0,01	<0,01

Además, ni la adición de Und-P exógenos a las incubaciones con membranas de PR4019, la cepa que sobre-expresa WecA, ni la adición de fracciones citosólicas de células 0157 resultaba en la formación de GalNAc-P-P-Und de UDP-GalNAc. Estos resultados demuestran que UDP-GalNAc no es un sustrato para la WecA y sugiere que el GalNAc-P-P-Und se forma por un mecanismo alternativo.

Cuando las membranas de la cepa K12 se incubaban con UDP- [^{3}H]GlcNAc, se sintetizaba [^{3}H] GlcNAc-P-P-Und como se esperaba (Rush, J. S., Rick, P. D. y Waechter, C. J. (1997) Glycobiology, 7, 315-322). Sin embargo, cuando se incubaban las membranas de la cepa 0157 con UDP- [^{3}H]GlcNAc, además de [^{3}H] GlcNAc-P-P-Und, se observaba un segundo lípido marcado que demostraba ser el [^{3}H] GalNAc-P-P-Und (véase posteriormente). Cuando el curso de tiempo para la formación de los dos glicolípidos se examinaban, la incorporación de radioactividad en [^{3}H] GlcNAc-P-P-Und (FIG. 1, O) se producía más rápidamente y en una extensión mayor que en [^{3}H] GalNAc-P-P-Und (FIG. 1, ·), compatible con la relación precursor-producto (FIG. 2).

La observación de que las membranas de *E. coli* 0157 no utilizan UDP-GalNAc como donante GalNAc-P para la síntesis de GalNAc-P-P-Und es un ejemplo que confirma la ruta biosintética parala formación de GalNAc-P-P-Und que se ilustra en la FIG. 2. En este esquema, se forma GlcNAc-P-P-Und por transferencia de GlcNAc-P a partir de UDP-GlcNAc, catalizada por WecA, y después se epimeriza el GlcNAc-P-P-Und por la acción de una 4-epimerasa previamente desconocida para producir el GalNAc-P-P-Und.

Ejemplo 3: Caracterización de GalNAc-P-P-Und formado *in vitro* con fracciones de membrana de la cepa 0157 de *E. coli*

Consistentemente con el producto glicolípido específico de 0157 adicional detectado en la FIG. 1, como GalNAc-P-P-Und, era estable en metanolisis alcalina moderada (tolueno/metanol 1:3, que contenía 0,1 N de KOH, 0 ºC, 60 min), mantenido en DEAE-celulosa equilibrada en $CHCl_3/CH_3OH/H_2O$ (10:10:3), y se eluyó con $CHCl_3/CH_3OH/H_2O$ (10:10:3) que contenía 20 mM de acetato amónico como se informó anteriormente para [^{3}H] GlcNAc,$_{-2}$-P-P-Dol (Waechter, C. J. y Harford, J. B. (1977) Arch. Biochem. Biophys. 181, 185-198).

El GalNAc-P-P-Und se resolvía claramente a partir del GlcNAc-P-P-Und por cromatografía de capa fina en gel de sílice G impregnado con borto (Kean, E. L. (1966) J. Lipid Res. 7, 449-452) y se purificó por TLC preparatoria como se muestra en las FIG. 3A y FIG. 3B.

Preparación de las placas de capa fina impregnadas con borato y papel Whatman Nº 1 – Las placas de capa fina con gen de sílice se impregnaron con borato sódico por inmersión breve de las placas en un 2,5 % de $Na_2B_4O_7$ 10 H_2O en un 95 % de metanol como describe Kean (Kean, E. L. (1966) J. Lipid Res. 7, 449-452). Las placas TLC impregnadas en borato se secaron durante una noche a temperatura ambiente y se almacenaron en un desecador al vacío sobre Drierita hasta su uso. Inmediatamente de la cromatografía, las placas se activaron calentando brevemente (10-15 min) a 100 ºC. El papel Whatman Nº 1 se impregnó con borato sódico por inmersión de láminas de 20 x 30 cm de papel Whatman 1 en 0,2 M de $Na_2B_4O_7$ 10 H_2O. Las láminas de papel Whatman Nº 1 se presionaron firmemente entre dos hojas de papel Whatman Nº 3MM y se dejó secar a temperatura ambiente durante varios días, como describen Cardini y Leloir (Cardini, C. E. y Leloir, L. F. (1957) J. Biol. Chem. 225, 317-324).

Caracterización de los productos de glicanos formados en las reacciones *in vitro* – Los glicanos de los glicolípidos individuales ([^{3}H] GlcNAc-P-P-Und y [^{3}H] GalNAc-P-P-Und) se caracterizando por cromatografía en papel descendente después de la liberación por hidrólisis ácida moderada. Los lípidos GlcNAc/GalNAc se secaron bajo una corriente de nitrógeno en un tubo cónico de tapa a rosca y se calentó a 100 ºC, 15 min en 0,2 ml de HCl 0,01 M.

Después de la hidrólisis se aplicaron las muestras en una columna de intercambio iónico en lecho mixto de 0,8 ml que contenía 0,4 ml de AG50WX8 (H^+) y 0,4 ml de AG 1x8 (forma de acetato) y se eluyeron con 1,5 ml de agua. El eluído se secó bajo una corriente de nitrógeno, se re-disolvió en un pequeño volumen de H_2O (0,02 ml), se dispersó en una tira de 30 cm de papel Whatman Nº 1 impregnado de borato, y se desarrolló en modo descendente con butanol/piridina/agua (6:4:3) durante 40-50 h. Después del secado, las tiras de papel se cortaron en zonas de 1 cm y se analizaron en cuanto a la radioactividad por espectrometría de centelleo. Las referencias de GlcNAc y GalNAc se detectaron utilizando un reactivo de inmersión de anilina-difenilamina (Schwimmer, S. y Benvenue, A. (1956) Science 123, 543-544).

Los productos de glicanos se convirtieron en sus correspondientes alditoles por reducción con 0,1 M de NaBttt en 0,1 M de NaOH (volumen final de 0,1 ml) a continuación de hidrólisis ácido moderada como se ha descrito anteriormente. Después de la incubación a temperatura ambiente durante una noche, las reacciones se inactivaron con varias gotas de ácido acético glacial y se secaron bajo una corriente de nitrógeno sin metanol que contenía una gota de ácido acético, varias veces. Los alditoles se disolvieron en agua, se des-salaron por el pasaje en columnas de 0,5 ml de AG50WX8 (H^+) y AG 1x8 (acetato), se secó bajo nitrógeno y se dispersó en tiras de 30 cm de papel de Whatman Nº 3MM. Las tiras de Whatman Nº 3MM se desarrollaron durante una noche en modo descendente con enil acetato, piridina, 0,1 M de ácido bórico (65:25:20), se secó y se cortó en zonas de 1 cm, y se analizaron en cuanto a radioactividad por espectrometría de centelleo. Las referencias de GlcNAcitol y GalNAcitol se visualizaron utilizando una modificación del procedimiento de inmersión en peryodato-bencidina (Gordon, H. T., Thornburg, W. y Werum, L. N. (1956) Anal. Chem. 28, 849-855). Las tiras de papel se sumergieron en acetona, 0,1 M de $NaIO_4$ (95:5), se permitió que se secara al aire durante 3 min, y entonces se sumergió en acetona/ácido acético/H_2O/o-toluidina (96:0,6:4,4:0.2 g). Los alditoles que contienen m-dioles se teñían como puntos amarillos en un fondo azul.

Espectrometría de masas ("MS") de glicolípidos – Los glicolípidos se analizaron utilizando un espectrómetro de masas ABI/MDS Sciex 4000 Q-Trap hybrid triple quadrupole lineal ion trap con una fuente de iones ABI Turbo V electrospray (ABI/MDS-Sciex, Toronto, Canadá). En resumen, si infundieron las muestras a 10 µl/min con los ajustes de la fuente de iones determinados empíricamente, y la información de la MS/MS (espectrometría de masas en una segunda dimensión) se obtuvo por fragmentación del ion molecular en modo de atrapamiento de iones lineal.

Cuando el glicolípido se trataba con ácido moderado (0,01 N HCl, 100 ºC, 15 min), el producto hidrosoluble se co-cromatografió con [^{3}H]GalNAc en cromatografía en papel descendente con papel Whatman Nº 1 impregnado con borato (FIG. 3). Además, cuando el azúcar marcado se reducía, se convertía en [^{3}H]alditol, GalNAc-OH (FIG. 3D). Además, el análisis MS de ion negativo daba lugar al [M-H]-ion de m/z = 1128, esperado para GalNAc-P-P-Und, y el espectro iónico hermano de MS/MS mostraba un ion prominente a m/z = 907, esperado para un glicolípido que contiene P-P-Und (Guan, Z., Breazeale, S. D. y Raetz, C. R. (2005) Anal. Biochem. 345, 336-339). La identificación del producto glicolipídico formado por la cepa 0157 como GalNAc-P-P-Und también se mantiene por su formación a partir de GlcNAc-P-P-Und exógeno (véase el Ejemplo 7).

Ejemplo 4: Marcado metabólico de [^{3}H] GalNAc-P-P-Und (*in vivo*) con [^{3}H] GlcNAc en células de *E. coli* que expresan el gen Z3206

Para investigar si la expresión del gen Z3206 de la *E. coli* 0157 hacia posible que las células sintetizaran GalNAc-P-P-Und, la cepa 21546 de *E. coli* (Meier-Dieter, U., Starman, R., Barr, K., Mayer, H. y Rick, P. D. (1990) J. Biol. Chem., 265, 13490- 13497) que expresa el gen Z3206 se marcó metabólicamente con [^{3}H]GlcNAc y se analizó en cuanto a la formación de [^{3}H] GlcNAc/ GalNAc-P-P-Und.

Marcado metabólico de células bacterianas – Las células de *E. coli* se cultivaron con agitado vigoroso en medio Luria-Bertani a 37 ºC hasta un Λ_0o θi 0,5-1.

Se añadió [^{3}H]GlcNAc a una concentración final de 1 mCi/ml y se continuó la incubación durante 5 min a 37 ºC. La incorporación del radiomarcador en glicolípidos se terminó por la adición de 0,5 g/ml de hielo picado, y los cultivos se mezclaron concienzudamente. Las células bacterianas se recuperaron por centrifugación a 4000 x g durante 10 min, y el sobrenadante se desechó. Las células se lavaron con solución salina tampón de fosfato enfriada en hielo, se resuspendieron por removido vigoroso mezclado en 10 volúmenes (del aglomerado celular) de metanol, y se sonicó brevemente con una sonda sonicadora y al 40 % de potencia total. Después de la sonicación, se añadieron 20 volúmenes de cloroformo, y se mezclaron vigorosamente los extractos y se les dejó en reposo a temperatura ambiente durante 15 min. El material insoluble se sedimentó por centrifugación, y el aglomerado se re-extrajo con un pequeño volumen de $CHCl_3/CH_3OH$ (2:1) dos veces. Los extractos orgánicos combinados se procesaron como se describe posteriormente.

La purificación de GlcNAc-P-P-Und y GalNAc-P-P-Und-GlcNAc/GalNAc-P-P-Und se extrajeron con $CHCl_3/CH_3OH$ (2:1) y se liberaron en material hidrosoluble por partición como se describe en otro sitio (Waechter, C. J., Kennedy, J. L. y Harford, J. B. (1976) Arch. Biochem. Biophys. 174, 726-737). El extracto orgánico se secó entonces bajo una corriente de nitrógeno, y el volumen de glicerofosfolípidos se destruyó por desacilación en tolueno/metanol (1:3) que contenía 0,1 N de KOH a 0 ºC durante 60 min. La reacción de desacilación se neutralizó con ácido acético, diluido con 4 volúmenes de $CHCl_3/CH_3OH$ (2:1), y se lavó con 1/5 volúmenes de 0,9 % de NaCl. La fase orgánica (inferior) se lavó con 1/3 de volúmenes de $CHCl_3/CH_3OH$ H_2O (10:10:3) y entonces se eluyeron con $CHCl_3/CH_3OH$ H_2O

(10:10:3) que contenía 20 mM de acetato amónico. Las fracciones (2 ml) se recolectaron y controlaron respecto a radioactividad, o GlcNAc/ GalNAc-P-P-Und utilizando un reactivo pulverizado de anisaldehído (Dunphy, P. J., Kerr, J. D., Pennock, J. F., Whittle, K. J., y Feeney, J. (1967) Biochim. Biophys. Acta 136, 136-147) después de la resolución por cromatografía de capa fina sobre placas de sílice impregnadas con borato (como se ha descrito anteriormente).

La cepa 21546 de *E. coli* se seleccionó como huésped para los estudios de expresión de Z3206 debido a una mutación en la síntesis de UDP-ManNAcA que da como resultado un bloqueo de la utilización de GlcNAc-P-P-Und para la síntesis del antígeno enterobacteriano común. Debido a que a *E. coli* 21546 se deriva de *E. coli* K12 no sintetiza una repetición del antígeno O tampoco (Stevenson, G., Neal, B., Liu, D., Hobbs, M., Packer, N. H., Batley, M., Redmond, J. W., Lindquist, L. y Reeves, P. (1994) J. Bacteriol., 176, 4144-4156), y por lo tanto, se acumulan grandes cantidades de GlcNAc-P-P-Und para la conversión en GalNAc-P-P-Und. Cuando la cepa 21546 y el transformante que expresa el gen Z3206 se marcaron con [^{3}H]GlcNAc y se analizaron los lípidos marcados por cromatografía de capa fina sobre placas de gel de sílice impregnado con borato, la cepa parental (FIG. 4A) sintetizaba solo un lípido marcado, el GlcNAc-P-P-Und. Sin embargo, las células 21546 que expresan el gen Z3206 (FIG. 4B) también sintetiza un lípido adicional que se demostró que era el GalNAc-P-P-Und.

Ejemplo 5: Fracciones de membrana de células de *E. coli* que expresan el gen Z3206 sintetizan GalNAc-P-P-Und *in vitro*

Para corroborar que la proteína codificada por el gen Z3206 de *E. coli* 0157 catalizaba la síntesis de GalNAc-P-P-Und, se incubaron fracciones de membrana de células de *E. coli* que expresaban Z3206 con [^{3}H]UDP-GlcNAc y se analizaron los productos [^{3}H]glicolípidos por cromatografía de capa fina (la preparación cromatográfica y los procedimientos de caracterización se describen en el Ejemplo 3) sobre placas de gel de sílice impregnado con borato como se muestra en la FIG. 5. Cuando las fracciones de membrana de *E. coli* K12 o las células huésped de la cepa 21546 de *E. coli* se incubaron con UDP- [^{3}H]GlcNAc. solo se observaba [^{3}H] GlcNAc-P-P-Und (FIG. 5A y FIG. 5C). Sin embargo, las fracciones de membrana de *E. coli* 0157 y *E. coli* 21546 que expresan Z3206 formaban GalNAc-P-P-Und también (FIG. 5B y FIG. 5D).

Ejemplo 6: Formación de GlcNAc-P-P-Und, pero no GalNAc-P-P-Und, se revierte en presencia de UMP

Para proporcionar una prueba adicional de que el GalNAc-P-P-Und se sintetiza a partir de GlcNAc-P-P-Und, y no por la acción de la WecA utilizando UDP-GalNAc como donante de glicosilo, el efecto de desecho endógeno, se examinaron [^{3}H] GlcNAc-P-P-Und y [^{3}H] GalNAc-P-P-Und con UMP. La reacción de GlcNAc-fosfotransferasa catalizada por WecA es libremente reversible por adición de un exceso de UMP re-sintetizando UDP-GlcNAc y liberando Und-P.

En este experimento las fracciones de membrana de la cepa 21546 de *E. coli* que expresan Z3206 se pre-marcaron durante 10 min con UDP- [^{3}H]GlcNAc seguido por la adición de 1 mM de UMP, y la cantidad de cada glicolípido marcado restante se determinó. Los resultados que se ilustran en la FIG. 6A muestra las cantidades relativas de [^{3}H] GlcNAc-P-P-Und y [^{3}H] GalNAc-P-P-Und al final del periodo de marcado de 10 min. Después de la incubación con 1 mM de UMP durante 1 min se puede ver que hay una perdida sustancial de [^{3}H] GlcNAc-P-P-Und, mientras que el pico de [^{3}H] GalNAc-P-P-Und está relativamente sin cambios (FIG. 6B) (la preparación cromatográfica y los procedimientos de caracterización se describen en el Ejemplo 5). Esta observación es consistente con los resultados de la Tabla 3 que indican que la WecA no cataliza la transferencia de GalNAc-P en GalNAc-P-P-Und a partir de UDP-GalNAc. Es digno de mención que durante el segundo minuto de incubación con UMP (FIG. 6C), la pérdida de GlcNAc-P-P-Und se enlentece, y hay una ligera reducción del pico de [^{3}H] GalNAc-P-P-Und, sugiriendo que el [^{3}H] GalNAc-P-P-Und se re-equilibra con el agrupamiento de [^{3}H] GlcNAc-P-P-Und por reversión de la reacción de epimerasa (véase el Ejemplo 7).

Ejemplo 7: Interconversión de [^{3}H] GlcNAc-P-P-Und purificada exógena y [^{3}H] GalNAc-P-P-Und catalizada por membranas de células de *E. coli* que expresan Z3206

Para proporcionar una prueba directa de que GlcNAc-P-P-Und y GalNAc-P-P-Und se pueden inter-convertir por fracciones de membrana de células de *E. coli* que expresan Z3206, [^{3}H] GlcNAc-P-P-Und purificado y [^{3}H] GalNAc-P-P-Und se ensayaron como sustratos exógenos.

El [^{3}H] GlcNAc-P-P-Und/ [^{3}H] GalNAc-P-P-Und se prepararon como en el ejemplo 4 (Marcado metabólico de células bacterianas y purificación de GlcNAc-P-P-Und y GalNAc-P-P-Und). Se incubaron [^{3}H] HexNAc-P-P-undecaprenoles (2000 dpm/pmol) dispersados en un 1 % de Triton X-100, a una concentración final del 0,1 %) con membranas de *E. coli* como en el Ejemplo 2 en Ensayo para la biosíntesis de [^{3}H] GlcNAc-P-P-Und y [^{3}H] GalNAc-P-P-Und en membranas de *E. coli in vitro*.

Los experimentos preliminares mostraban que la epimerasa era activa cuando se añadía [^{3}H] GlcNAc-P-P-Und exógeno a las mezclas de reacción dispersadas en Triton X-100, CHAPS (ácido 3-[3-colamidopropil)dimetilamonio]-l-propanosulfónico), Nonidet P-40, u octilglicósido y presentaba un pH óptimo en el intervalo de 7-8,5. La movilidad cromatográfica de [^{3}H] GlcNAc-P-P-Und y [^{3}H] GalNAc-P-P-Und purificados antes de la incubación con fracciones de membranas se muestra en la FIG. 7A y FIG. 7D. Como se ve en la FIG. 7B y FIG. 7E, los glicolípidos o se afectan por la incubación con fracciones de membrana de *E. coli* 21546. Sin embargo la incubación de los glicolípidos con

fracciones de membrana de *E. coli* 21546 que expresa Z3206 cataliza la conversión de [^{3}H] GlcNAc-P-P-Und exógeno en [^{3}H] GalNAc-P-P-Und (FIG.7C) y la conversión de [^{3}H] GalNAc-P-P-Und en [^{3}H] GlcNAc-P-P-Und (FIG. 7F). Estos resultados demuestran directamente que GlcNAc-P-P-Und y GalNAc-P-P-Und pueden inter-convertirse enzimáticamente en cepas de *E. coli* que expresan Z3206.

Ejemplo 8. La *E. coli* Z3206 no es una UDP-GlcNAc 4-epimerasa

Para determinar si Z3206 puede catalizar la formación de UDP-GalNAc, el aparato de N-glicosilación de *C. jejuni* se expresaba en *E. coli*. En este sistema indicador, la glicosilación de la proteína diana AcrA depende de la presencia del locus pgl (Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wrenn, B.W., Aebi, M. (2002) Science 298, 1790-1793), que incluye una epimerasa Gne UDP-Glc/UDP-GlcNAc (Bernatchez, S., Szymanski, CM., Ishiyama, N., Li, J., Jarrell, H.C., Lau, P.C., Berghuis, A.M., Young, N.M., Wakarchuk, W.W. (2005) J. Biol. Chem. 280, 4792-4802). La glicosilación de AcrA se pierde si el agrupamiento pgl contiene una eliminación de gne (Linton, D., Dorrell, N., Hitchen, P.G., Amber, S., Karlyshev, A.V., Morris, H.R., Dell, A., Valvano, M.A., Aebi, M. y Wren, B.W. (2005) Mol Microbiol. 55, 1695-1703). La capacidad de Z3206 para restaurar la glicosilación de AcrA en presencia del operón pgl Agne se investigó *in vivo* por la expresión de AcrA (pWA2) junto con el locus pgl Agne complementado por Gne (pMLBAD: gne) o Z3206 (pMLBAD:Z3206).

Los extractos celulares totales de *E. coli* se prepararon para el análisis de inmunodetección utilizando células a una concentración equivalente de 1 unidad A_{600} que se re-suspendieron en 100 µl de tampón de carga SDS (Laemmli, U. (1970) Nature 227, 680-685). Se cargaron alícuotas de 10 µl de la muestra final se cargaron en un 10 % SDS-PAGE. Los extractos periplásmicos de células de *E. coli* se prepararon por tratamiento con lisozima (Feldman, M.F., Wacker, M., Hernandez, M., Hitchen, P.G., Marolda, C.L., Kowarik, M., Morris, H.R., Dell, A., Valvano, M.A., Aebi, M. (2005) Proc Natl Acad Sci USA 102, 3016-3021), y 10 µl de la muestra final (que se corresponde con 0,2 A600 unidades de células) se analizó por SDS-PAGE. Después de la transferencia sobre una membrana de nitrocelulosa, se hizo una inmunotinción de la muestra con un antisuero específico (Aebi, M., Gassenhuber, J., Domdey, H., y te Heesen, S. (1996)Glycobiology 6, 439-444). Anti-AcrA (Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wrenn, B.W., Aebi, M. (2002) Science 298, 1790-1793) utilizando anticuerpos. Se utilizó un anti-IgG de conejo-HRP (Bio-Rad) como anticuerpo secundario. La detección se llevó a cabo con reactivos de detección de transferencia de Western ECL™ (Amersham Biosciences).

Como se muestra en la FIG. 8, la proteína glicosilada, que migra más lentamente que la forma no glicosilada, se formaba solo cuando las células que expresan el locus pgl Agne (calle 2). El Z3206 era incapaz de restaurar la glicosilación de la glicoproteína indicadora (FIG. 8, calle 1). En consecuencia, Z3206 no complementaria a la glicosilación de AcrA en un sistema de glicosilación dependiente de Gne. La expresión de Gne y el Z3206 asociado a la membrana se confirmaron por inmunodetección.

Ejemplo 9: Análisis de LPS de *S. flexneri* 6 +/- Z3206

En la FIG. 9 se representan algunos de los genes necesarios para la biosíntesis del antígeno O de *Shigella flexneri* 6: los genes que codifican las enzimas para la biosíntesis de precursores de nucleótidos de azúcares; los genes que codifican glicosil transferasas; genes que codifican proteínas procesadoras de antígeno O; y genes que codifican proteínas responsables de la O-acetilación. La estructura del antígeno O fue aclarada por Dmitriev, B.A. y col (Dmitriev, B.A., y col Somatic Antigens of *Shigella* Eur JBiochem, 1979. 98: p. 8; Liu B y col Structure and genetics of *Shigella* O antigens FEMS Microbiology Review, 2008. 32: p. 27).

Para identificar todos los genes necesarios para la biosíntesis del antígeno O de la *Shigella flexneri* 6 se construyó una biblioteca.

Clonación de ADN genómico de *S. flexneri* 6 – El ADN genómico de *S. flexneri* 6 se aisló utilizando el kit Macherey-Nagel NucleoSpin® Tissue siguiendo el protocolo para el aislamiento de ADN de bacterias. El ADN se aisló de cinco cultivos de una noche de *S. flexneri* 6 con 2 ml cada uno y se hizo una elución final con 100 µl de tampón de elución (5 mM Tris/HCl, pH 8,5). Las fracciones eluídas se agruparon, precipitaron en isopropanol y el aglomerado fina se re-suspendió en 53 µl de tampón TE del que el volumen total se sometió a reparación final de acuerdo con el protocolo que da el kit de producción de bibliotecas CopyControlTM Fosmid (EPICENTRE). El ADN de reparación final se purificó en un 1 % de gel de agarosa con punto bajo de fusión ejecutado con 1 x tampón TAE, recuperado y precipitado por etanol como se describe en el protocolo del kit. Se hizo la re-suspensión del ADN precipitado en 7 µl de tampón TE del que 0,15 µl del ADN se ligó en el pCCIFOS (SEQ ID NO: 27) de acuerdo con el protocolo EPICENTRE. El empaquetamiento del producto de ligadura en el fago se llevó a cabo de acuerdo con el protocolo y el fago empaquetado se diluyó 1:1 en tampón de dilución de fagos del que se utilizaron 10 µl para infectar 100 µl de células EP1300-T1 que se habían cultivado previamente como se describe por EPICENTRE. Las células (1 10 µl) se colocaron en placas seis veces con aproximadamente 100 colonias por placa de manera que las seis palcas contenían la biblioteca genómica de *S. flexneri* 6 completa. Las placas se desarrollaron por transferencia de colonias y se llevó a cabo una transferencia de Western con las colonias positivas/negativas y se tiñeron con plata.

Transferencia de colonias – Para las transferencias de colonias se dejó una membrana de nitrocelulosa sobre la placa de agar sólido, se removió, se lavó tres veces con 1x PST y se trataron de la misma manera. La membrana se

bloqueó primero con un 10 % de leche durante una hora a temperatura ambiente tras lo cual se incubó durante una hora a temperatura ambiente en 2 ml de un 1 % de leche (en PBST) con el antisuero anti-tipo VI (anticuerpo primario). después de tres lavados con PBST a 10 minutos cada uno, la membrana se incubó durante otra hora a temperatura ambiente en el anticuerpo secundario, 1:20000 anti-IgG de conejo de cabra conjugado con peroxidasa (BioRad) en 2 ml de leche al 1 % (en PBST). Después de tres lavados finales con PBST (de 10 min cada uno) se desarrolló la membrana en un sistema de creación de imágenes UVP Chemi Doc con una mezcla 1:1 de luminol y tampón de peróxido proporcionado por el Kit Super-Signal® West Dura Extended Duration Sustrate (Thermo Scientific).

El clon que reacciona con el antisuero de *S. flexneri* 6 después de la producción de una biblioteca genómica de *S. flexneri* 6 se secuenció mediante un cebador en avance en la región secuenciada previamente por Liu y col. (Liu y col., 2008) alcanzando desde rmlB a wfbZ (FIG. 9). Se hibridaron los cebadores rmlB_rev y wfbZ_fwd (S. flexneri - Z3206) y se utilizaron para secuenciar la inserción del clon hasta que se alcanzaran wcaM y hisI/F (S. flexneri + Z3206), respectivamente (FIG. 10).

Con el fin de establecer si la síntesis de antígeno O se mantiene en los clones que crecen de Z3206 (impidiendo de esta manera la epimerización de und-GlcNAc a und-GalNAc), se construyeron dos plásmidos (SEQ ID NO: 28 y SEQ ID NO: 29) (FIG. 10), se transformaron en células de *E. coli* y se analizaron mediante tinción con plata y transferencia de western.

Como se muestra en la FIG. 11, los LPS producidos en células de *E. coli* + o - a Z3206. El antígeno O se puede producir sin Z3206 sin embargo con un rendimiento de la producción menor, que indica que la eficacia de la producción de polisacáridos sin la epimerasa (Z3206) es menor.

Ejemplo 10: Análisis LLO de *S. flexneri* 6 +/- Z3206

Purificación de undecaprenol-PP-antígeno O por cromatografía en columna C18 – Las células de *E. coli* que expresan el antígeno +/- Z3206 se aglomeraron, se lavaron una vez en 50 ml de NaCl al 0,9 % y el aglomerado final se liofilizó durante una noche. El aglomerado se lavó una vez en 30 ml de metanol al 85-95 %, se re-extrajo con 10:10:3 cloroformo-metanol-agua (v/v/v) y los extractos se convirtieron en un sistema Bligh/Dyer de dos fases por la adición de agua, dando como resultado una relación final de 10:10:9 (C:M:A). Las fases se separaron por centrifugación y las fases acuosas superiores se cargaron cada una en un cartucho Sep-Pak C18 acondicionado con 10 ml de metanol y equilibrado con 10 ml de 3:48:47 (C:M:A). Después de la carga, los cartuchos se lavaron con 10 ml de 3:48:47 (C:M:A) y se eluyó con 5 ml de 10:10:3 (C:M:A). Se secaron las muestras de DO20 de las cargas, flujos continuos, lavados y eluciones de la columna C18 en un Eppendorf Concentrator Plus, se lavaron con 250 µl de metanol, se re-evaporaron y se lavaron otras tres veces con 20 µl de ddH_2O.

Hidrólisis de glicolípidos – Las muestras de glicolípidos del lavado de la columna C18 se hidrolizaron disolviendo las muestras secas en 2 ml de n-propanol: 2 M de ácido trifluoroacético (1:1), se calentó a 50 □C durante 15 minutos y se evaporó hasta el secado bajo N_2.

El marcado de oligosacáridos con 2-aminobenzoato y marcado HPLC se hicieron de acuerdo con Bigge y col. (Bigge, 1995) y la limpieza de glicano se llevó a cabo utilizando el procedimiento del disco de papel descrito en Merry y col. (2002) (Merry y col., 2002). La separación de glicanos marcados con 2-AB se llevó a cabo con HPLC utilizando una columna de fase normal GlicoSep-N de acuerdo con Royle y col. (Royle, 2002) pero modificado a un sistema de tres disolventes. El disolvente A era formato amónico 10 mM pH 4,4 en 80 % de acetonitrilo. El disolvente B era formato amónico 30 mM pH 4,4 en un 40 % de acetonitrilo. El disolvente C era ácido fórmico al 0,5 %. La temperatura de la columna era de 30 □C y los glicanos marcados con 2-AB se detectaron por fluorescencia (λ_{ε}X = 330 nm, $\lambda_{\varepsilon\pi}$| = 420 nm).

Las condiciones de gradiente eran un gradiente lineal del 100 % de A al 100 % de B durante 160 minutos con un caudal de 0,4 ml/min, seguido por 2 minutos al 100 % B al 100 % de C, aumentando el caudal a 1 ml/min. La columna se lavó durante 5 minutos con un 100 % de C, volviendo al 100 % de A durante 2 minutos y operando durante 15 minutos al 100 % de A con un caudal de 1 ml/min, volviendo entonces al caudal de 0,4 ml/min durante 5 minutos. Todas las muestras se inyectaron en agua.

Los plásmidos que expresan el antígeno O de *S. flexneri* con (SEQ ID NO: 29) o sin (SEQ ID NO: 28) Z3206 se transformaron en células SCM3 (FIG. 10). Las trazas de los volúmenes de la última elución muestran una diferencia entre las curvas de las dos muestras que contienen el antígeno O de *S. flexneri* +/- Z3206 (FIG. 12). Esta diferencia en el patrón de elución se puede explicar por una estructura diferente de oligosacárido que alberga un monosacárido diferente en el extremo reductor: GlcNAc o GalNAc dependiendo de la presencia de la epimerasa (Z3206).

Ejemplo 11: Análisis de la especificidad de pglB por la producción y caracterización de un bioconjugado producido por *S. flexneri* 6 +/- Z3206

Para evaluar si el PglB podía transferir oligosacáridos que tienen GlcNAc (antígeno O de *S. flexneri*) en el extremo reductor del vehículo proteico EPA. Se analizaron extractos purificados en níquel de células de *E. coli* que expresan EPA (SEQ ID NO: 25), PglB (SEQ ID NO: 26) y antígeno O de *S. flexneri* +/- Z3206 (SEQ ID NO: 29/SEQ ID NO: 28)

por transferencias de western utilizando anticuerpos anti EPA y anti- tipo VI. El antígeno de *S. flexneri* 06 con y sin GalNAc en el extremo reductor se transfirió a EPA por pgIB según se detecta por antisueros anti-EPA y anti-VI (FIG. 13).

El antígeno O se produce aun y se detecta, por con un rendimiento de producción menor, lo que indica que la eficacia de la producción de polisacárido sin la epimerasa es menor.

Mientras que esta invención se ha mostrado particularmente y se describe con referencia a las realizaciones de la misma, los expertos en la técnica entenderán que se pueden hacer distintos cambios en la misma en forma y detalle sin alejarse del ámbito de la invención englobada por las reivindicaciones. Dichos distintos cambios que entenderán los expertos en la técnica como cubiertos en el ámbito de la invención incluyen, en particular, proteínas N-glicosiladas y bioconjugados que comprenden un glicano distinto de los de *E. coli* y *S. flexneri* con GalNAc en el extremo reductor.

Listado de Secuencias

Solicitante: GlycoVaxyn AG

Título: Sistema biosintético que produce polisacáridos inmunogénicos en células procariotas

Número de SEQ ID NO: 29

SEQ ID NO: 1 – Secuencia de nucleótidos de Z3206 de *E. coli* 0157

Longitud: 993

Tipo: ADN

organismo: *E. coli* 0157

Secuencia:

ATGAACGATAACGTTTTGCTCATAGGAGCTTCCGGATTCGTAGGAACCCGACTA
CTTGAAACGG
CAATTGCTGACTTTAATATCAAGAACCTGGACAAACAGCAGAGCCACTTTTATCC
AGAAATCAC
ACAGATTGGCGATGTTCGCGATCAACAGGCACTCGACCAGGCGTTAGTCGGTT
TTGACACTGTT
GTACTACTGGCAGCGGAACACCGCGATGACGTCAGCCCTACTTCTCTCTATTAT
GATGTCAACG
TTCAGGGTACCCGCAATGTGCTGGCGGCCATGGAAAAAAATGGCGTTAAAAAT
ATCATCTTTAC

CAGTTCCGTTGCTGTTTATGGTTTGAACAAACACAACCCTGACGAAAACCATCC
ACACGACCCT
TTCAACCACTACGGCAAAAGTAAGTGGCAGGCAGAGGAAGTGCTGCGTGAATG
GTATAACAAAG
CACCAACAGAACGTTCATTAACCATCATCCGTCCTACCGTTATCTTCGGTGAAC
GCAACCGCGG
TAACGTCTATAACTTGCTGAAACAGATCGCTGGCGGCAAGTTTATGATGGTGGG
CGCAGGGACT
AACTATAAGTCCATGGCTTATGTTGGAAACATTGTTGAGTTTATCAAGTACAAAC
TGAAGAATG
TTGCCGCAGGTTATGAGGTTTATAACTACGTTGATAAGCCAGACCTGAACATGA
ACCAGTTGGT
TGCTGAAGTTGAACAAAGCCTGAACAAAAAGATCCCTTCTATGCACTTGCCTTA
CCCACTAGGA
ATGCTGGGTGGATATTGCTTTGATATCCTGAGCAAAATTACGGGCAAAAAATAC
GCTGTCAGCT
CAGTGCGCGTGAAAAAATTCTGCGCAACAACACAGTTTGACGCAACGAAAGTG
CATTCTTCAGG
TTTTGTGGCACCGTATACGCTGTCGCAAGGTCTGGATCGAACACTGCAGTATGA
ATTCGTTCAT GCCAAAAAAGACGACATAACGTTTGTTTCTGAG

SEQ ID NO: 2 - Secuencia de aminoácidos de Z3206

Longitud: 331

Tipo: PRT

Organismo: *E. coli* 0157

Secuencia:

MNDNVLLIGASGFVGTRLLETAIADFNIKNLDKQQSHFYPEITQIGDVRDQQALDQA
LV GFDTWLLAAEHRDDVSPTSLYYDVNVQGTR VLAAMEK GVK I
IFTSSVAWGLNKHNPDEN HPHDPFNHYGKSK QAEEVLRE YNKAPTERSLTI
IRPTVI FGERNRG VYNLLKQIAGGKF M VGAGTNYKSMAYVGNIVEFIKYKLK
VAAGYEVYNYVDKPDLN NQLVAEVEQSLNKKI PSMHL
PYPLGMLGGYCFDILSKITGKKYAVSSVRVKKFCATTQFDATKVHSSGFVAPYTLSQ
GLDRTLQ YE F VHAKKDD I T F VS E

SEQ ID NO: 3 - Secuencia de nucleótidos de gne de *E. coli* 055

Locus AF461121_1 BCT 02-MAY-2002

Definición (Gne de UDP-GlcNAc 4-epimerasa [Escherichia coli])

Acceso AAL67550

Longitud: 993

Tipo: ADN

Organismo: *E. coli* 055

Secuencia:

ATGAACGATA ACGTTTTGCT CATAGGAGCT TCCGGATTCG TAGGAACCCG

ACTACTTGAA ACGGCAATTG CTGACTTTAA TATCAAGAAC CTGGACAAAC

AGCAGAGCCA CTTTTATCCA GAAATCACAC AGATTGGTGA TGTTCGTGAT

CAACAGGCAC TCGACCAGGC GTTAGCCGGT TTTGACACTG TTGTGCTACT

GGCAGCGGAA CACCGCGATG ACGTCAGCCC TACTTCTCTC TATTATGATG

TCAACGTTCA GGGTACCCGC AATGTGCTGG CGGCCATGGA AAAAAATGGC

GTTAAAAATA TCATCTTTAC CAGTTCCGTT GCTGTTTATG GTTTGAACAA

ACACAACCCT GACGAAAAC C ATCCACACGA TCCTTTCAAC CACTACGGCA

AAAGTAAGTG GCAGGCAGAG GAAGTGCTGC GTGAATGGTA TAACAAAGCA

CCAACAGAAC GTTCATTAAC CATCATCCGT CCTACCGTTA TCTTCGGTGA

ACGGAACCGC GGTAACGTCT ATAACTTGCT GAAACAGATC GCTGGCGGCA

AGTTTATGAT GGTGGGCGCA GGGACTAACT ATAAGTCCAT GGCTTATGTT

GGAAACATTG TTGAGTTTAT CAAGTACAAA CTGAAGAATG TTGCCGCAGG
TTACGAGGTT TATAACTACG TTGATAAGCC AGACCTGAAC ATGAACCAGT
TGGTTGCTGA AGTTGAACAA AGCCTGAACA AAAAGATCCC TTCTATGCAC
TTGCCTTACC CACTAGGAAT GCTGGGTGGA TATTGCTTTG ATATCCTGAG
CAAAATTACG GGCAAAAAAT ACGCTGTCAG CTCTGTGCGC GTGAAAAAAT
TCTGCGCAAC AACACAGTTT GACGCAACGA AAGTGCATTC TTCAGGTTTT
GTGGCACCGT ATACGCTGTC GCAAGGTCTG GATCGAACTC TGCAGTATGA
ATTCGTCCAT GCCAAAAAAG ACGAGATAAC GTTTGTTTCT GAG

SEQ ID NO: 4 - Secuencia de aminoácidos de Gne de *E. coli* 055 UDP-GlcNAc 4-epimerasa

Locus AF461 121_1

Definición (Gne de UDP-GlcNAc 4-epimerasa [Escherichia coli])

Acceso AAL67550

Longitud: 331 aa lineal

Tipo: PRT

Organismo: *E. coli* 055

Secuencia:

mndnvlliga sgfvgtrlle taiadfnikn ldkqqshfyp eitqigdvrd qqaldqalag fdtwllaae hrddvsptsl
yydvnvqgtr nvlaamekng vkniiftssv avyglnkhnp denhphdpfn hygkskwqae evlrewynka
ptersltiir ptvifgernr gnvynllkqi aggkfmmvga gtnyksmayv gnivefikyk lknvaagyev
ynyvdkpdln mnqlvaeveq slnkkipsmh lpyplgmlgg ycfdilskit gkkyavssvr vkkfcattqf
datkvhssgf vapytlsqgl drtlqyefvh akkdditfvs e

SEQ ID NO: 5 - Secuencia de nucleótidos de gnel de *E. coli* 086

Locus AAO37706 BCT 06-DEC-2005

Definición UDP-GlcNAc C4-epimerasa [Escherichia coli 086].

Acceso AAO37706

Longitud: 993

Tipo: ADN

Organismo: *E. coli* 086

Secuencia:

ATGAACGATA ACGTTTTGCT CATAGGAGCT TCCGGATTCG TAGGAACCCG

ACTACTTGAA ACGGCAATTG CTGACTTTAA TATCAAGAAC CTGGACAAAC

AGCAGAGCCA CTTTTATCCA GAAATCACAC AGATTGGTGA TGTTCGTGAT

CAACAGGCAC TCGACCAGGC GTTAGCCGGT TTTGACACTG TTGTACTACT

GGCAGCGGAA CACCGCGATG ACGTCAGCCC TACTTCTCTC TATTATGATG

TCAACGTTCA GGGTACCCGC AATGTGCTGG CGGCCATGGA AAAAAATGGC

GTTAAAAATA TCATCTTTAC CAGTTCCGTT GCTGTTTATG GTTTGAACAA

ACACAACCCT GACGAAAACC ATCCACACGA CCCTTTCAAC CACTACGGCA

AAAGCAAGTG GCAGGCGGAG GAAGTGCTGC GTGAATGGTA TAACAAAGCA

CCAACAGAAC GTTCATTAAC TATCATCCGT CCTACCGTTA TCTTCGGTGA

ACGCAACCGC GGTAACGTCT ATAACTTGCT GAAACAGATC GCTGGCGGCA

AGTTTATGAT GGTGGGCGCA GGGACTAACT ATAAGTCCAT GGCTTATGTT

GGAAACATTG TTGAGTTTAT CAAGTACAAA CTGAAGAATG TTGCCGCAGG

TTACGAGGTT TATAACTACG TTGATAAGCC AGACCTGAAC ATGAACCAGT

TGGTTGCTGA AGTTGAACAA AGCCTGAACA AAAAGATCCC TTCTATGCAC

TTGCCTTACC CACTAGGAAT GCTGGGTGGA TATTGCTTTG ATATCCTGAG

CAAAATTACG GGCAAAAAAT ACGCTGTCAG CTCTGTGCGC GTGAAAAAAT

TCTGCGCAAC AACACAGTTT GACGCAACGA AAGTGCATTC TTCAGGTTTT

GTGGCACCGT ATACGCTGTC GCAAGGTCTG GATCGAACTC TGCAGTATGA

ATTCGTCCAT GCCAAAAAAG ACGACATAAC GTTTGTTTCT GAG

SEQ ID NO: 6 - Secuencia de aminoácidos de UDP-GlcNAc C4-epimerasa de *E. coli* 086

Locus AAO37706

Definición UDP-GlcNAc C4-epimerasa [Escherichia coli 086].

Acceso AAO37706

Longitud: 331 aa lineal

Tipo: PRT

Organismo: *E. coli* 086

Secuencia:

mndnvlliga sgfvgtrlle taiadfnikn ldkqqshfyp eitqigdvrd qqaldqalag fdtwllaae hrddvsptsl
yydvnvqgtr nvlaamekng vkniiftssv avyglnkhnp denhphdpfn hygkskwqae evlrewynka
ptersltiir ptvifgernr gnvynllkqi aggkfmmvga gtnyksmayv gnivefikyk lknvaagyev
ynyvdkpdln mnqlvaeveq slnkkipsmh lpyplgmlgg ycfdilskit gkkyavssvr vkkfcattqf
datkvhssgf vapytlsqgl drtlqyefvh akkdditfvs e

SEQ ID NO: 7 - Secuencia de nucleótidos de gne de *Shigella* boydii 018

Locus ACD09753 BCT 05-MAY-2008

Definición UDP-N-acetilglucosamina 4-epimerasa [*Shigella* boydii CDC 3083-94].

Acceso ACD09753

Longitud: 993

Tipo: ADN

Organismo: *Shigella* boydii 018

Secuencia:

ATGAACGATA ACGTTTTGCT CATAGGAGCT TCCGGATTCG TAGGAACCCG
ACTACTTGAA ACGGCAATTG CTGACTTTAA TATCAAGAAC CTGGACAAAC
AGCAGAGCCA TTTTTATCCA GCAATCACAC AGATTGGCGA TGTTCGTGAT
CAACAGGCAC TCGACCAGGC GTTAGCCGGT TTTGACACTG TTGTACTACT
GGCAGCGGAA CACCGCGATG ACGTCAGCCC TACTTCTCTC TATTATGATG
TCAACGTTCA GGGTACCCGC AATGTGCTGG CGGCCATGGA AAAAAATGGC
GTTAAAAATA TCATCTTTAC CAGTTCCGTT GCTGTTTATG GTTTGAACAA
ACACAACCCT GACGAAAACC ATCCACACGA CCCTTTCAAC CACTACGGCA
AAAGTAAGTG GCAGGCAGAG GAAGTGCTGC GTGAATGGTA TAACAAAGCA
CCAACAGAAC GTTCATTAAC CATCATCCGT CCTACCGTTA TCTTCGGTGA

ACGCAACCGC GGTAACGTCT ATAACTTGCT GAAACAGATC GCTGGCGGCA

AGTTTATGAT GGTGGGCGCA GGGACTAACT ATAAGTCCAT GGCTTATGTT
GGAAACATTG TTGAGTTTAT CAAGTACAAA CTGAAGAATG TTGCCGCAGG
TTATGAGGTT TATAACTATG TTGATAAGCC AGACCTGAAC ATGAACCAGT
TGGTTGCTGA AGTTGAACAA AGCCTGAACA AAAAGATCCC TTCTATGCAC
TTGCCTTACC CACTAGGAAT GCTGGGTGGA TATTGCTTTG ATATCCTGAG
CAAAATTACG GGCAAAAAAT ACGCTGTCAG CTCTGTGCGC GTGAAAAAAT
TCTGCGCAAC AACACAGTTT GACGCAACGA AAGTGCATTC TTCAGGTTTT
GTGGCACCGT ATACGCTGTC GCAAGGTCTG GATCGAACTC TGCAGTATGA
ATTCGTCCAT GCCAAAAAAG ACGACATAAC GTTTGTTTCT GAG

SEQ ID NO: 8 - Secuencia de aminoácidos de UDP-N-acetilglucosamina 4-epimerasa de *Shigella* boydii 018

Locus ACD09753

Definición UDP-N-acetilglucosamina 4-epimerasa [*Shigella* boydii CDC 3083- 94].

Acceso ACD09753

Longitud: 331 aa lineal

Tipo: PRT

Organismo: *Shigella* boydii 018

Secuencia:

mndnvlliga sgfvgtrlle taiadfnikn ldkqqshfyp aitqigdvrd

qqaldqalag fdtwllaae hrddvsptsl yydvnvqgtr nvlaamekng

vkniiftssv avyglnkhnp denhphdpfn hygkskwqae evlrewynka

pterslliir ptvifgernr gnvynllkqi aggkfmmvga gtnyksmayv

gnivefikyk lknvaagyev ynyvdkpdln mnqlvaeveq slrikkipsmh

lpyplgmlgg ycfdilskit gkkyavssvr vkkfcattqf datkvhssgf

vapytlsqgl drtlqyefvh akkdditfvs e

SEQ ID NO: 9 - Secuencia de nucleótidos de gne de *Salmonella enterica* 030

Locus AA V34516 BCT 25-OCT-2004

Definición UDP-GlcNAc 4-epimerasa [*Salmonella enterica* subesp. *salamae* serovar Greenside].

Acceso AAV34516

Longitud: 993

Tipo: ADN

Organismo: *Salmonella enterica* 030

Secuencia:

ATGAACGATA ACGTTTTGCT CATTGGTGCT TCCGGATTCG TAGGAACCCG
ACTCCTTGAA ACGGCAGTGG ATGATTTTAA TATCAAGAAC CTGGATAAAC
AGCAAAGCCA TTTCTACCCA GAGATTACAC ACATTGGCGA TGTTCGTGAC
CAACAAATCC TTGACCAGAC GTTGGTGGGT TTTGACACCG TAGTACTATT
GGCTGCGGAG CATCGTGATG ATGTTAGTCC TACCTCGCTT TATTATGATG
TCAACGTCCA GGGAACGCGT AATGTACTGG CGGCGATGGA AAAAAATGGT
GTAAAAAATA TCATTTTTAC CAGTTCCGTT GCAGTTTATG GACTCAACAA
GAAAAATCCT GACGAAACGC ACCCTCACGA TCCCTTTAAT CATTACGGAA
AAAGTAAATG GCAAGCAGAA GAAGTTCTGC GTGAGTGGCA TGCTAAAGCG
CCGAATGAGC GTTCTTTGAC CATAATTCGT CCTACCGTTA TTTTCGGGGA
GCGTAACCGC GGTAATGTAT ACAATCTCTT GAAACAGATC GCTGGTGGTA
AATTTGCGAT GGTTGGTCCG GGAACTAACT ATAAATCAAT GGCTTATGTT

GGTAATATCG TTGAGTTTAT CAAATTCAAA CTCAAGAATG TTACGGCGGG
CTATGAAGTT TATAATTATG TTGATAAACC TGATCTGAAT ATGAATCAAT
TGGTTGCTGA AGTAGAGCAG AGCCTGGGCA AAAAAATACC ATCGATGCAC
CTTCCATATC CATTAGGTAT GCTGGGGGGT TACTGTTTCG ATATCCTGAG
CAAAGTAACG GGCAAGAAGT ACGCTGTAAG TTCGGTTCGT GTTAAAAAAT
TCTGTGCGAC AACGCAGTTT GATGCAACAA AAGTGCATTC TTCTGGTTTT
GTTGCGCCAT ACACCTTATC TCAGGGGTTG GATCGTACAC TGCAATATGA
ATTTGTT CAT GCAAAGAAAG ATGACATTAC ATTCGTTTCA GAG

SEQ ID NO: 10 - Secuencia de aminoácidos de UDP-GlcNAc 4-epimerasa de *Salmonella enterica* 030

Locus AAV34516

Definición UDP-GlcNAc 4-epimerasa [*Salmonella enterica* subesp. salamae serovar Greenside].

Acceso AAV34516

Longitud: 331 aa lineal

Tipo: PRT

Organismo: *Salmonella enterica* 030

Secuencia:

mndnvlliga sgfvgtrlle tavddfnikn ldkqqshfyp eithigdvrd
qqildqtlvg fdtwllaae hrddvsptsl yydvnvqgtr nvlaamekng
vkniiftssv avyglnkknp dethphdpfn hygkskwqae evlrewhaka
pnersltiir ptvifgernr gnvynllkqi aggkfamvgp gtnyksmayv
gnivefikfk lknvtagyev ynyvdkpdln mnqlvaeveq slgkkipsmh
lpyplgmlgg ycfdilskvt gkkyavssvr vkkfcattqf datkvhssgf
vapytlsqgl drtlqyefvh akkdditfvs e

SEQ ID NO: 11 - Secuencia de nucleótidos de gne de *C. jejuni*

Locus YP_002344524 BCT 14-SEP-2010

Definición UDP-GlcNAc/Glc 4-epimerasa [*Campylobacter jejuni* subesp. jejuni

Acceso YP_002344524

Longitud: 987

Tipo: ADN

Organismo: *C. jejuni*

Secuencia:

ATGAAAATTCTTATTAGCGGTGGTGCAGGTTATATAGGTTCTCATACTTTAAGAC
AATT
TTTAAAAACAGATCATGAAATTTGTGTTTTAGATAATCTTTCTAAGGGTTCTAAAA
TCG
CAATAGAAGATTTGCAAAAAACAAGAGCTTTTAAATTTTTCGAACAAGATTTAAG
TGAT
TTTCAAGGCGTAAAAGCATTGTTTGAGAGAGAAAAATTTGACGCTATTGTGCATT
TTGC
AGCAAGCATTGAAGTTTTTGAAAGTATGCAAAATCCTTTAAAATATTATATGAACA
ACA
CTGTTAATACGACAAATCTCATCGAAACTTGTTTGCAAACTGGAGTGAATAAATT
TATA

TTTTCTTCAACGGCGGCCACTTATGGCGAACCACAAACTCCCGTTGTGAGCGAA
ACAAG
TCCTTTAGCACCTATTAATCCTTATGGGCGTAGTAAGCTTATGAGTGAAGAAGTT
TTGC
GTGATGCAAGTATGGCAAATCCTGAATTTAAGCATTGTATTTTAAGATATTTTAAT
GTT
GCAGGTGCTTGTATGGATTATACTTTAGGACAACGCTATCCAAAAGCGACTTTG
CTTAT
AAAAGTTGCAGCTGAATGTGCCGCAGGAAAACGTGATAAACTTTTCATATTTGG
CGATG
ATTATGATACAAAAGATGGTACTTGCATAAGAGATTTTATCCATGTAGATGATAT
TTCA
AGTGCACATTTAGCGGCTTTGGATTATTTAAAAGAGAATGAAAGCAATGTTTTTA
ATGT
AGGTTATGGACATGGTTTTAGCGTAAAAGAAGTGATTGAAGCGATGAAAAAAGT
TAGCG
GAGTGGATTTTAAAGTAGAACTTGCCCCACGCCGTGCGGGTGATCCTAGTGTAT
TGATT
TCTGATGCAAGTAAAATCAGAAATCTTACTTCTTGGCAGCCTAAATATGATGATT
TAGA GCTTATTTGTAAATCTGCTTTTGATTGGGAAAAACAGTGTTAA

SEQ ID NO: 12 - Secuencia de aminoácidos de UDP-GlcNAc/Glc 4-epimerasa de *C. jejuni*

Locus YP_002344524

Definición UDP-GlcNAc/Glc 4-epimerasa [*Campylobacter jejuni* subesp. jejuni]

Acceso YP_002344524

Longitud: 328 aa lineal

Tipo: PRT

Organismo: *C. jejuni*

Secuencia:

mkilisggag yigshtlrqf lktdheicvl dnlskgskia iedlqktraf kffeqdlsdf qgvkalfere kfdaivhfaa sievfesmqn plkyymnntv nttnlietcl qtgvnkfifs

staatygepq tpwsetspl apinpygrsk lmseevlrda smanpefkhc ilryfnvaga cmdytlgqry pkatllikva aecaagkrdk lfifgddydt kdgtcirdfi hvddissahl

aaldylkene snvfnvgygh gfsvkeviea mkkvsgvdfk velaprragd psvlisdask irnltswqpk yddlelicks

afdwekqc

SEQ ID NO: 13 - Secuencia de nucleótidos de galE de *E. coli* K12

Locus AP_001390 BCT 30-APR-2010

Definición UDP-galactosa-4-epimerasa [Escherichia coli cepa K-12 sustr. W3110].

Acceso AP_001390

Longitud: 1.017

Tipo: ADN

Organismo: *E. coli* K12

Secuencia:

ATGAGAGTTCTGGTTACCGGTGGTAGCGGTTACATTGGAAGTCATACCTGTGTGCAA

TTACTGCAAAACGGTCATGATGTCATCATTCTTGATAACCTCTGTAACAGTAAGCGC

AGCGTACTGCCTGTTATCGAGCGTTTAGGCGGCAAACATCCAACGTTTGTTGAA
GGC

GATATTCGTAACGAAGCGTTGATGACCGAGATCCTGCACGATCACGCTATCGAC
AC

CGTGATCCACTTCGCCGGGCTGAAAGCCGTGGGCGAATCGGTACAAAAACCGC
TGG

AATATTACGACAACAATGTCAACGGCACTCTGCGCCTGATTAGCGCCATGCGC
GCC

GCTAACGTCAAAAACTTTATTTTTAGCTCCTCCGCCACCGTTTATGGCGATCAG
CCC

AAAATTCCATACGTTGAAAGCTTCCCGACCGGCACACCGCAAAGCCCTTACGG
CAA

AAGCAAGCTGATGGTGGAACAGATCCTCACCGATCTGCAAAAAGCCCAGCCGG
ACT

GGAGCATTGCCCTGCTGCGCTACTTCAACCCGGTTGGCGCGCATCCGTCGGG
CGATA

TGGGCGAAGATCCGCAAGGCATTCCGAATAACCTGATGCCATACATCGCCCAG
GTT

GCTGTAGGCCGTCGCGACTCGCTGGCGATTTTTGGTAACGATTATCCGACCGA
AGAT

GGTACTGGCGTACGCGATTACATCCACGTAATGGATCTGGCGGACGGTCACGT
CGT

GGCGATGGAAAAACTGGCGAACAAGCCAGGCGTACACATCTACAACCTCGGCG
CTG

GCGTAGGCAACAGCGTGCTGGACGTGGTTAATGCCTTCAGCAAAGCCTGCGGC
AAA

CCGGTTAATTATCATTTTGCACCGCGTCGCGAGGGCGACCTTCCGGCCTACTG
GGCG

GACGCCAGCAAAGCCGACCGTGAACTGAACTGGCGCGTAACGCGCACACTCG
ATGA

AATGGCGCAGGACACCTGGCACTGGCAGTCACGCCATCCACAGGGATATCCC
GATT

AA

SEQ ID NO: 14 - Secuencia de aminoácidos de UDP-galactosa-4-epimerasa de *E. coli* K12

Locus AP_001390

Definición UDP-galactosa-4-epimerasa [Escherichia coli cepa K-12 sustr. W31 10].

Acceso AP_001390

Longitud: 338 aa lineal

Tipo: PRT

Organismo: *E. coli* 12

Secuencia:

mrvlvtggsgyigshtcvqllqnghdviildnlcnskrsvlpvierlggkhptfvegdi
mealmteilhdhaidtvihfaglkavgesvqkpleyydnnvngtlrlisamraanvkn f
ifsssatvygdqpkipyvesfptgtpqspygksklmveqiltdlqkaqpdwsiallry
fnpvgahpsgdmgedpqgipnnlmpyiaqvavgrrdslaifgndyptedgtgvrdyihv
mdladghwameklankpgvhiynlgagvgnsvldvvnafskacgkpvnyhfaprregd
lpaywadaskadrelnwrvtrtldemaqdtwhwqsrhpqgypd

SEQ ID NO: 15 - Secuencia de nucleótidos de gne2 de *E. coli* 086

Locus AAV85952 BCT 27-MAR-2005

Definición Gne [Escherichia coli 086].

Acceso AAV85952

Longitud: 1.020

Tipo: ADN

Organismo: *E. coli* 086

Secuencia:

ATGGTGATTT TCGTAACAGG CGGTGCAGGA TATATTGGAT CCCATACCAT
ACTTGAGTTA CTTAATAATG GTCATGATGT CGTTTCGATA GATAATTTTG
TCAATTCCTC TATAGAATCA TTAAAAAGAG TAGAGCAAAT AACTAATAAG
AAAATTATTT CTTATCAAGG TGATATCCGT GATAAAAATC TACTTGATGA
GATTTTTTCA AGACACCATA TCGATGCTGT AATTCACTTT GCATCGTTAA
AATCTGTAGG TGAGTCTAAG TTAAAGCCCT TAGAGTATTA TTCTAATAAT
GTTGGTGGAA CTTTAGTATT ACTTGAATGC ATGAAGAGAT ATAACATTAA

TAAAATGATA TTTAGCTCTT CTGCTACTGT TTATGGGAGT AACAGTATCC

CTCCCCATAC GGAAGATAGA CGAATTGGTG AAACTACAAA CCCATATGGG

ACATCGAAAT TTATAATAGA AATAATTTTG AGTGATTATT GTGATAGTGA

TAATAATAAA TCAGTAATTG CACTGCGTTA CTTTAATCCA ATCGGAGCAC

ATAAGTCCGG GATGATTGGT GAAAATCCTA ACGGGATCCC TAATAATCTG

GTTCCTTATA TATCTAAAGT TGCACAAAAT CAACTTCCTG TATTAAATAT

TTATGGCAAC GATTATCCAA CTAAAGATGG TACAGGAGTA AGAGACTATA

TACATGTCTG TGATTTGGCT AAAGGGCATG TTAAAGCATT AGAATATATG

TTTTTAAATG ATGTCAATTA TGAAGCTTTT AATTTAGGTA CTGGTCAAGG

TTATTCTGTT TTAGAGATTG TAAAAATGTT TGAGATAGTC ACTAAAAAGA

GTATAC CTGT TGCTATTTGT AATAGACGTG AGGGGGATGT TGCGGAGTCA

TGGGCGTCTG CTGATTTGGC ACATAAAAAG CTTTCCTGGA AAGCGGAAAA

AAATTTGAAA GAAATGATCG AAGATGTATG GCGTTGGCAA ACAAACAATC

CAAATGGATA TAAAAAATAA

SEQ ID NO: 16 - Secuencia de aminoácidos de Gne de *E. coli* 086

Locus AAV85952

Definición Gne [Escherichia coli 086].

Acceso AAV85952

Longitud: 339 aa (gne2) lineal

Tipo: PRT

Organismo: *E. coli* 086

Secuencia:

mvifvtggag yigshtilel lnnghdwsi dnfvnssies lkrveqitnk kiisyqgdir dknlldeifs rhhidavihf
aslksvgesk lkpleyysnn vggtlvliec mkryninkmi fsssatvygs nsipphtedr rigettnpyg
tskfiieiil sdycdsdnnk svialryfnp igahksgmig enpngipnnl vpyiskvaqn qlpvlniygn
dyptkdgtgv rdyihvcdla kghvkaleym . flndvnyeaf nlgtgqgysv leivkmfeiv tkksipvaic
nrregdvaes wasadlahkk lswkaeknlk emiedvwrwq tnnpngykk

SEQ ID NO: 17 - Secuencia de nucleótidos para el oligonucleótido sintético Z3206-Fw (cebador) que codifica un extremo de Z3206;

Sitios de restricción subrayados

Longitud: 30

Tipo: ADN

Secuencia: AAACCCGGGATGAACGATAACGTTTTGCTC

SEQ ID NO: 18 - Secuencia de nucleótidos para el oligonucleótido sintético Z3206-RvHA (cebador) que codifica un extremo de Z3206 con un marcador hemaglutinina (marcador HA); los sitios de restricción están subrayados

Longitud: 60

Tipo: ADN

Organismo:

Secuencia:

AAATCTAGATTAAGCGTAATCTGGAAC

ATCGTATGGGTACTCAGAAACAAACGTTATGTC

SEQ ID NO: 19 - Secuencia de nucleótidos para oligonucleótido sintético gne-Fw (cebador) con sitios de restricción subrayados

Longitud: 29

Tipo: ADN

Organismo:

Secuencia: AAACCATGGATGAAAATTCTTATTAGCGG

SEQ ID NO: 20 - Secuencia de nucleótidos para oligonucleótido sintético gne-RV (cebador) con sitios de restricción subrayados

Longitud: 57

Tipo: ADN

Organismo:

Secuencia:

AAATCTAGATTAAGCGTAATCTGGAAC

ATCGTATGGGTAGCACTGTTTTTCCCAATC

SEQ ID NO: 21 - Secuencia de nucleótidos para un oligonucleótido que contiene sitios de restricción para la enzima de restricción Nhel

Longitud: 1 1

Tipo: ADN

Organismo:

Secuencia: AAAAAGCTAGC

SEQ ID NO: 22 - Secuencia de nucleótidos para un oligonucleótido que contiene sitios de restricción para la enzima de restricción Ascl

Longitud: 8

Tipo: ADN

Organismo:

Secuencia: CCGCGCGG

SEQ ID NO: 23 - Secuencia de nucleótidos para el plásmido pMLBAD:Z3206 (*E. coli* 0157 inserción en el plásmido) que codifica Z3206 con un marcador de hemaglutinina en el extremo C

Definición Unión del producto dentro del plásmido Z3206-pMLB AD*

Características Localización/Calificadores

CDS 2105..3098

/marcador=Z3206

CDS 3098..3127

/marcador=HA

Longitud: 7794 pb

Tipo: ADN circular UNA

Secuencia:

1 TCTACGGGGT CTGACGCTCA GTGGAACGAA ATCGATGAGC TCGCACGAAC CCAGTTGACA

61 TAAGCCTGTT CGGTTCGTAA ACTGTAATGC AAGTAGCGTA TGCGCTCACG CAACTGGTCC

121 AGAACCTTGA CCGAACGCAG CGGTGGTAAC GGCGCAGTGG CGGTTTTCAT GGCTTGTTAT

181 GACTGTTTTT TTGTACAGTC TAGCCTCGGG CATCCAAGCT
AGCTAAGCGC GTTACGCCGT

241 GGGTCGATGT TTGATGTTAT GGAACAGCAA CGATGTTACG
CAGCAGGGTA GTCGCCCTAA

301 AACAAAGTTA GGCAGCCGTT GTGCTGGTGC TTTCTAGTAG
TTGTTGTGGG GTAGGCAGTC

361 AGAGCTCGAT TTGCTTGTCG CCATAATAGA TTCACAAGAA GGATTCGACA
TGGGTCAAAG

421 TAGCGATGAA GCCAACGCTC CCGTTGCAGG GCAGTTTGCG
CTTCCCCTGA GTGCCACCTT

481 TGGCTTAGGG GATCGCGTAC GCAAGAAATC TGGTGCCGCT
TGGCAGGGTC AAGTCGTCGG

541 TTGGTATTGC ACAAAACTCA CTCCTGAAGG CTATGCGGTC
GAGTCCGAAT CCCACCCAGG

601 CTCAGTGCAA ATTTATCCTG TGGCTGCACT TGAACGTGTG
GCCTAAGCGA TATCTTAGGA

661 TCTCCCATCG GTGATGTCGG CGATATAGGC GCCAGCAACC
GCACCTGTGG CGCCGGTGAT

721 GCCGGCCACG ATGCGTCCGG CGTAGAGGAT CTGCTCATGT
TTGACAGCTT ATCATCGATG

781 CATAATGTGC CTGTCAAATG GACGAAGCAG GGATTCTGCA
AACCCTATGC TACTCCGTCA

841 AGCCGTCAAT TGTCTGATTC GTTACCAATT ATGACAACTT GACGGCTACA
TCATTCACTT

901 TTTCTTCACA ACCGGCACGG AACTCGCTCG GGCTGGCCCC
GGTGCATTTT TTAAATACCC

961 GCGAGAAATA GAGTTGATCG TCAAAACCAA CATTGCGACC

GACGGTGGCG ATAGGCATCC

1021 GGGTGGTGCT CAAAAGCAGC TTCGCCTGGC TGATACGTTG
GTCCTCGCGC CAGCTTAAGA

1081 CGCTAATCCC TAACTGCTGG CGGAAAAGAT GTGACAGACG

CGACGGCGAC AAGCAAACAT

1141 GCTGTGCGAC GCTGGCGATA TCAAAATTGC TGTCTGCCAG
GTGATCGCTG ATGTACTGAC

1201 AAGCCTCGCG TACCCGATTA TCCATCGGTG GATGGAGCGA
CTCGTTAATC GCTTCCATGC

1261 GCCGCAGTAA CAATTGCTCA AGCAGATTTA TCGCCAGCAG
CTCCGAATAG CGCCCTTCCC

1321 CTTGCCCGGC GTTAATGATT TGCCCAAACA GGTCGCTGAA
ATGCGGCTGG TGCGCTTCAT

1381 CCGGGCGAAA GAACCCCGTA TTGGCAAATA TTGACGGCCA

GTTAAGCCAT TCATGCCAGT

1441 AGGCGCGCGG ACGAAAGTAA ACCCACTGGT GATACCATTC

GCGAGCCTCC GGATGACGAC

1501 CGTAGTGATG AATCTCTCCT GGCGGGAACA GCAAAATATC
ACCCGGTCGG CAAACAAATT

1561 CTCGTCCCTG ATTTTTCACC ACCCCCTGAC CGCGAATGGT
GAGATTGAGA ATATAACCTT

1621 TCATTCCCAG CGGTCGGTCG ATAAAAAAAT CGAGATAACC
GTTGGCCTCA ATCGGCGTTA

1681 AACCCGCCAC CAGATGGGCA TTAAACGAGT ATCCCGGCAG

CAGGGGATCA TTTTGCGCTT

1741 CAGCCATACT TTTCATACTC CCGCCATTCA GAGAAGAAAC
CAATTGTCCA TATTGCATCA

1801 GACATTGCCG TCACTGCGTC TTTTACTGGC TCTTCTCGCT
AACCAAACCG GTAACCCCGC

1861 TTATTAAAAG CATTCTGTAA CAAAGCGGGA CCAAAGCCAT

GACAAAAACG CGTAACAAAA

1921 GTGTCTATAA TCACGGCAGA AAAGTCCACA TTGATTATTT
GCACGGCGTC ACACTTTGCT

1981 ATGCCATAGC ATTTTTATCC ATAAGATTAG CGGATCCTAC CTGACGCTTT
TTATCGCAAC

2041 TCTCTACTGT TTCTCCATAC CCGTTTTTTT GGGCTAGCAG
GAGGAATTCA CCATGGTACC

2101 CGGGATGAAC GATAACGTTT TGCTCATAGG AGCTTCCGGA
TTCGTAGGAA CCCGACTACT

2161 TGAAACGGCA ATTGCTGACT TTAATATCAA GAACCTGGAC

AAACAGCAGA GCCACTTTTA

2221 TCCAGAAATC ACACAGATTG GCGATGTTCG CGATCAACAG
GCACTCGACC AGGCGTTAGT

2281 CGGTTTTGAC ACTGTTGTAC TACTGGCAGC GGAACACCGC
GATGACGTCA GCCCTACTTC

2341 TCTCTATTAT GATGTCAACG TTCAGGGTAC CCGCAATGTG
CTGGCGGCCA TGGAAAAAAA

2401 TGGCGTTAAA AATATCATCT TTACCAGTTC CGTTGCTGTT TATGGTTTGA
ACAAACACAA

2461 CCCTGACGAA AACCATCCAC ACGACCCTTT CAACCACTAC
GGCAAAAGTA AGTGGCAGGC

2521 AGAGGAAGTG CTGCGTGAAT GGTATAACAA AGCACCAACA

GAACGTTCAT TAACCATCAT

2581 CCGTCCTACC GTTATCTTCG GTGAACGCAA CCGCGGTAAC
GTCTATAACT TGCTGAAACA

2641 GATCGCTGGC GGCAAGTTTA TGATGGTGGG CGCAGGGACT

AACTATAAGT CCATGGCTTA

2701 TGTTGGAAAC ATTGTTGAGT TTATCAAGTA CAAACTGAAG
AATGTTGCCG CAGGTTATGA

2761 GGTTTATAAC TACGTTGATA AGCCAGACCT GAACATGAAC
CAGTTGGTTG CTGAAGTTGA

2821 ACAAAGCCTG AACAAAAAGA TCCCTTCTAT GCACTTGCCT
TACCCACTAG GAATGCTGGG

2881 TGGATATTGC TTTGATATCC TGAGCAAAAT TACGGGCAAA
AAATACGCTG TCAGCTCAGT

2941 GCGCGTGAAA AAATTCTGCG CAACAACACA GTTTGACGCA

ACGAAAGTGC ATTCTTCAGG

3001 TTTTGTGGCA CCGTATACGC TGTCGCAAGG TCTGGATCGA
ACACTGCAGT ATGAATTCGT

3061 TCATGCCAAA AAAGACGACA TAACGTTTGT TTCTGAGTAC
CCATACGATG TTCCAGATTA

3121 CGCTTAATCT AGAGTCGACC TGCAGGCATG CAAGCTTGGC
TGTTTTGGCG GATGAGAGAA

3181 GATTTTCAGC CTGATACAGA TTAAATCAGA ACGCAGAAGC
GGTCTGATAA AACAGAATTT

3241 GCCTGGCGGC AGTAGCGCGG TGGTCCCACC TGACCCCATG
CCGAACTCAG AAGTGAAACG

3301 CCGTAGCGCC GATGGTAGTG TGGGGTCTCC CCATGCGAGA

GTAGGGAACT GCCAGGCATC

3361 AAATAAAACG AAAGGCTCAG TCGAAAGACT GGGCCTTTCG
TTTTATCTGT TGTTTGTCGG

3421 TGAACGCTCT CCTGAGTAGG ACAAATCCGC CGGGAGCGGA
TTTGAACGTT GCGAAGCAAC

3481 GGCCCGGAGG GTGGCGGGCA GGACGCCCGC CATAAACTGC

CAGGCATCAA ATTAAGCAGA

3541 AGGCCATCCT GACGGATGGC CTTTTTGCGT TTCTACAAAC
TCTTCCACTC ACTACAGCAG

3601 AGCCATTTAA ACAACATCCC CTCCCCCTTT CCACCGCGTC
AGACGCCCGT AGCAGCCCGC

3661 TACGGGCTTT TTCATGCCCT GCCCTAGCGT CCAAGCCTCA
CGGCCGCGCT CGGCCTCTCT

3721 GGCGGCCTTC TGGCGCTGAG GTCTGCCTCG TGAAGAAGGT
GTTGCTGACT CATACCAGGC

3781 CTGAATCGCC CCATCATCCA GCCAGAAAGT GAGGGAGCCA

CGGTTGATGA GAGCTTTGTT

3841 GTAGGTGGAC CAGTTGGTGA TTTTGAACTT TTGCTTTGCC
ACGGAACGGT CTGCGTTGTC

3901 GGGAAGATGC GTGATCTGAT CCTTCAACTC AGCAAAAGTT
CGATTTATTC AACAAAGCCG

3961 CCGTCCCGTC AAGTCAGCGT AATGCTCTGC CAGTGTTACA
ACCAATTAAC CAATTCTGAT

4021 TAGAAAAACT CATCGAGCAT CAAATGAAAC TGCAATTTAT TCATATCAGG
ATTATCAATA

4081 CCATATTTTT GAAAAAGCCG TTTCTGTAAT GAAGGAGAAA
ACTCACCGAG GCAGTTCCAT

4141 AGGATGGCAA GATCCTGGTA TCGGTCTGCG ATTCCGACTC
GTCCAACATC AATACAACCT

4201 ATTAATTTCC CCTCGTCAAA AATAAGGTTA TCAAGCGAGA
AATCACCATG AGTGACGACT

4261 GAATCCGGTG AGAATGGCAA AAGCTAAAAA GGCCGTAATA

TCCAGCTGAA CGGTCTGGTT

4321 ATAGGTACAT TGAGCAACTG ACTGAAATGC CTCAAAATGT
TCTTTACGAT GCCATTGGGA

4381 TATATCAACG GTGGTATATC CAGTGATTTT TTTCTCCATT TTAGCTTCCT
TAGCTCCTGA

4441 AAATCTCGAT AACTCAAAAA ATACGCCCGG TAGTGATCTT ATTTCATTAT
GGTGAAAGTT

4501 GGAACCTCTT ACGTGCCGAT CAACGTCTCA TTTTCGCCAA
AAGTTGGCCC AGGGCTTCCC

4561 GGTATCAACA GGGACACCAG GATTTATTTA TTCTGCGAAG
TGATCTTCCG TCACAGGTAT

4621 TTATTCGAAG ACGAAAGGGC CTCGTGATAC GCCTATTTTT
ATAGGTTAAT GTCATGATAA

4681 TAATGGTTTC TTAGACGTCA GGTGGCACTT TTCGGGGAAA
TGTGCGCGCC CGCGTTCCTG

4741 CTGGCGCTGG GCCTGTTTCT GGCGCTGGAC TTCCCGCTGT
TCCGTCAGCA GCTTTTCGCC

4801 CACGGCCTTG ATGATCGCGG CGGCCTTGGC CTGCATATCC
CGATTCAACG GCCCCAGGGC

4861 GTCCAGAACG GGCTTCAGGC GCTCCCGAAG GTCTCGGGCC
GTCTCTTGGG CTTGATCGGC

4921 CTTCTTGCGC ATCTCACGCG CTCCTGCGGC GGCCTGTAGG
GCAGGCTCAT ACCCCTGCCG

4981 AACCGCTTTT GTCAGCCGGT CGGCCACGGC TTCCGGCGTC
TCAACGCGCT TTGAGATTCC

5041 CAGCTTTTCG GCCAATCCCT GCGGTGCATA GGCGCGTGGC
TCGACCGCTT GCGGGCTGAT

5101 GGTGACGTGG CCCACTGGTG GCCGCTCCAG GGCCTCGTAG

AACGCCTGAA TGCGCGTGTG

5161 ACGTGCCTTG CTGCCCTCGA TGCCCCGTTG CAGCCCTAGA
TCGGCCACAG CGGCCGCAAA

5221 CGTGGTCTGG TCGCGGGTCA TCTGCGCTTT GTTGCCGATG
AACTCCTTGG CCGACAGCCT

5281 GCCGTCCTGC GTCAGCGGCA CCACGAACGC GGTCATGTGC
GGGCTGGTTT CGTCACGGTG

5341 GATGCTGGCC GTCACGATGC GATCCGCCCC GTACTTGTCC
GCCAGCCACT TGTGCGCCTT

5401 CTCGAAGAAC GCCGCCTGCT GTTCTTGGCT GGCCGACTTC
CACCATTCCG GGCTGGCCGT

5461 CATGACGTAC TCGACCGCCA ACACAGCGTC CTTGCGCCGC
TTCTCTGGCA GCAACTCGCG

5521 CAGTCGGCCC ATCGCTTCAT CGGTGCTGCT GGCCGCCCAG
TGCTCGTTCT CTGGCGTCCT

5581 GCTGGCGTCA GCGTTGGGCG TCTCGCGCTC GCGGTAGGCG
TGCTTGAGAC TGGCCGCCAC

5641 GTTGCCCATT TTCGCCAGCT TCTTGCATCG CATGATCGCG
TATGCCGCCA TGCCTGCCCC

5701 TCCCTTTTGG TGTCCAACCG GCTCGACGGG GGCAGCGCAA
GGCGGTGCCT CCGGCGGGCC

5761 ACTCAATGCT TGAGTATACT CACTAGACTT TGCTTCGCAA
AGTCGTGACC GCCTACGGCG

5821 GCTGCGGCGC CCTACGGGCT TGCTCTCCGG GCTTCGCCCT
GCGCGGTCGC TGCGCTCCCT

5881 TGCCAGCCCG TGGATATGTG GACGATGGCC GCGAGCGGCC

ACCGGCTGGC TCGCTTCGCT

5941 CGGCCCGTGG ACAACCCTGC TGGACAAGCT GATGGACAGG

CTGCGCCTGC CCACGAGCTT

6001 GACCACAGGG ATTGCCCACC GGCTACCCAG CCTTCGACCA
CATACCCACC GGCTCCAACT

6061 GCGCGGCCTG CGGCCTTGCC CCATCAATTT TTTTAATTTT
CTCTGGGGAA AAGCCTCCGG

6121 CCTGCGGCCT GCGCGCTTCG CTTGCCGGTT GGACACCAAG
TGGAAGGCGG GTCAAGGCTC

6181 GCGCAGCGAC CGCGCAGCGG CTTGGCCTTG ACGCGCCTGG

AACGACCCAA GCCTATGCGA

6241 GTGGGGGCAG TCGAAGGCGA AGCCCGCCCG CCTGCCCCCC

GAGCCTCACG GCGGCGAGTG

6301 CGGGGGTTCC AAGGGGGCAG CGCCACCTTG GGCAAGGCCG

AAGGCCGCGC AGTCGATCAA

6361 CAAGCCCCGG AGGGGCCACT TTTTGCCGGA GGGGGAGCCG

CGCCGAAGGC GTGGGGGAAC

6421 CCCGCAGGGG TGCCCTTCTT TGGGCACCAA AGAACTAGAT

ATAGGGCGAA ATGCGAAAGA

6481 CTTAAAAATC AACAACTTAA AAAAGGGGGG TACGCAACAG

CTCATTGCGG CACCCCCCGC

6541 AATAGCTCAT TGCGTAGGTT AAAGAAAATC TGTAATTGAC TGCCACTTTT
ACGCAACGCA

6601 TAATTGTTGT CGCGCTGCCG AAAAGTTGCA GCTGATTGCG
CATGGTGCCG CAACCGTGCG

6661 GCACCCTACC GCATGGAGAT AAGCATGGCC ACGCAGTCCA

GAGAAATCGG CATTCAAGCC

6721 AAGAACAAGC CCGGTCACTG GGTGCAAACG GAACGCAAAG

CGCATGAGGC GTGGGCCGGG

6781 CTTATTGCGA GGAAACCCAC GGCGGCAATG CTGCTGCATC
ACCTCGTGGC GCAGATGGGC

6841 CACCAGAACG CCGTGGTGGT CAGCCAGAAG ACACTTTCCA

AGCTCATCGG ACGTTCTTTG

6901 CGGACGGTCC AATACGCAGT CAAGGACTTG GTGGCCGAGC

GCTGGATCTC CGTCGTGAAG

6961 CTCAACGGCC CCGGCACCGT GTCGGCCTAC GTGGTCAATG
ACCGCGTGGC GTGGGGCCAG

7021 CCCCGCGACC AGTTGCGCCT GTCGGTGTTC AGTGCCGCCG
TGGTGGTTGA TCACGACGAC

7081 CAGGACGAAT CGCTGTTGGG GCATGGCGAC CTGCGCCGCA
TCCCGACCCT GTATCCGGGC

7141 GAGCAGCAAC TACCGACCGG CCCCGGCGAG GAGCCGCCCA

GCCAGCCCGG CATTCCGGGC

7201 ATGGAACCAG ACCTGCCAGC CTTGACCGAA ACGGAGGAAT

GGGAACGGCG CGGGCAGCAG

7261 CGCCTGCCGA TGCCCGATGA GCCGTGTTTT CTGGACGATG
GCGAGCCGTT GGAGCCGCCG

7321 ACACGGGTCA CGCTGCCGCG CCGGTAGCAC TTGGGTTGCG

CAGCAACCCG TAAGTGCGCT

7381 GTTCCAGACT ATCGGCTGTA GCCGCCTCGC CGCCCTATAC
CTTGTCTGCC TCCCCGCGTT

7441 GCGTCGCGGT GCATGGAGCC GGGCCACCTC GACCTGAATG

GAAGCCGGCG GCACCTCGCT

7501 AACGGATTCA CCGTTTTTAT CAGGCTCTGG GAGGCAGAAT AAATGATCAT ATCGTCAATT

7561 ATTACCTCCA CGGGGAGAGC CTGAGCAAAC TGGCCTCAGG

CATTTGAGAA GCACACGGTC

7621 ACACTGCTTC CGGTAGTCAA TAAACCGGTA AACCAGCAAT

AGACATAAGC GGCTATTTAA

7681 CGACCCTGCC CTGAACCGAC GACCGGGTCG AATTTGCTTT CGAATTTCTG CCATTCATCC

7741 GCTTATTATC ACTTATTCAG GCGTAGCACC AGGCGTTTAA GTCGACCAAT AACC

SEQ ID NO: 24 - Secuencia de nucleótidos para pMLBADrgne (*E. coli* 0157 inserción en plásmido) que codifica Gne con un marcador hemaglutinina en el extremo C

Locus gne-pMLBAD

Definición: unión de dig galE en pMLBAD did (NcoI-XbaI)

Características Localización/Calificadores

CDS 2097..3080

/marcador=galE

CDS 3081..3107

/marcador=HA

Región 3108..31 10

/marcador=stop

Longitud: 7776 pb

Tipo: ADN circular UNA

Secuencia:

1 TCTACGGGGT CTGACGCTCA GTGGAACGAA ATCGATGAGC TCGCACGAAC CCAGTTGACA

61 TAAGCCTGTT CGGTTCGTAA ACTGTAATGC AAGTAGCGTA TGCGCTCACG CAACTGGTCC

121 AGAACCTTGA CCGAACGCAG CGGTGGTAAC GGCGCAGTGG
CGGTTTTCAT GGCTTGTTAT

181 GACTGTTTTT TTGTACAGTC TAGCCTCGGG CATCCAAGCT
AGCTAAGCGC GTTACGCCGT

241 GGGTCGATGT TTGATGTTAT GGAACAGCAA CGATGTTACG
CAGCAGGGTA GTCGCCCTAA

301 AACAAAGTTA GGCAGCCGTT GTGCTGGTGC TTTCTAGTAG
TTGTTGTGGG GTAGGCAGTC

361 AGAGCTCGAT TTGCTTGTCG CCATAATAGA TTCACAAGAA GGATTCGACA
TGGGTCAAAG

421 TAGCGATGAA GCCAACGCTC CCGTTGCAGG GCAGTTTGCG
CTTCCCCTGA GTGCCACCTT

481 TGGCTTAGGG GATCGCGTAC GCAAGAAATC TGGTGCCGCT
TGGCAGGGTC AAGTCGTCGG

541 TTGGTATTGC ACAAAACTCA CTCCTGAAGG CTATGCGGTC
GAGTCCGAAT CCCACCCAGG

601 CTCAGTGCAA ATTTATCCTG TGGCTGCACT TGAACGTGTG
GCCTAAGCGA TATCTTAGGA

661 TCTCCCATCG GTGATGTCGG CGATATAGGC GCCAGCAACC
GCACCTGTGG CGCCGGTGAT

721 GCCGGCCACG ATGCGTCCGG CGTAGAGGAT CTGCTCATGT
TTGACAGCTT ATCATCGATG

781 CATAATGTGC CTGTCAAATG GACGAAGCAG GGATTCTGCA
AACCCTATGC TACTCCGTCA

841 AGCCGTCAAT TGTCTGATTC GTTACCAATT ATGACAACTT GACGGCTACA
TCATTCACTT

901 TTTCTTCACA ACCGGCACGG AACTCGCTCG GGCTGGCCCC
GGTGCATTTT TTAAATACCC

961 GCGAGAAATA GAGTTGATCG TCAAAACCAA CATTGCGACC

GACGGTGGCG ATAGGCATCC

1021 GGGTGGTGCT CAAAAGCAGC TTCGCCTGGC TGATACGTTG
GTCCTCGCGC CAGCTTAAGA

1081 CGCTAATCCC TAACTGCTGG CGGAAAAGAT GTGACAGACG

CGACGGCGAC AAGCAAACAT

1141 GCTGTGCGAC GCTGGCGATA TCAAAATTGC TGTCTGCCAG
GTGATCGCTG ATGTACTGAC

1201 AAGCCTCGCG TACCCGATTA TCCATCGGTG GATGGAGCGA
CTCGTTAATC GCTTCCATGC

1261 GCCGCAGTAA CAATTGCTCA AGCAGATTTA TCGCCAGCAG
CTCCGAATAG CGCCCTTCCC

1321 CTTGCCCGGC GTTAATGATT TGCCCAAACA GGTCGCTGAA
ATGCGGCTGG TGCGCTTCAT

1381 CCGGGCGAAA GAACCCCGTA TTGGCAAATA TTGACGGCCA

GTTAAGCCAT TCATGCCAGT

1441 AGGCGCGCGG ACGAAAGTAA ACCCACTGGT GATACCATTC

GCGAGCCTCC GGATGACGAC

1501 CGTAGTGATG AATCTCTCCT GGCGGGAACA GCAAAATATC
ACCCGGTCGG CAAACAAATT

1561 CTCGTCCCTG ATTTTTCACC ACCCCCTGAC CGCGAATGGT
GAGATTGAGA ATATAACCTT

1621 TCATTCCCAG CGGTCGGTCG ATAAAAAAAT CGAGATAACC
GTTGGCCTCA ATCGGCGTTA

1681 AACCCGCCAC CAGATGGGCA TTAAACGAGT ATCCCGGCAG

CAGGGGATCA TTTTGCGCTT

1741 CAGCCATACT TTTCATACTC CCGCCATTCA GAG A AG A A AC
CAATTGTCCA TATTGCATCA

1801 GACATTGCCG TCACTGCGTC TTTTACTGGC TCTTCTCGCT
AACCAAACCG GTAACCCCGC

1861 TTATTAAAAG CATTCTGTAA CAAAGCGGGA CCAAAGCCAT

GACAAAAACG CGTAACAAAA

1921 GTGTCTATAA TCACGGCAGA AAAGTCCACA TTGATTATTT
GCACGGCGTC ACACTTTGCT

1981 ATGCCATAGC ATTTTTATCC ATAAGATTAG CGGATCCTAC CTGACGCTTT
TTATCGCAAC

2041 TCTCTACTGT TTCTCCATAC CCGTTTTTTT GGGCTAGCAG
GAGGAATTCA CCATGGATGA

2101 AAATTCTTAT TAGCGGTGGT GCAGGTTATA TAGGTTCTCA TACTTTAAGA
CAATTTTTAA

2161 AAACAGATCA TGAAATTTGT GTTTTAGATA ATCTTTCTAA GGGTTCTAAA
ATCGCAATAG

2221 AAGATTTGCA AAAAATAAGA ACTTTTAAAT TTTTTGAACA AGATTTAAGT
GATTTTCAAG

2281 GCGTAAAAGC ATTGTTTGAG AGAGAAAAAT TTGACGCTAT
TGTGCATTTT GCAGCGAGCA

2341 TTGAAGTTTT TGAAAGTATG CAAAACCCTT TAAAGTATTA TATGAATAAC
ACTGTTAATA

2401 CGACAAATCT CATCGAAACT TGTTTGCAAA CTGGAGTGAA TAAATTTATA
TTTTCTTCAA

2461 CGGC AGCC AC TTATGGCGA A CCACAAACTC CCGTTGTGAG

CGAAACAAGT CCTTTAGCAC

2521 CTATTAATCC TTATGGGCGT AGTAAGCTTA TGAGCGAAGA
GGTTTTGCGT GATGCAAGTA

2581 TGGCAAATCC TGAATTTAAG CATTGTATTT TAAGATATTT TAATGTTGCA
GGTGCTTGCA

2641 TGGATTATAC TTTAGGACAA CGCTATCCAA AAGCGACTTT GCTTATAAAA
GTTGCAGCTG

2701 AATGTGCCGC AGAAAAACGT AATAAACTTT TCATATTTGG CGATGATTAT
GATACAAAAG

2761 ATGGCACTTG CATAAGAGAT TTTATCCATG TGGATGATAT
TTCAAGTGCG CATTTATCGG

2821 CTTTGGATTA TTTAAAAGAG AATGAAAGCA ATGTTTTTAA TGTAGGTTAT
GGACATGGTT

2881 TTAGCGTAAA AGAAGTGATT GAAGCGATGA AAAAAGTTAG

CGGAGTGGAT TTTAAAGTAG

2941 AACTTGCCCC ACGCCGTGCG GGTGATCCTA GTGTATTGAT
TTCTGATGCA AGTAAAATCA

3001 GAAATCTTAC TTCTTGGCAG CCTAAATATG ATGATTTAGG GCTTATTTGT
AAATCTGCTT

3061 TTGATTGGGA AAAACAGTGC TACCCATACG ATGTTCCAGA
TTACGCTTAA TCTAGAGTCG

3121 ACCTGCAGGC ATGCAAGCTT GGCTGTTTTG GCGGATGAGA
GAAGATTTTC AGCCTGATAC

3181 AGATTAAATC AGAACGCAGA AGCGGTCTGA TAAAACAGAA

TTTGCCTGGC GGCAGTAGCG

3241 CGGTGGTCCC ACCTGACCCC ATGCCGAACT CAGAAGTGAA

ACGCCGTAGC GCCGATGGTA

3301 GTGTGGGGTC TCCCCATGCG AGAGTAGGGA ACTGCCAGGC

ATCAAATAAA ACGAAAGGCT

3361 CAGTCGAAAG ACTGGGCCTT TCGTTTTATC TGTTGTTTGT
CGGTGAACGC TCTCCTGAGT

3421 AGGACAAATC CGCCGGGAGC GGATTTGAAC GTTGCGAAGC

AACGGCCCGG AGGGTGGCGG

3481 GCAGGACGCC CGCCATAAAC TGCCAGGCAT CAAATTAAGC

AGAAGGCCAT CCTGACGGAT

3541 GGCCTTTTTG CGTTTCTACA AACTCTTCCA CTCACTACAG
CAGAGCCATT TAAACAACAT

3601 CCCCTCCCCC TTTCCACCGC GTCAGACGCC CGTAGCAGCC
CGCTACGGGC TTTTTCATGC

3661 CCTGCCCTAG CGTCCAAGCC TCACGGCCGC GCTCGGCCTC
TCTGGCGGCC TTCTGGCGCT

3721 GAGGTCTGCC TCGTGAAGAA GGTGTTGCTG ACTCATACCA
GGCCTGAATC GCCCCATCAT

3781 CCAGCCAGAA AGTGAGGGAG CCACGGTTGA TGAGAGCTTT

GTTGTAGGTG GACCAGTTGG

3841 TGATTTTGAA CTTTTGCTTT GCCACGGAAC GGTCTGCGTT
GTCGGGAAGA TGCGTGATCT

3901 GATCCTTCAA CTCAGCAAAA GTTCGATTTA TTCAACAAAG
CCGCCGTCCC GTCAAGTCAG

3961 CGTAATGCTC TGCCAGTGTT ACAACCAATT AACCAATTCT GATTAGAAAA
ACTCATCGAG

4021 CATCAAATGA AACTGCAATT TATTCATATC AGGATTATCA ATACCATATT
TTTGAAAAAG

4081 CCGTTTCTGT AATGAAGGAG AAAACTCACC GAGGCAGTTC

CATAGGATGG CAAGATCCTG

4141 GTATCGGTCT GCGATTCCGA CTCGTCCAAC ATCAATACAA
CCTATTAATT TCCCCTCGTC

4201 AAAAATAAGG TTATCAAGCG AGAAATCACC ATGAGTGACG

ACTGAATCCG GTGAGAATGG

4261 CAAAAGCTAA AAAGGCCGTA ATATCCAGCT GAACGGTCTG

GTTATAGGTA CATTGAGCAA

4321 CTGACTGAAA TGCCTCAAAA TGTTCTTTAC GATGCCATTG
GGATATATCA ACGGTGGTAT

4381 ATCCAGTGAT TTTTTTCTCC ATTTTAGCTT CCTTAGCTCC TGAAAATCTC
GATAACTCAA

4441 AAAATACGCC CGGTAGTGAT CTTATTTCAT TATGGTGAAA
GTTGGAACCT CTTACGTGCC

4501 GATCAACGTC TCATTTTCGC CAAAAGTTGG CCCAGGGCTT
CCCGGTATCA ACAGGGACAC

4561 CAGGATTTAT TTATTCTGCG AAGTGATCTT CCGTCACAGG TATTTATTCG
AAGACGAAAG

4621 GGCCTCGTGA TACGCCTATT TTTATAGGTT AATGTCATGA TAATAATGGT
TTCTTAGACG

4681 TCAGGTGGCA CTTTTCGGGG AAATGTGCGC GCCCGCGTTC
CTGCTGGCGC TGGGCCTGTT

4741 TCTGGCGCTG GACTTCCCGC TGTTCCGTCA GCAGCTTTTC
GCCCACGGCC TTGATGATCG

4801 CGGCGGCCTT GGCCTGCATA TCCCGATTCA ACGGCCCCAG
GGCGTCCAGA ACGGGCTTCA

4861 GGCGCTCCCG AAGGTCTCGG GCCGTCTCTT GGGCTTGATC GGCCTTCTTG CGCATCTCAC

4921 GCGCTCCTGC GGCGGCCTGT AGGGCAGGCT CATACCCCTG CCGAACCGCT TTTGTCAGCC

4981 GGTCGGCCAC GGCTTCCGGC GTCTCAACGC GCTTTGAGAT TCCCAGCTTT TCGGCCAATC

5041 CCTGCGGTGC ATAGGCGCGT GGCTCGACCG CTTGCGGGCT GATGGTGACG TGGCCCACTG

5101 GTGGCCGCTC CAGGGCCTCG TAGAACGCCT GAATGCGCGT GTGACGTGCC TTGCTGCCCT

5161 CGATGCCCCG TTGCAGCCCT AGATCGGCCA CAGCGGCCGC AAACGTGGTC TGGTCGCGGG

5221 TCATCTGCGC TTTGTTGCCG ATGAACTCCT TGGCCGACAG CCTGCCGTCC TGCGTCAGCG

5281 GCACCACGAA CGCGGTCATG TGCGGGCTGG TTTCGTCACG GTGGATGCTG GCCGTCACGA

5341 TGCGATCCGC CCCGTACTTG TCCGCCAGCC ACTTGTGCGC CTTCTCGAAG AACGCCGCCT

5401 GCTGTTCTTG GCTGGCCGAC TTCCACCATT CCGGGCTGGC CGTCATGACG TACTCGACCG

5461 CCAACACAGC GTCCTTGCGC CGCTTCTCTG GCAGCAACTC GCGCAGTCGG CCCATCGCTT

5521 CATCGGTGCT GCTGGCCGCC CAGTGCTCGT TCTCTGGCGT CCTGCTGGCG TCAGCGTTGG

5581 GCGTCTCGCG CTCGCGGTAG GCGTGCTTGA GACTGGCCGC CACGTTGCCC ATTTTCGCCA

5641 GCTTCTTGCA TCGCATGATC GCGTATGCCG CCATGCCTGC CCCTCCCTTT TGGTGTCCAA

5701 CCGGCTCGAC GGGGGCAGCG CAAGGCGGTG CCTCCGGCGG

GCCACTCAAT GCTTGAGTAT

5761 ACTCACTAGA CTTTGCTTCG CAAAGTCGTG ACCGCCTACG GCGGCTGCGG CGCCCTACGG

5821 GCTTGCTCTC CGGGCTTCGC CCTGCGCGGT CGCTGCGCTC
CCTTGCCAGC CCGTGGATAT

5881 GTGGACGATG GCCGCGAGCG GCCACCGGCT GGCTCGCTTC
GCTCGGCCCG TGGACAACCC

5941 TGCTGGACAA GCTGATGGAC AGGCTGCGCC TGCCCACGAG

CTTGACCACA GGGATTGCCC

6001 ACCGGCTACC CAGCCTTCGA CCACATACCC ACCGGCTCCA
ACTGCGCGGC CTGCGGCCTT

6061 GCCCCATCAA TTTTTTTAAT TTTCTCTGGG GAAAAGCCTC
CGGCCTGCGG CCTGCGCGCT

6121 TCGCTTGCCG GTTGGACACC AAGTGGAAGG CGGGTCAAGG

CTCGCGCAGC GACCGCGCAG

6181 CGGCTTGGCC TTGACGCGCC TGGAACGACC CAAGCCTATG

CGAGTGGGGG CAGTCGAAGG

6241 CGAAGCCCGC CCGCCTGCCC CCCGAGCCTC ACGGCGGCGA

GTGCGGGGGT TCCAAGGGGG

6301 CAGCGCCACC TTGGGCAAGG CCGAAGGCCG CGCAGTCGAT

CAACAAGCCC CGGAGGGGCC

6361 ACTTTTTGCC GGAGGGGGAG CCGCGCCGAA GGCGTGGGGG

AACCCCGCAG GGGTGCCCTT

6421 CTTTGGGCAC CAAAGAACTA GATATAGGGC GAAATGCGAA

AGACTTAAAA ATCAACAACT

6481 TAAAAAAGGG GGGTACGCAA CAGCTCATTG CGGCACCCCC

CGCAATAGCT CATTGCGTAG

6541 GTTAAAGAAA ATCTGTAATT GACTGCCACT TTTACGCAAC GCATAATTGT
TGTCGCGCTG

6601 CCGAAAAGTT GCAGCTGATT GCGCATGGTG CCGCAACCGT
GCGGCACCCT ACCGCATGGA

6661 GATAAGCATG GCCACGCAGT CCAGAGAAAT CGGCATTCAA

GCCAAGAACA AGCCCGGTCA

6721 CTGGGTGCAA ACGGAACGCA AAGCGCATGA GGCGTGGGCC

GGGCTTATTG CGAGGAAACC

6781 CACGGCGGCA ATGCTGCTGC ATCACCTCGT GGCGCAGATG

GGCCACCAGA ACGCCGTGGT

6841 GGTCAGCCAG AAGACACTTT CCAAGCTCAT CGGACGTTCT
TTGCGGACGG TCCAATACGC

6901 AGTCAAGGAC TTGGTGGCCG AGCGCTGGAT CTCCGTCGTG
AAGCTCAACG GCCCCGGCAC

6961 CGTGTCGGCC TACGTGGTCA ATGACCGCGT GGCGTGGGGC
CAGCCCCGCG ACCAGTTGCG

7021 CCTGTCGGTG TTCAGTGCCG CCGTGGTGGT TGATCACGAC
GACCAGGACG AATCGCTGTT

7081 GGGGCATGGC GACCTGCGCC GCATCCCGAC CCTGTATCCG

GGCGAGCAGC AACTACCGAC

7141 CGGCCCCGGC GAGGAGCCGC CCAGCCAGCC CGGCATTCCG

GGCATGGAAC CAGACCTGCC

7201 AGCCTTGACC GAAACGGAGG AATGGGAACG GCGCGGGCAG

CAGCGCCTGC CGATGCCCGA

7261 TGAGCCGTGT TTTCTGGACG ATGGCGAGCC GTTGGAGCCG
CCGACACGGG TCACGCTGCC

7321 GCGCCGGTAG CACTTGGGTT GCGCAGCAAC CCGTAAGTGC
GCTGTTCCAG ACTATCGGCT

7381 GTAGCCGCCT CGCCGCCCTA TACCTTGTCT GCCTCCCCGC
GTTGCGTCGC GGTGCATGGA

7441 GCCGGGCCAC CTCGACCTGA ATGGAAGCCG GCGGCACCTC

GCTAACGGAT TCACCGTTTT

7501 TATCAGGCTC TGGGAGGCAG AATAAATGAT CATATCGTCA ATTATTACCT CCACGGGGAG

7561 AGCCTGAGCA AACTGGCCTC AGGCATTTGA GAAGCACACG

GTCACACTGC TTCCGGTAGT

7621 CAATAAACCG GTAAACCAGC AATAGACATA AGCGGCTATT

TAACGACCCT GCCCTGAACC

7681 GACGACCGGG TCGAATTTGC TTTCGAATTT CTGCCATTCA TCCGCTTATT ATCACTTATT

7741 CAGGCGTAGC ACCAGGCGTT TAAGTCGACC AATAAC

SEQ ID NO: 25 - Secuencia de aminoácidos para EPA modificada con secuencia de señal

Desvelada en WO 2009/104074 (como SEQ ID NO: 6)

Tipo: PRT

Organismo: Artificial

/nota=" Descripción de Secuencia Artificial: Polipéptido sintético"

Longitud: 643

Secuencia

Met Lys Lys He Trp Leu Ala Leu Ala Gly Leu Val Leu Ala Phe Ser 1 5 10 15

Ala Ser Ala Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys

20 25 30

Ala Cys Val Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser

35 40 45

Val Asp Pro Ala He Ala Asp Thr Asn Gly Gin Gly Val Leu His Tyr 50 55 60

Met Val Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala He Asp

70 75 80

Asn Ala Leu Ser He Thr Ser Asp Gly Leu Thr He Arg Leu Glu Gly

85 90 95

Gly Val Glu Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gin Ala

100 105 110

Arg Gly Ser Trp Ser Leu Asn Trp Leu Val Pro He Gly His Glu Lys

115 120 125

Pro Ser Asn He Lys Val Phe He His Glu Leu Asn Ala Gly Asn Gin 130 135 140

Leu Ser His Met Ser Pro He Tyr Thr He Glu Met Gly Asp Glu Leu 145 150 155 160
Leu Ala Lys Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu

165 170 175

Ser Asn Glu Met Gin Pro Thr Leu Ala lie Ser His Ala Gly Val

180 185 190

Val Val Met Ala Gin Ala Gin Pro Arg Arg Glu Lys Arg Trp Ser Glu

195 200 205

Trp Ala Ser Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val 210 215 220

Tyr Asn Tyr Leu Ala Gin Gin Arg Cys Asn Leu Asp Asp Thr Trp Glu 225 230 235 24(

Gly Lys lie Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu

245 250 255

Asp lie Lys Asp Asn Asn Asn Ser Thr Pro Thr Val lie Ser His Arg

260 265 270

Leu His Phe Pro Glu Gly Gly Ser Leu Ala Ala Leu Thr Ala His Gin

275 280 285

Ala Cys His Leu Pro Leu Glu Ala Phe Thr Arg His Arg Gin Pro Arg 290 295 300

Gly Trp Glu Gin Leu Glu Gin Cys Gly Tyr Pro Val Gin Arg Leu Val 305 310 315 320

Ala Leu Tyr Leu Ala Ala Arg Leu Ser Trp Asn Gin Val Asp Gin Val

325 330 335

Arg Asn Ala Leu Ala Ser Pro Gly Ser Gly Gly Asp Leu Gly Glu

340 345 350

Ala lie Arg Glu Gin Pro Glu Gin Ala Arg Leu Ala Leu Thr Leu Ala 355 360 365

Ala Ala Glu Ser Glu Arg Phe Val Arg Gin Gly Thr Gly Asn Asp Glu 370 375 380

Gly Ala Ala Ser Ala Asp Val Val Ser Leu Thr Cys Pro Val Ala

390 395 400

Lys Asp Gin Asn Thr Lys Gly Glu Cys Ala Gly Pro Ala Asp Ser

410 415

Gly Asp Ala Leu Leu Arg Asn Tyr Pro Thr Gly Ala Glu Phe Leu

420 425 430

Gly Asp Gly Gly Asp Val Ser Phe Ser Thr Arg Gly Thr Gin Asn Trp

435 440 445

Thr Val Glu Arg Leu Leu Gin Ala His Arg Gin Leu Glu Glu Arg Gly 450 455 460

Tyr Val Phe Val Gly Tyr His Gly Thr Phe Leu Glu Ala Ala Gin Ser 465 470 475 480

He Val Phe Gly Gly Val Arg Ala Arg Ser Gin Asp Leu Asp Ala He

485 490 495

Trp Arg Gly Phe Tyr He Ala Gly Asp Pro Ala Leu Ala Tyr Gly Tyr

500 505 510

Ala Gin Asp Gin Glu Pro Asp Ala Arg Gly Arg He Arg Asn Gly Ala

515 520 525

Leu Leu Arg Val Tyr Val Pro Arg Trp Ser Leu Gly Phe Tyr Arg 530 535

Gly Leu Thr Leu Ala Ala Pro Glu Ala Ala Gly Glu Val Glu Arg

550 555 560 Leu lie Gly His Pro Leu Pro Leu Arg Leu Asp Ala lie Thr Gly Pro

565 570 575

Glu Glu Glu Gly Gly Arg Val Thr lie Leu Gly Trp Pro Leu Ala Glu

580 585 590

Arg Thr Val Val lie Pro Ser Ala lie Pro Thr Asp Pro Arg Asn Val

595 600 605

Gly Gly Asp Leu Asp Pro Ser Ser lie Pro Asp Lys Glu Gin Ala lie 610 615 620

Ser Ala Leu Pro Asp Tyr Ala Ser Gin Pro Gly Lys Pro Pro Arg Glu 625 630 635 640

Asp Leu Lys

SE

Q ID NO: 26 - Secuencia de aminoácidos para PgIB

Desvelada en WO 2009/104074 (como SEQ ID NO: 2)

Longitud: 722

Tipo: PRT

Organismo: *Campylobacter jejuni*

Met Leu Lys Lys Glu Tyr Leu Lys Asn Pro Tyr Leu Val Leu Phe Ala 1 5 10 15

Met lie lie Leu Ala Tyr Val Phe Ser Val Phe Cys Arg Phe Tyr Trp

20 25 30

Val Trp Trp Ala Ser Glu Phe Asn Glu Tyr Phe Phe Asn Asn Gin Leu

35 40 45

Met lie lie Ser Asn Asp Gly Tyr Ala Phe Ala Glu Gly Ala Arg Asp 50 55 60

Met lie Ala Gly Phe His Gin Pro Asn Asp Leu Ser Tyr Tyr Gly Ser 65 70 75 80

Ser Leu Ser Ala Leu Thr Tyr Trp Leu Tyr Lys lie Thr Pro Phe Ser

85 90 95

Phe Glu Ser lie lie Leu Tyr Met Ser Thr Phe Leu Ser Ser Leu Val

100 105 110

Val lie Pro Thr lie Leu Leu Ala Asn Glu Tyr Lys Arg Pro Leu Met

115 120 125

Gly Phe Val Ala Ala Leu Leu Ala Ser lie Ala Asn Ser Tyr Tyr Asn 130 135 140

Arg Thr Met Ser Gly Tyr Tyr Asp Thr Asp Met Leu Val lie Val Leu 145 150 155 160

Pro Met Phe lie Leu Phe Phe Met Val Arg Met lie Leu Lys Lys Asp

165 170 175

Phe Phe Ser Leu lie Ala Leu Pro Leu Phe lie Gly lie Tyr Leu Trp

180 185 190

Trp Tyr Pro Ser Ser Tyr Thr Leu Asn Val Ala Leu lie Gly Leu Phe

195 200 205

Leu lie Tyr Thr Leu lie Phe His Arg Lys Glu Lys lie Phe Tyr lie 210 215 220

Ala Val lie Leu Ser Ser Leu Thr Leu Ser Asn lie Ala Trp Phe Tyr 225 230 235 240

Gin Ser Ala He He Val He Leu Phe Ala Leu Phe Ala Leu Glu Gin

245 250 255 Arg Leu Asn Phe Met lie He Gly He Leu Gly Ser Ala Thr Leu 260 265 270

Phe Leu lie Leu Ser Gly Gly Val Asp Pro Leu Tyr Gin Leu 275 280 285

Lys Phe Tyr lie Phe Arg Ser Asp Glu Ser Ala Leu Thr Gin Gly 290 295

Met Tyr Phe Asn Val Asn Gin Thr He Gin Val Glu Asn Val

310 315 320

Asp Leu Ser Glu Phe Met Arg Arg He Ser Gly Ser Glu He Val Phe

325 330 335

Leu Phe Ser Leu Phe Gly Phe Val Trp Leu Leu Arg Lys His Lys Ser

340 345 350

Met lie Met Ala Leu Pro He Leu Val Leu Gly Phe Leu Ala Leu Lys

355 360 365

Gly Gly Leu Arg Phe Thr He Tyr Ser Val Pro Met Ala Leu Gly 370 375

Gly Phe Leu Leu Ser Glu Phe Lys Ala He Met Val Lys Lys Tyr

390 395 400

Ser Gin Leu Thr Ser Asn Val Cys He Val Phe Ala Thr He Leu Thr

405 410 415

Leu Ala Pro Val Phe He His He Tyr Asn Tyr Lys Ala Pro Thr Val

420 425 430

Phe Ser Gin Asn Glu Ala Ser Leu Leu Asn Gin Leu Lys Asn He Ala

435 440 445

Asn Arg Glu Asp Tyr Val Val Thr Trp Ala Ala Gly Tyr Pro Val 450 455

Arg Tyr Tyr Ser Asp Val Lys Thr Leu Val Asp Gly Gly Lys His Leu 465 470 475 480

Gly Lys Asp Asn Phe Phe Pro Ser Phe Ala Leu Ser Lys Asp Glu Gin

485 490 495

Ala Ala Ala Asn Met Ala Arg Leu Ser Val Glu Tyr Thr Glu Lys Ser

500 505 510

Phe Tyr Ala Pro Gin Asn Asp lie Leu Lys Thr Asp lie Leu Gin Ala

515 520 525

Met Met Lys Asp Tyr Asn Gin Ser Asn Val Asp Leu Phe Leu Ala Ser 530 535 540

Leu Ser Lys Pro Asp Phe Lys lie Asp Thr Pro Lys Thr Arg Asp lie 545 550 555 560

Tyr Leu Tyr Met Pro Ala Arg Met Ser Leu lie Phe Ser Thr Val Ala

565 570 575

Phe Ser Phe lie Asn Leu Asp Thr Gly Val Leu Asp Lys Pro Phe

580 585 590

Thr Phe Ser Thr Ala Tyr Pro Leu Asp Val Lys Asn Gly Glu lie Tyr

595 600 605

Leu Ser Asn Gly Val Val Leu Ser Asp Asp Phe Arg Ser Phe Lys lie 610 615 620

Gly Asp Asn Val Val Ser Val Asn Ser lie Val Glu lie Asn Ser lie 625 630 635 640

Lys Gin Gly Glu Tyr Lys lie Thr Pro lie Asp Asp Lys Ala Gin Phe

645 650 655 Tyr lie Phe Tyr Leu Lys Asp Ser Ala lie Pro Tyr Ala Gin Phe

660 665 670

Leu Met Asp Lys Thr Met Phe Asn Ser Ala Tyr Val Gin Met Phe Phe

675 680 685

Leu Gly Asn Tyr Asp Lys Asn Leu Phe Asp Leu Val lie Asn Ser Arg 690 695 700

Asp Ala Lys Val Phe Lys Leu Lys lie Tyr Pro Tyr Asp Val Pro Asp 705 710 715 720

Tyr Ala

SEQ ID NO: 27 - Secuencia de nucleótidos para pCCIFOS plásmido vacío

Locus pCCIFOS con casete MCS

Características Localización/Calificadores

Región 230..256

/marcador="pCCI/pEpiFOS fwd"

Región 311..330

/marcador=" promotor T7"

Región complementaria (504..529)

/marcador="pCCIpEpiFOS rv"

CDS complementaria (805..1464)

/marcador=cat

CDS 1683..2030

/marcador=redF

CDS 3425..4180

/marcador=repE

CDS 4759..5934

/marcador=parA

CDS 5934..6905

/marcador=parB

ORIGEN

Longitud: 8171 pb

Tipo: ADN circular UNA

Organismo: Artificial

Secuencia:

1 GCGGCCGCAA GGGGTTCGCG TCAGCGGGTG TTGGCGGGTG
TCGGGGCTGG

CTTAACTATG

61 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA
CCGCACAGAT

121 GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT
CAGCTGCGCA ACTGTTGGGA

181 AGGGCGATCG GTGCGGGCCT CTTCGCTATT ACGCCAGCTG
GCGAAAGGGG GATGTGCTGC

2 1 AAGGCGATTA AGTTGGGTAA CGC CAGGGTT TTCCCAGTCA
CGACGTTGTA AAACGACGGC

301 CAGTGAATTG TAATACGACT CACTATAGGG CGAATTCGAG
CTCGGTACCC GGGGATCCCA

361 CGTGGCGCGC CACTAGTGCT AGCGACGTCG TGGGATCCTC TAGAGT
CGAC CTGCAGGCAT

421 GCAAGCTTGA GTATTCTATA GTCTCACCTA AATAGCTTGG CGTAAT CATG
GTCATAGCTG

481 TTTCCTGTGT GAAATTGTTA TCCGCTCACA ATTCCACACA ACATACGAGC
CGGAAGCATA

541 AAGTGTAAAG CCTGGGGTGC CTAATGAGTG AGCTAACTCA
CATTAATTGC GTTGCGCTCA

601 CTGCCCGCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC
ATTAATGAAT CGGCCAACGC

661 GAACCCCTTG CGGCCGCCCG GGCCGTCGAC CAATTCTCAT
GTTTGACAGC TTATCATCGA

721 ATTTCTGCCA TTCATCCGCT TATTATCACT TATTCAGGCG TAGCAACCAG
GCGTTTAAGG

781 GCACCAATAA CTGCCTTAAA AAAATTACGC CCCGCCCTGC
CACTCATCGC AGTACTGTTG

841 TAATTCATTA AGCATTCTGC CGACATGGAA GCCATCACAA ACGGCATGAT
GAACCTGAAT

901 CGCCAGCGGC ATCAGCACCT TGTCGCCTTG CGTATAATAT
TTGCCCATGG TGAAAACGGG

961 GGCGAAGAAG TTGT C CAT AT TGGCCACGTT TAAATCAAAA
CTGGTGAAAC TCACCCAGGG

1021 ATTGGCTGAG ACGAAAAACA TATTCTCAAT AAACCCTTTA
GGGAAATAGG CCAGGTTTTC

1081 AC CGTAACAC GCCACATCTT GCGAATATAT GTGTAGAAAC
TGCCGGAAAT CGTCGTGGTA

1141 TTCACTCCAG AGCGATGAAA ACGTTTCAGT TTGCTCATGG
AAAACGGTGT AACAAGGGTG

1201 AACACTATCC CATATCACCA GCTCACCGTC TTTCATTGCC ATACGAAATT
CCGGATGAGC

1261 ATTCATCAGG CGGGCAAGAA TGTGAATAAA GGCCGGATAA
AACTTGTGCT TATTTTTCTT

1321 TACGGTCTTT AAAAAGGCCG TAATATCCAG CTGAACGGTC
TGGTTATAGG TACATTGAGC

1381 AACTGACTGA AATGCCTCAA AATGTTCTTT ACGATGCCAT TGGGATATAT
CAACGGTGGT

1441 ATATCCAGTG ATTTTTTTCT CCATTTTAGC TTCCTTAGCT CCTGAAAATC
TCGATAACTC

1501 AAAAAATACG CCCGGTAGTG ATCTTATTTC ATTATGGTGA
AAGTTGGAAC CTCTTACGTG

1561 CCGATCAACG TCTCATTTTC GCCAAAAGTT GGCCCAGGGC
TTCCCGGTAT CAACAGGGAC

1621 AC C AGGATTT ATTTATTCTG CGAAGTGATC TTCCGTCACA
GGTATTTATT CGCGATAAGC

1681 TCATGGAGCG GCGTAACCGT CGCACAGGAA GGACAGAGAA
AGCGCGGATC TGGGAAGTGA

1741 CGGACAGAAC GGTCAGGACC TGGATTGGGG AGGCGGTTGC
CGCCGCTGCT GCTGACGGTG

1801 TGACGTTCTC TGTTCCGGTC ACACCACATA CGTTCCGCCA
TTCCTATGCG ATGCACATGC

1861 TGTATGCCGG TATACCGCTG AAAGTTCTGC AAAGCCTGAT
GGGACATAAG TCCATCAGTT

1921 CAACGGAAGT CTACACGAAG GTTTTTGCGC TGGATGTGGC
TGCCCGGCAC CGGGTGCAGT

1981 TTGCGATGCC GGAGTCTGAT GCGGTTGCGA TGCTGAAACA
ATTATCCTGA GAATAAATGC

2041 CTTGGCCTTT ATATGGAAAT GTGGAACTGA GTGGATATGC
TGTTTTTGTC TGTTAAACAG

2101 AGAAGCTGGC TGTTATCCAC TGAGAAGCGA ACGAAACAGT
CGGGAAAATC TCCCATTATC

2161 GTAGAGATCC GCATTATTAA TCTCAGGAGC CTGTGTAGCG
TTTATAGGAA GTAGTGTTCT

2221 GTCATGATGC CTGCAAGCGG TAACGAAAAC GATTTGAATA
TGCCTTCAGG AACAATAGAA

2281 ATCTTCGTGC GGTGTTACGT TGAAGTGGAG CGGATTATGT
CAGCAATGGA CAGAACAACC

2341 TAATGAACAC AGAACCATGA TGTGGTCTGT CCTTTTACAG
CCAGTAGTGC TCGCCGCAGT

2401 CGAGCGACAG GGCGAAGCCC TCGGCTGGTT GCCCTCGCCG
CTGGGCTGGC GGCCGTCTAT

2461 GGCCCTGCAA ACGCGCCAGA AACGCCGTCG AAGCCGTGTG
CGAGACACCG CGGCCGGCCG

2521 CCGGCGTTGT GGATACCTCG CGGAAAACTT GGCCCTCACT
GACAGATGAG GGGCGGACGT

2581 TGACACTTGA GGGGCCGACT CACCCGGCGC GGCGTTGACA
GATGAGGGGC AGGCTCGATT

2641 TCGGCCGGCG ACGTGGAGCT GGCCAGCCTC GCAAATCGGC
GAAAACGCCT GATTTTACGC

2701 GAGTTTCCCA CAGATGATGT GGACAAGCCT GGGGATAAGT
GCCCTGCGGT ATTGACACTT

2761 GAGGGGCGCG ACTACTGACA GATGAGGGGC GCGATCCTTG
ACACTTGAGG GGCAGAGTGC

2821 TGACAGATGA GGGGCGCACC TATTGACATT TGAGGGGCTG
TCCACAGGCA GAAAATCCAG

2881 CATTTGCAAG GGTTTCCGCC CGTTTTTCGG CCACCGCTAA
CCTGTCTTTT AACCTGCTTT

2941 TAAAC CAATA TTTATAAACC TTGTTTTTAA CCAGGGCTGC
GCCCTGTGCG CGTGACCGCG

3001 CACGCCGAAG GGGGGTGCCC CCCCTTCTCG AACCCTCCCG
GTCGAGTGAG CGAGGAAGCA

3061 CCAGGGAACA GCACTTATAT ATTCTGCTTA CACACGATGC
CTGAAAAAAC TTCCCTTGGG

3121 GTTATCCACT TAT C CACGGG GATATTTTTA TAATTATTTT TTTTATAGTT
TTTAGATCTT

3181 CTTTTTTAGA GCGCCTTGTA GGCCTTTATC CATGCTGGTT
CTAGAGAAGG TGTTGTGACA

3241 AATTGCCCTT TCAGTGTGAC AAATCACCCT CAAATGACAG
TCCTGTCTGT GACAAATTGC

3301 CCTTAACCCT GTGACAAATT GCCCTCAGAA GAAGCTGTTT
TTTCACAAAG TTATCCCTGC

3361 TTATTGACTC TTTTTTATTT AGTGTGACAA TCTAAAAACT TGTCACACTT
CACATGGATC

3421 TGTCATGGCG GAAACAGCGG TTATCAATCA CAAGAAACGT
AAAAATAGCC CGCGAATCGT

3481 CCAGTCAAAC GACCTCACTG AGGCGGCATA TAGTCTCTCC
CGGGATCAAA AACGTATGCT

3541 GTATCTGTTC GTTGACCAGA TCAGAAAATC TGATGGCACC
CTACAGGAAC ATGACGGTAT

3601 CTGCGAGATC CATGTTGCTA AATATGCTGA AATATTCGGA
TTGACCTCTG CGGAAGCCAG

3661 TAAGGATATA CGGCAGGCAT TGAAGAGTTT CGCGGGGAAG
GAAGTGGTTT TTTATCGCCC

3721 TGAAGAGGAT GCCGGCGATG AAAAAGGCTA TGAATCTTTT
CCTTGGTTTA TCAAACGTGC

3781 GCACAGTCCA TCCAGAGGGC TTTACAGTGT ACATATCAAC
CCATATCTCA TTCCCTTCTT

3841 TATCGGGTTA CAGAACCGGT TTACGCAGTT TCGGCTTAGT
GAAACAAAAG AAATCACCAA

3901 TCCGTATGCC ATGCGTTTAT ACGAATCCCT GTGT CAGTAT CGTAAGC
CGG ATGGCTCAGG

3961 CATCGTCTCT CTGAAAATCG ACTGGATCAT AGAGCGTTAC
CAGCTGCCTC AAAGTTACCA

4021 GCGTATGCCT GACTTCCGCC GCCGCTTCCT GCAGGTCTGT
GTTAATGAGA TCAACAGCAG

4081 AACTCCAATG CGCCTCTCAT ACATTGAGAA AAAGAAAGGC
CGCCAGACGA CTCATATCGT

4141 ATTTTCCTTC CGCGATATCA CTTCCATGAC GACAGGATAG
TCTGAGGGTT ATCTGTCACA

4201 GATTTGAGGG TGGTTCGTCA CATTTGTTCT GACCTACTGA
GGGTAATTTG TCACAGTTTT

4261 GCTGTTTCCT TCAGCCTGCA TGGATTTTCT CATACTTTTT GAACTGTAAT
TTTTAAGGAA

4321 GCCAAATTTG AGGGCAGTTT GTCACAGTTG ATTTCCTTCT CTTTCCCTTC
GTCATGTGAC

4381 CTGATATCGG GGGTTAGTTC GTCATCATTG ATGAGGGTTG
ATTATCACAG TTTATTACTC

4441 TGAATTGGCT ATCCGCGTGT GTACCTCTAC CTGGAGTTTT
TCCCACGGTG GATATTTCTT

4501 CTTGCGCTGA GCGTAAGAGC TATCTGACAG AACAGTTCTT
CTTTGCTTCC TCGCCAGTTC

4561 GCTCGCTATG CTCGGTTACA CGGCTGCGGC GAGCGCTAGT
GATAATAAGT GACTGAGGTA

4621 TGTGCTCTTC TTATCTCCTT TTGTAGTGTT GCTCTTATTT TAAACAACTT
TGCGGTTTTT

4681 TGATGACTTT GCGATTTTGT TGTTGCTTTG CAGTAAATTG CAAGATTTAA
TAAAAAAACG

4741 CAAAGCAATG ATTAAAGGAT GTTCAGAATG AAACTCATGG
AAACACTTAA CCAGTGCATA

4801 AACGCTGGTC ATGAAATGAC GAAGGCTATC GCCATTGCAC
AGTTTAATGA TGACAGCCCG

4861 GAAGCGAGGA AAATAACCCG GCGCTGGAGA ATAGGTGAAG
CAGCGGATTT AGTTGGGGTT

4921 TCTTCTCAGG CTATCAGAGA TGCCGAGAAA GCAGGGCGAC
TACCGCACCC GGATATGGAA

4981 ATT CGAGGAC GGGTTGAGCA ACGTGTTGGT TATACAATTG
AACAAATTAA TCATATGCGT

5041 GATGTGTTTG GTACGCGATT GCGACGTGCT GAAGACGTAT
TTCCACCGGT GATCGGGGTT

5101 GCTGCCCATA AAGGTGGCGT TTACAAAACC TCAGTTTCTG
TTCATCTTGC TCAGGATCTG

5161 GCTCTGAAGG GGCTACGTGT TTTGCTCGTG GAAGGTAACG
ACCCCCAGGG AACAGCCTCA

5221 ATGTAT CACG GATGGGTACC AGATCTTCAT ATTCATGCAG
AAGACACTCT CCTGCCTTTC

5281 TATCTTGGGG AAAAGGACGA TGTCACTTAT GCAATAAAGC
CCACTTGCTG GCCGGGGCTT

5341 GACATTATTC CTTCCTGTCT GGCTCTGCAC CGTATTGAAA
CTGAGTTAAT GGGCAAATTT

5401 GATGAAGGTA AACTGCCCAC CGATCCACAC CTGATGCTCC
GACTGGCCAT TGAAACTGTT

5461 GCTCATGACT ATGATGTCAT AGTTATTGAC AGCGCGCCTA
ACCTGGGTAT CGGCACGATT

5521 AATGTCGTAT GTGCTGCTGA TGTGCTGATT GTTCCCACGC
CTGCTGAGTT GTTTGACTAC

5581 ACCTCCGCAC TGCAGTTTTT CGATATGCTT CGTGATCTGC
TCAAGAACGT TGATCTTAAA

5641 GGGTTCGAGC CTGATGTACG TATTTTGCTT ACCAAATACA
GCAATAGTAA TGGCTCTCAG

5701 TCCCCGTGGA TGGAGGAGCA AATTCGGGAT GCCTGGGGAA
GCATGGTTCT AAAAAATGTT

5761 GTACGTGAAA CGGATGAAGT TGGTAAAGGT CAGATCCGGA
TGAGAACTGT TTTTGAACAG

5821 GCCATTGATC AACGCTCTTC AACTGGTGCC TGGAGAAATG
CTCTTTCTAT TTGGGAACCT

5881 GTCTGCAATG AAATTTTCGA TCGTCTGATT AAACCACGCT GGGAGATTAG ATAATGAAGC

5941 GTGCGCCTGT TATTCCAAAA CATACGCTCA ATACTCAACC GGTTGAAGAT ACTTCGTTAT

6001 CGACACCAGC TGCCCCGATG GTGGATTCGT TAATTGCGCG CGTAGGAGTA ATGGCTCGCG

6061 GTAATGCCAT TACTTTGCCT GTATGTGGTC GGGATGTGAA GTTTACTCTT GAAGTGCTCC

6121 GGGGTGATAG TGTTGAGAAG ACCTCTCGGG TATGGTCAGG TAATGAACGT GACCAGGAGC

6181 TGCTTACTGA GGACGCACTG GATGATCTCA TCCCTTCTTT TCTACTGACT GGTCAACAGA

6241 CACCGGCGTT CGGTCGAAGA GTATCTGGTG TCATAGAAAT TGCCGATGGG AGTCGCCGTC

6301 GTAAAGCTGC TGCACTTACC GAAAGTGATT ATCGTGTTCT GGTTGGCGAG CTGGATGATG

6361 AGCAGATGGC TGCATTATCC AGATTGGGTA ACGATTATCG CCCAACAAGT GCTTATGAAC

6421 GTGGTCAGCG TTATGCAAGC CGATTGCAGA ATGAATTTGC TGGAAATATT TCTGCGCTGG

6481 CTGATGCGGA AAATATTTCA CGTAAGATTA TTACCCGCTG TATCAACACC GCCAAATTGC

6541 CTAAATCAGT TGTTGCTCTT TTTTCTCACC CCGGTGAACT ATCTGCCCGG TCAGGTGATG

6601 CACTTCAAAA AGCCTTTACA GATAAAGAGG AATTACTTAA GCAGCAGGCA TCTAACCTTC

6661 ATGAGCAGAA AAAAGCTGGG GTGATATTTG AAGCTGAAGA AGTTATCACT CTTTTAACTT

6721 CTGTGCTTAA AACGTCATCT GCATCAAGAA CTAGTTTAAG CTCACGACAT CAGTTTGCTC

6781 CTGGAGCGAC AGTATTGTAT AAGGGCGATA AAATGGTGCT TAACCTGGAC AGGTCTCGTG

6841 TTCCAACTGA GTGTATAGAG AAAATTGAGG CCATTCTTAA GGAACTTGAA AAGCCAGCAC

6901 CCTGATGCGA CCACGTTTTA GTCTACGTTT ATCTGTCTTT ACTTAATGTC
CTTTGTTACA

6961 GGCCAGAAAG CATAACTGGC CTGAATATTC TCTCTGGGCC
CACTGTTCCA CTTGTATCGT

7021 CGGTCTGATA ATCAGACTGG GACCACGGTC CCACTCGTAT
CGTCGGTCTG ATTATTAGTC

7081 TGGGACCACG GTCCCACTCG TATCGTCGGT CTGATTATTA
GTCTGGGACC ACGGTCCCAC

7141 TCGTATCGTC GGTGTGATAA TCAGACTGGG ACCACGGTCC
CACTCGTATC GTCGGTCTGA

7201 TTATTAGTCT GGGACCATGG TCCCACTCGT ATCGTCGGTC
TGATTATTAG TCTGGGACCA

7261 CGGTCCCACT CGTATCGTCG GTCTGATTAT TAGTCTGGAA
CCACGGTCCC ACTCGTATCG

7321 TCGGTCTGAT TATTAGTCTG GGACCACGGT CCCACTCGTA
TCGTCGGTCT GATTATTAGT

7381 CTGGGACCAC GATCCCACTC GTGTTGTCGG TCTGATTATC
GGTCTGGGAC CACGGTCCCA

7441 CTTGTATTGT CGATCAGACT ATCAGCGTGA GACTACGATT
CCATCAATGC CTGTCAAGGG

7501 CAAGTATTGA CATGTCGTCG TAACCTGTAG AACGGAGTAA
CCTCGGTGTG CGGTTGTATG

7561 CCTGCTGTGG ATTGCTGCTG TGTCCTGCTT AT C C AC AAC A
TTTTGCGCAC GGTTATGTGG

7621 ACAAAATACC TGGTTACCCA GGCCGTGCCG GCACGTTAAC
CGGGCTGCAT CCGATGCAAG

7681 TGTGTCGCTG TCGACGAGCT CGCGAGCTCG GACATGAGGT
TGCCCCGTAT TCAGTGTCGC

7741 TGATTTGTAT TGTCTGAAGT TGTTTTTACG TTAAGTTGAT GCAGATCAAT
TAATACGATA

7801 CCTGCGTCAT AATTGATTAT TTGACGTGGT TTGATGGCCT
CCACGCACGT TGTGATATGT

7861 AGATGATAAT CATTATCACT TTACGGGTCC TTTCCGGTGA
TCCGACAGGT TACGGGGCGG

7921 CGACCTCGCG GGTTTTCGCT ATTTATGAAA ATTTTCCGGT TTAAGGCGTT TCCGTTCTTC

7981 TTCGTCATAA CTTAATGTTT TTATTTAAAA TACCCTCTGA AAAGAAAGGA AACGACAGGT

8041 GCTGAAAGCG AGCTTTTTGG CCTCTGTCGT TTCCTTTCTC TGTTTTTGTC CGTGGAATGA

8101 ACAATGGAAG TCCGAGCTCA TCGCTAATAA CTTCGTATAG CATACATTAT ACGAAGTTAT

8161 ATTCGATCCA

SEQ ID NO: 28 - Secuencia de nucleótidos para pCCIFOS cortado (pFOS) y antígeno O de *S. flexneri* 6 sin Z3206

Locus pFOS cortado y Antígeno O cortado (-Z3206)

Definición unión de pCCIFOS invertido con un casete MCS cortado con NheI y AscI en agrupamiento de antígeno O de S. flexneri 6 amplificado con galFNheI y wzzAscI cortado con NheI y AscI

Características Localización/Calificadores

CDS 3..411

/marcador='galF

CDS 784..1869

/marcador=rmlB

CDS 1869..2768

/marcador=rmlD

CDS 2826..3704

/marcador=rmlA

CDS 3709..4266

/marcador=rmlC

CDS 4263..5495

/marcador=wzx

CDS 5551..6738

/marcador=wzy

CDS 6755..7624

/marcador=wfbY

CDS 7621..8454

/marcador=wfbZ

CDS 8559..9965

/marcador=gnd

CDS 10187..11380

/marcador=ugd

CDS complementaria (I 1446..12450)

/marcador=uge

CDS 12802..12828

/marcador=wzz’

Región complementaria 12868..12887)

/marcador="T7 promotor"

Región complementaria 12942..12968)

/marcador="pCCI/pEpiFOS fwd"

CDS complementaria (14460..15431)

/marcador=parB

CDS complementaria 15431..16606)

/marcador=parA

CDS complementaria 17185..17940)

/marcador=repE

CDS complementaria 19335..19682)

/marcador=redF

CDS 19901..20560

/marcador=cat

Región 20836..20861

/marcador="pCCIpEpiFOS rv"

Longitud: 20982 pb

Tipo: ADN circular UNA

Secuencia:

1 CTAGCGGCAA AACGTATGCC GGGTGACCTC TCTGAATACT CCGTCATCCA
GACCAAAGAA

61 CCGCTGGATC GCGAAGGTAA AGTCAGCCGC ATTGTTGAAT TTATCGAAAA
ACCGGATCAG

121 CCGCAGACGC TGGACTCAGA CATCATGGCC GTTGGTCGCT
ATGTGCTTTC TGCCGATATT

181 TGGCCGGAAC TTGAACGTAC TCAGCCTGGT GCATGGGGAC
GTATTCAGCT GACTGATGCC

241 ATTGCCGAGC TGGCGAAAAA ACAGTCCGTT GATGCAATGC
TGATGACCGG CGACAGCTAC

301 GACTGCGGTA AAAAAATGGG CTATATGCAG GCGTTTGTGA
AGTATGGGCT GCGCAACCTG

361 AAAGAAGGGG CGAAGTTCCG TAAAGGTATT GAGAAGCTGT

TAAGCGAATA ATGAAAATCT

421 GACCGGATGT AACGGTTGAT AAGAAAATTA TAACGGCAGT
GAAGATTCGT GGTGAAAGTA

481 ATTTGTTGCG AATATTCCTG CCGTTGTTTT ATATAAACAA TCAGAATAAC
AACGAGTTAG

541 CAATAGGATT TTAGTCAAAG TTTTCCAGGA TTTTCCTTGT TTCCAGAGCG
GATTGGTAAG

601 ACAATTAGCT TTTGAATTTT TCGGGTTTAG CGCGAGTGGG TAACGCTCGT
CACATCGTAG

661 GCATGCATGC AGTGCTCTGG TAGCTGTAAA GCCAGGGGCG
GTAGCGTGCA TTAATACTTC

721 TATTAATCAA ACTGAGAGCC GCTTATTTCA CAGCATGCTC TGAAGCAATA
TGGAATAAAT

781 TAGGTGAAAA TACTTGTTAC TGGTGGCGCA GGATTTATTG GTTTTGCTGT
AGTTCGTCAC

841 ATTATAAATA ATACGCAGGA TAGTGTTGTT AATGTCGATA AATTAACGTA
CGCCGGAAAC

901 CTGGAATCAC TTGCTGATGT TTCTGATTCT GAACGCTATG TTTTTGAACA
TGCGGATATT

961 TGCGATGCAG CTGCAATGGC ACGGATTTTT GCTCAGCATC
AGCCAGATGC AGTGATGCAC

1021 CTGGCTGCTG AAAGCCATGT TGACCGTTCA ATTACAGGTC
CTGCGGCATT TATTGAAACC

1081 AATATTGTTG GTACATATGT CCTTTTGGAA GCCGCTCGCA
ATTATTGGTC TGCTCTTGAT

1141 AGCGACAAGA AAACTAGATT CCGTTTTCAT CATATTTCTA CTGACGAAGT
CTATGGTGAT

1201 TTGCCTCATC CTGACGAGGT AAATAATACA GAAGAATTAC CCTTATTTAC
AGAGACAACA

1261 GCTTACGCGC CAAGCAGCCC TTATTCCGCT TCAAAAGCAT
CCAGCGATCA TTTAGTCCGC

1321 GCGTGGAAAC GTACCTATGG TTTACCAACC ATTGTGACTA
ATTGCTCTAA TAATTATGGT

1381 CCTTATCATT TCCCGGAAAA ATTGATTCCA TTGGTTATTC TGAATGCTCT
GGAAGGTAAG

1441 GCATTACCTA TTTATGGCAA AGGGGATCAA ATTCGTGACT
GGCTGTATGT TGAAGATCAT

1501 GCGCGTGCGT TATATACCGT CGTAACCGAA GGTAAAGCGG

GTGAAACTTA TAACATTGGT

1561 GGACACAACG AAAAGAAAAA CATCGATGTA GTGCTCACTA
TTTGTGATTT GCTGGATGAG

1621 ATTGTACCGA AAGAGAAATC TTACCGCGAG CAAATTACTT
ATGTTGCCGA TCGCCCGGGA

1681 CACGATCGCC GTTATGCGAT TGATGCAGAG AAGATTAGCC

GCGAATTGGG CTGGAAACCG

1741 CAGGAAACGT TTGAGAGCGG GATTCGGAAG ACATTGGAAT

GGTACCTGTC CAATACAAAA

1801 TGGGTTGATA ATGTGAAAAG TGGTGCTTAT CAATCGTGGA
TTGAACAGAA CTATGAGGGC

1861 CGCCAGTAAT GAATATCCTC CTTTTCGGCA AAACAGGGCA
GGTAGGTTGG GAACTACAGC

1921 GTGCTCTGGC ACCTTTGGGT AATTTGATTG CTCTTGATGT
TCACTCCACT GATTATTGTG

1981 GTGATTTTAG TAATCCTGAA GGTGTAGCTG AAACAGTCAA
AAGAATTCGA CCTGATGTTA

2041 TTGTTAATGC TGCGGCTCAC ACCGCAGTAG ATAAGGCTGA
GTCAGAACCC GAATTTGCAC

2101 AATTACTCAA TGCGACTAGT GTTGAATCAA TTGCAAAAGA
GGCTAATGAA GTTGGGGCTT

2161 GGGTAATTCA TTACTCAACT GACTACGTAT TCCCTGGAAA
TGGCGACACG CCATGGCTGG

2221 AGACGGATGC AACCGCACCG CTAAATGTTT ACGGTGAAAC

CAAGTTAGCC GGAGAAAAAG

2281 CGTTACAGGA ACATTGCGCG AAGCATCTTA TTTTCCGTAC
CAGCTGGGTA TACGCAGCTA

2341 AAGGAAATAA CTTCGCCAAA ACGATGTTGC GTCTGGCAAA

AGAGCGCGAA GAACTGGCTG

2401 TGATAAATGA TCAATTTGGT GCGCCAACAG GTGCTGAGCT
GCTGGCTGAT TGTACGGCAC

2461 ATGCTATTCG TGTGGCACTG AATAAACCGG AAGTCGCAGG
TTTGTACCAT CTGGTAGCCA

2521 GTGGTACCAC AACCTGGCAC GATTATGCTG CGCTGGTTTT
TGAAGAGGCG CGCAAAGCAG

2581 GTATTCCCCT TGCACTCAAC AAGCTCAACG CAGTACCAAC
AACAGCCTAT CCTACACCAG

2641 CTCGTCGTCC ACATAACTCT CGCCTTAATA CAGAAAAATT
TCAGCAGAAC TTTGCGCTTG

2701 TCTTGCCTGA CTGGCAGGTT GGTGTGAAAC GAATGCTCAA
CGAATTAATT ACGACTACAG

2761 CAATTTAATA GTTTTTGCAT CTTGTTCGTG ATGGTGGAGC AAGATGAATT
AAAAGGAATG

2821 ATGAAATGAA AACGCGTAAA GGTATTATTT TAGCGGGTGG
TTCTGGTACA CGTCTTTATC

2881 CTGTGACTAT GGCTGTCAGT AAACAGCTAT TACCTATTTA
TGATAAGCCG ATGATCTATT

2941 ACCCGCTCTC TACACTGATG TTGGCGGGTA TTCGCGATAT
TCTGATTATT AGTACGCCAC

3001 AGGATACTCC TCGTTTTCAA CAACTGCTAG GTGACGGTAG
CCAGTGGGGG CTAAATCTTC

3061 AGTACAAAGT GCAACCGACT CCAGATGGGC TTGCGCAGGC
GTTTATTATC GGTGAAGAGT

3121 TTATCGGTGG TGATGATTGT GCTTTGGTTC TTGGTGATAA TATCTTCTAC
GGTCATGATC

3181 TGCCGAAGTT AATGGATGTC GCTGTTAACA AAGAAAGTGG

TGCAACGGTA TTTGCCTATC

3241 ACGTTAATGA TCCTGAACGC TACGGCGTCG TTGAGTTTGA
TAAAAACGGT ACGGCAATAA

3301 GCCTGGAAGA AAAACCGCTA CAACCAAAAA GTAATTATGC

GGTAACCGGG CTTTATTTCT

3361 ATGATAACGA CGTTGTCGAA ATGGCGAAAA ACCTTAAGCC
TTCTGCCCGT GGTGAACTGG

3421 AAATTACCGA TATTAACCGT ATTTATATGG AACAGGGGCG TTTATCCGTT
GCCATGATGG

3481 GGCGTGGTTA TGCATGGCTG GATACGGGGA CACATCAGAG
TCTTATTGAA GCAAGCAACT

3541 TCATTGCCAC CATTGAAGAG CGCCAGGGAC TAAAGGTTTC
CTGCCCAGAA GAAATTGCTT

3601 ACCGTAAAGG GTTTATTGAT GCTGAACAGG TGAAAGCATT

AGCGGAGCCG CTGAAAAAAA

3661 ATGCTTATGG ACAGTATCTG CTGAAAATGA TTAAAGGTTA TTAATAAAAT
GAACGTAATT

3721 AAAACAGAAA TTCCTGATGT GTTAATTTTC GAGCCGAAAG TTTTTGGTGA
TGAGCGTGGT

3781 TTCTTTATGG AAAGCTTTAA TCAGAAAGTT TTCGAAGAAG
CTGTAGGACG TAAGGTTGAA

3841 TTTGTTCAGG ATAACCATTC GAAGTCTAGT AAAGGTGTTT
TACGCGGGCT GCATTATCAG

3901 TTAGAACCTT ATGCGCAAGG GAAACTGGTA CGTTGCGTTG
TTGGTGAGGT TTTTGATGTA

3961 GCTGTTGATA TTCGTAAATC GTCGCCTACC TTTGGTAAAT
GGGTTGGGGT GAATTTATCT

4021 GCTGAGAATA AGCGGCAATT GTGGATCCCT GAGGGATTTG
CACATGGTTT TTTGGTGCTG

4081 AGCGAGACTG CGGAATTTTT ATATAAAACG ACGAACTATT
ATCATCCTGA TAGTGATAGA

4141 GGGATTGTAT GGAATGATCC TATTCTGAGC ATAAAATGGC
CGACGATAGA ACATAATAAT

4201 TATATTTTAT CGATTAAAGA TGCAAGGGCT AAAGAATTGC ATAACATGAA
GGAATTATTT

4261 TTGTGAGTAT TGTAAAGAAT ACTTTATGGA ATATAAGTGG GTATATTATA
CCATCATTAA

4321 TAGCAATTCC TGCGTTAGGT ATACTGTCTA GAATTCTAGG
GACCGAGCAA TTTGGCCTTT

4381 TTACGTTAGC TATTGCCTTA GTTGGATATG CAAGTATTTT TGATGCTGGA
TTGACCAGAG

4441 CTGTTATAAG AGAAGTATCA ATATATAAAA ATGTTCATAA AGAATTAAGA
GCGATCATTT

4501 CAACTTCAAC GGTAATTCTA ACTATATTGG GCTTGATTGG
CGGTAGTGTA CTATTTTTGA

4561 GTAGCAATGT AATTGTTAAA TTATTAAACA TTAACGCGAA TCATGTTGTA GAATCTGTCA

4621 AAGCAATATA TATTATTTCA GCTACCATAC CCTTATACTT GTTAAACCAA GTCTGGTTGG

4681 GGATTTTTGA GGGGATGGAA AAGTTCAGAA AAGTAAATTT

AATAAAATCA ATTAACAACT

4741 CTTTTGTGGC TGGATTACCA GTGATTTTCT GT TTTTTCA TGGAGGATTA CTAAGTGCTA

4801 TATATGGTTT AGTTATGGCA AGAGTCTTAT CACTTATAGT GACCTTTATA TTTAGTCGAA

4861 AACTAATAAT ATCATCTGGG CTGTCTGTAA AAATTGTAAC AGTTAAAAGA TTAATCGGCT

4921 TTGGAAGCTG GATAACAGTT AGCAATATTA TTAGCCCTAT TATGACATAT ATGGATCGTT

4981 TTATTCTTTC ACACATTGTG GGGGCTGATA AAGTTTCTTT TTATACTGCT CCGTCTGAAG

5041 GTATACAACG CTTAACGATA TTACCAAGTG CGTTGTCCAG AGCTATTTTT CCAAGATTAA

5101 GTTCAGAATT GCAATCGGTA AAGCAAACTA AAATATTATC ATATTTTATA ATGGTTATTG

5161 GTATACTTCC AATTGTAATG TTGATAATTA TTTTATCAGA TTTTATAATG TCCGCTTGGA

5221 TGGGACCTAC ATATCATGGG ACGCCAGGTA TAGTATTAAA AATTCTTGCA ATAGGTTTCT

5281 TTTTTAATTG CATTGCACAA ATCCCATTTG TTTCAGTTCA GGCTAGTGGA AGATCAAAAA

5341 TTACAGCTAT TATTCATTTG CTCGAAGTTA TCCCATATTT ATGCATATTA TATATTTTTA

5401 TTTATCATTG GGGAATTGTT GGAGCCGCAA TAGCATGGTC TGTAAGAACA TCGTTAGATT

5461 TTTTGATATT ATTATTAATT GATACGAAAT ATTAATAGCG AATTGATTTT AGGGATTACT

5521 TCCTCAAGCC CATCTAATTA GAGTGCAAAC ATGACTTCTG A ГП ATAA
CTCAAAAGAC

5581 AAAAGTTTAA GTGTTCTTTT GTTTTTTGGG TTTATATTTT TCCTTACACG
TAGCTTTCCA

5641 TTTATTCAAT ATAGTTGGAT TATGGAGGGG TTTTTATGTC TTTGTATCAT
GTCATTTACA

5701 AAGAAAATTG CAAACGGAAT ATATCACTAT CCTGTTATTT TAATATTTCT
ATTAGCTCTT

5761 TTTATAAATT TTATTTATTC CTATATCAAG GGTAACGATA TAGCGATAAT
AATTAGGTTT

5821 TATATTATCA TATTATTTAT ATTATGTGCT TATTTCTGCT CTTATGGAAC
CATCTCGATT

5881 GTTAAAATAT TTTTATATTT AATGGTATTA CAGGCGGTTA TTATATCCAT
CATTAGTATT

5941 TATATGACAA AAACATATGG TATTGGTGAT TATTCAGCAC TAAGACATTA
TTTTTTGGAG

6001 AATGATTATG GTGATGTTTA TACATATGGA AGTGGTTTCT ATAGAGTTCA
AATTAAAGGA

6061 AATGCTCTCA TTCCATTTGC CTTTATGTTG CATATAGTCA TAAAAGATTA
TTTCTATTAT

6121 CGATTCAAAA ATACAATAAC CGTTATTCTG GCTATAGGTA
CTATAGTGGC TGGTAATTTT

6181 GCATATTTTG TTTCGATATG CTTGTTTTTT ATGTATATTA TACTATGTTC
TAAATCTAAC

6241 TCACGATACG CTAAATTAAG GAAAATTATT TTTGGGGTTT TTCTTACTGT
GATTCTCCCT

6301 TTTTTTATTA CATATTCAAT TGAGTTGATA ATCATGAAAT CAAATGGAGC
TGATTCTTCT

6361 TTAGGAGTTA GATGGGATCA GTTTACTGTA TTAATTAATG ATCTTACAGA
GTCTGTATCA

6421 AATTTTGTTA TAGGTTCTGG TTTGGGTAAT GTCATCAAAA TTCAAACTCC
TATCCGTGAT

6481 TATAGTGCAT ATATATATTA TGAATTGCAG TCAGTTTATT TTTTAAATCA
ACTTGGCGTT

6541 ATTTTATTTA CTTTGTTTTT ATTAATTAAT CTCCTTCTCA CGATTAAAAT CATAAAATAC

6601 AGTGAGTTGT GTGTGCTATA TTTTCTATAT GTTTCTTATG CAATTACTAA TCCTTATATT

6661 TTAGACTCTA ACCATGTTGC TGTAATAATT GTATTAGTGA CATTAAGTAA TGTTCTAAAA

6721 AAGATGAAAG CTAAATGAAG GTTTTAAGGT GAAGATGGAC

ACTGTATATG CCGTTTTGGT

6781 TGCTTACAAC CCAGAACATA ATGATTTAAA AAATGCGGTT GAATTATTGT TGAGACAAGT

6841 TACTAAAGTT GTCGTTTGCA ATAACTCTAC AAATGGTTAT AAATATGCTG AAAATTCTTC

6901 AGGCGATGTA AAAATATTCA ATTTCAATGA TAATTTAGGC ATAGCAGAAG CCCAAAGTAT

6961 AGGAATGAAA TGGGCTTTTG AAAATGGCGC TGATTTTATA TTGCAAATGG ATCAGGATAG

7021 TATTCCTGAT CCTAAGATGG TAGAGCAGTT ACTTACTTGT TACAAAAAAT TGCTTAAACA

7081 AAATGTCAAT GTTGGTTTAG TTGGTTCACA AGATTTTGAT AAAGTAACTG GTGAATTAAA

7141 TAAAGCAAGG GTAAAAAAAG GGAAACCACT TACAGAAGTT

TATTATGAGG TAGATAGTAC

7201 ATTAAGTTCT GGCAGTCTAA TACCAAAAAA TAGTTGGTTG ATTGTTGGAG GAATGAAAGA

7261 TGAGCTTTTT ATCGATGCGG TAGACCATGA ATATTGTTGG AGATTAAGAG CTGCTGGGTT

7321 TAAAGTAATT AGGAATAAAA ATGCGTTACT TGCACATAGA CTTGGAGATG GGCGATTTAA

7381 GATCTTAAAT ATTCTTTCTG TCGGTTTGCC AAGCCCATTT CGTCATTATT ATGCTACTCG

7441 AAATATCTTT CTTTTATTAA ATAAAAATTA TGTACCCATC TACTGGAAAA TTTCTAGTCT

7501 GGTTAAATTA ATTGGAAAGG TTTTTTTATA TCCTATTTTC CTTCCAAATG
GTAATAAAAG

7561 GTTATATTTT TTTTTAAAAG GCATTAATGA CGGTTTAATG GGTCGAAGTG
GTAAAATGAA

7621 ATGAATCATA GATTAGAAAA ATTCTCAGTT TTAATTAGCA TTTATAAAAA
TGATCTACCG

7681 CAATTTTTTG AGGTGGCTCT ACGCTCTATT TTTCACGATC AAACACTTAA
GCCAGATCAA

7741 ATAGTAATTG TTGCAGATGG AGAACTCCAT CAAACACACA TCGATATTAT
AAATTCATTC

7801 ATTGATGATG TTGGCAATAA AATAGTAACA TTTGTACCTT TACCTAGAAA
TGTTGGATTG

7861 GCTAATGCCT TAAATGAAGG ATTAAAGGCT TGTAGGAATG
AGTTAGTGGC AAGAATGGAT

7921 GCTGATGATA TTTCTTTGCC TCATCGGTTT GAGA AAC AAA
TTTCTTTTAT GATTAATAAT

7981 TCAGAAATAG ATGTATGTGG CAGTTTTATT GATGAAATTG AAACTGTTAC
TGAGGAGTTT

8041 ATTTCAACAC GCAAAGTGCC TCTCGAACAT AGAGAAATAG
TTAAATTCGC GAGGAAACGA

8101 AGCGCAGTTA GCCATCCTTC TGTAATTTTT AGAAAGAATA CAGTATTAGC
TGTTGGTGGT

8161 TATCCTCCAT TCAGAAAATC TCAAGATTTT GCATTGTGGA GCCTATTAAT
TGTACATAAT

8221 GCAAGATTTG CAAATCTTCC AGATATTTTA TTAAAAATGC GAACTGGTCG
TAATCTTATG

8281 GCTCGACGTG GATTGTCATA TTTATTGTAC GAGTATAAAG TATTGTATTA
TCAATATAAA

8341 ATTGGTTTTA TTCGAAAAAA TGAATTAATA AGTAATGCTA TGTTGAGAAC
ATTTTTTCGT

8401 ATAATGCCAT CTAAATTAAA GGAGCTGATG TATTCAATCG
TTAGGAATCG ATAATAATAA

8461 TTTTCTGATT AAGTGTTATG GATTTATTTT TATTAGGCAT ATTCTATAAT
TAAGCATAAC

8521 CCGCATACCA CCCAGCGGTA TCCTGACAGG AGTAAACAAT

GTCAAAGCAA CAGATCGGCG

8581 TCGTCGGTAT GGCAGTGATG GGGCGCAACC TTGCGCTCAA
TATCGAAAGC CGTGGTTATA

8641 CCGTCTCTAT TTTCAACCGT TCCCGTGAAA AGACCGAAGA
AGTGATTACC GAAAATCCAG

8701 GCAAGAAACT GGTTCCTTAC TATACGGTGA AAGAATTTGT
TGAATCTCTG GAAACGCCTC

8761 GTCGCATCCT GTTAATGGTG AAAGCAGGTG CTGGCACGGA
TGCTGCTATT GATTCCCTCA

8821 AGCCATACCT CGATAAAGGT GACATCATCA TTGATGGTGG
TAACACCTTC TTCCATGACA

8881 CCATTCGTCG TAACCGTGAG CTTTCTGCAG AAGGCTTTAA
CTTTATCGGT ACCGGTGTTT

8941 CCGGTGGTGA AGAAGGTGCG CTGAAAGGTC CTTCCATTAT
GCCTGGTGGG CAGAAAGAAG

9001 CTTATGAACT GATTGCGCCG ATCCTGACCA AAATCGCCGC
TGTGGCTGAA GACGGCGAAC

9061 CGTGCGTTAC CTATATTGGT GCCGATGGTG CAGGTCATTA
TGTGAAGATG GTTCACAACG

9121 GTATTGAATA CGGTGATATG CAGCTGATTG CTGAAGCCTA
TTCTCTGCTT AAAGGTGGCT

9181 TGAACCTCAC CAACGAAGAA CTGGCGCAGA CCTTTACCGA

GTGGAATAAC GGTGAACTGA

9241 GCAGCTACCT GATCGACATC ACCAAAGATA TCTTCACCAA

AAAAGATGAA GAGGGTAACT

9301 ACCTGGTTGA TGTGATTCTG GATGAAGCAG CAAACAAAGG

TACGGGCAAA TGGACCAGCC

9361 AGAGCGCGCT GGATCTCGGC GAACCGCTGT CGCTGATTAC
CGAGTCTGTG TTTGCACGTT

9421 ATATCTCTTC TCTGAAAGAG CAGCGTGTTG CCGCATCTAA
AGTTCTCTCT GGCCCGCAAG

9481 CGCAGCCAGC TGGCGACAAT GCTGAGTTCA TCGAAAAAGT

TCGCCGTGCG CTGTATCTGG

9541 GCAAAATCGT TTCTTACGCT CAGGGCTTCT CTCAGCTACG
CGCTGCGTCT GAAGAGTACA

9601 ACTGGGATCT GAACTACGGT GAAATCGCGA AGATTTTCCG
TGCTGGCTGC ATCATCCGTG

9661 CGCAGTTCCT GCAGAAAATC ACCGATGCTT ATGCCGAAAA
TCCGCAGATC GCTAACCTGT

9721 TGCTGGCTCC TTACTTCAAG CAAATTGCCG ATGACTACCA
GCAGGCGCTG CGCGATGTCG

9781 TCGCTTACGC AGTACAGAAC GGTATCCCGG TGCCGACCTT
CGCCGCTGCG GTTGCCTATT

9841 ACGACAGCTA CCGCGCCGCT GTTCTGCCTG CGAACCTGAT
CCAGGCACAG CGTGACTATT

9901 TCGGTGCGCA TACTTATAAG CGCATTGATA AAGAAGGTGT
GTTCCATACC GAATGGCTGG

9961 ATTAATCTGA TTTAAATCAA TTAATCAAAG CAAGGCCCGG
AGAAACCCTC CGGGCTTTTT

10021 TATTATACAA AGCGGCAGGT TAGGGCCTTT TTTTATAATT
TATAGTTAAA AACGCGATAT

10081 AATACAGCGC CGCACAGCAG GATCGCTGCC TTGACAGTTC

ATCTACATCA GCGTTAAAAA

10141 TCCCGCAGTA GATGAAGCTG TGGTGGTGGA TTAATGACCA
CTCTAAATGT TTAACCGGAA

10201 GAAGTCAGAG CTAATGAAAA TAACAATTTC AGGAACAGGT

TATGTTGGTC TTTCAAATGG

10261 TATTCTGATT GCGCAAAACC ACGAAGTGGT TGCACTGGAT
ATCGTTCAGG CCAAAGTGGA

10321 CATGCTTAAC AAGAGGCAGT CACCGCTTGT TGATAAGGAG

ATTGAAGAGT ATCTGGCGAC

10381 TAAAGATCTC AATTTCCGCG CTACGACAGA TAAGTATGAC

GCGTATAAAA ATGCCGATTA

10441 CGTTATTATT GCCACACCTA CCGATTATGA TCCGAAAACA
AATTATTTTA ATACCTCAAG

10501 CGTGGAAGCG GTCATTCGTG ATGTGACAGA AATTAATCCC

AACGCGGTAA TGATTATAAA

10561 ATCAACTATC CCTGTTGGTT TTACAGAGTC CATTAAAGAA
CGTTTTGGTA TTGAAAATGT

10621 GATCTTTTCG CCTGAGTTTT TGCGTGAAGG TAAAGCACTT
TATGATAACT TACACCCATC

10681 ACGCATTGTG ATTGGCGAGC AGTCTGAACG CGCTAAACGT
TTTGCTGCGT TATTACAGGA

10741 AGGCGCCATT AAGCAAGACA TACCAACATT GTTTACTGAC

TCAACCGAGG CTGAGGCGAT

10801 TAAACTTTTT GCGAACACTT ATCTGGCGAT GCGTGTAGCG
TATTTCAATG AACTTGATAG

10861 TTATGCTGAA AGCCTGGGAC TTAATTCACG CC AG ATT ATT
GAGGGCGTAT GCCTTGACCC

10921 GCGTATCGGT AATCACTACA ACAACCCGTC ATTCGGTTAT
GGTGGTTATT GTCTGCCGAA

10981 AGATACTAAG CAGTTACTGG CAAATTACCA GTCTGTGCCG

AATAACCTGA TCTCGGCAAT

11041 TGTTGACGCC AACCGCACGC GCAAAGATTT TATTGCCGAT
TCTATCCTTG CACGTAAACC

11101 GAAAGTTGTT GGCGTCTATC GTTTGATTAT GAAGAATGGT
TCAGACAATT TTCGTGCTTC

11161 CTCGATTCAG GGTATTATGA AGCGAATCAA GGCGAAAGGT

GTGCCTGTAA TCGTTTATGA

11221 GCCAGCTATG AAAGAGGACG ATTTTTTCCG GTCGCGCGTG GTACGTGATC TGGATGCGTT

11281 CAAACAAGAA GCTGATGTTA TTATTTCTAA CCGTATGTCT GCCGATCTGG CTGATGTAGC

1 1341 AGATAAAGTT TATACGCGCG ACTTGTTTGG CAATGATTAA TTATTTTGTT TCATTCTAAG

1 1401 AAAAGGCCCT AATAAATTAG GGCCTTTTCT TATGGTTTTG TAAAATCAAA CTTTATAGAA

1 1461 GTTACGATAC CATTCTACAA AGTTCTTTAC CCCTTCTTTA ACTGACGTTT CAGGTTTGAA

11521 TCCTATTACG TCATACAGTG CTTTTGTATC AGCACTGGTT TCCAGTACAT CACCGGGTTG

11581 GAGAGGCATC ATATTTTTGT TGGCTTCAAT ACCCAGAGCC TCTTCTAACG CATTGATATA

11641 GTCCATCAAC TCCACAGGCG AACTATTACC AATGTTATAG ACACGATATG GTGCTGAACT

11701 TGTTGCAGGC GAGCCTGTTT CTACAGCCCA CTGTGGGTTT TTTTCTGGAA TAACATCCTG

11761 TAAGCGAATA ATAGCTTCGG CAATATCATC AATGTAAGTA

AAGTCACGCT TCATTTTGCC

11821 GAAGTTGTAA ACATCAATGC TTTTACCTTC CAGCATGGCT TTAGTGAATT TAAATAATGC

11881 CATATCCGGA CGTCCCCATG GACCATAAAC CGTAAAGAAA

CGCAGCCCTG TGGTCGGTAA

11941 GCCATACAAA TGAGAATATG TATGGGCCAT GAGTTCATTC GCTTTTTTAG TTGCTGCATA

12001 AAGCGAAACA GGATGATCTA CAGAGTCATC TGTAGAGAAA

GGCATCTTGC GGTTCATGCC

12061 ATAAACAGAA CTGGAGGAAG CGTAAAGTAG ATGCTGAACA

TTATTATGGC GACATCCTTC

12121 TAGTATGTTC AGGAATCCAA TCAGGTTTGC ATCTGCATAT
GCATTGGGAT TTTCAAGAGA

12181 GTAACGTACA CCGGCTTGCG CAGCGAGGTT TATTACGCGT
TCGAACCGCT CGTCTGCAAA

12241 CAGTGCCGCC ATTTTCTCAC GATCGGCCAG GTCAATTTTA
TAAAAACTGA AGTTGTCGTG

12301 CTTGAGTAAA TCAAGTCGTG CTTGTTTGAG GTTGACATCG
TAATAATCAT TTAAGTTGTC

12361 AATGCCTACA ACCTGATGAC CAGCTGCAAG AAGCCGTTTA

CTTAGATAGA AACCGATAAA

12421 GCCAGCAGCT CCCGTAACCA GAAATTTCAT TTATAATCCT
CGCTCAGGCT AGAATATAGC

12481 CAATCTTCAT CTGGCATAAC TGAAAGTTAA ATTATACCGT TAG AC A AG
A A AAAAAGATAA

12541 TCGGTATCAG TTCTAAACTT GGCTGTTTTT TCTGGTAACG
TGCTCATTTT ACAATCAAAG

12601 CTGTTCTAAG CTGACTATAC AAGCCGACGT CATTATCTCC
AACCGTATGG CAGAAGAGCT

12661 TAAGGATGTG GCAGACAAAG TCTACACCCG CGATCTCTTT

GGCAGTGACT AACATCCTGT

12721 TATCATGGCG ATTTTCGCCC TGATTCTCTT ATGTTCCCTT
TGTAATAATT CATTATTTTT

12781 ATCATTTATC CTATAGCATT CATGGCGATT ATCGCTAAAC
TATGGCGGCG CGCCACGTGG

12841 GATCCCCGGG TACCGAGCTC GAATTCGCCC TATAGTGAGT
CGTATTACAA TTCACTGGCC

12901 GTCGTTTTAC AACGTCGTGA CTGGGAAAAC CCTGGCGTTA
CCCAACTTAA TCGCCTTGCA

12961 GCACATCCCC CTTTCGCCAG CTGGCGTAAT AGCGAAGAGG

CCCGCACCGA TCGCCCTTCC

13021 CAACAGTTGC GCAGCTGAAT GGCGAATGGC GCCTGATGCG
GTATTTTCTC CTTACGCATC

13081 TGTGCGGTAT TTCACACCGC ATATGGTGCA CTCTCAGTAC
AATCTGCTCT GATGCCGCAT

13141 AGTTAAGCCA GCCCCGACAC CCGCCAACAC CCGCTGACGC

GAACCCCTTG CGGCCGCATC

13201 GAATATAACT TCGTATAATG TATGCTATAC GAAGTTATTA
GCGATGAGCT CGGACTTCCA

13261 TTGTTCATTC CACGGACAAA AACAGAGAAA GGAAACGACA

GAGGCCAAAA AGCTCGCTTT

13321 CAGCACCTGT CGTTTCCTTT CTTTTCAGAG GGTATTTTAA
ATAAAAACAT TAAGTTATGA

13381 CGAAGAAGAA CGGAAACGCC TTAAACCGGA AAATTTTCAT

AAATAGCGAA AACCCGCGAG

13441 GTCGCCGCCC CGTAACCTGT CGGATCACCG GAAAGGACCC

GTAAAGTGAT AATGATTATC

13501 ATCTACATAT CACAACGTGC GTGGAGGCCA TCAAACCACG

TCAAATAATC AATTATGACG

13561 CAGGTATCGT ATTAATTGAT CTGCATCAAC TTAACGTAAA
AACAACTTCA GACAATACAA

13621 ATCAGCGACA CTGAATACGG GGCAACCTCA TGTCCGAGCT

CGCGAGCTCG TCGACAGCGA

13681 CACACTTGCA TCGGATGCAG CCCGGTTAAC GTGCCGGCAC

GGCCTGGGTA ACCAGGTATT

13741 TTGTCCACAT AACCGTGCGC AAAATGTTGT GGATAAGCAG

GACACAGCAG CAATCCACAG

13801 CAGGCATACA ACCGCACACC GAGGTTACTC CGTTCTACAG

GTTACGACGA CATGTCAATA

13861 CTTGCCCTTG ACAGGCATTG ATGGAATCGT AGTCTCACGC
TGATAGTCTG ATCGACAATA

13921 CAAGTGGGAC CGTGGTCCCA GACCGATAAT CAGACCGACA

ACACGAGTGG GATCGTGGTC

13981 CCAGACTAAT AATCAGACCG ACGATACGAG TGGGACCGTG

GTCCCAGACT AATAATCAGA

14041 CCGACGATAC GAGTGGGACC GTGGTTCCAG ACTAATAATC

AGACCGACGA TACGAGTGGG

14101 ACCGTGGTCC CAGACTAATA ATCAGACCGA CGATACGAGT

GGGACCATGG TCCCAGACTA

14161 ATAATCAGAC CGACGATACG AGTGGGACCG TGGTCCCAGT

CTGATTATCA GACCGACGAT

14221 ACGAGTGGGA CCGTGGTCCC AGACTAATAA TCAGACCGAC

GATACGAGTG GGACCGTGGT

14281 CCCAGACTAA TAATCAGACC GACGATACGA GTGGGACCGT

GGTCCCAGTC TGATTATCAG

14341 ACCGACGATA CAAGTGGAAC AGTGGGCCCA GAGAGAATAT

TCAGGCCAGT TATGCTTTCT

14401 GGCCTGTAAC AAAGGACATT AAGTAAAGAC AGATAAACGT

AGACTAAAAC GTGGTCGCAT

14461 CAGGGTGCTG GCTTTTCAAG TTCCTTAAGA ATGGCCTCAA
TTTTCTCTAT ACACTCAGTT

14521 GGAACACGAG ACCTGTCCAG GTTAAGCACC ATTTTATCGC
CCTTATACAA TACTGTCGCT

14581 CCAGGAGCAA ACTGATGTCG TGAGCTTAAA CTAGTTCTTG

ATGCAGATGA CGTTTTAAGC

14641 ACAGAAGTTA AAAGAGTGAT AACTTCTTCA GCTTCAAATA

TCACCCCAGC TTTTTTCTGC

14701 TCATGAAGGT TAGATGCCTG CTGCTTAAGT AATTCCTCTT TATCTGTAAA GGCTTTTTGA

14761 AGTGCATCAC CTGACCGGGC AGATAGTTCA CCGGGGTGAG

AAAAAAGAGC AACAACTGAT

14821 TTAGGCAATT TGGCGGTGTT GATACAGCGG GTAATAATCT TACGTGAAAT ATTTTCCGCA

14881 TCAGCCAGCG CAGAAATATT TCCAGCAAAT TCATTCTGCA ATCGGCTTGC ATAACGCTGA

14941 CCACGTTCAT AAGCACTTGT TGGGCGATAA TCGTTACCCA ATCTGGATAA TGCAGCCATC

15001 TGCTCATCAT CCAGCTCGCC AACCAGAACA CGATAATCAC TTTCGGTAAG TGCAGCAGCT

15061 TTACGACGGC GACTCCCATC GGCAATTTCT ATGACACCAG ATACTCTTCG ACCGAACGCC

15121 GGTGTCTGTT GACCAGTCAG TAGAAAAGAA GGGATGAGAT

CATCCAGTGC GTCCTCAGTA

15181 AGCAGCTCCT GGTCACGTTC ATTACCTGAC CATACCCGAG AGGTCTTCTC AACACTATCA

15241 CCCCGGAGCA CTTCAAGAGT AAACTTCACA TCCCGACCAC

ATACAGGCAA AGTAATGGCA

15301 TTACCGCGAG CCATTACTCC TACGCGCGCA ATTAACGAAT CCACCATCGG GGCAGCTGGT

15361 GTCGATAACG AAGTATCTTC AACCGGTTGA GTATTGAGCG TATGTTTTGG AATAACAGGC

15421 GCACGCTTCA TTATCTAATC TCCCAGCGTG GTTTAATCAG ACGATCGAAA ATTTCATTGC

15481 AGACAGGTTC CCAAATAGAA AGAGCATTTC TCCAGGCACC

AGTTGAAGAG CGTTGATCAA

15541 TGGCCTGTTC AAAAACAGTT CTCATCCGGA TCTGACCTTT ACCAACTTCA TCCGTTTCAC

15601 GTACAACATT TTTTAGAACC ATGCTTCCCC AGGCATCCCG AATTTGCTCC TCCATCCACG

15661 GGGACTGAGA GCCATTACTA TTGCTGTATT TGGTAAGCAA

AATACGTACA TCAGGCTCGA

15721 ACCCTTTAAG ATCAACGTTC TTGAGCAGAT CACGAAGCAT

ATCGAAAAAC TGCAGTGCGG

15781 AGGTGTAGTC AAACAACTCA GCAGGCGTGG GAACAATCAG

CACATCAGCA GCACATACGA

15841 CATTAATCGT GCCGATACCC AGGTTAGGCG CGCTGTCAAT AACTATGACA TCATAGTCAT

15901 GAGCAACAGT TTCAATGGCC AGTCGGAGCA TCAGGTGTGG

ATCGGTGGGC AGTTTACCTT

15961 CATCAAATTT GCCCATTAAC TCAGTTTCAA TACGGTGCAG AGCCAGACAG GAAGGAATAA

16021 TGTCAAGCCC CGGCCAGCAA GTGGGCTTTA TTGCATAAGT GACATCGTCC TTTTCCCCAA

16081 GATAGAAAGG CAGGAGAGTG TCTTCTGCAT GAATATGAAG

ATCTGGTACC CATCCGTGAT

16141 ACATTGAGGC TGTTCCCTGG GGGTCGTTAC CTTCCACGAG CAAAACACGT AGCCCCTTCA

16201 GAGCCAGATC CTGAGCAAGA TGAACAGAAA CTGAGGTTTT

GTAAACGCCA CCTTTATGGG

16261 CAGCAACCCC GATCACCGGT GGAAATACGT CTTCAGCACG

TCGCAATCGC GTACCAAACA

16321 CATCACGCAT ATGATTAATT TGTTCAATTG TATAACCAAC ACGTTGCTCA ACCCGTCCTC

16381 GAATTTCCAT ATCCGGGTGC GGTAGTCGCC CTGCTTTCTC GGCATCTCTG ATAGCCTGAG

16441 AAGAAACCCC AACTAAATCC GCTGCTTCAC CTATTCTCCA
GCGCCGGGTT ATTTTCCTCG

16501 CTTCCGGGCT GTCATCATTA AACTGTGCAA TGGCGATAGC
CTTCGTCATT TCATGACCAG

16561 CGTTTATGCA CTGGTTAAGT GTTTCCATGA GTTTCATTCT
GAACATCCTT TAATCATTGC

16621 TTTGCGTTTT TTTATTAAAT CTTGCAATTT ACTGCAAAGC AACAACAAAA
TCGCAAAGTC

16681 ATCAAAAAAC CGCAAAGTTG TTTAAAATAA GAGCAACACT

ACAAAAGGAG ATAAGAAGAG

16741 CACATACCTC AGTCACTTAT TATCACTAGC GCTCGCCGCA
GCCGTGTAAC CGAGCATAGC

16801 GAGCGAACTG GCGAGGAAGC AAAGAAGAAC TGTTCTGTCA

GATAGCTCTT ACGCTCAGCG

16861 C A AG A AGA AA T ATCC ACCGT GGG A AA A ACT CC AGGTAG AG

GTACACACGC GGATAGCCAA

16921 TTCAGAGTAA TAAACTGTGA TAATCAACCC TCATCAATGA

TGACGAACTA ACCCCCGATA

16981 TCAGGTCACA TGACGAAGGG AAAGAGAAGG AAATCAACTG

TGACAAACTG CCCTCAAATT

17041 TGGCTTCCTT A AAA ATT AC A GTTCAAAAAG TATGAGAAAA

TCCATGCAGG CTGAAGGAAA

17101 CAGCAAAACT GTGACAAATT ACCCTCAGTA GGTCAGAACA

AATGTGACGA ACCACCCTCA

17161 AATCTGTGAC AGATAACCCT CAGACTATCC TGTCGTCATG
GAAGTGATAT CGCGGAAGGA

17221 AAATACGATA TGAGTCGTCT GGCGGCCTTT CTTTTTCTCA
ATGTATGAGA GGCGCATTGG

17281 AGTTCTGCTG TTGATCTCAT TAACACAGAC CTGCAGGAAG

CGGCGGCGGA AGTCAGGCAT

17341 ACGCTGGTAA CTTTGAGGCA GCTGGTAACG CTCTATGATC CAGTCGATTT TCAGAGAGAC

17401 GATGCCTGAG CCATCCGGCT TACGATACTG ACACAGGGAT

TCGTATAAAC GCATGGCATA

17461 CGGATTGGTG ATTTCTTTTG TTTCACTAAG CCGAAACTGC GTAAACCGGT TCTGTAACCC

17521 GATAAAGAAG GGAATGAGAT ATGGGTTGAT ATGTACACTG

TAAAGCCCTC TGGATGGACT

17581 GTGCGCACGT TTGATAAACC AAGGAAAAGA TTCATAGCCT TTTTCATCGC CGGCATCCTC

17641 TTCAGGGCGA TAAAAAACCA CTTCCTTCCC CGCGAAACTC TTCAATGCCT GCCGTATATC

17701 CTTACTGGCT TCCGCAGAGG TCAATCCGAA TATTTCAGCA TATTTAGCAA CATGGATCTC

17761 GCAGATACCG TCATGTTCCT GTAGGGTGCC ATCAGATTTT CTGATCTGGT CAACGAACAG

17821 ATACAGCATA CGTTTTTGAT CCCGGGAGAG ACTATATGCC GCCTCAGTGA GGTCGTTTGA

17881 CTGGACGATT CGCGGGCTAT TTTTACGTTT CTTGTGATTG ATAACCGCTG TTTCCGCCAT

17941 GACAGATCCA TGTGAAGTGT GACAAGTTTT TAGATTGTCA CACTAAATAA AAAAGAGTCA

18001 ATAAGCAGGG ATAACTTTGT GAAAAAACAG CTTCTTCTGA

GGGCAATTTG TCACAGGGTT

18061 AAGGGCAATT TGTCACAGAC AGGACTGTCA TTTGAGGGTG

ATTTGTCACA CTGAAAGGGC

18121 AATTTGTCAC AACACCTTCT CTAGAACCAG CATGGATAAA

GGCCTACAAG GCGCTCTAAA

18181 AAAGAAGATC TAAAAACTAT AAAAAAAATA ATTATAAAAA

TATCCCCGTG GATAAGTGGA

18241 TAACCCCAAG GGAAGTTTTT TCAGGCATCG TGTGTAAGCA

GAATATATAA GTGCTGTTCC

18301 CTGGTGCTTC CTCGCTCACT CGACCGGGAG GGTTCGAGAA

GGGGGGGCAC CCCCCTTCGG

18361 CGTGCGCGGT CACGCGCACA GGGCGCAGCC CTGGTTAAAA

ACAAGGTTTA TAAATATTGG

18421 TTTAAAAGCA GGTTAAAAGA CAGGTTAGCG GTGGCCGAAA

AACGGGCGGA AACCCTTGCA

18481 AATGCTGGAT TTTCTGCCTG TGGACAGCCC CTCAAATGTC
AATAGGTGCG CCCCTCATCT

18541 GTCAGCACTC TGCCCCTCAA GTGTCAAGGA TCGCGCCCCT
CATCTGTCAG TAGTCGCGCC

18601 CCTCAAGTGT CAATACCGCA GGGCACTTAT CCCCAGGCTT
GTCCACATCA TCTGTGGGAA

18661 ACTCGCGTAA AATCAGGCGT TTTCGCCGAT TTGCGAGGCT
GGCCAGCTCC ACGTCGCCGG

18721 CCGAAATCGA GCCTGCCCCT CATCTGTCAA CGCCGCGCCG

GGTGAGTCGG CCCCTCAAGT

18781 GTCAACGTCC GCCCCTCATC TGTCAGTGAG GGCCAAGTTT
TCCGCGAGGT ATCCACAACG

18841 CCGGCGGCCG GCCGCGGTGT CTCGCACACG GCTTCGACGG
CGTTTCTGGC GCGTTTGCAG

18901 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG

CCGAGGGCTT CGCCCTGTCG

18961 CTCGACTGCG GCGAGCACTA CTGGCTGTAA AAGGACAGAC

CACATCATGG TTCTGTGTTC

19021 ATTAGGTTGT TCTGTCCATT GCTGACATAA TCCGCTCCAC
TTCAACGTAA CACCGCACGA

19081 AGATTTCTAT TGTTCCTGAA GGCATATTCA AATCGTTTTC
GTTACCGCTT GCAGGCATCA

19141 TG AC AG A AC A CTACTTCCTA TAAACGCTAC ACAGGCTCCT
GAGATTAATA ATGCGGATCT

19201 CTACGATAAT GGGAGATTTT CCCGACTGTT TCGTTCGCTT
CTCAGTGGAT AACAGCCAGC

19261 TTCTCTGTTT AACAGACAAA AACAGCATAT CCACTCAGTT
CCACATTTCC ATATAAAGGC

19321 CAAGGCATTT ATTCTCAGGA TAATTGTTTC AGCATCGCAA
CCGCATCAGA CTCCGGCATC

19381 GCAAACTGCA CCCGGTGCCG GGCAGCCACA TCCAGCGCAA

AAACCTTCGT GTAGACTTCC

19441 GTTGAACTGA TGGACTTATG TCCCATCAGG CTTTGCAGAA
CTTTCAGCGG TATACCGGCA

19501 TACAGCATGT GCATCGCATA GGAATGGCGG AACGTATGTG

GTGTGACCGG AACAGAGAAC

19561 GTCACACCGT CAGCAGCAGC GGCGGCAACC GCCTCCCCAA

TCCAGGTCCT GACCGTTCTG

19621 TCCGTCACTT CCCAGATCCG CGCTTTCTCT GTCCTTCCTG
TGCGACGGTT ACGCCGCTCC

19681 ATGAGCTTAT CGCGAATAAA TACCTGTGAC GGAAGATCAC

TTCGCAGAAT AAATAAATCC

19741 TGGTGTCCCT GTTGATACCG GGAAGCCCTG GGCCAACTTT
TGGCGAAAAT GAGACGTTGA

19801 TCGGCACGTA AGAGGTTCCA ACTTTCACCA TAATGAAATA

AGATCACTAC CGGGCGTATT

19861 TTTTGAGTTA TCGAGATTTT CAGGAGCTAA GGAAGCTAAA

ATGGAGAAAA AAATCACTGG

19921 ATATACCACC GTTGATATAT CCCAATGGCA TCGTAAAGAA
CATTTTGAGG CATTTCAGTC

19981 AGTTGCTCAA TGTACCTATA ACCAGACCGT TCAGCTGGAT
ATTACGGCCT TTTTAAAGAC

20041 CGTAAAGAAA AATAAGCACA AGTTTTATCC GGCCTTTATT
CACATTCTTG CCCGCCTGAT

20101 GAATGCTCAT CCGGAATTTC GTATGGCAAT GAAAGACGGT

GAGCTGGTGA TATGGGATAG

20161 TGTTCACCCT TGTTACACCG TTTTCCATGA GCAAACTGAA
ACGTTTTCAT CGCTCTGGAG

20221 TGAATACCAC GACGATTTCC GGCAGTTTCT ACACATATAT
TCGCAAGATG TGGCGTGTTA

20281 CGGTGAAAAC CTGGCCTATT TCCCTAAAGG GTTTATTGAG
AATATGTTTT TCGTCTCAGC

20341 CAATCCCTGG GTGAGTTTCA CCAGTTTTGA TTTAAACGTG
GCCAATATGG ACAACTTCTT

20401 CGCCCCCGTT TTCACCATGG GCAAATATTA TACGCAAGGC

GACAAGGTGC TGATGCCGCT

20461 GGCGATTCAG GTTCATCATG CCGTTTGTGA TGGCTTCCAT
GTCGGCAGAA TGCTTAATGA

20521 ATTACAACAG TACTGCGATG AGTGGCAGGG CGGGGCGTAA

TTTTTTTAAG GCAGTTATTG

20581 GTGCCCTTAA ACGCCTGGTT GCTACGCCTG AATAAGTGAT

AATAAGCGGA TGAATGGCAG

20641 AAATTCGATG ATAAGCTGTC AAACATGAGA ATTGGTCGAC

GGCCCGGGCG GCCGCAAGGG

20701 GTTCGCGTTG GCCGATTCAT TAATGCAGCT GGCACGACAG
GTTTCCCGAC TGGAAAGCGG

20761 GCAGTGAGCG CAACGCAATT AATGTGAGTT AGCTCACTCA

TTAGGCACCC CAGGCTTTAC

20821 ACTTTATGCT TCCGGCTCGT ATGTTGTGTG GAATTGTGAG
CGGATAACAA TTTCACACAG

20881 GAAACAGCTA TGACCATGAT TACGCCAAGC TATTTAGGTG

AGACTATAGA ATACTCAAGC

20941 TTGCATGCCT GCAGGTCGAC TCTAGAGGAT CCCACGACGT CG

SEQ ID NO: 29 - Secuencia de nucleótidos para pCCIFOS cortado (pFOS) y antígeno O de *S. flexneri* 6 con Z3206

Locus pFOS cortado y Antígeno O cortado (Z3206+)

Definición unión del agrupamiento de antígeno O invertido de *S. flexneri* 6 amplificado con Z3206 Nhe y wzzAscI cortados con NheI y AscI en pCCIFOS con el casete MCS cortado con NheI y AscI

Características Localización/Calificadores

CDS complementaria (370..396)

/marcador=wzz'

CDS 748..1752

/marcador=uge

CDS complementaria (1818..3011)

/marcador=ugd

CDS complementaria (3233..4639)

/marcador=gnd

CDS complementaria (4744..5577)

/marcador=wfbZ

CDS complementaria (5574..6443) [0429] /marcador=wfbY

CDS complementaria (6460..7647) [0431] /marcador=wzy

CDS complementaria (7703..8935) [0433] /marcador=wzx

CDS complementaria (8932..9489) [0435] /marcador=rmIC

CDS complementaria (9494..10372) [0437] /marcador=rmIA

CDS complementaria (10430..11329) [0439] /marcador=rmID

CDS complementaria (11329..12414) [0441] /marcador=rmIB

CDS complementaria (12787..13680) [0443] /marcador=galF

CDS complementaria 13912..14907) [0445] /marcador=Z3206

CDS complementaria (15065..15097) [0447] /marcador='wcaM

CDS complementaria 15525..16184) [0449] /marcador=cat

CDS 16403..16750

/marcador=redF

CDS 18145..18900

/marcador=repE

CDS 19479..20654

/marcador=parA

CDS 20654..21625

/marcador=parB

Longitud: 22887 pb

Tipo: ADN circular UNA

Secuencia:

1 GCGGCCGCAA GGGGTTCGCG TCAGCGGGTG TTGGCGGGTG
TCGGGGCTGG CTTAACTATG

61 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GT
CCGCACAGAT

121 GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT
CAGCTGCGCA ACTGTTGGGA

181 AGGGCGATCG GTGCGGGCCT CTTCGCTATT ACGCCAGCTG
GCGAAAGGGG GATGTGCTGC

241 AAGGCGATTA AGTTGGGTAA CGCCAGGGTT TTCCCAGTCA
CGACGTTGTA AAACGACGGC

301 CAGTGAATTG TAATACGACT CACTATAGGG CGAATTCGAG
CTCGGTACCC GGGGATCCCA

361 CGTGGCGCGC CGCCATAGTT TAGCGATAAT CGCCATGAAT
GCTATAGGAT AAATGATAAA

421 AATAATGAAT TATTACAAAG GGAACATAAG AGAATCAGGG

CGAAAATCGC CATGATAACA

481 GGATGTTAGT CACTGCCAAA GAGATCGCGG GTGTAGACTT
TGTCTGCCAC ATCCTTAAGC

541 TCTTCTGCCA TACGGTTGGA GATAATGACG TCGGCTTGTA TAGTCAGCTT
AGAACAGCTT

601 TGATTGTAAA ATGAGCACGT TACCAGAAAA AACAGCCAAG TTTAGAACTG
ATACCGATTA

661 TCTTTTTTTC TTGTCTAACG GTATAATTTA ACTTTCAGTT ATGCCAGATG
AAGATTGGCT

721 ATATTCTAGC CTGAGCGAGG ATTATAAATG AAATTTCTGG TTACGGGAGC
TGCTGGCTTT

781 ATCGGTTTCT ATCTAAGTAA ACGGCTTCTT GCAGCTGGTC ATCAGGTTGT
AGGCATTGAC

841 AACTTAAATG ATTATTACGA TGTCAACCTC AAACAAGCAC GACTTGATTT
ACTCAAGCAC

901 GACAACTTCA GTTTTTATAA AATTGACCTG GCCGATCGTG AGAAAATGGC
GGCACTGTTT

961 GCAGACGAGC GGTTCGAACG CGTAATAAAC CTCGCTGCGC
AAGCCGGTGT ACGTTACTCT

1021 CTTGAAAATC CCAATGCATA TGCAGATGCA AACCTGATTG
GATTCCTGAA CATACTAGAA

1081 GGATGTCGCC ATAATAATGT TCAGCATCTA CTTTACGCTT
CCTCCAGTTC TGTTTATGGC

1141 ATGAACCGCA AGATGCCTTT CTCTACAGAT GACTCTGTAG
ATCATCCTGT TTCGCTTTAT

1201 GCAGCAACTA AAAAAGCGAA TGAACTCATG GCCCATACAT
ATTCTCATTT GTATGGCTTA

1261 CCGACCACAG GGCTGCGTTT CTTTACGGTT TATGGTCCAT
GGGGACGTCC GGATATGGCA

1321 TTATTTAAAT TCACTAAAGC CATGCTGGAA GGTAAAAGCA TTGATGTTTA
CAACTTCGGC

1381 AAAATGAAGC GTGACTTTAC TTACATTGAT GATATTGCCG AAGCTATTAT
TCGCTTACAG

1441 GATGTTATTC CAGAAAAAAA CCCACAGTGG GCTGTAGAAA

CAGGCTCGCC TGCAACAAGT

1501 TCAGCACCAT ATCGTGTCTA TAACATTGGT AATAGTTCGC
CTGTGGAGTT GATGGACTAT

1561 ATCAATGCGT TAGAAGAGGC TCTGGGTATT GAAGCCAACA

AAAATATGAT GCCTCTCCAA

1621 CCCGGTGATG TACTGGAAAC CAGTGCTGAT ACAAAAGCAC
TGTATGACGT AATAGGATTC

1681 AAACCTGAAA CGTCAGTTAA AGAAGGGGTA AAGAACTTTG

TAGAATGGTA TCGTAACTTC

1741 TATAAAGTTT GATTTTACAA AACCATAAGA AAAGGCCCTA ATTTATTAGG
GCCTTTTCTT

1801 AGAATGAAAC AAAATAATTA ATCATTGCCA AACAAGTCGC

GCGTATAAAC TTTATCTGCT

1861 ACATCAGCCA GATCGGCAGA CATACGGTTA GAAATAATAA

CATCAGCTTC TTGTTTGAAC

1921 GCATCCAGAT CACGTACCAC GCGCGACCGG AAAAAATCGT
CCTCTTTCAT AGCTGGCTCA

1981 TAAACGATTA CAGGCACACC TTTCGCCTTG ATTCGCTTCA
TAATACCCTG AATCGAGGAA

2041 GCACGAAAAT TGTCTGAACC ATTCTTCATA ATCAAACGAT
AGACGCCAAC AACTTTCGGT

2101 TTACGTGCAA GGATAGAATC GGCAATAAAA TCTTTGCGCG
TGCGGTTGGC GTCAACAATT

2161 GCCGAGATCA GGTTATTCGG CACAGACTGG TAATTTGCCA
GTAACTGCTT AGTATCTTTC

2221 GGCAGACAAT AACCACCATA ACCGAATGAC GGGTTGTTGT
AGTGATTACC GATACGCGGG

2281 TCAAGGCATA CGCCCTCAAT AATCTGGCGT GAATTAAGTC
CCAGGCTTTC AGCATAACTA

2341 TCAAGTTCAT TGAAATACGC TACACGCATC GCCAGATAAG
TGTTCGCAAA AAGTTTAATC

2401 GCCTCAGCCT CGGTTGAGTC AGTAAACAAT GTTGGTATGT
CTTGCTTAAT GGCGCCTTCC

2461 TGTAATAACG CAGCAAAACG TTTAGCGCGT TCAGACTGCT
CGCCAATCAC AATGCGTGAT

2521 GGGTGTAAGT TATCATAAAG TGCTTTACCT TCACGCAAAA
ACTCAGGCGA AAAGATCACA

2581 TTTTCAATAC CAAAACGTTC TTTAATGGAC TCTGTAAAAC CAACAGGGAT
AGTTGATTTT

2641 ATAATCATTA CCGCGTTGGG ATTAATTTCT GTCACATCAC
GAATGACCGC TTCCACGCTT

2701 GAGGTATTAA AATAATTTGT TTTCGGATCA TAATCGGTAG
GTGTGGCAAT AATAACGTAA

2761 TCGGCATTTT TATACGCGTC ATACTTATCT GTCGTAGCGC GG A A ATTG
AG ATCTTTAGTC

2821 GCCAGATACT CTTCAATCTC CTTATCAACA AGCGGTGACT
GCCTCTTGTT AAGCATGTCC

2881 ACTTTGGCCT GAACGATATC CAGTGCAACC ACTTCGTGGT
TTTGCGCAAT CAGAATACCA

2941 TTTGAAAGAC CAACATAACC TGTTCCTGAA ATTGTTATTT TCATTAGCTC
TGACTTCTTC

3001 CGGTTAAACA TTTAGAGTGG TCATTAATCC ACCACCACAG
CTTCATCTAC TGCGGGATTT

3061 TTAACGCTGA TGTAGATGAA CTGTCAAGGC AGCGATCCTG
CTGTGCGGCG CTGTATTATA

3121 TCGCGTTTTT AACTATAAAT TATAAAAAAA GGCCCTAACC TGCCGCTTTG
TATAATAAAA

3181 AAGCCCGGAG GGTTTCTCCG GGCCTTGCTT TGATTAATTG
ATTTAAATCA GATTAATCCA

3241 GCCATTCGGT ATGGAACACA CCTTCTTTAT CAATGCGCTT
ATAAGTATGC GCACCGAAAT

3301 AGTCACGCTG TGCCTGGATC AGGTTCGCAG GCAGAACAGC

GGCGCGGTAG CTGTCGTAAT

3361 AGGCAACCGC AGCGGCGAAG GTCGGCACCG GGATACCGTT

CTGTACTGCG TAAGCGACGA

3421 CATCGCGCAG CGCCTGCTGG TAGTCATCGG CAATTTGCTT
GAAGTAAGGA GCCAGCAACA

3481 GGTTAGCGAT CTGCGGATTT TCGGCATAAG CATCGGTGAT
TTTCTGCAGG AACTGCGCAC

3541 GGATGATGCA GCCAGCACGG AAAATCTTCG CGATTTCACC
GTAGTTCAGA TCCCAGTTGT

3601 ACTCTTCAGA CGCAGCGCGT AGCTGAGAGA AGCCCTGAGC

GTAAGAAACG ATTTTGCCCA

3661 GATACAGCGC ACGGCGAACT TTTTCGATGA ACTCAGCATT
GTCGCCAGCT GGCTGCGCTT

3721 GCGGGCCAGA GAGAACTTTA GATGCGGCAA CACGCTGCTC

TTTCAGAGAA GAGATATAAC

3781 GTGCAAACAC AGACTCGGTA ATCAGCGACA GCGGTTCGCC

GAGATCCAGC GCGCTCTGGC

3841 TGGTCCATTT GCCCGTACCT TTGTTTGCTG CTTCATCCAG
AATCACATCA ACCAGGTAGT

3901 TACCCTCTTC ATCTTTTTTG GTGAAGATAT CTTTGGTGAT GTCGATCAGG
TAGCTGCTCA

3961 GTTCACCGTT ATTCCACTCG GTAAAGGTCT GCGCCAGTTC
TTCGTTGGTG AGGTTCAAGC

4021 CACCTTTAAG CAGAGAATAG GCTTCAGCAA TCAGCTGCAT
ATCACCGTAT TCAATACCGT

4081 TGTGAACCAT CTTCACATAA TGACCTGCAC CATCGGCACC
AATATAGGTA ACGCACGGTT

4141 CGCCGTCTTC AGCCACAGCG GCGATTTTGG TCAGGATCGG
CGCAATCAGT TCATAAGCTT

4201 CTTTCTGCCC ACCAGGCATA ATGGAAGGAC CTTTCAGCGC
ACCTTCTTCA CCACCGGAAA

4261 CACCGGTACC GATAAAGTTA AAGCCTTCTG CAGAAAGCTC
ACGGTTACGA CGAATGGTGT

4321 CATGGAAGAA GGTGTTACCA CCATCAATGA TGATGTCACC
TTTATCGAGG TATGGCTTGA

4381 GGGAATCAAT AGCAGCATCC GTGCCAGCAC CTGCTTTCAC
CATTAACAGG ATGCGACGAG

4441 GCGTTTCCAG AGATTCAACA AATTCTTTCA CCGTATAGTA
AGGAACCAGT TTCTTGCCTG

4501 GATTTTCGGT AATCACTTCT TCGGTCTTTT CACGGGAACG
GTTGAAAATA GAGACGGTAT

4561 AACC ACGGCT TTCGATATTG AGCGCAAGGT TGCGCCCCAT
CACTGCCATA CCGACGACGC

4621 CGATCTGTTG CTTTGACATT GTTTACTCCT GTCAGGATAC
CGCTGGGTGG TATGCGGGTT

4681 ATGCTTAATT ATAGAATATG CCTAATAAAA ATAAATCCAT AACACTTAAT
CAGAAAATTA

4741 TTATTATCGA TTCCTAACGA TTGAATACAT CAGCTCCTTT AATTTAGATG
GCATTATACG

4801 AAAAAATGTT CTCAACATAG CATTACTTAT TAATTCATTT TTTCGAATAA
AACCAATTTT

4861 ATATTGATAA TACAATACTT TATACTCGTA CAATAAATAT GACAATCCAC
GTCGAGCCAT

4921 AAGATTACGA CCAGTTCGCA TTTTTAATAA AATATCTGGA AGATTTGCAA
ATCTTGCATT

4981 ATGTACAATT AATAGGCTCC ACAATGCAAA ATCTTGAGAT TTTCTGAATG
GAGGATAACC

5041 ACCAACAGCT AATACTGTAT TCTTTCTAAA AATTACAGAA GGATGGCTAA
CTGCGCTTCG

5101 TTTCCTCGCG AATTTAACTA TTTCTCTATG TTCGAGAGGC ACTTTGCGTG
TTGAAATAAA

5161 CTCCTCAGTA ACAGTTTCAA TTTCATCAAT AAAACTGCCA CATACATCTA
TTTCTGAATT

5221 ATTAATCATA AAAGAAATTT GTTTCTCAAA CCGATGAGGC AAAGAAATAT CATCAGCATC

5281 CATTCTTGCC ACTAACTCAT TCCTACAAGC CTTTAATCCT TCATTTAAGG CATTAGCCAA

5341 TCCAACATTT CTAGGTAAAG GTACAAATGT TACTATTTTA TTGCCAACAT CATCAATGAA

5401 TGAATTTATA ATATCGATGT GTGTTTGATG GAGTTCTCCA TCTGCAACAA TTACTATTTG

5461 ATCTGGCTTA AGTGTTTGAT CGTGAAAAAT AGAGCGTAGA GCCACCTCAA AAAATTGCGG

5521 TAGATCATTT TTATAAATGC TAATTAAAAC TGAGAATTTT TCTAATCTAT GATTCATTTC

5581 ATTTTACCAC TTCGACCCAT TAAACCGTCA TTAATGCCTT TTAAAAAAAA ATATAACCTT

5641 TTATTACCAT TTGGAAGGAA AATAGGATAT AAAAAAACCT TTCCAATTAA TTTAACCAGA

5701 CTAGAAATTT TCCAGTAGAT GGGTACATAA TTTTTATTTA ATAAAAGAAA GATATTTCGA

5761 GTAGCATAAT AATGACGAAA TGGGCTTGGC AAACCGACAG

AAAGAATATT TAAGATCTTA

5821 AATCGCCCAT CTCCAAGTCT ATGTGCAAGT AACGCATTTT TATTCCTAAT TACTTTAAAC

5881 CCAGCAGCTC TTAATCTCCA ACAATATTCA TGGTCTACCG CATCGATAAA AAGCTCATCT

5941 TTCATTCCTC CAACAATCAA CCAACTATTT TTTGGTATTA GACTGCCAGA ACTTAATGTA

6001 CTATCTACCT CATAATAAAC TTCTGTAAGT GGTTTCCCTT TTTTTACCCT TGCTTTATTT

6061 AATTCACCAG TTACTTTATC AAAATCTTGT GAACCAACTA AACCAACATT GACATTTTGT

6121 TTAAGCAATT TTTTGTAACA AGTAAGTAAC TGCTCTACCA TCTTAGGATC AGGAATACTA

6181 TCCTGATCCA TTTGCAATAT AAAATCAGCG CCATTTTCAA AAGCCCATTT CATTCCTATA

6241 CTTTGGGCTT CTGCTATGCC TAAATTATCA TTGAAATTGA ATATTTTTAC ATCGCCTGAA

6301 GAATTTTCAG CATATTTATA ACCATTTGTA GAGTTATTGC AAACGACAAC TTTAGTAACT

6361 TGTCTCAACA ATAATTCAAC CGCATTTTTT AAATCATTAT GTTCTGGGTT GTAAGCAACC

6421 AAAACGGCAT ATACAGTGTC CATCTTCACC TTAAAACCTT CATTTAGCTT TCATCTTTTT

6481 TAGAACATTA CTTAATGTCA CTAATACAAT TATTACAGCA ACATGGTTAG AGTCTAAAAT

6541 ATAAGGATTA GTAATTGCAT AAGAAACATA TAGAAAATAT

AGCACACACA ACTCACTGTA

6601 TTTTATGATT TTAATCGTGA GAAGGAGATT A ATT A ATA AA AACAAAGTAA ATAAAATAAC

6661 GCCAAGTTGA TTTAAAAAAT AAACTGACTG CAATTCATAA TATATATATG CACTATAATC

6721 ACGGATAGGA GTTTGAATTT TGATGACATT ACCCAAACCA GAACCTATAA CAAAATTTGA

6781 TACAGACTCT GTAAGATCAT TAATTAATAC AGTAAACTGA TCCCATCTAA CTCCTAAAGA

6841 AGAATCAGCT CCATTTGATT TCATGATTAT CAACTCAATT GAATATGTAA TAAAAAAAGG

6901 GAGAATCACA GTAAGAAAAA CCCCAAAAAT AATTTTCCTT AATTTAGCGT ATCGTGAGTT

6961 AGATTTAGAA CATAGTATAA TATACATAAA AAACAAGCAT

ATCGAAACAA AATATGCAAA

7021 ATTACCAGCC ACTATAGTAC CTATAGCCAG AATAACGGTT ATTGTATTTT TGAATCGATA

7081 ATAGAAATAA TCTTTTATGA CTATATGCAA CATAAAGGCA AATGGAATGA GAGCATTTCC

7141 TTTAATTTGA ACTCTATAGA AACCACTTCC ATATGTATAA ACATCACCAT AATCATTCTC

7201 CAAAAAATAA TGTCTTAGTG CTGAATAATC ACCAATACCA TATGTTTTTG TCATATAAAT

7261 ACTAATGATG GATATAATAA CCGCCTGTAA TACCATTAAA TATAAAAATA TTTTAACAAT

7321 CGAGATGGTT CCATAAGAGC AGAAATAAGC ACATAATATA

AATAATATGA TAATATAAAA

7381 CCTAATTATT ATCGCTATAT CGTTACCCTT GATATAGGAA TAAATAAAAT TTATAAAAAG

7441 AGCTAATAGA AATATTAAAA TAACAGGATA GTGATATATT CCGTTTGCAA TTTTCTTTGT

7501 AAATGACATG ATACAAAGAC ATAAAAACCC CTCCATAATC CAACTATATT GAATAAATGG

7561 AAAGCTACGT GTAAGGAAAA ATATAAACCC AAAAAACAAA

AGAACACTTA AACTTTTGTC

7621 TTTTGAGTTA TAAAAATCAG AAGTCATGTT TGCACTCTAA TTAGATGGGC TTGAGGAAGT

7681 AATCCCTAAA ATCAATTCGC TATTAATATT TCGTATCAAT TAATAATAAT ATCAAAAAAT

7741 CTAACGATGT TCTTACAGAC CATGCTATTG CGGCTCCAAC AATTCCCCAA TGATAAATAA

7801 AAATATATAA TATGCATAAA TATGGGATAA CTTCGAGCAA

ATGAATAATA GCTGTAATTT

7861 TTGATCTTCC ACTAGCCTGA ACTGAAACAA ATGGGATTTG TGCAATGCAA TTAAAAAAGA

7921 AACCTATTGC AAGAATTTTT AATACTATAC CTGGCGTCCC ATGATATGTA GGTCCCATCC

7981 AAGCGGACAT TATAAAATCT GATAAAATAA TTATCAACAT TACAATTGGA AGTATACCAA

8041 TAACCATTAT AAAATATGAT AATATTTTAG TTTGCTTTAC CGATTGCAAT TCTGAACTTA

8101 ATCTTGGAAA AATAGCTCTG GACAACGCAC TTGGTAATAT
CGTTAAGCGT TGTATACCTT

8161 CAGACGGAGC AGTATAAAAA GAAACTTTAT CAGCCCCCAC

AATGTGTGAA AGAATAAAAC

8221 GATCCATATA TGTCATAATA GGGCTAATAA TATTGCTAAC TGTTATCCAG
CTTCCAAAGC

8281 CGATTAATCT TTTAACTGTT ACAATTTTTA CAGACAGCCC AGATGATATT
ATTAGTTTTC

8341 GACTAAATAT AAAGGTCACT ATAAGTGATA AGACTCTTGC CATAACTAAA
CCATATATAG

8401 CACTTAGTAA TCCTCCATGA AAA AA AC AG A AAATCACTGG
TAATCCAGCC ACAAAAGAGT

8461 TGTTAATTGA TTTTATTAAA TTTACTTTTC TGAACTTTTC CATCCCCTCA
AAAATCCCCA

8521 ACCAGACTTG GTTTAACAAG TATAAGGGTA TGGTAGCTGA
AATAATATAT ATTGCTTTGA

8581 CAGATTCTAC AACATGATTC GCGTTAATGT TTAATAATTT AACAATTACA
TTGCTACTCA

8641 AAAATAGTAC ACTACCGCCA ATCAAGCCCA ATATAGTTAG
AATTACCGTT GAAGTTGAAA

8701 TGATCGCTCT TAATTCTTTA TGAACATTTT TATATATTGA TACTTCTCTT
ATAACAGCTC

8761 TGGTCAATCC AGCATCAAAA ATACTTGCAT ATCCAACTAA
GGCAATAGCT AACGTAAAAA

8821 GGCCAAATTG CTCGGTCCCT AGAATTCTAG ACAGTATACC
TAACGCAGGA ATTGCTATTA

8881 ATGATGGTAT AATATACCCA CTTATATTCC ATAAAGTATT CTTTACAATA
CTCACAAAAA

8941 TAATTCCTTC ATGTTATGCA ATTCTTTAGC CCTTGCATCT TTAATCGATA
AAATATAATT

9001 ATTATGTTCT ATCGTCGGCC ATTTTATGCT CAGAATAGGA TCATTCCATA
CAATCCCTCT

9061 ATCACTATCA GGATGATAAT AGTTCGTCGT TTTATATAAA AATTCCGCAG
TCTCGCTCAG

9121 CACCAAAAAA CCATGTGCAA ATCCCTCAGG GATCCACAAT
TGCCGCTTAT TCTCAGCAGA

9181 TAAATTCACC CCAACCCATT TACCAAAGGT AGGCGACGAT
TTACGAATAT CAACAGCTAC

9241 ATCAAAAACC TCACCAACAA CGCAACGTAC CAGTTTCCCT
TGCGCATAAG GTTCTAACTG

9301 ATAATGCAGC CCGCGTAAAA CACCTTTACT AGACTTCGAA
TGGTTATCCT GAACAAATTC

9361 AACCTTACGT CCTACAGCTT CTTCGAAAAC TTTCTGATTA AAGCTTTCCA
TAAAGAAACC

9421 ACGCTCATCA CCAAAAACTT TCGGCTCGAA AATTAACACA
TCAGGAATTT CTGTTTTAAT

9481 TACGTTCATT TTATTAATAA CCTTTAATCA TTTTCAGCAG ATACTGTCCA
TAAGCATTTT

9541 TTTTCAGCGG CTCCGCTAAT GCTTTCACCT GTTCAGCATC
AATAAACCCT TTACGGTAAG

9601 CAATTTCTTC TGGGCAGGAA ACCTTTAGTC CCTGGCGCTC
TTCAATGGTG GCAATGAAGT

9661 TGCTTGCTTC AATAAGACTC TGATGTGTCC CCGTATCCAG
CCATGCATAA CCACGCCCCA

9721 TCATGGCAAC GGATAAACGC CCCTGTTCCA TATAAATACG
GTTAATATCG GTAATTTCCA

9781 GTTCACCACG GGCAGAAGGC TTAAGGTTTT TCGCCATTTC
GACAACGTCG TTATCATAGA

9841 AATAAAGCCC GGTTACCGCA TAATTACTTT TTGGTTGTAG CGGTTTTTCT
TCCAGGCTTA

9901 TTGCCGTACC GTTTTTATCA AACTCAACGA CGCCGTAGCG
TTCAGGATCA TTAACGTGAT

9961 AGGCAAATAC CGTTGCACCA CTTTCTTTGT TAACAGCGAC
ATCCATTAAC TTCGGCAGAT

10021 CATGACCGTA GAAGATATTA TCACCAAGAA CCAAAGCACA

ATCATCACCA CCGATAAACT

10081 CTTCACCGAT AATAAACGCC TGCGCAAGCC CATCTGGAGT
CGGTTGCACT TTGTACTGAA

10141 GATTTAGCCC CCACTGGCTA CCGTCACCTA GCAGTTGTTG

AAAACGAGGA GTATCCTGTG

10201 GCGTACTAAT AATCAGAATA TCGCGAATAC CCGCCAACAT

CAGTGTAGAG AGCGGGTAAT

10261 AGATCATCGG CTTATCATAA ATAGGTAATA GCTGTTTACT
GACAGCCATA GTCACAGGAT

10321 AAAGACGTGT ACCAGAACCA CCCGCTAAAA TAATACCTTT
ACGCGTTTTC ATTTCATCAT

10381 TCCTTTTAAT TCATCTTGCT CCACCATCAC GAACAAGATG
CAAAAACTAT TAAATTGCTG

10441 TAGTCGTAAT TAATTCGTTG AGCATTCGTT TCACACCAAC
CTGCCAGTCA GGCAAGACAA

10501 GCGCAAAGTT CTGCTGAAAT TTTTCTGTAT TAAGGCGAGA
GTTATGTGGA CGACGAGCTG

10561 GTGTAGGATA GGCTGTTGTT GGTACTGCGT TGAGCTTGTT
GAGTGCAAGG GGAATACCTG

10621 CTTTGCGCGC CTCTTCAAAA ACCAGCGCAG CATAATCGTG
CCAGGTTGTG GTACCACTGG

10681 CTACCAGATG GTACAAACCT GCGACTTCCG GTTTATTCAG
TGCCACACGA ATAGCATGTG

10741 CCGTACAATC AGCCAGCAGC TCAGCACCTG TTGGCGCACC

AAATTGATCA TTTATCACAG

10801 CCAGTTCTTC GCGCTCTTTT GCCAGACGCA ACATCGTTTT
GGCGAAGTTA TTTCCTTTAG

10861 CTGCGTATAC CCAGCTGGTA CGGAAAATAA GATGCTTCGC
GCAATGTTCC TGTAACGCTT

10921 TTTCTCCGGC TAACTTGGTT TCACCGTAAA CATTTAGCGG
TGCGGTTGCA TCCGTCTCCA

10981 GCCATGGCGT GTCGCCATTT CCAGGGAATA CGTAGTCAGT

TGAGTAATGA ATTACCCAAG

11041 CCCCAACTTC ATTAGCCTCT TTTGCAATTG ATTCAACACT
AGTCGCATTG AGTAATTGTG

11101 CAAATTCGGG TTCTGACTCA GCCTTATCTA CTGCGGTGTG
AGCCGCAGCA TTAACAATAA

11161 CATCAGGTCG AATTCTTTTG ACTGTTTCAG CTACACCTTC AGG ATT
ACTA AAATCACCAC

11221 AATAATCAGT GGAGTGAACA TCAAGAGCAA TCAAATTACC

CAAAGGTGCC AGAGCACGCT

11281 GTAGTTCCCA ACCTACCTGC CCTGTTTTGC CGAAAAGGAG
GATATTCATT ACTGGCGGCC

11341 CTCATAGTTC TGTTCAATCC ACGATTGATA AGCACCACTT
TTCACATTAT CAACCCATTT

11401 TGTATTGGAC AGGTACCATT CCAATGTCTT CCGAATCCCG
CTCTCAAACG TTTCCTGCGG

11461 TTTCCAGCCC AATTCGCGGC TAATCTTCTC TGCATCAATC
GCATAACGGC GATCGTGTCC

11521 CGGGCGATCG GCAACATAAG TAATTTGCTC GCGGTAAGAT
TTCTCTTTCG GTACAATCTC

11581 ATCCAGCAAA TCACAAATAG TGAGCACTAC ATCGATGTTT
TTCTTTTCGT TGTGTCCACC

11641 AATGTTATAA GTTTCACCCG CTTTACCTTC GGTTACGACG
GTATATAACG CACGCGCATG

11701 ATCTTCAACA TACAGCCAGT CACGAATTTG ATCCCCTTTG
CCATAAATAG GTAATGCCTT

11761 ACCTTCCAGA GCATTCAGAA TAACCAATGG AATCAATTTT
TCCGGGAAAT GATAAGGACC

11821 ATAATTATTA GAGCAATTAG TCACAATGGT TGGTAAACCA
TAGGTACGTT TCCACGCGCG

11881 GACTAAATGA TCGCTGGATG CTTTTGAAGC GGAATAAGGG

CTGCTTGGCG CGTAAGCTGT

11941 TGTCTCTGTA AATAAGGGTA ATTCTTCTGT ATTATTTACC
TCGTCAGGAT GAGGCAAATC

12001 ACCATAGACT TCGTCAGTAG AAATATGATG AAAACGGAAT
CTAGTTTTCT TGTCGCTATC

12061 AAG AGC AG AC C AATA ATTGC GAGCGGCTTC C A AAAGG AC A

TATGTACCAA CAATATTGGT

12121 TTCAATAAAT GCCGCAGGAC CTGTAATTGA ACGGTCAACA
TGGCTTTCAG CAGCCAGGTG

12181 CATCACTGCA TCTGGCTGAT GCTGAGCAAA AATCCGTGCC
ATTGCAGCTG CATCGCAAAT

12241 ATCCGCATGT TCAAAAACAT AGCGTTCAGA ATCAGAAACA

TCAGCAAGTG ATTCCAGGTT

12301 TCCGGCGTAC GTTAATTTAT CGACATTAAC AACACTATCC
TGCGTATTAT TTATAATGTG

12361 ACGAACTACA GCAAAACCAA TAAATCCTGC GCCACCAGTA

ACAAGTATTT TCACCTAATT

12421 TATTCCATAT TGCTTCAGAG CATGCTGTGA AATAAGCGGC
TCTCAGTTTG ATTAATAGAA

12481 GTATTAATGC ACGCTACCGC CCCTGGCTTT ACAGCTACCA
GAGCACTGCA TGCATGCCTA

12541 CGATGTGACG AGCGTTACCC ACTCGCGCTA AACCCGAAAA

ATTCAAAAGC TAATTGTCTT

12601 ACCAATCCGC TCTGGAAACA AGGAAAATCC TGGAAAACTT

TGACTAAAAT CCTATTGCTA

12661 ACTCGTTGTT ATTCTGATTG TTTATATAAA ACAACGGCAG
GAATATTCGC AACAAATTAC

12721 TTTCACCACG AATCTTCACT GCCGTTATAA TTTTCTTATC
AACCGTTACA TCCGGTCAGA

12781 TTTTCATTAT TCGCTTAACA GCTTCTCAAT ACCTTTACGG
AACTTCGCCC CTTCTTTCAG

12841 GTTGCGCAGC CCATACTTCA CAAACGCCTG CATATAGCCC ATTTTTTTAC CGCAGTCGTA

12901 GCTGTCGCCG GTCATCAGCA TTGCATCAAC GGACTGTTTT TTCGCCAGCT CGGCAATGGC

12961 ATCAGTCAGC TGAATACGTC CCCATGCACC AGGCTGAGTA CGTTCAAGTT CCGGCCAAAT

13021 ATCGGCAGAA AGCACATAGC GACCAACGGC CATGATGTCT

GAGTCCAGCG TCTGCGGCTG

13081 ATCCGGTTTT TCGATAAATT CAACAATGCG GCTGACTTTA CCTTCGCGAT CCAGCGGTTC

13141 TTTGGTCTGG ATGACGGAGT ATTCAGAGAG GTCACCCGGC ATACGTTTTG CCAGCACCTG

13201 GCTACGGCCC GTTTCATTGA AGCGCGCAAT CATGGCAGCA

AGGTTGTAGC GTAGCGGGTC

13261 GGCGCTGGCG TCGTCGATCA CAACGTCTGG CAGCACCACG

ACAAATGGAT TGTCACCAAT

13321 GGCGGGTCGT GCACACAAAA TGGAGTGACC TAAACCTAAA

GGTTCGCCCT GACGCACGTT

13381 CATAATAGTC ACGCCCGGCG GGCAGATAGA TTGCACTTCC

GCCAGTAGTT GACGCTTCAC

13441 GCGCTGCTCA AGGAGAGATT CTAATTCATA AGAGGTGTCG AAGTGGTTTT CGACCGCGTT

13501 CTTGGACGCA TGAGTTACCA GGAGGATTTC TTTGATCCCT GCAGCCACAA TCTCGTCAAC

13561 AATGTACTGA ATCATTGGCT TGTCGACGAT CGGTAGCATC TCTTTGGGTA TCGCCTTAGT

13621 GGCAGGCAAC ATATGCATCC CAAGACCCGC TACCGGTATA ACTGCTTTTA AATTCGTCAT

13681 TATTTTCCTA CCTCTAAGGG GCTGATAGTG CGTAAATTAT TGTCATAGGT TAGCCAAACG

13741 GTATGGCTAT ATACCAAGCA TAACTTTGAT TAAACCTTAC
GATAACACTA CACACCATCA

13801 GCATCTGGGT TACTCGGATT ACTCGGAAAT CCACATACTG
ATAATTTAAT CAGTACCTCT

13861 TTCCGAATAA TCGTAGTCCA ACCTGGTCCT TTTTTCTCTG
ACTCGTCTGC ATTACTCAGA

13921 AACAAACGTT ATGTCGTCTT TTTTGGCATG GACGAATTCA
TACTGCAGAG TTCGATCCAG

13981 ACCTTGCGAC AGCGTATACG GTGCAACAAA ACCTGAAGAA

TGCACTTTCG TTGCGTCAAA

14041 CTGTGTTGTT GCGCAGAATT TTTTCACGCG CACAGAGCTG
ACAGCGTATT TTTTGCCCGT

14101 AATTTTGCTC AGGATATCAA AGCAATATCC ACCCAGCATT
CCTAGTGGGT AAGGCAAGTG

14161 CATAGAAGGG ATCTTTTTGT TCAGGCTTTG TTCAACTTCA
GCAACCAACT GGTTCATGTT

14221 CAGGTCTGGC TTATCAACAT AGTTATAAAC CTCATAACCT
GCGGCAACAT TCTTCAGTTT

14281 GTACTTGATA AACTCAACAA TGTTTCCAAC ATAAGCCATG
GACTTATAGT TAGTCCCTGC

14341 GCCCACCATC ATAAACTTGC CGCCAGCGAT CTGTTTCAGC
AAGTTATAGA CGTTACCGCG

14401 GTTGCGTTCA CCGAAGATAA CGGTAGGACG GATGATGGTT

AATGAACGTT CTGTTGGTGC

14461 TTTGTTATAC CATTCACGCA GCACTTCCTC TGCCTGCCAC
TTACTTTTGC CGTAGTGGTT

14521 GAAAGGGTCG TGTGGATGGT TTTCGTCAGG GTTGTGTTTG
TTCAAACCAT AAACAGCAAC

14581 GGAACTGGTA AAGATGATAT TTTTAACGCC ATTTTTTTCC
ATGGCCGCCA GCACATTGCG

14641 GGTACCCTGA ACGTTGACAT CATAATAGAG AGAAGTAGGG

CTGACGTCAT CGCGGTGTTC

14701 CGCTGCCAGT AGTACAACAG TGTCAAAACC GGCTAACGCC

TGGTCGAGTG CCTGTTGATC

14761 ACGAACATCA CCAATCTGTG TGATTTCTGG ATAAAAGTGG
CTCTGCCGTT TGTCCAGGTT

14821 CTTGATATTA AAGTCAGCAA TTGCCGTTTC AAGTAGTCGG
GTTCCTACGA ATCCGGAAGC

14881 TCCTATGAGC AAAACGTTAT TGTTCATAAA TCACTTTAGT
CTGGTTGTTA CGTAAGAAAC

14941 ACAAGATAAA GATGAGTACC TTCCCTGAGT AGTCAATGCT

GCCCAGCCCC AGCTTTAACA

15001 GTTAGTGTGA GGATTATAAT CTTTTAGAAC ATTATATCCA GTAAGTTTAT
GAATGGTCGC

15061 AAATCTACTC TCTCCGTTCC GGCAATCTAA AGTTAATGCT
AGCGACGTCG TGGGATCCTC

15121 TAGAGTCGAC CTGCAGGCAT GCAAGCTTGA GTATTCTATA
GTCTCACCTA AATAGCTTGG

15181 CGTAATCATG GTCATAGCTG TTTCCTGTGT GAAATTGTTA
TCCGCTCACA ATTCCACACA

15241 ACATACGAGC CGGAAGCATA AAGTGTAAAG CCTGGGGTGC

CTAATGAGTG AGCTAACTCA

15301 CATTAATTGC GTTGCGCTCA CTGCCCGCTT TCCAGTCGGG
AAACCTGTCG TGCCAGCTGC

15361 ATTAATGAAT CGGCCAACGC GAACCCCTTG CGGCCGCCCG

GGCCGTCGAC CAATTCTCAT

15421 GTTTGACAGC TTATCATCGA ATTTCTGCCA TTCATCCGCT
TATTATCACT TATTCAGGCG

15481 TAGCAACCAG GCGTTTAAGG GCACCAATAA CTGCCTTAAA

AAAATTACGC CCCGCCCTGC

15541 CACTCATCGC AGTACTGTTG TAATTCATTA AGCATTCTGC
CGACATGGAA GCCATCACAA

15601 ACGGCATGAT GAACCTGAAT CGCCAGCGGC ATCAGCACCT

TGTCGCCTTG CGTATAATAT

15661 TTGCCCATGG TGAAAACGGG GGCGAAGAAG TTGTCCATAT

TGGCCACGTT TAAATCAAAA

15721 CTGGTGAAAC TCACCCAGGG ATTGGCTGAG ACGAAAAACA

TATTCTCAAT AAACCCTTTA

15781 GGGAAATAGG CCAGGTTTTC ACCGTAACAC GCCACATCTT

GCGAATATAT GTGTAGAAAC

15841 TGCCGGAAAT CGTCGTGGTA TTCACTCCAG AGCGATGAAA
ACGTTTCAGT TTGCTCATGG

15901 AAAACGGTGT AACAAGGGTG AACACTATCC CATATCACCA

GCTCACCGTC TTTCATTGCC

15961 ATACGAAATT CCGGATGAGC ATTCATCAGG CGGGCAAGAA

TGTGAATAAA GGCCGGATAA

16021 AACTTGTGCT TATTTTTCTT TACGGTCTTT AAAAAGGCCG
TAATATCCAG CTGAACGGTC

16081 TGGTTATAGG TACATTGAGC AACTGACTGA AATGCCTCAA
AATGTTCTTT ACGATGCCAT

16141 TGGGATATAT CAACGGTGGT ATATCCAGTG ATTTTTTTCT
CCATTTTAGC TTCCTTAGCT

16201 CCTGAAAATC TCGATAACTC AAAAAATACG CCCGGTAGTG
ATCTTATTTC ATTATGGTGA

16261 AAGTTGGAAC CTCTTACGTG CCGATCAACG TCTCATTTTC
GCCAAAAGTT GGCCCAGGGC

16321 TTCCCGGTAT CAACAGGGAC ACCAGGATTT ATTTATTCTG
CGAAGTGATC TTCCGTCACA

16381 GGTATTTATT CGCGATAAGC TCATGGAGCG GCGTAACCGT

CGCACAGGAA GGACAGAGAA

16441 AGCGCGGATC TGGGAAGTGA CGGACAGAAC GGTCAGGACC

TGGATTGGGG AGGCGGTTGC

16501 CGCCGCTGCT GCTGACGGTG TGACGTTCTC TGTTCCGGTC
ACACCACATA CGTTCCGCCA

16561 TTCCTATGCG ATGCACATGC TGTATGCCGG TATACCGCTG
AAAGTTCTGC AAAGCCTGAT

16621 GGGACATAAG TCCATCAGTT CAACGGAAGT CTACACGAAG

GTTTTTGCGC TGGATGTGGC

16681 TGCCCGGCAC CGGGTGCAGT TTGCGATGCC GGAGTCTGAT

GCGGTTGCGA TGCTGAAACA

16741 ATTATCCTGA GAATAAATGC CTTGGCCTTT ATATGGAAAT
GTGGAACTGA GTGGATATGC

16801 TGTTTTTGTC TGTTAAACAG AGAAGCTGGC TGTTATCCAC
TGAGAAGCGA ACGAAACAGT

16861 CGGGAAAATC TCCCATTATC GTAGAGATCC GCATTATTAA
TCTCAGGAGC CTGTGTAGCG

16921 TTTATAGGAA GTAGTGTTCT GTCATGATGC CTGCAAGCGG

TAACGAAAAC GATTTGAATA

16981 TGCCTTCAGG AACAATAGAA ATCTTCGTGC GGTGTTACGT

TGAAGTGGAG CGGATTATGT

17041 CAGCAATGGA CAGAACAACC TAATGAACAC AGAACCATGA

TGTGGTCTGT CCTTTTACAG

17101 CCAGTAGTGC TCGCCGCAGT CGAGCGACAG GGCGAAGCCC

TCGGCTGGTT GCCCTCGCCG

17161 CTGGGCTGGC GGCCGTCTAT GGCCCTGCAA ACGCGCCAGA

AACGCCGTCG AAGCCGTGTG

17221 CGAGACACCG CGGCCGGCCG CCGGCGTTGT GGATACCTCG

CGGAAAACTT GGCCCTCACT

17281 GACAGATGAG GGGCGGACGT TGACACTTGA GGGGCCGACT

CACCCGGCGC GGCGTTGACA

17341 GATGAGGGGC AGGCTCGATT TCGGCCGGCG ACGTGGAGCT

GGCCAGCCTC GCAAATCGGC

17401 GAAAACGCCT GATTTTACGC GAGTTTCCCA CAGATGATGT

GGACAAGCCT GGGGATAAGT

17461 GCCCTGCGGT ATTGACACTT GAGGGGCGCG ACTACTGACA

GATGAGGGGC GCGATCCTTG

17521 ACACTTGAGG GGCAGAGTGC TGACAGATGA GGGGCGCACC

TATTGACATT TGAGGGGCTG

17581 TCCACAGGCA GAAAATCCAG CATTTGCAAG GGTTTCCGCC
CGTTTTTCGG CCACCGCTAA

17641 CCTGTCTTTT AACCTGCTTT TAAACCAATA TTTATAAACC TTGTTTTTAA
CCAGGGCTGC

17701 GCCCTGTGCG CGTGACCGCG CACGCCGAAG GGGGGTGCCC

CCCCTTCTCG AACCCTCCCG

17761 GTCGAGTGAG CGAGGAAGCA CCAGGGAACA GCACTTATAT

ATTCTGCTTA CACACGATGC

17821 CTGAAAAAAC TTCCCTTGGG GTTATCCACT TATCCACGGG
GATATTTTTA TAATTATTTT

17881 TTTTATAGTT TTTAGATCTT CTTTTTTAGA GCGCCTTGTA GGCCTTTATC
CATGCTGGTT

17941 CTAGAGAAGG TGTTGTGACA AATTGCCCTT TCAGTGTGAC
AAATCACCCT CAAATGACAG

18001 TCCTGTCTGT GACAAATTGC CCTTAACCCT GTGACAAATT
GCCCTCAGAA GAAGCTGTTT

18061 TTTCACAAAG TTATCCCTGC TTATTGACTC TTTTTTATTT AGTGTGACAA
TCTAAAAACT

18121 TGTCACACTT CACATGGATC TGTCATGGCG GAAACAGCGG
TTATCAATCA CAAGAAACGT

18181 AAAAATAGCC CGCGAATCGT CCAGTCAAAC GACCTCACTG

AGGCGGCATA TAGTCTCTCC

18241 CGGGATCAAA AACGTATGCT GTATCTGTTC GTTGACCAGA

TCAGAAAATC TGATGGCACC

18301 CTACAGGAAC ATGACGGTAT CTGCGAGATC CATGTTGCTA

AATATGCTGA AATATTCGGA

18361 TTGACCTCTG CGGAAGCCAG TAAGGATATA CGGCAGGCAT

TGAAGAGTTT CGCGGGGAAG

18421 GAAGTGGTTT TTTATCGCCC TGAAGAGGAT GCCGGCGATG

AAAAAGGCTA TGAATCTTTT

18481 CCTTGGTTTA TCAAACGTGC GCACAGTCCA TCCAGAGGGC
TTTACAGTGT ACATATCAAC

18541 CCATATCTCA TTCCCTTCTT TATCGGGTTA CAGAACCGGT
TTACGCAGTT TCGGCTTAGT

18601 GAAACAAAAG AAATCACCAA TCCGTATGCC ATGCGTTTAT

ACGAATCCCT GTGTCAGTAT

18661 CGTAAGCCGG ATGGCTCAGG CATCGTCTCT CTGAAAATCG
ACTGGATCAT AGAGCGTTAC

18721 CAGCTGCCTC AAAGTTACCA GCGTATGCCT GACTTCCGCC
GCCGCTTCCT GCAGGTCTGT

18781 GTTAATGAGA TCAACAGCAG AACTCCAATG CGCCTCTCAT

ACATTGAGAA AAAGAAAGGC

18841 CGCCAGACGA CTCATATCGT ATTTTCCTTC CGCGATATCA
CTTCCATGAC GACAGGATAG

18901 TCTGAGGGTT ATCTGTCACA GATTTGAGGG TGGTTCGTCA
CATTTGTTCT GACCTACTGA

18961 GGGTAATTTG TCACAGTTTT GCTGTTTCCT TCAGCCTGCA
TGGATTTTCT CATACTTTTT

19021 GAACTGTAAT TTTTAAGGAA GCCAAATTTG AGGGCAGTTT GTCACAGTTG ATTTCCTTCT

19081 CTTTCCCTTC GTCATGTGAC CTGATATCGG GGGTTAGTTC GTCATCATTG ATGAGGGTTG

19141 ATTATCACAG TTTATTACTC TGAATTGGCT ATCCGCGTGT GTACCTCTAC CTGGAGTTTT

19201 TCCCACGGTG GATATTTCTT CTTGCGCTGA GCGTAAGAGC TATCTGACAG AACAGTTCTT

19261 CTTTGCTTCC TCGCCAGTTC GCTCGCTATG CTCGGTTACA CGGCTGCGGC GAGCGCTAGT

19321 GATAATAAGT GACTGAGGTA TGTGCTCTTC TTATCTCCTT TTGTAGTGTT GCTCTTATTT

19381 TAAACAACTT TGCGGTTTTT TGATGACTTT GCGATTTTGT TGTTGCTTTG CAGTAAATTG

19441 CAAGATTTAA TAAAAAAACG CAAAGCAATG ATTAAAGGAT

GTTCAGAATG AAACTCATGG

19501 AAACACTTAA CCAGTGCATA AACGCTGGTC ATGAAATGAC

GAAGGCTATC GCCATTGCAC

19561 AGTTTAATGA TGACAGCCCG GAAGCGAGGA AAATAACCCG

GCGCTGGAGA ATAGGTGAAG

19621 CAGCGGATTT AGTTGGGGTT TCTTCTCAGG CTATCAGAGA TGCCGAGAAA GCAGGGCGAC

19681 TACCGCACCC GGATATGGAA ATTCGAGGAC GGGTTGAGCA

ACGTGTTGGT TATACAATTG

19741 AACAAATTAA TCATATGCGT GATGTGTTTG GTACGCGATT GCGACGTGCT GAAGACGTAT

19801 TTCCACCGGT GATCGGGGTT GCTGCCCATA AAGGTGGCGT TTACAAAACC TCAGTTTCTG

19861 TTCATCTTGC TCAGGATCTG GCTCTGAAGG GGCTACGTGT TTTGCTCGTG GAAGGTAACG

19921 ACCCCCAGGG AACAGCCTCA ATGTATCACG GATGGGTACC

AGATCTTCAT ATTCATGCAG

19981 AAGACACTCT CCTGCCTTTC TATCTTGGGG AAAAGGACGA
TGTCACTTAT GCAATAAAGC

20041 CCACTTGCTG GCCGGGGCTT GACATTATTC CTTCCTGTCT
GGCTCTGCAC CGTATTGAAA

20101 CTGAGTTAAT GGGCAAATTT GATGAAGGTA AACTGCCCAC

CGATCCACAC CTGATGCTCC

20161 GACTGGCCAT TGAAACTGTT GCTCATGACT ATGATGTCAT
AGTTATTGAC AGCGCGCCTA

20221 ACCTGGGTAT CGGCACGATT AATGTCGTAT GTGCTGCTGA
TGTGCTGATT GTTCCCACGC

20281 CTGCTGAGTT GTTTGACTAC ACCTCCGCAC TGCAGTTTTT
CGATATGCTT CGTGATCTGC

20341 TCAAGAACGT TGATCTTAAA GGGTTCGAGC CTGATGTACG
TATTTTGCTT ACCAAATACA

20401 GCAATAGTAA TGGCTCTCAG TCCCCGTGGA TGGAGGAGCA

AATTCGGGAT GCCTGGGGAA

20461 GCATGGTTCT AAAAAATGTT GTACGTGAAA CGGATGAAGT

TGGTAAAGGT CAGATCCGGA

20521 TGAGAACTGT TTTTGAACAG GCCATTGATC AACGCTCTTC
AACTGGTGCC TGGAGAAATG

20581 CTCTTTCTAT TTGGGAACCT GTCTGCAATG AAATTTTCGA
TCGTCTGATT AAACCACGCT

20641 GGGAGATTAG ATAATGAAGC GTGCGCCTGT TATTCCAAAA

CATACGCTCA ATACTCAACC

20701 GGTTGAAGAT ACTTCGTTAT CGACACCAGC TGCCCCGATG
GTGGATTCGT TAATTGCGCG

20761 CGTAGGAGTA ATGGCTCGCG GTAATGCCAT TACTTTGCCT
GTATGTGGTC GGGATGTGAA

20821 GTTTACTCTT GAAGTGCTCC GGGGTGATAG TGTTGAGAAG
ACCTCTCGGG TATGGTCAGG

20881 TAATGAACGT GACCAGGAGC TGCTTACTGA GGACGCACTG

GATGATCTCA TCCCTTCTTT

20941 TCTACTGACT GGTCAACAGA CACCGGCGTT CGGTCGAAGA

GTATCTGGTG TCATAGAAAT

21001 TGCCGATGGG AGTCGCCGTC GTAAAGCTGC TGCACTTACC

GAAAGTGATT ATCGTGTTCT

21061 GGTTGGCGAG CTGGATGATG AGCAGATGGC TGCATTATCC

AGATTGGGTA ACGATTATCG

21 121 CCCAACAAGT GCTTATGAAC GTGGTCAGCG TTATGCAAGC

CGATTGCAGA ATGAATTTGC

21181 TGGAAATATT TCTGCGCTGG CTGATGCGGA AAATATTTCA
CGTAAGATTA TTACCCGCTG

21241 TATCAACACC GCCAAATTGC CTAAATCAGT TGTTGCTCTT
TTTTCTCACC CCGGTGAACT

21301 ATCTGCCCGG TCAGGTGATG CACTTCAAAA AGCCTTTACA

GATAAAGAGG AATTACTTAA

21361 GCAGCAGGCA TCTAACCTTC ATGAGCAGAA AAAAGCTGGG

GTGATATTTG AAGCTGAAGA

21421 AGTTATCACT CTTTTAACTT CTGTGCTTAA AACGTCATCT
GCATCAAGAA CTAGTTTAAG

21481 CTCACGACAT CAGTTTGCTC CTGGAGCGAC AGTATTGTAT
AAGGGCGATA AAATGGTGCT

21541 TAACCTGGAC AGGTCTCGTG TTCCAACTGA GTGTATAGAG

AAAATTGAGG CCATTCTTAA

21601 GGAACTTGAA AAGCCAGCAC CCTGATGCGA CCACGTTTTA
GTCTACGTTT ATCTGTCTTT

21661 ACTTAATGTC CTTTGTTACA GGCCAGAAAG CATAACTGGC
CTGAATATTC TCTCTGGGCC

21721 CACTGTTCCA CTTGTATCGT CGGTCTGATA ATCAGACTGG GACCACGGTC CCACTCGTAT

21781 CGTCGGTCTG ATTATT AGTC TGGGACCACG GTCCCACTCG TATCGTCGGT CTGATTATTA

21841 GTCTGGGACC ACGGTCCCAC TCGTATCGTC GGTCTGATAA TCAGACTGGG ACCACGGTCC

21901 CACTCGTATC GTCGGTCTGA TTATTAGTCT GGGACCATGG TCCCACTCGT ATCGTCGGTC

21961 TGATTATTAG TCTGGGACCA CGGTCCCACT CGTATCGTCG GTCTGATTAT TAGTCTGGAA

22021 CCACGGTCCC ACTCGTATCG TCGGTCTGAT TATTAGTCTG GGACCACGGT CCCACTCGTA

22081 TCGTCGGTCT G ATTATT AGT CTGGGACCAC GATCCCACTC GTGTTGTCGG TCTGATTATC

22141 GGTCTGGGAC CACGGTCCCA CTTGTATTGT CGATCAGACT ATCAGCGTGA GACTACGATT

22201 CCATCAATGC CTGTCAAGGG CAAGTATTGA CATGTCGTCG TAACCTGTAG AACGGAGTAA

22261 CCTCGGTGTG CGGTTGTATG CCTGCTGTGG ATTGCTGCTG TGTCCTGCTT ATCCACAACA

22321 TTTTGCGCAC GGTTATGTGG ACAAAATACC TGGTTACCCA GGCCGTGCCG GCACGTTAAC

22381 CGGGCTGCAT CCGATGCAAG TGTGTCGCTG TCGACGAGCT

CGCGAGCTCG GACATGAGGT

22441 TGCCCCGTAT TCAGTGTCGC TGATTTGTAT TGTCTGAAGT TGTTTTTACG TTAAGTTGAT

22501 GCAGATCAAT TAATACGATA CCTGCGTCAT AATTGATTAT TTGACGTGGT TTGATGGCCT

22561 CCACGCACGT TGTGATATGT AGATGATAAT CATTATCACT TTACGGGTCC TTTCCGGTGA

22621 TCCGACAGGT TACGGGGCGG CGACCTCGCG GGTTTTCGCT ATTTATGAAA ATTTTCCGGT

22681 TTAAGGCGTT TCCGTTCTTC TTCGTCATAA CTTAATGTTT TTATTTAAAA TACCCTCTGA

22741 AAAGAAAGGA AACGACAGGT GCTGAAAGCG AGCTTTTTGG

CCTCTGTCGT TTCCTTTCTC

22801 TGTTTTTGTC CGTGGAATGA ACAATGGAAG TCCGAGCTCA TCGCTAATAA CTTCGTATAG

22861 CATACATTAT ACGAAGTTAT ATTCGAT

REIVINDICACIONES

1. Un organismo procariota recombinante que es una *E. coli* que comprende ácidos nucleicos que codifican:

i) una epimerasa que sintetiza N-acetilgalactosamina sobre pirofosfato de undecaprenilo, en el que dicho ácido nucleico que codifica dicha epimerasa es al menos un 95 % idéntico a la SEQ ID NO: 1;
ii) glicosil transferasas que ensamblan un oligo o polisacárido en un vehículo lipídico;
iii) una oligosacaril transferasa de *Campylobacter jejuni*; y
iv) una proteína que comprende una o más de una secuencia de consenso introducida recombinantemente, D/E - X - N - Z - S/T, en la que X y Z son independientemente cualquier aminoácido natural excepto prolina

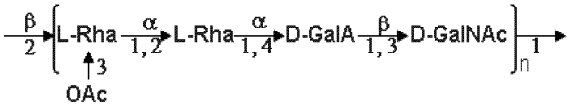
en el que el oligosacárido ensamblado por dichas glicosiltransferasas se puede transferir desde el vehículo lipídico a la proteína diana que tiene una o más secuencias de consenso optimizadas, por la oligosacaril transferasa.

2. El sistema biosintético procariota recombinante de la reivindicación 1, en el que dicha proteína es una exoproteína de *P. aeruginosa*.

3. El sistema biosintético procariota recombinante de una cualquiera de las reivindicaciones 1-2, en el que dicho oligo o polisacárido es de una bacteria Gram-negativa.

4. El sistema biosintético procariota recombinante de la reivindicación 3, en el que dicho oligo o polisacárido es de *Shigella flexneri*, preferentemente de *Shigella flexneri* 6; o de *E. coli* preferentemente *E. coli* 0157.

5. El sistema biosintético procariota recombinante de una cualquiera de las reivindicaciones 1 a 4, en el que dicho oligo o polisacárido comprende una estructura:



O (ii) la siguiente estructura: α-D-PerNAc-α-L-Fuc-(β-D-Glc-α-D-GalNAc.

6. Un metodo de producción de una proteína glicosilada unida a N que comprende el cultivo del organismo procariota recombinante de una cualquiera de las reivindicaciones 1-5 en condiciones adecuadas para la producción de proteínas y el aislamiento de la proteína N-glicosilada del cultivo.

7. Un procedimiento de producción de una proteína glicosilada unida a N, que comprende:

a.) introducir en un organismo huésped de *E. coli* ácidos nucleicos que codifican:

i.) una epimerasa que sintetiza N-acetilgalactosamina sobre pirofosfato de undecaprenilo, en el que dicho ácido nucleico que codifica dicha epimerasa es al menos un 95 % idéntico a la SEQ ID NO: 1;
ii.) glicosil transferasas que ensamblan un oligo o polisacárido en un vehículo lipídico;
iii.) una oligosacaril transferasa de *Campylobacter jejuni*; y
iv.) una proteína que comprende una o más de una secuencia de consenso introducida, D/E - X - N - Z - S/T, en la que X y Z pueden ser independientemente cualquier aminoácido natural excepto prolina; y

b.) cultivar dicho organismo huesped hasta que se produzca al menos una proteína N-glicosilada.

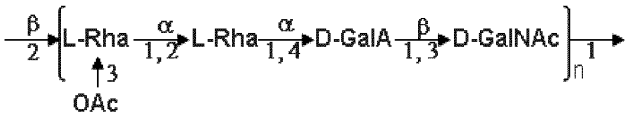
8. El procedimiento de la reivindicación 7, en el que dicha proteína es una exoproteína de *P. aeruginosa*.

9. El procedimiento de una cualquiera de las reivindicaciones 7 a 8, en el que dicho oligo o polisacarido es de una bacteria Gram-negativa.

10. El procedimiento de una cualquiera de las reivindicaciones 7 a 9, en el que dicho oligo o polisacárido es de *Shigella flexneri*, preferentemente de *Shigella flexneri* 6.

11. El procedimiento de una cualquiera de las reivindicaciones 7 a 10, en el que dicho oligo o polisacárido es de *E. coli*, preferentemente de *E. coli* 0157.

12. El procedimiento de una cualquiera de las reivindicaciones 7 a 11, en el que dicho oligo o polisacárido comprende (i) la siguiente estructura:



O (ii) la siguiente estructura: α-D-PerNAc-α-L-Fuc-(β-D-Glc-α-D-GalNAc.

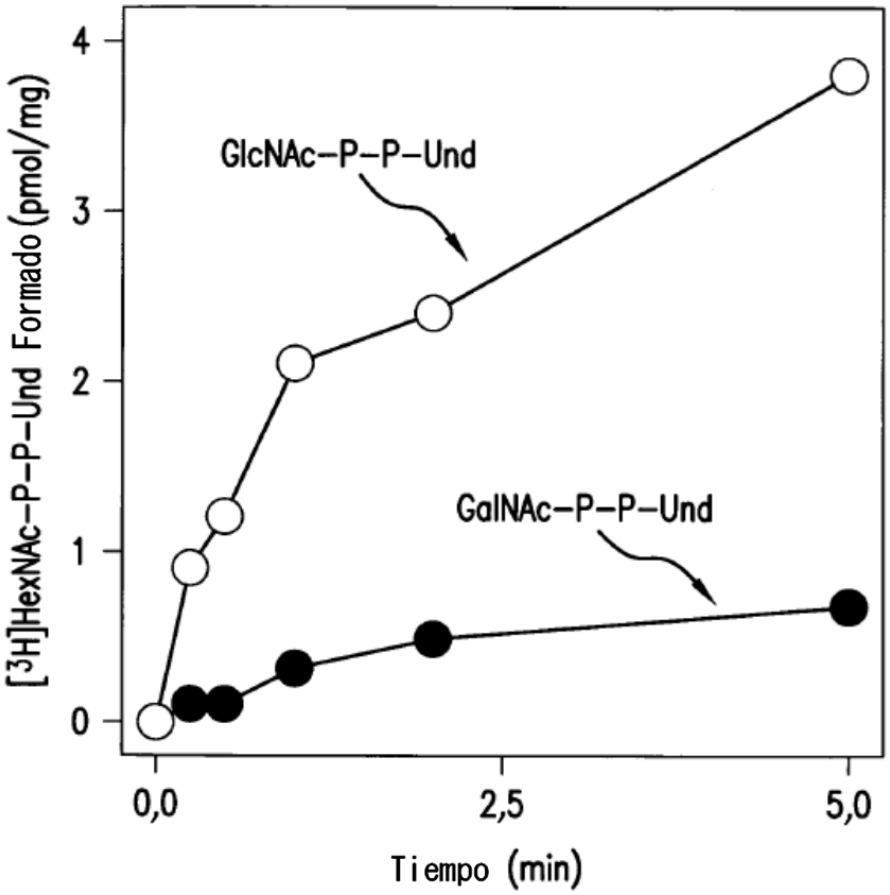

[3H]HexNAc-P-P-Und Formado (pmol/mg)
4
3
2
1
0
GlcNAc-P-P-Und
GalNAc-P-P-Und
0,0
2,5
5,0
Tiempo (min)

FIG. 1

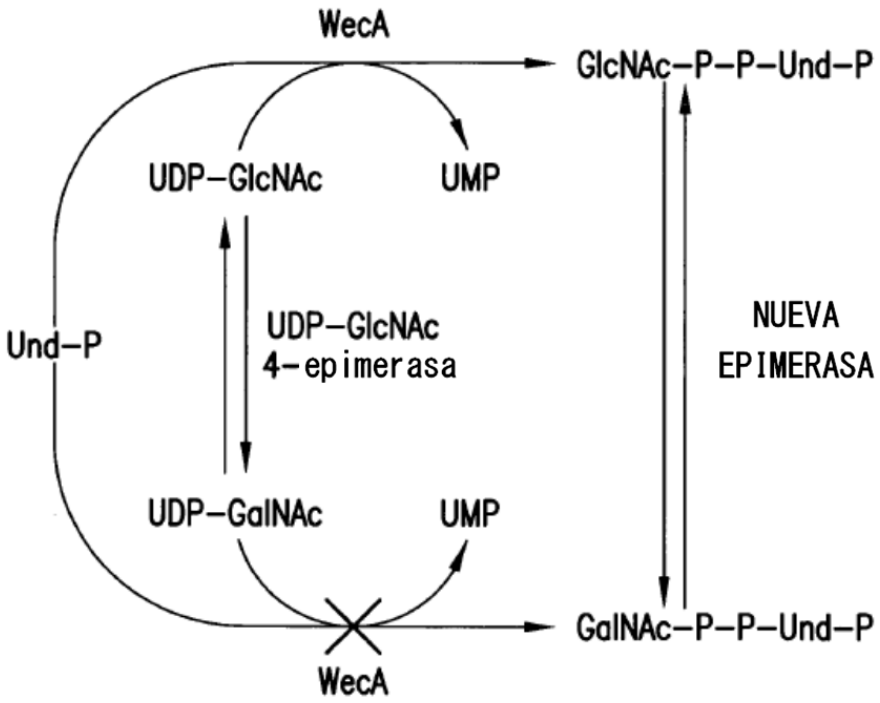

WecA
GlcNAc-P-P-Und-P
UDP-GlcNAc
UMP
Und-P
UDP-GlcNAc
4-epimerasa
NUEVA
EPIMERASA
UDP-GalNAc
UMP
GalNAc-P-P-Und-P
WecA

FIG.2

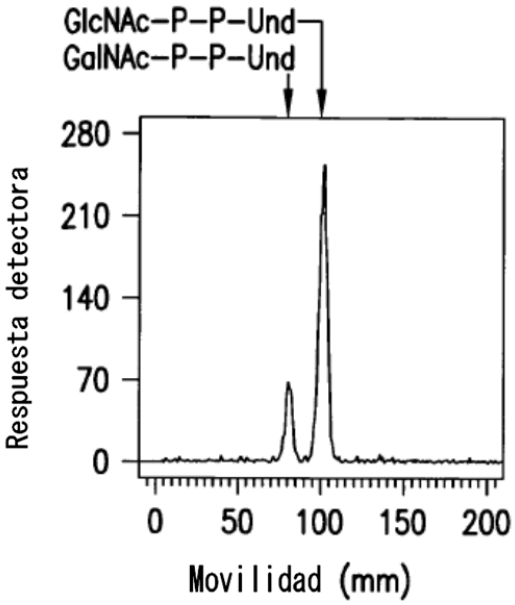


FIG.3A

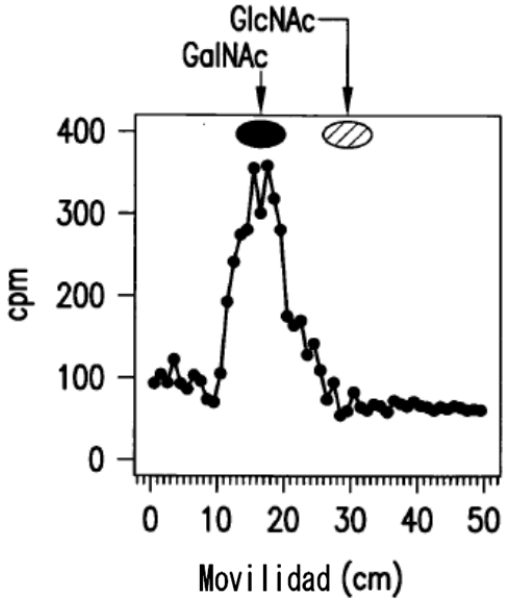


FIG.3C

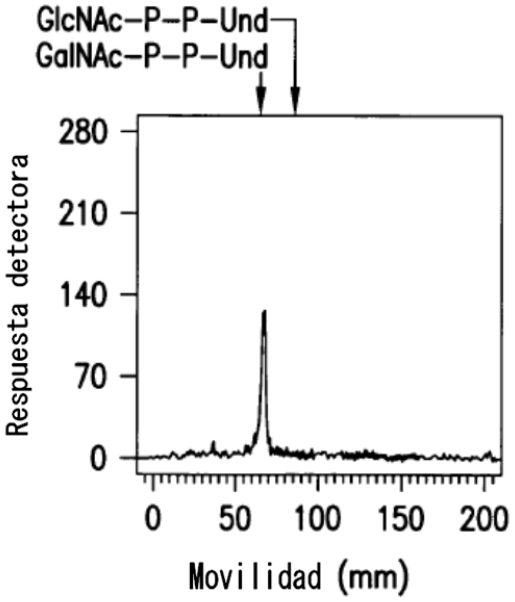


FIG.3B

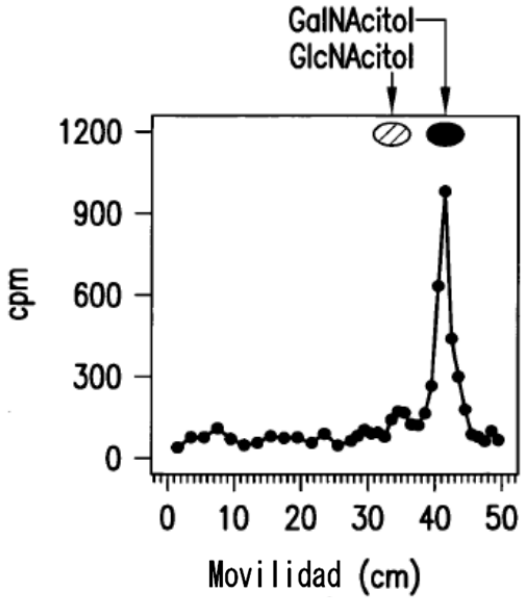


FIG.3D

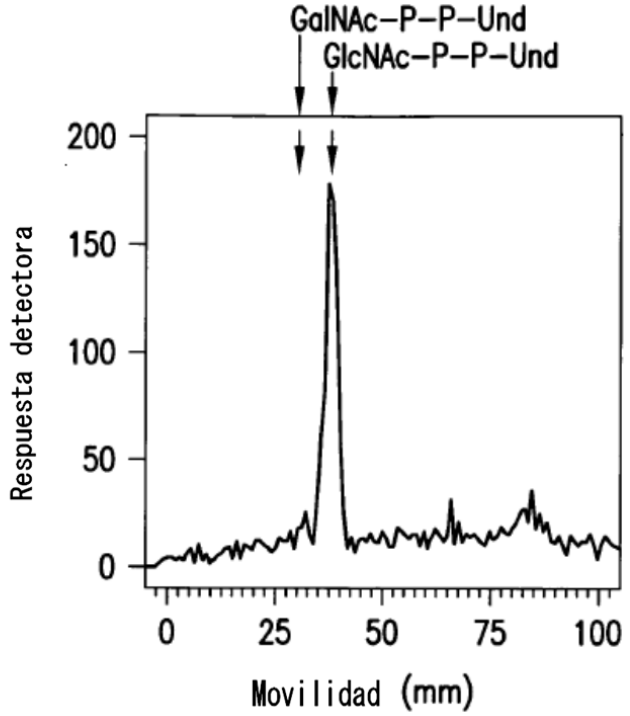

GalNAc–P–P–Und
GlcNAc–P–P–Und
Respuesta detectora
200
150
100
50
0
0
25
50
75
100
Movilidad (mm)

FIG.4A

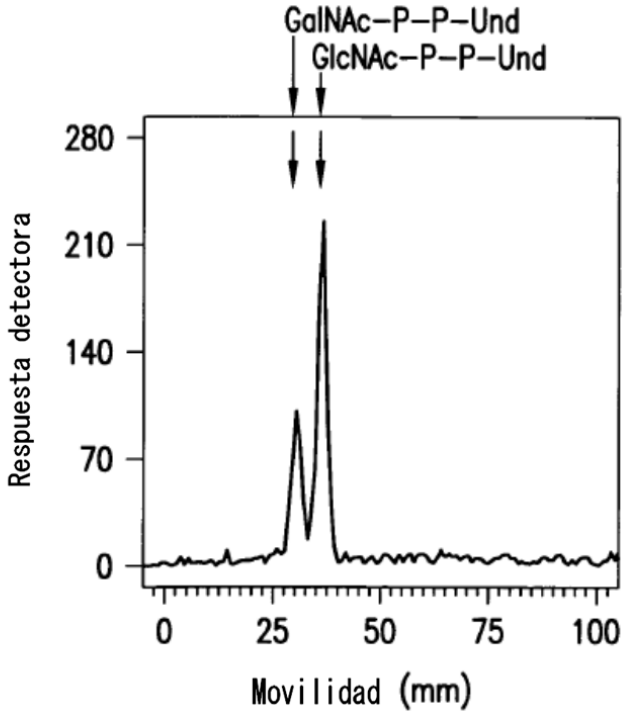

GalNAc–P–P–Und
GlcNAc–P–P–Und
Respuesta detectora
280
210
140
70
0
0
25
50
75
100
Movilidad (mm)

FIG.4B

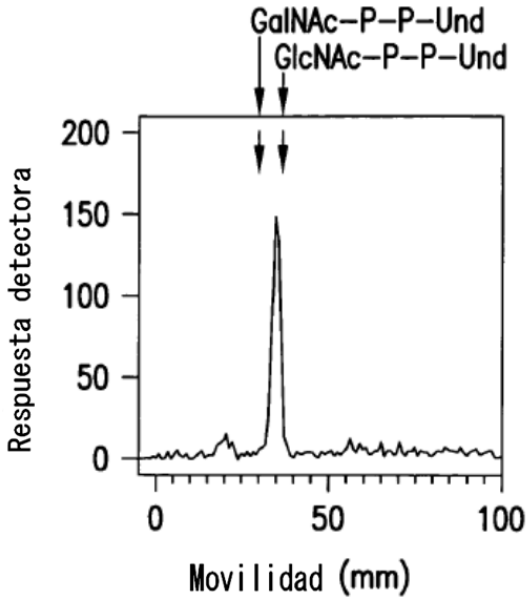


FIG.5A

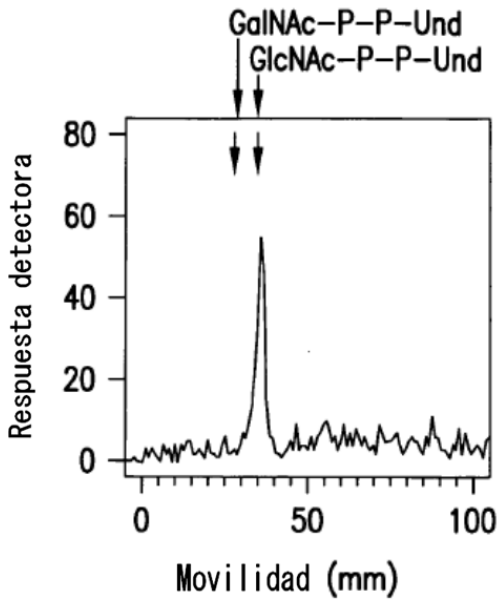


FIG.5C

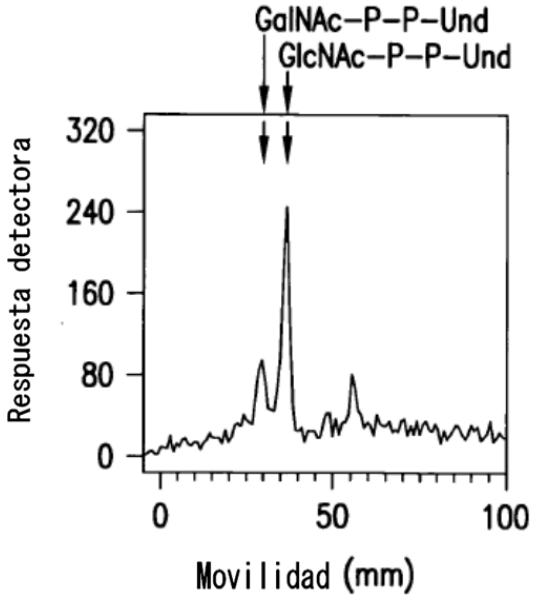


FIG.5B

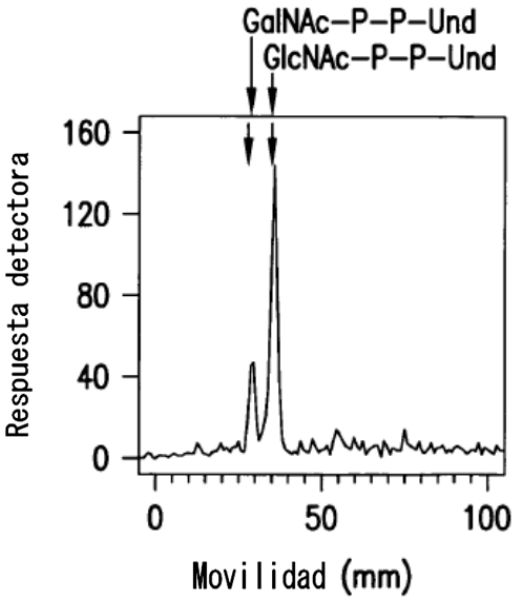


FIG.5D

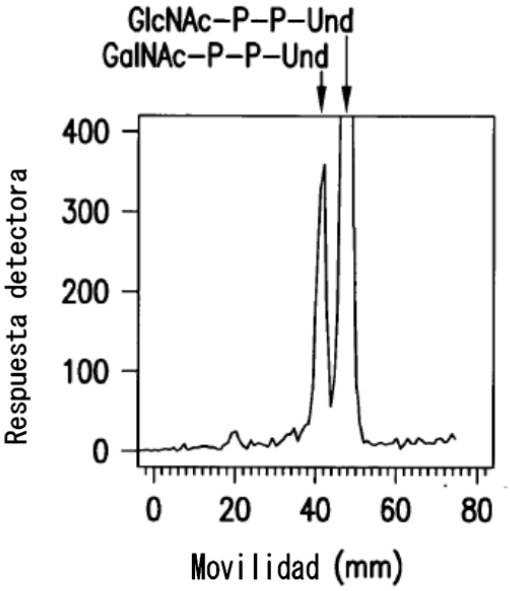


FIG.6A

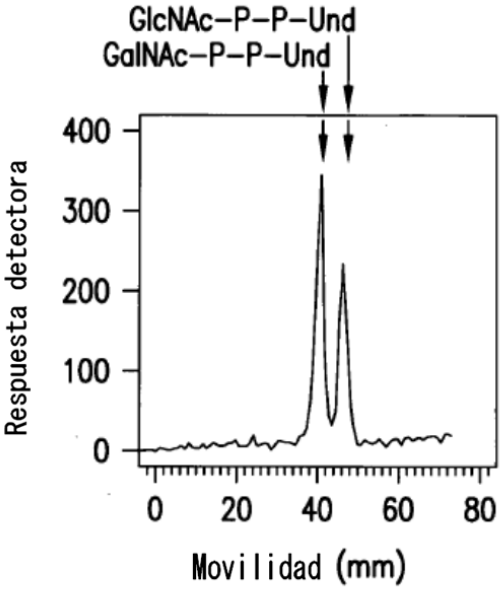


FIG.6B

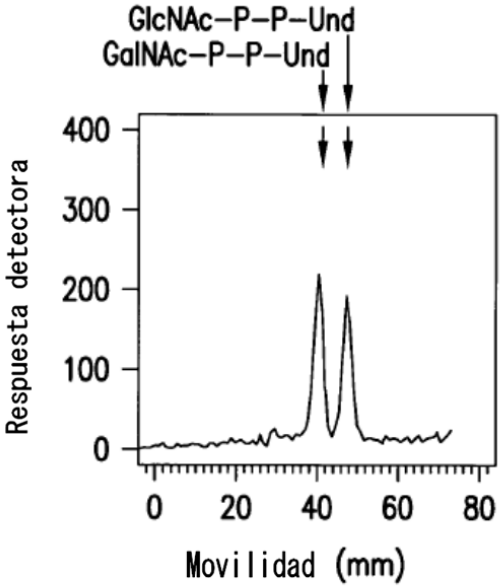


FIG.6C

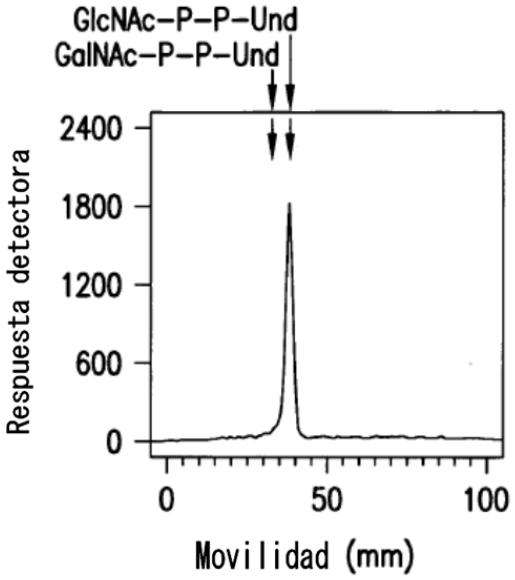


FIG.7A

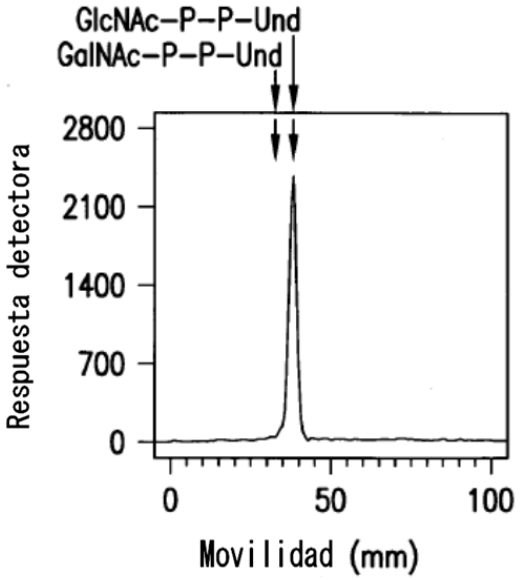


FIG.7B

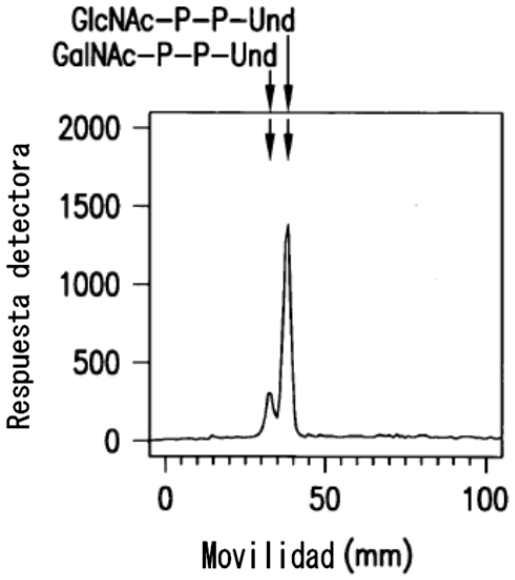


FIG.7C

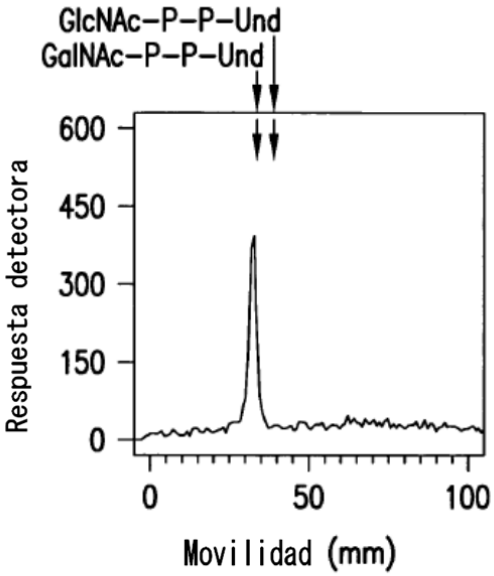


FIG.7D

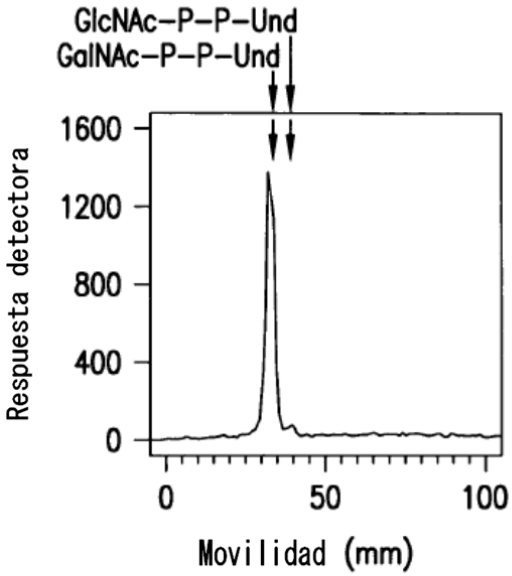


FIG.7E

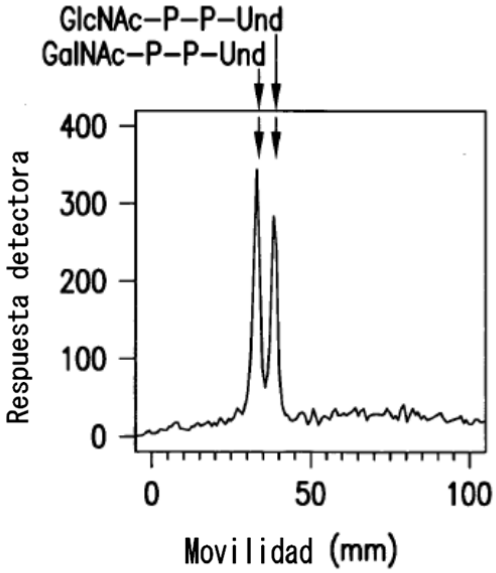


FIG.7F

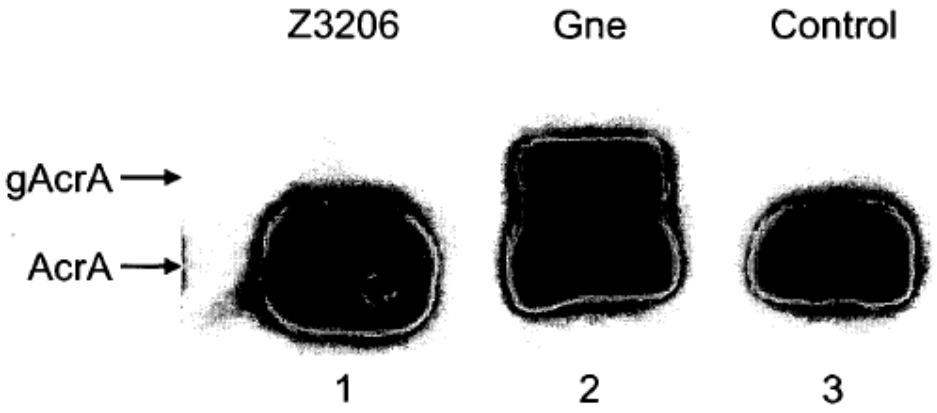


FIG.8

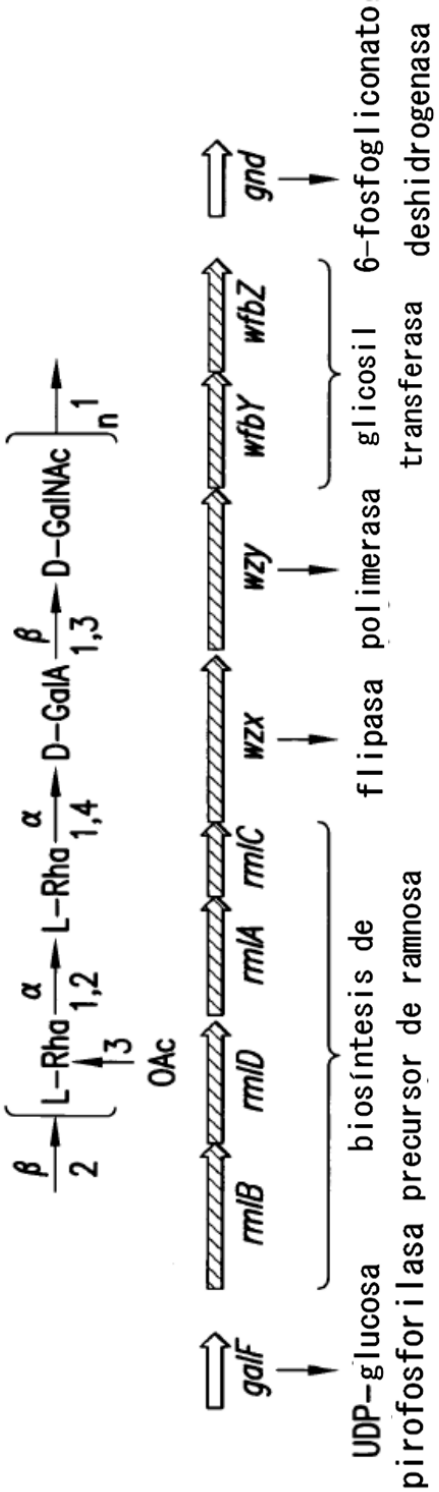

β
2
L-Rha
3
OAc
α
1,2
L-Rha
α
1,4
D-GalA
β
1,3
D-GalNAc
1
n
galF
rmlB
rmlD
rmlA
rmlC
wzx
wzy
wfbY
wfbZ
gnd
UDP-glucosa pirofosforilasa
biosíntesis de precursor de ramnosa
flipasa
polimerasa
glicosil transferasa
6-fosfogliconato deshidrogenasa

FIG.9

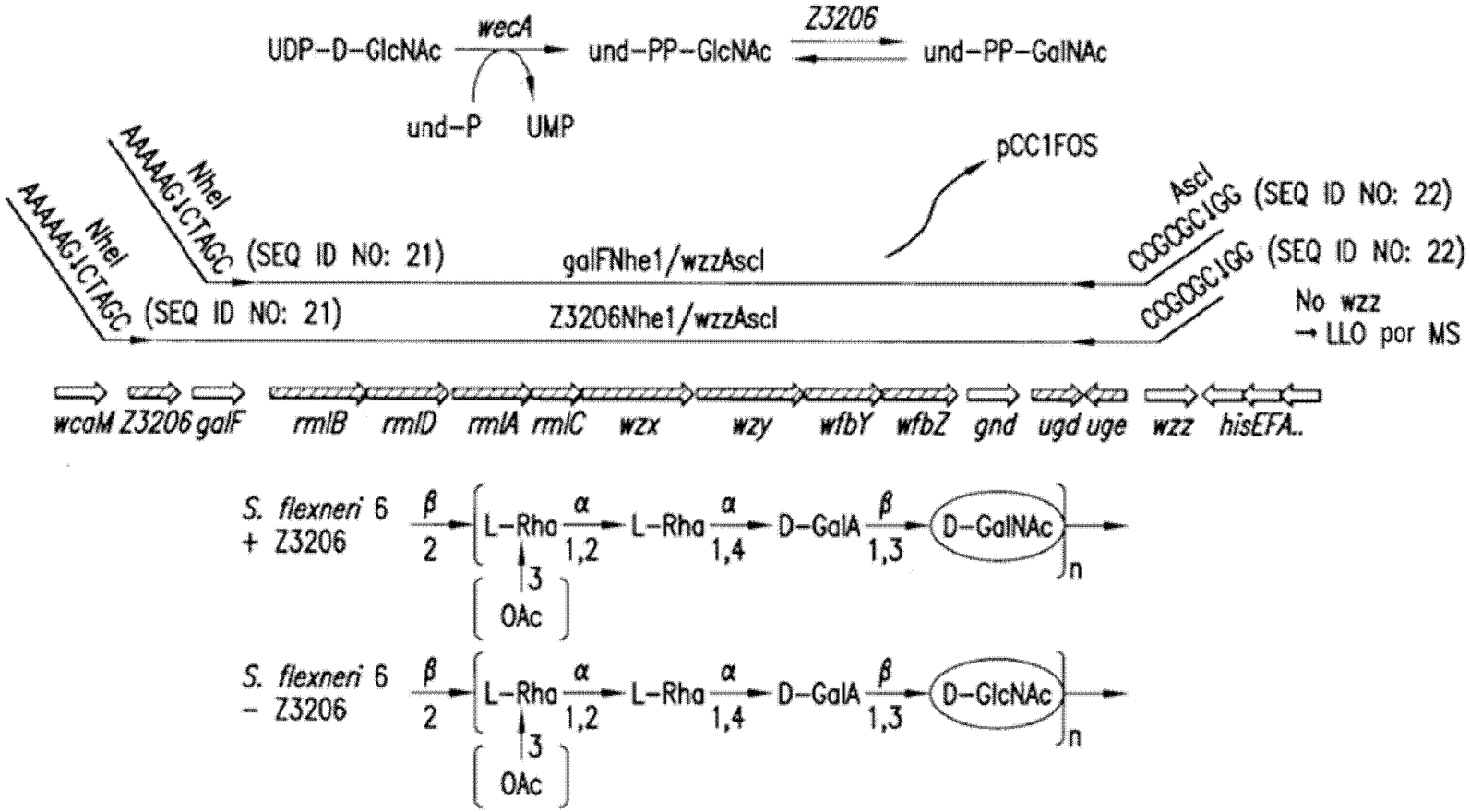
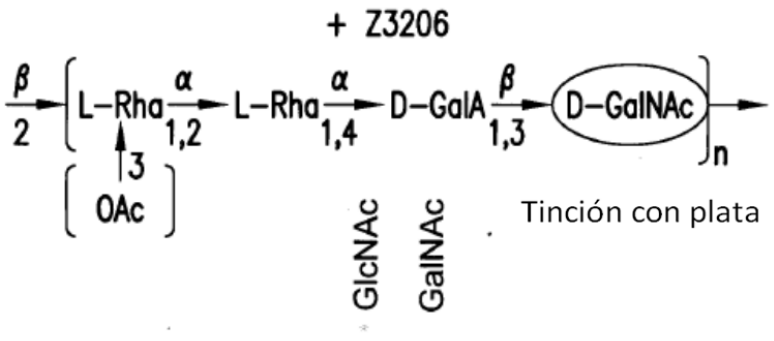
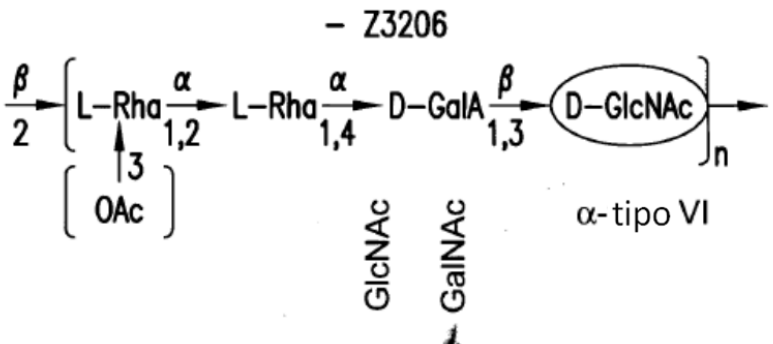

UDP-D-GlcNAc
wecA
und-PP-GlcNAc
Z3206
und-PP-GalNAc
und-P
UMP
pCC1FOS
NheI
AAAAAG↓CTAGC
(SEQ ID NO: 21)
galFNhe1/wzzAscI
AscI
CCGCGC↓GG
(SEQ ID NO: 22)
Z3206Nhe1/wzzAscI
No wzz
→ LLO por MS
wcaM Z3206 galF rmlB rmlD rmlA rmlC wzx wzy wfbY wfbZ gnd ugd uge wzz hisEFA..
S. flexneri 6
+ Z3206
L-Rha
L-Rha
D-GalA
D-GalNAc
OAc
S. flexneri 6
− Z3206
D-GlcNAc

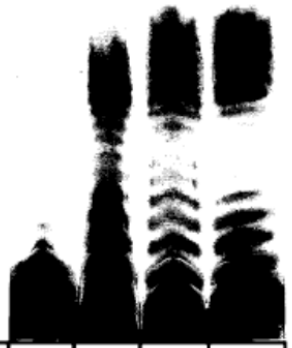
FIG.10





Tinción con plata

GlcNAc GalNAc



	1	2	3	4
S.Flex 6 antígeno O	X	✓	✓	✓
Z3206	X	X	✓	✓
wzz	X	X	X	✓

α-tipo VI

GlcNAc GalNAc

	1	2	3	4
S.Flex 6 antígeno O	X	✓	✓	✓
Z3206	X	X	✓	✓
wzz	X	X	X	✓

10 µl carga

12% Bis Tris

FIG. 11

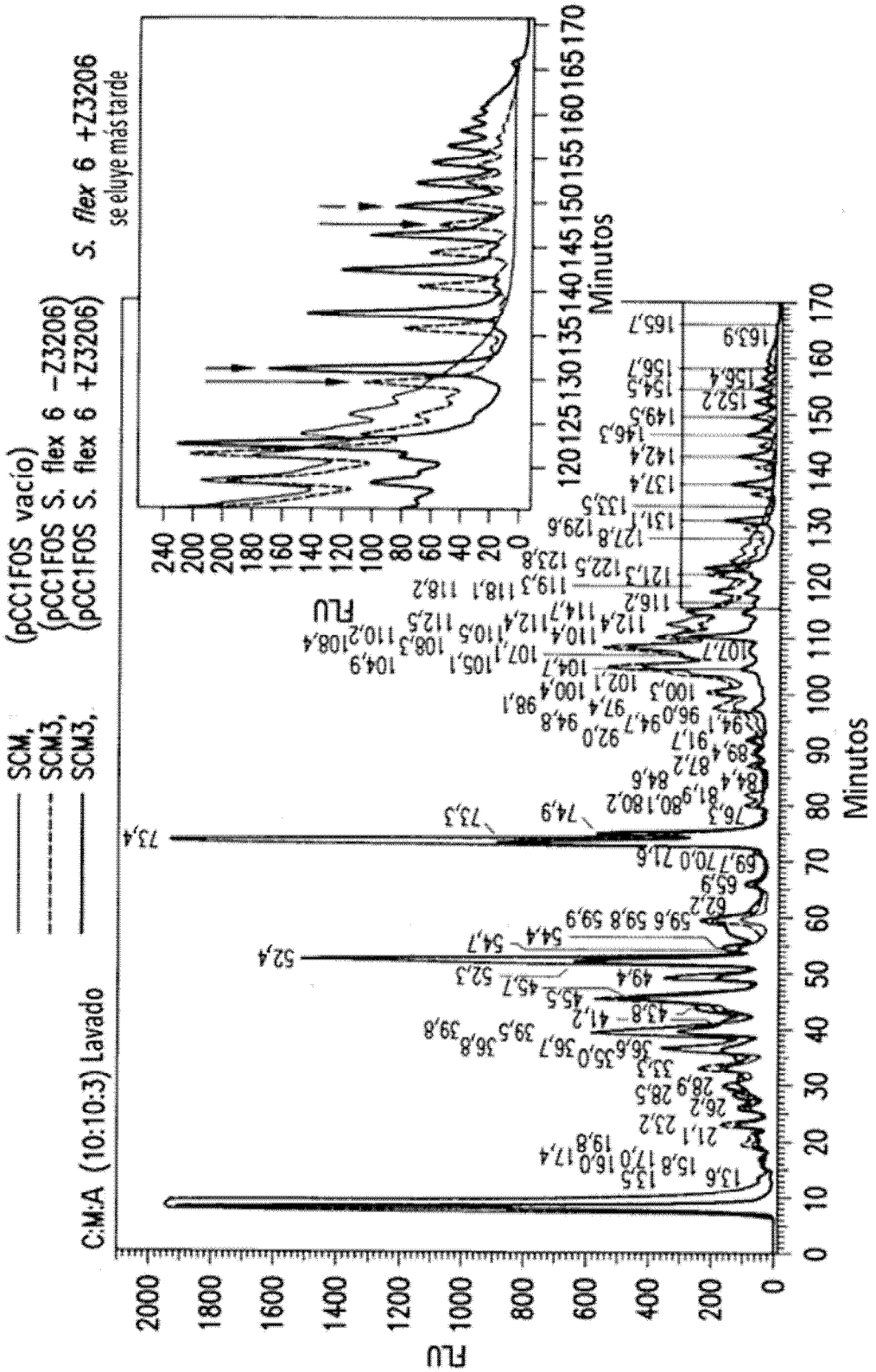
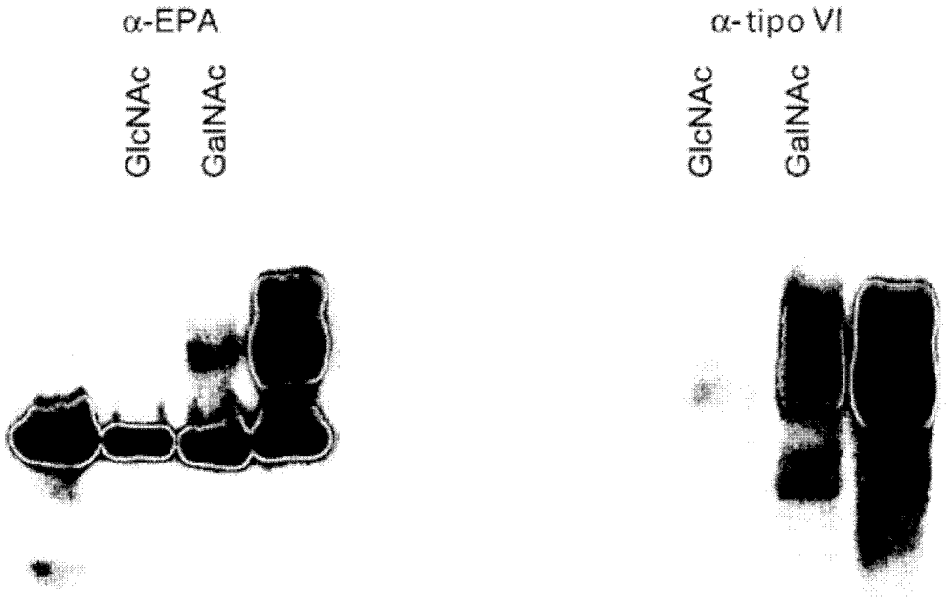

C:M:A (10:10:3) Lavado
SCM,
SCM3,
SCM3,
(pCC1FOS vacío)
(pCC1FOS S. flex 6 −Z3206)
(pCC1FOS S. flex 6 +Z3206)
S. flex 6 +Z3206
se eluye más tarde
FLU
Minutos

FIG. 12



	1	2	3	4
S.Flex 6 antígeno O	X	✓	✓	✓
Z3206	X	X	✓	✓
wzz	X	X	X	✓
EPA + pglB	✓	✓	✓	✓

	1	2	3	4
S.Flex 6 antígeno O	X	✓	✓	✓
Z3206	X	X	✓	✓
wzz	X	X	X	✓
EPA + pglB	✓	✓	✓	✓

FIG.13