

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 984 862**

51 Int. Cl.:

C12N 9/10 (2006.01)

C12P 21/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **06.08.2020 PCT/EP2020/072102**

87 Fecha y número de publicación internacional: **18.02.2021 WO21028303**

96 Fecha de presentación y número de la solicitud europea: **06.08.2020 E 20753722 (6)**

97 Fecha y número de publicación de la concesión europea: **24.04.2024 EP 4010467**

54 Título: **Enzimas PGLB oligosacariltransferasa mutadas**

30 Prioridad:

09.08.2019 US 201962884791 P
06.11.2019 US 201962931265 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
31.10.2024

73 Titular/es:

GLAXOSMITHKLINE BIOLOGICALS SA (100.0%)
rue de l'Institut 89
1330 Rixensart, BE

72 Inventor/es:

IHNKEN, LEIGH ANNE;
KARPIAK, JOEL;
KEMMLER, STEFAN JOCHEN;
KOWARIK, MICHAEL THOMAS;
MELBY, JOEL;
OLLIS, ANNE y
QUEBATTE, JULIEN LAURENT

74 Agente/Representante:

ARIZTI ACHA, Monica

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 984 862 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

Enzimas PGLB oligosacaryltransferasa mutadas

5 Campo de la invención

La presente invención se refiere a PglB oligosacaryltransferasas mejoradas y su uso en la glucosilación de proteínas y/o la producción de proteínas portadoras conjugadas con cadenas de sacáridos. La invención también incluye polinucleótidos que codifican para las PglB oligosacaryltransferasas mejoradas y células hospedadoras que comprenden los polinucleótidos.

Antecedentes de la invención

Las vacunas de glucoconjugados son ampliamente reconocidas por su capacidad para prevenir muchas infecciones bacterianas potencialmente mortales. Las vacunas de glicoconjugado en general se consideran eficaces y seguras y se han usado en humanos durante más de 30 años. La producción convencional de glicovacunas frecuentemente implica la modificación química de proteínas portadoras inmunogénicas con antígenos polisacáridos de bacterias patógenas. Sin embargo, más recientemente, han surgido procesos biotecnológicos para producir vacunas de glucoconjugados que se espera que reduzcan los costos de producción e incrementen aún más la homogeneidad y posiblemente la potencia y seguridad de las preparaciones de vacunas de glucoconjugados.

En células eucariotas, la glucosilación N-enlazada es un mecanismo de modificación de proteínas postraduccional clave que implica varias enzimas. En células procariotas, la glicosilación N-enlazada se cataliza por ciertas N-oligosacaryltransferasas bacterianas (N-OST). El cúmulo de genes de glicosilación de proteínas de *Campylobacter jejuni* (*C. jejuni*) incluye el gen *pglB*, que codifica para una N-OST unida a membrana (PglB_{Cj}). PglB se puede expresar en hospedadores bacterianos estándar, tal como *Escherichia coli* (*E. coli*), y puede glucosilar proteínas periplásmicas coexpresadas que portan al menos un motivo de glicosilación D/E-Z₁-N-Z₂-S/T (Z₁ y Z₂ ≠ P) expuesto en la superficie. PglB puede transferir antígenos de polisacáridos bacterianos a determinadas proteínas de *C. jejuni*, así como a proteínas portadoras inmunogénicas de otros organismos que contienen sitios de glucosilación modificados. WO 16/107818 y WO 16/023018 describe aspectos de esto. PglB puede transferir oligosacáridos y hasta cierto grado, estructuras de lipopolisacáridos de antígeno O de bacterias Gram-negativas y polisacáridos de antígeno capsular de bacterias Gram-positivas. Sin embargo, la eficiencia de la actividad de oligosacaryltransferasa de PglB puede variar dependiendo de la naturaleza del sacárido que se une covalentemente a una proteína que contiene la secuencia de consenso requerida. Por lo tanto, existe la necesidad de proteínas PglB mejoradas que sean capaces de catalizar la transferencia eficiente de sacáridos con una estructura diferente a aquellas transferidas en *C. jejuni*, a una proteína que contiene el motivo de glicosilación requerido.

La presente divulgación proporciona PglB oligosacaryltransferasas modificadas que se han modificado para mejorar la eficiencia de PglB cuando se transfiere una variedad de sacáridos que no se transfieren a una proteína en una célula de *C. jejuni*.

Por consiguiente, se proporciona un polipéptido de PglB oligosacaryltransferasa que comprende una secuencia de aminoácidos que es al menos 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntica a la secuencia de aminoácidos expuesta en SEQ ID NO: 1 o 2 o un fragmento funcional de la misma, donde la secuencia de aminoácidos de polipéptido de PglB oligosacaryltransferasa incluye la característica de que: al menos un residuo seleccionado del grupo que consiste en el aminoácido X57, X63, X94, X101, X172, X176, X191, X193, X233, X234, X255, X286, X295, X301, X319, X397, X402, X425, X435, X446, X462, X479, X523, X532, X601, X605, X606, X610, X645, X676 y X695 se sustituye por un aminoácido diferente al que se encuentra en esa posición en SEQ ID NO: 1 o fragmento funcional de lo mismo, donde el residuo que corresponde al aminoácido 57 de SEQ ID NO: 1 se muta a R o K o T, preferentemente a R o T, más preferentemente a R_y donde el polipéptido de PglB oligosacaryltransferasa es capaz de incrementar el rendimiento de glucosilación de la proteína con el sacárido para producir una proteína glicosilada por al menos 1,5 veces en comparación con una PglB oligosacaryltransferasa correspondiente que tiene la secuencia de SEQ ID NO: 1.

En una realización preferida, la secuencia de aminoácidos es al menos 90% idéntica a SEQ ID NO:2.

En una segunda realización, se proporciona un polinucleótido que codifica para un polipéptido de PglB oligosacaryltransferasa mutado de la invención.

En una tercera realización, se proporciona una composición o célula hospedadora (por ejemplo, una célula hospedadora procariota o una célula hospedadora de *E. coli*) que comprende al menos una PglB oligosacaryltransferasa de la invención o un polinucleótido que codifica para al menos una PglB OST de la invención.

En una cuarta realización se proporciona un proceso para preparar una proteína glicosilada, que comprende los pasos de:

(a) cultivar una célula hospedadora de la invención, que comprende la PglB de la invención y/o un polinucleótido que codifica para la PglB de la invención, bajo condiciones adecuadas para la producción de proteínas; y

5 (b) aislar la proteína glucosilada de la célula hospedadora.

En una quinta realización se proporciona un proceso *in vitro* para preparar una proteína glucosilada, que comprende los pasos de;

10 i) mezclar conjuntamente:

a) una PglB oligosacaryltransferasa de la invención;

15 b) una proteína que comprende al menos una secuencia de consenso de glucosilación que comprende la secuencia de aminoácidos Asp/Glu-Z₁-Asn-Z₂-Ser/Thr donde Z₁ y Z₂ pueden ser cualquier aminoácido natural excepto Pro; y

c) una cadena de sacárido en un portador lipídico reconocido por la PglB;

20 ii) incubar bajo condiciones adecuadas para la actividad enzimática de PglB para transferir la cadena de sacárido a la al menos una secuencia de consenso de glucosilación de la proteína para lograr una proteína glucosilada; y

iii) aislar la proteína glucosilada.

En una sexta realización, se proporciona una proteína glucosilada que se elabora por el proceso de la invención.

25

En una séptima realización se proporciona un uso de la PglB oligosacaryltransferasa o fragmento funcional de la misma de la invención en la producción de una proteína glucosilada en la cual un sacárido se une a un residuo N de una secuencia de consenso de glucosilación, que comprende la secuencia de aminoácidos Asp/Glu-Z₁-Asn-Z₂-Ser/Thr donde Z₁ y Z₂ pueden ser cualquier aminoácido natural excepto Pro, de una proteína para formar la proteína glucosilada.

30

Breve descripción de las figuras

35 Figura 1 - Las mutaciones en PglB dan por resultado mayor actividad de oligosacaryltransferasa cuando se transfiere el sacárido capsular de *S. pneumoniae* serotipo 8 a una proteína. El panel A muestra el cambio múltiple de compuesto de la producción cuando se usa PglB de la ronda 1, 2, 3, 4, 5, 6 y 7 para transferir el sacárido capsular de *S. pneumoniae* a una proteína portadora de EPA. El panel B muestra una transferencia Western y un gel teñido con Coomassie que demuestra el incremento de EPA glucosilada con sacárido de *S. pneumoniae* serotipo 8 después de diferentes rondas de mutación. El panel C muestra el gel teñido con Coomassie y transferencias Western sondadas con un anticuerpo anti-*S. pneumoniae* serotipo 8, que demuestra la producción incrementada de conjugado que contiene sacárido de *S. pneumoniae* serotipo 8 después de la bioconjugación usando PglB_{Cj} mutada puntual.

40

45 Figura 2 - Las PglB mutadas también tienen actividad incrementada para transferencia de *S. pneumoniae* serotipo 22F a una proteína. El panel A muestra los resultados de ELISA que muestran actividad incrementada de las PglB de ronda 3, 4 y 5 en la transferencia del sacárido capsular de *S. pneumoniae* serotipo 22F a una proteína. El panel B muestra los resultados de transferencia Western del incremento en la producción de EPA glucosilado con un sacárido capsular de serotipo 22F de las PglB de la ronda 3, 4 y 5.

45

50 Figura 3 - Las PglB mutadas tienen actividad incrementada para transferencia de *S. pneumoniae* serotipos 23A y 35B a una proteína. El panel A muestra los resultados de un gel de coomassie y una transferencia Western que muestra glucosilación incrementada de EPA con polisacárido de *S. pneumoniae* serotipo 23A cuando se usó PglB de ronda 4, 5 y 7 para catalizar la transferencia. El panel B muestra glucosilación incrementada de EPA cuando se usó una PglB de ronda 3 para transferir el sacárido capsular de *S. pneumoniae* serotipo 35B, en comparación con el uso de PglB tipo silvestre.

50

55 Figura 4 - Las PglB mutadas tienen actividad incrementada para transferencia de *S. pneumoniae* serotipo 19A a una proteína. Resultados de tinción de Coomassie y transferencia Western para la cantidad de EPA glucosilada con *S. pneumoniae* serotipo 19A usando: carril 1 - PglB inactiva, carril 2 - PglB que contiene mutaciones Y77R, N311V y H479M, carril 3 - PglB de ronda 6, carriles 4-10 - diferentes mutaciones de PglB de ronda 7.

55

60 Figura 5 - Estructura de unidades de repetición de polisacárido capsular de *S. pneumoniae*. Los círculos oscuros indican residuos de glucosa, círculos claros indican residuos de galactosa, círculos claros con "f" en el interior indican galactofurano, cuadrados oscuros indican N-acetilglucosamina, cuadrados claros indican N-acetilgalactosamina, triángulos indican ramnosa, diamantes divididos horizontalmente con la parte superior oscura indican ácido glucurónico, los óvalos indican glicerol, estrellas indican ribitol, cuadrados oscuros indican N-acetilfucosamina, cuadrados medios indican N-acetilmanosamina, cuadrados divididos diagonalmente indican 4-amino-N-acetilfucosamina y diamantes

60

divididos horizontalmente indican ácido galacturónico.

Figura 6 - Estructura de unidades de repetición de sacárido capsular de estreptococo de grupo B.

5 Figura 7 - Alineación de secuencias de aminoácidos de secuencias de OST de PglB de cepas de *Campylobacter coli* que muestran que EFM37568 es algo divergente, en tanto que EIA90085 y CDG57218 son más similares.

Descripción detallada de la invención

10 La presente divulgación proporciona PglB oligosacariltransferasas altamente eficientes capaces de catalizar la adición de sacárido a una proteína que contiene al menos una secuencia de consenso de glucosilación de D/E-Z₁-N-Z₂-S/T (Z₁ y Z₂ ≠ P). Estas PglB oligosacariltransferasas altamente eficientes se logran al modificar la PglB. Por ejemplo, a través de la selección de mutaciones puntuales ventajosas, ya sea individualmente o en combinación, que dan por resultado producción incrementada de proteína N-glucosilada.

15 Los ensayos para confirmar la actividad de las OST de PglB descritas en la presente se conocen bien por los expertos en la técnica (por ejemplo, ELISA, transferencia Western) e incluyen los ensayos descritos en los ejemplos 2 y 3. En algunas realizaciones, la OST de PglB es una PglB modificada que se expresa opcionalmente en una célula hospedadora, opcionalmente una célula hospedadora heteróloga (es decir, una célula hospedadora que no es una célula de *Campylobacter*).

Los oligosacáridos y polisacáridos pueden incluir cualquier oligosacárido o polisacárido descrito en la presente.

Las proteínas portadoras pueden comprender cualquier proteína portadora descrita en la presente.

25 En algunas realizaciones, la OST de PglB comprende modificaciones (por ejemplo, sustituciones de aminoácidos o sustituciones de nucleótidos en el polinucleótido que codifica para PglB) en, por ejemplo, dos o más, tres o más, cuatro o más, cinco o más, seis o más, siete o más, ocho o más, nueve o más, o diez o más aminoácidos, por ejemplo, 2-30, 3-25, 4-20, 5-20, 6-20, 7-20 o 10-20 sustituciones de aminoácidos. En una realización, las sustituciones están en las posiciones correspondientes a los siguientes residuos de SEQ ID NO:1:57; 57 y 77; ;57, 77 y 311; 57 y 462; 57, 462 y 479; 57, 462, 479, 77 y 311; 57, 462, 479, 300, 301, 308 y 570; 57, 462, 479, 300, 301, 308, 570 y 77; 57, 462, 479, 300, 301, 308, 570, 77 y 311.

35 En una realización, las siguientes sustituciones están presentes en las posiciones correspondientes a los siguientes residuos de SEQ ID NO: 1 (donde / denota o):A57R/T; A57R/T y Y77R; A57R y Y77R; A57R/T, Y77R y N311V; A57R, Y77R y N311V; A57R/T y Y462P/C/W/T/N; A57R/T y Y462W/T/N; A57R y Y462W; A57R/T, Y462P/C/W/T/N y H479M; A57R/T, Y462W/T/N y H479M; A57R, Y462W y H479M; A57R/T, Y462P/C/W/T/N, H479M, Y77R y N311V; A57R/T, Y462W/T/N, H479M, Y77R y N311V; A57R, Y462W, H479M, Y77R y N311V; A57R/T, Y462P/C/W/T/N, H479M, L301P/G/F y L570R/V; A57R, Y462W, H479M, L301P y L570R; A57R/T, Y462P/C/W/T/N, H479M, N300L, L301P/G/F, F308W/F y L570R/V; A57R, Y462W, H479M, N300L, L301P, F308W y L570R; A57R/T, Y462P/C/W/T/N, H479M, L301P/G/F, L570R/V y Y77R; A57R, Y462W, H479M, L301P, L570R y Y77R; A57R/T, Y462P/C/W/T/N, H479M, N300L, L301P/G/F, F308W/F, L570R/V y Y77R; A57R, Y462W, H479M, N300L, L301P, F308W, L570R y Y77R; A57R/T, Y462P/C/W/T/N, H479M, L301P/G/F, L570R/V, Y77R y N311V; A57R, Y462W, H479M, L301P, L570R, Y77R y N311V; A57R/T, Y462P/C/W/T/N, H479M, N300L, L301P/G/F, F308W/F, L570R/V, Y77R y N311V; A57R, Y462W, H479M, N300L, L301P, F308W, L570R, Y77R y N311V; A57R, Y462W, H479M, N300L, L301P, F308W, L570R, Y77R, S80D y N311V.

40 En algunas realizaciones, una mutación puntual individual en la PglB modificada puede incrementar la eficiencia de glucosilación de una proteína portadora con un polisacárido que conduce a un incremento en el rendimiento comparativo de proteína glucosilada. Un incremento en el rendimiento comparativo se puede determinar al dividir el rendimiento de la proteína glucosilada producida por una OST de PglB modificada que contiene una sustitución de aminoácidos particular con el rendimiento de la proteína glucosilada producida por una OST de PglB correspondiente que no contiene esa sustitución de aminoácidos. En una realización, la introducción de una mutación puntual individual en la PglB modificada puede incrementar el rendimiento comparativo entre aproximadamente 1,1 veces y aproximadamente 10 veces, entre aproximadamente 1,2 veces y aproximadamente 7 veces, entre aproximadamente 1,3 veces y aproximadamente 5 veces, entre aproximadamente 1,5 veces y aproximadamente 2,5 veces o entre aproximadamente 1,5 veces y aproximadamente 6 veces en comparación con la tasa de una PglB correspondiente que carece de esa mutación puntual particular. El efecto positivo de las mutaciones puntuales individuales puede multiplicarse cuando se combinan como múltiples mutaciones puntuales en una PglB. En algunas realizaciones, la PglB modificada puede contener al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 sustituciones de aminoácidos que incrementan el rendimiento comparativo de proteína glucosilada por más de 2 veces, más de 5 veces, más de 10 veces, más de 50 veces, más de 100 veces, más de 200 veces, más de 500 veces, más de 700 veces, más de 1.000 veces, más de 2.000 veces, más de 3.000 veces, más de 4.000 veces, más de 5.000 veces, más de 6.000 veces, más de 7.000 veces, más de 8.000 veces, más de 10.000 veces, en comparación con el rendimiento obtenido usando la PglB correspondiente en la cual no se ha introducido la combinación de mutaciones puntuales (la PglB original).

5 En algunas realizaciones, el rendimiento de glucosilación de proteínas de la PglB modificada y la forma no mutada de la PglB modificada se puede comparar al comparar el rendimiento de glucosilación de la PglB modificada y de la PglB original de una proteína con un polisacárido u oligosacárido que carece de un azúcar N-acetilo en el extremo reductor. Por ejemplo, un azúcar con una glucosa en el extremo reductor, por ejemplo, un polisacárido capsular de *Streptococcus pneumoniae* o un sacárido capsular de estreptococo de grupo B, por ejemplo, aquellos representados en las figuras 5 y 6. En una realización, el polisacárido u oligosacárido que carece de un azúcar N-acetilo en el extremo reductor es el sacárido capsular de *S. pneumoniae* serotipo 8.

10 En algunas realizaciones, el rendimiento de glucosilación de la PglB modificada de una proteína con un polisacárido u oligosacárido que carece de un azúcar N-acetilo en el extremo reductor se compara con el rendimiento de una PglB no mutada de una proteína glucosilada con un polisacárido u oligosacárido que tiene un azúcar N-acetilo en el extremo reductor. Por ejemplo, un azúcar con una glucosa en el extremo reductor, por ejemplo, un polisacárido capsular de *Streptococcus pneumoniae* o un sacárido capsular de estreptococo de grupo B, por ejemplo, aquellos representados en las figuras 5 y 6. En una realización, el polisacárido u oligosacárido que carece de un azúcar N-acetilo en el extremo reductor es el sacárido capsular de *S. pneumoniae* serotipo 8.

20 En algunas realizaciones, el rendimiento de glucosilación de la PglB modificada de una proteína glucosilada con un sacárido capsular de *S. pneumoniae* serotipo 8 se incrementa por al menos 1,5 veces, 1,7 veces, 2 veces, 2,5 veces, 3 veces, 5 veces, 10 veces, 100 veces, 500 veces o 1.000 veces con respecto a una PglB no mutada.

25 En algunas realizaciones, la PglB modificada puede incrementar el rendimiento *in vivo* o *in vitro* de glucosilación de una proteína con un sacárido capsular de *S. pneumoniae* serotipo 8 entre aproximadamente 1,5 veces y aproximadamente 2.000 veces, entre aproximadamente 2 veces y aproximadamente 1.000 veces, entre aproximadamente 2 veces y aproximadamente 20 veces o entre aproximadamente 20 veces y aproximadamente 2.000 veces en comparación con el rendimiento logrado con una forma no mutada de la PglB modificada.

30 En algunas realizaciones, la PglB modificada puede producir un nivel de glucosilación *in vivo* o un nivel de glucosilación *in vitro* de la proteína de entre aproximadamente 1 % a aproximadamente 70 %, de entre aproximadamente 3 % a aproximadamente 65 %, de entre aproximadamente 5 % a aproximadamente 60 %, de entre aproximadamente 5 % a aproximadamente 55 %, de entre aproximadamente 10 % a aproximadamente 50 %, de entre aproximadamente 15 % a aproximadamente 45 %, de entre aproximadamente 20 % a aproximadamente 40 %, o de entre aproximadamente 25 % a aproximadamente 35 %. En algunas realizaciones, la PglB modificada puede producir un nivel de glucosilación *in vivo* o un nivel de glucosilación *in vitro* de la proteína portadora de al menos 1 %, al menos 3 %, al menos 5 %, al menos 10 %, al menos 15 %, al menos 20 %, al menos 25 %, al menos 30 %, al menos 35 %, al menos 40 %, al menos 45 %, al menos 50 %, al menos 55 %, al menos 60 %, al menos 65 % o al menos 70 %.

40 La PglB modificada puede ser de cualquier organismo que tiene una PglB. En algunas realizaciones, la PglB modificada es de un organismo procarionta. En algunas realizaciones, la PglB modificada es de *Campylobacter jejuni* (*C. jejuni*), *Campylobacter coli* (*C. coli*), *Campylobacter lari* (*C. lari*), *Campylobacter upsaliensis* (*C. upsaliensis*), *Campylobacter curvus* (*C. curvus*), *Campylobacter concisus* (*C. concisus*), *Campylobacter hominis* (*C. hominis*), *Campylobacter gracilis* (*C. gracilis*), *Campylobacter showae* (*C. showae*), *Sulfurimonas autotrophica* (*S. autotrophica*), *Sulfurimonas denitrificans* (*S. denitrificans*), *Sulfurospirillum deleyianum* (*S. deleyianum*), *Sulfuricumrum kujiense* (*S. kujiense*), *Nautiliaprofundicola* (*N. profundicola*), *Sulfuorum* sp. NBC37-1, *Wolinella succinogenes* (*W. succinogenes*), *Caminiabacter mediatlanticus* (*C. mediatlanticus*), *Nitratiruptor* sp. SB155-2, *Helicobacter pullorum* (*H. pullorum*), *Helicobacter Canadensis* (*H. Canadensis*), *Helicobacter winghamensis* (*Helicobacter winghamensis*), *Desulfurobacterium thermolithotr* (*D. thermolithotr*), *Desulfomicrobium baculatum* (*D. baculatum*), *Desulfovibrio vulgaris* (*D. vulgaris*), *Desulfovibrio alkaliphilus* (*D. alkaliphilus*), *Desulfohalobium retbaense* (*D. retbaense*), *Deferribacter desulfuricans* (*D. desulfuricans*), *Desulfovibrio salexigenes* (*D. salexigenes*), *Desulfovibrio piger* (*D. salexigenes*), *Desulfovibrio aespoensis* (*D. aespoensis*), *Cand. Puniceispirillum marinum*, *Calditerrivibrio nitroreducens* (*C. nitroreducens*) o *Methanothermus fervidus* (*M. fervidus*).

55 Las mutaciones puntuales identificadas en la presente son habitualmente aquellas que se encuentran en la secuencia de PglB de *C. jejuni* (por ejemplo, PglB de *C. jejuni* de SEQ ID NO: 1). Un aspecto adicional de la invención es una PglB de cualquiera de las especies identificadas anteriormente que contiene la mutación correspondiente a aquellas descritas para PglB de *C. jejuni* de SEQ ID NO: 1.

60 En algunas realizaciones, el polipéptido OST de PglB es una PglB modificada, un homólogo de PglB modificado o una versión modificada de una variante de PglB que se presenta de manera natural. Los homólogos de PglB_{Cj} pueden comprender homólogos de PglB_{Cj} que se presentan de manera natural y homólogos de PglB_{Cj} que no se presentan de manera natural. Los homólogos de PglB_{Cj} pueden comprender proteínas que tienen al menos 80 %, al menos 85 %, al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 % o al menos 99 % de identidad de secuencia a una PglB_{Cj} de SEQ ID NO: 1. El grado de identidad de secuencia se puede determinar usando el algoritmo de alineación de homología de Needleman y Wunsch, el programa ClustaW o el algoritmo BLASTP. Se prefiere un algoritmo que use

alineación global (Needleman y Wunsch).

5 En algunas realizaciones, la PglB modificada es una PglB_{C1} modificada, un homólogo de PglB_{C1} modificado o una versión modificada de una variante de PglB_{C1} que se presenta de manera natural. Los homólogos de PglB_{C1} pueden comprender homólogos de PglB_{C1} que se presentan de manera natural y homólogos de PglB_{C1} que no se presentan de manera natural. Los homólogos de PglB_{C1} pueden comprender proteínas que tienen al menos 80 %, al menos 85 %, al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 % o al menos 99 % de identidad de secuencia a una PglB_{C1} de SEQ ID NO: 2. El grado de identidad de secuencia se puede determinar por el algoritmo de alineación de homología de Needleman y Wunsch, el programa ClustalW o un algoritmo BLASTP. Se prefiere un algoritmo que use alineación global (Needleman y Wunsch).

15 En algunas realizaciones, la PglB modificada comprende un fragmento de PglB, por ejemplo, un fragmento de PglB_{C1}. En algunas realizaciones, el fragmento de PglB comprende al menos 100, al menos 150, al menos 200, al menos 250, al menos 300, al menos 350, al menos 400, al menos 450, al menos 500, al menos 550, al menos 600 o al menos 650 aminoácidos contiguos de una PglB de longitud completa, por ejemplo, de SEQ ID NO: 1.

Modificaciones de PglB_{C1}

20 En algunas realizaciones, las OST de PglB modificadas descritas en la presente son N-OST tipo silvestre modificadas, por ejemplo, PglB_{C1} tipo silvestre. En algunas realizaciones, la PglB_{C1} tipo silvestre es una PglB_{C1} tipo silvestre de SEQ ID NO: 9, o de una variante que se presenta de manera natural de la misma:

```

MLKKEYLKNP YLVLFAMIIL AYVFSVFCRF YVWVWASEFN EYFFNNQLMI
ISNDGYAFAE
GARDMIAGFH QPNDSYYGS SLSALTYWLY KITPFSFESI ILYMSTFLSS
LVVIPTILLA
NEYKRPLMGF VAALLASIAN SYNRTMSGY YDTDMLVIVL PMFILFFMVR
MILKKDFEFL
IALPLFIGIY LWWYPSSYTL NVALIGLFLI YTLIFHRKEK IFYIAVILSS
LTLNSIAWFY
QSAIIVILFA LFALEQKRLN FMIIGILGSA TLIFLILSGG VDPILYQLKF
YIFRSDESAN
LTQGFMYFNV NQTIQEVENV DLSEFMRRIS GSEIVFLFSL FGFVWLLRKH
KSMIMALPIL
VLGFLALKGG LRFTIYSVPV MALGFGFLLS EFKAIMVKKY SQLTSNVCIV
FATILTLAPV
FIHIYNYKAP TVFSQNEASL LNQLKNIANR EDYVVTWWDY GYPVRYSDV
KTLVDGKHL
GKDNFFPSFA LSKDEQAAAN MARLSVEYTE KSFYAPQNDI LKTDILQAMM
KDYNSQSNVDL
FLASLSKPDF KIDTPKTRDI YLYMPARMSL IFSTVASFSF INLDTGVLDK
PTFESTAYPL
DVKNGEIYLS NGVLSDDFR SFKIGDNVVS VNSIVEINSI KQGEYKITPI
DDKAQFYIFY
LKDSAIPYAQ FILMDKTMFN SAYVQMFFLG NYDKNLFDLV INSRDAKVEK
LKIYPYDVPD
YA
    
```

25 En algunas realizaciones, uno o más de los aminoácidos X57, X63, X94, X101, X172, X176, X191, X193, X233, X234, X255, X286, X295, X301, X319, X397, X402, X425, X435, X446, X462, X479, X523, X532, X601, X605, X606, X610, X645, X676 y X695 de SEQ ID NO: 1, o cualquier combinación de los mismos, se modifican, donde X57 se muta a R o K o T. En una realización, uno o más de X57, X301, X319, X462, X479 y X523, o cualquier combinación de los mismos se modifica, por ejemplo, por sustitución, donde X57 se muta a R o K o T. En una realización, uno o más de X57, X462 y X479 se modifican, donde X57 se muta a R o K o T. En una realización, X57 se modifica por mutación a R o K o T. En una realización, X462 y X479 se sustituyen, donde X57 se muta a R o T. En una realización, se realiza la sustitución de A57R. En una realización, se realizan sustituciones de A57R, Y462W y H479M.

35 (a) Modificaciones de PglB_{C1}

En algunas realizaciones, las OST de PglB modificadas descritas en la presente son OST de PglB tipo silvestre modificadas, por ejemplo, PglB_{C1} tipo silvestre (PglB de *Campylobacter lari*). En algunas realizaciones, la PglB_{C1} tipo silvestre es una PglB_{C1} tipo silvestre de SEQ ID NO: 10, o de una variante que se presenta de manera natural de la misma:

```

MKLQQNFTDN NSIKYTCILI LIAFAFSVLC RLYWVAWASE
FYEFFFNDQL
MITTNDGYAF AEGARDMIAG FHQPNDLSYF GSSLSTLTYW
LYSILPFSFE SIILYMSAFF ASLIVVPIIL IAREYKLTYY
GFIAALLGSI ANSYNRTMS GYYDTDMLVL VLPMLILLTF
IRLTINKDIF TLLSPVFIM IYLWWYPSSY SLNFAMIGLF
GLYTLVFHRK EKIFYLTIAL MIIALSMLAW QYKLALIVLL
FAIFAFKEEK INFYMIWALI FISILILHLS GGLDPVLYQL
KFYVFKASDV QNLKDAAFMY FNVNETIMEV NTIDPEVFMQ
RISSSVLVFI LSFIFGILLC KDHKSMLLAL PMLALGFMAL
RAGLRFTIYA VPVMALGFGY FLYAFFNFLE KKQIKLSLRN
KNILLILIAF FSISPALMHI YYYKSSTVFT SYEASILNDL
KNKAQREDYV VAWWDYGYPI RYYSVKTILI DGGKHLGKDN
FFSSFVLSKE QIPAAANMARL SVEYTEKSKF ENYPDVLKAM
VKDYNKTSK DFLESLNDKD FKFDTNKTRD VYIYMPYRML
RIMPVVAQFA NTNPDNGEQE KSLFFSQANA IAQDKTTGSV
MLDNGVEIIN DFRALKVEGA SIPLKAFVDI ESITNGKFYY
NEIDSKAQIY LLFLREYKSF VILDESLYNS SYIQMFLLNQ
YDQDLFEQIT NDTRAKIYRL KR

```

5 En una realización adicional, se proporciona una PglB_{C1} modificada que comprende una sustitución de A59R, Y468W y H485M. En una realización adicional, se proporciona una PglB_{C1} modificada que comprende una sustitución de N314V, Y79R, A59R, Y468W y H485M.

PglB de *Campylobacter coli*

10 Un aspecto adicional de la invención son PglB nativas particularmente de *Campylobacter coli* que se ha encontrado que tienen mayor actividad para la transferencia de ciertos sacáridos a una proteína portadora que contiene la secuencia de consenso de glucosilación Asp/Glu-Z₁-Asn-Z₂-Ser/Thr (donde Z₁ y Z₂ pueden ser cualquier aminoácido natural excepto Pro). En el caso de PglB de la cepa EFM37568 de *C. coli* (SEQ ID NO: 12) se ha encontrado que la actividad de OST es mayor que la de PglB de *C. jejuni* nativa (SEQ ID NO: 1) para la transferencia de al menos *S. flexneri* 2a, 3a y 6, más *E. coli* 018.

15 Se probó PglB de la cepa CDG57218 de *C. coli* (SEQ ID NO: 13) para su actividad en la transferencia del sacárido de *E. coli* 018 a una proteína que contiene una secuencia de consenso de glucosilación. Esta actividad de OST fue la misma entre PglB de *C. jejuni* y PglB de SEQ ID NO: 13.

20 Se probó PglB de la cepa EIA90085 de *C. coli* (SEQ ID NO: 14) para su actividad en la transferencia del sacárido de *E. coli* 018 a una proteína que contiene una secuencia de consenso de glucosilación. Esta actividad de OST fue ligeramente mayor que la de PglB de *C. jejuni*. Se concluyó que las PglB tipo silvestre de varias cepas de *C. coli* son equivalentes o superiores a PglB de *C. jejuni* para catalizar la transferencia de sacárido a una proteína que contiene la secuencia de consenso de glucosilación para PglB. Las secuencias de tres tinciones de PglB de *C. coli* se muestran en la figura 7 y la secuencia de PglB de cepa EFM37568 es ligeramente divergente de la secuencia de aminoácidos de las otras dos PglB probadas. La PglB de *C. coli* EFM37568 dio resultados particularmente buenos.

30 Una realización adicional de la invención proporciona un uso de una PglB oligosacariltransferasa (OST) o fragmento funcional de la misma de *Campylobacter coli* (PglB_{C.coli}) en la producción de una proteína glucosilada en la cual un sacárido se une a un residuo N de una secuencia de consenso de glucosilación, que comprende la secuencia de aminoácidos Asp/Glu-Z₁-Asn-Z₂-Ser/Thr donde Z₁ y Z₂ pueden ser cualquier aminoácido natural excepto Pro. En una realización, el uso es de una PglB de *C. coli* que tiene una secuencia de aminoácidos que es al menos 80 %, 85 %, 90 %, 95 %, 97 %, 99 % o 100 % idéntica a SEQ ID NO: 12, 13 o 14. Opcionalmente, la PglB de *C. coli* tiene una secuencia de aminoácidos que es al menos 80 %, 85 %, 90 %, 95 %, 97 %, 99 % o 100 % idéntica a SEQ ID NO: 12.

35 Una realización adicional de la invención divulga la introducción de las sustituciones de aminoácidos descritas en la presente en una PglB de *C. coli*. En una realización, la PglB de *C. coli* contiene al menos un residuo seleccionado del grupo que consiste en el aminoácido X57, X63, X94, X101, X172, X176, X191, X193, X233, X234, X255, X286, X295, X301, X319, X397, X402, X426, X436, X447, X463, X480, X524, X533, X602, X606, X607, X611, X646, X677 y X696 se sustituye por un aminoácido diferente al que se encuentra en esa posición en SEQ ID NO: 12, donde X57 se muta a R o K o T. En una realización, se proporciona un polipéptido de PglB_{C.coli} oligosacariltransferasa o fragmento funcional que contiene una sustitución en cualquiera de los residuos que corresponden a los aminoácidos X57, X463 y X480 de SEQ ID NO: 12, donde X57 se muta a R o K o T; opcionalmente, la sustitución es al menos una de A57R, Y463 y H480M. En

una realización, la PglB *C. coli* contiene al menos una sustitución de aminoácidos en el o los residuos X57, X77, X311, X462 y/o X480, donde X57 se muta a R o K o T. En una realización, las sustituciones de aminoácidos están en 1, 2, 3, 4 o 5 de estas posiciones. En una realización, las sustituciones son al menos una sustitución de aminoácidos en el o los residuos A57R, Y77R, N311V, Y463W y/o H480M, donde A57 se muta a R. En una realización, una, dos, tres, cuatro o cinco de estas mutaciones se introducen en la secuencia de aminoácidos de PglB de *C. coli*, que tiene opcionalmente la secuencia de SEQ ID NO: 12.

Oligosacáridos y polisacáridos

Los oligosacáridos que se pueden enlazar a una proteína mediante las OST de PglB modificadas proporcionadas en la presente pueden tener de aproximadamente 2 a aproximadamente 100 unidades de monosacáridos, por ejemplo, 2, 4, 6, 8, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 o 100 o de 4 a 90, 6 a 80, 8 a 70, 10 a 60, 15 a 50, 20 a 40 o 25 a 40 unidades de monosacáridos. Los polisacáridos que se pueden enlazar a una proteína por las OST de PglB modificadas proporcionadas en la presente pueden tener más de 100 unidades de monosacáridos, por ejemplo, al menos 101, 110, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1.000 unidades de monosacáridos o más. Por ejemplo, de 100 a 500, de 100 a 300 o de 100 a 200 unidades de monosacáridos.

Las proteínas u OST de PglB pueden comprender cualquier OST de PglB o cualquier proteína divulgada en la presente.

En algunas realizaciones, el azúcar en el extremo reductor del oligosacárido o polisacárido es una pentosa, hexosa o heptosa. En algunas realizaciones, el azúcar en el extremo reductor del oligosacárido o polisacárido es una aldopentosa o una cetopentosa. En algunas realizaciones, la pentosa es una D-arabinosa, una D-lixosa, una D-ribosa, una D-xilosa, una D-ribulosa o una D-xilulosa. En algunas realizaciones, el azúcar en el extremo reductor del oligosacárido o polisacárido es una aldohexosa o una cetoheptosa. En algunas realizaciones, la hexosa es, por ejemplo, una D-alosa, D-altrosa, D-glucosa, D-manosa, D-gulosa, D-idosa, D-galactosa, D-talosa, D-psicosa, D-fructosa, D-sorbosa o D-tagatosa. En algunas realizaciones, el azúcar en el extremo reductor del oligosacárido o polisacárido es un azúcar desoxi o di-desoxi, tal como, por ejemplo, una ramnosa, una fucosa o una abequosa. En algunas realizaciones, el azúcar en el extremo reductor del oligosacárido o polisacárido es una aldohexosa o una cetoheptosa. En algunas realizaciones, la heptosa es una manohexulosa. En una realización preferida, el azúcar en el extremo reductor del oligosacárido de polisacárido es glucosa.

Los oligosacáridos y polisacáridos que se pueden enlazar a un residuo N de una proteína por las OST de PglB modificadas proporcionadas en la presente pueden ser aquellos que se encuentran normalmente en cualquier organismo, por ejemplo, un organismo procarionta o un organismo eucarionta. En algunas realizaciones, el oligosacárido o polisacárido es de un organismo patógeno, por ejemplo, un patógeno humano o un patógeno animal (por ejemplo, un animal de granja o una mascota). En algunas realizaciones, el oligosacárido o polisacárido es de un organismo bacteriano. En algunas realizaciones, el oligosacárido o polisacárido puede ser de *E. coli*, *Shigella sonnei*, *Shigella flexneri*, *Shigella dysenteriae*, *Salmonella* sp (por ejemplo, *S. enterica* subsp. *Enterica*, *S. enterica* subsp. *Salamae*, *S. enterica* subsp. *arizonae*, *S. enterica* subsp. *Diarizonae*, *S. enterica* subsp. *Houtenae*, *S. bongori*, y *S. enterica* subsp. *Indica*, *Pseudomonas* sp (*P. aeruginosa*), *Klebsiella* sp. (por ejemplo, *K. pneumonia*), *Acinetobacter*, *Chlamydia trachomatis*, *Vibrio cholera*, *Listeria* sp., por ejemplo, *L. monocytogenes*, *Legionella pneumophila*, *Bordetella parapertussis*, *Burkholderia mallei* and *pseudomallei*, *Francisella tularensis*, *Campylobacter* sp. (*C. jejuni*); *Clostridium difficile*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *E. coli*, *Streptococcus agalacticae*, *Neisseria meningitidis*, *Candida albicans*, *Haemophilus influenza*, *Enterococcus faecalis*, *Borrelia burgdorferi*, *Neisseria gonorrhoea*, *Haemophilus influenza*, *Leishmania major*.

En algunas realizaciones, el oligosacárido o polisacárido comprende un antígeno, por ejemplo, un epítipo que es inmunogénico en un humano o un animal (por ejemplo, un animal de granja o una mascota). En algunas realizaciones, el oligosacárido o el polisacárido comprende un antígeno O de *E. coli* (por ejemplo, O1, O2, O3, O4, O5, O6, O7, O8, O9, O10, O11, O12, O13, O14, O15, O16, O17, O18, O19, O20, O21, O22, O23, O24, O25, O26, O27, O28, O29, O30, O32, O33, O34, O35, O36, O37, O38, O39, O40, O41, O42, O43, O44, O45, O46, O48, O49, O50, O51, O52, O53, O54, O55, O56, O57, O58, O59, O60, O61, O62, O63, O64, O65, O66, O68, O69, O70, O71, O73, O74, O75, O76, O77, O78, O79, O80, O81, O82, O83, O84, O85, O86, O87, O88, O89, O90, O91, O92, O93, O95, O96, O97, O98, O99, O100, O101, O102, O103, O104, O105, O106, O107, O108, O109, O110, O111, O112, O113, O114, O115, O116, O117, O118, O119, O120, O121, O123, O124, O125, O126, O127, O128, O129, O130, O131, O132, O133, O134, O135, O136, O137, O138, O139, O140, O141, O142, O143, O144, O145, O146, O147, O148, O149, O150, O151, O152, O153, O154, O155, O156, O157, O158, O159, O160, O161, O162, O163, O164, O165, O166, O167, O168, O169, O170, O171, O172, O173, O174, O175, O176, O177, O178, O179, O180, O181, O182, O183, O184, O185, O186, O187, *Shigella flexneri* O1A, O1B, O2A, O3A, O6, antígeno O de *Shigella sonnei*, *Shigella dysenteriae* O1, *Salmonella* sp (antígeno de *S. enterica* subsp. *Enterica*, *S. enterica* subsp. *Salamae*, *S. enterica* subsp. *arizonae*, *S. enterica* subsp. *diarizonae*, *S. enterica* subsp. *houtenae*, *S. bongori*, o *S. enterica* subsp. *indica* y O tipos 1-67, como se detalla en [44], *Pseudomonas* sp. (O de *P. aeruginosa* serotipos 1-20 [45]), *Klebsiella* sp. (por ejemplo, *K. pneumonia* serotipos O1, O2 (y subserotipos), O3, O4, O5, O6, O7, O8, O9, O10, O11, O12, [46]), antígenos O de *Acinetobacter* (por ejemplo, antígenos O de *A. baumannii* identificados en [47]), antígenos O de *Chlamydia trachomatis* (serotipos A, B, C, D, E, F, G, H, I, J, K, L1, L2, L3), antígenos O de *Vibrio cholera* O1 a

- 155, *Listeria* sp., en particular *L. monocytogenes* tipo 1, 2, 3, 4 y subserotipos de la misma, antígenos O de *Legionella pneumophila* serotipos 1 a 15, antígenos O de *Bordetella parapertussis*, antígenos O de *Burkholderia mallei* y *pseudomallei*, *Francisella tularensis*, *Campylobacter* sp. (*C. jejuni*); polisacáridos capsulares de *Clostridium difficile* (serotipos A, G, H, K, S1, S4, D, Cd-5, K Toma *et al.* 1988, y *C. perfringens* serotipos A, B, C, D o E), sacáridos capsulares de *Staphylococcus aureus* de serotipo 5 o serotipo 8, sacáridos capsulares de *Streptococcus pneumoniae* del serotipo 1, 2, 3, 4, 5, 6A, 6B, 6C, 7F, 8, 9A, 9L, 9N, 9V, 10A, 11A, 12F, 14, 15A, 15B, 16F, 18C, 19A, 19F, 22F, 23F, 33F, 35B, *Streptococcus pyogenes* (polisacáridos de serotipo capsular de estreptococo de grupo B del serotipo Ia, Ib, II, III, IV, V, VI, VII, o VIII), *Streptococcus agalacticae* (polisacáridos capsulares de estreptococo de grupo A), *Neisseria meningitidis* (serogrupos A, B, C, W, Y, X), *Candida albicans*, *Haemophilus influenzae*, polisacáridos capsulares de *Enterococcus faecalis* tipo I-V; y otras estructuras de polisacáridos de superficie, *por ejemplo*, los glucolípidos de *Borrelia burgdorferi*, pilina O glicano y lipooligosacárido (LOS) de *Neisseria meningitidis*, LOS de *Haemophilus influenzae*, lipofosfoglicano de *Leishmania major*, antígenos de carbohidratos asociados a tumores (glicosilfosfatidilinositol de malaria, arabinomanano de *Mycobacterium tuberculosis*).
- 15 En algunas realizaciones, el oligosacárido o polisacárido es un polisacárido de *Staphylococcus aureus* (*S. aureus*) o de *Salmonella enterica* sv. (*S. enterica* sv.). En algunas realizaciones, el polisacárido es un CP5 o CP8 de *S. aureus* o un Polisacárido de *S. enterica* sv. Typhimurium LT2.
- 20 En algunas realizaciones, el oligosacárido o polisacárido comprende un azúcar N-acetilo en el extremo reductor. En algunas realizaciones, el oligosacárido o polisacárido que comprende el azúcar N-acetilo en el extremo reductor puede comprender, *por ejemplo*, un antígeno O de *E. coli* (*por ejemplo*, O1, O2, O3, O4, O5, O6, O7, O8, O9, O10, O11, O12, O13, O14, O15, O16, O17, O18, O19, , O21, O22, O23, O24, O25, O26, O27, O28, O29, O30, O32, O33, O34, O35, O36, O37, O38, O39, O40, O41, O42, O43, O44, O45, O46, O48, O49, O50, O51, O52, O53, O54, O55, O56, O57, O58, O59, O60, O61, O62, O63, O64, O65, O66, O68, O69, O70, O71, O73, O74, O75, O76, O77, O78, O79, O80, O81, O82, O83, O84, O85, O86, O87, O88, O89, O90, O91, O92, O93, O95, O96, , O98, O99, O100, O101, O102, O103, O104, O105, O106, O107, O108, O109, O110, O111, O112, O113, O114, O115, O116, O117, O118, O119, O120, O121, O123, O124, O125, O126, O127, O128, O129, O130, O131, O132, O133, O134, O135, O136, O137, O138, O139, O140, O141, O142, O143, O144, O145, O146, O147, O148, O149, O150, O151, O152, O153, O154, O155, O156, O157, O158, O159, O160, O161, O162, O163, O164, O165, O166, O167, O168, O169, O170, O171, O172, O173, O174, O175, O176, O177, O178, O179, O180, O181, O182, O183, O184, O185, O186, O187), un polisacárido capsular de *Staphylococcus aureus* (*S. aureus*) (*por ejemplo*, CP5 o CP8), un polisacárido capsular de *Francisella tularensis* Schu4, un polisacárido capsular de *S. pneumoniae* capsules (*por ejemplo*, CP1, 4, 5, 12, 25, 38, 44, 45 o 46), un O glicano de pilina de *Neisseria meningitidis*, un antígeno O de *Burkholderia mallei* y *pseudomallei*, un antígeno O de *Bordetella parapertussis*, un antígeno O de *Legionella pneumophila* serotipos 1 a 15, un antígeno O de *Listeria* sp., en particular un antígeno O de *L. monocytogenes* tipo 1, 2, 3, 4, un antígeno O de *Pseudomonas* sp. (O de *P. aeruginosa* serotipos 1-20), un antígeno O de *Klebsiella* sp. (*por ejemplo*, *K. pneumoniae* serotipos O1, O2 (y subserotipos), O3, O4, O5, O6, O7, O8, O9, O10, O11, O12), un antígeno O de *Shigella* sp. (*por ejemplo*, *S. dysenteriae*, *S. sonnei*, *S. flexneri*, *S. boydii*), un antígeno O de *Acinetobacter* (*por ejemplo*, antígenos O de *A. baumannii*, o un antígeno O de *Listeria* sp.).
- 30
- 35
- 40 Los azúcares N-acetilo pueden comprender un sustituyente de amino-acetilo (N-acetilo) en uno o más átomos de carbono del azúcar. *Por ejemplo*, un azúcar N-acetilo puede comprender un sustituyente de N-acetilo en el átomo C2 de una unidad de monosacárido, tal como una unidad de glucosa (N-acetilglucosamina).
- 45 En algunas realizaciones, el oligosacárido o polisacárido comprende un azúcar en el extremo reductor que no está N-acetilado. En algunas realizaciones, el oligosacárido o polisacárido que comprende el azúcar no N-acetilado en el extremo reductor puede comprender, *por ejemplo*, *E. coli* O20, un antígeno de *Salmonella* sp (*por ejemplo*, *S. enterica* subsp. *Enterica*, *S. enterica* subsp. *Salamae*, *S. enterica* subsp. *arizonae*, *S. enterica* subsp. *diarizonae*, *S. enterica* subsp. *houtenae*, *S. bongori* o *S. enterica* subsp. *Indica* o *S. Typhi*), un antígeno O de tipo 1-67, un polisacárido capsular de *Streptococcus* de grupo A (*S. pyogenes*), *Streptococcus* de grupo B y de los serotipos CPS de *S. pneumoniae* (*que codifican para wchA, wcjG o wcjH en sus cúmulos de genes capsulares, es decir, todos los serotipos excepto CP1, 4, 5, 12, 25, 38, 44, 45, 46*) o un antígeno O de *Salmonella enterica* sv. (*S. enterica* sv.).
- 50
- 55 En algunas realizaciones, el oligosacárido o polisacárido comprende un polisacárido de Typhimurium LT2 CP5 o CP8 de *S. aureus* o *S. enterica* sv., un antígeno O de *Vibrio cholera* (*por ejemplo*, O1 a 155), o un antígeno O de *Listeria* sp. (*por ejemplo*, *L. monocytogenes* tipo 1, 2, 3, 4).
- 60 En algunas realizaciones, el oligosacárido o polisacárido comprende un residuo de D-N-acetilfucosamina (D-FucNAc) en su extremo reductor, tal como, *por ejemplo*, polisacáridos capsulares de *S. aureus* serotipos 5, 8 o antígeno O de *P. aeruginosa* serotipos O2, O5, O11, O16.
- En algunas realizaciones, el oligosacárido o polisacárido comprende un residuo de 4-amino-d-N-acetilfucosamina (D-FucNAc4N) en su extremo reductor, tal como, *por ejemplo*, ciertos oligosacáridos o polisacáridos de *S. pneumoniae*, como serotipo 1, antígeno O de *Shigella sonnei* o *Plesiomonas shigelloides* O17.

En algunas realizaciones, el oligosacárido o polisacárido comprende un residuo de D-N-acetilquinosamina (D-QuiNAc) en su extremo reductor, tal como, por ejemplo, como antígeno O de *P. aeruginosa* serotipos O6, O1 o *Francisella tularensis* serotipo Schu4.

- 5 En algunas realizaciones, el oligosacárido o polisacárido comprende un residuo de galactosa en su extremo reductor, tal como, por ejemplo, *S. enterica* LT2.

En algunas realizaciones, el oligosacárido o polisacárido comprende un polisacárido capsular de *S. pneumoniae* serotipo 5, *E. coli* O1, O2, *Cronobacter sakazakii* O5, es decir, poli- y oligosacárido con un extremo reductor D-GlcNAc enlazado a 1-4 a una L-ramnosa en configuración beta.

Proteínas

15 La proteína que se va a glucosilar por una OST de PglB modificada de la invención es una proteína que contiene al menos un residuo N situado en una secuencia de consenso de glucosilación D/E-Z₁-N-Z₂-S/T (Z₁ y Z₂ ≠ P). En algunas realizaciones, la proteína es una proteína portadora. La actividad de OST de PglB modificada es capaz de unir covalentemente un inmunógeno de sacárido a la proteína portadora. La proteína portadora proporciona epítomos de células T que permiten que el inmunógeno de sacárido unido genere una respuesta inmunitaria dependiente de T. En algunas realizaciones, la proteína es una proteína biofarmacéutica, a la cual la OST de PglB modificada de la invención 20 adiciona al menos un sacárido para lograr la glucosilación correcta o plegamiento correcto o estabilidad incrementada de la proteína biofarmacéutica. En una realización, la proteína biofarmacéutica es un anticuerpo monoclonal, un fragmento de un anticuerpo monoclonal capaz de unirse a un antígeno. En una realización, la proteína biofarmacéutica es una eritropoyetina, hormona de crecimiento, insulina humana, factor VIII, factor IX, activador de plasminógeno tisular, glucagón, gonadotropina, factor estimulante de colonias, interferón α, β o γ, una interleucina, por ejemplo, interleucina 2 o factor de necrosis tumoral.

Las proteínas portadoras se pueden enlazar a oligosacáridos o polisacáridos por las OST de PglB modificadas proporcionadas en la presente.

30 La proteína portadora puede ser cualquier proteína portadora natural (del mismo organismo que la OST de PglB) o cualquier proteína portadora heteróloga (de un organismo diferente a OST de PglB). En algunas realizaciones, la proteína portadora es un inmunógeno. Las proteínas portadoras pueden ser proteínas de longitud completa o fragmentos de las mismas. Los ejemplos de proteínas portadoras comprenden, sin limitación, exotoxina A de *P. aeruginosa* (EPA), CRM197, toxoide diftérico, toxoide tetánico, hemolisina A destoxificada de *S. aureus*, factor de aglutinación A, factor de aglutinación B, *E. coli* FimH, *E. coli* FimHC, enterotoxina termolábil de *E. coli*, variantes destoxificadas de enterotoxina termolábil 35 de *E. coli*, subunidad de toxina B del cólera (CTB), toxina del cólera, variantes destoxificadas de toxina del cólera, proteína sat de *E. coli*, el dominio pasajero de proteína sat de *E. coli*, AcrA de *C. jejuni* y glucoproteínas naturales de *C. jejuni*. En algunas realizaciones, la proteína portadora es exotoxina A de *P. aeruginosa* (EPA).

40 En algunas realizaciones, las proteínas portadoras N-glucosiladas por una OST de PglB modificada descrita en la presente se modifican, por ejemplo, se modifican de tal manera que la proteína es menos tóxica y/o más susceptible a la glucosilación, etc. En algunas realizaciones, las proteínas portadoras se modifican de modo que el número de sitios de glucosilación en las proteínas portadoras se incrementa al máximo de una manera que permite que se administren concentraciones más bajas de la proteína, por ejemplo, en una composición inmunogénica, en su forma de bioconjugado. 45 Por consiguiente, en ciertas realizaciones, las proteínas portadoras descritas en la presente se modifican para comprender 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más sitios de glucosilación que normalmente se asociarían con la proteína portadora (por ejemplo, con respecto al número de sitios de glucosilación asociados con la proteína portadora en su estado nativo/natural, por ejemplo, "tipo silvestre"). En algunas realizaciones, la introducción de sitios de glucosilación se logra por la inserción de secuencias de consenso de glucosilación (por ejemplo, (i) la secuencia de consenso Asn-Z-Ser (Thr), donde Z se selecciona independientemente de cualquier aminoácido excepto Pro; o (ii) la secuencia de consenso D/E-Z₁-N-Z₂-S/T, donde Z₁ y Z₂ se seleccionan independientemente de cualquier aminoácido excepto Pro) en cualquier parte en la estructura primaria de la proteína. La introducción de estos sitios de glucosilación se puede lograr, por ejemplo, al 50 adicionar nuevos aminoácidos a la estructura primaria de la proteína (los sitios de glucosilación se adicionan, en su totalidad o en parte), o al modificar los aminoácidos existentes en la proteína a fin de generar los sitios de glucosilación (los aminoácidos no se adicionan a la proteína, pero los aminoácidos seleccionados de la proteína se mutan para formar 55 sitios de glucosilación). Aquellos expertos en la técnica reconocerán que la secuencia de aminoácidos de una proteína se puede modificar fácilmente usando enfoques conocidos en la técnica, por ejemplo, enfoques modificados que comprenden la modificación de la secuencia de ácido nucleico que codifica para la proteína. En realizaciones específicas, las secuencias de consenso de glucosilación se introducen en regiones específicas de la proteína portadora, por ejemplo, estructuras de superficie de la proteína, en los extremos N-terminal o C-terminal de la proteína, y/o en bucles que se estabilizan por puentes de disulfuro en la base de la proteína. En ciertas realizaciones, la secuencia de consenso de glucosilación de 5 aminoácidos clásica se puede extender por residuos de lisina para una glucosilación más eficiente y por lo tanto, la secuencia de consenso insertada puede codificar para 5, 6 o 7 aminoácidos que se deben insertar o que reemplazan los aminoácidos de proteína aceptora.

Las OST de PglB pueden comprender cualquier N-OST divulgada en la presente.

5 En algunas realizaciones, las proteínas portadoras comprenden una "etiqueta", una secuencia de aminoácidos que permite el aislamiento y/o identificación de la proteína portadora. Por ejemplo, adicionar una etiqueta a una proteína portadora descrita en la presente puede ser útil en la purificación de esa proteína y por lo tanto, la purificación de vacunas conjugadas que comprenden la proteína portadora etiquetada. Las etiquetas de ejemplo que se pueden usar en la presente comprenden, sin limitación, etiquetas de histidina (HIS) (por ejemplo, etiqueta de histidina hexa o etiqueta 6XHis), etiquetas FLAG-TAG y HA. En ciertas realizaciones, las etiquetas usadas en la presente son removibles, por
10 ejemplo, remoción por agentes químicos o por medios enzimáticos, una vez que ya no son necesarias, por ejemplo, después de que la proteína se ha purificado.

Ácidos nucleicos

15 En otro aspecto, en la presente se proporcionan polinucleótidos que codifican para una oligosacariiltransferasa de PglB mutada de cualquier especie de *Campylobacter* que incluye *C. jejuni*, *C. lari* y *C. coli*, como se proporciona en la presente.

En algunas realizaciones, los ácidos nucleicos codifican para una PglB_{Cj} modificada donde uno o más de los aminoácidos 57, 63, 94, 101, 176, 191, 193, 233, 234, 286, 301, 319, 397, 402, 435, 446, 462, 479, 523, 532, 605, 610, 645, 676 o 695
20 se modifican, donde 57 se muta a R o K o T.

En una realización, el polinucleótido que codifica para OST de PglB tiene una subestación en el aminoácido que corresponde a 57 en SEQ ID NO: 1, por ejemplo, por un residuo de arginina (A57R). La sustitución 57 está en combinación con las sustituciones 311 y 77 o se sustituye independientemente. En una realización, el polinucleótido codifica para una
25 OST de PglB con sustituciones de N311V, Y77R y A57R.

En una realización, se proporciona un polinucleótido que codifica para una OST de PglB que contiene sustituciones en el residuo 462 o 479 o 462 y 479 de SEQ ID NO: 1, por ejemplo, Y462W o H479M o Y462W y H479M en combinación con una sustitución de A57R. Estas mutaciones se combinan opcionalmente con mutaciones que corresponden a los residuos
30 311, 77 y/o 57 de SEQ ID NO: 1, por ejemplo, N311V, Y77R y A57R. En una realización, se proporcionan ácidos nucleicos que codifican para OST de PglB con mutaciones en 57, 462 y 479, por ejemplo, A57R, Y462W y H479M. En una realización, se proporciona un polinucleótido que codifica para una OST de PglB con sustituciones en los residuos que corresponden a 77, 57, 462 y 479 de SEQ ID NO: 1, por ejemplo, Y77R, A57R, Y462W y H479M.

35 En algunas realizaciones, el polinucleótido que codifica para la OST de PglB que comprende modificaciones (por ejemplo, sustituciones de aminoácidos o sustituciones de nucleótidos en el polinucleótido que codifica para PglB) en, por ejemplo, dos o más, tres o más, cuatro o más, cinco o más, seis o más, siete o más, ocho o más, nueve o más, o diez o más aminoácidos, por ejemplo, 2-30, 3-25, 4-20, 5-20, 6-20, 7-20 o 10-20 sustituciones de aminoácidos. En una realización, las sustituciones están en las posiciones que codifican los siguientes residuos de SEQ ID NO:1:57; 57 y 77; 57, 77 y 311;
40 57 y 462; 57, 462 y 479; 57, 462, 479, 77 y 311; 57, 462, 479, 300, 301, 308 y 570; 57, 462, 479, 300, 301, 308, 570 y 77; 57, 462, 479, 300, 301, 308, 570, 77 y 311.

En una realización, las siguientes sustituciones están presentes en las posiciones que codifican los siguientes residuos de SEQ ID NO: 1 (donde / denota o):A57R/T; A57R/T y Y77R; A57R y Y77R; A57R/T, Y77R y N311V; A57R, Y77R y
45 N311V; A57R/T y Y462P/C/W/T/N; A57R /T y Y462W/T/N; A57R y Y462W; A57R/T, Y462P/C/W/T/N y H479M; A57R/T, Y462W/T/N y H479M; A57R, Y462W y H479M; A57R/T, Y462P/C/W/T/N, H479M, Y77R y N311V; A57R/T, Y462W/T/N, H479M, Y77R y N311V; A57R, Y462W, H479M, Y77R y N311V; A57R/T, Y462P/C/W/T/N, H479M, L301P/G/F y L570R/V; A57R, Y462W, H479M, L301P y L570R; A57R/T, Y462P/C/W/T/N, H479M, N300L, L301P/G/F, F308W/F y L570R/V; A57R, Y462W, H479M, N300L, L301P, F308W y L570R; A57R/T, Y462P/C/W/T/N, H479M, L301P/G/F, L570R/V y Y77R; A57R,
50 Y462W, H479M, L301P ,L570R y Y77R; A57R/T, Y462P/C/W/T/N, H479M, N300L, L301P/G/F, F308W/F, L570R/V y Y77R; A57R, Y462W, H479M, N300L, L301P, F308W, L570R y Y77R; A57R/T, Y462P/C/W/T/N, H479M, L301P/G/F, L570R/V, Y77R y N311V; A57R, Y462W, H479M, L301P ,L570R, Y77R y N311V; A57R/T, Y462P/C/W/T/N, H479M, N300L, L301P/G/F, F308W/F, L570R/V, Y77R y N311V; A57R, Y462W, H479M, N300L, L301P, F308W, L570R, Y77R y N311V.

55 Células hospederas

En otro aspecto, en la presente se proporciona una célula hospedadora que comprende una OST de PglB mutada o una OST de PglB de *C. coli* tipo silvestre o mutada proporcionada en la presente. En una realización, la célula hospedadora es una célula hospedadora heteróloga (por ejemplo, no *Campylobacter*). En una realización, la célula hospedadora es *A. coli*. En algunas realizaciones, la célula hospedadora comprende dos o más OST de PglB modificadas proporcionadas en la presente (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más OST de PglB modificadas).
60

En otro aspecto, en la presente se proporciona una célula hospedadora que comprende un ácido nucleico proporcionado en la presente (por ejemplo, que codifica para una OST de PglB modificada proporcionada en la presente). En algunas

realizaciones, la célula hospedadora comprende dos o más ácidos nucleicos proporcionados en la presente (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más ácidos nucleicos).

5 En algunas realizaciones, la célula hospedadora comprende una o más enzimas adicionales útiles para la producción de bioconjugado o N-glucosilación de proteínas (por ejemplo, una glicosiltransferasa). En algunas realizaciones, al menos una de las enzimas adicionales útiles para la producción de bioconjugado es una enzima recombinante. En algunas realizaciones, la célula hospedadora comprende dos o más enzimas adicionales útiles para la producción de bioconjugados (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más enzimas adicionales).

10 En algunas realizaciones, la célula hospedadora es una célula procariota. En algunas realizaciones, la célula hospedadora es una célula de *E. coli*. En algunas realizaciones, la célula hospedadora comprende una OST de PglB modificada proporcionada en la presente. En algunas realizaciones, la célula hospedadora comprende una proteína que contiene al menos una secuencia de consenso de glucosilación y una N-OST modificada proporcionada en la presente. En algunas realizaciones, la célula hospedadora comprende una proteína que contiene al menos una secuencia de consenso de glucosilación, una OST de PglB modificada proporcionada en la presente y una glicosiltransferasa recombinante. En algunas realizaciones, la OST de PglB modificada es una PglB de *C. jejuni* modificada.

20 En ciertas realizaciones, las células hospedadoras usadas para producir los bioconjugados descritos en la presente se modifican para comprender ácidos nucleicos heterólogos, por ejemplo, ácidos nucleicos heterólogos que codifican para una o más proteínas que contienen al menos una secuencia de consenso de glucosilación y/o ácidos nucleicos heterólogos que codifican para una o más proteínas, por ejemplo, genes que codifican para una o más enzimas. En algunas realizaciones, los ácidos nucleicos heterólogos que codifican para proteínas implicadas en las vías de glucosilación (por ejemplo, vías de glucosilación procariotas y/o eucariotas) se introducen en las células hospedadoras descritas en la presente. Estos ácidos nucleicos pueden codificar para proteínas que incluyen, sin limitación, oligosacaril transferasas y/o glicosiltransferasas. Los ácidos nucleicos heterólogos (por ejemplo, ácidos nucleicos que codifican para proteínas que contienen al menos una secuencia de consenso de glucosilación y/o ácidos nucleicos que codifican para otras proteínas, por ejemplo, proteínas implicadas en la glucosilación) se pueden introducir en las células hospedadoras descritas en la presente usando cualquier método conocido por aquellos expertos en la técnica, por ejemplo, electroporación, transformación química por choque térmico, transformación natural, transducción de fagos y conjugación.

30 En algunas realizaciones, los ácidos nucleicos heterólogos se introducen en las células hospedadoras descritas en la presente usando un plásmido, por ejemplo, los ácidos nucleicos heterólogos se expresan en las células hospedadoras por un plásmido (por ejemplo, un vector de expresión). En algunas realizaciones, los ácidos nucleicos heterólogos se introducen en las células hospedadoras descritas en la presente usando el método de inserción descrito en la publicación de solicitud de patente internacional No. WO 2014/057109.

35 En ciertas realizaciones, se pueden introducir modificaciones adicionales (por ejemplo, usando técnicas recombinantes) en las células hospedadoras descritas en la presente. Por ejemplo, los ácidos nucleicos de célula hospedadora (por ejemplo, genes) que codifican para proteínas que forman parte de una vía de glucosilación posiblemente competitiva o interferente (por ejemplo, competir o interferir con uno o más genes heterólogos implicados en la glucosilación que se introducen de forma recombinante en la célula hospedadora) se pueden eliminar o modificar en el fondo de célula hospedadora (genoma) de una manera que los hace inactivos/disfuncionales (es decir, los ácidos nucleicos de célula hospedadora que se eliminan/modifican no codifican para una proteína funcional o no codifican para una proteína en absoluto). En ciertas realizaciones, cuando los ácidos nucleicos se eliminan del genoma de las células hospedadoras proporcionadas en la presente, se reemplazan por una secuencia deseable, por ejemplo, una secuencia que es útil para la producción de glucoproteínas.

50 Los genes de ejemplo que se pueden eliminar en células hospedadoras (y, en algunos casos, reemplazar con otras secuencias de ácido nucleico deseadas) incluyen genes de células hospedadoras implicadas en la biosíntesis de glucolípidos, tal como *waaL* (ver, por ejemplo, Feldman et al., 2005, PNAS USA 102:3016-3021), el cúmulo de biosíntesis de núcleo de lípido A (*waa*), cúmulo de galactosa (*gal*), cúmulo de arabinosa (*ara*), cúmulo de ácido colánico (*wca*), cúmulo de polisacáridos capsulares, enzimas metabólicas implicadas en la biosíntesis de azúcar activada por nucleótidos, cúmulo de antígeno común enterobacteriano (*wec*) y cúmulos de modificación de antígeno profago O como el cúmulo *gtrABS*.

55 Las células hospedadoras descritas en la presente pueden producir las proteínas portadoras N-glucosiladas descritas en la presente. En algunas realizaciones, las proteínas portadoras N-glucosiladas producidas por las células hospedadoras descritas en la presente son antígenos, por ejemplo, antígenos virales o bacterianos que se pueden usar en vacunas. En algunas realizaciones, las proteínas portadoras N-glucosiladas producidas por las células hospedadoras descritas en la presente pueden ser cualquier proteína que contiene una secuencia de consenso de glucosilación descrita en la presente, donde estas proteínas se modifican por las células hospedadoras descritas en la presente para poseer una o más características beneficiosas, por ejemplo, la proteína está N-glucosilada.

60

Algunos de los ejemplos más adelante describen la aplicación de métodos descritos en la presente en células hospedadoras de *E. coli* Gram-negativas; sin embargo, cualquier célula hospedadora conocida por aquellos expertos en

la técnica se puede usar para producir proteínas portadoras N-glucosiladas, incluidas arqueas, células hospedadoras procariontas diferentes de *E. coli* y células hospedadoras eucariotas.

5 Los ejemplos de células hospedadoras procariontas que se pueden usar de acuerdo con los métodos descritos en la presente comprenden, sin limitación, especies de *Escherichia*, especies de *Shigella*, especies de *Klebsiella*, especies de *Xhantomonas*, especies de *Salmonella*, especies de *Yersinia*, especies de *Lactococcus*, especies de *Lactobacillus*, especies de *Pseudomonas*, especies de *Corynebacterium*, especies de *Streptomyces*, especies de *Streptococcus*, especies de *Staphylococcus*, especies de *Bacillus* y especies de *Clostridium*.

10 En ciertas realizaciones, las células hospedadoras descritas en la presente comprenden un genoma en el cual se ha introducido una o más secuencias de ADN, donde las secuencias de ADN codifican para una proteína o comprenden un cúmulo de operones/genes implicados en la N-glucosilación de proteínas. Por ejemplo, en algunas realizaciones, una célula hospedadora descrita en la presente comprende un genoma en el cual uno o más de los siguientes se han insertado:
 15 ADN que codifica para una OST de PglB modificada, ADN que codifica para una glucosiltransferasa, ADN que codifica para una proteína que contiene al menos una secuencia de glucosilación, ADN que comprende un cúmulo de genes *rtb*, ADN que comprende un cúmulo de genes de polisacárido capsular y/o ADN que codifica para una epimerasa.

Las células hospedadoras pueden incluir OST de PglB modificadas u OST de PglB de *C. coli* proporcionadas en la presente o ácidos nucleicos que codifican para las OST de PglB modificadas o las OST de PglB de *C. coli* proporcionadas
 20 en la presente, por lo que las OST de PglB modificadas pueden ser de cualquier organismo que tenga N-OST, incluido un organismo procarionta. En algunas realizaciones, la proteína OST de PglB o el ácido nucleico que codifica para OST de PglB es del género *Campylobacter* (por ejemplo, el gen *pglB* de *C. jejuni*).

Las células hospedadoras descritas en la presente pueden comprender una glucosiltransferasa conocida en la técnica o una secuencia de ácido nucleico que codifica para una glucosiltransferasa conocida en la técnica. En algunas realizaciones, la glucosiltransferasa es una glucosiltransferasa descrita en la publicación de solicitud de patente internacional No. WO 2011/138361, la divulgación de la cual se incorpora por referencia en la presente en su totalidad. En algunas realizaciones, la glucosiltransferasa es de una bacteria Gram-positiva, por ejemplo, la glucosiltransferasa es de *S. pneumoniae*, por ejemplo, de *S. pneumoniae* serotipo 1, 2, 3, 4, 5, 6A, 6B, 6C, 7F, 8, 9A, 9L, 9N, 9V, 10A, 11A, 12F, 14, 15A, 15B, 16F, 18C, 19A, 19F, 22F, 23F, 33F, 35B (preferentemente del serotipo 8), o de *Streptococcus pyogenes* (polisacáridos de serotipo capsular de estreptococo de grupo B del serotipo Ia, Ib, II, III, IV, V, VI, VII o VIII). En algunas realizaciones, la glucosiltransferasa es polisacárido capsular 5 de *S. aureus*. En algunas realizaciones, la glucosiltransferasa es polisacárido capsular 8 de *S. aureus*. En algunas realizaciones, la glucosiltransferasa es de una bacteria Gram-negativa, por ejemplo, *E. coli* o *Shigella flexneri* o *Shigella sonnei*.
 35

Las células hospedadoras descritas en la presente pueden comprender o producir una proteína que contiene al menos un sitio de consenso de glucosilación o comprender una secuencia de ácido nucleico que codifica para una proteína que contiene al menos una secuencia de consenso de glucosilación conocida en la técnica. Las proteínas producidas por las células hospedadoras descritas en la presente comprenden al menos una secuencia de consenso de N-glucosilación, por ejemplo, ya sea la secuencia de consenso (i) Asn-Z-Ser (Thr), donde Z se selecciona independientemente de cualquier aminoácido excepto Pro; o (ii) D/E-Z₁-N-Z₂-S/T, donde Z₁ y Z₂ se seleccionan independientemente de cualquier aminoácido excepto Pro. Por consiguiente, la célula hospedadora puede comprender secuencias de ADN que codifican para una secuencia de consenso de N-glucosilación. La célula hospedadora puede incluir cualquier proteína conocida en la técnica, incluidas las descritas en la presente. En algunas realizaciones, la proteína es una proteína portadora tal como una exotoxina A de *P. aeruginosa* (EPA), que incluye EPA que se ha modificado para comprender al menos una secuencia consenso de N-glucosilación. En algunas realizaciones, la proteína que es la proteína portadora es la toxina B de cólera. En algunas realizaciones, la proteína portadora es AcrA. En algunas realizaciones, la proteína portadora es HIA. En algunas realizaciones, la proteína portadora es CifA. En algunas realizaciones, la proteína portadora es CRM197.
 45

50 Bioconjugados

Los bioconjugados descritos en la presente son conjugados entre una proteína (por ejemplo, cualquier proteína portadora descrita en la presente) y un oligosacárido o un polisacárido (por ejemplo, cualquier oligosacárido o polisacárido descrito en la presente) preparados en una célula hospedadora, donde la maquinaria de célula hospedadora enlaza el oligosacárido o polisacárido a la proteína (por ejemplo, enlaces N). En algunas realizaciones, el oligosacárido o polisacárido es un antígeno (por ejemplo, cualquier antígeno descrito en la presente). Los glucoconjugados pueden incluir bioconjugados, así como conjugados de antígeno de azúcar (por ejemplo, oligo y polisacáridos)-proteína preparados por otros medios, por ejemplo, por enlace químico de la proteína y antígeno de azúcar.
 55

Las OST de PglB modificadas o las OST de PglB de *C. coli* descritas en la presente se pueden introducir en una célula (por ejemplo, una célula de bacteria Gram negativa) para producir células hospedadoras que producen bioconjugados que comprenden una proteína portadora N-glucosilada. En algunas realizaciones, en la presente se proporcionan bioconjugados que incluyen una proteína portadora N-glucosilada con un antígeno (por ejemplo, un oligosacárido o un polisacárido) descrito en la presente. En algunas realizaciones, la proteína portadora es EPA. Los bioconjugados descritos
 60

en la presente pueden, por ejemplo y sin limitación, comprender cualquier proteína portadora descrita en la presente. Los bioconjugados descritos en la presente pueden, por ejemplo y sin limitación, comprender cualquier oligosacárido o polisacárido descrito en la presente.

5 En una realización, la proteína portadora glucosilada heteróloga es *Streptococcus pneumoniae* serotipo 12F-EPA, *Streptococcus pneumoniae* serotipo 8-EPA, *S. pneumoniae* serotipo 19A-EPA, *S. pneumoniae* serotipo 22F-EPA, *S. pneumoniae* serotipo 23A-EPA o *S. pneumoniae* serotipo 35B-EPA. En una realización, la proteína portadora glucosilada heteróloga es un sacárido capsular de *S. pneumoniae* o un sacárido capsular de estreptococo de grupo B unido a una proteína portadora, por ejemplo, EPA, CRM197, DT o TT.

10 En una realización, la proteína portadora glucosilada heteróloga es polisacárido capsular tipo 5 de *S. aureus* conjugado con EPA, CRM197, DT, TT o una proteína estafilocócica tal como ClfA o Hla, o un polisacárido capsular tipo 8 de *S. aureus* conjugado con EPA, CRM197, DT, TT o una proteína estafilocócica tal como ClfA o Hla o *S. sonnei*-EPA.

15 En algunas realizaciones, en la presente se proporciona un bioconjugado que incluye EPA y uno o más oligosacáridos o polisacáridos diferentes descritos en la presente.

20 En algunas realizaciones, en la presente se proporciona un bioconjugado que incluye una proteína portadora conjugada con uno o más de *E. coli* O1, O2, O4, O6, O7, O8, O11, O15, O16, O17, O18, O20, O22, O25, O73, O75 y/u O83. En algunas realizaciones, la proteína portadora es EPA.

En algunas realizaciones, en la presente se proporciona un bioconjugado que incluye una proteína portadora conjugada con uno o más polisacáridos de *P. aeruginosa* diferentes. En algunas realizaciones, la proteína portadora es EPA.

25 En algunas realizaciones, en la presente se proporciona un bioconjugado que comprende una proteína portadora conjugada con uno o más polisacáridos de *K. pneumoniae* diferentes. En una realización específica, la proteína portadora es EPA.

30 Métodos para producir un bioconjugado

En algunas realizaciones, las OST de PglB modificadas y PglB de *C. coli* proporcionadas en la presente se pueden usar para producir un bioconjugado proporcionado en la presente, tal como un glucoconjugado. En algunas realizaciones, las OST de PglB modificadas y PglB de *C. coli* proporcionadas en la presente se pueden usar para producir vacunas conjugadas, es decir, vacunas que contienen un oligosacárido o polisacárido y un antígeno proteico del patógeno contra el que se diseña la vacuna.

35 En otro aspecto, en la presente se proporciona un método para producir un bioconjugado que incluye cultivar una célula hospedadora proporcionada en la presente, en un medio de cultivo celular. En algunas realizaciones, la célula hospedadora comprende un ácido nucleico que codifica para una OST de PglB modificada o PglB de *C. coli* proporcionada en la presente. En algunas realizaciones, la célula hospedadora comprende un ácido nucleico que codifica para una proteína portadora descrita en la presente. En algunas realizaciones, la proteína portadora tiene una o más secuencias consenso de N-glucosilación. En algunas realizaciones, la célula hospedadora comprende un ácido nucleico que codifica para una glucosiltransferasa.

45 En algunas realizaciones, el bioconjugado es una proteína portadora N-glucosilada. La proteína portadora N-glucosilada puede comprender un componente de oligosacárido o polisacárido que incluye cualquier oligosacárido o polisacárido descrito en la presente. La proteína portadora N-glucosilada puede comprender cualquier proteína portadora descrita en la presente. En algunas realizaciones, el bioconjugado es un polipéptido N-glucosilado de *C. jejuni* natural (que incluye un componente de oligosacárido o polisacárido de *C. jejuni* y una proteína portadora de *C. jejuni*). En algunas realizaciones, el bioconjugado es un polipéptido glucosilado de *C. jejuni* heterólogo (que incluye un componente de polisacárido y/o una proteína portadora que no es de *C. jejuni*). En algunas realizaciones, el polipéptido glucosilado no tiene un azúcar N-acetilo en su extremo reductor. En algunas realizaciones, el polipéptido glucosilado tiene una glucosa o galactosa en su extremo reductor.

55 En algunas realizaciones, los métodos comprenden además purificar el bioconjugado del cultivo de células hospedadoras. Los métodos para purificar bioconjugados, tal como proteínas portadoras N-glucosiladas, de cultivos de células hospedadoras se conocen en la técnica. Véase, por ejemplo, Jan-Christer Janson, Protein Purification: Principles, High Resolution Methods, and Applications. Wiley; 3 edición (22 de marzo de 2011).

60 Métodos analíticos

Se pueden usar diferentes métodos para analizar las composiciones estructurales y las longitudes de cadena de azúcar de los bioconjugados o proteínas portadoras N-glucosiladas descritas en la presente.

En una realización, se puede usar hidrazinolisis para analizar glucanos. En primer lugar, los polisacáridos se liberan de sus portadores de proteína por incubación con hidrazina de acuerdo con las instrucciones del fabricante (Ludger Liberate Hydrazinolysis Glycan Release Kit, Oxfordshire, Reino Unido). La hidrazina nucleófila ataca el enlace glucosídico entre el polisacárido y la proteína portadora y permite la liberación de los glucanos unidos. Los grupos N-acetilo se pierden durante este tratamiento y se tienen que reconstituir por re-N-acetilación. Los glucanos libres se purifican en columnas de carbono y posteriormente se etiquetan en el extremo reductor con el fluoróforo 2-amino benzamida (Bigge JC, Patel TP, Bruce JA, Goulding PN, Charles SM, Parekh RB. Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid. *Anal Biochem.* 1995 Sep 20;230(2):229-38). Los polisacáridos etiquetados se separan en una columna GlycoSep-N (GL Sciences) de acuerdo con el protocolo de HPLC de Royle *et al.* (Royle L, Mattu TS, Hart E, Langridge JI, Merry AH, Murphy N, Harvey DJ, Dwek RA, Rudd PM. Una base de datos analítica y estructural proporciona una estrategia para secuenciar O-glucanos a partir de cantidades de microgramos de glucoproteínas. *Anal Biochem.* 2002 May 1;304(1):70-90). El cromatograma de fluorescencia resultante indica la longitud de polisacárido y el número de unidades de repetición. La información estructural se puede recopilar por recolección de picos individuales y posterior realización de análisis MS/MS. De este modo, se pudo confirmar la composición de monosacáridos y la secuencia de la unidad de repetición y adicionalmente, se pudo identificar la homogeneidad de la composición de polisacáridos. Los picos específicos de bajo peso molecular se pueden analizar por MALDI-MS/MS y el resultado se usa para confirmar la secuencia de glucano. Cada pico corresponde a un polímero que consiste en un cierto número de unidades de repetición y fragmentos de las mismas. Por lo tanto, el cromatograma permite medir la distribución de longitud de polímero. El tiempo de elución es una indicación de la longitud de polímero, la intensidad de fluorescencia se correlaciona con la abundancia molar para el polímero respectivo.

En otra realización, se puede usar SDS-PAGE o electroforesis en gel capilar para valorar glucanos y glucoconjugados. La longitud de polímero para los glucanos de antígeno O que se sintetizan aquí se define por el número de unidades de repetición que se montan linealmente. Esto significa que el patrón habitual tipo escalera es una consecuencia de diferentes números de unidades de repetición que componen el glucano. Por lo tanto, dos bandas una al lado de la otra en SDS PAGE u otras técnicas que se separan por tamaño difieren por una unidad de repetición individual. Estas diferencias discretas se explotan cuando se analizan las glucoproteínas para determinar el tamaño de glucanos: La proteína portadora no glucosilada y el glucoconjugado con diferentes longitudes de cadena polimérica se separan de acuerdo con sus movilidades electroforéticas. Se miden el primer número de unidades de repetición detectables (m) y el número de unidades de repetición promedio (n_{promedio}) presentes en un glucoconjugado. Estos parámetros se pueden usar para demostrar la consistencia de lote a lote o estabilidad de polisacárido.

En otra realización, se puede aplicar MS de masa alta y HPLC de exclusión por tamaño para medir el tamaño de los glucoconjugados completos.

En otra realización, se puede usar un ensayo de antrona-ácido sulfúrico para medir los rendimientos de polisacáridos (Leyva A, Quintana A, Sánchez M, Rodríguez EN, Cremata J, Sánchez JC. Rapid and sensitive anthrone-sulfuric acid assay in microplate format to quantify carbohydrate in biopharmaceutical products: method development and validation. *Biologicals.* 2008 Mar;36(2): 134-41. Epub 2007 Nov 26).

40 *Cambio en el uso de sitio de glucosilación*

Para mostrar que el uso de sitio en una proteína específica se cambia, se puede cuantificar el uso de sitio de glucosilación. Los métodos para hacerlo se listan más adelante.

45 LC-MS/MS de glucopéptido: los glucoconjugados se digieren con proteasas, y los péptidos se separan por un método cromatográfico adecuado (C18, HPLC de interacción hidrófila HILIC, columnas GlycoSepN, SE HPLC, AE HPLC), y los diferentes péptidos se identifican usando MS/MS Este método se puede usar con o sin acortamiento previo de cadena de azúcar por métodos químicos (degradación de Smith) o enzimáticos. La cuantificación de los picos de glucopéptido usando detección UV a 215 a 280 nm permite la determinación relativa del uso de sitio de glucosilación.

50 HPLC de exclusión por tamaño: El mayor uso del sitio de glucosilación se refleja en un tiempo de elución más temprano de una columna SE HPLC. Ver también (a).

55 *Homogeneidad*

La homogeneidad de glucoconjugado (la homogeneidad de los residuos de azúcar unidos) se puede valorar usando métodos que miden la longitud de glucano y el radio hidrodinámico.

60 *Beneficios*

Las OST de PglB modificadas y PglB de *C. coli* proporcionadas en la presente y los métodos proporcionados en la presente para usar la OST de PglB modificada o PglB de *C. coli* proporcionadas en la presente son de particular importancia y relevancia comercial, ya que permiten una fermentación rápida, de alto rendimiento, a gran escala y de bajo

costo de preparaciones altamente homogéneas de proteínas glucosiladas (incluido, por ejemplo, preparación de glucoconjugado o preparaciones de vacuna conjugada). Las OST de PglB modificadas y PglB de *C. coli* proporcionadas en la presente permiten una producción económicamente viable de proteínas glucosiladas comercial y terapéuticamente valiosas, tal como vacunas conjugadas. Además, las OST de PglB modificadas y PglB de *C. coli* de la invención proporcionan una N-glucosilación más eficiente de proteínas, particularmente en casos donde el glucano no se transfiere naturalmente por PglB tipo silvestre. Por ejemplo, cuando el oligosacárido que se va a transferir al residuo N de la proteína no tiene un residuo GlcNAc en el extremo reductor, por ejemplo, cuando el extremo reductor del oligosacárido es un residuo de glucosa. Se espera que la reproducibilidad y robustez de los métodos de producción de bioconjugados biotecnológicos que usan las OST de PglB modificadas proporcionadas en la presente contribuyan a una reducción de los costos de producción. En general, se cree que la homogeneidad de las vacunas conjugadas especialmente bioterapéuticas afecta la seguridad clínica de los productos farmacéuticos.

Abreviaturas y definiciones

15 Para los propósitos de las descripciones en la presente, las abreviaturas usadas para los aminoácidos genéticamente codificados son convencionales y son las siguientes en la tabla 1:

Tabla 1

AMINOÁCIDO	TRES LETRAS	UNA LETRA
ALANINA	ALA	A
ARGININA	ARG	R
ASPARAGINA	ASN	N
ASPARTATO	ASP	D
CISTEÍNA	CYS	C
GLUTAMATO	GLU	E
GLUTAMINA	GLN	Q
GLICINA	GLY	G
HISTIDINA	HIS	H
ISOLEUCINA	ILE	I
LEUCINA	LEU	L
LISINA	LYS	K
METIONINA	MET	M
FENILALANINA	PHE	F
PROLINA	PRO	P
SERINA	SER	S
TREONINA	THR	T
TRIPTÓFANO	TRP	W
TIROSINA	TYR	Y
VALINA	VAL	V

20 Cuando se usan las abreviaturas de tres letras, a menos que estén precedidas específicamente por una "L" o una "D" o sean claras a partir del contexto en el cual se usa la abreviatura, el aminoácido puede estar en la configuración L o D sobre el carbono α (C α). Por ejemplo, en tanto que "Ala" designa alanina sin especificar la configuración sobre el carbono α , "D-Ala" y "L-Ala" designan D-alanina y L-alanina, respectivamente. Cuando se usan las abreviaturas de una letra, las letras mayúsculas designan aminoácidos en la configuración L sobre el carbono α y las letras minúsculas designan aminoácidos en la configuración D sobre el carbono α . Por ejemplo, "A" designa L-alanina y "a" designa D-alanina. Cuando las secuencias peptídicas se presentan como una cadena de abreviaturas de una letra o tres letras (o mezclas de las mismas), las secuencias se presentan en la dirección N→C de acuerdo con la convención.

30 Los términos técnicos y científicos usados en las descripciones en la presente tendrán los significados comúnmente

entendidos por un experto en la técnica, a menos que se defina específicamente de otro modo. Por consiguiente, se propone que los siguientes términos tengan los siguientes significados. Todas las patentes de Estados Unidos y las solicitudes de patente de Estados Unidos publicadas, incluidas todas las secuencias divulgadas dentro de estas patentes y solicitudes de patente, referidas en la presente se incorporan expresamente por referencia.

5 "Aminoácido o residuo ácido" se refiere a un aminoácido o residuo hidrófilo que tiene una cadena lateral que exhibe un valor de pK menor de aproximadamente 6 cuando el aminoácido se incluye en un péptido o polipéptido. Los aminoácidos ácidos habitualmente tienen cadenas laterales negativamente cargadas a pH fisiológico debido a la pérdida de un ion hidrógeno. Genéticamente, los aminoácidos ácidos codificados incluyen L-Glu (E) y L-Asp (D).

10 "Aminoácido" o "residuo" como se usa en el contexto de los polipéptidos divulgados en la presente se refiere al monómero específico en una posición de secuencia (por ejemplo, P5 indica que el "aminoácido" o "residuo" en la posición 5 de SEQ ID NO: 2 es una prolina).

15 "Diferencia de aminoácidos" o "diferencia de residuos" se refiere a un cambio en el residuo en una posición especificada de una secuencia de polipéptidos cuando se compara con una secuencia de referencia. La posición de secuencia polipeptídica en la cual está presente un aminoácido o cambio de aminoácido particular ("diferencia de residuo") algunas veces se describe en la presente como "Xn" o "posición n", donde n se refiere a la posición de residuo con respecto a la secuencia de referencia.

20 Por ejemplo, una diferencia de residuos en la posición X8, donde la secuencia de referencia tiene una serina, se refiere a un cambio del residuo en la posición X8 a cualquier residuo diferente de serina. Como se divulga en la presente, una enzima puede incluir una o más diferencias de residuos con respecto a una secuencia de referencia, donde las múltiples diferencias de residuos habitualmente se indican mediante una lista de las posiciones especificadas donde se realizan cambios con respecto a la secuencia de referencia (por ejemplo, "una o más diferencias de residuos en comparación con SEQ ID NO: 1 en las siguientes posiciones de residuos: X27, X30, X35, X37, X57, X75, X103, X185, X207, X208, X271, X286, o X296.").

30 Una mutación de sustitución específica, que es un reemplazo del residuo específico en una secuencia de referencia con un residuo especificado diferente se puede denotar por la notación convencional "X(número) X", donde X es el identificador de una sola letra del residuo en la secuencia de referencia, "número" es la posición de residuo en la secuencia de referencia, y X' es el identificador de una sola letra de la sustitución de residuo en la secuencia modificada.

35 "Aminoácido o residuo alifático" se refiere a un aminoácido o residuo hidrófobo que tiene una cadena lateral de hidrocarburo alifático. Los aminoácidos alifáticos genéticamente codificados incluyen L-Ala (A), L-Val (V), L-Leu (L) y L-Ile (I).

40 "Aminoácido o residuo aromático" se refiere a un aminoácido o residuo hidrófilo o hidrófobo que tiene una cadena lateral que incluye al menos un anillo aromático o heteroaromático. Los aminoácidos aromáticos genéticamente codificados incluyen L-Phe (F), L-Tyr (Y) y L-Trp (W). Aunque debido al pKa de su átomo de nitrógeno heteroaromático L-His (H) algunas veces se clasifica como un residuo básico, o como un residuo aromático ya que su cadena lateral incluye un anillo heteroaromático, en la presente la histidina se clasifica como un residuo hidrófilo o como un "residuo restringido" (ver más adelante).

45 "Aminoácido o residuo básico" se refiere a un aminoácido o residuo hidrófilo que tiene una cadena lateral que exhibe un valor de pKa mayor de aproximadamente 6 cuando el aminoácido se incluye en un péptido o polipéptido. Los aminoácidos básicos habitualmente tienen cadenas laterales positivamente cargadas a pH fisiológico debido a la asociación con el ion hidronio. Los aminoácidos básicos genéticamente codificados incluyen L-Arg (R) y L-Lys (K).

50 "Optimizado por codones" se refiere a cambios en los codones del polinucleótido que codifica para una proteína con respecto a aquellos usados preferentemente en un organismo particular de modo que la proteína codificada se exprese de manera eficiente en el organismo de interés. Aunque el código genético es degenerado en el sentido de que la mayoría de los aminoácidos se representan por varios codones, llamados "sinónimos" o codones "sinónimos", se conoce bien que el uso de codones por organismos particulares no es aleatorio y está sesgado hacia tripletes de codones particulares. Este sesgo de uso de codones puede ser mayor con referencia a un gen dado, genes de función común u origen ancestral, proteínas altamente expresadas versus proteínas con bajo número de copias y las regiones codificadoras de proteínas agregadas del genoma de un organismo. En algunas realizaciones, los polinucleótidos que codifican para las enzimas PglB oligosacaryltransferasa se pueden optimizar por codones para producción óptima a partir del organismo hospedador seleccionado para expresión.

60 "Ventana de comparación" se refiere a un segmento conceptual de al menos aproximadamente 20 posiciones de nucleótidos o residuos de aminoácidos contiguos donde una secuencia se puede comparar con una secuencia de referencia de al menos 20 nucleótidos o aminoácidos contiguos y donde la porción de la secuencia en la ventana de comparación puede comprender adiciones o deleciones (es decir, separaciones) de 20 por ciento o menos en

comparación con la secuencia de referencia (que no comprende adiciones o deleciones) para alineación óptima de las dos secuencias. La ventana de comparación puede ser más larga que 20 residuos contiguos e incluye, opcionalmente, 30, 40, 50, 100 o ventanas más largas.

5 Las sustituciones o mutaciones de aminoácidos “conservadoras” se refieren a la intercambiabilidad de residuos que tienen cadenas laterales similares y por lo tanto, habitualmente implican la sustitución del aminoácido en el polipéptido con aminoácidos dentro de la misma clase de aminoácidos definida o similar. Sin embargo, como se usa en la presente, en algunas realizaciones, las mutaciones conservadoras no incluyen sustituciones de un residuo hidrófilo a hidrófilo, hidrófobo a hidrófobo, que contiene hidroxilo a que contiene hidroxilo, o pequeño a pequeño, si la mutación conservadora puede ser en cambio una sustitución de un residuo alifático a uno alifático, no polar a no polar, polar a polar, ácido a ácido, básico a básico, aromático a aromático, o restringido a restringido. Además, como se usa en la presente, A, V, L o I se pueden mutar de manera conservadora a otro residuo alifático o a otro residuo no polar. La tabla 2 a continuación muestra sustituciones conservadoras de ejemplo.

15 **Tabla 2**

Residuo	Posibles mutaciones conservadoras
A, L, V, I	Otros alifáticos (A, L, V, I) Otros no polares (A, L, V, I, G, M)
G, M	Otros no polares (A, L, V, I, G, M)
D, E	Otros ácidos (D, E)
K, R	Otros básicos (K, R)
P	ninguno
N, Q, S, T	Otros polares
H, Y, W, F	Otros aromáticos (H, Y, W, F)
C	Ninguno

“Aminoácido o residuo restringido” se refiere a un aminoácido o residuo que tiene una geometría restringida. En la presente, los residuos restringidos incluyen L-Pro (P) y L-His (H). La histidina tiene una geometría restringida debido a que tiene un anillo de imidazol relativamente pequeño. La prolina tiene una geometría restringida, debido a que también tiene un anillo de cinco miembros.

“Secuencia de control2 se define en la presente para incluir todos los componentes, que son necesarios o ventajosos para la expresión de un polinucleótido y/o polipéptido de la presente divulgación. Cada secuencia de control puede ser nativa o extraña al polinucleótido de interés. Estas secuencias de control incluyen, pero no se limitan a, un líder, secuencia de poliadenilación, secuencia de propéptido, promotor, secuencia de péptido de señal y terminador de transcripción.

“Que corresponde a”, “referencia a” o “con respecto a” cuando se usa en el contexto de la numeración de una secuencia de aminoácidos o polinucleótidos dada se refiere a la numeración de los residuos de una secuencia de referencia especificada cuando la secuencia de aminoácidos o polinucleótidos dada se compara con la secuencia de referencia. En otras palabras, el número de residuo o posición de residuo de un polímero dado se designa con respecto a la secuencia de referencia en lugar de por la posición numérica real del residuo dentro de la secuencia de aminoácidos o polinucleótidos dada de la secuencia no de referencia. Por ejemplo, una secuencia de aminoácidos dada, tal como la de una PglB oligosacaryltransferasa modificada, se puede alinear con una secuencia de referencia por la introducción de separaciones para optimizar las coincidencias de residuos entre las dos secuencias. En estos casos, aunque las separaciones están presentes, la numeración del residuo en la secuencia de aminoácidos o polinucleótidos dada se realiza con respecto a la secuencia de referencia con la cual se ha alineado.

“Cisteína” o L-Cys (C) es inusual ya que puede formar puentes de disulfuro con otros aminoácidos L-Cys (C) u otros aminoácidos que contienen sulfanilo o sulfhidrilo. Los “residuos tipo cisteína” incluyen cisteína y otros aminoácidos que contienen porciones de sulfhidrilo que están disponibles para formación de puentes de disulfuro. La capacidad de L-Cys (C) (y otros aminoácidos con cadenas laterales que contienen -SH) para existir en un péptido en la forma de -SH libre reducida o con puente de disulfuro oxidado afecta si L-Cys (C) contribuye con un carácter hidrófobo o hidrófilo neto a un péptido. En tanto que L-Cys (C) exhibe una hidrofobicidad de 0,29 de acuerdo con la escala de consenso normalizada de Eisenberg (Eisenberg *et al.*, 1984, *supra*), se va a entender que para los propósitos de la presente divulgación L-Cys (C) se clasifica en su propio grupo único.

“Deleción” se refiere a la modificación del polipéptido por la remoción de uno o más aminoácidos del polipéptido de referencia. Las deleciones pueden comprender la remoción de 1 o más aminoácidos, 2 o más aminoácidos, 5 o más

5 aminoácidos, 10 o más aminoácidos, 15 o más aminoácidos, o 20 o más aminoácidos, hasta 10 % del número total de aminoácidos, hasta 20 % del número total de aminoácidos, o hasta 30 % del número total de aminoácidos que componen el polipéptido en tanto que se retiene la actividad enzimática y/o se retienen las propiedades mejoradas de una enzima PglB oligosacariltransferasa modificada. Las deleciones se pueden dirigir a las porciones internas y/o porciones terminales del polipéptido. En diferentes realizaciones, la deleción puede comprender un segmento continuo o puede ser discontinua.

10 "Derivado de", como se usa en la presente en el contexto de enzimas modificadas, identifica la enzima de origen, y/o el gen que codifica para la enzima, en el cual se basó la modificación. Por ejemplo, la enzima oligosacariltransferasa modificada de SEQ ID NO: 4 se obtuvo al mutar la PglB oligosacariltransferasa de SEQ ID NO: 2. Por lo tanto, esta enzima PglB oligosacariltransferasa modificada de SEQ ID NO: 4 se "deriva del" polipéptido de SEQ ID NO: 2

15 Una "PglB oligosacariltransferasa modificada", como se usa en la presente, se refiere a una proteína de tipo PglB oligosacariltransferasa que se ha modificado sistemáticamente, a través de la inserción de nuevos aminoácidos en su secuencia de referencia, la deleción de aminoácidos presentes en su secuencia de referencia, o la mutación de aminoácidos en su secuencia de referencia en aminoácidos alternativos, ya sea a través de un proceso de mutagénesis aleatoria seguido por la selección de mutantes que tienen una propiedad particular o a través de la introducción intencional de cambios de aminoácidos particulares en la secuencia de proteína.

20 "Fragmento", como se usa en la presente, se refiere a un polipéptido que tiene una deleción amino-terminal y/o carboxi-terminal, pero donde la secuencia de aminoácidos restante es idéntica a las posiciones correspondientes en la secuencia. Los fragmentos pueden tener al menos 14 aminoácidos de longitud, al menos 20 aminoácidos de longitud, al menos 50 aminoácidos de longitud o más, y hasta 70 %, 80 %, 90 %, 95 %, 98 % y 99 %, o más, del polipéptido de PglB oligosacariltransferasa de longitud completa, por ejemplo, el polipéptido de SEQ ID NO: 4.

25 Un "fragmento funcional" o un "fragmento biológicamente activo", usado indistintamente, en la presente se refiere a un polipéptido que tiene una o más deleciones amino-terminales y/o carboxi-terminales y/o deleciones internas, pero donde la secuencia de aminoácidos restante es idéntica a las posiciones correspondientes en la secuencia con la cual se compara y que retiene sustancialmente toda la actividad del polipéptido de longitud completa. Un fragmento funcional contiene al menos 100, 200, 300, 400 o 500 aminoácidos contiguos

30 El polinucleótido "heterólogo" se refiere a cualquier polinucleótido que se introduce en una célula hospedadora por técnicas de laboratorio, e incluye polinucleótidos que se remueven de una célula hospedadora, se someten a manipulación de laboratorio y entonces se reintroducen en una célula hospedadora.

35 "Rigurosidad de hibridación" se refiere a condiciones de hibridación, tal como condiciones de lavado, en la hibridación de ácidos nucleicos. en general, las reacciones de hibridación se realizan bajo condiciones de menor rigor, seguido de lavados de rigor variable pero mayor. El término "hibridación moderadamente rigurosa" se refiere a condiciones que permiten que el ADN diana se una a un ácido nucleico complementario que tiene aproximadamente 60 % de identidad, preferentemente, aproximadamente 75 % de identidad, aproximadamente 85 % de identidad con el ADN diana, o con más de aproximadamente 90 % de identidad con el polinucleótido diana. Las condiciones moderadamente rigurosas de ejemplo son condiciones equivalentes a la hibridación en formamida al 50 %, 5x solución de Denhart, 5x solución salina-fosfato de sodio-EDTA (SSPE), dodecilsulfato de sodio (SDS) al 0,2 % a 42 °C, seguido de lavado en 0,2xSSPE, SDS al 0,2 %, a 42 °C. "Hibridación de alta rigurosidad" se refiere en general a condiciones que son de aproximadamente 10 °C o menos de la temperatura de fusión térmica T_m como se determina bajo la condición de solución para una secuencia de polinucleótidos definida. En algunas realizaciones, una condición de alta rigurosidad se refiere a condiciones que permiten la hibridación de solo aquellas secuencias de ácido nucleico que forman híbridos estables en NaCl 0,018 M a 65 °C (es decir, si un híbrido no es estable en NaCl 0,018 M a 65 °C, no será estable bajo condiciones de alta rigurosidad, como se contempla en la presente). Se pueden proporcionar condiciones de alta rigurosidad, por ejemplo, por hibridación en condiciones equivalentes a 50 % de formamida, 5x solución de Denhart, 5xSSPE, 0,2 % de SDS a 42 °C, seguido de lavado en 0,1x SSPE y 0,1 % de SDS a 65 °C. Otra condición de alta rigurosidad es la hibridación en condiciones equivalentes a la hibridación en 5X SSC que contiene SDS al 0,1 % (p:v) a 65°C y lavado en 0,1x SSC que contiene SDS al 0,1 % a 65 °C. Otras condiciones de hibridación de alta rigurosidad, así como condiciones moderadamente rigurosas, se describen en las referencias citadas anteriormente.

55 "Aminoácido o residuo hidrófilo" se refiere a un aminoácido o residuo que tiene una cadena lateral que exhibe una hidrofobicidad menor de cero de acuerdo con la escala de hidrofobicidad de consenso normalizada de Eisenberg et al., 1984, J. Mol. Biol. 179:125-142. Los aminoácidos hidrófilos genéticamente codificados incluyen L-Thr (T), L-Ser (S), L-His (H), L-Glu (E), L-Asn (N), L-Gln (Q), L-Asp (D), L-Lys (K) y L-Arg (R). "Aminoácido o residuo hidrófobo" se refiere a un aminoácido o residuo que tiene una cadena lateral que exhibe una hidrofobicidad mayor que cero de acuerdo con la escala de hidrofobicidad de consenso normalizada de Eisenberg et al., 1984, J. Mol. Biol. 179:125-142. Genéticamente, los aminoácidos hidrófobos codificados incluyen L-Pro (P), L-Ile (I), L-Phe (F), L-Val (V), L-Leu (L), L-Trp (W), L-Met (M), L-Ala (A) y L-Tyr (Y).

"Aminoácido o residuo que contiene hidroxilo" se refiere a un aminoácido que contiene una porción de hidroxilo (-OH). Los

aminoácidos que contienen hidroxilo genéticamente codificados incluyen L-Ser (S) L-Thr (T) y L-Tyr (Y).

5 “Propiedad enzimática mejorada” se refiere a cualquier propiedad enzimática que se hace mejor o más deseable para un propósito particular en comparación con la propiedad que se encuentra en una enzima de referencia. Para los polipéptidos de PglB oligosacilariltransferasa modificados descritos en la presente, la comparación se hace en general con una enzima PglB oligosacilariltransferasa de referencia que no contiene la mutación particular que mejora la eficiencia enzimática, aunque en algunas realizaciones, la PglB oligosacilariltransferasa de referencia puede ser otra PglB oligosacilariltransferasa modificada mejorada. Las propiedades enzimáticas para las cuales se puede realizar una mejora incluyen, pero no se limitan a, actividad enzimática (que se puede expresar en términos de rendimiento de proteína N-glucosilada), estabilidad
10 térmica, estabilidad de solvente, perfil de actividad de pH, requisitos de coenzima, refractariedad a inhibidores (por ejemplo, inhibición del producto), estereoespecificidad y supresión de la producción de productos secundarios ácidos.

15 “Inserción” se refiere a la modificación del polipéptido por la adición de uno o más aminoácidos al polipéptido de referencia. En algunas realizaciones, las enzimas PglB oligosacilariltransferasa modificadas mejoradas comprenden inserciones de uno o más aminoácidos al polipéptido PglB oligosacilariltransferasa que se presentan de manera natural, así como inserciones de uno o más aminoácidos a otros polipéptidos de PglB oligosacilariltransferasa mejorados. Las inserciones pueden ser en las porciones internas del polipéptido, o en el extremo carboxi o amino-terminal. Las inserciones como se usan en la presente incluyen proteínas de fusión como se conoce en la técnica. La inserción puede ser un segmento contiguo de aminoácidos o separado por uno o más de los aminoácidos en el polipéptido de origen natural.
20

25 “Polipéptido aislado” se refiere a un polipéptido que se separa sustancialmente de otros contaminantes que lo acompañan naturalmente, por ejemplo, proteína, lípidos y polinucleótidos. El término abarca polipéptidos que se han eliminado o purificado de su entorno o sistema de expresión de origen natural (por ejemplo, célula hospedadora o síntesis *in vitro*). Las enzimas PglB oligosacilariltransferasa mejoradas pueden estar presentes dentro de una célula, presentes en el medio celular o preparadas en diferentes formas, tal como productos lisados o preparaciones aisladas. Como tal, en algunas realizaciones, la enzima PglB oligosacilariltransferasa mejorada puede ser un polipéptido aislado.

30 “Sustitución no conservadora” se refiere a la sustitución o mutación de un aminoácido en el polipéptido con un aminoácido con propiedades de cadena lateral significativamente diferentes. Las sustituciones no conservadoras pueden usar aminoácidos entre, en lugar de dentro de, los grupos definidos listados anteriormente. En una realización, una mutación no conservadora afecta: (a) la estructura de la cadena principal del péptido en el área de la sustitución (por ejemplo, prolina por glicina); (b) la carga o hidrofobicidad; o (c) el volumen de la cadena lateral.

35 “Aminoácido no polar” o “Residuo no polar” se refiere a un aminoácido o residuo hidrófobo que tiene una cadena lateral que no está cargada a pH fisiológico y que tiene enlaces en los cuales el par de electrones compartidos en común por dos átomos en general se mantiene por igual por cada uno de los dos átomos (es decir, la cadena lateral no es polar). Los aminoácidos no polares genéticamente codificados incluyen L-Gly (G), L-Leu (L), L-Val (V), L-Ile (I), L-Met (M) y L-Ala (A).

40 “Porcentaje de identidad de secuencia”, “por ciento de identidad” y “por ciento idéntico” se usan en la presente para referirse a comparaciones entre secuencias de polinucleótidos o secuencias de polipéptidos, y se determinan al comparar dos secuencias alineadas de manera óptima a través de una ventana de comparación, donde la porción de la secuencia de polinucleótidos o polipéptidos en la ventana de comparación puede comprender adiciones o deleciones (es decir, separaciones) en comparación con la secuencia de referencia para alineación óptima de las dos secuencias. El porcentaje se calcula al determinar el número de posiciones en las cuales se presenta el residuo aminoácido o base de ácido nucleico idéntica en ambas secuencias o se alinea un residuo aminoácido o base de ácido nucleico con una separación para producir el número de posiciones coincidentes, al dividir el número de posiciones coincidentes entre el número total de posiciones en la ventana de comparación y al multiplicar el resultado por 100 para producir el porcentaje de identidad de secuencia. La determinación de la alineación óptima y el porcentaje de identidad de secuencia se realiza usando los algoritmos BLAST Y BLAST 2.0 (ver, por ejemplo, Altschul, et al., 1990, J. Mol. Biol. 215: 403-410 y Altschul, et al., 1977, Nucleic Acids Res. 3389-3402). El software para realizar análisis de BLAST está disponible públicamente a través del sitio web del Centro Nacional de Información Biotecnológica.
45
50

55 Brevemente, los análisis por BLAST implican primero identificar pares de secuencias de alto puntaje (HSP, por sus siglas en inglés) identificando palabras cortas de longitud W en la secuencia de consulta, que coinciden o satisfacen algún puntaje umbral de valor positivo T cuando se alinean con una palabra de la misma longitud en una secuencia de la base de datos. T se conoce como el umbral de puntaje de palabras cercanas (Altschul, *et al, supra*). Estos resultados iniciales de palabras del vecindario actúan como semillas para iniciar búsquedas para encontrar HSP más largos que los contengan. Por lo tanto, los aciertos de palabras se extienden en ambas direcciones a lo largo de cada secuencia en la medida en que el puntaje de alineamiento acumulativo puede ser aumentado. Los puntajes acumulativos se calculan utilizando, para las secuencias de nucleótidos, los parámetros M (puntaje de recompensa para un par de residuos coincidentes; siempre >0) y N (puntaje de penalización para residuos de apareamiento erróneo; siempre <0). Para las secuencias de aminoácidos, se utiliza una matriz de puntaje para calcular el puntaje acumulado. La extensión de los aciertos de palabra en cada dirección se detiene cuando: el puntaje de alineamiento acumulativo cae por la cantidad X de
60

su valor máximo alcanzado; el puntaje acumulado llega a cero o por debajo, debido a la acumulación de uno o más alineamientos de residuos de puntuación negativa; o se alcanza el extremo de cualquiera de las secuencias. Los parámetros del algoritmo de BLAST W, T y X determinan la sensibilidad y la velocidad del alineamiento. El programa BLASTN (para secuencias de nucleótidos) utiliza por defecto una longitud de palabra (W) de 11, una expectativa (E) o 10, M=5, N=4 y una comparación de ambas cadenas. Para secuencias de aminoácidos, el programa BLASTP usa por defecto una longitud de palabra (W) de 3, una expectativa (E) de 10 y la matriz de puntuación BLOSUM62 (ver Henikoff y Henikoff, 1989, Proc Natl Acad Sci USA 89:10915).

Hay muchos otros algoritmos disponibles que funcionan de manera similar a BLAST para proporcionar un por ciento de identidad para dos secuencias. La alineación óptima de secuencias para comparación se puede llevar a cabo, por ejemplo, por el algoritmo de homología local de Smith y Waterman, 1981, Adv. Appl. Math. 2:482, por el algoritmo de alineación de homología de Needleman and Wunsch, 1970, J. Mol. Biol. 48:443, por el método de búsqueda de similitud de Pearson y Lipman, 1988, Proc. Natl. Acad. Sci. USA 85:2444, por implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Paquete de Software GCG Wisconsin), o por inspección visual (ver en general, Current Protocols in Molecular Biology, F. M. Ausubel, et al., eds. Current Protocols, una empresa conjunta entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (1995 Supplement) (Ausubel)). Además, la determinación de la alineación de secuencias y el por de identidad de secuencia puede emplear los programas BESTFIT o GAP en el paquete de software GCG Wisconsin (Accelrys, Madison WI), utilizando los parámetros predeterminados proporcionados. El programa ClustalW también es adecuado para determinar la identidad.

“Aminoácido o residuo polar” se refiere a un aminoácido o residuo hidrófilo que tiene una cadena lateral que no está cargada a pH fisiológico, pero que tiene al menos un enlace en el que el par de electrones compartidos en común por dos átomos se mantiene más estrechamente por uno de los átomos. Los aminoácidos polares codificados genéticamente incluyen L-Asn (N), L-Gln (Q), L-Ser (S) y L-Thr (T).

“Codones de desviación de uso de codones altos, óptimos, preferidos” se refiere, indistintamente, a codones que se usan a mayor frecuencia en las regiones codificadoras de proteínas que otros codones que codifican para el mismo aminoácido. Los codones preferidos se pueden determinar en relación con el uso de codones en un gen individual, un conjunto de genes de función u origen común, genes altamente expresados, la frecuencia de codones en las regiones codificadoras de proteínas agregadas del organismo completo, la frecuencia de codones en las regiones codificadoras de proteínas agregadas de organismos relacionados, o combinaciones de los mismos. Los codones cuya frecuencia incrementa con el nivel de expresión génica son habitualmente codones óptimos para la expresión. Se conoce una variedad de métodos para determinar la frecuencia de codón (por ejemplo, uso de codón, uso de codón sinónimo relativo) y preferencia de codón en organismos específicos, incluido análisis multivariante, por ejemplo, usando análisis de grupo o análisis de correspondencia, y el número efectivo de codones usados en un gen (ver GCG CodonPreference, Genetics Computer Group Wisconsin Package; CodonW, John Peden, University of Nottingham; McInerney, J. O, 1998, Bioinformatics 14:372-73; Stenico, et al., 1994, Nucleic Acids Res. 22:2437-46; Wright, F., 1990, Gene 87:23-29). Las tablas de uso de codones están disponibles para una lista creciente de organismos (ver, por ejemplo, Wada, et al., 1992, Nucleic Acids Res. 20:2111-2118; Nakamura, et al., 2000, Nucl. Acids Res. 28:292; Duret, *et al.*, *supra*; Henaut and Danchin, "Escherichia coli and Salmonella," 1996, Neidhardt, et al., Eds., ASM Press, Washington D.C., p. 2047-2066. La fuente de datos para obtener el uso de codones puede basarse en cualquier secuencia de nucleótidos disponible capaz de codificar para una proteína. Estos conjuntos de datos incluyen secuencias de ácido nucleico que se sabe que codifican para proteínas expresadas (*por ejemplo*, secuencias codificadoras de proteínas completas-CDS), etiquetas de secuencia expresada (EST) o regiones codificadoras predichas de secuencias genómicas (ver, por ejemplo, Mount, D., Bioinformatics: Sequence and Genome Analysis, Capítulo 8, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001; Uberbacher, E. C., 1996, Methods Enzymol. 266:259-281; Tiwari, et al., 1997, Comput. Appl. Biosci. 13:263-270).

“Proteína”, “polipéptido” y “péptido” se usan indistintamente en la presente para indicar un polímero de al menos dos aminoácidos enlazados covalentemente por un enlace amida, independientemente de la longitud o modificación postraducciona (por ejemplo, glucosilación, fosforilación, lipidación, miristilación, ubiquitinación, etc.). Dentro de esta definición se incluyen D- y L-aminoácidos y mezclas de D- y L-aminoácidos.

“Secuencia de referencia” se refiere a una secuencia definida con la cual se compara otra secuencia (por ejemplo, alterada). Una secuencia de referencia puede ser un subconjunto de una secuencia más grande, por ejemplo, un segmento de una secuencia de polipéptido o gen de longitud completa. En general, una secuencia de referencia es de al menos 20 residuos de nucleótidos o aminoácidos de longitud, al menos 25 residuos de longitud, al menos 50 residuos de longitud, o la longitud completa del ácido nucleico o polipéptido. Debido a que cada uno de dos polinucleótidos o polipéptidos puede (1) comprender una secuencia (es decir, una porción de la secuencia completa) que es similar entre las dos secuencias y (2) puede comprender además una secuencia que es divergente entre las dos secuencias, las comparaciones de secuencia entre dos (o más) polinucleótidos o polipéptidos se realizan habitualmente al comparar secuencias de los dos polinucleótidos a través de una ventana de comparación para identificar y comparar regiones locales de similitud de secuencia.

El término “secuencia de referencia” no se propone que se limite a secuencias de tipo silvestre, y puede incluir secuencias

modificadas o alteradas. Por ejemplo, en algunas realizaciones, una "secuencia de referencia" puede ser una secuencia de aminoácidos modificada o alterada previamente. Por ejemplo, una "secuencia de referencia basada en SEQ ID NO: 2 que tiene un residuo de glicina en la posición X12" se refiere a una secuencia de referencia que corresponde a SEQ ID NO: 2 con un residuo de glicina en X12 (la versión inalterada de SEQ ID NO: 2 tiene un aspartato en X12).

"Aminoácido pequeño" o "residuo pequeño" se refiere a un aminoácido o residuo que tiene una cadena lateral que se compone de un total de tres o menos carbonos y/o heteroátomos (excluyendo el carbono α y los hidrógenos). Los aminoácidos o residuos pequeños se pueden clasificar adicionalmente como aminoácidos o residuos pequeños alifáticos, no polares, polares o ácidos, de acuerdo con las definiciones anteriores. Los aminoácidos pequeños genéticamente codificados incluyen L-Ala (A), L-Val (V), L-Cys (C), L-Asn (N), L-Ser (S), L-Thr (T) y L-Asp (D).

"Identidad sustancial" se refiere a una secuencia de polinucleótidos o polipéptidos que tiene al menos 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o más por ciento de identidad de secuencia, en comparación con una secuencia de referencia en una ventana de comparación de al menos 20 posiciones de residuos, frecuentemente en una ventana de al menos 30-50 residuos, donde el porcentaje de identidad de secuencia se calcula al comparar la secuencia de referencia con una secuencia que incluye deleciones o adiciones que totalizan 20 por ciento o menos de la secuencia de referencia en la ventana de comparación. En realizaciones específicas aplicadas a los polipéptidos, el término "identidad sustancial" significa que dos secuencias de polipéptidos, cuando se alinean de manera óptima, tal como por los programas GAP o BESTFIT usando ponderaciones de separación predeterminadas, comparten al menos 80 por ciento de identidad de secuencia, preferentemente al menos 89 por ciento de identidad de secuencia, al menos 95 por ciento de identidad de secuencia o más (por ejemplo, 99 por ciento de identidad de secuencia). Preferentemente, las posiciones de residuo que no son idénticas difieren por sustituciones de aminoácidos conservadoras.

"Polipéptido sustancialmente puro" se refiere a una composición en la cual la especie de polipéptido es la especie predominante presente (es decir, en una base molar o en peso es más abundante que cualquier otra especie macromolecular individual en la composición), y es en general una composición sustancialmente purificada cuando la especie objeto comprende al menos aproximadamente 50 por ciento de la especie macromolecular presente en moles o % en peso. En general, una composición de PglB oligosacariltransferasa sustancialmente pura comprenderá aproximadamente 60 % o más, aproximadamente 70 % o más, aproximadamente 80 % o más, aproximadamente 90 % o más, aproximadamente 95 % o más, aproximadamente 96 % o más, aproximadamente 97 % o más, aproximadamente 98 % o más o aproximadamente 99 % o más de todas las especies macromoleculares por mol o % en peso presentes en la composición. En algunas realizaciones, la presente especie se purifica hasta una homogeneidad esencial (es decir, las especies contaminantes no se pueden detectar en la composición mediante métodos de detección convencionales) donde la composición consiste esencialmente en una especie macromolecular individual. Las especies de solventes y las especies de iones elementales no se consideran especies macromoleculares. En algunas realizaciones, el polipéptido de PglB oligosacariltransferasa mejorado aislado es una composición de polipéptido sustancialmente pura.

Ejemplos

Ejemplo 1

Plásmidos y cepa usados para valorar la actividad de las variantes de PglB

La actividad de las variantes de PglB se probó en un derivado de *E. coli* W3110 con deleciones de genes como se describe en WO 14/057109A1. Los loci que codifican para polisacáridos capsulares se integraron de manera estable en el cromosoma de *E. coli* con base en los métodos detallados en WO2014057109A1. Otros elementos requeridos, incluida PglB, proteínas portadoras, enzimas biosintéticas suplementarias y proteínas reguladoras se expresaron de manera diversa a partir de los plásmidos pEXT21 (resistente a Spec, inducible por IPTG), pEXT22 (resistente a Kan, inducible por IPTG), pEC415 (resistente a Kan, inducible por arabinosa) o pLAFR (resistente a tet, constitutivo) o derivados de los mismos.

Plásmidos:

Nombre	Fuente
pEXT21	Dykhooome et al., 1996 Gene 177:133-136
pEXT22	Dykhooome et al., 1996 Gene 177:133-136
pEC415	Schulz et al., 1998 Science 281,1197-1200
pLAFR1	Friedman et al., 1982 Gene 18(3):289-96

Ejemplo 2

Detección de glucosilación por ELISA a partir de cultivos a pequeña escala

Las variantes de PglB se probaron para su capacidad para catalizar la glucosilación de la exoproteína A de *Pseudomonas aeruginosa* (EPA) que contiene sitios de glucosilación D/E-Z₁-N-Z₂-S/T (donde Z₁ y Z₂ no son P) usando un polisacárido que corresponde al de *S. pneumoniae* serotipo 8. Por lo tanto, se transformó una célula hospedadora de *E. coli* con plásmidos que codifican para genes de glucosiltransferasa requeridos para la construcción de un polisacárido capsular de *S. pneumoniae* serotipo 8, un gen PglB variante y sitios de glucosilación que contienen EPA. La expresión de los genes se indujo usando IPTG y arabinosa y las células hospedadoras de *E. coli* se cultivaron durante la noche para permitir la expresión de glucosiltransferasas, PglB y EPA y la glucosilación de EPA de la siguiente manera.

Los pozos de una placa de 96 pozos profundos se llenaron con 1 ml de medio TB y cada pozo se inoculó con una colonia individual de *E. coli* de célula hospedadora y se incubó a 37 grados C durante la noche. Las muestras de cada pozo se usaron para inocular cultivos principales en una placa de 96 pozos profundos que contiene 1 ml de TB complementada con MgCl₂ 10 mM y antibióticos apropiados y se cultivaron hasta que se alcanzó una OD600 de 1,3-1,5. Las células se incubaron con IPTG 1 mM y arabinosa al 0,1 % durante la noche a 37 grados C.

Los extractos periplásmicos se elaboraron al centrifugar las placas, al remover el sobrenadante y al adicionar 0,2 ml de Tris-HCl 50 mM pH 7,5, NaCl 175 mM, EDTA 5 mM seguido de agitación a 4 grados C para suspender las células. Se adicionaron 10 µl de 10 mg/ml de polimixina B a cada pozo y las células se incubaron durante 1 hora a 4 grados C. La placa se centrifugó y el sobrenadante se removió.

A fin de aislar la proteína glucosilada del extracto periplásmico, se adicionaron 120 µl de una suspensión espesa al 25 % de resina IMAC en Tris 30 mM pH 8,0, imidazol 10 mM, NaCl 500 mM a cada pozo de una placa de filtro de 96 pozos (Acroprep Advance) y se colocaron en la parte superior de una placa ELISA Nunc. La placa se centrifugó y el flujo continuo se descartó. Se adicionaron 150 µl de extracto periplásmico y 37,5 ml de amortiguador de unión 5x (Tris 150 mM pH 8,0, imidazol 50 mM, NaCl 2,5 M a cada pozo. Las muestras se incubaron durante 30 minutos a temperatura ambiente. La placa se centrifugó y el flujo continuo se descartó y se llevaron a cabo tres pasos de lavado más. Finalmente, la proteína glucosilada se eluyó con Tris 30 mM pH 8,0, imidazol 500 mM, NaCl 200 mM, listo para uso en un ensayo ELISA.

Se realizó un ELISA sándwich al recubrir los pozos de una placa de 96 pozos con un anticuerpo que reconoce la parte de sacárido de la proteína glucosilada (por ejemplo, un anticuerpo monoclonal contra *S. pneumoniae* serotipo 8) diluido en PBS. La placa se incubó durante la noche a 4 grados C para permitir el recubrimiento. La placa entonces se lavó con PBS que contiene Tween al 0,1 %. La placa entonces se bloqueó durante 2 horas a temperatura ambiente usando albúmina de suero bovino al 5 % en PBST. La placa se lavó en PBST. La muestra se diluyó en PBST que contiene BSA al 1 % y se incubó en los pozos recubiertos durante una hora a temperatura ambiente. Después de lavar un anticuerpo de detección, por ejemplo, se adicionó peroxidasa de rábano picante anti-Histag diluida en PBST que contiene BSA al 1 % a cada pozo y se incubó durante una hora a temperatura ambiente. La placa entonces se lavó antes de adicionar sustrato líquido de 3,3',5,5'-Tetrametilbencidina, Supersensible, para ELISA (Sigma-Aldrich). Después de unos minutos, la reacción se detuvo por adición de ácido sulfúrico 2M. Los resultados se obtuvieron al leer la OD a 450 nm.

Resultados

Como punto de partida, se generaron mutaciones en una PglB que ya contenía una mutación en N31 1V. Una primera ronda de generación de variantes identificó Y77R como una mutación que incrementó aún más la actividad OST de PglB (ver figura 1). Se sometió a mutación una PglB que contiene mutaciones en N31 1V e Y77R y se seleccionaron, secuenciaron y analizaron variantes prometedoras para determinar la actividad de OST como se describió anteriormente. El cambio múltiple en la actividad de oligosacariiltransferasa de cada variante se calculó y los resultados se muestran en la tabla 3.

Tabla 3 Mejora en la actividad de OST de PglB modificada en la transferencia del sacárido de *S. pneumoniae* 8 a una proteína como se determina por la RONDA 2 de ELISA

Mutación de variante de PglB	Sustitución de aminoácidos	Cambio múltiple en actividad de OST
57	T R	T- 1,047, R-2,2755
63	L Q	L-1,132, Q-1,146
94	N	N-1,103
101	W E H P R M G	W- 1,2635, E - 1,263, H-1,26, P-1,2595, R- 1,11325, M - 2,263, G - 1,235
176	E G	E - 1,2805
191	H D R Y	H-1,798, D-1,763, R-1,516, Y-2,017

Mutación de variante de PglB	Sustitución de aminoácidos	Cambio múltiplo en actividad de OST
193	T H G F	T -1,2375, H-1,291, G-1,508, F-1,5055
233	V	V-1,431
234	H C W	H-1,761, C-1,302, W-1,838
286	A Q L	A-1,296, Q-1,091, L-1,132
301	P G F	P-1,210, G-1,3, F-1,135
319	A Q L T	A-1,329, Q-1,423, L-1,724, T-1,626
397	N L Q	N -1,093, L - 1,1245,
402	R H	R-1,434, H-1,191
435	A L F H R	A-1,214, L-1,442, F-1,719, H-1,213, R-1,471
446	G	G-1,4375
462	P C W T N	P-2,221, C-1,2895, W-2,112, T-1,362, N-1,247
479	M	M-1,582
523	R	R-1,171
532	H	H-1,4085
605	D	D-1,209
610	P L R D A	P-1,227, I-1,34, R-1,518, D-1,245, A-1,467
645	L S H	L-1,377, 1,225, H-1,481
676	Q W G	Q-1,247, W-1,84, G-1,279
695	I Q	I-1,218

5 Las mutaciones 57T, 57R, 63L, 63Q, 94N, 101W, 101E, 101H, 101P, 101R, 101M, 101G, 172E, 176E, 176G, 191H, 191D, 101R, 101Y, 193T, 193H, 193G, 193F, 233V, 234H, 234C, 234W, 255H, 286A, 286Q, 286L, 301P, 301G, 301F, 319A, 319Q, 319L, 319T, 397N, 397L, 397Q, 402R, 402H, 425S, 435A, 435L, 435F, 435H, 435R, 446G, 462,P, 462C, 462W, 462T, 462N 479M, 523R, 532H, 601G, 605D, 606P, 610P, 610L, 610R, 610D, 610A, 645L, 645S, 645H, 676Q, 676W, 676G, 695I y 695Q se observaron como mutaciones que aparecieron en varias variantes de PglB que fueron capaces de catálisis mejorada de la adición de un *sacárido de S. pneumoniae* serotipo 8 a una proteína portadora. De estos, la mutación A57R se seleccionó como una mutación prometedora para avanzar en rondas adicionales debido a su alto incremento en la actividad de OST, su frecuencia de aparición en variantes prometedoras y su posición en la estructura de PglB.

10 La capacidad de los residuos 462 y 479 para sinergizar a fin de producir mayores cambios múltiplo en la actividad de OST cuando ambos residuos están mutados se demuestra en la siguiente tabla que muestra los resultados para individuo :

Mutación de variante de PglB	Cambio múltiplo en actividad de OST	Número de muestra
H479M	7,483	S00099450
H479M	6,8	S00099887
Y462W	2,377	S00109538
Y462W	2,164	S00109534
Y462W + H479M	12,404	S00099448
Y462W + H479M + A84W + M155Q + T523R	15,071	S00099611

15

En una ronda adicional de experimentos, se adicionó la mutación A57R favorable a las mutaciones Y77R y N311V. Se adicionaron mutaciones adicionales a PglB con mutaciones A57R, Y77R y N311V. Las nuevas mutaciones se probaron para actividad incrementada de PglB por ELISA usando *S. pneumoniae* PS8 como el sacárido adicionado a EPA. Los resultados se muestran en la Tabla 4 a continuación.

5

Tabla 4 - Mejora en la actividad de OST de PglB modificada en la transferencia del sacárido de *S. pneumoniae* 8 a una proteína como se determina por la RONDA 3 de ELISA

Mutación de variante de PglB	Sustitución de aminoácidos	Cambio múltiple en actividad de OST
78	T	T-1,2455
84	W	W-1,361
300	L	L-1,123
301	G P	G-1,419. P-2,22
306	H	H-1,375
462	N T W	N-1,76, T-1,4246, W-1,681,
479	M	M-5,5548
570	R V	R-1,083, V-1,419
462W + 479M		12,404, 8,368

10 A partir de esta ronda, la combinación de 462W y 479M se encontró en 52 variantes de PglB separadas, produciendo cambios múltiple en la actividad de OST de hasta 15 veces, u 8-12 veces donde estas mutaciones fueron las únicas mutaciones nuevas presentes. Estas mutaciones se consideran importantes para mejorar la eficiencia de PglB para glucosilación de proteínas con sacáridos que contienen un residuo de glucosa en el extremo reductor del sacárido.

15 Las mutaciones N300L, L301P, F308W y L570R también se observaron como mutaciones que aparecieron en varias variantes de PglB que fueron capaces de catálisis mejorada de la adición de un sacárido de *S. pneumoniae* serotipo 8 a una proteína portadora.

20 N300L, L301P, F308W, Y462W, H479M y L570R se seleccionaron como mutaciones prometedoras para avanzar en rondas adicionales de mutación/selección debido a su capacidad para incrementar la actividad de OST, su frecuencia de aparición en variantes de PglB eficientes y su posición en la estructura molecular de PglB. Estos residuos se adicionaron a las mutaciones Y77R, N311V y A57R y se probaron mutaciones puntuales adicionales para su capacidad para mejorar la actividad de PglB para adicionar un sacárido de *S. pneumoniae* serotipo 8 a una proteína. Los resultados se muestran en la Tabla 5 a continuación.

25 **Tabla 5 - Mejora en la actividad de OST de PglB modificada en la transferencia del sacárido de *S. pneumoniae* 8 a una proteína como se determina por la RONDA 4 de ELISA**

Mutación de variante de PglB	Sustitución de aminoácidos	Cambio múltiple en actividad de OST
218	A	A-3,084
308	F	F-1,519
319	L	L-1,218
382	S	S-2,77
523	R	R-1,086
L191Y, Y286Q, S295L, A382S, K482R + T523R		9,166

30 A partir de esta ronda, se adicionaron las mutaciones L191Y, Y286Q, S295L, A382S, K482R y T523R a las mutaciones probadas previamente, dando por resultado una PglB de referencia que contiene las siguientes mutaciones: Y77R, N311V, A57R, N300L, L301P, F308W, Y462W, H479M, L570R, L191Y, Y286Q, S295L, A382S, K482R y T523R. En la siguiente ronda, se probaron mutaciones adicionales para su capacidad para incrementar adicionalmente la eficiencia de un PglB que contiene Y77R, N311V, A57R, N300L, L301P, F308W, Y462W, H479M, L570R, round L191Y, Y286Q, S295L, A382S, K482R y T523R adicionalmente y los resultados se muestran en la tabla 6 más adelante.

35

Tabla 6 - Mejora en la actividad de OST de PglB modificada en la transferencia del sacárido de *S. pneumoniae* 8 a una proteína como se determina por la RONDA 5 de ELISA

Mutación de variante de PglB	Sustitución de aminoácidos	Cambio múltiplo en actividad de OST
21	S L	S - 1,7; L - 4,415
27	S A M W	S - 2,01; A - 2,156; M-2,21; W - 2,505
42	W C	W - 1,32; C - 3,566
44	M H	M - 1,348; H - 1,76
53	S I H	S - 1,556; I - 1,611; H 2,298
80	A D	A - 1,647; D - 1,788
97	I	I - 3,456
297	K Y S R	K - 1,96; Y - 2,033; S-2,1635; R - 2,51
317	S A	S - 1,3805; A - 2,283
341	L	L - 2,103
383	M	M - 2,208
388	I M	I - 1,34; M - 2,261
410	I	1-2,143
421	G	G - 2,235
480	W Q I N M T	W - 1,246; Q - 1,715; I-1,727; N - 1,864; M-1,924; T - 4,098
486	N V C L	N - 1,231; V - 1,479; C-2,315; L - 4,4

De esta ronda se seleccionó un E297R debido al incremento en la actividad producida donde esta mutación estaba presente y la frecuencia de aparición de esta mutación. También se seleccionó una mutación S80D como un residuo prometedor debido a su frecuencia de aparición en variantes con niveles más altos de actividad de OST.

5

Breve descripción de la evolución

En el transcurso de este estudio, la actividad de PglB en el contexto de la adición de un sacárido de *S. pneumoniae* serotipo 8 a una proteína que contiene una secuencia de consenso de glucosilación se incrementó por más de tres órdenes de magnitud (figura 1). La introducción de A57R en una PglB en una enzima PglB que ya contenía mutaciones Y77R y N311V condujo a un cambio múltiplo de 24 en la actividad de PglB. La adición adicional de N300L, L301P, F308W, Y462W, H479M y L750R condujo a un incremento acumulativo en la actividad de 360 veces. La incorporación adicional de L191Y, Y286Q, S295L, A382S, K482R y T523R condujo a un incremento acumulativo de la actividad de 2520 veces y la inclusión adicional de una mutación E297R condujo a un cambio múltiplo acumulativo de 5040 en la actividad. Otras rondas de evolución permitieron pequeños incrementos en la actividad de PglB, sin embargo, los incrementos más grandes en la actividad se lograron en las rondas 1-3 (ver figura 1).

10

15

Ejemplo 3

20

Medición de actividad de PglB en volumen de matraz de agitación

Algunas de las enzimas PglB mutadas se usaron en ensayos a mayor escala a fin de confirmar incrementos en la actividad en la glucosilación de una proteína EPA que contiene 3 secuencias de consenso de glucosilación con sacárido de *S. pneumoniae* serotipo 8.

25

Las cepas de *E. coli* electrocompetentes se transformaron con los plásmidos requeridos por electroporación. Las células se dejaron recuperar durante 1 hora y se colocaron en placas de agar que contienen antibióticos apropiados y MgCl₂ 2 mM. Las placas se incubaron durante la noche. Se realizó un precultivo al inocular medios de TB que contienen antibióticos apropiados y MgCl₂ 10 mM con células de la placa e incubando durante la noche.

30

El cultivo principal se inició al diluir el precultivo en medio de TB que contiene antibióticos apropiados y MgCl₂ 10 mM a una OD_{600nm} de 0,1. El cultivo se cultivó a una OD_{600nm} de 0,8-1,0 y las células entonces se indujeron usando un inductor apropiado (por ejemplo, arabinosa de IPTG). Las células entonces se incubaron durante la noche.

35

Se realizó un extracto periplásmico por el siguiente proceso. El cultivo se centrifugó para sedimentar la *E. coli*. El sobrenadante se desechó y el sedimento se resuspendió en Tris-HCl 30 mM pH 8,5, EDTA 1 mM, sacarosa al 20 %. Se

adicionó lisozima a una concentración final de 1 mg/ml y las células se incubaron con la lisozima durante 25 minutos a 4 grados C con agitación. Después de la centrifugación, se retuvo el extracto periplásmico.

5 Se mezcló 1 ml de extracto periplásmico con 0,25ml de Tris 150 mM pH 8,0, imidazol 50 mM, NaCl 2,5 M y MgCl₂ 20 mM. Se adicionaron 0,2 ml de una suspensión espesa al 50 % o agarosa NiNTA preequilibrada (Qiagen) y la muestra se incubó durante 20 minutos a temperatura ambiente con agitación. La resina IMAC se centrifugó y el sobrenadante se descartó. Se adicionaron 0,5 ml de Tris 30 mM pH 8,0, imidazol 10 mM, NaCl 500 mM con n-dodecil-B-maltosa al 0,1 %, se centrifugó la resina y se descartó el sobrenadante. La resina se lavó adicionalmente tres veces con Tris 30 mM pH 8,0, imidazol 10 mM, NaCl 500 mM. Se adicionaron 0,2 ml de amortiguador de elución (Tris 30 mM pH 8,0, imidazol 10 mM, NaCl 50 mM) a la resina y se incubó durante 5 minutos a temperatura ambiente. El eluato se recuperó y se usó para análisis adicional.

15 La cantidad de glicosilación se valoró por SDS-PAGE y transferencia Western. Después de correr las muestras en un SDS-PAGE, las proteínas se transfirieron a una membrana de nitrocelulosa. La membrana se bloqueó con leche al 10 % durante al menos 10 minutos. Después de bloquear, la membrana se incubó con un primer anticuerpo (por ejemplo, un Mab de ratón contra *S. pneumoniae* serotipo 8) en PBS-T que contiene leche al 1 % durante 1 hora. Después de lavar con PBS-T, la membrana se incubó con un segundo conjugado de anticuerpo - HRP (IgG Fc anti-ratón HRP) en PBS-T con leche al 1 % durante una hora. La membrana se lavó en PBS-T y se desarrolló usando el sustrato de membrana BioFX TMB One Component HRP.

20 Resultados

La figura 1 muestra un gel en el cual se obtuvieron niveles mejorados de EPA glucosilada con sacárido capsular de *S. pneumoniae* serotipo 8. Los incrementos más importantes en la actividad de OST se lograron en las rondas 1-3, y se lograron cambios múltiplo más pequeños en la actividad en las rondas posteriores. A escala de matraz de agitación, se lograron incrementos de rendimiento de más de 1.000 veces.

Ejemplo 4

30 Las PglB oligosacariltransferasas mutadas muestran una eficiencia mejorada para catalizar la glucosilación con sacáridos adicionales

Se llevaron a cabo experimentos adicionales para investigar si las OST de PglB modificadas de cada ronda pueden producir mayores rendimientos de bioconjugados adicionales donde diferentes sacáridos se unieron a la proteína EPA.

35 Los protocolos de los ejemplos 2 y 3 se usaron para elaborar bioconjugados de *S. pneumoniae* serotipo 22F unidos covalentemente a EPA modificada. Los resultados de ELISA y transferencia Western muestran que se pudieron lograr buenos rendimientos de conjugado de *S. pneumoniae* serotipo 22F-EPA usando las PglB modificadas generadas a partir de las rondas 3, 4 y 5 del ejemplo 2. El rendimiento usando una PglB de la ronda 3 con mutaciones en A57R, Y77R y N311V es bueno, pero se mejora adicionalmente usando la PglB de la ronda 4 que contiene mutaciones puntuales adicionales en N300L, L301P, F308W, Y462W, H479M y L570R. El rendimiento se mejora adicionalmente al usar la ronda 5 PglB que contiene mutaciones adicionales en L191Y, Y286Q, S295L y A382S.

45 Las OST de PglB mutadas también fueron eficientes para catalizar la adición del sacárido de *S. pneumoniae* serotipo 23A a una proteína como se muestra en la figura 3A. Las OST de PglB mutadas también fueron eficientes para catalizar la adición de *S. pneumoniae* serotipo 35B a una proteína como se muestra en la figura 3B. La inclusión de sustituciones en A57R, Y77R y N311V condujo a la actividad mejorada mostrada en la figura 3B.

50 Las OST de PglB mutadas también fueron eficientes para catalizar la adición de *S. pneumoniae* serotipo 19A a una proteína como se muestra en la figura 4.

Listado de secuencias

SEQ ID NO:
1PglB tipo silvestre de *C. jejuni*. (pLMTB1937)

55

ES 2 984 862 T3

MLKKEYLKNPYL.VLFAMIILAYVFSVFCRFYVWWWASEFNEYFFNNQLMIIISNDGYAFAEGARDM
IAGFHQPNDLSYYGSSLSALTYWLYKITPPFSFESHLYMSTFLSSLVVIPTILLANEYKRPLMGFVAAL
L.ASIANSYYNRTMSGYYDIDML.VIVL.PMFILFFMVRMILKKDFFSLIAL.PLFIGIYLWWYPSSYTLNV
ALIGLFLIYTLIFHRKEKIFYIAVILSSL.TLSNIAWFYQSAIIVILFALFALEQKRL.NFMHIGILGSATLIFLI
LSGGVDPILYQLKFYIFRSDDESANLTQGFMYFNVNQTIQEVENVDLSEFMRRISGSEIVFLFSLFGPV
WLLRKHKSMIMALPILVLGFLALKGGLRFTTYSVPVMALGFGFLLSEFKAIMVKKYSQLTSNVCIVF
ATIL.TLAPVFIHIYNYKAPT VFSQNEASLLNQLKNIANREDYVVTWWDYGYPVRYYSDEVKTLVDGG
KHLGKDNFFPSFALS.KDEQAAANMARLSVEYTEKSFYAPQNDILKTDILQAMMKDYNQSNVDLFL
ASL.SKPDFKIDTPKTRDIYLYMPARMSLIFSTVA SFSFINLDTGVLDKPFSTAYPLDVKNGEIYLSN
GVVLSDDFRSFKIGDNVVSVNSIVEINSIKQGEYKIPIDDKAQFYIFYLKDSAIPYAQFILMDKTMFN
SAYVQMPFLGNYDKNLFDLVINSRDAKVPKLEKI

SEQ ID NO:2

5 PgIB de *C. jejuni* que contiene las mutaciones N311V y Y77R (pLMTB4028)

MLKKEYLKNPYL.VLFAMIILAYVFSVFCRFYVWWWASEFNEYFFNNQLMIIISNDGYAFAEGARDM
IAGFHQPNDLSRYYGSSLSALTYWLYKITPPFSFESHLYMSTFLSSLVVIPTILLANEYKRPLMGFVAALL
ASIANSYYNRTMSGYYDIDML.VIVL.PMFILFFMVRMILKKDFFSLIAL.PLFIGIYLWWYPSSYTLNV
ALIGLFLIYTLIFHRKEKIFYIAVILSSL.TLSNIAWFYQSAIIVILFALFALEQKRL.NFMHIGILGSATLIFLI
LSGGVDPILYQLKFYIFRSDDESANLTQGFMYFNVNQTIQEVENVDLSEFMRRISGSEIVFLFSLFGPV
WLLRKHKSMIMALPILVLGFLALKGGLRFTTYSVPVMALGFGFLLSEFKAIMVKKYSQLTSNVCIVF
ATIL.TLAPVFIHIYNYKAPT VFSQNEASLLNQLKNIANREDYVVTWWDYGYPVRYYSDEVKTLVDGG
KHLGKDNFFPSFALS.KDEQAAANMARLSVEYTEKSFYAPQNDILKTDILQAMMKDYNQSNVDLFL
ASL.SKPDFKIDTPKTRDIYLYMPARMSLIFSTVA SFSFINLDTGVLDKPFSTAYPLDVKNGEIYLSN
GVVLSDDFRSFKIGDNVVSVNSIVEINSIKQGEYKIPIDDKAQFYIFYLKDSAIPYAQFILMDKTMFN
SAYVQMPFLGNYDKNLFDLVINSRDAKVPKLEKI

SEQ ID NO:3

10 PgIB de *C. jejuni* con las mutaciones A57R, Y77R y N311V (pLMTB4768)

MLKKEYLKNPYL.VLFAMIILAYVFSVFCRFYVWWWASEFNEYFFNNQLMIIISNDGYRFAEGARDM
IAGFHQPNDLSRYYGSSLSALTYWLYKITPPFSFESHLYMSTFLSSLVVIPTILLANEYKRPLMGFVAALL
ASIANSYYNRTMSGYYDIDML.VIVL.PMFILFFMVRMILKKDFFSLIAL.PLFIGIYLWWYPSSYTLNV
ALIGLFLIYTLIFHRKEKIFYIAVILSSL.TLSNIAWFYQSAIIVILFALFALEQKRL.NFMHIGILGSATLIFLI
LSGGVDPILYQLKFYIFRSDDESANLTQGFMYFNVNQTIQEVENVDLSEFMRRISGSEIVFLFSLFGPV
WLLRKHKSMIMALPILVLGFLALKGGLRFTTYSVPVMALGFGFLLSEFKAIMVKKYSQLTSNVCIVF
ATIL.TLAPVFIHIYNYKAPT VFSQNEASLLNQLKNIANREDYVVTWWDYGYPVRYYSDEVKTLVDGG
KHLGKDNFFPSFALS.KDEQAAANMARLSVEYTEKSFYAPQNDILKTDILQAMMKDYNQSNVDLFL
ASL.SKPDFKIDTPKTRDIYLYMPARMSLIFSTVA SFSFINLDTGVLDKPFSTAYPLDVKNGEIYLSN
GVVLSDDFRSFKIGDNVVSVNSIVEINSIKQGEYKIPIDDKAQFYIFYLKDSAIPYAQFILMDKTMFN
SAYVQMPFLGNYDKNLFDLVINSRDAKVPKLEKI

15 SEQ ID NO:4

PgIB de *C. jejuni* con las mutaciones A57R, Y77R, N300L, L301P, F308W, N311V, Y462W, H479M, L570R (pLMTB5298)

MLKKEYLKNPYLVLFAMILAYVFSVFCRFYVWWWASEFNEYFFNNQLMHSNDGYRFAEGARDM
IAGFHQPNDLSRYGSSLSALTYWLYKITPFSFESHLYMSTFLSSLVVIPTILLANEYKRPLMGFVAALL
ASIANSYYNRTMSGYYDMLVIVLPMFILFFMVRMILKKDFFSLIALPLFIGIYLWWYPSSYTLNV
ALIGLFLIYTLIFHRKEKIFYIAVILSSLTLSNIAWFYQSAHVILFALFALEQKRLNFMHIGILGSATLIFLI
LSGGVDPILYQLKFIYFRSDESALPTQGFMYWNVQTIQEVENVDLSEFMRRISGSEIVFLSFLGFV
WLLRKHKSMIMALPILVLGFLALKGGLRFTIYSVPVMSLGFGLLSEFKAIMVKKYSQLTNSVCIVF
ATILTLAPVFIHIYNYKAPT VFSQNEASLLNQLKNIANREDYVVTWWDYGWVPRYYSDVKTLVDG
GKMLGKDNFFPSFALSDEQAAANMARLSVEYTEKSFYAPQNDILKTDILQAMMKDYNQSNVDLF
LASLSKPDFKIDTPKTRDIYLYMPARMSRIFSTVASFSFINLDTGVLDKPFTFSTAYPLDVKNGEIYLS
NGVVLSDDFRSFKIGDNVVSNSIVEINSIKQGEYKITPIDDKAQFYIFYLKDSAIPYAQFILMDKTMF
NSAYVQMFFLGNYDKNLFDLVINSRDAKVFKLKI

5 SEQ ID NO:5

PgIB de *C. jejuni* con las mutaciones A57R, Y77R, L191Y, Y286Q, S295L, N300L, L301P, F308W, N311V, A382S, Y462W, H479M, K482R, T523R, L570R (pLMTB5685)

MLKKEYLKNPYLVLFAMILAYVFSVFCRFYVWWWASEFNEYFFNNQLMHSNDGYRFAEGARDM
IAGFHQPNDLSRYGSSLSALTYWLYKITPFSFESHLYMSTFLSSLVVIPTILLANEYKRPLMGFVAALL
ASIANSYYNRTMSGYYDMLVIVLPMFILFFMVRMILKKDFFSLIALPLFIGIYWWYPSSYTLNV
ALIGLFLIYTLIFHRKEKIFYIAVILSSLTLSNIAWFYQSAHVILFALFALEQKRLNFMHIGILGSATLIFLI
LSGGVDPILQQLKFIYFRIDESALPTQGFMYWNVQTIQEVENVDLSEFMRRISGSEIVFLSFLGFV
WLLRKHKSMIMALPILVLGFLALKGGLRFTIYSVPVMSLGFGLLSEFKAIMVKKYSQLTNSVCIVF
ATILTLAPVFIHIYNYKAPT VFSQNEASLLNQLKNIANREDYVVTWWDYGWVPRYYSDVKTLVDG
GKMLGRDNFFPSFALSDEQAAANMARLSVEYTEKSFYAPQNDILKRDILQAMMKDYNQSNVDLF
LASLSKPDFKIDTPKTRDIYLYMPARMSRIFSTVASFSFINLDTGVLDKPFTFSTAYPLDVKNGEIYLS
NGVVLSDDFRSFKIGDNVVSNSIVEINSIKQGEYKITPIDDKAQFYIFYLKDSAIPYAQFILMDKTMF
NSAYVQMFFLGNYDKNLFDLVINSRDAKVFKLKI

10

SEQ ID NO:6

15 PgIB de *C. jejuni* con las mutaciones A57R, Y77R, L191Y, Y286Q, S295L, E297R, N300L, L301P, F308W, N311V, A382S, Y462W, H479M, K482R, T523R, L570R (pLMTB5967)

MLKKEYLKNPYLVLFAMILAYVFSVFCRFYVWWWASEFNEYFFNNQLMHSNDGYRFAEGARDM
IAGFHQPNDLSRYGSSLSALTYWLYKITPFSFESHLYMSTFLSSLVVIPTILLANEYKRPLMGFVAALL
ASIANSYYNRTMSGYYDMLVIVLPMFILFFMVRMILKKDFFSLIALPLFIGIYWWYPSSYTLNV
ALIGLFLIYTLIFHRKEKIFYIAVILSSLTLSNIAWFYQSAHVILFALFALEQKRLNFMHIGILGSATLIFLI
LSGGVDPILQQLKFIYFRIDESALPTQGFMYWNVQTIQEVENVDLSEFMRRISGSEIVFLSFLGFV
WLLRKHKSMIMALPILVLGFLALKGGLRFTIYSVPVMSLGFGLLSEFKAIMVKKYSQLTNSVCIVF
ATILTLAPVFIHIYNYKAPT VFSQNEASLLNQLKNIANREDYVVTWWDYGWVPRYYSDVKTLVDG
GKMLGRDNFFPSFALSDEQAAANMARLSVEYTEKSFYAPQNDILKRDILQAMMKDYNQSNVDLF
LASLSKPDFKIDTPKTRDIYLYMPARMSRIFSTVASFSFINLDTGVLDKPFTFSTAYPLDVKNGEIYLS
NGVVLSDDFRSFKIGDNVVSNSIVEINSIKQGEYKITPIDDKAQFYIFYLKDSAIPYAQFILMDKTMF
NSAYVQMFFLGNYDKNLFDLVINSRDAKVFKLKI

20 SEQ ID NO:7

PgIB de *C. jejuni* con las mutaciones A57R, Y77R, S80D, L191Y, Y286Q, S295L, E297R, N300L, L301P, F308W, N311V,

A382S, Y462W, H479M, K482R, T523R, L570R (pLMTB6274)

MLKKEYLKNPYLVLFAMIILAYVFSVFCRFYVWVWASEFNEYFFNNQLMIISNDGYRFAEGARDM
 IAGFHQPNDLSRYGDSL SALTYWLYKTI PFSFESIILYMSTFLSSLVVIPTILLANEYKRPLMGFVAAL
 LASIANSYYNRTMSGYD TDMLVIVLPMFILFFMVRMILKKDFFSLIALPLFIGIYYWYYPSSYTLN
 VALIGLFLIYTLIFHRKEKIFYIAVILSSLTLSNIAWFYQSAIIVILFALFALEQKRLNFMIIHILGSA TLIF
 LILSGGVDPILQQLKFYIFRLDRSALPTQGFMYWNVVQTIQEVENVDLSEFMRRISGSEIVFLFSLFGF
 VWLLRKHKSMIMALPILVLGFLALKGGLRFTIYSVPVMSLGFGLLSEFKAIMVKKYSQLTSNVCIV
 FATILTLAPVFIHIYNYKAPT VFSQNEASLLNQLKNIANREDYVVTWWDYGWVPRYYS DVKTLVD
 GGKMLGRDNFFPSFALS KDEQAAANMARLSVEYTEKSFYAPQNDILKRDILQAMMKDYNQSNVD
 LFLASLSKPDFKIDTPKTRDIYLYMPARMSRIFSTVASFSFINLDTGVLDKPFIFSTAYPLDVKNGEIY
 LSNGVVLSDDFRSFKIGDNVVSNSIVEINSIKQGEYKITPIDDKAQFYIFYLKDSAIPYAQFILMDKT
 MFNSAYVQMPFLGNYDKNLFDLVINSRDAKVFKLKI

5 SEQ ID NO:8

PgIB de *C. jejuni* con las mutaciones A57R, Y77R, S80D, 1187V, L191Y, Y286Q, S295L, E297R, N300L, L301P, F308W, N311V, 13590, A382S, N4061, Y462W, H479M, K482R, T523R, L570R (pLMTB6526)

MLKKEYLKNPYLVLFAMIILAYVFSVFCRFYVWVWASEFNEYFFNNQLMIISNDGYRFAEGARDM
 IAGFHQPNDLSRYGDSL SALTYWLYKTI PFSFESIILYMSTFLSSLVVIPTILLANEYKRPLMGFVAAL
 LASIANSYYNRTMSGYD TDMLVIVLPMFILFFMVRMILKKDFFSLIALPLFVGIYYWYYPSSYTLN
 VALIGLFLIYTLIFHRKEKIFYIAVILSSLTLSNIAWFYQSAIIVILFALFALEQKRLNFMIIHILGSA TLIF
 LILSGGVDPILQQLKFYIFRLDRSALPTQGFMYWNVVQTIQEVENVDLSEFMRRISGSEIVFLFSLFGF
 VWLLRKHKSMIMALPQLVLGFLALKGGLRFTIYSVPVMSLGFGLLSEFKAIMVKKYSQLTSIVCIV
 FATILTLAPVFIHIYNYKAPT VFSQNEASLLNQLKNIANREDYVVTWWDYGWVPRYYS DVKTLVD
 GGKMLGRDNFFPSFALS KDEQAAANMARLSVEYTEKSFYAPQNDILKRDILQAMMKDYNQSNVD
 LFLASLSKPDFKIDTPKTRDIYLYMPARMSRIFSTVASFSFINLDTGVLDKPFIFSTAYPLDVKNGEIY
 LSNGVVLSDDFRSFKIGDNVVSNSIVEINSIKQGEYKITPIDDKAQFYIFYLKDSAIPYAQFILMDKT
 MFNSAYVQMPFLGNYDKNLFDLVINSRDAKVFKLKI

10

SEQ ID NO:9 PgIB de *C. jejuni*

MLKKEYLKNP YLVLFAMIIL AYVFSVFCRF YWVWASEFN EYFFNNQLMI
 ISNDGYAFAE
 GARDMIAGFH QPNDLSYYGS SLSALTYWLY KITPFSFESI ILYMSTFLSS
 LVVIPTILLA
 NEYKRPLMGF VAALLASIAN SYYNRTMSGY YDTDMLVIVL PMFILFFMVR
 MILKKDFFSL
 IALPLFIGIY LWYYPSSYTL NVALIGLFLI YTLIFHRKEK IFYIAVILSS
 LTLNIAWFY
 QSAIIVILFA LFALEQKRLN FMIIGILGSA TLIFLILSGG VDPILYQLKF
 YIFRSDESAN
 LTQGFMYFNV NQTIQEVENV DLSEFMRRIS GSEIVFLFSL FGFVWLLRKH
 KSMIMALPIL
 VLGFLALKGG LRFTIYSVPV MALGFGLLS EFKAIMVKKY SQLTSNVCIV
 FATILTLAPV

15

ES 2 984 862 T3

FIHIYNYKAP TVFSQNEASL LNQLKNIANR EDYVVTWWDY GYPVRYSDV
KTLVDGGKHL
GKDNFFPSFA LSKDEQAAAN MARLSVEYTE KSFYAPQNDI LKTDILQAMM
KDYNQSNVDL
FLASLSKPDF KIDTPKTRDI YLYMPARMSL IFSTVASFSF INLDTGVLDK
PFTFSTAYPL
DVKNGEIYLS NGVVLSDDFR SFKIGDNVVS VNSIVEINSI KQGEYKITPI
DDKAQFYIFY
LKDSAIPYAQ FILMDKTMFN SAYVQMFFLG NYDKNLFDLV INSRDAKVFK
LKIYPYDVPD
YA

SEQ ID NO: 10 PglB de *C. lari*

MKLQQNFTDN NSIKYTCILI LIAFAFSVLC RLYWVAWASE
FYEFFFNDQL
MITTNDGYAF AEGARDMIAG FHQPNDLSYF GSSLSTLTYW
LYSILPFSFE SIILYMSAFF ASLIVVPIIL IAREYKLTTY
GFIAALLGSI ANSYNRTMS GYDMDMLVL VLPMLILLTF
IRLTINKDIF TLLSPVFIM IYLWWYPSSY SLNFAMIGLF
GLYTLVFHRK EKIFYLTIAL MIIALSMLAW QYKLALIVLL
FAIFAFKEEK INFYMIWALI FISILILHLS GGLDPVLYQL
KFYVFKASDV QNLKDAAFMY FNVNETIMEV NTIDPEVFMQ
RISSSVLVFI LSFIFILLC KDHSMLLAL PMLALGFMAL
RAGLRFTIYA VPVMALGFGY FLYAFFNFLE KKQIKLSLRN
KNILLILIAF FSISPALMHI YYYKSSTVFT SYEASILNDL
KNKAQREDYV VAWWDYGYPI RYSDVKTLI DGGKHLGKDN
FFSSFVLSKE QIPAANMARL SVEYTEKSKF ENYPDVLKAM
VKDYNKTSK DFLESNDKD FKFDTNKTRD VYIYMPYRML
RIMPVVAQFA NTNPDNGEQE KSLFFSQANA IAQDKTTGVS
MLDNGVEIIN DFRALKVEGA SIPLKAFVDI ESITNGKFY
NEIDSKAQIY LLFLREYKSF VILDESLYNS SYIQMFLLNQ
YDQDLFEQIT NDTRAKIYRL KR

5

SEQ ID NO:11

MLKKEYLKNPYL.VLFAMILAYVFSVFCRFYVWWWASEFNEYFFNNQLMHSNDGYAFAEGARDM
IAGFHQPNDLSHYGSSLSALTYWLKIFFSFESHLYMSTFLSSLVVIPTILLANEYKRPLMGFVAAL
LASIANSYNNRTMSGYDMDMLVIVLPMFILFFMVRMILKKDFFSLIALPLFIGIYLWWYPSSYTLNV
ALIGLFLIYTLIFHRKEKIFYIAVILSSLTSLNIAWFYQSAHVILFALFALEQKRLNFMHIGILGSATLIFLI
LSGGVDPILYQLKFYIFRSDESANLTQGFMYFNVVQTIQEVENVDLSEFMRRISGSEIVFLSFLGFPV
WLLRKHKSMIMALPILVLGFLALKGGLRFTTYSVPVMALGFGFLLEFKAIMVKKYSQLTSNVCIVF
ATILTLAPVFIHIYNYKAPT VFSQNEASLLNQLKNIANREDYVVTWWDYGYPVRYSDVKTLDVGG
KHLGKDNFFPSFALS.KDEQAAANMARLSVEYTEKSFYAPQNDILKTDILQAMMKDYNQSNVDLEL
ASLSKPDFKIDTPKTRDIYLYMPARMSLIFSTVASFSFINLDTGVLDKPPFTFSTAYPLDVKNGEIYLSN
GVVLSDDFRSFKIGDNVVS VNSIVEINSIKQGEYKITPIDDKAQFYIFYLKDSAIPYAQFILMDKTMFN
SAYVQMFFLGNYDKNLFDLVINSRDAKVFKLKI

10

SEQ ID NO: 12 PglB de *Campylobacter_coli_jv20.ASM14683v1 EFM37568*

MLKKEYFKNPTFILLAFHILAYVFSVLCRFYWIFWASEFNEYFFNNELMIISNDGYAFAEGARDMIA
FHQPNDLSYYGSSLSTLTYWFKITPPSLESIFIFIYSTFLSSLVVIPLILIANEYKRPLMGFVAALLASIA
NSYYNRTMSGYYDTDMLVIVLAMMIVFFMIRLILKKDLLSLITLPLFVGIYLWWYPPSSYTLNVALLG
LFFIYTLVFHIKEKTLYMAHLASITLSNIAWIFYQSAHVILFSLFVLQNKRFSEALLGILGLATLVFLILS
GGIDPILYQLKFYIFRSDSANLAQGFMYFNVNQTIQEVESIDL SIFMQRISGSELVFFVSLIGFIFLVR
KHKSMILALPMLALGFLALKSGLRFTIYAVPVLALGFGFLMSLLQERKQKNNNTYWWANIGVFIF
FLSLIPMFYHINNYKAPT VFSQNEATKLDLKKIAQREYVVTWWDYGYPIRYSDVKTLADGGK
HLGKDNFFPSFVLSKDQVAAANMARLSVEYTEKSFYAPLNDILKNDLLQAMMKDYNQNNVDLFL
ASLSKPDFKINMPKTRDVYTYMPARMSLIFSTVASFSFVDLETGEINKPFTFSAAYPLDVKNGEIYLS
NGIALSDDFRSFKINNSTISVNSHEINSIKQGEYKITPIDDMAQFYIFYLKDSTIPYAQFILMDKTMFNS
AYVQMFFLGN YDKNLYDLVINARDAKVFELKI

SEQ ID NO: 13 PgIB de Campylobacter_coli_76339.3S CDG57218

MLKKEYFKNPTFILLTLIILAYAFSVLCRFYWVFWASEFNEYFFNNELMIISNDGYAFAEGARDMIA
GFHQPNDSL YYGSSLSTLTYWFKITPPSLESIFIFIYSTFLSSLVVVPLILIANEYKRPLMGFVAALLASI
ANSYYNRTMSGYYDTDMLVIVLAMMIVFFMIRLILKKDLLSLIALPLFVGIYLWWYPPSSYTLNVAL
LGLLFFIYTLVFHIKEKTLYMAHLASITLSNIAWIFYQSAHVILFSLFVLQNKRFSEALLGFLGLATLVFL
ILSGGVDPIYQLKFYIFRSDSANLAQGFMYFNVNQTIQEVESIDL SIFMKRISGSELVFFISLIGFIFL
VRKHKSMILALPMLALGFLALKGGLRFTIYAVPVLALGFGFLMSLLQERKWKNNKNIYASVSIFTF
LSLLIPMFYHINNYKAPT VFSQNEASKLDELKKIAQREYVVAWWDYGYPIRYSDVKTLADGGKH
LQKDNFFPSFVLSKDQTAANMARLSVEYTEKSFYAPSNLILKNDLLKAMMKDYKQNNVDLFLAS
LSKPDFKINTPKTRDVYTYMPARMSLIFSTVASFSFVDLDTGKIDKPFSTFSAAYPLDVKNGEIYLSNGI
VLSDDFRSFKINHNTIPVNSHEVSSIKQGEYKITPIDDTAQFYILYLKDSTIPYAQFILMDKAMFNSAY
VQMFFLGN YDKNLYDLVINTRDTKVFELKI

5

SEQ ID NO:14 PgIB de Campylobacter_coli_317_04.ASM25395v3 EIA90085

MLKKEYFKNPTFILLAFIILAYVFSVLCRFYWVFWASEFNEYFFNNELMIISNDGYAFAEGARDMIA
GFHQPNDSL YYGSSLSTLTYWFKITPPSLESIFIFIYSTFLSSLVVIPLILIANEYKRPLMGFVAALLASI
ANSYYNRTMSGYYDTDMLVIVLAMMIVFFMIRLILKKDLLSLIALPLFVGIYLWWYPPSSYTLNVAL
LGLFFIYTLVFHIKEKTLYMAHLASITLSNIAWIFYQSAHVILFSLFVLQNKRFSEALLGILGLATLVFLI
LSGGIDPILYQLKFYIFRSDSANLAQGFMYFNVNQTIQEVESIDL SIFMQRISGSELVFFVSLIGFIFL
RKHKSMILALPMLALGFLALKSGLRFTIYAVPVLALGFGFLMSLLQERKQKNNNTYWWANIGVFIF
TFLSLIPMFYHINNYKAPT VFSQNEATKLDLKKIAQREYVVAWWDYGYPIRYSDVKTLADGG
KHLGKDNFFPSFVLSKDQVAAANMARLSVEYTEKSFYAPLNDILKNDLLQAMMKDYNQNNVDLFL
LALLSKPDFKINTPKTRDVYTYMPARMSLIFSTVASFSFVDLGTGEINKPFTFSAAYPLDVKNGEIYLS
NGIVLSDDFRSFKINNSTISVNSHEINSIKQGEYKITPIDDTAQFYIFYLKDSTIPYAQFILMDKTMFNS
AYVQMFFLGN YDKNLYDLVINARDAKVFELKF

REIVINDICACIONES

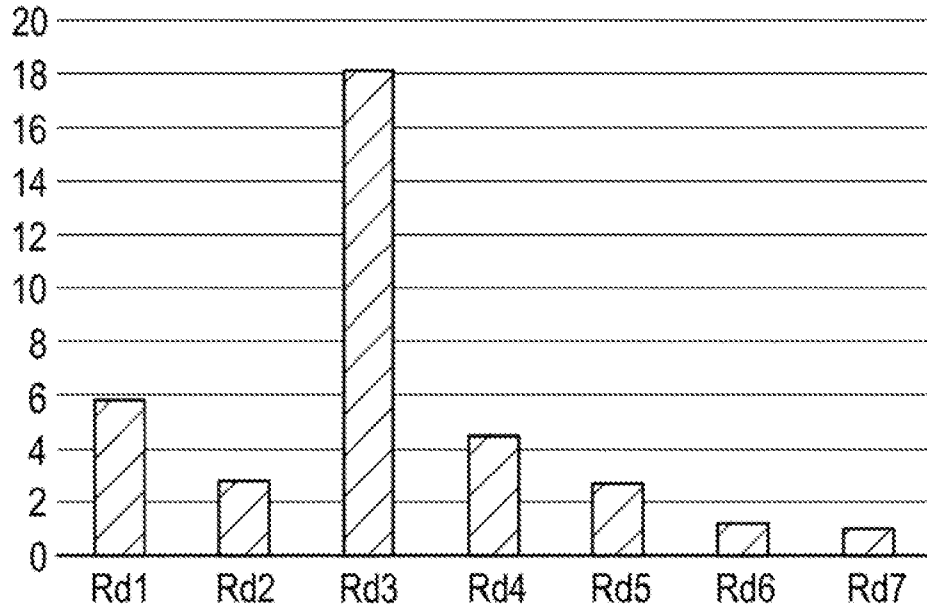
1. Un polipéptido de PglB oligosacaryltransferasa que comprende una secuencia de aminoácidos que es al menos 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntica a la secuencia de aminoácidos expuesta en SEQ ID NO: 1 o 2 o un fragmento funcional de la misma, donde la secuencia de aminoácidos de polipéptido de PglB oligosacaryltransferasa incluye la característica de que: al menos un residuo seleccionado del grupo que consiste en el aminoácido X57, X63, X94, X101, X172, X176, X191, X193, X233, X234, X255, X286, X295, X301, X319, X397, X402, X425, X435, X446, X462, X479, X523, X532, X601, X605, X606, X610, X645, X676 y X695 se sustituye por un aminoácido diferente al que se encuentra en esa posición en SEQ ID NO: 1 o fragmento funcional del mismo, donde el residuo que corresponde al aminoácido 57 de SEQ ID NO: 1 se muta a R o K o T, preferentemente a R o T, más preferentemente a R_y donde el polipéptido de PglB oligosacaryltransferasa es capaz de incrementar el rendimiento de glucosilación de la proteína con el sacárido a producir una proteína glucosilada por al menos 1,5 veces en comparación con una PglB oligosacaryltransferasa correspondiente que tiene la secuencia de SEQ ID NO: 1.
2. El polipéptido de PglB oligosacaryltransferasa o fragmento funcional del mismo de la reivindicación 1, donde la secuencia de aminoácidos de polipéptido de PglB oligosacaryltransferasa incluye la característica de que: al menos un residuo seleccionado del grupo que consiste en el aminoácido X78, X84, A155, X293, X300, X301, X306, X308, X462, X464, X479, X523 y X570 se sustituye por un aminoácido diferente al que se encuentra en esa posición en SEQ ID NO: 1.
3. El polipéptido OST de PglB o fragmento funcional del mismo de la reivindicación 2, donde la secuencia de aminoácidos incluye al menos una característica seleccionada del grupo que consiste en: X78 se muta a T; X84 se muta a W; X155 se muta a Q; X293 se muta a C; X300 se muta a L; X301 se muta a P o G; X306 se muta a H; X308 se muta a W; X462 se muta a W, N o T; X464 se muta a L; X479 se muta a M; X523 se muta a R; X570 se muta a R o V.
4. La OST de PglB de cualquiera de las reivindicaciones 2-3, donde la secuencia de aminoácidos incluye al menos una característica seleccionada de la lista que consiste en: el residuo que corresponde al aminoácido X301 de SEQ ID NO: 1 se muta a P; el residuo que corresponde al aminoácido X462 de SEQ ID NO: 1 se muta a N o W y el residuo que corresponde al aminoácido X479 de SEQ ID NO: 1 se muta a M.
5. La OST de PglB de cualquiera de las reivindicaciones 2-4 que contiene al menos, 2, 3, 4, 5 o 6 de las características del aminoácido que corresponde a X300 de SEQ ID NO: 1 se muta a L; el aminoácido que corresponde a X301 de SEQ ID NO: 1 se muta a P, el aminoácido que corresponde a X308 de SEQ ID NO: 1 se muta a W, el aminoácido que corresponde a X462 de SEQ ID NO: 1 se muta a W, el aminoácido que corresponde a X479 de SEQ ID NO: 1 se muta a M y el aminoácido que corresponde a X570 de SEQ ID NO: 1 se muta a R.
6. El polipéptido de PglB oligosacaryltransferasa o fragmento funcional del mismo de cualquiera de reivindicaciones 1-5, donde el residuo que corresponde al aminoácido X77 de SEQ ID NO: 1 es R y el residuo que corresponde al aminoácido X311 de SEQ ID NO: 1 es V.
7. El polipéptido de PglB oligosacaryltransferasa o fragmento funcional del mismo de cualquiera de las reivindicaciones 1-6, donde los residuos que corresponden a los aminoácidos X57, X462 y X479 de SEQ ID NO: 1 se sustituyen con un aminoácido diferente al que se encuentra en esa posición en SEQ ID NO: 1; opcionalmente a A57R, Y462W y H479M.
8. El polipéptido de PglB oligosacaryltransferasa de cualquiera de las reivindicaciones 1-7, donde el polipéptido de PglB oligosacaryltransferasa es de longitud completa, opcionalmente con una longitud de 712, 713 o 714 aminoácidos.
9. Una PglB de *Campylobacter coli* (PglB_{C. coli}) que tiene una secuencia de aminoácidos al menos 85 %, 90 %, 95 %, 97 %, 98 %, 99 % o 100 % idéntica a SEQ ID NO: 12, 13 o 14, donde el residuo que corresponde al aminoácido X57 de SEQ ID NO: 12 se sustituye con un aminoácido diferente al que se encuentra en esa posición en SEQ ID NO: 12; opcionalmente a A57R y donde la PglB es capaz de incrementar el rendimiento de glucosilación de la proteína con el sacárido para producir una proteína glucosilada por al menos 1,5 veces, en comparación con una PglB oligosacaryltransferasa correspondiente que tiene la secuencia de SEQ ID NO: 1.
10. Un polinucleótido que codifica para un polipéptido de PglB oligosacaryltransferasa mutado como se reivindica en cualquiera de las reivindicaciones anteriores.
11. Una composición o célula hospedadora (por ejemplo, una célula hospedadora procariota o una célula hospedadora de *E. coli*) que comprende al menos una PglB oligosacaryltransferasa de cualquiera de las reivindicaciones 1-9 o el polinucleótido de la reivindicación 10.
12. Un proceso para preparar una proteína glucosilada, que comprende los pasos de:
- (a) cultivar la célula hospedadora de la reivindicación 11 bajo condiciones adecuadas para la producción de proteínas; y

(b) aislar la proteína glucosilada de la célula hospedadora.

- 5 13. Un uso de la PglB oligosacaryltransferasa o fragmento funcional de la misma de cualquiera de las reivindicaciones 1-9 en la producción de una proteína glucosilada en la cual un sacárido se une a un residuo N de una secuencia de consenso de glucosilación, que comprende la secuencia de aminoácidos Asp/Glu-Z₁-Asn-Z₂-Ser/Thr donde Z₁ y Z₂ pueden ser cualquier aminoácido natural excepto Pro, de una proteína.
- 10 14. El uso de la reivindicación 13, donde la PglB oligosacaryltransferasa o fragmento funcional de la misma es capaz de incrementar el rendimiento de glucosilación de la proteína con el sacárido para producir una proteína glucosilada por al menos 1,5 veces, 2 veces, 3 veces, 5 veces, 10 veces, 20 veces, 50 veces, 100 veces, 200 veces, 500 veces, 700 veces, 1000 veces en comparación con una PglB oligosacaryltransferasa correspondiente que tiene la secuencia de SEQ ID NO: 1.
- 15 15. El uso de la PglB oligosacaryltransferasa de la reivindicación 13 o 14, donde un residuo de glucosa en el sacárido se une covalentemente al residuo N de la secuencia de glucosilación.

Figura 1A

Veces de mejora para cada ronda



Trayectoria de Evolución

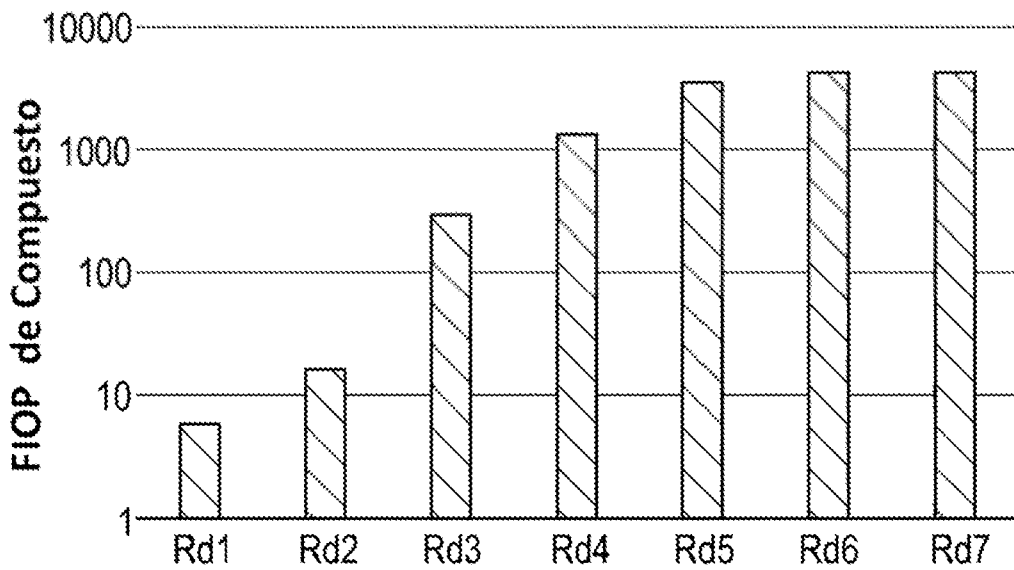


Figura 1B

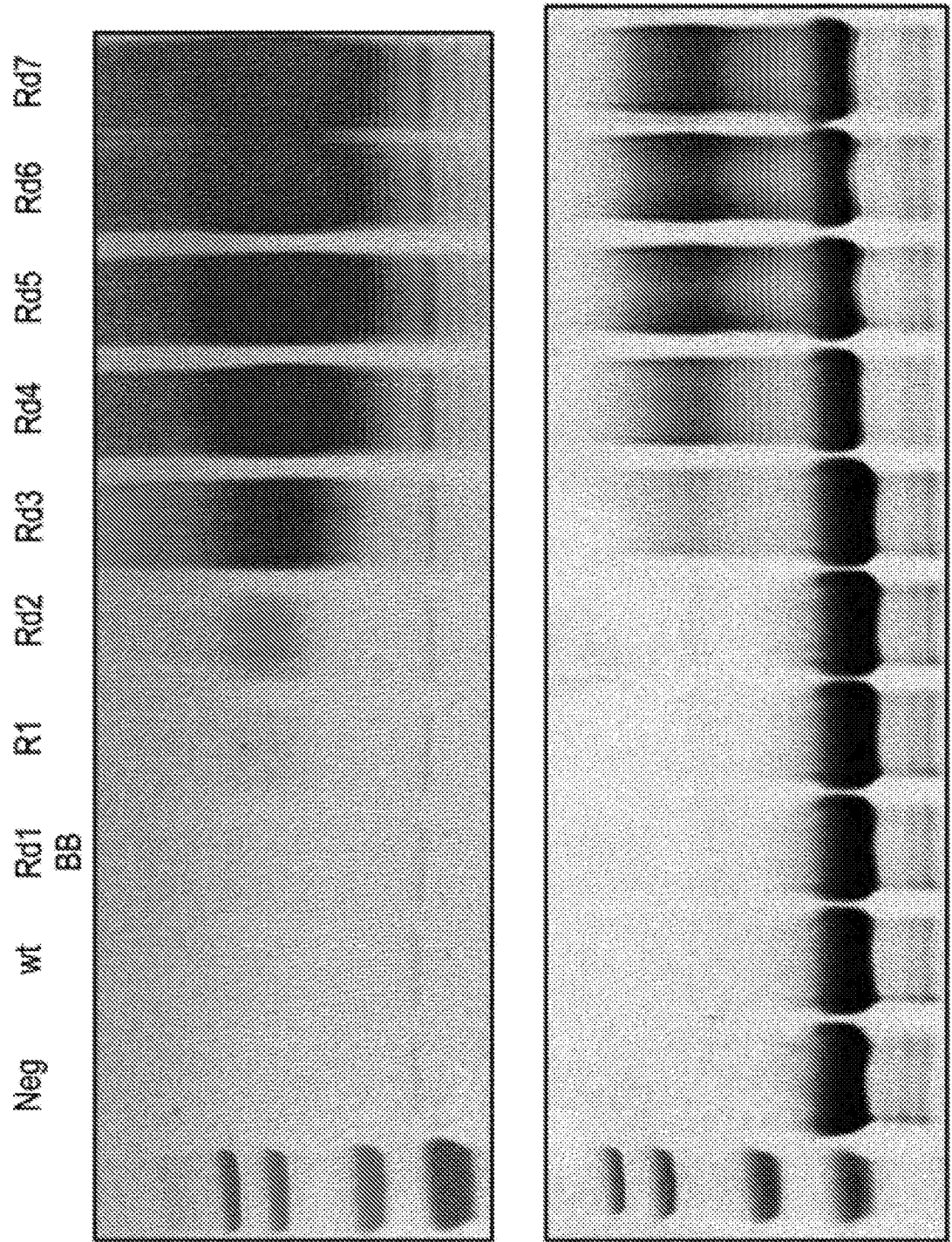


Figura 2

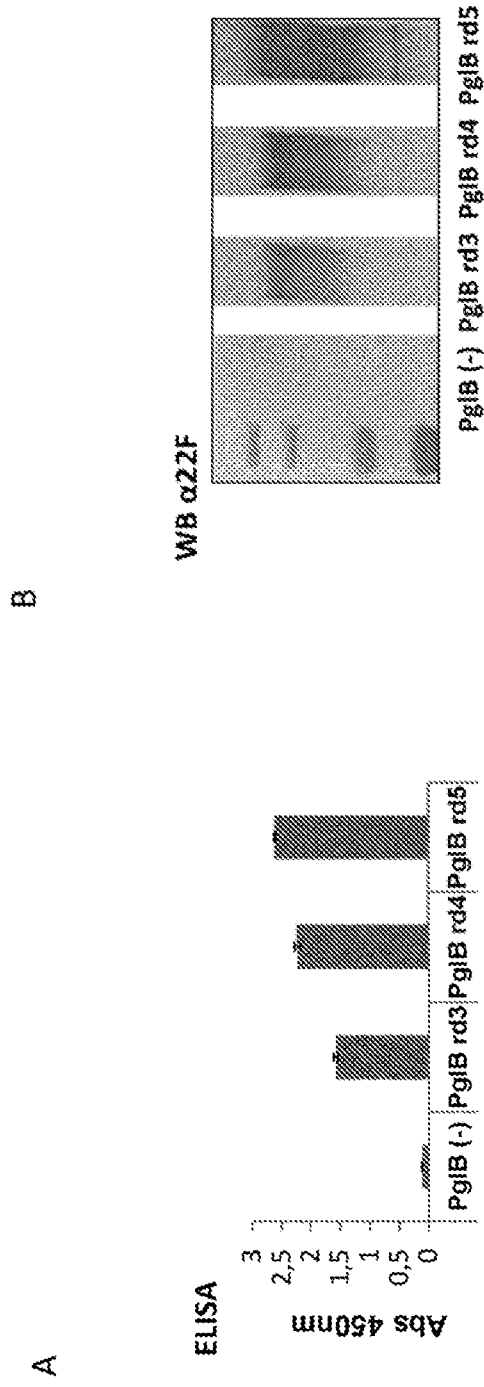


Figura 3

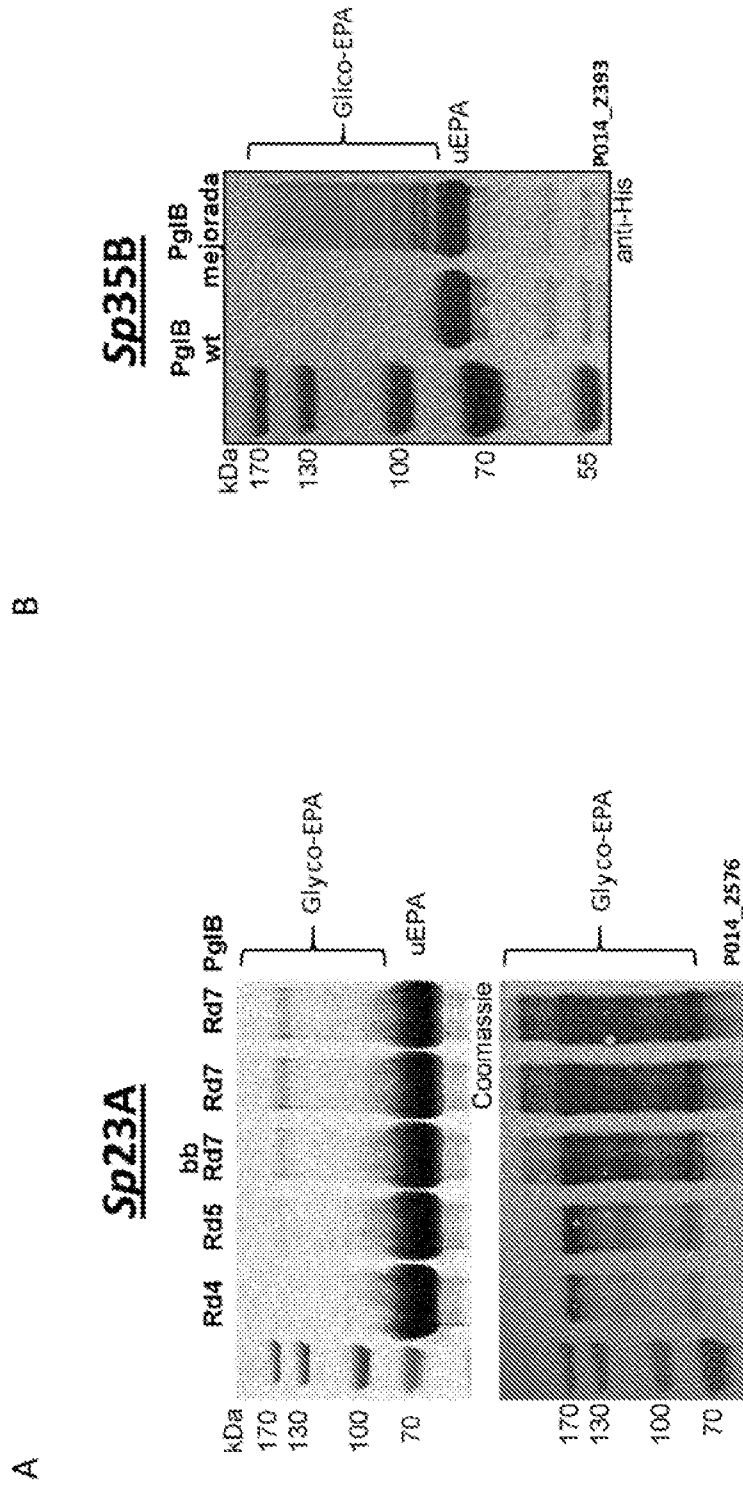


Figura 4

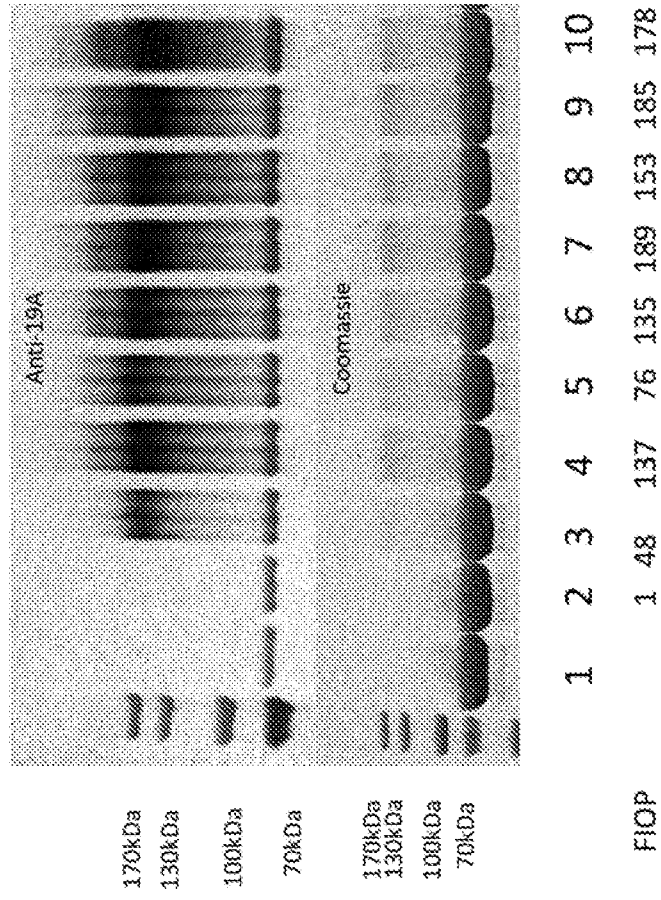


Figura 5

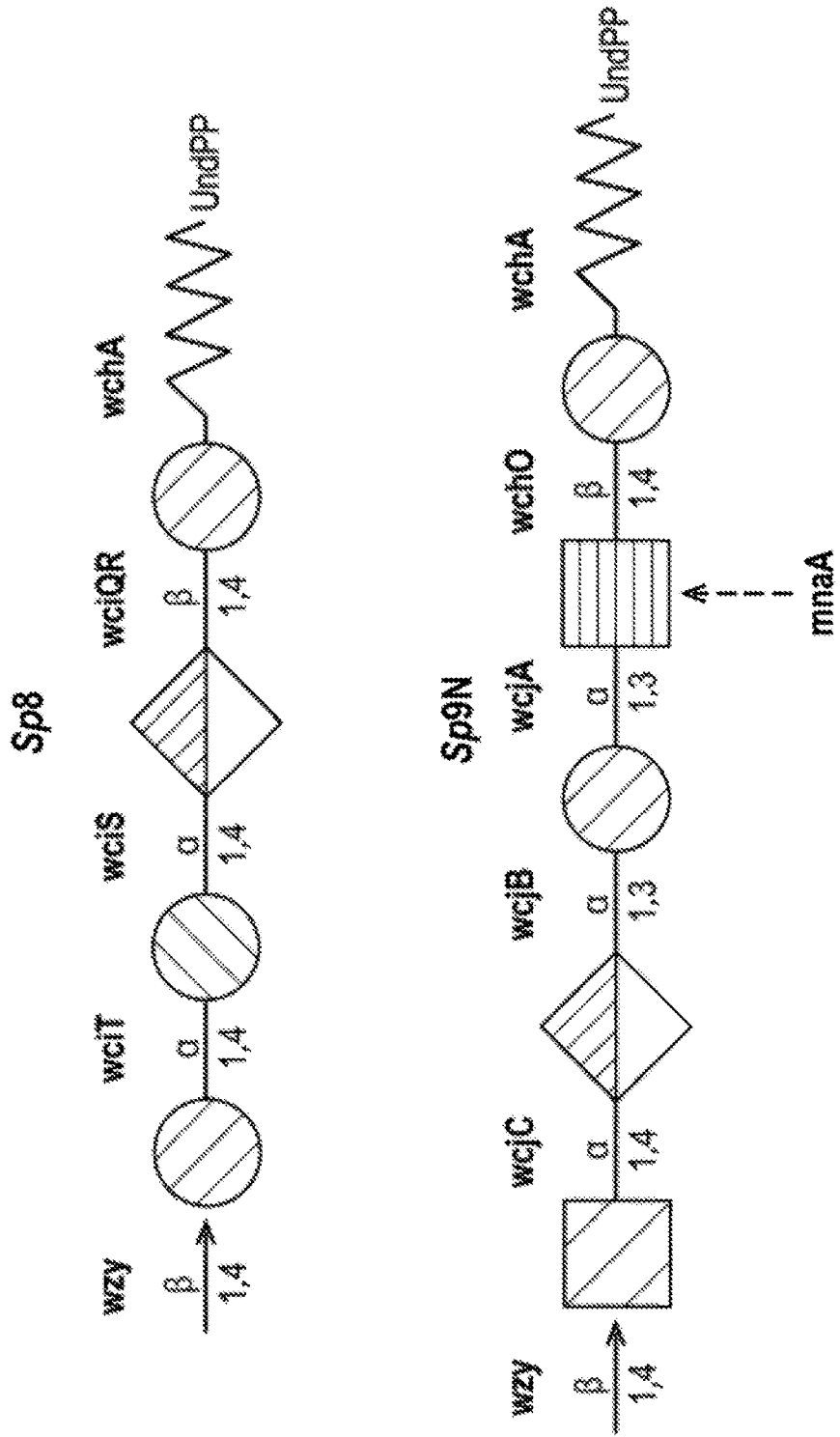
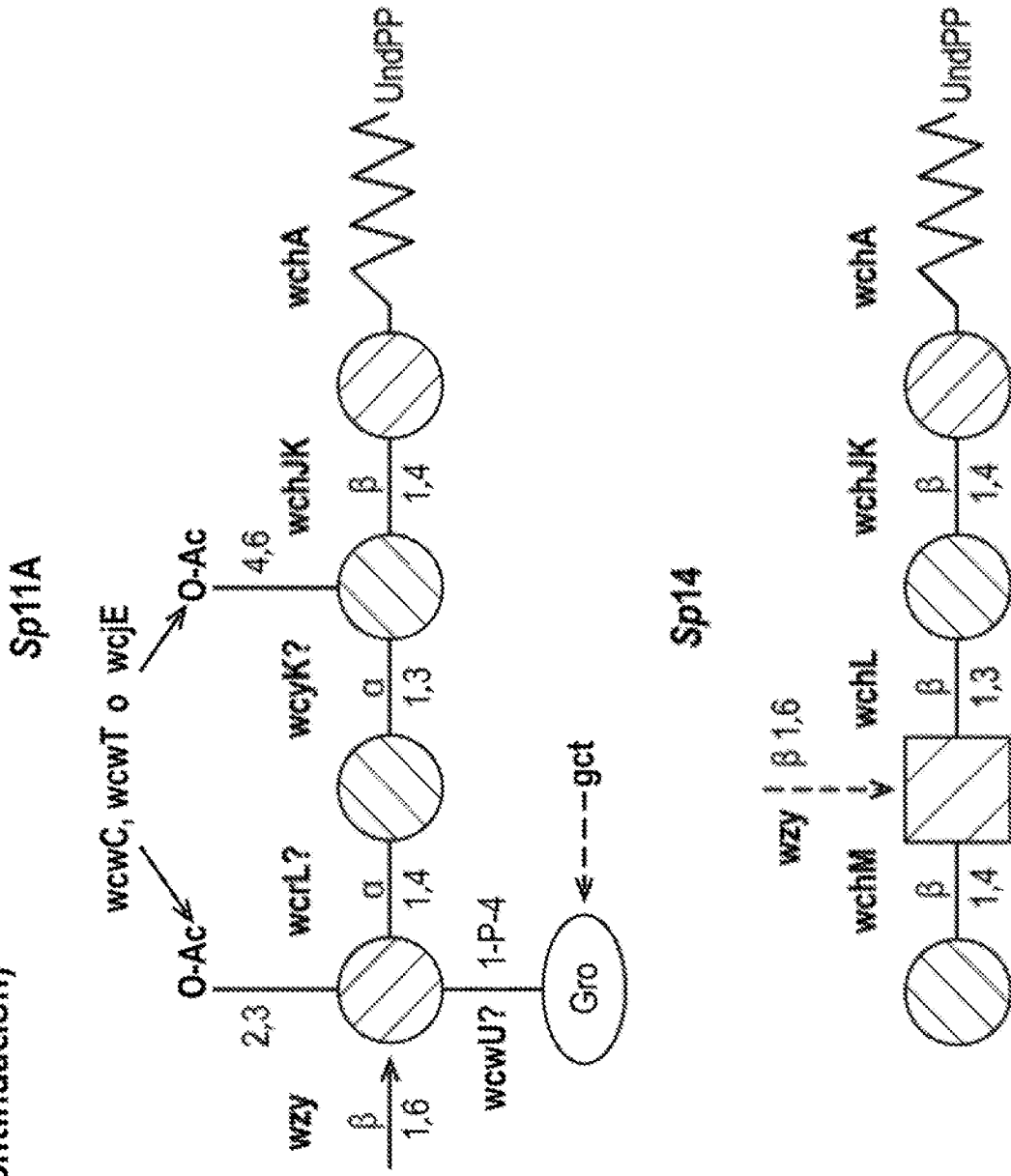


Figura 5 (Continuación)



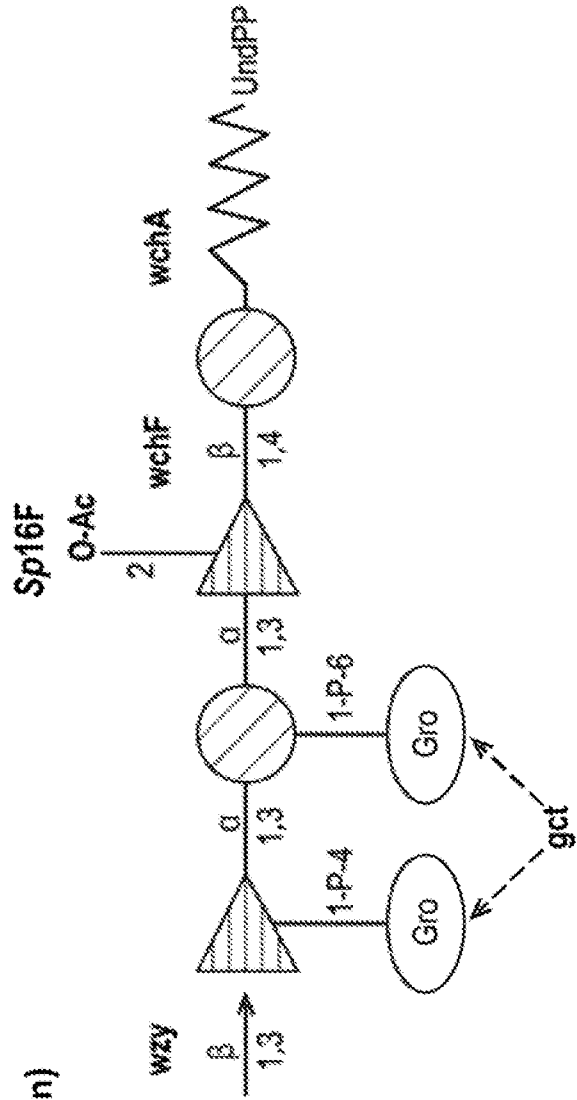
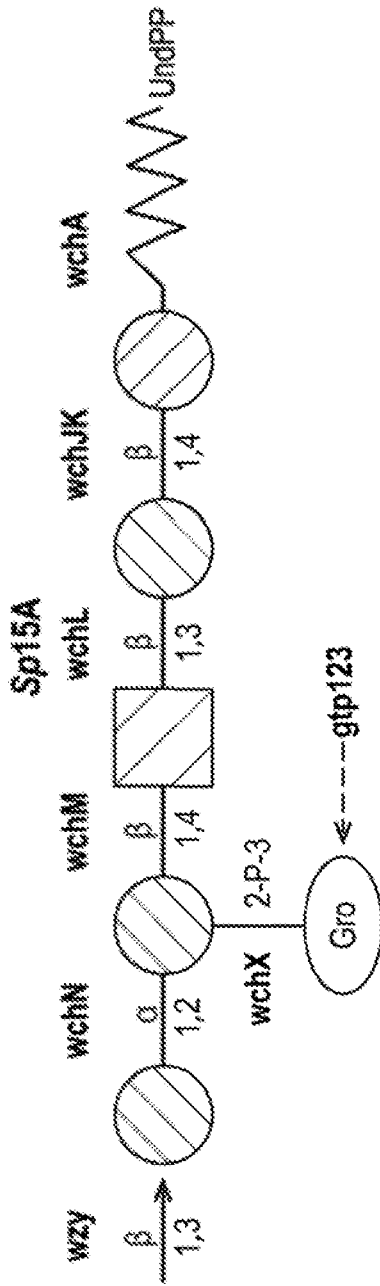


Figura 5 (Continuación)

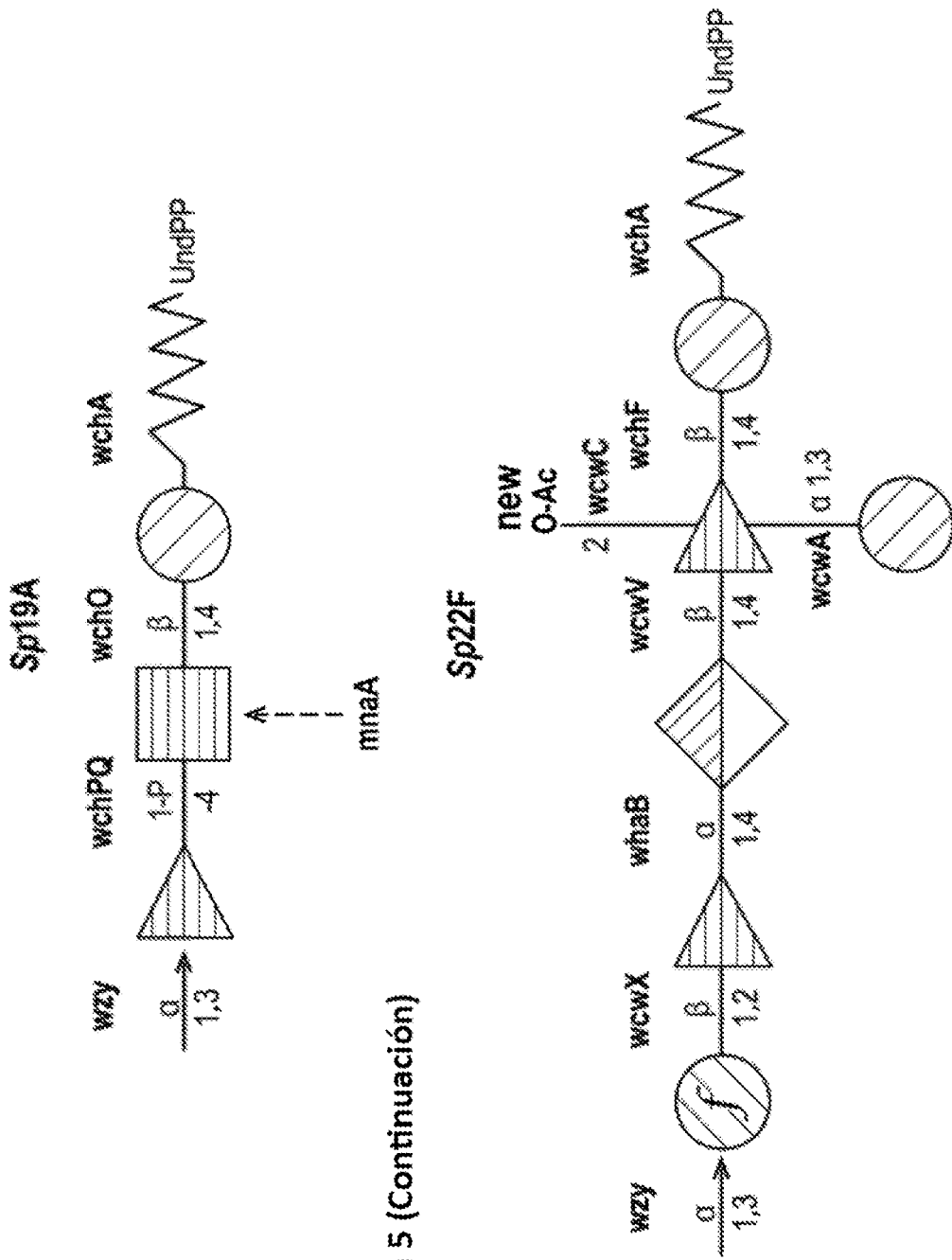


Figura 5 (Continuación)

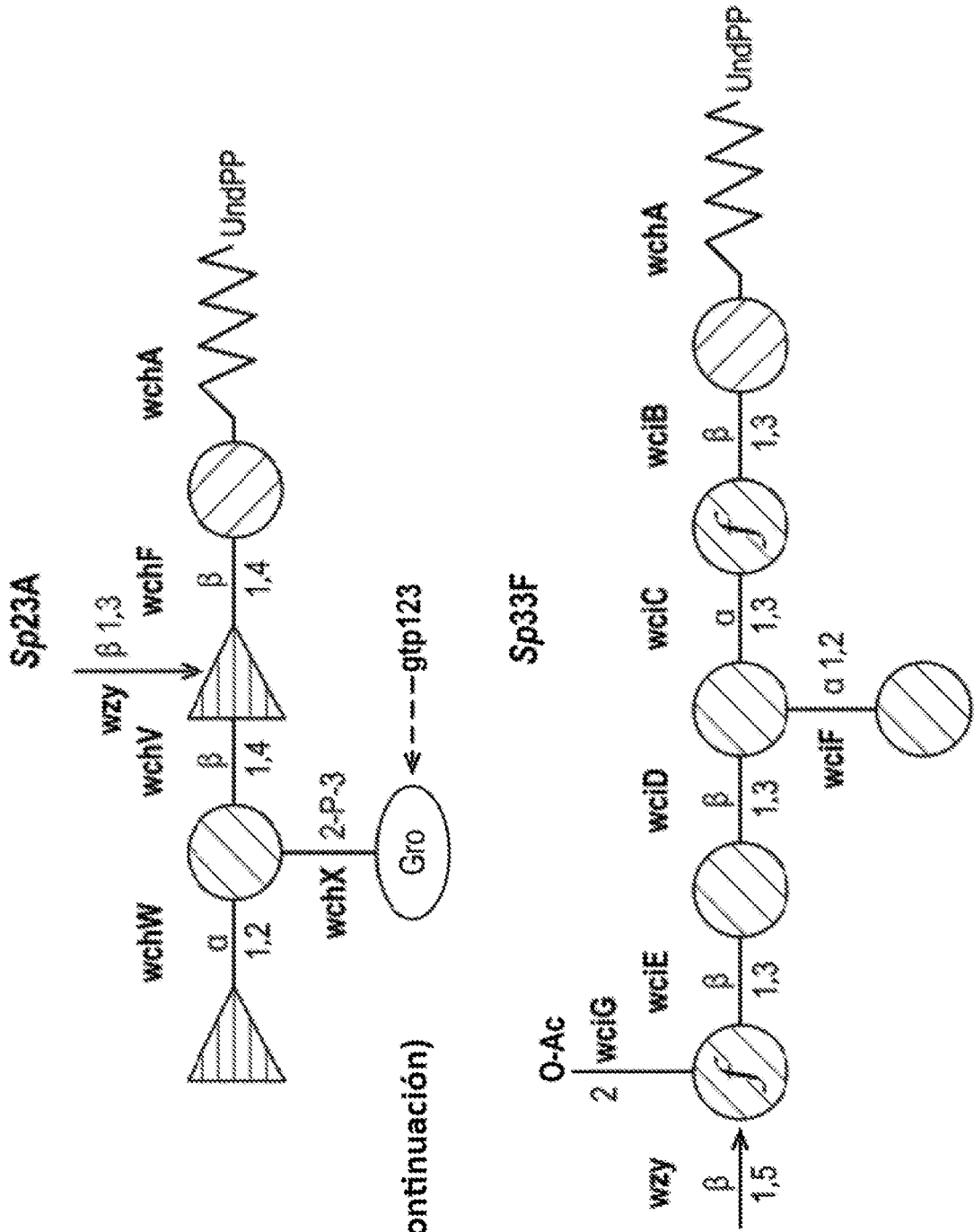


Figura 5 (Continuación)

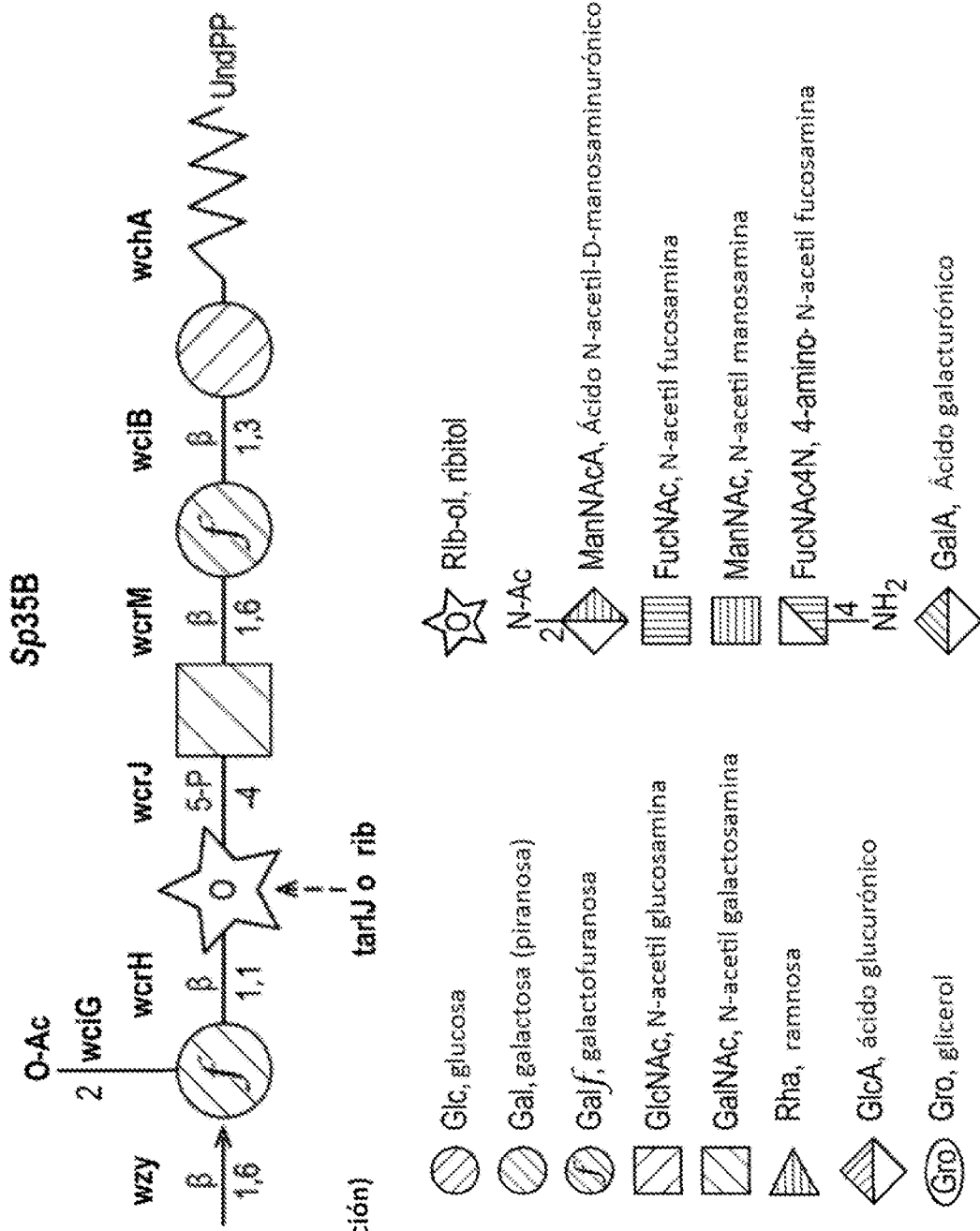


Figura 5 (Continuación)

Figura 6

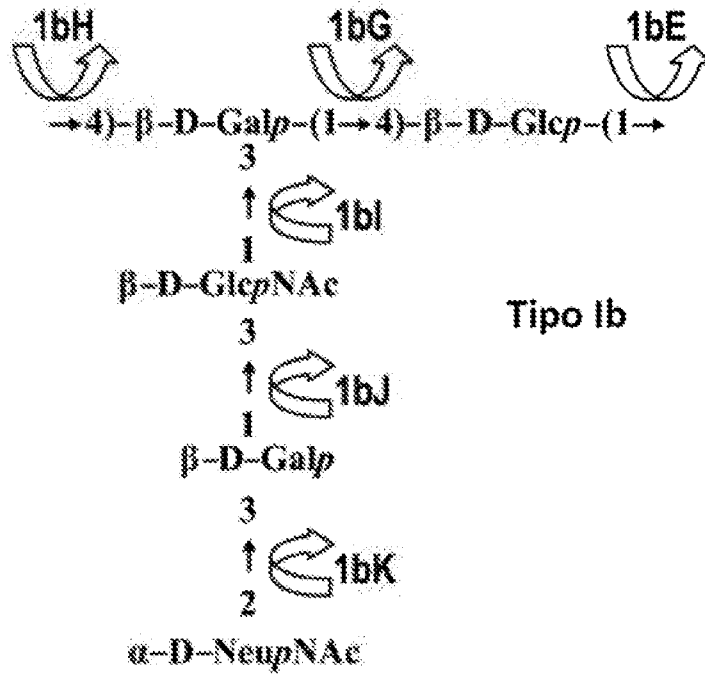
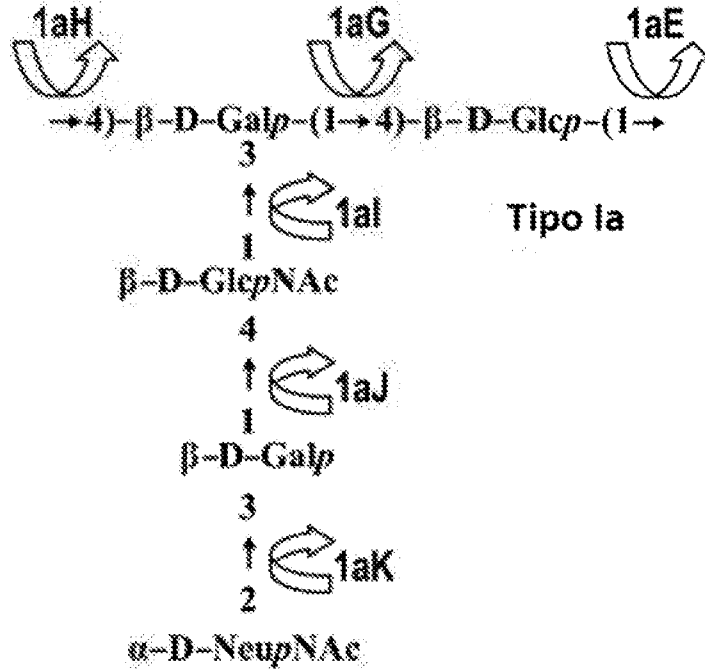


Figura 6 (Continuación)

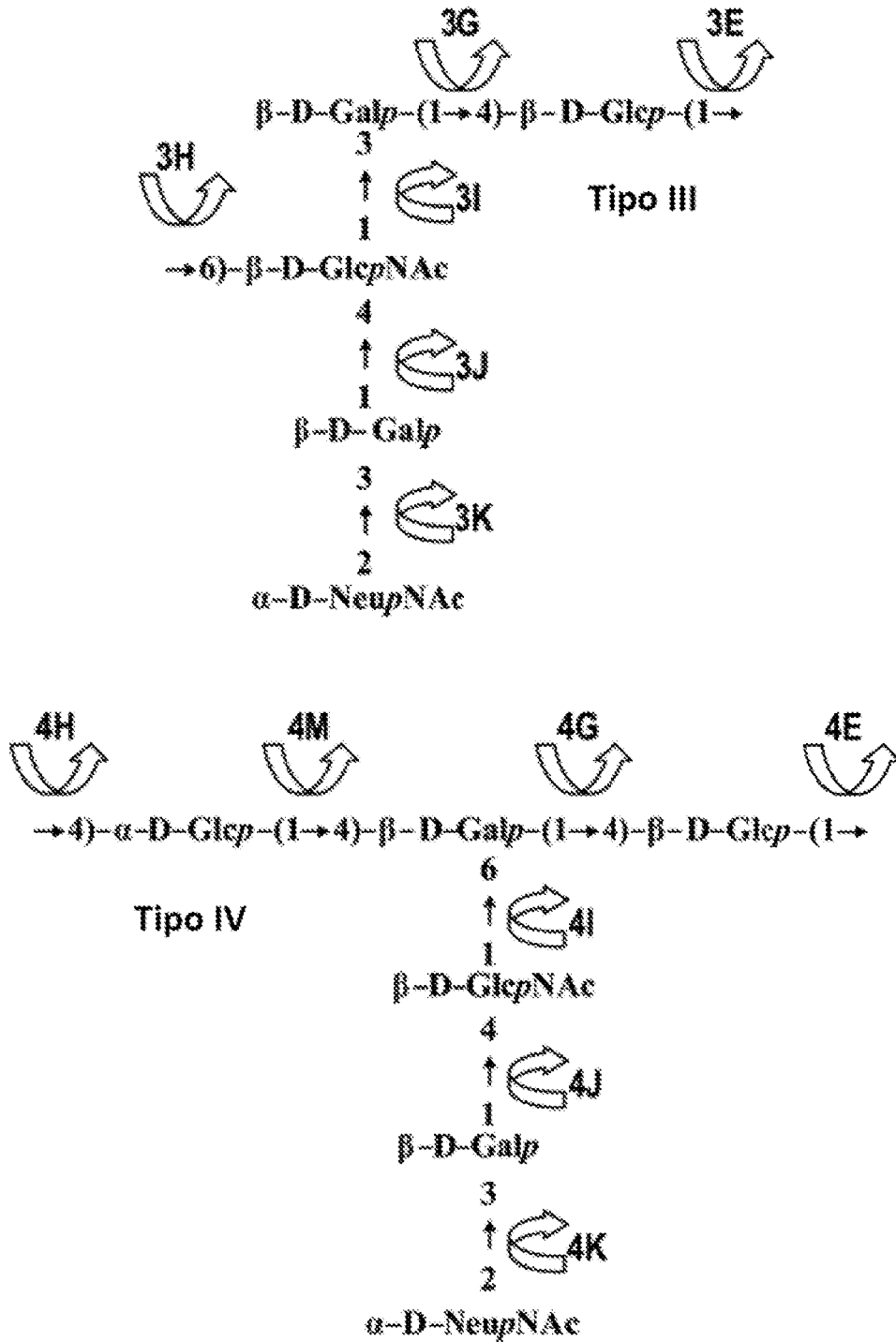


Figura 6 (Continuación)

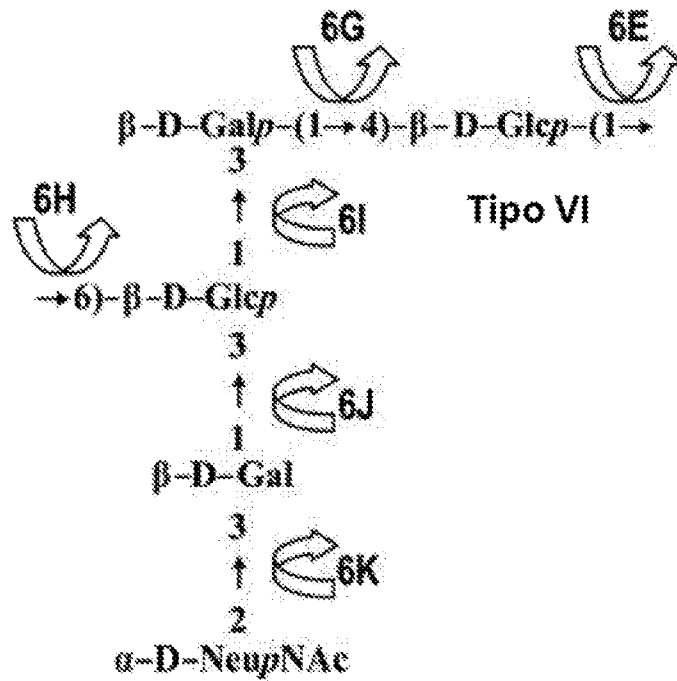
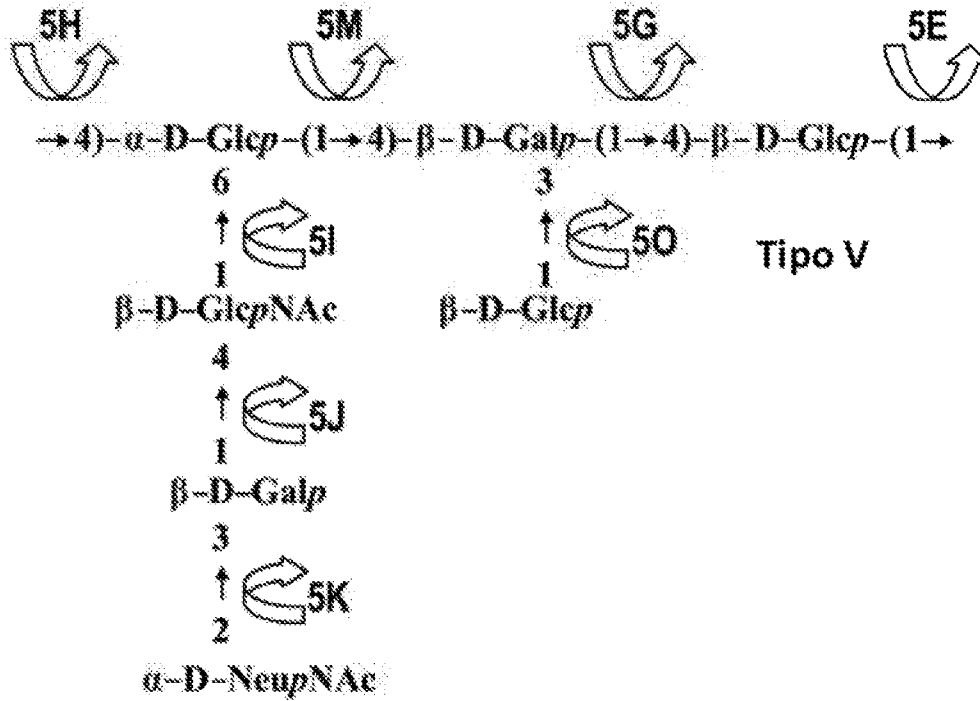
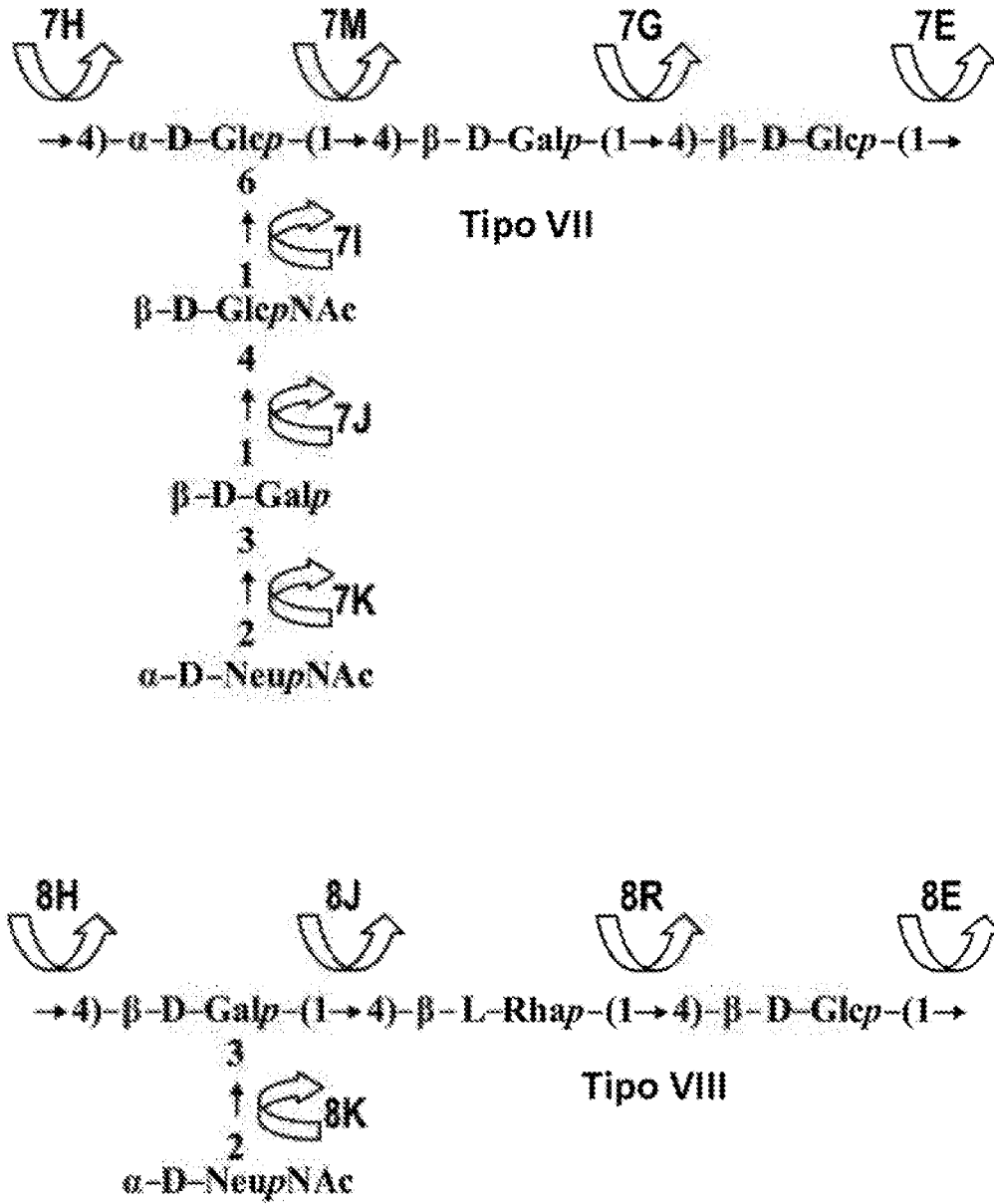


Figura 6 (Continuación)



Cieslewicz et al., Structural and Genetic Diversity of Group B Streptococcus Capsular Polysaccharides, I&I, 2005 May; 73(5):3096-103.
 DOI: <https://doi.org/10.1128/IAI.73.5.3096-3103.2005>

Figura 7 (Continuación)

Campylobacter_coli_76339.3S	QSAIVTLFSLVLPQKRFESFALLGILGLATLVFLLSGVDFPILYQLKFIYFRSDESAN	300
Campylobacter_coli_jv20.ASM14683v1	QSAIVTLFSLVLPQKRFESFALLGILGLATLVFLLSGGIDPILYQLKFIYFRSDESAN	300
Campylobacter_coli_317_04.ASM25395v3	QSAIVTLFSLVLPQKRFESFALLGILGLATLVFLLSGGIDPILYQLKFIYFRSDESAN	300
	*****;*****;*****;*****;*****;*****	
Campylobacter_coli_76339.3S	LAQGFHYFNVNTIQEVESIDLSIFNQRISGSELVFFISLIGFIFLVRKHKSMILALPML	360
Campylobacter_coli_jv20.ASM14683v1	LAQGFHYFNVNTIQEVESIDLSIFNQRISGSELVFFISLIGFIFLVRKHKSMILALPML	360
Campylobacter_coli_317_04.ASM25395v3	LAQGFHYFNVNTIQEVESIDLSIFNQRISGSELVFFISLIGFIFLVRKHKSMILALPML	360
	*****;*****;*****;*****;*****;*****	
Campylobacter_coli_76339.3S	ALGFLALKGGLRFTIYAVPVIALGFGFLMSLLQERKQKQNYNA--SVSIFITLSLIP	417
Campylobacter_coli_jv20.ASM14683v1	ALGFLALKSGLRFTIYAVPVIALGFGFLMSLLQERKQKQNTYWNANIGVFFITLSLIP	420
Campylobacter_coli_317_04.ASM25395v3	ALGFLALKSGLRFTIYAVPVIALGFGFLMSLLQERKQKQNTYWNANIGVFFITLSLIP	420
	*****,* *****;*****;***** **;* * * * *****;*	
Campylobacter_coli_76339.3S	MFYHINNYKAPTFSQNEASKLDELKKTAAQREDYVVAWWDYGYPIRYYSVDVKTLADGSKH	477
Campylobacter_coli_jv20.ASM14683v1	MFYHINNYKAPTFSQNEATKLDLAKTAAQREDYVVTWWDYGYPIRYYSVDVKTLADGSKH	480
Campylobacter_coli_317_04.ASM25395v3	MFYHINNYKAPTFSQNEATKLDLAKTAAQREDYVVAWWDYGYPIRYYSVDVKTLADGSKH	480
	**** * *****;*****;*****;*****;*****;*****	

