

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 943 019**

51 Int. Cl.:

C07K 14/195 (2006.01)

C12N 15/63 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **11.07.2018 PCT/EP2018/068847**

87 Fecha y número de publicación internacional: **24.01.2019 WO19016052**

96 Fecha de presentación y número de la solicitud europea: **11.07.2018 E 18740570 (9)**

97 Fecha y número de publicación de la concesión europea: **29.03.2023 EP 3655422**

54 Título: **Promotor para la expresión heteróloga**

30 Prioridad:

21.07.2017 EP 17182615

22.12.2017 EP 17209978

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

08.06.2023

73 Titular/es:

BASF SE (100.0%)

Carl-Bosch-Strasse 38

67056 Ludwigshafen am Rhein, DE

72 Inventor/es:

FELLE, MAX FABIAN y

TOKOVENKO, BOGDAN

74 Agente/Representante:

GONZÁLEZ PECES, Gustavo Adolfo

Observaciones:

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

ES 2 943 019 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

Promotor para la expresión heteróloga

Campo de la invención

5 La presente invención se refiere a una construcción de ácido nucleico que comprende una secuencia promotora de un operón que comprende un gen *secA*, que es adecuado para conducir la expresión de polinucleótidos heterólogos. Además, la presente invención se relaciona con un vector de expresión y una célula hospedante que comprende dicha construcción de ácido nucleico y procedimientos para expresar polinucleótidos utilizando la secuencia promotora.

Antecedentes de la invención

10 Hoy en día, los microorganismos se aplican con profusión en la industria aprovechando su capacidad de fermentación. Los microorganismos se utilizan especialmente como hospedantes para la producción fermentativa de diversas sustancias, tales como enzimas, proteínas, productos químicos, azúcares y polímeros. Para ello, los microorganismos se someten a ingeniería genética con el fin de adaptar su expresión génica a las exigencias del proceso de producción específico.

15 Un elemento clave para dirigir la expresión génica en una célula hospedante es la secuencia promotora. Para que se produzca la expresión de un gen, la ARN polimerasa debe unirse a la secuencia promotora próxima a un gen. Así, los promotores contienen secuencias de ADN específicas que proporcionan un sitio de unión para la ARN polimerasa y también para otras proteínas que reclutan a la ARN polimerasa hacia la secuencia de reconocimiento (es decir, los factores de transcripción). En las bacterias, el promotor suele ser reconocido por la ARN polimerasa y un factor sigma asociado, que son guiados hasta el ADN promotor por la unión de una proteína activadora a su propio sitio de unión
20 al ADN cercano. En los eucariotas, el proceso es más complicado y se necesitan diversos factores para la unión de una ARN polimerasa al promotor. Influidos por la secuencia de ácidos nucleicos, los promotores pueden conferir niveles de expresión bajos, moderados o altos y pueden ser constitutivos o inducibles.

Debido a su papel central en la expresión génica, existe una necesidad constante de secuencias nuevas que sean adecuadas como promotor para la expresión de transgenes en una célula hospedante en condiciones específicas.

25 En concreto, para los productos transgénicos que no son tolerados por la célula hospedante en cantidades elevadas, se necesitan secuencias promotoras que proporcionen un nivel de expresión moderado. Los ejemplos de estos productos transgénicos son los componentes del sistema de modificación por restricción, tales como las endonucleasas de restricción y las metiltransferasas, las recombinasas y las proteínas que confieren resistencia a los antibióticos.

30 Los presentes inventores han descubierto que el uso de un promotor *secA* para la expresión de un transgén en una célula hospedante mejora enormemente los resultados de expresión, en concreto si se desean niveles de expresión moderados del transgén.

Breve resumen de la invención

35 La presente invención se dirige a un tipo concreto de secuencias promotoras y a construcciones de ácido nucleico que comprenden dichas secuencias promotoras.

Específicamente, la presente invención se dirige a una construcción de ácido nucleico que comprende un polinucleótido unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en una célula hospedante, en la que al menos una secuencia de control comprende una secuencia promotora de un operón que comprende un gen *secA* y en la que dicha secuencia promotora es heteróloga con respecto al polinucleótido, en la que el polinucleótido no comprende un gen indicador y la secuencia promotora tiene al menos un
40 70 % de identidad de secuencia con SEQ ID NO:9.

En otra realización, la invención se dirige a un vector de expresión y a una célula hospedante que comprende la construcción de ácido nucleico que comprende la secuencia promotora descrita en el presente documento.

45 Otra realización de la presente invención es un procedimiento para expresar un polinucleótido, que comprende las etapas de (a) proporcionar una célula hospedante que comprende una construcción de ácido nucleico heteróloga que comprende un promotor como se describe en el presente documento; (b) cultivar la célula hospedante recombinante de la etapa (a) en condiciones que conducen a la expresión del polinucleótido; y (c) opcionalmente, recuperar una proteína de interés codificada por el polinucleótido.

50 La secuencia promotora descrita en el presente documento es especialmente útil para la expresión de polinucleótidos que requieren un nivel de expresión moderado en una célula hospedante.

Breve descripción de las figuras

La figura 1 muestra secciones pertinentes de un alineamiento múltiple de secuencias basado en la estructura de las estructuras predichas de las metiltransferasas divulgadas en el presente documento, que pueden crearse utilizando programas convencionales de modelización de la homología, tal como el servidor web SWISS-MODEL (Biasini M., Bienert S., Waterhouse A., Arnold K., Studer G., Schmidt T., Kiefer F., Cassarino T.G., Bertoni M., Bordoli L., Schwede T. (2014), SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information, *Nucleic Acids Research*, 2014 (1 de julio de 2014), 42 (W1): W252-W258) utilizando los parámetros por defecto y las siguientes plantillas estructurales de la base de datos RCSB PDB (Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. (2000), *The Protein Data Bank Nucleic Acids Research*, 28: 235-242): 2uyc_A (utilizado para SEQ ID NO:33 (M.Fnu4HI), SEQ ID NO:33 (M.RBH3250), SEQ ID NO:41 (M.CocII)), 2i9k_C (utilizado para SEQ ID NO:33 (M.Bsp6I), SEQ ID NO:43 (M.LlaDII)), 3swr_A (utilizado para SEQ ID NO:33 (M.Cdi13307II), SEQ ID NO:38 (M.Cdi630IV)), 1mht_C (utilizado para SEQ ID NO:39 (M.Ckr177III)), 2z6u_A (utilizado para SEQ ID NO:40 (M.CmaLM2II)) y 9mht_C (utilizado para SEQ ID NO:42 (M.Fsp4HI)). Las estructuras predichas se alinearon estructuralmente con la estructura predicha de SEQ ID NO:33 (M.Fnu4HI) con TAlign, versión 20160521 (Y. Zhang, J. Skolnick (2005), TM-align: A protein structure alignment algorithm based on TM-score, *Nucleic Acids Research*, 33: 2302-2309) utilizando los parámetros por defecto. Los alineamientos estructurales por pares se combinaron en un alineamiento múltiple de secuencias utilizando MAFFT, versión 7.221 (Kato, S. (2013), MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Molecular Biology and Evolution*, 30:772-780) utilizando los parámetros por defecto del modo de fusión. La anotación de la estructura secundaria se añadió a la figura como consenso de las predicciones estructurales. Las posiciones conservadas importantes están resaltadas en columnas grises e identificadas con letras griegas.

La figura 2 muestra un gel de agarosa al 0,8 % de una digestión de restricción SatI de plásmidos aislados de diferentes cepas de *E. coli* que contienen MTasa (ejemplo 1). Carril 1: Marcador de ADN - Escalera de ADN Generuler de 1 kb (ThermoFisher Scientific). Carril 2: Ec n.º 83, Carril 3: Ec n.º 84, Carril 4: Ec n.º 85, Carril 5: Ec n.º 86, Carril 6: Ec n.º 87, Carril 7: Ec n.º 88. Carril 8: Ec n.º 82. Carril 10: Marcador de ADN.

La figura 3 muestra las eficiencias relativas de transformación en células de *B. licheniformis* ATCC 53926 de ADN plasmídico aislado de diferentes cepas de *E. coli*, tal como se describe en el ejemplo 2. La cepa de *E. coli* Ec n.º 083 porta la ADN metiltransferasa de *B. licheniformis* ATCC 53926 y se ajustó al 100 %. La cepa de *E. coli* Ec n.º 084 porta una variante codones optimizados de la ADN metiltransferasa de *B. licheniformis* ATCC 53926 y actúa como control para la expresión génica. El ADN plasmídico de *E. coli* Ec n.º 082 que no fue metilado por una ADN metiltransferasa específica de GCNGC no recuperó ninguna transformante. El ADN plasmídico aislado de cepas de *E. coli* portadoras de MTasas de estructura similar y heterólogas con *B. licheniformis* ATCC 53926 (Ec n.º 85-87) transformado en *B. licheniformis* ATCC 53926 dio lugar a eficiencias de transformación significativamente mayores. El ADN plasmídico aislado de la cepa de *E. coli* portadora de la MTasa homóloga (Ec n.º 88) de *B. licheniformis* ATCC 53926 con una delección de los aminoácidos 103-108 de SEQ ID NO:34 (se truncaron 6 aminoácidos en total, dando como resultado SEQ ID NO:35) también dio lugar a una eficiencia de transformación significativamente mayor en comparación con Ec n.º 83.

La figura 4 muestra las eficiencias relativas de transformación en células de *B. licheniformis* ATCC 53926 de ADN del plásmido pUK56 aislado de cepas de *B. subtilis* Bs n.º 54 y Bs n.º 55 como se describe en el ejemplo 3. La eficiencia de transformación del ADN del plásmido pUK56 aislado de *B. licheniformis* Bli n.º 112, portador del patrón de metilación del ADN de *B. licheniformis* ATCC 53926, se fijó en el 100 %.

La figura 5 muestra las eficiencias relativas de transformación en células de *B. licheniformis* ATCC 53926 de ADN plasmídico que comprende diferentes promotores aislados de diferentes cepas de *E. coli*, tal como se describe en el ejemplo 3. La eficiencia de transformación del ADN plasmídico de la cepa Ec n.º 85 de *E. coli* se fijó en el 100%. Las construcciones que incluían promotores SecA tuvieron la mayor eficiencia de transformación.

Descripción detallada de la invención

La presente invención puede entenderse más fácilmente remitiéndose a la siguiente descripción detallada de las realizaciones de la invención y los ejemplos incluidos en el presente documento.

Definiciones

A menos que se indique lo contrario, los términos y las expresiones utilizados en el presente documento deben entenderse de acuerdo con el uso convencional por parte de los expertos en la materia.

Debe entenderse que, tal como se utiliza en la memoria descriptiva y en las reivindicaciones, "un" o "una" puede significar uno o más, dependiendo del contexto en el que se utilice. Así, la referencia a "una célula" puede significar que puede utilizarse al menos una célula.

A lo largo de esta solicitud, se remite a diversas publicaciones con el fin de describir con más detalle el estado de la técnica al que pertenece la presente invención.

La secuencia "parental" (por ejemplo, "enzima parental" o "proteína parental") es la secuencia de partida para la introducción de cambios (por ejemplo, la introducción de una o más sustituciones de aminoácidos) de la secuencia que dan lugar a "variantes" de las secuencias parentales. Así, las expresiones "variante de enzima", "variante de secuencia" o "variante de proteína" se utilizan en referencia a las enzimas parentales que son el origen de las respectivas enzimas variantes. Por lo tanto, las enzimas parentales incluyen enzimas de tipo salvaje y variantes de enzimas de tipo salvaje que se utilizan para el desarrollo de otras variantes. Las enzimas variantes difieren de las enzimas parentales en su secuencia de aminoácidos hasta cierto punto; sin embargo, las variantes mantienen al menos las propiedades enzimáticas de la enzima parental respectiva. En una realización, las propiedades enzimáticas se mejoran en las enzimas variantes en comparación con la enzima parental respectiva. En una realización, las enzimas variantes tienen al menos la misma actividad enzimática en comparación con la enzima parental respectiva o las enzimas variantes tienen una mayor actividad enzimática en comparación con la enzima parental respectiva.

Cuando se describen variantes de proteínas, se utilizan las abreviaturas convencionales de aminoácidos de una o tres letras de la IUPAC.

Las "sustituciones" se describen proporcionando el aminoácido original seguido del número de la posición dentro de la secuencia de aminoácidos, seguido del aminoácido sustituido. Por ejemplo, la sustitución de histidina en la posición 120 por alanina se indica como "His120Ala" o "H120A".

Las "delecciones" se describen proporcionando el aminoácido original, seguido del número de la posición dentro de la secuencia de aminoácidos, seguido de *. En consecuencia, la delección de glicina en la posición 150 se indica como "Gly150*" o G150*. Como alternativa, las delecciones se indican, por ejemplo, como "delección de D183 y G184".

Las "inserciones" se describen proporcionando el aminoácido original, seguido del número de la posición dentro de la secuencia de aminoácidos, seguido del aminoácido original y del aminoácido adicional. Por ejemplo, una inserción en la posición 180 de la lisina junto a la glicina se indica como "Gly180GlyLys" o "G180GK". Cuando se inserta más de un residuo de aminoácido, como, por ejemplo, una Lys y una Ala después de Gly180, esto puede indicarse como: Gly180GlyLysAla o G195GKA.

En los casos en los que se produce una sustitución y una inserción en la misma posición, esto puede indicarse como S99SD+S99A o, abreviado, S99AD. En los casos en los que se inserta un residuo de aminoácido idéntico al residuo de aminoácido existente, es evidente que se produce una degeneración en la nomenclatura. Si, por ejemplo, se inserta una glicina después de la glicina en el ejemplo anterior, se indicaría con G180GG. Las variantes que comprenden alteraciones múltiples se separan por "+", por ejemplo, "Arg170Tyr+Gly195Glu" o "R170Y+G195E" que representan una sustitución de arginina y glicina en las posiciones 170 y 195 por tirosina y ácido glutámico, respectivamente. Como alternativa, las alteraciones múltiples pueden separarse por un espacio o una coma, por ejemplo, R170Y G195E o R170Y, G195E respectivamente. Cuando pueden introducirse diferentes alteraciones en una posición, las diferentes alteraciones se separan por una coma, por ejemplo, "Arg170Tyr, Glu" y R170T, E, respectivamente, representan una sustitución de arginina en la posición 170 por tirosina o ácido glutámico. Como alternativa, pueden indicarse entre paréntesis diferentes alteraciones o sustituciones opcionales, por ejemplo, Arg170[Tyr, Gly] o Arg170{Tyr, Gly}, o abreviado R170 [Y, G] o R170 {Y, G}.

La numeración de los residuos de aminoácidos de la ADN metiltransferasa descrita en el presente documento está de acuerdo con la numeración de la ADN metiltransferasa Fnu4HI de *Fusobacterium nucleatum* 4HI como se muestra en SEQ ID NO:33 (es decir, según la numeración de SEQ ID NO:33).

Para las secuencias de nucleótidos, por ejemplo, las secuencias consenso, se utiliza la nomenclatura de nucleótidos de la IUPAC (Comité de Nomenclatura de la Unión Internacional de Bioquímica (NC-IUB), 1984, "Nomenclatura para bases incompletamente especificadas en secuencias de ácidos nucleicos"), con las siguientes definiciones de nucleótido y de ambigüedad de nucleótido, pertinentes para la presente invención: A, adenina; C, citosina; G, guanina; T, timina; K, guanina o timina; R, adenina o guanina; W, adenina o timina; M, adenina o citosina; Y, citosina o timina; D, no es una citosina; N, cualquier nucleótido.

Además, la notación "N(3-5)" significa que la posición de consenso indicada puede tener de 3 a 5 nucleótidos cualesquiera (N). Por ejemplo, una secuencia consenso "AWN(4-6)" representa 3 posibles variantes, con 4, 5 o 6 nucleótidos cualesquiera, que serían: AWNNNN, AWNNNNN, AWNNNNNN.

Las variantes de las moléculas enzimáticas parentales pueden tener una secuencia de aminoácidos que sea al menos n por ciento idéntica a la secuencia de aminoácidos de la enzima parental respectiva que tiene actividad enzimática, siendo n un número entero entre 50 y 100, preferentemente 50, 55, 60, 65, 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98 o 99 en comparación con la secuencia polipeptídica de longitud completa. Las enzimas variantes descritas en el presente documento que son n por ciento idénticas en comparación con una enzima parental tienen actividad enzimática.

Las variantes de enzimas pueden definirse por su identidad de secuencia cuando se comparan con una enzima parental. La identidad de secuencia se suele indicar como "% de identidad de secuencia" o "% de identidad". Para determinar el porcentaje de identidad entre dos secuencias de aminoácidos, en una primera etapa se genera un alineamiento de secuencias por pares entre esas dos secuencias, en el que las dos secuencias se alinean en toda su

5 longitud (es decir, un alineamiento global por pares). El alineamiento se genera con un programa que utiliza el algoritmo de Needleman y Wunsch (J. Mol. Biol. (1979), 48, págs. 443-453), utilizando preferentemente el programa "NEEDLE" (The European Molecular Biology Open Software Suite (EMBOSS)) con los parámetros por defecto del programa (apertura de hueco = 10,0, extensión de hueco = 0,5 y matriz = EBLOSUM62). El alineamiento preferido para los fines de la presente invención es el alineamiento a partir del cual puede determinarse la mayor identidad de secuencia.

10 Tras alinear las dos secuencias, en una segunda etapa se determinará un valor de identidad a partir del alineamiento. Por lo tanto, según la presente invención se aplica el siguiente cálculo de porcentaje de identidad: % de identidad = (residuos idénticos / longitud de la región de alineamiento que muestra la secuencia respectiva de la presente invención en toda su longitud) * 100. Así, la identidad de secuencia en relación con la comparación de dos secuencias de aminoácidos según esta realización se calcula dividiendo el número de residuos idénticos por la longitud de la región de alineamiento que muestra la secuencia respectiva de la presente invención en toda su longitud. Este valor se multiplica por 100 para obtener el "% de identidad".

15 Para calcular el porcentaje de identidad de dos secuencias de ADN se aplica lo mismo que para el cálculo del porcentaje de identidad de dos secuencias de aminoácidos con algunas condiciones. En el caso de las secuencias de ADN que codifican una proteína, el alineamiento por pares se realizará en toda la longitud de la región codificante, desde el codón de inicio hasta el codón de terminación, excluyendo los intrones. Para las secuencias de ADN que no codifican proteínas, el alineamiento por pares se realizará en toda la longitud de la secuencia de la presente invención, de modo que la secuencia completa de la presente invención se compara con otra secuencia o con regiones de otra secuencia. Además, el programa de alineamiento preferido que utiliza el algoritmo de Needleman y Wunsch (J. Mol. Biol. (1979), 48, págs. 443-453) es "NEEDLE" (The European Molecular Biology Open Software Suite (EMBOSS)) con los parámetros por defecto del programa (apertura de hueco = 10,0, extensión de hueco = 0,5 y matriz = EDNA-FULL).

20 El siguiente ejemplo pretende ilustrar las realizaciones de la invención y se realiza con dos secuencias de nucleótidos, pero los mismos cálculos se aplican a las secuencias de proteínas:

25 Secuencia A: Longitud de AAGATACTG: 9 bases
 Secuencia B: Longitud de GATCTGA: 7 bases

Por lo tanto, la secuencia más corta es la secuencia B.

La producción de un alineamiento global por pares que muestre ambas secuencias en toda su longitud da como resultado:

```

    Secuencia A: AAGATACTG-
                ||| |||
    Secuencia B: --GAT-CTGA
    
```

30 El símbolo "|" en el alineamiento indica residuos idénticos (lo que significa bases para el ADN o aminoácidos para las proteínas). El número de residuos idénticos es 6.

El símbolo "-" en el alineamiento indica huecos. El número de huecos introducidos por el alineamiento dentro de la secuencia B es 1. El número de huecos introducidos por el alineamiento en los bordes de la secuencia B es 2, y en los bordes de la secuencia A es 1.

35 La longitud de alineamiento que muestra las secuencias alineadas en toda su longitud es 10.

Por tanto, la producción de un alineamiento por pares que muestra la secuencia más corta en toda su longitud según la invención da como resultado:

```

    Secuencia A: GATACTG-
                ||| |||
    Secuencia B: GAT-CTGA
    
```

40 Por tanto, la producción de un alineamiento por pares que muestre la secuencia A en toda su longitud según la invención da como resultado:

```

    Secuencia A: AAGATACTG
                ||| |||
    Secuencia B: --GAT-CTG
    
```

Por tanto, la producción de un alineamiento por pares que muestre la secuencia B en toda su longitud según la invención da como resultado:

Secuencia A: GATACTG-

 ||| |||
 Secuencia B: GAT-CTGA

La longitud de alineamiento que muestra la secuencia más corta en toda su longitud es 8 (hay un hueco que se tiene en cuenta en la longitud de alineamiento de la secuencia más corta).

5 En consecuencia, la longitud de alineamiento que muestra la secuencia A en toda su longitud sería 9 (lo que significa que la secuencia A es la secuencia de la invención).

En consecuencia, la longitud de alineamiento que muestra la secuencia B en toda su longitud sería 8 (lo que significa que la secuencia B es la secuencia de la invención).

Según el ejemplo anterior, el porcentaje de identidad es: para la secuencia A que es la secuencia de la invención $(6/9) * 100 = 66,7 \%$; para la secuencia B que es la secuencia de la invención $(6/8) * 100 = 75 \%$.

10 Las variantes de enzimas pueden definirse por su similitud de secuencia en comparación con una enzima parental. La similitud de secuencias suele indicarse como "% de similitud de secuencias" o "% de similitud". Para calcular la similitud de secuencias, en una primera etapa debe generarse un alineamiento de secuencias como se ha descrito anteriormente. En una segunda etapa, hay que calcular el porcentaje de similitud, advirtiendo que el porcentaje de similitud de secuencia tiene en cuenta que determinados conjuntos de aminoácidos comparten propiedades similares, por ejemplo, su tamaño, su hidrofobicidad, su carga u otras características. En este caso, el intercambio de un aminoácido por otro similar se denomina "mutación conservadora". Las variantes de enzimas que comprenden mutaciones conservadoras parecen tener un efecto mínimo sobre el plegamiento de la proteína, lo que da lugar a que se mantengan sustancialmente determinadas propiedades enzimáticas en comparación con las propiedades enzimáticas de la enzima parental.

20 Para la determinación del porcentaje de similitud según la presente invención se aplica lo siguiente, que también está de acuerdo con la matriz BLOSUM62, que es una de las matrices de similitud de aminoácidos más utilizadas para la búsqueda en bases de datos y alineamientos de secuencias

El aminoácido A es similar al aminoácido S

El aminoácido D es similar a los aminoácidos E; N

25 El aminoácido E es similar a los aminoácidos D; K; Q

El aminoácido F es similar a los aminoácidos W; Y

El aminoácido H es similar a los aminoácidos N; Y

El aminoácido I es similar a los aminoácidos L; M; V

El aminoácido K es similar a los aminoácidos E; Q; R

30 El aminoácido L es similar a los aminoácidos I; M; V

El aminoácido M es similar a los aminoácidos I; L; V

El aminoácido N es similar a los aminoácidos D; H; S

El aminoácido Q es similar a los aminoácidos E; K; R

El aminoácido R es similar a los aminoácidos K; Q

35 El aminoácido S es similar a los aminoácidos A; N; T

El aminoácido T es similar al aminoácido S

El aminoácido V es similar a los aminoácidos I; L; M

El aminoácido W es similar a los aminoácidos F; Y

El aminoácido Y es similar a los aminoácidos F; H; W.

40 Las sustituciones conservadoras de aminoácidos pueden producirse en toda la longitud de la secuencia de una secuencia polipeptídica de una proteína funcional, tal como una enzima. En una realización, tales mutaciones no pertenecen a los dominios funcionales de una enzima. En otra realización, las mutaciones conservadoras no pertenecen a los centros catalíticos de una enzima.

Por lo tanto, según la presente invención se aplica el siguiente cálculo del porcentaje de similitud:
 % de similitud = [(residuos idénticos + residuos similares)/longitud de la región de alineamiento que muestra la
 secuencia respectiva de la presente invención en toda su longitud] * 100. Por lo tanto, en el presente documento, la
 similitud de secuencia en relación con la comparación de dos secuencias de aminoácidos se calcula dividiendo el
 número de residuos idénticos más el número de residuos similares por la longitud de la región de alineamiento que
 muestra la secuencia respectiva de la presente invención en toda su longitud. Este valor se multiplica por 100 para
 obtener el "% de similitud".

En especial, se prevé que las enzimas variantes que comprenden mutaciones conservadoras que son al menos un m
 por ciento similares a las secuencias parentales respectivas, siendo m un número entero entre 50 y 100,
 preferentemente 50, 55, 60, 65, 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98 o 99 en comparación con la secuencia
 polipeptídica de longitud completa, tengan propiedades enzimáticas fundamentalmente inalteradas. Las enzimas
 variantes descritas en el presente documento con un m por ciento de similitud cuando se comparan con una enzima
 parental tienen actividad enzimática.

El término "hibridación", tal como se define en el presente documento, es un proceso en el que secuencias de
 nucleótidos sustancialmente complementarias se hibridan entre sí. El proceso de hibridación puede producirse
 completamente en solución, es decir, ambos ácidos nucleicos complementarios están en solución. El proceso de
 hibridación también puede producirse con uno de los ácidos nucleicos complementarios inmovilizado sobre una matriz,
 tales como esferas magnéticas, esferas de Sepharose o cualquier otra resina. El proceso de hibridación puede
 producirse además con uno de los ácidos nucleicos complementarios inmovilizado sobre un soporte sólido, tal como
 una membrana de nitrocelulosa o nailon, o inmovilizado, por ejemplo, mediante fotolitografía sobre un soporte, incluido,
 entre otros, un soporte de vidrio silíceo (estos últimos se conocen como matrices o micromatrices de ácidos nucleicos
 o como chips de ácidos nucleicos). Para que se produzca la hibridación, las moléculas de ácido nucleico se suelen
 desnaturar térmica o químicamente para fundir una cadena doble en dos cadenas sencillas y/o para eliminar las
 horquillas u otras estructuras secundarias de los ácidos nucleicos monocatenarios.

Esta formación o fusión de híbridos depende de diversos parámetros, entre ellos la temperatura. Un aumento de la
 temperatura favorece la fusión, mientras que una disminución de la temperatura favorece la hibridación. Sin embargo,
 este proceso de formación de híbridos no sigue de forma lineal un cambio aplicado en la temperatura: el proceso de
 hibridación es dinámico, y los pares de nucleótidos ya formados favorecen también el apareamiento de los nucleótidos
 adyacentes. Así, puede decirse que la hibridación es un proceso de sí o no, y existe una temperatura que básicamente
 define la frontera entre la hibridación y la no hibridación. Esta temperatura es la temperatura de fusión (Tf). La Tf es la
 temperatura en grados Celsius en la que el 50 % de todas las moléculas de una secuencia de nucleótidos determinada
 se hibridan en una cadena doble y el 50 % están presentes como cadenas sencillas.

La temperatura de fusión (Tf) depende de las propiedades físicas de la secuencia de ácido nucleico analizada y, por
 lo tanto, puede indicar la relación entre dos secuencias distintas. Sin embargo, la temperatura de fusión (Tf) también
 se ve influida por otros parámetros, que no están directamente relacionados con las secuencias, y deben tenerse en
 cuenta las condiciones aplicadas en el experimento de hibridación. Por ejemplo, un aumento de sales (por ejemplo,
 cationes monovalentes) da lugar a una Tf más alta.

La Tf para una condición de hibridación concreta puede determinarse realizando un experimento de hibridación física,
 pero la Tf también puede calcularse de modo informático para un par concreto de secuencias de ADN. En esta
 realización, se usa la ecuación de Meinkoth y Wahl (Anal. Biochem., 138:267-284, 1984) para tramos de 50 bases o
 más: $T_f = 81,5 \text{ °C} + 16,6 (\log M) + 0,41 (\% \text{ de GC}) - 0,61 (\% \text{ de forma}) - 500/L$.

M es la molaridad de los cationes monovalentes, % de GC es el porcentaje de nucleótidos de guanosina y citosina en
 el tramo de ADN, % de forma es el porcentaje de formamida en la solución de hibridación, y L es la longitud del híbrido
 en pares de bases. La ecuación es para intervalos de sal de 0,01 a 0,4 M y % de GC en intervalos del 30 % al 75 %.

Aunque una temperatura por encima de la Tf corresponde a una sonda perfectamente apareada, la Tf se reduce en
 aproximadamente 1 °C por cada 1 % de desapareamiento (Bonner *et al.*, J. Mol. Biol., 81: 123-135, 1973): $T_f = [81,5 \text{ °C}$
 $+ 16,6(\log M) + 0,41 (\% \text{ de GC}) - 0,61 (\% \text{ de formamida}) - 500/L] - \% \text{ de no identidad}$.

Esta ecuación es útil para sondas que tienen 35 o más nucleótidos y está ampliamente referenciada en la bibliografía
 de procedimientos científicos (por ejemplo, en: "Recombinant DNA Principles and Methodologies", James Greene,
 capítulo "Biochemistry of Nucleic acids", Paul S. Miller, página 55; 1998, CRC Press), en muchas solicitudes de
 patentes (por ejemplo, en: documento US 7026149), y también en hojas de datos de empresas comerciales (por
 ejemplo, "Equations for Calculating Tm" de www.genomics.agilent.com).

Otras fórmulas para el cálculo de la Tf, que son menos preferidas en esta realización, podrían utilizarse únicamente
 para los casos indicados:

Para híbridos de ADN-ARN (Case, J. y Davidson, N. (1977), Nucleic Acids Res., 4:1539):

$$T_f = 79,8 \text{ °C} + 18,5 (\log M) + 0,58 (\% \text{ de GC}) + 11,8 (\% \text{ de GC} * \% \text{ de GC}) - 0,5 (\% \text{ de form}) - 820/L.$$

ES 2 943 019 T3

Para híbridos de ARN-ARN (Bodkin D.K. y Knudson D.L. (1985), J. Virol. Methods, 10: 45):

$$T_f = 79,8 \text{ }^\circ\text{C} + 18,5 (\log M) + 0,58 (\% \text{ de GC}) + 11,8 (\% \text{ de GC} * \% \text{ de GC}) - 0,35 (\% \text{ de form}) - 820/L.$$

Para sondas oligonucleotídicas de menos de 20 bases (Wallace, R.B., *et al.* (1979) Nucleic Acid Res., 6: 3535): $T_f = 2 \times n(A + T) + 4 \times n(G + C)$, siendo n el número de bases respectivas de la sonda que forman un híbrido.

- 5 Para sondas oligonucleotídicas de 20 a 35 nucleótidos, podría aplicarse un cálculo de Wallace modificado: $T_f = 22 + 1,46 n(A + T) + 2,92 n(G + C)$, siendo n el número de bases respectivas de la sonda que forman un híbrido.

Para otros oligonucleótidos, debe utilizarse el modelo del vecino más cercano para el cálculo de la temperatura de fusión junto con los datos termodinámicos apropiados:

$$T_f = (\Sigma(\Delta H_d) + \Delta H_i) / (\Sigma(\Delta S_d)) + \Delta S + \Delta S_{\text{auto}} + R \times \ln(cT/b) + 16,6 \log[Na^+] - 273,15$$

- 10 (Breslauer, K.J., Frank, R., Blöcker, H., Marky, L.A. 1986, Predicting DNA duplex stability from the base sequence, Proc. Natl Acad. Sci. USA, 833746-833750; Alejandro Panjkovich, Francisco Melo, 2005, Comparison of different melting temperature calculation methods for short DNA sequences, Bioinformatics, 21 (6): 711-722), en la que:

T_f es la temperatura de fusión en grados Celsius;

- 15 $\Sigma(\Delta H_d)$ y $\Sigma(\Delta S_d)$ son las sumas de entalpía y entropía (respectivamente), calculadas sobre todos los dobletes internos del vecino más próximo;

ΔS_{auto} es la penalización entrópica para secuencias autocomplementarias;

ΔH_i y ΔS_i son las sumas de entalpías y entropías de iniciación, respectivamente;

R es la constante de gases (fijada en 1,987 cal/K mol);

cT es la concentración total de cadenas en unidades molares;

- 20 la constante b adopta el valor de 4 para las secuencias no autocomplementarias o igual a 1 para los dúplex de cadenas autocomplementarias o para los dúplex cuando una de las cadenas está en un exceso significativo.

Los cálculos termodinámicos suponen que la hibridación se produce en una solución tamponada a un pH cercano a 7,0 y que se produce una transición de dos estados.

- 25 Los valores termodinámicos para el cálculo pueden obtenerse de la tabla 1 en Alejandro Panjkovich, Francisco Melo, 2005, Comparison of different melting temperature calculation methods for short DNA sequences, Bioinformatics, 21 (6): 711-722, o de los trabajos de investigación originales de Breslauer, K.J., Frank, R., Blöcker, H., Marky, L.A. 1986, Predicting DNA duplex stability from the base sequence, Proc. Natl Acad. Sci. USA, 833746-833750; SantaLucia, J., Jr, Allawi, H.T., Seneviratne, P.A. 1996, Improved nearest-neighbor parameters for predicting DNA duplex stability, Biochemistry, 353555-353562; Sugimoto, N., Nakano, S., Yoneyama, M., Honda, K., 1996, Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes, Nucleic Acids Res., 244501-244505.

- 35 Para el cálculo informático de la T_f según esta realización, primero se genera un conjunto de alineamientos de secuencias bioinformáticas entre las dos secuencias. Dichos alineamientos pueden generarse mediante diversas herramientas conocidas por un experto en la materia, tales como los programas "Blast" (NCBI), "Water" (EMBOSS) o "Matcher" (EMBOSS), que producen alineamientos locales, o "Needle" (EMBOSS), que produce alineamientos globales. Estas herramientas deben aplicarse con su parametrización por defecto, pero también con algunas variaciones de los parámetros. Por ejemplo, el programa "MATCHER" puede aplicarse con diversos parámetros para apertura/extensión de hueco (tales como 14/4; 14/2; 14/5; 14/8; 14/10; 20/2; 20/5; 20/8; 20/10; 30/2; 30/5; 30/8; 30/10; 40/2; 40/5; 40/8; 40/10; 10/2; 10/5; 10/8; 10/10; 8/2; 8/5; 8/8; 8/10; 6/2; 6/5; 6/8; 6/10) y el programa "WATER" puede aplicarse con diversos parámetros de apertura/extensión de hueco (tales como 10/0,5; 10/1; 10/2; 10/3; 10/4; 10/6; 15/1; 15/2; 15/3; 15/4; 15/6; 20/1; 20/2; 20/3; 20/4; 20/6; 30/1; 30/2; 30/3; 30/4; 30/6; 45/1; 45/2; 45/3; 45/4; 45/6; 60/1; 60/2; 60/3; 60/4; 60/6), y también estos programas se aplicarán utilizando ambas secuencias de nucleótidos tal cual, pero también con una de las secuencias en su forma de complemento inverso. Por ejemplo, BlastN (NCBI) puede aplicarse con un valor de corte e incrementado (por ejemplo, e+1 o incluso e+10) para identificar también alineamientos muy cortos, en especial en bases de datos de pequeño tamaño.

- 45 Es importante que se tengan en cuenta los alineamientos locales, ya que la hibridación puede no producirse necesariamente en toda la longitud de las dos secuencias, sino que puede ser mejor en regiones diferenciadas, que son las que determinan la temperatura de fusión real. Por lo tanto, a partir de todos los alineamientos creados, hay que determinar la longitud del alineamiento, el contenido en % de GC del alineamiento (de forma más precisa, el contenido en % de GC de las bases que se aparean dentro del alineamiento) y la identidad del alineamiento. A continuación, hay que calcular la temperatura de fusión (T_f) prevista para cada alineamiento. La T_f calculada más alta se utiliza para predecir la temperatura de fusión real.

La expresión "hibridación en toda la secuencia de la invención", tal como se define en el presente documento, significa que, para secuencias de más de 300 bases, cuando la secuencia de la invención se fragmenta en trozos de aproximadamente 300 a 500 bases de longitud, cada fragmento debe hibridarse. Por ejemplo, un ADN puede fragmentarse en trozos utilizando una enzima de restricción o una combinación de las mismas. A continuación, se realiza un cálculo bioinformático de la T_f mediante el mismo procedimiento descrito anteriormente, pero para cada fragmento. La hibridación física de los fragmentos individuales puede analizarse mediante el análisis Southern convencional o mediante procedimientos comparables conocidos por los expertos en la materia.

El término "rigurosidad", tal como se define en el presente documento, describe la facilidad con la que puede tener lugar la formación de híbridos entre dos secuencias de nucleótidos. Las condiciones de una "mayor rigurosidad" requieren que más bases de una secuencia se apareen con la otra secuencia (la temperatura de fusión T_f se reduce en condiciones de "mayor rigurosidad"), y las condiciones de "menor rigurosidad" permiten que algunas bases más no se apareen. Por tanto, el grado de relación entre dos secuencias pueda calcularse por las condiciones reales de rigurosidad en las que aún son capaces de formar híbridos. Se puede conseguir un aumento de la rigurosidad manteniendo constante la temperatura de hibridación experimental y disminuyendo las concentraciones de sales, o manteniendo constantes las sales y aumentando la temperatura de hibridación experimental, o una combinación de estos parámetros. También un aumento de formamida aumentará la rigurosidad. Los expertos conocen otros parámetros que pueden modificarse durante la hibridación y que mantendrán o modificarán las condiciones de rigurosidad (Sambrook *et al.*, (2001), Molecular Cloning: a laboratory manual, 3^a edición, Cold Spring Harbor Laboratory Press, CSH, Nueva York; o Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. 1989, y actualizaciones anuales).

Un experimento típico de hibridación se realiza mediante una etapa inicial de hibridación, que va seguida de una a varias etapas de lavado. Las soluciones utilizadas para estas etapas pueden contener otros componentes que impidan la degradación de las secuencias analizadas y/o eviten la unión de fondo inespecífica de la sonda, tales como EDTA, SDS, ADN espermático fragmentado o reactivos similares, conocidos por los expertos en la materia (Sambrook *et al.* (2001), Molecular Cloning: a laboratory manual, 3^a edición, Cold Spring Harbor Laboratory Press, CSH, Nueva York; o Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. 1989, y actualizaciones anuales).

Una sonda típica para un experimento de hibridación se genera mediante el procedimiento de marcaje de cebado aleatorio, desarrollado inicialmente por Feinberg y Vogelstein (Anal. Biochem, 132 (1), 6-13 (1983); Anal. Biochem., 137 (1), 266-267 (1984) y que se basa en la hibridación de una mezcla de todos los hexanucleótidos posibles con el ADN que se desea marcar. El producto de la sonda marcada será en realidad una colección de fragmentos de longitud variable, que suele oscilar entre los 100 y los 1 000 nucleótidos de longitud, con la mayor concentración de fragmentos generalmente en torno a los 200 a 400 pb. El intervalo de tamaño real de los fragmentos de la sonda, que finalmente se utilizarán como sondas para el experimento de hibridación, también puede verse influido por el parámetro del procedimiento de marcaje utilizado, la purificación posterior de la sonda generada (por ejemplo, gel de agarosa) y el tamaño del ADN de plantilla utilizado para el marcaje (las plantillas grandes, por ejemplo, pueden ser digeridas por restricción utilizando un cortador de 4 pb, por ejemplo, Haelll, antes del marcaje).

Para la presente invención, la secuencia descrita en el presente documento se analiza mediante un experimento de hibridación, en el que la sonda se genera a partir de la otra secuencia, y esta sonda se genera mediante un procedimiento convencional de marcado con cebado aleatorio. Para la presente invención, la sonda consiste en un conjunto de oligonucleótidos marcados que tienen tamaños de aproximadamente 200 a 400 nucleótidos. Una hibridación entre la secuencia de la presente invención y la otra secuencia significa que la hibridación de la sonda se produce en toda la secuencia de la presente invención, tal como se ha definido anteriormente. El experimento de hibridación se realiza alcanzando la mayor rigurosidad mediante la rigurosidad de la etapa de lavado final. La etapa final de lavado tiene condiciones de rigurosidad comparables a las condiciones de rigurosidad de al menos la condición de lavado 1: 1,06 x SSC, SDS al 0,1 %, formamida al 0 % a 50 °C, en otra realización de al menos la condición de lavado 2: 1,06 x SSC, SDS al 0,1 %, formamida al 0 % a 55 °C, en otra realización de al menos la condición de lavado 3: 1,06 x SSC, SDS al 0,1 %, formamida al 0 % a 60 °C, en otra realización de al menos la condición de lavado 4: 1,06 x SSC, SDS al 0,1 %, formamida al 0 % a 65 °C, en otra realización de al menos la condición de lavado 5: 0,52 x SSC, SDS al 0,1 %, formamida al 0 % a 65 °C, en otra realización de al menos la condición de lavado 6: 0,25 x SSC, SDS al 0,1 %, formamida al 0 % a 65 °C, en otra realización de al menos la condición de lavado 7: 0,12 x SSC, SDS al 0,1 %, formamida al 0 % a 65 °C, en otra realización de al menos la condición de lavado 8: 0,07 x SSC, SDS al 0,1 %, formamida al 0 % a 65 °C.

Un "lavado poco riguroso" tiene unas condiciones de rigurosidad comparables a las condiciones de rigurosidad de al menos la condición de lavado 1, pero no más rigurosas que la condición de lavado 3, siendo las condiciones de lavado las descritas anteriormente.

Un "lavado muy riguroso" tiene condiciones de rigurosidad comparables a las condiciones de rigurosidad de al menos la condición de lavado 4, en otra realización de al menos la condición de lavado 5, en otra realización de al menos la condición de lavado 6, en otra realización de al menos la condición de lavado 7, en otra realización de al menos la condición de lavado 8, siendo las condiciones de lavado las descritas anteriormente.

El término polipéptido "heterólogo" (o exógeno o extraño o recombinante o no nativo) se define en el presente documento como un polipéptido que no es nativo de la célula hospedante, un polipéptido nativo de la célula hospedante en el que se han realizado modificaciones estructurales, por ejemplo, deleciones, sustituciones y/o inserciones, mediante técnicas de ADN recombinante para alterar el polipéptido nativo, o un polipéptido nativo de la célula hospedante cuya expresión está cuantitativamente alterada o cuya expresión está dirigida desde una ubicación genómica diferente de la célula hospedante nativa como resultado de la manipulación del ADN de la célula hospedante mediante técnicas de ADN recombinante, por ejemplo, un promotor más fuerte. Del mismo modo, el término polinucleótido "heterólogo" (o exógeno o extraño o recombinante o no nativo) se refiere a un polinucleótido que no es nativo de la célula hospedante, un polinucleótido nativo de la célula hospedante en el que se han realizado modificaciones estructurales, por ejemplo, deleciones, sustituciones y/o inserciones, mediante técnicas de ADN recombinante para alterar el polinucleótido nativo, o un polinucleótido nativo de la célula hospedante cuya expresión está cuantitativamente alterada como resultado de la manipulación de los elementos reguladores del polinucleótido mediante técnicas de ADN recombinante, por ejemplo, un promotor más fuerte, o un polinucleótido nativo de la célula hospedante, pero que no está integrado dentro de su entorno genético natural como resultado de la manipulación genética mediante técnicas de ADN recombinante.

Con respecto a dos o más secuencias de polinucleótidos o a dos o más secuencias de aminoácidos, el término "heterólogo" se utiliza para indicar que dichas dos o más secuencias de polinucleótidos o dos o más secuencias de aminoácidos no se aparecen en la naturaleza en la combinación específica entre sí.

A los efectos de la invención, "recombinante" (o transgénico) con respecto a una célula o un organismo significa que la célula o el organismo contiene un polinucleótido heterólogo que ha sido introducido por el ser humano mediante tecnología génica y, con respecto a un polinucleótido, incluye todas aquellas construcciones realizadas por el ser humano mediante tecnología génica/técnicas de ADN recombinante, en las que:

(a) la secuencia del polinucleótido o de una parte del mismo, o bien

(b) una o más secuencias de control genético que están unidas operativamente con el polinucleótido, incluidas, entre otras, un promotor, o bien

(c) tanto a) como b)

no se encuentran en su entorno genético de tipo salvaje o han sido modificadas.

Los términos célula u organismo "nativo" (o de tipo salvaje o endógeno) y polinucleótido o polipéptido "nativo" (o de tipo salvaje o endógeno) se refieren a la célula u organismo tal como se encuentra en la naturaleza y al polinucleótido o polipéptido en cuestión tal como se encuentra en una célula en su forma natural y en su entorno genético, respectivamente (es decir, sin que haya intervención humana).

Un "ADN con patrón de metilación extraño a una célula" se refiere a un ADN que comprende un patrón de metilación que no aparece de forma natural en la célula y que, por lo tanto, puede ser reconocido y escindido por una o más enzimas de restricción de la célula.

El término "polinucleótido" y las expresiones "secuencia de ácido nucleico", "secuencia de nucleótidos", "ácido nucleico" y "molécula de ácido nucleico" se utilizan indistintamente en el presente documento y se refieren a nucleótidos, ya sean ribonucleótidos o desoxirribonucleótidos o una combinación de ambos, en una forma polimérica no ramificada de cualquier longitud. Los "polinucleótidos" están compuestos por monómeros, que son "nucleótidos" formados por tres componentes: un azúcar pentosa, un grupo fosfato y una base nitrogenada.

La expresión "construcción de ácido nucleico", tal como se utiliza en el presente documento, se refiere a una molécula de ácido nucleico, monocatenaria o bicatenaria, que se aísla a partir de un gen natural o se modifica para que contenga segmentos de ácidos nucleicos de una manera que, de otro modo, no existiría en la naturaleza o es sintética. La expresión "construcción de ácido nucleico" es sinónima de la expresión "casete de expresión" cuando la construcción de ácido nucleico contiene las secuencias de control necesarias para la expresión de un polinucleótido.

La expresión "secuencia de control" se define en el presente documento para incluir todas las secuencias que afectan a la expresión de un polinucleótido, incluida, entre otras, la expresión de un polinucleótido que codifica un polipéptido. Cada secuencia de control puede ser nativa o extraña al polinucleótido o nativa o extraña entre sí. Dichas secuencias de control incluyen, entre otras, la secuencia promotora, la 5'-UTR (también denominada secuencia líder), el sitio de unión ribosómico (RBS, secuencia Shine-Dalgarno), la 3'-UTR y los sitios de inicio y terminación de la transcripción.

Las expresiones "unión funcional" o "unido operativamente" con respecto a los elementos reguladores deben entenderse en el sentido de la disposición secuencial de un elemento regulador (incluido, entre otros, un promotor) con una secuencia de ácido nucleico que debe expresarse y, si procede, otros elementos reguladores (incluidos, entre otros, un terminator) de tal manera que cada uno de los elementos reguladores pueda cumplir su función prevista de permitir, modificar, facilitar o influir de otro modo en la expresión de dicha secuencia de ácido nucleico. Por ejemplo, se sitúa una secuencia de control en una posición adecuada con respecto a la secuencia codificante de la secuencia polinucleotídica, de forma que la secuencia de control dirija la expresión de la secuencia codificante de un polipéptido.

- Las disposiciones preferidas son aquellas en las que la secuencia de ácido nucleico que debe expresarse recombinantemente se sitúa detrás de la secuencia que actúa como promotor, de modo que las dos secuencias estén unidas covalentemente entre sí. En una realización, la secuencia de ácido nucleico a transcribir se sitúa detrás del promotor, de tal forma que el inicio de la transcripción es idéntico al inicio deseado del ARN. El enlace funcional, y una construcción de expresión, pueden generarse mediante las técnicas habituales de recombinación y clonación descritas (por ejemplo, Sambrook, J. y Russell, D.W., *Molecular Cloning: A laboratory manual*, 3ª ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001; Silhavy *et al.* (1984), *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor (NY); Ausubel *et al.* (1987), *Current Protocols in Molecular Biology*, Greene Publishing Assoc. and Wiley Interscience; Gelvin *et al.* (eds.) (1990), *Plant Molecular Biology Manual*; Kluwer Academic Publisher, Dordrecht, Países Bajos; *Plant Molecular Biology Labfax* (1993) de R.D.D. Croy, publicado por BIOS Scientific Publications Ltd (Reino Unido) y Blackwell Scientific Publications (Reino Unido)). Sin embargo, también pueden situarse entre las dos secuencias otras secuencias, incluidas, entre otras, una secuencia que actúe como conector con sitios de corte específicos para enzimas de restricción, o como un péptido señal. La inserción de secuencias también puede dar lugar a la expresión de proteínas de fusión.
- 5
- 10
- 15 Un "promotor" o "secuencia promotora" es una secuencia de nucleótidos situada cadena arriba de un gen en la misma cadena que el gen que permite la transcripción de ese gen. Al promotor le sigue el sitio de inicio de la transcripción del gen. El promotor es reconocido por la ARN polimerasa (junto con los factores de transcripción necesarios), que inicia la transcripción. Un fragmento funcional o una variante funcional de un promotor es una secuencia de nucleótidos reconocible por la ARN polimerasa y capaz de iniciar la transcripción.
- 20 La expresión "secuencia promotora que comprende una secuencia consenso y en la que la secuencia consenso va seguida inmediatamente de un sitio de inicio de la transcripción" o la expresión "secuencia promotora que comprende una secuencia consenso seguida inmediatamente de un sitio de inicio de la transcripción" se entienden en el presente documento como el sitio de inicio de la transcripción directamente adyacente a la secuencia consenso, es decir, sin ningún nucleótido adicional de conexión entre la secuencia consenso y el sitio de inicio de la transcripción.
- 25 La expresión "sitio de inicio de la transcripción" o "sitio de inicio transcripcional" se entenderá como la ubicación donde comienza la transcripción en el extremo 5' de una secuencia genética. En procariontes, el primer nucleótido, denominado +1, es en general un nucleótido de adenosina (A) o guanosina (G). En este contexto, los términos "sitios" y "señal" pueden utilizarse indistintamente.
- 30 Cuando se utiliza en el presente documento, la expresión "secuencia codificante" significa una secuencia de nucleótidos que especifica directamente la secuencia de aminoácidos de su producto proteico. Los límites de la secuencia codificante suelen estar determinados por un marco de lectura abierto, que suele comenzar con el codón de inicio ATG o codones de inicio alternativos, tales como GTG, CTG o TTG, y termina con un codón de terminación, tal como TAA, TAG o TGA. La secuencia codificante puede ser una secuencia de nucleótidos de ADN, ADNc, sintética o recombinante. El codón de inicio también puede denominarse "señal de inicio de la traducción" o "sitio de inicio de la traducción". El codón de terminación también puede denominarse "señal de terminación de la traducción" o "sitio de terminación de la traducción".
- 35 El término "expresión" o la expresión "expresión génica" significa la transcripción de un gen específico o genes específicos o una construcción de ácido nucleico específica. El término "expresión" o la expresión "expresión génica" se refiere en concreto a la transcripción de un gen o genes o una construcción genética en ARN estructural (por ejemplo, ARNr, ARNt) o ARNm con o sin traducción posterior de este último en una proteína. El proceso incluye la transcripción del ADN y el procesamiento del producto de ARNm resultante.
- 40 La expresión "vector de expresión" se define en el presente documento como una molécula de ADN lineal o circular que comprende un polinucleótido que está unido operativamente a una o más secuencias de control que proporcionan la expresión del polinucleótido.
- 45 La expresión "expresión moderada" de un gen se define en el presente documento como un nivel de expresión de un gen determinado que no perjudica el crecimiento o la viabilidad celular y que permite el cultivo continuo de la célula hospedante.
- 50 La expresión "célula hospedante", tal como se utiliza en el presente documento, incluye cualquier tipo de célula que sea susceptible de transformación, transfección, transducción, conjugación y similares con una construcción de ácido nucleico o vector de expresión.
- El término "introducción" y sus variaciones se definen en el presente documento como la transferencia de un ADN a una célula hospedante. La introducción de un ADN en una célula hospedante puede llevarse a cabo por cualquier procedimiento conocido en la técnica, incluidos, entre otros, la transformación, la transfección, la transducción, la conjugación y similares.
- 55 La expresión "célula donante" se define en el presente documento como una célula que es la fuente del ADN introducido por cualquier medio en otra célula.
- La expresión "célula receptora" se define en el presente documento como una célula en la que se introduce ADN.

La expresión "fermentación a escala industrial" (también llamado fermentación a gran escala) se refiere a procesos de fermentación con volúmenes del fermentador superiores o iguales a 20 litros.

5 La expresión "ADN metiltransferasa que metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC" se define en el presente documento como una ADN (citosina-5)-metiltransferasa (EC 2.1.1.37) que cataliza la transferencia de un grupo metilo desde S-adenosil-L-metionina a citosina dentro de la secuencia GCNGC, dando como resultado S-adenosil-L-homocisteína y ADN que contiene 5-metilcitosina. A efectos de la presente invención, la actividad de la ADN metiltransferasa se determina según el procedimiento descrito por Pfeifer *et al.*, 1983, *Biochim. Biophys. Acta*, 740: 323-330. Una unidad de actividad de la ADN metiltransferasa es la cantidad necesaria para proteger 1 µg de ADN lambda en 1 hora en un volumen total de reacción de 20 µl frente a la ruptura por la endonucleasa de restricción correspondiente.

10 La expresión "sistema de modificación por restricción" se define en el presente documento como una endonucleasa de restricción, una ADN metiltransferasa correspondiente que protege el ADN de la ruptura por la endonucleasa de restricción, y los genes que codifican al menos estas dos enzimas.

15 El término "operón" se entiende en el presente documento como una unidad de ADN genómico que contiene un único promotor y uno o más genes, todos los cuales se transcriben a partir de ese único promotor. Los genes del operón pueden solaparse o tener regiones no traducidas ("untranslated regions", UTR) entre sí. Estas UTR pueden tener opcionalmente otros elementos de control, lo cual afecta a la eficiencia traslacional.

20 La expresión "operón secA" se entiende en el presente documento como un operón que comprende un gen secA y una secuencia promotora endógenamente unida al mismo, es decir, el promotor de tipo silvestre de este operón secA ("promotor secA"). Sin limitarse a ello, un ejemplo de operón que contiene secA es una construcción que consiste en un promotor, una 5'UTR, un gen secM (control de secreción, SecM regulador de SecA), una UTR, y un gen secA, en el que el promotor es la secuencia promotora nativa del operón que contiene el gen secA.

25 La proteína SecA es una proteína multifuncional implicada en el proceso de secreción de proteínas (translocación de proteínas) a través de la membrana celular interna bacteriana (Green, Erin R. y Joan Meccas, "Bacterial Secretion Systems - An Overview", *Microbiology spectrum*, 4.1 (2016)). El gen secA, que codifica la proteína SecA, suele indicarse como "subunidad de translocasa SecA", "subunidad de preproteína translocasa SecA", "subunidad de proteína translocasa SecA", "subunidad de unión a translocasa (ATPasa)" o "preproteína de translocasa; proteína de secreción". Algunos organismos tienen dos proteínas SecA homólogas, una de las cuales es fundamental y la otra no (Braunstein M., Brown A.M., Kurtz S., Jacobs W.R. Jr., "Two nonredundant SecA homologues function in mycobacteria", *Journal of Bacteriology*, 1 de diciembre de 2001, 183(24):6979-6990; Feltcher M.E., Braunstein M., "Emerging themes in SecA2-mediated protein export", *Nature Reviews Microbiology*, 24 de septiembre de 2012, 10(11):779-789). En los organismos que tienen dos proteínas translocasa similares a SecA, la indicación del gen de la subunidad de proteína translocasa SecA pertinente para esta solicitud de patente se suele marcar con un índice adicional 1, por ejemplo, "subunidad de proteína translocasa SecA1", e indica la translocasa fundamental.

35 La expresión "gen indicador" se entiende en el presente documento como un gen cuyo producto puede ensayarse fácilmente tras la introducción en una célula hospedante y que puede utilizarse como marcador para seleccionar células transformadas con éxito, para estudiar la regulación de la expresión génica o para actuar como control para normalizar las eficiencias de transformación. Los genes indicadores habituales codifican proteínas fluorescentes, por ejemplo, entre otras, la proteína fluorescente verde GFP, la proteína fluorescente roja RFP, y genes que codifican enzimas que catalizan una reacción química que da lugar a un producto detectable, por ejemplo, entre otros, la galactosidasa, y genes de luciferasa. La proteína codificada por el gen indicador también se denomina en el presente documento "proteína indicadora".

Descripción detallada

Construcciones de ácido nucleico

45 En una realización, la presente invención se dirige a una construcción de ácido nucleico que comprende un polinucleótido unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en una célula hospedante, en la que al menos una secuencia de control comprende una secuencia promotora de un operón secA que comprende un gen secA (dicha secuencia promotora en el presente documento también se denomina "promotor secA") y en la que dicha secuencia promotora es heteróloga al polinucleótido, en la que el polinucleótido no comprende un gen indicador y la secuencia promotora tiene al menos un 70 % de identidad de secuencia con SEQ ID NO:9.

En una realización, la secuencia promotora de un operón que comprende un gen secA o un fragmento funcional o variante funcional del mismo confiere un nivel de expresión moderado.

55 En una realización preferida, la secuencia promotora tiene al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 92 %, al menos un 95 %, al menos un 98 %, al menos un 99 % o al menos un 99 % de identidad de secuencia con SEQ ID NO:9. Preferentemente, la secuencia promotora tiene al menos un 90 % o al menos un 95 % de identidad de secuencia con SEQ ID NO:9.

En otra realización, la secuencia promotora de un operón que comprende un gen *secA* procede de una especie de *Bacillus*. Preferentemente, la especie de *Bacillus* es *Bacillus licheniformis*.

5 En una realización preferida, la secuencia promotora tiene al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 92 %, al menos un 95 %, al menos un 98 %, al menos un 99 % o incluso un 100 % con SEQ ID NO:9 y en la que la secuencia promotora procede de una especie de *Bacillus*, en una realización específica procede de *Bacillus licheniformis*.

10 En una realización, la construcción de ácido nucleico y/o el vector de expresión descritos en el presente documento comprenden una o más secuencias de control adicionales. Dichas secuencias de control incluyen, entre otras, la secuencia promotora, la 5'-UTR (también denominada secuencia líder), el sitio de unión ribosómico (RBS, secuencia Shine-Dalgarno), la 3'-UTR y el terminador de la transcripción y la traducción. En una realización, las secuencias de control pueden estar provistas de conectores con el fin de introducir sitios de restricción específicos que faciliten la conexión de las secuencias de control con el polinucleótido a expresar.

15 La construcción de ácido nucleico comprende una secuencia adecuada de inicio y terminación de la transcripción. En la presente invención puede utilizarse cualquier iniciador o terminador de transcripción que sea funcional en la célula hospedante elegida.

20 En una realización, la construcción de ácido nucleico comprende una secuencia UTR (región no traducida) adecuada. En una realización, la construcción de ácido nucleico descrita en el presente documento comprende una secuencia 5'UTR y/o 3'UTR. En una realización, dichas una o más secuencias de control de la construcción de ácido nucleico comprenden una 5'UTR, también denominada secuencia líder. En otra realización, dichas una o más secuencias de control de la construcción de ácido nucleico comprenden una secuencia 5'UTR que comprende un sitio de unión al ribosoma, también denominado secuencia Shine-Dalgarno. En la presente invención puede utilizarse cualquier secuencia líder que sea funcional en la célula hospedante elegida. La UTR puede ser natural o artificial. En una realización, la 5'UTR tiene al menos un 90 %, al menos un 92 %, al menos un 95 %, al menos un 98 % o incluso un 100% de identidad de secuencia con SEQ ID NO:13 o SEQ ID NO:99 a 116.

25 Las construcciones de ácido nucleico descritas en el presente documento pueden utilizarse para la expresión de una proteína de interés. Por lo tanto, en una realización, el polinucleótido de la construcción de ácido nucleico unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en una célula hospedante, en el que al menos una secuencia de control comprende una secuencia promotora de un operón que comprende un gen *secA*, es un polinucleótido que codifica una proteína de interés. En una realización, la proteína de interés se selecciona del grupo que consiste en una metiltransferasa, una endonucleasa, una serina recombinasa, una tirosina recombinasa y una proteína que confiere resistencia a antibióticos.

30 La construcción de ácido nucleico y/o el vector de expresión descritos en el presente documento pueden utilizarse para proporcionar un nivel de expresión moderado de un polinucleótido, preferentemente un polinucleótido que codifica una proteína de interés, en una célula hospedante.

35 En una realización, la construcción de ácido nucleico y/o el vector de expresión descritos en el presente documento pueden utilizarse para la expresión de un polinucleótido, preferentemente un polinucleótido que codifica una proteína de interés en una célula hospedante, proporcionando un nivel de expresión de dicho polinucleótido que permite el cultivo continuo de la célula hospedante.

Vectores de expresión

40 En otra realización, la presente invención se dirige a un vector de expresión que comprende la construcción de ácido nucleico descrita en el presente documento. Así, en una realización, la presente invención se dirige a un vector de expresión que comprende una construcción de ácido nucleico que comprende un polinucleótido unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en una célula hospedante, en el que al menos una secuencia de control comprende una secuencia promotora de un operón que comprende un gen *secA*, en el que dicha secuencia promotora es heteróloga al polinucleótido, en el que el polinucleótido no comprende un gen indicador y la secuencia promotora tiene al menos un 70 % de identidad de secuencia con SEQ ID NO:9.

En una realización, el vector de expresión se sitúa fuera del ADN cromosómico de la célula hospedante. En otra realización, el vector de expresión se integra en el ADN cromosómico de la célula hospedante. El vector de expresión puede ser lineal o circular. En una realización, el vector de expresión es un vector viral o un plásmido.

50 Para la replicación autónoma, el vector de expresión puede comprender además un origen de replicación que permita al vector replicarse autónomamente en la célula hospedante en cuestión. Los orígenes de replicación bacterianos incluyen, entre otros, los orígenes de replicación de los plásmidos pBR322, pUC19, pSC101, pACYC177 y pACYC184 que permiten la replicación en *E. coli* (Sambrook, J. y Russell, D.W., Molecular Cloning: A laboratory manual, 3ª ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001; Cohen, S. N., Chang, A. C. Y., Boyer, H. W., y Helling, R. B. (1973), Construction of Biologically Functional Bacterial Plasmids In Vitro, Proceedings of the National Academy of Sciences of the United States of America, 70(11), 3240-3244), y pUB110, pC194, pTB19, pAMβ1 y pTA1060 que permiten la replicación en *Bacillus* (Janniére, L., Bruand, C. y Ehrlich, S.D. (1990), Structurally stable

Bacillus subtilis cloning vectors, Gene, 87, 53-56; Ehrlich, S.D., Bruand, C., Sozhamannan, S., Dabert, P., Gros, M.F., Janniere, L. y Gruss, A. (1991), Plasmid replication and structural stability in *Bacillus subtilis*, Res. Microbiol., 142, 869-873), y pE194 (Dempsey, L.A. y Dubnau, D.A. (1989), Localization of the replication origin of plasmid pE194, J. Bacteriol., 171, 2866-2869). El origen de la replicación puede ser un origen que tenga una mutación para que su función sea sensible a la temperatura en la célula hospedante (véase, por ejemplo, Ehrlich, 1978, Proceedings of the National Academy of Sciences USA, 75:1433-1436).

En una realización, el vector de expresión contiene uno o más marcadores seleccionables que permiten una fácil selección de las células transformadas. Un marcador seleccionable es un gen que codifica un producto que proporciona resistencia a biocidas, resistencia a metales pesados, prototrofia a auxótrofos y similares. Los marcadores seleccionables bacterianos incluyen, entre otros, los genes *dal* de *Bacillus subtilis* o *Bacillus licheniformis*, o marcadores que confieren resistencia a antibióticos, tales como la ampicilina, la kanamicina, la eritromicina, el cloranfenicol o la tetraciclina. Además, la selección puede llevarse a cabo por cotransformación, por ejemplo, como se describe en el documento WO91/09129, en el que el marcador seleccionable se encuentra en un vector separado.

Proteínas de interés

Las construcciones de ácido nucleico descritas en el presente documento pueden utilizarse para la expresión de una proteína de interés. Así, en una realización, la construcción de ácido nucleico descrita en el presente documento comprende un polinucleótido, en el que el polinucleótido codifica una proteína de interés. Por lo tanto, en una realización, el polinucleótido de la construcción de ácido nucleico unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en una célula hospedante, en el que al menos una secuencia de control comprende una secuencia promotora de un operón que comprende un gen *secA* o un fragmento funcional o variante funcional del mismo, es un polinucleótido que codifica una proteína de interés.

Como la presente invención está dirigida al uso de una secuencia promotora de un operón que comprende un gen *secA* para la expresión heteróloga de un polinucleótido, la secuencia promotora de un operón que comprende un gen *secA* no está asociada en la naturaleza con el polinucleótido. Por lo tanto, en una realización preferida, el polinucleótido que va a ser expresado por la secuencia promotora de un operón que comprende un gen *secA* no comprende un gen *secA*.

En una realización, el polinucleótido no comprende un gen indicador. Así, en una realización, la proteína de interés no es una proteína indicadora. En una realización preferida, la proteína de interés no es una proteína indicadora seleccionada del grupo que consiste en β -galactosidasa, cloranfenicol acetiltransferasa, fosfatasa alcalina, β -glucuronidasa, luciferasa, aequorina y proteína fluorescente, preferentemente proteína fluorescente verde, proteína fluorescente roja y proteína fluorescente amarilla.

En una realización, el polinucleótido no comprende un gen indicador seleccionado del grupo que consiste en *lacZ*, *spoVG-lacZ*, *cat*, *phoA*, *gusA*, gen de la luciferasa, gen de la aequorina, *gfp*, *yfp* y *rfp*.

En una realización concreta, la presente invención se dirige además a procedimientos de producción de un polipéptido nativo o extraño que comprenden: (a) cultivar una célula hospedante recombinante que comprende la construcción de ácido nucleico descrita en el presente documento en condiciones propicias para la producción de la proteína de interés; y (b) opcionalmente, recuperar la proteína de interés.

En una realización, la proteína de interés es tolerada por la célula hospedante solo en cantidades limitadas. Así, en una realización, la expresión de la proteína de interés mediante el uso del promotor *secA* descrito en el presente documento permite el cultivo continuo de la célula hospedante sin efectos negativos sobre el crecimiento. Así, en una realización, el nivel de expresión de la proteína de interés conferido por el promotor *secA* descrito en el presente documento no es tóxico para la célula hospedante. En una realización, el promotor *secA* descrito en el presente documento proporciona a la célula la proteína de interés en un nivel de expresión moderado. En una realización, la proteína de interés es tóxica para la célula hospedante cuando se expresa en grandes cantidades. En una realización, la proteína de interés es una enzima. En otra realización, la proteína de interés es una enzima que es tolerada por la célula hospedante solo en cantidades limitadas.

En una realización, la proteína de interés se selecciona del grupo que consiste en una enzima modificadora de ADN o ARN, y una proteína que confiere resistencia a antibióticos. En una realización, la proteína de interés se selecciona del grupo que consiste en una enzima modificadora de ADN o ARN. En una realización, la proteína de interés es una enzima modificadora de ADN o ARN, seleccionada de la lista de una metiltransferasa, una endonucleasa, una serina recombinasa y una tirosina recombinasa.

En otra realización, la endonucleasa es *i-SceI* o *Cas9*. En otra realización, la serina recombinasa es *cre* del fago P1 o *FLP* de *S. cerevisiae*. En otra realización, la tirosina recombinasa es *lambda* integrasa. En otra realización, la proteína que confiere resistencia a antibióticos está codificada por un gen seleccionado del grupo que consiste en un gen de resistencia a la blasticidina S, un gen de resistencia a la bleomicina, un gen de acetiltransferasa de cloranfenicol, una enzima modificadora de aminoglucósidos, incluida, entre otras, aminoglucósido acetiltransferasas, aminoglucósido acetiltransferasas, aminoglucósido nucleotidiltransferasas, un gen de resistencia a la tetraciclina, incluido, entre otros,

la protección ribosomal a través de Tet(M), TetO-Q), Tet(S), OtrA, transportadores de eflujo de tetraciclina Tet(A-E, G, H, K, L, P [tetA(P)]), y OtrB.

En otra realización, la proteína de interés es una ADN metiltransferasa.

5 La figura 1 muestra un alineamiento de secuencias múltiples basado en la estructura de las secuencias de aminoácidos de varias ADN metiltransferasas que comprenden estas características estructurales. De la figura 1 se deduce que entre el residuo de aminoácido 72 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, hay 33 aminoácidos en las SEQ ID NO:33, 36, 37 y 38, hay 32 aminoácidos en las SEQ ID NO:39-42 y hay 34 aminoácidos en SEQ ID NO:43. Sin embargo, hay 38 aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, en SEQ ID NO:34. Por tanto, en comparación con SEQ ID NO:34, hay una delección en SEQ ID NO:33 y SEQ ID NO:36 a 43.

10 Por lo tanto, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención metila el ADN dentro de la secuencia de reconocimiento GCNGC dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y en la que la ADN metiltransferasa comprende menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33.

15 En una realización, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención metila el ADN dentro de la secuencia de reconocimiento GCNGC dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y en la que la ADN metiltransferasa comprende menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33, en la que la ADN metiltransferasa comprende al menos 22, preferentemente al menos 28 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33. Por lo tanto, preferentemente, la ADN metiltransferasa comprende entre el residuo de aminoácido 72 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, de 22 a 34, preferentemente de 28 a 34, más preferentemente de 22 a 33, aún más preferentemente de 28 a 33, lo más preferentemente de 30 a 34 o de 30 a 33 residuos de aminoácidos. Lo más preferido es que haya 33 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33.

20 En otra realización, la ADN metiltransferasa utilizada en estos procedimientos de la presente comprende menos de 23 residuos de aminoácidos entre el residuo de aminoácido 84 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33.

25 En otra realización, la ADN metiltransferasa comprende menos de 23 residuos de aminoácidos entre el residuo de aminoácido 84 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33, en la que la ADN metiltransferasa comprende al menos 12, preferentemente al menos 18 residuos de aminoácidos entre el residuo de aminoácido 84 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33. Por lo tanto, preferentemente, la ADN metiltransferasa comprende entre el residuo de aminoácido 84 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, de 12 a 22, preferentemente de 18 a 22, más preferentemente de 12 a 21, aún más preferentemente de 18 a 21 residuos de aminoácidos.

30 En otra realización, la ADN metiltransferasa utilizada en los procedimientos de la presente invención metila el ADN dentro de la secuencia de reconocimiento GCNGC dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 5 residuos de aminoácidos entre el residuo de aminoácido 101 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33.

35 En una realización preferida, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención metila el ADN dentro de la secuencia de reconocimiento GCNGC dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 5 residuos de aminoácidos entre el residuo de aminoácido 101 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33, en la que la ADN metiltransferasa comprende de 0 a 4, preferentemente de 2 a 4, más preferentemente de 3 a 4, 0, 1, 2, 3, 4, preferentemente 4 residuos de aminoácidos entre el residuo de aminoácido 101 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33.

40 En otra realización, la ADN metiltransferasa comprende menos de 11 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 83 según la numeración de SEQ ID NO:33.

45 En otra realización, la ADN metiltransferasa comprende menos de 11 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 83 según la numeración de SEQ ID NO:33, en la que la ADN metiltransferasa comprende al menos 7, preferentemente al menos 8 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 83 según la numeración de SEQ ID NO:33. Por lo tanto, preferentemente, la ADN metiltransferasa comprende entre el residuo de aminoácido 72 y el residuo de aminoácido 83, según la numeración de SEQ ID NO:33, de 7 a 10, preferentemente de 8 a 10, más preferentemente de 9 a 10 residuos de aminoácidos.

50 En otra realización, la ADN metiltransferasa utilizada en los procedimientos de la presente invención metila el ADN dentro de la secuencia de reconocimiento GCNGC dando como resultado un ADN que contiene 5-metilcitosina dentro

de la secuencia de reconocimiento GCNGC y comprende menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33 y comprende además menos de 23 residuos de aminoácidos entre el residuo de aminoácido 84 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33.

5 En otra realización, la ADN metiltransferasa utilizada en los procedimientos de la presente invención metila el ADN dentro de la secuencia de reconocimiento GCNGC dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33 y comprende además menos de 5 residuos de aminoácidos entre el residuo de aminoácido 101 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33.

En otra realización, la ADN metiltransferasa comprende además menos de 11 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 83 según la numeración de SEQ ID NO:33.

15 En una realización preferida, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención metila el ADN dentro de la secuencia de reconocimiento GCNGC dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 5 residuos de aminoácidos entre el residuo de aminoácido 101 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33. Preferentemente, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención comprende de 0 a 4, preferentemente de 2 a 4, más preferentemente de 3 a 4 residuos de aminoácidos entre el residuo de aminoácido 101 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33. Preferentemente, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención comprende 0, 1, 2, 3 o 4, preferentemente 4 residuos de aminoácidos entre el residuo de aminoácido 101 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33.

25 Como se muestra en la figura 1, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención comprende, en una realización, las siguientes características estructurales con respecto a las posiciones de aminoácidos indicadas correspondientes a SEQ ID NO:33.

Identificador			α	β	γ	δ	ϵ		
Posición en SEQ ID NO:33	I66	G70	P72	K83	G84	I101	P106	V108	V114
Comentarios	inicio de la lámina	final de la lámina	inicio del bucle de interacción con el ADN	final del bucle de interacción con el ADN	inicio de la hélice	inicio de la región de inserción	final de la hélice/final de la región de inserción	inicio de la lámina	final de la lámina

30 Así, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención que metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, comprende entre el residuo de aminoácido 72 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, una región de bucle de interacción con el ADN y una región de hélice alfa. Preferentemente, la región del bucle de interacción con el ADN se encuentra entre el residuo de aminoácido 72 y el residuo de aminoácido 83 según la numeración de SEQ ID NO:33, y la región de hélice alfa se encuentra entre los residuos 84 y 106 según la numeración de SEQ ID NO:33.

35 En una realización, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención que metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC comprende al menos un 70 %, al menos un 80 %, al menos un 90 %, o al menos un 95 % de identidad de secuencia con SEQ ID NO:34 y comprende una deleción entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33, de modo que haya menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33. En una realización, la deleción se encuentra en la región de la hélice alfa entre los residuos de aminoácidos 84 y 106, según la numeración de SEQ ID NO:33. Preferentemente, hay una deleción entre el residuo de aminoácido 72 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, de 1 a 20 aminoácidos, preferentemente de 3 a 12, más preferentemente de 4 a 8. Preferentemente, la deleción en la región de la hélice entre el residuo de aminoácido 84 y 106, según la numeración de SEQ ID NO:33, es una deleción de 1 a 15, preferentemente de 2 a 10, más preferentemente de 4 a 8 residuos de aminoácidos. En otra realización, la deleción se encuentra en la región del bucle de interacción con el ADN entre los residuos de aminoácidos 72 y 83 según la numeración de SEQ ID NO:33. Preferentemente, la deleción en la región del bucle de interacción con el ADN entre los residuos de aminoácidos 72 y 83, según la numeración de SEQ ID NO:33, es una deleción de 1 a 6, preferentemente de 2 a 4, más preferentemente

de 1 a 2 residuos de aminoácidos. En una realización, la delección se encuentra en la región de la hélice alfa entre los residuos de aminoácidos 84 y 106, según la numeración de SEQ ID NO:33, y en la región del bucle de interacción con el ADN entre los residuos de aminoácidos 72 y 83, según la numeración de SEQ ID NO:33. Preferentemente, la delección se encuentra en la región de la hélice alfa entre los residuos de aminoácidos 84 y 106, según la numeración de SEQ ID NO:33, y en la región del bucle de interacción con el ADN entre los residuos de aminoácidos 72 y 83, según la numeración de SEQ ID NO:33, es una delección de 1 a 20, preferentemente de 2 a 12, más preferentemente de 4 a 10, lo más preferentemente de 4 a 8 residuos de aminoácidos. Preferentemente, la delección entre el residuo de aminoácido 72 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, no suprime la función de la región del bucle de interacción con el ADN entre los residuos de aminoácidos 72 y 83 según la numeración de SEQ ID NO:33, y no elimina completamente la región de hélice alfa entre los residuos de aminoácidos 84 y 106 según la numeración de SEQ ID NO:33.

La variante de ADN metiltransferasa de SEQ ID NO:34 comprende una región de bucle de interacción con el ADN y una región de hélice alfa entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33.

En una realización, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención que metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, se selecciona del grupo que consiste en:

(a) una ADN metiltransferasa que tiene al menos un 90 % de identidad con SEQ ID NO:33, 35, 36, 37, 38, 39, 40, 41, 42 o 43;

(b) una ADN metiltransferasa codificada por un polinucleótido que tiene al menos un 80 % de identidad con SEQ ID NO:18, 19, 20, 21, 24, 25, 26, 27, 28, 29, 30, 31 o 32;

(c) una ADN metiltransferasa codificada por un polinucleótido que se hibrida en condiciones de alta rigurosidad con (i) un polinucleótido que comprende SEQ ID NO:18, 19, 20, 21, 24, 25, 26, 27, 28, 29, 30, 31 o 32, o (ii) el complemento de longitud completa de (i);

(d) una variante de la ADN metiltransferasa de SEQ ID NO:33, 35, 36, 37, 38, 39, 40, 41, 42 o 43 que comprenden una sustitución, delección y/o inserción en una o más posiciones y que tiene actividad ADN metiltransferasa;

(e) una ADN metiltransferasa codificada por un polinucleótido que difiere de SEQ ID NO:18, 19, 20, 21, 24, 25, 26, 27, 28, 29, 30, 31 o 32 debido a la degeneración del código genético; y

(f) un fragmento de la ADN metiltransferasa de (a), (b), (c), (d) o (e) que tiene actividad ADN metiltransferasa.

En una realización, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención que metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33 seleccionado del grupo que consiste en:

(a) una ADN metiltransferasa que tiene al menos un 55 % de identidad con SEQ ID NO:33;

(b) una ADN metiltransferasa codificada por un polinucleótido que tiene al menos un 70 % de identidad con SEQ ID NO:19;

(c) una ADN metiltransferasa codificada por un polinucleótido que se hibrida en condiciones de alta rigurosidad con (i) un polinucleótido que comprende SEQ ID NO:19, o (ii) el complemento de longitud completa de (i);

(d) una variante de la ADN metiltransferasa de SEQ ID NO:33 que comprende una sustitución, en una realización una sustitución conservadora, delección y/o inserción en una o más posiciones y que tiene actividad ADN metiltransferasa;

(e) una ADN metiltransferasa codificada por un polinucleótido que difiere de SEQ ID NO:19 debido a la degeneración del código genético; y

(f) un fragmento de la ADN metiltransferasa de (a), (b), (c), (d) o (e) que tiene actividad ADN metiltransferasa.

En una realización, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención que metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 35 residuos de aminoácidos entre el residuo de

aminoácido 72 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, se selecciona del grupo que consiste en:

(a) una ADN metiltransferasa que tiene al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 98 % o un 100 % de identidad con SEQ ID NO:33; y

5 (b) una ADN metiltransferasa codificada por un polinucleótido que tiene al menos un 90 %, al menos un 95 %, al menos un 98 % o un 100 % de identidad con SEQ ID NO:18.

En una realización preferida, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención que metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, se selecciona del grupo que consiste en:

10

(a) una ADN metiltransferasa que tiene al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 98 % o un 100 % de identidad con SEQ ID NO:33; y

15 (b) una ADN metiltransferasa codificada por un polinucleótido que tiene al menos un 90 %, al menos un 95 %, al menos un 98 % o un 100 % de identidad con SEQ ID NO:18;

en la que la ADN metiltransferasa comprende entre el residuo de aminoácido 72 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, una región de bucle de interacción con el ADN y una región de hélice alfa. Preferentemente, la región del bucle de interacción con el ADN se encuentra entre el residuo de aminoácido 72 y el residuo de aminoácido 83 según la numeración de SEQ ID NO:33, y la región de hélice alfa se encuentra entre los residuos 84 y 106 según la numeración de SEQ ID NO:33.

20

En una realización, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención que metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 23 residuos de aminoácidos entre el residuo de aminoácido 84 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, se selecciona del grupo que consiste en:

25

(a) una ADN metiltransferasa que tiene al menos un 90 %, al menos un 95 %, al menos un 98 % o un 100 % de identidad con SEQ ID NO:33; y

(b) una ADN metiltransferasa codificada por un polinucleótido que tiene al menos un 90 %, al menos un 95 %, al menos un 98 % o un 100 % de identidad con SEQ ID NO:18.

30 En otra realización preferida, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención que metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 23 residuos de aminoácidos entre el residuo de aminoácido 84 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, se selecciona del grupo que consiste en:

35 (a) una ADN metiltransferasa que tiene al menos un 90 %, al menos un 95 %, al menos un 98 % o un 100 % de identidad con SEQ ID NO:33; y

(b) una ADN metiltransferasa codificada por un polinucleótido que tiene al menos un 90 %, al menos un 95 %, al menos un 98 % o un 100 % de identidad con SEQ ID NO:18;

en la que la ADN metiltransferasa comprende entre el residuo de aminoácido 84 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, una región de hélice alfa.

40

En una realización más preferida, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención que metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 5 residuos de aminoácidos entre el residuo de aminoácido 101 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, se selecciona del grupo que consiste en:

45

(a) una ADN metiltransferasa que tiene al menos un 90 %, al menos un 95 %, al menos un 98 % o un 100 % de identidad con SEQ ID NO:33; y

(b) una ADN metiltransferasa codificada por un polinucleótido que tiene al menos un 90 %, al menos un 95 %, al menos un 98 % o un 100 % de identidad con SEQ ID NO:18.

50 En una realización, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención que metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 5 residuos de aminoácidos entre el residuo de

aminoácido 101 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, se selecciona del grupo que consiste en:

(a) una ADN metiltransferasa que tiene al menos un 90 %, al menos un 95 %, al menos un 98 % o un 100 % de identidad con SEQ ID NO:33; y

5 (b) una ADN metiltransferasa codificada por un polinucleótido que tiene al menos un 90 %, al menos un 95 %, al menos un 98 % o un 100 % de identidad con SEQ ID NO:18;

10 en la que la ADN metiltransferasa comprende entre el residuo de aminoácido 72 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, una región de bucle de interacción con el ADN y una región de hélice alfa. Preferentemente, la región del bucle de interacción con el ADN se encuentra entre el residuo de aminoácido 72 y el residuo de aminoácido 83 según la numeración de SEQ ID NO:33, y la región de hélice alfa se encuentra entre los residuos 84 y 106 según la numeración de SEQ ID NO:33.

15 En una realización, la variante de la ADN metiltransferasa de SEQ ID NO:33, 35, 36, 37, 38, 39, 40, 41, 42 o 43 que comprende una sustitución en una o más posiciones y que tiene actividad ADN metiltransferasa comprende 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30 sustituciones. En otra realización, la variante de la ADN metiltransferasa de SEQ ID NO:33, 35, 36, 37, 38, 39, 40, 41, 42 o 43 que comprende una sustitución en una o más posiciones y que tiene actividad ADN metiltransferasa comprende 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30 sustituciones conservadoras.

20 La ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención es, en una realización, una ADN metiltransferasa que tiene al menos un 55 %, al menos un 60 %, al menos un 65 %, al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 %, al menos un 99 % o incluso un 100 % de identidad de secuencia con SEQ ID NO:33, 35, 36, 37, 38, 39, 40, 41, 42 o 43, en la que la ADN metiltransferasa metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33.

30 En una realización, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende además menos de 5 residuos de aminoácidos entre el residuo de aminoácido 101 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, y es una ADN metiltransferasa que tiene al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 %, al menos un 99 % o un 100 % de identidad de secuencia con una secuencia de aminoácidos mostrada en SEQ ID NO:33.

35 En una realización, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende además menos de 5 residuos de aminoácidos entre el residuo de aminoácido 101 y el residuo de aminoácido 106, según la numeración de SEQ ID NO:33, y es una ADN metiltransferasa que tiene al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 %, al menos un 99 % o un 100 % de identidad de secuencia con una secuencia de aminoácidos mostrada en SEQ ID NO:33, en la que la ADN metiltransferasa comprende de 0 a 4, preferentemente de 2 a 4, más preferentemente de 3 a 4, 0, 1, 2, 3 o 4, preferentemente 4 residuos de aminoácidos entre el residuo de aminoácido 101 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33.

45 En una realización, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención comprende además menos de 5 residuos de aminoácido entre el residuo de aminoácido 101 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33 y, en otra realización, comprende menos de 11 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 83 según la numeración de SEQ ID NO:33, y es una ADN metiltransferasa que tiene al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 %, al menos un 99 % o incluso un 100 % de identidad de secuencia con una secuencia de aminoácidos mostrada en SEQ ID NO:33.

50 En una realización, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención comprende además menos de 5 residuos de aminoácidos entre el residuo de aminoácido 101 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33, y comprende además menos de 11 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 83 según la numeración de SEQ ID NO:33, y es una ADN metiltransferasa que tiene al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %

%, al menos un 96 %, al menos un 97 %, al menos un 98 %, al menos un 99 % o un 100 % de identidad de secuencia con una secuencia de aminoácidos mostrada en SEQ ID NO:33.

5 En una realización, la ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención comprende menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33 y comprende además menos de 5 residuos de aminoácidos entre el residuo de aminoácido 101 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33, y comprende menos de 11 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 83 según la numeración de SEQ ID NO:33, y es una ADN metiltransferasa que tiene al menos un 80 %, al menos un 85 %
10 %, al menos un 90 %, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 %, al menos un 99 % o un 100 % de identidad de secuencia con una secuencia de aminoácidos mostrada en SEQ ID NO:33, en la que la ADN metiltransferasa metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC.

15 En otra realización, la ADN metiltransferasa utilizada en la presente invención se define como anteriormente, mientras que la identidad de secuencia indicada se intercambia con la similitud de secuencia tal como se define en el presente documento.

20 En otra realización, la ADN metiltransferasa es una variante de la ADN metiltransferasa de SEQ ID NO:33, 35, 36, 37, 38, 39, 40, 41, 42 o 43 que comprende una sustitución conservadora en una o más posiciones y que tiene actividad ADN metiltransferasa. En otra realización, la ADN metiltransferasa es una variante de la ADN metiltransferasa de SEQ ID NO:33, 35, 36, 37, 38, 39, 40, 41, 42 o 43 que comprende, en comparación con la secuencia parental, únicamente sustituciones conservadoras en una o más posiciones y que tiene actividad ADN metiltransferasa.

En otra realización, la ADN metiltransferasa es un fragmento de una ADN metiltransferasa que tiene actividad de ADN metiltransferasa. En una realización, la ADN metiltransferasa es un fragmento de SEQ ID NO:33, 35, 36, 37, 38, 39, 40, 41, 42 o 43 que tiene actividad ADN metiltransferasa.

25 En otra realización, el fragmento de la ADN metiltransferasa descrito en el presente documento y utilizado en el procedimiento de la presente invención tiene uno o más aminoácidos delecionados del extremo amino y/o carboxilo de SEQ ID NO:33, 35, 36, 37, 38, 39, 40, 41, 42 o 43, en la que el fragmento metila el ADN dentro de la secuencia de reconocimiento GCNGC dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33. En otra realización, el fragmento de la ADN metiltransferasa descrito en el presente documento y utilizado en el procedimiento de la presente invención tiene 1, 2,
30 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 aminoácidos delecionados del extremo amino y/o carboxilo y/o truncamientos de regiones de bucle intermedias. En una realización, un fragmento de SEQ ID NO:33, 35, 36, 37, 38, 39, 40, 41, 42 o 43 contiene al menos 300 residuos de aminoácidos.

35 La ADN metiltransferasa descrita en el presente documento y utilizada en los procedimientos de la presente invención, en una realización, está codificada por un polinucleótido que tiene al menos un 70%, al menos un 75%, al menos un 80%, al menos un 85%, al menos un 90%, al menos un 91%, al menos un 92%, al menos un 93%, al menos un 94%, al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99% o incluso un 100% de identidad de secuencia con SEQ ID NO:18, 19, 20, 21, 24, 25, 26, 27, 28, 29, 30, 31 o 32, en la que la ADN metiltransferasa metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y comprende menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33.

40 En otra realización, la ADN metiltransferasa está codificada por un polinucleótido que se hibrida con (i) un polinucleótido que comprende SEQ ID NO:19, 24, 25 o 27, o (ii) el complemento de longitud completa de (i). En una realización, la ADN metiltransferasa está codificada por un polinucleótido que se hibrida en condiciones muy rigurosas con (i) un polinucleótido que comprende SEQ ID NO:19, 24, 25 o 27, o (ii) el complemento de longitud completa de (i).
45 En otra realización, las condiciones rigurosas son las descritas anteriormente.

En otra realización, la ADN metiltransferasa está codificada por un polinucleótido que difiere de SEQ ID NO:19, 24, 25 o 27 debido a la degeneración del código genético. En una realización, la ADN metiltransferasa está codificada por un polinucleótido que difiere de SEQ ID NO:19 solo debido a la degeneración del código genético.

50 *Células hospedantes*

En una realización, la presente invención se dirige a una célula hospedante que comprende la construcción de ácido nucleico descrita en el presente documento. En una realización, la construcción de ácido nucleico descrita en el presente documento está comprendida en la célula hospedante en un vector de expresión como se describe en el presente documento.

55 La introducción de un ácido nucleico en una célula hospedante puede llevarse a cabo, por ejemplo, entre otras técnicas, por transformación de protoplastos (véase, por ejemplo, Chang y Cohen, 1979, Molecular General Genetics, 168: 111-115), mediante el uso de células competentes (véase, por ejemplo, Young y Spizzen, 1961, Journal of

Bacteriology, 81: 823-829; o Dubnau y Davidoff-Abelson, 1971, Journal of Molecular Biology, 56: 209-221), por electroporación (véase, por ejemplo, Shigekawa y Dower, 1988, Biotechniques, 6: 742-751) o por conjugación (véase, por ejemplo, Koehler y Thorne, 1987, Journal of Bacteriology, 169: 5271-5278). Los protocolos específicos de transformación son conocidos en la técnica para diversos tipos de células hospedantes (véase, por ejemplo, para la transformación de protoplastos de *E. coli*, véase Hanahan, 1983, J. Mol. Biol., 166: 557-580).

Pueden utilizarse diversas células hospedantes para expresar la construcción de ácido nucleico descrita en el presente documento. Las células hospedantes que comprenden las construcciones genéticas descritas en el presente documento pueden obtenerse mediante uno de los procedimientos descritos en el presente documento para introducir los polinucleótidos en dichas células hospedantes.

En una realización, la célula hospedante es un procarionta o un eucarionta. En otra realización, la célula hospedante es una bacteria, una arquea, una célula fúngica, una célula de levadura o una célula eucarionta. En otra realización, la célula hospedante es una célula hospedante no humana.

En una realización, la célula hospedante es una célula bacteriana. La célula hospedante bacteriana puede ser cualquier bacteria grampositiva o gramnegativa. Las bacterias grampositivas incluyen, entre otras, *Bacillus*, *Brevibacterium*, *Corynebacterium*, *Streptococcus*, *Streptomyces*, *Staphylococcus*, *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Clostridium*, *Geobacillus* y *Oceanobacillus*. Las bacterias gramnegativas incluyen, entre otras, *Escherichia*, *Pseudomonas*, *Salmonella*, *Campylobacter*, *Helicobacter*, *Acetobacter*, *Flavobacterium*, *Fusobacterium*, *Gluconobacter*. En una realización específica, la célula hospedante bacteriana es una célula de *Escherichia coli*. En una realización, la célula hospedante es una célula bacteriana. En una realización específica, la célula hospedante es del género *Escherichia* o *Bacillus*.

En los procedimientos de la presente invención, la célula hospedante bacteriana puede ser cualquier célula de *Bacillus*. Las células de *Bacillus* útiles en la práctica de la presente invención incluyen, entre otras, *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus firmus*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus pumilus*, *Bacillus stearothermophilus*, *Bacillus methylotrophicus*, *Bacillus cereus*, *Bacillus paralicheniformis*, *Bacillus subtilis* y *Bacillus thuringiensis*. En una realización, la célula hospedante bacteriana es una célula de *Bacillus amyloliquefaciens*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus stearothermophilus* o *Bacillus subtilis*. En otra realización, la célula hospedante bacteriana es una célula de *Bacillus licheniformis* o una célula de *Bacillus subtilis*, y en una realización específica es una célula de *Bacillus licheniformis*. Preferentemente, la célula hospedante bacteriana es una célula de *Bacillus licheniformis*. Más preferentemente, la célula hospedante es una célula de *Bacillus licheniformis* ATCC 53926.

En los procedimientos de la presente invención, la célula hospedante bacteriana puede ser *Lactobacillus acidophilus*, *Lactobacillus plantarum*, *Lactobacillus gasseri*, *Lactobacillus bulgaricus*, *Lactobacillus reuteri*, *Escherichia coli*, *Staphylococcus aureus*, *Corynebacterium glutamicum*, *Corynebacterium acetoglutamicum*, *Corynebacterium acetoacidophilum*, *Corynebacterium callunae*, *Corynebacterium ammoniagenes*, *Corynebacterium thermoaminogenes*, *Corynebacterium melassecola*, *Corynebacterium efficiens*, *Corynebacterium efficiens*, *Corynebacterium deserti*, *Brevibacterium flavum*, *Brevibacterium lactofermentum*, *Brevibacterium divarecatum*, *Pseudomonas putida*, *Pseudomonas syringae*, *Streptomyces coelicolor*, *Streptomyces lividans*, *Streptomyces albus*, *Streptomyces avermitilis*, *Gluconobacter oxydans*, *Gluconobacter morbifer*, *Gluconobacter thailandicus*, *Acetobacter acetii*, *Clostridium acetobutylicum*, *Clostridium saccharobutylicum*, *Clostridium beijerinckii*, *Streptococcus equisimilis*, *Streptococcus pyogenes*, *Streptococcus uberis*, *Streptococcus equi* subsp., *Zooepidemicus* o *Basfia succiniciproducens*.

En una realización, la célula hospedante no expresa de modo natural la proteína de interés. Así, en una realización, la construcción de ácido nucleico descrita en el presente documento es heteróloga para la célula hospedante.

En otra realización, la célula hospedante bacteriana puede contener además modificaciones, por ejemplo, deleciones o inactivaciones, de otros genes que pueden ser perjudiciales para la producción, la recuperación o la aplicación de un polipéptido de interés. En una realización, una célula hospedante bacteriana es una célula deficiente en proteasas. En otra realización, la célula hospedante bacteriana, por ejemplo, una célula de *Bacillus*, comprende una inactivación o deleción de genes de proteasas extracelulares, incluidas, entre otras, aprE, mpr, vpr, bpr y/o epr. En una realización, la célula hospedante bacteriana no produce esporas. En otra realización, la célula hospedante bacteriana, por ejemplo, una célula de *Bacillus*, comprende una inactivación o deleción de spollAC, sigE y/o sigG. En una realización, la célula hospedante bacteriana, por ejemplo, una célula de *Bacillus*, comprende una inactivación o deleción de uno de los genes implicados en la biosíntesis de surfactina, por ejemplo, srfA, srfB, srfC y/o srfD. Véase, por ejemplo, la patente de EE. UU. n.º 5 958 728. En otra realización, la célula hospedante bacteriana comprende una inactivación o deleción de uno de los genes implicados en la biosíntesis del ácido poliglútamico. Otros genes, incluidos, entre otros, el gen amyE, que son perjudiciales para la producción, la recuperación o la aplicación de un polipéptido de interés, también pueden ser inactivados o eliminados.

En otra realización, la célula hospedante bacteriana es una célula hospedante de *E. coli* defectuosa en los sistemas de modificación-restricción de EcoKI, algunas además son defectuosas en las restrictasas dependientes de la metilación mcrA, mcrB, mcrC, mrr, algunas además son defectuosas en ADN metiltransferasas dam y dcm. En una

realización, la célula hospedante de *Escherichia coli* es deficiente en la metilación de dam y/o dcm. En otra realización, la célula hospedante de *Escherichia coli* es recA positiva. En otra realización, la célula hospedante de *Escherichia coli* es deficiente en la metilación de dam y/o dcm y es recA positiva.

5 En otra realización, la célula hospedante bacteriana es una célula hospedante de clonación convencional de *E. coli*, incluidas, entre otras, DH5alpha (Invitrogen), DH10B (Invitrogen), Omnimax (Invitrogen), INV110 (Invitrogen), TOP10 (Invitrogen), HB101 (Promega), SURE (Stratagene), XL1-Blue (Stratagene), TG1 (Lucigen) y JM109 (NEB). En otra realización, la célula hospedante bacteriana es una célula hospedante de clonación convencional de *Bacillus subtilis*, incluidas, entre otras, *B. subtilis* que porta un locus hsd(RI)R-M- defectuoso, tal como *B. subtilis* IG-20 (BGSC 1A436) o una mutación hsdRM1 defectuosa, tal como *B. subtilis* 1012 WT (Mobitec).

10 En una realización, la construcción de ácido nucleico y/o el vector de expresión pueden formar parte del ADN cromosómico o pueden ser un ADN extracromosómico de la célula hospedante. En una realización, la construcción de ácido nucleico y/o el vector de expresión se encuentran en la célula hospedante como ADN plasmídico, ADN viral o ADN lineal.

Procedimientos de la invención

15 En otra realización, la presente invención se dirige a un procedimiento de expresión de un polinucleótido, que comprende las etapas de:

(a) proporcionar una célula hospedante que comprende una construcción de ácido nucleico heteróloga que comprende un polinucleótido mediante la introducción de la construcción de ácido nucleico que comprende el polinucleótido en la célula hospedante;

20 (b) cultivar la célula hospedante recombinante de la etapa (a) en condiciones que conducen a la expresión del polinucleótido; y

(c) opcionalmente, recuperar una proteína de interés codificada por el polinucleótido,

25 en el que el polinucleótido de la construcción de ácido nucleico está unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en la célula hospedante, en el que al menos una secuencia de control comprende una secuencia promotora de un operón que comprende un gen *secA* tal como se describe en el presente documento y en el que dicha secuencia promotora es heteróloga al polinucleótido tal como se describe en el presente documento, en el que el polinucleótido no comprende un gen indicador y la secuencia promotora tiene al menos un 70 % de identidad de secuencia con SEQ ID NO:9.

30 En una realización preferida, la secuencia promotora utilizada en el procedimiento tiene al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 92 %, al menos un 95 %, al menos un 98 %, al menos un 99 % o incluso un 100 % de identidad de secuencia con SEQ ID NO:9, y en la que la secuencia promotora procede preferentemente de una especie de *Bacillus*, en una realización específica preferentemente de *Bacillus licheniformis*. Preferentemente, la secuencia promotora utilizada en el procedimiento tiene al menos un 70 % o un 95 % de identidad de secuencia con SEQ ID NO:9 y en la que la secuencia promotora procede de una especie de *Bacillus*.

35 Preferentemente, el polinucleótido no comprende un gen indicador como se describe en el presente documento. Preferentemente, la proteína de interés es tolerada por la célula hospedante solo en cantidades limitadas como se describe en el presente documento. Preferentemente, la proteína de interés se selecciona del grupo que consiste en una enzima modificadora de ADN o ARN como se describe en el presente documento.

40 En otra realización, la presente invención se dirige a un procedimiento de expresión de un polinucleótido, que comprende las etapas de:

(a) proporcionar una célula hospedante que comprende una construcción de ácido nucleico heteróloga que comprende un polinucleótido mediante la introducción de la construcción de ácido nucleico que comprende el polinucleótido en la célula hospedante, en el que el polinucleótido codifica una proteína de interés, en el que la proteína de interés no es una proteína indicadora, preferentemente en el que la proteína de interés es una enzima modificadora del ADN o ARN;

45 (b) cultivar la célula hospedante recombinante de la etapa (a) en condiciones que conducen a la expresión del polinucleótido; y

50 en el que el polinucleótido de la construcción de ácido nucleico está unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en la célula hospedante, en el que al menos una secuencia de control comprende una secuencia promotora de un operón que comprende un gen *secA* tal como se describe en el presente documento y en el que dicha secuencia promotora es heteróloga al polinucleótido tal como se describe en el presente documento, en el que el polinucleótido no comprende un gen indicador y la secuencia promotora tiene al menos un 70 % de identidad de secuencia con SEQ ID NO:9.

En otra realización, la presente invención se dirige a un procedimiento de expresión de un polinucleótido, que comprende las etapas de:

5 (a) proporcionar una célula hospedante que comprende una construcción de ácido nucleico heteróloga que comprende un polinucleótido mediante la introducción de la construcción de ácido nucleico que comprende el polinucleótido en la célula hospedante, en la que el polinucleótido codifica una enzima modificadora de ADN o ARN;

(b) cultivar la célula hospedante recombinante de la etapa (a) en condiciones que conducen a la expresión del polinucleótido; y

10 en el que el polinucleótido de la construcción de ácido nucleico está unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en la célula hospedante, en el que al menos una secuencia de control comprende una secuencia promotora de un operón que comprende un gen *secA* tal como se describe en el presente documento y en el que dicha secuencia promotora es heteróloga al polinucleótido tal como se describe en el presente documento, en el que el polinucleótido no comprende un gen indicador y la secuencia promotora tiene al menos un 70 % de identidad de secuencia con SEQ ID NO:9.

15 En otra realización, la presente invención se dirige a un procedimiento de expresión de un polinucleótido, que comprende las etapas de:

20 (a) proporcionar una célula hospedante que comprende una construcción de ácido nucleico heteróloga que comprende un polinucleótido mediante la introducción de la construcción de ácido nucleico que comprende el polinucleótido en la célula hospedante, en la que el polinucleótido codifica una enzima modificadora de ADN o ARN;

(b) cultivar la célula hospedante recombinante de la etapa (a) en condiciones que conducen a la expresión del polinucleótido; y

25 en el que el polinucleótido de la construcción de ácido nucleico está unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en la célula hospedante, en el que al menos una secuencia de control comprende una secuencia promotora de un operón que comprende un gen *secA* y en el que dicha secuencia promotora es heteróloga al polinucleótido tal como se describe en el presente documento, en el que el polinucleótido no comprende un gen indicador y la secuencia promotora tiene al menos un 70 % de identidad de secuencia con SEQ ID NO:9.

30 La presente invención también puede utilizarse para mejorar la transformación de una célula hospedante con una segunda construcción de ácido nucleico heteróloga o para mejorar o facilitar el cultivo de una célula hospedante que ha sido transformada con una segunda construcción de ácido nucleico.

En otra realización, la presente invención se dirige a un procedimiento de expresión de un polinucleótido, que comprende las etapas de:

35 (a) introducir en una célula hospedante una primera construcción de ácido nucleico heteróloga que comprende un polinucleótido mediante la introducción de la primera construcción de ácido nucleico que comprende el polinucleótido en la célula hospedante, en el que el polinucleótido de la primera construcción de ácido nucleico está unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en la célula hospedante, en el que al menos una secuencia de control comprende una secuencia promotora de un operón *secA* que comprende un gen *secA* tal como se describe en el presente documento y en el que dicha secuencia promotora es heteróloga al polinucleótido tal como se describe en el presente documento, en el que el polinucleótido no comprende un gen indicador y la secuencia promotora tiene al menos un 70 % de identidad de secuencia con SEQ ID NO:9;

(b) opcionalmente, cultivar la célula hospedante recombinante de la etapa (a) en condiciones que conducen a la expresión del polinucleótido de la primera construcción de ácido nucleico; y

45 (c) introducir en la célula hospedante una segunda construcción de ácido nucleico heteróloga que comprende un polinucleótido que codifica una segunda proteína de interés mediante la introducción de la segunda construcción de ácido nucleico en la célula hospedante, y

(d) cultivar la célula hospedante recombinante de la etapa (c) en condiciones que conducen a la expresión del polinucleótido de la primera y segunda construcción de ácido nucleico; y

50 (e) opcionalmente, recuperar una proteína de interés codificada por el polinucleótido de la segunda construcción de ácido nucleico.

En una realización, la expresión del polinucleótido de la primera construcción de ácido nucleico mejora la introducción de la segunda construcción de ácido nucleico en la célula hospedante o mejora o facilita el cultivo de la célula hospedante tras la transformación con la segunda construcción de ácido nucleico.

En una realización, el polinucleótido de la primera construcción de ácido nucleico codifica una proteína de interés que es tolerada por la célula hospedante solo en cantidades limitadas como se describe en el presente documento. Preferentemente, la proteína de interés se selecciona del grupo que consiste en una enzima modificadora de ADN o ARN como se describe en el presente documento.

- 5 En una realización, el polinucleótido de la segunda construcción de ácido nucleico codifica una proteína de interés comercial, preferentemente, seleccionada del grupo que consiste en amilasa, proteasa, lipasa, mananasa, fitasa y celulasa.

En otra realización, la presente invención se dirige a un procedimiento de producción de una proteína de interés que comprende las etapas de:

- 10 (a) proporcionar una célula hospedante que comprende una construcción de ácido nucleico heteróloga que comprende un polinucleótido que codifica la proteína de interés, en el que el polinucleótido está unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en la célula hospedante, en el que al menos una secuencia de control comprende una secuencia promotora de un operón que comprende un gen *secA* como se describe en el presente documento y en el que dicha secuencia
- 15 promotora es heteróloga al polinucleótido como se describe en el presente documento, en el que el polinucleótido no comprende un gen indicador y la secuencia promotora tiene al menos un 70 % de identidad de secuencia con SEQ ID NO:9 mediante la introducción de la construcción de ácido nucleico en la célula hospedante;
- 20 (b) cultivar la célula hospedante recombinante de la etapa (a) en condiciones que conducen a la producción de la proteína de interés; y
- (c) opcionalmente, recuperar una proteína de interés.

En una realización preferida, la secuencia promotora utilizada en los procedimientos descritos en el presente documento tiene una identidad de secuencia de al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 92 %, al menos un 95 %, al menos un 98 %, al menos un 99 % o incluso un 100 % con SEQ ID

25 NO:9 y la secuencia promotora preferentemente procede de una especie de *Bacillus*, en una realización específica procede de *Bacillus licheniformis*. Preferentemente, la secuencia promotora utilizada en el procedimiento tiene al menos un 70 % o un 95 % de identidad de secuencia con SEQ ID NO:9 y la secuencia promotora procede de una especie de *Bacillus*.

30 La célula hospedante que comprende la construcción de ácido nucleico descrita en el presente documento puede obtenerse, sin estar limitada a ello, por cualquiera de los procedimientos para obtener células hospedantes transgénicas descritos en el presente documento.

En otra realización, la presente invención se dirige a un procedimiento para proporcionar un nivel de expresión moderado de un polinucleótido, en una realización específica un polinucleótido que codifica una proteína de interés, en una célula hospedante que comprende las etapas de:

- 35 (a) proporcionar una célula hospedante que comprende una construcción de ácido nucleico heteróloga que comprende un polinucleótido que codifica la proteína de interés, en el que el polinucleótido está unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en la célula hospedante, en el que al menos una secuencia de control comprende una secuencia promotora de un operón que comprende un gen *secA* como se describe en el presente documento y en el que dicha secuencia
- 40 promotora es heteróloga al polinucleótido como se describe en el presente documento, en el que el polinucleótido no comprende un gen indicador y la secuencia promotora tiene al menos un 70 % de identidad de secuencia con SEQ ID NO:9 mediante la introducción de la construcción de ácido nucleico en la célula hospedante;
- 45 (b) cultivar la célula hospedante recombinante de la etapa (a) en condiciones que conducen a la expresión del polinucleótido con un nivel de expresión moderado.

En otra realización, la presente invención se dirige al uso de un promotor *secA* de un operón que comprende un gen *secA* y en el que la secuencia promotora tiene al menos un 70 % de identidad de secuencia con SEQ ID NO:9, para proporcionar un nivel de expresión moderado de un polinucleótido, en una realización específica un polinucleótido que codifica una proteína de interés, en una célula hospedante.

50 En otra realización, la presente invención se dirige al uso de un promotor *secA* como se describe en el presente documento para proporcionar un nivel de expresión moderado de un polinucleótido, en una realización específica un polinucleótido que codifica una proteína de interés, en una célula hospedante que comprende las etapas de:

- 55 (a) proporcionar una célula hospedante que comprende una construcción de ácido nucleico heteróloga que comprende un polinucleótido que codifica la proteína de interés, en el que el polinucleótido está unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en la célula

hospedante, en el que al menos una secuencia de control comprende una secuencia promotora de un operón que comprende un gen *secA* como se describe en el presente documento y en el que dicha secuencia promotora es heteróloga al polinucleótido como se describe en el presente documento, en el que el polinucleótido no comprende un gen indicador y la secuencia promotora tiene al menos un 70 % de identidad de secuencia con SEQ ID NO:9;

(b) cultivar la célula hospedante recombinante de la etapa (a) en condiciones que conducen a la expresión del polinucleótido con un nivel de expresión moderado.

En una realización, la proteína de interés es una metiltransferasa. Así, otra realización es un procedimiento de producción de una ADN metiltransferasa, que comprende las etapas de:

(a) proporcionar una célula hospedante que comprende un polinucleótido heterólogo que codifica una ADN metiltransferasa, en el que la ADN metiltransferasa metila el ADN dando como resultado un ADN que contiene 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC y en el que la ADN metiltransferasa comprende menos de 35 residuos de aminoácidos entre el residuo de aminoácido 72 y el residuo de aminoácido 106 según la numeración de SEQ ID NO:33 como se describe en el presente documento;

(b) cultivar la célula hospedante recombinante de la etapa (a) en condiciones que conducen a la producción de la ADN metiltransferasa; y

(c) opcionalmente, recuperar la ADN metiltransferasa.

El cultivo de la célula hospedante recombinante y la recuperación de la proteína de interés pueden llevarse a cabo mediante procedimientos convencionales de la técnica anterior, que se describen con más detalle en el presente documento.

En una realización, la construcción de ácido nucleico y/o el vector de expresión y la célula hospedante son como se ha descrito anteriormente.

Las células hospedantes se cultivan en un medio nutritivo que favorece el crecimiento de la célula hospedante. En una realización, las células hospedantes se cultivan en un medio nutritivo que favorece el crecimiento de la célula hospedante y durante un tiempo y en condiciones adecuadas para expresar el polinucleótido utilizando procedimientos conocidos en la técnica. Por ejemplo, la célula puede cultivarse en un matraz de agitación o mediante fermentación a pequeña o gran escala (que incluyen fermentaciones continuas, discontinuas, de alimentación discontinua o en estado sólido) en fermentadores de laboratorio o industriales realizadas en un medio adecuado y en condiciones que permitan conseguir la expresión del polinucleótido. En una realización, el cultivo de la célula hospedante es por fermentación a escala industrial. El cultivo tiene lugar en un medio nutritivo adecuado que comprende fuentes de carbono y nitrógeno y sales inorgánicas, utilizando procedimientos conocidos en la técnica. Los medios adecuados pueden adquirirse en proveedores comerciales o prepararse según composiciones publicadas (por ejemplo, en los catálogos de la American Type Culture Collection).

La proteína de interés codificada por el polinucleótido puede acumularse en la célula o puede ser secretada fuera de la célula. El polipéptido secretado de interés puede recuperarse directamente del medio. La expresión del polinucleótido puede detectarse utilizando procedimientos conocidos en la técnica que sean específicos para el polinucleótido. Estos procedimientos de detección pueden incluir el uso de procedimientos basados en la PCR, procedimientos basados en la hibridación, anticuerpos específicos contra la proteína codificada, cromatografía líquida de alta resolución, cromatografía capilar, formación de un producto enzimático, desaparición de un sustrato enzimático o SDS-PAGE. Por ejemplo, puede utilizarse un ensayo enzimático para determinar la actividad de la enzima. Los procedimientos para determinar la actividad enzimática son conocidos en la técnica para muchas enzimas (véase, por ejemplo, D. Schomburg y M. Salzman (eds.), *Enzyme Handbook*, Springer-Verlag, Nueva York, 1990). En el presente documento se describen ensayos para determinar la actividad de una endonucleasa de restricción o de una ADN metiltransferasa.

Una proteína de interés codificada por el polinucleótido puede aislarse por procedimientos conocidos en la técnica. Por ejemplo, una proteína de interés puede aislarse del caldo de fermentación mediante procedimientos convencionales que incluyen, entre otros, centrifugación, filtración, extracción, secado por pulverización, evaporación o precipitación. En función del constructo de expresión utilizado, la proteína de interés puede ser secretada hacia el caldo de fermentación o permanecer en el interior de la célula hospedante. En este último caso, la proteína de interés puede recuperarse del caldo de fermentación aplicando una etapa en la que se lisan las células. A continuación, el polipéptido aislado puede purificarse aún más mediante una diversidad de procedimientos conocidos en la técnica que incluyen, entre otros, cromatografía (por ejemplo, de intercambio iónico, de afinidad, hidrófoba, de cromatografía de exclusión por tamaño), procedimientos electroforéticos (por ejemplo, enfoque isoeléctrico preparativo ("preparative isoelectric focusing", IEF), solubilidad diferencial (por ejemplo, precipitación con sulfato de amonio) o extracción (véase, por ejemplo, *Protein Purification*, J.-C. Janson y Lars Ryden, editores, VCH Publishers, Nueva York, 1989). A continuación, el polipéptido purificado puede concentrarse mediante procedimientos conocidos en la técnica que incluyen, entre otros, la ultrafiltración y la evaporación, en conocido, la evaporación de película fina. En otra realización,

la proteína de interés no se purifica a partir del caldo de fermentación. En una realización específica, la proteína de interés no se secreta hacia el caldo de fermentación y no se recupera del caldo de fermentación.

Ejemplos

5 Los siguientes ejemplos solo pretenden ilustrar la presente divulgación. Las numerosas variaciones posibles que son obvias para un experto en la materia también entran dentro del alcance de la invención.

A menos que se indique lo contrario, los siguientes experimentos se han realizado aplicando equipos, procedimientos, productos químicos y bioquímicos convencionales utilizados en ingeniería genética y en la producción fermentativa de compuestos químicos mediante el cultivo de microorganismos. Véase también Sambrook *et al.* (Sambrook, J. y Russell, D.W. Molecular Cloning: A laboratory manual, 3ª ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001) y Chmiel *et al.* (Bioprozesstechnik 1. Einführung in die Bioverfahrenstechnik, Gustav Fischer Verlag, Stuttgart, 1991).

Células electrocompetentes de Bacillus licheniformis y electroporación

15 La transformación de ADN en *B. licheniformis* ATCC 53926 se realizó mediante electroporación. La preparación de células electrocompetentes de *B. licheniformis* ATCC 53926 y la transformación del ADN se llevan a cabo según lo descrito fundamentalmente por Brigidi *et al.* (Brigidi, P., Mateuzzi, D. (1991), Biotechnol. Techniques, 5, 5) con la siguiente modificación: Tras la transformación del ADN, las células se recuperan en 1 ml de tampón LBSPG y se incuban durante 60 min a 37 °C (Vehmaanperä J., 1989, FEMS Microbio. Lett., 61: 165-170) tras la siembra en placas selectivas de LB-agar. Si no se indica lo contrario, el ADN extraño al ADN de *B. licheniformis* ATCC 53926 se metila *in vitro* según el procedimiento descrito en la patente DE4005025.

Aislamiento de plásmidos

20 El ADN plasmídico se aisló de células de *Bacillus* y *E. coli* mediante procedimientos convencionales de biología molecular descritos en Sambrook, J. y Russell, D.W. Molecular Cloning: A laboratory manual, 3ª ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001, o el procedimiento de lisis alcalina (Birnboim, H. C., Doly, J. (1979), Nucleic Acids Res., 7(6): 1513-1523). En comparación con *E. coli*, las células de *Bacillus* se trataron con lisozima 10 mg/ml durante 30 min a 37 °C antes de la lisis celular.

Plásmidos

Plásmido pUK56 y pUK56S: Plásmido de expresión de proteasa

30 El casete de expresión de proteasa del plásmido pCB56C (documento US5352604) se amplificó por PCR con los oligonucleótidos de SEQ ID NO:1 y SEQ ID NO:2 y el esqueleto del plásmido pUB110 que comprende repU y el gen de resistencia a la kanamicina se amplificó por PCR con los oligonucleótidos de SEQ ID NO:3 y SEQ ID NO:4. Los fragmentos de PCR se cortaron con las enzimas de restricción SmaI y SnaBI, se ligaron con T4-DNA ligasa (NEB) tras la transformación en células competentes de *Bacillus subtilis* 168 según el protocolo de Spizizen (Anagnostopoulos, C. y Spizizen, J. (1961), J. Bacteriol., 81, 741-746). Los clones correctos del plásmido final pUK56 se analizaron mediante digestión con enzimas de restricción y secuenciación. El plásmido pUK56 se cortó con SnaBI y el fragmento de pBR322 se amplificó por PCR con los oligonucleótidos de SEQ ID NO:5 y SEQ ID NO:6 cortado con SnaBI/EcoRV (número de registro pBR322 J01749.1) y se clonó en pUK56 tras su transformación en células competentes *E. coli* XL1-Blue (Stratagene). Se recuperó el plásmido lanzadera de *E. coli/Bacillus* pUK56S con sitios RE SnaBI funcionales.

Plásmido pLCS3: plásmido de expresión

40 El origen de replicación de bajo número de copias pSC101 del plásmido pZS4-Int-1 (Lutz, R. y Bujard, H. (1997), Nucleic Acids Res., 25, 1203-1210; número de registro U66308) se recuperó mediante digestión con las endonucleasas de restricción XbaI y ScaI y se clonó en pZA3PLtetO-1 luc (número de registro U66309) cortado con las endonucleasas de restricción ScaI y AvrII para sustituir el origen de replicación dando lugar al plásmido pLCS3.

Plásmido pLCS31: plásmido de expresión con represor lac

45 El fragmento del gen represor lac de *E. coli* se amplificó por PCR a partir de pZS4-int1 (Lutz, R. y Bujard, H. (1997), Nucleic Acids Res., 25, 1203-1210, número de registro U66308) con los oligonucleótidos de SEQ ID NO:7 y SEQ ID NO:8. El fragmento PCR se cortó con las enzimas de restricción KpnI y XbaI y se clonó en pLCS3 digerido con KpnI y XbaI para crear el plásmido pLCS31.

Plásmido pLCS4: plásmido de expresión

50 El fragmento del gen de resistencia a la kanamicina (SmaI/XhoI) de pZE2 PLtetO-1 MCS2 (Lutz, R. y Bujard, H. (1997), Nucleic Acids Res., 25, 1203-1210; número de registro de gen U66312) se clonó en el plásmido pLCS3 cortado con las endonucleasas de restricción SmaI/XhoI para sustituir al gen de resistencia al cloranfenicol, dando lugar al plásmido pLCS4.

Plásmido pEDS3: plásmido de expresión

5 El fragmento de ADN sintético que comprende un fragmento de la región de control del gen *secA* de *B. licheniformis* (SEQ ID NO:9), el terminador lambda T0 (Stueber, D. y Bujard, H. (1982), EMBO J., 1, 1399-1404) que comprende dos sitios Bsal RE, flanqueados por los sitios de restricción XhoI y XbaI (SEQ ID NO:10) se clonó en el plásmido pLCS3 cortado con XhoI y XbaI dando lugar al plásmido pEDS3.

Plásmido pEDT31: plásmido de expresión

10 El fragmento de ADN sintético que comprende el promotor T5/lac (SEQ ID NO:11) de pDS56 (Stueber, D., Garotta, G. (1990), Immunological Methods, vol. IV, Academic Press, Nueva York, págs. 121-152), el terminador lambda T0 (Stueber, D. y Bujard, H. (1982), EMBO J., 1, 1399-1404) que comprende dos sitios Bsal RE, flanqueados por los sitios de restricción XhoI y KpnI (SEQ ID NO:12) se clonó en el plásmido pLCS31 cortado con XhoI y KpnI dando lugar al plásmido pEDT31.

15 Plásmido pMDS001-006, pMDT002-003: Construcciones de expresión génica de ADN metiltransferasas Los genes de las ADN metiltransferasas se encargaron como fragmentos génicos sintéticos que comprendían la 5'UTR/RBS del gen *secA* de *B. licheniformis* (SEQ ID NO:13), la secuencia codificante (*cds*) del gen de la MTasa (véase el listado de secuencias), flanqueada por sitios de restricción Bsal con proyecciones compatibles con la restricción para su posterior clonación en los plásmidos pEDS3 y pEDT31. Los sitios de restricción Bsal internos se eliminaron por variación del codón-triplete.

Tabla 1: Construcciones de expresión de ADN metiltransferasa

Polinucleótido de MTasa (SEQ ID NO:)	Plásmido de destino	Plásmido de MTasa
SEQ ID NO:22	pEDS3	pMDS001
SEQ ID NO:23	pEDS3	pM DS002
SEQ ID NO:19	pEDS3	pM DS002
SEQ ID NO:25	pEDS3	pM DS002
SEQ ID NO:27	pEDS3	pM DS002
SEQ ID NO:24	pEDS3	pM DS002
SEQ ID NO:19	pEDT31	pM DS002
SEQ ID NO:21	pEDT31	pM DS002

20 *Plásmido pBIL009: Plásmido de integración de Bacillus subtilis*

25 El plásmido de integración pBS1C amyE (Radeck, J. *et al.* (2013), J. Biol. Eng., 7, 29) para *B. subtilis* se amplificó mediante PCR con los oligonucleótidos de SEQ ID NO:14 y SEQ ID NO:15 restringidos con Bsal tras su clonación en el esqueleto del plásmido pLCS4 que comprende el origen de replicación pSC101 y el gen de resistencia a la kanamicina recuperado como fragmento de digestión de restricción XbaI/XhoI. La mezcla de ligamiento se transformó en células *E. coli* XL1-Blue (Stratagene) y los clones se recuperaron en placas de LB-agar que contenían kanamicina 20 µg/ml. Los clones positivos que produjeron el plásmido pBIL009 se analizaron mediante digestión de restricción y el gen funcional de resistencia al cloranfenicol.

Plásmido pMIS012: ADN metiltransferasa - Construcción de expresión génica de B. subtilis

30 La construcción de expresión de ADN metiltransferasa de pMDS003 se amplificó por PCR con los oligonucleótidos de SEQ ID NO:16 y SEQ ID NO:17 restringidos con BamHI/XbaI tras su clonación en el esqueleto del plásmido pBIL009 recuperado como fragmento de digestión de restricción BamHI/XbaI. La mezcla de ligamiento se transformó en células *E. coli* XL1-Blue (Stratagene) y los clones se recuperaron en placas de LB-agar que contenían kanamicina 20 µg/ml. Los clones positivos que produjeron el plásmido pMIS012 se analizaron mediante digestión de restricción y el gen funcional de resistencia al cloranfenicol.

35

Predicción de estructuras

Las estructuras de las metiltransferasas se predijeron utilizando la herramienta de modelización de homología SWISS-MODEL (Biasini M., Bienert S., Waterhouse A., Arnold K., Studer G., Schmidt T., Kiefer F., Cassarino T.G., Bertoni M., Bordoli L., Schwede T. (2014), SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information, *Nucleic Acids Research*, 2014 (1 de julio de 2014), 42 (W1): W252-W258) utilizando los parámetros por defecto y las siguientes plantillas estructurales de la base de datos RCSB PDB (Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. (2000), *The Protein Data Bank Nucleic Acids Research*, 28: 235-242): 2uyc_A (utilizado para SEQ ID NO:33 (M_Fnu4HI), SEQ ID NO:33 (M_RBH3250), SEQ ID NO:41 (M_CocII)), 2i9k_C (utilizado para SEQ ID NO:33 (M_Bsp6I), SEQ ID NO:43 (M_LlaDII)), 3swr_A (utilizado para SEQ ID NO:33 (M_Cdi13307II), SEQ ID NO:38 (M_Cdi630IV)), 1mht_C (utilizado para SEQ ID NO:39 (M_Ckr177III)), 2z6u_A (utilizado para SEQ ID NO:40 (M_CmaLM2II)) y 9mht_C (utilizado para SEQ ID NO:42 (M_Fsp4HI).

Alineamiento estructural

Las estructuras predichas se alinearon estructuralmente con la estructura predicha de SEQ ID NO:33 (M_Fnu4HI) con TAlign, versión 20160521 (Y. Zhang, J. Skolnick (2005), TM-align: A protein structure alignment algorithm based on TM-score, *Nucleic Acids Research*, 33: 2302-2309) utilizando los parámetros por defecto.

15 *Alineamiento múltiple de secuencias basada en estructuras*

Los alineamientos estructurales por pares se combinaron en un alineamiento múltiple de secuencias utilizando MAFFT, versión 7.221 (Kato, S. (2013), MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Molecular Biology and Evolution*, 30:772-780) utilizando los parámetros por defecto del modo de fusión. La anotación de la estructura secundaria se añadió a la figura como consenso de las predicciones estructurales.

20 *Selección de secuencias*

Todas las ADN (citosina-5)-metiltransferasas con una secuencia de reconocimiento de GCNGC y que se verificaron experimentalmente con una técnica de vanguardia (determinada por tener 'PacBio' en su campo de comentarios) se extrajeron de REBASE (Roberts R.J., Vincze T., Posfai J., Macelis D. (2015), REBASE-a database for DNA restriction and modification: enzymes, genes and genomes, *Nucleic Acids Research*, 43: D298-D299). A este conjunto de metiltransferasas se le añadieron secuencias para las que se disponía de datos internos (M.Fnu4HI y M.RBH03250) o datos procedentes de la bibliografía (M.Fsp4HI (Chmuzh, E.V. y Degtiarev, S.K., 2007, *Mol. Biol. (Mosk)*, 41, 43-50), M.Bsp6I (Lubys *et al.*, 1995, *Gene*, 157: 25-29), M.LlaDII (Madsen *et al.*, 1995, *Applied and Environmental Microbiology*, 64(7): 2424-2431)) que confirman las secuencias de reconocimiento.

Ejemplo 1: Generación de ADN metilado *in vivo* en células de *E. coli*

30 Se transformaron células *E. coli* INV110 competentes (Invitrogen/Life technologies) con el plásmido pUK65S y se seleccionaron en placas de LB con kanamicina 20 µg/ml, obteniéndose la cepa de *E. coli* Ec n.º 082. La cepa Ec n.º 082 de *E. coli* se hizo competente según el procedimiento de Chung (Chung, C.T., Niemela, S.L., y Miller, R.H. (1989), One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution, *Proc. Natl. Acad. Sci. U. S. A.*, 86, 2172-2175) y se transformaron con plásmidos de expresión que codifican la ADN metiltransferasa (tabla 1) tras la selección en placas de LB-agar que contenían kanamicina 20 µg/ml y cloranfenicol 30 µg/ml. Los plásmidos de expresión de MTasa que utilizan el promotor *secA* se construyeron como se ha descrito anteriormente (tabla 1). Se indican el nombre de la cepa de *E. coli*, los nombres de los plásmidos y los genes de la MTasa.

Tabla 2

Nombre de <i>E. coli</i>	Plásmidos	Polinucleótido de MTasa (SEQ ID NO)	Promotor
Ec n.º 082	pUK56S	-	-
Ec n.º 083	pUK56S, pMDS001	22	<i>secA</i>
Ec n.º 084	pUK56S, pMDS002	23	<i>secA</i>
Ec n.º 085	pUK56S, pMDS003	19	<i>secA</i>
Ec n.º 086	pUK56S, pMDS004	25	<i>secA</i>
Ec n.º 087	pUK56S, pMDS005	27	<i>secA</i>
Ec n.º 088	pUK56S, pMDS006	24	<i>secA</i>

El ADN plasmídico total se aisló de las diferentes cepas de *E. coli* de acuerdo con procedimientos convencionales en biología molecular (Sambrook, J. y Russell, D.W., Molecular Cloning: A laboratory manual, 3ª ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001.) y se determinó la eficiencia de la metilación *in vivo* por restricción de 1 µg de ADN plasmídico total con *SatI* (ThermoFisher Scientific), cuya ruptura es inhibida por la 5-metilcitosina dentro de la secuencia de reconocimiento GCNGC. Las reacciones de restricción se analizaron mediante electroforesis en gel de agarosa con tinción de bromuro de etidio para su visualización. Se utilizó la escalera de ADN de 1 kb *Generuler* (ThermoFisher Scientific) para calcular el tamaño de los fragmentos de ADN (figura 2). El ADN plasmídico aislado a partir de cepas de *E. coli* con metilación del ADN de 5-metilcitosina específica de GCNGC está protegido frente a la restricción por *SatI*, mientras que pUK56S de la cepa *Ec n.º 082* de *E. coli* no lo está.

Ejemplo 2: Transformación de ADN metilado de células de *E. coli* en *Bacillus licheniformis*

Se aisló ADN plasmídico de células de *E. coli* *Ec n.º 082-Ec n.º 088* como se describe en el ejemplo 1 y se transformó 1 µg de ADN plasmídico en células electrocompetentes de *B. licheniformis* ATCC 53926 como se describe fundamentalmente en Brigidi *et al.* (Brigidi, P., Mateuzzi, D. (1991), *Biotechnol. Techniques*, 5, 5) con la siguiente modificación: Tras la transformación del ADN, las células se recuperan en 1ml de tampón LBSPG- y se incuban durante 60min a 37°C (Vehmaanperä J., 1989, *FEMS Microbio. Lett.*, 61: 165-170) tras la siembra en placas de LB-agar con kanamicina 20 µg/ml. Las placas de LB-agar se incuban toda la noche a 37 °C y se determina la eficiencia de la transformación en unidades formadoras de colonias (ufc). Las eficiencias de transformación del ADN plasmídico de diferentes cepas de *E. coli* se normalizaron con respecto a la cepa de *E. coli* *Ec n.º 083*, que se fijó en el 100 %. Obsérvese que la cepa de *E. coli* *Ec n.º 083* es portadora de la ADN metiltransferasa de *B. licheniformis* ATCC 53926. La cepa de *E. coli* *Ec n.º 084* porta una variante codones optimizados de la ADN metiltransferasa de *B. licheniformis* ATCC 53926 y actúa como control para la expresión génica. El ADN del plásmido pUK56S de *E. coli* *Ec n.º 082* que no fue metilado por una ADN metiltransferasa específica de GCNGC no recuperó ninguna transformante. Sorprendentemente, el ADN plasmídico aislado de cepas de *E. coli* portadoras de MTAsas (*Ec n.º 85-87*) heterólogas a *B. licheniformis* ATCC 53926 transformado en *B. licheniformis* ATCC 53926 dio lugar a eficiencias de transformación significativamente mayores (figura 3). Además, el ADN plasmídico aislado de la cepa de *E. coli* portadora de la MTasa homóloga (*Ec n.º 88*) a *B. licheniformis* ATCC 53926 con una delección de los aminoácidos 103-108 de SEQ ID NO:34 (se truncaron 6 aminoácidos en total, dando como resultado SEQ ID NO:35) también dio lugar a una eficiencia de transformación significativamente mayor en comparación con *Ec n.º 83*.

Ejemplo 3: Metilación *in vivo* en *B. subtilis*

El plásmido de expresión de MTasa pMIS012 para su integración en el gen *amyE* de *B. subtilis* se linealizó con la enzima de restricción *SacI* tras la transformación de 2 µg de ADN de plásmido linealizado en células *B. subtilis* 168 que se hicieron competentes según el procedimiento de Spizizen (Anagnostopoulos, C. y Spizizen, J. (1961), *J. Bacteriol.*, 81, 741-746). Las células se extendieron y se incubaron durante la noche a 37 °C en placas de LB-agar que contenían cloranfenicol 10 µg/ml. Se recogieron las colonias crecidas y se sembraron en estrías tanto en placas de LB-agar que contenían cloranfenicol 10 µg/ml como en placas de LB-agar que contenían cloranfenicol 10 µg/ml y almidón soluble al 0,5 % soluble (Sigma) tras incubarlas toda la noche a 37 °C. Las placas de almidón se cubrieron con una solución de Lugol yodada y se identificaron clones de integración positiva con actividad amilasa negativa. El ADN genómico de los clones positivos se aisló mediante procedimientos convencionales de extracción con fenol/cloroformo después de 30 minutos de tratamiento con lisozima (10 mg/ml) a 37 °C, tras analizar la correcta integración del casete de expresión de la MTasa mediante PCR. La cepa de *B. subtilis* resultante se denominó *Bs n.º 053*.

El plásmido pUK56 se transformó en células de *B. subtilis* 168 y *B. subtilis* *Bs n.º 053* que se hicieron competentes según el procedimiento de Spizizen (Anagnostopoulos, C. y Spizizen, J. (1961), *J. Bacteriol.*, 81, 741-746). El ADN del plásmido pUK56 fue metilado *in vitro* como se describe en la patente DE4005025 tras la transformación en células electrocompetentes de *B. licheniformis* ATCC 53926 como se describe en el ejemplo 2. Las transformantes se extendieron y se incubaron durante la noche a 37 °C en placas de LB-agar que contenían kanamicina 20 µg/ml y leche desnatada al 1 % generando las cepas de *B. subtilis* *Bs n.º 54* y *Bs n.º 55* y la cepa de *B. licheniformis* *Bli n.º 112*, respectivamente.

El ADN del plásmido pUK56 se aisló de las cepas *Bs n.º 54* y *Bs n.º 55* de *B. subtilis* y de la cepa *Bli n.º 112* de *B. licheniformis* como se describe en el ejemplo 1 después de 30 min de tratamiento con lisozima (10 mg/ml) a 37 °C. Se transformó 1 µg de ADN de cada uno de los plásmidos en células electrocompetentes de *B. licheniformis* ATCC 53926 como se describe en el ejemplo 2. Las eficiencias de transformación del plásmido pUK56 procedente de *B. subtilis* *Bs n.º 54* y *Bs n.º 55* se normalizaron frente a la eficiencia de transformación del plásmido pUK56 aislado de *B. licheniformis* *Bli n.º 112*, que se fijó en el 100 %. Sorprendentemente, el ADN del plásmido pUK56 aislado de *B. subtilis* *Bs n.º 55*, portador de una MTasa heteróloga a *B. licheniformis* ATCC 53926, en comparación con el ADN del plásmido pUK56 aislado de *B. licheniformis* *Bli n.º 112*, portador del patrón de metilación del ADN nativo de *B. licheniformis* ATCC 53926, dio lugar a una eficiencia de transformación significativamente mayor (figura 4). Por el contrario, casi no se recuperaron colonias tras la transformación del plásmido pUK56 de *B. subtilis* *Bs n.º 54*, que actuó como control.

Ejemplo 4: Estudios de promotores para la expresión de MTAsas heterólogas

Las ADN metiltransferasas de SEQ ID NO:19 y SEQ ID NO:21 se encargaron como fragmentos de genes sintéticos que comprendían la región 5'UTR/RBS de SEQ ID NO:13 y los sitios de restricción Bsal flanqueantes tras clonación en el vector pEDT31 restringido con la endonucleasa de restricción Bsal (NEB) y la transformación en células competentes de *E. coli* XL1-Blue (Stratagene), dando lugar a los plásmidos de expresión de MTasa pMDT002 y pMDT003, que contienen los genes que codifican las metiltransferasas bajo el control del promotor T5. La cepa de *E. coli* competente Ec n.º 082 (ejemplo 1) se transformó con plásmidos de expresión que codifican la ADN metiltransferasa tras la selección en placas de LB-agar que contenían kanamicina 20 µg/ml y cloranfenicol 30 µg/ml. Las cepas de *E. coli* resultantes se enumeran en la siguiente tabla 3.

5

Tabla 3

Nombre de <i>E. coli</i>	Plásmidos	MTasa (SEQ ID NO)	Promotor
Ec n.º 089	pUK56S, pMDT002	19	T5
Ec n.º 090	pUK56S, pMDT003	21	T5

Para la preparación del plásmido a partir de células de *E. coli* Ec n.º 089 y Ec n.º 090 con el gen de la MTasa bajo el control del promotor T5 que se utiliza habitualmente (SEQ ID NO:11), las células se cultivaron durante la noche en medio LB y la expresión de MTasa se indujo con IPTG 100 µM 5 h antes de la recolección de las células.

15 Para comparar el promotor T5 con el promotor *secA* se aisló ADN plasmídico de células de *E. coli* Ec n.º 082 que no expresaban la metiltransferasa, de células de *E. coli* que expresaban la metiltransferasa bajo el promotor *secA* (Ec n.º 085) y de células de *E. coli* que expresaban la metiltransferasa bajo el promotor T5 (Ec n.º 089 y Ec n.º 090) como se describe en el ejemplo 1 y 1 µg de ADN plasmídico total transformado en células electrocompetentes de *B. licheniformis* ATCC 53926 como se describe en el ejemplo 2. Las eficiencias de transformación del ADN plasmídico de las cepas de *E. coli* indicadas se normalizaron con respecto al plásmido pUK56S aislado de la cepa Ec n.º 085 de *E. coli*, que se fijó en el 100 %. El ADN del plásmido pUK56S procedente de la cepa Ec n.º 082 de *E. coli* no recuperó ningún transformante. El ADN del plásmido pUK56S de las cepas de *E. coli* Ec n.º 089 y Ec n.º 090, en comparación con Ec n.º 085, dio lugar a una eficiencia de transformación relativa del 6 % y el 2 %, respectivamente (figura 5).

20 Estos resultados demuestran que, en comparación con promotores inducibles bien conocido para la expresión de genes en *E. coli*, tales como el promotor T5, el uso del promotor *secA* para la expresión heteróloga de una metiltransferasa sorprendentemente dio como resultado un nivel de expresión que no perjudicaba la viabilidad de las células y permitía continuar el cultivo de las células con un nivel de expresión de metiltransferasa beneficioso para la metilación del ADN y, posteriormente, la eficiencia de la transformación.

Ejemplo 5: Extracción de promotores de especies de *Bacillus* y construcción de consenso

30 Se realizó una búsqueda blast traducida utilizando tblastn 2.5.0+ (Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., y Madden T.L. (2008), "BLAST+: architecture and applications", BMC Bioinformatics, 10:421) utilizando la secuencia de la proteína *SecA* procedente de *Bacillus licheniformis* (ID de la proteína WP_071583862) como consulta en las bases de datos Genbank y Genbank WGS (Whole Genome Shotgun), con las opciones: -evalue 1e-20, -db_gencode 11, -max_target_seqs 60000. Se recuperaron los registros completos de GenBank para los resultados de BLAST superiores al 60 % de identidad de proteína mínima.

Utilizando la información de localización de éxitos de BLAST procedente de los resultados de la búsqueda BLAST, se extrajeron las secuencias cadena arriba de los genes que codifican *secA*, en las siguientes condiciones:

40 1. El tamaño de la extracción cadena arriba fue de 160 nucleótidos. Si había una anotación de gen/CDS cadena arriba a menos de 160 nucleótidos, se extrajo un fragmento más corto. Si la longitud del fragmento era inferior a 100 nucleótidos, no se extrajo dicho fragmento.

2. Las secuencias cadena arriba extraídas se agruparon por puntuación ("bitscore") de éxitos de BLAST y se clasificaron en orden descendente por la misma puntuación. Para evitar sesgos, se desduplicaron las secuencias cadena arriba idénticas del mismo grupo de puntuación.

45 3. Para cada uno de los grupos de secuencias cadena arriba por puntuación, se realizó (y guardó por separado) un alineamiento múltiple acumulado utilizando MAFFT versión 7.307 (Kato, Standley, "MAFFT multiple sequence alignment software version 7: improvements in performance and usability", Molecular Biology and Evolution, 30:772-780, 2013), con la opción keeplength. Los alineamientos múltiples de nucleótidos generados se visualizaron como logotipos de secuencias y se examinaron para identificar el umbral de puntuación en el que la conservación de las secuencias reguladoras cadena arriba es más

evidente: los fragmentos conservados siguen teniendo un alto contenido de información, mientras que los fragmentos no conservados tienen un bajo contenido de información.

5 4. Basándose en el umbral identificado, todas las secuencias cadena arriba con puntuación por encima del umbral se sometieron a un alineamiento múltiple utilizando MAFFT. El alineamiento se refinó de modo manual y se construyó una secuencia de consenso a partir del alineamiento.

10 5. Para las especies de *Bacillus*, se identificaron importantes componentes del promotor con la ayuda de "Compilation and analysis of *Bacillus subtilis* sigma-A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA", John D. Helmann, Nucleic Acids Research, 1995, vol. 23, n.º 13, 2351-2360 y "Temporal Expression of the *Bacillus subtilis* secA Gene, Encoding a Central Component of the Preprotein Translocase", Markus Herbort, Michael Klein, Erik H. Manting, Arnold J. M. Driessen, Roland Freudl, Journal of Bacteriology, enero de 1999, vol. 181, n.º 2, págs. 493-500.

Extracción de promotores de especies de Enterobacteriaceae y construcción de consensos

Para *Enterobacteriaceae*, el proceso fue el mismo que para *Bacillus* (véase anteriormente), con las siguientes diferencias:

15 1. En *Escherichia coli*, el gen secA forma parte del operón secM-secA, con un promotor común cadena arriba de secM (proteína de control de la secreción). Así, la proteína SecM SECM_ECOLI se utilizó como consulta para la búsqueda BLAST traducida.

2. El tamaño de la extracción cadena arriba se fijó en 350 y la longitud cadena arriba mínima extraída en 50.

20 3. Para las especies de *Enterobacteriaceae*, se identificaron importantes componentes del promotor con la ayuda de "Promoter element spacing controls basal expression and light inducibility of the cyanobacterial secA gene", K. Mazouni, S. Bulteau, C. Cassier-Chauvat, F. Chauvat, Molecular Microbiology, 1998, vol. 30, n.º 5, págs. 1113-1122.

REIVINDICACIONES

- 1.- Una construcción de ácido nucleico recombinante que comprende un polinucleótido unido operativamente a una o más secuencias de control que dirigen la expresión del polinucleótido en una célula hospedante, en la que al menos una secuencia de control comprende una secuencia promotora *secA* y en la que dicha secuencia promotora es heteróloga al polinucleótido, en la que el polinucleótido no comprende un gen indicador; y en la que la secuencia promotora tiene al menos un 70 % de identidad de secuencia con SEQ ID NO:9.
- 5
- 2.- La construcción de ácido nucleico de la reivindicación 1, en la que la secuencia promotora procede de una especie de *Bacillus*.
- 3.- La construcción de ácido nucleico de cualquiera de las reivindicaciones anteriores, en la que dichas una o más secuencias de control comprenden una secuencia 5'UTR que comprende un sitio de unión al ribosoma.
- 10
- 4.- La construcción de ácido nucleico de la reivindicación 1, en la que la secuencia promotora es de un operón que comprende un gen *secA* procedente de un *Bacillus licheniformis*.
- 5.- La construcción de ácido nucleico de cualquiera de las reivindicaciones anteriores, en la que el polinucleótido codifica una proteína de interés.
- 15
- 6.- La construcción de ácido nucleico de la reivindicación 5, en la que la proteína de interés se selecciona del grupo que consiste en una metiltransferasa, una endonucleasa, una serina recombinasa o una tirosina recombinasa, y una proteína que confiere resistencia a antibióticos.
- 7.- La construcción de ácido nucleico de cualquiera de las reivindicaciones anteriores, en la que la célula hospedante es una célula de *Bacillus* o de *Escherichia*.
- 20
- 8.- Un vector de expresión recombinante que comprende la construcción de ácido nucleico de cualquiera de las reivindicaciones anteriores.
- 9.- Una célula hospedante recombinante que comprende la construcción de ácido nucleico de cualquiera de las reivindicaciones anteriores.
- 10.- Un procedimiento de expresión de un polinucleótido, que comprende las etapas de_
- 25
- (a) proporcionar una célula hospedante que comprende una construcción de ácido nucleico heteróloga de cualquiera de las reivindicaciones anteriores mediante la introducción de la construcción de ácido nucleico en la célula hospedante;
- (b) cultivar la célula hospedante recombinante de la etapa (a) en condiciones que conducen a la expresión del polinucleótido; y
- 30
- (c) opcionalmente, recuperar una proteína de interés codificada por el polinucleótido.
- 11.- El uso de una construcción de ácido nucleico de cualquiera de las reivindicaciones 1 a 7 para proporcionar un nivel de expresión moderado de un polinucleótido, preferentemente un polinucleótido que codifica una proteína de interés, en una célula hospedante.

Figura 1

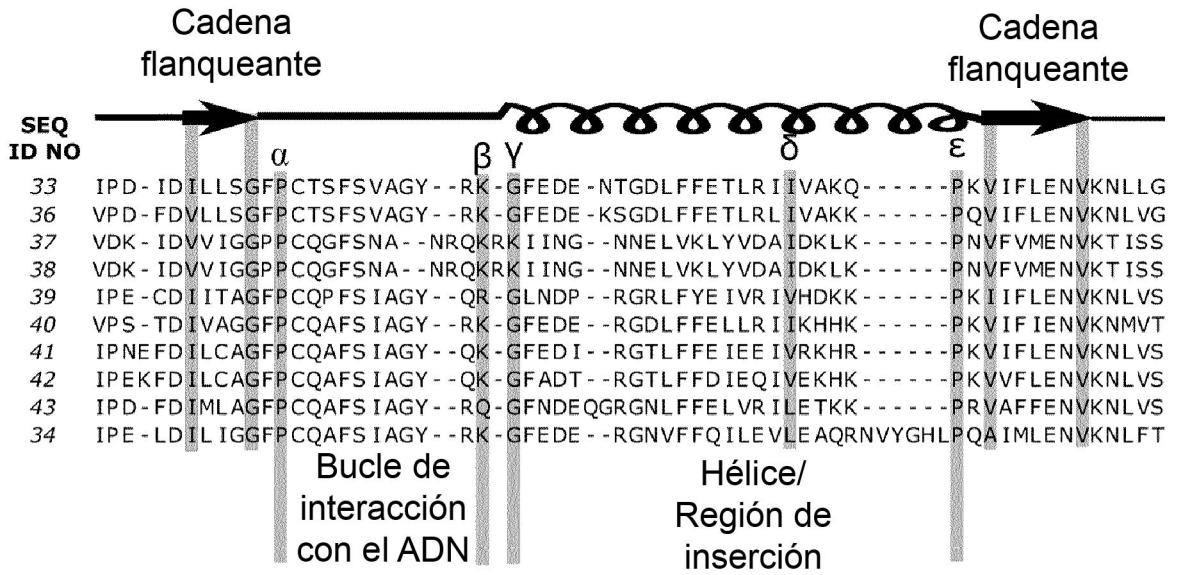


Figura 2

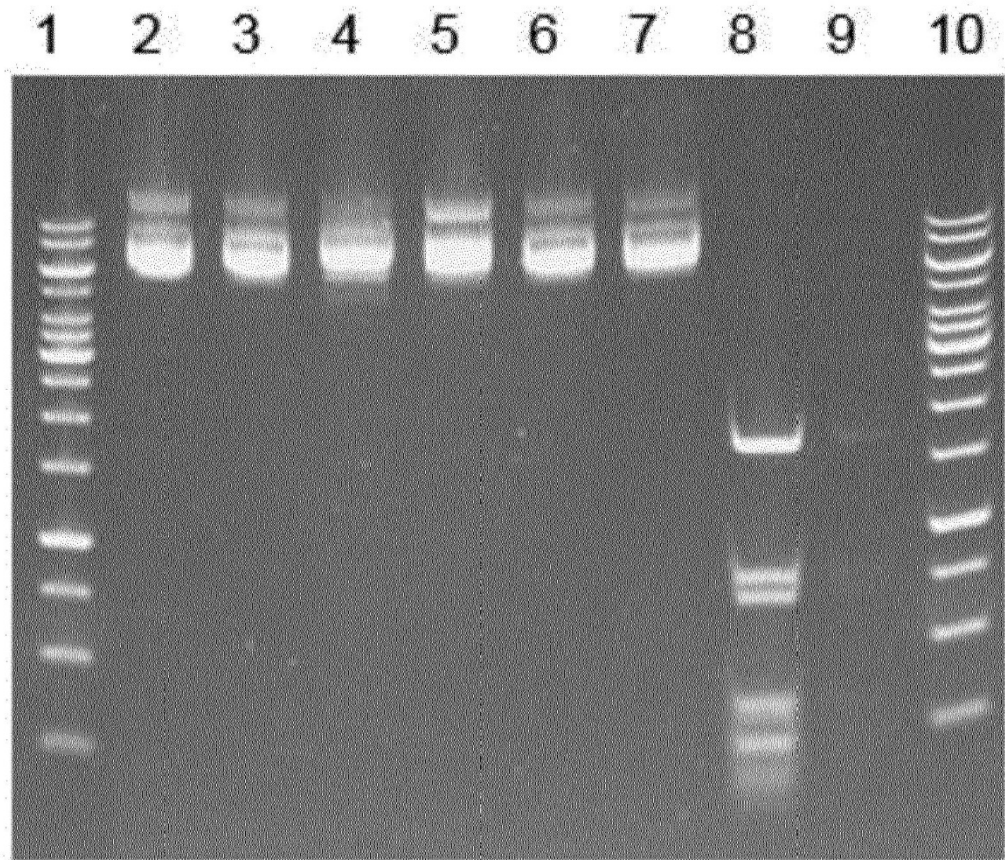


Figura 3

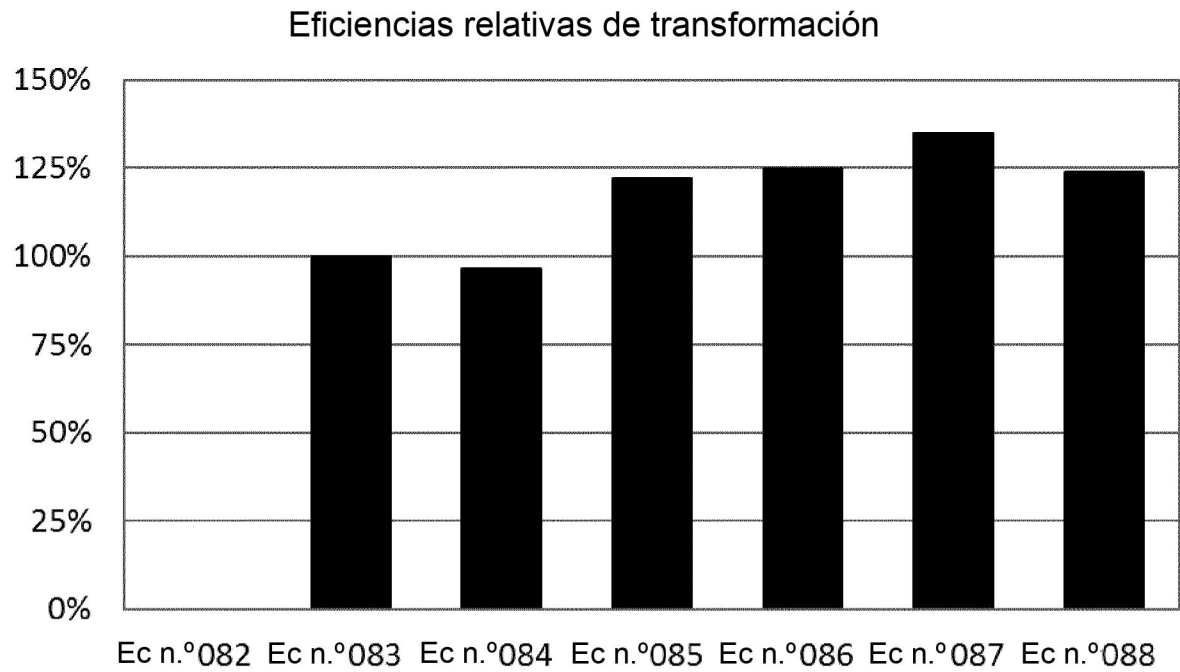


Figura 4

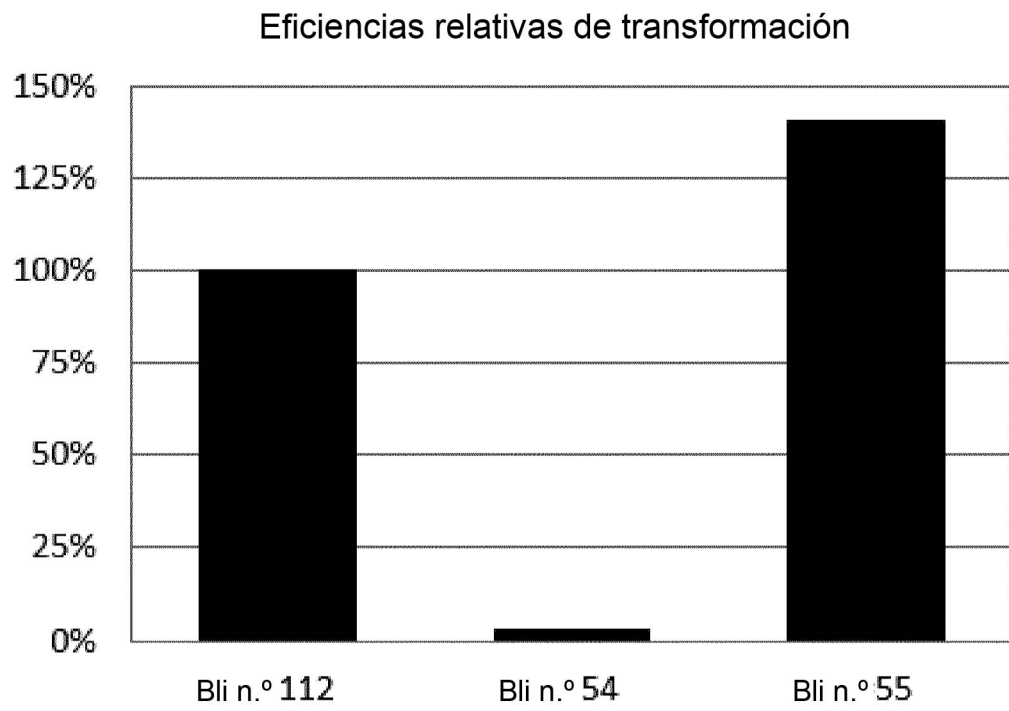


Figura 5

