

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 895 625**

51 Int. Cl.:

C12Q 1/6869 (2008.01)

C07K 14/35 (2006.01)

G01N 33/487 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **16.04.2015 PCT/US2015/026239**

87 Fecha y número de publicación internacional: **22.10.2015 WO15161117**

96 Fecha de presentación y número de la solicitud europea: **16.04.2015 E 15779198 (9)**

97 Fecha y número de publicación de la concesión europea: **01.09.2021 EP 3131920**

54 Título: **Nanoporos de MSP y usos de los mismos**

30 Prioridad:

16.04.2014 US 201461980415 P

16.04.2014 US 201461980393 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
22.02.2022

73 Titular/es:

**THE UAB RESEARCH FOUNDATION (100.0%)
1720 2nd Avenue South, Administration Building
770
Birmingham, AL 35294-0107, US**

72 Inventor/es:

**NIEDERWEIS, MICHAEL y
PAVLENOK, MIKHAIL**

74 Agente/Representante:

CONTRERAS PÉREZ, Yahel

ES 2 895 625 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

Nanoporos de MSP y usos de los mismos

5 DECLARACIÓN DE DERECHOS GUBERNAMENTALES

La presente invención se realizó con soporte del gobierno con la subvención n. R01HG005115, otorgada por los National Institutes of Health. El gobierno posee ciertos derechos sobre la presente invención.

10 Antecedentes

La identificación y caracterización de analitos con frecuencia implica una metodología costosa y que requiere mucho tiempo. Por ejemplo, las tecnologías actuales requieren reactivos hechos a medida y/o métodos de detección costosos para determinar secuencias de ácidos nucleicos. De forma análoga, los métodos para determinar secuencias de proteínas, por ejemplo, huellas de péptidos mediante tecnologías de espectrometría de masas, pueden resultar laboriosos y costosos. El documento WO 2012/107778 divulga una larga lista de formas mutantes de porinas de *Mycobacterium smegmatis* (Msp, por sus siglas en inglés). En una de las porinas Msp mutantes, la serina ha reemplazado a la prolina en la posición 97. El documento WO 2011/106545 A2 divulga una larga lista de formas mutantes de porinas de *Mycobacterium smegmatis* (Msp). En una de las porinas Msp mutantes, la cisteína ha reemplaza la serina en la posición 73. En otra mutante, la cisteína ha reemplazado a la prolina en la posición 97. Estas formas mutantes no se han caracterizado adicionalmente. Por lo tanto, son necesarios métodos eficaces de detección y análisis de ácidos nucleicos, proteínas y otros analitos.

Sumario

En el presente documento se proporcionan secuencias de los ácidos nucleicos que codifican una porina de *Mycobacterium smegmatis* (Msp) de cadena sencilla mutante como se define en las reivindicaciones. La secuencia de los ácidos nucleicos comprende opcionalmente una primera y una segunda secuencia de nucleótidos, en donde la primera secuencia de nucleótidos codifica una primera secuencia de monómero de Msp y la segunda secuencia de nucleótidos codifica una segunda secuencia de monómero de Msp; y una tercera secuencia de nucleótidos que codifica una secuencia engarce de aminoácidos. En las Msp de cadena sencilla mutantes proporcionadas en el presente documento, al menos una de las secuencias de monómeros de Msp primera y segunda es una secuencia de monómero de Msp mutante que comprende una mutación P97F.

Opcionalmente, una secuencia de un ácido nucleico que codifica la Msp de cadena sencilla mutante puede comprender una primera, segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencias de nucleótidos o cualquier subconjunto de las mismas y la primera, segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencias de nucleótidos codifican una primera, segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencias de monómeros de Msp, respectivamente. La secuencia del ácido nucleico comprende además una novena secuencia de nucleótidos que codifica una secuencia engarce de aminoácidos que está opcionalmente presente en múltiples sitios (por ejemplo, entre las secuencias de monómeros de Msp). Al menos una de las secuencias de monómeros de Msp primera y segunda es la secuencia de monómero de Msp mutante que comprende una mutación P97F.

Se proporciona además en el presente documento un ácido nucleico que codifica un monómero de Msp mutante, en donde el monómero de Msp comprende una mutación P97F.

Se proporcionan además polipéptidos que comprenden las Msp mutantes descritas en el presente documento y polipéptidos codificados por los ácidos nucleicos descritos en el presente documento. También se proporciona en el presente documento un sistema que comprende una Msp mutante descrita en el presente documento, en donde la Msp tiene un vestíbulo y una zona de constricción que definen un túnel, en donde el túnel se coloca entre un primer medio líquido conductor y un segundo medio líquido conductor, en donde al menos un medio líquido conductor comprende un analito y en donde el sistema es operativo para detectar el analito, cuando el sistema se somete a un campo eléctrico suficiente para translocar el analito de un medio líquido conductor al otro.

Se proporcionan además métodos para detectar la presencia de un analito. Los métodos incluyen aplicar un campo eléctrico suficiente para translocar un analito de un primer medio conductor a un segundo medio conductor en comunicación de líquidos a través de una Msp mutante descrita en el presente documento y medir una corriente de iones, en donde una reducción en la corriente de iones indica la presencia del analito en el primer medio.

60 Descripción de los dibujos

Las figuras relacionadas con la materia objeto que no se incluyen en las reivindicaciones son de referencia.

La figura 1 muestra una alineación de los monómeros de MspA, MspB, MspC y MspD de *Mycobacterium smegmatis*. La numeración de cada proteína comienza con el primer aminoácido de la porción madura de la

secuencia. Las secuencias de monómeros de MspA, MspB, MspC y MspD sin una secuencia señal/líder se proporcionan como la SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 y SEQ ID NO: 4, respectivamente.

La figura 2 muestra representaciones esquemáticas de los plásmidos utilizados para la construcción de ML712 *M. smegmatis*. pML1611 - *mshB* es un vector de eliminación: *mshBup*, *mshBdown* - regiones homólogas cadena arriba y cadena abajo del gen cromosómico *mshB*; *loxP*-sitio de recombinación para Cre recombinasa, *hyg* resistencia a higromicina; *sacB*-levansacarasa; *xylE* - catecol-2,3-dioxigenasa; *gfp2+* - proteína verde fluorescente; *tsPAL5000*: origen de replicación sensible a la temperatura de micobacterias; *ColEI*- origen de replicación de *E. coli*. pML2622 - plásmido integrador para la expresión del dominio formador de canales N-terminal (d1) de *cpxT*. *hyg*: gen de resistencia a higromicina; *pBR322-Ori*: origen de replicación de *E. coli*; *pNIT1* y *pNIT2* son necesarios para la expresión inducible por nitrilo de *cpxT*, *attP*: sitio de unión cromosómica del fago L5; *int*: integrasa L5; *FRT*: sitio de recombinación Fip.

La figura 3 es una transferencia Southern de locus *msh* en cepas de *M. smegmatis*. Se aisló ADN cromosómico de las cepas SMR5 (1), ML16 (2) y ML712 (3) de *M. smegmatis*. Los fragmentos de ADN se separaron en gel de agarosa al 1 %, se transfirieron a una membrana de nitrocelulosa y se detectaron utilizando sondas específicas marcadas con digoxigenina. Los tamaños de los fragmentos de ADN que se hibridaron con la sonda fueron coherentes con los valores predichos. Tamaños de fragmentos esperados: locus *mshA*: 1 - 1500 pb, 2,3 - 1250 pb; locus *mshB*: 1,2 - 1140 pb, 3 - 1850 pb; locus *mshC*: 1 - 1200 pb, 2,3 - 2100 pb; locus *mshD*: 1 - 1730 pb, 2,3 - 1090 pb.

La figura 4 es una transferencia Western de la expresión de porina Msp en mutantes con eliminación del gen de porina de *M. smegmatis*. Las porinas Msp se extrajeron con óxido de octilpolietileno al 0,5 % y se cargaron 10 µl en cada carril. Las proteínas se separaron en SDS-PAGE al 8 % y se detectaron mediante transferencia Western utilizando un anticuerpo policlonal contra MspA. Carriles: 1, SMR5 (ts; expresión de *mshA*); 2, ML16 (triple mutante con eliminación de porina; expresión de *mshB*); 3, ML712/pMS2 (cuádruple mutante con eliminación de porina; vector vacío); 4, ML712/pMN016 (plásmido de expresión + *mshA*); 5, ML712/pML904 (plásmido de expresión + M1 *mshA*).

La figura 5 es una representación esquemática del plásmido de expresión de micobacterias pML3213 que codifica M1₈-MspA de cadena sencilla. El plásmido comprende lo siguiente: *hyg*, gen de resistencia a higromicina; *ory myc*, un origen de replicación micobacteriano; *COLE1 ORI*, un origen de replicación de *E. coli*; *psmyc*, promotor constitutivo micobacteriano; *m1-1*, *m1-2*, etc., genes *m1 mshA* con un número de índice correspondiente a la posición en la construcción génica de cadena sencilla. La tétrada A está flanqueada por PacI al comienzo del primer *m1-mshA* y por MluI al final del cuarto *m1-mshA*; la tétrada B está flanqueada por EcoRV al comienzo del quinto *m1-mshA* y por HindIII al final del octavo *m1-mshA*. Los genes *m1-mshA* individuales dentro de las tétradas están flanqueados por KpnI, NsiI, NdeI, PstI, ScaI, NheI. Los enlazadores de conexión (GGGGS)₃ (SEQ ID NO: 5) están ubicados entre sitios de restricción, pero no se muestran.

La figura 6 es una transferencia de Western que muestra el análisis de la expresión de MspA de cadena sencilla en *M. smegmatis* ML714. Las porinas Msp se extrajeron con óxido de octilpolietileno al 0,5 % y se cargaron 10 µl en cada carril. Las proteínas se separaron en SDS-PAGE al 10 % seguido de transferencia a una membrana de PVDF y se sondaron con anticuerpos monoclonales contra αMspA. Los carriles están marcados de la siguiente manera: M, escalera de proteínas; 1, expresión de MspA de ts (SMR5); 2, M1₂-MspA; 3, M1₄-MspA; 4, M1₈-MspA (M1-MspA de cadena sencilla) La cantidad de MspA se determinó mediante análisis cuantitativo por imágenes.

La figura 7 muestra la estabilidad térmica de M1₈-MspA de cadena sencilla. Se extrajeron porinas M1 MspA de cadena sencilla con óxido de n-octilpolietileno al 0,5 %. Se incubaron 15 µl de la muestra en un tampón que contenía SDS al 2 % durante 15 min a las temperaturas indicadas encima de cada carril. Las proteínas se separaron en SDS PAGE al 8 % seguido de transferencia a una membrana de PVDF y se sondaron con anticuerpos policlonales contra αMspA. Los carriles están marcados de la siguiente manera: M, escalera de proteínas; TA, temperatura ambiente; 40, 50, etc. indican la temperatura de incubación. Se cargaron cantidades iguales de muestras de proteína en cada carril.

La figura 8A muestra los resultados de los experimentos de bicapa lipídica con M1₈-MspA. Se añadieron aproximadamente 70 ng de proteína a la cámara bicapa. Se registraron señales de corriente de M1₈-MspA en una bicapa lipídica de difitanoilfosfatidilcolina en KCl 1 M a un potencial de -10 mV. Esto dio como resultado el aumento escalonado de la corriente a través de la bicapa lipídica que indica la inserción de canales de M1₈-MspA hacia la membrana.

La figura 8B es un histograma de la conductancia de M1₈-MspA. Se analizaron 269 poros en cuatro membranas diferentes. M1₈-MspA de cadena sencilla mostró un pico de conductancia predominante a 1,1 nS.

La figura 9 muestra los resultados de los experimentos de activación por voltaje. M1₈-MspA se añadió al lado *cis* de una membrana de difitanoilfosfatidilcolina. Se aplicaron voltajes cada vez más positivos (*señales superiores*) y negativos (*señales inferiores*) a la membrana cuando se reconstituyeron ~ 220 canales en la membrana. La corriente de la membrana se registró a cada voltaje aplicado. El voltaje crítico al que los canales comenzaron a cerrarse. (*V_c*) se determinó como el voltaje al que la conductancia de la bicapa lipídica disminuyó después de su valor máximo inicial. El voltaje crítico *V_c* de M1₈-MspA fue de +90 mV. El canal M1₈-MspA se mantuvo estable en todos los voltajes negativos aplicados.

La figura 10A muestra un ejemplo de una rampa positiva creada en una Msp de cadena sencilla que comprende un primer monómero de Msp mutante que comprende mutaciones en las posiciones D56, I68, S73, D118, D134 y E139 y un séptimo monómero mutante de Msp que comprende una mutación en las posiciones L88 y I105.

La figura 10B muestra una vista superior de una rampa positiva creada en una Msp de cadena sencilla que

comprende un primer monómero de Msp mutante que comprende mutaciones en las posiciones D56, I68, S73, D118, D134 y E139 y un séptimo monómero mutante de Msp que comprende una mutación en las posiciones L88 y 1105.

La figura 10C es un esquema de una Msp de cadena sencilla. Los números bajo las subunidades n.º 1 y n.º 7 representan ubicaciones de la rampa positiva.

La figura 11 es una representación esquemática del plásmido pML3216, para la expresión de *m2-mspA* de cadena sencilla en *E. coli*. *bla*, gen de resistencia a ampicilina; *pBR322-Ori*, origen de replicación de *E. coli*; *lacI*, proteína represora lac; *T7P*, promotor de T7; *SD*, secuencia Shine-Dalgarno; *m2-1 - m2-8*, codón de *m2 mspA*.

La figura 12 muestra que *scm2-mspA* se puede producir en *E. coli*. Las células *E. coli* Omp8 se indujeron con IPTG 1,5 mM a una DO₆₀₀ de 0,5. En diferentes puntos temporales, se recogieron y lisaron las células. Se cargaron cantidades iguales de muestra de proteína en gel de poliacrilamida al 10 %. Tras la electroforesis, el gel se tiñó con Coomassie y se exploró para cuantificación mediante el programa informático LabWorks (Waltham, Massachusetts). Carriles: *M*, marcador de peso molecular con masas indicadas a la izquierda (kDa); *Un*, células no inducidas; 2, 4 y 6 horas después de la inducción con IPTG.

La figura 13 muestra una transferencia Western del replegamiento de scMspA M2. Las muestras antes y después del replegamiento se cargaron en gel de poliacrilamida al 8 % seguido de transferencia durante la noche a una membrana de PVDF. La membrana se tiñó con anticuerpos de conejo específicos de MspA. Carriles: *M*, marcador de peso molecular con masas indicadas a la izquierda (kDa); *Un*, muestra sin plegar después de la cromatografía de intercambio aniónico; *F*, scMspA M2 plegada.

Las figuras 14 (A-B) muestran las propiedades de formación de canales de scMspA M2. (A) Registro de señales de corriente de scMspA M2 en bicapa lipídica plana. Después de la adición de proteína scMspA M2 replegada como se muestra en la figura 3, se observa un aumento de corriente escalonado indicativo de inserciones de canales. (B) Histograma de distribución de conductancia de un solo canal. Se analizaron un total de 392 canales de 8 membranas. La línea de puntos representa un ajuste gaussiano de la distribución de datos. La conductancia principal de scMspA M2 es 2,3 nS.

La figura 15 muestra el cierre de canal dependiente del voltaje de scMspA M2. Se añadió proteína al lado *cis* de la cubeta. Después de la inserción de aproximadamente 200 poros, el voltaje positivo o negativo se incrementó en incrementos de 10 mV y la corriente se registró durante 2 minutos. Una línea plana representa los poros abiertos, mientras que una línea inclinada representa el cierre de los poros. Se determinó que los voltajes críticos para scMspA M2 eran aproximadamente +80 mV/-70 mV.

La figura 16 muestra una representación esquemática de plásmido pML3222, para la expresión de *scmspA PN1* (*scmspA PN1*) de cadena sencilla en *E. coli*. *bla*, gen de resistencia a ampicilina; *pBR322-Ori*, origen de replicación de *E. coli*; *lacI*, proteína represora lac; *T7 P*, promotor de T7; *SD*, secuencia Shine-Dalgarno; *m2-97-1 - m2-97-8*, genes con codones optimizados de *m2 mspA* con una mutación P97F.

La figura 17 muestra que *scmspA PN1* se puede producir en *E. coli*. Las células *E. coli* Omp8 se indujeron con IPTG 1,5 mM a una DO₆₀₀ de 0,5. A las 2 horas después de la inducción, se recogieron y lisaron las células. Se cargaron cantidades iguales de muestra de proteína en gel de poliacrilamida al 8 %. Tras la electroforesis, el gel se tiñó con Coomassie y se exploró para cuantificación mediante el programa informático LabWorks (Waltham, Massachusetts). Carriles: *M*, marcador de peso molecular con masas indicadas a la izquierda (kDa); *Un*, células no inducidas; 2 horas después de la inducción con IPTG.

La figura 18 muestra una transferencia Western del replegamiento de scMspA PN1. Las muestras se cargaron en gel de poliacrilamida al 8 % seguido de tinción con Coomassie y se exploraron para cuantificación mediante el programa informático LabWorks. Carriles: *M*, marcador de peso molecular con masas indicadas a la izquierda (kDa); *IB*, cuerpos de inclusión purificados de *E. coli* Omp8; *AE*, muestra después de cromatografía de intercambio aniónico; *D*, muestra después de diálisis; *R*, proteína scMspA PN1 plegada.

La figura 19 muestra que scMspA PN1 se inserta en membranas lipídicas. Los liposomas de DPhPC se cargaron con colorante de carboxifluoresceína 30 mM. La liberación de colorante se monitorizó a una emisión de 517 nm (excitación de 492 nm) utilizando fluorímetro Varian Cary (Palo Alto, CA). Símbolos: cuadrados abiertos - Triton X-100 (0,1 % v/v); círculos cerrados - LDAO (0,1 % v/v); círculos abiertos - OPOE (0,5 % v/v); triángulos cerrados - wtMspA (60 ng/ml); cuadrados cerrados - scMspA PN1 (60 ng/ml); triángulos abiertos - scMspA M2 (120 ng/ml). No se observó una liberación de colorante significativa cuando los liposomas estaban solo en tampón PBS.

La figura 20 muestra la distribución del tiempo de reconstitución de los poros en la membrana de DPhPC artificial. Se llenó una cubeta con KCl en las concentraciones indicadas. Se añadieron muestras de proteína de 100 ng/ml final a ambos lados. Se aplicó un potencial de -10mV y los datos se registraron con el programa informático TestPoint. Cada punto representa el tiempo de la primera inserción de los poros. El tiempo medio de inserción para scMspA PN1 en KCl 1 M fue de 399 segundos (9 membranas con 89 % de eventos de inserción exitosos analizados). El tiempo medio de inserción para scMspA PN1 en KCl 0,3 M fue de 859 segundos (8 membranas con 50 % de eventos de inserción exitosos analizados). El tiempo medio de inserción para scMspA M2 en KCl 1 M fue de 695 segundos (10 membranas con 40 % de eventos de inserción exitosos analizados). En KCl 0,3 M, scMspA M2 tuvo un tiempo de inserción de 1270 segundos (8 membranas, 12 % de inserción exitosa). * - P = 0,028 según lo determinado por la prueba de suma de rangos de Mann-Whitney.

Las figuras 21 (A-D) muestran las conductancias de un solo canal de scMspA PN1 y scMspA M2 en KCl 1,0 M. (A) Registro de señales de corriente de scMspA PN1 en bicapa lipídica plana. Después de la adición de la proteína scMspA PN1 replegada, se observa un aumento de corriente escalonado indicativo de inserciones de canales. (B) Histograma de distribución de conductancia de un solo canal. Se analizaron un total de 137 canales de 4

membranas. La línea de puntos representa un ajuste gaussiano de la distribución de datos. La conductancia principal de scMspA PN1 es 2,0 nS. (C) Registro de señales de corriente de scMspA M2 en bicapa lipídica plana. (D) Histograma de distribución de conductancia de un solo canal. Se analizaron un total de 238 canales de 6 membranas. La línea de puntos representa un ajuste gaussiano de la distribución de datos. La conductancia principal de scMspA M2 es 1,3 nS.

La figura 22 muestra la conductancia de un solo canal de MspA PN1 de cadena sencilla a KCl 0,3 M y KCl 1,0 M y el aumento de las inserciones de canales de scMspA PN1 en membranas a KCl 0,3 M después de estar en contacto con las membranas.

10 Descripción detallada

En el presente documento se proporcionan porinas de *Mycobacterium smegmatis* (Msp) mutantes como se definen en las reivindicaciones. Una Msp mutante puede ser un complejo multimérico compuesto por dos o más monómeros de Msp, en donde al menos uno de los monómeros es un monómero de Msp mutante como se define en las reivindicaciones. Un monómero de Msp está codificado por un gen en *Mycobacterium smegmatis*. *Mycobacterium smegmatis* tiene cuatro genes Msp identificados, denominados MspA, MspB, MspC y MspD. En la figura 1 se muestra una alineación de las secuencias polipeptídicas de tipo silvestre para los monómeros de MspA, MspB, MspC y MspD de *Mycobacterium smegmatis*. La numeración de cada proteína comienza con el primer aminoácido de la porción madura de la secuencia, como lo indica el número "1" encima del primer aminoácido de la secuencia de aminoácidos madura. Las secuencias de aminoácidos de un monómero de MspA, MspB, MspC y MspD sin secuencia señal, es decir, la parte madura de la secuencia, se proporcionan como la SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 y SEQ ID NO: 4, respectivamente. Las secuencias de aminoácidos de un monómero MspA, MspB, MspC y MspD con una secuencia señal/líder se proporcionan como la SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8 y SEQ ID NO: 9, respectivamente.

Además, las secuencias de monómeros de Msp de tipo silvestre que pueden modificarse se divulgan en GenBank, y estas secuencias y otras se incorporan en el presente documento como referencia en su totalidad, como subsecuencias individuales o fragmentos contenidos en las mismas. Por ejemplo, las secuencias de nucleótidos y aminoácidos de un monómero de MspA de tipo silvestre se pueden encontrar en los números de registro de GenBank AJ001442 y CAB56052, respectivamente. Se pueden encontrar las secuencias de nucleótidos y aminoácidos de un monómero de MspB de tipo silvestre, por ejemplo, en los números de registro de GenBank NC_008596.1 (del nucleótido 600086 al 600730) e YP_884932.1, respectivamente. Se pueden encontrar las secuencias de nucleótidos y aminoácidos de un monómero de MspC de tipo silvestre, por ejemplo, en los números de registro de GenBank AJ299735 y CAC82509, respectivamente. Se pueden encontrar las secuencias de nucleótidos y aminoácidos de un monómero de MspD de tipo silvestre, por ejemplo, en los números de registro de GenBank AJ300774 y CAC83628, respectivamente.

Un monómero de Msp mutante puede ser un monómero de longitud completa o un fragmento funcional del mismo codificado por un ácido nucleico que codifica MspA, MspB, MspC o MspD, por ejemplo, un ARNm o una secuencia genómica que codifica MspA, MspB, MspC o MspD, en donde el monómero comprende una o más modificaciones.

Opcionalmente, una Msp mutante es una Msp de cadena sencilla mutante o es un multímero de varias Msp de cadena sencilla, en donde el multímero comprende al menos una Msp de cadena sencilla mutante. Una Msp mutante también puede ser un multímero de varios monómeros de Msp en donde al menos un monómero de Msp es un monómero de Msp mutante.

Una Msp de cadena sencilla puede comprender, por ejemplo, un multímero formado por dos o más monómeros de Msp (por ejemplo, ocho monómeros) conectados por uno o más péptidos enlazadores de aminoácidos. Una Msp de cadena sencilla parcial se refiere a un complejo multimérico de cadena sencilla que se dimeriza, trimeriza, o similar para formar una porina. Una porina Msp de cadena sencilla completa se refiere a un complejo multimérico de cadena sencilla que forma una porina sin necesidad de dimerizar, trimerizar o similares para formar una porina. Dicho de manera diferente, los pliegues de cadena sencilla para formar una porina, pero todos los componentes, están en una cadena de aminoácidos, en comparación con una porina que debe asociarse con otros monómeros de Msp de cadena sencilla parcial o monómeros de Msp monoméricos para formar una porina.

Msp de cadena sencilla mutante y los ácidos nucleicos que las codifican

En el presente documento se proporcionan secuencias de los ácidos nucleicos que codifican Msp de cadena sencilla mutantes como se define en las reivindicaciones. Por ejemplo, la secuencia de un ácido nucleico que codifica una Msp de cadena sencilla mutante comprende: (a) una primera y segunda secuencias de nucleótidos, en donde la primera secuencia de nucleótidos codifica una primera secuencia de monómero de Msp y la segunda secuencia de nucleótidos codifica una segunda secuencia de monómero de Msp; y (b) una tercera secuencia de nucleótidos que codifica una secuencia engarce de aminoácidos, en donde al menos una de las secuencias de monómeros de Msp primera y segunda es una secuencia de monómero de Msp mutante que comprende una mutación P97F. Como se muestra en los ejemplos, restos hidrófobos adicionales, por ejemplo, fenilalanina, ubicados en el bucle 6 de scMspA (aminoácidos

91-103) promueven una inserción más rápida y eficaz de los poros en las bicapas lipídicas. Para obtener una descripción del bucle 6 de MspA y los restos que contiene, véase Hufte *et al.*, J. Biol. Chem. 284: 10223-10231 (2009). Por lo tanto, se proporciona en el presente documento una Msp de cadena sencilla que comprende una o más sustituciones hidrófobas en el bucle 6 (aminoácidos 91-103) de Msp.

5

Por ejemplo, se proporciona en el presente documento una secuencia de un ácido nucleico que codifica una porina de *Mycobacterium smegmatis* (Msp) de cadena sencilla mutante, en donde la secuencia del ácido nucleico comprende (a) una primera y una segunda secuencias de nucleótidos, en donde la primera secuencia de nucleótidos codifica una primera secuencia de monómero de Msp y la segunda secuencia de nucleótidos codifica una segunda secuencia de monómero de Msp y (b) una tercera secuencia de nucleótidos que codifica una secuencia engarce de aminoácidos, en donde al menos una de las secuencias de monómeros de Msp primera y segunda es una secuencia de monómero de Msp mutante que comprende una mutación P97F.

Como se usa a lo largo de la presente memoria descriptiva, una mutación en un aminoácido específico se indica mediante el código de una sola letra para el aminoácido en una posición, seguido del número de la posición del aminoácido en una secuencia polipeptídica de Msp (por ejemplo, una posición del aminoácido en la SEQ ID NO: 1) y el código de una sola letra para la sustitución del aminoácido en esta posición. Por lo tanto, se entiende que una mutación P97F es una sustitución de prolina por fenilalanina en el aminoácido 97 de la SEQ ID NO: 1. De forma análoga, una mutación D90N es una sustitución de ácido aspártico por arginina en el aminoácido 90 de la SEQ ID NO: 1, una mutación D91N es una sustitución de aspártico por arginina en el aminoácido 91 de la SEQ ID NO: 1, etc. También se entiende que los aminoácidos correspondientes a las posiciones en la SEQ ID NO: 1 también se proporcionan en el presente documento (véase la figura 1). Por ejemplo, y sin ser limitante, un experto en la materia entenderá que, el aminoácido correspondiente para E139 de la SEQ ID NO: 1 en MspB (SEQ ID NO: 2), MspC (SEQ ID NO: 3) y MspD (SEQ ID NO: 4) es A139, A139 y K138, respectivamente.

25

Opcionalmente, la secuencia de monómero de Msp mutante como se reivindica puede comprender además una mutación en la posición de aminoácido D118, una mutación en la posición D134 o una mutación en la posición E139. Opcionalmente, una mutación en la posición E139 puede ser una sustitución de E por R (arginina) o de E por K (lisina). Opcionalmente, una mutación en la posición D118 puede ser una sustitución de D por R o una sustitución de D por K. Opcionalmente, una mutación en la posición D134 puede ser una sustitución de D por R o una sustitución de D por K. Por ejemplo, la secuencia de monómero de Msp mutante como se reivindica puede comprender una o más mutaciones seleccionadas del grupo que consiste en: una mutación D118R, una mutación D134R y una mutación E139K. Opcionalmente, la secuencia de monómero de Msp mutante como se reivindica puede comprender además al menos una de (i) una mutación en la posición 93 y (ii) una mutación en la posición D90, posición D91 o ambas posiciones D90 y D91. Opcionalmente, el aminoácido en la posición 90, 91 o 93 se sustituye con arginina, lisina, histidina, glutamina, metionina, treonina, fenilalanina, tirosina o triptófano. Opcionalmente, la secuencia de monómero de Msp mutante como se reivindica puede comprender además una mutación D90N, una D91N y una D93N.

Por ejemplo, la secuencia de monómero de Msp mutante que comprende una mutación P97F puede comprender además (i) una mutación en la posición de aminoácido D118, D134 y/o E139 (ii) una mutación en la posición D93, y/o (iii) una mutación en la posición D90, posición D91 o ambas posiciones D90 y D91. Por ejemplo, la secuencia de monómero de MspA mutante como se reivindica puede comprender una mutación D90N, una mutación D91N, una mutación D93N, una mutación P97F, una mutación D118R, una mutación D134R y una mutación E139K. La secuencia de monómero de MspA mutante como se reivindica también puede comprender una mutación D90N, una mutación D91N, una mutación D93N, una mutación P97F, una mutación D118R, una mutación D134R y una mutación E139K.

En las Msp mutantes de cadena sencilla proporcionadas en el presente documento, la primera secuencia de monómero puede ser cualquier secuencia de monómero mutante como se reivindica. El segundo monómero se puede seleccionar del grupo que consiste en un monómero de Msp de tipo silvestre, un segundo monómero de Msp mutante, un monómero de Msp de tipo silvestre parálogo u homólogo y un monómero de Msp mutante parálogo u homólogo. Se entiende que el segundo monómero de Msp mutante puede ser igual o diferente al primer monómero de Msp mutante. Estos incluyen, pero sin limitación, MspA/Msmeg0965, MspB/Msmeg0520, MspC/Msmeg5483, MspD/Msmeg6057, MppA, PorM1, PorM2, PorM1, Mmcs4296, Mmcs4297, Mmcs3857, Mmcs4382, Mmcs4383, Mjls3843, Mjls3857, Mjls3931 Mjls4674, Mjls4675, Mjls4677, Map3123c, Mav3943, Mvan1836, Mvan4117, Mvan4839, Mvan4840, Mvan5016, Mvan5017, Mvan5768, MUL_2391, Mflv1734, Mflv1735, Mflv2295, Mflv1891, MCH4691c, MCH4689c, MCH4690c, MAB1080, MAB1081, MAB2800, RHA1 ro08561, RHA1 ro04074 y RHA1 ro03127. Un monómero de MspA de tipo silvestre parálogo u homólogo puede ser un monómero de MspB de tipo silvestre. Los monómeros de MspA de tipo silvestre parálogos u homólogos son bien conocidos en la técnica. La tabla 1 proporciona un listado no limitante de dichos parálogos y homólogos.

60

Tabla 1. Monómeros de MspA de tipo silvestre y parálogos u homólogos de MspA de tipo silvestre

N.º de proteína	Organismo	Identidad/similitud con MspA (%)	Longitud (aa)	Referencia
MspA/Msmeg0965	<i>M. smegmatis</i>	100/100	211	gb ABK74363.1 , (Stahl <i>et al.</i> , 2001)*

(continuación)

N.º de proteína	Organismo	Identidad/similitud con MspA (%)	Longitud (aa)	Referencia
MspB/Msmeg0520	<i>M. smegmatis</i>	94/95	215	gb ABK73437.1 , (Stahl <i>et al.</i> , 2001)*
MspC/Msmeg5483	<i>M. smegmatis</i>	93/95	215	gb ABK74976.1 , (Stahl <i>et al.</i> , 2001)*
MspD/Msmeg6057	<i>M. smegmatis</i>	82/89	207	gb ABK72453.1 , (Stahl <i>et al.</i> , 2001)*
MppA	<i>M. phlei</i>	100/100	211	AJ812030, (Dorner <i>et al.</i> , 2004)**
PorMI	<i>M. fortuitum</i>	95/96	211	emb CAI54228.1
PorM2	<i>M. fortuitum</i>	91/93	215	emb CAL29811.1
PorMI	<i>M. peregrinum</i>	94/96	211	emb CA154230.1
Mmcs4296	<i>Mycobacterium sp. MCS</i>	85/91	216	gb ABG10401.1
Mmcs4297	<i>Mycobacterium sp. MCS</i>	85/91	216	gb ABG10402.1
Mmcs3857	<i>Mycobacterium sp. MCS</i>	30/44	235	gb ABG09962.1
Mmcs4382	<i>Mycobacterium sp. MCS</i>	85/91	216	gb ABL93573.1
Mmcs4383	<i>Mycobacterium sp. MCS</i>	85/91	216	gb ABL93574.1
Mjls3843	<i>Mycobacterium sp. JLS</i>	26/40	235	gb ABN99619.1
Mjls3857	<i>Mycobacterium sp. JLS</i>	26/40	235	gb ABG09962.1
Mjls3931	<i>Mycobacterium sp. JLS</i>	26/40	235	gb ABL93123.1
Mjls4674	<i>Mycobacterium sp. JLS</i>	85/89	216	gb ABO00440.1
Mjls4675	<i>Mycobacterium sp. JLS</i>	83/89	216	gb ABO00441.1
Mjls4677	<i>Mycobacterium sp. JLS</i>	84/89	216	gb ABO00443.1
Map3123c	<i>M. avium paratuberculosis</i>	24/39	220	gb AAS05671.1
Mav3943	<i>M. avium</i>	24/39	227	gb ABK66660.1
Mvan1836	<i>M. vanbaalenii</i> PYR-1	82/88	209	gb ABM12657.1
Mvan4117	<i>M. vanbaalenii</i> PYR-1	32/43	239	gb ABM14894.1
Mvan4839	<i>M. vanbaalenii</i> PYR-1	83/88	209	gb ABM15612.1
Mvan4840	<i>M. vanbaalenii</i> PYR-1	83/89	209	gb ABM15613.1
Mvan5016	<i>M. vanbaalenii</i> PYR-1	30/41	238	gb ABM15788.1
Mvan5017	<i>M. vanbaalenii</i> PYR-1	25/35	227	gb ABM15789.1
Mvan5768	<i>M. vanbaalenii</i> PYR-1	21/32	216	gb ABM16533.1
MUL_2391	<i>M. ulcerans</i> Agy99	21/34	233	gb ABL04749.1
Mflv1734	<i>M. gilvum</i> PYR-GCK	21/32	225	gb ABP44214.1
Mflv1735	<i>M. gilvum</i> PYR-GCK	32/41	226	gb ABP44215.1
Mflv2295	<i>M. gilvum</i> PYR-GCK	25/40	250	gb ABP44773.1
Mflv1891	<i>M. gilvum</i> PYR-GCK	84/90	217	gb ABP44371.1
MCH4691c	<i>M. chelonae</i>	70/80	223	gb ACV04474.1
MCH4689c	<i>M. chelonae</i>	66/78	223	gb ACV04472.1
MCH4690c	<i>M. chelonae</i>	72/81	217	gb ACV04473.1
MAB1080	<i>M. abscessus</i>	69/79	223	emb CAM61170.1
MAB1081	<i>M. abscessus</i>	68/78	222	emb CAM61171.1
MAB2800	<i>M. abscessus</i>	27/44	246	emb CAM62879.1
RHA1 ro08561	<i>Rhodococcus jostii</i> RHA1	34/51	233	gb ABG99605.1
n.d.	<i>Rhodococcus opacus</i> B4	34/51	233	gb BAH52196.1
RHA1 ro04074	<i>Rhodococcus sp. RHA1</i>	34/50	233	gb ABG95871.1
RHA1 ro03127	<i>Rhodococcus sp. RHA1</i>	34/50	233	gb ABG94930.1
n.d.	<i>Rhodococcus erythropolis</i> PR4	35/50	229	gb BAH30938.1

Solo se incluyeron proteínas con similitudes significativas de aminoácidos en toda la longitud de la proteína. Los datos se obtuvieron mediante el algoritmo PSI-Blast (matriz BLOSUM62) utilizando la base de datos de NIH GenBank en el sitio web ncbi.nlm.nih.gov/blast/Blast.cgi.

n. d.: "no determinado"

*Stahl *et al.*, Mol. Microbial. 40:451 (2001)

**Dorner *et al.*, Biochim. Biophys. Acta. 1667:47-55 (2004)

Como se usa en el presente documento, una Msp de cadena sencilla mutante es un polipéptido que comprende al

menos dos monómeros de Msp, o fragmentos funcionales de los mismos, conectados mediante uno o más péptidos enlazadores de aminoácidos en donde al menos uno de los monómeros de Msp es un monómero de Msp mutante. Por ejemplo, la Msp de cadena sencilla mutante puede comprender dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, once, doce o más monómeros de Msp conectados mediante uno o más péptidos enlazadores de aminoácidos, en donde al menos uno de los monómeros de Msp es un monómero de Msp mutante. Como se ha expuesto anteriormente, una Msp de cadena sencilla mutante puede formar una porina, por ejemplo, mediante plegamiento, sin necesidad de dimerizar, trimerizar o similares para formar una porina. Como alternativa, una Msp de cadena sencilla mutante puede ser una Msp de cadena sencilla mutante parcial que comprende al menos dos monómeros de Msp o fragmentos de los mismos conectados por uno o más péptidos unidos a aminoácidos, que puede dimerizar, trimerizar o similares para formar una porina.

Opcionalmente, una porina Msp que comprende una Msp de cadena sencilla mutante puede comprender, por ejemplo, dos o más dímeros de porina Msp de cadena sencilla, dos o más trímeros de porina Msp de cadena sencilla, dos o más cuadrímeros de porina Msp de cadena sencilla, dos o más pentámeros de porina Msp de cadena sencilla, uno o más hexámeros de porina Msp de cadena sencilla, uno o más septámeros de porina Msp de cadena sencilla, uno o más octámeros de porina Msp de cadena sencilla o combinaciones de los mismos. Por ejemplo, una porina Msp puede comprender un dímero de porina Msp de cadena sencilla y dos trímeros de porina Msp de cadena sencilla. A modo de otro ejemplo, una porina Msp puede comprender un cuadrímero de porina Msp de cadena sencilla y dos dímeros de porina Msp de cadena sencilla.

Las secuencias engarce de aminoácidos se describen en el presente documento. En cualquier Msp de cadena sencilla descrita en el presente documento, una secuencia engarce puede comprender, por ejemplo, de 10 a 20 aminoácidos. Por ejemplo, una secuencia engarce de aminoácidos comprende 15 aminoácidos. Opcionalmente, la secuencia engarce de aminoácidos comprende una secuencia peptídica (GGGGS)₃ (SEQ ID NO: 5). Se puede proporcionar la misma o diferente secuencia engarce que codifica el ácido nucleico entre secuencias de los ácidos nucleicos que codifican más de dos monómeros de Msp. Opcionalmente, se puede proporcionar una secuencia engarce entre todas o algunas de las secuencias de los ácidos nucleicos que codifican los monómeros de Msp en las Msp de cadena sencilla proporcionadas en el presente documento.

Se proporciona además una secuencia del ácido nucleico que codifica una Msp de cadena sencilla mutante, en donde la secuencia del ácido nucleico comprende (a) una primera, segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencias de nucleótidos o cualquier subconjunto de las mismas, en donde la primera, segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencias de nucleótidos codifican una primera, segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencias de monómeros de Msp, respectivamente; y (b) una novena secuencia de nucleótidos que codifica una secuencia engarce de aminoácidos, en donde la primera secuencia de monómero de Msp es una secuencia de monómero de Msp mutante que comprende una mutación en la posición P97.

La secuencia de monómero de Msp mutante comprende una mutación P97F. Como se ha expuesto anteriormente, la secuencia de monómero de Msp mutante como se reivindica puede comprender además una mutación en la posición de aminoácido D118, una mutación en la posición D134 o una mutación en la posición E139. Por ejemplo, cualquier secuencia de monómero de Msp mutante descrita en el presente documento puede comprender una mutación D118R, una mutación D134R y/o una mutación E139K. La secuencia de monómero de Msp mutante como se reivindica puede comprender además (i) una mutación en la posición 93 y/o (ii) una mutación en la posición D90, posición D91 o ambas posiciones D90 y D91. Opcionalmente, el aminoácido en la posición 90, 91 o 93 se sustituye con arginina, lisina, histidina, glutamina, metionina, treonina, fenilalanina, tirosina o triptófano. La secuencia de monómero de Msp mutante como se reivindica puede comprender además una mutación D90N, una D91N y una D93N. Por ejemplo, en el presente documento se proporciona una secuencia de un ácido nucleico que codifica una Msp de cadena sencilla mutante, en donde la secuencia del ácido nucleico comprende (a) una primera, segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencias de nucleótidos o cualquier subconjunto de las mismas, en donde la primera, segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencias de nucleótidos codifican una primera, segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencias de monómeros de Msp, respectivamente; y (b) una novena secuencia de nucleótidos que codifica una secuencia engarce de aminoácidos, en donde la primera secuencia de monómero de Msp es una secuencia de monómero de Msp mutante que comprende una mutación P97F y puede comprender además (i) una mutación en la posición de aminoácido D118, D134 y/o E139 (ii) una mutación en la posición D93, y/o (iii) una mutación en la posición D90, posición D91 o ambas posiciones D90 y D91. Por ejemplo, la primera secuencia de monómero de Msp puede ser una secuencia de monómero de Msp mutante que comprende una mutación D90N, una mutación D91N, una mutación D93N, una mutación P97F, una mutación D118R, una mutación D134R y una mutación E139K.

Se entiende que la primera secuencia de nucleótidos y la séptima secuencia de nucleótidos se pueden disponer, aunque no están necesariamente dispuestas como la primera secuencia de nucleótidos y la séptima secuencia de nucleótidos en la secuencia de ácido nucleico que comprende una primera, segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencia de nucleótidos en ese orden. En este contexto, la primera secuencia de nucleótidos es una secuencia de nucleótidos que codifica el primer monómero o el monómero inicial de una Msp de cadena sencilla y puede ser la primera, segunda, tercera, cuarta, quinta, sexta, séptima u octava secuencia de nucleótidos de la Msp

de cadena sencilla. La secuencia de nucleótidos inicial se denomina primera secuencia de nucleótidos independientemente del lugar donde aparezca en la Msp de cadena sencilla.

Las figuras 10A y 10B muestran un ejemplo de referencia no limitante de una rampa positiva creada en una Msp de cadena sencilla que comprende un primer monómero de Msp mutante que comprende una mutación en las posiciones D56, I68, S73, D118, D134 y E139 y un séptimo monómero mutante de Msp que comprende una mutación en las posiciones L88 y 1105. Esta rampa cargada positivamente dentro del vestíbulo de la MspA guía los ácidos nucleicos monocatenarios, por ejemplo, ADN, a través de la Msp. Las interacciones electrostáticas entre el ácido nucleico y la rampa permiten la translocación controlada del ADN a través del poro. Esto reduce el movimiento browniano del ácido nucleico y la tasa de translocación. Esto también aumenta la precisión y la interacción entre las bases de los ácidos nucleicos y los aminoácidos en la zona de constricción. La figura 10C es un esquema de una Msp de cadena sencilla. Los números bajo las subunidades n.º 1 y n.º 7 representan ubicaciones de la rampa positiva.

En cualquiera de las Msp de cadena sencilla mutantes establecidas en el presente documento, la zona de constricción se puede modificar para aumentar las propiedades de reconocimiento de nucleobases, proteínas o analitos de MspA. Las modificaciones a la zona de constricción pueden crear una cabeza de lectura que aumenta, por ejemplo, las interacciones específicas de base. Se puede crear una cabeza de lectura introduciendo un aminoácido con una cadena lateral más larga que sobresalga en la ruta del ADN u otro analito. Por ejemplo, y sin ser limitante, para crear una o más cabezas de lectura, el aminoácido en la posición 90 y/o 91 en cualquiera de los monómeros de Msp mutantes de las Msp de cadena sencilla descritas en el presente documento se puede sustituir con arginina, lisina, histidina, glutamina, metionina, treonina, fenilalanina, tirosina, triptófano o un aminoácido no natural. También se pueden crear cabezas de posicionamiento para aumentar la eficacia de una o más cabezas de lectura. Por ejemplo, se pueden introducir aminoácidos con cadenas laterales más largas, preferentemente hidrófobos o cargados negativamente, en la zona opuesta a la cabeza de lectura, para reducir los movimientos de escape del ADN u otro analito en la zona de constricción. Los aminoácidos que son adecuados incluyen, pero sin limitación, aspartato, glutamato, valina, leucina, isoleucina, fenilalanina, tirosina, triptófano y aminoácidos no naturales. Para ralentizar aún más las tasas de translocación, se puede crear una lámina de apilamiento mutando uno o más de los aminoácidos en las posiciones 83, 86, 88 y 105. Por ejemplo, y sin ser limitante, uno o más de los aminoácidos en las posiciones 83, 86, 88 y 105 se pueden sustituir con triptófano, tirosina o fenilalanina. Opcionalmente, la lámina de apilamiento se coloca de manera que esté ubicada cerca de una rampa positiva.

Una o más de la segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencias de monómeros de Msp o cualquier subconjunto de las mismas, puede seleccionarse de manera independiente del grupo que consiste en un monómero de MspA de tipo silvestre, un monómero de MspA mutante, un monómero de MspA de tipo silvestre parólogo u homólogo y un monómero de MspA mutante parólogo u homólogo. Se entiende que, cuando la segunda, tercera, cuarta, quinta, sexta, séptima y/u octava secuencias de monómeros de Msp es una secuencia de monómero de MspA mutante, la secuencia de monómero de MspA mutante puede ser la misma o diferente que la primera secuencia de monómero de MspA mutante. Opcionalmente, la segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencias de monómeros de Msp, o cualquier subconjunto de las mismas, es un monómero de MspA de tipo silvestre parólogo u homólogo. Estos incluyen, pero sin limitación, MspA/Msmeg0965, MspB/Msmeg0520, MspC/Msmeg5483, MspD/Msmeg6057, MppA, PorM1, PorM2, PorM1, Mmcs4296, Mmcs4297, Mmcs3857, Mmcs4382, Mmcs4383, Mjls3843, Mjls3857, Mjls3931, Mjls4674, Mjls4675, Mjls4677, Map3123c, Mav3943, Mvan1836, Mvan4117, Mvan4839, Mvan4840, Mvan5016, Mvan5017, Mvan5768, MUL_2391, Mflv1734, Mflv1735, Mflv2295, Mflv1891, MCH4691c, MCH4689c, MCH4690c, MAB1080, MAB1081, MAB2800, RHA1 ro08561, RHA1 ro04074 y RHA1 ro03127. Un monómero de MspA de tipo silvestre parólogo u homólogo puede ser un monómero de MspB de tipo silvestre.

Monómeros de Msp mutantes y los ácidos nucleicos que los codifican

Se proporciona además un ácido nucleico que codifica un monómero de Msp mutante, en donde el monómero de Msp comprende una mutación P97F. Opcionalmente, el monómero de Msp mutante puede comprender además una mutación en la posición de aminoácido D118, una mutación en la posición D134 o una mutación en la posición E139. Por ejemplo, el monómero de Msp mutante que comprende la mutación en la posición 97 puede comprender además una mutación D118R, una mutación D134R y/o una mutación E139K. Opcionalmente, el monómero de Msp mutante que comprende la mutación en la posición 97 puede comprender además (i) una mutación en la posición 93 y/o (ii) una mutación en la posición D90, posición D91 o ambas posiciones D90 y D91. Opcionalmente, el aminoácido en la posición 90 o 91 está sustituido con arginina, lisina, histidina, glutamina, metionina, treonina, fenilalanina, tirosina o triptófano. Opcionalmente, el monómero de Msp mutante puede comprender además una mutación D90N, una D91N y una D93N. Por ejemplo, y sin ser limitante, una secuencia de monómero de MspA mutante puede comprender una mutación D90N, una mutación D91N, una mutación D93N, una mutación P97F, una mutación D118R, una mutación D134R y una mutación E139K.

Como se usa en el presente documento, un monómero de Msp mutante se refiere a un monómero de Msp que tiene al menos o como máximo un 70, 75, 80, 85, 90, 95, 98 o 99 por ciento o más de identidad, o cualquier intervalo que pueda derivarse a partir de los mismos, pero menos del 100 % de identidad, en comparación con un monómero de

Msp de tipo silvestre, y conserva la capacidad de formación de túneles cuando se asocia con uno o más monómeros de Msp diferentes (de tipo silvestre o mutante). Por lo tanto, además de las mutaciones descritas en el presente documento, cualquier Msp mutante proporcionada en el presente documento puede comprender además modificaciones adicionales tales como sustituciones, inserciones, eliminaciones y/o adiciones, siempre que el monómero de Msp mutante tenga al menos o como máximo un 70, 75, 80, 85, 90, 95, 98 o 99 por ciento o más de identidad, o cualquier intervalo que pueda derivarse a partir de los mismos, pero menos del 100 %, con respecto a un monómero de Msp de tipo silvestre y conserva la capacidad de formación de túneles cuando se asocia con uno o más monómeros de Msp diferentes.

- 10 Cualquier Msp mutante descrita en el presente documento puede comprender 2-15 monómeros de Msp que son iguales o diferentes, en donde al menos uno de los monómeros de Msp es un monómero de Msp mutante. Opcionalmente, una Msp mutante comprende 7-9 monómeros de Msp que son iguales o diferentes. Opcionalmente, al menos un segundo monómero se selecciona del grupo que consiste en un monómero de Msp de tipo silvestre, un segundo monómero de Msp mutante, un monómero de Msp de tipo silvestre parólogo u homólogo y un monómero de Msp mutante parólogo u homólogo, en donde el segundo monómero de Msp mutante puede ser el mismo que o diferente al primer monómero de Msp mutante. Por ejemplo, cualquier Msp mutante descrita en el presente documento puede comprender 2-15 monómeros de Msp en donde al menos uno de los monómeros de Msp es un monómero de MspA mutante. Opcionalmente, al menos un segundo monómero se selecciona del grupo que consiste en un monómero de MspA de tipo silvestre, un segundo monómero de MspA mutante, un monómero de MspA de tipo silvestre parólogo u homólogo y un monómero de MspA mutante parólogo u homólogo, en donde el segundo monómero de MspA mutante puede ser el mismo que o diferente al primer monómero de MspA mutante. Opcionalmente, el segundo monómero es un monómero de MspA de tipo silvestre parólogo u homólogo.

Por ejemplo, una Msp mutante puede comprender uno o más monómeros de Msp que comprenden una mutación P97F.

- Las modificaciones en la secuencia de aminoácidos pueden surgir como variaciones alélicas (p. ej., debido a polimorfismo genético), puede surgir debido a la influencia ambiental (por ejemplo, debido a la exposición a la radiación ultravioleta) u otra intervención humana (por ejemplo, por mutagénesis de secuencias de ADN clonadas), tal como mutantes puntuales inducidos, por eliminación, por inserción y por sustitución. Estas modificaciones pueden dar como resultado cambios en la secuencia de aminoácidos, proporcionar mutaciones silenciosas, modificar un sitio de restricción o proporcionar otras mutaciones específicas. Las modificaciones de la secuencia de aminoácidos se encuentran normalmente en una o más de tres clases: modificaciones por sustitución, inserción o eliminación. Las inserciones incluyen fusiones amino y/o terminales, así como inserciones intrasecuencia de uno solo o de múltiples restos de aminoácidos. Las inserciones normalmente serán inserciones más pequeñas que las de las fusiones amino o carboxilo terminales, por ejemplo, en el orden de uno a cuatro restos. Las eliminaciones se caracterizan por la eliminación de uno o más restos de aminoácidos de la secuencia de proteína. Normalmente, se eliminan no más de aproximadamente 2 a aproximadamente 6 restos en un sitio cualquiera en la molécula de proteína. Las sustituciones de aminoácidos normalmente son de restos individuales, pero pueden producirse en una serie de ubicaciones diferentes a la vez; las inserciones normalmente serán en el orden de aproximadamente 1 a aproximadamente 10 restos de aminoácidos; y las eliminaciones variarán de aproximadamente 1 a aproximadamente 30 restos. Las eliminaciones o inserciones se realizan preferentemente en pares adyacentes, es decir, una eliminación de 2 restos o inserción de 2 restos. Pueden combinarse sustituciones, eliminaciones o cualquier combinación de las mismas para obtener una construcción final. Las mutaciones pueden o no colocar la secuencia fuera del marco de lectura y pueden o no crear regiones complementarias que podrían producir una estructura de ARNm secundaria. Las modificaciones por sustitución son aquellas en las que se ha eliminado al menos un resto y se ha insertado un resto diferente en su lugar.

- Las modificaciones, incluyendo las sustituciones de aminoácidos específicas descritas en el presente documento, se realizan mediante métodos conocidos. A modo de ejemplo, las modificaciones se realizan mediante mutagénesis de sitio específico de nucleótidos en el ADN que codifica la proteína, produciendo de ese modo un ADN que codifica la modificación y, posteriormente, expresando el ADN en un cultivo de células recombinantes para producir los monómeros o los multímeros de cadena sencilla de Msp. Las técnicas para realizar mutaciones por sustitución en sitios predeterminados en ADN que tiene una secuencia conocida son bien conocidas, por ejemplo, mutagénesis de cebador M13 y mutagénesis de PCR.

- Los aminoácidos en las proteínas Msp descritas en el presente documento pueden ser cualquiera de los 20 aminoácidos naturales, D-estereoisómeros de los aminoácidos naturales, aminoácidos no naturales y aminoácidos químicamente modificados. Los aminoácidos no naturales (es decir, los que no se encuentran de manera natural en las proteínas) también se conocen en la técnica, tal como se establece en, por ejemplo, Williams *et al.*, Mol. Cell. Biol. 9:2574 (1989); Evans *et al.*, J. Amer. Chem. Soc. 112:4011-4030 (1990); Pu *et al.*, J. Amer. Chem. Soc. 56:1280-1283 (1991); Williams *et al.*, J. Amer. Chem. Soc. 113:9276-9286 (1991); y todas las referencias citadas en los mismos. Los aminoácidos B y γ se conocen en la técnica y también se contemplan en el presente documento como aminoácidos no naturales.

Como se usa en el presente documento, un aminoácido químicamente modificado se refiere a un aminoácido cuya cadena lateral se ha modificado químicamente. Por ejemplo, una cadena lateral se puede modificar para que comprenda un resto de señalización, tal como un fluoróforo o un radiomarcador. También se puede modificar una cadena lateral para que comprenda un nuevo grupo funcional, tal como un tiol, ácido carboxílico o grupo amino. Los aminoácidos modificados postraduccionalmente también se incluyen en la definición de aminoácidos modificados químicamente.

También se contemplan las sustituciones de aminoácidos conservativas. A modo de ejemplo, se pueden realizar sustituciones conservadoras de aminoácidos en uno o más de los restos de aminoácidos de cualquier monómero de Msp proporcionado en el presente documento. Un experto en la materia sabría que una sustitución conservativa es la sustitución de un resto de aminoácido por otro que es biológica y/o químicamente similar. Los ocho grupos siguientes contienen, cada uno, aminoácidos que son sustituciones conservativas entre sí:

- 1) Alanina (A), Glicina (G);
- 2) Ácido aspártico (D), Ácido glutámico (E);
- 3) Asparagina (N), Glutamina (Q);
- 4) Arginina (R), Lisina (K);
- 5) Isoleucina (I), Leucina (L), Metionina (M), Valina (V);
- 6) Fenilalanina (F), Tirosina (Y), Triptófano (W);
- 7) Serina (S), Treonina (T); y
- 8) Cisteína (C), Metionina (M)

También se contemplan sustituciones no conservativas, por ejemplo, la sustitución de una prolina por glicina.

Los expertos en la materia comprenderán fácilmente cómo determinar la identidad de dos polipéptidos o ácidos nucleicos. Por ejemplo, la identidad puede calcularse después de alinear las dos secuencias, de tal forma que la identidad se encuentra en su nivel más elevado. Puede llevarse a cabo otra forma para calcular la identidad mediante algoritmos publicados. La alineación óptima de secuencias para la comparación se puede realizar utilizando el algoritmo de Smith y Waterman, *Adv. Appl. Math.* 2: 482 (1981); mediante el algoritmo de alineación de Needleman y Wunsch, *J. Mol. Biol.* 48: 443 (1970); mediante el método de búsqueda de similitud de Pearson y Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85: 2444 (1988); mediante implementaciones informatizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el paquete informático de Wisconsin Genetics, Genetics Computer Group, 575 Science Dr., Madison, WI; el algoritmo BLAST de Tatusova y Madden *FEMS Microbiol. Lett.* 174: 247-250 (1999), disponible en el National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html>); o mediante inspección.

Pueden obtenerse los mismos tipos de identidad para ácidos nucleicos mediante, por ejemplo, los algoritmos divulgados en Zuker, *Science* 244:48-52, 1989; Jaeger *et al.* *Proc. Natl. Acad. Sci. USA* 86:7706-7710, 1989; Jaeger *et al.* *Methods Enzymol.* 183:281-306, 1989. Se entiende que normalmente se puede utilizar cualquiera de los métodos y que, en determinados casos, los resultados de estos diversos métodos pueden diferir, pero el experto en la materia entenderá que, si se encuentra identidad con al menos uno de estos métodos, se diría que las secuencias tienen la identidad establecida.

Por ejemplo, como se usa en el presente documento, una secuencia que se indica que tiene un porcentaje particular de identidad con otra secuencia se refiere a secuencias que tienen la identidad indicada calculada mediante uno cualquiera o más de los métodos de cálculo descritos anteriormente. Por ejemplo, una primera secuencia tiene un 80 por ciento de identidad, como se define en el presente documento, con respecto a una segunda secuencia si se calcula que la primera secuencia tiene un 80 por ciento de identidad con la segunda secuencia utilizando el método de cálculo de Zuker incluso si la primera secuencia no tiene un 80 por ciento de identidad con respecto a la segunda secuencia calculada por cualquiera de los otros métodos de cálculo. Como otro ejemplo más, una primera secuencia tiene un 80 por ciento de identidad, como se define en el presente documento, con respecto a una segunda secuencia si se calcula que la primera secuencia tiene un 80 por ciento de identidad con la segunda secuencia utilizando cada uno de los métodos de cálculo (aunque, en la práctica, los diferentes métodos de cálculo con frecuencia darán como resultado diferentes porcentajes de identidad calculados).

Además, cualquier monómero de Msp o Msp también puede modificarse química o biológicamente. Por ejemplo, se puede modificar un monómero de Msp o Msp con productos químicos para producir puentes disulfuro, como conocen los expertos en la materia.

Una Msp puede comprender un sitio de unión a nucleótidos. Como se usa en el presente documento, un sitio de unión a nucleótidos se refiere a un sitio en una Msp donde un nucleótido permanece en contacto con, o reside en, un aminoácido durante un período de tiempo más largo que el atribuible al movimiento de difusión, tal como más de un picosegundo o un nanosegundo. Se pueden emplear cálculos de dinámica molecular para evaluar estos tiempos de reposo temporales.

También se proporcionan polipéptidos codificados por ácidos nucleicos descritos en el presente documento. Por lo tanto, se proporcionan polipéptidos que comprenden un monómero de Msp mutante o un fragmento funcional del mismo. Los ejemplos no limitantes de monómeros de Msp mutantes incluyen, pero sin limitación, polipéptidos que comprenden la SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 y SEQ ID NO: 4 que comprenden cualquiera de las mutaciones descritas en el presente documento. Se proporciona además un monómero de Msp que comprende una secuencia de aminoácidos que tiene al menos un 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 99 % de identidad o cualquier porcentaje intermedio con un polipéptido que comprende la SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 y SEQ ID NO: 4, en donde el polipéptido comprende cualquiera de las mutaciones descritas en el presente documento. También se proporcionan polipéptidos que comprenden una Msp de cadena sencilla mutante o un fragmento funcional de la misma. También se proporcionan polipéptidos que comprenden una Msp de cadena sencilla mutante que comprende cualquiera de los monómeros de Msp mutantes descritos en el presente documento o un fragmento de los mismos.

En la tabla 2 se proporcionan ejemplos no limitantes de monómeros de Msp mutantes que comprenden las mutaciones expuestas en el presente documento. Las mutaciones no incluidas en las reivindicaciones son de referencia. Cada monómero de Msp mutante ilustrativo comprende todas las mutaciones enumeradas para cada monómero. Por ejemplo, D90N/D91N/D93N/P97F indica que están presentes las cuatro mutaciones. También se proporcionan en el presente documento Msp de cadena sencilla que comprenden cualquiera de las secuencias de monómeros de Msp mutantes ilustrativas proporcionadas en la tabla 2. Se entiende que los aminoácidos enumerados entre paréntesis se enumeran como alternativas para sustituciones en esa posición. Por ejemplo, P97 (A/V/L/I/F/M/Y/W o G) significa que P97 se puede sustituir por A, V, L, I, F, M, Y W o G. Sin embargo, la invención se refiere a mutantes que comprenden la mutación P97F.

Tabla 2

D90N/D91N/D93N/P97(A/V/L/I/F/M/Y/W o G)
D90N/ D91N/D93N/P97F/D118R/D134R/E139K
D90N/D91N/D93N/P97F
D90N/D91N/D93N/G92 (A/V/L/I/P/F/M/Y o W)
D90N/D91N/D93N/I94 (A/V/L/P/F/M/Y/W o G)
D90N/D91N/D93NAT95 (A/V/L/P/F/M/Y/W o G)
D90N/D91N/D93N/A96 (V/L/P/F/I/M/Y/W o G)
D90N/D91N/D93N/P98 (A/V/L/F/I/M/Y/W o G)
D90N/D91N/D93N/F99 (A/V/L/P/I/M/Y/W o G)
D90N/D91N/D93N/G100 (A/V/L/P/F/I/M/Y o W)
D90N/D91N/D93N/L101 (A/V/I/F/M/Y/W o G)
D90N/D91N/D93N/N102 ((A/V/L/I/P/F/M/Y/W o G)
D90N/D91N/D93N/S103 ((A/V/L/I/P/F/M/Y/W o G)
D90N/G92F/ D91N /D93N/P97F
D90N/T95F/ D91N /D93N/P97F
D90N/A96F/ D91N /D93N/P97F
D90N/A96F/ D91N /D93N/P97F/P98F
D90N/G92F/ D91N /D93N/P97F/D118R/D134R/E139K
I68 (R/K/F/W/Y o H)/D90N/D91N/D93N
I68 (R/K/F/W/Y o H)/D90N/D91N/D93N/D118R/D134R/E139K
I68 (K/R/F/W/Y o H)/D90N/D91N/D93N
I68 (K/R/F/W/Y o H)/D90N/D91N/D93N/D118R/D134R/E139K
S73 (K/R/F/W/Y o H)/D90N/D91N/D93N
S73 (K/R/F/W/Y o H)/D90N/D91N/D93N/D118R/D134R/E139K
S73 (K/R/F/W/Y o H)/D90N/D91N/D93N
S73 (K/R/F/W/Y o H)/D90N/D91N/D93N/D118R/D134R/E139K
D90N/D91N/D93N/S116R
D90N/D91N/D93N/S116R/D118R/D134R/E139K
D90N/D91N/D93N/S116K
D90N/D91N/D93N/S116K/D118R/D134R/E139K
D90N/D91N/D93N/P123R
D90N/D91N/D93N/P123R/D118R/D134R/E139K
D90N/D91N/D93N/P123K
D90N/D91N/D93N/P123K/D118R/D134R/E139K
D90N/D91N/D93N/L88 (K/R/F/W/H o Y)
D90N/D91N/D93N/I105 (K/R/F/W/H o Y)
D90N/D91N/D93N/L88 (K/R/F/W/H o Y)/D118R/D134R/E139K
D90N/D91N/D93N/I105 (K/R/F/W/H o Y)/D118R/D134R/E139K
D90N/D91N/D93N/L88 (K/R/F/W/H o Y)/I105 (K/R/F/W/H o Y)

(continuación)

D90N/D91 N/D93N/L88 (K/R/F/W/H o Y)/1105 (K/R/F/W/H o Y)/D118R/D134R/E139K
D90N/D91N/D93N/L88 (K/R/F/W/H o Y)
D90N/D91 N/D93N/T83 (K/R/F/W/H o Y)/D118R/D134R/E139K
D90N/D91N/D93N/T83 (K/R/F/W/H o Y)
D90N/D91N/D93N/N86 (K/R/F/W/H o Y)
D90N/D91 N/D93N/N86 (K/R/F/W/H o Y)/D118R/D134R/E139K

Proteínas formadoras de túneles

- 5 Los métodos para determinar si una proteína es una proteína formadora de túnel son bien conocidos en la técnica. Se puede determinar si una Msp forma un túnel determinando si la proteína se inserta en una bicapa, tal como se describe en el ejemplo 2 de la publicación de patente de Estados Unidos N.º 20120055792. Todos los métodos de preparación y uso de porinas descritos en la publicación de patente de Estados Unidos n.º 20120055792 pueden emplearse para preparar y utilizar las porinas Msp descritas en el presente documento. Si la proteína se inserta en la bicapa, entonces
- 10 la porina es una proteína formadora de túnel. Normalmente, la formación de túneles se detecta al observar un cambio discreto en la conductividad. Véanse, la publicación de patente de Estados Unidos. N.º 20120055792 y Niederweis *et al.*, Mol. Microbiol. 33:933 (1999).

- Las bicapas se describen en el presente documento. Normalmente, una Msp podrá insertarse en una bicapa lipídica u otra película delgada, que son bien conocidas en la técnica. Un ejemplo de inserción de una MspA mutante en una bicapa lipídica se proporciona en la publicación de patente de Estados Unidos N.º 20120055792; esta técnica también se puede aplicar a otras proteínas Msp. De manera adicional, la patente de Estados Unidos N.º 6.746.594 describe una variedad de bicapas lipídicas y películas delgadas, incluyendo materiales inorgánicos, que se pueden emplear con respecto a las Msp analizadas en el presente documento. También se pueden emplear métodos, aparatos y
- 15 técnicas descritos en la patente de Estados Unidos n.º 6.267.872, con respecto a las Msp analizadas en el presente documento. Por otra parte, más de una Msp puede estar comprendida en una bicapa lipídica. Por ejemplo, 2, 3, 4, 5, 10, 20, 200, 2000 o más pueden estar comprendidas en una bicapa lipídica. Opcionalmente, en cualquier lugar se pueden emplear de 2 a 10^{10} Msp en los métodos descritos en el presente documento. Dicha pluralidad de Msp puede estar en forma de agrupaciones de Msp. Las agrupaciones se pueden ensamblar al azar o pueden adoptar un patrón.
- 20 Como se usa en el presente documento, una agrupación se refiere a moléculas que se agrupan y se mueven como una unidad, pero no están unidas covalentemente entre sí.

- Opcionalmente, las Msp no se abren de forma espontánea. Como se usa en el presente documento, activar o activación se refiere al cambio espontáneo de la conductancia eléctrica a través del túnel de la proteína que suele ser
- 25 temporal (p. ej., que dura desde 1 a 10 milisegundos hasta un segundo). Los eventos de activación de larga duración con frecuencia se pueden revertir cambiando la polaridad. En la mayoría de circunstancias, la probabilidad de activación aumenta con la aplicación de voltajes más altos. La activación y el grado de conductancia a través del cambio de túnel son muy variables entre las Msp, dependiendo de, por ejemplo, la composición del vestíbulo y la zona de constricción, así como de las propiedades del medio líquido en el que se sumerge la proteína. Normalmente, la
- 30 proteína se vuelve menos conductora durante la activación y la conductancia puede detenerse permanentemente (es decir, el túnel puede cerrarse permanentemente) como resultado, de manera que el proceso sea irreversible. Opcionalmente, la activación se refiere a la conductancia a través del túnel de una proteína que cambia espontáneamente a menos del 75 % de su corriente en estado abierto.

- 40 Varias condiciones tales como medio ligero y líquido, incluyendo su pH, composición del tampón, composición y temperatura del detergente, pueden afectar el comportamiento de una Msp, particularmente con respecto a su conductancia a través del túnel, así como al movimiento de un analito con respecto al túnel, ya sea de manera temporal o permanente.

- 45 Como se usa a lo largo de la presente memoria descriptiva, un túnel se refiere a la porción central, vacía de una Msp que está definida por el vestíbulo y la zona de constricción, a través del cual puede pasar un gas, líquido, ion o analito. Como se usa en el presente documento, "cis" se refiere al lado de un túnel de Msp a través del cual un analito entra en el túnel o por cuya cara se mueve el analito. Como se usa en el presente documento, "trans" se refiere al lado de un túnel de Msp a través del cual un analito (o fragmentos del mismo) sale del túnel o por cuya cara no se mueve el
- 50 analito.

- Cualquier Msp mutante descrita en el presente documento, por ejemplo, una MspA mutante, puede comprender un vestíbulo y una zona de constricción que definen un túnel. Además, el diámetro de un parálogo u homólogo de Msp mutante o Msp mutante puede ser menor que el diámetro de la zona de constricción de un parálogo u homólogo de
- 55 Msp de tipo silvestre o Msp de tipo silvestre correspondiente. Una Msp mutante o un parálogo u homólogo de Msp mutante pueden tener una mutación en el vestíbulo o en la zona de constricción que permite que un analito se transloque, electroforéticamente o de otro modo, a través del túnel del parálogo u homólogo de Msp mutante o Msp mutante con una velocidad de translocación o una velocidad de translocación promedio que es menor que la velocidad

de translocación o la velocidad de translocación promedio a la que el analito se transloca a través del túnel de una Msp de tipo silvestre o de un parólogo u homólogo de Msp de tipo silvestre. Adicionalmente, cualquier Msp mutante descrita en el presente documento puede comprender un vestíbulo que tiene una longitud de aproximadamente 2 a aproximadamente 6 nm y un diámetro de aproximadamente 2 a aproximadamente 6 nm, y una zona de constricción 5 que tiene una longitud de aproximadamente 0,3 a aproximadamente 3 nm y un diámetro de aproximadamente 0,3 a aproximadamente 3 nm, en donde el vestíbulo y la zona de constricción juntos definen un túnel. Se entiende que, pueden realizarse una o más mutaciones, en el vestíbulo o en la zona de constricción de cualquiera de las Msp descritas en el presente documento para de aumentar o disminuir la conductancia a través del túnel de una Msp. Por ejemplo, cualquiera de las Msp mutantes descritas en el presente documento puede comprender además una 10 eliminación, sustitución o inserción de un aminoácido en el vestíbulo y/o en la zona de constricción para modificar la conductancia.

Como se usa a lo largo de la presente memoria descriptiva, un vestíbulo se refiere a la parte en forma de cono del interior de una Msp cuyo diámetro generalmente disminuye de un extremo al otro a lo largo de un eje central, donde 15 la parte más estrecha del vestíbulo está conectada a la zona de constricción. Un vestíbulo también puede denominarse cáliz. El vestíbulo y la zona de constricción juntos definen el túnel de una Msp. Al referirse a un diámetro del vestíbulo, se entiende que debido a que el vestíbulo tiene forma de cono, el diámetro cambia a lo largo de la trayectoria de un eje central, donde el diámetro es mayor en un extremo que en el extremo opuesto. El diámetro de los poros puede variar de aproximadamente 2 a aproximadamente 6 nm. Opcionalmente, el diámetro es aproximadamente, al menos 20 aproximadamente, o como máximo aproximadamente 2, 2,1, 2,2, 2,3, 2,4, 2,5, 2,6, 2,7, 2,8, 2,9, 3,0, 3,1, 3,2, 3,3, 3,4, 3,5, 3,6, 3,7, 3,8, 3,9, 4,0, 4,1, 4,2, 4,3, 4,4, 4,5, 4,6, 4,7, 4,8, 4,9, 5,0, 5,1, 5,2, 5,3, 5,4, 5,5, 5,6, 5,7, 5,8, 5,9 o 6,0 nm, o cualquier intervalo que pueda derivarse a partir de los mismos. La longitud del eje central puede variar entre aproximadamente 2 nm y aproximadamente 6 nm. Opcionalmente, la longitud es aproximadamente, al menos aproximadamente, o como máximo aproximadamente 2, 2,1, 2,2, 2,3, 2,4, 2,5, 2,6, 2,7, 2,8, 2,9, 3,0, 3,1, 3,2, 3,3, 3,4, 25 3,5, 3,6, 3,7, 3,8, 3,9, 4,0, 4,1, 4,2, 4,3, 4,4, 4,5, 4,6, 4,7, 4,8, 4,9, 5,0, 5,1, 5,2, 5,3, 5,4, 5,5, 5,6, 5,7, 5,8, 5,9 o 6,0 nm, o cualquier intervalo que pueda derivarse a partir de los mismos. Al referirse al diámetro en el presente documento, se puede determinar un diámetro midiendo distancias de centro a centro o distancias atómicas de superficie a superficie.

Como se usa a lo largo de la presente memoria descriptiva, una zona de constricción se refiere a la parte más estrecha del túnel de una Msp, en términos de diámetro, que está conectada al vestíbulo. La longitud de la zona de constricción puede variar entre aproximadamente 0,3 nm y aproximadamente 2 nm. Opcionalmente, la longitud es aproximadamente, como máximo aproximadamente, o al menos aproximadamente 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1,0, 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2 o 3 nm, o cualquier intervalo que pueda derivarse a partir de los mismos. El 30 diámetro de la zona de constricción puede variar entre aproximadamente 0,3 nm y aproximadamente 2 nm. Opcionalmente, el diámetro es aproximadamente, como máximo aproximadamente, o al menos aproximadamente 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1,0, 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2 o 3 nm, o cualquier intervalo que pueda derivarse a partir de los mismos.

Cualquier Msp analizada en el presente documento puede estar comprendida en una bicapa lipídica. Opcionalmente, un analito se transloca electroforéticamente o de otro modo desde el lado cis a través de un túnel hacia el lado trans. Opcionalmente, un analito se transloca electroforéticamente o de otro modo desde el lado trans a través de un túnel hacia el lado cis. Opcionalmente, un analito se conduce electroforéticamente o de otra manera desde el lado cis o el lado trans hacia un túnel y permanece en el túnel o después se retrae hacia el lado cis o el lado trans, respectivamente. 40 Se entiende que los analitos pueden translocarse a través del túnel en presencia o ausencia de un campo eléctrico.

Las Msp de cadena sencilla funcionan en un amplio intervalo de concentración de electrolitos, por ejemplo, KCl de aproximadamente 0,3-1 M (véase la figura 22). Para optimizar la actividad de canal, la asociación de lípidos se puede realizar antes de la inserción de Msp en una membrana o bicapa lipídica. En un ejemplo no limitante, la figura 22 50 muestra que no se observó actividad de canal en un tampón que contenía solamente KCl 0,3 M a pH 8,0. Sin embargo, la rotura de la membrana y el posterior repintado de la membrana da lugar a un aumento de la actividad de canal de scMspA PN1 en el electrolito que contiene KCl 0,3 M a pH 8,0. Por lo tanto, en cualquiera de los métodos expuestos en el presente documento, una Msp puede ponerse en contacto o incubarse previamente con uno o más lípidos para optimizar la actividad de canal.

55 **Vectores y células**

También se proporciona un vector que comprende un ácido nucleico que codifica un polipéptido descrito en el presente documento. El vector puede comprender además una secuencia promotora, por ejemplo, un promotor constitutivo o 60 un promotor inducible. Los ejemplos de promotores constitutivos incluyen, pero sin limitación, el promotor p_{smyc} y Phsp60. Los ejemplos de promotores inducibles incluyen, pero sin limitación, un promotor inducible por acetamida y un promotor inducible por tetraciclina.

También se proporcionan células cultivadas transfectadas con cualquier vector descrito en el presente documento la 65 descendencia de las mismas, en donde la célula es capaz de expresar una Msp (ya sea como una Msp de cadena

sencilla, una Msp que comprende monómeros de Msp o un monómero de Msp). También se proporciona una cepa de *Mycobacterium smegmatis* que comprende cualquier vector descrito en el presente documento. También se contempla una cepa de *Mycobacterium smegmatis* sin porinas endógenas y puede comprender además cualquier vector descrito en el presente documento. Por "sin" se entiende que no se puede detectar una porina endógena en una
 5 inmunotransferencia cuando se usa un antisuero específico de Msp apropiado o que comprende menos del 1 % de porinas endógenas.

Cualquiera de los monómeros de Msp o Msp de cadena sencilla descritos en el presente documento se puede producir transformando una cepa bacteriana mutante que comprende una eliminación de una MspA de tipo silvestre, una MspB
 10 de tipo silvestre, una MspC de tipo silvestre, una MspD de tipo silvestre, con un vector que comprende un promotor inducible unido operativamente a una secuencia de un ácido nucleico que codifica el monómero de Msp o la porina Msp de cadena sencilla; y purificando el monómero de Msp o la porina Msp de cadena sencilla (véase, por ejemplo, la patente de Estados Unidos. n.º 6.746.594). Opcionalmente, la cepa bacteriana mutante comprende una eliminación de un gen *recA*. Opcionalmente, el vector comprende cualquiera de los ácidos nucleicos que codifican un monómero
 15 de Msp o Msp de cadena sencilla descritos en el presente documento. La cepa bacteriana puede comprender además las cepas de *M. smegmatis* ML16, ML714 o ML712.

Sistemas y métodos de uso

20 También se proporciona un sistema que comprende una Msp mutante descrita en el presente documento que tiene un vestíbulo y una zona de constricción que definen un túnel, en donde el túnel se coloca entre un primer medio líquido y un segundo medio líquido, en donde al menos un medio líquido comprende un analito y en donde el sistema es operativo para detectar una propiedad del analito. Un sistema puede ser operativo para detectar una propiedad de cualquier analito que comprende someter una Msp a un campo eléctrico de manera que el analito interactúe con la
 25 Msp. Un sistema puede ser operativo para detectar una propiedad del analito que comprende someter la Msp a un campo eléctrico de modo que el analito se transloque electroforéticamente a través del túnel de la Msp. También se proporciona un sistema que comprende una Msp que tiene un vestíbulo y una zona de constricción que define un túnel, en donde el túnel se coloca en una bicapa lipídica entre un primer medio líquido y un segundo medio líquido y en donde el único punto de comunicación de líquidos entre el primer y segundo medio líquido se produce en el túnel.
 30 Por otra parte, cualquier sistema descrito en el presente documento puede comprender cualquier Msp mutante descrita en el presente documento.

El primer y segundo medio líquido pueden ser iguales o diferentes y uno o ambos pueden comprender una o más sales, detergentes o tampones. En realidad, cualquier medio líquido descrito en el presente documento puede
 35 comprender uno o más de una sal, un detergente o un tampón. Opcionalmente, al menos un medio líquido es conductor. Opcionalmente, al menos un medio líquido no es conductor. Cualquier medio líquido descrito en el presente documento puede comprender una sustancia que altere la viscosidad o una sustancia que altere la velocidad. El medio líquido puede comprender cualquier analito descrito en el presente documento.

40 Una propiedad de un analito puede ser una propiedad eléctrica, química o física. Una Msp puede estar comprendida en una bicapa lipídica en un sistema o en cualquier otra realización descrita en el presente documento. Un sistema puede comprender una pluralidad de Msp. Un sistema puede comprender cualquier Msp descrita en el presente documento, tal como una MspA mutante de cadena sencilla o una Msp mutante que comprende al menos 2-15 monómeros, en donde al menos uno de los monómeros es un monómero de MspA mutante. Una Msp mutante
 45 comprendida en un sistema puede comprender un vestíbulo que tiene una longitud de aproximadamente 2 a aproximadamente 6 nm y un diámetro de aproximadamente 2 a aproximadamente 6 nm, y una zona de constricción que tiene una longitud de aproximadamente 0,3 a aproximadamente 3 nm y un diámetro de aproximadamente 0,3 a aproximadamente 3 nm, en donde el vestíbulo y la zona de constricción juntos definen un túnel.

50 Cualquier Msp descrita en el presente documento, incluyendo una Msp comprendida en un sistema, puede comprender además un motor molecular. El motor molecular de un sistema es capaz de mover un analito hacia o a través de un túnel con una velocidad de translocación o una velocidad de translocación promedio que es menor que la velocidad de translocación o la velocidad de translocación promedio a la que el analito se transloca hacia o a través del túnel en ausencia del motor molecular. El motor molecular puede ser, por ejemplo, una enzima, tal como una
 55 polimerasa, una exonucleasa, o una helicasa tal como DnaB o motores de empaquetamiento de ácidos nucleicos de fagos (véase, por ejemplo, Serwer Viruses 3(7): 1249-80 (2011)). Cualquier sistema descrito en el presente documento puede comprender además un amplificador de pinzamiento zonal o un dispositivo de adquisición de datos. Un sistema puede comprender además uno o más dispositivos de regulación de temperatura en comunicación con el primer medio líquido, el segundo medio líquido, o ambos. Cualquier sistema descrito en el presente documento puede ser operativo
 60 para translocar un analito a través de un túnel de Msp de forma electroforética o de otro modo.

La MspA mutante puede tener una mutación en el vestíbulo o en la zona de constricción que permita que un analito se transloque, por ejemplo, electroforéticamente, a través del túnel con una velocidad de translocación promedio de menos de 0,5 nm/μs o menos de 0,05 nm/μs. El analito se puede seleccionar del grupo que consiste en un o unos
 65 nucleótidos, un ácido nucleico, aminoácido o aminoácidos, un péptido, una proteína, un polímero, un fármaco, un ion,

un agente de guerra biológica, un contaminante, un objeto nanoscópico o una combinación o agrupación de los mismos. Opcionalmente, el analito se define además como un ácido nucleico. El ácido nucleico puede translocarse, electroforéticamente o de otro modo, a través del túnel con una velocidad de translocación promedio de menos de 1 nucleótido/μs, o menos de 0,1 nucleótidos/μs. Un ácido nucleico se puede definir además como ADNmc, ADNbc, ARN o una combinación de los mismos.

Como se usa en el presente documento, translocación electroforética de un analito, se refiere a aplicar un campo eléctrico a una porina Msp que está en contacto con una o más soluciones (por ejemplo, sumergida en una solución), de manera que la corriente fluya a través del túnel de Msp. El campo eléctrico mueve un analito de manera que interactúa con el túnel. Como se usa en el presente documento, "interactúa" significa que el analito se mueve hacia y, opcionalmente, a través del túnel, donde "a través del túnel de Msp" (o "trasloca") significa entrar en un lado del túnel y moverse hacia y fuera del otro lado del túnel. Se contempla específicamente que cualquier analito analizado en el presente documento pueda translocarse a través de un túnel de Msp, ya sea electroforéticamente o de otro modo, en cualquier realización analizada en el presente documento. A este respecto, se contempla específicamente que cualquier realización del presente documento que comprenda translocación puede referirse a translocación electroforética o translocación no electroforética, a no ser que se indique específicamente. Opcionalmente, se contemplan métodos que no emplean translocación electroforética.

Como se usa a lo largo de la presente memoria descriptiva, un medio líquido incluye medios líquidos acuosos, orgánicos acuosos y orgánicos únicamente. Los medios orgánicos incluyen, por ejemplo, metanol, etanol, dimetilsulfóxido y mezclas de los mismos. Los líquidos que se pueden emplear en los métodos descritos en el presente documento son bien conocidos en la técnica. Descripciones y ejemplos de dichos medios, incluidos los medios líquidos conductores, se proporcionan en la patente de Estados Unidos nº 7.189.503, por ejemplo. Se pueden añadir sales, detergentes o tampones a dichos medios. Dichos agentes pueden emplearse para alterar el pH o la fuerza iónica del medio líquido. Se pueden incluir sustancias que alteran la viscosidad, tal como glicerol o diversos polímeros (p. ej., polivinilpirrolidona, polietilenglicol, alcohol polivinílico, polímeros de celulosa) y mezclas de los mismos, en los medios líquidos. Los métodos para medir la viscosidad son bien conocidos en la técnica.

Cualquier agente que se pueda añadir a un medio líquido también puede alterar la velocidad de un analito que se está estudiando. Se este modo, un agente que modifica la velocidad puede ser una sal, un detergente, un tampón, una sustancia que altere la viscosidad o cualquier otro agente añadido a un medio líquido que aumente o disminuya la velocidad de un analito. Normalmente, un analito empleado en el presente documento es soluble o parcialmente soluble en al menos un medio líquido que está en contacto con una Msp descrita en el presente documento.

Como se usa en el presente documento, ácido nucleico se refiere a un polímero de desoxirribonucleótidos o ribonucleótidos en forma de cadena sencilla o doble y, a menos que se limite de otro modo, abarca análogos conocidos de nucleótidos naturales que se hibridan con ácidos nucleicos de una manera similar a los nucleótidos naturales, tal como los ácidos nucleicos peptídicos (PNA, por sus siglas en inglés) y el ADN fosforotioato. A menos que se indique lo contrario, una secuencia de un ácido nucleico particular incluye la secuencia complementaria de la misma. Los nucleótidos incluyen, pero sin limitación, ATP, dATP, CTP, dCTP, GTP, dGTP, UTP, TTP, dUTP, 5-metil-CTP, 5-metildCTP, E.T.C, diTP, 2-amino-adenosina-TP, 2-amino-desoxiadenosina-TP, 2-tiotimidina trifosfato, pirrolopirimidina trifosfato y 2-tiocitidina, así como los alfatiotriposfosatos para todos los anteriores, y 2'-O-metilribonucleótido trifosfatos para todas las bases anteriores. Las bases modificadas incluyen, pero sin limitación, 5-Br-UTP, 5-BrdUTP, 5-F-UTP, 5-F-dUTP, 5-propinil dCTP y 5-propinil-dUTP.

Como se usa en el presente documento, un fármaco se refiere a cualquier sustancia que pueda alterar un proceso biológico de un sujeto. Los fármacos se pueden diseñar o utilizar para o en el diagnóstico, tratamiento o prevención de una enfermedad, trastorno, síndrome u otra afección de la salud de un sujeto. Los fármacos pueden ser de naturaleza recreativa, es decir, utilizados simplemente para alterar un proceso biológico y no utilizados para o en el diagnóstico, tratamiento o prevención de una enfermedad, trastorno, síndrome u otra afección de la salud de un sujeto. Los agentes biológicos, que se refieren a sustancias producidas mediante mecanismos biológicos que implican tecnología de ADN recombinante, también se abarcan en el término fármaco. Los fármacos incluyen, por ejemplo, antibacterianos, antiinflamatorios, anticoagulantes, antivíricos, antihipertensivos, antidepresivos, antimicrobianos, analgésicos, anestésicos, betabloqueantes, bisfosfonatos, quimioterápicos, agentes de contraste, medicamentos para la fertilidad, alucinógenos, hormonas, narcóticos, opiáceos, sedantes, estatinas, esteroides y vasodilatadores. También se pueden encontrar ejemplos no limitantes de fármacos en el Merck Index: an Encyclopedia of Chemicals, Drugs, and Biologicals, 15ª ed. Nueva Jersey: Merck, 2013. Los fármacos antibacterianos utilizados en el tratamiento de la tuberculosis, por ejemplo, incluyen isoniazida, rifampicina, pirazinamida y etambutol. Los métodos que emplean un fármaco como analito pueden comprender además la exploración de fármacos. Por ejemplo, la captación de un fármaco en una célula o un organismo se puede investigar utilizando una Msp mediante la observación de bloqueos de corrientes de iones. Se pueden construir zonas de constricción y/o vestíbulos de porinas Msp específicos con varios tamaños, propiedades electrostáticas y propiedades químicas para emular de manera muy similar la vía deseada para que los fármacos entren o salgan de una célula u organismo. Estos métodos podrían acelerar en gran medida la exploración de fármacos, así como el diseño de fármacos (véase, por ejemplo, Pagel *et al.*, J. Bacteriology 189:8593 (2007)).

Como se usa en el presente documento, un agente de guerra biológica se refiere a cualquier organismo o cualquier componente de origen natural, modificado por bioingeniería o sintetizado de cualquier microorganismo capaz de causar la muerte o enfermedad en plantas o animales (incluidos los seres humanos) o la degradación de los suministros de alimentos o agua, o la degradación del medio ambiente. Los ejemplos no limitantes incluyen los virus del Ébola, virus de Marburg, *Bacillus Anthracis* y *Clostridium botulinum*, *Variola mayor*, *Variola menor*, ántrax y ricina.

Como se usa en el presente documento, un contaminante se refiere a un material que contamina el aire, agua o suelo. Los ejemplos no limitantes de contaminantes incluyen fertilizantes, plaguicidas, insecticidas, detergentes, hidrocarburos de petróleo, humo y sustancias que contienen metales pesados, tal como los que contienen zinc, cobre o mercurio (por ejemplo, metilmercurio).

Puede utilizarse cualquier analito en el presente documento, incluyendo, por ejemplo, un o unos nucleótidos, un ácido nucleico, un o unos aminoácidos, un péptido, una proteína, un polímero, un fármaco, un ion, un agente de guerra biológica, un contaminante, un objeto nanoscópico, o cualquier otra molécula que comprenda uno de estos analitos o una combinación de los mismos. Un analito puede ser una agrupación de moléculas (p. ej., 2-10 nucleótidos o aminoácidos), en el sentido de que la agrupación en su conjunto se considera un analito. Normalmente, el tamaño de un analito no será tan grande que no pueda entrar en un túnel de una Msp. En otras palabras, un analito típico tendrá un tamaño más pequeño que la apertura de un túnel de una Msp. Sin embargo, se puede emplear un analito que tenga un tamaño mayor que la abertura de un túnel, y se puede determinar que el tamaño del analito es demasiado grande para entrar en el túnel. Opcionalmente, el peso molecular del analito es inferior a un millón de Da. Opcionalmente, el peso molecular del analito es aproximadamente, como máximo aproximadamente, o al menos aproximadamente 1.000.000, 950.000, 900.000, 850.000, 800.000, 750.000, 700.000, 650.000, 600.000, 550.000, 500.000, 450.000, 400.000, 350.000, 300.000, 250.000, 200.000, 150.000, 100.000, 75.000, 50.000, 25.000, 20.000, 15.000, 10.000, 7.500, 5.000, 2.500, 2.000, 1.500, 1.000 o 500 Da o menos, o cualquier intervalo que pueda derivarse a partir de los mismos.

Un analito también puede ser un objeto nanoscópico, que es un objeto es menor de 100 nm en dos de sus dimensiones. Como se usa en el presente documento, un analito puede comprender además una perla magnética. Una perla magnética se puede definir además como una perla magnética recubierta de estreptavidina. Un analito puede comprender además una perla óptica. Cualquier analito descrito en el presente documento puede ser un ion o puede ser neutro. Un analito puede comprender biotina.

Las perlas que se pueden emplear incluyen perlas magnéticas y perlas ópticas. Por ejemplo, se pueden usar perlas magnéticas recubiertas de estreptavidina para aplicar una fuerza opuesta a las fuerzas electrostáticas que tiran del ADN a través del túnel de una Msp. En esta última técnica, se fija una perla magnética al ADN biotinilado y se aplicaría una fuerza comparable a la fuerza impulsora electrostática (-10 pN) utilizando un fuerte gradiente de campo magnético. Véase Gosse y Croquette, *Biophys. J.* 82:3314 (2002). De este modo, la lectura de la corriente de bloqueo no se vería afectada, pero las fuerzas sobre el ADN podrían controlarse de forma independiente. Decenas o cientos de lecturas completas, independientes de cada ADN podrían correlacionarse y ensamblarse para reconstruir una secuencia de ADN precisa.

También se conocen en la técnica perlas ópticas manipuladas mediante "pinzas ópticas", y dichos métodos se pueden aplicar a las Msp descritas en el presente documento. Las pinzas ópticas son una herramienta común que se utiliza para ejercer una fuerza sobre un objeto nanoscópico. Se adjunta un analito en un extremo de la perla, mientras que el otro extremo se puede insertar en el túnel de la porina. La posición y la fuerza de la perla se controlan y miden con las pinzas ópticas. Dichos métodos controlan el paso del analito al túnel y permiten un mayor control de la lectura del analito, tal como la lectura de las unidades de un polímero. Véanse, por ejemplo, Trepagnier *et al.*, *Nano Lett.* 7:2824 (2007) para obtener una descripción de dichos métodos en el contexto de los nanoporos artificiales. La patente de Estados Unidos n.º 5.795.782 también analiza el uso de pinzas ópticas.

Puede emplearse transferencia de energía de resonancia de fluorescencia (FRET, por sus siglas en inglés), una técnica bien conocida, en los métodos analíticos descritos en el presente documento. Por ejemplo, se puede incorporar un aceptor de FRET fluorescente o una molécula donante de FRET en una Msp. A continuación, el analito se marca con un donante de FRET o un aceptor de FRET coincidente. Cuando el donante de FRET coincidente se encuentra dentro de la distancia de Forster al aceptor de FRET, es probable que se produzca una transferencia de energía. La señal resultante podría usarse con fines analíticos en lugar de o además de los métodos que utilizan corriente de iones como se describe en el presente documento. Por consiguiente, los métodos de detección, identificación o secuenciación pueden comprender tecnología FRET. Otros métodos ópticos que pueden emplearse incluyen la introducción de moléculas ópticamente activas en el interior de una Msp (tal como en el vestíbulo o en la zona de constricción). Se aplicaría luz externa para afectar el interior de la proteína. Dichos métodos podrían usarse para afectar la velocidad de traslocación de un analito o podrían permitir la entrada o salida del analito del túnel, ofreciendo el paso controlado del analito. Como alternativa, se podrían usar pulsos ópticos enfocados en el poro para calentar el poro y afectar la forma en la que interactúa con el analito. Dicho control podría ser muy rápido ya que el calor de un pequeño volumen de un punto focal se disiparía rápidamente. Por tanto, los métodos para controlar la velocidad de

translocación de un analito pueden emplear tales moléculas ópticamente activas o pulsos ópticos. La manipulación de la velocidad de translocación también se puede lograr fijando un objeto a un extremo de un analito, y el otro extremo del analito interactúa después con la Msp. El objeto puede ser una perla (por ejemplo, una perla de poliestireno), una célula, una molécula grande tal como estreptavidina, neutravidina, ADN, etc., o un objeto nanoscópico. A continuación, el objeto podría someterse a un flujo de fluido o podría someterse a un arrastre viscoso pasivo.

Los motores moleculares son bien conocidos en la técnica y se refieren a una molécula (por ejemplo, una enzima) que interactúa físicamente con un analito, tal como un polímero (por ejemplo, un polinucleótido de 15), y es capaz de mover físicamente el analito con respecto a una ubicación fija, tal como el vestíbulo, zona de constricción, o túnel de una Msp. Aunque sin pretender quedar ligados a teoría alguna, los motores moleculares utilizan energía química para generar fuerza mecánica. Un motor molecular puede interactuar con cada unidad (o "mero") de un polímero de manera secuencial. Los ejemplos no limitantes de motores moleculares incluyen ADN polimerasas, ARN polimerasas, helicasas, ribosomas y exonucleasas. También se conocen motores no enzimáticos, tal como motores de virus que empaquetan ADN. Véase Smith *et al.*, Nature 413:748 (2001). Una variedad de motores moleculares y propiedades deseables de tales motores se describen en la patente de Estados Unidos n.º 7.238.485.

Se puede disponer un motor molecular en el lado cis o en el lado trans de una porina Msp y opcionalmente se puede inmovilizar, tal como se describe en la patente '485. Los métodos para incorporar un motor molecular en una Msp se pueden realizar utilizando los métodos descritos en la patente '485. Los sistemas y aparatos descritos en la patente '485 también se pueden emplear con respecto a una Msp descrita en el presente documento. De hecho, cualquier realización analizada en la patente '485 puede emplearse utilizando una Msp, como se describe en el presente documento. Los motores moleculares también se analizan en, por ejemplo, Cockroft *et al.*, J. Amer. Chem. Soc. 130:818 (2008); Benner *et al.*, Nature Nanotech. 2:718 (2007); y Gyrfas *et al.*, ACS Nano 3:1457 (2009).

Por lo general, se emplea un motor molecular para regular la tasa o la velocidad de translocación a la que un analito interactúa con una Msp. Cualquier Msp descrita en el presente documento puede comprender un motor molecular. Opcionalmente, se emplea un motor molecular para disminuir la tasa a la que un analito entra en un túnel de porina Msp o para disminuir la velocidad de translocación a la que un analito se transloca a través de un túnel de Msp. Opcionalmente, la velocidad de translocación o la velocidad de translocación promedio es inferior a 0,5 nm/μs. Opcionalmente, la velocidad de translocación o la velocidad de translocación promedio es inferior a 0,05 nm/μs. Opcionalmente, la velocidad de translocación o la velocidad de translocación promedio es inferior a 1 nucleótido/μs. Opcionalmente, la velocidad de translocación o la velocidad de translocación promedio es inferior a 0,1 nucleótido/μs.

Opcionalmente, la tasa de movimiento de un analito varía de más de 0 Hz a 2000 Hz. En este caso, tasa se refiere al número de subunidades (o "meros") de un polímero regular que avanza en un segundo (Hz). Opcionalmente, el intervalo es de aproximadamente 50-1500Hz, 100-1500 Hz o 350-1500 Hz. Opcionalmente, la tasa de movimiento es aproximadamente, como máximo aproximadamente o al menos aproximadamente 25, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950 o 2000 Hz, o cualquier intervalo que pueda derivarse a partir de los mismos. La tasa se puede controlar mediante el uso de un motor molecular que mueve un analito a una tasa sustancialmente constante, al menos durante una parte del tiempo durante una caracterización. De manera adicional, el intervalo de tasa de movimiento puede depender del motor molecular. Por ejemplo, para una ARN polimerasa, un intervalo puede ser 350-1500 Hz; para una ADN polimerasa, un intervalo puede ser 75-1500 Hz; y para los ribosomas, helicasas y exonucleasas, un intervalo puede ser 50-1500 Hz.

Pueden emplearse técnicas de registro y detección en los métodos descritos en el presente documento. De manera adicional, las patentes de Estados Unidos números 5.795.782 y 7.189.503 también describen métodos de registro e instrumentación que se pueden emplear con respecto a Msp, así como métodos para optimizar las lecturas de conductancia. La patente de Estados Unidos n.º 6.746.594 describe un soporte para películas delgadas que contienen nanoporos y métodos para usar dichos soportes que pueden emplearse con respecto a las Msp descritas en el presente documento.

Método para preparar una Msp de cadena sencilla

El poro de Msp, por ejemplo, MspA, es actualmente la mejor proteína de canal disponible para la secuenciación de ácidos nucleicos por nanoporos. Sin embargo, su composición de ocho subunidades imposibilita introducir cambios asimétricos en el poro que pudieran optimizar las propiedades de la Msp para la secuenciación de ácidos nucleicos. Para superar dicho inconveniente, en el presente documento se proporcionan métodos para preparar una Msp de cadena sencilla. Estos métodos se pueden utilizar para producir una Msp de cadena sencilla total o parcial. Generalmente, el método comprende transformar una cepa bacteriana mutante. La cepa mutante comprende una eliminación de una MspA de tipo silvestre, una MspB de tipo silvestre, una MspC de tipo silvestre, una MspD de tipo silvestre y, opcionalmente, una eliminación del gen *recA*. La cepa mutante se transforma con un vector que comprende una secuencia de un ácido nucleico que codifica una porina Msp de cadena sencilla. A continuación, la porina Msp de cadena sencilla se purifica de las bacterias. Opcionalmente, la porina Msp de cadena sencilla comprende una porina MspA de cadena sencilla. Opcionalmente, el vector comprende cualquiera de los ácidos nucleicos descritos en el

presente documento.

Como se describe en los ejemplos, para combinar las capacidades superiores de secuenciación de MspA con una mayor capacidad para adaptar las propiedades del vestíbulo y la constricción a la secuenciación del ADN, se construyó un octámero de porina MspA de cadena sencilla que permite las propiedades óptimas del vestíbulo y la zona de constricción para la secuenciación del ADN. Los extremos de la cadena de MspA están muy juntos en la porina MspA y están conectados mediante un enlazador peptídico corto. El péptido (GGGS)₃ (SEQ ID NO: 3), por ejemplo, se utiliza para conectar el extremo carboxi del monómero (o multímero) de MspA anterior al extremo amino del siguiente monómero (o multímero) de MspA, que carece de péptido señal. Para crear un vector que comprenda la secuencia de porina MspA, cada secuencia de monómero de MspA está flanqueada por un sitio de restricción exclusivo, lo que permite la capacidad de mutar cualquier monómero individual. Para crear una secuencia de porina MspA, cada secuencia de MspA se puede ensamblar por etapas para formar una MspA de cadena sencilla dimérica, tetramérica y octamérica utilizando los sitios de restricción exclusivos. Para evitar problemas de recombinación en la creación del multímero de MspA de cadena sencilla, se sintetizan siete genes MspA con diferentes usos de codones, es decir, los genes codifican exactamente la misma secuencia de aminoácidos, sin embargo, la secuencia de ADN se ha alterado a partir de la secuencia de nucleótidos de MspA natural (SEQ ID NO: 10). Para crear la secuencia de porina MspA, la secuencia de nucleótidos que codifica el primer monómero de Msp puede contener opcionalmente una secuencia de un ácido nucleico que codifica una secuencia líder (por ejemplo, los aminoácidos 1 a 27 de la SEQ ID NO: 6). Cada una de las siete secuencias de monómeros de Msp que siguen a la primera secuencia de monómero de Msp puede comprender la SEQ ID NO: 1 o la SEQ ID NO: 1 con una o más mutaciones descritas en el presente documento. El vector que comprende la secuencia de porina MspA se transforma en la cepa bacteriana cuádruple mutante de porina, como se describe en los ejemplos. Opcionalmente, las Msp de cadena sencilla pueden purificarse y someterse a un procedimiento de replegamiento. Por ejemplo, puede usarse cromatografía de intercambio aniónico en presencia de urea 8 M para obtener una fracción pura de una Msp de cadena sencilla que se dializa frente a un tampón para eliminar la urea. Después de la diálisis, un tampón de replegamiento que comprende un agente de replegamiento, por ejemplo, L-arginina y detergente, y se obtiene una Msp de cadena sencilla purificada y replegada. Los agentes de replegamiento son conocidos por los expertos en la materia. Estos incluyen, pero sin limitación, arginina, clorhidrato de arginina, argininamida, glicinamida, prolina, glicerol y ciclodextrinas (véase, por ejemplo, Yamaguchi *et al.* Biomolecules 4: 235-251 (2014); y los niveles de expresión y el estado oligomérico de la porina MspA pueden comprobarse mediante transferencia Western u otras técnicas inmunohistoquímicas conocidas por los expertos en la materia. La actividad del túnel de la porina MspA se puede determinar mediante experimentos de bicapa lipídica, como se describe en los ejemplos y como es conocido por los expertos en la materia.

Los poros de M1₈-MspA de cadena sencilla se insertan con mucha más frecuencia en las bicapas lipídicas que una cantidad similar de M1-MspA octamérica. La inserción de MspA octamérica es un procedimiento tedioso. Por tanto, las Msp de cadena sencilla, tales como las descritas en el presente documento facilitan la configuración de sistemas y métodos de uso de Msp para detectar e identificar analitos, por ejemplo, para la secuenciación de ácidos nucleicos.

Método para aumentar la inserción de Msp en una bicapa lipídica

En el presente documento se proporciona un método para aumentar el número de inserciones de Msp en una bicapa lipídica, que comprende poner en contacto cualquier Msp descrita en el presente documento con un lípido para formar una Msp asociada a lípidos e insertar la Msp asociada a lípidos de la etapa en una bicapa lipídica. Opcionalmente, la etapa de poner en contacto comprende insertar la Msp en una bicapa lipídica y romper la bicapa lipídica para formar una Msp asociada a lípidos. Por ejemplo, se puede insertar una Msp en una bicapa lipídica que posteriormente se rompe. La bicapa lipídica rota comprende una o varias Msp. Por lo tanto, la o las Msp están asociadas a lípidos. La Msp asociada a lípidos puede entonces ponerse en contacto con otros lípidos para formar otra bicapa lipídica que comprende la Msp asociada a lípidos. Como se usa en el presente documento, una bicapa lipídica es una membrana delgada que comprende moléculas lipídicas, por ejemplo, fosfolípidos, que se puede utilizar para insertar cualquier Msp proporcionada en el presente documento. Por lo tanto, en los métodos proporcionados en el presente documento, la Msp puede ponerse en contacto con fosfolípidos, ya sea como parte de una bicapa lipídica o no, para formar Msp asociada a lípidos.

Como se ha expuesto anteriormente, un experto en la materia puede determinar si una Msp se inserta en una bicapa, utilizando técnicas tales como las descritas en el ejemplo 2 de la publicación de patente de Estados Unidos N.º 20120055792. Todos los métodos de preparación y uso de porinas descritos en la publicación de patente de Estados Unidos N.º 20120055792 pueden emplearse para preparar y utilizar las porinas Msp descritas en el presente documento. Si la proteína se inserta en la bicapa, entonces la porina es una proteína formadora de túnel. Normalmente, la formación de túneles se detecta al observar un cambio discreto en la conductividad. Véanse, la publicación de patente de Estados Unidos N.º 20120055792 y Niederweis *et al.*, Mol. Microbiol. 33:933 (1999). El aumento en las inserciones de Msp puede ser un aumento de aproximadamente un 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 200 %, 300 %, 400 % o más.

Métodos de detección

Se proporciona además un método para detectar la presencia de un analito, que comprende: (a) aplicar un campo eléctrico suficiente para translocar un analito desde un primer medio conductor a un segundo medio conductor en comunicación de líquidos a través de cualquier Msp mutante descrita en el presente documento; y (b) medir una corriente de iones, en donde una reducción en la corriente de iones indica la presencia del analito en el primer medio.

- 5 Opcionalmente, los medios conductores de líquidos primero y segundo son iguales. Opcionalmente, los medios conductores de líquidos primero y segundo son diferentes. La porina Msp mutante puede ser cualquier porina Msp mutante analizada en el presente documento. Por ejemplo, la porina Msp puede ser una Msp de cadena sencilla mutante, una Msp mutante que comprende 2-15 monómeros o combinaciones de las mismas. Como se ha expuesto anteriormente, también se puede usar una pluralidad de Msp en los métodos descritos en el presente documento.

10

En los métodos divulgados en el presente documento, una Msp puede comprender además un motor molecular. El motor molecular puede ser capaz de mover un analito hacia o a través de un túnel con una velocidad de translocación o una velocidad de translocación promedio que es menor que la velocidad de translocación o la velocidad de translocación promedio a la que el analito se transloca electroforéticamente hacia o a través del túnel en ausencia del motor molecular. Por consiguiente, en cualquier realización en el presente documento que comprenda la aplicación de un campo eléctrico, el campo eléctrico puede ser suficiente para hacer que el analito se transloque electroforéticamente a través del túnel. Cualquier medio líquido analizado en el presente documento, tal como un medio líquido conductor, puede comprender un analito. En los métodos que comprenden medir una corriente de iones, el analito interactúa con un túnel de porina Msp para proporcionar un patrón de corriente, en donde la aparición de un

20

bloqueo en el patrón de corriente indica la presencia del analito.

Los métodos divulgados en el presente documento pueden comprender además identificar el analito. Por ejemplo, tales métodos pueden comprender comparar el patrón de corriente obtenido con respecto a un analito desconocido con el de un patrón de corriente conocido obtenido utilizando un analito conocido en las mismas condiciones. En otro

25

ejemplo, y sin pretender ser limitante, identificar el analito puede comprender (a) medir la corriente de iones para proporcionar un patrón de corriente, en donde una reducción en la corriente define un bloqueo en el patrón de corriente, y (b) comparar uno o más bloqueos en el patrón de corriente con (i) uno o más bloqueos en el patrón de corriente, o (ii) uno o más bloqueos en un patrón de corriente conocido obtenido utilizando un analito conocido.

30

El analito puede ser cualquier analito descrito en el presente documento. Por ejemplo, el analito puede ser un o unos nucleótido, un ácido nucleico, un o unos aminoácidos, un péptido, una proteína, un polímero, un fármaco, un ion, un contaminante, un objeto nanoscópico o un agente de guerra biológica. En los métodos proporcionados en el presente documento, opcionalmente, al menos uno de los medios líquidos conductores primero o segundo comprende una pluralidad de analitos diferentes.

35

En métodos donde el analito es un polímero, por ejemplo, una proteína, un péptido o un ácido nucleico, el método puede comprender además identificar una o más unidades del polímero. Por ejemplo, identificar una o más unidades del polímero puede comprender medir la corriente de iones para proporcionar un patrón de corriente que comprende un bloqueo para cada unidad de polímero, y comparar uno o más bloqueos en el patrón de corriente con (i) uno o más

40

bloqueos en el patrón de corriente o (ii) uno o más bloqueos en un patrón de corriente obtenido utilizando un polímero que tiene unidades conocidas. Estos métodos pueden comprender identificar unidades secuenciales del polímero, por ejemplo, y sin ser limitante, nucleótidos secuenciales o consecutivos en un ácido nucleico. En otro ejemplo, los aminoácidos secuenciales o consecutivos en un polipéptido pueden identificarse utilizando los métodos descritos en el presente documento.

45

Los métodos proporcionados en el presente documento pueden comprender distinguir al menos una primera unidad dentro de un polímero de al menos una segunda unidad dentro del polímero. La distinción puede comprender medir la corriente de iones producida cuando la primera y la segunda unidades se translocan por separado a través de un túnel para producir un primer y un segundo patrón de corriente, respectivamente, donde los patrones de corriente

50

primero y segundo difieren entre sí.

Los métodos proporcionados en el presente documento pueden comprender además secuenciar un polímero. La secuenciación puede comprender medir la corriente de iones o las señales ópticas a medida que cada unidad del polímero se transloca por separado a través del túnel para proporcionar un patrón de corriente que se asocia con cada

55

unidad, y comparar cada patrón de corriente con el patrón de corriente de una unidad conocida obtenida en la mismas condiciones, de manera que se secuencie el polímero.

Se proporciona además un método de secuenciación de ácidos nucleicos o polipéptidos utilizando cualquiera de las Msp mutantes proporcionadas en el presente documento. El método comprende crear una bicapa lipídica que comprende un primer y un segundo lado, añadir una Msp purificada al primer lado de la bicapa lipídica, aplicar un voltaje positivo al segundo lado de la bicapa lipídica, translocar una secuencia experimental de un ácido nucleico o polipéptido a través de la porina Msp, comparar la corriente de bloqueo experimental con un patrón de corriente de bloqueo y determinar la secuencia experimental.

60

65 Cualquiera de los métodos de detección proporcionados en el presente documento puede comprender además

determinar la concentración, tamaño, peso molecular, forma u orientación del analito o cualquier combinación de los mismos.

Como se usa en el presente documento, un polímero se refiere a una molécula que comprende dos o más unidades lineales (también conocidas como "meros"), donde cada unidad puede ser igual o diferente. Los ejemplos no limitantes de polímeros incluyen ácidos nucleicos, péptidos y proteínas, así como una variedad de polímeros de hidrocarburos (p. ej., polietileno, poliestireno) y polímeros de hidrocarburos funcionalizados, en donde la cadena principal del polímero comprende una cadena de carbono (p. ej., cloruro de polivinilo, polimetacrilatos). Los polímeros incluyen copolímeros, copolímeros en bloque y polímeros ramificados tales como polímeros estrella y dendrímeros.

En el presente documento se describen métodos de secuenciación de polímeros utilizando Msp. De manera adicional, los métodos de secuenciación se pueden realizar en métodos análogos a los descritos en la patente de Estados Unidos n.º 7.189.503. Véase también la patente de Estados Unidos n.º 6.015.714. Se puede realizar más de una lectura en dichos métodos de secuenciación para mejorar la precisión. Métodos para analizar las características de los polímeros (p. ej., tamaño, longitud, concentración, identidad) e identificación de unidades discretas (o "meros") también se analizan en la patente '503, y pueden emplearse con respecto a las Msp de la presente invención. De hecho, se puede emplear una Msp con respecto a cualquier método analizado en la patente '503.

En la actualidad, se pueden utilizar varios tipos de señales observables como mecanismos de lectura en la secuenciación por nanoporos y la detección de analitos. Un método de lectura ilustrativo se basa en una corriente de bloqueo iónico o en una corriente de paso conjunto, determinadas únicamente por la identidad de un nucleótido u otro analito que ocupa la constricción más estrecha en el poro. Este método se conoce como secuenciación con nanoporos por corriente de bloqueo o BCNS. Se ha demostrado la detección de corrientes de bloqueo y la caracterización de ácidos nucleicos tanto en la proteína de poros ahemolisina (aHL) como en nanoporos en estado sólido.

Se ha demostrado que la detección y caracterización de la corriente de bloqueo proporciona una gran cantidad de información sobre la estructura del ADN que pasa a través de o se mantiene en un nanoporo en varios contextos. En general, un bloqueo se evidencia por un cambio en la corriente de iones que se distingue claramente de las fluctuaciones de ruido y generalmente se asocia con la presencia de una molécula de analito en la abertura central del poro. La fuerza del bloqueo dependerá del tipo de analito presente. Más particularmente, un bloqueo se refiere a un intervalo en el que la corriente de iones desciende por debajo de un umbral de aproximadamente un 5-100 % del nivel de corriente sin bloqueo, permanece allí durante al menos 1,0 μ s y vuelve espontáneamente al nivel sin bloqueo. Por ejemplo, la corriente de iones puede descender por debajo de un umbral de aproximadamente, al menos aproximadamente o como máximo aproximadamente un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % o cualquier intervalo que pueda derivarse a partir de los mismos. Los bloqueos se rechazan si la señal sin bloqueo que la precede o la sigue tiene una corriente promedio que se desvía del nivel sin bloqueo habitual en más del doble del ruido rms de la señal sin bloqueo. Los bloqueos profundos se identifican como intervalos en los que la corriente de iones desciende <50 % del nivel sin bloqueo. Los intervalos en los que la corriente permanece entre el 80 % y el 50 % del nivel sin bloqueo se identifican como bloqueos parciales.

Se desvelan materiales, composiciones y componentes que se pueden usar para, se pueden usar junto con, se pueden usar en la preparación de, o son productos de los métodos y composiciones divulgados. Estos y otros materiales se divulgan en el presente documento, y se entiende que cuando las combinaciones, subconjuntos, interacciones, grupos, etc. de estos materiales se divulgan, aunque la referencia específica a cada una de las diversas combinaciones y permutaciones colectivas e individuales de estos compuestos pueden no divulgarse explícitamente, contempla cada una de ellas de manera específica y se describe en el presente documento. Por ejemplo, si un método se divulga y se analiza y se analiza una serie de modificaciones que pueden realizarse a un número de composiciones incluidas en el método, se contemplan de manera específica todas y cada una de las combinaciones y permutaciones del método y las modificaciones que sean posibles a menos que se indique específicamente lo contrario. Del mismo modo, cualquier subconjunto o combinación de estas también se contempla y se divulga de manera específica. Este concepto se aplica a todos los aspectos de la presente divulgación incluyendo, pero sin limitación, las etapas en los métodos. Por tanto, si existe una diversidad de etapas adicionales que puedan realizarse, se entiende que cada una de estas etapas adicionales pueden realizarse con cualquier etapa específica del método o combinación de etapas específicas de los métodos divulgados y que se contempla de manera específica cada una de dicha combinación o subconjunto de combinaciones y deberá considerarse divulgada. También se contempla que cualquier realización analizada en la presente memoria descriptiva se puede implementar con respecto a cualquier método, compuesto, proteína, porina, péptido, polipéptido, multímero, monómero, ácido nucleico, vector, cepa, célula cultivada, sistema o composición, etc., descritos en el presente documento y *viceversa*.

Se han descrito varias realizaciones. No obstante, se entenderá que se pueden realizar diversas modificaciones.

Ejemplos

Se construyó una MspA de cadena sencilla que consistía de ocho monómeros conectados covalentemente. Como se

muestra en el presente documento, un plásmido de expresión que codifica M1-MspA de cadena sencilla no es estable en *M. smegmatis* de tipo silvestre pero, se puede propagar en un *M. smegmatis* que carece del gen *recA*, que se necesita para la recombinación homóloga. El gen *recA* se eliminó en el cuádruple mutante de porina *M. smegmatis* ML712, que carece de las cuatro porinas Msp conocidas. Esta cepa permitió la expresión de M1-MspA de cadena sencilla. Los túneles realizados mediante M1-MspA de cadena sencilla tenían una conductancia similar a los túneles de M1-MspA octamérica, pero había mejorado drásticamente la resistencia a la activación por voltaje. Esta ventaja imprevista de MspA de cadena sencilla es de gran importancia para la secuenciación de ácidos nucleicos con nanoporos, por ejemplo, ADN.

10 **Construcción de un cuádruple mutante por eliminación de Msp de *M. smegmatis***

Para el aislamiento de porinas MspA mutantes, se utilizó una cepa ML16 de *Mycobacterium smegmatis* triple mutante con eliminación de porina ($\Delta mspA::FRT$, $\Delta mspC::FRT$, $\Delta mspD::FRT$) (véase Stephan *et al.*, Mol. Microbiol. 58: 714-730 (2005)). Sin embargo, todavía podían detectarse bajos niveles de MspB en esta cepa en inmunotransferencias con antisuero de conejo específico de MspA. La presencia de MspB puede contribuir a la heterogeneidad observada en experimentos de un solo canal y complica el análisis de datos. Para superar este problema y mejorar las preparaciones de MspA, se construyó una cepa de *M. smegmatis* que carecía de los cuatro genes *msp*.

Debido a que la eliminación de los cuatro genes *msp* de *M. smegmatis* es letal, la primera etapa fue integrar un casete de expresión para el dominio de formación de canales aminoterminal de CpnT (1) en el triple mutante de porina *M. smegmatis* ML16. Para este fin, se construyó el plásmido pML2622, que lleva el dominio aminoterminal de CpnT etiquetado con His6 y HA bajo el control de un promotor inducible por nitrilo (figura 2). El dominio aminoterminal de CpnT formó canales en experimentos de bicapa plana y complementó la cepa ML16 de *M. smegmatis* mutante de porina en experimentos de captación de glicerol. El plásmido de rescate pML2622 se integró en el sitio L5 del micobacteriófago en el cromosoma del triple mutante de porina *M. smegmatis* ML16. La integración se confirmó mediante PCR con los conjuntos de cebadores complementarios a la región genómica L5. Esta cepa se nombró *M. smegmatis* ML709. Después de la integración de pML2622, el esqueleto del plásmido se escindió del cromosoma mediante recombinasa Flp como se describe en Stephan *et al.* (Gene 343: 181-190 (2004)) para eliminar los genes que codifican la higromicina fosfotransferasa y la integrasa L5. Esta cepa se nombró *M. smegmatis* ML709-234.

Para eliminar el gen *mspB* restante en *M. smegmatis* ML711 se construyó el vector de eliminación de *mspB* pML1611 que contiene los dos genes indicadores *gfp* y *xylE* como marcadores para la integración y el reemplazo alélico (figura 2). El vector de eliminación de *mspB* pML1611 lleva 863 pb y 946 pb de longitud cadena arriba y cadena abajo de las regiones de *mspB* y se utilizó para eliminar el gen *mspB* en *M. smegmatis* ML709-234 mediante intercambio alélico. Este cuádruple mutante de porina Msp se nombró *M. smegmatis* ML711. El plásmido pCreSacB que contenía Cre recombinasa se utilizó para escindir los genes *gfp* e *hyg* flanqueados por sitios *loxP* del cromosoma. La eliminación de los cuatro genes *msp* se confirmaron mediante PCR con ADN cromosómico utilizando conjuntos específicos de cebadores y mediante transferencia Southern con ADN cromosómico utilizando una sonda específica (figura 3).

Este cuádruple mutante de porina Msp sin marcar se nombró *M. smegmatis* ML712 (genotipo relevante: $\Delta mspA::FRT$, $\Delta mspB::loxP$, $\Delta mspC::FRT$, $\Delta mspD::FRT$, $attB$ L5::FRT-pNIT-cpnT_{del}-FRT). El crecimiento de la cepa ML712 en placas de agar Middlebrook 7H10 se vio afectado en comparación con las cepas de ts y ML16. A continuación, se evaluó la expresión de los genes *msp* en la cepa ML712 mediante la extracción de células de *M. smegmatis* que utiliza el detergente óxido de octilpolietileno como se describe en Heinz *et al.* (Anal. Biochem. 285: 113-120 (2000)). El cuádruple mutante con eliminación de porina Msp *M. smegmatis* ML712 cultivado en medio Middlebrook 7H9 no produce ninguna proteína Msp en comparación con el triple mutante de porina ML16 (figura 4). Esto demostró que la eliminación de todos los genes *msp* se logró en *M. smegmatis* ML712. Los niveles de expresión de la MspA fueron similares a los de *M. smegmatis* de tipo silvestre, cuando se expresaron *mspA* de ts o M1 *mspA* en ML712 utilizando los plásmidos pMN016 y pML904, respectivamente (figura 4).

50 **Construcción de M1-MspA de cadena sencilla**

Previamente, se construyó un dímero de las subunidades M1₂-MspA (Pavlenok *et al.* PLoS One 7(6): e38726). Como siguiente etapa hacia la MspA de cadena sencilla, se fusionaron cuatro genes *mspA-M1* para codificar un tetrámero de subunidades M1₄-MspA. El plásmido resultante pML2647 se transformó en la cepa con cuádruple eliminación de porina de *M. smegmatis* ML712 para la producción y purificación de proteínas. Sin embargo, el plásmido con *mspA* tetramérica era inestable. Para evitar la recombinación, el gen *recA* se eliminó en el cuádruple mutante de porina ML712 y se creó la ML714 de *M. smegmatis*.

Posteriormente, un gen que codifica MspA de cadena sencilla, en el que se unen ocho subunidades M1-MspA (M1₈-MspA), se clonó en *E. coli*. Cada una de las subunidades tiene una mutación D90N, una mutación D91N y una mutación D93N. Esto se logró fusionando dos genes que codifican M1 MspA tetramérica (M1₄-MspA) juntos utilizando pML2647 como plantilla. Las subunidades individuales de MspA están separadas por enlazadores (GGGG)₃. En el plásmido pML3213 resultante, las dos construcciones con M1-*mspA* tetramérica están flanqueadas por sitios de restricción exclusivos (tétrada A: PacI, MluI; tétrada B: EcoRV, HindIII) (figura 5). Los genes en los tetrámeros están flanqueados

por los mismos sitios de restricción con la excepción de los genes primero y último de las tétradas.

El gen *m1₈-mspA* resultante se colocó bajo el control del promotor constitutivo *p_{smyc}* (figura 5). El plásmido pML3213 se transformó en *M. smegmatis* ML714 (cuádruple mutante de porina con eliminación de *recA*) para la producción y purificación de proteínas. Los experimentos de transferencia Western mostraron que el nivel de expresión de M1₈-MspA de cadena sencilla en *M. smegmatis* ML714 es más bajo en comparación con M1₂-MspA y se reduce a aproximadamente un 7 % de los niveles de MspA de ts (figura 6).

Estabilidad de M1-MspA de cadena sencilla

El poro de MspA es muy resistente a la desnaturalización térmica y química (Heinz *et al.* J. Biol. Chem. 278: 8678-8685 (2003)). Para probar la estabilidad térmica de MspA de cadena sencilla, la proteína M1₈-MspA se sometió a temperaturas crecientes durante 15 min en presencia de SDS al 2 %. (Figura 7) Una cantidad significativa de M1₈-MspA es estable incluso después de calentar la muestra de proteína a 100 °C (figura 7). Este resultado muestra que M1₈-MspA es al menos tan estable frente a la desnaturalización térmica como la proteína MspA de ts.

Propiedades de canal de M1-MspA de cadena sencilla

Para examinar si M1₈-MspA forma canales funcionales, se realizaron experimentos de bicapa lipídica *in vitro*. No se registraron inserciones cuando se añadió únicamente tampón n-octil-POE a la bicapa lipídica. La adición de aproximadamente 70 ng de proteína M1₈-MspA dio como resultado el aumento escalonado de la corriente a través de la bicapa lipídica que indica la inserción de canales de M1₈-MspA en la membrana (figura 8A). El análisis de los registros de corriente de M1₈-MspA mostró un pico principal de 1,1 nS (figura 8B). Esta conductancia de canal es similar a la de los poros formados por el dímero M1-MspA (1,3 nS) y los monómeros M1 MspA (1,4 nS) (véase Pavlenok *et al.*).

Activación por voltaje

La activación por voltaje se define como un cierre espontáneo del canal a un determinado umbral de voltaje y es una propiedad intrínseca de las proteínas de los canales de barril β bacterianos (Bainbridge *et al.* FEBS Lett 431 (3): 305-308 (1998)). La resistencia a la activación por voltaje es muy importante para los experimentos de secuenciación con nanoporos, ya que se utilizan voltajes de hasta +180 mV para translocar el ADNmc a través del poro de MspA (Manrao *et al.* Nat. Biotechnol. 30(4): 349-353 (2012); Derrington *et al.* Proc. Natl. Acad. Sci. USA 107(37): 16060-16065 (2010); Butler *et al.* Proc. Natl. Acad. Sci. USA 105(52): 20647-20652 (2008)). Por lo tanto, la activación por voltaje de M1₈-MspA se analizó en experimentos de bicapa lipídica. Después de la inserción de aproximadamente 220 poros de M1₈-MspA, el voltaje a través de la bicapa lipídica se incrementó secuencialmente en incrementos de 10 mV, y se midió la corriente de iones que pasa a través de los poros durante tres minutos. El voltaje crítico *V_c* se define como el voltaje al que los poros comienzan a cerrarse y se mide en estos experimentos como la disminución de la corriente de iones. Los canales de M1₈-MspA comenzaron a cerrarse a +90 mV y se mantuvieron completamente estables en todos los voltajes negativos aplicados (figura 9). En un segundo experimento con proteína M1₈-MspA purificada en gel no se observó activación por voltaje, hasta voltajes de ± 100 mV. Por tanto, el voltaje crítico *V_c* de M1₈-MspA es dos veces mayor que el de M1-MspA o M1₂-MspA (*V_{crítico}* +40 mV, - 50mV para ambas proteínas). Estos resultados muestran que la unión de las ocho subunidades en un único polipéptido aumentó drásticamente la resistencia de MspA de cadena sencilla a la activación por voltaje. Esta ventaja imprevista de MspA de cadena sencilla es de gran importancia para la secuenciación de ácidos nucleicos con nanoporos, por ejemplo, ADN.

Construcción de una MspA de cadena sencilla mutante (scMspA M2)

Como se describe en el presente documento, las mutaciones en MspA son útiles para mejorar sus interacciones con el ADN, sus propiedades de reconocimiento de bases y sus interacciones con membranas y proteínas accesorias como, por ejemplo, ADN polimerasa Phi29. Utilizando el enfoque descrito anteriormente para M1 MspA de cadena sencilla, se construyó una MspA de cadena sencilla mutante (MspA M2), en la que se unen entre sí ocho monómeros de MspA mutantes. La expresión de las construcciones de MspA M1 de cadena sencilla y MspA M2 de cadena sencilla en *M. smegmatis* ML712 se mostró mediante transferencias Western utilizando un anticuerpo contra MspA que demuestra que la producción de scMspA en *M. smegmatis* es factible. Como se muestra en el presente documento, las Msp de cadena sencilla se puede expresar en *E. coli*. La proteína M2 MspA de cadena sencilla (scMspA M2) se produce en cantidades de mg, pero no está plegada. Se ha creado un protocolo de plegado que permite el aislamiento de scMspA M2 activa.

Una *m2-mspA de cadena sencilla (scm2-mspA)* en la que ocho genes *m2-mspA* (que contienen las mutaciones D90N/D91N/D93N/D118R/D134R/E139K como se describe en Butler *et al.* (PNAS 105: 20647-20652 (2008)) se conectaron mediante fragmentos de ADN que codificaban enlazadores polipeptídicos (GGGS)₃. De manera adicional, cada gen estaba flanqueado por sitios de restricción únicos para permitir modificaciones específicas de cada subunidad de MspA. Los genes de la secuencia se denominan *m2-1* hasta *m2-8* comenzando desde el codón de inicio

ATG (figura 11 y tabla 3). Para la producción de proteínas y la purificación de la proteína MspA M2 de cadena sencilla en células de *E. coli* se eliminó el péptido señal de MspA. La secuencia de *scm2-mspA* tenía codones optimizados para una expresión óptima en *E. coli* y se sintetizó por GenScript. El gen *scm2-mspA* resultante estaba flanqueado por EcoRI y HindIII y se obtuvo en un plásmido pUC57 de GenScript. A continuación, el *scm2-mspA* completo se escindió y se clonó en el vector pET-21(a)+. El gen *scm2-mspA* está bajo el control del promotor de T7 en el plásmido pML3216 resultante (figura 11).

Para la producción y purificación de la proteína scMspA M2, el plásmido pML3216 se transformó en la cepa de *E. coli* BL21 (DE3)Omp8 que carece de 3 porinas principales (véase Prilipov *et al.* FEMSMicrobiol. Lett 163: 65-72 (1998)). Se eligió la cepa BL21(DE3)Omp8 para evitar la contaminación de scMspA M2 con porinas endógenas de *E. coli*. Después de la inducción de la expresión de *scm2-mspA* con 1,5 mM de células IPTG se cultivaron a 37 °C en medio LB complementado con ampicilina. Se observó la máxima expresión de la proteína diana dos horas después de la inducción, lo que representa aproximadamente un 4 % de la proteína total en el lisado celular (figura 12). Una banda de proteína correspondiente a scMspA M2 tenía una masa aparente de 170 kDa que es coherente con su masa molecular predicha de 165,6 kDa (figura 12). A continuación, se aisló y purificó scMspA M2 a partir de cuerpos de inclusión como se describe en Sambrook *et al.* (CSH Protocols 2006) Los cuerpos de inclusión que contenían predominantemente proteína scMspA M2 se solubilizaron en urea 8 M. Esta muestra se sometió posteriormente a cromatografía de intercambio aniónico utilizando una columna HiTrap QFF (GE HealthCare, Reino Unido) en presencia de urea 8 M. El perfil de elución de la proteína scMspA M2 fue muy similar al de MspA de ts publicado anteriormente (Heinz *et al.*, 2003). Es probable que esta proteína no esté plegada y no tenga actividad de canal.

Posteriormente, scMspA M2 se purificó y se sometió a un procedimiento de replegamiento. Después de la cromatografía de intercambio aniónico, se diluyó una fracción pura de scMspA M2 con una concentración de 50 µg/ml por un factor de 10 en un tampón que contenía NaCl 10 mM, HEPES 25 mM, L-arginina 0,6 M, LDAO al 0,1 % (v/v), pH 8,0 para alcanzar un volumen final de 1 ml. La mezcla se incubó durante la noche a temperatura ambiente (aproximadamente 21 °C) en un mezclador rotatorio. Posteriormente, la muestra se transfirió a un tubo de diálisis con MWCO de 3,5 kDa y se dializó frente a 2 l de un tampón que contenía NaCl 10 mM, HEPES 25 mM (pH 8,0), LDAO al 0,023 % (v/v) durante la noche a temperatura ambiente. La proteína dializada se transfirió a un microtubo y se incubó a temperatura ambiente durante un día más. A continuación, la eficacia de replegamiento se evaluó mediante análisis de transferencia Western utilizando antisuero de conejo específico de MspA. Después del procedimiento de replegamiento, la banda que reacciona con los anticuerpos policlonales contra MspA migró de 170 kDa a aproximadamente 130 kDa, lo que indica que se había producido el plegamiento de scMspA M2 a una forma más compacta con una mayor movilidad electroforética (figura 13). Se ha observado un cambio de movilidad electroforética de este tipo tras el plegamiento para las proteínas de la membrana externa de *E. coli* previamente. Sin embargo, no estaba claro si MspA mostraría un fenómeno similar.

Para examinar si scMspA M2 forma canales funcionales *in vitro* después del procedimiento de replegamiento se realizaron experimentos de bicapa lipídica. No se observó actividad de canal cuando solamente se añadió tampón LDAO al 0,023 % a la bicapa plana. Por el contrario, la adición de proteína scMspA M2 después de la etapa de replegamiento dio como resultado un aumento de corriente escalonado indicativo de inserciones de canales en la bicapa lipídica (figura 14). El análisis de las señales de corriente mostró una conductancia promedio de 2,3 nS (figura 14). El análisis de interés de MspA M2 elaborado a partir de monómeros mostró dos picos a 1,2 nS y 2,4 nS, lo que sugiere dos conformaciones proteicas diferentes. De manera adicional, un experimento multicanal con scMspA M2 mostró una mejor resistencia de activación por voltaje con un voltaje crítico de +80 mV/-70 mV (figura 15). El aumento de la resistencia al voltaje es beneficioso, por ejemplo, para experimentos de ADNmc realizados a voltajes relativamente altos.

Tabla 3. Sitios de restricción de scMspA M2

Número de SR	SR	Secuencia	Extremos	Aminoácidos	Gene flanqueado
1	EcoRI	GAATTC	cohesivos	EF	m2-1
2	KpnI	GGTACC	cohesivos	GT	
3	NsiI	ATGCAT	cohesivos	MH	m2-2
4	Seal	AGTACT	romos	ST	
5	NheI	GCTAGC	cohesivos	AS	m2-3
6	HpaI	GTTAAC	romos	VN	
7	XbaI	TCTAGA	cohesivos	SR	m2-4
8	NdeI	CATATG	cohesivos	HM	
9	EcoRV	GATATC	romos	DI	m2-5
10	PstI	CTGCAG	cohesivos	LQ	
11	BstBI	TTCGAA	cohesivos	FE	m2-6
12	BamHI	GGATCC	cohesivos	GS	
13	MluI	ACGCGT	cohesivos	TR	m2-7
14	PvuII	CAGCTG	romos	QL	
15	Afl II	CTTAAG	cohesivos	LK	

(continuación)

Tabla 3. Sitios de restricción de <i>scMspA M2</i>					
Número de SR	SR	Secuencia	Extremos	Aminoácidos	Gene flanqueado
16	HindIII	AAGCTT	cohesivos	KL	m2-8

Construcción de MspA de cadena sencilla mutante (MspA PN1)

- 5 Un gen *mspA pn1* (*scmspA PN1*) de cadena sencilla donde ocho genes *mspA* (que contienen una mutación P97F y mutaciones D90N/D91N/D93N/D118R/D134R/E139K como se describe en Butler *et al.* (PNAS 105: 20647-20652 (2008)) se conectaron mediante fragmentos de ADN que codifican enlazadores polipeptídicos (GGGGS)₃. De manera adicional, cada gen estaba flanqueado por sitios de restricción únicos para permitir modificaciones específicas de cada subunidad de MspA. Los genes en la secuencia se denominan *m2-97-1* hasta *m2-97-8* comenzando desde el codón de inicio ATG (figura 16 y tabla 3). Para la producción de proteínas y la purificación de la proteína MspA PN1 de cadena sencilla en células de *E. coli* se eliminó el péptido señal de MspA. La secuencia de *cmmspA PN1* tenía codones optimizados para una expresión óptima en *E. coli* y se sintetizó por GenScript. El gen *cmmspA PN1* resultante estaba flanqueado por EcoRI y HindIII y se obtuvo en un plásmido pUC57 de GenScript. A continuación, el *scmspA PN1* completo se escindió y se clonó en el vector pET-21(a)+. El gen *scmspA PN1* está bajo el control del promotor T7 en el plásmido pML3216 resultante (figura 16).

- Para la producción y purificación de la proteína scMspA PN1, el plásmido pML3216 se transformó en la cepa de *E. coli* BL21(DE3)Omp8 que carece de 3 porinas principales (véase Prilipov *et al.* FEMSMicrobiol. Lett 163: 65-72 (1998)). Se eligió la cepa BL21(DE3)Omp8 para evitar la contaminación de scMspA PN1 con porinas endógenas de *E. coli*. Después de la inducción de la expresión de *scmspA PN1* con IPTG 1,5 mM, las células se cultivaron a 37 °C en medio LB complementado con ampicilina. Se observó la máxima expresión de la proteína diana dos horas después de la inducción, lo que representa aproximadamente un 5 % de la proteína total en el lisado celular (figura 17). Una banda de proteína correspondiente a scMspA PN1 tenía una masa aparente de 170 kDa que es coherente con su masa molecular predicha de 165,6 kDa (figura 17). A continuación, se aisló y purificó scMspA PN1 a partir de cuerpos de inclusión como se describe en Sambrook *et al.* (CSH Protocols 2006) Los cuerpos de inclusión que contenían predominantemente proteína scMspA PN1 se solubilizaron en urea 8 M. Esta muestra se sometió posteriormente a cromatografía de intercambio aniónico utilizando una columna HiTrap QFF (GE Healthcare, Reino Unido) en presencia de urea 8 M. El perfil de elución de la proteína scMspA PN1 fue muy similar al de MspA de ts publicado anteriormente (Heinz *et al.*, 2003). Es probable que esta proteína no esté plegada y no tenga actividad de canal.

- Posteriormente, scMspA PN1 se purificó y se sometió a un procedimiento de replegamiento. Después de la cromatografía de intercambio aniónico, se dializó una fracción pura de scMspA PN1 frente a 2 l de tampón que contenía NaCl 140 mM, K₂HPO₄/KH₂PO₄ 10 mM, KCl 2 mM (pH 7,5) para eliminar la urea. La mezcla se incubó durante la noche a temperatura ambiente (aproximadamente 21 °C). Después de la diálisis, se añadieron L-arginina y LDAO a la muestra para alcanzar una concentración final de 0,6 M y 0,1 % (v/v), respectivamente. La muestra de proteína en el tampón de replegamiento (NaCl 140 mM, K₂HPO₄/KH₂PO₄ 10 mM, KCl 2 mM, L-arginina 0,6 M, LDAO al 0,1 % (v/v), pH 7,5) se incubó durante la noche en un agitador orbital (figura 18). Se calculó que la concentración de la muestra purificada era 1,37 mg/ml, determinada por absorbancia a 280 nm. El rendimiento de proteínas fue de 0,45 mg por 1 litro de cultivo bacteriano.

- Para probar el efecto de la fenilalanina en la posición 97 sobre la incorporación de la MspA de cadena sencilla en membranas lipídicas artificiales, se midió la actividad de inserción de diferentes construcciones de MspA monitorizando la liberación de colorante de carboxifluoresceína fluorescente a partir de los liposomas como se describe (véase Schwarz *et al.*, Biophys. J. 58(3):577-83 (1990); Schwarz *et al.*, Biochim. Biophys. Acta 1239(1): 51-7 (1995)). Brevemente, los liposomas DPhPC se prepararon por extrusión en presencia de carboxifluoresceína 30 mM. La carboxifluoresceína se desactiva automáticamente cuando se encierra en vesículas lipídicas. Después de la inserción del poro de MspA en el liposoma cargado de colorante, el flujo de salida del colorante mediado por difusión da como resultado el aumento de la fluorescencia en la mezcla de reacción. La figura 19 muestra los resultados de estos experimentos de liberación de carboxifluoresceína. La adición de tampones que contienen LDAO (0,1 % v/v) u OPOE (0,5 % v/v) dio como resultado una liberación mínima de colorante de los liposomas, a diferencia del tampón Triton X-100 (1 % v/v) que se utilizó como control positivo. Es importante destacar que, la adición de scMspA PN1 (60 ng/ml, final) da lugar a una liberación más rápida y mayor de carboxifluoresceína que la adición de scMspA M2 (120 ng/ml, final). Cabe destacar que, MspA de ts (60 ng/ml, final) dio como resultado una difusión de colorante más lenta desde los liposomas que scMspA PN1 (figura 19). Estos datos indican que las fenilalaninas adicionales ubicadas en el bucle 6 de scMspA promueven una inserción más rápida y eficaz de los poros en las bicapas lipídicas.

- A continuación, se midió el tiempo de la primera inserción de poros en la membrana de DPhPC en una configuración de bicapa. Se planteó la hipótesis de que los poros con capacidades de inserción potenciadas requerirían menos tiempo para insertarse en la membrana lipídica. Para examinar el efecto de las fenilalaninas en el bucle 6 en el momento de la inserción en la membrana de scMspA, scMspA PN1 se comparó con scMspA M2. Brevemente, la cubeta bicapa se llenó de electrolito, se aplicó un potencial de -10 mV y los datos se adquirieron y registraron utilizando el programa informático TestPoint. Siempre se utilizó la misma cubeta en estos experimentos. La proteína se añadió

a ambos lados a una concentración final de 100 ng/ml. Es importante destacar que, se observaron eventos de inserción exitosos en el 89 % de los experimentos para scMspA PN1, pero solamente en el 40 % de los experimentos para scMspA M2. Esto es coherente con los resultados de los experimentos de liberación de carboxifluoresceína. Aunque el tiempo medio de inserción para scMspA PN1 fue de 399 segundos en comparación con los 695 segundos para scMspA M2, esta diferencia no fue significativa. Sorprendentemente, la tasa de inserción disminuyó cuando se analizó scMspA PN1 en una solución de KCl 0,3 M (tiempo medio: 859 segundos, 50 %). Sin embargo, la mitad de los experimentos dieron como resultado inserciones exitosas con scMspA PN1, mientras que solo se observó una inserción exitosa con scMspA M2 en KCl 0,3 M con un tiempo de 1270 segundos (8 membranas analizadas, 12 % de inserciones exitosas) (figura 20). Este resultado muestra el efecto beneficioso de las fenilalaninas en el bucle 6 para la inserción de la membrana por MspA de cadena sencilla.

Para examinar si scMspA PN1 forma canales funcionales *in vitro* después del procedimiento de replegamiento, se realizaron experimentos con bicapas lipídicas. No se observó actividad de canal cuando solamente se añadió tampón LDAO al 0,1 % a la bicapa plana. Por el contrario, la adición de proteína scMspA PN1 después de la etapa de replegamiento dio como resultado un aumento de corriente escalonado indicativo de inserciones de canales en la bicapa lipídica (figura 21). El análisis de las señales de corriente mostró una conductancia promedio de 2,0 nS (figura 21). Esto podría traducirse en corrientes residuales más grandes para cada nucleótido y mejores señales en los experimentos de secuenciación de ADN.

20 **Efecto de los lípidos sobre la actividad de canal de MspA PN1 de cadena sencilla**

Se almacenó scMspA PN1 durante más de un mes a temperatura ambiente en cantidades de 1 µg/mg y 0,2 µg/ml. La scMspA PN1 se diluyó en LDAO al 0,1 %, NaCl 140 mM, K₂HPO₄/KH₂PO₄ 10 mM (pH 7,5), KCl 2 mM. Se conocen en la técnica métodos para fabricar bicapas horizontales para experimentos de canal. Véanse, por ejemplo, Butler *et al.* (2008) y Akeson *et al.* Biophysical Journal; 77: 3227-3233 (1999), ambos incorporados en el presente documento en su totalidad. Para los experimentos de canal, se utilizó diftanoilfosfatidilcolina (DiphPC) al 2 % en cloroformo para formar bicapas de membrana para la inserción de MspA esencialmente como se describe en Butler *et al.* y Akeson *et al.* Después de la inserción de la MspA en las bicapas, la membrana se rompió y la membrana se volvió a aplicar utilizando DiphPC al 1 % en n-decano. El electrolito utilizado en estos experimentos fue KCl 0,3 o 1 M, Hepes 10 mM, pH 8,0 o pH 7,0, respectivamente.

Como se muestra en la figura 22, las MspA de cadena sencilla funcionan en un amplio intervalo de concentración de electrolitos, por ejemplo, KCl de aproximadamente 0,3-1 M. Para optimizar la actividad de canal, la asociación de lípidos se puede realizar antes de la inserción de MspA en una membrana o bicapa lipídica. Por lo tanto, en cualquiera de los métodos expuestos en el presente documento, una MspA puede ponerse en contacto o incubarse previamente con uno o más lípidos para optimizar la actividad de canal. En un ejemplo no limitante, la figura 22 muestra que no se observó actividad de canal en un tampón que contenía solamente KCl 0,3 M a pH 8,0. Sin embargo, la rotura de la membrana y el posterior repintado de la membrana da lugar a un aumento de la actividad de canal de scMspA PN1 en el electrolito que contiene KCl 0,3 M a pH 8,0.

40 LISTADO DE SECUENCIAS

<110> UAB Research Foundation Niederweis, Michael Pavlenok, Mikhail

45 <120> NANOPOROS DE MSP Y USOS DE LOS MISMOS

<130> 35979-941614 (16WO1)

<150> US 61/980,415

50 <151> 16/04/2014

<150> US 61/980,393

<151> 16/04/2014

55 <160> 10

<170> PatentIn versión 3.5

<210> 1

60 <211> 184

<212> PRT

<213> *Mycobacterium smegmatis*

<400> 1

65

ES 2 895 625 T3

Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu
 1 5 10 15
 Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp
 20 25 30
 Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr
 35 40 45
 Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu
 50 55 60
 Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe
 65 70 75 80
 Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Ala
 85 90 95
 Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly
 100 105 110
 Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val
 115 120 125
 Ala Thr Phe Ser Val Asp Val Ser Gly Ala Glu Gly Gly Val Ala Val
 130 135 140
 Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu
 145 150 155 160
 Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr
 165 170 175
 Tyr Gly Glu Pro Trp Asn Met Asn
 180

<210> 2
 <211> 184
 <212> PRT
 <213> *Mycobacterium smegmatis*
 <400> 2

ES 2 895 625 T3

Gly	Leu	Asp	Asn	Glu	Leu	Ser	Leu	Val	Asp	Gly	Gln	Asp	Arg	Thr	Leu	1	5	10	15
Thr	Val	Gln	Gln	Trp	Asp	Thr	Phe	Leu	Asn	Gly	Val	Phe	Pro	Leu	Asp	20	25	30	
Arg	Asn	Arg	Leu	Thr	Arg	Glu	Trp	Phe	His	Ser	Gly	Arg	Ala	Lys	Tyr	35	40	45	
Ile	Val	Ala	Gly	Pro	Gly	Ala	Asp	Glu	Phe	Glu	Gly	Thr	Leu	Glu	Leu	50	55	60	
Gly	Tyr	Gln	Ile	Gly	Phe	Pro	Trp	Ser	Leu	Gly	Val	Gly	Ile	Asn	Phe	65	70	75	80
Ser	Tyr	Thr	Thr	Pro	Asn	Ile	Leu	Ile	Asp	Asp	Gly	Asp	Ile	Thr	Ala	85	90	95	
Pro	Pro	Phe	Gly	Leu	Asn	Ser	Val	Ile	Thr	Pro	Asn	Leu	Phe	Pro	Gly	100	105	110	
Val	Ser	Ile	Ser	Ala	Asp	Leu	Gly	Asn	Gly	Pro	Gly	Ile	Gln	Glu	Val	115	120	125	
Ala	Thr	Phe	Ser	Val	Asp	Val	Ser	Gly	Pro	Ala	Gly	Gly	Val	Ala	Val	130	135	140	
Ser	Asn	Ala	His	Gly	Thr	Val	Thr	Gly	Ala	Ala	Gly	Gly	Val	Leu	Leu	145	150	155	160
Arg	Pro	Phe	Ala	Arg	Leu	Ile	Ala	Ser	Thr	Gly	Asp	Ser	Val	Thr	Thr	165	170	175	
-																			
Tyr Gly Glu Pro Trp Asn Met Asn																			
180																			

<210> 3
 <211> 184
 5 <212> PRT
 <213> *Mycobacterium smegmatis*
 <400> 3

ES 2 895 625 T3

Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu
 1 5 10 15
 Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp
 20 25 30
 Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr
 35 40 45
 Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu
 50 55 60
 Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe
 65 70 75 80
 Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Gly
 85 90 95
 Pro Pro Phe Gly Leu Glu Ser Val Ile Thr Pro Asn Leu Phe Pro Gly
 100 105 110
 Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val
 115 120 125
 Ala Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val
 130 135 140
 Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu
 145 150 155 160
 Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr
 165 170 175
 Tyr Gly Glu Pro Trp Asn Met Asn
 180

<210> 4
 <211> 176
 <212> PRT
 <213> *Mycobacterium smegmatis*
 <400> 4

ES 2 895 625 T3

Val Asp Gly Gln Gly Arg Thr Leu Thr Val Gln Gln Ala Glu Thr Phe
1 5 10 15

Leu Asn Gly Val Phe Pro Leu Asp Arg Asn Arg Leu Thr Arg Glu Trp
20 25 30

Phe His Ser Gly Arg Ala Thr Tyr His Val Ala Gly Pro Gly Ala Asp
35 40 45

Glu Phe Glu Gly Thr Leu Glu Leu Gly Tyr Gln Val Gly Phe Pro Trp
50 55 60

Ser Leu Gly Val Gly Ile Asn Phe Ser Tyr Thr Thr Pro Asn Ile Leu
65 70 75 80

Ile Asp Gly Gly Asp Ile Thr Gln Pro Pro Phe Gly Leu Asp Thr Ile
85 90 95

Ile Thr Pro Asn Leu Phe Pro Gly Val Ser Ile Ser Ala Asp Leu Gly
100 105 110

Asn Gly Pro Gly Ile Gln Glu Val Ala Thr Phe Ser Val Asp Val Lys
115 120 125

Gly Ala Lys Gly Ala Val Ala Val Ser Asn Ala His Gly Thr Val Thr
130 135 140

Gly Ala Ala Gly Gly Val Leu Leu Arg Pro Phe Ala Arg Leu Ile Ala
145 150 155 160

Ser Thr Gly Asp Ser Val Thr Thr Tyr Gly Glu Pro Trp Asn Met Asn
165 170 175

<210> 5
<211> 15
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 5

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
1 5 10 15

<210> 6
<211> 211
<212> PRT
<213> *Mycobacterium smegmatis*

ES 2 895 625 T3

<400> 6

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

5

<210> 7
 <211> 215
 <212> PRT

<213> *Mycobacterium smegmatis*

<400> 7

```

Met Thr Ala Phe Lys Arg Val Leu Ile Ala Met Ile Ser Ala Leu Leu
 1          5          10          15

Ala Gly Thr Thr Gly Met Phe Val Ser Ala Gly Ala Ala His Ala Gly
 20          25          30

Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu Thr
 35          40          45

Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp Arg
 50          55          60

Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr Ile
 65          70          75          80

Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu Gly
 85          90          95

Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe Ser
 100          105          110

Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Ala Pro
 115          120          125

Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly Val
 130          135          140

Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val Ala
 145          150          155          160

Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val Ser
 165          170          175

Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu Arg
 180          185          190

Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr Tyr
 195          200          205

Gly Glu Pro Trp Asn Met Asn
 210          215

```

5

<210> 8

<211> 215
 <212> PRT
 <213> *Mycobacterium smegmatis*

5 <400> 8

```

Met Lys Ala Ile Ser Arg Val Leu Ile Ala Met Ile Ser Ala Leu Ala
 1          5          10          15

Ala Ala Val Ala Gly Leu Phe Val Ser Ala Gly Thr Ser His Ala Gly
 20          25          30

Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu Thr
 35          40          45

Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp Arg
 50          55          60

Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr Ile
 65          70          75          80

Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu Gly
 85          90          95

Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe Ser
100          105          110

Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Gly Pro
115          120          125

Pro Phe Gly Leu Glu Ser Val Ile Thr Pro Asn Leu Phe Pro Gly Val
130          135          140

Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val Ala
145          150          155          160

Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val Ser
165          170          175

Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu Arg
180          185          190

Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr Tyr
195          200          205

Gly Glu Pro Trp Asn Met Asn
210          215
    
```

10 <210> 9
 <211> 207

<212> PRT

<213> *Mycobacterium smegmatis*

<400> 9

5

```

Met Arg Tyr Leu Val Met Met Phe Ala Leu Leu Val Ser Val Thr Leu
1          5          10          15

Val Ser Pro Arg Pro Ala Asn Ala Val Asp Asn Gln Leu Ser Val Val
          20          25          30

Asp Gly Gln Gly Arg Thr Leu Thr Val Gln Gln Ala Glu Thr Phe Leu
          35          40          45

Asn Gly Val Phe Pro Leu Asp Arg Asn Arg Leu Thr Arg Glu Trp Phe
          50          55          60

His Ser Gly Arg Ala Thr Tyr His Val Ala Gly Pro Gly Ala Asp Glu
65          70          75          80

Phe Glu Gly Thr Leu Glu Leu Gly Tyr Gln Val Gly Phe Pro Trp Ser
          85          90          95

Leu Gly Val Gly Ile Asn Phe Ser Tyr Thr Thr Pro Asn Ile Leu Ile
          100          105          110

Asp Gly Gly Asp Ile Thr Gln Pro Pro Phe Gly Leu Asp Thr Ile Ile
          115          120          125

Thr Pro Asn Leu Phe Pro Gly Val Ser Ile Ser Ala Asp Leu Gly Asn
          130          135          140

Gly Pro Gly Ile Gln Glu Val Ala Thr Phe Ser Val Asp Val Lys Gly
145          150          155          160

Ala Lys Gly Ala Val Ala Val Ser Asn Ala His Gly Thr Val Thr Gly
          165          170          175

Ala Ala Gly Gly Val Leu Leu Arg Pro Phe Ala Arg Leu Ile Ala Ser
          180          185          190

Thr Gly Asp Ser Val Thr Thr Tyr Gly Glu Pro Trp Asn Met Asn
          195          200          205

```

<210> 10

<211> 549

<212> ADN

<213> *Mycobacterium smegmatis*

10

ES 2 895 625 T3

<400> 10

ggcctggaca acgagctgag cctcgttgat ggccaggacc gcaccctcac cgtgcagcag	60
tgggacacct tcctcaatgg tgtgttcccc ctggaccgca accgtcttac ccgtgagtgg	120
ttccactccg gtcgcgcaa gtacatcgtg gccggccccg gtgccgacga gttcgagggc	180
acgctggaac tcggctacca gatcggcttc ccgtggtcgc tgggtgtggg catcaacttc	240
agctacacca cccgaacat cctgatcgac gacggtgaca tcaccgctcc gccgttcggc	300
ctgaactcgg tcatcacccc gaacctgttc cccggtgtgt cgatctcggc agatctgggc	360
aacggccccg gcatccagga agtcgcaacg ttctcggtcg acgtctccgg cggcgagggc	420
ggcgtggccg tgtcgaacgc ccacggcacc gtgaccggtg cggccggcgg tgtgctgctg	480
cgtccgttcg cccgcctgat cgcctcgacc ggtgactcgg tcaccaccta cggcgaaccc	540
tggaacatg	549

REIVINDICACIONES

1. Ácido nucleico que codifica un monómero de porina de *Mycobacterium smegmatis* (Msp) mutante, en donde el monómero de Msp comprende una mutación P97F.
- 5 2. Secuencia de ácido nucleico que codifica una porina de *Mycobacterium smegmatis* (Msp) de cadena sencilla mutante, en donde la secuencia de ácido nucleico comprende:
 - (a) una primera y segunda secuencia de nucleótidos, en donde la primera secuencia de nucleótidos codifica una primera secuencia de monómero de Msp y la segunda secuencia de nucleótidos codifica una segunda secuencia de monómero de Msp; y
 - (b) una tercera secuencia de nucleótidos que codifica una secuencia engarce de aminoácidos, en donde al menos una de la primera y segunda secuencias de monómero de Msp está codificada por el ácido nucleico de la reivindicación 1.
- 10 3. Una secuencia de ácido nucleico que codifica una Msp de cadena sencilla mutante, en donde la secuencia de ácido nucleico comprende:
 - (a) una primera, segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencia de nucleótidos o cualquier subconjunto de las mismas, en donde la primera, segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencias de nucleótidos codifican una primera, segunda, tercera, cuarta, quinta, sexta, séptima y octava secuencias de monómeros de Msp, respectivamente; y
 - (b) una novena secuencia de nucleótidos que codifica una secuencia engarce de aminoácidos, en donde la primera secuencia de monómero de Msp está codificada por el ácido nucleico de la reivindicación 1.
- 15 4. El ácido nucleico de cualquiera de las reivindicaciones 1 a 3, en donde la secuencia de monómero de Msp mutante comprende además (i) una mutación en una o más posiciones de aminoácidos D118, D134 o E139, (ii) una mutación en la posición 93, y/o (iii) una mutación en la posición 90, posición 91 o ambas posiciones 90 y 91, en donde opcionalmente el aminoácido en la posición 91 o el aminoácido en la posición 90 está sustituido con arginina, lisina, histidina, glutamina, metionina, treonina, fenilalanina, tirosina o triptófano.
- 20 5. Una Msp mutante o un monómero de Msp mutante codificado por cualquiera de los ácidos nucleicos de cualquiera de las reivindicaciones 1-4.
- 25 6. Un vector que comprende el ácido nucleico de cualquiera de las reivindicaciones 1-4.
- 30 7. Un sistema que comprende una Msp de cadena sencilla codificada por cualquiera de los ácidos nucleicos de cualquiera de las reivindicaciones 2 y 3, o un sistema que comprende la Msp mutante de la reivindicación 5, en donde la Msp tiene un vestíbulo y una zona de constricción que definen un túnel, en donde el túnel se coloca entre un primer medio líquido conductor y un segundo medio líquido conductor, en donde al menos un medio líquido conductor comprende un analito y en donde el sistema está operativo para detectar el analito, cuando el sistema se somete a un campo eléctrico suficiente para translocar el analito de un medio líquido conductor al otro.
- 35 8. Un método para detectar la presencia de un analito, que comprende:
 - a) aplicar un campo eléctrico suficiente para translocar un analito desde un primer medio conductor a un segundo medio conductor en comunicación de líquidos a través de la Msp mutante de la reivindicación 5; y
 - b) medir una corriente de iones, en donde una reducción en la corriente de iones indica la presencia del analito en el primer medio.
- 40 9. Un método para aumentar el número de inserciones de Msp en una bicapa lipídica, que comprende
 - (a) poner en contacto la Msp de la reivindicación 5 con un lípido para formar una Msp asociada a lípidos, en donde la etapa de poner en contacto comprende opcionalmente insertar la Msp en una bicapa lipídica y romper la bicapa lipídica para formar una Msp asociada a lípidos; y
 - (b) insertar la Msp asociada a lípidos de la etapa (a) en una bicapa lipídica.
- 45 50 55

	I										
		*	*	*	*	*	*	*	*	*	
MspA	MKAISRVLIAMV----	AAIAALETSTGTSHAGLDNELSLVDGQDR	TLTVQQWDTFLNGVFPLDRNRLTREW								40
MspB	MAFRRVLIAMISALLAGTTGMEV	SACAAHAGLDNELSLVDGQDR	TLTVQQWDTFLNGVFPLDRNRLTREW								40
MspC	MKAISRVLIAMISALAAAVAGLEV	SAGTSHAGLDNELSLVDGQDR	TLTVQQWDTFLNGVFPLDRNRLTREW								40
MspD	----MFMVLMF----	ALLAVSVTLAVSPRPANA-VDNQLS	VVDGQGR	TLTVQQAETFLNGVFPLDRNRLTREW							39
	*	*	*	*	*	*	*	*	*	*	
MspA	FHSGRAKYIVAGPGADEFE	GTLELGYQIGFPWSLGVGINESY	TTPNILIDDG	DITAPPEGINSVITPNLF							110
MspB	FHSGRAKYIVAGPGADEFE	GTLELGYQIGFPWSLGVGINESY	TTPNILIDDG	DITAPPEGINSVITPNLF							110
MspC	FHSGRAKYIVAGPGADEFE	GTLELGYQIGFPWSLGVGINESY	TTPNILIDDG	DITAPPEGINSVITPNLF							110
MspD	FHSGRAKYIVAGPGADEFE	GTLELGYQIGFPWSLGVGINESY	TTPNILIDDG	DITAPPEGINSVITPNLF							109
	*	*	*	*	*	*	*	*	*	*	
MspA	PGVSIADLGNPGCIQEVATF	SVDSGAEGGVAVSNAHGT	VTGAAGVLLRPFAR	LIASTGDSVTTYGEPWNN							184
MspB	PGVSIADLGNPGCIQEVATF	SVDSGAEGGVAVSNAHGT	VTGAAGVLLRPFAR	LIASTGDSVTTYGEPWNN							184
MspC	PGVSIADLGNPGCIQEVATF	SVDSGAEGGVAVSNAHGT	VTGAAGVLLRPFAR	LIASTGDSVTTYGEPWNN							184
MspD	PGVSIADLGNPGCIQEVATF	SVDSGAEGGVAVSNAHGT	VTGAAGVLLRPFAR	LIASTGDSVTTYGEPWNN							183

FIGURA 1

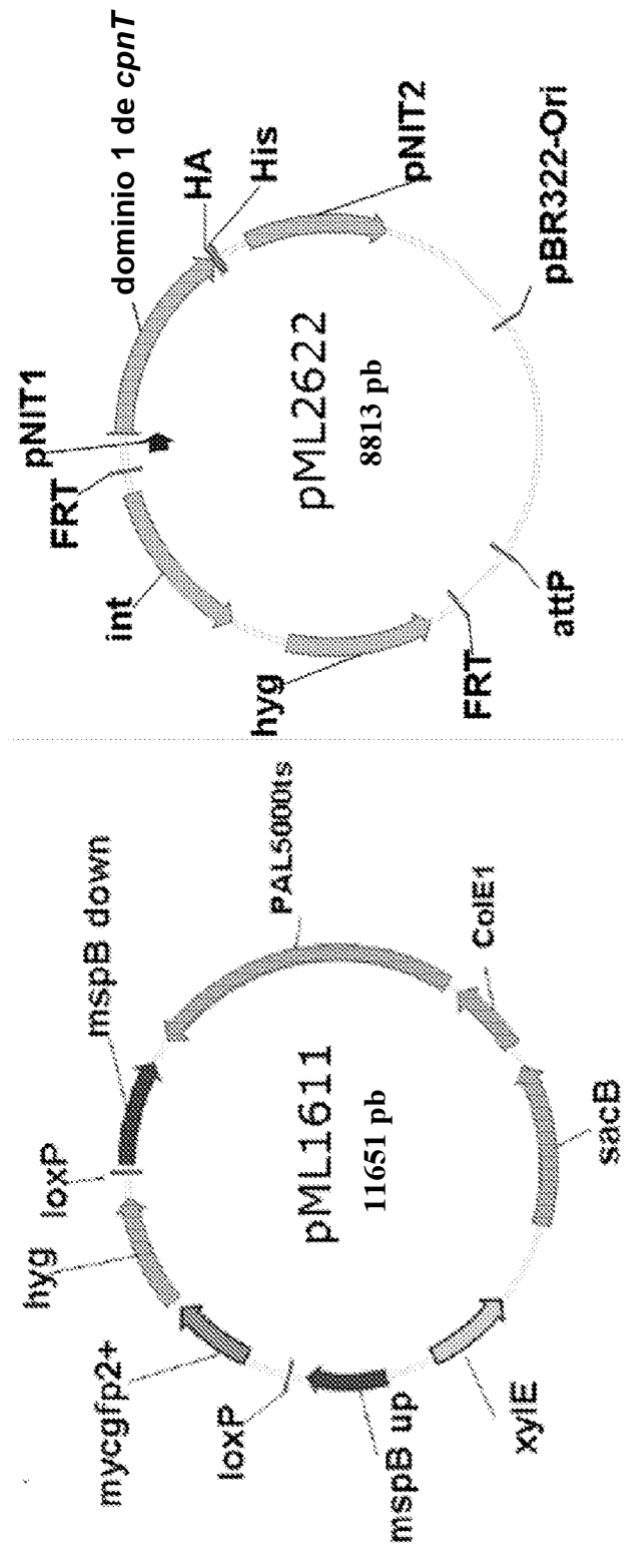


FIGURA 2

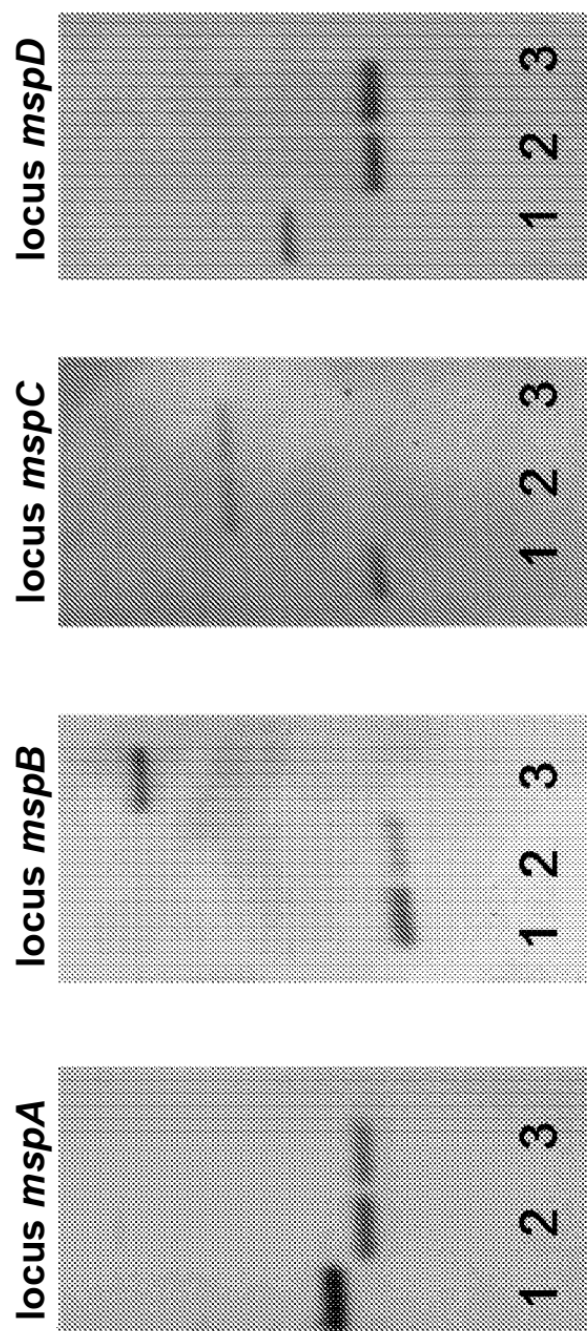


FIGURA 3

1 2 3 4 5

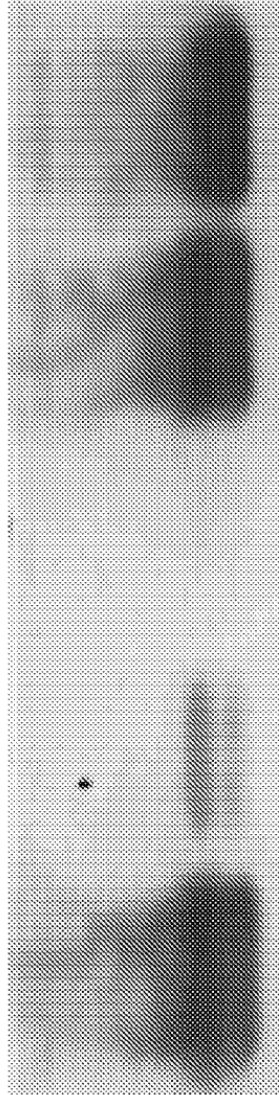


FIGURA 4

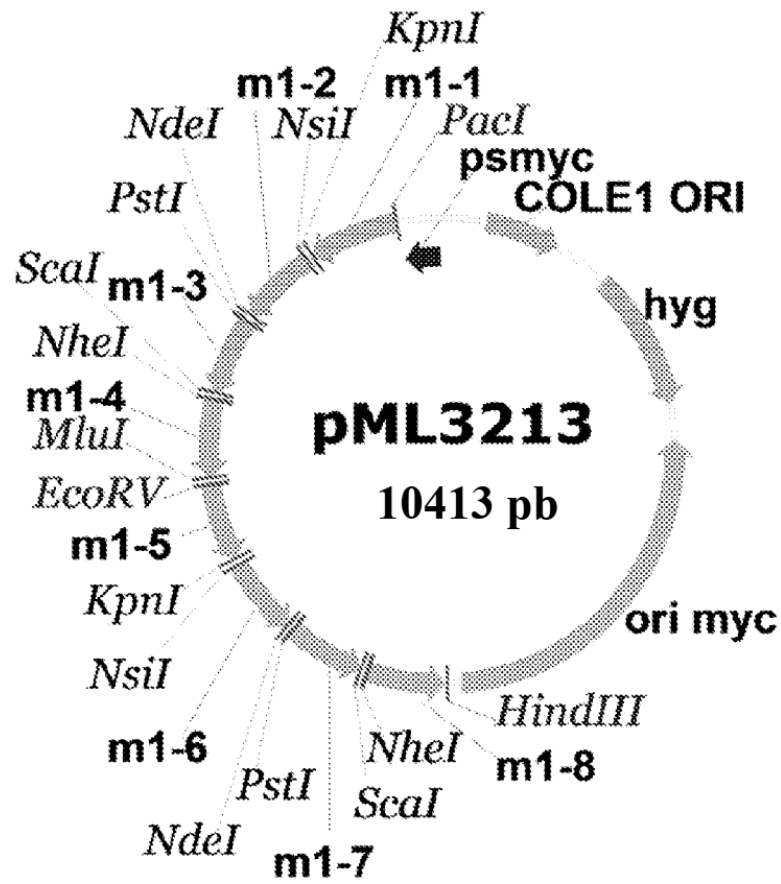


FIGURA 5

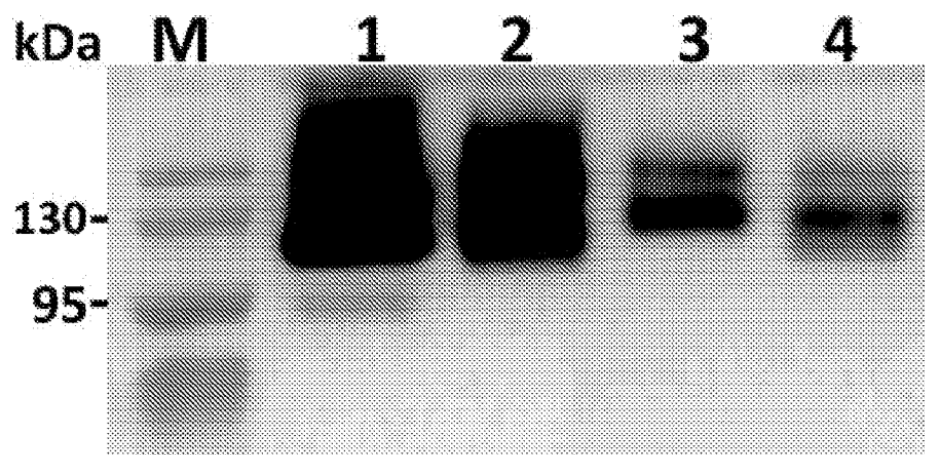


FIGURA 6

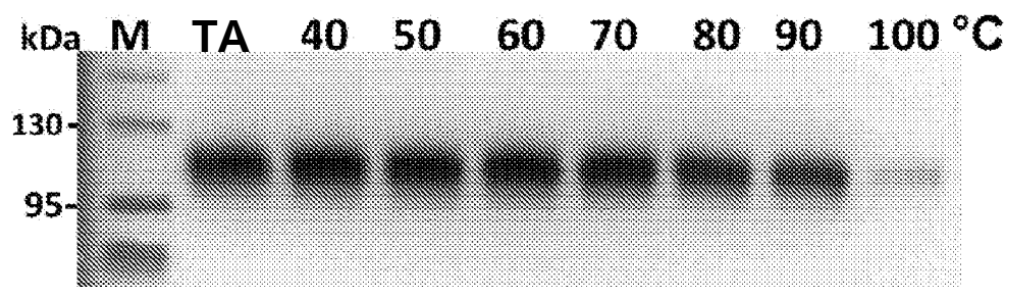


FIGURA 7

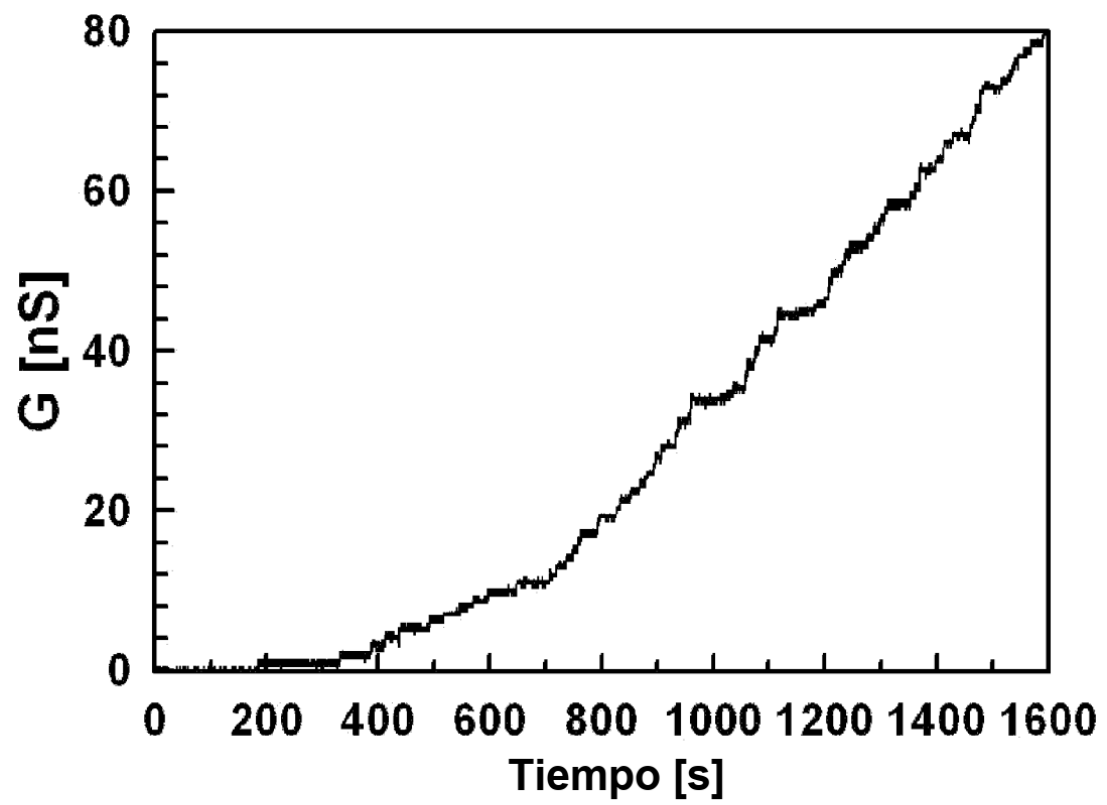


FIGURA 8A

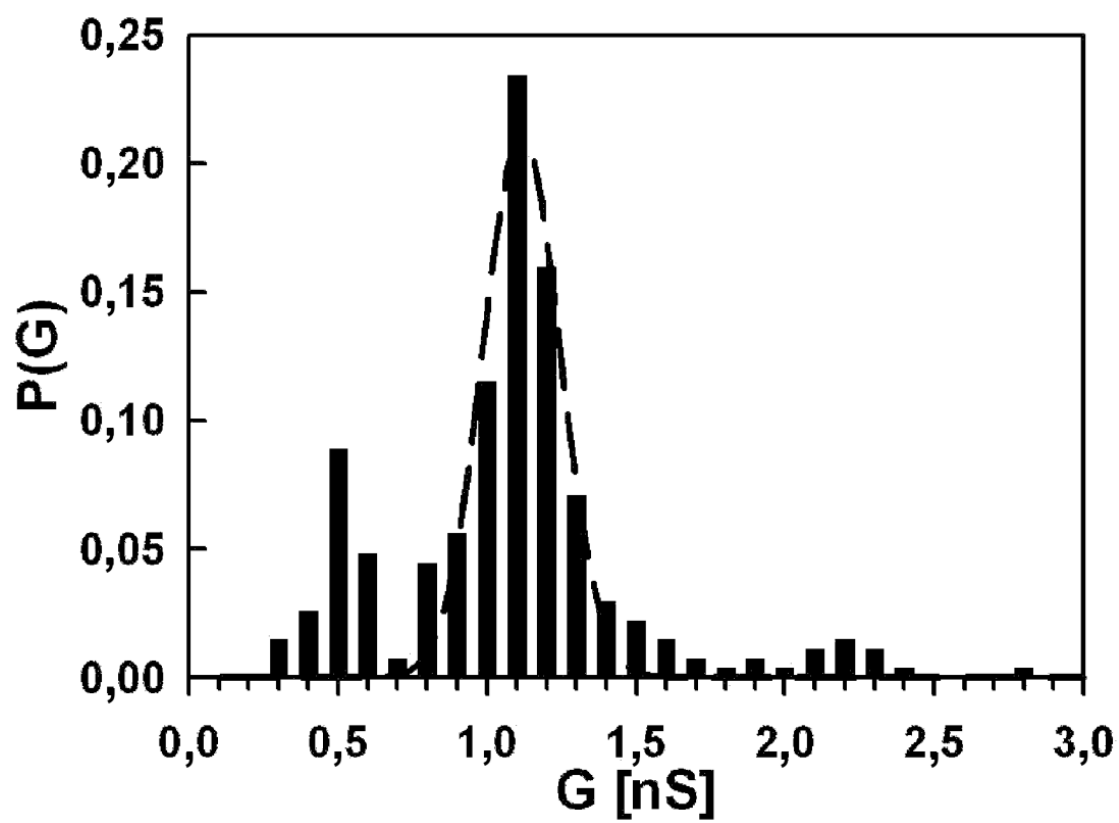


FIGURA 8B

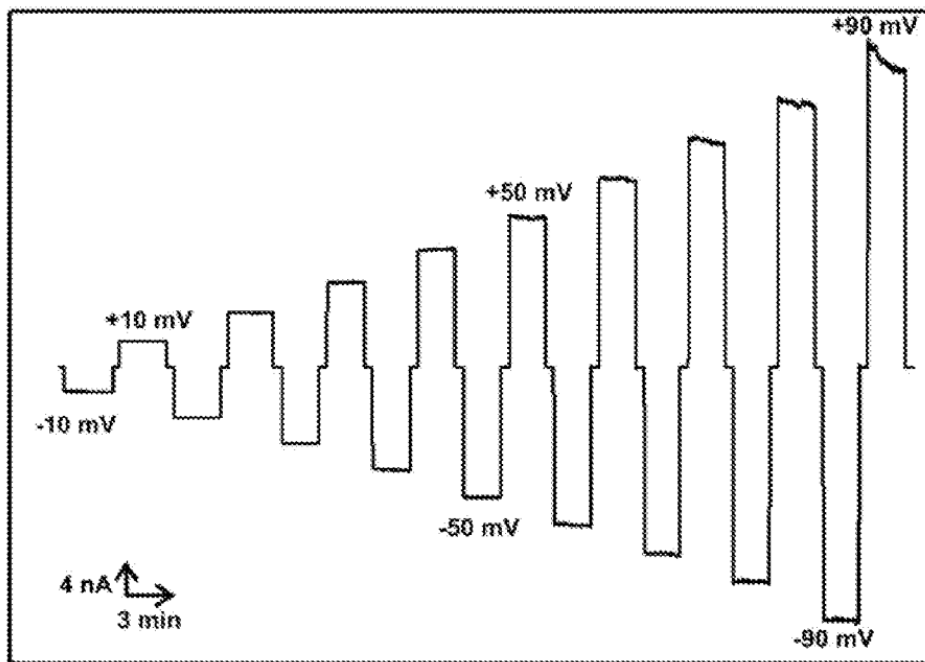


FIGURA 9

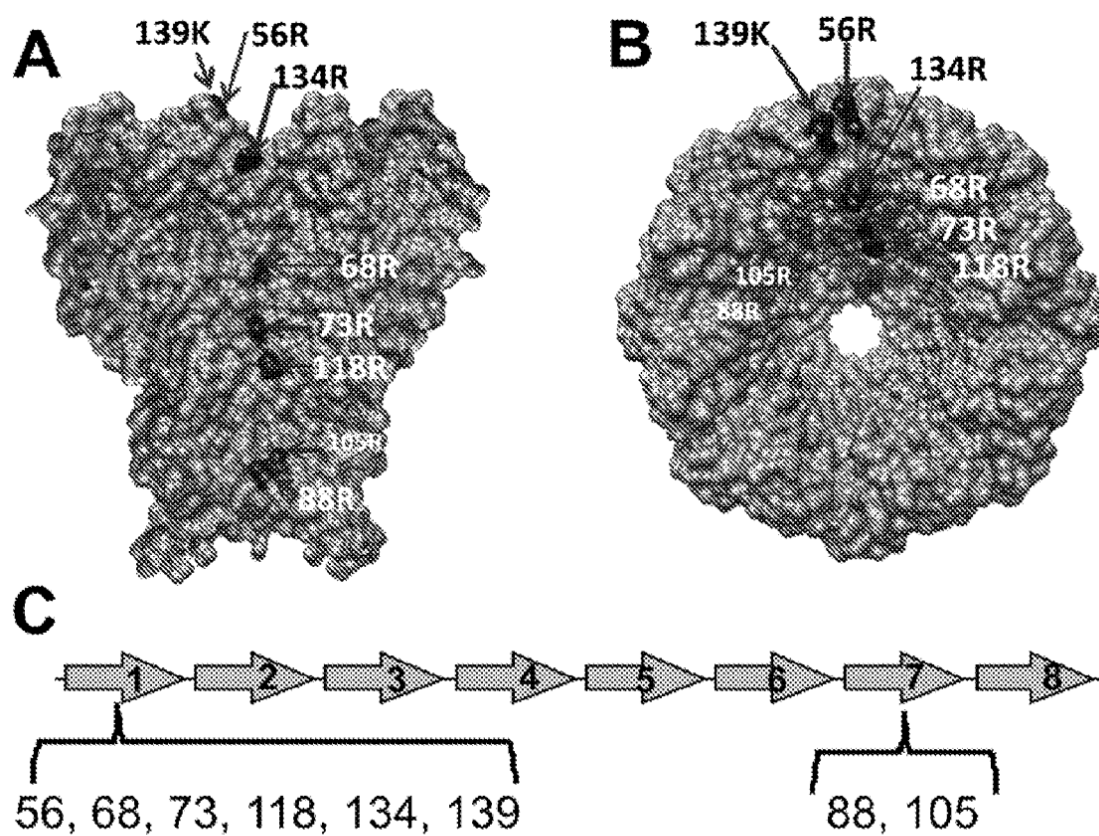


FIGURA 10

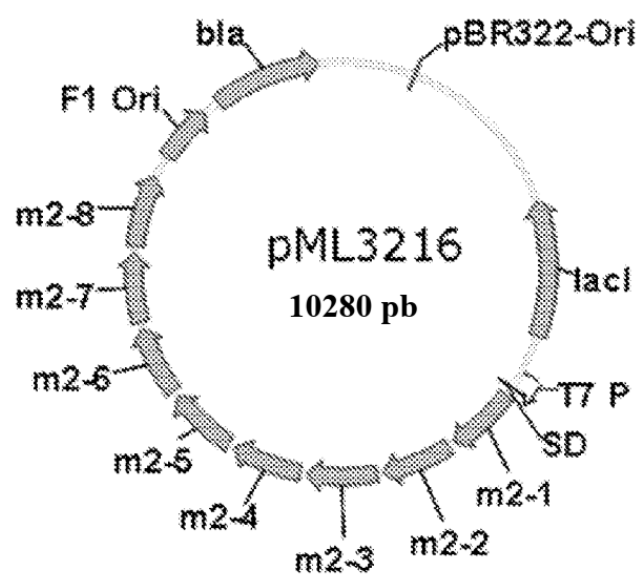


FIGURA 11

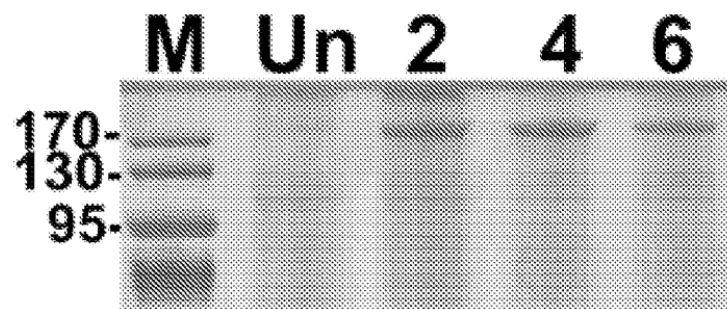


FIGURA 12

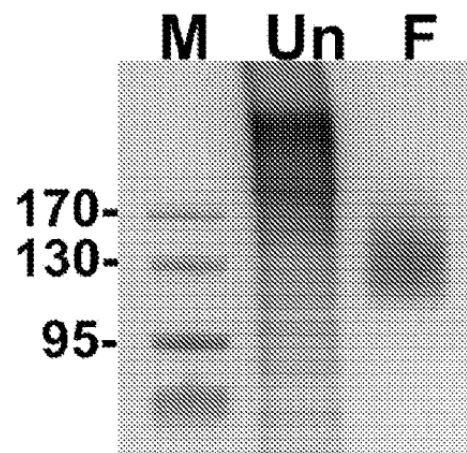


FIGURA 13

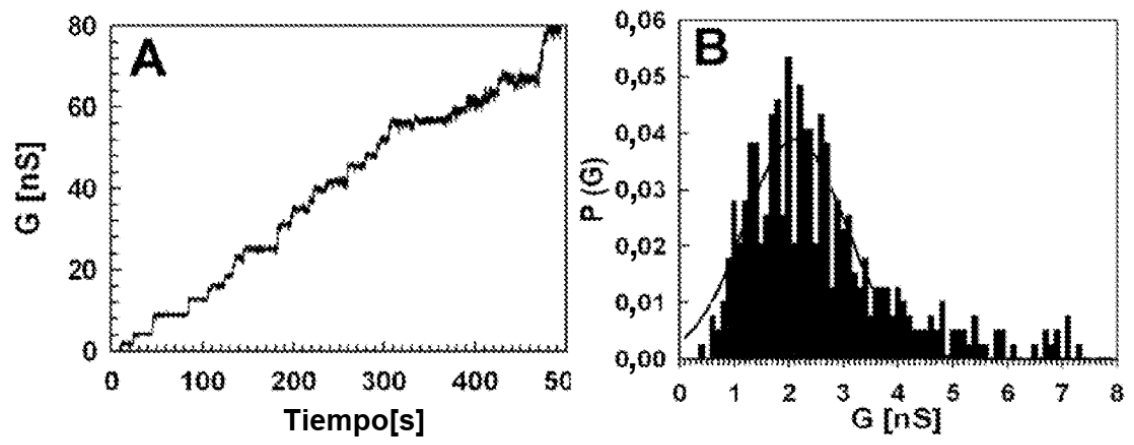


FIGURA 14

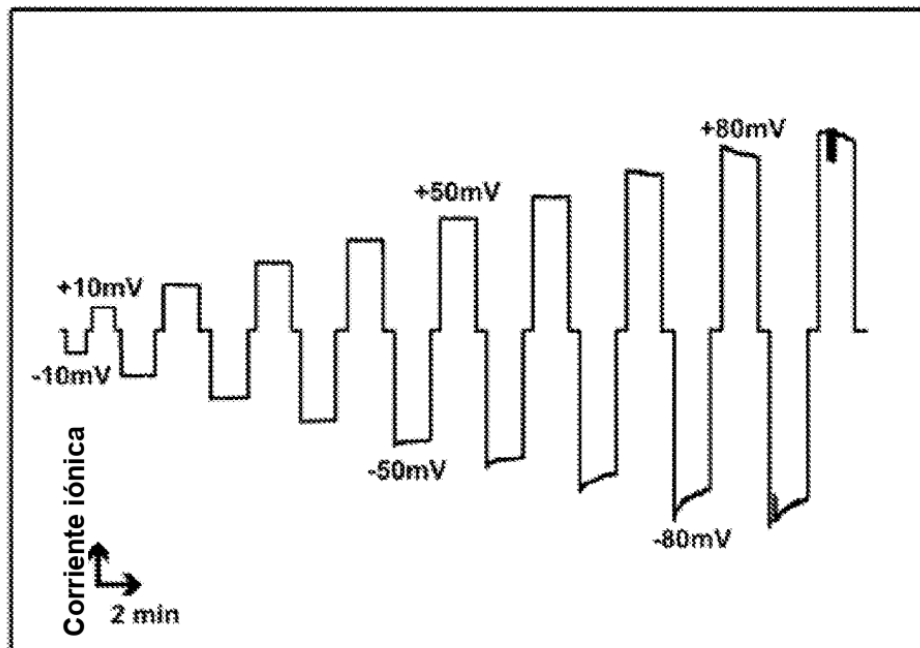


FIGURA 15

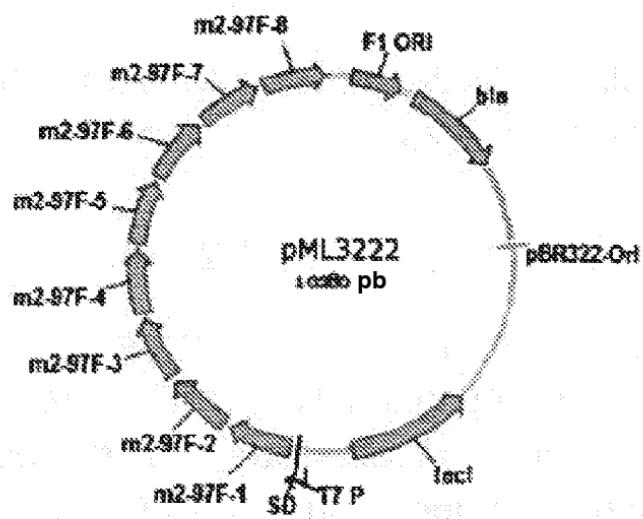


FIGURA 16

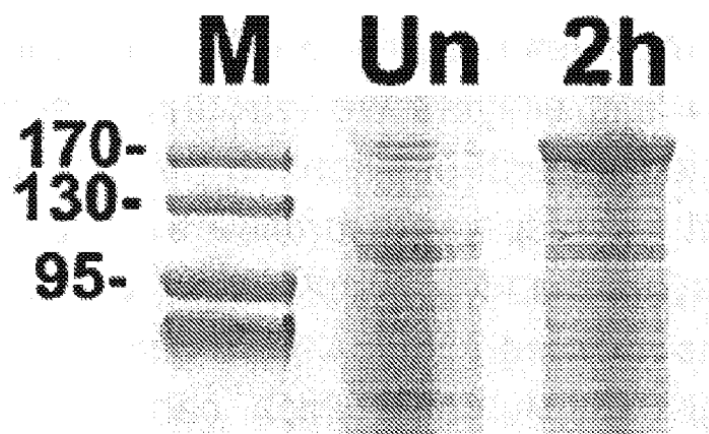


FIGURA 17

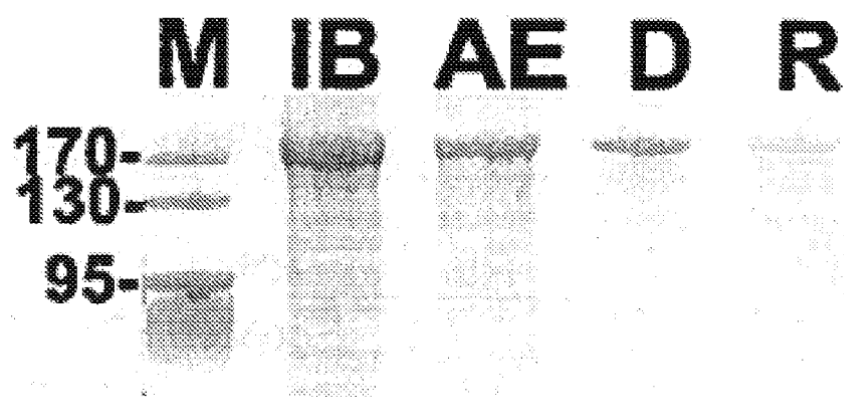


FIGURA 18

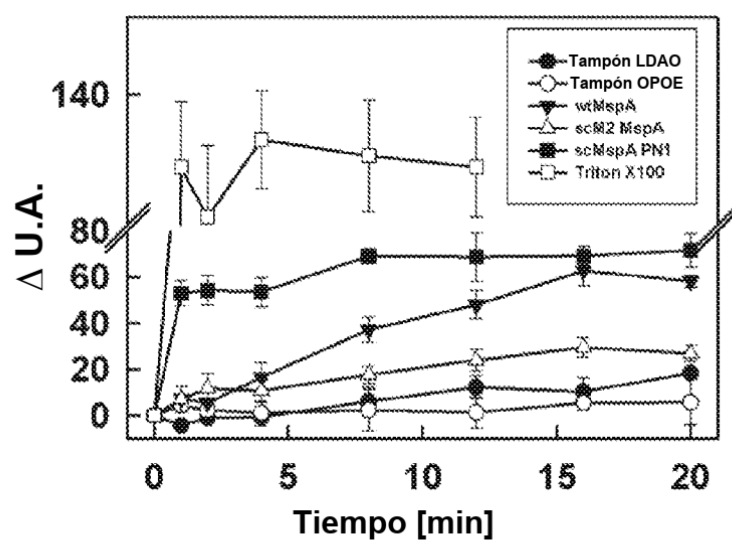


FIGURA 19

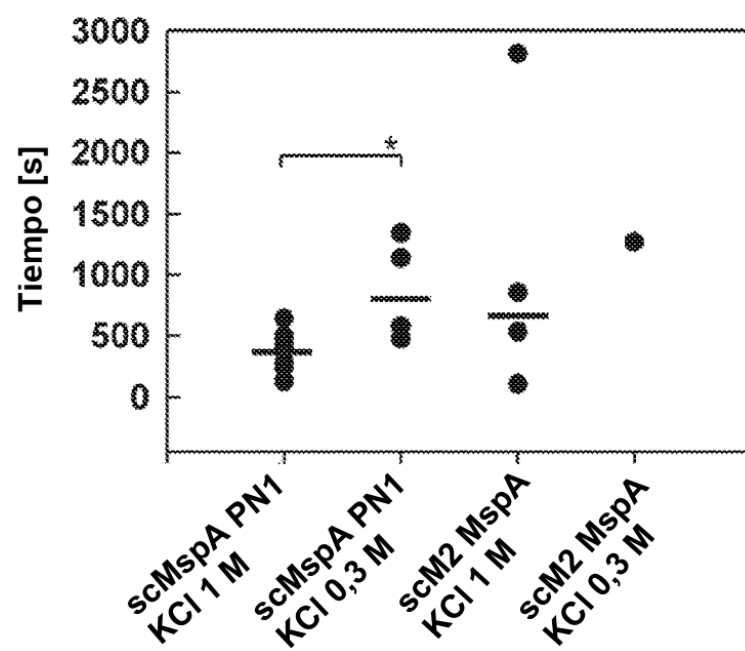


FIGURA 20

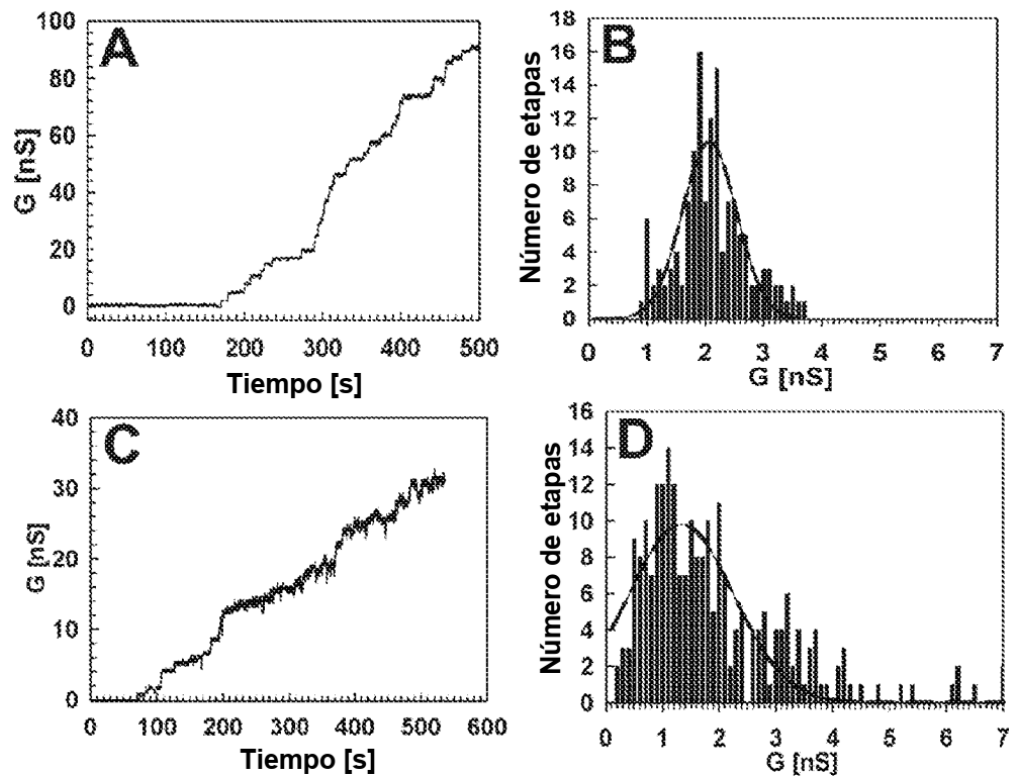


FIGURA 21

Actividad de canal de MspA PN1 de cadena sencilla

Registros de señal de corriente:

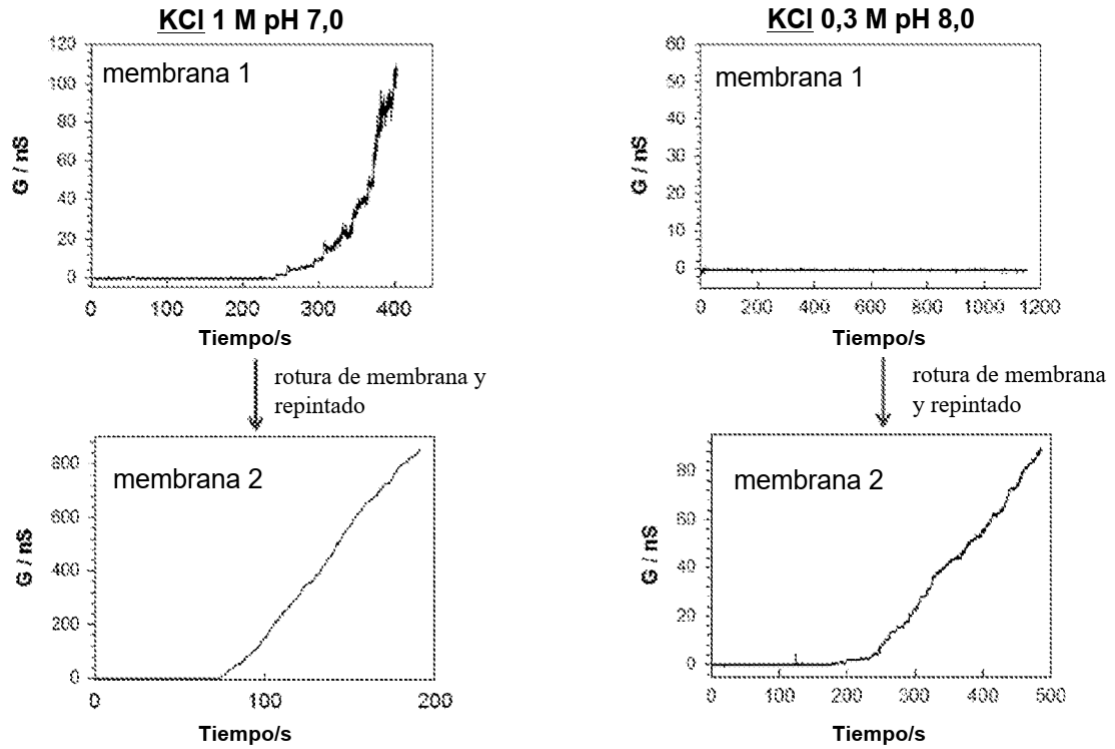


FIGURA 22