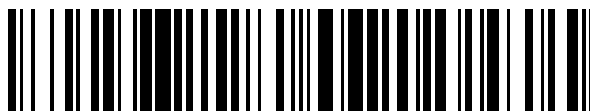


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 774 802**

51 Int. Cl.:

**C12Q 1/68** (2008.01)

**C07K 14/31** (2006.01)

**C12Q 1/6869** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.10.2015 PCT/US2015/057902**

87 Fecha y número de publicación internacional: **06.05.2016 WO16069806**

96 Fecha de presentación y número de la solicitud europea: **28.10.2015 E 15855632 (4)**

97 Fecha y número de publicación de la concesión europea: **01.01.2020 EP 3212810**

54 Título: **Variantes de hemolisina alfa con características alteradas**

30 Prioridad:

**31.10.2014 US 201462073936 P**

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

**22.07.2020**

73 Titular/es:

**GENIA TECHNOLOGIES, INC. (100.0%)  
2841 Scott Boulevard  
Santa Clara, CA 95050, US**

72 Inventor/es:

**DORWART, MICHAEL y  
KORENBLUM, DANIEL**

74 Agente/Representante:

**LINAGE GONZÁLEZ, Rafael**

ES 2 774 802 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

Variantes de hemolisina alfa con características alteradas

### 5 LISTADO DE SECUENCIAS

Se adjunta a esto un listado de secuencias que comprende SEQ ID NO: 1-8. Cada secuencia proporcionada en el listado de secuencias se incorpora en el presente documento por referencia, en su totalidad, para todos los propósitos. Dicha copia ASCII, creada el 26 de octubre de 2015, se llama 20-04338.P519WO1\_SL.txt y tiene un tamaño de 20.325 bytes.

### CAMPO TÉCNICO

Se divulgan composiciones y procedimientos relativos a variantes de hemolisina alfa de *Staphylococcal aureaus*. Las variantes de hemolisina alfa ( $\alpha$ -HL) son útiles, por ejemplo, como nanoporo en un dispositivo para determinar la información de la secuencia polimérica. Los nanoporos, procedimientos y sistemas descritos en el presente documento proporcionan la detección cuantitativa de ácidos nucleicos monocatenarios, tales como ADN, ARN, etc., empleando tecnología de molécula única basada en nanoporos con características mejoradas.

### 20 ANTECEDENTES

Las hemolisinas son miembros de una familia de toxinas proteicas que se producen por una amplia variedad de organismos. Algunas hemolisinas, por ejemplo hemolisinas alfa, pueden romper la integridad de una membrana celular (por ejemplo, una membrana de célula huésped) formando un poro o canal en la membrana. Los poros o canales que se forman en una membrana por proteínas formadoras de poros se pueden usar para transportar determinados polímeros (por ejemplo, polipéptidos o polinucleótidos) de un lado de una membrana al otro.

La hemolisina alfa ( $\alpha$ -HL, a-HL o alfa-HL) es una toxina autoensamblable que forma un canal acuoso en la membrana de una célula huésped. La alfa-HL se ha convertido en un componente principal para la comunidad de secuenciación de nanoporos. Tiene muchas propiedades ventajosas incluyendo alta estabilidad, autoensamblaje y un diámetro de poro que es lo suficientemente ancho como para alojar ADN monocatenario pero no ADN bicatenario (Kasianowicz *et al.*, 1996).

El trabajo previo sobre detección de ADN en el poro de a-HL se ha centrado en analizar la firma de corriente iónica a medida que el ADN se transloca a través del poro (Kasianowicz *et al.*, 1996, Akeson *et al.*, 1999, Meller *et al.*, 2001), una tarea muy difícil dada la tasa de translocación ( $\sim 1$  nt/ $\mu$ s a 100 mV) y el consiguiente ruido en la señal de corriente iónica. Se ha logrado una mayor especificidad en sensores basados en nanoporos por la incorporación de moléculas de sonda atadas permanentemente al interior del poro (Howorka *et al.*, 2001a y Howorka *et al.*, 2001b; Movileanu *et al.*, 2000).

La a-HL natural da como resultado un número significativo de errores de delección, es decir, no se miden las bases. Por lo tanto, se desean nanoporos de  $\alpha$ -HL con propiedades mejoradas.

### 45 BREVE SUMARIO DE LA INVENCION

La invención presenta un polipéptido de hemolisina alfa ( $\alpha$ HL) estafilocócica mutante que contiene una variación aminoacídica que potencia el tiempo de enhebrado, por ejemplo, disminuye el tiempo de captura de la molécula de interés.

Las variantes actualmente divulgadas reducen el tiempo de enhebrado de la molécula de interés, por ejemplo, diversos nucleótidos marcados o un nucleótido que se va a secuenciar.

En el presente documento se divulgan variantes de hemolisina  $\alpha$  ( $\alpha$ HL). Las variantes de hemolisina  $\alpha$  ( $\alpha$ HL) se derivan de un polipéptido de  $\alpha$ -HL original o una secuencia que tiene al menos un 80 %, 90 %, 95 %, 98 % o más de identidad de secuencia con SEQ ID NO: 8, y comprende una sustitución en una posición correspondiente a la posición 12 o 17 de SEQ ID NO:3 (a-HL madura). En algunos modos de realización, la variante comprende además H144A. En algunos modos de realización, la sustitución comprende una o más cargas positivas. En algunos modos de realización, la variante comprende una sustitución en una posición correspondiente a uno o más de los residuos T12 y/o N17. En algunos modos de realización, la variante comprende una sustitución seleccionada de T12K, T12R, N17K, N17R y combinaciones de las mismas. En algunos modos de realización, la variante tiene un tiempo de enhebrado (TTT) alterado relativo a la hemolisina  $\alpha$  original. En algunos modos de realización, el TTT se disminuye. En algunos modos de realización, la variante comprende una sustitución en una posición correspondiente a un residuo seleccionado del grupo que consiste en T12R o K, y/o N17R o K en hemolisina  $\alpha$  ( $\alpha$ HL) de *Staphylococcus aureus* (SEQ ID NO: 1). En algunos modos de realización, la sustitución es T12K. En algunos modos de realización, la sustitución es T12R. En algunos modos de realización, la sustitución es N17K. En algunos modos de realización, la sustitución es N17R. En algunos modos de realización, la variante de a-HL que tiene una característica alterada en comparación con una

hemolisina  $\alpha$  original (por ejemplo, AAA26598) comprende H144A y al menos una mutación adicional seleccionada de

a. T12K/R;

5 b. N17K/R;  
o combinaciones de las mismas.

En todos los modos de realización, la hemolisina alfa tiene una secuencia que tiene al menos un 90 %, preferentemente un 95 %, 98 % o más de identidad de secuencia con SEQ ID NO: 8.

10 En algunos modos de realización, la sustitución aminoacídica permite la adición de moléculas heterólogas, por ejemplo, PEG. En algún modo de realización, la variante de  $\alpha$ -HL tiene modificaciones postraduccionales.

15 En algunos modos de realización, la sustitución es un aminoácido no natural que es básico o positivamente cargado a un pH de aproximadamente 5 a aproximadamente 8,5.

En algunos casos, una polimerasa se asocia con el nanoporo (por ejemplo, se une covalentemente al nanoporo) y la polimerasa realiza acontecimientos de incorporación de nucleótidos.

20 En un aspecto, se proporciona un ensamblaje de poros heptomérico que comprende al menos una variante de hemolisina  $\alpha$  ( $\alpha$ HL) como se describe en el presente documento. En un modo de realización, la invención proporciona un ensamblaje de poros heteromérico que contiene un polipéptido de  $\alpha$ HL mutante (M), por ejemplo, un ensamblaje de poros que contiene un polipéptido de  $\alpha$ HL estafilocócico natural (WT) y un polipéptido de  $\alpha$ HL mutante en el que una variante aminoacídica (como se proporciona en el presente documento) del polipéptido de  $\alpha$ HL mutante ocupa una posición en un canal transmembranario de la estructura de poros. Por ejemplo, la proporción de polipéptidos de  $\alpha$ HL WT y variantes se expresa por la fórmula  $WT_{7-n}M_n$ , donde n es 1, 2, 3, 4, 5, 6 o 7; preferentemente la proporción de polipéptidos de  $\alpha$ HL en el heteroheptámero es  $WT_{7-n}M_n$ ; lo más preferentemente, la proporción es  $WT_6M_1$ . Los poros homoméricos en los que cada subunidad del heptómero es un polipéptido de  $\alpha$ HL mutado (es decir, donde n=7) también se engloban por la invención.

30 En un aspecto, se proporciona un ácido nucleico que codifica una variante de  $\alpha$ -HL como se describe en el presente documento.

35 En un aspecto, se proporciona un vector que comprende un ácido nucleico que codifica una variante de hemolisina alfa como se describe en el presente documento.

En un aspecto, se proporciona una célula huésped transformada con el vector que comprende un ácido nucleico que codifica una variante de hemolisina alfa como se describe en el presente documento.

40 En un aspecto, se proporciona un procedimiento de producción de una variante de hemolisina alfa que comprende las etapas de: (a) cultivar una célula huésped que comprende un ácido nucleico que codifica una variante de hemolisina alfa como se describe en el presente documento en un medio de cultivo adecuado en condiciones adecuadas para producir la variante de hemolisina alfa; y (b) obtener dicha variante de hemolisina alfa producida.

45 En un aspecto, se proporciona un procedimiento para detectar una molécula diana, que comprende: (a) proporcionar un chip que comprende un nanoporo como se describe en el presente documento en una membrana que se dispone adyacente o próxima a un electrodo sensor; (b) dirigir una molécula de ácido nucleico a través de dicho nanoporo, en el que dicha molécula de ácido nucleico se asocia con una molécula indicadora, en el que dicha molécula de ácido nucleico comprende una región de dirección y una región de sonda, en el que dicha molécula indicadora se asocia con dicha molécula de ácido nucleico en dicha región de sonda, y en el que dicha molécula indicadora se acopla a una molécula diana; (c) secuenciar dicha región de dirección mientras dicha molécula de ácido nucleico se dirige a través de dicho nanoporo para determinar una secuencia de ácido nucleico de dicha región de dirección; y (d) identificar, con la ayuda de un procesador informático, dicha molécula diana en base a una secuencia de ácido nucleico de dicha región de dirección determinada en (c).

55 Otros objetivos, rasgos característicos y ventajas de la presente invención serán evidentes a partir de la siguiente descripción detallada.

## 60 BREVE DESCRIPCIÓN DE LOS DIBUJOS

Las figuras 1-5 comprenden cada una dos figuras, por ejemplo, figura 1A y 1B. La figura A para cada figura es un histograma del número de acontecimientos de captura que tuvieron un "tiempo de enhebrado" igual al intervalo de tiempo mostrado en el eje x. La figura B para cada figura es una parte de los datos brutos para la correspondiente figura A.

65 La figura 1 muestra los resultados para el nanoporo de hemolisina  $\alpha$  natural. La figura 1A (panel superior) muestra

datos de "tiempo de enhebrado". Estos datos se combinan de muchos poros que capturaron los nucleótidos marcados lo que indica que el poro tuvo una molécula tanto de ADN molde como polimerasa. Los valores de media y mediana, junto con la desviación estándar para aHL natural son 20,7 ms, 16,1 ms y 1,5 ms respectivamente, y el número total de ondas cuadradas usadas para los cálculos es de 41910.

5 La figura 1B (panel inferior) muestra algunos datos brutos con cinco ondas cuadradas consecutivas mostradas. Los puntos de datos entre las líneas verdes representan el canal abierto (donde no se enhebra ningún nucleótido marcado en el poro) y los datos entre las líneas rojas representan cuando el nucleótido marcado se ha enhebrado en el poro y está bloqueando los iones que se mueven a través del canal. El electrodo se cicla entre 100 mV positivo y negativo, y en el presente sistema los puntos de datos no se registran cuando se aplica un voltaje negativo. Por tanto, todos los puntos de datos se recopilan del potencial aplicado positivamente, y el tiempo donde existe una ausencia de puntos de datos (entre 1716,9-1717 s, por ejemplo) es cuando los electrodos tienen un voltaje negativo aplicado a estos. En este ejemplo, la medición del "tiempo de enhebrado" se calcula a partir de ondas cuadradas que tienen un nivel enhebrado observable y la onda cuadrada previa tuvo un nivel enhebrado al final del voltaje positivo (lo que indica que la marca se enhebró en el poro y se unió por la polimerasa).

La figura 2 muestra los resultados para el nanoporo hemolisina  $\alpha$  que comprende una mutación T12K. La figura 2A (panel superior) son datos combinados de muchos poros que capturaron los nucleótidos marcados lo que indica que el poro tuvo una molécula tanto de ADN molde como polimerasa. Los valores de media y mediana, junto con la desviación estándar para T12K aHL son 19,7 ms, 14,5 ms y 1,5 ms respectivamente, y el número total de ondas cuadradas usadas para los cálculos es 4311.

La figura 2B (panel inferior) muestra algunos datos brutos con cinco ondas cuadradas consecutivas mostradas. Los puntos de datos entre las líneas verdes representan el canal abierto (donde no se enhebra ningún nucleótido marcado en el poro) y los datos entre las líneas rojas representan cuando el nucleótido marcado se ha enhebrado en el poro y está bloqueando los iones que se mueven a través del canal. El electrodo se cicla entre 100 mV positivo y negativo, y en el presente sistema los puntos de datos no se registran cuando se aplica un voltaje negativo. Por tanto, todos los puntos de datos se recopilan del potencial aplicado positivamente, y el tiempo donde existe una ausencia de puntos de datos (entre 1600,4-1601,2 s, por ejemplo) es cuando los electrodos tienen un voltaje negativo aplicado a estos. En este ejemplo, la medición del "tiempo de enhebrado" se calcula a partir de ondas cuadradas que tienen un nivel enhebrado observable y la onda cuadrada previa tuvo un nivel enhebrado al final del voltaje positivo (lo que indica que la marca se enhebró en el poro y se unió por la polimerasa).

La figura 3 muestra los resultados para el nanoporo hemolisina  $\alpha$  que comprende una mutación T12R. La figura 3A son datos combinados de muchos poros que capturaron los nucleótidos marcados lo que indica que el poro tuvo una molécula tanto de ADN molde como polimerasa. Los valores de media y mediana, junto con la desviación estándar para T12R aHL son 16,9 ms, 10,5 ms y 1,5 ms respectivamente, y el número total de ondas cuadradas usadas para los cálculos es 4138.

La figura 3B (panel inferior) muestra algunos datos brutos con cinco ondas cuadradas consecutivas mostradas. Los puntos de datos entre las líneas verdes representan el canal abierto (donde no se enhebra ningún nucleótido marcado en el poro) y los datos entre las líneas rojas representan cuando el nucleótido marcado se ha enhebrado en el poro y está bloqueando los iones que se mueven a través del canal. El electrodo se cicla entre 100 mV positivo y negativo, y en el presente sistema los puntos de datos no se registran cuando se aplica un voltaje negativo. Por tanto, todos los puntos de datos se recopilan del potencial aplicado positivamente, y el tiempo donde existe una ausencia de puntos de datos (entre 267,2 - 268,2 s, por ejemplo) es cuando los electrodos tienen un voltaje negativo aplicado a estos. En este ejemplo, la medición del "tiempo de enhebrado" se calcula a partir de ondas cuadradas que tienen un nivel enhebrado observable y la onda cuadrada previa tuvo un nivel enhebrado al final del voltaje positivo (lo que indica que la marca se enhebró en el poro y se unió por la polimerasa).

La figura 4 muestra los resultados para el nanoporo hemolisina  $\alpha$  que comprende una mutación N17R. La figura 4A (panel superior) son datos combinados de muchos poros que capturaron los nucleótidos marcados lo que indica que el poro tuvo una molécula tanto de ADN molde como polimerasa. Los valores de media y mediana, junto con la desviación estándar para N17R aHL son 17,5 ms, 10,5 ms y 1,7 ms respectivamente, y el número total de ondas cuadradas usadas para los cálculos es 3877.

La figura 4B (panel inferior) muestra algunos datos brutos con cinco ondas cuadradas consecutivas mostradas. Los puntos de datos entre las líneas verdes representan el canal abierto (donde no se enhebra ningún nucleótido marcado en el poro) y los datos entre las líneas rojas representan cuando el nucleótido marcado se ha enhebrado en el poro y está bloqueando los iones que se mueven a través del canal. El electrodo se cicla entre 100 mV positivo y negativo, y en el presente sistema los puntos de datos no se registran cuando se aplica un voltaje negativo. Por tanto, todos los puntos de datos se recopilan del potencial aplicado positivamente, y el tiempo donde existe una ausencia de puntos de datos (entre 344 - 344,9 s, por ejemplo) es cuando los electrodos tienen un voltaje negativo aplicado a estos. En este ejemplo, la medición del "tiempo de enhebrado" se calcula a partir de ondas cuadradas que tienen un nivel enhebrado observable y la onda cuadrada previa tuvo un nivel enhebrado al final del voltaje positivo (lo que indica que la marca se enhebró en el poro y se unió por la polimerasa).

La figura 5 muestra los resultados para el nanoporo hemolisina  $\alpha$  que comprende una mutación N17K. La figura 5A (panel superior) muestra datos combinados de muchos poros que capturaron los nucleótidos marcados lo que indica que el poro tuvo una molécula tanto de ADN molde como polimerasa. Los valores de media y mediana, junto con la desviación estándar para N17K aHL son 5,7 ms, 2,4 ms y 0,7 ms respectivamente, y el número total de ondas cuadradas usadas para los cálculos es 2424.

La figura 5B (panel inferior) muestra algunos datos brutos con cinco ondas cuadradas consecutivas mostradas. Los puntos de datos entre las líneas verdes representan el canal abierto (donde no se enhebra ningún nucleótido marcado en el poro) y los datos entre las líneas rojas representan cuando el nucleótido marcado se ha enhebrado en el poro y está bloqueando los iones que se mueven a través del canal. El electrodo se cicla entre 100 mV positivo y negativo, y en el presente sistema los puntos de datos no se registran cuando se aplica un voltaje negativo. Por tanto, todos los puntos de datos se recopilan del potencial aplicado positivamente, y el tiempo donde existe una ausencia de puntos de datos (entre 79,5 - 80,5 s, por ejemplo) es cuando los electrodos tienen un voltaje negativo aplicado a estos. En este ejemplo, la medición del "tiempo de enhebrado" se calcula a partir de ondas cuadradas que tienen un nivel enhebrado observable y la onda cuadrada previa tuvo un nivel enhebrado al final del voltaje positivo (lo que indica que la marca se enhebró en el poro y se unió por la polimerasa).

## DESCRIPCIÓN DETALLADA

La invención se describirá ahora en detalle a modo de referencia usando solo las siguientes definiciones y ejemplos.

A menos que se defina de otro modo en el presente documento, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que se entiende comúnmente por un experto en la técnica a la que pertenece la presente invención. Singleton, *et al.*, DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2.<sup>a</sup> ed., John Wiley y Sons, Nueva York (1994), y Hale y Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) proporcionan a un experto un diccionario general de muchos de los términos usados en la presente invención. Aunque cualquier procedimiento y material similar o equivalente a los descritos en el presente documento se puede usar en la práctica o prueba de la presente invención, se describen los procedimientos y materiales preferentes. Los profesionales sanitarios se dirigen en particular a Sambrook *et al.*, 1989, y Ausubel *et al.*, 1993, para definiciones y términos de la técnica. Se debe entender que la presente invención no está limitada a la metodología, protocolos y reactivos particulares descritos, ya que estos pueden variar.

Los intervalos numéricos incluyen los números que definen el intervalo. El término aproximadamente en el presente documento quiere decir más o menos un diez por ciento (10 %) de un valor. Por ejemplo, "aproximadamente 100" se refiere a cualquier número entre 90 y 110.

A menos que se indique de otro modo, los ácidos nucleicos se escriben de izquierda a derecha en una orientación de 5' a 3'; las secuencias de aminoácidos se escriben de izquierda a derecha en una orientación de amino a carboxilo, respectivamente.

Los títulos proporcionados en el presente documento no son limitaciones de los diversos aspectos o modos de realización de la invención que se pueden tener por referencia a la memoria descriptiva como un todo. En consecuencia, los términos definidos inmediatamente a continuación se definen más completamente por referencia a la memoria descriptiva como un todo.

### Definiciones

**Hemolisina alfa:** como se usa en el presente documento, "hemolisina alfa", "hemolisina  $\alpha$ ", "a-HL" y " $\alpha$ -HL" se usan de manera intercambiable y se refieren a la proteína monomérica que se autoensambla en un canal transmembranario lleno de agua heptamérico (es decir, nanoporo). Dependiendo del contexto, el término también se puede referir al canal transmembranario formado por siete proteínas monoméricas.

**Aminoácido:** como se usa en el presente documento, el término "aminoácido", en su sentido más amplio, se refiere a cualquier compuesto y/o sustancia que se puede incorporar en una cadena polipeptídica. En algunos modos de realización, un aminoácido tiene la estructura general  $H_2N-C(H)(R)-COOH$ . En algunos modos de realización, un aminoácido es un aminoácido natural. En algunos modos de realización, el aminoácido es un aminoácido sintético; en algunos modos de realización, un aminoácido es un D-aminoácido; en algunos modos de realización, un aminoácido es un L-aminoácido. "Aminoácido estándar" se refiere a cualquiera de los veinte L-aminoácidos estándar encontrados comúnmente en péptidos naturales. "Aminoácido no estándar" se refiere a cualquier aminoácido, distinto de los aminoácidos estándar, independientemente de si se prepara sintéticamente o se obtiene de una fuente natural. Como se usa en el presente documento, "aminoácido sintético" o "aminoácido no natural" engloba aminoácidos modificados químicamente, incluyendo pero sin limitarse a sales, derivados aminoacídicos (tales como amidas) y/o sustituciones. Los aminoácidos, incluyendo aminoácidos carboxi y/o aminoterminales en péptidos, se pueden modificar por metilación, amidación, acetilación y/o sustitución con otros grupos químicos sin afectar negativamente a su actividad. Los aminoácidos pueden participar en un enlace disulfuro. El término "aminoácido" se usa de manera intercambiable

con "residuo aminoacídico" y se puede referir a un aminoácido libre y/o a un residuo aminoacídico de un péptido. Será evidente a partir del contexto en el que se usa el término si se refiere a un aminoácido libre o un residuo de un péptido. Cabe destacar que todas las secuencias de residuos aminoacídicos se representan en el presente documento por fórmulas con una orientación izquierda y derecha que está en el sentido convencional del extremo aminoterminal al carboxiterminal.

**Par de bases (pb):** como se usa en el presente documento, par de bases se refiere a una asociación de adenina (A) con timina (T), o de citosina (C) con guanina (G) en una molécula de ADN bicatenario.

**Complementario:** como se usa en el presente documento, el término "complementario" se refiere al concepto amplio de complementariedad de secuencia entre regiones de dos hebras polinucleotídicas o entre dos nucleótidos a través de emparejamiento de bases. Es conocido que un nucleótido adenina puede formar enlaces de hidrógeno específicos ("emparejamiento de bases") con un nucleótido que sea timina o uracilo. De forma similar, es conocido que un nucleótido citosina puede realizar emparejamiento de bases con un nucleótido guanina.

**Casete de expresión:** un "casete de expresión" o "vector de expresión" es una construcción de ácido nucleico generada de forma recombinante o sintética, con una serie de elementos de ácido nucleico especificados que permiten la transcripción de un ácido nucleico particular en una célula diana. El casete de expresión recombinante se puede incorporar en un plásmido, cromosoma, ADN mitocondrial, ADN de plástido, virus o fragmento de ácido nucleico.

Típicamente, la porción de casete de expresión recombinante de un vector de expresión incluye, entre otras secuencias, una secuencia de ácido nucleico que se va a transcribir y un promotor.

**Heterólogo:** una construcción o secuencia de ácido nucleico "heterólogo" tiene una porción de la secuencia que no es natural para la célula en la que se expresa. Heterólogo, con respecto a una secuencia de control se refiere a una secuencia de control (es decir, promotor o potenciador) que no funciona en la naturaleza para regular el mismo gen con una expresión que se está regulando actualmente. En general, las secuencias de ácido nucleico heterólogo no son endógenas a la célula o parte del genoma en el que están presentes, y se han añadido a la célula, por infección, transfección, transformación, microinyección, electroporación o similares. Una construcción de ácido nucleico "heterólogo" puede contener una combinación de secuencia de control/secuencia codificante de ADN que es igual a, o diferente de una combinación de secuencia de control/secuencia codificante de ADN encontrada en la célula natural.

**Célula huésped:** por el término "célula huésped" se quiere decir una célula que contiene un vector y soporta la replicación, y/o transcripción o transcripción y traducción (expresión) de la construcción de expresión. Las células huésped para su uso en la presente invención pueden ser células procariotas, tales como *E. coli* o *Bacillus subtilis*, o células eucariotas tales como levadura, células de planta, insecto, anfibio o mamífero. En general, las células huésped son procariotas, por ejemplo, *E. coli*.

**Aislado:** una molécula "aislada" es una molécula de ácido nucleico que se separa de al menos otra molécula con la que se asocia normalmente, por ejemplo, en su entorno natural. Una molécula de ácido nucleico aislada incluye una molécula de ácido nucleico contenida en células que expresan habitualmente la molécula de ácido nucleico, pero la molécula de ácido nucleico está presente de forma extracromosómica o en una localización cromosómica que es diferente de su localización cromosómica natural.

**Hemolisina alfa modificada:** como se usa en el presente documento, el término "hemolisina alfa modificada" se refiere a una hemolisina alfa originada de otra hemolisina alfa (es decir, original) y contiene una o más alteraciones aminoacídicas (por ejemplo, sustitución, delección o inserción aminoacídica) en comparación con la hemolisina alfa original. En algunos modos de realización, una hemolisina alfa modificada de la invención se origina o modifica a partir de una hemolisina alfa natural. En algunos modos de realización, una hemolisina alfa modificada de la invención se origina o modifica a partir de una hemolisina alfa recombinante o genomanipulada incluyendo, pero sin limitarse a, hemolisina alfa quimérica, hemolisina alfa de fusión u otra hemolisina alfa modificada. Típicamente, una hemolisina alfa modificada tiene al menos un fenotipo cambiado en comparación con la hemolisina alfa original.

**Mutación:** como se usa en el presente documento, el término "mutación" se refiere a un cambio introducido en una secuencia original incluyendo, pero sin limitarse a, sustituciones, inserciones, delecciones (incluyendo truncamientos). Las consecuencias de una mutación incluyen, pero no se limitan a, la creación de un nuevo carácter, propiedad, función, fenotipo o rasgo no encontrado en la proteína codificada por la secuencia original.

**Nanoporo:** el término "nanoporo", como se usa en el presente documento, se refiere en general a un poro, canal o paso formado o provisto de otro modo en una membrana. Una membrana puede ser una membrana orgánica, tal como una bicapa lipídica, o una membrana sintética, tal como una membrana formada de un material polimérico. La membrana puede ser un material polimérico. El nanoporo se puede disponer adyacente o próximo a un circuito sensor o un electrodo acoplado a un circuito sensor, tal como, por ejemplo, un semiconductor de óxido metálico complementario (CMOS) o circuito de transistor de efecto de campo (FET). En algunos ejemplos, un nanoporo tiene un ancho o diámetro característico del orden de 0,1 nanómetros (nm) a aproximadamente 1000 nm. Algunos nanoporos son proteínas. La hemolisina alfa es un ejemplo de un nanoporo proteico.

**Molécula de ácido nucleico:** el término "molécula de ácido nucleico" incluye moléculas de ARN, ADN y ADNc. Se entenderá que, como resultado de la degeneración del código genético, se puede producir una multitud de secuencias de nucleótidos que codifican una proteína dada tal como hemolisina alfa y/o variantes de la misma. La presente invención contempla cada posible secuencia de nucleótidos variante, que codifica hemolisina alfa variante, de las que todas son posibles dada la degeneración del código genético.

**Promotor:** como se usa en el presente documento, el término "promotor" se refiere a una secuencia de ácido nucleico que funciona para dirigir la transcripción de un gen hacia 3'. El promotor será apropiado en general para la célula huésped en la que se expresa el gen diana. El promotor conjuntamente con otras secuencias de ácido nucleico reguladoras transcripcionales y traduccionales (también denominadas "secuencias de control") son necesarios para expresar un gen dado. En general, las secuencias reguladoras transcripcionales y traduccionales incluyen, pero no se limitan a, secuencias promotoras, sitios de unión ribosómicos, secuencias de inicio y parada transcripcionales, secuencias de inicio y parada traduccionales y secuencias potenciadoras o activadoras.

**Purificado:** como se usa en el presente documento, "purificado" quiere decir que una molécula está presente en una muestra a una concentración de al menos un 95 % en peso, o al menos un 98 % en peso de la muestra en la que está contenida.

**Purificar:** como se usa en el presente documento, el término "purificar" en general se refiere a someter a las células que contienen proteína o ácido nucleico transgénico a purificación bioquímica y/o cromatografía en columna.

**Marca:** como se usa en el presente documento, el término "marca" se refiere a un resto detectable que puede ser átomos o moléculas, o una colección de átomos o moléculas. Una marca puede proporcionar una firma óptica, electroquímica, magnética o electrostática (por ejemplo, inductiva, capacitiva), que se puede detectar con la ayuda de un nanoporo. Típicamente, cuando un nucleótido se une a la marca, se llama "nucleótido marcado". La marca puede estar unida al nucleótido por medio del resto fosfato.

**Tiempo de enhebrado:** el término "tiempo de enhebrado" o "TTT" quiere decir el tiempo que tarda el complejo polimerasa-marca o una hebra de ácido nucleico en enhebrar la marca en el cuerpo del nanoporo.

**Variante:** como se usa en el presente documento, el término "variante" se refiere a una proteína modificada que presenta características alteradas cuando se compara con la proteína original, por ejemplo, conductancia iónica alterada.

**Hemolisina variante:** el término "gen de hemolisina variante" o "hemolisina variante" quiere decir, respectivamente, que la secuencia de ácido nucleico del gen de hemolisina alfa de *Staphylococcus aureus* se ha alterado retirando, añadiendo y/o manipulando la secuencia codificante o la secuencia de aminoácidos de la proteína expresada se ha modificado consecuentemente con la invención descrita en el presente documento.

**Vector:** como se usa en el presente documento, el término "vector" se refiere a una construcción de ácido nucleico diseñada para la transferencia entre diferentes células huésped. Un "vector de expresión" se refiere a un vector que tiene la capacidad de incorporar y expresar fragmentos de ADN heterólogos en una célula exógena. Muchos vectores de expresión procariotas y eucariotas están disponibles comercialmente. La selección de vectores de expresión apropiados está dentro del conocimiento de los expertos en la técnica.

**Natural:** como se usa en el presente documento, el término "natural" se refiere a un gen o producto génico que tiene las características de ese gen o producto génico cuando se aísla de una fuente natural.

**Porcentaje de homología:** el término "% de homología" se usa de manera intercambiable en el presente documento con el término "% de identidad" en el presente documento y se refiere al nivel de identidad de secuencia de ácido nucleico o aminoácidos entre la secuencia de ácido nucleico que codifica uno cualquiera de los polipéptidos según la invención o la secuencia de aminoácidos del polipéptido según la invención, cuando se alinea usando un programa de alineación de secuencias.

Por ejemplo, como se usa en el presente documento, un 80 % de homología quiere decir lo mismo que un 80 % de identidad de secuencia determinada por un algoritmo definido y, en consecuencia, un homólogo de una secuencia dada tiene más de un 80 % de identidad de secuencia en una longitud de la secuencia dada. Los niveles ejemplares de identidad de secuencia incluyen, pero no se limitan a, 80, 85, 90, 95, 98 % o más de identidad de secuencia para una secuencia dada, por ejemplo, la secuencia codificante para uno cualquiera de los polipéptidos según la invención, como se describe en el presente documento.

Los programas informáticos ejemplares que se pueden usar para determinar la identidad entre dos secuencias incluyen, pero no se limitan a, el paquete de programas BLAST, por ejemplo, BLASTN, BLASTX y TBLASTX, BLASTP y TBLASTN, disponibles públicamente en Internet. Véanse también, Altschul, *et al.*, 1990 y Altschul, *et al.*, 1997.

Las búsquedas de secuencias se llevan a cabo típicamente usando el programa BLASTN cuando se evalúa una secuencia de ácido nucleico dada relativa a las secuencias de ácido nucleico en las secuencias de ADN de GenBank y otras bases de datos públicas. El programa BLASTX es preferente para buscar secuencias de ácido nucleico que se han traducido en todos los marcos de lectura frente a secuencias de aminoácidos en las secuencias de proteínas de GenBank y otras bases de datos públicas. Tanto BLASTN como BLASTX se ejecutan usando parámetros predeterminados de una penalización por hueco abierto de 11,0 y una penalización por hueco extendido de 1,0, y utilizan la matriz BLOSUM-62. (Véase, por ejemplo, Altschul, S. F., *et al.*, Nucleic Acids Res. 25:3389-3402, 1997.)

Se realiza una alineación preferente de secuencias seleccionadas para determinar el "% de identidad" entre dos o más secuencias, usando por ejemplo, el programa CLUSTAL-W en MacVector versión 13.0.7, operado con parámetros predeterminados, incluyendo una penalización por hueco abierto de 10,0, una penalización por hueco extendido de 0,1 y una matriz de similitud BLOSUM 30.

### Nomenclatura

En la presente descripción y reivindicaciones, se usan los códigos de una letra y tres letras convencionales para los residuos aminoácidos.

Para facilitar la referencia, las variantes de la solicitud se describen por el uso de la siguiente nomenclatura:

Aminoácido(s) original(es): posición/posiciones: aminoácido(s) sustituido(s). De acuerdo con esta nomenclatura, por ejemplo, la sustitución de treonina por una arginina en la posición 17 se muestra como:

Thr17Arg o T17R

Las mutaciones múltiples se separan por signos de más, es decir:

Thr17Arg+Glu34Ser o T17R+E34S

representando mutaciones en las posiciones 30 y 34 sustituyendo alanina y ácido glutámico por asparagina y serina, respectivamente.

Cuando uno o más residuos aminoácidos alternativos se pueden insertar en una posición dada, se indica como: T17R/K o T17R o T17K.

### Mutagénesis dirigida a sitio de hemolisina alfa

En el presente documento se proporcionan secuencias naturales de hemolisina alfa de *Staphylococcus aureus* (SEQ ID NO:1, región codificante de ácido nucleico; SEQ ID NO:3, región codificante de proteína) y están disponibles en otros lugares (National Center for Bioinformatics o GenBank con números de acceso M90536 y AAA26598).

Se pueden introducir mutaciones puntuales usando el kit QuikChange Lightning 2 (Stratagene/Agilent) siguiendo las instrucciones del fabricante.

Se pueden encargar cebadores a empresas comerciales, por ejemplo, ADN de IDT.

### Ensamblaje e inserción de nanoporos

Los procedimientos descritos en el presente documento pueden usar un nanoporo que tiene una polimerasa unida al nanoporo. En algunos casos, es deseable tener una y solo una polimerasa por nanoporo (por ejemplo, de modo que solo se secuencia una molécula de ácido nucleico en cada nanoporo). Sin embargo, muchos nanoporos, incluyendo hemolisina alfa (aHL), pueden ser proteínas multiméricas que tienen una pluralidad de subunidades (por ejemplo, 7 subunidades para aHL). Las subunidades pueden ser copias idénticas del mismo polipéptido. En el presente documento se proporcionan proteínas multiméricas (por ejemplo, nanoporos) que tienen una proporción definida de subunidades modificadas (por ejemplo, variantes de a-HL) con respecto a subunidades no modificadas (por ejemplo, a-HL). En el presente documento también se proporcionan procedimientos para producir proteínas multiméricas (por ejemplo, nanoporos) que tienen una proporción definida de subunidades modificadas con respecto a subunidades no modificadas.

Con referencia a la figura 27 del documento WO2014/074727, un procedimiento para ensamblar una proteína que tiene una pluralidad de subunidades comprende proporcionar una pluralidad de primeras subunidades 2705 y proporcionar una pluralidad de segundas subunidades 2710, donde las segundas subunidades se modifican en comparación con las primeras subunidades. En algunos casos, las primeras subunidades son naturales (por ejemplo, purificadas de fuentes naturales o producidas de forma recombinante). Las segundas subunidades se pueden modificar de cualquier forma adecuada. En algunos casos, las segundas subunidades tienen una proteína (por ejemplo, una polimerasa) unida (por ejemplo, como una proteína de fusión).



Las subunidades modificadas pueden comprender un resto químicamente reactivo (por ejemplo, una acida o un grupo alquino adecuado para formar un enlace). En algunos casos, el procedimiento comprende además realizar una reacción (por ejemplo, una cicloadición química *Click*) para unir una entidad (por ejemplo, una polimerasa) al resto químicamente reactivo.

El procedimiento puede comprender además poner en contacto las primeras subunidades con las segundas subunidades 2715 en una primera proporción para formar una pluralidad de proteínas 2720 que tienen las primeras subunidades y las segundas subunidades. Por ejemplo, una parte de las subunidades de aHL modificadas que tienen un grupo reactivo adecuado para unirse a una polimerasa se puede mezclar con seis partes de subunidades de aHL naturales (es decir, siendo la primera proporción 1:6). La pluralidad de proteínas puede tener una pluralidad de proporciones de las primeras subunidades con respecto a las segundas subunidades. Por ejemplo, las subunidades mixtas pueden formar varios nanoporos que tienen una distribución de estequiometrías de subunidades modificadas con respecto a no modificadas (por ejemplo, 1:6, 2:5, 3:4).

En algunos casos, las proteínas se forman simplemente mezclando las subunidades. En el caso de nanoporos de aHL, por ejemplo, un detergente (por ejemplo, ácido desoxicólico) puede provocar que el monómero de aHL adopte la conformación de poro. Los nanoporos también se pueden formar usando un lípido (por ejemplo, 1,2-difitanoiil-sn-glicero-3-fosfocolina (DPhPC) o 1,2-di-0-fitanil-sn-glicero-3-fosfocolina (DoPhPC)) y temperatura moderada (por ejemplo, menos de aproximadamente 100 °C). En algunos casos, mezclar DPhPC con una solución tampón crea grandes vesículas multilaminares (LMV), y añadir subunidades de aHL a esta solución e incubar la mezcla a 40 °C durante 30 minutos da como resultado la formación de poro.

Si se usan dos tipos diferentes de subunidades (por ejemplo, la proteína natural y un segundo monómero de aHL que puede contener una única mutación puntual), las proteínas resultantes pueden tener una estequiometría mixta (por ejemplo, de las proteínas naturales y mutantes). La estequiometría de estas proteínas puede seguir una fórmula que es dependiente de la proporción de las concentraciones de las dos proteínas usadas en la reacción de formación de poros. Esta fórmula es como sigue:

$$100 P_m = 100[n!/m!(n-m)!] \cdot f_{mut}^m \cdot f_{wt}^{n-m}, \text{ donde}$$

$P_m$  = probabilidad de que un poro tenga un número  $m$  de subunidades mutantes

$n$  = número total de subunidades (por ejemplo, 7 para aHL)

$m$  = número de subunidades "mutantes"

$f_{mut}$  = fracción o proporción de subunidades mutantes mezcladas juntas

$f_{wt}$  = fracción o proporción de subunidades naturales mezcladas juntas

El procedimiento puede comprender además fraccionar la pluralidad de proteínas para enriquecer proteínas que tienen una segunda proporción de las primeras subunidades con respecto a las segundas subunidades 2725. Por ejemplo, se pueden aislar proteínas de nanoporo que tienen una y solo una subunidad modificada (por ejemplo, una segunda proporción de 1:6). Sin embargo, cualquier segunda proporción es adecuada. También se puede fraccionar una distribución de segundas proporciones, tal como enriquecer proteínas que tienen una o bien dos subunidades modificadas. El número total de subunidades que forman la proteína no siempre es 7 (por ejemplo, se puede usar un nanoporo diferente o se puede formar un nanoporo de hemolisina alfa que tiene seis subunidades) como se representa en la figura 27 del documento WO2014/074727. En algunos casos, se enriquecen las proteínas que tienen solo una subunidad modificada. En dichos casos, la segunda proporción es 1 segunda subunidad por  $(n-1)$  primeras subunidades donde  $n$  es el número de subunidades que comprenden la proteína.

La primera proporción puede ser la misma que la segunda proporción, sin embargo, esto no se requiere. En algunos casos, las proteínas que tienen monómeros mutados se pueden formar menos eficazmente que las que no tienen subunidades mutadas. Si este es el caso, la primera proporción puede ser mayor que la segunda proporción (por ejemplo, si se desea una segunda proporción de 1 subunidad mutada con respecto a 6 no mutadas en un nanoporo, formar un número adecuado de 1:6 proteínas puede requerir mezclar las subunidades en una proporción mayor que 1:6).

Las proteínas que tienen diferentes segundas proporciones de subunidades se pueden comportar de forma diferente (por ejemplo, tienen diferentes tiempos de retención) en una separación. En algunos casos, las proteínas se fraccionan usando cromatografía, tal como cromatografía de intercambio iónico o cromatografía de afinidad. Puesto que las primeras y segundas subunidades pueden ser idénticas aparte de la modificación, el número de modificaciones en la proteína puede servir como base para la separación. En algunos casos, las primeras o bien segundas subunidades tienen una marca de purificación (por ejemplo, además de la modificación) para permitir o mejorar la eficacia del fraccionamiento. En algunos casos, se usa una marca polihistidina (marca His), una marca estreptavidina (marca

Strep) u otra marca peptídica. En algunos casos, las primeras y segundas subunidades comprenden cada una diferentes marcas y la etapa de fraccionamiento fracciona en base a cada marca. En el caso de una marca His, se crea una carga en la marca a pH bajo (los residuos de histidina se cargan positivamente por debajo del pKa de la cadena lateral). Con una diferencia significativa de carga en una de las moléculas de aHL en comparación con las otras, se puede usar cromatografía de intercambio iónico para separar los oligómeros que tienen 0, 1, 2, 3, 4, 5, 6 o 7 de las subunidades de aHL "marcadas con carga". En principio, esta marca de carga puede ser una cadena de cualquier aminoácido que porte una carga uniforme. La figura 28 y figura 29 muestran ejemplos de fraccionamiento de nanoporos en base a una marca His. La figura 28 muestra una curva de absorbancia ultravioleta a 280 nanómetros, absorbancia ultravioleta a 260 nanómetros y conductividad. Los picos corresponden a nanoporos con diversas proporciones de subunidades modificadas y no modificadas. La figura 29 del documento WO2014/074727 muestra el fraccionamiento de nanoporos de aHL y mutantes de los mismos usando marca His y marcas Strep.

En algunos casos, una entidad (por ejemplo, una polimerasa) se une a la proteína después del fraccionamiento. La proteína puede ser un nanoporo y la entidad puede ser una polimerasa. En algunos casos, el procedimiento comprende además insertar las proteínas que tienen las segundas subunidades de proporción en una bicapa.

En algunas situaciones, un nanoporo puede comprender una pluralidad de subunidades. Se puede unir una polimerasa a una de las subunidades y al menos una y menos que todas las subunidades comprenden una primera marca de purificación. En algunos ejemplos, el nanoporo es hemolisina alfa o una variante de la misma. En algunos casos, todas las subunidades comprenden una primera marca de purificación o una segunda marca de purificación. La primera marca de purificación puede ser una marca polihistidina (por ejemplo, en la subunidad que tiene la polimerasa unida).

#### **Polimerasa unida a nanoporo**

En algunos casos, una polimerasa (por ejemplo, ADN polimerasa) se une y/o se localiza próxima al nanoporo. La polimerasa se puede unir al nanoporo antes o después de que el nanoporo se incorpore en la membrana. En algunos casos, el nanoporo y la polimerasa son una proteína de fusión (es decir, una única cadena polipeptídica).

La polimerasa se puede unir al nanoporo de cualquier forma adecuada. En algunos casos, la polimerasa se une al monómero proteico de nanoporo (por ejemplo, hemolisina) y a continuación se ensambla el heptámero de nanoporo completo (por ejemplo, en una proporción de un monómero con una polimerasa unida con respecto a 6 monómeros de nanoporo (por ejemplo, hemolisina) sin una polimerasa unida). El heptámero de nanoporo se puede insertar a continuación en la membrana.

Otro procedimiento para unir una polimerasa a un nanoporo implica unir una molécula conectora a un monómero de hemolisina o mutar un monómero de hemolisina para tener un sitio de unión y a continuación ensamblar el heptámero de nanoporo completo (por ejemplo, en una proporción de un monómero con conector y/o sitio de unión con respecto a 6 monómeros de hemolisina sin conector y/o sitio de unión). A continuación se puede unir una polimerasa al sitio de unión o conector de unión (por ejemplo, en bloque, antes de insertarlo en la membrana). La polimerasa también se puede unir al sitio de unión o conector de unión después de que el nanoporo (por ejemplo, heptámero) se forme en la membrana. En algunos casos, se inserta una pluralidad de pares nanoporo-polimerasa en una pluralidad de membranas (por ejemplo, dispuestas sobre los pocillos y/o electrodos) del biochip. En algunos casos, la unión de la polimerasa al complejo de nanoporo se produce en el biochip sobre cada electrodo.

La polimerasa se puede unir al nanoporo con cualquier química adecuada (por ejemplo, enlace covalente y/o conector). En algunos casos, la polimerasa se une al nanoporo con grapas moleculares. En algunos casos, las grapas moleculares comprenden tres secuencias de aminoácidos (indicados como conectores A, B y C). El conector A se puede extender desde un monómero de hemolisina, el conector B se puede extender desde la polimerasa, y el conector C puede unir a continuación los conectores A y B (por ejemplo, envolviendo ambos conectores A y B) y por tanto, la polimerasa al nanoporo. El conector C también se puede construir para formar parte del conector A o conector B, reduciendo así el número de moléculas conectoras.

En algunos casos, la polimerasa se une al nanoporo usando la química Solulink™. Solulink™ puede ser una reacción entre HyNic (ácido 6-hidracino-nicotínico, una hidracina aromática) y 4FB (4-formilbenzoato, un aldehído aromático).

En algunos casos, la polimerasa se une al nanoporo usando la química *Click* (disponible de LifeTechnologies, por ejemplo). En algunos casos, las mutaciones de dedos de cinc se introducen en la molécula de hemolisina y a continuación se usa una molécula (por ejemplo, una molécula intermedia de ADN) para unir la polimerasa a los sitios de dedos de cinc en la hemolisina.

#### **Configuración del aparato**

El nanoporo se puede formar o incluirse de otro modo en una membrana dispuesta adyacente a un electrodo sensor de un circuito sensor, tal como un circuito integrado. El circuito integrado puede ser un circuito integrado específico de la aplicación (ASIC). En algunos ejemplos, el circuito integrado es un transistor de efecto de campo o un semiconductor de óxido metálico complementario (CMOS). El circuito sensor puede estar situado en un chip u otro dispositivo que tenga el nanoporo, o fuera del chip o dispositivo, tal como en una configuración fuera del chip. El semiconductor puede

ser cualquier semiconductor, incluyendo, sin limitación, semiconductores del grupo IV (por ejemplo, silicio) y grupo III-V (por ejemplo, arseniuro de galio). Véase, por ejemplo, el documento WO 2013/123450, para la configuración del aparato y dispositivo para detectar un nucleótido o marca.

5 Se pueden usar sensores basados en poros (por ejemplo, biochips) para consulta electrónica de moléculas individuales. Un sensor basado en poros puede incluir un nanoporo de la presente divulgación formado en una membrana que se dispone adyacente o próxima a un electrodo sensor. El sensor puede incluir un contraelectrodo. La membrana incluye un lado *trans* (es decir, un lado orientado hacia el electrodo sensor) y un lado *cis* (es decir, un lado orientado hacia el contraelectrodo).

10 En la divulgación experimental que sigue, se aplican las siguientes abreviaturas: eq (equivalentes); M (Molar);  $\mu\text{M}$  (micromolar); N (normal); mol (moles); mmol (milimoles);  $\mu\text{mol}$  (micromoles); nmol (nanomoles); g (gramos); mg (miligramos); kg (kilogramos);  $\mu\text{g}$  (microgramos); l (litros); ml (mililitros);  $\mu\text{l}$  (microlitros); cm (centímetros); mm (milímetros);  $\mu\text{m}$  (micrómetros); nm (nanómetros); °C (grados centígrados); h (horas); min (minutos); s (segundos); ms (milisegundos).

## EJEMPLOS

20 La presente invención se describe en mayor detalle en los siguientes ejemplos que de ningún modo pretenden limitar el alcance de la invención como se reivindica. Las figuras adjuntas se deben considerar como partes integrales de la memoria descriptiva y la descripción de la invención.

Se ofrecen los siguientes ejemplos para ilustrar, pero no para limitar la invención reivindicada.

### 25 Ejemplo 1

#### EXPRESIÓN Y RECUPERACIÓN

30 Este ejemplo ilustra la expresión y recuperación de proteínas de células huésped bacterianas, por ejemplo, *E. coli*.

Se adquirió el ADN que codifica la a-HL natural de una fuente comercial. Se verificó la secuencia por secuenciación.

Construcción de plásmido. Se insertó el gen que codifica una hemolisina  $\alpha$  natural o bien variante en un plásmido pPR-IBA2 (IBA Life Sciences, Alemania) bajo el control del promotor T7.

35 Transformación. Se transformaron células BL21 DE3 de *E. coli* (de Life Technologies) con el vector de expresión que comprende el ADN que codifica la hemolisina  $\alpha$  natural o variante usando técnicas bien conocidas en la técnica. En resumen, se descongelaron las células en hielo (si se congelaron). A continuación, se añadió el ADN deseado (en un vector/plásmido adecuado) directamente a las células competentes (no debe exceder un 5 % del de las células competentes) y se mezcló moviendo el tubo. Se colocaron los tubos en hielo durante 20 minutos. A continuación, se colocaron las células en un baño de agua a 42 °C durante 45 segundos sin mezclar, seguido de colocación de los tubos en hielo durante 2 min. A continuación se transfirieron las células a un tubo de cultivo esterilizado de 15 ml que contenía 0,9 ml de medio SOC (precalentado a temperatura ambiente) y se cultivaron a 37 °C durante 1 h en un agitador. Finalmente, se extendió una alícuota de las células sobre una placa de agar LB que contenía el antibiótico apropiado y se incubaron las placas a 37 °C durante la noche.

40 Expresión de proteínas. Después de la transformación, se recogieron las colonias y se inocularon en un pequeño volumen (por ejemplo, 3 ml) de medio de crecimiento (por ejemplo, caldo LB) que contenía el antibiótico apropiado con agitación a 37 °C, durante la noche.

50 A la mañana siguiente, transferir 1 ml del cultivo durante la noche a 100 ml nuevos de medio de autoinducción, por ejemplo, Magic Media (Life Technologies) que contiene un antibiótico apropiado para seleccionar el plásmido de expresión. Cultivar el cultivo con agitación a 25 °C aproximadamente 16 h, pero esto dependió de los plásmidos de expresión. Se recogieron las células por centrifugación a 3.000 g durante 20 min a 4 °C y se almacenaron a -80 °C hasta su uso.

Purificación. Se lisaron las células por sonicación. Se purificó la hemolisina alfa hasta homogeneidad por cromatografía en columna de afinidad.

### 60 Ejemplo 2

#### VARIANTES T12 Y/O N17

El siguiente ejemplo detalla la introducción de una mutación en un residuo deseado.

65 Mutaciones. La mutagénesis dirigida a sitio se lleva a cabo usando un kit de mutagénesis dirigida a sitio múltiple

QuikChange (Stratagene, La Jolla, CA) para preparar las variantes T12 y/o N17.

Se expresaron las variantes y se purificaron como en el ejemplo 1.

### 5 **Ejemplo 3**

#### **ENSAMBLAJE DEL NANOPORO**

10 Este ejemplo describe el ensamblaje de un nanoporo que comprende seis subunidades variantes de a-HL y una subunidad natural.

Se expresó la a-HL natural como se describe en el ejemplo 1 con SpyTag y marca His y se purificó en una columna de afinidad de cobalto usando un tampón de elución de cobalto (NaCl 200 mM, imidazol 300 mM, tris 50 mM, pH 8). Se expresó la variante de a-HL deseada como se describe en el ejemplo 1 con una marca Strep y se purificó usando una columna de afinidad de StrepTactin en la cromatografía de líquidos de proteína rápida (FPLC) usando un tampón de elución (tris 50 mM, destiobiotina 5 mM, NaCl 200 mM, pH 8). Se almacenaron las proteínas a 4 °C si se usaban en 5 días, de lo contrario se añadió trehalosa al 8 % y se almacenó a -80 °C.

20 Usando aproximadamente 20 mg de proteína total, se mezclaron conjuntamente las soluciones de a-HL natural a las variantes de a-HL deseadas en una proporción de 1:6. Se solubilizó el lípido de difitanoilfosfatidilcolina (DPhPC) en Tris 50 mM, NaCl 200 mM, pH 8 o KCl 150 mM, HEPES 30 mM, pH 7,5 hasta una concentración final de 50 mg/ml y se añadió a la mezcla de monómeros de a-HL hasta una concentración final de 5 mg/ml. Se incubó la mezcla de los monómeros de a-HL a 40 °C durante al menos 10 min. Se aplica la mezcla de hemolisina lipídica a una columna de cromatografía de exclusión por tamaño para separar el lípido de las proteínas oligomerizadas.

### 25 **Ejemplo 4**

#### **UNIÓN DE UNA POLIMERASA**

30 Este ejemplo proporciona la unión de una polimerasa a un nanoporo.

Se puede acoplar la polimerasa al nanoporo por cualquier medio adecuado. Véanse, por ejemplo, los documentos PCT/US2013/068967 (publicado como WO2014/074727; Genia Technologies), PCT/US2005/009702 (publicado como WO2006/028508) y PCT/US2011/065640 (publicado como WO2012/083249; Columbia Univ).

35 Se acopló la polimerasa, por ejemplo, ADN polimerasa phi29, a un nanoporo proteico (por ejemplo, hemolisina alfa), a través de una molécula conectora. Específicamente, se usó el sistema SpyTag y SpyCatcher, que forma espontáneamente enlaces isopeptídicos covalentes en condiciones fisiológicas. Véase, por ejemplo, Li *et al*, J Mol Biol., 23 enero 2014;426(2):309-17.

40 Se expresó la marca His de Sticky phi29 SpyCatcher de acuerdo con el ejemplo 1 y se purificó usando una columna de afinidad de cobalto. Se incubaron la polimerasa SpyCatcher y la proteína oligomerizada SpyTag durante la noche a 4 °C en SrCl<sub>2</sub> 3 mM. A continuación se purifica el complejo 1:6-polimerasa-molde usando cromatografía de exclusión por tamaño.

### 45 **Ejemplo 5**

#### **ACTIVIDAD DE LAS VARIANTES**

50 Este ejemplo muestra la actividad de los nanoporos como se proporciona en el ejemplo 3 (nanoporos con una polimerasa unida).

Se sometieron a ensayo los nanoporos naturales y variantes para determinar el efecto de una mutación en una o más posiciones. Se diseñó el ensayo para medir el tiempo que se tarda en capturar una molécula marcada por una ADN polimerasa unida al nanoporo usando voltajes alternos, es decir, ondas cuadradas.

55 Se formaron las bicapas y se insertaron poros como se describe en el documento WO2015/061510 presentado el 23 de octubre de 2014. Se configuró el dispositivo (o sensor) de nanoporos usado para detectar una molécula (y/o secuenciar un ácido nucleico) como se describe en el documento WO2013123450.

60 Para medir el tiempo que se tarda en capturar un nucleótido marcado por una ADN polimerasa en este complejo de secuenciación, se ha ideado un ensayo que usa voltajes positivos y negativos alternos (ondas cuadradas) para determinar el tiempo que esto tarda. Este complejo de secuenciación se compone de un nanoporo proteico (αHL) que se une a una única ADN polimerasa (véase el ejemplo 4). Los nucleótidos marcados están cargados negativamente y, por lo tanto, se atraen por el nanoporo cuando el voltaje aplicado es de naturaleza positiva, y se repelen cuando el voltaje aplicado al complejo de secuenciación de nanoporo es negativo. Por tanto, se puede medir el tiempo que tarda

65

una marca en enhebrarse en el poro alternando el voltaje entre los potenciales positivo y negativo y determinar cuánto tiempo la corriente del nanoporo está sin obstrucciones (canal abierto) frente a cuando se enhebra la marca (flujo de corriente reducido).

5 Para llevar a cabo este ensayo de "tiempo de enhebrado", se usa el dispositivo de secuenciación Genia con un chip de secuenciación Genia. Se acondicionan los electrodos y se establecen bicapas de fosfolípidos en el chip como se explica en el documento PCT/US2013/026514. Se inserta el complejo de secuenciación Genia en las bicapas siguiendo el protocolo descrito en el documento PCT/US2013/026514 (publicado como WO2013/123450). Se  
10 recopilaron los datos de tiempo de enhebrado mostrados en la presente patente usando un sistema tampón compuesto de HEPES 20 mM pH 7,5, KCl 300 mM, nucleótido marcado 3  $\mu$ M,  $Ca^{2+}$  3 mM, con un voltaje aplicado de  $\pm$  100 mV con un ciclo de trabajo de 5 Hz. Después de que se recopilaran los datos, se analizaron para determinar las ondas cuadradas que mostraban la captura de un nucleótido marcado (nivel enhebrado) que duró hasta el final de la porción positiva de la onda cuadrada, y se siguió de otra captura de marca en la onda cuadrada posterior. Se midió el tiempo de enhebrado determinando cuánto tiempo la segunda onda cuadrada informó de corriente de canal abierto sin  
15 obstrucciones. Como ejemplo, si 10 ondas cuadradas consecutivas mostraron capturas de nucleótidos marcados que duraron hasta el final de la porción positiva de la onda cuadrada, entonces el parámetro de tiempo de enhebrado se calcularía a partir de las ondas cuadradas 2-10 (la primera onda cuadrada no se tiene en cuenta en el cálculo porque la polimerasa no tenía una marca unida a ella en la onda cuadrada anterior). A continuación se recopilaron esos números de tiempo de enhebrado para todos los poros en el experimento y se extrajeron parámetros estadísticos de ellos (tales como una media, mediana, desviación estándar, etc.).  
20

Los resultados se muestran en las figuras 1-5.

25 Se entiende que los ejemplos y modos de realización descritos en el presente documento son solo para propósitos ilustrativos.

TEXTO LIBRE DEL LISTADO DE SECUENCIAS

**SEQ ID NO: 1 (ADN de WT aHL)**

ATGGCAGATC	TCGATCCCGC	GAAATTAATA	CGACTCACTA	TAGGGAGGCC	50
ACAACGGTTT	CCCTCTAGAA	ATAATTTTGT	TTAACTTTAA	GAAGGAGATA	100
TACAAATGGA	TTCAGATATT	AATATTAATA	CAGGTACAAC	AGATATTGGT	150
TCAAATACAA	CAGTAAAAAC	TGGTGATTTA	GTAACCTATG	ATAAAGAAAA	200
TGGTATGCAT	AAAAAAGTAT	TTTATTCTTT	TATTGATGAT	AAAAATCATA	250
ATAAAAAATT	GTTAGTTATT	CGTACAAAAG	GTACTATTGC	AGGTCAATAT	300
AGAGTATATA	GTGAAGAAGG	TGCTAATAAA	AGTGGTTTAG	CATGGCCATC	350
TGCTTTTAAA	GTTCAATTAC	AATTACCTGA	TAATGAAGTA	GCACAAATTT	400
CAGATTATTA	TCCACGTAAT	AGTATTGATA	CAAAGAATA	TATGTCAACA	450
TTAACTTATG	GTTTTAATGG	TAATGTAACA	GGTGATGATA	CTGGTAAAAT	500
TGGTGGTTTA	ATTGGTGCTA	ATGTTTCAAT	TGGTCATACA	TTAAAATATG	550
TACAACCAGA	TTTTAAAACA	ATTTTAGAAA	GTCCTACTGA	TAAAAAAGTT	600
GGTTGAAAAG	TAATTTTTAA	TAATATGGTT	AATCAAAATT	GGGGTCCTTA	650
TGATCGTGAT	AGTTGGAATC	CTGTATATGG	TAATCAATTA	TTTATGAAAA	700
CAAGAAATGG	TTCTATGAAA	GCAGCTGATA	ATTTCTTAGA	TCCAAATAAA	750
GCATCAAGTT	TATTATCTTC	AGGTTTTTCT	CCTGATTTTG	CAACAGTTAT	800
TACTATGGAT	AGAAAAGCAT	CAAAACAACA	AACAAATATT	GATGTTATTT	850
ATGAACGTGT	AAGAGATGAT	TATCAATTAC	ATTGGACATC	AACTAATTGG	900
5 AAAGGTACAA	ATACTAAAGA	TAAATGGACA	GATAGAAGTT	CAGAAAGATA	950
TAAAATTGAT	TGGGAAAAAG	AAGAAATGAC	AAATGGTCTC	AGCGCTTGGA	1000
GCCACCCGCA	GTTCGAAAAA	TAA	1023		

**SEQ ID NO: 2 (aminoácidos de WT aHL) [como se expresa en E. coli]**

MADSDINIKT	GTTDIGSNTT	VKTGDLVTYD	KENGMHKKVF	YSFIDDKNHN	50
KKLLVIRTKG	TIAGQYRVYS	EEGANKSGLA	WPSAFKVQLQ	LPDNEVAQIS	100
DYYPRNSIDT	KEYMSTLTYG	FNGNVTGDDT	GKIGGLIGAN	VSIGHTLKYY	150
QPDFKTILES	PTDKKVGWKV	IFNNMVNQNW	GPYDRDSWNP	VYGNQLFMKT	200
RNGSMKAADN	FLDPNKASSL	LSSGFSPDFA	TVITMDRKAS	KQQTNIIDVIY	250
10 ERVRDDYQLH	WTSTNWKGTN	TKDKWTRSS	ERYKIDWEKE	EMTNGLSAWS	300
HPQFEK					306

**SEQ ID NO: 3 (secuencia de WT aHL madura para numeración)**

ADSDINIKTG TTDIGSNTTV KTGDLVITYDK ENGMHKKVIFY SFIDDKNHNK 50  
 KLLVIRTKGT IAGQYRVYSE EGANKSGLAW PSAFKVQLQL PDNEVAQISD 100  
 YYPRNSIDTK EYMSTLTYGF NGNVTGDDTG KIGGLIGANV SIGHTLKYVQ 150  
 PDFKTILESP TDKKVGWKVI FNNMVNQNWG PYDRDSWNPV YGNQLFMKTR 200  
 NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK QQTNIDVIYE 250  
 RVRDDYQLHW TSTNWKGTNT KDKWTDRSSE RYKIDWEKEE MTNGLSAWSH 300  
 PQFEK 305

5 **SEQ ID NO: 4 (aminoácidos de N17K aHL)**

ADSDINIKTG TTDIGSKTTV KTGDLVITYDK ENGMHKKVIFY SFIDDKNHNK 50  
 KLLVIRTKGT IAGQYRVYSE EGANKSGLAW PSAFKVQLQL PDNEVAQISD 100  
 YYPRNSIDTK EYMSTLTYGF NGNVTGDDTG KIGGLIGANV SIGHTLKYVQ 150  
 PDFKTILESP TDKKVGWKVI FNNMVNQNWG PYDRDSWNPV YGNQLFMKTR 200  
 NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK QQTNIDVIYE 250  
 RVRDDYQLHW TSTNWKGTNT KDKWTDRSSE RYKIDWEKEE MTNGLSAWSH 300  
 PQFEK 305

10 **SEQ ID NO: 5 (aminoácidos de N17R aHL)**

ADSDINIKTG TTDIGSRTTV KTGDLVITYDK ENGMHKKVIFY SFIDDKNHNK 50  
 KLLVIRTKGT IAGQYRVYSE EGANKSGLAW PSAFKVQLQL PDNEVAQISD 100  
 YYPRNSIDTK EYMSTLTYGF NGNVTGDDTG KIGGLIGANV SIGHTLKYVQ 150  
 PDFKTILESP TDKKVGWKVI FNNMVNQNWG PYDRDSWNPV YGNQLFMKTR 200  
 NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK QQTNIDVIYE 250  
 RVRDDYQLHW TSTNWKGTNT KDKWTDRSSE RYKIDWEKEE MTNGLSAWSH 300  
 PQFEK 305

**SEQ ID NO: 6 (aminoácidos de T12K aHL)**

ADSDINIKTG TKDIGSNTTV KTGDLVITYDK ENGMHKKVIFY SFIDDKNHNK 50  
 15 KLLVIRTKGT IAGQYRVYSE EGANKSGLAW PSAFKVQLQL PDNEVAQISD 100  
 YYPRNSIDTK EYMSTLTYGF NGNVTGDDTG KIGGLIGANV SIGHTLKYVQ 150  
 PDFKTILESP TDKKVGWKVI FNNMVNQNWG PYDRDSWNPV YGNQLFMKTR 200  
 NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK QQTNIDVIYE 250  
 RVRDDYQLHW TSTNWKGTNT KDKWTDRSSE RYKIDWEKEE MTNGLSAWSH 300  
 PQFEK 305

**SEQ ID NO: 7 (aminoácidos de T12R aHL)**

ADSDINIKTG **TR**DIGSNTTV KTGDLVITYDK ENGMHKKVIFY SFIDDKNHNK 50  
 KLLVIRTKGT IAGQYRVYSE EGANKSGLAW PSFAFKVQLQL PDNEVAQISD 100  
 YYPRNSIDTK EYMSTLTYGF NGNVTGDDTG KIGGLIGANV SIGHTLKYVQ 150  
 PDFKTILESP TDKKVGWKVI FNNMVNQNWG PYDRDSWNPV YGNQLFMKTR 200  
 NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK QQTNIDVIYE 250  
 RVRDDYQLHW TSTNWKGTNT KDKWTDRSSE RYKIDWEKEE MTNGLSAWSH 300  
 PQFEK 305

5 **SEQ ID NO: 8 (WT aHL madura; AAA26598)**

ADSDINIKTG **TT**DIGSNTTV KTGDLVITYDK ENGMHKKVIFY SFIDDKNHNK 50  
 KLLVIRTKGT IAGQYRVYSE EGANKSGLAW PSFAFKVQLQL PDNEVAQISD 100  
 YYPRNSIDTK EYMSTLTYGF NGNVTGDDTG KIGGLIGANV SIGHTLKYVQ 150  
 PDFKTILESP TDKKVGWKVI FNNMVNQNWG PYDRDSWNPV YGNQLFMKTR 200  
 NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK QQTNIDVIYE 250  
 RVRDDYQLHW TSTNWKGTNT KDKWTDRSSE RYKIDWEKEE MTN 293

**LISTA DE CITAS**

10 Literatura de patente

[1] Documento PCT/US2013/026514 (publicado como WO2013/123450) titulado "Methods for Creating Bilayers for Use with Nanopore Sensors"

15 [2] Documento PCT/US2013/068967 (publicado como WO 2014/074727) titulado "Nucleic Acid Sequencing Using Tags"

20 [3] Documento PCT/US14/61853 presentado el 23 de octubre de 2014 titulado "Methods for Forming Lipid Bilayers on Biochips"

Literatura distinta de patente

25 [4] Aksimentiev y Schulten, *Imaging  $\alpha$ -Hemolysin with Molecular Dynamics: Ionic Conductance, Osmotic Permeability, and the Electrostatic Potential Map*, Biophysical Journal (2005) 88: 3745-3761.

[5] Butler *et al.*, *Single-molecule DNA detection with an engineered MspA protein nanopore*, PNAS (2008) 105(52): 20647-20652.

30 [6] Korchev *et al.*, *Low Conductance States of a Single Ion Channel are not 'Closed'*, J. Membrane Biol. (1995) 147:233-239.

[7] Krasilnikov y Sabirov, *Ion Transport Through Channels Formed in Lipid Bilayers by Staphylococcus aureus Alpha-Toxin*, Gen. Physiol. Biophys. (1989) 8:213-222.

35 [8] Nakane *et al.*, *A Nanosensor for Transmembrane Capture and Identification of Single Nucleic Acid Molecules*, Biophys. J. (2004) 87:615-621.

[9] Rhee y Burns, *Nanopore sequencing technology: nanopore preparations*, TRENDS in Biotech. (2007) 25(4):174-181.

40 [10] Song *et al.*, *Structure of Staphylococcal  $\alpha$ -Hemolysin, a Heptameric Transmembrane Pore*, Science (1996) 274:1859-1866.

45 [11] Kasianowicz *et al.*, *Nanometer-scale pores: potential applications for analyte detection and DNA characterization*,



Proc. Natl. Acad. Sci. USA (1996) 93:13770-13773.

**[12]** Akesson *et al.*, *Microsecond timescale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules*, *Biophys. J.* (1999) 77:3227-3233.

5 **[13]** Meller *et al.*, *Voltage-driven DNA translocations through a nanopore*, *Phys. Rev. Lett.*, 86 (2001), pp. 3435-3438.

**[14]** Howorka *et al.*, *Sequence-specific detection of individual DNA strands using engineered nanopores*, *Nat. Biotechnol.*, 19 (2001a), pp. 636-639.

10 **[15]** Howorka *et al.*, *Kinetics of duplex formation for individual DNA strands within a single protein nanopore*, *Proc. Natl. Acad. Sci. USA*, 98 (2001b), pp. 12996-13001.

15 **[16]** Movileanu *et al.*, *Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore*, *Nat. Biotechnol.*, 18 (2000), pp. 1091-1095.

LISTADO DE SECUENCIAS

<110> GENIA TECHNOLOGIES, INC.

5 <120> VARIANTES DE ALFA-HEMOLISINA CON CARACTERÍSTICAS ALTERADAS

<130> 20-04338.P519WO1

10 <140>  
<141>

<150> 62/073.936  
<151> 31/10/2014

15 <160> 8

<170> PatentIn versión 3.5

20 <210> 1  
<211> 1023  
<212> ADN  
<213> Staphylococcus aureus

<400> 1

atggcagatc	tcgatcccgc	gaaattaata	cgactcacta	tagggaggcc	acaacggttt	60
ccctctagaa	ataattttgt	ttaactttaa	gaaggagata	tacaaatgga	ttcagatatt	120
aatattaana	caggtacaac	agatattggt	tcaaatacaa	cagtaaaaac	tggtgattta	180
gtaacttatg	ataaagaaaa	tggtatgcat	aaaaaagtat	tttattcttt	tattgatgat	240
aaaaatcata	ataaaaaaatt	gttagtatt	cgtacaaaag	gtactattgc	aggtcaatat	300
agagtatata	gtgaagaagg	tgctaataaa	agtggtttag	catggccatc	tgcttttaaa	360
gttcaattac	aattacctga	taatgaagta	gcacaaattt	cagattatta	tccacgtaat	420
agtattgata	caaaagaata	tatgtcaaca	ttaacttatg	gttttaatgg	taatgtaaca	480
ggtgatgata	ctggtaaaat	tggtggttta	attggtgcta	atgtttcaat	tggtcataca	540
ttaaaatatg	tacaaccaga	ttttaaaaca	attttagaaa	gtcctactga	taaaaaagtt	600
ggttggaag	taatttttta	taatatgggt	aatcaaaaatt	ggggtcotta	tgatcgtgat	660
agttggaatc	ctgtatatgg	taatcaatta	tttatgaaaa	caagaaatgg	ttctatgaaa	720
gcagctgata	atctcttaga	tccaaataaa	gcatcaagtt	tattatcttc	aggtttttct	780
cctgattttg	caacagttat	tactatggat	agaaaagcat	caaaacaaca	aacaaatatt	840
gatgttattt	atgaacgtgt	aagagatgat	tatcaattac	attggacatc	aactaattgg	900
aaaggtacaa	atactaaaga	taaatggaca	gatagaagtt	cagaaagata	taaaattgat	960
tgggaaaaag	aagaaatgac	aatgggtctc	agcgcttgga	gccaccgca	gttcgaaaaa	1020
taa						1023

25

<210> 2  
<211> 306  
<212> PRT

30 <213> Staphylococcus aureus

ES 2 774 802 T3

<400> 2

Met Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly  
 1 5 10 15

Ser Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu  
 20 25 30

Asn Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn  
 35 40 45

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

Gln Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala  
 65 70 75 80

Trp Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val  
 85 90 95

Ala Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu  
 100 105 110

Tyr Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp  
 115 120 125

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

His Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser  
 145 150 155 160

Pro Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val  
 165 170 175

Asn Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr  
 180 185 190

Gly Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala  
 195 200 205

Asp Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly  
 210 215 220

Phe Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser  
 225 230 235 240

ES 2 774 802 T3

Lys Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp  
 245 250 255

Tyr Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys  
 260 265 270

Asp Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu  
 275 280 285

Lys Glu Glu Met Thr Asn Gly Leu Ser Ala Trp Ser His Pro Gln Phe  
 290 295 300

Glu Lys  
 305

<210> 3

<211> 305

<212> PRT

<213> Staphylococcus aureus

<400> 3

Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser  
 1 5 10 15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn  
 20 25 30

Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His  
 35 40 45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln  
 50 55 60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp  
 65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala  
 85 90 95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr  
 100 105 110

Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp  
 115 120 125

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

5

ES 2 774 802 T3

Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro  
145 150 155 160

Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn  
165 170 175

Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly  
180 185 190

Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp  
195 200 205

Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe  
210 215 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys  
225 230 235 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr  
245 250 255

Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp  
260 265 270

Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys  
275 280 285

Glu Glu Met Thr Asn Gly Leu Ser Ala Trp Ser His Pro Gln Phe Glu  
290 295 300  
Lys  
305

- 5 <210> 4
- <211> 305
- <212> PRT
- <213> Staphylococcus aureus

<400> 4  
Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser  
1 5 10 15

Lys Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn  
20 25 30

Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His  
35 40 45

10 Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln  
50 55 60

ES 2 774 802 T3

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

<210> 5

ES 2 774 802 T3

<211> 305  
 <212> PRT  
 <213> Staphylococcus aureus

5 <400> 5  
 Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser  
 1 5 10 15  
 Arg Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn  
 20 25 30  
 Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His  
 35 40 45  
 Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln  
 50 55 60  
 Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp  
 65 70 75 80  
 Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala  
 85 90 95  
 Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr  
 100 105 110  
 Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp  
 115 120 125  
 Thr Gly Lys Ile Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His  
 130 135 140  
 Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro  
 145 150 155 160  
 Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn  
 165 170 175  
 Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly  
 180 185 190  
 Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp  
 195 200 205  
 Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe  
 210 215 220

ES 2 774 802 T3

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys  
225 230 235 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr  
245 250 255

Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp  
260 265 270

Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys  
275 280 285

Glu Glu Met Thr Asn Gly Leu Ser Ala Trp Ser His Pro Gln Phe Glu  
290 295 300  
Lys  
305

- 5 <210> 6
- <211> 305
- <212> PRT
- <213> Staphylococcus aureus

<400> 6  
Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Lys Asp Ile Gly Ser  
1 5 10 15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn  
20 25 30

Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His  
35 40 45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln  
50 55 60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp  
65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala  
85 90 95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr  
100 105 110

10 Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp  
115 120 125



ES 2 774 802 T3

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

Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro  
 145 150 155 160

Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn  
 165 170 175

Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly  
 180 185 190

Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp  
 195 200 205

Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe  
 210 215 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys  
 225 230 235 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr  
 245 250 255

Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp  
 260 265 270

Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys  
 275 280 285

Glu Glu Met Thr Asn Gly Leu Ser Ala Trp Ser His Pro Gln Phe Glu  
 290 295 300

Lys  
 305

<210> 7

<211> 305

5 <212> PRT

<213> Staphylococcus aureus

<400> 7

Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Arg Asp Ile Gly Ser  
 1 5 10 15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn  
 20 25 30

ES 2 774 802 T3

Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His  
 35 40 45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln  
 50 55 60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp  
 65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala  
 85 90 95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr  
 100 105 110

Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp  
 115 120 125

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

Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro  
 145 150 155 160

Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn  
 165 170 175

Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly  
 180 185 190

Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp  
 195 200 205

Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe  
 210 215 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys  
 225 230 235 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr  
 245 250 255

Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp  
 260 265 270

Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys  
 275 280 285

Glu Glu Met Thr Asn Gly Leu Ser Ala Trp Ser His Pro Gln Phe Glu  
 290 295 300

Lys  
305

<210> 8  
<211> 293  
<212> PRT  
<213> Staphylococcus aureus

5

<400> 8  
Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser  
1 5 10 15  
  
Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn  
20 25 30  
  
Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His  
35 40 45  
  
Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln  
50 55 60  
  
Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp  
65 70 75 80  
  
Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala  
85 90 95  
  
Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr  
100 105 110  
  
Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp  
115 120 125  
  
Thr Gly Lys Ile Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His  
130 135 140  
  
Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro  
145 150 155 160  
  
Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn  
165 170 175  
  
Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly  
180 185 190

ES 2 774 802 T3

Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp  
195 200 205

Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe  
210 215 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys  
225 230 235 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr  
245 250 255

Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp  
260 265 270

Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys  
275 280 285

Glu Glu Met Thr Asn  
290

## REIVINDICACIONES

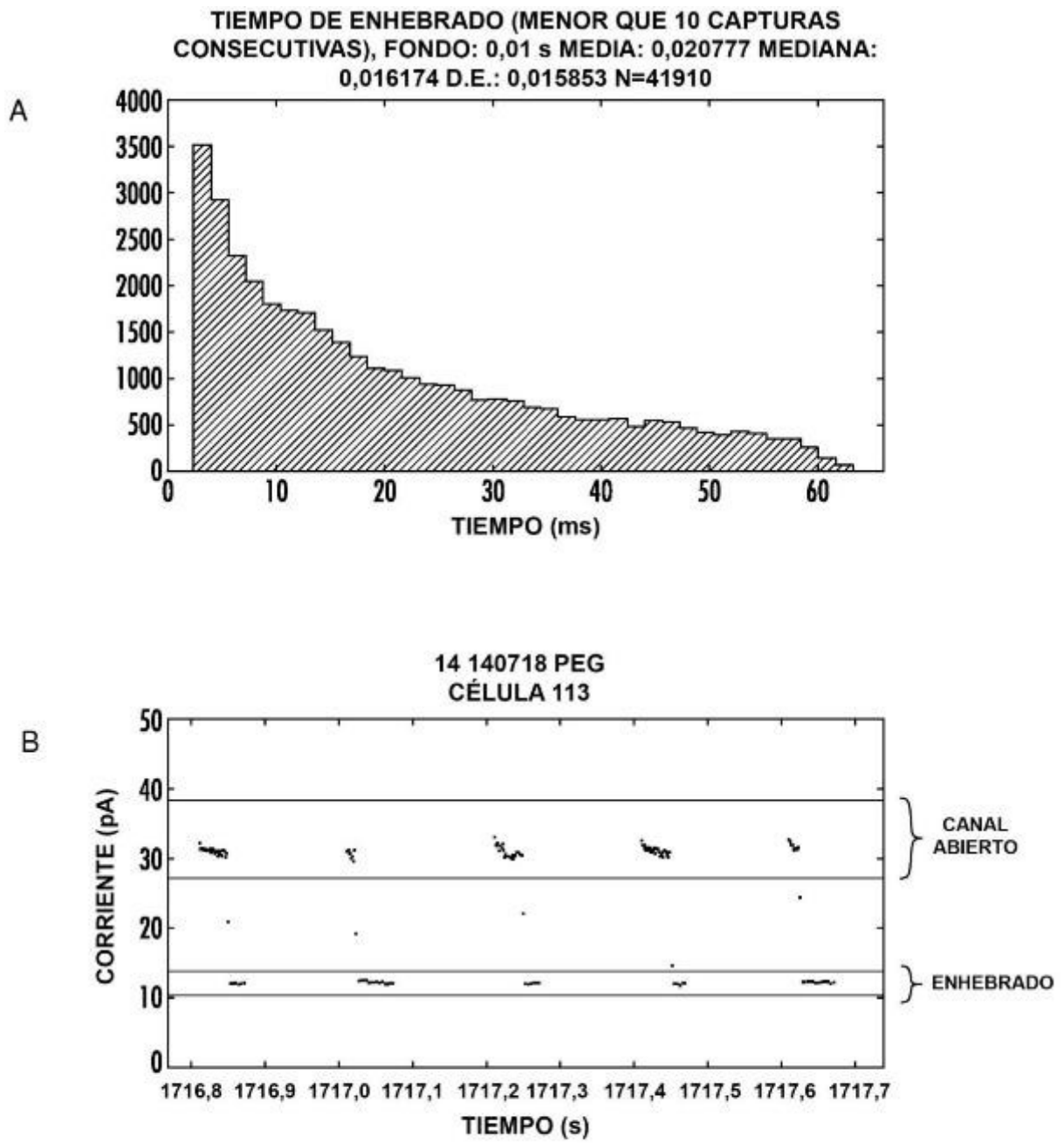
- 5 1. Una variante de hemolisina  $\alpha$  ( $\alpha$ -HL) que comprende una sustitución en una posición correspondiente a la posición 12 o 17 de SEQ ID NO:3, en la que la sustitución comprende una o más cargas positivas.
2. La variante de hemolisina  $\alpha$  de acuerdo con la reivindicación 1, en la que dicha variante comprende además H144A.
- 10 3. La variante de hemolisina  $\alpha$  de acuerdo con la reivindicación 1 o 2, en la que dicha variante comprende una sustitución en una posición correspondiente a los residuos T12 y N17.
4. La variante de hemolisina  $\alpha$  de acuerdo con la reivindicación 1 o 2, en la que dicha variante comprende una sustitución seleccionada de T12K, T12R, N17K, N17R y combinaciones de las mismas.
- 15 5. La variante de hemolisina  $\alpha$  ( $\alpha$ -HL) de acuerdo con la reivindicación 1, en la que la sustitución es T12K.
6. La variante de hemolisina  $\alpha$  ( $\alpha$ -HL) de acuerdo con la reivindicación 1, en la que la sustitución es T12R.
- 20 7. La variante de hemolisina  $\alpha$  ( $\alpha$ -HL) de acuerdo con la reivindicación 1, en la que la sustitución es N17K.
8. La variante de hemolisina  $\alpha$  ( $\alpha$ -HL) de acuerdo con la reivindicación 1, en la que la sustitución es N17R.
9. La variante de hemolisina  $\alpha$  ( $\alpha$ -HL) de acuerdo con la reivindicación 1, en la que la variante tiene una secuencia que tiene al menos un 80 %, 90 %, 95 %, 98 % o más de identidad de secuencia con SEQ ID NO: 8.
- 25 10. Una variante de hemolisina alfa de acuerdo con las reivindicaciones 1-9, que se une covalentemente a una ADN polimerasa.
11. Una variante de hemolisina alfa de acuerdo con la reivindicación 10, en la que dicha variante se une a la ADN polimerasa por medio de un enlace isopeptídico.
- 30 12. Un ensamblaje de nanoporos heptamérico que comprende al menos una variante de hemolisina  $\alpha$  ( $\alpha$ -HL) de acuerdo con una cualquiera de las reivindicaciones 1-11.
- 35 13. Un ensamblaje de nanoporos heptamérico de acuerdo con la reivindicación 12, en el que dicho ensamblaje de poros tiene un tiempo de enhebrado (TTT) alterado relativo a un complejo de poros que consiste en hemolisina alfa natural.
- 40 14. El ensamblaje de nanoporos heptamérico de acuerdo con la reivindicación 13, en el que el TTT se disminuye.
15. Un ácido nucleico que codifica una variante de alfa-HL de acuerdo con una cualquiera de las reivindicaciones 1-9.
- 45 16. La molécula de ácido nucleico de la reivindicación 15, en la que dicha molécula de ácido nucleico se deriva de *Staphylococcus aureus* (SEQ ID NO: 1).
17. Un vector que comprende un ácido nucleico que codifica una variante de hemolisina alfa de acuerdo con una cualquiera de las reivindicaciones 15-16.
- 50 18. Una célula huésped transformada con el vector de la reivindicación 17.
19. Un procedimiento de producción de una variante de hemolisina alfa que comprende las etapas de:
- 55 (a) cultivar la célula huésped de acuerdo con la reivindicación 18 en un medio de cultivo adecuado en condiciones adecuadas para producir la variante de hemolisina alfa;
- (b) obtener dicha variante de hemolisina alfa producida.
- 60 20. Un procedimiento para detectar una molécula diana, que comprende:
- (a) proporcionar un chip que comprende un nanoporo de acuerdo con las reivindicaciones 12-14 en una membrana que se dispone adyacente o próxima a un electrodo sensor;
- 65 (b) dirigir una molécula de ácido nucleico a través de dicho nanoporo, en el que dicha molécula de ácido nucleico se asocia con una molécula indicadora, en el que dicha molécula de ácido nucleico comprende una región de dirección y una región de sonda, en el que dicha molécula indicadora se asocia con dicha molécula de ácido nucleico en dicha

región de sonda, y en el que dicha molécula indicadora se acopla a una molécula diana;

(c) secuenciar dicha región de dirección mientras dicha molécula de ácido nucleico se dirige a través de dicho nanoporo para determinar una secuencia de ácido nucleico de dicha región de dirección; e

5

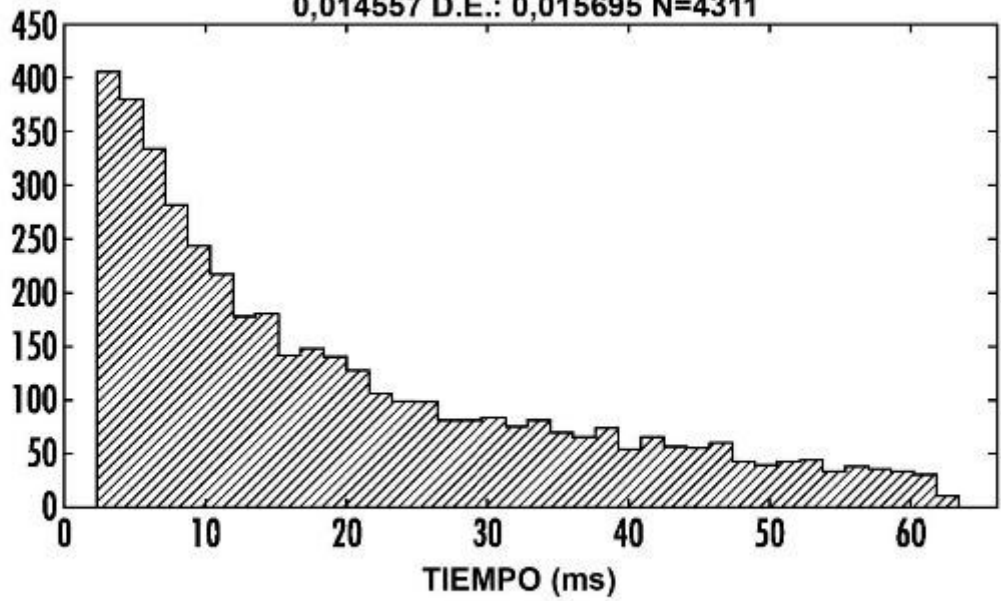
(d) identificar, con la ayuda de un procesador informático, dicha molécula diana en base a una secuencia de ácido nucleico de dicha región de dirección determinada en (c).



**FIG. 1**

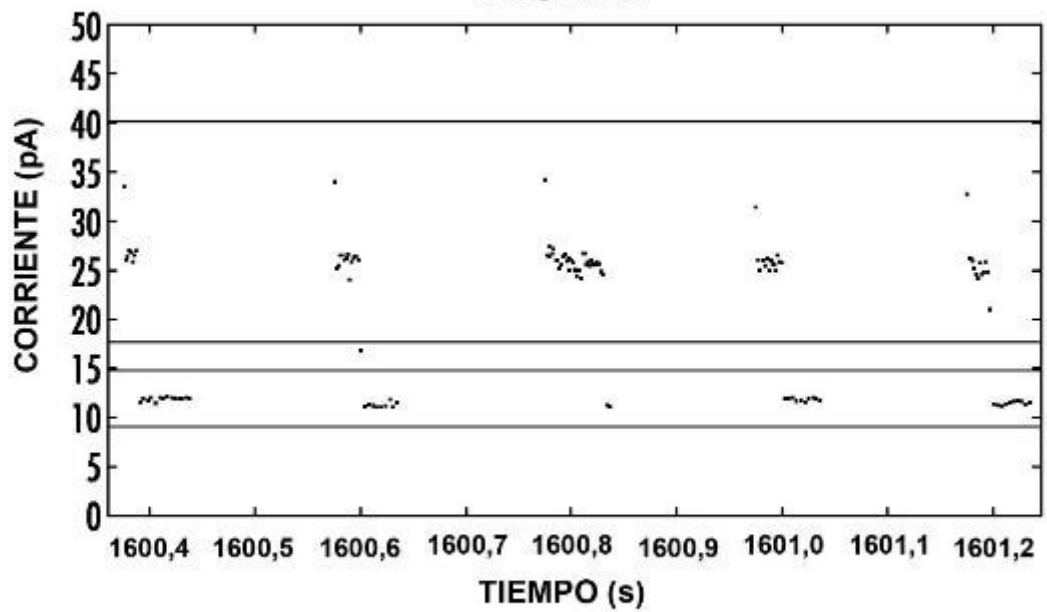
A

TIEMPO DE ENHEBRADO (MENOR QUE 10 CAPTURAS  
CONSECUTIVAS), FONDO: 0,01 s MEDIA: 0,01975 MEDIANA:  
0,014557 D.E.: 0,015695 N=4311



B

03 140725 PEG  
CÉLULA 197

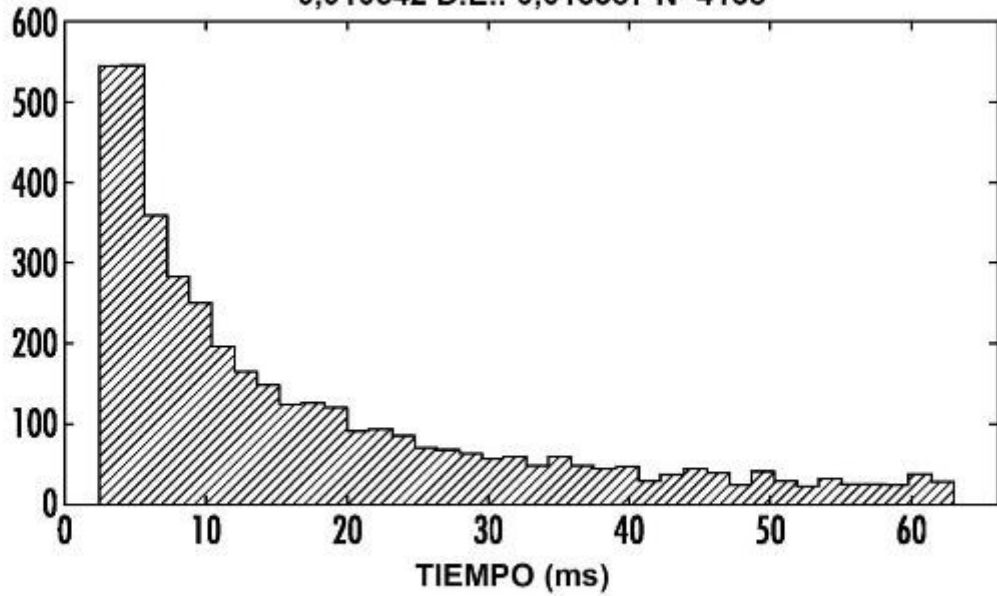


**FIG. 2**

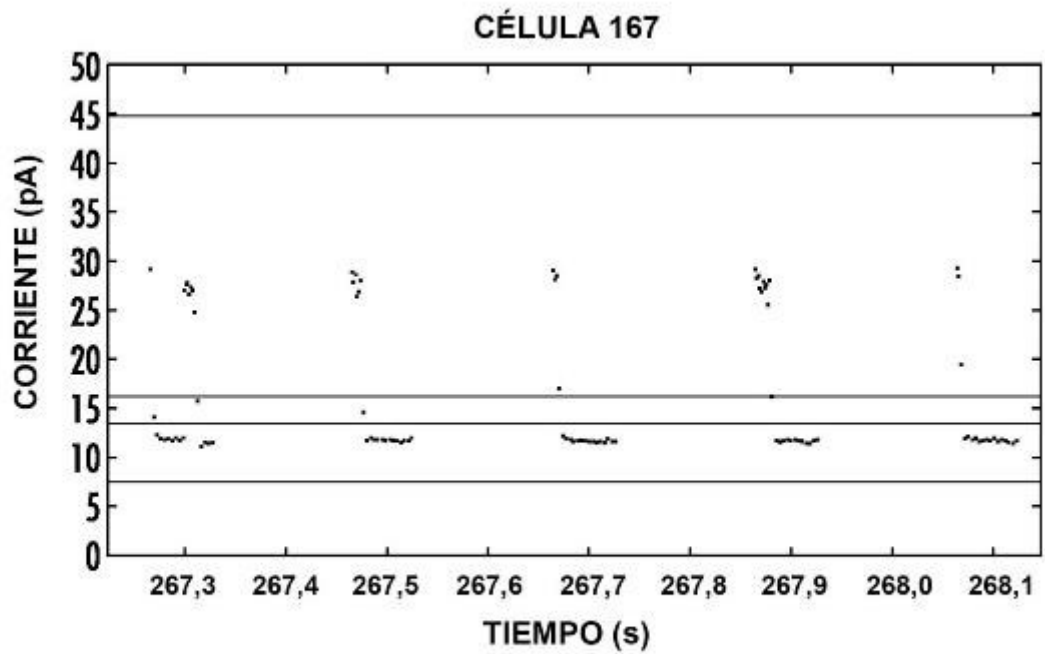


A

TIEMPO DE ENHEBRADO (MENOR QUE 10 CAPTURAS  
CONSECUTIVAS), FONDO: 0,01 s MEDIA: 0,0169 MEDIANA:  
0,010542 D.E.: 0,015387 N=4138



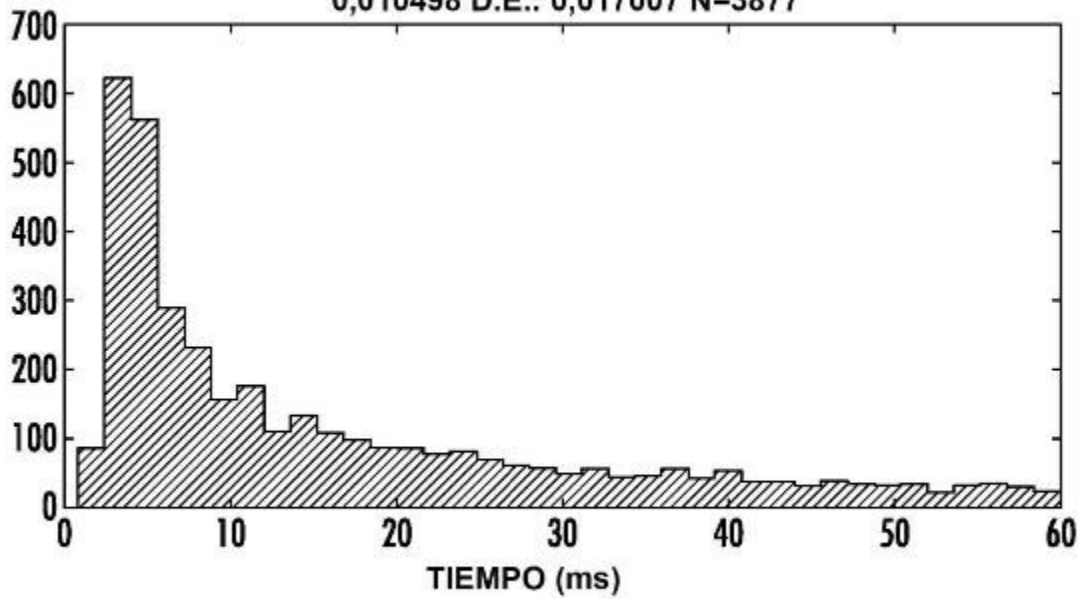
B



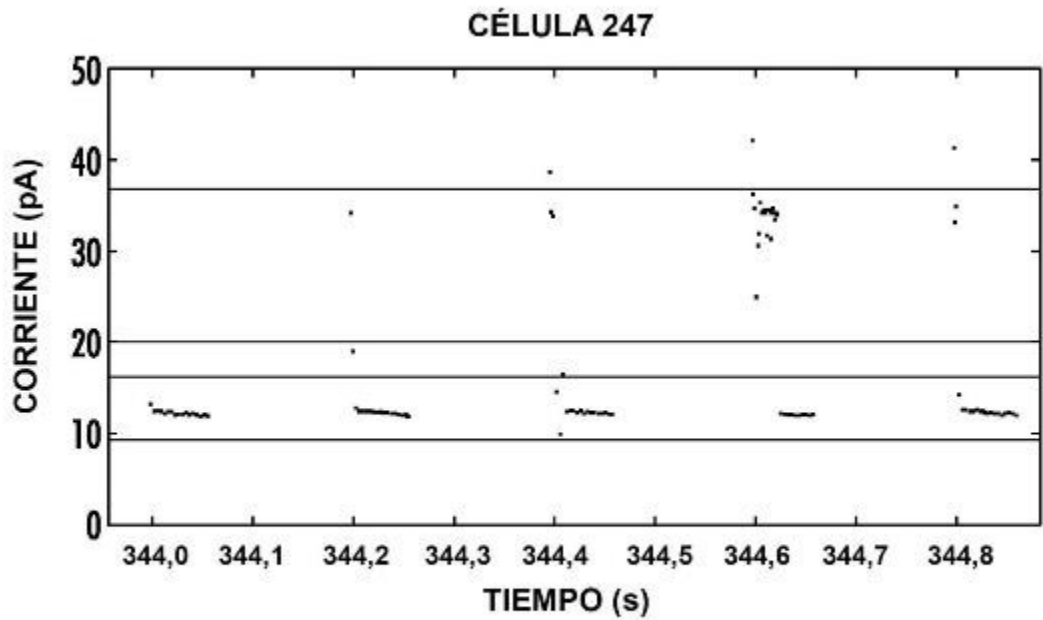
**FIG. 3**

A

TIEMPO DE ENHEBRADO (MENOR QUE 10 CAPTURAS  
CONSECUTIVAS), FONDO: 0,01 s MEDIA: 0,017481 MEDIANA:  
0,010498 D.E.: 0,017007 N=3877



B

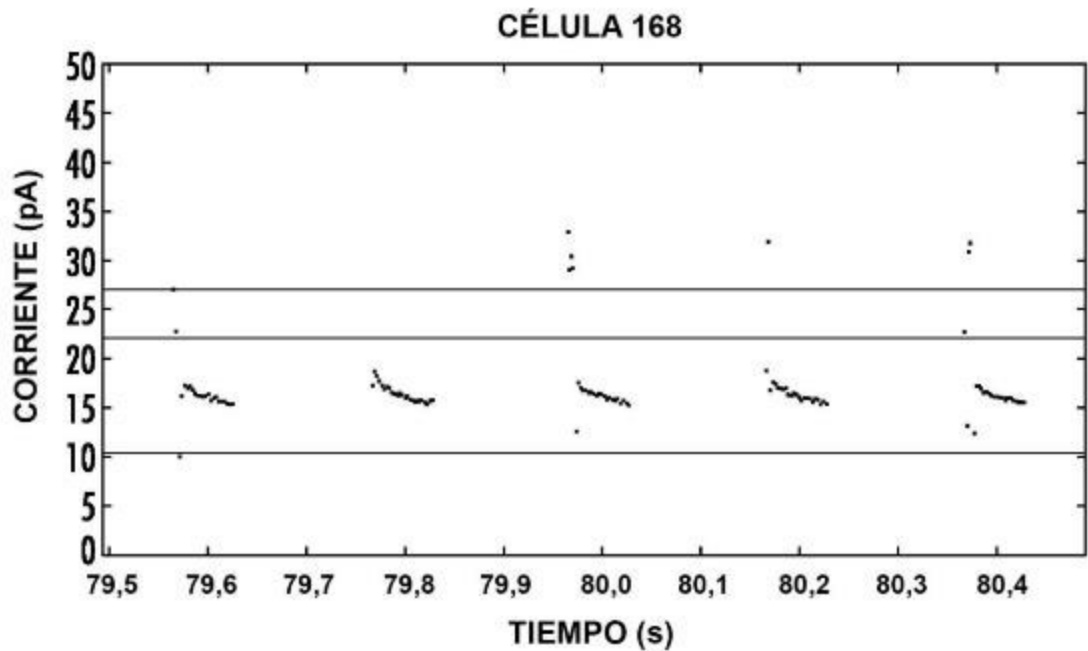


**FIG 4**

A



B



**FIG. 5**