

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 689 676**

51 Int. Cl.:

**G01N 33/557** (2006.01)

**C12Q 1/68** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **22.01.2014 PCT/EP2014/051272**

87 Fecha y número de publicación internacional: **31.07.2014 WO14114687**

96 Fecha de presentación y número de la solicitud europea: **22.01.2014 E 14701960 (8)**

97 Fecha y número de publicación de la concesión europea: **04.07.2018 EP 2948774**

54 Título: **Procedimiento para la detección de modificaciones de ADN por manipulación de una sola molécula**

30 Prioridad:

**22.01.2013 EP 13305074**

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

**15.11.2018**

73 Titular/es:

**CENTRE NATIONAL DE LA RECHERCHE  
SCIENTIFIQUE (CNRS) (33.3%)**

**3, rue Michel Ange**

**75016 Paris, FR;**

**ECOLE NORMALE SUPÉRIEURE (33.3%) y  
SORBONNE UNIVERSITÉ (33.3%)**

72 Inventor/es:

**BENSIMON, DAVID;**

**CROQUETTE, VINCENT;**

**GOUET, HAROLD;**

**ALLEMAND, JEAN-FRANÇOIS y**

**DING, FANG-YUAN**

74 Agente/Representante:

**CURELL AGUILÁ, Mireia**

ES 2 689 676 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

Procedimiento para la detección de modificaciones de ADN por manipulación de una sola molécula.

### 5 Antecedentes de la invención

Se proporciona en la presente memoria un método para determinar si una proteína se une a una secuencia de ADN específica. Este método es útil en particular para identificar modificaciones en la secuencia de ADN (por ejemplo metilaciones) vía la unión de proteínas que reconocen específicamente aquellas modificaciones (por ejemplo anticuerpos), pero también para identificar la secuencia de unión en el ADN de una variedad de proteínas.

La unión de proteínas al ADN es un fenómeno importante en biología; ejerce una función fundamental para regular las funciones celular y vírica. Incluyen procesos celulares fundamentales tales como replicación del ADN, transcripción, reparación del ADN, y recombinación del ADN, pero también modificación del ADN o el mantenimiento de la arquitectura cromosómica.

Existen varias proteínas que se unen a sitios específicos en el genoma para regular la expresión y mantenimiento del genoma. Las proteínas que se unen al ADN constituyen una gran familia de proteínas con funciones biológicas diversas e importantes. La familia de proteínas que se unen al ADN es una de las más pobladas y estudiadas entre los diversos genomas de bacterias, arqueas y eucariotas. La mayoría de estas proteínas, tales como factores de transcripción eucariotas y procariotas, contienen unidades plegadas independientemente (dominios) a fin de lograr su reconocimiento con los contornos del ADN. Incluyen proteínas importantes reguladoras de genes conocidas como factores de transcripción, y proteínas del procesamiento del ADN, tales como, por ejemplo, ADN y ARN polimerasas, ADN ligasas, ADN helicasas, ADN endonucleasas y exonucleasas, y proteínas de reparación y recombinación del ADN.

La identificación de los sitios unidos por estas proteínas ha demostrado ser una tarea desalentadora. Por ejemplo, en el genoma humano, hay más de 700 factores de transcripción de dedos de cinc  $C_2H_2$  predichos (Tadepally et al., BMC Evol. Biol., 8: 176, 2008), pero solamente alrededor de 10% de éstos tienen motivos de unión conocidos (Matys et al., Nucleic Acids Res., 34: D108-D110, 2006). Además, mientras que las propiedades de equilibrio termodinámico de la unión de proteínas al ADN son bien conocidas, la medida de la cinética de su unión y desunión es un problema más desafiante.

Las interacciones de ADN-proteínas se estudian usando una variedad de métodos tales como los ensayos de desplazamiento en gel, huella de pisada, y activación transcripcional (Carey et al., Cold Spring Harb Protoc, 2012(7): 733-57, 2012). Aunque cada uno de estos métodos puede contribuir una información distinta sobre la localización o el efecto de la unión, no proporcionan una manera simple de medir cuantitativamente la unión específica. La polarización/anisotropía fluorescente proporciona un método rápido, no radioactivo, para medir exactamente la unión de ADN-proteínas directamente en disolución sin usar unión al filtro, electroforesis, o etapas de precipitación (Guest et al., 1991; Heyduk y Lee, 1990; LeTilly y Royer, 1993; Lundblad et al., 1996; Royer et al., 1992).

Los mecanismos moleculares mediante los cuales la información genómica dirige la síntesis de diferentes biomoléculas ha sido el foco de gran parte de la investigación de biología molecular a lo largo de las últimas tres décadas. Estudios previos se han concentrado típicamente en genes individuales, proporcionando entonces los principios generales resultantes comprensiones en la transcripción, remodelado de la cromatina, el ajuste del ARN mensajero, la replicación del ADN, y otros numerosos procesos genómicos. Aunque muchos de tales principios parecen válidos a medida que se investigan genes adicionales, generalmente no han proporcionado conocimientos de todo el genoma sobre la función biológica. Por otro lado, los análisis sistemáticos de transcritos e información reguladora son esenciales para la identificación de genes y regiones reguladoras, y son una fuente importante para el estudio de la biología y enfermedad humanas. Tales análisis también pueden proporcionar visiones amplias de la organización y variabilidad de genes e información reguladora a lo largo de contextos celulares, especies e individuos.

Los esfuerzos a nivel de todo el genoma, tal como el proyecto Encode (Enciclopedia de Elementos de ADN) para identificar, por ejemplo, todos los sitios de unión a factores de transcripción en el genoma humano, han demostrado ser engorrosos y que exigen muchísimo trabajo (The ENCODE Project Consortium, Nature, 489: 57-74, 2012). Aunque en la técnica anterior se han descrito métodos de secuenciación que usan la manipulación física, difieren del método descrito en la presente memoria. De hecho, el documento EP 2 390 351 describe un método para la secuenciación mediante hibridación, particularmente detectando el bloqueo debido a la presencia de oligonucleótidos, mientras que Manosas et al., 2009, se refiere a un método para detectar y monitorizar la actividad de T4 primasa. Ninguno de estos documentos describe el método que se refiere al estudio de la unión de proteínas, y más especialmente a la detección de bases modificadas, como se describe en la presente memoria. De este modo, todavía existe la necesidad de un método simple y fiable para detectar interacciones de proteínas/ácidos nucleicos.

## Sumario de la invención

La presente invención se define en las reivindicaciones 1 a 16.

## Descripción detallada de la invención

Se describe en la presente memoria un método para la determinación de la unión de una proteína a una molécula de ácido nucleico por manipulación física.

El método como se describe en la presente memoria, basado en técnicas físicas y tratamientos electrónicos, difiere de los enfoques actuales, que son químicos o bioquímicos. Ofrece numerosas ventajas con respecto a la técnica anterior:

- 1) Es muy sensible, puesto que se basa en la detección de una sola proteína o en una molécula de un complejo proteico frente a una única molécula de ácido nucleico. El uso de una única molécula ofrece la capacidad para medir no solo el tiempo requerido para que una proteína encuentre su diana de ácido nucleico, y el tiempo que permanece en su diana, sino también la localización exacta del suceso de unión.
- 2) No usa nucleótidos marcados caros (ya sea con fluoróforos o algunos otros grupos).
- 3) Permite determinar la localización precisa (en pb) del sitio de unión de la proteína a lo largo de un ácido nucleico bicatenario midiendo la distancia entre los dos extremos de dicha molécula de ácido nucleico bicatenario.
- 4) La medida se puede repetir periódicamente en una escala de tiempo de segundos, conduciendo así a la eliminación de falsos positivos, a una estadística mejorada, y a una reducción significativa en las desviaciones instrumentales.
- 5) El experimento se puede repetir muchas veces en la misma molécula, mejorando así la estadística y la fiabilidad de la medida.
- 6) Permite la detección de cualquier proteína que se une al ácido nucleico. De este modo, se pueden identificar proteínas que reconocen específicamente la modificación estructural del ácido nucleico, conduciendo a la detección de los sitios de la modificación estructural.

Más particularmente, se describe en la presente memoria un método para la detección de la unión de una proteína a una secuencia de ácido nucleico basado en la localización física en la molécula de ácido nucleico secuenciada de los sitios en los que se une la proteína.

En el contexto de la presente invención, "unión" se refiere a una interacción no covalente entre macromoléculas (por ejemplo, entre una proteína y un ácido nucleico). Tales interacciones se caracterizan generalmente por una constante de disociación ( $K_d$ ) de  $10^{-6}$  M<sup>-1</sup> o menor. "Afinidad" se refiere a la fortaleza de la unión: estando una mayor afinidad de unión correlacionada con una menor  $K_d$ .

Por "detección de la unión de una proteína a una molécula de ácido nucleico", se quiere decir en la presente memoria todas las actividades que conduzcan directa o indirectamente a la obtención de alguna información sobre la presencia o ausencia de una interacción entre dicha proteína y dicha molécula de ácido nucleico. La detección de dicha unión puede implicar o no la determinación de información adicional, tal como, por ejemplo, los parámetros cinéticos de la reacción de unión, o la secuencia del sitio unido por la proteína. Como resulta evidente para el experto en la materia, el método descrito en la presente memoria permite que tal determinación se lleve a cabo fácilmente.

El método se basa en la observación de que las dos hebras de un ácido nucleico bicatenario desnaturalizado se rehibridarán en condiciones apropiadas. Si una molécula está unida a cualquiera de las hebras de dicha molécula de ácido nucleico bicatenario desnaturalizado durante la etapa de renaturalización, la rehibridación será solamente parcial. Ahora se ha encontrado que, en ciertas condiciones, esta pausa en la rehibridación, sea esta permanente o transitoria, se puede usar para detectar una interacción entre una proteína y dicha molécula de ácido nucleico bicatenario desnaturalizado. Según el método descrito en la presente memoria, es posible detectar un bloqueo de la rehibridación de la molécula de ácido nucleico bicatenario; los parámetros físicos (por ejemplo la duración del bloqueo, la posición del bloqueo en la molécula de ácido nucleico bicatenario) asociados con este bloqueo permiten entonces la detección de una interacción entre una proteína y la secuencia del ácido nucleico.

De este modo, se describe en la presente memoria un método para la determinación de la unión de una proteína a una molécula de ácido nucleico, comprendiendo dicho método una etapa de detectar un bloqueo de la

renaturalización de la molécula de ácido nucleico bicatenario desnaturalizado.

Por “desnaturalización”, se quiere decir en la presente memoria el proceso de separación de las dos hebras de una molécula de ácido nucleico bicatenario que se produce cuando se rompe la mayoría de los enlaces de hidrógeno entre las mencionadas hebras. El proceso de desnaturalización produce una molécula de ácido nucleico desnaturalizado, mediante lo cual se quiere decir en la presente memoria las dos hebras complementarias separadas que resultan de la desnaturalización de una molécula de ácido nucleico bicatenario. Por “renaturalización”, se refiere en la presente memoria al proceso mediante el cual se vuelven a formar a través de la hibridación en una doble hélice dos hebras complementarias separadas. Como se usa en la presente memoria, “hibridación” es el proceso de establecer una interacción no covalente, específica de la secuencia, entre dos o más hebras complementarias de ácidos nucleicos en un único híbrido.

Existen varias posibilidades conocidas por el experto en la materia para desnaturalizar el ácido nucleico. En una manera más preferida, las dos hebras se separan sometiéndolas a una fuerza física. Una “fuerza física”, según la invención, es cualquier influencia que provoca que un objeto sufra cierto cambio, ya sea relativo a su movimiento, dirección, o construcción geométrica. Resulta evidente para el experto en la materia que una fuerza según la invención es diferente de otros parámetros físicos tales como, por ejemplo, la temperatura (que es una propiedad directa de la materia en lugar de una influencia ejercida sobre ella). Las fuerzas físicas según la invención comprenden fuerzas tales como fricción, tensión, fuerza normal, fuerza de resistencia del aire, fuerza aplicada, y fuerza elástica. Lo más preferible, la fuerza física según la invención es una fuerza de tensión. Según esta forma de realización, los extremos libres del mencionado ácido nucleico bicatenario se pueden separar tirando de ellos, rompiendo así todos los enlaces entre las bases emparejadas, y abriendo el ácido nucleico bicatenario.

El método se aplica a cualquier tipo de ácido nucleico bicatenario. Lo más a menudo, el ácido nucleico bicatenario será ADN, pero se entiende que el método también se aplica a dúplex de ADN monocatenario-ADN bicatenario, perfectamente emparejados o no perfectamente emparejados, o alternativamente, a dúplex de ADN monocatenario-ARN monocatenario, perfectamente emparejados o no perfectamente emparejados, o como alternativa, a dúplex de ARN monocatenario-ARN monocatenario, perfectamente emparejados o no perfectamente emparejados. Además, el dúplex puede consistir en por lo menos el reemparejamiento parcial de dos hebras individuales obtenidas de muestras de orígenes diferentes. Finalmente, el método también se aplica a las estructuras secundarias de un único ADN monocatenario o de un único ARN monocatenario.

De este modo, se describe en la presente memoria un método para la detección de la unión de una proteína a una molécula de ácido nucleico, comprendiendo dicho método las etapas de:

- desnaturalizar una molécula de ácido nucleico bicatenario aplicando una fuerza física a dicha molécula; y
- detectar un bloqueo de la renaturalización del ácido nucleico bicatenario.

Ventajosamente, el mencionado método comprende la etapa adicional de determinar la posición del bloqueo.

En este tipo de método para evaluar la unión de una proteína a una molécula de ADN, puede ser ventajoso, a fin de facilitar el reemparejamiento, disponer los extremos libres de la molécula de ADN bicatenario (es decir, los extremos que no están unidos a los soportes) para que se unan entre sí covalente o casi covalentemente antes de la separación. En una forma de realización preferida, la molécula de ácido nucleico bicatenario es una horquilla. Si se desea que el ácido nucleico bicatenario se represente en forma de diagrama en el contexto del presente método, es posible compararlo con un “cierre de cremallera”, que se abre (o se cierra): la desnaturalización del ácido nucleico bicatenario es la apertura de la cremallera, la renaturalización es el cierre nuevamente de la cremallera.

Los inventores han observado que, en ciertas condiciones, cuando una molécula se une a la molécula de ácido nucleico bicatenario desnaturalizado, la renaturalización de dicha molécula de ácido nucleico bicatenario se bloquea. La molécula unida puede ser cualquier tipo de molécula con una afinidad por una secuencia específica en dicha molécula de ácido nucleico bicatenario desnaturalizado, por ejemplo un ácido nucleico, una proteína o una molécula pequeña.

En un primer aspecto del método descrito en la presente memoria, una proteína se usa para bloquear la renaturalización del mencionado ácido nucleico bicatenario.

Los términos “proteína”, “proteínas”, “polipéptido” y “polipéptidos”, como se usan en la presente memoria, son sinónimos, y se refieren a polímeros de aminoácidos enlazados covalentemente a través de enlaces peptídicos en una cadena. Los enlaces peptídicos se forman entre el grupo carboxilo de un aminoácido y el grupo amino del siguiente aminoácido. Los términos también se aplican a polímeros de aminoácidos en los que uno o más aminoácidos son análogos químicos o derivados modificados de aminoácidos de origen natural correspondientes. Los términos “aminoácidos” y “aminoácido” se refieren a todos los alfa-aminoácidos de origen natural, tanto en sus formas estereoisómeras D como L, y sus análogos y derivados. Un análogo se define como

una sustitución de un átomo en el aminoácido por un átomo diferente que habitualmente tiene propiedades similares. Un derivado se define como un aminoácido que tiene otra molécula o átomo unido a él. Los derivados incluirían, por ejemplo, la acetilación de un grupo amino, la aminación de un grupo carboxilo, u oxidación de los restos de azufre de dos moléculas de cisteína para formar cistina.

5 Las proteínas pueden tener varias funciones. Una "proteína de unión" es una proteína que resulta apta para unirse no covalentemente a otra molécula. Una proteína de unión se puede unir a, por ejemplo, una molécula de ADN (proteína de unión a ADN), una molécula de ARN (proteína de unión a ARN), y/o una molécula proteínica (una proteína de unión a proteína). En el caso de una proteína de unión a proteína, se puede unir a sí misma (para formar multímeros), y/o se puede unir a una o más moléculas de una proteína o proteínas diferentes. Una proteína de unión puede tener más de un tipo de actividad de unión. Por ejemplo, las proteínas de dedo de cinc tienen actividad de unión a ADN, de unión a ARN y de unión a proteína. Una "proteína de unión a ácido nucleico", según el método descrito en la presente memoria, es de este modo una proteína apta para interactuar con un ácido nucleico. Una "proteína de unión a ácido nucleico monocatenario", según el método descrito en la presente memoria, es de este modo una proteína que resulta apta para interactuar con un ácido nucleico monocatenario, mientras que una "proteína de unión a ácido nucleico bicatenario", según el método descrito en la presente memoria, es de este modo una proteína que resulta apta para interactuar con un ácido nucleico bicatenario.

20 Según este ejemplo, se describe un método para la determinación de la unión de una proteína a una molécula de ácido nucleico que comprende una secuencia de ácido nucleico, comprendiendo dicho método las etapas de:

- 25 a) desnaturalizar una mencionada molécula de ácido nucleico bicatenario que comprende dicha secuencia aplicando una fuerza física a dicha molécula;
- b) proporcionar dicha proteína;
- c) renaturalizar dicha molécula de ácido nucleico bicatenario en presencia de dicha proteína, y
- 30 d) detectar un bloqueo de la renaturalización del ácido nucleico bicatenario.

Ventajosamente, el mencionado método comprende la etapa adicional de determinar la posición del bloqueo.

35 Como es bien conocido en el campo, las proteínas de unión a ácido nucleico se pueden distinguir en si son aptas para unirse a ácidos nucleicos monocatenarios (ADNmc y ARNmc), o si son aptas para unirse a ácidos nucleicos bicatenarios (ADNbc, ARNbc, híbridos de ADN/ARN, etc.).

40 En un primer ejemplo del método, la proteína que se usa para bloquear la renaturalización del ácido nucleico bicatenario desnaturalizado es una proteína apta para unirse a ácido nucleico monocatenario.

45 Las proteínas de unión a ácido nucleico con afinidad por ácido nucleico monocatenario serán aptas para interactuar per se con la molécula bicatenaria desnaturalizada, conduciendo así a un bloqueo de la renaturalización del ácido nucleico bicatenario. El experto en la materia apreciará que el presente método permite la determinación fácil y precisa de los parámetros de la cinética de la reacción de unión, incluso si la proteína no se une a una secuencia específica. De hecho, las proteínas de unión a ácido nucleico monocatenario muy a menudo no tienen afinidad por una secuencia específica, sino más bien por ácidos nucleicos en general. Por ejemplo, se sabe que las helicasas se unen a espacios del ADNmc, a fin de desenrollar ADNbc. Las proteínas de unión a ADN monocatenario bacteriano, o SSB, se unen a regiones monocatenarias de ADN para prevenir la recombinación prematura, para proteger al ADN monocatenario de ser digerido por nucleasas, y para eliminar la estructura secundaria del ADN. La proteína Rad52, una proteína importante para la reparación de rupturas de la cadena doble del ADN y para la recombinación homóloga, se une a extremos de ADN monocatenario, y media la interacción ADN-ADN necesaria para la recombinación de hebras de ADN complementarias.

55 Estas proteínas de unión a ácido nucleico monocatenario tienen una afinidad general por ácidos nucleicos, lo que significa en el contexto del presente método que las proteínas son aptas para unirse a un ácido nucleico monocatenario, independientemente de la secuencia del mencionado ácido nucleico. Tal proteína de unión a ácido nucleico no específica de la secuencia se une a una pluralidad de secuencias de ADN no relacionadas, con una constante de disociación que varía en menos de 100 veces, habitualmente menos de diez veces, con respecto a las diferentes secuencias.

60 Por otro lado, algunas proteínas de unión a ácido nucleico tienen afinidad por moléculas de ácido nucleico que contienen una secuencia específica, es decir, solamente reconocen y se unen al ácido nucleico que comprende dicha secuencia. No todos los componentes de una interacción de unión tienen que ser específicos de la secuencia (por ejemplo, contactos con restos de fosfato en una cadena principal de ADN), en tanto que la interacción con el conjunto sea específica de la secuencia. De hecho, aunque un gran número de proteínas de

unión a ácido nucleico monocatenario tienen solamente una afinidad general por ácidos nucleicos, algunas de estas proteínas son aptas para unirse a ácidos nucleicos monocatenarios en secuencias específicas. De este modo, una proteína de unión a ácido nucleico, específica de la secuencia, se une a una secuencia específica o familia de secuencias específicas mostrando un grado elevado de identidad de secuencia entre sí (por ejemplo, por lo menos alrededor de 80% de identidad de secuencia), con una afinidad por lo menos 100 veces mayor que con respecto a secuencias no relacionadas. La constante de disociación de una proteína de unión a ácido nucleico, específica de la secuencia, con respecto a su secuencia o secuencias específicas es habitualmente menor que alrededor de 100 nM, y puede ser tan baja como 10 nM, 1 nM, 1 pM, o 1 fM.

Un gran número de proteínas de unión a ácido nucleico no son aptas para unirse a ácidos nucleicos monocatenarios. Estas proteínas, que en su lugar poseen afinidad por ácidos nucleicos bicatenarios, no serán aptas para interactuar con la molécula bicatenaria desnaturalizada *per se*. Estas proteínas muy probablemente no accionarán un bloqueo de la renaturalización del ácido nucleico bicatenario en estas condiciones.

La mayoría de estas proteínas reconocen y se unen a secuencias de ácidos nucleicos bicatenarios específicas. Por ejemplo, las proteínas de unión a ADN bicatenario desempeñan un papel importante en la regulación de la expresión de nuevas proteínas. Estas proteínas interactúan con ADN por medio de diversos motivos estructurales, y pueden estimular o reprimir la transcripción de ARN mensajero, dependiendo de las propiedades y localización de la secuencia de ADN a la que se unen.

En este caso, puede ser ventajoso proporcionar una molécula de ácido nucleico monocatenario con dicha proteína de unión a ácido nucleico bicatenario, tras desnaturalizar dicha molécula bicatenaria. De hecho, es bien conocido en la técnica que el mencionado ácido nucleico monocatenario se puede hibridar con una secuencia complementaria en una de las hebras del ácido nucleico bicatenario desnaturalizado, formando así un híbrido de ácido nucleico bicatenario que se puede unir por la proteína. Este ácido nucleico monocatenario puede tener cualquier longitud, con la condición de que sea suficientemente largo para bloquear el proceso de renaturalización. Preferentemente, la longitud del ácido nucleico monocatenario estará comprendida entre 3 y 50 nucleótidos; más preferentemente, entre 3 y 45 nucleótidos, entre 3 y 40 nucleótidos, entre 3 y 35 nucleótidos, entre 3 y 30 nucleótidos, entre 3 y 25 nucleótidos, entre 3 y 20 nucleótidos, entre 3 y 15, e incluso más preferentemente entre 3 y 12. El ácido nucleico monocatenario de la invención puede ser en particular una molécula de ADN o de ARN, ya sea natural o modificada. El mencionado ácido nucleico monocatenario también puede estar hecho de nucleótidos modificados, tales como ácido nucleico bloqueado (LNA), que son nucleótidos en los que el resto de ribosa se modifica con un puente extra que conecta el oxígeno de 2' y el carbono de 4', o ácido nucleico peptídico (PNA), en el que la cadena principal está compuesta de unidades de N-(2-aminoetil)-glicina que se repiten enlazadas mediante enlaces peptídicos.

Cuando una molécula de ácido nucleico monocatenario se añade de este modo a un ácido nucleico bicatenario desnaturalizado antes de la renaturalización, un bloqueo de la rehibridación indica que la secuencia de la molécula de ácido nucleico monocatenario es complementaria a por lo menos parte de la secuencia de la molécula de ácido nucleico bicatenario.

En el contexto de la presente invención se ha demostrado que cuando está presente una proteína de unión a ácido nucleico bicatenario, es apta para unirse al híbrido formado entre el ácido nucleico bicatenario desnaturalizado y la molécula de ácido nucleico monocatenario. Esta interacción entre la proteína y el híbrido de ácido nucleico conduce a una alteración de la duración del bloqueo. La mayoría del tiempo, esta interacción conduce a un incremento del bloqueo de la renaturalización. Por ejemplo, una primasa estabilizará oligos de ADN que de otro modo no habrían sido suficientemente estables para bloquear la rehibridación de la horquilla durante un tiempo suficientemente largo para ser detectado. Igualmente, la unión de una ADN polimerasa al extremo 3' de un pequeño oligonucleótido usado como cebador incrementa su estabilidad. Como alternativa, la duración del bloqueo también se puede reducir. De hecho, los presentes inventores han encontrado que la unión de algunas helicasas provoca una desestabilización del mencionado híbrido, lo que se traduce en un tiempo de bloqueo más corto.

Según este ejemplo, el método comprende así las etapas de:

- a) desnaturalizar una molécula de ácido nucleico bicatenario que comprende una secuencia específica aplicando una fuerza física a dicha molécula;
- b) proporcionar dicha proteína y una molécula de ácido nucleico monocatenario que corresponde a dicha secuencia de ácido nucleico;
- c) renaturalizar dicha molécula de ácido nucleico bicatenario en presencia de dicha proteína y de dicha molécula de ácido nucleico monocatenario; y
- d) detectar un bloqueo de la renaturalización del ácido nucleico bicatenario.

Este ejemplo es particularmente ventajoso debido a que permite la determinación de la unión de dicha proteína a la secuencia comprendida dentro del ácido nucleico bicatenario.

En una configuración típica, las moléculas de ácido nucleico bicatenario se pueden anclar específicamente en dos sustratos sólidos (por ejemplo, portaobjetos de microscopio, micropipeta, micropartícula). Uno de los extremos se puede unir directa o indirectamente a la superficie, mientras que el otro extremo se une directa o indirectamente a una superficie móvil. En este ejemplo, se aplica una tensión en ambos extremos del ácido nucleico bicatenario cuando los soportes se alejan. Cuando la tensión es mayor que un valor umbral, las dos hebras se separan y la molécula de ácido nucleico se desnaturaliza. La tensión aplicada está preferentemente por encima de o es igual a 15 pN; más preferentemente, está por encima o es igual a 16 pN; incluso más preferentemente, está por encima o es igual a 17 pN; en un aspecto mucho más preferido, está por encima o es igual a 18 pN. Esta fuerza puede variar con la temperatura, el tipo de nucleótido y amortiguador, pero el experto en la materia adaptará fácilmente dicha fuerza con respecto a estos parámetros a fin de obtener la separación de las dos hebras. Por otro lado, cuando se disminuye la tensión por debajo de un valor mínimo, las dos hebras del ácido nucleico bicatenario desnaturalizado se pueden volver a hibridar. Para obtener una rehibridación de las mencionadas dos hebras, preferentemente se aplica una tensión menor o igual a 12 pN; más preferentemente, es menor o igual a 11 pN; incluso más preferentemente, es menor o igual a 10 pN.

Lo más preferible, el ácido nucleico bicatenario es una horquilla. Como se usa en la presente memoria, "horquilla" significa una hélice doble en la que el extremo 5' de una hebra está enlazado físicamente al extremo 3' de la otra hebra a través de un bucle desemparejado. El mencionado enlace físico puede ser covalente o no covalente. Preferentemente, el mencionado enlace físico es un enlace covalente. De este modo, una horquilla consiste en un tallo bicatenario y un bucle monocatenario desemparejado. En una horquilla, los extremos de las dos hebras que no están acoplados en el bucle están libres, y de este modo se pueden separar. Esto da como resultado el desemparejamiento del ácido nucleico bicatenario, produciendo así una molécula de ácido nucleico bicatenario desnaturalizada. Es posible abrir completamente una molécula de ácido nucleico bicatenario de horquilla al tirar en cada extremo de dicha molécula de ácido nucleico con una fuerza mayor que un valor umbral. Cuando la tensión aplicada a la molécula se disminuye hasta menos de un valor mínimo, la molécula de ácido nucleico se vuelve a hibridar para volver a formar una horquilla. La presencia de una proteína unida a dicha molécula de ácido nucleico desnaturalizado (por ejemplo, ADNmc) conduce a una pausa en la rehibridación. Igualmente, la presencia de una molécula de ácido nucleico bicatenario hibridada a una de las hebras del ácido nucleico de la horquilla abierta conduce a una pausa en la rehibridación, modificándose la duración de dicha pausa (es decir, incrementándose o disminuyéndose) cuando una proteína de unión a ácido nucleico bicatenario se une al complejo. Por lo tanto, la detección de un cambio en la duración de tal pausa indica que una proteína está unida a por lo menos parte del tallo bicatenario.

A este respecto, es ventajoso diseñar la secuencia y longitud del bucle de manera que la horquilla se vuelva a plegar después de una transición corta, por ejemplo 1 segundo. Los métodos de este efecto se han descrito en la técnica anterior, por ejemplo en Woodside et al., Proc. Natl. Acad. Sci. U.S.A., 103 (16): 6190-6195, 2006). Cuando la fuerza se disminuye desde el valor de apertura hasta el valor de ensayo, la extensión de la horquilla abierta varía debido a la elasticidad del ADN monocatenario. El pequeño retraso antes de que la horquilla se vuelva a plegar permite al usuario determinar la extensión de la horquilla a la misma fuerza que la usada para detectar el estado de bloqueo.

El uso de una horquilla hace posible, en particular, llevar a cabo ciclos de apareamiento y desemparejamiento, y de este modo mejorar la relación señal/ruido.

Las técnicas que permiten que los extremos libres del ácido nucleico bicatenario se unan juntos son conocidas, y algunas se describirán con mayor detalle en lo siguiente.

Por determinación del bloqueo, se quiere decir en la presente memoria la determinación de los parámetros físicos asociados con el bloqueo. Un parámetro útil es la posición del bloqueo en la molécula de ácido nucleico bicatenario, correspondiendo dicha posición a la posición de unión de la proteína a la molécula de ácido nucleico bicatenario abierta, o a la hibridación de la molécula de ácido nucleico bicatenario en dicha molécula de ácido nucleico bicatenario abierta. De hecho, se ha encontrado que la posición en el ácido nucleico bicatenario en la que se produce la pausa en la renaturalización se puede determinar de forma precisa: el uso de una horquilla da a la persona un medio para determinar la distancia física entre los dos extremos libres de la horquilla en cualquier momento durante el proceso de desnaturalización/renaturalización.

De este modo, según la presente invención, es particularmente ventajoso que el mencionado método comprenda una etapa adicional de determinar la posición del bloqueo.

Según este ejemplo, se proporciona un método para la determinación de la unión de una proteína a una molécula de ácido nucleico que comprende una secuencia de ácido nucleico, comprendiendo dicho método las etapas de:

- 5
- a) desnaturalizar una molécula de ácido nucleico bicatenario que comprende una secuencia de ácido nucleico aplicando una fuerza física a dicha molécula;
- b) proporcionar dicha proteína;
- c) renaturalizar dicha molécula de ácido nucleico bicatenario en presencia de dicha proteína;
- d) detectar un bloqueo de la renaturalización del ácido nucleico bicatenario; y
- 10 e) determinar la posición del mencionado bloqueo en dicha molécula de ácido nucleico bicatenario.

15 Por "extremo libre" se quiere decir en la presente memoria el extremo de una hebra que no está enlazado covalentemente a una extremidad de la otra hebra; estos extremos libres pueden estar unidos cada uno a una superficie diferente. Por ejemplo, una de estas superficies puede ser móvil, mientras que la otra puede estar inmóvil. El experto en la materia apreciará así que, para medir la distancia entre los extremos libres del ácido nucleico bicatenario de horquilla, es posible medir simplemente la distancia entre las dos superficies.

20 Esta distancia es máxima ( $Z_{\text{elevada}} (F_{\text{abierta}})$ ) cuando la molécula de horquilla está completamente desnaturalizada, puesto que el ácido nucleico de horquilla está entonces completamente extendido; es mínima ( $Z_{\text{baja}} (F_{\text{ensayo}})$ ) cuando dicha molécula de horquilla está completamente renaturalizada. Es ventajoso llevar a cabo todas las comparaciones de longitudes a la misma fuerza  $F_{\text{ensayo}}$ , de manera que el ácido nucleico monocatenario presente las mismas propiedades elásticas. Usando el retraso en el cierre del bucle, el usuario experto puede medir  $Z_{\text{elevada}} (F_{\text{ensayo}})$ . Igualmente, se puede medir la distancia entre los dos extremos libres cuando se pausa temporalmente el proceso de renaturalización: como era de esperar, esta distancia  $z$  está comprendida entre  $Z_{\text{elevada}}$  y  $Z_{\text{baja}}$  (midiéndose todas las  $z$  con  $F = F_{\text{ensayo}}$ ). Inmediatamente está claro que la distancia  $z$  varía con la localización en la molécula de horquilla del sitio de unión de la proteína de unión a ácido nucleico monocatenario, o de la secuencia a la que el ácido nucleico monocatenario es complementario. Si dicha proteína se une a una secuencia que está localizada próxima a los extremos libres de la horquilla, el proceso de autorrehibridación se bloquea justo antes de que se vuelva a formar la horquilla completa; en este caso,  $Z_{\text{pausa}}$  es mínima. Por otro lado, si dicha proteína se une a una parte de la horquilla cuando está próxima al bucle desemparejado, el proceso de renaturalización se detendrá en una situación en la que la horquilla está completamente, o casi completamente, desnaturalizada; en este caso,  $Z_{\text{pausa}}$  es máxima. Igualmente, si el mencionado ácido nucleico monocatenario se hibrida con una secuencia que está situada próxima a los extremos libres de la horquilla, el proceso de autorrehibridación se bloquea justo antes de que se vuelva a formar la horquilla completa; en este caso,  $Z_{\text{pausa}}$  es mínima. Por otro lado, si el mencionado ácido nucleico monocatenario se hibrida con una parte de la horquilla que está próxima al bucle desemparejado, el proceso de renaturalización se detendrá en una situación en la que la horquilla está completamente, o casi completamente, desnaturalizada; en este caso,  $Z_{\text{pausa}}$  es máxima (figura 1).

40 Es posible correlacionar de forma precisa una distancia física en una molécula de ácido nucleico bicatenario con un número de bases. Por ejemplo, una distancia de 0,8 nm corresponde a la distancia abarcada por dos nucleótidos sucesivos (1 pb) en un ácido nucleico monocatenario bajo una fuerza de 10 pN. El ajuste exacto de la extensión frente a la fuerza se da por la elasticidad del ácido nucleico monocatenario. Por lo tanto, midiendo simplemente la distancia entre los dos extremos libres de la molécula de ácido nucleico bicatenario que se ha vuelto a cerrar parcialmente (o cualesquiera dos posiciones de referencia en la molécula), es posible determinar de forma precisa dónde se bloquea la renaturalización.

50 De este modo, en un ejemplo, el método consiste en la determinación de la unión de una proteína a una molécula de ácido nucleico, en el que dicha molécula de ácido nucleico bicatenario se desnaturaliza en primer lugar aplicando una fuerza física, después se vuelve a hibridar en presencia de dicha proteína, y opcionalmente de un ácido nucleico monocatenario, y se detecta la presencia de un bloqueo en la rehibridación. En un aspecto, la distancia entre los dos extremos de la molécula bicatenaria parcialmente renaturalizada se determina cuando el proceso de renaturalización se bloquea. Preferentemente, la distancia entre los dos extremos de dicha molécula se determina cuando la molécula está completamente desnaturalizada. Más preferentemente, las dos distancias se comparan, y se determina la posición del bloqueo. Más preferentemente, se mide la distancia entre el bucle completamente extendido y una posición de hibridación de referencia, y se usa para determinar la posición del bloqueo. Incluso más preferentemente, se mide la distancia entre dos posiciones de hibridación de referencia, y se usa para determinar la posición del bloqueo.

60 Aparte de su posición a lo largo de la molécula, el parámetro más útil asociado con el bloqueo en la renaturalización es el período de tiempo durante el cual la renaturalización está bloqueada (denominado en la presente memoria como la duración de la pausa en la renaturalización). De hecho, es posible medir el período de tiempo durante el cual está bloqueada la rehibridación. Por ejemplo, un experto en la materia puede determinar el período de tiempo durante el cual la distancia entre los dos extremos del ácido nucleico bicatenario es  $z$  como se define anteriormente, es decir, un valor intermedio comprendido entre  $Z_{\text{elevada}}$  y  $Z_{\text{baja}}$ .

Cuando el bloqueo está causado por la hibridación entre el ácido nucleico bicatenario desnaturalizado y el ácido nucleico monocatenario complementario, la duración del bloqueo depende del grado de complementariedad entre las dos secuencias. Cuanto mayor es la complementariedad, mayor es el número de enlaces establecidos entre las dos moléculas, y por lo tanto mayor es la duración. También está claro que el tiempo de bloqueo dependerá de la longitud de la región de complementariedad entre las dos secuencias. Cuanto mayor sea la región, mayor es el número de enlaces establecidos entre las dos moléculas, y por lo tanto mayor es la duración. Por lo tanto, es concebible fácilmente que en ciertas condiciones la duración de la pausa de renaturalización será casi permanente. En particular, cuando el ácido nucleico monocatenario comprende más de 20, preferentemente más de 25, incluso más preferentemente más de 30 nucleótidos aptos para hibridarse con el ácido nucleico bicatenario desnaturalizado, el ácido nucleico monocatenario permanece hibridado a la horquilla bicatenaria (durante muchos minutos) incluso cuando la fuerza aplicada al mencionado ácido nucleico bicatenario se disminuye hasta  $F_{\text{ensayo}}$ , previniendo de este modo la autorrehibridación de dicha horquilla bicatenaria. En tal caso, puede ser ventajoso usar una enzima para eyectar la molécula de ácido nucleico monocatenario o añadir una tercera fase en la que la fuerza se reduce hasta 0,5 o 1 pN durante unos pocos segundos que expele eficientemente oligonucleótidos hibridados. La eyección de dicha molécula de ácido nucleico monocatenario hace posible de este modo llevar a cabo ciclos de apareamiento y desemparejamiento, y mejorar de este modo la relación de señal/ruido.

La duración de la pausa también puede variar con las condiciones de la reacción. Dicha duración disminuirá a medida que aumenta la temperatura. Igualmente, las condiciones del amortiguador también pueden modular la duración de la pausa: por ejemplo, magnesio, betaína y cloruro de tetrametilamonio (TMAC usado a concentración molar) aumentan el tiempo de bloqueo. Estos compuestos refuerzan pares AT más que GC, reduciendo así la diferencia en la fortaleza entre estos pares. Sin embargo, cuando se fija la temperatura y el amortiguador, la duración de la pausa dependerá solamente de la fuerza que tira del ácido nucleico bicatenario desnaturalizado, y de su complementariedad con el ácido nucleico monocatenario. De hecho, los inventores han mostrado que el tiempo de bloqueo disminuye exponencialmente a medida que se reduce la fuerza.

Finalmente, la duración de la pausa también dependerá de las propiedades del complejo formado entre la proteína, el ácido nucleico bicatenario desnaturalizado y el ácido nucleico monocatenario complementario. La presencia de la proteína de unión a ácido nucleico bicatenario puede estabilizar el complejo. Cuanto mayor es su afinidad por el ácido nucleico bicatenario, más prolongada parece la pausa. También es posible que la proteína desestabilice el ácido nucleico bicatenario (como es el caso, por ejemplo, para el complejo abierto de una ARN-polimerasa), conduciendo a una pausa más corta.

Igualmente, la presencia de una proteína apta para unirse al ácido nucleico bicatenario desnaturalizado bloqueará transitoriamente la renaturalización de dicha molécula de ácido nucleico. La duración de este bloqueo también dependerá de la afinidad de la proteína por el ácido nucleico. Está claro que una proteína con una afinidad elevada por dicha molécula conducirá a una pausa más prolongada que una proteína con una afinidad más débil.

El experto en la materia apreciará inmediatamente de que la medida de la pausa permite la determinación del tiempo medio de bloqueo, y por tanto de los parámetros cinéticos de la reacción de unión, como se explica en la sección experimental.

De este modo, en un aspecto particular, el método comprende las etapas de:

- a) desnaturalizar dicha molécula de ácido nucleico bicatenario aplicando una fuerza física a dicha molécula;
- b) proporcionar una proteína y, opcionalmente, una molécula de ácido nucleico monocatenario,
- c) renaturalizar la molécula de ácido nucleico bicatenario en presencia de dicha proteína y, opcionalmente, de dicha molécula de ácido nucleico monocatenario; y
- d) detectar un bloqueo de la renaturalización de dicha molécula de ácido nucleico bicatenario, y
- e) determinar la duración de la pausa.

Preferentemente, el mencionado método comprende la etapa adicional de determinar la posición del bloqueo.

En este ejemplo, la duración de la pausa se puede comparar con un control. En particular, cuando dicha proteína es una proteína de unión a ácido nucleico bicatenario, puede ser ventajoso comparar dicha pausa con una pausa medida cuando el método se lleva a cabo en ausencia de la proteína. Como se explica anteriormente, la unión de la proteína al complejo formado entre el ácido nucleico bicatenario desnaturalizado y el ácido nucleico monocatenario complementario altera la duración del bloqueo de la renaturalización. Dicho bloqueo se traslada como un incremento, o disminución (dependiendo de la proteína específica), en la duración de la pausa.

De este modo, en un ejemplo preferido, el método comprende las etapas de:

- 5 a) desnaturalizar dicha molécula de ácido nucleico bicatenario aplicando una fuerza física a dicha molécula;
- b) proporcionar una proteína y, opcionalmente, una molécula de ácido nucleico monocatenario,
- c) renaturalizar la molécula de ácido nucleico bicatenario en presencia de dicha proteína y, opcionalmente, de dicha molécula de ácido nucleico monocatenario; y
- 10 d) detectar el bloqueo de la renaturalización de dicha molécula de ácido nucleico bicatenario, y
- e) determinar la duración de la pausa; y
- 15 f) comparar con la duración en ausencia de proteína.

Ventajosamente, el mencionado método comprende la etapa adicional de determinar la posición del bloqueo.

Aunque es posible detectar y medir la unión de la proteína a un ácido nucleico sin buscar información sobre la secuencia del sitio de unión, en algunas aplicaciones puede ser útil determinar dicha secuencia. Por ejemplo, puede ser interesante identificar mutaciones del mencionado sitio de unión que suprimen la unión de dicha proteína.

De este modo, en un ejemplo preferido, el método se refiere así a un método para la determinación de la unión de una proteína a una molécula de ácido nucleico bicatenario que comprende una secuencia de ácido nucleico, comprendiendo dicho método las etapas de:

- a) desnaturalizar dicha molécula de ácido nucleico bicatenario aplicando una fuerza física a dicha molécula;
- 30 b) proporcionar dicha proteína y opcionalmente una molécula de ácido nucleico monocatenario complementaria a por lo menos parte de dicha molécula de ácido nucleico bicatenario;
- c) renaturalizar dicha molécula de ácido nucleico bicatenario en presencia de dicha proteína y opcionalmente el mencionado ácido nucleico monocatenario;
- 35 d) detectar un bloqueo de la renaturalización del ácido nucleico bicatenario; y
- e) secuenciar la secuencia de ácido nucleico unida por dicha proteína.

Ventajosamente, la detección del bloqueo de la renaturalización es seguida de una etapa de determinar la posición del bloqueo.

Preferentemente, dicha proteína y dicha molécula de ácido nucleico monocatenario se eliminan por lavado de la molécula de ácido nucleico bicatenario antes de que se secuencie el sitio de unión.

Puesto que el método de la invención se basa en la detección de una sola molécula, sería conveniente usar un método que pueda secuenciar una sola molécula sin amplificación previa. Tales métodos de identificación y secuenciación de moléculas individuales se ha descrito previamente (documentos (WO 2011/147931; WO2011/147929; Ding et al., Nature Met, 9(4): 367-372, 2012). Estos métodos de secuenciación se basan en la detección de un bloqueo de la renaturalización de una molécula de ácido nucleico bicatenario desnaturalizado. De este modo, un método de secuenciación como se describe en la presente memoria comprende preferentemente las etapas de:

- 55 a) desnaturalizar una molécula de ácido nucleico bicatenario que corresponde a dicha secuencia de ácido nucleico aplicando una fuerza física a dicha molécula;
- b) proporcionar una molécula de ácido nucleico monocatenario;
- c) renaturalizar dicha molécula de ácido nucleico bicatenario en presencia de dicha molécula de ácido nucleico monocatenario; y
- 60 d) detectar un bloqueo de la renaturalización del ácido nucleico bicatenario.

Ventajosamente, el mencionado método comprende la etapa adicional de determinar la posición del bloqueo.

Estos métodos de secuenciación se pueden combinar fácilmente con el método de la invención, puesto que usan el mismo aparato que el presente método. Estirando de perlas magnéticas atadas por una horquilla a la

superficie, la molécula se puede abrir. En este estado abierto, se puede hibridar con ácidos nucleicos monocatenarios complementarios, que bloquean transitoriamente el cierre de nuevo de la horquilla cuando se reduce la fuerza de estirado. Midiendo la distancia desde la superficie a la perla de una horquilla bloqueada, se puede determinar la posición del híbrido a lo largo de la molécula con una precisión de casi una única base, estableciendo por tanto cuál es la secuencia local (el complemento de la secuencia de los ácidos nucleicos monocatenarios conocidos en disolución). De este modo es posible secuenciar directamente la molécula unida por dicha proteína, sin alterar el montaje del experimento, sustituyendo solamente el amortiguador que contiene la proteína, y opcionalmente un ácido nucleico monocatenario complementario, por un amortiguador adecuado para la secuenciación según los mencionados métodos.

La identificación eficiente de elementos reguladores en cis del ADN es un reto importante de la biología post-genómica. La identificación de todos los sitios de unión de una proteína de unión a ácido nucleico específica en el genoma es particularmente útil, puesto que identifica todos los genes cuya expresión está regulada potencialmente por dicha proteína. La identificación completa de los elementos cis-reguladores del ADN es crucial para una comprensión predictiva de la dinámica de la red transcripcional.

La confluencia de datos de secuencias del ADN del genoma completo, tecnologías de alto rendimiento, y nuevos algoritmos está fomentando nuestra capacidad para identificar y caracterizar elementos reguladores transcripcionales (Eisen et al., Proc. Natl. Acad. Sci., 95: 14863-14868, 1998; Tavazoie et al., Nat. Genet., 22: 281-285, 1999; Bussemaker et al., Nat. Genet., 27: 167-171, 2001; Lee et al., Science, 298: 799-804, 2002). Sin embargo, estos enfoques tienen limitaciones inherentes. Por ejemplo, el éxito de métodos híbridos que usan agrupamiento de la expresión génica y el descubrimiento de motivos reguladores en cis está limitado por el intervalo de perturbaciones fisiológicas usadas en el laboratorio. Lo mismo es cierto para enfoques *in vivo*, tales como la inmunoprecipitación con cromatina a base de chip (ChIP), en la que las interacciones ADN-proteína, en virtud de su papel regulador, solo se producen en condiciones medioambientales específicas (Lee et al., Science, 298: 799-804, 2002). Estas limitaciones son incluso más severas para eucariotas metazoicas, en las que los datos experimentales son más difíciles de adquirir.

El presente método ofrece una alternativa a los métodos de la técnica anterior, tales como ChIP (inmunoprecipitación cromosómica) y obtención de la huella de ADNasa I para cartografiar las localizaciones de la unión en el genoma de factores de transcripción (The ENCODE Project Consortium, Nature, 489: 57-74, 2012).

De este modo, según otro aspecto, se describe en la presente memoria un método para identificar moléculas de ácido nucleico que comprenden una secuencia apta para unirse a una proteína de unión a ácido nucleico específica, comprendiendo dicho método las etapas de:

- a) proporcionar una población de moléculas de ácido nucleico bicatenario;
- b) evaluar la unión de dicha proteína a dicha molécula de ácido nucleico mediante el método descrito anteriormente; y
- c) seleccionar las moléculas de ácido nucleico aptas para unirse a dicha proteína.

Preferentemente, el método implica la provisión de un ácido nucleico monocatenario complementario al sitio de unión de dicha molécula de ácido nucleico.

Según este ejemplo, el método comprende de este modo las etapas de:

- a) proporcionar una población de moléculas de ácido nucleico bicatenario;
- b) desnaturalizar dicha molécula de ácido nucleico bicatenario aplicando una fuerza física a dicha molécula;
- c) proporcionar dicha proteína y una molécula de ácido nucleico monocatenario complementaria al mencionado sitio de unión;
- d) renaturalizar dicha molécula de ácido nucleico bicatenario en presencia de dicha proteína y dicha molécula de ácido nucleico monocatenario; y
- e) detectar o no un bloqueo de la renaturalización del ácido nucleico bicatenario; y
- f) seleccionar las moléculas de ácido nucleico en las que la renaturalización está transitoria o permanentemente bloqueada.

Ventajosamente, el mencionado método comprende la etapa adicional de determinar la posición del bloqueo.

Las moléculas de ácido nucleico a aislar de este modo corresponden a una población de moléculas de ácido

nucleico, que comprenden dicha secuencia de unión específica. De este modo difieren de otras moléculas de ácido nucleico por cuanto contienen esta secuencia específica. Aunque estas moléculas comparten todas ellas esta secuencia, de otro modo pueden ser idénticas o no. En cierto ejemplo, puede ser preferible que el experto en la materia identifique la secuencia de cada una de las moléculas de ácido nucleico que difieren fuera de dicha secuencia de unión específica. De hecho, cuando se identifican moléculas de ácido nucleico que contienen uno o más sitios de unión para una proteína de unión a ácido nucleico específica, puede ser ventajoso secuenciar las moléculas identificadas, por ejemplo con el método de secuenciación descrito anteriormente. La información obtenida mediante esta etapa puede permitir la localización de dicha molécula en todo el genoma, e identificar así las unidades de expresión que pueden estar o no reguladas por este sitio de unión. Esto se puede lograr fácilmente usando cuidadosamente la información obtenida mediante la etapa de secuenciación para inspeccionar las bases de datos: el experto en la materia conoce cómo buscar clones que contienen las secuencias obtenidas mediante secuenciación, con la ayuda de bases de datos de secuencias disponibles públicamente (por ejemplo, Genbank), y esto no necesita ser detallado adicionalmente en la presente memoria.

En un ejemplo particular, la población de moléculas de ácido nucleico bicatenario representa el genoma completo.

La población de moléculas de ácido nucleico bicatenario se obtiene ventajosamente digiriendo en primer lugar los cromosomas mediante una enzima de restricción de corte raro. Como el experto en la materia conoce, una enzima de restricción de corte raro es una enzima de restricción con una secuencia de reconocimiento que aparece solo raramente en un genoma, por ejemplo una secuencia de reconocimiento que comprende 7 u 8 bases. Los ejemplos de tales enzimas de corte raro incluyen *SfiI*, *XmaI*, *AscI*, *AsiS I* (isoesquizómero *Sgf I*), *Not I* (isoesquizómero *CciN I*), *Sbf I* (isoesquizómeros *Sse8387 I*, *Sda I*), *Fse I*, *Pac I*, etc. Todas estas enzimas están comercialmente disponibles. En una segunda etapa, los fragmentos de restricción así obtenidos se digieren con una enzima de restricción común de 6 bases, tal como *EcoRI*, *BamHI*, *XhoI*, etc. Los fragmentos bicatenarios lineales resultantes se pueden transformar entonces en horquillas. Las técnicas que permiten unir juntos los extremos libres de bicatenarias son conocidas, y algunas se describen con mayor detalle en lo siguiente.

Otra aplicación particular del método de la invención es la detección de modificaciones epigenéticas. Tales ensayos son actualmente muy difíciles de realizar, y pierden muchas modificaciones del ADN. También, las modificaciones epigenéticas son extremadamente importantes en una variedad de patologías, incluyendo infección microbiana y oncología. Ventajosamente, el método mencionado anteriormente se puede usar para cribar modificaciones en ADN genómico en regiones completas o en regiones seleccionadas.

Las modificaciones epigenéticas para el ADN están presentes en los genomas de casi cualquier organismo vivo. Su tipo y localización varían a lo largo de los organismos, tejidos, y tipos celulares; a lo largo del tiempo; y a través de interacción con el entorno. Algunas de estas modificaciones suceden a través de procesos celulares cuidadosamente controlados. Otras son el resultado de daño al ADN.

Tales modificaciones expanden enormemente la cantidad de información que se puede almacenar en el ADN. Por ejemplo, el gen *dam* de *Escherichia coli* codifica una ADN metiltransferasa que metila adenina en secuencias -GATC- en ADN bicatenario, regulando de este modo la expresión génica (véase, por ejemplo, Calmann y Marinus, J. Bacteriol., 185(16): 5012-5014, 2003). Por otro lado, el marcador epigenético más común en eucariotas es 5-metilcitosina (5mC). Esta modificación específica es requerida para controlar y regular una amplia variedad de procesos celulares importantes y fisiológicos más amplios, y problemas con la metilación del ADN en seres humanos se han visto implicados en una variedad de enfermedades, muy notablemente ciertos tipos de cáncer. Además de 5mC, existe en eucariotas una amplia variedad de otras modificaciones del ADN (Korlach y Turner, Curr. Opin. Struct. Biol., 22: 251-261, 2012).

Actualmente, el patrón de oro para la determinación de 5mC es la "conversión del bisulfito", en la que todos los restos de citosina se convierten en uracilo, excepto aquellos que han sido metilados, que permanecen sin cambios. La amplificación subsiguiente del producto de ADN convierte uracilo en timina. Estos cambios de conversión se pueden detectar entonces a través de la secuenciación del ADN (Song et al., Nature Biotechnol, 30(11): 1107-1116, 2012). Sin embargo, esto es un proceso complicado, que consume tiempo, y caro, con tasas de error de 5-34% (Beck, Nature Biotechnol, 10: 1026- 1028, 2010).

La presente invención proporciona un método fácil para detectar modificaciones epigenéticas de ácidos nucleicos. Por "modificaciones epigenéticas", se hace referencia en la presente memoria a modificaciones de las bases que constituyen una molécula de ácido nucleico, que tiene lugar tras la síntesis de dicha molécula de ácido nucleico. Tales modificaciones epigenéticas incluyen, entre otras, 4-metilcitosina (m4C), 5-metilcitosina (5mC), 5-hidroximetilcitosina (5hmC), 5-formilcitosina (5fC) y 5-carboxilcitosina (5caC), así como 6-metiladenosina (m6A) en ADN, y 5-hidroximetiluracilo (5hmU) y N6-metiladenosina (m6A) en ARN.

De este modo, en un aspecto particular, la presente invención proporciona un método para detectar por lo menos una base modificada comprendida en una molécula de ácido nucleico bicatenario, comprendiendo dicho método las etapas de:

- a) proporcionar el mencionado ácido nucleico bicatenario;
- b) proporcionar una proteína apta para unirse a dicha base modificada; y
- c) evaluar la unión de dicha proteína a dicha molécula de ácido nucleico mediante el método descrito anteriormente.

Opcionalmente, el método puede comprender una etapa adicional de evaluar la hibridación de un simple oligonucleótido que reconoce el sitio de posible modificación, para validar mejor los resultados. Por ejemplo, tras detectar la metilación de 5mC con su anticuerpo, se puede detectar la secuencia ATGC con un oligo NNTACGNN.

Este método es particularmente ventajoso, debido a que usa en un proceso reversible moléculas de unión no modificadas. Por ejemplo, cuando se usan para detectar 5mC, no se requiere ninguna reacción química (bisulfato de sodio) en el ADN. Además, el método de la invención es mucho más sensible que cualquiera de los métodos de la técnica anterior, puesto que permite la detección de una base modificada sobre la base de una sola molécula.

En una forma de realización preferida, la base modificada se selecciona de entre el grupo constituido por 5-metilcitosina (5mC), 5-hidroximetilcitosina (5hmC), 5-formilcitosina (5fC), 5-carboxilcitosina (5caC), 5-hidroximetiluracilo (5hmU), y N6-metiladenosina (m6A). En una forma de realización más preferida, dicha base se escoge entre 5mC y 5hmC. En una forma de realización incluso más preferida, dicha base es 5mC. Se han descrito proteínas que reconocen y se unen específicamente a estas bases modificadas. Por ejemplo, los anticuerpos dirigidos contra 5mC se han descrito y usado tiñendo esta modificación para visualización a base de células (Ito et al., *Nature*, 466: 1129-1133, 2010; Ko et al., *Nature*, 468: 839-843, 2010; Szulwach et al., *Nature Neurosci*, 14: 1607-1611, 2011; Haffner et al., *Oncotarget*, 2: 627-637, 2011; Inoue et al., *Science*, 334: 194, 2011; Inoue et al., *Cell Res*, 21: 1670-1676, 2011). Tales anticuerpos están comercialmente disponibles (por ejemplo clon 33D3; ref: 39649 de Active Motif). Aparte de los anticuerpos, se han identificado enzimas que reconocen y reaccionan específicamente con el nucleótido de interés (Song et al., *Nature Biotechnol*, 30(11): 1107-1116, 2012). Por ejemplo, la enzima  $\beta$ -glucosiltransferasa (BGT) del bacteriófago T4 transfiere un resto de glucosa sobre 5hmC. Las proteínas Tet1-3 son responsables de la conversión de 5mC en 5hmC. La proteína 2 de unión a metil-CpG, (MeCP2), se identificó por primera vez mediante su afinidad por ADN que contiene 5mC. Preferentemente, dicha proteína es un anticuerpo dirigido contra dicha base modificada, o una enzima que reconoce específicamente dicha base. Más preferentemente, dicha proteína es un anticuerpo.

Resulta evidente que se podría aplicar el mismo método a la detección de otras modificaciones de ácidos nucleicos. Por ejemplo, es posible detectar un error de apareamiento presente en una molécula de ácido nucleico bicatenario. Se sabe desde hace mucho tiempo que las proteínas tales como la MutS bacteriana reconocen la base emparejada erróneamente en la hebra hija, y se unen al ADN mutado. Tal propiedad se puede poner en uso para detectar e identificar cualquier apareamiento erróneo en una molécula de ácido nucleico bicatenario.

Por lo tanto, también es un aspecto de la presente invención proporcionar un método para detectar por lo menos un apareamiento erróneo en un ácido nucleico bicatenario, comprendiendo dicho método las etapas de:

- a) proporcionar el mencionado ácido nucleico bicatenario;
- b) proporcionar una proteína apta para unirse a una base emparejada erróneamente; y
- c) estudiar la unión de dicha proteína a dicha molécula de ácido nucleico mediante el método descrito anteriormente.

Puesto que se sabe que MutS se une como dímero a un apareamiento erróneo, es ventajoso usar un dímero de MutS en el método de la invención. En eucariotas, los homólogos de MutS forman dos heterodímeros principales: Msh2/Msh6 (MutS $\alpha$ ) y Msh2/Msh3 (MutS $\beta$ ). Preferentemente, dicha proteína se selecciona entre un dímero de MutS, Msh2/Msh6 (MutS $\alpha$ ), y Msh2/Msh3 (MutS $\beta$ ).

Un polimorfismo de un solo nucleótido (SNP, pronunciado snip; plural, snips) es una variación de secuencia de ADN que ocurre cuando un solo nucleótido – A, T, C o G – en el genoma (u otra secuencia compartida) difiere entre miembros de una especie biológica o cromosomas emparejados en un ser humano. De media, los SNPs ocurren en la población humana más de 1 por ciento de las veces. Debido a que solamente alrededor de 3 a 5 por ciento de una secuencia de ADN de una persona codifica la producción de proteínas, la mayoría de los SNPs se encuentran fuera de las secuencias codificantes. Los SNPs encontrados dentro de una secuencia codificante son de particular interés debido a que muy probablemente alteran la función biológica de una proteína.

Una molécula que comprende un SNP formará un apareamiento erróneo cuando se hibride con una molécula

que comprende la secuencia encontrada en la mayoría de la población. La presente invención permite así la detección fácil de los SNPs.

De este modo, esta forma de realización se refiere a un método para detectar un SNP en una secuencia  
5 contenida en un ácido nucleico, comprendiendo dicho método las etapas de:

- a) hibridar el mencionado ácido nucleico con un ácido nucleico monocatenario que comprende la secuencia encontrada en la mayoría de la población; y
- 10 b) detectar el apareamiento erróneo resultante mediante el método anterior.

Si el ácido nucleico que se debe someter a ensayo es un ácido nucleico bicatenario, puede ser ventajoso desnaturalizar el mencionado ácido nucleico, antes de la etapa a).

15 Resulta evidente que estos métodos se pueden llevar a cabo en una escala de genoma completo, mediante una adaptación simple del método descrito anteriormente. Esto conducirá a la identificación de todos los sitios en el genoma que contienen, por ejemplo, una base modificada particular. Los genes cuya expresión es susceptible para que se vea afectada por las mencionadas bases modificadas se pueden identificar, secuenciando las moléculas de ácido nucleico que contienen tales bases modificadas. Además, entonces se puede evaluar la  
20 transmisión de las mencionadas bases modificadas a la progenie. Esta información puede ser de interés en los campos como la selección de animales o de plantas, en los que es importante asegurar que algunos genes permanecen en silencio mientras que otros siguen siendo expresados a lo largo de las generaciones.

Más específicamente, la presente invención se refiere a un método que comprende las etapas definidas en las reivindicaciones 1 a 16.  
25

En todavía otro aspecto, se proporciona un método para identificar compuestos que interfieren con la unión de una proteína a su secuencia de unión específica. Estos compuestos disminuyen o anulan la unión de dicha proteína a su sitio de unión. Tales compuestos pueden ser útiles como sustancias terapéuticas. Por ejemplo, los  
30 compuestos que prevengan la interacción de las formas oncogénicas de cMyc con su sitio de unión serían útiles para tratar el cáncer.

Según este ejemplo, se describe un método para identificar por lo menos un compuesto apto para prevenir la interacción entre una proteína y su sitio de unión, comprendiendo dicho método las etapas de:  
35

- a) proporcionar dicha proteína y una molécula de ácido nucleico que comprende una secuencia que corresponde al mencionado sitio de unión;
- b) proporcionar un compuesto; y
- 40 c) estudiar la unión de dicha proteína a dicha molécula de ácido nucleico mediante el método descrito anteriormente.

En un ejemplo preferido, un compuesto se selecciona cuando la unión de dicha proteína a dicha molécula de ácido nucleico se disminuye o anula.  
45

Resulta evidente que la mayoría de proteínas de unión a ácido nucleico que están implicadas en el cáncer son factores de transcripción que se unen a ácidos nucleicos bicatenarios. Por lo tanto, en otro ejemplo, dicha molécula de ácido nucleico es una molécula de ácido nucleico bicatenario. En un ejemplo preferido adicional, el método comprende además proporcionar un ácido nucleico monocatenario complementario a la secuencia de dicha molécula de ácido nucleico bicatenario. Por supuesto, estas moléculas se proporcionan antes de que tenga lugar el ensayo de la unión.  
50

La implementación del método ha sido posible, en particular, por la existencia de dispositivos diseñados para sondear la interacción de ácidos nucleicos en tiempo real a nivel de una sola molécula. Tal dispositivo se describe, por ejemplo, en las patentes U.S. n<sup>os</sup> 7.052.650 y 7.244.391. El aparato descrito allí usa trampas magnéticas para aplicar una fuerza a escala de piconewtons sobre una perla superparamagnética de tamaño micrométrico. De forma breve, el mencionado aparato comprende un microscopio óptico, imanes y un PC. Las moléculas de ácido nucleico bicatenario están ancladas en múltiples puntos en un extremo a un elemento sin movimiento, por ejemplo una superficie, y en el otro extremo a una superficie móvil, en este caso una perla magnética. Se proporcionan imanes para actuar sobre la perla. En particular, los imanes se pueden usar para tirar de la perla lejos de la superficie. Sin embargo, la implementación del método no está restringida al aparato anterior. Para implementar el método, se puede usar cualquier dispositivo que permita extender completamente, y después volver a plegar, una molécula de ácido nucleico bicatenario, a la vez que monitoriza al mismo tiempo la extensión de dicha molécula. Por ejemplo, se pueden usar pinzas ópticas; sin embargo, requieren una calibración previa de la fuerza, y no se pueden paralelizar fácilmente para medidas de alto rendimiento. Otros inconvenientes son la  
55  
60  
65

complejidad de ajustar el control torsional del ácido nucleico y el posible calentamiento local de la disolución por el láser enfocado, que puede alterar las condiciones de hibridación.

5 El ácido nucleico bicatenario se incuba durante unos pocos minutos en una disolución de perlas adecuadas (por ejemplo, las revestidas con estreptavidina), a las que se une mediante uno de sus extremos marcados (por ejemplo biotina). Las perlas pueden ser transparentes si las pinzas ópticas se usan más tarde para la manipulación, o magnéticas si se usan trampas magnéticas o pinzas para manipulación.

10 El montaje de perla-ácido nucleico se inyecta en una cámara fluidica, cuya superficie se ha tratado para unirse al otro extremo marcado de la molécula (por ejemplo una superficie revestida con anti-Dig, para unirse al extremo marcado con Dig del ácido nucleico). De este modo, las perlas se anclan a la superficie vía una horquilla de ácido nucleico; véase la figura 1a. Entonces se monitoriza la distancia de la perla a la superficie por diversos medios conocidos por el experto en la materia: por ejemplo, se pueden usar los anillos de difracción de su imagen en una cámara para deducir su distancia, o se puede usar la intensidad de luz que dispersa (o emite mediante fluorescencia) cuando se ilumina en un modo evanescente para medir su distancia. Alternativamente, el campo magnético que generan se puede medir (usando un sensor magnético tal como los sensores GMR o Hall) para deducir su distancia a un sensor en la superficie que ancla.

20 Para tirar de la molécula de ácido nucleico que ancla las perlas a la superficie, se han descrito diversas técnicas. Una puede usar la luz de un haz de láser enfocado para atrapar una perla transparente cerca del punto focal. Mediante la traslación relativa del haz con respecto a la superficie que ancla, se puede aplicar una fuerza sobre la molécula que ancla (un ensayo de pinzas ópticas típico). La fuerza ejercida que es proporcional al desplazamiento de la perla desde su posición de equilibrio, para ejercer una fuerza constante sobre la molécula que ata, requiere un bucle de retroalimentación sobre el haz que atrapa.

25 Para ejercer una fuerza constante sobre una perla, se ha descrito el uso del arrastre hidrodinámico generado por un flujo alrededor de la perla, pero habitualmente produce una exactitud espacial baja ( $> 100$  nm). El ejemplo particular usa una trampa magnética para tirar sobre perlas superparamagnéticas ancladas a una superficie mediante una horquilla de ácido nucleico como se describe anteriormente. En esta configuración, se usan pequeños imanes colocados encima de la muestra para aplicar una fuerza constante sobre la perla anclada, cuya posición se puede determinar con una exactitud de  $< 1$  nm (dependiendo de la fuerza de tracción y la disipación debido al arrastre hidrodinámico).

35 En cada caso, uno se da cuenta de que la horquilla que ata se puede abrir totalmente de forma mecánica haciendo tracción sobre las perlas con una fuerza mayor que alrededor de 16 pN. La reducción de la tensión sobre la molécula por debajo de alrededor de 11 pN permite que la horquilla se vuelva a cerrar espontáneamente (la transición de la apertura de la cremallera es reversible aunque histerética). Si, durante la fase de cremallera abierta, algunas moléculas en disolución (tales como proteínas u oligonucleótidos complementarios de ADN, ARN, LNA o PNA) se han unido al ácido nucleico monocatenario estirado, estas moléculas bloquearán el cierre nuevamente de la cremallera de la horquilla cuando la fuerza se reduzca hasta por debajo de 11 pN. El principio del ensayo es de este modo cambiar entre dos fuerzas: una grande,  $F_{abierta}$ , para abrir la horquilla, y una más pequeña  $F_{ensayo}$ , usada para permitir el cierre nuevamente de la cremallera y medir la extensión de la molécula en bloques transitorios. La posición de bloqueo está relacionada con la secuencia mediante una relación lineal entre la extensión total y la bloqueada. Para una mejor exactitud, la extensión total se mide preferentemente a la fuerza de ensayo  $F_{ensayo}$ . Esto se logra diseñando el bucle de la horquilla de manera que requiera una fracción de un segundo para volverse a plegar una vez que la fuerza se reduce desde  $F_{abierta}$  hasta  $F_{ensayo}$ .

50 Para unir los ácidos nucleicos a las superficies o a los soportes, se puede hacer uso de una cualquiera de las técnicas conocidas en el campo. Esencialmente, el ácido nucleico se ancla directamente al soporte, por ejemplo la microperla, que implica una funcionalización de esta superficie, por ejemplo revistiéndola con estreptavidina, un grupo COOH, y similar, que puede reaccionar con el extremo funcionalizado del ácido nucleico.

55 Tales métodos necesitan, en general, funcionalizar el ácido nucleico, especialmente los extremos 3' y 5', es decir, injertar en ellos grupos químicos apropiados. Además, es preferible unir los otros dos extremos libres de la molécula mediante un bucle a fin de evitar que las hebras se disocien al final de la operación, de manera que esta última se pueda repetir si es apropiada. Para este fin, se pueden adoptar diferentes procedimientos.

60 Lo más simple es funcionalizar, usando oligonucleótidos sintéticos, uno de los extremos de un ácido nucleico bicatenario con dos funciones diferentes (biotina y amina, por ejemplo), lo que permite el anclaje a dos superficies previamente tratadas diferentes. Las dos hebras en el otro extremo se pueden unir usando un nucleótido sintético parcialmente emparejado en forma de un bucle. De esta manera, un ácido nucleico monocatenario, emparejado, es decir, una horquilla, es producido a partir de un ácido nucleico bicatenario. La ventaja de este método se basa en su capacidad para funcionalizar una población heterogénea de fragmentos grandes de ácido nucleico (como se obtienen mediante fraccionamiento de un gen o cromosoma), que entonces se puede analizar simultáneamente. En este caso, la muestra de ácido nucleico se fracciona usando dos (o más) enzimas de restricción, lo que permite que se obtenga una subpoblación con dos sitios de restricción diferentes

en sus extremos, que son similares a lo largo de todos los fragmentos. Esto permite que los dos extremos sean tratados diferentemente (por ejemplo uniendo un extremo a un oligonucleótido en forma de un bucle que posee el sitio de restricción apropiado en su extremo). El inconveniente de este método reside en la interferencia estérica entre los dos grupos funcionales adyacentes, lo que puede hacer difícil el acoplamiento a las superficies. Para resolver este problema, puede ser ventajoso añadir en cada extremo libre de la molécula de horquilla una secuencia "espaciadora" de bases, al final de la cual se añade entonces un grupo funcional; las dos secuencias espaciadoras son no complementarias, produciendo cada grupo funcional suficiente espacio para unirse a su superficie destinada. Más ventajosamente, la secuencia de cada secuencia espaciadora se diseña a fin de usar cebadores de secuenciación monocatenarios de secuencia conocida en el método de secuenciación de la invención. La adición de un bucle y/o espaciadores a las moléculas de ácido nucleico bicatenario se puede llevar a cabo con cualquiera de los métodos usados habitualmente en biología molecular. Estos métodos son bien conocidos por el experto en la materia, y de este modo no es necesario detallarlos en la presente memoria.

Con respecto a las técnicas actuales de anclaje, existen muchas de éstas, y derivan de las técnicas para anclar macromoléculas (proteínas, ADN, y similares) a superficies comercialmente disponibles previamente tratadas. La mayoría de estas técnicas se han desarrollado para ensayos inmunológicos, y enlazan proteínas (inmunoglobulinas) a superficies que portan grupos ( $-\text{COOH}$ ,  $-\text{NH}_2$ ,  $-\text{OH}$ , y similares) aptas para reaccionar con los extremos carboxílico ( $-\text{COOH}$ ) o aminínico ( $-\text{NH}_2$ ) de las proteínas.

El anclaje covalente de ácido nucleico se puede lograr directamente, vía el fosfato libre del extremo 5' de la molécula, que reacciona con una amina secundaria (Covalink  $-\text{NH}$  surface comercializado por Polyabo en Estrasburgo) para formar un enlace covalente. También es posible funcionalizar ADN con un grupo amínico, y entonces proceder de la misma manera que con una proteína.

También existen superficies revestidas con estreptavidina (perlas Dynal, y similares), que permiten un anclaje casi covalente entre la estreptavidina y una molécula de ADN biotinilada. Finalmente, injertando un anticuerpo dirigido contra digoxigenina sobre una superficie (mediante los métodos mencionados anteriormente), se le puede anclar un ácido nucleico funcionalizado con digoxigenina. Esto representa meramente una muestra de las muchas técnicas de anclaje posibles.

Entre las técnicas de unión y anclaje, también se deberían de mencionar, por ejemplo, las técnicas descritas en la patente EP 152886, que usan un acoplamiento enzimático para la unión de ADN a un soporte sólido tal como celulosa.

La patente EP 146815 también describe diversos métodos de unión de ADN a un soporte.

De forma similar, la solicitud de patente WO 92/16659 propone un método que usa un polímero para unir ADN.

Naturalmente, el ácido nucleico se puede unir directamente al soporte, pero, cuando sea necesario, especialmente con vistas a limitar la influencia de las superficies, el ácido nucleico se puede unir al extremo de un brazo inerte de péptido u otra naturaleza, como se describe, por ejemplo, en la patente EP 329198.

Los ejemplos a continuación permitirán poner más claramente de manifiesto otras características y ventajas de la presente invención.

#### Leyendas de las figuras:

Figura 1. Principio de detección de la hibridación de un oligonucleótido a su secuencia complementaria en un ADN de horquilla. El ADN de horquilla que ancla la perla a la superficie (a) se abre momentáneamente incrementando la fuerza que tira de la perla hasta un valor por encima de 16 pN. En esa fase, el fragmento complementario en disolución se hibrida a su diana en la horquilla de ADN abierta, evitando así transitoriamente el cierre de nuevo de la horquilla (b) cuando la fuerza se reduce nuevamente a su valor inicial. A partir del cambio en la extensión de la molécula ( $Z_{\text{elevada}}-Z$ ) entre el punto de bloqueo y la longitud inicial de la horquilla, se deduce dónde, a lo largo de la horquilla, se ha emparejado la secuencia complementaria. A partir del período de tiempo promedio del bloqueo, se puede saber la posible existencia de emparejamientos erróneos y su posición a lo largo del híbrido. (c) Traza del tiempo de la extensión de una horquilla a medida que la fuerza se incrementa desde 11,5 pN hasta 17,8 pN, y después disminuye nuevamente hasta su valor inicial. Se observa la presencia de una pausa durante la rehibridación de alrededor de 10 s. Esa pausa se observa solamente en presencia en disolución de oligómeros complementarios (o casi complementarios) de longitud  $> 7$  nucleótidos (en la presente memoria, la señal es debida a un 10mero).

Figura 2: a) Distribución exponencial del tiempo de bloqueo de un oligonucleótido de 10 nts obtenida a  $F_{\text{ensayo}} = 9$  pN. b) Dependencia exponencial de  $T_{\text{off}}$  frente a  $F_{\text{ensayo}}$  obtenida para un oligonucleótido de 9 nts.

Figura 3: Evolución de la probabilidad de bloqueo  $P_{\text{bloqueo}} = \text{Nb.ciclos bloqueados}/\text{Nb. Ciclos con la duración}$

de la fase abierta  $T_{abierta}$  para a por un oligonucleótido de 12-nt para encontrar su sitio complementario. Un ajuste demuestra que  $T_{on}$  en el tiempo requerido para que la molécula encuentre su diana es típicamente 15 s cuando la concentración de oligonucleótido es 20 nM. Este tiempo no depende de la fuerza usada en la fase de ensayo. El parámetro  $a(F)$  sería igual a 1 si se detectasen todos los sucesos, pero puesto que se pierden sucesos cortos,  $a(F)$  es menor que 1, especialmente cuando  $F_{ensayo}$  es pequeña.

Figura 4: La probabilidad de bloqueo aumenta y se satura con la concentración de oligonucleótido. en la presente memoria, un oligonucleótido de 12 nt, a una concentración de 27,5 nM, conduce a un bloqueo que sucede una vez cada dos ciclos para una fase abierta que dura 10 s y  $F_{ensayo} = 8$  pN. Como se puede observar en la figura 3, la saturación de  $P_{bloqueo}$  no alcanza suficientemente 1; esto es debido a que estamos perdiendo bloqueos muy cortos.

Figura 5: Parámetros cinéticos que definen la propiedad de unión de un oligonucleótido de 12 nts a su sustrato complementario como función de la fuerza iónica del amortiguador.  $k_{off}$  varía poco con la fuerza iónica, mientras que  $k_{on}$  presenta una fuerte dependencia.  $k_{on}$  aumenta en un factor de 3 añadiendo  $Mg^{2+}$ . La constante de equilibrio  $k_d$  se puede calcular para ambos parámetros cinéticos.

Figura 6: Bloqueo de una horquilla por una primasa que estabiliza un oligonucleótido de ARN de 5 nt complementario a la secuencia de ADN cebadora. b) Posición del suceso de bloqueo a lo largo de la secuencia. c) Distribución del tiempo de bloqueo producido por la T4 primasa que estabiliza un oligonucleótido de ARN pentamérico en el proceso de cebado observado con  $F_{ensayo} = 9$  pN. El oligonucleótido de ARN de 5 nt no bloquea de manera visible el replegamiento de la horquilla. Con la T4 primasa WT, el bloqueo se produce en la posición esperada a lo largo de la secuencia, y el tiempo de bloqueo es 5 s. Con el mutante E248Q, observamos el mismo fenómeno, pero el tiempo de bloqueo se reduce significativamente.

Figura 7: Serie de ciclos que evalúan la unión de helicasa RecQ a ADNmc con tres fases: abierta en  $F_{abierta} = 20$  pN, ensayo a  $F_{ensayo} = 10$  pN, y una fase de limpieza a  $F_{limpieza} = 0,5$  pN. Se muestran 10 trazas, presentando unas pocas un suceso de bloqueo para un ciclo. La fase de limpieza a fuerza baja asegura que se elimine cualquier unión enzimática del molde. En ausencia de ATP, RecQ solamente se une y bloquea el replegamiento, la presión de la horquilla de replegamiento produce un deslizamiento de la helicasa, disminuyendo la posición de bloqueo por las etapas sucesivas.

Figura 8: Evolución de la probabilidad de bloqueo de RecQ frente a su concentración.  $P_{bloqueo}$  aumenta y se satura a medida que aumenta la concentración; esto define una concentración característica en la presente memoria de 226 pM.

Figura 9: Distribución de la posición de bloqueo de la helicasa RecQ sin ATP a lo largo del molde.

Figura 10: Señal bruta del bloqueo producida por el anticuerpo frente a la metilación de citosina a lo largo de una horquilla de ADN de 1,2 kb. Tres trazas presentan la extensión de las horquillas a lo largo de 5 ciclos. Cada ciclo comienza abriendo la horquilla durante 5,5 s con una fuerza de 20 pN, seguido de una fase de ensayo que dura 37 s a  $F = 6,5$  pN.

La mayoría del tiempo, el ciclo no presenta bloqueo (1), una horquilla puede presentar bloqueo sucesivo durante el mismo ciclo (2), y el bloqueo se puede prolongar a lo largo de varios ciclos. [Ac] la concentración de anticuerpo es 35 nM, el amortiguador es Tris 100 mM con 0,2% de BSA para evitar la unión no específica.

Las trazas se han desplazado en y por claridad.

Figura 11: Histograma de la posición de metilación a lo largo de la secuencia de una horquilla de 1,2 kb después de que se ha metilado por una ADN metiltransferasa humana. Se presentan cuatro histogramas de diferentes perlas. Existe un consenso sobre las posiciones de unión; se observan las cuatro posiciones de unión esperadas con relación a la metilación eucariótica, así como aquella en 882, que corresponde a la metilación realizada por *E. coli* en la que se produjo originalmente el ADN.

Figura 12: Registro de 30 ciclos de cierre y apertura de la horquilla con una pendiente suave en la fuerza con 2 s hacia arriba y 2 s hacia abajo. Los puntos representativos en el ciclo giran en sentido contrario a las agujas del reloj (véanse las flechas) partiendo en  $F = 1,5$  pN y  $Z = 0$ ; a medida que aumenta la fuerza, la extensión permanece muy pequeña hasta que la fuerza alcanza 15 pN, en ese momento la molécula se abre y  $Z$  alcanza  $1,3 \mu\text{m}$ . Cuando la fuerza se disminuye con una pendiente,  $Z$  disminuye lentamente hasta  $F = 11$  pN, en este punto la horquilla se vuelve a plegar hasta que golpea en el oligonucleótido de 12 nt. A medida que la fuerza continúa disminuyendo, también lo hace la  $Z$  del bloqueo, pero a medida que la fuerza disminuye alcanza pronto el punto en el que el oligonucleótido es expelido, como se puede observar por la rápida disminución en la marca de  $Z$  mediante un símbolo de diamante. La distribución de la fuerza que corresponde al desprendimiento del oligonucleótido se presenta a la derecha; su máximo, alrededor de 7 pN,

corresponde a la fuerza a la que  $T_{off}$  es igual a una fracción de un segundo.

Figura 13: Detección de los sitios de metilación en un ADN humano obtenido a partir de células humanas. Se preparó un ADN de horquilla a partir de una molécula de ADN genómico humano de 2,5 kb. A) Variación de la fuerza aplicada a lo largo de los ciclos de medida: la horquilla se abre durante 5 segundos mediante una fuerza de 19 pN; la fuerza se reduce entonces hasta 8,5 pN durante 10 segundos. B) Superposición de las señales obtenidas en aprox. 20 ciclos en presencia de anticuerpos dirigidos contra 5mC, que muestra la apertura de cremallera de la molécula, seguido por su cierre nuevamente de la cremallera, interrumpido por bloqueos transitorios. Estos bloqueos están causados por la unión del anticuerpo a 5mC. C) Un histograma de las posiciones de bloqueo muestra posiciones bien definidas que corresponden a la presencia de 5mC. Existen aproximadamente 20 posiciones, lo que sugiere una metilación cada aprox. 100 bases.

### Ejemplos experimentales

#### Antecedentes de la invención

La unión de proteína a ADN es un fenómeno importante en biología; es un proceso muy general que controla muchas reacciones. Aunque las propiedades de equilibrio termodinámico de este mecanismo son bien conocidas, la medida de su cinética es un problema más desafiante. El uso de moléculas individuales ofrece la capacidad de medir el tiempo requerido para que una proteína encuentre su diana de ADN, pero también la localización exacta del suceso de unión. Describimos en la presente memoria un nuevo ensayo de una sola molécula que logra estas metas.

Aunque el ensayo es amplio, ilustramos en primer lugar su aplicabilidad a la unión de un oligonucleótido específico, y a la unión no específica de una helicasa a un ADNmc. Finalmente, explicamos la unión específica de un anticuerpo que reconoce sitios metilados en el ADN.

#### Sumario

Este método se refiere a un nuevo procedimiento para la detección de una amplia variedad de modificaciones del ADN y sucesos de unión de proteína a ADN, que se basa en la detección mecánica de la obstrucción de la rehibridación de una horquilla de ADN. El ensayo se basa en una serie de ciclos que proporcionan información estadística de la unión de una sola molécula. Durante un ciclo, se comienza mediante una fase de apertura de la cremallera, en la que una única horquilla de ADN se despliega durante un tiempo  $T_{abierta}$  tirando de sus extremos con una fuerza  $F_{abierta}$  mayor que alrededor de 16 pN. En una segunda fase de ensayo, que dura  $T_{ensayo}$ , la tensión  $F_{ensayo}$  se reduce por debajo de hasta alrededor de 11 pN, que permite que la horquilla vuelva a cerrar la cremallera. Si una molécula presente en disolución se puede unir a una secuencia definida, o no específicamente, en la horquilla abierta (por ejemplo, una proteína apta para reconocer una secuencia específica mono- o bicatenaria, modificada o no), se unirá al ADN con una probabilidad  $P_{bloqueo}$  y, en ese caso, bloqueará transitoriamente el cierre nuevamente de la cremallera cuando la fuerza se reduce por debajo de alrededor de 11 pN. Esta obstrucción es fácilmente detectable como una pausa que ocurre en una posición definida durante la rehibridación de la horquilla, que conduce a tres parámetros:

- la posición  $Z_{bloqueo}$  de esta pausa a lo largo del ADN estirado es característica de la secuencia que se reconoce;
- la duración del bloqueo  $T_{off}$  caracteriza el tiempo durante el cual la molécula ha permanecido unida a ADN; y
- la probabilidad del bloqueo  $P_{bloqueo}$ , que está relacionado con el tiempo  $T_{on}$  requerido para que la molécula encuentre su sitio de unión.

$T_{on}$  y  $T_{off}$  son ambos característicos de la fuerza de la interacción entre el ADN y la molécula que bloquea. De este modo, sondando una secuencia de ADN (unida como una horquilla a una perla en un extremo y a una superficie en el otro) con una proteína o anticuerpo que reconoce la metilación, se puede identificar mediante ciclos repetidos de apertura y cierre de la horquilla la presencia del sitio de metilación sondado (vía la presencia de un bloqueo de algunas de las horquillas durante la rehibridación). De forma similar, se puede medir la unión de una proteína a un sitio de ADNbc putativo midiendo el incremento en la estabilidad del híbrido entre un oligonucleótido complementario en presencia frente a la ausencia de la proteína.

Este método permite la detección de modificaciones del ADN en ADN genómico sin pasar a través de la reacción del bisulfito y etapas de amplificación mediante PCR. Requiere algún procesamiento previo del ADN, necesario para procesarlo en fragmentos de horquilla que se pueden usar para unir perlas a una superficie (fragmentación y ligación con fragmentos adecuados). La presente invención no requiere marcaje fluorescente de las proteínas o del ADN. En su presente forma de realización, la técnica necesita un (microscopio) óptico para detectar el bloqueo de la horquilla durante la rehibridación.

Descripción técnica detallada

- 5 Un fragmento de ADN bicatenario (bc) de un tamaño comprendido entre unas pocas decenas y unos pocos millares de pares de bases (obtenido, por ejemplo, a partir de cizallamiento mecánico o cortes de restricción de ADN genómico) se liga a uno de los extremos de un bucle de ADN. Sus otros extremos se ligan a un fragmento de ADNbc que permite la unión de sus dos hebras a superficies revestidas diferentemente. Por ejemplo, el extremo 3' libre de una hebra se puede marcar con biotina, que permite la unión a perlas revestidas con estreptavidina, mientras que el extremo 5' en la hebra opuesta se puede marcar con digoxigenina, que permite su unión a superficies revestidas con un anticuerpo anti-Dig. Este marcaje de los extremos se puede realizar de 10 diversas maneras conocidas por el experto en la materia, tal como el uso de transferasa terminal para añadir nucleótidos modificados con biotina (o dig), o la hibridación con oligonucleótidos marcados adecuadamente.
- 15 Este constructo de ADN se incuba durante unos pocos minutos en una disolución de perlas adecuadas (por ejemplo las revestidas con estreptavidina), a las cuales se une mediante uno de sus extremos marcados (por ejemplo biotina). Las perlas pueden ser transparentes si se usan más tarde pinzas ópticas para la manipulación, o magnéticas si se usan trampas o pinzas magnéticas para la manipulación.
- 20 El ensamblaje de perla-ADN se inyecta en una cámara fluidica, cuya superficie se ha tratado para unirse al otro extremo marcado de la molécula (por ejemplo una superficie revestida con anti-Dig, para unir el extremo con Dig del ADN). Las perlas se anclan así a la superficie vía una horquilla de ADN (ver la figura 1a a continuación). La distancia de la perla a la superficie se monitoriza entonces por diversos medios. Por ejemplo, para deducir su distancia, se pueden usar los anillos de difracción de la imagen de las perlas en una cámara.
- 25 Para medir su distancia, también se podría usar la intensidad de luz dispersada por las perlas (o emitida como fluorescencia) cuando se iluminan en un modo evanescente. Como alternativa, cuando se usan perlas magnéticas, el campo magnético generado se puede medir (usando sensores GMR o Hall) para deducir la distancia de las perlas a la superficie, enviado a un sensor en la superficie de anclaje.
- 30 Para tirar de la molécula de ADN que ancla las perlas a la superficie, se han descrito diversas técnicas. Se puede usar la luz de un haz de láser enfocado, para atrapar una perla transparente cerca del punto focal. Mediante la translación relativa del haz con respecto a la superficie de anclaje, se puede aplicar una fuerza en la molécula anclante (un ensayo de pinzas ópticas típico). La fuerza ejercida, que es proporcional al desplazamiento de la perla desde su posición de equilibrio, para ejercer una fuerza constante sobre la molécula anclante requiere un 35 bucle de retroalimentación en el haz que atrapa.
- Para ejercer una fuerza constante en una perla, se ha descrito el uso del arrastre hidrodinámico generado por un flujo alrededor de la perla, pero habitualmente produce una baja exactitud espacial (>100 nm). El ejemplo particular usa una trampa magnética para tirar de las perlas superparamagnéticas ancladas a una superficie 40 mediante una horquilla de ADN como se describe anteriormente. En esta configuración, se usan pequeños imanes colocados encima de la muestra para aplicar una fuerza constante sobre la perla anclada, cuya posición se puede determinar con una exactitud de  $\sim 1$  nm (dependiendo de la fuerza de tracción y la disipación debido al arrastre hidrodinámico).
- 45 En cada caso, se observa que la horquilla anclante se puede abrir mecánicamente de forma total tirando de las perlas con una fuerza mayor que alrededor de 16 pN. La reducción de la tensión en la molécula por debajo de  $\sim 11$  pN permite que la horquilla vuelva a cerrarse espontáneamente (la transición de la apertura de la cremallera es reversible, aunque histerética).
- 50 Si, durante la fase de cremallera abierta, algunas moléculas en disolución (tales como proteínas y/u oligonucleótidos complementarios de ADN, ARN, LNA o PNA) se han unido a un ADN(bc) monocatenario estirado, estas moléculas bloquearán transitoriamente el cierre nuevamente de la cremallera de la horquilla cuando la fuerza se reduce hasta por debajo de  $\sim 11$  pN.
- 55 Midiendo la extensión de la molécula de ADN  $Z(t)$  (la distancia de la perla a la superficie) a lo largo de una serie de ciclos durante una de estas pausas de reapertura de la cremallera, se puede determinar la posición del bloqueo con una precisión de aproximadamente 1 nm (que corresponde a la distancia abarcada por dos nucleótidos (1 pb) en un ADNmc bajo una fuerza de 10 pN). Además, midiendo el tiempo medio de bloqueo, se puede determinar  $T_{off} = 1/k_{off}$ . Midiendo  $P_{bloqueo}$ , y conociendo la concentración de la molécula  $[M]$ , es posible 60 obtener acceso a  $T_{on}$ , y de este modo a  $k_{on}$ . Uno o ambos de estos parámetros ayuda a caracterizar la naturaleza de la unión. Por ejemplo, es posible determinar si es debido a una hibridación perfecta con un oligonucleótido complementario o no, o si una proteína estabiliza la hibridación o no, o si hay un error de apareamiento y dónde está (por ejemplo en el centro del oligonucleótido hibridado o cerca de uno de sus extremos).
- 65 Estas observaciones sugieren diversas formas de realización para aplicaciones en la detección de modificaciones del ADN, y más generalmente en la detección de la interacción de proteínas con ADNmc o

ADNbc.

Detección de modificaciones del ADN mediante detección mecánica de bloqueos durante la rehibridación

5 Si hay oligonucleótidos (de una longitud mayor que siete nucleótidos) en disolución cuando la horquilla de ADN se abre mecánicamente, estos oligonucleótidos se pueden emparejar con su secuencia complementaria en el ADN y evitar transitoriamente el cierre nuevamente total de la cremallera de la horquilla cuando la fuerza se reduce por debajo de 11 pN; véase la figura 1b. Se puede llevar a cabo fácilmente una serie de ciclos de apertura/cierre de nuevo en la misma molécula y detectar los bloqueos (pausa) sobre el cierre nuevamente de la cremallera debido al apareamiento de los oligonucleótidos con un ADN en la fase de cremallera abierta.

10 La duración del tiempo de bloqueo presenta típicamente una distribución exponencial cuyo valor medio  $T_{off}$  disminuye exponencialmente con  $F_{ensayo}$ . Esta distribución de probabilidad es reminiscente de la naturaleza de la molécula única de este ensayo. Tiene algunas consecuencias: el tiempo de bloqueo más probable es 0, lo que significa que existe una fracción sustancial de bloqueo que no detectaremos debido a que son más cortos que nuestra resolución experimental. La molécula que bloquea el replegamiento de la horquilla está bajo la presión de la horquilla de ADN. Si  $F_{ensayo}$  está próxima a 15 pN, (la fuerza mecánica de desplegamiento de la horquilla), esta presión es débil; por el contrario, si  $F_{ensayo}$  es reducida, la presión de la horquilla aumenta drásticamente, expeliendo la molécula. Encontramos que  $T_{off}$  disminuye exponencialmente con  $F_{ensayo}$ , como se muestra en la figura 2. Esta dependencia es tan fuerte que solamente podemos medir  $T_{off}$  en un intervalo de unos pocos pN. Obsérvese también que  $T_{off}(F)$  solamente coincidiría con el  $T_{off}$  clásico de una molécula que se separa espontáneamente cuando  $F_{ensayo} = F_{apertura} = 15$  pN, lo que no es lograble en la presente memoria.

15 La probabilidad de bloqueo  $P_{bloqueo}$  aumenta con la duración de la fase abierta  $T_{abierta}$ , con un comportamiento exponencial:  $P_{bloqueo} = a(F) \cdot [1 - \exp(-T_{abierta}/T_{on})]$ , como se muestra en la figura 3.

20 Como se puede esperar,  $P_{bloqueo}$  aumenta con la concentración de la molécula; en la figura 4 mostramos que para un oligonucleótido de 12 nt,  $P_{bloqueo}$  aumenta y se satura con  $[M]$ .

30 Conociendo  $T_{abierta}$  y la concentración de la molécula  $[M]$ , es posible deducir  $k_{on}$  a partir de  $P_{bloqueo}$  usando la siguiente relación:

$$k_{on} = -\text{Log}(1 - P_{bloqueo}/a(F)) / ([M]T_{abierta}).$$

35 La fuerza de la unión (véase la figura 5) se puede caracterizar por:

$$k_d^{-1} = -(T_{off} \text{Log}(1 - P_{bloqueo}/a(F))) / ([M]T_{abierta})$$

40 El tiempo medio de bloqueo  $T_{off}$  depende del tamaño del oligonucleótido, de la fuerza  $F_{ensayo}$  aplicada durante el cierre de nuevo de la cremallera, de la temperatura, y no significativamente de la fuerza iónica del amortiguador usado.

45 El  $T_{on}$  depende también del tamaño del oligonucleótido, de la temperatura, de la fuerza iónica del amortiguador, pero no significativamente de  $F_{ensayo}$ . Como se muestra en la Figura 5, los emparejamientos erróneos entre el oligonucleótido y el sustrato también se pueden caracterizar midiendo estas constantes cinéticas. Por ejemplo, un oligonucleótido de 12 nts completamente complementario presenta una  $k_{on}$  de  $1,5 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ ; la introducción de un único apareamiento erróneo 3 bases alejado de un extremo no altera mucho  $k_{on}$ .

50 Mover el apareamiento erróneo a la mitad del oligonucleótido reduce  $k_{on}$  por un factor de 10.

55  $T_{off}$  también depende de la presencia de proteínas de unión a ADNbc que pueden estabilizar el híbrido. Por ejemplo, hemos mostrado que una primasa estabilizará oligos de ADN que de otro modo no habrían sido suficientemente estables para bloquear la rehibridación de la horquilla durante un tiempo suficientemente prolongado para ser detectada; véase la figura 6. De manera similar, la unión de una polimerasa al extremo 3' de un pequeño oligonucleótido usado como cebador incrementará su estabilidad; este ensayo se puede usar para determinar la afinidad de la polimerasa por su sitio de cebador. De forma similar, si una proteína se une a un sitio de ADNmc específico (por ejemplo, una base metilada), bloqueará el cierre de nuevo de la cremallera en un sitio específico y durante un tiempo suficiente para ser detectado.

60 La técnica se puede usar para identificar modificaciones del ADN a lo largo de un ADNmc o un ADNbc. De este modo, sondando las horquillas de ADN que anclan las perlas a la superficie con un anticuerpo (Ab) dirigido contra una modificación específica de una de sus bases, se puede detectar la existencia y posición de esta base modificada a lo largo de la cadena vía el bloqueo transitorio que resultará de la unión del Ab con la rehibridación de la horquilla. Sondar el sitio de unión con un conjunto de oligonucleótidos complementarios permitirá la identificación del fragmento de ADN que exhibe esa modificación.

65

Detección de la afinidad de unión de RecQ a un molde de ADNmc.

Las helicasas se unen a espacios de ADNmc a fin de desenrollar ADNbc. La actividad de estas enzimas depende directamente de su afinidad por ADNmc. Proponemos en la presente memoria medir este parámetro directamente con nuestro ensayo. Esto se puede hacer con o sin ATP o ADP, u otros análogos. Se presentan en la presente memoria algunos resultados relativos a la helicasa RecQ procedente de *E. coli* sin ATP. En la figura 7 se puede observar la señal de unión típica; permite medir  $P_{\text{bloqueo}}$  para una concentración de helicasa. En la Figura 8 se presenta la evolución de  $P_{\text{bloqueo}}$  frente [RecQ]. Se observa que la concentración característica de [RecQ] es igual a 226 pM. En la figura 9, se observa que la helicasa se une de forma no específica. Finalmente, el bloqueo por la enzima presenta un comportamiento de descenso: la posición Z no es realmente constante, sino que disminuye mediante múltiples etapas. Con este comportamiento, es difícil definir un valor real de  $T_{\text{off}}$ , y de este modo solamente podemos medir  $T_{\text{on}}$  y  $k_{\text{on}}$ .

El pico en  $Z = 0$  no corresponde a un bloqueo, sino justamente al replegamiento directo. El bloqueo por RecQ se encuentra uniforme a lo largo del molde; el decaimiento en  $0,9 \mu\text{m}$  es debido al promedio de las moléculas que tienen una extensión ligeramente diferente.

Detección de la metilación

Figura 11: Histograma del tiempo de bloqueo por el anticuerpo contra 5mC. La mayoría de los bloqueos son cortos, y se pueden ajustar razonablemente bien a una distribución exponencial con un tiempo característico de 1,3 s. Sin embargo, un número sustancial de bloqueos, 17,5%, supera los 30 s. En esta condición, no es muy fácil determinar el  $T_{\text{off}}$  de la enzima; se cree que están compitiendo dos mecanismos de unión diferentes, siendo uno más fuerte que el otro.

Alternativamente, se puede sondear la existencia de modificaciones del ADN conocidas mediante hibridación de un oligonucleótido complementario al sitio modificado putativo en presencia (o no) de una proteína que reconoce la modificación (tal como la proteína 1 del dominio de unión de metilo (MBD1), que reconoce citosinas metiladas, o un Ab apropiado provocado contra un ADNbc modificado específicamente). El tiempo de bloqueo en presencia de la proteína aumentará significativamente, conduciendo a una identificación y localización fáciles de la base modificada.

Usando proteínas que reconocen emparejamientos erróneos, se podría usar de forma similar el método mencionado anteriormente para identificar emparejamientos erróneos (es decir, SNPs) a lo largo del ADN. También se puede usar ese ensayo para detectar proteínas (o fármacos) que afectarán a la estabilidad de un complejo dado de proteína/ADN.

Parámetros que influyen el ensayo.

$F_{\text{abierta}}$ : un valor de 20 pN es una buena elección debido a que esto asegura que un gran número de perlas se abrirán simultáneamente (variando su magnetización, y de este modo su fuerza, en 10 a 20%).

$T_{\text{abierta}}$  aparece como un parámetro importante en combinación con la concentración de moléculas: para observar el bloqueo, se debe usar una combinación de ambos parámetros, que conduce a un valor sustancial de  $P_{\text{bloqueo}}$  según la fórmula:

$$P_{\text{bloqueo}} = a(F) \cdot [1 - \exp(-T_{\text{abierta}} \cdot k_{\text{on}} \cdot [M])].$$

Si se quiere medir  $k_{\text{on}}$ , es sensato evitar saturar  $P_{\text{bloqueo}}$ , el ajuste de  $[M]$  y  $T_{\text{abierta}}$  para lograr una  $P_{\text{bloqueo}}$  en el intervalo 0,2 a 0,5 asegurará un número mínimo de ciclos para lograr estadísticas razonables. Obsérvese que  $T_{\text{abierta}}$  se puede modificar simplemente ajustando un parámetro en el programa de adquisición, el cambio de la concentración de enzima requiere cambiar el amortiguador en la cámara de flujo. Por otro lado, si  $k_{\text{on}}$  no se va a medir, merece la pena saturar  $P_{\text{bloqueo}}$ , esto producirá la mejor estadística de bloqueo. La concentración de moléculas se puede limitar por su suministro o mediante unión no deseada, por ejemplo en el estudio usando anticuerpo frente a 5mC; a una concentración elevada, esta enzima se une al ADN bicatenario de la horquilla en su estado cerrado, evitando su despliegue. Se ha encontrado que limitando la concentración de enzima por debajo de 35 nM se resuelve este aspecto. En estos experimentos, incrementar  $T_{\text{abierta}}$  es la única vía para incrementar  $P_{\text{bloqueo}}$ .

El parámetro  $a(F)$  está en principio próximo a 1; la mejor forma de evaluar este valor es llevar a cabo un ensayo de saturación que varía ya sea  $T_{\text{abierta}}$  o  $[M]$  hasta que  $P_{\text{bloqueo}}$  alcanza asintóticamente  $a(F)$ , como en las Figuras 3 y 4. Como alternativa, es posible estimar  $a(F)$  con la siguiente fórmula:

$$a(F) = \exp(-T_{\text{muerto}}/T_{\text{off}})$$

en la que  $T_{\text{muerto}}$  es el tiempo muerto del sistema de detección, y  $T_{\text{off}}$  es el tiempo de bloqueo medio. Típicamente,

$T_{\text{muerto}}$  es del orden de 0,1 s.

$F_{\text{ensayo}}$  es un parámetro muy importante que se debe ajustar: su intervalo depende de la horquilla usada, pero abarca típicamente [12 pN, 2 pN]. Para fuerzas mayores que 12 pN, el repliegamiento de la horquilla presenta ya algunos bloqueos debido a estructura secundaria que forma el ADN<sub>mc</sub>, que enmascara señales interesantes. A fuerzas bajas, la extensión del ADN se hace muy pequeña, y el ruido aumenta drásticamente. La presión de horquilla de la horquilla empuja la molécula a deshidratarse muy eficientemente, y se observa que el  $T_{\text{off}} = T_0 \exp F/F_0$ ; de este modo,  $T_{\text{off}}$  disminuye muy rápidamente a medida que se reduce  $F_{\text{ensayo}}$ . Por ejemplo, un oligonucleótido de 9 nts producirá un bloqueo de 1 s a alrededor de  $F_{\text{ensayo}} = 11$  pN; a una fuerza por debajo de 9 pN, el bloqueo es apenas visible ( $a(F)$  se hace muy pequeño). Para un oligonucleótido de 12 nts, el intervalo de observación es [10 pN, 6 pN]. Para un oligonucleótido de 37 nts, el bloqueo dura para siempre a 6 pN, pero cae hasta unos pocos segundos a  $F_{\text{ensayo}} = 3$  pN. La misma observación es verdad para la proteína de unión: cuanto más fuerte es la unión, menor es la fuerza a la que se observa bloqueo.

Se ajustó  $F_{\text{ensayo}}$  de manera que el tiempo de bloqueo es medible ( $a(F) - 1$ ), pero no demasiado prolongado de manera que  $T_{\text{ensayo}}$  es relativamente corto, permitiendo que se realicen muchos ciclos.

En este ensayo, se puede medir  $T_{\text{off}}$  en un intervalo de 0,2 s a 20 s. Se pudo observar un tiempo más corto con un dispositivo de medida más rápido, como una videocámara más rápida; tiempos más prolongados conducen a una adquisición muy prolongada, puesto que se necesita lograr algunos ciclos para promediar la distribución. Para oligonucleótido,  $T_{\text{off}}$  varía exponencialmente con  $F_{\text{ensayo}}$ ; de este modo, se puede ajustar  $F_{\text{ensayo}}$  para llevar a  $T_{\text{off}}$  al intervalo utilizable. Para proteína, la variación de  $T_{\text{off}}$  con  $F_{\text{ensayo}}$  no es conocida, pero se observó que la disminución de  $F_{\text{ensayo}}$  habitualmente disminuye drásticamente  $T_{\text{off}}$ . Sin embargo, a priori,  $T_{\text{off}}$  es desconocido, y puede variar en un amplio intervalo. Para tener una idea del valor típico de  $F_{\text{ensayo}}$ , se ha encontrado que es conveniente lograr primeramente una serie de ciclos con la fuerza aumentando y disminuyendo, seguido de una pendiente a lo largo de unos pocos segundos como se hace en la Figura 12. El extremo de la fase de bloqueo corresponde a una fuerza  $F_c$ . La distribución de  $F_c$  alcanza un máximo para un valor en el que  $T_{\text{off}}$  es del orden de la duración de la pendiente.

Entonces se puede proceder con los ciclos que tienen mesetas en la fuerza ( $F_{\text{abierta}}$  y  $F_{\text{ensayo}}$ ), siendo  $F_{\text{ensayo}}$  ligeramente más grande que  $\langle F_c \rangle$ , para obtener un  $T_{\text{off}}$  en el intervalo medible.

$T_{\text{ensayo}}$  y  $N_{\text{ciclos}}$ :  $T_{\text{ensayo}}$  debería ser 2 o 3 veces más grande que  $T_{\text{off}}$ . Finalmente, el número de ciclos define la exactitud global de la medida. Para lograr un X% de exactitud, se necesita  $X/100 = 1/N_{\text{bloqueo}}^{1/2}$  ya que  $P_{\text{bloqueo}} = N_{\text{bloqueo}}/N_{\text{ciclo}}$ ; En un  $N_{\text{ciclo}} = 10000/(X^2 P_{\text{bloqueo}})$ .

Mejorando el ensayo: frecuentemente surgen diversos problemas; la unión de una enzima puede presentar sucesos cortos y también sucesos muy largos (figura 9); esta situación última dará como resultado que los bloqueos todavía sean activos mientras el final de la fase de ensayo y el comienzo del nuevo ciclo empieza (figura 7). Puesto que el bloqueo se oculta durante la fase abierta, el bloqueo que se extiende a lo largo de ciclos sucesivos es probable pero nunca un suceso demostrado. Para evitar esta situación complicada, es posible aprovechar el hecho de que el bloqueo es habitualmente muy corto a fuerzas bajas. De este modo, añadiendo una tercera fase después del ensayo, alguien con una fuerza baja puede limpiar la horquilla de cualquier molécula unida, con  $F_{\text{limpio}} = 0,5$  pN y  $T_{\text{limpio}} = 2$  s, eliminamos cualquier molécula unida y preparamos una horquilla limpia para el siguiente ciclo. Una molécula también puede estar presente en varios sitios de unión, y de este modo la señal de bloqueo tendrá un aspecto de escalera en la que tras un primer bloqueo la molécula se bloquea en el segundo sitio de unión, etc. (figura 10). Para el segundo bloqueo, la fase abierta eficaz es  $T_{\text{abierta}} + T_{\text{bloqueo1}}$  (figura 10); si  $T_{\text{bloqueo1}}$  es mayor que  $T_{\text{abierta}}$ , probablemente se observará un segundo bloqueo tras un primero, complicando la medida de los parámetros cinéticos. Entonces es mejor usar un  $T_{\text{abierta}}$  grande, en comparación con  $T_{\text{ensayo}}$ , para minimizar este efecto.

#### Listado de secuencias

<110> CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)  
ECOLE NORMALE SUPERIEURE

UNIVERSITE PIERRE ET MARIE CURIE (PARIS 6)

David BENSIMON, Vincent CROQUETTE, Harold GOUET, Jean-Francois ALLEMAND, Fang-Yuan DING

<120> PROCEDIMIENTO PARA LA DETECCIÓN DE MODIFICACIONES DE ADN Y UNIÓN DE PROTEÍNAS POR MANIPULACIÓN DE UNA SOLA MOLÉCULA

<130> 364874D31906

<140> EP 13305074.0

<141> 2013-01-22

<160> 1

5 <170> PatentIn version 3.5

<210> 1

<211> 12

<212> ADN

10 <213> Secuencia artificial

<220>

<223> Oligonucleótido

15 <400> 1

acagccagcc ga 12

**REIVINDICACIONES**

1. Método para detectar por lo menos una base modificada en una molécula de ácido nucleico bicatenario que comprende una secuencia de ácido nucleico, en el que dicha molécula de ácido nucleico bicatenario es una horquilla, y en el que uno de los extremos del ácido nucleico bicatenario está unido directa o indirectamente a un soporte, y en el que el otro extremo del ácido nucleico bicatenario está unido a un soporte móvil, comprendiendo dicho método las etapas de:
- 5
- 10 a) desnaturalizar dicha molécula de ácido nucleico bicatenario que comprende una secuencia de ácido nucleico aplicando una fuerza física a dicha molécula;
- b) proporcionar una proteína apta para unirse a dicha base modificada;
- 15 c) renaturalizar dicha molécula de ácido nucleico bicatenario en presencia de dicha proteína;
- d) detectar un bloqueo de la renaturalización del ácido nucleico bicatenario; y
- 20 e) determinar la posición de dicho bloqueo en dicha molécula de ácido nucleico bicatenario, comprendiendo dicha determinación las etapas de:
- 25
- medir la distancia (z) entre los dos extremos de la molécula de ácido nucleico bicatenario que están unidos al soporte;
  - medir la distancia ( $Z_{\text{elevada}}$ ) entre los dos extremos de la molécula de ácido nucleico bicatenario que están unidos al soporte, cuando dicha molécula de ácido nucleico bicatenario se desnaturaliza; y
  - comparar z y  $Z_{\text{elevada}}$ .
- 30 2. Método según la reivindicación 1, en el que en la etapa b) se proporciona además una molécula de ácido nucleico monocatenario que corresponde a dicha secuencia.
3. Método según cualquiera de las reivindicaciones 1 o 2, en el que el ácido nucleico bicatenario se desnaturaliza en la etapa a) alejando los soportes.
- 35 4. Método según cualquiera de las reivindicaciones 1 a 3, en el que se aplica una fuerza física superior o igual a 15 pN, preferentemente superior o igual a 17 pN, más preferentemente superior o igual a 18 pN, a la molécula bicatenaria alejando los soportes.
- 40 5. Método según cualquiera de las reivindicaciones 1 a 4, en el que el ácido nucleico bicatenario desnaturalizado se renaturaliza en la etapa c) juntando los soportes.
- 45 6. Método según la reivindicación 5, en el que la fuerza aplicada a la molécula bicatenaria se reduce hasta menos de o igual a 12 pN, preferentemente menos de o igual a 11 pN, más preferentemente menos de o igual a 10 pN, juntando los soportes.
- 50 7. Método según cualquiera de las reivindicaciones 1 a 6, en el que las etapas a) a d) se repiten varias veces (para acumular unas medidas y aumentar la relación de señal/ruido).
8. Método según cualquiera de las reivindicaciones 1 a 7, que comprende una etapa adicional de medir la duración del bloqueo.
9. Método según la reivindicación 8, que comprende una etapa adicional de comparar la duración del bloqueo con un valor de referencia.
- 55 10. Método según cualquiera de las reivindicaciones 1 a 9, que comprende además una etapa de secuenciar la secuencia de ácido nucleico unida por dicha proteína.
- 60 11. Método según cualquiera de las reivindicaciones 1 a 10, en el que por lo menos una base modificada se selecciona de entre una modificación epigenética, un apareamiento erróneo, o un SNP.
- 65 12. Método según cualquiera de las reivindicaciones 1 a 11, en el que la base modificada se selecciona de entre el grupo que consiste en 4-metilcitosina (4mC), 5-metilcitosina (5mC), 5-hidroximetilcitosina (5hmC), 5-formilcitosina (5fC), 5-carboxilcitosina (5caC), 5-hidroximetiluracilo (5hmU), y N6-metiladenosina (m6A).
13. Método según la reivindicación 12, en el que la base modificada es 5mC.

14. Método según cualquiera de las reivindicaciones 1 a 13, en el que dicha proteína es un anticuerpo o una enzima apto/a para unirse a dicha base modificada.

5 15. Método según cualquiera de las reivindicaciones 1 a 11, en el que dicha proteína se selecciona de entre el grupo que consiste en un dímero de MutS, Msh2/Msh6 (MutS $\alpha$ ), y Msh2/Msh3 (MutS $\beta$ ).

16. Método según cualquiera de las reivindicaciones 1 a 11 o 14, en el que la detección de un SNP en una secuencia contenida en una molécula de ácido nucleico comprende las etapas de:

10 a) hibridar dicho ácido nucleico con un ácido nucleico monocatenario que comprende la secuencia encontrada en la mayoría de la población; y

b) detectar el apareamiento erróneo resultante mediante el método según cualquiera de las reivindicaciones 1 a 11 o 14.

15

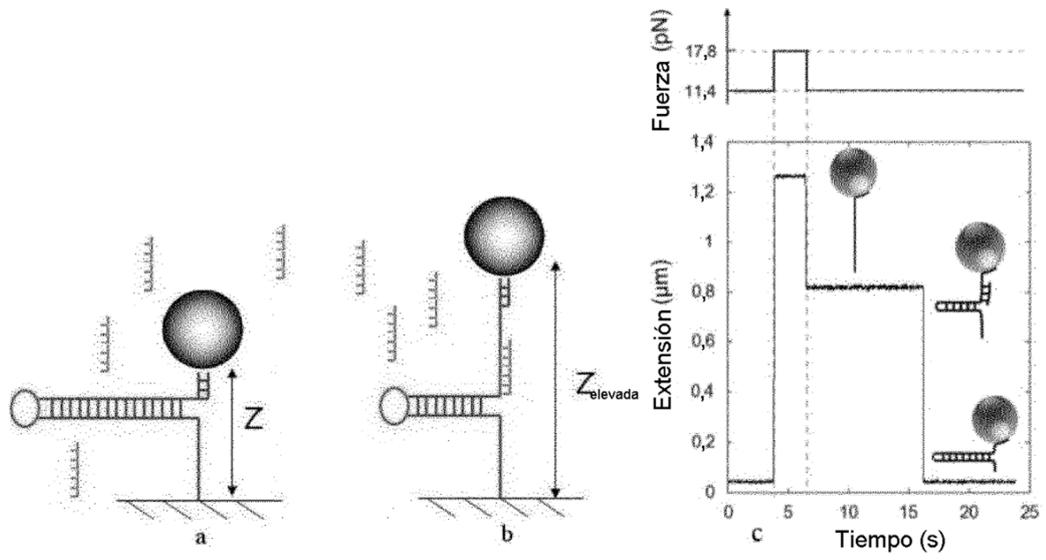


Figura 1

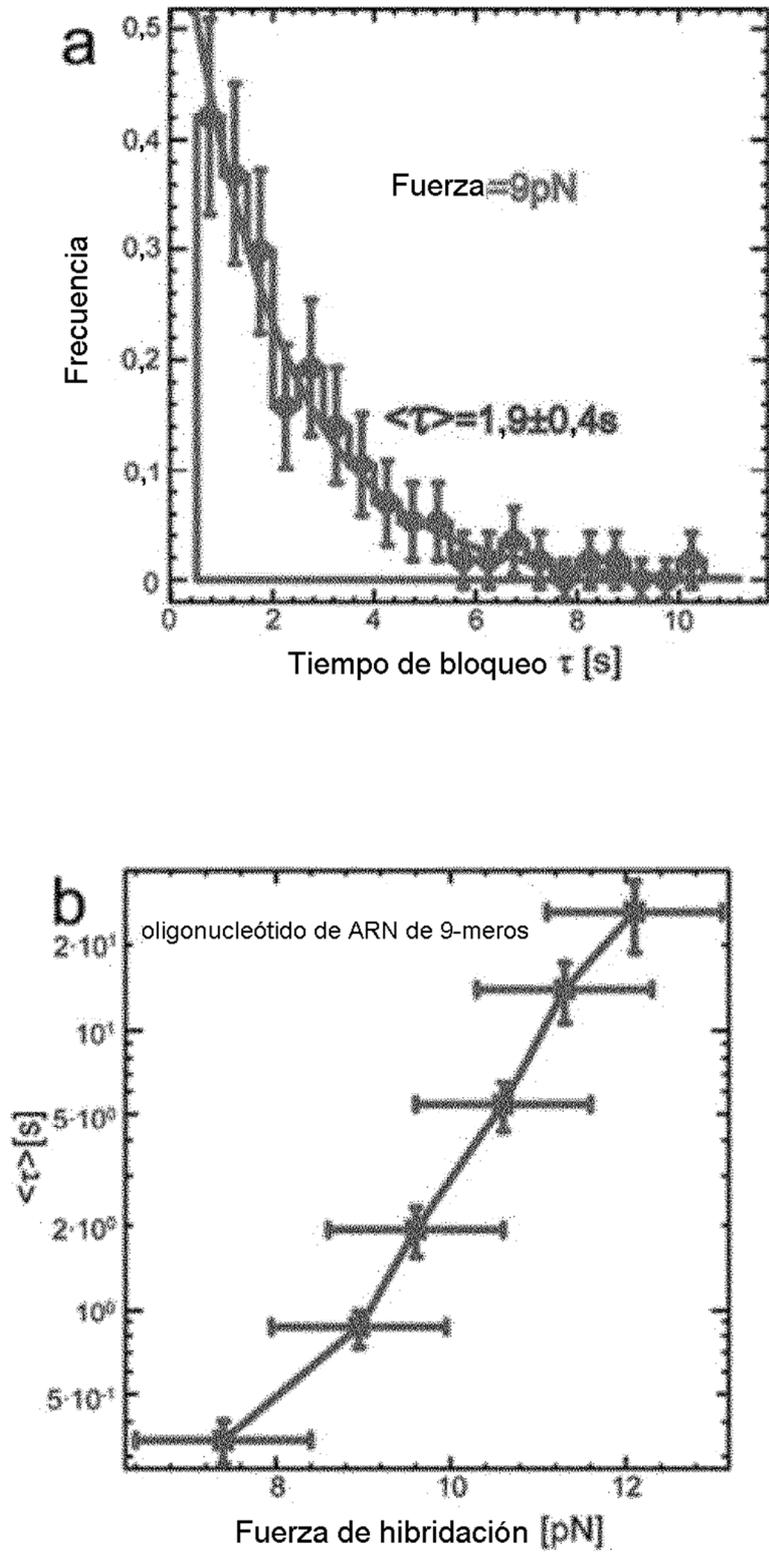


Figura 2

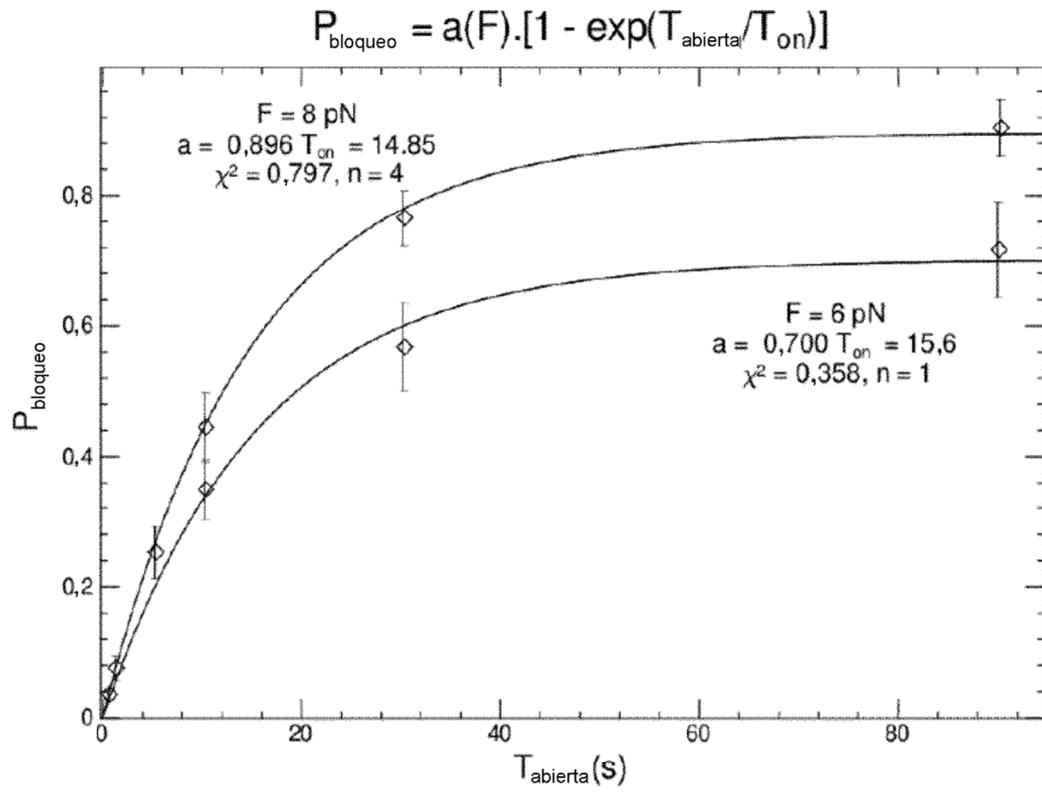


Figura 3

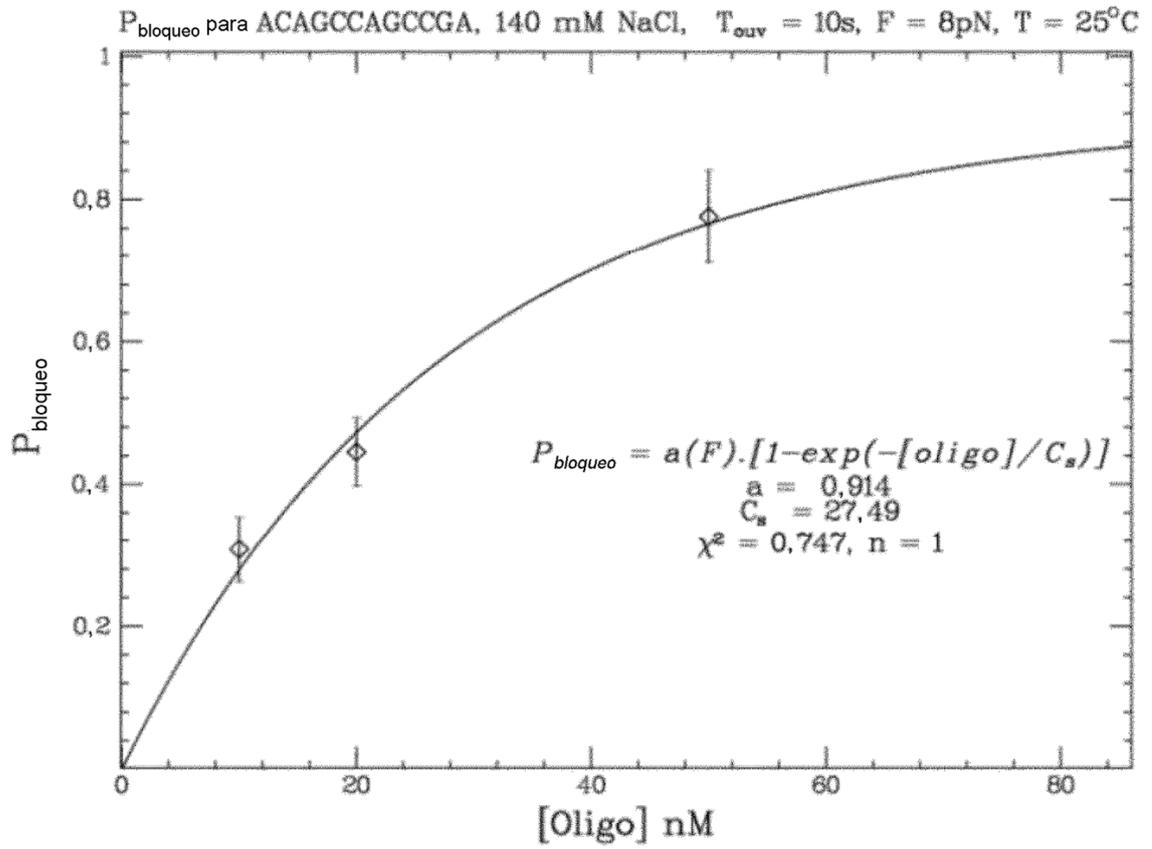


Figura 4

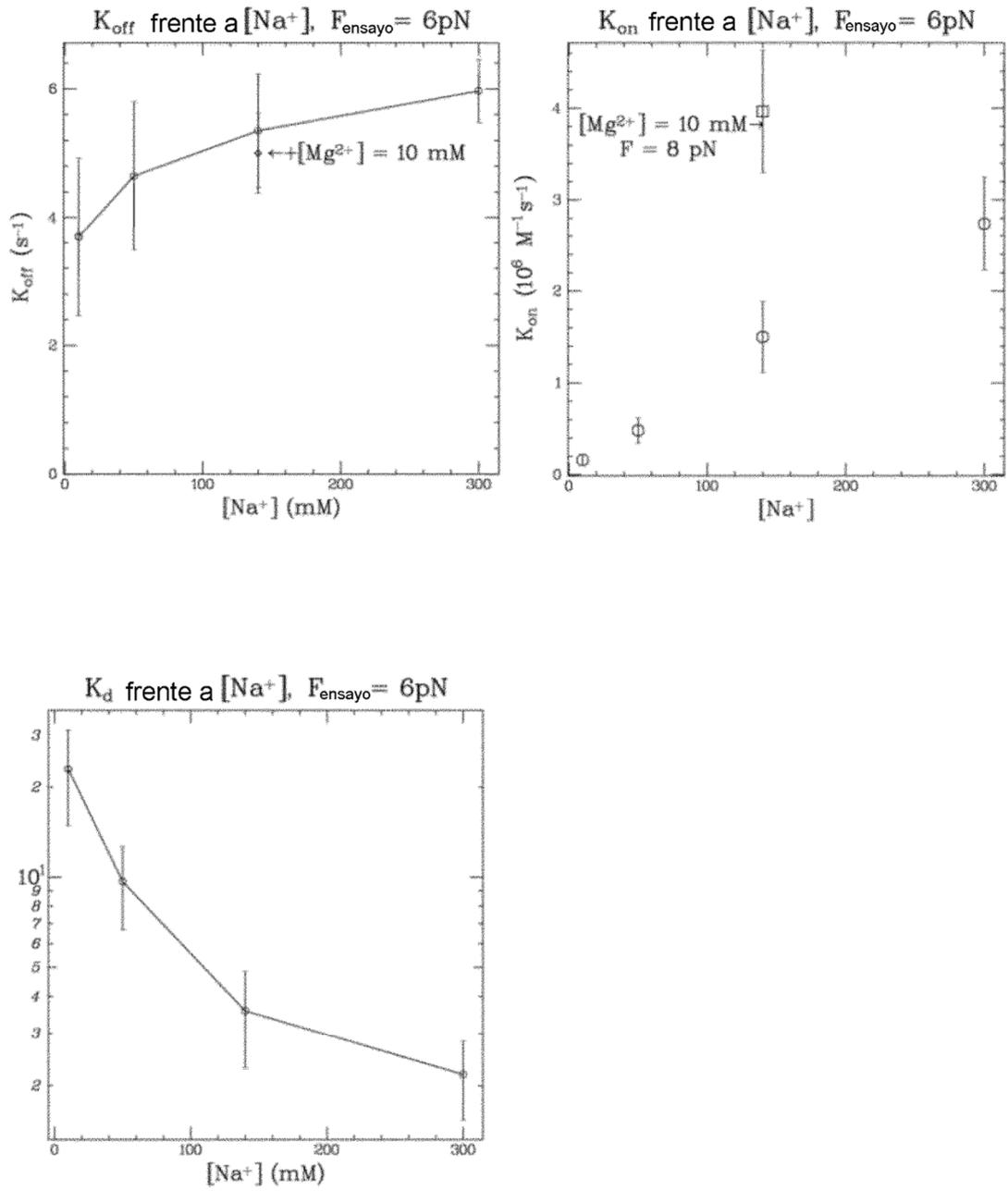


Figura 5

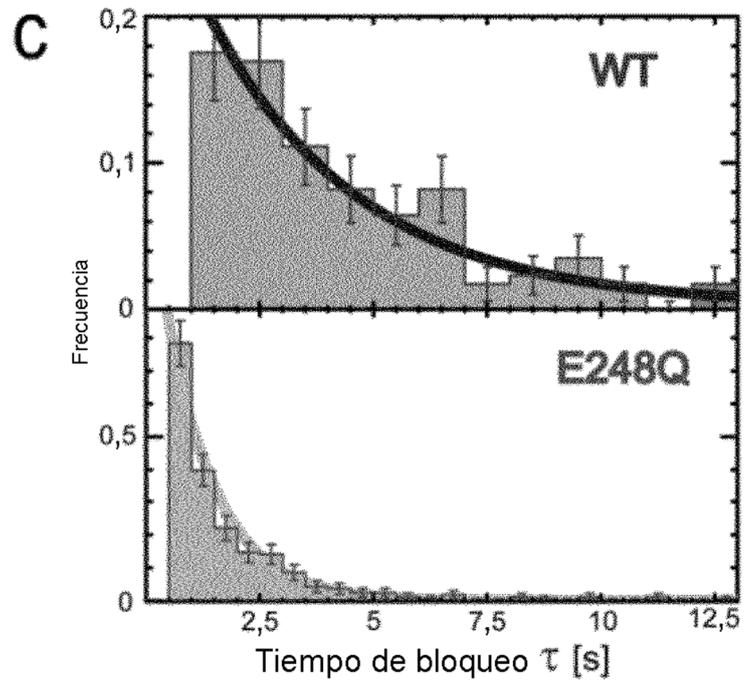
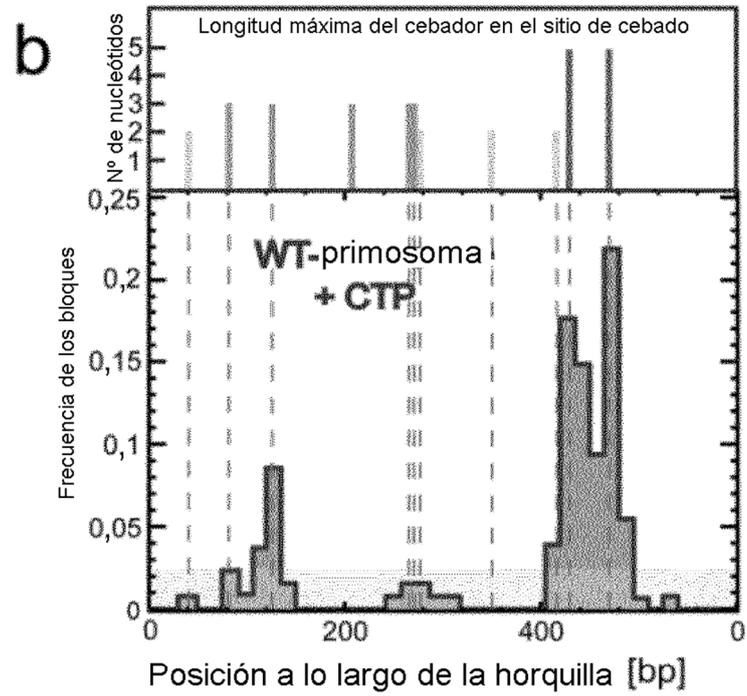


Figura 6

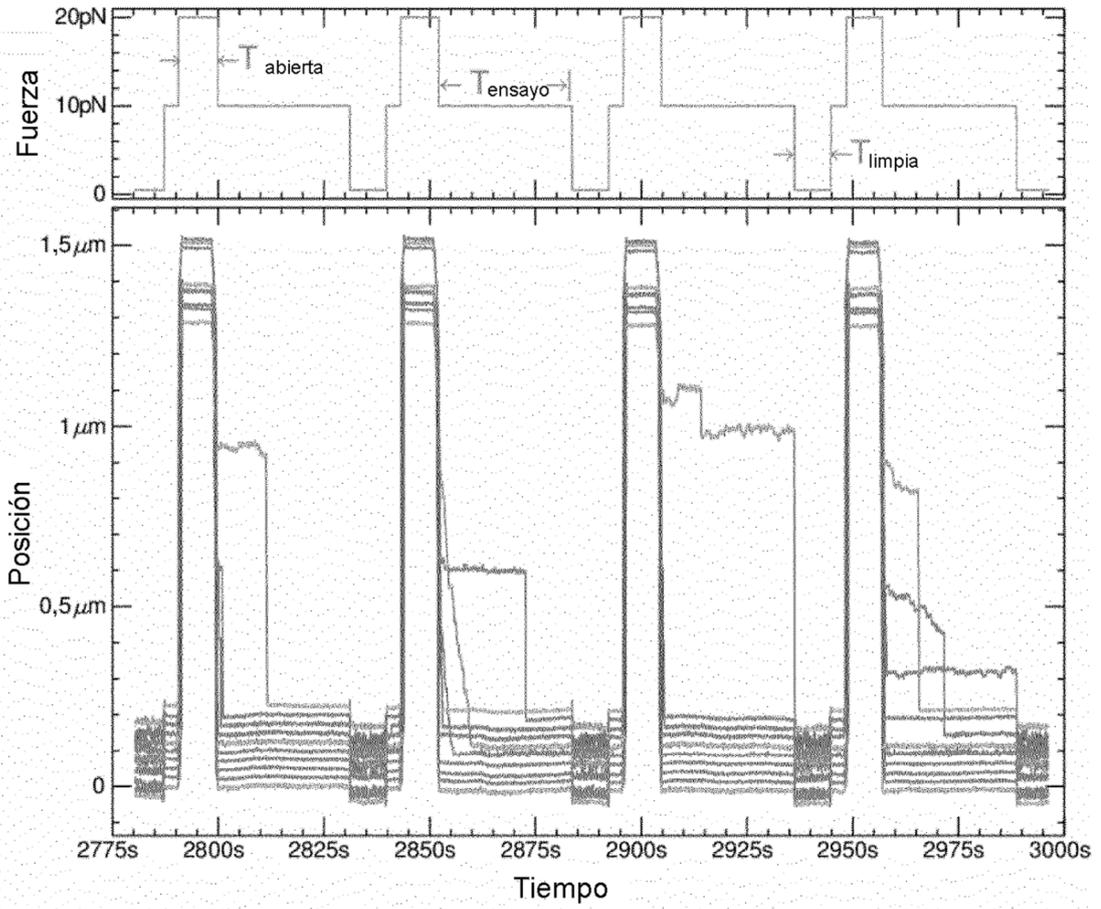


Figura 7

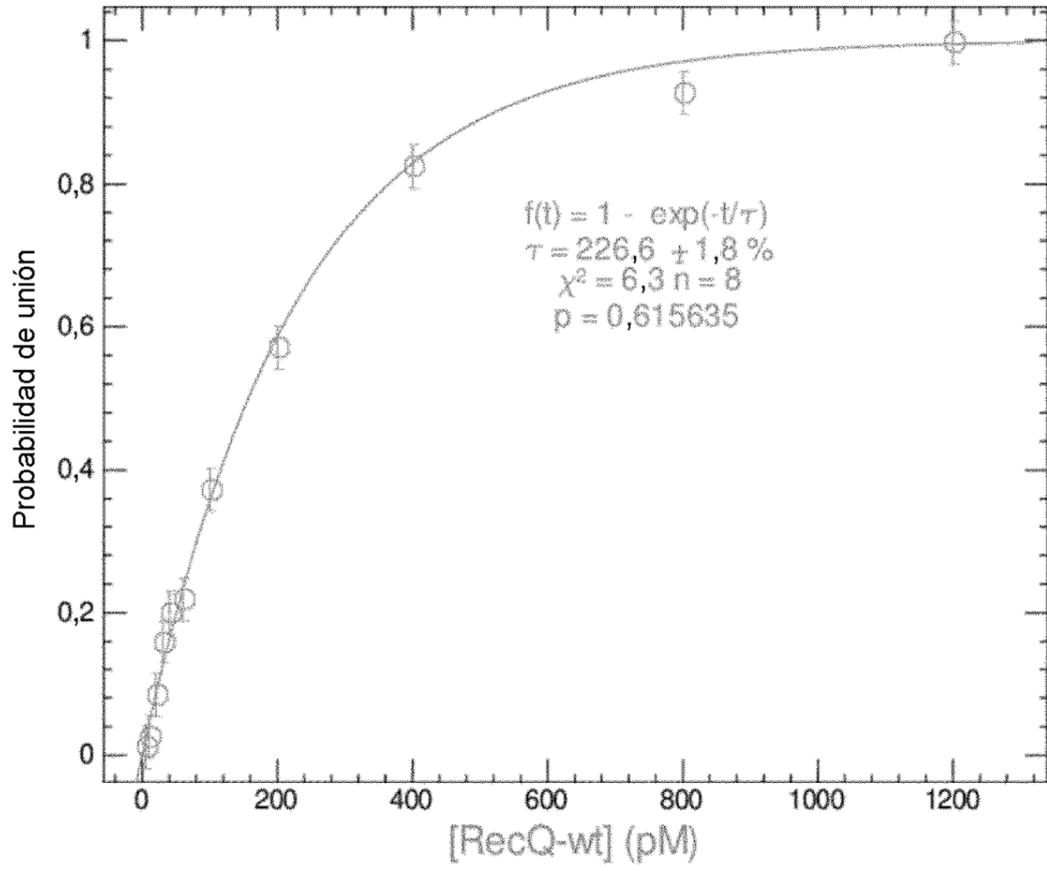


Figura 8

Posición de bloqueo de RecQ  
bin: 0,015 con 229817 pts sin ATP

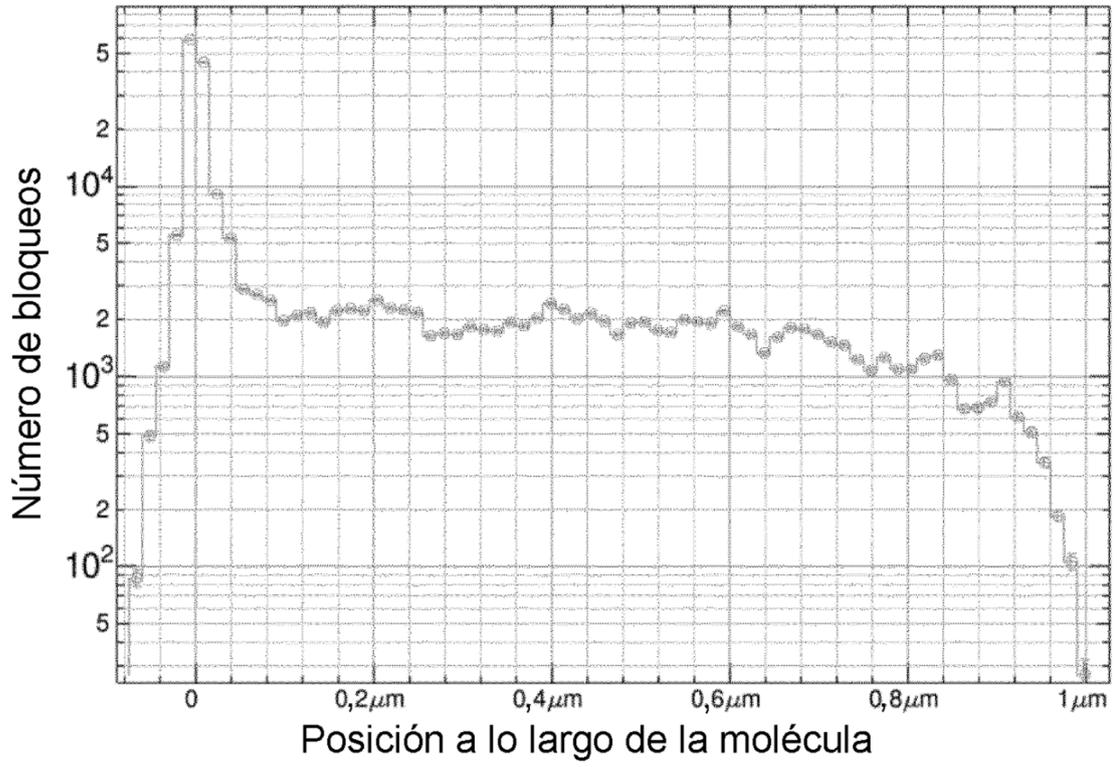


Figura 9

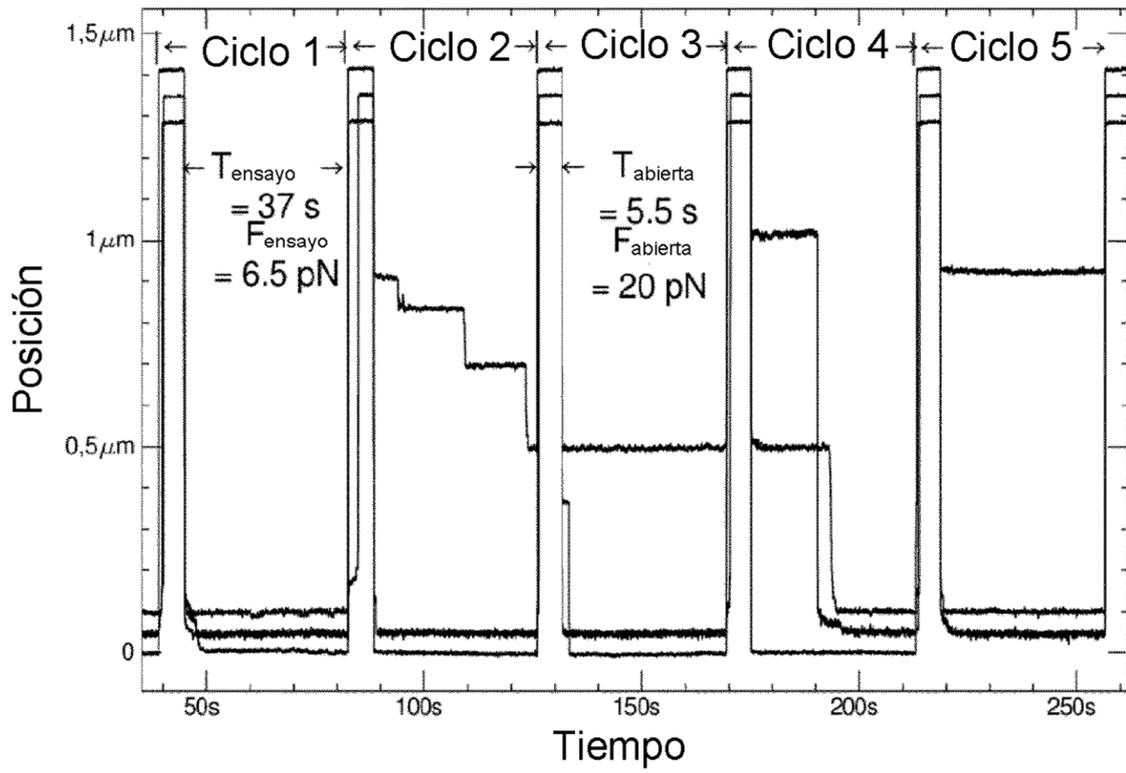


Figura 10

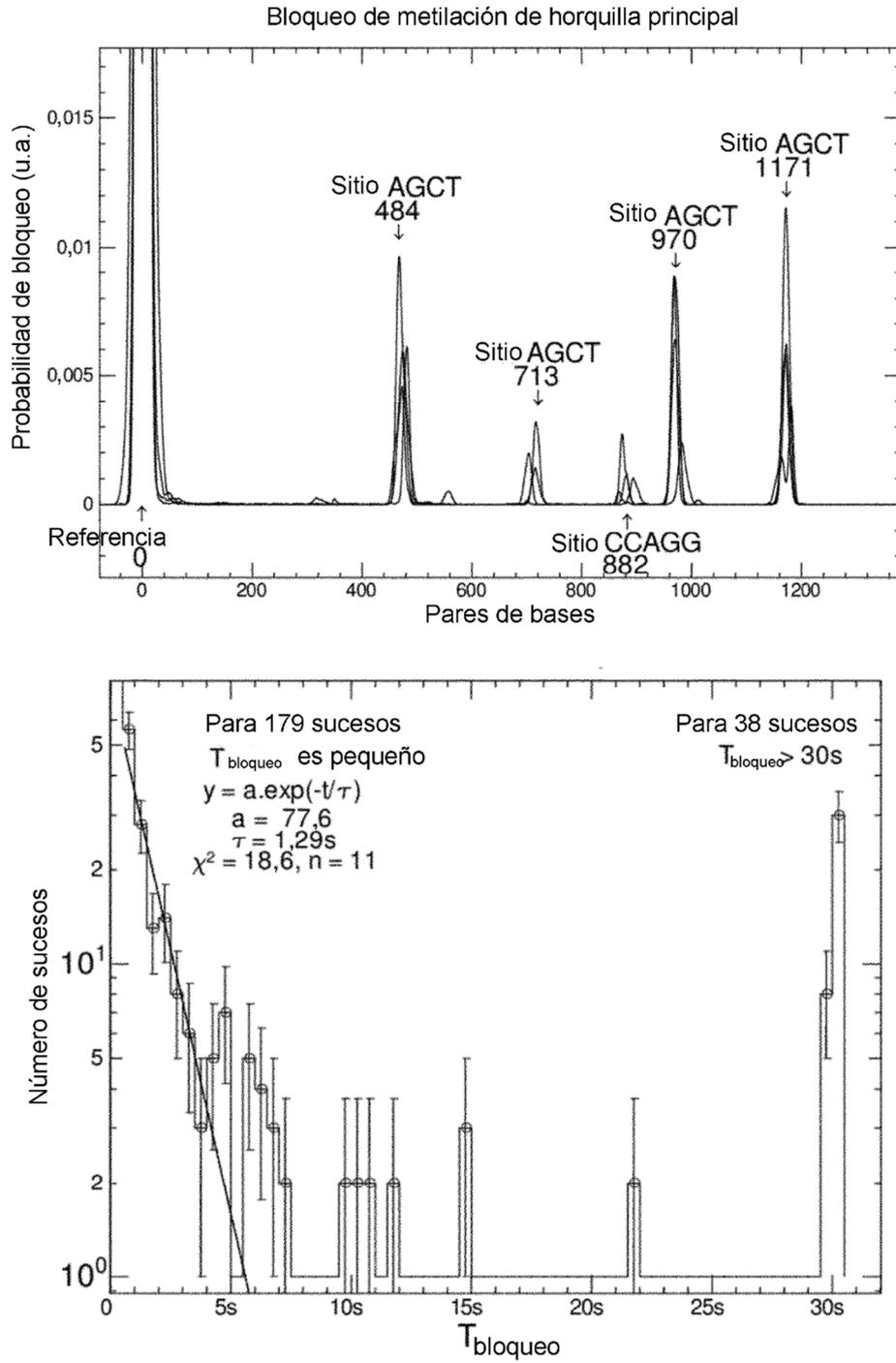
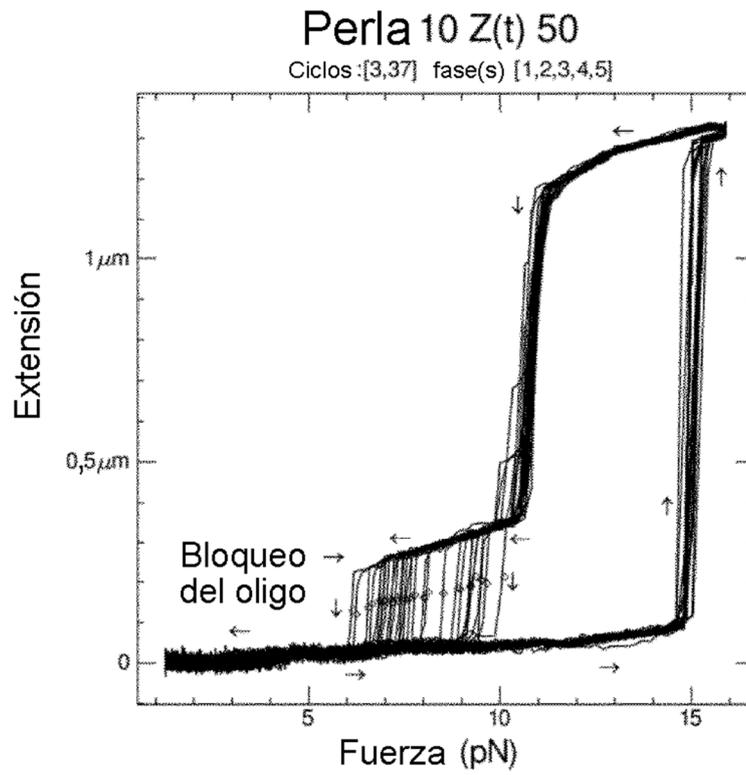


Figura 11



histograma de Z(t)bd10track50.gr  
bin: 1pN con 35 puntos

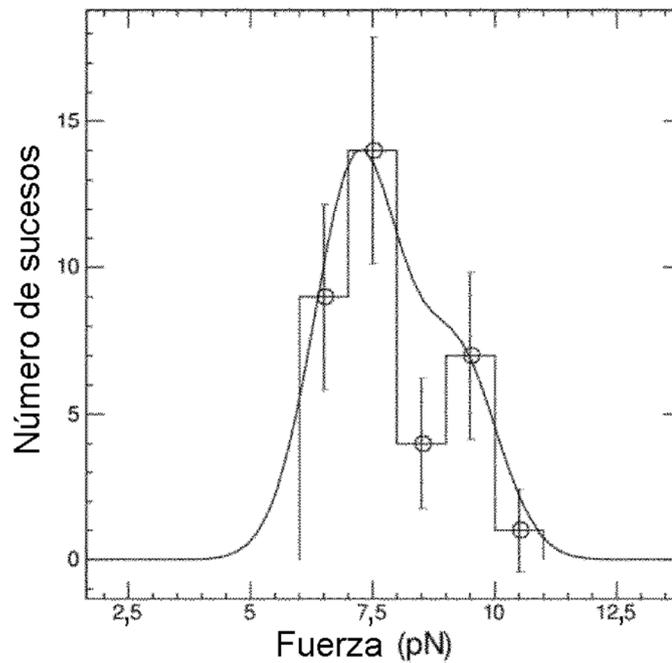


Figura 12

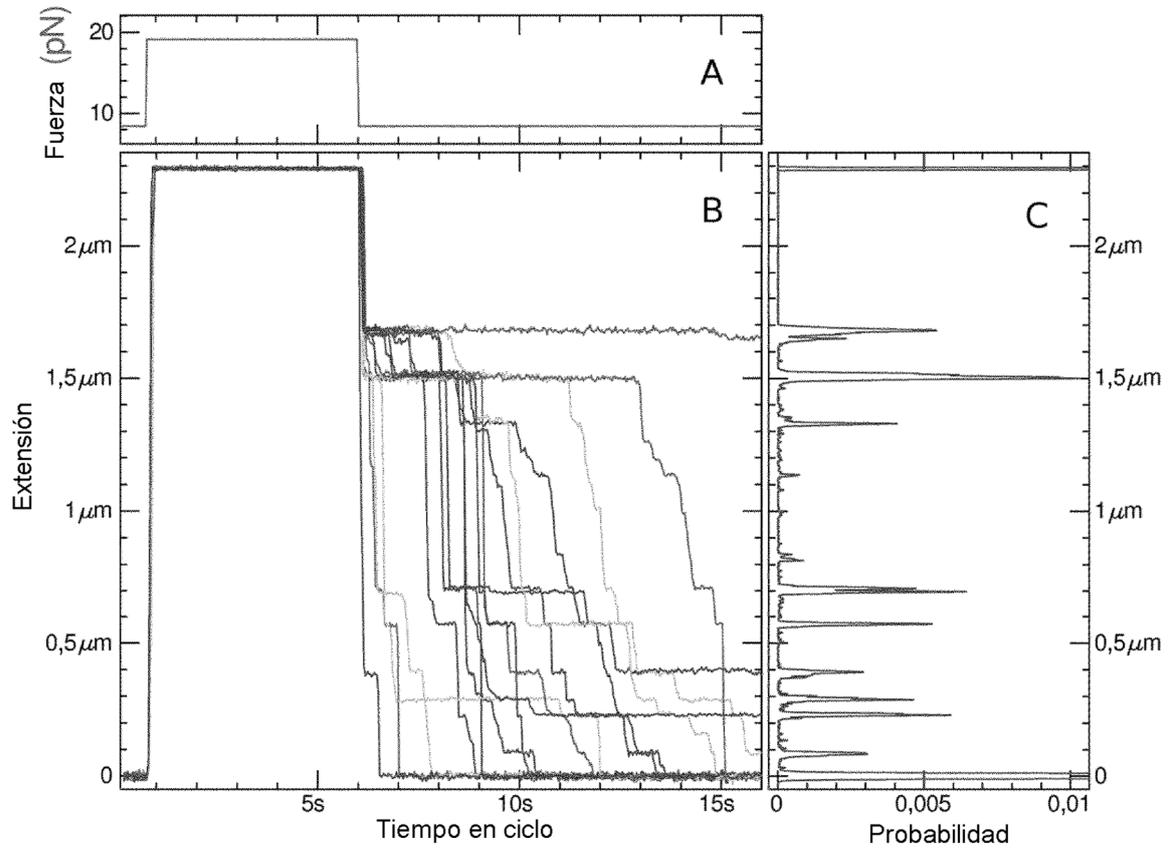


Figura 13