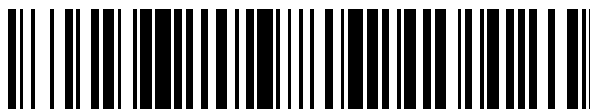


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 686 043**

51 Int. Cl.:

**C12Q 1/68**

(2008.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **17.09.2012** **PCT/EP2012/068250**

87 Fecha y número de publicación internacional: **21.03.2013** **WO13038010**

96 Fecha de presentación y número de la solicitud europea: **17.09.2012** **E 12759455 (4)**

97 Fecha y número de publicación de la concesión europea: **06.06.2018** **EP 2756098**

54 Título: **Método para preparar una biblioteca de moléculas de ácido nucleico**

30 Prioridad:

**16.09.2011 EP 11181546**

**24.07.2012 EP 12177647**

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

**16.10.2018**

73 Titular/es:

**LEXOGEN GMBH (100.0%)**

**Campus-Vienna-Biocenter 5, Helmut Qualtinger-  
Gasse 6**

**1030 Vienna, AT**

72 Inventor/es:

**SEITZ, ALEXANDER;**

**MOLL, PAMELA y**

**NAPORA, MAGDALENA ANNA**

74 Agente/Representante:

**ARIAS SANZ, Juan**

ES 2 686 043 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**

Método para preparar una biblioteca de moléculas de ácido nucleico

**5 Campo de la invención**

La presente invención se refiere al campo de la amplificación o el análisis de muestras de ácidos nucleicos mediante amplificación de partes de secuencias definidas.

**10 Antecedentes**

Se conocen bien y se han establecido en la técnica numerosos métodos basados en amplificación para la amplificación y detección de ácidos nucleicos diana. La reacción en cadena de la polimerasa, habitualmente denominada PCR, usa múltiples ciclos de desnaturalización, emparejamiento de pares de cebadores con cadenas opuestas y extensión de cebadores para aumentar exponencialmente los números de copias de la secuencia diana (documentos US 4.683.195; US 4.683.202; US 4.800.159; US 5.804.375). En una variación denominada RT-PCR, se usa transcriptasa inversa (RT) para preparar un ADN complementario (ADNc) a partir de ARN y el ADNc se amplifica después mediante PCR para producir múltiples copias de ADN (documentos US 5.322.770; US 5.310.652).

20 Las reacciones de PCR comprenden en general llevar a cabo múltiples ciclos de:

(A) hibridación (emparejamiento) de un primer cebador con un sitio en una cadena de ácido nucleico en un extremo de la secuencia de ácido nucleico diana e hibridación de un segundo cebador con un sitio correspondiente al extremo opuesto de la secuencia diana en la cadena de ácido nucleico complementaria;

25 (B) síntesis (extensión) de una secuencia de ácido nucleico desde cada cebador respectivo; y

(C) desnaturalización del ácido nucleico bicatenario producido en la etapa (B) para formar ácido nucleico monocatenario.

30 La desnaturalización se lleva a cabo en general a de 80 a 100 °C, la hibridación (emparejamiento) se lleva a cabo en general a de 40 a 80 °C y la extensión se lleva a cabo en general a de 50 a 80 °C. Un ciclo típico es desnaturalización: aproximadamente 94 °C durante aproximadamente 1 min, hibridación: aproximadamente 58 °C durante aproximadamente 2 min y extensión: aproximadamente 72 °C durante aproximadamente 1 min. El protocolo exacto depende de factores tales como la longitud y secuencia de los cebadores y la secuencia diana, y la enzima usada.

35 La PCR se ha adoptado para diversas aplicaciones. Por ejemplo, el documento GB 2293238 describe métodos para reducir los cebadores no específicos y amplificar secuencias de ácido nucleico. Se desvelan "cebadores" (u oligonucleótidos) de bloqueo que producen alineamiento erróneo y reducen los cebadores no específicos creando reacciones de emparejamiento de cebadores competitivas a los cebadores de amplificación. Por ejemplo, se usan mezclas de cebadores de bloqueo aleatorios que comprenden un ddNTP en la posición 3' para prevenir el inicio de reacciones de extensión. Solamente los cebadores de amplificación correctos desplazan sus cebadores de bloqueo y pueden iniciar la reacción de amplificación.

45 Se han establecido métodos para usar cebadores de bloqueo que se unen específicamente con moléculas oligonucleotídicas diana no deseadas en una muestra para prevenir la amplificación de las mismas en una reacción de PCR de moléculas oligonucleotídicas desbloqueadas. Los oligonucleótidos desbloqueados pueden amplificarse sin medidas adicionales para asegurar la especificidad diana - en ausencia de moléculas oligonucleotídicas competitivas amplificables que no se pretende amplificar (documentos US 2002/0076767 A1 y US 6.391.592 B1; WO 99/61661).

50 Se desvela un método similar en el documento WO 02/086155, en el que oligonucleótidos de bloqueo unidos con un molde no deseado dan como resultado terminación prematura de una reacción de elongación. Los oligonucleótidos de bloqueo se unen específicamente con un molde en una mezcla dejando otros moldes libres para amplificación.

55 El documento US 5.849.497 describe el uso de oligonucleótidos de bloqueo durante un método de PCR con una ADN polimerasa que carece de actividad 5' exonucleasa. Esta ADN polimerasa no puede digerir los oligonucleótidos de bloqueo que previenen la amplificación. Dicho sistema se ha seleccionado para evitar el uso de PNA (ácidos nucleicos peptídicos) como oligonucleótidos de bloqueo. Se describe un sistema similar en el documento WO 2009/019008 que contempla sin embargo el uso de PNA y LNA, entre otros, como oligonucleótidos de bloqueo.

60 Todos estos métodos tienen en común que la amplificación de moldes no deseados se suprime específicamente por hibridación de un oligonucleótido de bloqueo específico.

65 En la solicitud de patente WO 98/02449 A1 (documento US 6.090.552) se describe un método de amplificación de ADN de "triamplicación". Se basa en el uso de un cebador en horquilla que se extiende y se liga con un bloqueador. Tanto el cebador como el bloqueador se unen con una cadena de ADN molde. El segundo cebador se une con la

cadena de ADN complementaria. El bloqueador y un cebador son parcialmente complementarios, conteniendo el cebador un donante y conteniendo el bloqueador un resto aceptor para TERF (transferencia de energía por resonancia de fluorescencia). Una extensión del cebador y un ligamiento de un producto de extensión con el bloqueador conduce a una reducción de la fluorescencia, debido a que ya no están en proximidad estrecha en un híbrido de cebador y bloqueador. Este método de triamplificación está limitado al uso de ADN molde y no está relacionado con métodos de ARN.

El documento WO 94/17210 A1 se refiere a un método de PCT usando múltiples cebadores para cadena tanto antisentido como con sentido de un ADN diana.

El documento WO 93/13216 A1 describe un método para introducir mutaciones puntuales en un ADN diana que comprende ligar cebadores con la mutación puntual con una cadena diana y extender los cebadores.

Li *et al.* [35] describen un método similar al documento WO 93/13216 A1, sin embargo, usando múltiples cebadores para introducir múltiples mutaciones puntuales en un producto de amplificación cíclica.

Seyfang *et al.* [1] describen el uso de múltiples oligonucleótidos fosforilados para introducir mutaciones en una cadena de ADN. Se usa la ADN polimerasa T4, que carece de ninguna actividad de desplazamiento de cadena detectable o actividad exonucleasa 5'-3', lo que es inadecuado para moldes de ARN.

Hogrefe *et al.* [2] describen la generación de bibliotecas de aminoácidos aleatorios con el kit de mutagénesis dirigida a múltiples sitios QuikChange. Se usan cebadores específicos que contienen 3 nucleótidos degradados en el centro complementarios de un ADN diana monocatenario conocido. El kit descrito usa ADN polimerasa *PfuTurbo* que es inadecuada para moldes de ARN.

El análisis de ARN comienza regularmente con transcripción inversa de ARN a ADNc ya que el ADN es más estable que el ARN y existen muchos métodos para analizar ADN. Sea cual sea el protocolo usado para analizar el ADNc, es importante que el ADNc generado durante la transcripción inversa (RT) representa el ARN que es necesario analizar en secuencia y concentración tan estrechamente como sea posible.

Se lleva a cabo en general transcripción inversa usando transcriptasas inversas. Estas enzimas requieren un cebador oligonucleotídico que hibrida con el ARN para iniciar (cebar) la polimerización dependiente de molde del ADNc. Las dos estrategias de cebado más comunes usadas son cebado con oligo dT y cebado aleatorio.

El cebado con oligo dT se usa para RT de ARNm que tienen una cola de poli A en su extremo 3'. El oligo dT ceba el ARN en el extremo 3' y la transcriptasa inversa copia el ARNm hasta su extremo 5'. Un inconveniente de este enfoque es que es necesario ARNm de alta calidad ya que cualquier degradación de ARNm conducirá a una fuerte sobrerrepresentación de los extremos 3' de ARNm.

Incluso si se usa ARNm no degradado, las moléculas de ADNc aún pueden truncarse debido a acontecimientos de detención de la polimerización prematura. Una causa frecuente es la formación de estructura secundaria y terciaria en regiones de ARN altamente estructuradas. Especialmente cuando el contenido de GC es alto, la transcriptasa inversa podría no leer estas regiones completas y por lo tanto el ADNc se trunca. La probabilidad de que se produzcan dichos acontecimientos aumenta cuanto más largo sea el ARNm que es necesario copiar. Por lo tanto el ADN cebado con oligo dT puede mostrar un fuerte sesgo hacia sobrerrepresentación de los extremos 3' de ARN. Por lo tanto, el cebado del extremo 3' adolece de un sesgo de concentración que conduce a un aumento de secuencias en o cerca del extremo 3' con representación gradualmente reducida de secuencias en la dirección del extremo 5' (véase Fig. 15, triángulos, para medición por qPCR del sesgo). Esto es problemático en enfoques cuantitativos, por ejemplo en la determinación del grado de expresión de un gen particular, en análisis de diferencias o en realización de perfiles de expresión completos de una célula.

Se han desarrollado enfoques para superar la terminación de la estructura secundaria de ARN especialmente cuando es necesario transcribir de forma inversa ARNm largos a ADNc de longitud completa. Uno de dichos métodos implica por ejemplo una mezcla de 2 transcriptasas inversas, una altamente procesiva tal como MMLV o AMV y mutantes de las mismas, incubando en primer lugar la mezcla de reacción a un intervalo de temperaturas normal para permitir la síntesis de primera cadena más el uso de una composición de enzima termoestable que tiene actividad transcriptasa inversa e incubando después la mezcla de reacción a una temperatura que inhibe la presencia de estructuras de ARNm secundarias para generar una primera cadena (patente de los Estados Unidos US 6.406.891). Sin embargo, los tampones para transcriptasas inversas contienen altas concentraciones de  $MgCl_2$  (3-10 mM) o  $Mn^{2+}$  (por ejemplo para ADN polimerasa Tth) y el ARN es altamente inestable y susceptible a roturas y/o degradación a mayores temperaturas especialmente en presencia de estos cationes divalentes. El método de ciclación entre dos temperaturas para evitar estructuras secundarias podría conducir también a cebado aleatorio por fragmentos de ARN cortos que se generaron durante altas temperaturas. Dichos fragmentos de ARN cortos serán usados por MMLV-H u otras transcriptasas inversas víricas como un cebador [3]. De nuevo esto conduciría a un sesgo en el ADNc sintetizado.

Otro enfoque es cebado aleatorio que tiene la ventaja de hibridar en múltiples localizaciones a lo largo del ARN y por lo tanto también bloquear la participación de esas secuencias en la formación de estructura secundaria. En el cebado aleatorio se usa una población de oligonucleótidos de secuencia aleatoria, habitualmente un hexámero aleatorio, para cebar la RT en cualquier parte dentro de la cadena de ácido nucleico molde. El cebado aleatorio se usa para ambas, transcripción inversa o transcripción regular usando ADN como molde. Cuando se analizó el ADN producto se descubrió que el cebado aleatorio no da como resultado eficacias iguales de transcripción inversa para todas las dianas en la muestra [4, 5]. Además no hay correlación lineal entre la cantidad de aporte de ácido nucleico molde y producción de ADN producto cuando se miden dianas específicas [4, 5]. De hecho, se ha mostrado que el uso de cebadores aleatorios puede conducir a sobreestimar algunos números de copias del molde hasta 19 veces en comparación con moldes cebados de forma específica de secuencia [6]. Aunque se ha especulado mucho acerca de las causas subyacentes para estos fenómenos, no se ha expuesto ninguna explicación concluyente.

### Objetivo

Los presentes inventores han observado que hay un sesgo general en dichas bibliotecas de ADN con cebadores aleatorios, en la forma en que las partes de secuencia en los extremos 5' de moléculas de ARN están sobrerrepresentadas en comparación con las partes en los extremos 3'. La razón de este fenómeno se encuentra en la combinación de cebado aleatorio y la fuerte actividad de desplazamiento de cadena de las transcriptasas inversas. Ya que el ARN tiene un alto grado de estructura secundaria, las transcriptasas inversas tuvieron que desarrollar una fuerte actividad de desplazamiento de cadena para superar esta estructura secundaria y generar eficazmente ADNc. Dada la fuerte actividad de desplazamiento de cadena de la transcriptasa inversa, el lado 5' de cualquier ARN estará representado varias veces en una biblioteca de ADNc cuando se usen oligonucleótidos aleatorios para cebado. Se produce un efecto similar durante la extensión de combinaciones de cebadores (por ejemplo aleatorios) que tienen cebadores que se emparejan con una posición más 3' en el ARN molde (cebadores cadena arriba en la dirección de los productos de extensión), en el que la transcriptasa inversa desplazará los productos de extensión de todos los cebadores que han hibridado con una posición más 5' en el ARN molde (cebadores cadena abajo en la dirección de los productos de extensión). Por lo tanto el cebado aleatorio es un método de síntesis de ADN con un fuerte sesgo para sobrerrepresentar los extremos 5' de un ácido nucleico molde dado (véase también Fig. 1 para una representación esquemática del problema y Fig. 15 para mediciones por qPCR del sesgo). Además de usar oligómeros aleatorios (por ejemplo: hexámeros aleatorios) para preparación de bibliotecas de ADNc y en el radiomarcado de sondas de ADN [7, 8], también se usan para detectar polimorfismos de un único nucleótido (SNP) así como acontecimientos de cromosomas a pequeña escala, principalmente inserciones o supresiones [5, 6]. Se ha desarrollado hibridación genómica comparativa (CGH) para dilucidar la variación del número de copias (CNV) de secuencias de todo el genoma entre diferentes genomas, tal como la amplificación o supresión diferencial de regiones genéticas entre ADN tumoral y ADN normal de tejido no afectado adyacente [9, 10].

En la actualidad uno de los métodos de análisis más completos para bibliotecas de ADN es la secuenciación de siguiente generación (NGS) [para revisión véase 11]. NGS es un término genérico para secuenciación en paralelo mediante polimerización de una manera de alto rendimiento. La NGS se basa en la obtención de lecturas de secuenciación de fragmentos pequeños. En la generación de bibliotecas de ADNc, el ARNm se fragmenta antes de la síntesis de ADNc o se fragmenta ADNc monocatenario o bicatenario. Sin embargo, cualquier fragmentación de ácidos nucleicos molde (química o física) introduce un sesgo indefinido e imprevisible y agotará el molde. En NGS, la secuencia completa se obtiene por alineamiento de esas lecturas lo que es una tarea difícil debido al gran número de lecturas pequeñas que hay que ensamblar en una secuencia completa. Hasta la fecha muchas lecturas proporcionan solamente información limitada. Por ejemplo muchas de las lecturas no pueden asignarse de forma única y por lo tanto se descartan. La generación de secuencias está obstaculizada adicionalmente por el sesgo de representación de fragmentos de secuencia.

Por lo tanto existe la necesidad de métodos mejorados para amplificar ácidos nucleicos molde que produzcan menos sesgo en la cantidad amplificada, por ejemplo para NGS o para la generación de bibliotecas de ADN para mejorar la representación de la concentración de secuencia del molde original.

### Sumario de la invención

Por lo tanto la presente invención proporciona un método para generar una biblioteca de partes de ácido nucleico amplificadas de un ácido nucleico molde, que comprende  
A) obtener dicho ácido nucleico molde, que es ARN, emparejar al menos un cebador oligonucleotídico con dicho ARN molde, emparejar al menos un finalizador oligonucleotídico y/o cebador adicional con dicho ARN molde, preferentemente en el que el finalizador oligonucleotídico se hibrida con al menos un cebador oligonucleotídico adicional por un par de marcadores de secuencia complementarios en el finalizador y cebador, en el que uno cualquiera de los finalizadores oligonucleotídicos comprende un marcador de secuencia o se hibrida con un marcador de secuencia, y uno cualquiera de los cebadores oligonucleotídicos comprende un marcador de secuencia o se hibrida con un marcador de secuencia, y en el que el marcador de secuencia no se empareja con el molde, elongar el al menos un cebador oligonucleotídico de una manera específica de molde hasta que el ácido nucleico

producto de elongación alcanza la posición de un finalizador oligonucleotídico o cebador adicional emparejado, por lo que se detiene la reacción de elongación, en el que en dicha reacción de elongación dicho finalizador oligonucleotídico opcional no se elonga y/o dicho cebador oligonucleotídico adicional se elonga de una manera específica de molde,

5 en el que el ácido nucleico producto elongado se liga con (i) el extremo 5' de dicho finalizador oligonucleotídico que se hibrida con la cadena molde o (ii) el extremo 5' de dicho cebador adicional que se hibrida con la cadena molde o con (iii) un marcador de secuencia que se hibrida con dicho finalizador o cebador adicional en las cercanías del extremo 3' del producto elongado proporcionando el marcador de secuencia hibridado a una región complementaria del finalizador oligonucleotídico o cebador adicional que está unido al extremo 5' de la región hibridada con el molde

10 del finalizador o cebador adicional; o

B) obtener dicho ácido nucleico molde, que es preferentemente ARN o ADN,

emparejar al menos un cebador oligonucleotídico con dicho ácido nucleico molde,

15 emparejar al menos un finalizador oligonucleotídico con dicho ácido nucleico molde, opcionalmente en el que el finalizador oligonucleotídico se hibrida con al menos un cebador oligonucleotídico adicional por un par de marcadores de secuencia complementarios en el finalizador y cebador, en el que uno cualquiera de los finalizadores oligonucleotídicos comprende un marcador de secuencia o se hibrida con un marcador de secuencia, y uno cualquiera de los cebadores oligonucleotídicos comprende un marcador de secuencia o se hibrida con un marcador de secuencia, y en el que el marcador de secuencia no se empareja con el molde,

20 elongar el al menos un cebador oligonucleotídico de una manera específica de molde hasta que el ácido nucleico producto de elongación alcanza la posición de un finalizador oligonucleotídico, por lo que se detiene la reacción de elongación, en el que en dicha reacción de elongación dicho finalizador oligonucleotídico no se elonga, y

25 en el que el ácido nucleico producto elongado se liga con (i) el extremo 5' de dicho finalizador oligonucleotídico que se hibrida con la cadena molde o con (ii) un marcador de secuencia que se hibrida con dicho finalizador en las cercanías del extremo 3' del producto elongado proporcionando el marcador de secuencia hibridado a una región complementaria del finalizador oligonucleotídico que está unido al extremo 5' de la región hibridada con el molde del finalizador; o

C) obtener dicho ácido nucleico molde, que es preferentemente ARN o ADN,

30 emparejar un primer cebador oligonucleotídico con dicho ácido nucleico molde,

emparejar al menos un cebador oligonucleotídico adicional con dicho ácido nucleico molde,

en el que uno cualquiera de los cebadores oligonucleotídicos y uno cualquiera de los cebadores oligonucleotídicos adicionales comprende un marcador de secuencia o se hibrida con un marcador de secuencia, y en el que el marcador de secuencia no se empareja con el molde,

35 elongar dicho primer cebador oligonucleotídico de una manera específica de molde hasta que el ácido nucleico producto de elongación alcanza la posición de uno de dichos cebadores oligonucleotídicos adicionales, por lo que se detiene la reacción de elongación, y se elonga al menos un cebador oligonucleotídico adicional de una manera específica de molde, en el que el ácido nucleico producto elongado detenido se liga con (i) el extremo 5' de dicho cebador oligonucleotídico adicional que se hibrida con la cadena molde o con (ii) un marcador de secuencia que se hibrida con dicho cebador en las cercanías del extremo 3' del producto elongado proporcionando el marcador de secuencia hibridado a una región complementaria del cebador adicional que está unido al extremo 5' de la región hibridada con el molde del cebador adicional;

45 en el que en A, B y C el producto de elongación se selecciona o amplifica usando un cebador que hibrida con el marcador de secuencia en el finalizador o cebador adicional.

En el método C) dicho cebador oligonucleotídico adicional actúa tanto como un finalizador, que previene la elongación adicional de una reacción de amplificación que alcanza la posición del finalizador emparejado, como un cebador en sí mismo, es decir como un iniciador de la elongación.

50 El ácido nucleico molde puede comprender o consistir sustancialmente en ARN o ADN, en realizaciones preferidas dicho molde es ARN.

La presente invención proporciona los métodos de la invención para generar una biblioteca de secuencias de uno o más ácidos nucleicos molde que comprenden una mezcla de partes de ácido nucleico amplificadas, preferentemente solapantes, de dichos ácidos nucleicos molde. Una biblioteca de secuencias es una colección de fragmentos de ADN que pueden almacenarse y copiarse mediante cualquier proceso conocido en la técnica. Por ejemplo puede obtenerse una biblioteca de secuencias mediante el proceso de clonación molecular. Las bibliotecas de secuencias pueden amplificarse también mediante por ejemplo PCR usando secuencias universales en los extremos de los fragmentos de ADN.

60 La invención también se refiere a un kit A) para generar una biblioteca de partes de ácido nucleico amplificadas de un ácido nucleico molde según la invención o para generar una biblioteca de secuencias según la invención que comprende una ADN polimerasa, cebadores oligonucleotídicos aleatorios que comprenden una modificación que aumenta la T<sub>m</sub> y finalizadores oligonucleotídicos aleatorios que son inadecuados para extensión de nucleótidos y comprenden una modificación que aumenta la T<sub>m</sub>, y una ligasa, opcionalmente además uno o más de tampones de reacción que comprenden Mn<sup>2+</sup> o Mg<sup>2+</sup>, un agente aglutinante, tal como PEG, en el que uno cualquiera de los

finalizadores oligonucleotídicos comprende un marcador de secuencia y uno cualquiera de los cebadores oligonucleotídicos comprende un marcador de secuencia; o

B) para generar una biblioteca de partes de ácido nucleico amplificadas de un ácido nucleico molde según la invención o para generar una biblioteca de secuencias según la invención que comprende una ADN polimerasa y una ligasa, cebadores oligonucleotídicos, finalizadores oligonucleotídicos, en el que uno cualquiera de los finalizadores oligonucleotídicos comprende un marcador de secuencia y uno cualquiera de los cebadores oligonucleotídicos comprende un marcador de secuencia, preferentemente un agente aglutinante, especialmente PEG; o

C) para generar una biblioteca de partes de ácido nucleico amplificadas de un ácido nucleico molde que contiene a) cebadores oligonucleotídicos aleatorios que comprenden una modificación que aumenta la  $T_m$ , b) finalizadores oligonucleotídicos aleatorios que son inadecuados para extensión de nucleótidos y comprenden una modificación que aumenta la  $T_m$ , y c) una ligasa, preferentemente que comprende además uno o más de tampones de reacción que comprenden  $Mn^{2+}$  o  $Mg^{2+}$ , un agente aglutinante, tal como PEG y/o preferentemente que comprende además una ADN polimerasa y/o una ligasa, en el que uno cualquiera de los finalizadores oligonucleotídicos comprende un marcador de secuencia y uno cualquiera de los cebadores oligonucleotídicos comprende un marcador de secuencia.

La siguiente divulgación detallada se refiere a todos los aspectos y realizaciones de la presente invención. El alcance de la presente invención se define en las reivindicaciones adjuntas.

## 20 Divulgación detallada de la invención

La presente invención proporciona un método para generar una biblioteca de ácidos nucleicos amplificados o amplificar ácidos nucleicos. Esta generación puede estar relacionada también con una única reacción de amplificación, por ejemplo un ciclo de transcripción, o más. Incluye la generación de ARN o ADN mediante polimerización dependiente de ARN o ADN. Por lo tanto, la amplificación de ácidos nucleicos incluyó polimerización de nucleótidos de ARN basándose en un ácido nucleico molde de ARN o ADN o polimerización de nucleótidos de ADN basándose en un ácido nucleico de molde de ARN o ADN. Preferentemente el método incluye una etapa o un ciclo de transcripción inversa, polimerización de ADN dependiente de ARN.

Los métodos de la invención incluyen el uso de al menos dos oligonucleótidos cortos que hibridan con el ácido nucleico molde. Al menos un oligonucleótido tiene una función de cebador, es decir puede actuar como iniciador de polimerización de nucleótidos para amplificación dependiente de polimerasa, es decir transcripción. La extensión de cebadores mediante la adición de nucleótidos de una manera dependiente de molde se denomina en el presente documento elongación o extensión. Los productos de dichas reacciones se denominan productos de elongación o productos de extensión. Las ARN o ADN polimerasas añaden nucleótidos a una cadena oligonucleotídica dada que forma pares de bases con una base nucleotídica de una cadena molde. La hibridación y el emparejamiento se entienden como formación de pares de bases de nucleótidos complementarios. Los nucleótidos o bases complementarios son los capaces de formar pares de bases tales como A y T (o U); G y C; G y U.

Al menos un oligonucleótido adicional tiene una función de finalizador. Esto significa que a medida que una reacción de elongación (extensión) que se ha iniciado en un cebador cadena arriba (en la dirección de los productos de extensión) alcanza un oligonucleótido cadena abajo (en la dirección de los productos de extensión) con una función de finalizador, se previene la elongación adicional de dicha reacción de elongación. En relación con la posición de nucleótidos en el ácido nucleico molde esto significa que una vez que una reacción de elongación que se ha iniciado de un cebador que se ha emparejado más hacia el extremo 3' del ácido nucleico molde en relación con el oligonucleótido con la función de finalizador, alcanza ese oligonucleótido (finalizador), se previene la elongación adicional de dicha reacción de elongación. La reacción de elongación puede detenerse por hibridación fuerte del oligonucleótido con función de finalizador con los ácidos nucleicos molde de modo que no sea desplazado por la polimerasa.

El oligonucleótido con función de finalizador también puede ser un cebador. Puede hibridar cadena abajo (en la dirección de la reacción de elongación, cadena arriba en la dirección del molde) con un primer cebador y detener la reacción de elongación de dicho primer cebador. A su vez (o simultáneamente) actúa como iniciador de la elongación en sí mismo para producir un producto de transcripción - que a su vez también puede detenerse en la posición de un oligonucleótido cadena abajo adicional (en relación con la dirección de las reacciones de elongación) con función de finalizador.

"Cadena arriba" se refiere a la dirección hacia el extremo 5' (3'-5') de un ácido nucleico u oligonucleótido dado. "Cadena abajo" se refiere a la dirección hacia el extremo 3' (5'-3') de un ácido nucleico u oligonucleótido dado. Ya que los oligonucleótidos hibridan de manera inversa cadena abajo para un cebador se refiere a la dirección cadena arriba de un ácido nucleico molde hibridado. Esto significa que un oligonucleótido (u oligo, o cebador o finalizador o bloqueador) cadena abajo es un oligo que hibridó con una parte más cadena arriba de un ácido nucleico molde en relación con un oligonucleótido (u oligo, o cebador o finalizador o bloqueador) cadena arriba que se ha hibridado con una parte más cadena abajo de un ácido nucleico molde. Por lo tanto, cuando se usa la expresión "oligonucleótido cadena abajo" u "oligonucleótido cadena arriba" la direccionalidad siempre se refiere a la del producto o los productos de extensión, excepto cuando se indica otra cosa. Las reacciones de elongación dependientes de

polimerasa de la invención están en la dirección 5'-3', cadena abajo.

Como se usa en el presente documento, se entenderá que "que comprende" se refiere a una definición abierta, que permite miembros adicionales de elementos similares u otros. Se entenderá "que consiste en" como una definición cerrada relacionada con una serie limitada de elementos.

Como se usa en el presente documento "cebador" también puede referirse a un cebadores oligonucleotídico. "Finalizador" también se refiere a finalizador oligonucleotídico. "Oligo" se usa tanto para cebadores oligonucleotídicos como para finalizadores oligonucleotídicos.

Un finalizador oligonucleotídico es un oligonucleótido que puede detener una reacción de elongación como se ha descrito anteriormente y no inicia una reacción de elongación adicional, por ejemplo el oligonucleótido es incapaz de aceptar (unirse covalentemente con) un nucleótido adicional en su posición 3'. Dichos nucleótidos se conocen en la técnica y habitualmente carecen de un 3' OH, como por ejemplo en ddNTS (didesoxinucleótidos).

Por lo tanto, la presente invención proporciona esencialmente dos métodos, uno que utiliza finalizadores oligonucleotídicos y otro cebadores oligonucleotídicos adicionales. Aparte de esta diferencia, los métodos comprenden la similitud de la invención para proporcionar productos de amplificación limitados y bien definidos que pueden amplificarse de manera controlada por pocillo y, lo que es más importante, con una distribución de concentración unitaria sobre la longitud del molde de ácido nucleico. Los productos de amplificación de la invención pueden caracterizarse adicionalmente y usarse. Son los productos deseados que se obtienen y no se descartan, como en métodos de supresión de moldes. La presente invención genera productos de ácido nucleico amplificados que representan mejor las moléculas molde que es necesario analizar y las prepara para integración ininterrumpida con métodos de análisis posteriores tales como secuenciación de siguiente generación o como biblioteca de secuencias. Si los finalizadores son al mismo tiempo cebadores, es posible proporcionar una secuencia continua basándose solo en una molécula molde. Esta secuencia continua del producto amplificado puede proporcionarse como una molécula individual cuando las secuencias de productos individuales están conectadas de forma covalente, por ejemplo ligadas. Dicha reacción de conexión o ligamiento puede realizarse mientras está hibridada a la cadena molde para asegurar que los productos están conectados en el mismo orden que la cadena molde (siendo por supuesto al mismo tiempo la cadena de complemento inverso).

Las ventajas de la presente invención son más prominentes con ácidos nucleicos molde largos que se amplifican según la invención. Dichos moldes pueden ser de al menos 100 bases, al menos 1000 bases (1kb), al menos 2 kb, al menos 4 kb, al menos 6 kb, al menos 10 kb, al menos 20 kb, al menos 30 kb, al menos 40 kb, al menos 50 kb, de longitud.

En el caso de una transcripción inversa una realización de la presente invención se refiere a un método de transcripción inversa de una molécula de ARN que comprende hibridar al menos dos cebadores con una molécula de ARN molde y extender cebadores utilizando una ADN polimerasa dependiente de ARN, en el que el producto de extensión de un cebador cadena arriba (cadena abajo en la dirección del molde) no desplaza el producto de extensión de un cebador cadena abajo (cadena arriba en la dirección del molde). Esta realización en general comprende esencialmente una reacción de extensión de cebadores usando un primer cebador que se extiende pero se detiene en (no desplaza) un segundo cebador que también se extiende.

Sin embargo, para una representación completa de un molde dado por el producto de elongación, no se requiere necesariamente conexión directa de los nucleótidos elongados. Una muestra de ácidos nucleicos molde contiene habitualmente muchas moléculas molde que tienen la misma secuencia. Mediante el uso de muchas, más de una, combinaciones de cebador y finalizador es posible tener una representación completa de dicha secuencia por los muchos productos de elongación. Tener muchos productos de elongación cortos es con frecuencia un requisito de una biblioteca de ácidos nucleicos que represente completamente el molde. Por lo tanto la presente invención proporciona el método de preparación de una biblioteca de ácidos nucleicos proporcionando los productos de elongación. La biblioteca puede contener los productos de elongación en una mezcla. Preferentemente los productos de elongación, especialmente de moldes de la misma secuencia, contienen partes de secuencias solapantes. Las partes de secuencias solapantes son más fáciles para ensamblaje de secuencia completo, como se requiere por ejemplo en métodos de NGS y son convenientes para proporcionar una biblioteca que puede usarse para clonar cualquier secuencia específica sin o con interrupciones limitadas alcanzando los límites del tamaño de un ácido nucleico elongado individual.

En el caso de una transcripción inversa una realización tal de la presente invención se refiere a un método de transcripción inversa de una molécula de ARN que comprende hibridar al menos dos oligonucleótidos, uno un cebador y el otro un finalizador, con una molécula de ARN molde y extender el cebador que utiliza una ADN polimerasa dependiente de ARN, en el que el producto de extensión del cebador cadena arriba (cadena abajo en la dirección del molde) se detiene en (no desplaza) la posición del finalizador cadena abajo (cadena arriba en la dirección del molde). Dicho finalizador no se extiende. Esta realización en general comprende esencialmente una reacción de extensión de cebadores usando un primer cebador que se extiende pero se detiene en (no desplaza) un finalizador oligonucleotídico que no se extiende.

Es posible combinar las realizaciones de la invención, por ejemplo usando al menos dos cebadores y un finalizador, un primer cebador cadena arriba que se extiende cadena abajo, deteniéndose dicha reacción de extensión en la posición de un segundo cebador cadena abajo que también se extiende cadena abajo, deteniéndose dicha reacción de extensión a su vez en la posición del finalizador oligonucleotídico.

5 En realizaciones especiales no se usan finalizadores oligonucleotídicos que no puedan extenderse. En dichas realizaciones solamente pueden usarse cebadores extensibles durante la amplificación/transcripción.

10 En otras realizaciones se desea obtener productos de amplificación que se han detenido en finalizadores oligonucleotídicos. Para este fin un experto en la materia puede por ejemplo seleccionar dichos ácidos nucleicos elongados por métodos bien conocidos, tal como por marcaje, incluyendo inmovilización en una fase sólida (por ejemplo perlas o una superficie sólida) o uniendo una secuencia de código de barras o marcador de secuencia. Los productos elongados también pueden ligarse al finalizador oligonucleotídico - similar al ligamiento con cebadores con función de finalizador para proporcionar una molécula de producto largo como se ha mencionado anteriormente.

15 Pueden combinarse el marcaje y el ligamiento con el finalizador oligonucleotídico, por ejemplo marcando el finalizador oligonucleotídico y ligando el producto de extensión con un finalizador oligonucleotídico marcado. Por supuesto, el marcaje del finalizador oligonucleotídico puede realizarse después de ligamiento con el producto de extensión. Dicho marcaje permite la manipulación, selección y/o amplificación fácil de los productos elongados. Por ejemplo un marcador de secuencia puede usarse para amplificación seleccionada adicional de dichos productos de elongación elongados y de este modo marcados usando cebadores que hibridan con dicho marcador y puede iniciar una reacción de amplificación, por ejemplo PCR, de dicho producto de ácido nucleico (previamente) elongado). Esto es particularmente ventajoso si se obtiene una multitud de productos de elongación diferentes y muchos se marcan con el mismo marcador de secuencia. En el caso de bibliotecas de secuencias este método permite una amplificación fácil de una biblioteca completa - aún con una representación uniforme de la cantidad de todas las

20 secuencias sobre la longitud completa del molde original.

25

En realizaciones preferidas se usa más de un cebador oligonucleotídico, que actúa de forma similar al primer cebador pero tiene una secuencia de cebador diferente, la secuencia que se hibrida con un molde diana. En particular el método de la invención según esta realización comprende además emparejar un cebador oligonucleotídico adicional con dicho ácido nucleico molde y elongar dicho cebador oligonucleotídico adicional hasta que el ácido nucleico producto de elongación alcanza la posición de otro cebador oligonucleotídico o un finalizador oligonucleotídico. En realizaciones especialmente preferidas se usan al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 15, al menos 20, al menos 30, al menos 40, al menos 50 o más cebadores oligonucleotídicos diferentes. Se entiende que "cebadores oligonucleotídicos diferentes" difieren en una secuencia de cebador pero, por supuesto, pueden compartir otras partes de secuencia similares, tales como marcadores de secuencia. Dichos marcadores de secuencia se usan preferentemente en una reacción de amplificación de los productos elongados usando cebadores que se hibridan con los marcadores de secuencia. Por lo tanto, con un cebador todos los productos potencialmente diferentes pueden amplificarse. En una realización especial los cebadores oligonucleotídicos son cebadores aleatorios. Debe entenderse "cebadores aleatorios" como una mezcla de diferentes cebadores con partes de secuencia de cebador diferentes, con una alta variación debido a una síntesis aleatoria de al menos una parte de la secuencia de cebador. Los cebadores aleatorios abarcan potencialmente el área combinatoria completa para dicha secuencia. La parte de cebador de secuencia aleatoria del cebador aleatorio puede abarcar 1, 2, 3, 4, 5, 6, 7, 8 o más nucleótidos que se seleccionan de forma aleatoria de A, G, C o T (U). Con respecto a secuencias de hibridación de secuencias de cebador se usan T y U indistintamente en el presente documento. El área posible combinatoria completa para una parte de secuencia aleatoria es  $m^n$ , en el que m es el número de tipos de nucleótidos usados (preferentemente los cuatro de A, G, C, T(U) y n es el número de los nucleótidos aleatorios. Por lo tanto un hexámero aleatorio, en el que está representada cada secuencia posible, consiste en  $4^6 = 4096$  secuencias diferentes.

30

35

40

45

50 De forma similar también es posible usar más de un finalizador oligonucleotídico en uno cualquiera de los métodos de la invención. Dichos finalizadores adicionales actúan de forma similar al primer finalizador, pero difieren en la secuencia alineada con el molde. Lo mismo que se ha descrito anteriormente para cebadores adicionales se aplica para el finalizador adicional - por supuesto con la diferencia de que los finalizadores no son adecuados para reacciones de elongación. Por lo tanto, por coherencia la región del finalizador oligonucleotídico que hibrida con el molde también se denomina "secuencia de cebador". Dicha secuencia de cebador puede ser de 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 o más nucleótidos de longitud. La invención proporciona por lo tanto un método como se ha definido anteriormente que comprende emparejar un finalizador oligonucleotídico adicional con dicho ácido nucleico molde y en el que en dicha reacción de elongación dicho finalizador oligonucleotídico adicional no se elonga. En realizaciones preferidas se usan al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 15, al menos 20, al menos 30, al menos 40, al menos 50 o más finalizadores oligonucleotídicos diferentes. Los finalizadores oligonucleotídicos pueden ser finalizadores oligonucleotídicos aleatorios, que comprenden una secuencia de cebador aleatoria que se empareja con el molde. Como se ha descrito anteriormente para cebadores oligonucleotídicos, además los finalizadores oligonucleotídicos pueden compartir partes de secuencias similares, tales como marcadores o códigos de barras que pueden usarse para amplificación o identificación de los productos. Dichos tipos de marcadores permiten la identificación fácil en una biblioteca de secuencias que comprende los productos de amplificación de la invención.

55

60

65

Como se ha descrito en la introducción una biblioteca de ADNc con cebadores (aleatorios) sin control de la reacción de elongación (y su terminación) distorsiona la representación de ARN real de partes de secuencias. Incluso por ejemplo en moldes de ARNm cortos (200-1000 nt) pueden producirse múltiples acontecimientos de cebado y con el desplazamiento de cadena el lado 5' de moléculas de ARN estará presente en más copias que el lado 3' (véase también Fig. 1). Por lo tanto cuando se mide la concentración de un determinado transcrito génico se obtendrán valores diferentes cuando se explore con respecto a partes de secuencias en el extremo 5' o 3'. Se usan habitualmente sondas aleatorias cortas en micromatriz y análisis de cuantitativo PCR. Esto conduce a distorsiones graves de la concentración medida. Además, esta distorsión será mayor al comparar transcritos largos y cortos, ya que los extremos 5' de transcritos largos muy abundantes estarán incluso más representados que los extremos 5' de transcritos más cortos. Cuando se analiza globalmente la concentración de transcritos tal como en secuenciación de alto rendimiento (por ejemplo secuenciación de siguiente generación) además de distorsión de las concentraciones, la detección de transcritos cortos poco habituales será incluso menos probable que la detección de transcritos largos poco habituales. Ya que la expresión diferencial de transcritos génicos y sus variantes de corte y empalme es una parte importante del análisis de fenotipos, es importante que cada parte de secuencia de un transcrito esté representada en la biblioteca de ADNc generada en la abundancia correcta.

Estos problemas son resueltos por la presente invención. La aplicación de los métodos de la invención para proporcionar transcritos bien definidos, finalizados en una posición de cebador o finalizador (preferentemente con desplazamiento de cadena de la transcriptasa inversa inhibido) durante la generación de ADNc de una multitud de moléculas de ARN como por ejemplo en la generación de una biblioteca de ADNc de moléculas de ARNm, permite la representación igual de cada parte de las moléculas de ARN. Por ejemplo el ARNm puede cebarse con cebadores aleatorios tales como hexámeros aleatorios que se modifican para asegurar que el cebador aleatorio no sea desplazado por la transcriptasa inversa. Puede usarse cualquier cebador aleatorio que pueda iniciar la transcripción inversa desde múltiples sitios de las moléculas de ARN molde.

La presente invención también proporciona métodos que permiten la unión covalente (por ejemplo: mediante ligamiento) de los productos elongados obtenidos como se proporcionan en un híbrido con el molde (véase también Fig. 2c). Esto permite la generación de una molécula de ácido nucleico amplificada de longitud completa.

Usando cebadores que proporcionan un fosfato 5', se ha descubierto que los ácidos nucleicos elongados "cortos" - que son esencialmente fragmentos de la cadena complementaria para la cadena molde - pueden estar ligados estando al mismo tiempo hibridados con el molde. Esto dará como resultado moléculas de ácido nucleico amplificadas largas y en la mayoría de los casos de longitud completa que son una representación directa del molde. Ser capaz de conservar la información de la secuencia de longitud completa de un molde es, por ejemplo, importante cuando es necesario analizar variantes de corte y empalme de un gen. Especialmente cuando el corte y empalme de genes multiexónicos sea complejo el análisis de puntos de unión de corte y empalme individuales solo no producirá información inequívoca para la secuencia de longitud completa de la variante de transcrito implicada. Sin embargo, solamente el cebado con oligo dT en el caso de ARNm o cebado de un enlazador universal ligado a 3' de cualquier molde de ácido nucleico conducirá a una infrarrepresentación del lado 5' del molde. En otras palabras, cuanto más largo sea el molde menos probable es que la molécula se transcriba de forma inversa hasta su longitud completa. Como se puede ver en la Fig. 15, cuando se usa una RT cebada con oligo dT convencional en promedio después de 6 kb solamente ~10 % de las moléculas de ARN se transcriben de forma inversa hasta su longitud completa. Ya que la presente invención proporciona - en una realización - métodos que comenzarán la polimerización desde dos o más posiciones de una molécula molde con una reacción de elongación que se detiene en la posición de un cebador cadena abajo, el producto de extensión del cebador cadena arriba puede ligarse con el extremo 5' del cebador cadena abajo. Uniendo covalentemente ambos productos de extensión cuando estén en híbrido con el molde se crea una secuencia más larga. En continuación cuando se usan muchos cebadores diferentes todos los moldes presentes en la muestra pueden transcribirse y ligando los productos de extensión cortos puede crearse una copia amplificada de longitud completa del molde (véase también Fig. 2c, d).

Asimismo, según las realizaciones de la invención que utilizan finalizadores oligonucleotídicos, puede realizarse un ligamiento con el producto elongado para obtener ácidos nucleicos amplificados bien definidos de una amplificación de una secuencia entre el cebador y el finalizador.

En realizaciones preferidas los cebadores oligonucleotídicos y/o finalizadores oligonucleotídicos se fosforilan o adenilan en el extremo 5'. Esta medida ayuda a ligar fácilmente los cebadores o finalizadores oligonucleotídicos con otro ácido nucleico, tal como el producto de elongación, en una o varias etapas usando una ligasa. En algunas realizaciones, especialmente cuando se usan mezclas de muchos cebadores y finalizadores, es posible proporcionar solamente los finalizadores con dicha modificación para asegurar el ligamiento de ácidos nucleicos elongados con finalizadores y prevenir el ligamiento con otros cebadores.

Se pueden usar cualquier ligasa conocida en la técnica, tal como ARN ligasa T4, ADN ligasa T4, ARN ligasa T4 2, ADN ligasa Taq y ligasa de *E. coli*.

Para prevenir que cualquiera de los cebadores o finalizadores no hibridados se ligen una realización preferida de la invención usa una ligasa específica de doble cadena tal como ARN ligasa T4 2 o ADN ligasa T4. El ADN bicatenario

es el sustrato natural de ADN ligasa T4 y los híbridos ADN-ARN son malos sustratos [29]. Para superar esta ineficacia puede reemplazarse  $Mg^{2+}$  con  $Mn^{2+}$  como un ion divalente para la enzima [30]. En una realización de la presente invención se muestra que la adición de PEG a la reacción de ligamiento aumenta la eficacia de ligamiento de moléculas de ADN en un híbrido de ARN incluso en un tampón de ligasa que contenga  $Mg^{2+}$ .

Opcionalmente una ARN ligasa T4, que es deficiente en adenilación (por ejemplo truncada), puede usarse para ligamiento que se basa en la presencia de fragmentos de ADN adenilados que están en el híbrido con el ARN (se produce adenilación con ADN ligasa T4 exclusivamente en ácidos nucleicos bicatenarios). Puede añadirse adicionalmente a la reacción de ligamiento para aumentar más la eficacia. El ligamiento se produce exclusivamente en el híbrido y se consideró previamente una forma muy ineficaz de ligamiento [31]. En el documento US 6.368.801 se describe un ligamiento de 2 moléculas de ADN (con ribonucleótidos en sus extremos 3' y 5') en un híbrido de ARN por ARN ligasa T4. Sin embargo, la ARN ligasa T4 no es específica para híbridos ya que se liga con todos los ácidos nucleicos monocatenarios que contienen desoxirribonucleótidos en su extremo 5' o 3' (ARN a ADN, ARN a ARN, ADN a ADN siempre que haya un desoxirribonucleótido en el extremo 3' y 5') también. La presente invención incluye la adición de PEG en la reacción de ligamiento, que permite que tenga lugar el ligamiento en el híbrido de ARN. Se ha usado PEG en reacciones de ligamiento monocatenario como un agente de aglutinamiento molecular, aumentando la probabilidad de que el complejo de ligasa y oligonucleótido donante interacciones con el aceptor (3' OH) reduciendo el volumen reactivo eficaz [32]. Aquí por el contrario el complejo de ligasa y oligonucleótido donante ya está junto al aceptor y PEG actúa para cambiar la conformación del híbrido ARN:ADN condensando la doble hélice en una conformación que es más reminiscente de una hélice de ADN:ADN, de modo que la ADN ligasa T4 que normalmente es específica para ligar dos moléculas de ADN en un híbrido con una cadena de ADN, puede ligar dos moléculas de ADN que están en un híbrido con una molécula de ARN. Como alternativa puede usarse la ARN ligasa T4 2, que es una ligasa específica de doble cadena.

Podrían añadirse otros aditivos tales como Tween-20, NP-40 adicionalmente o en lugar de PEG para ligamiento eficaz en el híbrido de ARN. Dentro del alcance de la invención la reacción de ligamiento requiere PEG-8000 12 %-25 % final (v/v). Puede usarse una diversidad de pesos moleculares y compuestos de PEG, y el experimentador experto apreciará que la identidad y concentración del aditivo puede variarse para optimizar los resultados. En el contexto de la presente invención, una cantidad eficaz de PEG es una cantidad suficiente para permitir la actividad de ligamiento en un híbrido de ARN. En una reacción de RT de 20  $\mu$ l se descubrió que la concentración de PEG óptima para un ligamiento en un híbrido de ARN era del 20 %. Sin embargo, es evidente para el experto en la materia que opcionalmente el volumen de reacción y la concentración de PEG pueden aumentarse o reducirse (por ejemplo reducirse su volumen, aumentarse la cantidad o concentración de PEG) para optimizar adicionalmente de forma potencial la eficacia de ligamiento de 2 moléculas de ADN en un híbrido de ARN por ADN ligasa T4 o ARN ligasa T4 2. Otros aditivos tales como HCC 1 mM y/o pirofosfatasa pueden aumentar adicionalmente la eficacia de ligamiento en un híbrido de ARN.

Pueden insertarse oligonucleótidos preadenilados en la reacción y usarse con ARN ligasa T4, siempre que se haya retirado cualquier oligonucleótido no hibridado antes de la reacción de ligamiento. Como se ha mencionado anteriormente en otra realización de la presente invención puede añadirse ARN ligasa T4 2 además de ADN ligasa T4 para reforzar adicionalmente el ligamiento de fragmentos de ADN en un híbrido de ARN.

Por lo tanto se prefiere que la ligasa sea una ligasa específica de doble cadena tal como ADN ligasa T4 o ARN ligasa T4 2 y en la que se prefiere usar polietilenglicol a una concentración entre 12 % y 25 %.

Aparte de una síntesis de amplificado de molde representativo un beneficio adicional de la presente invención es la eficacia mejorada de generación de ácidos nucleicos de producto largo, especialmente síntesis de ADNc, que da como resultado mayor sensibilidad de detección y productos más largos (véase ejemplo 5 y 6).

En realizaciones preferidas uno cualquiera de los cebadores oligonucleotídicos puede comprender un marcador de secuencia. Dicho marcador de secuencia es una etiqueta en forma de una secuencia de ácido nucleico preseleccionada única que puede usarse para detectar, reconocer o amplificar una secuencia marcada con dicho marcador. Preferentemente un marcador de secuencia uniforme se une a más de uno de los cebadores oligonucleotídicos, especialmente de forma preferente a todos los cebadores oligonucleotídicos. Se previene preferentemente que dicho marcador de secuencia se empareje con el molde, por ejemplo hibridando con un ácido nucleico complementario. Pueden unirse marcadores de secuencia al extremo 5' del finalizador oligonucleotídico - para no prevenir la reacción de elongación del cebador en su extremo 3'.

En realizaciones preferidas uno cualquiera de los finalizadores oligonucleotídicos puede comprender un marcador de secuencia. Dicho marcador de secuencia es una etiqueta en forma de una secuencia de ácido nucleico preseleccionada única que puede usarse para detectar, reconocer o amplificar una secuencia marcada con dicho marcador. Preferentemente un marcador de secuencia uniforme se une a más de uno de los finalizadores oligonucleotídicos, especialmente de forma preferente a todos los finalizadores oligonucleotídicos. Se previene preferentemente que dicho marcador de secuencia se empareje con el molde, por ejemplo hibridando con un ácido nucleico complementario. El marcador de secuencia puede unirse al extremo 3' del finalizador oligonucleotídico. En esta posición el marcador no obstaculiza el contacto del ácido nucleico producto de elongación con el extremo 5' del

finalizador - dando como resultado productos bien definidos. También permite el ligamiento del producto elongado con el finalizador oligonucleotídico. Como alternativa el marcador puede estar en el extremo 5' de dicho finalizador, sin embargo, se evita su hibridación con el molde de modo que la reacción de elongación todavía alcance el extremo 5' de la región cebadora del finalizador. Dicho marcador comprende preferentemente un extremo 5' libre de modo que el producto elongado pueda ligarse con el marcador para marcaje del producto elongado por dicho marcador. Un extremo 5' libre del marcador que puede ligarse fácilmente con el producto se proporciona preferentemente en las cercanías del extremo 3' del producto elongado. Esto puede conseguirse proporcionando por ejemplo el marcador hibridado con una región complementaria del finalizador oligonucleotídico, estando dicha región complementaria hibridada con el marcador y unida con el extremo 5' de la región cebadora del finalizador oligonucleotídico (para un ejemplo véase Fig. 4).

La etapa de marcaje de la invención del producto elongado cuando alcanza la posición del finalizador oligonucleotídico (u otro cebador), preferentemente un finalizador, puede conseguirse con cualquier medio conocido fácilmente disponible en la técnica. Dicho medio incluye unir un cromóforo, fluoróforo o separación de fase sencilla mediante unión con una fase sólida y lavado de dicha fase sólida para retirar todos los ácidos nucleicos no unidos, aislando de este modo el ácido nucleico marcado. En realizaciones preferidas dicha etapa de marcado comprende ligamiento con un marcador de secuencia. Puede unirse un marcador de secuencia con dichos cebadores oligonucleotídicos o finalizadores oligonucleotídicos. El marcaje también puede comprender ligamiento con dichos cebadores oligonucleotídicos o finalizadores oligonucleotídicos, que comprenden un marcador de secuencia.

Para muchos análisis corriente abajo se prefiere que oligonucleótidos, tales como cebadores, finalizadores, bloqueadores o enlazadores se agoten. Especialmente, cuando se desea o es necesaria una amplificación de la biblioteca utilizando las secuencias enlazadoras universales es preferible agotar los enlazadores no ligados.

Esto puede conseguirse por ejemplo mediante una exclusión por tamaños, conservando los oligonucleótidos (más cortos) en un lecho apropiado y recuperando la biblioteca (más largos). Otra posibilidad es unir la biblioteca más larga con un vehículo basado en sílice sin conservar los oligos más cortos. Dicha diferenciación también puede llevarse a cabo para distinguir entre el estado monocatenario del oligo y el estado bicatenario de la biblioteca cuando todavía está en un híbrido con el molde. Otra posibilidad para purificación basada en longitud son métodos basados en precipitación con PEG. Cuanto mayor sea la concentración de PEG más cortos serán los ácidos nucleicos que pueden precipitarse. Por ejemplo, se descubrió que usando una precipitación de PEG de 12,5 %, todos los fragmentos pequeños (por debajo de 60 nt) permanecerán en la solución y solamente se precipitarán el ADNc y la biblioteca ligada a enlazadores.

En una realización preferida se usa un enfoque de limpieza basado en perlas. Están disponibles en el mercado perlas acopladas a oligo dT o perlas de estreptavidina para aislar cebadores de oligodT marcados con biotina. Una limpieza basada en perlas tiene la ventaja adicional de que pueden realizarse tanto RT como ligamiento en las perlas. Después de hibridar el ARNm con el oligo dT (marcado con biotina o en perlas) y con el iniciador y los finalizadores, cualquier exceso de iniciador y finalizador no hibridado puede retirarse por lavado antes de comenzar la reacción de RT. Por lo tanto en realizaciones preferidas de la invención el ácido nucleico molde se inmoviliza en una fase sólida o un soporte sólido, preferentemente perlas. Los ácidos nucleicos amplificados hibridados con dicho ácido nucleico molde pueden después lavarse.

Otra realización de la invención es que pueden realizarse RT y ligamiento en una etapa de reacción, simplemente añadiendo ligasa (por ejemplo: ADN ligasa T4 o ARN ligasa T4 2), PEG 10 % y ATP 0,4  $\mu$ M a la reacción de transcripción inversa en un tampón de reacción de transcripción inversa regular. Aunque puede usarse incubación de 30 min, se obtuvo un mejor rendimiento usando 2 h de incubación a 37 °C. Después de otra etapa de lavado (4 lavados) el ARN puede después hidrolizarse para obtener la biblioteca de ADNc con dos marcadores, que después puede insertarse en una reacción de PCR. También pueden realizarse RT y ligamiento en dos etapas de reacción sucesivas, volviendo a tamponar la reacción simplemente mediante el lavado de las perlas. Sin embargo, en una realización preferida de la invención se realiza una reacción simultánea de RT/ligamiento ya que esto dio como resultado un rendimiento similar al protocolo de dos etapas, pero tiene la ventaja de tiempos reducidos de manipulación e incubación. Pueden realizarse RT y ligamiento simultáneos mediante la adición de una ADN polimerasa y ligasa en una mezcla de reacción con el ácido nucleico molde.

En métodos preferidos de la invención dichos productos elongados se amplifican. La amplificación puede comprender usar cebadores específicos de marcador para amplificar productos elongados que comprenden un marcador 5' y/o 3', partiendo del cebador y/o finalizador oligonucleotídico marcado con marcador, respectivamente. La amplificación es preferentemente por PCR. Los marcadores de secuencia adecuados para hibridación de cebadores en un ciclo de amplificación siguiente también se denominan en el presente documento "enlazadores".

En la actualidad cualquier preparación de fragmentos de ADNc pequeños para secuenciación de alto rendimiento tal como NGS implica fragmentación de ARN y en la mayoría de los casos un procedimiento multietapa para introducir los marcadores enlazadores 5' y 3' para amplificación e identificación por código de barras. Por ejemplo el kit de preparación de bibliotecas de ARNm-Sec ScriptSeq™ de Epicentre usa tecnología de marcadores terminales (documento US 2009/0227009 A1) y cebadores aleatorios en ARN fragmentado químicamente (con ARN ribosómico

agotado). Sin embargo, cualquier fragmentación de ARNm (químico o físico) introduce un sesgo indefinido e imprevisible. Además durante los protocolos de fragmentación una parte del ARN se degrada o se pierde. Por lo tanto la presente invención está idealmente adaptada para generar bibliotecas de ADNc de un tamaño corto más bien definido sin necesidad de fragmentación de ARN.

Además cuando el ARN se fragmenta se crean muchos extremos 5' adicionales. Las transcriptasas inversas tienen tendencia a añadir varios nucleótidos no de molde cuando alcanzan el extremo 5' del ARN molde y usan estos nucleótidos para cebar la síntesis de segunda cadena. Por lo tanto cuando se fragmenta ARN se generan más extremos 5' y por lo tanto se inicia más síntesis de segunda cadena. Una cuestión importante durante la secuenciación de ARN es la cuestión de qué cadena de ADN se transcribió un ARN. Especialmente en el análisis de transcripción con sentido y antisentido se requiere una alta conservación catenaria (conservación de información de cadena) de la biblioteca secuenciada. Por lo tanto ya que no es necesaria digestión de ARN en la presente invención puede conseguirse un grado mucho mayor de conservación catenaria en la biblioteca de ADNc generada (véase también ejemplo 9) en comparación con métodos de preparación de bibliotecas que incluyen fragmentación de ARN.

Por supuesto este método puede emplearse para la generación de fragmentos para cualquier tipo de ácido nucleico molde y no se limita a ARN. Dichos fragmentos son proporcionados por los productos elongados de la invención o partes de ácido nucleico amplificados. La secuencia de nucleótidos molde podría ser también ADN. Por lo tanto también puede iniciarse una preparación de bibliotecas usando las técnicas descritas en la presente invención a partir de ADN genómico o productos de PCR. Ya que las transcriptasas inversas también aceptan moldes de ADN [33], todas las etapas pueden realizarse básicamente como se describe en la presente invención. Opcionalmente también pueden usarse polimerasas dependientes de ADN si se usa ADN como molde.

Como para muchos métodos de análisis tales como NGS, es preferible haber definido secuencias enlazadoras universales presentes en el extremo 3' y/o 5' del ADNc. Dichas secuencias enlazadoras pueden por lo tanto actuar como sitios de cebadores para amplificación por PCR para enriquecer con respecto a una biblioteca o sitios de cebadores para amplificación de enlace en una superficie sólida o para cebar una reacción de secuenciación. Por lo tanto está dentro del alcance de la invención que secuencias enlazadoras se ligen con las partes de ácido nucleico amplificadas. Sin embargo se prefiere que la secuencia enlazadora 5' se introduzca directamente con los cebadores (véase también Fig. 3 (L1)). Aquí la secuencia enlazadora 5' es una extensión 5' a la secuencia que se usa para cebar la polimerasa.

Como alternativa o además una secuencia enlazadora puede introducirse en el extremo 3' del producto de ácido nucleico elongado, por ejemplo usando un oligo finalizador (Fig. 3. (S1, Sm)) que tiene una secuencia enlazadora añadida en su extremo 3' (Fig. 3. (L2)). El extremo 5' del oligo finalizador puede ligarse con el extremo 3' del producto de extensión (Fig. 3b). La reacción de ligamiento específico asegura que el extremo 5' del oligo finalizador no tenga desplazamiento de cadena. Por lo tanto en una realización preferida se añade un oligonucleótido finalizador a la reacción de transcripción inversa y en el que oligonucleótido finalizador tiene una extensión de secuencia enlazadora 3'.

En una versión alternativa - como se ilustra en las Figuras 5 y 8 - el oligonucleótido iniciador y finalizador se hibridan al menos parcialmente entre sí. El oligonucleótido finalizador fosforilado 5' puede aún hibridar con la cadena molde (como se muestra en la Fig. 8a-d,f-h) o no tener ninguna hibridación con la cadena molde (véase Fig. 8e). En este caso el oligo extendido denominado iniciador tiene que ser el mismo oligo que detiene el desplazamiento de cadena y por lo tanto contienen preferentemente modificaciones que detienen el desplazamiento de cadena. La cadena extendida de un iniciador más cadena arriba se detendrá en el siguiente iniciador y el oligo fosforilado denominado previamente el finalizador que se hibrida con el iniciador se ligarán con la cadena de ADNc extendida.

En 8e se muestra una combinación de iniciador/finalizador más elaborado. Los detalles en la secuencia enlazadora L1 y L2 pueden encontrarse en la descripción de la figura 9.

Finalmente el marcador de secuencia ("enlazador") del oligo iniciador y finalizador puede unirse (véase fig. 8h) para introducir marcadores de secuencia en el ADNc, pero mantiene la secuencia de los productos de extensión individuales en orden ya que ahora están unidos covalentemente entre sí mediante sus marcadores de secuencia iniciadores y finalizadores.

Se prefiere que se use una polimerasa durante elongación que tiene actividad transferasa terminal baja o no tiene, para no añadir nucleótidos sin moldes al extremo 3' del producto de extensión tras alcanzar la posición del extremo 5' del oligo finalizador, ya que un saliente 3' reduciría la especificidad y eficacia de la reacción de ligamiento. Son transcriptasas inversas con baja actividad transferasa terminal por ejemplo Superscript III (Invitrogen); son RT sin actividad transferasa terminal por ejemplo AMV-RT.

Como alternativa o además la actividad transferasa terminal y por lo tanto también síntesis de segunda cadena puede inhibirse mediante la adición de actinomicina D. Puede añadirse actinomicina D a la reacción de polimerización en cantidades suficientes para evitar síntesis de segunda cadena y/o reducir el desplazamiento de cadena de la polimerasa en comparación con ausencia de adición de actinomicina. En una realización preferida se

añade actinomicina D a la reacción de RT a una concentración final de aproximadamente 50 µg/ml, también pueden usarse concentraciones mayores o menores, tales como por ejemplo de 5 µg/ml a 200 µg/ml.

Además o como alternativa puede digerirse un saliente por una nucleasa específica de cadena individual, preferentemente una exonucleasa 3'-5', tal como pero sin limitación exonucleasa I (digestión de ADNmc 3'-5'), Exo T5 (digestión de ADNmc o bc 3'-5').

Uno de los objetivos de la invención es controlar cualquier sesgo introducido en la biblioteca de secuencias obtenida por los productos de ácido nucleico amplificados, y esto significa en la mayoría de los casos minimizar el sesgo. Las secuencias enlazadoras que se introducen como una extensión para secuencias cebadoras aleatorias o semialeatorias pueden participar también en la hibridación con el molde. Esto añadirá un sesgo a la biblioteca generada. Se prefiere por lo tanto que se inhiba la participación de al menos los nucleótidos del enlazador que queda junto al cebador o la secuencia finalizadora en la hibridación con el molde. Esto puede conseguirse mediante diferentes medios. Por ejemplo puede añadirse a la reacción un oligonucleótido con una secuencia del complemento inversa para la secuencia enlazadora (véase también Fig. 5c-f; Fig. 6c-f). En ese caso el complemento inverso competirá con el molde por la hibridación con la secuencia enlazadora y usando exceso de complemento inverso puede detenerse eficazmente la participación de la secuencia enlazadora en el cebado de la RT o hibridación del oligo de terminación.

Se prefiere proporcionar el complemento inverso con modificaciones de nucleótidos que potencia la estabilidad de híbrido de enlazador:complemento inverso, usando por ejemplo LNA. Sin embargo, puede usarse cualquier otra modificación que potencie la energía de unión (véase también Fig. 6e; Fig. 7e).

Además, se prefiere añadir el cebador y/o finalizador como un adaptador previamente preparado a la transcripción inversa. Esto significa que esencialmente todas las secuencias enlazadoras están en un híbrido con sus cadenas de complemento inverso y así se inhibe la participación de la secuencia enlazadora en la reacción de hibridación.

Por lo tanto, en una realización preferida se añade a la reacción una secuencia de complemento inverso para la parte de secuencia de los oligonucleótidos, preferentemente ya en un híbrido con el cebador oligonucleotídico y se prefiere además que la T<sub>m</sub> del oligo de complemento inverso aumente mediante por ejemplo la introducción de modificaciones tales como LNA, 2' fluoronucleótidos o PNA.

En una realización preferida la secuencia de complemento inverso se une covalentemente con la secuencia enlazadora (véase Fig. 6g; 7g). Esto puede estar en continuación directa con la secuencia enlazadora o mediante una horquilla de nucleótidos o espaciadores tales como C3, C6, C12 o cualquier otro resto. Además este resto puede ser una modificación que potencia la formación de adaptadores. Como se ha descrito antes se prefiere que los nucleótidos del complemento inverso se modifiquen para potenciar la hibridación con la secuencia enlazadora. Por lo tanto se prefiere que el complemento inverso para la secuencia enlazadora se conecte directamente con la secuencia enlazadora o mediante una horquilla o un espaciador de nucleótidos tal como C3, C6, C12 o cualquier otro resto.

En una realización más preferida un marcador de secuencia en el cebador (L1) y un marcador de secuencia en el finalizador (L2) comprenden secuencias al menos parcialmente complementarias que permiten al iniciador y al finalizador formar híbridos (véase Fig. 5 y 8). De esa manera las secuencias de marcadores de secuencia ("enlazador") no hibridarán con la cadena molde y como un beneficio añadido se producirá un ligamiento de finalizador en proximidad inmediata con el siguiente acontecimiento de inicio, minimizando cualquier hueco entre los acontecimiento de inicio y finalización. Por lo tanto; en realizaciones preferidas de la invención al menos uno o más o todos los finalizadores oligonucleotídicos se hibridan con al menos un cebador oligonucleotídico adicional. En una realización alternativa o realización adicional en combinación, al menos uno o más, preferentemente todos, cebadores oligonucleotídicos hibridan con un finalizador oligonucleotídico. Se prefiere especialmente que uno cualquiera de los finalizadores oligonucleotídicos y uno cualquiera de los cebadores oligonucleotídicos comprenda un marcador de secuencia cada uno, y preferentemente en el que el marcador de secuencia de los cebadores oligonucleotídicos es al menos parcialmente complementario del marcador de secuencia del finalizador oligonucleotídico permitiendo de este modo la hibridación de los finalizadores oligonucleotídicos y los cebadores oligonucleotídicos entre sí al menos en parte del marcador de secuencia respectivo.

Ya que cualquier OH 3' libre puede actuar potencialmente como un aceptor durante la polimerización o el ligamiento, se prefiere que cualquier OH 3' libre (excepto el que esté en el cebador oligonucleotídico) esté bloqueado (véase también Fig. 6f; Fig. 7f). Se conocen en la técnica muchos grupos de bloqueo. Se proporcionan por referencia pero no como limitación didesoxinucleótidos, C-espaciadores y grupos fosfato. Además, el extremo 3' del complemento inverso en el adaptador de cebado también puede estar provisto de un saliente (véase Fig. 6d). De forma correspondiente, la secuencia enlazadora del adaptador de finalización puede estar provisto de un saliente 3' (véase Fig. 7d). Por lo tanto se prefiere que el OH 3' de los oligonucleótidos que no participan en la reacción de extensión de cebadores se bloqueen y/o cuando se proporcionen en un híbrido que el extremo 3' tenga un saliente sobre el extremo 5'.

Quando se introducen las secuencias enlazadoras junto con los oligonucleótidos cebador y/o finalizador la secuencia entre las secuencias enlazadoras refleja la secuencia de ARN molde. Ya que es mucho más probable una hibridación errónea del cebador o el finalizador con el molde que la incorporación de un nucleótido erróneo durante la polimerización, es mucho más probable que contengan un error la secuencia del cebador y/o secuencia del finalizador que la secuencia polimerizada. Por lo tanto, cuando por ejemplo se secuencian la biblioteca en un experimento de NGS, sería preferible que la secuencia enlazadora que contiene el cebador de secuenciación esté junto al producto de extensión de cebadores. Una solución a este problema se muestra en la Fig. 4. Aquí la secuencia L2 está en un híbrido con su complemento inverso que se introducido en el lado 5' del oligonucleótido finalizador. De esta manera la secuencia L2 puede ligarse con el producto de extensión de cebadores una vez que la polimerasa alcanza y se detiene en el oligo de finalización. Sin embargo, la secuencia de la secuencia de finalización del oligo no está incluida en la biblioteca. Por lo tanto, el inicio de cualquier reacción de secuenciación a partir de la secuencia L2 no incluirá una secuencia finalizadora con hibridación potencialmente errónea. Por lo tanto, se prefiere que la secuencia enlazadora esté en el extremo 5' del oligonucleótido de finalización y el extremo 5' de un complemento inverso de la secuencia enlazadora se ligue con el extremo 3' de un producto de extensión finalizado por desplazamiento de cadena como una secuencia marcadora. El mismo principio se aplica para la combinación de iniciador-"finalizador" mostrada en la Fig 8e con la única diferencia de que el oligo responsable de la finalización es de hecho el iniciador del siguiente fragmento de biblioteca.

Por lo tanto en realizaciones preferidas el cebador oligonucleotídico o finalizador oligonucleotídico se hibrida con un marcador de secuencia como etiqueta oligonucleotídica. Se prefiere especialmente que dicho marcador de secuencia hibride preferentemente con una parte del extremo 5' de dicho cebador oligonucleotídico o finalizador oligonucleotídico. Los siguientes, por ejemplo 1, 2, 3, 4, 5, 6 o más, nucleótidos de dicho cebador o finalizador en dirección 3' a continuación de los nucleótidos de dicho cebador o finalizador que se hibridan con dicho marcador de secuencia se hibridan con el molde. Esto permite situar el extremo 5' del marcador cerca del extremo 3' de un producto de elongación de otro cebador, que se localiza cadena arriba del cebador oligonucleotídico o finalizador oligonucleotídico con el marcador de secuencia hibridado de modo que el extremo 3' elongado de dicho cebador adicional puede ligarse con el extremo 5' de dicho marcador de secuencia. Dichos marcadores también pueden usarse para reacciones de amplificación posteriores de los productos ligados y también se denominan "enlazadores" en el presente documento.

En realizaciones preferidas el cebador oligonucleotídico y/o finalizador oligonucleotídico comprende una modificación de nucleótidos que aumenta la Tm o que refuerza la cadena principal de fosfato del azúcar de dicho oligonucleótido. Estas modificaciones para aumentar la Tm son para fortalecer la hibridación con el molde para asegurar la finalización de la reacción de elongación y evitar el desplazamiento del cebador o finalizador. Dichas modificaciones se conocen en la técnica a partir de bloqueadores oligonucleotídicos, tal como se describe en los documentos GB 2293238, US 2002/0076767 A1, EP 6.391.592 B1, WO 99/61661, WO 02/086155, US 5.849.497, WO 2009/019008. Las modificaciones adecuadas incluyen una o más de las modificaciones seleccionadas de nucleósidos 2' fluoro, LNA (ácido nucleico bloqueado), ZNA (ácidos nucleicos zip), PNA (ácido nucleico peptídico). Además la Tm puede aumentarse usando intercaladores o aditivos que se unen específicamente con ácidos nucleicos, tales como bromuro de etidio, Sybr Green. Los intercaladores preferidos son específicos para híbridos de ARN:ADN. El número de nucleótidos modificados puede variar, dependiendo de otras medidas tomadas para aumentar la Tm. Preferentemente se modifican 1, 2, 3, 4, 5 o 6 nucleótidos. Preferentemente los ácidos nucleicos modificados están en el lado 5' de la parte de secuencia de cebador, la parte del cebador oligonucleotídico o finalizador que pueden hibridar o emparejarse con el molde. Preferentemente se modifican 1, 2 o 3 nucleótidos 5'. Las ADN polimerasas pueden tener una actividad de desplazamiento de cadena intrínseca, especialmente transcriptasas inversas para desnaturalizar estructuras secundarias de ARN. Puede usarse una polimerasa que tiene actividad de desplazamiento de cadena de nucleótidos para la elongación.

Ya que ADN polimerasas, especialmente transcriptasas inversas, pueden desplazar un oligonucleótido de ADN de una cadena molde de ARN al menos tan bien como la disolución de la estructura secundaria o terciaria, la hibridación del oligonucleótido tiene que potenciarse para detener el desplazamiento de cadena de la transcriptasa inversa. Esto puede conseguirse usando modificaciones del oligonucleótido en sí mismo o usando aditivos que estabilizan la hibridación del oligonucleótido o que detienen la transcriptasa inversa. Son modificaciones de los oligonucleótidos que reducen o inhiben la actividad de desplazamiento de cadena de la transcriptasa inversa por ejemplo nucleósidos 2' fluoro [15], PNA [16, véase FIG. 2; 17], ZNA [18, 19], abrazaderas G (documento US 6.335.439, un análogo de citosina capaz de unirse en abrazadera con guanina) o LNA (documentos US 2003/0092905; US 7.084.125) [20, 21]. Estas modificaciones en general aumentan la temperatura de fusión del oligonucleótido, aumentando la energía de hibridación local del oligonucleótido con la cadena de ARN molde. Algunos también refuerzan la cadena principal de fosfato del azúcar los que inhibe adicionalmente desplazamiento de cadena por la transcriptasa inversa.

Como alternativa o además, la hibridación del oligo con el molde de ARN puede alterarse usando diferentes aditivos que se unen o intercalan con los ácidos nucleicos. Por ejemplo, pueden usarse bromuro de etidio, SybrGreen (documentos US 5.436.134; US 5.658.751; US 6.569.627) o acridina. Otros compuestos que pueden unirse con ANbc son actinomicina D y análogos [22]. Sin embargo potencialmente estabilizan también estructura de ARN secundario.

Por lo tanto, se prefiere que dichos intercaladores o aditivos se unan específicamente con híbridos de ARN:ADN. Son ejemplos aminoglucósidos de la familia de neomicina (neomicina, ribostamicina, paromomicina y frameticina [23]). Los aditivos que alteran las propiedades de hibridación del oligonucleótido también pueden incluirse de forma covalente en la estructura oligonucleotídica [23].

La energía y cinética de hibridación pueden cambiarse para inhibir el desplazamiento de cadena por la transcriptasa inversa mediante la adición de proteínas de unión a ácido nucleico tales como proteína de unión monocatenaria tal como Tth SSB [24] o Tth RecA [25].

Resultará evidente para los expertos en la materia que esos aditivos son solamente ejemplos y puede usarse cualquier otro compuesto, modificación de bases o enzima que conducen a un aumento de la hibridación del oligonucleótido con ARN para aumentar la  $T_m$  y de este modo inhibir el desplazamiento de cadena.

El aumento de la  $T_m$  debería ser suficientemente fuerte para evitar un desplazamiento de uno cualquiera de los nucleótidos del extremo 5' de la región del cebador emparejada con el molde por una polimerasa de elongación. En particular, el aumento de la  $T_m$  de la invención evita el desplazamiento del 3<sup>er</sup>, 2<sup>o</sup> y/o 1<sup>er</sup> nucleótido cadena abajo del extremo 5' de la región de cebador.

En determinadas realizaciones de la invención es necesario detener el desplazamiento de la cadena justo en el primer nucleótido 5' del cebador cadena abajo, especialmente cuando los fragmentos de ADNc se ligan entre sí o con un enlazador como se describe posteriormente. El desplazamiento de cadena de un oligonucleótido se reduce o inhibe usando modificaciones de nucleótidos que aumentan la  $T_m$  o refuerzan la cadena principal de fosfato de azúcar del oligonucleótido, incluyendo, por ejemplo, nucleósidos 2' fluoro LNA, PNA, ZNA o PNA o usando intercaladores o aditivos que se unen específicamente con ácidos nucleicos tales como bromuro de etidio, Sybr gold, SybrGreen, preferentemente intercaladores que son específicos para híbridos de ARN:ADN.

Por lo tanto se prefiere que la unión de los cebadores oligonucleotídicos tenga  $T_m$  potenciada específicamente en sus extremos 5' para evitar que la polimerasa de elongación los desplace. Dichas modificaciones incluyen, pero sin limitación, LNA, PNA, ZNA, acridina o fluoróforos.

Los oligonucleótidos con una  $T_m$  aumentada en su extremo 5' tales como oligonucleótidos modificados por LNA permiten una terminación al comienzo del siguiente cebador. Está dentro del alcance de la invención combinar la terminación de desplazamiento de cadena usando los oligos modificados por LNA junto con los mutantes deficientes en síntesis de desplazamiento tales como M-MLV Y64A o VIH F61W o cualquier otra transcriptasa inversa con síntesis de desplazamiento alterada así como reduciendo la temperatura de reacción y usando diferentes aditivos para aumentar la unión de oligonucleótidos con el ARN.

Preferentemente se modifican nucleótidos C y/o G. Incluso sin modificar estos nucleótidos tienen una  $T_m$  mayor que A o T debido a la mayor formación de enlaces de hidrógeno cuando se emparejan de forma complementaria. En realizaciones preferidas el cebador oligonucleotídico y/o finalizador oligonucleotídico comprende al menos uno, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6 nucleótidos modificados que se seleccionan de G o C. Estos nucleótidos modificados están preferentemente en el extremo 5' de la secuencia del cebador como se ha mencionado anteriormente.

Se consigue detención del desplazamiento de cadena más eficaz mediante bases G o C a medida que aumentan la  $T_m$  local del cebador o finalizador. Por lo tanto cebadores o finalizadores semialeatorios (hexámeros, heptámeros, octámeros, nonámeros, etc.) contienen al menos dos, más preferentemente tres o más G o C o una combinación de G y C. Se prefiere más que estos G o C se modifiquen para aumentar la temperatura de fusión local, como sucede cuando se usan bases modificadas por LNA. Se prefiere más usar al menos 1, al menos 2 o al menos 3 bases modificadas por LNA en el extremo 5' del cebador. Por lo tanto, se prefiere usar al menos dos, al menos 3 nucleótidos modificados opcionalmente elegidos de G o C.

Existen varios métodos y medios para asegurar que se detenga la reacción de elongación cuando la reacción de elongación alcance la posición de un finalizador oligonucleotídico o cebador más o adicional emparejado con el molde. Esta detención también se denomina prevención del desplazamiento de cadena en el presente documento. El desplazamiento de cadena es un problema particular para transcripción inversa debido a actividades de desplazamiento de cadena aumentadas o altas de transcriptasas inversas. La etapa de la invención de inhibir la transcriptasa inversa para desplazar la cadena del ADNc de una parte de ARN ya copiada asegura que cualquier parte de una molécula de ARN que ya se haya copiado no se copie de nuevo. Por lo tanto, ninguna parte copiada del ARN está sobrerrepresentada en la biblioteca de ADNc sintetizada. Esta inhibición del desplazamiento de cadena puede conseguirse mediante diferentes medios, tal como reducir la temperatura de reacción, usar una transcriptasa inversa sin actividad de desplazamiento de cadena, aumentar la temperatura de fusión o la energía de hibridación del híbrido de ARN del cebador o aumentar la rigidez del ARN o cebador o estabilizar la hélice. En la práctica, habitualmente se selecciona una combinación de estos medios para conseguir condiciones de reacción óptimas sin desplazamiento de cadena. Un experto en la materia es bien capaz de seleccionar parámetros

adecuados como se describe en el presente documento o se conoce en la técnica para adaptarse a un molde y unas condiciones de reacción particulares.

Una opción es modificar la temperatura de reacción. En general, se favorece una temperatura de reacción por encima de 37 °C durante la RT para disolver mejor estructuras secundarias en el molde de ARN lo que conduce a una síntesis de desplazamiento más eficaz. En una realización se consigue detención del desplazamiento de cadena del cebador reduciendo la temperatura de reacción. Se usan temperaturas de reacción por debajo de 37 °C y hasta 4 °C para reducir el desplazamiento de cadena. Se prefiere que la polimerización durante la RT se lleve a cabo entre 20 °C y 37 °C. Sin embargo, incluso a estas temperaturas de reacción bajas la detención del desplazamiento de cadena no se completará cuando se usen transcriptasas inversas que tengan actividad de desplazamiento de cadena y/o se use un oligonucleótido finalizador sencillo que no tenga modificaciones que alteren su temperatura de fusión.

En una realización en lugar de o además de reducir la temperatura de reacción para conseguir una mejor detención de la elongación en dicha posición de un cebador adicional o un finalizador (y reducir el desplazamiento de cadena) pueden usarse transcriptasas inversas que son deficientes en desplazamiento de cadena tales como el mutante de M-MLV Y64A [12] o el mutante de VIH F61W (Phe-61-Trp) [13, 14]. Los mutantes deficientes en desplazamiento de cadena son capaces de desplazar el siguiente cebador o finalizador hasta 3 nt cuando no están modificados. Está dentro del alcance de la invención combinar la detención del desplazamiento de cadena reduciendo la temperatura de reacción con el uso de mutantes deficientes en síntesis de desplazamiento tales como M-MLV Y64A o VIH F61W o cualquier otra transcriptasa inversa con síntesis de desplazamiento alterada.

Un inconveniente de reducir la temperatura de reacción durante la RT o usar una transcriptasa inversa que tiene una actividad de desplazamiento de cadena reducida o alterada es que la transcriptasa inversa tendrá dificultades para leer regiones de estructura secundaria de ARN. Cuanto más estable sea la estructura secundaria menos probable será que esta parte del ARN se copie a ADNc. Esto significa que aunque ninguna parte de la molécula de ARN se copia más de una vez partes del ARN que forman una estructura secundaria no se copiarán. Esto significa que algunas partes del ARN estarán infrarrepresentadas en la biblioteca de ADNc producida.

Por lo tanto, en una realización preferida se usa una transcriptasa inversa que tiene actividad de desplazamiento de cadena y/o la transcripción inversa se lleva a cabo a temperaturas elevadas que disuelven lo suficiente la estructura de ARN secundaria. En estas condiciones cada parte del ARN es accesible a la transcriptasa inversa. Sin embargo ya que los híbridos de ARN:ARN son en general más estables que los híbridos de ARN:ADN, también la copia de ADNc puede de nuevo someterse a desplazamiento de cadena si no se usan modificaciones adicionales. Por lo tanto, en realizaciones preferidas la transcripción inversa se lleva a cabo en condiciones que no permiten la formación de estructuras secundarias o terciarias del molde de ARN (híbridos de ARN:ARN) o en condiciones que permitan el desplazamiento de cadena de estas estructuras secundarias por la transcriptasa inversa, mientras que al mismo tiempo el producto de extensión de cebadores (copia de ADNc) no puede desplazarse.

El aumento de la concentración de contraiones monovalentes también estabilizará el híbrido (pero también la estructura secundaria de ARN), ya que se ha indicado para RT de VIH que a KCl 75 mM la actividad de desplazamiento de cadena está alterada aunque no inhibida [26, 27]. En una realización preferida la concentración de iones positivos monovalentes se selecciona preferentemente de al menos 20 mM, al menos 30 mM, al menos 40 mM, al menos 50 mM, al menos 60 mM, al menos 70 mM. Pueden seleccionarse concentraciones similares de forma independiente para iones con carga negativa individual.

La transcriptasa inversa usada durante la reacción de elongación puede ser una transcriptasa inversa vírica y puede seleccionarse del grupo que consiste en RT de AMV (y mutantes de la misma tales como RT Thermoscript), RT de MMLV (y mutantes de la misma incluyendo, pero sin limitación, Superscript I, II o III, RT Maxima, RevertAid, RevertAid Premium, Omniscript, GoScript), RT de VIH, RT de VSR, RT de EIAV, RT de RAV2, ADN polimerasa Tth, ADN polimerasa de *C. hydrogeniformans*, virus de la leucosis del sarcoma aviar (ASLV) y mutantes de RNasa H del mismo. Pueden usarse mezclas de cualquiera de estas transcriptasas inversas. En particular, pueden usarse mezclas de enzimas RT víricas, tales como mezclas de MMLV y ASLV, y/o pueden usarse sus análogos de RNasa H reducida o RNasa H menos. En cualquiera de estos métodos y composiciones, pueden usarse una o más transcriptasas inversas, incluyendo cualquier transcriptasa inversa como se ha descrito anteriormente.

Está dentro del alcance de la invención combinar la detención del desplazamiento de cadena reduciendo la temperatura de reacción con el uso de mutantes deficientes en síntesis de desplazamiento tales como M-MLV Y64A o VIH F61W o cualquier otra transcriptasa inversa con síntesis de desplazamiento alterada y aumentar la T<sub>m</sub> del oligonucleótido para el ARN.

En cualquiera de estos métodos y composiciones también puede usarse una ADN polimerasa termoestable, aunque con respecto a estabilidad y cinética de hibridación de ARN de los cebadores esto no está recomendado en todas las situaciones.

Especialmente, pero sin limitación, para representación óptima en una biblioteca de ADNc con cebadores aleatorios

los cebadores aleatorios pueden estar presentes en una concentración de 50 nM a 100  $\mu$ M, y más preferentemente a aproximadamente 2,5  $\mu$ M pero también puede ser al menos 300 nM. En realizaciones preferidas la relación (p/p) de cebador con respecto a ácidos nucleicos molde está entre 5:1 y 1:1000, preferentemente entre 2:1 y 1:500, preferentemente entre 1:1 y 1:300, preferentemente entre 1:2 y 1:250, preferentemente entre 1:5 y 1:150, preferentemente entre 1:10 y 1:100, preferentemente entre 1:12 y 1:50. La relación molar de cebador con respecto a ácidos nucleicos molde puede estar entre 1000:1 y 5.000.000:1, preferentemente entre 5.000:1 y 1.000.000:1, entre 10.000:1 y 500.000:1 o entre 20.000:1 y 300.000:1. En un ejemplo, usando 150 ng de material de partida de ARNm y suponiendo una longitud de ARNm de 500-5000 nt esto significaría que se añaden cebadores añadidos a una concentración final de 2,5  $\mu$ M en un exceso molar de 1:280.000-1-28.000. En realizaciones preferidas la relación molar o (p/p) de cebadores con respecto a finalizadores está entre 2:1 y 1:10, preferentemente entre 1:1 y 1:5, especialmente aproximadamente 1:1.

Sin embargo es posible reducir la concentración de oligo por ejemplo: usando 34 ng de material de partida de ARNm y suponiendo una longitud de ARNm de 500-5000 nt esto significaría que se añaden cebadores añadidos a una concentración final de 300 nM en una relación molar de 1:33-1:3,3, estando el molde en un exceso molar. Una concentración de nucleótidos reducida preferida durante la reacción de polimerización puede ayudar a reducir el desplazamiento de cadena y falsa síntesis de segunda cadena de la polimerasa. En realizaciones preferidas la relación molar o (p/p) de cebadores con respecto a finalizadores está entre 2:1 y 1:10, preferentemente entre 1:1 y 1:5, especialmente aproximadamente 1:1.

Quando se transcribe ARNm de forma inversa puede añadirse un cebador de oligo dT para abarcar mejor también la cola de poli A del ARNm. Opcionalmente, la adición del cebador de oligo dT puede omitirse ya que es parte de la mezcla de cebadores aleatorios, aunque en una realización preferida para garantizar representación igual del ARNm debería añadirse. La longitud de oligo dT puede variar de 8 bases a 27 bases, pero preferentemente se usa un cebador de oligo dT de 25 nt. Otros tipos de cebadores con diferente composición pueden usarse en lugar de oligo dT. Los ejemplos de dichas composiciones incluyen, pero sin limitación, oligo dT donde la base 3' es A, o C, o G (dT anclado). Además, también pueden usarse otras secuencias o restos que pueden formar pares de bases con secuencias de poli A de ARNm. Un ejemplo, sin limitación, es desoxiuridina, (dU).

El oligo(dT) puede consistir esencialmente en entre aproximadamente 12 y aproximadamente 25 restos de dT, y puede ser una molécula de oligo(dT) anclada que contiene un nucleótido distinto de T terminal. El oligo(dT) puede ser oligo (dT) 18-25 nt de longitud o equivalentes anclados del mismo.

El oligo(dT) puede estar presente en una concentración de entre aproximadamente 20 nM y aproximadamente 1  $\mu$ M, se usan más preferentemente 500 nM. Los cebadores aleatorios pueden ser de entre 5 y 15 nucleótidos de longitud, y pueden ser hexámeros aleatorios con al menos 3 LNA o al menos tres bases modificadas con 2' fluoro en el sitio 5' para detener eficazmente el desplazamiento de cadena.

Una característica adicional es que dependiendo de la concentración del cebador aleatorio puede influirse en la longitud promedio de los ácidos nucleicos producto. Aumentando la concentración de los cebadores aleatorios puede reducirse el tamaño del producto elongado resultante. El tamaño de fragmento deseado depende de la aplicación posterior. Si las partes de ácido nucleico amplificadas deseadas debieran estar por debajo de 300 nt (como se desea en la actualidad para secuenciación de próxima generación) las concentraciones de cebadores aleatorios deberían variar de 125 nM a 350 nM de concentración final.

Según los métodos de la invención la concentración de oligo dT puede ser de 44 nM a 750 nM, por ejemplo, o valores intermedios. Es evidente para los expertos en la materia que pueden usarse diversas relaciones de cebadores aleatorios y oligo dT.

Quando se investiga ARNm, preferentemente deberían usarse muestras de ARN enriquecidas en ARNm. Varios métodos para enriquecimiento de ARNm están bien documentados y son conocidos por los expertos en la materia. El usado más habitualmente es enriquecimiento de poli A+ mediante perlas paramagnéticas de oligodT o celulosa de oligodT [28]. Están disponibles varios kits comerciales para enriquecimiento de ARNm. Como alternativa el ARNm puede enriquecerse mediante tratamiento con Terminator (Epicentre). Adicionalmente, varias compañías ofrecen ARNm disponibles en el mercado de varios organismos y tejidos.

La concentración y combinaciones de cebadores aleatorios modificados y oligo dT usados en la presente invención proporcionan conversión eficaz y representativa de secuencias de ARNm en ADNc. Este método proporciona conversión superior y sin sesgos de secuencias de ARNm en ADNc independientemente de la distancia desde el extremo 3' del ARNm. Adicionalmente garantiza que cualquier molécula de ARN se transcriba de forma inversa solamente una vez inhibiendo el desplazamiento de cadena de la transcriptasa inversa.

Se desvelan métodos para aumentar la representación igual de ARNm y, más particularmente, aumentar la precisión de cuantificación de expresión génica. Por lo tanto, se proporciona síntesis de ADNc mejorada útil en descubrimiento de genes, investigación genómica, diagnóstico e identificación de genes expresados diferencialmente e identificación de genes importantes para enfermedad.

En otra realización pueden usarse finalizadores oligonucleotídicos para bloquear específicamente la transcripción inversa de transcritos no deseados tales como, pero sin limitación, altamente abundantes o ribosómicos durante el proceso de transcripción inversa. Esta etapa puede usarse para suprimir moldes específicos para generar una mezcla normalizada de productos. Un uso práctico es en la supresión de ARNm abundante para generar una biblioteca de ADNc normalizada. Los finalizadores oligonucleotídicos que se usan para suprimir un molde, es decir, no se usan para generar un producto de elongación bien definido, se denominan en el presente documento oligonucleótidos de bloqueo o bloqueadores. Se conocen oligonucleótidos de bloqueo de los documentos GB 2293238, US 2002/0076767 A1, EP 6.391.592 B1, WO 99/61661, WO 02/086155, US 5.849.497, WO 2009/019008 y pueden emplearse según la presente invención. Preferentemente dichos oligonucleótidos de bloqueo serían altamente específicos para los transcritos que debería evitarse que se transcriban de forma inversa, por lo tanto tienen preferentemente una longitud de más de 15 nt. Los oligonucleótidos más largos tienen una T<sub>m</sub> mayor, por lo tanto son más difíciles de desplazar mediante la transcriptasa inversa. En una realización preferida los oligonucleótidos de bloqueo también tienen una T<sub>m</sub> aumentada en el extremo 5' mediante la introducción de modificaciones tales como pero sin limitación LNA, ZNA, PNA o acridina. Es necesario que estos oligonucleótidos de bloqueo se bloqueen en el extremo 3' para evitar que se extiendan. Dichos bloqueos incluyen, pero sin limitación, espaciadores C3, C6, C12 o didesoxinucleótidos. Los oligonucleótidos de bloqueo deberían diseñarse preferentemente contra una secuencia que se localice cerca del extremo 3' del ARN, pero cadena arriba de la cola de poli A.

Está dentro del alcance de la invención combinar los métodos de la invención usando los oligos de bloqueo anteriormente descritos junto con los mutantes deficientes en síntesis de desplazamiento de cadena tales como M-MLV Y64A o VIH F61W o cualquier otra transcriptasa inversa con síntesis de desplazamiento alterada así como reduciendo la temperatura de reacción y usando diferentes aditivos para aumentar la unión de oligonucleótidos con el ARN.

Como alternativa, si las secuencias de los transcritos altamente abundantes se desconocen, por lo tanto no pueden diseñarse oligos de bloqueo específicos, la detención de desplazamiento de cadena también puede usarse para normalización de biblioteca. Una biblioteca de impulsores se sintetizaría con oligonucleótidos de SDS (preferentemente aquellos con T<sub>m</sub> local aumentada en el extremo 5' tales como oligonucleótidos modificados por LNA, PNA o ZNA).

Con una transferasa terminal pueden añadirse didesoxinucleótidos a las partes de ácido nucleico amplificadas para generar los oligos de bloqueo. Esta biblioteca puede después hibridarse con una muestra molde de ensayo. Los transcritos altamente abundantes tendrán una ventaja y la parte de ácido nucleico amplificada monocatenaria del impulsor será más rápida y más eficaz en la hibridación con el molde correspondiente. La síntesis de productos pueden iniciarse después usando un cebador, tal como cebador de oligodT para ARNm y ya que la mayoría de los transcritos altamente abundantes se hibridan con oligonucleótidos de bloqueo, los transcritos poco abundantes tienen una probabilidad mucho mayor de transcribirse de forma inversa. Si también se usa un protocolo de cambio de molde o recubrimiento de oligo, el ADNc resultante de los transcritos poco abundantes puede incluso insertarse en una reacción de PCR con los cebadores correspondientes al cebador de oligo-dT y el cambio de molde o nucleótido de recubrimiento de oligo. Como alternativa puede usarse marcaje terminal selectivo (documento US 2009/0227009 A1) para marcar el extremo 3' del ADNc de nueva síntesis. Ya que es más probable que los transcritos altamente abundantes nunca hayan alcanzado longitud completa debido a los oligonucleótidos de bloqueo solamente se amplificarán transcritos poco abundantes en una PCR con cebadores correspondientes al cebador de oligodT y el marcador de ADNc 3'.

Como alternativa, si el extremo 3' del ADNc no está marcado, podría usarse un cebador de oligo-dT con un promotor de ARN polimerasa T7 bicatenario para cebar la reacción de RT de las bibliotecas de ADNc normalizadas y puede usarse una amplificación lineal mediante transcripción *in vitro*.

Por lo tanto puede añadirse un oligonucleótido de bloqueo a la elongación para detener partes de secuencia no deseadas de una molécula molde para elongar. Se seleccionan secuencias preseleccionadas de la secuencia molde no deseada para su uso como bloqueadores.

En realizaciones preferidas de la invención el ácido nucleico molde, por ejemplo el molde de ARN, es monocatenario. En particular es posible que el ácido nucleico molde obtenido carece de una cadena complementaria sobre al menos 30 % de su longitud y/o carece de una cadena complementaria de al menos 100 nucleótidos de longitud, preferentemente carece de un ácido nucleico complementario sobre su longitud completa. El método de la invención también abarca separar cadenas complementarias posiblemente existentes del ácido nucleico molde de la invención e introducir el ácido nucleico molde purificado de este modo sin una cadena complementaria en el método de la invención.

En particular se prefiere que los uno o más cebadores de la invención y uno o más finalizadores se unan con la cadena molde, y particularmente no con una cadena complementaria.

La presente invención se refiere a los métodos descritos anteriormente para generar uno o más ácidos nucleicos molde para generar una biblioteca de secuencia que comprende una mezcla de partes de ácido nucleico amplificadas de dichos ácidos nucleicos molde. Preferentemente, las partes de ácido nucleico amplificadas proporcionan partes de secuencia solapantes del molde. Esto puede facilitarse por ejemplo usando una multitud de

La invención proporciona además un kit para generar una biblioteca de partes de ácido nucleico amplificadas de un ácido nucleico molde como se describe en el presente documento o para generar una biblioteca de secuencias como se describe en el presente documento que comprende una ADN polimerasa, cebadores oligonucleotídicos aleatorios que comprenden una modificación que aumenta la  $T_m$  y finalizadores oligonucleotídicos aleatorios que son inadecuados para extensión de nucleótidos y comprenden una modificación que aumenta la  $T_m$ , y una ligasa, preferentemente ADN ligasa o ARN ligasa con actividad de ligamiento de ADN, opcionalmente además uno o más de tampones de reacción que comprenden  $Mn^{2+}$  o  $Mg^{2+}$ , un agente aglutinante adecuado para reacciones de ligasa, como PEG. La ligasa también puede ser una ARN ligasa, especialmente una ARN ligasa que tiene actividad de ligamiento de ADN tal como ARN ligasa T4 2. Los agentes aglutinantes son moléculas inertes que pueden usarse en altas concentraciones para imitar los efectos de aglutinamiento macromolecular dentro de una célula. Son ejemplos PEG (polietilenglicol), PVP (polivinilpirrolidona), trehalosa, ficoll y dextrano. Se desvelan agentes aglutinantes por ejemplo en los documentos US 5.554.730 o US 8.017.339. El kit puede comprender como alternativa una ADN polimerasa y una ligasa, cebadores oligonucleotídicos, finalizadores oligonucleotídicos, en el que uno cualquiera de los finalizadores oligonucleotídicos comprende un marcador de secuencia y uno cualquiera de los cebadores oligonucleotídicos comprende un marcador de secuencia, preferentemente un agente aglutinante, especialmente PEG.

Kits adicionales de la invención que son adecuados para generar partes de ácido nucleico amplificadas de un ácido nucleico molde comprenden o contienen a) cebadores oligonucleotídicos aleatorios que comprenden una modificación que aumenta la  $T_m$  y b) finalizadores oligonucleotídicos aleatorios que son inadecuados para extensión de nucleótidos y comprenden una modificación que aumenta la  $T_m$ , y c) una ligasa. El kit puede comprender además uno o más de tampones de reacción que comprenden  $Mn^{2+}$  o  $Mg^{2+}$ , una ligasa, un agente aglutinante, tal como PEG. El kit puede comprender además una ADN polimerasa.

Los kits para uso de acuerdo con la invención también pueden comprender un medio de transporte, tal como una caja o un cartón, que tiene en confinamiento estrecho en el mismo uno o más medios recipientes, tales como viales, tubos, frascos y similares. El kit puede comprender (en el mismo recipiente o recipientes separados) uno o más de los siguientes: una o más transcriptasas inversas, tampones adecuados, uno o más nucleótidos especialmente dNTP y/o uno o más cebadores (por ejemplo, oligo(dT), iniciadores, finalizadores, complementos inversos, cebadores de PCR) para transcripción inversa y reacciones de PCR posteriores. Los kits abarcados por este aspecto de la presente invención pueden comprender además reactivos y compuestos adicionales necesarios para llevar a cabo protocolos de transcripción inversa de ácidos nucleicos según la invención, tales como perlas de oligodT o perlas acopladas con estreptavidina. Además, los cebadores o finalizadores pueden inmovilizarse en una superficie sólida.

La presente invención se ilustra adicionalmente en las siguientes figuras y ejemplos, sin limitarse a estas realizaciones de la invención.

## Figuras

Figura 1: Representación esquemática del problema principal que la invención busca resolver.

a) Los cebadores P1, P2 hasta Pn se hibridan con un ARN molde. El cebador P2 ha hibridado con una posición más cadena arriba (5') del ARN molde que el cebador P1 y más en general el cebador Pn ha hibridado con una posición más cadena arriba en el ARN molde que el cebador P(n-1). O, en otras palabras, el cebador P1 ha hibridado con una posición más cadena abajo (3') en el molde que P2 y más en general P(n-1) ha hibridado con una posición más cadena abajo en el ARN molde que Pn. La extensión de cada cebador se inicia por una transcriptasa inversa, b) Cuando la transcriptasa inversa mientras polimeriza el producto de extensión de P2 alcanza un cebador P3, el cebador P3 y su producto de extensión se someten a desplazamiento de cadena por la transcriptasa inversa que continúa extendiendo el producto de extensión de cebador P2. Lo mismo sucede para el producto de extensión de P1 que desplaza el cebador P2 y su producto de extensión, c) Por lo tanto, cuando todos los productos de extensión están terminados, está presente una copia de ADNc de la secuencia molde entre P1 y P2, pero están presentes dos copias de ADNc entre P2 y P3. Este fenómeno conduce a una sobrerrepresentación de los extremos 5' de ARN que se ceba múltiples veces, como sucede durante la transcripción inversa convencional usando cebadores aleatorios tales como hexámeros aleatorios. Más en general, cuando n cebadores han hibridado y fueron extendidos por la polimerasa en el extremo 5' del ARN molde, entonces el extremo 5' del ARN estará representado n veces mientras que el extremo 3' del ARN estará representado solamente una vez.

Figura 2: Representación esquemática de una realización de la invención para crear una biblioteca de ADNc

equilibrada 5' - 3'.

a) La hibridación de los cebadores P1, P2 a Pn se refuerza usando por ejemplo: ácidos nucleicos bloqueados (LNA) para inhibir la actividad de desplazamiento de cadena de la transcriptasa inversa. Aquí se usan tres modificaciones al final del extremo 5' de los cebadores. b) Ahora tras extender un cebador que queda más cadena abajo en el ARN molde la transcriptasa inversa no puede desplazar el cebador que ha hibridado con una posición más cadena arriba en el molde. Por lo tanto, cada parte del ARN solamente está representada una vez como ADNc. De esa manera no se producirá ninguna sobrerepresentación de extremos 5' de ARN. c) Cuando se desee ADNc de longitud completa los productos de extensión de cebadores individuales se ligan para producir d) una copia de ADNc de longitud completa del ARN molde.

Figura 3: Representación esquemática de la creación de una biblioteca de ADNc corto marcada con enlazador.

Para muchos análisis cadena abajo de bibliotecas de ADNc es necesaria una secuencia enlazadora universal en uno o ambos extremos del ADNc, para por ejemplo amplificar la biblioteca o iniciar una reacción de secuenciación. Aquí se muestra la preparación de una biblioteca que tiene dos enlazadores. a) Los cebadores P1, P2 a Pn tienen una extensión de secuencia enlazadora universal 5' (L1). Además se usan oligos finalizadores S1, S2 a Sm que tienen una extensión de secuencia enlazadora universal 3' (L2). Los oligos finalizadores también se modifican (por ejemplo LNA) de modo que tampoco puedan ser desplazados por la transcriptasa inversa, b) En una segunda etapa de reacción el extremo 3' del producto de extensión se liga con el extremo 5' del oligo finalizador. De esa manera se crea una biblioteca de ADNc que tiene dos secuencias enlazadoras (L1, L2) disponibles para c) amplificar la biblioteca completa durante por ejemplo una PCR posterior.

Figura 4: Representación esquemática de la creación de una biblioteca de ADNc corto marcada con enlazador usando un concepto de oligo finalizador alternativo.

Ya que la tasa de error que se introduce en la biblioteca de ADNc mediante una hibridación errónea de la secuencia del oligo finalizador es mayor que si esta parte se hubiera transcrito por una polimerasa, se muestra un concepto de oligo finalizador alternativo. a) Aquí en comparación con la Fig. 3 el oligo finalizador Sm se extiende en su lado 5' por una secuencia L2rc (complemento inverso). Este L2rc se hibrida con L2 para formar un adaptador. b) De nuevo como en la Fig. 3 la transcriptasa inversa extiende los cebadores Pn hasta que alcanzan los oligos finalizadores Sm. Durante el ligamiento el producto de extensión se liga ahora con la cadena L2 del adaptador. De esta manera se crea de nuevo una biblioteca de ADNc que tiene dos secuencias enlazadoras (L1, L2) presentes en sus extremos. Aquí, sin embargo, no se introduce ninguna secuencia de oligo finalizador. Por lo tanto, cuando se secuencia desde el lado L2 de la biblioteca una hibridación errónea potencial del oligo finalizador no introduce ninguna ambigüedad con respecto a la identidad de los primeros nucleótidos. c) Finalmente, se realiza una PCR.

Figura 5: Representación esquemática de la creación de una biblioteca de ADNc corto marcada con enlazador usando un concepto de oligo finalizador alternativo.

En un concepto alternativo el iniciador y el oligo forman un heterodímero sobre una secuencia complementaria en sus secuencias enlazadoras respectivas L1 y L2. Los huecos de secuencias se reducirán ya que en cualquier detención también hay una polimerización iniciada por el iniciador. a) Los híbridos de iniciador/finalizador se extienden en el extremo 3' de la secuencia iniciadora P1 a Pn hasta que se encuentra el siguiente híbrido de iniciador/finalizador. Los oligos finalizadores S1, S2 a Sm tienen una extensión de secuencia enlazadora universal 3' (L2) que está en un híbrido con la secuencia L1 del iniciador. Es suficiente si el finalizador se modifica para evitar el desplazamiento de cadena aunque en un subconjunto alternativo de dichos híbridos de iniciador/finalizador mostrados en la Figura 8e es necesario que la modificación se localice en el iniciador si la secuencia L2 no tiene extensión de secuencia 5' que hibrida con el ácido nucleico molde. b) En una segunda etapa de reacción el extremo 3' del producto de extensión se liga con el extremo 5' del oligo finalizador. De esa manera se crea una biblioteca de ADNc que tiene dos secuencias enlazadoras (L1, L2) disponibles para c) amplificar la biblioteca completa durante por ejemplo una PCR posterior.

Figura 6: Representación esquemática de modificaciones de cebadores preferidas.

Se representan modificaciones de cebadores preferidas. Estas modificaciones también pueden combinarse. a) Estructura general de oligo de cebado, con secuencia de cebador y enlazador. Cuando se genera una biblioteca de ADNc con cebadores aleatorios la secuencia de cebador es típicamente una secuencia aleatoria tal como, por ejemplo, un hexámero aleatorio. b) La parte de secuencia de cebador contiene nucleótidos modificados. La modificación reduce o inhibe la actividad de desplazamiento de cadena de la transcriptasa inversa, c) Se introduce un complemento inverso (L1rc) que inhibe la parte de secuencia L1 del oligo para participar en la hibridación con la cadena de ARN molde. Por lo tanto, se bloquea un sesgo hacia la secuencia L1. d) Para evitar que la transcriptasa inversa se asocie con los adaptadores enlazadores y de este modo reduzca la eficacia de la transcripción inversa se usa un saliente 3' en el extremo 3' de L1rc. e) La secuencia L1rc se modifica para aumentar la fuerza de hibridación con la secuencia L1 para reducir adicionalmente la probabilidad de sesgo en la biblioteca hacia la secuencia L1. f) Bloqueo de todos los extremos que no participan en extensión de polimerasa o reacción de ligasa. g) El extremo 5'

de la secuencia L1 y el extremo 3' de la secuencia L1rc se conectan mediante un enlace covalente, por ejemplo un espaciador Cn o una secuencia en horquilla.

Figura 7: Representación esquemática de modificaciones de finalizadores oligo preferidas.

Se representan modificaciones de oligonucleótidos finalizadores preferidas. Estas modificaciones también pueden combinarse. a) Estructura general de oligo finalizador, con secuencia de finalizador y enlazador. La secuencia finalizadora es típicamente una secuencia aleatoria tal como por ejemplo un nonúmero aleatorio. b) La parte de secuencia de finalizador contiene nucleótidos modificados. La modificación reduce o inhibe la actividad de desplazamiento de cadena de la transcriptasa inversa. c) Se introduce un complemento inverso (L2rc) que inhibe la parte de secuencia L2 del oligo para participar en la hibridación con la cadena de ARN molde. Por lo tanto, se bloquea un sesgo hacia la secuencia L2. d) Para evitar que la transcriptasa inversa se asocie con los adaptadores de enlazadores y por lo tanto reduzca la eficacia de transcripción inversa, se usa un saliente 3' en el extremo 3' de L2. e) La secuencia L2rc se modifica para aumentar la fuerza de hibridación con la secuencia L2 para reducir adicionalmente la probabilidad de sesgo en la biblioteca hacia la secuencia L2. f) Bloqueo de todos los extremos que no participan en extensión de polimerasa o reacción de ligasa. g) El extremo 5' de la secuencia L1 y el extremo 3' de la secuencia L1rc se conectan mediante un enlace covalente, por ejemplo un espaciador Cn o una secuencia en horquilla. h) El extremo 5' del oligo finalizador se fosforila para poder actuar como un donante en la reacción de ligamiento. Como alternativa el extremo 5' se adenila. i) Representa la estructura general del oligo finalizador usado en el concepto de finalizador oligo alternativo en la Fig. 4. El extremo 5' de la secuencia L2 puede fosforilarse para poder actuar como un donante en una reacción de ligamiento, o como alternativa adenilarse.

Figura 8: Representación esquemática de las combinaciones de iniciador y finalizador oligo más preferidas.

Se representan combinaciones de iniciador/finalizador preferidas. Los iniciadores y oligonucleótido finalizador contienen un tramo de secuencia complementaria (por ejemplo 14 nt) en su secuencia enlazadora, que permite que hibriden entre sí para hibridar entre sí haciendo de este modo la adición de complementos inversos adicionales obsoletos. Los diseños de iniciador y finalizador de la Figura 6 y 7 aún son válidos. El iniciador se bloquea mediante un OH 5'. El oligonucleótido finalizador se bloquea opcionalmente en el extremo 3' por ejemplo mediante didesoxinucleótidos, dioxinucleótidos, espaciadores o nucleótidos invertidos. a) Estructura general de un oligonucleótido finalizador en un híbrido con un oligonucleótido iniciador. Los oligos se hibridan entre sí sobre un tramo de por ejemplo 14 nt. Los oligonucleótidos tanto iniciadores como finalizadores hibridan con la cadena molde juntos. Los iniciadores y finalizadores pueden tener una base de hibridación con el ácido nucleico molde de diferentes longitudes. El oligonucleótido finalizador puede tener una base de hibridación más larga o más corta en comparación con el iniciador o viceversa. b) Tanto el iniciador como el finalizador se modifican para inhibir el desplazamiento de cadena por la polimerasa. c) Aquí solamente se modifican los nucleótidos del extremo 5' del finalizador. Esto es suficiente ya que es necesario que la polimerasa se inhiba para desplazar la cadena solamente en el extremo 5' del finalizador. En efecto una única modificación (tal como LNA o 2'Fluoro) es suficiente. d) muestra también una modificación del iniciador, ya que no todos los iniciadores que se extienden podrían estar en un híbrido con un finalizador que también ha hibridado con el molde. e) Cuando la secuencia del finalizador que hibrida con el molde se reduce a 0 entonces es necesario que el iniciador también proporcione la detención y por lo tanto es deseable una modificación que detenga el desplazamiento de cadena. f) La secuencia L1 y L2 puede tener una parte que no hibrida. g) muestra una configuración alternativa en la que las secuencias enlazadoras tienen partes que hibridan entre sí y partes que no. h) muestra una configuración en la que el oligo iniciador y finalizador se unen entre sí formando en efecto un oligonucleótido. Por supuesto existen muchas más variaciones de iniciadores y finalizadores que pueden ser usadas por alguien experto en la técnica.

Figura 9: Representación esquemática de las estructuras de iniciador y finalizador oligo.

a) Los oligonucleótidos iniciadores se representan de 3' a 5' y b) los oligonucleótidos finalizadores se representan de 5' a 3'. El iniciador (a) consiste en  
(I) una secuencia de cebador tal como una secuencia hexamérica aleatoria que se une con la cadena molde en el lado 3' que preferentemente se modifica para proteger contra el desplazamiento de cadena.  
(II) Opcionalmente una secuencia de código de barras puede localizarse 5' de la secuencia de cebador aleatoria. La secuencia de código de barras consiste preferentemente en 3-9 nucleótidos, lo que permite una identificación única y específica de la biblioteca. Dichos códigos de barras permiten por ejemplo mezclar bibliotecas de diferentes muestras y secuenciarlas juntas en una celda de flujo. Después del ciclo de secuenciación las lecturas pueden demultiplexarse según el código de barras específico.  
(III) El sitio de unión a cebador de secuenciación. Esta es la secuencia usada para unir el cebador de secuenciación durante la secuenciación.  
(IV) una secuencia para amplificación de enlace (por ejemplo: para secuenciación por Illumina NGS) o para unión a una superficie sólida tal como perlas (por ejemplo para secuenciación por SOLiD NGS).  
(V) un marcador de secuencia que proporciona una base de hibridación para oligo iniciador y finalizador.

Un iniciador mínimo puede consistir en una parte de cebado aleatorio (I) y un sitio de unión a cebador de secuenciación (III), en cuyo caso se prefiere un complemento inverso como se describe en la Figura 6c-g para evitar

que tenga lugar la parte de cebador de secuenciación en el proceso de hibridación. Como alternativa un iniciador puede consistir en una parte de cebado aleatorio (I) y un sitio de unión a cebador de secuenciación (III) y una base de hibridación con el finalizador, de modo que III y V podrían ser completa o parcialmente idénticos por ejemplo si el sitio de unión a cebador de secuenciación es complementario de una parte de secuencia III en la región finalizadora.

5 Como alternativa los iniciadores pueden consistir en I, II, III y V, de modo que de nuevo III y V podrían ser completa o parcialmente idénticos por ejemplo si el sitio de unión a cebador de secuenciación es complementario de una parte de secuencia III en la región finalizadora. Los iniciadores también pueden consistir en I, II, III, IV y V.

10 Si se usa un iniciador con una secuencia enlazadora más corta pueden introducirse secuencias correspondientes a sitio de unión a cebador de secuencia (III), códigos de barras (II) y/o unión a superficie (IV) durante la amplificación de las bibliotecas generadas por ejemplo: durante una reacción de PCR introduciendo estos marcadores de secuencia con los cebadores de PCR.

15 Se representan oligonucleótidos finalizadores de 5' a 3' en b). El finalizador (b) puede consistir en (I) una secuencia aleatoria que se une con el molde en el lado 5' que se modifica preferentemente para proteger contra el desplazamiento de cadena.

(II) Opcionalmente una secuencia de código de barras o una secuencia que es complementaria inversa de la secuencia de código de barras en el iniciador (II) puede localizarse 3' de la secuencia de cebador aleatoria. La secuencia de código de barras consiste preferentemente en 3-9 nucleótidos, lo que permite una identificación única y específica de la biblioteca.

20 (III) El sitio de unión a cebador de secuenciación.

(IV) Opcionalmente un marcador de secuencia para unión a superficie para amplificación de enlace (por ejemplo: para secuenciación por Illumina NGS) o para unión a una superficie sólida tal como perlas (por ejemplo para secuenciación por SOLiD NGS) así como

25 (V) un marcador de secuencia que proporciona una base de hibridación para oligo iniciador y finalizador. Un finalizador mínimo puede consistir en una parte de cebado aleatorio (I) y un sitio de unión a cebador de secuenciación (III) en cuyo caso se prefiere un complemento inverso como se describe en la Figura 7c-g para evitar que tenga lugar la parte de cebador de secuenciación en el proceso de hibridación. Como alternativa un finalizador puede consistir en una parte de cebado aleatorio (I) y un sitio de unión a cebador de secuenciación (III) y una base de hibridación con el iniciador, de modo que III y V podrían ser completa o parcialmente idénticos: si el sitio de unión a cebador de secuenciación es complementario de una parte de secuencia III en la región finalizadora. En una realización alternativa este finalizador mínimo también puede carecer de la secuencia aleatoria como se muestra en la Figura 4 u 8e. Como alternativa los finalizadores pueden consistir en I, II, III y V, de modo que III y V podrían ser completa o parcialmente idénticos por ejemplo: si el sitio de unión a cebador de secuenciación es complementario de una parte de secuencia III en la región finalizadora. Los iniciadores también pueden consistir en I, II, III, IV y V.

35 Si se usa un finalizador con una secuencia enlazadora más corta pueden introducirse secuencias correspondientes a sitio de unión a cebador de secuencia (III), códigos de barras (II) y/o unión a superficie (IV) durante la amplificación de las bibliotecas generadas por ejemplo: durante una reacción de PCR introduciendo estos marcadores de secuencia con los cebadores de PCR.

Figura 10: Detención del desplazamiento de cadena durante la transcripción inversa.

45 a) Ilustra el ensayo preparado para determinar la capacidad de oligonucleótidos modificados o no modificados (Seq ID No: 5-8) para inhibir la actividad de desplazamiento de cadena de las transcriptasas inversas que se usan además de un cebador de oligo dT (Seq ID No: 9). La extensión del oligo de detención de desplazamiento de cadena es de 35 nt (Seq ID No: 10) y el producto de detención de desplazamiento de cadena es de 103 nt (Seq ID No: 11) y el producto de ADNc de longitud completa es de 138 nt (Seq ID No: 12).

50 b) Muestra los resultados del Ejemplo 1.

Figura 11: Regulación del tamaño del fragmento de ADNc por la cantidad de oligos de detención insertados en la RT.

55 Muestra los resultados del Ejemplo 2. Se obtienen fragmentos de ADNc mediante síntesis de ADNc con cebadores aleatorios con o sin oligos de detención de desplazamiento de cadena y se analizan en una inmunotransferencia.

Figura 12: Detención del desplazamiento de cadena durante la transcripción inversa más ligamiento de los fragmentos de ADNc con un producto de longitud completa

60 Muestra los resultados del Ejemplo 3. Aquí la detención del desplazamiento de cadena es inducida por un oligo con 3 LNA en el extremo 5' y el ligamiento posterior de los 2 fragmentos de ADN resultantes (35 nt Seq ID No: 10 y 103 nt, Seq ID No: 11) con el producto de longitud completa de 138 nt (Seq ID No: 12) usando ADN ligasa T4 (carril 3), ARN ligasa T4 2 truncada (carril 5) o una combinación de ambas (carril 4). En el carril 6 se muestra una reacción de control que omite el oligo SDS. En el carril 2 se añadió el oligo SDS, pero no se añadió ligasa a la reacción de ligamiento.

Figura 13: Validación de SDS/ligamiento en ARNm.

Muestra los resultados del Ejemplo 4. La inmunotransferencia muestra un aumento significativo de la longitud si los fragmentos de ADNc de detención del desplazamiento de cadena cortos se ligan usando ADN ligasa T4.

Figura 14: La detención del desplazamiento de cadena durante la transcripción inversa más ligamiento de los fragmentos de ADNc con un producto de longitud completa da como resultado más producto de un ADNc seleccionado.

Se ilustran los resultados del Ejemplo 5. Se muestran resultados de qPCR de un fragmento de 5 kb de Dync1h1 (NM\_030238) amplificado a partir de diferentes transcripciones inversas (RT). Se realizó una RT usando un oligo SDS (Seq ID No: 15) y un cebador de oligo dT (Seq ID No: 17) o cebador de oligo dT (Seq ID No: 17) por sí sola. Se añadió ADN ligasa T4 o en reacciones de control no se añadió y se realizó una qPCR en los ADNc resultantes. Solamente en caso de ligamiento los oligos SDS fueron capaces de producir el fragmento de 5 kb en una reacción de PCR (cebadores Seq ID No: 18 y Seq ID No: 19). Hay más producto de PCR en reacciones realizadas en SDS/ADNc de ligamiento (es decir: más molde de ADNc disponible) que en un ADNc con cebador de oligo dT regular (compárese carril 2 con carril 6 y 8, respectivamente).

Figura 15: Comparación de la longitud de ADNc de SDS/ligamiento frente a cebado con oligo dT en un ADNc de 15 kb.

La Fig. 15 muestra los resultados del Ejemplo 6, un ensayo de qPCR diseñado para valorar la longitud de ADNc generada en una reacción de RT. Triángulos: transcripción inversa cebada con oligodT, cuadrados: transcripción cebada con hexámero aleatorio, círculos: transcripción de la invención con elongación detenida en cebadores cadena abajo más ligamiento.

Figura 16: Generación de una biblioteca de ADN con doble marcador a partir de ARNm.

Muestra una biblioteca de ADN generada usando el enfoque de SDS/ligamiento (véase carril 2), mientras que el control sin ligamiento permanece vacío (véase carril 3). Sin la adición de molde de ARN pueden generarse algunos productos secundarios enlazador-enlazador ya que el cebador de oligo dT actuará después como un molde para hibridación (véase carril 4). Los controles de PCR sin molde están limpios (véase carril 5).

Figura 17: Transferencia de descubrimiento que compara una preparación de biblioteca usando el nuevo protocolo de detención del desplazamiento de cadena y ligamiento (SDS-ligamiento) con un protocolo de sec de ARNm convencional (kit de prep de muestras de ARN TruSeq™, n.º de catálogo RS-930-20 01). Ambos se secuenciaron en un secuenciador Illumina GAIIx (lectura individual, 72 pb). El eje X muestra el número de lecturas que se mapearon de forma única en el genoma anotado frente al número de genes descubiertos en el eje Y. La comparación de los gráficos muestra que el protocolo de SDS-ligamiento es factible y necesita de hecho menos lecturas para detectar la misma cantidad de genes que el protocolo convencional.

#### Abreviaturas

qPCR: reacción en cadena de la polimerasa cuantitativa  
 SDS: detención del desplazamiento de cadena  
 RT: transcripción inversa o transcriptasa inversa (dependiendo del contexto)  
 LNA: ácido nucleico bloqueado  
 PNA: ácido nucleico peptídico  
 rc: complemento inverso  
 Tm: temperatura de fusión  
 SNP: polimorfismo de un único nucleótido  
 CGH: hibridación genómica comparativa  
 CNV: variación del número de copias  
 PTO: enlace de fosforotioato  
 Fos: fosforilación

#### Definiciones

Iniciador: Los iniciadores son moléculas que pueden cebar una reacción de extensión de polimerasa con molde. Habitualmente los iniciadores tienen una secuencia oligonucleotídica denominada cebador. Los iniciadores pueden tener extensiones de secuencia 5' tales como secuencias enlazadoras universales descritas en las Fig. 6, 7, 8 y 9.

## Ejemplos

**Ejemplo 1: Detención del desplazamiento de cadena de transcriptasas inversas usando oligonucleótidos modificados por LNA.**

Secuencias:

Un asterisco indica un enlace de fosforotioato (PTO), un signo más "+" delante de un nucleótido indica un ácido nucleico bloqueado (LNA). "Fos" indica una fosforilación,

SEQ ID No. 1: 5'-GCTAATACGACTCACTATAGTTGTCACCAGCATCCC-3'

SEQ ID No. 2: 5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGAATGGGCCGCAGGA-3'

| SEQ                                                                         | ID | No. | 3: | 5'- |
|-----------------------------------------------------------------------------|----|-----|----|-----|
| GCTAATACGACTCACTATAGTTGTCACCAGCATCCCTAGACCCGTACAGTGCCCACTCCCCTTCCCAGTTTCCGA |    |     |    |     |
| CTGTCCCCGGCCTCCTGCGGCCCATTCGAAAAAAAAAAAAAAAAAAAAAAAAAAAA-3'                 |    |     |    |     |

| SEQ                                                                                            | ID | No. | 4: | 5'- |
|------------------------------------------------------------------------------------------------|----|-----|----|-----|
| guugucaccagcaucccuagaccguacagugccacucccuucccaguuuuccgacuguccccggccuccugcgcccauucgaaaaaaaaaaaaa |    |     |    |     |
| aaaaaaaaaaaaa-3' (ARN)                                                                         |    |     |    |     |

SEQ ID No. 5: (Fos)(5'-+GGGCACTGTACG-3')

SEQ ID No. 6: (Fos) (5'-GGGCACTGTACG-3')

SEQ ID No. 7: (Fos) (5'-G\*GGCACTGTAC\*G-3')

SEQ ID No. 8: (Fos)(5'-+G+G+GCACTGTAC\*G-3')

SEQ ID No. 9: 5'-A\*CGGAGCCTATCTATATGTTCTTGACATTTTTTTTTTTTTTTTTTTTTTTTTTTT\*V-3'

SEQ ID No. 10: 5'-GGGCACTGTACGGGTCTAGGGATGCTGGTGACAAC-3'

| SEQ                                                                             | ID | No. | 11: | 5'- |
|---------------------------------------------------------------------------------|----|-----|-----|-----|
| A*CGGAGCCTATCTATATGTTCTTGACATTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGAATGGGCCGCAGGAGGCCGG |    |     |     |     |
| GGACAGTCGGAACTGGGAAGGGGAGT-3'                                                   |    |     |     |     |

| SEQ                                                                             | ID | No. | 12: | 5'- |
|---------------------------------------------------------------------------------|----|-----|-----|-----|
| A*CGGAGCCTATCTATATGTTCTTGACATTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGAATGGGCCGCAGGAGGCCGG |    |     |     |     |
| GGACAGTCGGAACTGGGAAGGGGAGTGGGCACTGTACGGGTCTAGGGATGCTGGTGACAAC-3'                |    |     |     |     |

Para investigar la viabilidad de la inhibición de la actividad de desplazamiento de cadena de la transcriptasa inversa en condiciones que permitan hibridación de cebadores y polimerización de ADNc, se muestra un experimento de prueba de concepto. Para un perfil de la preparación del ensayo véase Fig. 10a y para los resultados véase Fig. 10b.

Generación de molde de ARN de 111 nt transcrito *in vitro*:

Se amplificó por PCR un fragmento de 75 pb de GAPDH (NC\_000072, 48 nt-122 nt) con cebadores que contenían una secuencia promotora de T7 (Seq ID No: 1) o una cola de T27 (Seq ID No: 2). El producto de PCR resultante (SEQ ID No: 3) actuó como molde para transcripción de T7 *in vitro* usando el kit de transcripción de T7 AmpliScribe Flash de Epicentre. El ARN de 111 nt transcrito *in vitro* (Seq ID No: 4) actuó como molde para el ensayo de detención del desplazamiento de cadena durante la transcripción inversa.

Síntesis de ADNc:

Se llevó a cabo síntesis de ADNc de primera cadena usando MMLV-H de Promega (reacción de 250 U/20 µl). La preparación de ensayo de desplazamiento de cadena se representa en la Fig. 10a. Cebadores para síntesis de ADNc, oligos modificados por LNA (1 LNA-G (Seq ID No: 5), oligos no modificados (Seq ID No: 6), oligos modificados por PTO (Seq ID No: 7), oligos 3 LNA-G (Seq ID No: 8) u oligo dT (Seq ID No: 9) se obtuvieron de Microsynth AG (Balgach Suiza) o Eurogentec (Seraing, Bélgica). Se calentaron 800 ng de ARN de 111 nt transcrito *in vitro* (Seq ID No: 4) y oligos (Seq ID No: 5-9; cebador de oligodT 50 nM SEQ ID No: 9 y SEQ ID No: 5-8 1,5 µM) hasta 70 °C durante 2 min con todos los componente necesarios (excepto por la transcriptasa inversa) incluyendo: tampón (Bis-Tris-metano 50 mM pH 7,9, KCl 75 mM, MgCl<sub>2</sub> 4 mM, trehalosa 0,6 M y glicerol 7,5 %), 0,5 mM de cada dNTP, ditiotreititol (DTT) 10 mM, y se emparejó lentamente reduciendo la temperatura hasta 40 °C con reposos de 30 s en cada reducción de 2 °C. A 40 °C se añadieron 250 unidades de transcriptasa inversa y la temperatura se elevó lentamente hasta 45 °C con reposos de 1 min en cada aumento de 1 °C seguido de una incubación de 30 minutos a 46 °C. Después de síntesis de primera cadena, las muestras se calentaron hasta 95 °C durante 5 min en NaOH 0,1 M, se neutralizaron con molaridades iguales de HCl y se purificaron mediante precipitación con EtOH. Después de lavar con EtOH 75 % los sedimentos se disolvieron en 5 µl de Tris 10 mM, pH 8,0 y se mezclaron con un volumen igual de tampón de carga de formamida 100 %, se desnaturalizaron a 95 °C durante 2 min, se enfriaron en hielo y se resolvieron por electroforesis en una acrilamida 15 %/urea 7 M.

Los resultados se muestran en la Fig. 10b.

En los diferentes carriles se varió el segundo oligo cadena abajo (que ha hibridado con una parte más cadena arriba del molde) (Seq ID No: 5-8) usando un oligo que tenía una modificación por LNA (Seq ID No: 5) en el carril 2, sin modificación (Seq ID No: 6) en el carril 3, un PTO (Seq ID No: 7) en el carril 4 o tres modificaciones por LNA (Seq ID

No: 8) en el carril 5 y sin segundo oligo en el carril 6. Los carriles 1 y 7 muestran un marcador de tamaño (marcador de 10 pb, Invitrogen). Puede verse que los oligonucleótidos no modificados (carril 3) o los modificados por PTO (carril 4) tienen desplazamiento completo de cadena por transcriptasa inversa de MMLV-H ya que solo el producto de longitud completa (Seq ID No: 12) a 138 nt es visible y no hay producto de detención del desplazamiento de cadena a 103 nt (Seq ID No: 11). El segundo oligo también se ha extendido hasta 35 nt ((Seq ID No: 10) y por lo tanto el extremo 5' del ARN de 111 nt está representado dos veces (una vez en forma del producto de longitud completa (138 nt) y una vez por el producto de extensión de 35 nt del segundo oligo). La introducción de 1 LNA en la secuencia 5' del segundo oligo ya provoca cierta detención del desplazamiento de cadena (son visibles productos de 138 nt y 103 nt), mientras que 3 LNA conducen a una detención casi completa del desplazamiento de cadena (ningún producto de 138 nt más, solamente el producto de detención del desplazamiento de cadena de 103 nt (Seq ID No: 11).

## Ejemplo 2: Regulación del tamaño de los fragmentos de ADNc.

Este ejemplo muestra que el tamaño de una biblioteca de ADNc que se genera a partir de ARNm puede regularse por la concentración de un cebador aleatorio que detiene una reacción de elongación. Para los resultados véase Fig. 11.

### Aislamiento y purificación de ARN:

Se aisló ARN total de hígado de ratón usando columnas PeqLab Gold en combinación con extracción de fenol ácido (PeqLab, PEQLAB Biotechnologie GMBH, -91052 Erlangen) según la recomendación del fabricante. Se midió la cantidad de ARN mediante absorbancia óptica a 260 nm y se comprobó la integridad en un gel de agarosa de formaldehído o un bioanalizador Agilent.

### Tratamiento con Terminator de ARN total para enriquecer con respecto a ARNm:

se trataron 2-5 µg de ARN total con exonucleasa dependiente de fosfato 5' Terminator™ (Epicentre Biotechnologies, Madison, WI 53713), según las instrucciones del fabricante.

### Síntesis e inmunotransferencia de ADNc:

se llevó a cabo síntesis de ADNc en reacciones de 20 µl con Tris-HCl 50 mM (pH 8,3 a 25 °C), KCl 75 mM, MgCl 3 mM, DTT 10 mM, dNTP 0,75 mM con Digoxigenina-11-2'-desoxi-uridina-5'-trifosfato, estable en alcalinidad. Se incubaron 135 ng de ARNm (ARN total tratado con terminator) a 70 °C durante 2 min en presencia de cebadores y todos los componentes de reacción aparte de la enzima y los dNTP. Se eligió un programa de emparejamiento lento con reposos de 30 segundos a las siguientes temperaturas: 45 °C, 43 °C, 40 °C, 38 °C, 35 °C, 30 °C, 28 °C y 1 minuto a 25 °C. Se añadieron 200 U de MMLV-H, mutante puntual (Promega) y dNTP por cada 20 µl de reacción y se incubaron durante 2 min cada uno a las siguientes temperaturas: 25 °C, 28 °C, 30 °C y 35 °C antes de una extensión final a 37 °C durante 10 min. Después de síntesis de primera cadena, las muestras se calentaron hasta 55 °C durante 15 min en NaOH 0,1 M, se neutralizaron con molaridades iguales de HCl, y se purificaron por precipitación con EtOH. Los fragmentos de ADNc se separaron mediante electroforesis en gel de agarosa de formaldehído (0,8 %), se transfirieron a membranas de transferencia genómica Zeta-Probe GT (BioRAD) mediante electrotransferencia durante 1 h a 50 V según el manual de instrucciones "Mini Trans-BlotR Electrophoretic Transfer Cell" (BioRAD), y después se reticularon mediante luz UV.

Las membranas se equilibraron en tampón de bloqueo lx (ácido maleico 100 mM/NaCl 150 mM, pH 7,5) durante 5 min, antes de bloquear sitios de unión inespecíficos de la membrana en solución de bloqueo (leche 5 % en tampón de bloqueo) durante 30 min. Se incubaron anticuerpos anti Fab diluidos 1:2000 (fragmentos FAB anti digoxigenina-AP, Roche cat n.º 11 093 274 910); 30 ml de solución de bloqueo durante 30 min a temperatura ambiente con agitación. Las membranas se lavaron 2x durante 15 min en tampón de bloqueo lx y después se equilibraron durante 5 min en tampón de tinción 1x (Tris-HCl 0,1 M, pH 9,5 (20 °C), NaCl 0,1 M). Se realizó tinción en solución de tinción (sustrato líquido BCIP®/NBT, de 800 µl a 30 ml de tampón de tinción lx) a temperatura ambiente (en oscuridad) durante una noche sin agitación.

Los resultados se pueden ver en la Fig. 11.

En los carriles 1-5 se usó un oligo (SDS-oligo) con tres LNA en su lado 5' (SEQ ID No: 13: (Fos)(5'-+N+N+NNNN-3')) para aumentar las concentraciones (carril 2: 0,25 µM; carril 3: 2,5 µM; carril 4: 25 µM, carril 5: 50 µM y carril 6: 100 µM). En el carril 1 se muestra una reacción de control con un hexámero aleatorio no modificado (SEQ ID No: 14: (Fos)(5'-NNNNNN-3')) a 2 µM. Aumentando la cantidad de oligos SDS puede reducirse el tamaño de los fragmentos de ADNc generados.

## Ejemplo 3: Ligamiento usando un molde artificial.

El ejemplo muestra que un producto de extensión de un cebador más cadena arriba (P1) puede detenerse mediante

un cebador más cadena abajo (P2) que tiene tres modificaciones por LNA en su extremo 5' y que el producto de extensión del cebador cadena arriba (P1) puede ligarse con el cebador cadena abajo (P2) cuando todos están en un híbrido con el molde. Para una visión de conjunto de las secuencias implicadas véase Fig. 10a. Para los resultados véase Fig. 12. Se llevó a cabo transcripción inversa usando oligos, moldes y condiciones como se describen en el Ejemplo 1.

Después de la RT las muestras se precipitaron con etanol y se insertaron en una reacción de ligamiento de 15 µl con HCC 1 mM, PEG-8000 20 %, Tris-HCl 30 mM (pH 7,8 a 25 °C), MgCl<sub>2</sub> 10 mM, DTT 10 mM, ATP 1 mM y 1,5 µl de ADN ligasa T4 (1-3 unidades de Weiss/µl, Promega) y/o 1 µl de ARN ligasa T4 2 troncada (10 unidades/µl, NEB) y se incubaron 2 h a 37 °C. Los fragmentos pequeños no ligados y los oligos restantes se retiraron mediante precipitación con PEG. Por lo tanto el volumen de reacción aumentó hasta obtener 11,5 % de concentración final de PEG y se añadieron 2 µl de poliacrilamida lineal (10 mg/ml) como vehículo. Las reacciones se agitaron vorticialmente de forma exhaustiva antes de la centrifugación a 20.000xg durante 15 min a 18 °C. El sedimento se lavó 2x en EtOH 75 %, se disolvió en 5 µl de Tris 10 mM (pH 8,0), se mezcló con un volumen igual de tampón de carga de formamida 100 %, se desnaturalizó a 95 °C durante 2 min, se enfrió en hielo y se cargó en un gel de acrilamida 15 %/urea 7 M. Los resultados se pueden ver en la Fig. 12. En el carril 6 se transcribió de forma inversa el ARN artificial usando solamente un cebador de oligo dT (SEQ ID No: 9). Cuando se añade un oligo SDS (SEQ ID No: 8) además del cebador de oligo dT (SEQ ID No: 9), hay una detención completa del desplazamiento de cadena como se ve por la aparición de un producto de SDS de 103 nt (Seq ID No: 11, véase Fig. 12, carril 2). Además se genera un producto de extensión de oligo SDS de 35 nt (Seq ID No: 10). ADN ligasa T4 (Fig. 12, carril 3) o ARN ligasa T4 2, troncada (Fig. 12, carril 5) así como una combinación de ambas enzimas (Fig. 12, carril 4) se ensayaron con respecto a su eficacia para ligar los dos fragmentos de ADNc en un híbrido de ARN. ARN ligasa T4 2, troncada, es deficiente en una función de adenilación y por lo tanto puede ligarse solamente con oligos ya adenilados es decir: oligos que se adenilaron previamente mediante ADN ligasa T4 en el híbrido. No se usó ARN ligasa T4 1 debido al ligamiento preferido de moléculas monocatenarias que después también daría como resultado ligamiento de oligos no hibridados preferentemente con el molde de ARN. Como puede verse en los carriles 3-5 los productos detenidos (103 nt, Seq ID No: 11) y el producto de extensión de oligo SDS (35 nt, Seq ID No: 10) pueden ligarse en el híbrido de ARN, dando como resultado un ADNc de 138 nt de longitud completa (Seq ID No: 12).

#### **Ejemplo 4: Ligamiento de fragmentos de ADNc sintetizados a partir de poli A-ARNm.**

El ejemplo muestra que ligando fragmentos de ADNc cortos que se generaron por el método de la invención se produce un desplazamiento de tamaño en la inmunotransferencia, lo que indica que se creó ADNc más largo por el proceso de ligamiento. Para los métodos véase Ejemplo 2 y 3. Para los resultados véase Fig. 13. Se usó ARNm seleccionado por poli A (poliA-ARNm de hígado de ratón, Stratagene) como un molde. Los fragmentos de ADNc cortos (véase Fig. 13, carril 1, fragmentos entre 100 y 700 nt) que se generaron usando combinación de cebador de oligo dT (SEQ ID No: 9) y un dodecámero aleatorio con nucleótidos modificados por 3 LNA (SEQ ID No: 15: (Fos)(5'-G+G+GHHNNNNNN-3')) se ligaron en el híbrido de ARN usando ADN ligasa T4, que da como resultado ADNc largos (de longitud completa) aun mayores de 6000 nt (véase Fig. 13, carril 2).

#### **Ejemplo 5: En una qPCR específica de gen la transcripción inversa de SDS/ligamiento produce más producto que una RT con cebador de oligodT regular.**

Se llevó a cabo una PCR específica de gen en una reacción de 20 µl que contiene 1 µl de ADNc (sintetizado a partir de 800 ng de ARN total, disuelto en 42 µl de Tris 10 mM, pH 8,0 después de purificación), Tris 50 mM-Cl pH 9,2, sulfato de amonio 16 mM, Tween 20 0,1 % y MgCl<sub>2</sub> 5,1 mM, Betaína 1,5 M, DMSO 1,3 %, SYBRGreen 0,5x, 0,2 mM de cada dNTP, 0,3 µM de cada cebador (SEQ ID No: 18: 5'-CTGGATGAATGGCTTGAGTGT-3' y SEQ ID No: 19: 5'-GCAACTCCACGCTCATAGAAG-3', cebadores diseñados para NM\_030238), 0,8 unidades de polimerasa KlenTaq AC y 0,2 unidades de polimerasa Pfu. Las muestras se desnaturalizaron a 95,8 °C durante 15 s y se ciclaron 20 veces a 95,8 °C durante 15 s, 55 °C durante 30 s, 74 °C durante 20 min (aumento de velocidad a ABI9700: 50 %). Posteriormente, siguieron 19 ciclos de amplificación a 95,8 °C durante 15 s, 58 °C durante 30 s, 74 °C durante 20 min (aumento de velocidad a ABI9700: 10%) con una etapa de extensión final a 72 °C durante 3 min. Los productos de PCR se purificaron usando columna de sílice y se cargaron en un gel de agarosa 0,7 %. Los resultados se muestran en la Fig. 14. El carril 8 muestra el producto de PCR de 5096 pb generado a partir de un ADNc sintetizado con un cebador de oligo dT (Seq ID NO: 17: 5'-G\*GCGTTTTTTTTTTTTTTTTT\*V-3'). Como control también se sometió un ADNc con cebador de oligodT al protocolo de ligamiento que habitualmente se aplica para los oligos de detención del desplazamiento de cadena (véase carril 6). Cuando Seq ID No: 15 (oligos SDS) y Seq ID No: 17 (oligo dT) se usaron para cebar la reacción de RT seguido de un protocolo de ligamiento como se describe en el Ejemplo 3, la cantidad de producto de PCR generado a partir de dicho ADNc era aún más que de un ADNc con cebador de oligo dT regular (compárese Fig. 14, carril 2 con carril 6 y 8, respectivamente). Esto puede explicarse por los oligos SDS que evitan la formación de estructura secundaria debido a hibridación, mientras que esas estructuras secundarias conducen a acontecimientos de detención de polimerización prematura si solo se usa un cebador de oligo dT. Con el protocolo de SDS/ligamiento la RT se inicia en múltiples lugares y los fragmentos de ADNc resultantes se ligan después para proporcionar productos de ADNc de longitud completa o en este caso el fragmento de 5 kb seleccionado de un transcrito específico seleccionado aleatoriamente. Sin ligamiento no se genera producto de PCR en la PCR posterior de ADNc sintetizado con oligos SDS (véase carril 4). Esto muestra claramente que el

desplazamiento de cadena estaba efectivamente detenido y los fragmentos de ADNc generados no estaban conectados, por lo tanto no puede obtenerse producto de PCR, ya que no hay ningún molde que contenga ambos sitios de unión a cebador de PCR. Los carriles 3, 5 y 7 son controles en los que no se añadió ninguna transcriptasa inversa a la reacción de RT, mostrando de este modo que no hubo contaminación de ADN genómico y que los oligos SDS no dan como resultado ningún fondo inespecífico.

**Ejemplo 6: Los transcritos largos se transcriben de forma inversa más eficazmente usando un protocolo de RT de SDS/ligamiento.**

Se eligió un transcrito largo (componente de ubiquitina proteína ligasa E3 n-recognina 4; Ubr4; NM\_001160319) y se diseñaron amplicones de 200 pb a lo largo del ADNc. Los amplicones se separaron a una distancia de aproximadamente 2 kb entre sí (cebadores Seq ID No: 20-29).

SEQ ID NO. 20: 5'-CCTTCCAGGAGGAGTTCATGCCAGT-3'  
 SEQ ID NO. 21: 5'-CACACGGAGAGATGAATGAGGGGAGA-3'  
 SEQ ID NO. 22: 5'-GCCTTCATGGCTGTGTGCATTGA-3'  
 SEQ ID NO. 23: 5'-CATCCTGCCCTGTAGAAAGTCCTCTTG-3'  
 SEQ ID NO. 24: 5'-CCAGTGTCAAGTGCAGGTCCATC-3'  
 SEQ ID NO. 25: 5'-GCGGTCAGCTTTGTCCAGAAGTGTGT-3'  
 SEQ ID NO. 26: 5'-GTAAGATGGTGGATGGGGTGGGTGT-3'  
 SEQ ID NO. 27: 5'-TCGCTCTGAAATGCTGACTCCTTCA-3'  
 SEQ ID NO. 28: 5'-ACCCAGGTTCTACTGCGTCCTGTCC-3'  
 SEQ ID NO. 29: 5'-CCTCCAGGGCTGTCACGTTCTTCTT-3'

El Ct delta entre los amplicones más 3' y el amplicón más 5' (complementario del extremo 3' del ARNm, cerca de la cola de poli A) se usa para calcular el factor de diferencia (según Pfaffl [34]) y por lo tanto muestra la reducción relativa de ADNc generado sobre la longitud del molde de ARNm. Se comparó ADNc con cebador de oligo dT (Seq ID No: 17) con un ADNc con cebador de Seq ID: 15 y Seq ID No: 14 seguido de un ligamiento de los fragmentos de ADNc resultantes en el híbrido de ARN. Cada una de las reacciones se realizó por triplicado y las medias se representan en la Fig. 15. Los círculos representan valores que se midieron usando SDS-oligos (línea discontinua); Cuadrados cuando se usaron hexámeros aleatorios (línea continua) y triángulos cuando se usó oligo dT (línea punteada-discontinua). Solamente el protocolo de SDS/ligamiento garantiza una síntesis de ADNc igual sobre la longitud del transcrito que tiene una relación de 3' con respecto a 5' casi perfecta de 1.

El ADNc con cebador de oligo dT (triángulos, línea punteada-discontinua) muestra una reducción constante de ADNc generado sobre la longitud del ARNm de 15 kb. Solamente aproximadamente el 10 % de la síntesis de ADNc iniciada originalmente alcanza 6 kb. El cebado aleatorio sin desplazamiento de cadena conduce a una sobrerrepresentación del extremo 5' del ARNm (cuadrados, línea continua).

**Ejemplo 7: Generación de una biblioteca de ADN con doble marcador a partir de ARNm.**

Cebador de oligo dT marcado con biotina 0,11 µM (SEQ ID NO. 30: (Biotina-TEG) (5'-TTTTTTTTTTTTTTTTTTTTTTTTTTT-3'), cebador marcado 2,5 µM (SEQ ID NO. 31: (C12-espaciador) (5'-TCCCTACACGACGCTCTTCCGATCTGACTG+G+G+GNNN-3')+ complemento inverso 2,5 µM del marcador de cebador (SEQ ID NO. 32: (C3-espaciador) (5'-CAGTCAGATCG+GAA+GA+GC+GTC+GT+GTAGGGA-3') (C3-espaciador)), finalizador marcado 5 µM (SEQ ID NO. 33: (Fos)(5'-+G+G+GHHNNNAGATCGGAAGAGCGGTTCAGCAGGA-3')(C3Spacer)) + complemento inverso 5 µM del marcador de finalizador (SEQ ID NO. 34: (C12-espaciador)(5'-TCCT+GCT+GAACC+GCTCTTCC+GATC-desoxyT-3'), desoxiT indica un desoxiT bloqueado 3'), se hibridaron con 150 ng de ARNm poliA+ (BioCat Heidelberg, Alemania) hibridado en Tris, pH 7,0 (70 °C, 1 min, después enfriado lentamente en hielo). Suponiendo una longitud de ARNm de 500-5000 nt esto significaría que los iniciadores se añaden en un exceso molar de 1:280.000-1:28.000, mientras que los finalizadores se añaden en un exceso molar de 1:560.000-1:56.000. Los ácidos nucleicos hibridados se unieron después (20 min a temperatura ambiente) con Dynabeads recubiertas con 1,1 µl de estreptavidina prelavadas (M-280, 10 mg/ml). Los ácidos nucleicos no hibridados se retiraron por lavado (4 lavados según las instrucciones del fabricante). A continuación se añadió tampón de RT a una concentración final de 1x (Tris-HCl 50 mM (pH 8,3 a 25 °C), KCl 75 mM, MgCl<sub>2</sub> 3 mM, DTT 10 mM), dNTP 0,5 mM, 200 unidades de MMLV-H así como 8U de ADN ligasa T4 más PEG 10 % y ATP 0,4 µM. La reacción de transcripción inversa-ligamiento se realizó en perlas calentando la reacción elevando lentamente la temperatura (25 °C durante 2 min, 28 °C durante 1 min, 31 °C durante 1 min, 34 °C durante 1 min) antes de incubar durante 2 horas a 37 °C. De nuevo se lavaron las perlas (4x), antes de hidrolizar el ARN (55 °C durante 15 min en NaOH 0,1 N). Después de neutralización con HCl 0,1 N, las muestras se precipitaron, se disolvieron en 20 µl y después se insertaron 4 µl en una qPCR Illumina (cebadores: SEQ ID NO. 35: 5'-A\*ATGATACGGCGACACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATC\*T-3' y SEQ ID NO. 36: 5'-C\*AAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCTGCTGAACCGCTCTTCCGATC\*T-3'). Los resultados se muestran en la Fig. 16. Se genera una mancha de biblioteca usando el enfoque de SDS/ligamiento (véase carril 2), mientras que el control sin ligamiento permanece vacío (véase carril 3). Sin la adición de molde de ARN pueden generarse algunos productos secundarios enlazador-enlazador ya que el cebador de oligo dT actuará después como

un molde para hibridación (véase carril 4). Los controles de PCR sin molde están limpios (véase carril 5).

Ejemplo 8: Se prepararon muestras de SDS-ligamiento como se describe en el Ejemplo 7. Para comparación se preparó una biblioteca de ARNm-Sec convencional según un protocolo de prep de biblioteca de Illumina (kit de prep de muestras de ARN TruSeq™, n.º de catálogo RS-930-20 01). Ambas bibliotecas se enviaron para secuenciación NGS en una máquina Illumina GAllx (lectura individual, 72 pb). Para comparar el rendimiento de ambas bibliotecas se calcularon las manchas de transferencia de descubrimiento que mostraban el número de genes detectados en relación con un número dado de lecturas que se mapearon de forma única en el genoma anotado. Las manchas de transferencia de descubrimiento se muestran en la Fig. 17. Muestran la viabilidad del protocolo de SDS-ligamiento que de hecho rinde mejor que el protocolo de preparación de bibliotecas de ARNm Sec convencional.

Ejemplo 9: Se prepararon bibliotecas de doble marcador a partir de ARN de referencia humano universal (Agilent Technologies, n.º de catálogo 740000) con adiciones de ARN de ERCC en mezcla de control (n.º de catálogo 4456740) según las instrucciones del fabricante. Se usaron dos métodos de preparación de muestras de NGS: kit ScriptSeq V2 (cat n.º SSV21106, Epicentre, WI) según las instrucciones del fabricante o una preparación de muestras como se describe en el ejemplo 7 con las siguientes modificaciones: se usaron perlas magnéticas oligodT25 de Dynazyme (tomadas del kit directo de ARNm n.º de catálogo 610-12) así como iniciadores modificados por LNA-N

SEQ ID No. 37: (5Sp9)(5'-TCCCTACACGACGCTCTTCCGATCTAGC+N+N+NNNN-3') y finalizadores

SEQ ID No. 38: (fos)(5'-N+N+NNNNNNNAGATCGGAAGAGCGGTTCAGCAGGA-3') (espaciador C3).

En la Tabla 1 se enumeran los resultados del proceso de secuenciación NGS.

| Método       | lecturas totales | % de ERCC | mediana del % de conservación catenaria |
|--------------|------------------|-----------|-----------------------------------------|
| a) SDS_Lig   | 43399527         | 1,09 %    | 100 %                                   |
| b) ScriptSeq | 32709274         | 0,85 %    | 97,52 %                                 |

Tabla 1: Determinación de la conservación catenaria de la biblioteca con adición de ERCC en transcritos.

La conservación catenaria (conservación de la información de cadena) se determinó para ambos métodos usando la adición de ARN de ERCC en controles. Estos controles proporcionan una medida absoluta de la conservación catenaria ya que solamente existen en una orientación es decir: no hay antisentido. Si un método detectara transcritos antisentido del ERCC esto es una medida directa de la tasa de error inherente de dicho método. La conservación catenaria se calcula como la mediana de 1000 lecturas/ERCC para ambos métodos. El método de SDS-ligamiento mostró 100 % de especificidad de cadena sin ninguna lectura en la dirección errónea mientras que con ScriptSeq se determina que la conservación catenaria es del 97,52 %.

## Referencias

- [1] Seyfang A, Jin JH. Multiple site-directed mutagenesis of more than 10 sites simultaneously and in a single round. *Anal Biochem.* 15 ene 2004;324 (2): 285-91.
- [2] Hogrefe HH, Cline J, Youngblood GL, Allen RM. Creating randomized amino acid libraries with the QuikChange Multi Site-Directed Mutagenesis Kit. *Biotechniques.* Nov 2002;33(5):1158-60, 1162, 1164-5.
- [3] Winshell J, Paulson BA, Buelow BD, Champoux JJ. Requirements for DNA unpairing during displacement synthesis by HIV-1 reverse transcriptase. *J Biol Chem.* 17 dic 2004;279 (51):52924-33. Epub 30 sep 2004.
- [4] Bustin SA, Nolan T. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J Biomol Tech.* sep 2004; 15(3): 155-66. Revisión.
- [5] Lacey HA, Nolan T, Greenwood SL, Glazier JD, Sibley CP. Gestational profile of Na<sup>+</sup>/H<sup>+</sup> exchanger and Cl<sup>-</sup>/HCO<sub>3</sub><sup>-</sup> anion exchanger mRNA expression in placenta using real-time QPCR. *Placenta.* ene 2005; 26(1):93-8.
- [6] Zhang J, Byrne CD. Differential priming of RNA templates during cDNA synthesis markedly affects both accuracy and reproducibility of quantitative competitive reverse-transcriptase PCR. *Biochem J* 1999;337:231-241.
- [7] Feinberg, A.P. y Vogelstein, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry*, 132: 6-13 (1983).
- [8] Feinberg, A.P. y Vogelstein, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Addendum. *Analytical Biochemistry*, 137: 266-7 (1984).
- [9] Houldsworth, J. y Chaganti, R. Comparative Genomic Hybridization: an Overview. *American Journal of Pathology* 145: 1253-1260 (1994).
- [10] Gresham, D., Dunham, M.J. y Botstein, D Comparing whole genomes using DNA microarrays. *Nature Reviews Genetics* 9(4):291-302 (2008).
- [11] Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet.* ene 2010; 11(1):31-46. Epub 8 dic 2009. Revisión
- [12] Paulson BA, Zhang M, Schultz SJ, Champoux JJ. Substitution of alanine for tyrosine-64 in the fingers subdomain of M-MuLV reverse transcriptase impairs strand displacement synthesis and blocks viral replication *in vivo*. *Virology.* 30 sep 2007; 366 (2):361-76. Epub 29 may 2007.

- [13] Fisher TS, Darden T, Prasad VR. Substitutions at Phe61 in the beta3-beta4 hairpin of HIV-1 reverse transcriptase reveal a role for the Fingers subdomain in strand displacement DNA synthesis. *J Mol Biol.* 17 ene 2003;325 (3):443-59.
- [14] Fisher, T. S. y Prasad, V. R. (2002). Substitutions of Phe61 located in the vicinity of template 50-overhang influence polymerase fidelity and nucleoside analog sensitivity of HIV-1 reverse transcriptase. *J. Biol. Chem.* 277, 22345-22352.
- [15] Kawasaki, A.M., et al., Uniformly modified 2'-deoxy-2'-fluoro phosphorothioate oligonucleotides as nuclease resistant antisense compounds with high affinity and specificity for RNA targets, *Journal of Medicinal Chemistry* (1993), 36: 831-841.
- [16] Nielsen PE, Egholm M, Berg RH, Buchardt O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science.* 6 dic 1991; 254 (5037):1497-500.
- [17] Egholm M, Buchardt O, Christensen L, Behrens C, Freier SM, Driver DA, Berg RH, Kim SK, Norden B, Nielsen PE. PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature.* 7 oct 1993; 365 (6446):566-8.
- [18] Voirin, E. *et al.* (2007) Versatile synthesis of oligodeoxyribonucleotide-oligospermine conjugates. *Nat Protoc*, 2, 1360-1367
- [19] Moreau *et al.* (2009) Zip nucleic acids (ZNAs): new high affinity oligonucleotides as potent primers for PCR and reverse transcription. *Nucl. Acids Res.*, 37: e130; doi:10.1093/nar/gkp661.
- [20] Nielsen, P., Pfundheller, H. M., Olsen, C. E. y Wengel, J., Synthesis of 2'-O,3'-C-Linked Bicyclic Nucleosides and Bicyclic Oligonucleotides, *J. Chem. Soc., Perkin Trans. 1*, 1997, 3423;
- [21] Singh, S. K.; Nielsen, P.; Koshkin, A. A.; Wengel, J. LNA (Locked Nucleic Acids): Synthesis and High-Affinity Nucleic Acid Recognition. *Chem. Commun.* 1998, 455-456.
- [22] Takusagawa, F. (1997) Selectivity of F8-actinomycin D for RNA:DNA hybrids and its anti-leukemia activity. *Bioorg. Med. Chem.* 5, 1197-1207.
- [23] Shaw Nicholas N., Arya Dev P. Recognition of the unique structure of DNA:RNA hybrid; revisión *Biochimie* 90 (2008), 1026e1039
- [24] Perales C, Cava F, Meijer WJ, Berenguer J. Enhancement of DNA, cDNA synthesis and fidelity at high temperatures by a dimeric single-stranded DNA-binding protein. *Nucleic Acids Res.* 15 nov 2003; 31 (22):6473-80.
- [25] Shigemori Y, Mikawa T, Shibata T, Oishi M. Multiplex PCR: use of heat-stable *Thermus thermophilus* RecA protein to minimize non-specific PCR products. *Nucleic Acids Res.* 8 ago 2005;33 (14):e126.
- [26] Boyer PL, Sarafianos SG, Arnold E, Hughes SH. Analysis of mutations at positions 115 and 116 in the dNTP binding site of HIV-1 reverse transcriptase. *Proc Natl Acad Sci USA.* 28 mar 2000; 97 (7):3056-61.
- [27] Fuentes GM, Rodriguez-Rodriguez L, Palaniappan C, Fay PJ, Bambara RA. Strand displacement synthesis of the long terminal repeats by HIV reverse transcriptase. *J Biol Chem.* 26 ene 1996;271 (4):1966-71.
- [28] Sambrook J. y Russell D., *Molecular Cloning: A Laboratory Manual* (tercera edición, libro 1, capítulo 7.20) US6335439: Method of preparing phosphoramidites
- [29] Engler, M. J. y Richardson, C. C. (1982) DNA ligases. En *The Enzymes*, vol. XV (Boyer, P. D., ed.), págs. 3-29, Academic Press, Nueva York
- [30] Hsuih TC, Park YN, Zaretsky C, Wu F, Tyagi S, Kramer FR, Sperling R, Zhang DY. Novel, ligation-dependent PCR assay for detection of hepatitis C in serum. *J Clin Microbiol.* mar 1996; 34 (3):501-7.
- [31] Bullard DR, Bowater RP. Direct comparison of nick-joining activity of the nucleic acid ligases from bacteriophage T4. *Bio- chem J.* 15 ago 2006;398(1):135-44.
- [32] Zimmerman SB, Pfeiffer BH. Macromolecular crowding allows blunt-end ligation by DNA ligases from rat liver or *Escherichia coli*. *Proc Natl Acad Sci USA.* oct 1983;80(19):5852-6.
- [33] Gerard G. F. y D' Alessio J. M., capítulo 6 (73-93) de: *Methods in Molecular Biology*, Vol.16: *Enzymes of Molecular Biology* editado por: M. M. Burell 1993 Humana Press Inc. Totowa, NJ
- [34] Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 1 may 2001;29(9):e45.
- [35] Li *et al.*, *PNAS* 105 (51) (2008): 20179-20184
- US6335439. Alessandra Eleuteri *et al.* (2002): Method of preparing phosphoramidites
- US20030092905. Alexei Kochkine (2003): Synthesis of [2.2.1]bicyclo nucleosides
- US7084125. Jesper Wengel (2006): Xylo-LNA analogues
- US5436134. Richard P. Haugland *et al.* (1995): Cyclic-substituted unsymmetrical cyanine dyes.
- US5658751. Stephen T. Yue *et al.* (1997): Substituted unsymmetrical cyanine dyes with selected permeability.
- Colorante N.º 211.
- US6569627. Carl T. Wittwer (2003): Monitoring hybridization during PCR using SYBR™ Green I
- US2009/0227009 A1. Roy R. Sooknanan (2009): SELECTIVE TERMINAL TAGGING OF NUCLEIC ACIDS
- US 4.683.195
- US 4.683.202
- US 4.800.159
- US 5.804.375
- US 5.322.770
- US 5.310.652
- US 2002/0076767 A1
- US 6.391.592 B1
- WO 93/13216 A1

WO 94/17210 A1  
WO 98/02449 A1  
WO 99/61661  
WO 02/086155  
5 US 5.849.497  
WO 2009/019008

LISTADO DE SECUENCIAS

|    |                                                                         |            |
|----|-------------------------------------------------------------------------|------------|
| 10 | <110> Lexogen GmbH                                                      |            |
|    | <120> Método de transcripción de ácido nucleico                         |            |
|    | <130> r62429                                                            |            |
| 15 | <150> EP11181546.0                                                      |            |
|    | <151> 16/09/2011                                                        |            |
|    | <150> EP12177647.0                                                      |            |
| 20 | <151> 24/07/2012                                                        |            |
|    | <160> 38                                                                |            |
|    | <170> PatentIn versión 3.3                                              |            |
| 25 | <210> 1                                                                 |            |
|    | <211> 36                                                                |            |
|    | <212> ADN                                                               |            |
|    | <213> Artificial                                                        |            |
| 30 | <220>                                                                   |            |
|    | <223> oligonucleótido cebador                                           |            |
|    | <400> 1                                                                 |            |
| 35 | gctaatacga ctactatag ttgtcaccag catccc                                  | 36         |
|    | <210> 2                                                                 |            |
|    | <211> 43                                                                |            |
|    | <212> ADN                                                               |            |
| 40 | <213> Artificial                                                        |            |
|    | <220>                                                                   |            |
|    | <223> oligonucleótido cebador                                           |            |
| 45 | <400> 2                                                                 |            |
|    | tttttttt tttttttt ttttttcga atgggccgca gga                              | 43         |
|    | <210> 3                                                                 |            |
|    | <211> 130                                                               |            |
| 50 | <212> ADN                                                               |            |
|    | <213> Artificial                                                        |            |
|    | <220>                                                                   |            |
|    | <223> Producto de PCR                                                   |            |
| 55 | <400> 3                                                                 |            |
|    | <b>gctaatacga ctactatag ttgtcaccag catccctaga ccgtacagt gccactccc</b>   | <b>60</b>  |
|    | <b>cttcccagtt tccgactgtc cccggcctcc tgcggcccat tcgaaaaaa aaaaaaaaaa</b> | <b>120</b> |
|    | <b>aaaaaaaaaa</b>                                                       | <b>130</b> |
| 60 | <210> 4                                                                 |            |
|    | <211> 111                                                               |            |

|    |                                                                          |            |
|----|--------------------------------------------------------------------------|------------|
|    | <212> ARN                                                                |            |
|    | <213> Artificial                                                         |            |
|    | <220>                                                                    |            |
|    | <223> Producto de PCR                                                    |            |
| 5  | <400> 4                                                                  |            |
|    | <b>guugucacca gcaucccuag acccguacag ugcccacucc ccuucccagu uuccgacugu</b> | <b>60</b>  |
|    | <b>ccccggccuc cugcggccca uucgaaaaaa aaaaaaaaaa aaaaaaaaaa a</b>          | <b>111</b> |
| 10 | <210> 5                                                                  |            |
|    | <211> 12                                                                 |            |
|    | <212> ADN                                                                |            |
|    | <213> Artificial                                                         |            |
| 15 | <220>                                                                    |            |
|    | <223> oligonucleótido finalizador                                        |            |
|    | <220>                                                                    |            |
|    | <221> LNA                                                                |            |
| 20 | <222> (1)..(1)                                                           |            |
|    | <400> 5                                                                  |            |
|    | gggcactgta cg                                                            | 12         |
| 25 | <210> 6                                                                  |            |
|    | <211> 12                                                                 |            |
|    | <212> ADN                                                                |            |
|    | <213> Artificial                                                         |            |
| 30 | <220>                                                                    |            |
|    | <223> oligonucleótido finalizador                                        |            |
|    | <400> 6                                                                  |            |
|    | gggcactgta cg                                                            | 12         |
| 35 | <210> 7                                                                  |            |
|    | <211> 11                                                                 |            |
|    | <212> ADN                                                                |            |
|    | <213> Artificial                                                         |            |
| 40 | <220>                                                                    |            |
|    | <223> oligonucleótido finalizador                                        |            |
|    | <220>                                                                    |            |
| 45 | <221> cadena principal de fosforotioato                                  |            |
|    | <222> (1)..(2)                                                           |            |
|    | <220>                                                                    |            |
|    | <221> cadena principal de fosforotioato                                  |            |
| 50 | <222> (10)..(11)                                                         |            |
|    | <400> 7                                                                  |            |
|    | gggcactgta c                                                             | 11         |
| 55 | <210> 8                                                                  |            |
|    | <211> 11                                                                 |            |
|    | <212> ADN                                                                |            |
|    | <213> Artificial                                                         |            |
| 60 | <220>                                                                    |            |
|    | <223> oligonucleótido finalizador                                        |            |

|    |                                                                                   |     |
|----|-----------------------------------------------------------------------------------|-----|
|    | <220><br><221> LNA<br><222> (1)..(3)                                              |     |
| 5  | <220><br><221> cadena principal de fosforotioato<br><222> (10)..(11)              |     |
| 10 | <400> 8<br>gggcactgta c                                                           | 11  |
|    | <210> 9<br><211> 55<br><212> ADN<br><213> Artificial                              |     |
| 15 | <220><br><223> oligonucleótido cebador                                            |     |
| 20 | <400> 9<br>acggagccta tctatatgtt cttgacattt tttttttt tttttttt ttttv               | 55  |
|    | <210> 10<br><211> 35<br><212> ADN<br><213> Artificial                             |     |
| 25 | <220><br><223> Producto de PCR                                                    |     |
| 30 | <400> 10<br>gggcactgta cgggtctagg gatgctggtg acaac                                | 35  |
|    | <210> 11<br><211> 103<br><212> ADN<br><213> Artificial                            |     |
| 35 | <220><br><223> Producto de PCR                                                    |     |
| 40 | <400> 11<br><br>acggagccta tctatatgtt cttgacattt tttttttttt tttttttttt ttttcgaatg | 60  |
|    | ggccgcagga ggccggggac agtcggaaac tgggaagggg agt                                   | 103 |
| 45 | <210> 12<br><211> 138<br><212> ADN<br><213> Artificial                            |     |
| 50 | <220><br><223> Producto de PCR                                                    |     |
|    | <400> 12                                                                          |     |
| 55 | acggagccta tctatatgtt cttgacattt tttttttttt tttttttttt ttttcgaatg                 | 60  |
|    | ggccgcagga ggccggggac agtcggaaac tgggaagggg agtgggcact gtacgggtct                 | 120 |
|    | agggatgctg gtgacaac                                                               | 138 |
|    | <210> 13<br><211> 6                                                               |     |

|    |                                             |    |
|----|---------------------------------------------|----|
|    | <212> ADN                                   |    |
|    | <213> Artificial                            |    |
|    | <220>                                       |    |
| 5  | <223> oligonucleótido de hexámero aleatorio |    |
|    | <220>                                       |    |
|    | <221> LNA                                   |    |
|    | <222> (1)..(3)                              |    |
| 10 | <220>                                       |    |
|    | <221> misc_feature                          |    |
|    | <222> (1)..(6)                              |    |
|    | <223> n es a, c, g o t                      |    |
| 15 | <400> 13                                    |    |
|    | nnnnnn                                      | 6  |
|    | <210> 14                                    |    |
|    | <211> 6                                     |    |
| 20 | <212> ADN                                   |    |
|    | <213> Artificial                            |    |
|    | <220>                                       |    |
|    | <223> oligonucleótido de hexámero aleatorio |    |
| 25 | <220>                                       |    |
|    | <221> misc_feature                          |    |
|    | <222> (1)..(6)                              |    |
|    | <223> n es a, c, g o t                      |    |
| 30 | <400> 14                                    |    |
|    | nnnnnn                                      | 6  |
|    | <210> 15                                    |    |
| 35 | <211> 12                                    |    |
|    | <212> ADN                                   |    |
|    | <213> Artificial                            |    |
|    | <220>                                       |    |
| 40 | <223> oligonucleótido cebador               |    |
|    | <220>                                       |    |
|    | <221> LNA                                   |    |
|    | <222> (1)..(3)                              |    |
| 45 | <220>                                       |    |
|    | <221> misc_feature                          |    |
|    | <222> (7)..(12)                             |    |
| 50 | <223> n es a, c, g o t                      |    |
|    | <400> 15                                    |    |
|    | ggghhhnnnn nn                               | 12 |
| 55 | <210> 16                                    |    |
|    | <211> 12                                    |    |
|    | <212> ADN                                   |    |
|    | <213> Artificial                            |    |
| 60 | <220>                                       |    |
|    | <223> oligonucleótido cebador               |    |
|    | <220>                                       |    |
|    | <221> misc_feature                          |    |
| 65 | <222> (7)..(12)                             |    |
|    | <223> n es a, c, g o t                      |    |

|    |                                                       |    |
|----|-------------------------------------------------------|----|
|    | <400> 16<br>ggghhhnnnn nn                             | 12 |
| 5  | <210> 17<br><211> 23<br><212> ADN<br><213> Artificial |    |
| 10 | <220><br><223> oligonucleótido cebador                |    |
| 15 | <400> 17<br>ggcgtttttt tttttttt ttv                   | 23 |
| 20 | <210> 18<br><211> 21<br><212> ADN<br><213> Artificial |    |
|    | <220><br><223> oligonucleótido cebador                |    |
| 25 | <400> 18<br>ctggatgaat ggcttgagtg t                   | 21 |
| 30 | <210> 19<br><211> 21<br><212> ADN<br><213> Artificial |    |
|    | <220><br><223> oligonucleótido cebador                |    |
| 35 | <400> 19<br>gcaactccac gctcatagaa g                   | 21 |
| 40 | <210> 20<br><211> 25<br><212> ADN<br><213> Artificial |    |
| 45 | <220><br><223> oligonucleótido cebador                |    |
|    | <400> 20<br>ccttcagga ggagttcatg ccagt                | 25 |
| 50 | <210> 21<br><211> 26<br><212> ADN<br><213> Artificial |    |
| 55 | <220><br><223> oligonucleótido cebador                |    |
|    | <400> 21<br>cacacggaga gatgaatgag gggaga              | 26 |
| 60 | <210> 22<br><211> 23<br><212> ADN<br><213> Artificial |    |
| 65 | <220><br><223> oligonucleótido cebador                |    |

|    |                                                       |    |
|----|-------------------------------------------------------|----|
|    | <400> 22<br>gccttcacg ctgtgtcat tga                   | 23 |
| 5  | <210> 23<br><211> 27<br><212> ADN<br><213> Artificial |    |
| 10 | <220><br><223> oligonucleótido cebador                |    |
| 15 | <400> 23<br>catcctgccc ttagaaagt cctctg               | 27 |
| 20 | <210> 24<br><211> 25<br><212> ADN<br><213> Artificial |    |
|    | <220><br><223> oligonucleótido cebador                |    |
| 25 | <400> 24<br>ccagtgtcac aagtcaggt ccatc                | 25 |
| 30 | <210> 25<br><211> 26<br><212> ADN<br><213> Artificial |    |
|    | <220><br><223> oligonucleótido cebador                |    |
| 35 | <400> 25<br>gcggtcagct ttgtccagaa gtgtgt              | 26 |
| 40 | <210> 26<br><211> 25<br><212> ADN<br><213> Artificial |    |
| 45 | <220><br><223> oligonucleótido cebador                |    |
|    | <400> 26<br>gtaagatggt ggatggggtg ggtgt               | 25 |
| 50 | <210> 27<br><211> 25<br><212> ADN<br><213> Artificial |    |
| 55 | <220><br><223> oligonucleótido cebador                |    |
| 60 | <400> 27<br>tcgctctgaa atgctgactc ctca                | 25 |
|    | <210> 28<br><211> 25<br><212> ADN<br><213> Artificial |    |
| 65 | <220><br><223> oligonucleótido cebador                |    |

|    |                                                                           |    |
|----|---------------------------------------------------------------------------|----|
|    | <400> 28<br>accaggttc tactgcgtcc tgtcc                                    | 25 |
| 5  | <210> 29<br><211> 25<br><212> ADN<br><213> Artificial                     |    |
| 10 | <220><br><223> oligonucleótido cebador                                    |    |
| 15 | <400> 29<br>cctccagggc tgcacgttc ttctt                                    | 25 |
| 20 | <210> 30<br><211> 27<br><212> ADN<br><213> Artificial                     |    |
| 25 | <220><br><223> oligonucleótido cebador                                    |    |
| 30 | <400> 30<br>tttttttt tttttttt tttttt                                      | 27 |
| 35 | <210> 31<br><211> 36<br><212> ADN<br><213> Artificial                     |    |
| 40 | <220><br><223> oligonucleótido cebador                                    |    |
| 45 | <220><br><221> misc_feature<br><222> (34)..(36)<br><223> n es a, c, g o t |    |
| 50 | <400> 31<br>tcctacacg acgctctcc gatctgactg gggnnn                         | 36 |
| 55 | <210> 32<br><211> 30<br><212> ADN<br><213> Artificial                     |    |
| 60 | <220><br><223> oligonucleótido cebador                                    |    |
| 65 | <400> 32<br>cagtcagatc ggaagagcgt cgtgtaggga                              | 30 |
|    | <210> 33<br><211> 34<br><212> ADN<br><213> Artificial                     |    |
|    | <220><br><223> oligonucleótido finalizador                                |    |
|    | <220><br><221> misc_feature<br><222> (6)..(9)<br><223> n es a, c, g o t   |    |

|    |                                                                                     |          |
|----|-------------------------------------------------------------------------------------|----------|
|    | <400> 33<br>ggghnnnnna gatcggaaga gcggttcagc agga                                   | 34       |
|    | <210> 34                                                                            |          |
|    | <211> 25                                                                            |          |
| 5  | <212> ADN                                                                           |          |
|    | <213> Artificial                                                                    |          |
|    | <220>                                                                               |          |
| 10 | <223> oligonucleótido finalizador                                                   |          |
|    | <400> 34<br>tcctgctgaa ccgctcttcc gatct                                             | 25       |
|    | <210> 35                                                                            |          |
| 15 | <211> 58                                                                            |          |
|    | <212> ADN                                                                           |          |
|    | <213> Artificial                                                                    |          |
|    | <220>                                                                               |          |
| 20 | <223> oligonucleótido cebador                                                       |          |
|    | <400> 35<br>aatgatacgg cgaccaccga gatctacact cttccctac acgacgctct tccgatct          | 58       |
| 25 | <210> 36                                                                            |          |
|    | <211> 61                                                                            |          |
|    | <212> ADN                                                                           |          |
|    | <213> Artificial                                                                    |          |
| 30 | <220>                                                                               |          |
|    | <223> oligonucleótido cebador                                                       |          |
|    | <400> 36<br>caagcagaag acggcatacg agatcggctc cggcattcct gctgaaccgc tcttccgatac<br>t | 60<br>61 |
|    | <210> 37                                                                            |          |
|    | <211> 34                                                                            |          |
| 40 | <212> ADN                                                                           |          |
|    | <213> Artificial                                                                    |          |
|    | <220>                                                                               |          |
|    | <223> oligonucleótido cebador                                                       |          |
| 45 | <220>                                                                               |          |
|    | <221> misc_feature                                                                  |          |
|    | <222> (29)..(34)                                                                    |          |
|    | <223> n es a, c, g o t                                                              |          |
| 50 | <400> 37<br>tccttacacg acgctcttcc gatctagcnn nnnn                                   | 34       |
|    | <210> 38                                                                            |          |
|    | <211> 34                                                                            |          |
| 55 | <212> ADN                                                                           |          |
|    | <213> Artificial                                                                    |          |
|    | <220>                                                                               |          |
| 60 | <223> finalizador oligonucleotídico                                                 |          |
|    | <220>                                                                               |          |
|    | <221> misc_feature                                                                  |          |
|    | <222> (1)..(9)                                                                      |          |
|    | <223> n es a, c, g o t                                                              |          |
| 65 | <400> 38                                                                            |          |

nnnnnnnnna gatcggaaga gcggttcagc agga

34

## REIVINDICACIONES

1. Método para generar una biblioteca de partes de ácido nucleico amplificadas de un ácido nucleico molde, que comprende

- 5 A) obtener dicho ácido nucleico molde, que es ARN, emparejar al menos un cebador oligonucleotídico con dicho ARN molde, emparejar al menos un finalizador oligonucleotídico y/o cebador adicional con dicho ARN molde, preferentemente en el que el finalizador oligonucleotídico se hibrida con al menos un cebador oligonucleotídico adicional por un par de marcadores de secuencia complementarios en el finalizador y cebador,
- 10 en el que uno cualquiera de los finalizadores oligonucleotídicos comprende un marcador de secuencia o se hibrida con un marcador de secuencia, y uno cualquiera de los cebadores oligonucleotídicos comprende un marcador de secuencia o se hibrida con un marcador de secuencia, y en el que el marcador de secuencia no se empareja con el molde,
- 15 elongar el al menos un cebador oligonucleotídico de una manera específica de molde hasta que el ácido nucleico producto de elongación alcanza la posición de un finalizador oligonucleotídico o cebador adicional emparejado, por lo que se detiene la reacción de elongación, en el que en dicha reacción de elongación dicho finalizador oligonucleotídico opcional no se elonga y/o dicho cebador oligonucleotídico adicional se elonga de una manera específica de molde,
- 20 en el que el ácido nucleico producto elongado se liga con (i) el extremo 5' de dicho finalizador oligonucleotídico que se hibrida con la cadena molde o (ii) el extremo 5' de dicho cebador adicional que se hibrida con la cadena molde o con (iii) un marcador de secuencia que se hibrida con dicho finalizador o cebador adicional en las cercanías del extremo 3' del producto elongado proporcionando el marcador de secuencia hibridado a una región complementaria del finalizador oligonucleotídico o cebador adicional que está unido al extremo 5' de la región hibridada con el molde del finalizador o cebador adicional; o

- 25 B) obtener dicho ácido nucleico molde, que es preferentemente ARN o ADN, emparejar al menos un cebador oligonucleotídico con dicho ácido nucleico molde, emparejar al menos un finalizador oligonucleotídico con dicho ácido nucleico molde, opcionalmente en el que el finalizador oligonucleotídico se hibrida con al menos un cebador oligonucleotídico adicional por un par de marcadores de secuencia complementarios en el finalizador y cebador, en el que uno cualquiera de los finalizadores oligonucleotídicos comprende un marcador de secuencia o se hibrida con un marcador de secuencia, y uno cualquiera de los cebadores oligonucleotídicos comprende un marcador de secuencia o se hibrida con un marcador de secuencia, y en el que el marcador de secuencia no se empareja con el molde,
- 30 elongar el al menos un cebador oligonucleotídico de una manera específica de molde hasta que el ácido nucleico producto de elongación alcanza la posición de un finalizador oligonucleotídico, por lo que se detiene la reacción de elongación, en el que en dicha reacción de elongación dicho finalizador oligonucleotídico no se elonga, y en el que el ácido nucleico producto elongado se liga con (i) el extremo 5' de dicho finalizador oligonucleotídico que se hibrida con la cadena molde o con (ii) un marcador de secuencia que se hibrida con dicho finalizador en las cercanías del extremo 3' del producto elongado proporcionando el marcador de secuencia hibridado a una región complementaria del finalizador oligonucleotídico que está unido al extremo 5' de la región hibridada con el molde del finalizador; o
- 40

- C) obtener dicho ácido nucleico molde, que es preferentemente ARN o ADN, emparejar un primer cebador oligonucleotídico con dicho ácido nucleico molde,
- 45 emparejar al menos un cebador oligonucleotídico adicional con dicho ácido nucleico molde, en el que uno cualquiera de los cebadores oligonucleotídicos y uno cualquiera de los cebadores oligonucleotídicos adicionales comprende un marcador de secuencia o se hibrida con un marcador de secuencia, y en el que el marcador de secuencia no se empareja con el molde,
- 50 elongar dicho primer cebador oligonucleotídico de una manera específica de molde hasta que el ácido nucleico producto de elongación alcanza la posición de uno de dichos cebadores oligonucleotídicos adicionales, por lo que se detiene la reacción de elongación, y se elonga al menos un cebador oligonucleotídico adicional de una manera específica de molde, en el que el ácido nucleico producto elongado detenido se liga con (i) el extremo 5' de dicho cebador oligonucleotídico adicional que se hibrida con la cadena molde o con (ii) un marcador de secuencia que se hibrida con dicho cebador en las cercanías del extremo 3' del producto elongado proporcionando el marcador de secuencia hibridado a una región complementaria del cebador adicional que está unido al extremo 5' de la región hibridada con el molde del cebador adicional;
- 55

en el que en A, B y C el producto de elongación se selecciona o amplifica usando un cebador que hibrida con el marcador de secuencia en el finalizador o cebador adicional.

- 60 2. El método de la reivindicación 1, en el que al menos uno, preferentemente todos, del cebador o los cebadores oligonucleotídicos hibridan con un finalizador oligonucleotídico.

- 65 3. El método según la reivindicación 1 o 2, que comprende además emparejar un cebador oligonucleotídico adicional con dicho ácido nucleico molde y elongar dicho cebador oligonucleotídico adicional hasta que el ácido nucleico producto de elongación alcanza la posición de otro cebador oligonucleotídico o un finalizador oligonucleotídico,

preferentemente en el que se usan 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50 o más cebadores oligonucleotídicos diferentes, especialmente preferido en el que los cebadores oligonucleotídicos son cebadores aleatorios.

4. El método según una cualquiera de las reivindicaciones 1 a 3, que comprende además emparejar un finalizador oligonucleotídico adicional con dicho ácido nucleico molde y en el que en dicha reacción de elongación dicho finalizador oligonucleotídico adicional no se elonga, preferentemente en el que se usan 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50 o más finalizadores oligonucleotídicos diferentes, especialmente preferido en el que los finalizadores oligonucleotídicos son finalizadores aleatorios.

5. El método según una cualquiera de las reivindicaciones 1 a 4, caracterizado porque el marcador de secuencia del finalizador oligonucleotídico y/o los cebadores oligonucleotídicos es un marcador de secuencia uniforme que se une con más de uno de los finalizadores oligonucleotídicos y/o cebadores oligonucleotídicos, preferentemente con todos los finalizadores oligonucleotídicos y/o cebadores oligonucleotídicos, o en el que dicho marcador de secuencia se une con el extremo 3' del finalizador oligonucleotídico, o en el que dicho marcador de secuencia se une con el extremo 5' del cebador oligonucleotídico; preferentemente que comprende además amplificar dichos productos elongados que comprenden dicho marcador ligado, preferentemente por PCR usando cebadores específicos de marcador.

6. El método de una cualquiera de las reivindicaciones 1 a 5, en el que uno cualquiera de los finalizadores oligonucleotídicos y uno cualquiera de los cebadores oligonucleotídicos comprende un marcador de secuencia cada uno, y en el que el marcador de secuencia de los cebadores oligonucleotídicos es al menos parcialmente complementario del marcador de secuencia del finalizador oligonucleotídico permitiendo de este modo la hibridación de los finalizadores oligonucleotídicos y los cebadores oligonucleotídicos entre sí al menos en parte del marcador de secuencia respectivo.

7. El método según una cualquiera de las reivindicaciones 1 a 6, caracterizado porque dicha etapa de marcaje comprende ligamiento con un marcador de secuencia unido con dichos cebadores oligonucleotídicos o finalizadores oligonucleotídicos o comprende ligamiento con dichos cebadores oligonucleotídicos o finalizadores oligonucleotídicos, que comprenden un marcador de secuencia, preferentemente en el que dicho marcador de secuencia se hibrida con el extremo 5' de dicho cebador oligonucleotídico o finalizador oligonucleotídico, preferentemente en el que dicho marcador de secuencia está ligado con el producto de elongación de un cebador cadena arriba.

8. El método según una cualquiera de las reivindicaciones 1 a 7, caracterizado porque dicha etapa de marcaje comprende ligamiento con un cebador oligonucleotídico o finalizador oligonucleotídico que comprende un marcador de secuencia, preferentemente en el que dicho marcador de secuencia se une covalentemente con dicho cebador oligonucleotídico o finalizador oligonucleotídico o en el que el ácido nucleico que comprende dicho marcador de secuencia se hibrida con dicho cebador oligonucleotídico o finalizador oligonucleotídico, preferentemente en el que dicho marcador de secuencia se liga o hibrida con el extremo 3' de un finalizador oligonucleotídico.

9. El método según una cualquiera de las reivindicaciones 1 a 8, en el que un ácido nucleico con una secuencia de complemento inverso para la secuencia de marcador se añade durante la reacción de emparejamiento o elongación o se hibrida con el marcador de secuencia, preferentemente en el que la temperatura de fusión del marcador de secuencia con el ácido nucleico con la secuencia de complemento inverso aumenta en comparación con ácido nucleico no modificado por nucleótidos modificados, preferentemente seleccionados de uno o más nucleótidos de LNA, 2' fluoronucleótidos o nucleótidos de PNA, y/o preferentemente en el que el ácido nucleico con la secuencia de complemento inverso se une covalentemente con el marcador de secuencia, preferentemente mediante un espaciador o un bucle en horquilla.

10. El método según una cualquiera de las reivindicaciones 1 a 9, caracterizado porque los cebadores oligonucleotídicos y/o finalizadores oligonucleotídicos se fosforilan o adenilan en el extremo 5' y/o en el que el ácido nucleico molde se inmoviliza en una fase sólida o un soporte sólido.

11. El método según una cualquiera de las reivindicaciones 1 a 10, que comprende además la etapa de lavar ácidos nucleicos producto de elongación, preferentemente hibridados con un ácido nucleico molde, especialmente hibridados con un ácido nucleico molde inmovilizado según la reivindicación 10 y/o en el que se realizan reacciones de elongación y ligamiento simultáneamente en una etapa de reacción y/o en el que dicho método se realiza mediante la adición por una ADN polimerasa y/o una ligasa, preferentemente en el que dicha ADN polimerasa y ligasa se añaden en una mezcla de reacción con el ácido nucleico molde.

12. El método según una cualquiera de las reivindicaciones 1 a 11, caracterizado porque dicho cebador oligonucleotídico y/o finalizador oligonucleotídico comprende una modificación de nucleótidos que aumenta la T<sub>m</sub> o refuerza la cadena principal de fosfato del azúcar de dicho oligonucleótido, preferentemente seleccionado de 2'fluoro nucleósidos, LNA, ZNA, PNA o usando intercaladores o aditivos que se unen específicamente con ácidos nucleicos, tales como bromuro de etidio, Sybr Green, preferentemente intercaladores que son específicos para híbridos de ARN:ADN, preferentemente en el que dicho cebador oligonucleotídico y/o finalizador oligonucleotídico comprende al

menos uno, preferentemente al menos 2, más preferentemente al menos 3 nucleótidos modificados que se seleccionan de G o C, preferentemente en el extremo 5' de la secuencia de cebador y/o en el que se usa para la elongación una polimerasa que tiene actividad de desplazamiento de cadena de nucleótidos.

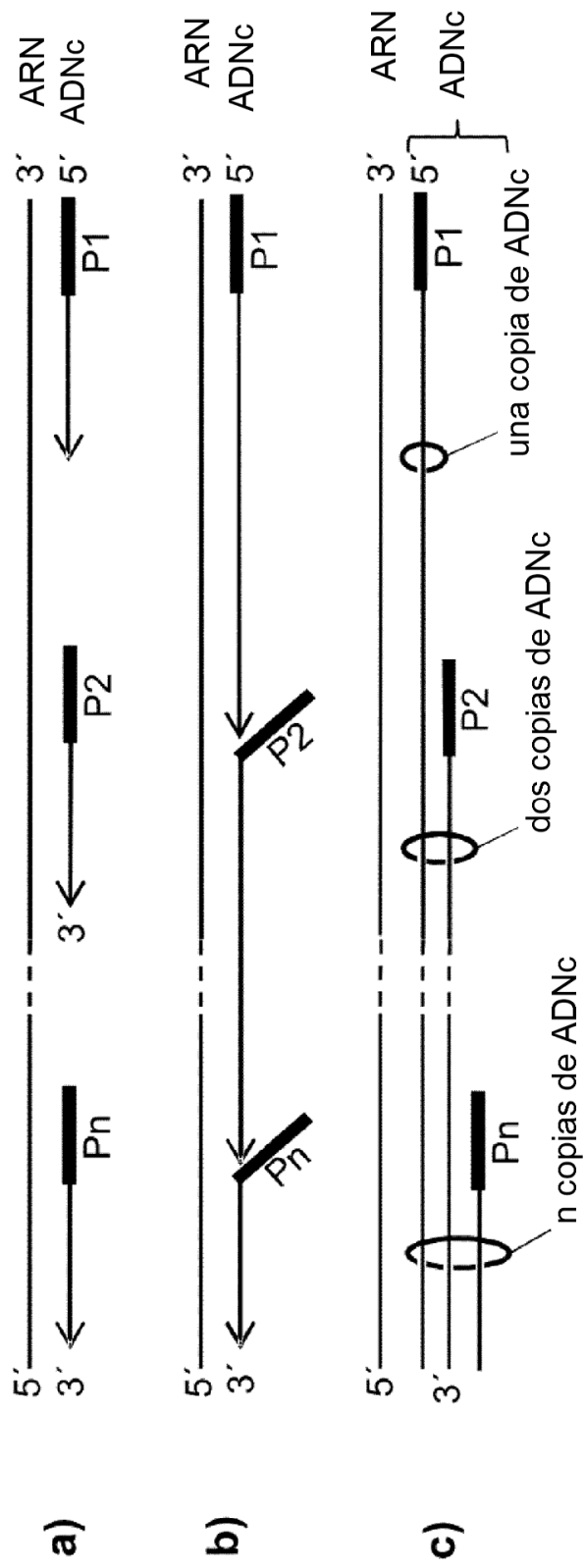
- 5 13. El método según una cualquiera de las reivindicaciones 1 a 12, caracterizado porque la concentración de nucleótidos durante la reacción de polimerización es menor que la concentración de molde, preferentemente la relación molar de la concentración de nucleótidos con respecto a concentración de molde está en el intervalo de 1:33 a 1:3,3 y/o en el que se añade actinomicina D a la reacción de polimerización en cantidades suficientes para evitar síntesis de segunda cadena y para reducir el desplazamiento de cadena de la polimerasa y/o dicho ácido nucleico
- 10 molde es monocatenario y/o en el que el ácido nucleico molde obtenido carece de una cadena complementaria sobre al menos 30 % de su longitud y/o carece de una cadena complementaria de al menos 100 nucleótidos de longitud, preferentemente carece de un ácido nucleico complementario sobre su longitud completa.

- 15 14. El método de una cualquiera de las reivindicaciones 1 a 13 que comprende realizar una mezcla de partes de ácido nucleico amplificadas, preferentemente solapantes, de uno o más ácidos nucleicos molde generando de este modo una biblioteca de secuencias de dichos ácidos nucleicos molde.

#### 15. Kit

- 20 A) para generar una biblioteca de partes de ácido nucleico amplificadas de un ácido nucleico molde según una cualquiera de las reivindicaciones 1 a 13 o para generar una biblioteca de secuencias según la reivindicación 14 que comprende una ADN polimerasa, cebadores oligonucleotídicos aleatorios que comprenden una modificación que aumenta la T<sub>m</sub> y finalizadores oligonucleotídicos aleatorios que son inadecuados para extensión de nucleótidos y comprenden una modificación que aumenta la T<sub>m</sub>, y una ligasa, opcionalmente además uno o más de tampones de
- 25 reacción que comprenden Mn<sup>2+</sup> o Mg<sup>2+</sup>, un agente aglutinante, tal como PEG, en el que uno cualquiera de los finalizadores oligonucleotídicos comprende un marcador de secuencia y uno cualquiera de los cebadores oligonucleotídicos comprende un marcador de secuencia; o
- 30 B) para generar una biblioteca de partes de ácido nucleico amplificadas de un ácido nucleico molde según una cualquiera de las reivindicaciones 1 a 13 o para generar una biblioteca de secuencias según la reivindicación 14 que comprende una ADN polimerasa y una ligasa, cebadores oligonucleotídicos, finalizadores oligonucleotídicos, en el que uno cualquiera de los finalizadores oligonucleotídicos comprende un marcador de secuencia y uno cualquiera de los cebadores oligonucleotídicos comprende un marcador de secuencia, preferentemente además un agente aglutinante, especialmente PEG; o
- 35 C) para generar una biblioteca de partes de ácido nucleico amplificadas de un ácido nucleico molde que contiene a) cebadores oligonucleotídicos aleatorios que comprenden una modificación que aumenta la T<sub>m</sub>, b) finalizadores oligonucleotídicos aleatorios que son inadecuados para extensión de nucleótidos y comprenden una modificación que aumenta la T<sub>m</sub>, y c) una ligasa, preferentemente además que comprende además uno o más de tampones de reacción que comprenden Mn<sup>2+</sup> o Mg<sup>2+</sup>, un agente aglutinante, tal como PEG y/o preferentemente que comprende además una ADN polimerasa y/o una ligasa, en el que uno cualquiera de los finalizadores oligonucleotídicos
- 40 comprende un marcador de secuencia y uno cualquiera de los cebadores oligonucleotídicos comprende un marcador de secuencia.

FIG. 1



**FIG. 2**

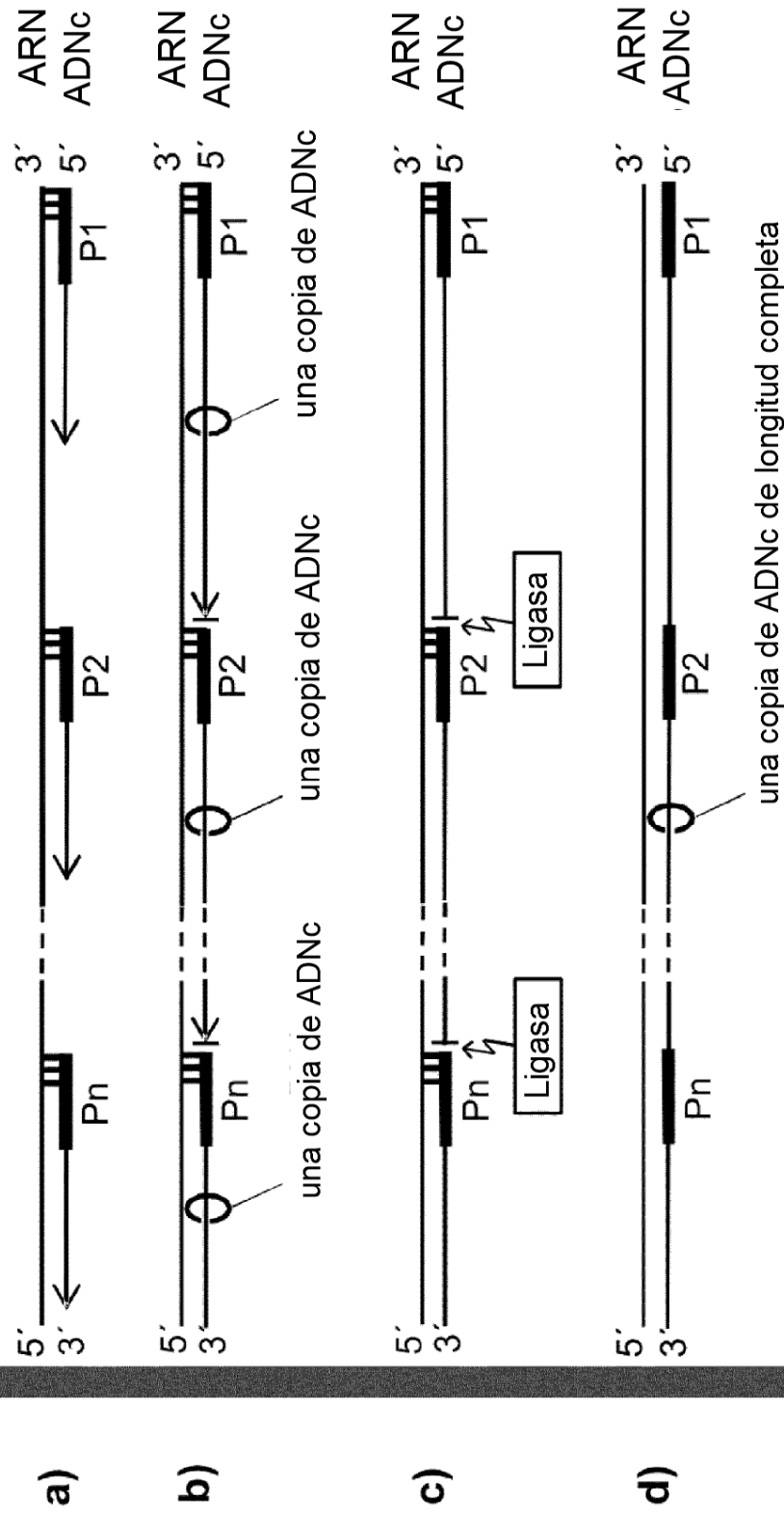
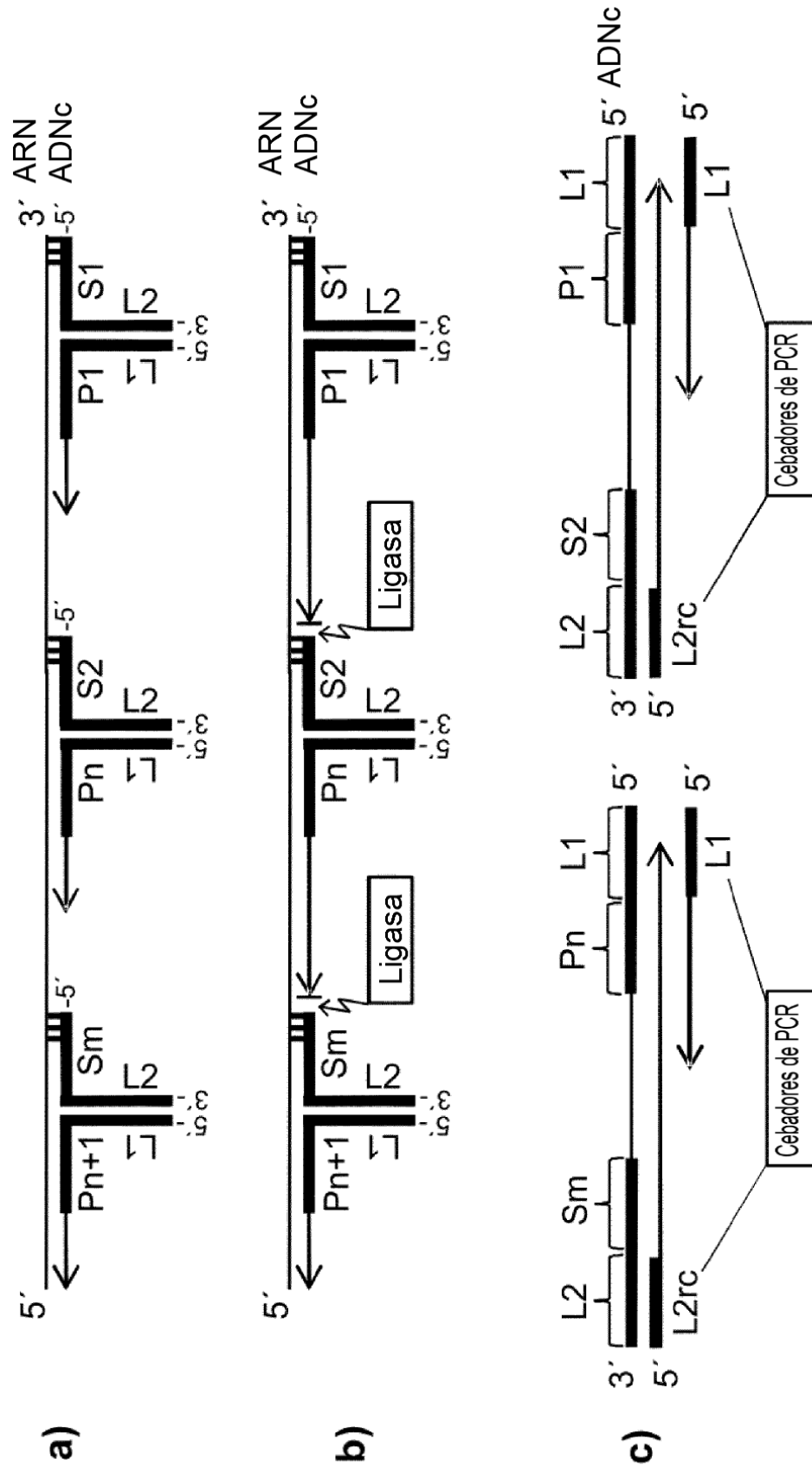


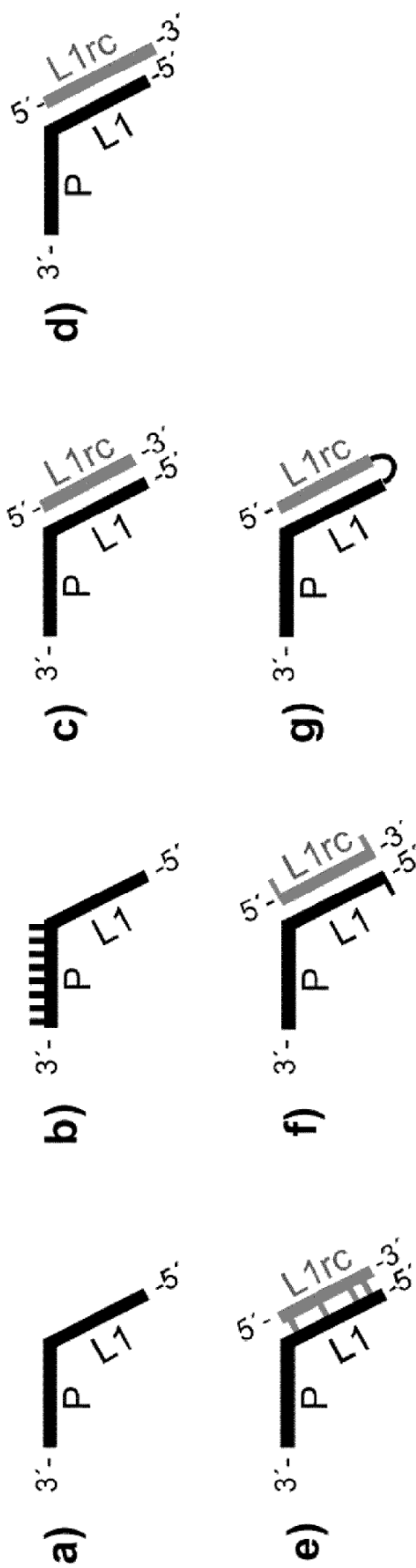




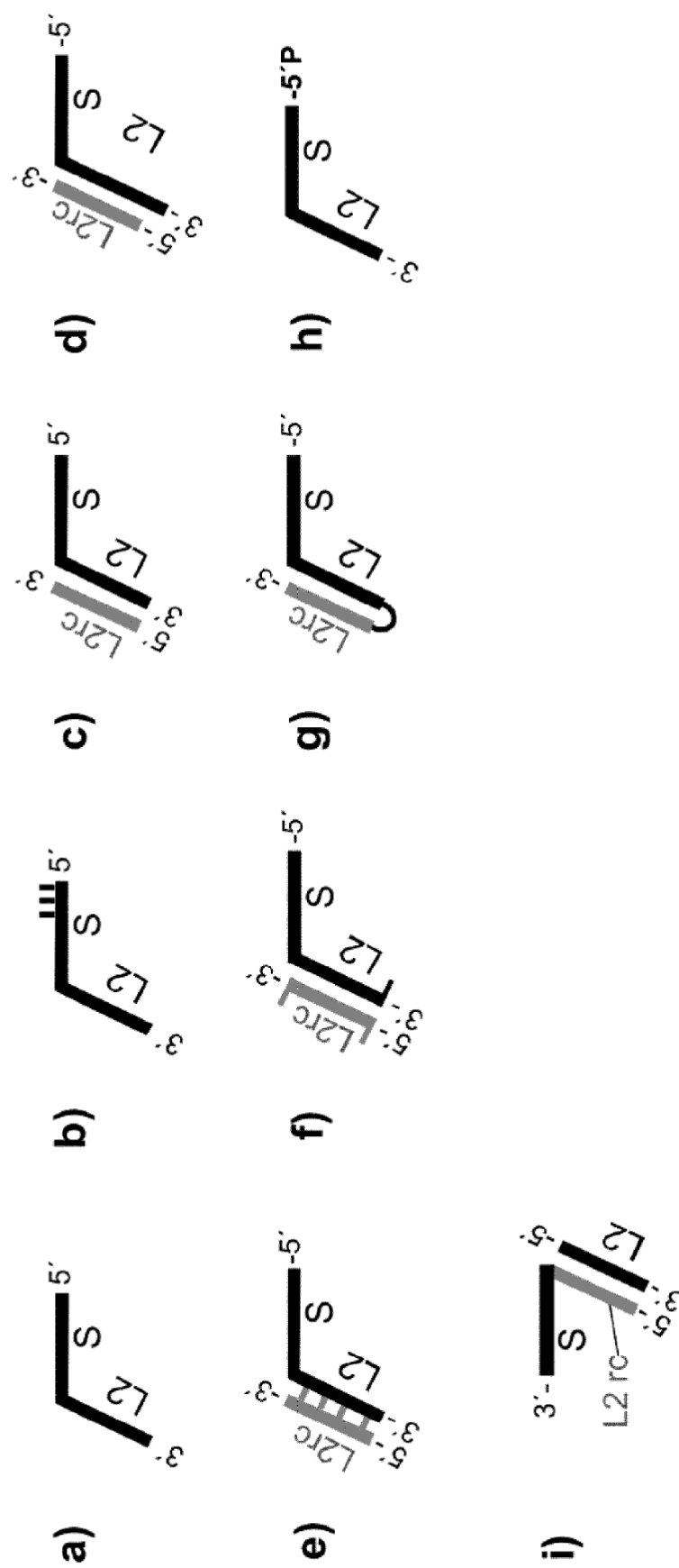
FIG. 5



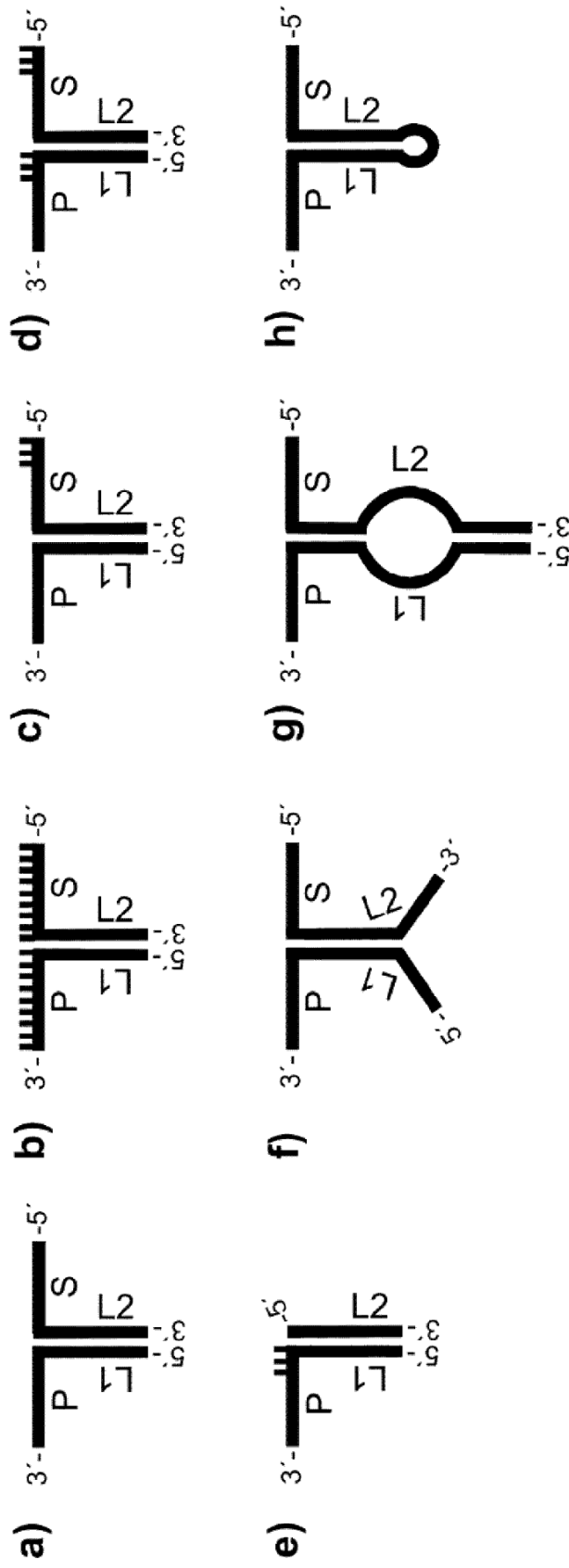
**FIG. 6**



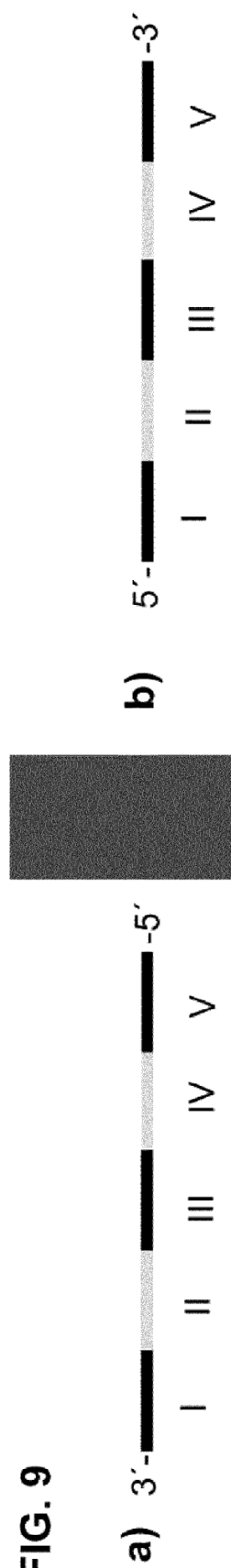
**FIG. 7**



**FIG. 8**

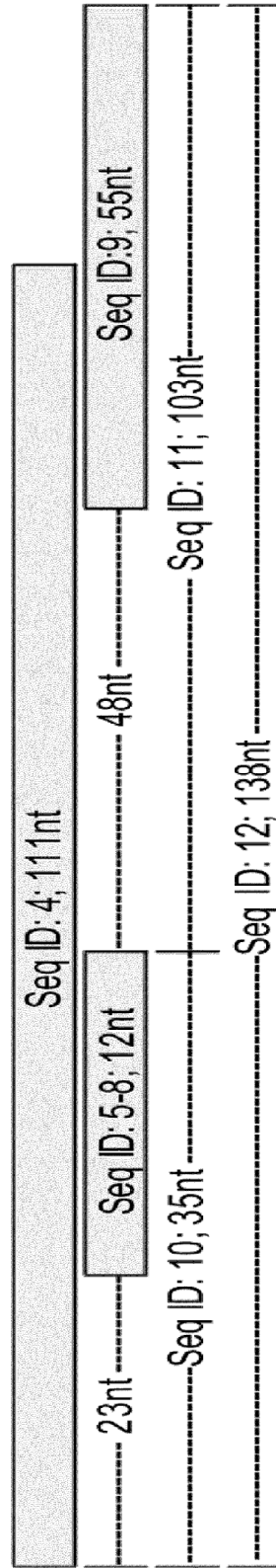


**FIG. 9**

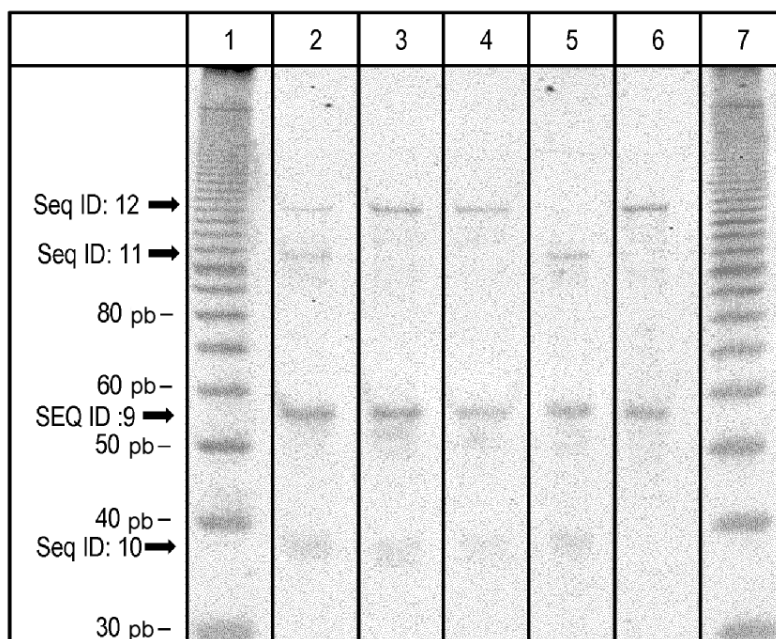


**FIG. 10**

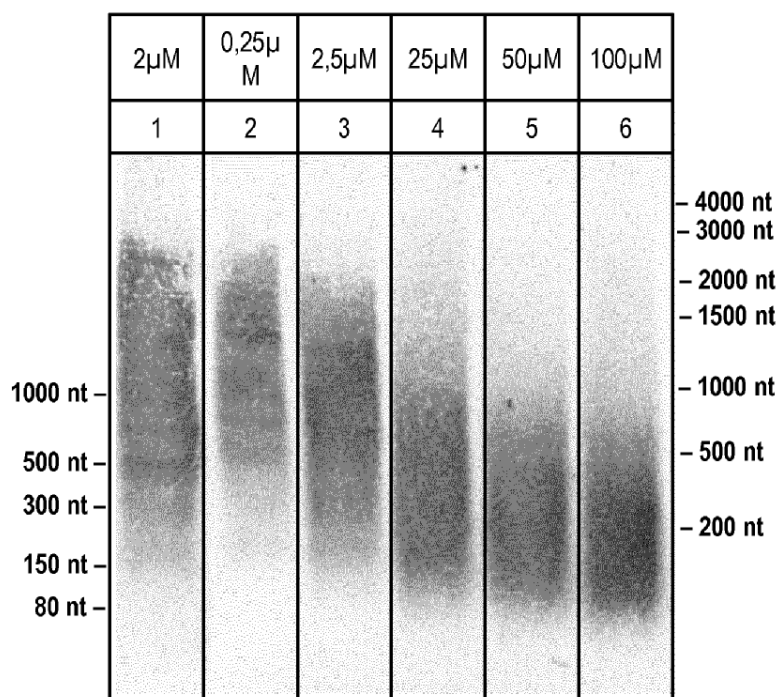
a)



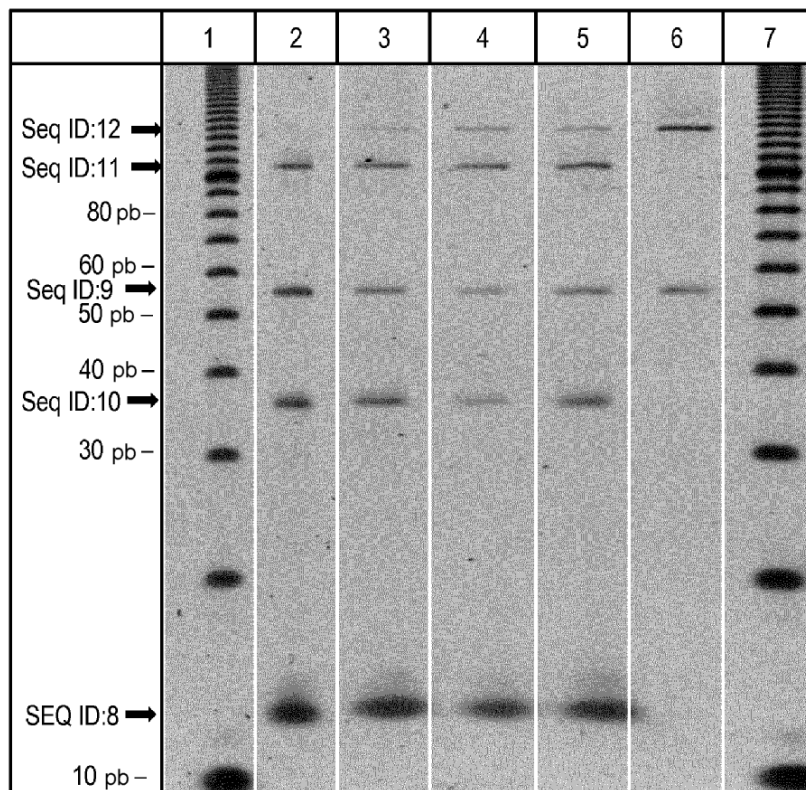
**b)**



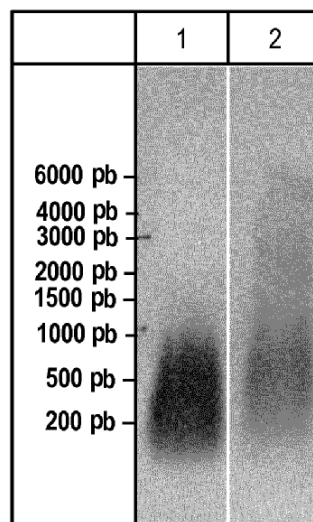
**FIG. 11**



**FIG. 12**



**FIG. 13**



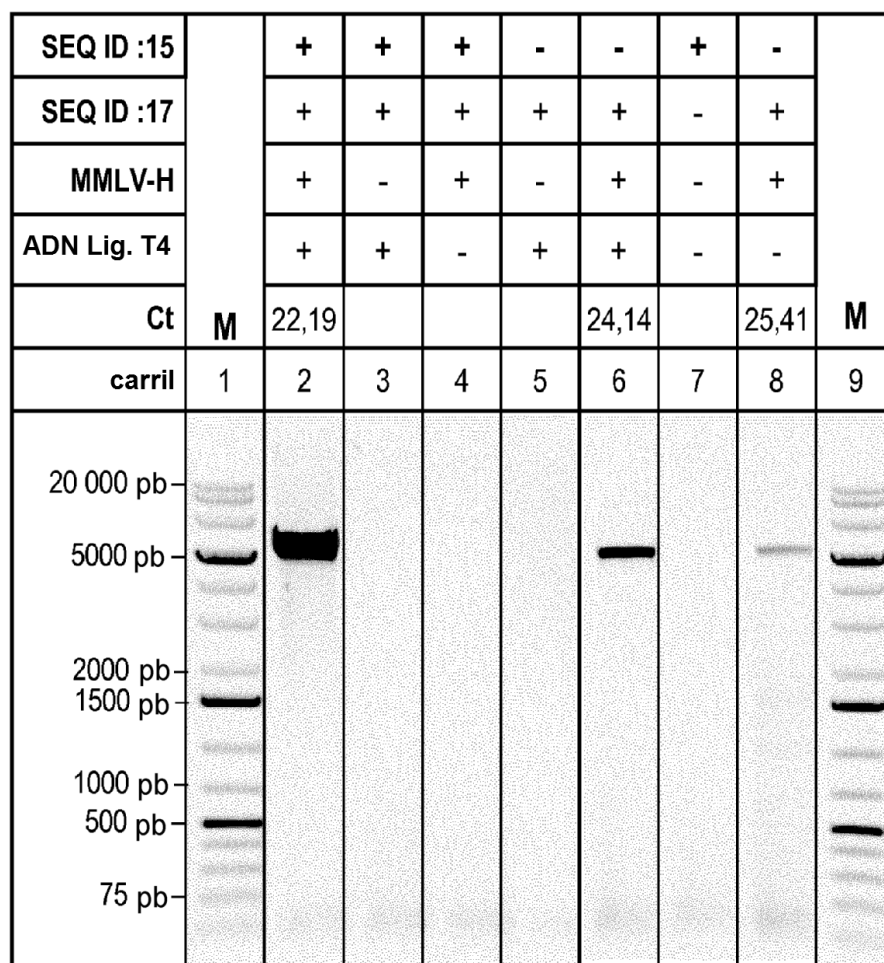
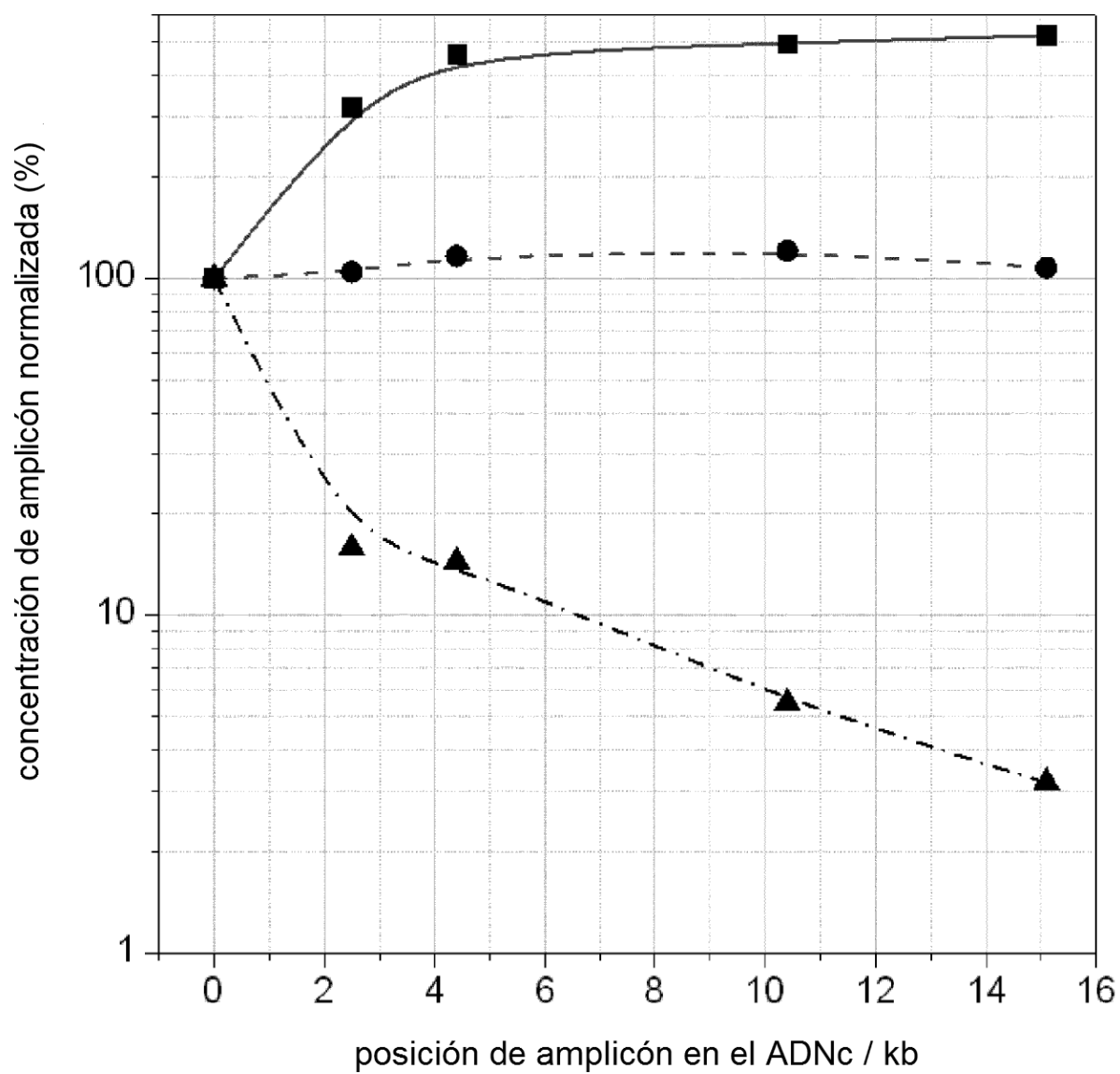
**FIG. 14**

FIG. 15



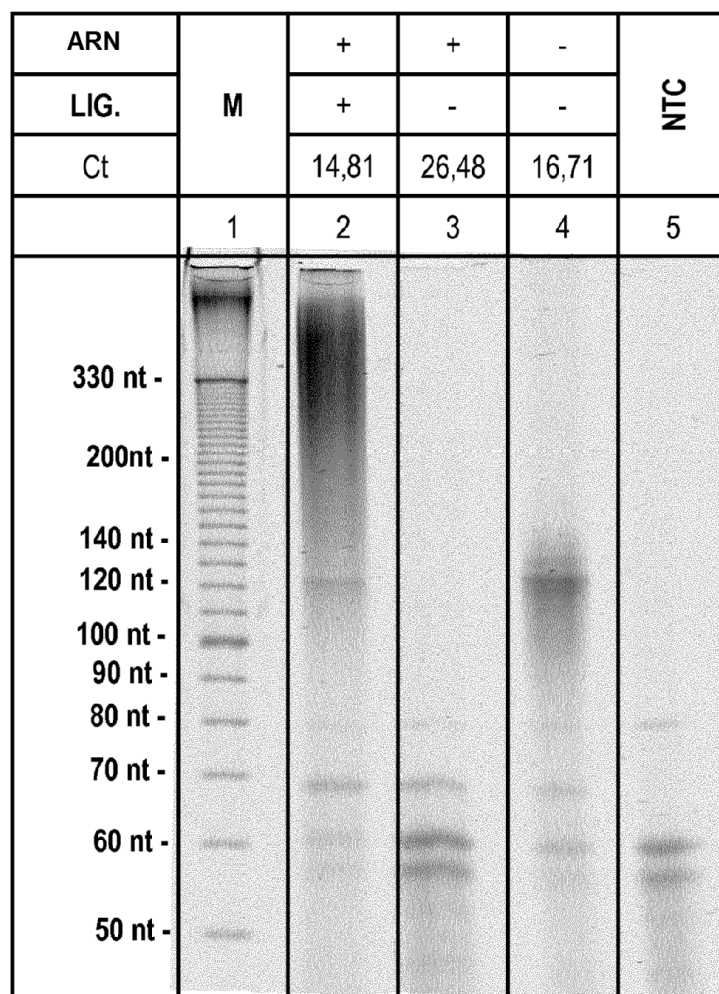
**FIG. 16**

FIG. 17

