

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 675 167**

51 Int. Cl.:

**C12N 15/10** (2006.01)

**C40B 40/06** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **12.07.2013 PCT/US2013/050303**
- 87 Fecha y número de publicación internacional: **16.01.2014 WO14012010**
- 96 Fecha de presentación y número de la solicitud europea: **12.07.2013 E 13816635 (0)**
- 97 Fecha y número de publicación de la concesión europea: **04.04.2018 EP 2872680**

54 Título: **Bibliotecas codificadas por ADN que tienen enlaces oligonucleotídicos codificantes no legibles por polimerasas**

30 Prioridad:

**13.07.2012 US 201261671406 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**09.07.2018**

73 Titular/es:

**X-CHEM, INC. (100.0%)  
100 Beaver Street, Suite 101  
Waltham, Massachusetts 02453, US**

72 Inventor/es:

**KEEFE, ANTHONY, D.;  
LITOVCHICK, ALEXANDER y  
CLARK, MATTHEW, A.**

74 Agente/Representante:

**CARVAJAL Y URQUIJO, Isabel**

ES 2 675 167 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

Bibliotecas codificadas por ADN que tienen enlaces oligonucleotídicos codificantes no legibles por polimerasas

Referencia cruzada con solicitudes relacionadas

5 Esta solicitud reivindica el beneficio de la solicitud provisional de Estados Unidos No. 61/671.406, presentada el 13 de julio de 2012

Antecedentes de la invención

En general, esta invención se refiere a bibliotecas de compuestos codificados por ADN y a métodos para usar y crear tales bibliotecas. La invención también se refiere a composiciones para uso en tales bibliotecas.

10 Las bibliotecas combinatorias codificadas por ADN proporcionan muchos beneficios para el descubrimiento de fármacos. Estas bibliotecas pueden proporcionar una gran cantidad de compuestos diversos que pueden evaluarse e interrogarse rápidamente. Para aumentar aún más la complejidad, se pueden programar y automatizar varias etapas del proceso de descubrimiento. Estas etapas incluyen el uso de la síntesis por división y agrupamiento de múltiples etapas para agregar bloques de construcción a andamios atómicos o poliatómicos y el uso de ligación enzimático y/o químico para agregar etiquetas de oligonucleótidos que codifican las etapas de síntesis como los  
15 bloques de construcción.

A pesar de estos beneficios, pueden surgir numerosos problemas cuando deben sintetizarse y deconvolucionarse bibliotecas muy grandes o complejas. A medida que aumenta el tamaño de la biblioteca, pueden ser necesarios métodos mejorados para proporcionar altos rendimientos de ligación de la etiqueta de oligonucleótido codificante usando metodologías robustas y rápidas. Para crear bibliotecas bajo diversas condiciones de reacción, las  
20 construcciones estables de oligonucleótidos serían beneficiosas, tales como construcciones que son estables en condiciones de pH alto y temperatura elevada. Para simplificar la deconvolución de las etiquetas, la secuencia de las etiquetas podría reconocerse mediante polimerasas dependientes de ADN o ARN, de forma que los datos demográficos de la población de etiquetas se puedan determinar mediante polimerización dependiente de la plantilla y determinación de la secuencia. Pueden surgir dificultades al crear una biblioteca que tenga todos estos atributos  
25 beneficiosos. De acuerdo con esto, existe la necesidad de métodos mejorados, más robustos para seleccionar e identificar pequeños compuestos en bibliotecas codificadas por oligonucleótidos.

Sumario de la invención

La presente invención presenta complejos para usar en bibliotecas codificadas por ADN, donde estos complejos tienen uno o más enlaces para los cuales una polimerasa tiene capacidad reducida para leer o translocarse a través  
30 de ellos. Con el fin de descubrir la identidad de regiones codificadas (por ejemplo, cabeza de dominio, una o más etiquetas, y/o cola de dominio), el enlace se invierte antes de exponerse a la polimerasa o alternativamente es pasado por alto por la polimerasa. Este enlace puede ser pasado por alto de cualquier manera útil que puede no capturar toda la información de la secuencia de las etiquetas (por ejemplo, como en los conectores 5' o 3', como se describe aquí) pero captura la información de la secuencia de codificación de las etiquetas. Dichos enlaces podrían  
35 dar como resultado una menor incidencia de errores de etiquetado, ampliar el número y tipo de enlaces que se utilizarán para crear complejos y bibliotecas, y proporcionar métodos no enzimáticos para crear y seleccionar complejos codificados por ADN. Las ventajas adicionales de estos complejos y métodos se describen en este documento.

De acuerdo con esto, en un aspecto, la invención presenta un complejo que incluye: una entidad química que incluye uno o más andamios o uno o más bloques de construcción; una primera etiqueta de oligonucleótido que codifica la identidad de al menos uno de los uno o más andamios o bloques de construcción; y una cabeza de dominio que tiene un primer grupo funcional y un segundo grupo funcional, donde el primer grupo funcional se asocia operativamente con la entidad química y el segundo grupo funcional se asocia operativamente con la primera  
40 etiqueta a través de un primer enlace para el cual una polimerasa tiene capacidad reducida para leer o translocarse a través de ese enlace.  
45

En otro aspecto, la invención presenta un complejo que incluye: una entidad química que incluye uno o más andamios o uno o más bloques de construcción; un número de etiquetas de oligonucleótidos que tienen n-1 enlaces, donde es un número entero entre 1 y aproximadamente 10, y donde cada uno de los enlaces están entre dos  
50 etiquetas adyacentes y cada etiqueta codifica la identidad de al menos uno de los uno o más andamios o bloques de construcción; y una cabeza de dominio que tiene un primer grupo funcional asociado operativamente con la entidad química y un segundo grupo funcional asociado operativamente con al menos uno del número de n etiquetas a través de un primer enlace, donde una polimerasa tiene capacidad reducida para leer o translocarse a través de al menos uno del primer enlace y n-1 enlaces.

En algunas realizaciones, n está entre 1 y 2, 1 y 3, 1 y 4, 1 y 5, 1 y 6, 1 y 7, 1 y 8, 1 y 9, 1 y 10, 1 y 12, 1 y 15, 1 y 18, 1 y 20, 2 y 3, 2 y 4, 2 y 5, 2 y 6, 2 y 7, 2 y 8, 2 y 9, 2 y 10, 2 y 12, 2 y 15, 2 y 18, 2 y 20, 3 y 4, 3 y 5, 3 y 6, 3 y 7, 3 y 8, 3 y 9, 3 y 10, 3 y 12, 3 y 15, 3 y 18, 3 y 20, 4 y 5, 4 y 6, 4 y 7, 4 y 8, 4 y 9, 4 y 10, 4 y 12, 4 y 15, 4 y 18, o 4 y 20.

5 En algunas realizaciones, la polimerasa tiene una capacidad reducida para leer o translocarse a través de al menos aproximadamente 10% (por ejemplo, aproximadamente 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% o incluso 100%, en comparación con el control) del primer enlace y n-1 enlaces. En realizaciones particulares, la polimerasa tiene una capacidad reducida para leer o translocarse a través de aproximadamente 10% a aproximadamente 100% del primer enlace y n-1 enlaces (por ejemplo, 20% a 100%, 25% a 100%, 50% a 100%, 75% a 100%, 90% a 100%, 95% a 100%, 10% a 95%, 20% a 95%, 25% a 95%, 50% a 95%, 75% a 95%, 90% a 95%, 10% a 90%, 20% a 90%, 25% a 90%, 50% a 90%, o 75% a 90%, en comparación con el control (por ejemplo, en comparación con un oligonucleótido de control que carece del enlace)).

15 En algunas realizaciones, una o más etiquetas incluyen un conector 5' en el extremo 5' de las una o más etiquetas y un conector 3' en el extremo 3' de las una o más etiquetas. En realizaciones particulares, cada conector 5' y/o cada conector 3' incluyen la misma secuencia. En otras realizaciones, cada conector 5' y/o cada conector 3' incluyen una secuencia diferente.

20 En algunas realizaciones, el primer enlace y/o n-1 enlaces tienen al menos aproximadamente 3 angstroms de longitud (por ejemplo, al menos aproximadamente 5, 8, 10, 15, 20, 30, 35, 40, 50, 60, 65 o 70 angstroms). En algunas realizaciones, el primer enlace y/o n-1 enlaces tienen menos de aproximadamente 30 angstroms de longitud (por ejemplo, menos de aproximadamente 25, 20, 15 o 10 angstroms).

En algunas realizaciones, una ADN polimerasa y/o una ARN polimerasa (por ejemplo, como se describe en este documento) tienen una capacidad reducida para leer o transponer a través del primer enlace y/o n-1 enlaces.

25 En algunas realizaciones, menos de aproximadamente 10% (por ejemplo, aproximadamente 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% o 95%) del primer enlace y n-1 enlaces incluyen un enlace enzimático. En algunas realizaciones, el primer enlace y los n-1 enlaces incluyen entre 0% y 90% de enlaces enzimáticos (por ejemplo, aproximadamente 0% a 40%, 0% a 45%, 0% a 50%, 0% a 55%, 0% a 60%, 0% a 65%, 0% a 70%, 0% a 75%, 0% a 80%, 0% a 85%, 0% a 90%, 0% a 95%, 0% a 96%, 0% a 97%, 0% a 98%, 0% a 99%, 5% a 40%, 5% a 45%, 5% a 50%, 5% a 55%, 5% a 60%, 5% a 65%, 5% a 70%, 5% a 75%, 5% a 80%, 5% a 85%, 5% a 90%, 5% a 95%, 5% a 96%, 5% a 97%, 5% a 98%, 5% a 99%, 10% a 40%, 10% a 45%, 10% a 50%, 10% a 55%, 10% a 60%, 10% a 65%, 10% a 70%, 10% a 75%, 10% a 80%, 10% a 85%, 10% a 90%, 10% a 95%, 10% a 96%, 10% a 97%, 10% a 98%, 10% a 99%, 15% a 40%, 15% a 45%, 15% a 50%, 15% a 55%, 15% a 60%, 15% a 65%, 15% a 70%, 15% a 75%, 15% a 80%, 15% a 85%, 15% a 90%, 15% a 95%, 15% a 96%, 15% a 97%, 15% a 98%, 15% a 99%, 20% a 40%, 20% a 45%, 20% a 50%, 20% a 55%, 20% a 60%, 20% a 65%, 20% a 70%, 20% a 75%, 20% a 80%, 20% a 85%, 20% a 90%, 20% a 95%, 20% a 96%, 20% a 97%, 20% a 98%, o 20% a 99%).

40 En algunas realizaciones, el primer enlace y/o n-1 enlaces incluyen un enlace químico (por ejemplo, un grupo químico reactivo, un grupo fotorreactivo, una fracción de intercalación o un oligonucleótido de entrecruzamiento). En realizaciones particulares, al menos uno (por ejemplo, dos, tres, cuatro, cinco o más) grupo químico reactivo, grupo fotorreactivo o fracción de intercalación está presente en un conector 5' en o cerca del extremo 5' de la etiqueta y/o en un conector 3' en o cerca del extremo 3' de la etiqueta. En otras realizaciones, la secuencia de al menos uno del conector 5' es complementaria a la secuencia del conector 3' adyacente o idéntica o suficientemente similar para permitir la hibridación con un oligonucleótido complementario. En algunas realizaciones, al menos 10% (por ejemplo, aproximadamente 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% o incluso 100%) del primer enlace y/o n-1 enlaces son enlaces químicos. En otras realizaciones, aproximadamente 10% a aproximadamente 100% del primer enlace y n-1 enlaces (por ejemplo, del 20% al 100%, del 25% al 100%, del 50% al 100%, del 75% al 100%, del 90% al 100%, del 95% al 100%, del 10% al 95%, del 20% al 95%, del 25% al 95%, del 50% al 95%, del 75% al 95%, del 90% al 95%, del 10% al 90%, del 20% al 90%, del 25% al 90%, del 50% al 90%, o del 75% al 90%) son enlaces químicos.

50 En algunas realizaciones, el grupo químico reactivo se selecciona de un par de un grupo alquilino opcionalmente sustituido y un grupo azido opcionalmente sustituido; un par de un dieno opcionalmente sustituido que tiene un sistema de electrones  $4\pi$  y un dienófilo opcionalmente sustituido o un heterodienófilo opcionalmente sustituido que tiene un sistema de electrones  $2\pi$ ; un par de un nucleófilo y un electrófilo heterocíclico deformado; un par de un grupo amino opcionalmente sustituido y un grupo aldehído o una cetona; un par de un grupo amino opcionalmente sustituido y un grupo ácido carboxílico; un par de una hidracina opcionalmente sustituida y un grupo aldehído o una cetona; un par de una hidroxilamina opcionalmente sustituida y un grupo aldehído o una cetona; un par de un nucleófilo y un haluro de alquilo opcionalmente sustituido; un complejo de platino; un agente alquilante; o un nucleótido modificado con furano (por ejemplo, cualquiera descrito en este documento).

- 5 En algunas realizaciones, el grupo fotorreactivo incluye una fracción de intercalación, un derivado de psoraleno, un grupo cianovinilcarbazol opcionalmente sustituido (por ejemplo, un grupo 3-cianovinilcarbazol, tal como 3-cianovinilcarbazol-1'- $\beta$ -desoxirribósido-5'-trifosfato), un grupo vinilcarbazol opcionalmente sustituido (por ejemplo, un grupo amidovinil carbazol, un grupo carboxivinilcarbazol, o un grupo alcoxicarbonilvinilcarbazol  $C_{2-7}$ , como se describe en la presente memoria), un grupo cianovinilo opcionalmente sustituido, un grupo acrilamida opcionalmente sustituido, un grupo diazirina opcionalmente sustituido, una benzofenona opcionalmente sustituida, o un grupo azida opcionalmente sustituido (por ejemplo, cualquiera descrito en este documento).
- 10 En algunas realizaciones, la fracción de intercalación es un derivado de psoraleno (por ejemplo, psoraleno, 8-metoxipsoraleno o 4-hidroxiometil-4,5,8-trimetil-psoraleno (HMT-psoraleno)), un derivado de alcaloide (por ejemplo, berberina, palmatina, coralina, sanguinarina (por ejemplo, formas de iminio o alcanolamina de los mismos, o aristololactama- $\beta$ -D-glucósido), un catión de etidio (por ejemplo, bromuro de etidio), un derivado de acridina (por ejemplo, proflavina, acriflavina o amsacrina), un derivado de antraciclina (por ejemplo, doxorubicina, epirubicina, daunorubicina (daunomicina), idarrubicina y aclarrubicina) o talidomida.
- 15 En algunas realizaciones, el enlace químico incluye el oligonucleótido de entrecruzamiento, donde la secuencia de al menos cinco nucleótidos en el extremo 5' del oligonucleótido de entrecruzamiento es complementaria a la secuencia de al menos cinco nucleótidos en el extremo 3' de una o más etiquetas o idéntico o suficientemente similar para permitir la hibridación con un oligonucleótido complementario, y donde la secuencia de al menos cinco nucleótidos en el extremo 3' del oligonucleótido de entrecruzamiento es complementaria a la secuencia de al menos cinco nucleótidos en el extremo 5' de una o más etiquetas o idéntico o suficientemente similar para permitir la hibridación con un oligonucleótido complementario. En realizaciones particulares, el extremo 3' de una o más etiquetas incluye un conector 3'. En realizaciones particulares, el extremo 5' de una o más etiquetas incluye un conector 5'.
- 20 En algunas realizaciones, el extremo 5' y/o el extremo 3' del oligonucleótido de entrecruzamiento incluyen un grupo correactivo reversible (por ejemplo, un grupo cianovinilcarbazol, un grupo cianovinilo, un grupo acrilamida, un grupo tiol, o un grupo vinilsulfona, como se describe en este documento).
- 25 En algunas realizaciones, el conector 3' y/o el conector 5' incluyen un grupo correactivo reversible (por ejemplo, un grupo cianovinilcarbazol, un grupo cianovinilo, un grupo acrilamida, un grupo tiol o un grupo vinilsulfona, como se describe en este documento).
- 30 En algunas realizaciones, la entidad química está operativamente asociada a la cabeza de dominio a través de un espaciador bifuncional (por ejemplo, cualquiera descrito en la presente memoria). En otras realizaciones, la entidad química está unida covalentemente a la cabeza de dominio. En realizaciones particulares, la cabeza de dominio incluye un oligonucleótido seleccionado del grupo que consiste en un oligonucleótido de cadena doble, un oligonucleótido de cadena sencilla o un oligonucleótido en horquilla. En algunas realizaciones, la cabeza de dominio incluye una región de unión del cebador.
- 35 En cualquiera de las realizaciones descritas en este documento, el complejo o método incluye además una o más primeras etiquetas identificadoras de bibliotecas, etiquetas de uso y/o etiquetas de origen. En algunas realizaciones, el complejo o método incluye entre 2 y 20 etiquetas (por ejemplo, entre 2 y 10 bloques de construcción o etiquetas de andamio, una primera etiqueta de identificación de biblioteca, una etiqueta de uso opcional y una etiqueta de origen). En algunas realizaciones, el complejo o método incluye de 5 a 75 nucleótidos (por ejemplo, como se describe en la presente memoria, tal como aproximadamente de 40 o 50 nucleótidos).
- 40 En cualquiera de las realizaciones descritas en este documento, cada una de las etiquetas dentro de un conjunto de etiquetas individuales incluye aproximadamente la misma masa.
- En cualquiera de las realizaciones descritas en este documento, el complejo incluye ARN, ADN, ADN modificado y/o ARN modificado (por ejemplo, PNA, LNA, GNA, TNA o una mezcla de los mismos dentro del mismo oligonucleótido, como se describe en este documento).
- 45 En cualquiera de las realizaciones descritas en este documento, el complejo o método incluye además una cola de dominio.
- 50 En otro aspecto, la invención presenta una biblioteca que incluye uno o más complejos descritos en este documento. En algunas realizaciones, la biblioteca incluye una pluralidad de cabezas de dominio. En otras realizaciones, cada cabeza de dominio de la pluralidad de cabezas de dominio incluye una región de secuencia idéntica (por ejemplo, una región de unión del cebador) y una región de codificación diferente (por ejemplo, una primera etiqueta que codifica el uso de la biblioteca, el origen de la biblioteca, la identidad de la biblioteca, el historial de la biblioteca, un enlace, un espaciador o la adición de un primer componente o una secuencia de oligonucleótidos que facilita las tecnologías de hibridación, amplificación o secuenciación). En realizaciones particulares, la biblioteca incluye entre aproximadamente  $10^2$  a  $10^{20}$  complejos (por ejemplo, de aproximadamente  $10^2$  a  $10^3$ ,  $10^2$  a  $10^4$ ,  $10^2$  a  $10^5$ ,  $10^2$  a  $10^6$ ,

5  $10^2$  a  $10^7$ ,  $10^2$  a  $10^8$ ,  $10^2$  a  $10^9$ ,  $10^2$  a  $10^{10}$ ,  $10^2$  a  $10^{11}$ ,  $10^2$  a  $10^{12}$ ,  $10^2$  a  $10^{13}$ ,  $10^2$  a  $10^{14}$ ,  $10^2$  a  $10^{15}$ ,  $10^2$  a  $10^{16}$ ,  $10^2$  a  $10^{17}$ ,  $10^2$  a  $10^{18}$ ,  $10^2$  a  $10^{19}$ ,  $10^4$  a  $10^5$ ,  $10^4$  a  $10^6$ ,  $10^4$  a  $10^7$ ,  $10^4$  a  $10^8$ ,  $10^4$  a  $10^9$ ,  $10^4$  a  $10^{10}$ ,  $10^4$  a  $10^{11}$ ,  $10^4$  a  $10^{12}$ ,  $10^4$  a  $10^{13}$ ,  $10^4$  a  $10^{14}$ ,  $10^4$  a  $10^{15}$ ,  $10^4$  a  $10^{16}$ ,  $10^4$  a  $10^{17}$ ,  $10^4$  a  $10^{18}$ ,  $10^4$  a  $10^{19}$ ,  $10^4$  a  $10^{20}$ ,  $10^5$  a  $10^6$ ,  $10^5$  a  $10^7$ ,  $10^5$  a  $10^8$ ,  $10^5$  a  $10^9$ ,  $10^5$  a  $10^{10}$ ,  $10^5$  a  $10^{11}$ ,  $10^5$  a  $10^{12}$ ,  $10^5$  a  $10^{13}$ ,  $10^5$  a  $10^{14}$ ,  $10^5$  a  $10^{15}$ ,  $10^5$  a  $10^{16}$ ,  $10^5$  a  $10^{17}$ ,  $10^5$  a  $10^{18}$ ,  $10^5$  a  $10^{19}$ , or  $10^5$  to  $10^{20}$  complejos). En algunas realizaciones, cada complejo es diferente.

10 En otro aspecto, la invención presenta un método de etiquetar una primera biblioteca que incluye una entidad química codificada, incluyendo el método: (a) proporcionar una cabeza de dominio que tiene un primer grupo funcional y un segundo grupo funcional (por ejemplo, donde la cabeza de dominio opcionalmente codifica información); (b) unir el primer grupo funcional de la cabeza de dominio a un primer componente de la entidad química, donde la cabeza de dominio se conecta directamente al primer componente o la cabeza de dominio se conecta indirectamente al primer componente mediante un espaciador bifuncional; y (c) unir el segundo grupo funcional de la cabeza de dominio a una primera etiqueta de oligonucleótido a través de un primer enlace para formar un complejo, donde una polimerasa tiene una capacidad reducida para leer o translocarse a través del primer enlace; donde las etapas (b) y (c) pueden realizarse en cualquier orden y donde la primera etiqueta codifica la reacción de unión de la etapa (b), proporcionando de ese modo una biblioteca etiquetada.

20 En otro aspecto más, la invención presenta un método para etiquetar una primera biblioteca que incluye una entidad química codificada, incluyendo el método: (a) proporcionar una cabeza de dominio que tiene un primer grupo funcional y un segundo grupo funcional (por ejemplo, donde la cabeza de dominio opcionalmente codifica información); (b) unir el primer grupo funcional de la cabeza de dominio a un primer componente de la entidad química, donde la cabeza de dominio se conecta directamente al primer componente o la cabeza de dominio se conecta indirectamente al primer componente mediante un espaciador bifuncional; (c) unir el segundo grupo funcional de la cabeza de dominio a una primera etiqueta de oligonucleótido mediante un primer enlace; (d) unir un número  $n_c$  de componentes adicionales de la entidad química, donde  $n_c$  es un número entero entre 1 y 10; y (e) unir un número  $n_t$  de etiquetas de oligonucleótidos adicionales que tienen  $n_t$  enlaces para formar un complejo, donde  $n_t$  es un número entero entre 1 y 10, donde cada uno de los enlaces está entre dos etiquetas adyacentes y donde cada etiqueta codifica la identidad de al menos uno de los componentes; donde una polimerasa tiene una capacidad reducida para leer o translocarse a través de al menos uno del primer enlace y  $n_t$  enlaces; y donde las etapas (b) y (c) se pueden realizar en cualquier orden y donde la primera etiqueta codifica la reacción de unión de la etapa (b); donde las etapas (d) y (e) se pueden realizar en cualquier orden y donde cada etiqueta adicional codifica la reacción de unión de cada componente adicional de la etapa (d), proporcionando de ese modo una biblioteca etiquetada.

30 En algunas realizaciones,  $n_c$  y  $n_t$  son cada uno independientemente un número entero entre 1 y 2, 1 y 3, 1 y 4, 1 y 5, 1 y 6, 1 y 7, 1 y 8, 1 y 9, 1 y 10, 1 y 12, 1 y 15, 1 y 18, 1 y 20, 2 y 3, 2 y 4, 2 y 5, 2 y 6, 2 y 7, 2 y 8, 2 y 9, 2 y 10, 2 y 12, 2 y 15, 2 y 18, 2 y 20, 3 y 4, 3 y 5, 3 y 6, 3 y 7, 3 y 8, 3 y 9, 3 y 10, 3 y 12, 3 y 15, 3 y 18, 3 y 20, 4 y 5, 4 y 6, 4 y 7, 4 y 8, 4 y 9, 4 y 10, 4 y 12, 4 y 15, 4 y 18, o 4 y 20.

35 En algunas realizaciones, la polimerasa tiene una capacidad reducida para leer o translocarse a través de al menos aproximadamente 10% del primer enlace y  $n_t$  enlaces (por ejemplo, aproximadamente 20%, 25%, 30%, 35%, 40%, 45%, 50 %, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% o incluso 100%, en comparación con el control, del primer enlace y  $n_t$  enlaces). En realizaciones particulares, la capacidad reducida para leer o translocarse a través del primer enlace y  $n_t$  enlaces está entre aproximadamente 10 y aproximadamente 100% (por ejemplo, 20% a 100%, 25% a 100%, 50% a 100%, 75% a 100%, 90% a 100%, 95% a 100%, 10% a 95%, 20% a 95%, 25% a 95%, 50% a 95%, 75% a 95%, 90% a 95%, 10 % a 90%, 20% a 90%, 25% a 90%, 50% a 90%, o 75% a 90%, en comparación con el control (por ejemplo, en comparación con un oligonucleótido de control que carece del enlace).

45 En algunas realizaciones, el primer componente y/o los componentes adicionales incluyen un andamio o un bloque de construcción.

En algunas realizaciones, la etapa (b) incluye unir indirectamente la cabeza de dominio al primer componente a través de un espaciador bifuncional.

50 En algunas realizaciones, los métodos incluyen además unir una etiqueta de uso, una etiqueta de origen y/o una primera etiqueta identificadora de biblioteca al complejo. En realizaciones particulares, el método incluye además proporcionar una segunda biblioteca y combinar la primera biblioteca con la segunda biblioteca. En algunas realizaciones, el método incluye además unir una cola de dominio al complejo.

55 En algunas realizaciones, la etapa de unión incluye un grupo químico reactivo, un grupo fotorreactivo, una fracción de intercalación o un oligonucleótido de entrecruzamiento (por ejemplo, como se describe en la presente memoria para cualquiera de los complejos de la presente invención). En realizaciones particulares, el método incluye además realizar síntesis de división y agrupamiento en una o más etapas, separar uno o más complejos, y/o purificar uno o más complejos.

En cualquiera de los métodos descritos en este documento, el método da como resultado cualquiera de los complejos descritos en este documento o cualquiera de las bibliotecas descritas en este documento.

5 En otro aspecto más, la invención presenta un método para seleccionar una pluralidad de entidades químicas, incluyendo el método: (a) poner en contacto un objetivo con cualquier complejo descrito en la presente memoria y/o una biblioteca descrita en este documento; y (b) seleccionar uno o más complejos que tienen una característica predeterminada para el objetivo, en comparación con un control, detectando de ese modo la entidad química.

En algunas realizaciones, la característica predeterminada incluye una mayor unión con el objetivo (por ejemplo, un objetivo biológico, como se describe en este documento), en comparación con un control.

10 En algunas realizaciones, antes de la etapa (a) se realizan una o más operaciones seleccionadas de la lista que consiste en: hibridación de uno o más cebadores de relevo con dicho complejo, donde dichos uno o más cebadores de relevo abarcan dicho primer enlace y/o n-1 enlaces; extender dichos cebadores de relevo usando una polimerasa para producir fragmentos de oligonucleótidos; y ligar dichos fragmentos de oligonucleótidos para producir una plantilla.

15 En otro aspecto, la invención presenta un método para determinar la secuencia de nucleótidos de cualquier complejo descrito en la presente memoria, incluyendo el método: (a) hibridación de uno o más cebadores de relevo con el complejo, donde los uno o más cebadores de relevo abarcan el primer enlace y/o n-1 enlaces; (b) extender los cebadores de relevo usando una polimerasa para producir fragmentos de oligonucleótidos; (c) ligar los fragmentos de oligonucleótidos para producir una plantilla; (d) amplificar opcionalmente la plantilla mediante reacción en cadena de la polimerasa para producir una mezcla amplificada; y (e) secuenciar la mezcla opcionalmente amplificada para  
20 determinar la secuencia del complejo.

25 En algunas realizaciones, el complejo incluye un conector 5' en el extremo 5' de una o más etiquetas y un conector 3' en el extremo 3' de una o más etiquetas, y donde cada uno de los uno o más cebadores de relevo se hibridan con conectores adyacentes 5' y 3'. En realizaciones particulares, cada conector 5' incluye la misma secuencia, y/o cada conector 3' incluye la misma secuencia. En algunas realizaciones, la secuencia de al menos uno del conector 5' es complementaria a la secuencia del conector 3' adyacente (por ejemplo, para formar un dúplex entre los conectores 5' y 3' en condiciones de hibridación) o idéntica o suficientemente similar para permitir la hibridación con un oligonucleótido complementario. En realizaciones adicionales, el método incluye hibridar el conector 5' con el conector 3' adyacente. En algunas realizaciones, cada cebador de relevo para una unión específica (por ejemplo, formando así una unión de tres hélices) entre los conectores 5' y 3' incluye la misma secuencia.

30 En otro aspecto, la invención presenta un método para determinar la secuencia de nucleótidos de cualquier complejo descrito en el presente documento, incluyendo el método: (a) proporcionar el complejo que incluye el enlace químico, donde el enlace químico es un oligonucleótido de entrecruzamiento que extiende la unión entre dos etiquetas adyacentes y que se asocia operativamente con las dos etiquetas adyacentes a través de uno o más grupos correactivos reversibles; (b) liberar el oligonucleótido de entrecruzamiento (por ejemplo, cualquiera descrito  
35 en este documento) para producir una plantilla; (c) amplificar opcionalmente la plantilla mediante reacción en cadena de la polimerasa para producir una mezcla amplificada; y (d) secuenciar la mezcla opcionalmente amplificada para determinar la secuencia del complejo. En algunas realizaciones, una porción del oligonucleótido de entrecruzamiento se hibrida con los extremos de dos etiquetas adyacentes, produciendo de ese modo un corte o un espacio entre las dos etiquetas adyacentes.

40 En realizaciones adicionales, el método incluye ligar las etiquetas usando un proceso químico o un proceso enzimático (por ejemplo, 5'-fosforilación) antes de la etapa (b).

45 En algunas realizaciones, el método incluye además reparar la plantilla antes de la etapa de amplificación opcional o la etapa de secuenciación. En realizaciones particulares, la etapa de reparación incluye la modificación de al menos uno del primer enlace y/o n-1 enlaces para ser un enlace reparado que puede ser leído o translocado a través de una polimerasa. En algunas realizaciones, la polimerasa puede leer o translocarse a través de al menos 50% (por ejemplo, al menos 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% o incluso 100%) de los enlaces reparados (por ejemplo, en comparación con un control, tal como un oligonucleótido de control que tiene el primer enlace y n-1 enlaces que no se han reparado). En realizaciones particulares, la polimerasa puede leer o translocarse a través de aproximadamente 10% a aproximadamente 100% de los enlaces reparados (por ejemplo, 20% a 100%, 25% a 100%, 50% a 100%, 75% a 100%, 90% a 100%, 95% a 100%, 10% a 95%, 20% a 95%, 25% a 95%, 50% a 95%, 75% a 95%, 90% a 95%, 10% a 90%, 20% a 90%, 25% a 90%, 50% a 90%, y 75% a 90%, en comparación con el control (por ejemplo, en comparación con un oligonucleótido de control que tiene el primer enlace y n-1 enlaces que no han sido reparados)). En realizaciones particulares, la etapa de reparación incluye proporcionar la plantilla con una enzima de reparación (por ejemplo, una fotoliasa, una glicosilasa, una endonucleasa, una endonucleasa Flap, una endonucleasaapurínica/apirimidínica (AP), una poli ADP ribosa polimerasa o una metiltransferasa, como se describe en este documento).  
55

En cualquiera de las realizaciones anteriores, los métodos o complejos pueden incluir solo moléculas de cadena sencilla, donde la cabeza de dominio, la primera etiqueta y/o las una o más etiquetas adicionales son de cadena sencilla.

5 En cualquiera de las realizaciones anteriores, el método comprende adicionalmente una o más etapas opcionales para diversificar la biblioteca o interrogar a los miembros de la biblioteca, como se describe en el presente documento. En algunas realizaciones, el método comprende además identificar un pequeño miembro de la biblioteca similar a un fármaco que se une o inactiva una proteína de interés terapéutico. En otras realizaciones, el método comprende además contactar a un miembro de la biblioteca con un objetivo biológico en condiciones adecuadas para que al menos un miembro de la biblioteca se una al objetivo, eliminando uno o más miembros de la biblioteca  
10 que no se unen al objetivo, y analizar las una o más etiquetas de oligonucleótidos asociadas con el objetivo.

Como se describe en este documento, el uso de moléculas de cadena sencilla podría tener numerosos beneficios. Por consiguiente, en cualquiera de las realizaciones descritas en este documento, los métodos y complejos incluyen una cabeza de dominio, una o más etiquetas, un complejo, una entidad química, una molécula o cualquier miembro de una biblioteca etiquetada que tiene una masa disminuida, mayor solubilidad (por ejemplo, en un disolvente orgánico), costo reducido, mayor reactividad, mayor accesibilidad al objetivo, menor radio hidrodinámico, y/o mayor exactitud de evaluaciones analíticas, en comparación con un método que incluye una o más moléculas de cadena doble (por ejemplo, una cabeza de dominio de doble cadena o una etiqueta de cadena doble). En algunas realizaciones, cada una de las etiquetas dentro de un conjunto de etiquetas (por ejemplo, la primera etiqueta o una etiqueta posterior, si está presente) tiene aproximadamente la misma masa (por ejemplo, cada etiqueta tiene una masa que es aproximadamente +/- 10% de la masa promedio entre dos o más etiquetas). En realizaciones particulares, la etiqueta tiene una masa disminuida (por ejemplo, menos de aproximadamente 15.000 Daltons, aproximadamente 14.000 Daltons, aproximadamente 13.000 Daltons, aproximadamente 12.000 Daltons, aproximadamente 11.000 Daltons, aproximadamente 10.000 Daltons, aproximadamente 9.000 Daltons, aproximadamente 8.000 Daltons, aproximadamente 7.500 Daltons, aproximadamente 7.000 Daltons, aproximadamente 6.000 Daltons, aproximadamente 5.500 Daltons, aproximadamente 5.000 Daltons, aproximadamente 4.500 Daltons, o aproximadamente 3.000 Daltons en comparación con una etiqueta de cadena doble (por ejemplo, una etiqueta de cadena doble que tiene una masa de aproximadamente 15.000 Daltons, aproximadamente 14.000 Daltons, aproximadamente 13.000 Daltons, o aproximadamente 12.000 Daltons). En otras realizaciones, la etiqueta tiene una longitud reducida en comparación con una etiqueta de doble cadena (por ejemplo, una etiqueta de doble cadena que tiene una longitud de menos de aproximadamente 20 nucleótidos, menos de aproximadamente 19 nucleótidos, menos de aproximadamente 18 nucleótidos, menos que aproximadamente 17 nucleótidos, menos de aproximadamente 16 nucleótidos, menos de aproximadamente 15 nucleótidos, menos de aproximadamente 14 nucleótidos, menos de aproximadamente 13 nucleótidos, menos de aproximadamente 12 nucleótidos, menos de aproximadamente 11 nucleótidos, menos de aproximadamente 10 nucleótidos, menos de aproximadamente 9 nucleótidos, menos de aproximadamente 8 nucleótidos, o menos de aproximadamente 7 nucleótidos). En algunas realizaciones, una o más etiquetas o miembros de la biblioteca carecen de una región de unión al cebador y/o una región constante (por ejemplo, durante una etapa de selección, tal como una selección usando cromatografía de exclusión por tamaño). En algunas realizaciones, una o más etiquetas o miembros de la biblioteca tienen una región constante reducida (por ejemplo, una longitud inferior a aproximadamente 30 nucleótidos, inferior a aproximadamente 25 nucleótidos, inferior a aproximadamente 20 nucleótidos, inferior a aproximadamente 19 nucleótidos, inferior a aproximadamente 18 nucleótidos, inferior a aproximadamente 17 nucleótidos, inferior a aproximadamente 16 nucleótidos, inferior a aproximadamente 15 nucleótidos, inferior a aproximadamente 14 nucleótidos, inferior a aproximadamente 13 nucleótidos, inferior a aproximadamente 12 nucleótidos, inferior a aproximadamente 11 nucleótidos, inferior a aproximadamente 10 nucleótidos, inferior a aproximadamente 9 nucleótidos, inferior a aproximadamente 8 nucleótidos o inferior a aproximadamente 7 nucleótidos). En otras realizaciones, los métodos incluyen una cabeza de dominio que codifica una molécula, una porción de una entidad química, una reacción de unión (por ejemplo, ligación química o enzimática) de una etapa, o la identidad de una biblioteca, donde la cabeza de dominio codificadora elimina la necesidad de una etiqueta adicional para codificar dicha información.  
50

En cualquiera de las realizaciones anteriores, un oligonucleótido (por ejemplo, la cabeza de dominio, la primera etiqueta y/o una o más etiquetas adicionales, si están presentes) codifica la identidad de la biblioteca. En algunas realizaciones, el oligonucleótido (por ejemplo, la cabeza de dominio, la primera etiqueta y/o una o más etiquetas adicionales, si están presentes) incluye una primera secuencia de identificación de la biblioteca, donde la secuencia codifica la identidad de la primera biblioteca. En realizaciones particulares, el oligonucleótido es una primera etiqueta identificadora de biblioteca. En algunas realizaciones, el método incluye proporcionar una primera etiqueta de identificación de biblioteca, donde la etiqueta incluye una secuencia que codifica para una primera biblioteca y/o unir la primera etiqueta de identificación de biblioteca al complejo. En algunas realizaciones, el método incluye proporcionar una segunda biblioteca y combinar la primera biblioteca con una segunda biblioteca. En realizaciones adicionales, el método incluye proporcionar una segunda etiqueta de identificación de biblioteca, donde la etiqueta incluye una secuencia que codifica para una segunda biblioteca. En algunas realizaciones, se combinan más de dos bibliotecas (por ejemplo, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más bibliotecas).  
60

En cualquiera de las realizaciones anteriores, la información codificada se proporciona en una o más etiquetas o en una combinación de más de una etiqueta. En algunas realizaciones, la información codificada está representada por más de una etiqueta (por ejemplo, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más etiquetas). En algunas realizaciones, la información codificada está representada por más de una etiqueta, donde todas las etiquetas de codificación están contenidas dentro de la secuencia de codificación (por ejemplo, mediante el uso de una combinación de etiqueta específica para codificar información). En algunas realizaciones, la información codificada está representada por más de una etiqueta, donde menos de la totalidad de las etiquetas de codificación están contenidas dentro de la secuencia de codificación (por ejemplo, usando una etiqueta de un conjunto de más de una etiqueta individual para codificar dentro de una secuencia de codificación individual). En algunas realizaciones, la información codificada se representa ortogonalmente, donde la información codificada se representa mediante una combinación de más de una etiqueta con menos de toda la información de codificación contenida dentro de un miembro de biblioteca individual, de modo que más de un miembro de biblioteca correspondiente necesita ser secuenciado para deconvolucionar la información codificada. En algunas realizaciones, más de un bloque de construcción químico está representado por una sola etiqueta (por ejemplo, para un bloque de construcción racémico, tal como dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más bloques de construcción representados por una sola etiqueta).

En cualquiera de las realizaciones anteriores, un oligonucleótido (por ejemplo, una cabeza de dominio y/o uno o más bloques de construcción) codifica para el uso del miembro de la biblioteca (por ejemplo, uso en una etapa de selección o una etapa de enlace, como se describe en la presente memoria). En algunas realizaciones, el oligonucleótido (por ejemplo, la cabeza de dominio, la primera etiqueta y/o una o más etiquetas adicionales, si están presentes) incluye una secuencia de uso, donde la secuencia codifica para el uso de un subconjunto de miembros en la biblioteca en una o más etapas (por ejemplo, una etapa de selección y/o una etapa de unión). En realizaciones particulares, el oligonucleótido es una etiqueta de uso que incluye una secuencia de uso. En algunas realizaciones, un oligonucleótido (por ejemplo, una cabeza de dominio y/o una o más etiquetas de oligonucleótidos) codifica el origen del miembro de la biblioteca (por ejemplo, en una parte particular de la biblioteca). En algunas realizaciones, el oligonucleótido (por ejemplo, la cabeza de dominio, la primera etiqueta y/o una o más etiquetas adicionales, si están presentes) incluye una secuencia de origen (por ejemplo, una secuencia aleatoria o degenerada que tiene una longitud de aproximadamente 10, 9, 8, 7 o 6 nucleótidos), donde la secuencia codifica el origen del miembro en la biblioteca, en comparación con otros miembros de la biblioteca. En realizaciones particulares, el oligonucleótido es una etiqueta de origen que incluye una secuencia de origen. En algunas realizaciones, el método incluye además unir, ligar o asociar operativamente una etiqueta de uso y/o una etiqueta de origen al complejo.

En cualquiera de las realizaciones de la presente memoria, los métodos, composiciones y complejos incluyen opcionalmente una cola de dominio, donde la cola de dominio incluye una o más de una secuencia de identificación de la biblioteca, una secuencia de uso o una secuencia de origen, como se describe en este documento. En realizaciones particulares, los métodos incluyen además unir, enlazar o asociar operativamente la cola de dominio (por ejemplo, que incluye una o más de una secuencia identificadora de la biblioteca, una secuencia de uso o una secuencia de origen) al complejo.

En cualquiera de las realizaciones anteriores, los métodos, composiciones y complejos, o partes de los mismos (por ejemplo, la cabeza de dominio, la primera etiqueta y/o las una o más etiquetas adicionales, si están presentes), incluyen una modificación que soporta solubilidad en condiciones semiacuosas, reducidas en agua o no acuosas (por ejemplo, orgánicas). En algunas realizaciones, el espaciador bifuncional, la cabeza de dominio o una o más etiquetas se modifica para aumentar la solubilidad de un miembro de dicha biblioteca química codificada por ADN en condiciones orgánicas. En algunas realizaciones, la modificación es una o más de una cadena de alquilo, una unidad de polietilenglicol, una especie ramificada con cargas positivas, o una estructura anular hidrófoba. En algunas realizaciones, la modificación incluye uno o más nucleótidos modificados que tienen una fracción hidrófoba (por ejemplo, modificada en las posiciones C5 de las bases T o C con cadenas alifáticas, tal como en 5'-dimetoxitritil-N4-diisobutilaminometilideno-5-(1-propinil)-2'-desoxicidina, 3'-[(2-cianoetil)-(N,N-diisopropil)]-fosforamidita; 5'-dimetoxitritil-5-(1-propinil)-2'-desoxiuridina, 3'-[(2-cianoetil)-(N,N-diisopropil)]-fosforamidita; 5'-dimetoxitritil-5-fluoro-2'-desoxiuridina, 3'-[(2-cianoetil)-(N,N-diisopropil)]-fosforamidita y 5'-dimetoxitritil-5-(piren-1-il-etinil)-2'-desoxiuridina, o 3'-[(2-cianoetil)-(N,N-diisopropil)]-fosforamidita) o una inserción que tiene una fracción hidrófoba (por ejemplo, un azobenceno). En algunas realizaciones, el miembro de la biblioteca tiene un coeficiente de octanol:agua de aproximadamente 1,0 a aproximadamente 2,5 (por ejemplo, aproximadamente 1,0 a aproximadamente 1,5, aproximadamente 1,0 a aproximadamente 2,0, aproximadamente 1,3 a aproximadamente 1,5, aproximadamente 1,3 a aproximadamente 2,0, aproximadamente 1,3 a aproximadamente 2,5, aproximadamente 1,5 a aproximadamente 2,0, aproximadamente 1,5 a aproximadamente 2,5, o aproximadamente 2,0 a aproximadamente 2,5).

En cualquiera de las realizaciones anteriores, la cabeza de dominio, la cola de dominio, la primera etiqueta, las una o más etiquetas adicionales, la etiqueta que identifica la biblioteca, la etiqueta de uso y/o la etiqueta de origen, si está presente, pueden incluir desde aproximadamente 5 a aproximadamente 75 nucleótidos (por ejemplo, de 5 a 7 nucleótidos, de 5 a 8 nucleótidos, de 5 a 9 nucleótidos, de 5 a 10 nucleótidos, de 5 a 11 nucleótidos, de 5 a 12 nucleótidos, de 5 a 13 nucleótidos, de 5 a 14 nucleótidos, de 5 a 15 nucleótidos, de 5 a 16 nucleótidos, de 5 a 17 nucleótidos, de 5 a 18 nucleótidos, de 5 a 19 nucleótidos, de 5 a 20 nucleótidos, de 5 a 30 nucleótidos, de 5 a 40

- nucleótidos, de 5 a 50 nucleótidos, de 5 a 60 nucleótidos, de 5 a 70 nucleótidos, de 6 a 7 nucleótidos, de 6 a 8 nucleótidos, de 6 a 9 nucleótidos, de 6 a 10 nucleótidos, de 6 a 11 nucleótidos, de 6 a 12 nucleótidos, de 6 a 13 nucleótidos, de 6 a 14 nucleótidos, de 6 a 15 nucleótidos, de 6 a 16 nucleótidos, de 6 a 17 nucleótidos, de 6 a 18 nucleótidos, de 6 a 19 nucleótidos, de 6 a 20 nucleótidos, de 7 a 8 nucleótidos, de 7 a 9 nucleótidos, de 7 a 10 nucleótidos, de 7 a 11 nucleótidos, de 7 a 12 nucleótidos, de 7 a 13 nucleótidos, de 7 a 14 nucleótidos, de 7 a 15 nucleótidos, de 7 a 16 nucleótidos, de 7 a 17 nucleótidos, de 7 a 18 nucleótidos, de 7 a 19 nucleótidos, de 7 a 20 nucleótidos, de 8 a 9 nucleótidos, de 8 a 10 nucleótidos, de 8 a 11 nucleótidos, de 8 a 12 nucleótidos, de 8 a 13 nucleótidos, de 8 a 14 nucleótidos, de 8 a 15 nucleótidos, de 8 a 16 nucleótidos, de 8 a 17 nucleótidos, de 8 a 18 nucleótidos, de 8 a 19 nucleótidos, de 8 a 20 nucleótidos, de 9 a 10 nucleótidos, de 9 a 11 nucleótidos, de 9 a 12 nucleótidos, de 9 a 13 nucleótidos, de 9 a 14 nucleótidos, de 9 a 15 nucleótidos, de 9 a 16 nucleótidos, de 9 a 17 nucleótidos, de 9 a 18 nucleótidos, de 9 a 19 nucleótidos, de 9 a 20 nucleótidos, de 10 a 11 nucleótidos, de 10 a 12 nucleótidos, de 10 a 13 nucleótidos, de 10 a 14 nucleótidos, de 10 a 15 nucleótidos, de 10 a 16 nucleótidos, de 10 a 17 nucleótidos, de 10 a 18 nucleótidos, de 10 a 19 nucleótidos, de 10 a 20 nucleótidos, de 10 a 30 nucleótidos, de 10 a 40 nucleótidos, de 10 a 50 nucleótidos, de 10 a 60 nucleótidos, de 10 a 70 nucleótidos, de 10 a 75 nucleótidos, de 11 a 12 nucleótidos, de 11 a 13 nucleótidos, de 11 a 14 nucleótidos, de 11 a 15 nucleótidos, de 11 a 16 nucleótidos, de 11 a 17 nucleótidos, de 11 a 18 nucleótidos, de 11 a 19 nucleótidos, de 11 a 20 nucleótidos, de 12 a 13 nucleótidos, de 12 a 14 nucleótidos, de 12 a 15 nucleótidos, de 12 a 16 nucleótidos, de 12 a 17 nucleótidos, de 12 a 18 nucleótidos, de 12 a 19 nucleótidos, de 12 a 20 nucleótidos, de 13 a 14 nucleótidos, de 13 a 15 nucleótidos, de 13 a 16 nucleótidos, de 13 a 17 nucleótidos, de 13 a 18 nucleótidos, de 13 a 19 nucleótidos, de 13 a 20 nucleótidos, de 14 a 15 nucleótidos, de 14 a 16 nucleótidos, de 14 a 17 nucleótidos, de 14 a 18 nucleótidos, de 14 a 19 nucleótidos, de 14 a 20 nucleótidos, de 15 a 16 nucleótidos, de 15 a 17 nucleótidos, de 15 a 18 nucleótidos, de 15 a 19 nucleótidos, de 15 a 20 nucleótidos, de 16 a 17 nucleótidos, de 16 a 18 nucleótidos, de 16 a 19 nucleótidos, de 16 a 20 nucleótidos, de 17 a 18 nucleótidos, de 17 a 19 nucleótidos, de 17 a 20 nucleótidos, de 18 a 19 nucleótidos, de 18 a 20 nucleótidos, de 19 a 20 nucleótidos, de 20 a 30 nucleótidos, de 20 a 40 nucleótidos, de 20 a 50 nucleótidos, de 20 a 60 nucleótidos, de 20 a 70 nucleótidos, de 20 a 75 nucleótidos, de 30 a 40 nucleótidos, de 30 a 50 nucleótidos, de 30 a 60 nucleótidos, de 30 a 70 nucleótidos, de 30 a 75 nucleótidos, de 40 a 50 nucleótidos, de 40 a 60 nucleótidos, de 40 a 70 nucleótidos, de 40 a 75 nucleótidos, de 50 a 60 nucleótidos, de 50 a 70 nucleótidos, de 50 a 75 nucleótidos, de 60 a 70 nucleótidos, de 60 a 75 nucleótidos, y de 70 a 75 nucleótidos). En realizaciones particulares, la cabeza de dominio, la primera etiqueta, la segunda etiqueta, las una o más etiquetas adicionales, la etiqueta que identifica la biblioteca, la etiqueta de uso y/o la etiqueta de origen, si está presente, tienen una longitud de menos de 20 nucleótidos (por ejemplo, menos de 19 nucleótidos, menos de 18 nucleótidos, menos de 17 nucleótidos, menos de 16 nucleótidos, menos de 15 nucleótidos, menos de 14 nucleótidos, menos de 13 nucleótidos, menos de 12 nucleótidos, menos de 11 nucleótidos, menos que 10 nucleótidos, menos de 9 nucleótidos, menos de 8 nucleótidos o menos de 7 nucleótidos).
- 35 En cualquiera de las realizaciones anteriores, la secuencia de codificación (por ejemplo, la cabeza de dominio, la cola de dominio, la primera etiqueta, las una o más etiquetas adicionales, la etiqueta que identifica la biblioteca, la etiqueta de uso y/o la etiqueta de origen, si está presente) puede incluir más de 20 nucleótidos (por ejemplo, más de 25, 30, 35, 40, 45, 50, 55, 60, 65, 70 o 75 nucleótidos).

#### Definiciones

- 40 Por "nucleótido sustituido en 2'" se entiende una base de nucleótidos que tiene una sustitución en la posición 2' de ribosa.
- Por "aproximadamente" se entiende +/- 10% del valor indicado.
- Por "bifuncional" se entiende que tiene dos grupos reactivos que permiten la unión de dos fracciones químicas.
- 45 Por "espaciador bifuncional" se entiende una fracción espaciadora que tiene dos grupos reactivos que permiten la unión de una entidad química y la información de codificación del complejo. En un ejemplo no limitante, el espaciador bifuncional se proporciona entre la entidad química y una etiqueta. En otro ejemplo no limitativo, el espaciador bifuncional se proporciona entre la entidad química y una cabeza de dominio. Ejemplos de espaciadores bifuncionales se proporcionan en este documento.
- 50 Por "unión" se entiende la unión mediante un enlace covalente o un enlace no covalente. Los enlaces no covalentes incluyen los formados por fuerzas de van der Waals, enlaces de hidrógeno, enlaces iónicos, atrapamiento o encapsulación física, absorción, adsorción y/o otras fuerzas intermoleculares. La unión se puede llevar a cabo por cualquier medio útil, tal como mediante unión enzimática (por ejemplo, ligación enzimática para proporcionar un enlace enzimático) o mediante unión química (por ejemplo, ligación química para proporcionar un enlace químico).
- 55 Por "bloque de construcción" se entiende una unidad estructural de una entidad química, donde la unidad está directamente unida a otras unidades estructurales químicas o está unida indirectamente a través del andamio. Cuando la entidad química es polimérica u oligomérica, los bloques de construcción son las unidades monoméricas del polímero u oligómero. Los bloques de construcción pueden tener uno o más nodos de diversidad que permiten la

adición de uno o más bloques de construcción o andamios. En la mayoría de los casos, cada nodo de diversidad es un grupo funcional capaz de reaccionar con uno o más bloques de construcción o andamios para formar una entidad química. En general, los bloques de construcción tienen al menos dos nodos de diversidad (o grupos funcionales reactivos), pero algunos bloques de construcción pueden tener un nodo de diversidad (o grupo funcional reactivo).

5 Alternativamente, las etapas químicas o de unión codificadas pueden incluir varios componentes químicos (por ejemplo, reacciones de condensación de múltiples componentes o procesos de múltiples etapas). Los grupos reactivos en dos bloques de construcción diferentes deberían ser complementarios, es decir, capaces de reaccionar juntos para formar un enlace covalente o no covalente.

10 Por "entidad química" se entiende un compuesto que comprende uno o más bloques de construcción y, opcionalmente, uno o más andamios. La entidad química puede ser cualquier molécula pequeña o fármaco peptídico o candidato a fármaco diseñado o construido para tener una o más características deseadas, por ejemplo, capacidad para unir un objetivo biológico, solubilidad, disponibilidad de donantes y aceptores de enlaces de hidrógeno, grados de libertad de rotación de los enlaces, carga positiva, carga negativa y similares. En ciertas realizaciones, la entidad química puede hacerse reaccionar adicionalmente como una entidad bifuncional o trifuncional (o mayor).

15 Por "grupo químico reactivo" se entiende un grupo reactivo que participa en una reacción modular, produciendo así un enlace. Los ejemplos de reacciones y los grupos reactivos incluyen los seleccionados de una reacción de cicloadición 1,3-dipolar de Huisgen con un par de un grupo alquínico opcionalmente sustituido y un grupo azido opcionalmente sustituido; una reacción de Diels-Alder con un par de un dieno opcionalmente sustituido que tiene un sistema de electrones  $4\pi$  y un dienófilo opcionalmente sustituido o un heterodienófilo opcionalmente sustituido que tiene un sistema de electrones  $2\pi$ ; una reacción de apertura de anillo con un nucleófilo y un electrófilo heterocíclico deformado; una reacción de ligación de soporte con un grupo fosforotioato y un grupo yodo; y una reacción de aminación reductiva con un grupo aldehído y un grupo amino, como se describe en este documento.

20 Por "complementaria" se entiende una secuencia capaz de hibridar, como se define en este documento, para formar una estructura secundaria (una porción dúplex o de cadena doble de una molécula de ácido nucleico). La complementariedad no necesita ser perfecta, pero puede incluir una o más faltas de correspondencia en uno, dos, tres o más nucleótidos. Por ejemplo, la secuencia complementaria puede contener nucleobases que pueden formar enlaces de hidrógeno de acuerdo con las reglas de apareamiento de bases de Watson-Crick (por ejemplo, G con C, A con T o A con U) u otros motivos de enlace de hidrógeno (por ejemplo, diaminopurina con T, 5-metil C con G, 2-tiotimidina con A, inosina con C, pseudoisocitosina con G). La secuencia y su secuencia complementaria pueden estar presentes en el mismo oligonucleótido o en diferentes oligonucleótidos.

25 Por "complejo" o "complejo ligado" se entiende una cabeza de dominio que se asocia operativamente con una entidad química y/o una o más etiquetas de oligonucleótidos mediante un enlace covalente o un enlace no covalente. El complejo puede incluir opcionalmente un espaciador bifuncional entre la entidad química y la cabeza de dominio.

30 Por "componente" de una entidad química se entiende un andamio o un bloque de construcción.

35 Por "conector" de una etiqueta de oligonucleótido se entiende una porción de la etiqueta en o cerca del extremo 5' o 3' que tiene una secuencia fija. Un conector 5' está ubicado en o cerca del extremo 5' de un oligonucleótido, y un conector 3' está ubicado en o cerca del extremo 3' de un oligonucleótido. Cuando están presentes en un complejo, cada conector 5' puede ser el mismo o diferente, y cada conector 3' puede ser el mismo o diferente. En un complejo a modo de ejemplo no limitante que tiene más de una etiqueta, cada etiqueta puede incluir un conector 5' y un conector 3', donde cada conector 5' tiene la misma secuencia y cada conector 3' tiene la misma secuencia (por ejemplo, cuando la secuencia del conector 5' puede ser la misma o diferente de la secuencia del conector 3'). En otro complejo a modo de ejemplo no limitante, la secuencia del conector 5' está diseñada para ser complementaria, como se define en este documento, a la secuencia del conector 3' (por ejemplo, para permitir la hibridación entre los conectores 5' y 3'). El conector puede incluir opcionalmente uno o más grupos que permiten un enlace (por ejemplo, un enlace para el cual una polimerasa tiene capacidad reducida para leer o translocarse a través de ese enlace, tal como un enlace químico).

40 Por secuencia "constante" o "constante fija" se entiende una secuencia de un oligonucleótido que no codifica información. Las porciones a modo de ejemplo no limitativas de un complejo que tiene una secuencia constante incluyen una región de unión al cebador, un conector 5' o un conector 3'. La cabeza de dominio de la invención puede codificar información (por lo tanto, una etiqueta) o, alternativamente, no codificar información (por lo tanto, una secuencia constante). Del mismo modo, la cola de dominio de la invención puede codificar o no codificar información.

45 Por "oligonucleótido de entrecruzamiento" se entiende un oligonucleótido que se asocia operativamente, como se define en el presente documento, en una unión particular entre dos etiquetas adyacentes en un complejo. En un ejemplo no limitante, un extremo del oligonucleótido de entrecruzamiento se hibrida con el conector 3' de una

primera etiqueta, y el otro extremo del oligonucleótido de entrecruzamiento se hibrida con el conector 5' de una segunda etiqueta que es adyacente a la primera etiqueta. Las realizaciones ilustrativas y no limitantes de oligonucleótidos de entrecruzamiento incluyen aquellas que tienen uno o más grupos reactivos (por ejemplo, un grupo químico reactivo, un grupo fotorreactivo, una fracción de intercalación, o un grupo correactivo reversible, o cualquiera de los descritos aquí) que se asocia operativamente con etiquetas o conectores adyacentes de etiquetas adyacentes.

Por "nodo de diversidad" se entiende un grupo funcional en una posición en el andamio o en el bloque de construcción que permite agregar otro bloque de construcción.

Por "cabeza de dominio" se entiende una estructura química para la síntesis de bibliotecas que está operativamente unida a un componente de una entidad química y a una etiqueta, por ejemplo, un oligonucleótido de partida. Opcionalmente, una cabeza de dominio puede contener unos pocos o ningún nucleótido, pero puede proporcionar un punto en el que puedan estar asociados operativamente. Opcionalmente, un espaciador bifuncional conecta la cabeza de dominio con el componente.

Por "hibridar" se entiende emparejar para formar una molécula de cadena doble entre oligonucleótidos complementarios, o porciones de los mismos, bajo diversas condiciones de rigurosidad. (Véase, por ejemplo, Wahl, GM y SL Berger (1987) *Methods Enzymol.* 152: 399; Kimmel, AR (1987) *Methods Enzymol.* 152: 507). Por ejemplo, se puede obtener una hibridación de alta rigurosidad con una concentración de sal normalmente menor a aproximadamente 750 mM de NaCl y 75 mM de citrato trisódico, menos de aproximadamente 500 mM de NaCl y 50 mM de citrato trisódico, o menos de aproximadamente 250 mM de NaCl y 25 mM de citrato trisódico. Puede obtenerse una hibridación de baja rigurosidad en ausencia de disolvente orgánico, por ejemplo, formamida, mientras que puede obtenerse una hibridación de alta rigurosidad en presencia de al menos aproximadamente 35% de formamida o al menos aproximadamente 50% de formamida. Las condiciones de temperatura de hibridación de alta rigurosidad incluirán normalmente temperaturas de al menos aproximadamente 30°C, 37°C o 42°C. Los expertos en la técnica conocen diversos parámetros adicionales, tales como el tiempo de hibridación, la concentración de detergente, por ejemplo, dodecilsulfato de sodio (SDS) y la inclusión o exclusión del ADN del portador. Se logran diversos niveles de rigurosidad combinando estas diversas condiciones según sea necesario. En una realización, la hibridación tendrá lugar a 30°C en 750 mM de NaCl, 75 mM de citrato trisódico y 1% de SDS. En una realización alternativa, la hibridación tendrá lugar a 37°C en 500 mM de NaCl, 50 mM de citrato trisódico, 1% de SDS, 35% de formamida, y 100 µg/mL de ADN de esperma de salmón (ADNes) desnaturalizado. En una realización alternativa adicional, la hibridación tendrá lugar a 42°C en 250 mM de NaCl, 25 mM de citrato trisódico, 1% de SDS, 50% de formamida, y 200 µg/mL de ADNes. Las variaciones útiles en estas condiciones serán evidentes para los expertos en la técnica.

Para la mayoría de las aplicaciones, las etapas de lavado que siguen a la hibridación también variarán en rigurosidad. Las condiciones de rigurosidad de lavado pueden definirse por concentración de sal y por temperatura. Como se indicó anteriormente, la rigurosidad del lavado se puede aumentar disminuyendo la concentración de sal o aumentando la temperatura. Por ejemplo, las concentraciones de sal de alta rigurosidad para las etapas de lavado pueden ser, por ejemplo, menos de aproximadamente 30 mM de NaCl y 3 mM de citrato trisódico, o menos de aproximadamente 15 mM de NaCl y 1,5 mM de citrato trisódico. Las condiciones de temperatura de alta rigurosidad para las etapas de lavado normalmente incluirán una temperatura de, por ejemplo, al menos aproximadamente 25°C, 42°C o 68°C. En una realización, las etapas de lavado se producirán a 25°C en 30 mM de NaCl, 3 mM de citrato trisódico y SDS al 0,1%. En una realización alternativa, las etapas de lavado se producirán a 42°C en 15 mM de NaCl, 1,5 mM de citrato trisódico y SDS al 0,1%. En una realización alternativa adicional, las etapas de lavado se producirán a 68°C en 15 mM de NaCl, 1,5 mM de citrato trisódico y SDS al 0,1%. Las variaciones adicionales en estas condiciones serán fácilmente evidentes para los expertos en la técnica. Las técnicas de hibridación son bien conocidas por los expertos en la técnica y se describen, por ejemplo, en Benton y Davis (*Science* 196: 180, 1977); Grunstein y Hogness (*Proc. Natl. Acad. Sci., USA* 72: 3961, 1975); Ausubel et al. (*Current Protocols in Molecular Biology*, Wiley Interscience, Nueva York, 2001); Berger y Kimmel (*Guide to Molecular Cloning Techniques*, 1987, Academic Press, Nueva York); y Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Nueva York.

Por "fracción de intercalación" se entiende un grupo reactivo que da como resultado la inclusión de una fracción entre dos o más nucleótidos. En un ejemplo no limitante, la fracción de intercalación reacciona con uno o más nucleótidos para formar enlaces cruzados intercatenario o dentro de la cadena entre oligonucleótidos dúplex o triples. Ejemplos de fracciones de intercalación no limitantes se describen en este documento.

Por "unión" se entiende un corte (falta de un enlace internucleotídico) o un espacio (falta de uno o más nucleótidos) entre dos etiquetas adyacentes en un complejo. La unión también puede ser entre dos conectores adyacentes presentes en dos etiquetas adyacentes (por ejemplo, entre el conector 3' de una primera etiqueta y el conector 5' de una segunda etiqueta que es adyacente a la primera etiqueta).

Por "biblioteca" se entiende una colección de moléculas o entidades químicas. Opcionalmente, las moléculas o entidades químicas están unidas a uno o más oligonucleótidos que codifican las moléculas o porciones de la entidad química.

5 Por "enlace" se entiende una entidad de conexión química que permite asociar operativamente dos o más estructuras químicas, donde el enlace está presente entre la cabeza de dominio y una o más etiquetas, entre dos etiquetas, o entre una etiqueta y una cola de dominio. La entidad de conexión química puede ser un enlace no covalente (por ejemplo, como se describe en la presente memoria), un enlace covalente o un producto de reacción entre dos grupos funcionales. Por "enlace químico" se entiende un enlace formado por una reacción química no enzimática entre dos grupos funcionales. Los ejemplos de grupos funcionales no limitantes incluyen un grupo químico reactivo, un grupo fotorreactivo, una fracción de intercalación o un oligonucleótido de entrecruzamiento (por ejemplo, como se describe en este documento). Por "enlace enzimático" se entiende un enlace internucleotídico o internucleósido formado por una enzima. Las enzimas ilustrativas, no limitantes incluyen una quinasa, una polimerasa, una ligasa o combinaciones de las mismas. Mediante un enlace "para el que una polimerasa tiene capacidad reducida para leer o translocarse a través de ese enlace" se entiende un enlace, cuando está presente en una plantilla de oligonucleótido, que proporciona una cantidad reducida de productos alargados y/o amplificados por una polimerasa, en comparación con un oligonucleótido de control que carece del enlace. Los métodos ilustrativos, no limitantes para determinar tal enlace incluyen la extensión del cebador según se evalúa por análisis de PCR (por ejemplo, PCR cuantitativa), análisis por RT-PCR, cromatografía líquida-espectrometría de masas, datos demográficos de la secuencia u otros métodos. Las polimerasas ilustrativas, no limitantes incluyen ADN polimerasas y ARN polimerasas, tales como ADN polimerasa I, ADN polimerasa II, ADN polimerasa III, ADN polimerasa VI, ADN polimerasa Taq, ADN polimerasa Deep Vent<sup>MR</sup> (ADN polimerasa termófila de alta fidelidad, disponible a través de New England Biolabs), ADN polimerasa T7, ADN polimerasa T4, ARN polimerasa I, ARN polimerasa II, ARN polimerasa III o ARN polimerasa T7.

25 Por "catión multivalente" se entiende un catión capaz de formar más de un enlace con más de un ligando o anión. El catión multivalente puede formar un complejo iónico o un complejo de coordinación. Ejemplos de cationes multivalentes incluyen los de metales alcalinotérreos (por ejemplo, magnesio) y metales de transición (por ejemplo, manganeso (II) o cobalto (III)) y aquellos que están opcionalmente unidos a uno o más aniones y/o uno o más compuestos univalentes o ligandos polidentados, tales como cloruro, amina y/o etilendiamina.

30 Por "oligonucleótido" se entiende un polímero de nucleótidos que tiene un extremo 5', un extremo 3' y uno o más nucleótidos en la posición interna entre los extremos 5' y 3'. El oligonucleótido puede incluir ADN, ARN o cualquier derivado de los mismos conocido en la técnica que pueda sintetizarse y usarse para el reconocimiento de pares de bases. El oligonucleótido no tiene que tener bases contiguas, pero se puede intercalar con fracciones enlazadoras. El polímero oligonucleotídico y el nucleótido (por ejemplo, ADN o ARN modificado) pueden incluir bases naturales (por ejemplo, adenosina, timidina, guanosa, citidina, uridina, desoxiadenosina, desoxitimidina, desoxiguanosina, desoxicitidina, inosina o diamino purina), análogos de bases (por ejemplo, 2-aminoadenosina, 2-tiotimidina, inosina, pirrolo-pirimidina, 3-metil adenosina, C5-propiniluridina, C5-bromouridina, C5-fluorouridina, C5-yodouridina, C5-metilcitidina, 7-desazaadenosina, 7-desazaguanosina, 8-oxoadenosina, 8-oxoguanosina, O(6)-metilguanina y 2-tiocitidina), bases modificadas (por ejemplo, nucleótidos sustituidos en 2', tales como bases 2'-O-metiladas y bases 2'-fluoro), bases intercaladas, azúcares modificados (por ejemplo, 2'-fluororibosa; ribosa; 2'-desoxirribosa; arabinosa; hexosa; anhidrohexitol; alritol; manitol; ciclohexanilo; ciclohexenilo; morfolino que también tiene una cadena principal de fosforamido; ácidos nucleicos bloqueados (LNA, por ejemplo, donde el 2'-hidroxilo de la ribosa está conectado por un alquileo C<sub>1-6</sub> o un puente de heteroalquileo C<sub>1-6</sub> al carbono 4' del mismo azúcar de ribosa, donde los puentes ilustrativos incluyen puentes de metileno, propileno, éter o amino); ácido nucleico de glicol (GNA, por ejemplo, R-GNA o S-GNA, en el que la ribosa se reemplaza por unidades de glicol unidas a enlaces fosfodiéster); ácido nucleico de treosa (TNA, donde la ribosa es reemplazada con α-L-treofuranosil-(3'→2')); y/o reemplazo del oxígeno en la ribosa (por ejemplo, con S, Se o alquileo, como metileno o etileno)), cadenas principales modificadas (por ejemplo, ácido nucleico peptídico (PNA), donde los enlaces 2-amino-etil-glicina reemplazan la cadena principal de ribosa y fosfodiéster), y/o grupos fosfato modificados (por ejemplo, fosforotioatos, 5'-N-fosforamiditas, fosforoselenatos, boranofosfatos, ésteres de boranofosfato, hidrógeno fosfonatos, fosforamidatos, fosforodiamidatos, fosfonatos de alquilo o arilo, fosfotriésteres, fosforamidatos puenteados, fosforotioatos puenteados y metileno-fosfonatos puenteados). El oligonucleótido puede ser de cadena sencilla (por ejemplo, horquilla), de cadena doble, o poseer otras estructuras secundarias o terciarias (por ejemplo, estructuras de tallo-bucle, hélices dobles, triplex, cuádruplex, etc.).

55 Por "enlazada operativamente" o "asociada operativamente" se entiende que dos o más estructuras químicas están unidas directa o indirectamente entre sí de tal manera que permanecen unidas a través de las diversas manipulaciones que se espera que experimenten. Típicamente, la entidad química y la cabeza de dominio están asociadas operativamente de una manera indirecta (por ejemplo, de forma covalente a través de un espaciador apropiado). Por ejemplo, el espaciador puede ser una fracción bifuncional con un sitio de unión para la entidad química y un sitio de unión para la cabeza de dominio.

Por "grupo fotorreactivo" se entiende un grupo reactivo que participa en una reacción causada por la absorción de radiación ultravioleta, visible o infrarroja, produciendo así un enlace. Ejemplos de grupos fotorreactivos no limitantes se describen en este documento.

5 Por "grupo protector" se entiende un grupo destinado a proteger el extremo 3' o el extremo 5' de un oligonucleótido o para proteger uno o más grupos funcionales de la entidad química, andamio o bloque constitutivo frente a reacciones indeseables durante una o más etapas de unión de fabricación, etiquetado o uso de una biblioteca codificada por oligonucleótidos. Los grupos protectores usados comúnmente se describen en Greene, "Protective Groups in Organic Synthesis", 4ª Edición (John Wiley & Sons, Nueva York, 2007).

10 Ejemplos de grupos protectores para oligonucleótidos incluyen grupos protectores irreversibles, tales como didesoxinucleótidos y didesoxinucleósidos (ddNTP o ddN), y, más preferiblemente, grupos protectores reversibles para grupos hidroxilo, tales como grupos éster (por ejemplo, O-( $\alpha$ -metoxietil)éster, O-isovaleril éster, y O-levulinil éster), grupos tritilo (por ejemplo, dimetoxitritilo y monometoxitritilo), grupos xantenilo (por ejemplo, 9-fenilxanten-9-ilo y 9-(p-metoxifenil)xanten-9-ilo), grupos acilo (por ejemplo, fenoxiacetilo y acetilo) y grupos sililo (por ejemplo, t-butildimetilsililo). Ejemplos de grupos protectores no limitantes para entidades químicas, andamios y bloques de construcción incluyen grupos protectores de N para proteger un grupo amino frente a reacciones indeseables durante el procedimiento de síntesis (por ejemplo, grupos acilo, ariloilo, carbamilo, tales como formilo, acetilo, propionilo, pivaloilo, t-butilacetilo, 2-cloroacetilo, 2-bromoacetilo, trifluoroacetilo, tricloroacetilo, ftalilo, o-nitrofenoxiacetilo,  $\alpha$ -clorobutirilo, benzoilo, 4-clorobenzoilo, 4-bromobenzoilo, 4-nitrobenzoilo y auxiliares quirales, como D, L o D,L-aminoácidos no protegidos, tales como alanina, leucina, fenilalanina y similares, grupos que contienen sulfonilo, tales como bencenosulfonilo, p-toluenosulfonilo y similares, grupos que forman carbamato, tales como benciloxicarbonilo, p-clorobenciloxicarbonilo, p-metoxibenciloxicarbonilo, p-nitrobenciloxicarbonilo, 2-nitrobenciloxicarbonilo, p-bromobenciloxicarbonilo, 3,4-dimetoxibenciloxicarbonilo, 3,5-dimetoxibenciloxicarbonilo, 2,4-dimetoxibenciloxicarbonilo, 4-metoxibenciloxicarbonilo, 2-nitro-4,5-dimetoxibenciloxicarbonilo, 3,4,5-trimetoxibenciloxicarbonilo, 1-(p-bifenilil)-1-metiletoxicarbonilo,  $\alpha,\alpha$ -dimetil-3,5-dimetoxibenciloxicarbonilo, benzhidriloxicarbonilo, t-butiloxicarbonilo, diisopropilmetoxicarbonilo, isopropiloxicarbonilo, etoxicarbonilo, metoxicarbonilo, aliloxicarbonilo, 2,2,2-tricloroetoxicarbonilo, fenoxicarbonilo, 4-nitrofenoxicarbonilo, fluorenil-9-metoxicarbonilo, ciclopentiloxicarbonilo, adamantiloxicarbonilo, ciclohexiloxicarbonilo, feniltiocarbonilo y similares; grupos alcarilo, tales como bencilo, trifenilmetilo, benciloximetilo y similares; y grupos sililo tales como trimetilsililo, y similares; donde los grupos protectores de N preferidos son formilo, acetilo, benzoilo, pivaloilo, t-butilacetilo, alaniilo, fenilsulfonilo, bencilo, t-butiloxicarbonilo (Boc) y benciloxicarbonilo (Cbz)); grupos protectores de O para proteger un grupo hidroxilo frente a reacciones indeseables durante el procedimiento de síntesis (por ejemplo, grupos alquilcarbonilo, tales como acilo, acetilo, pivaloilo, y similares; grupos arilcarbonilo opcionalmente sustituidos, tales como benzoilo; grupos sililo, tales como trimetilsililo (TMS), tertbutildimetilsililo (TBDMS), tri-iso-propilsililoximetilo (TOM), triisopropilsililo (TIPS) y similares; grupos formadores de éter con el hidroxilo, tales como metilo, metoximetilo, tetrahidropirano, bencilo, p-metoxibencilo, tritilo y similares; alcoxicarbonilos, tales como metoxicarbonilo, etoxicarbonilo, isopropoxicarbonilo, n-isopropoxicarbonilo, n-butiloxicarbonilo, isobutiloxicarbonilo, sec-butiloxicarbonilo, t-butiloxicarbonilo, 2-etilhexiloxicarbonilo, ciclohexiloxicarbonilo, metiloxicarbonilo, y similares; grupos alcoxialcoxicarbonilo, tales como metoximetoxicarbonilo, etoximetoxicarbonilo, 2-metoxietoxicarbonilo, 2-etoxietoxicarbonilo, 2-butoxietoxicarbonilo, 2-metoxietoximetoxicarbonilo, aliloxicarbonilo, propargiloxicarbonilo, 2-butenoxicarbonilo, 3-metil-2-butenoxicarbonilo y similares; haloalcoxicarbonilos, tales como 2-cloroetoxicarbonilo, 2-cloroetoxicarbonilo, 2,2,2-tricloroetoxicarbonilo y similares; grupos arilalcoxicarbonilo opcionalmente sustituidos, tales como benciloxicarbonilo, p-metilbenciloxicarbonilo, p-metoxibenciloxicarbonilo, p-nitrobenciloxicarbonilo, 2,4-dinitrobenciloxicarbonilo, 3,5-dimetilbenciloxicarbonilo, p-clorobenciloxicarbonilo, p-bromobenciloxicarbonilo y similares; y grupos ariloxicarbonilo opcionalmente sustituidos, tales como fenoxicarbonilo, p-nitrofenoxicarbonilo, o-nitrofenoxicarbonilo, 2,4-dinitrofenoxicarbonilo, p-metil-fenoxicarbonilo, m-metilfenoxicarbonilo, o-bromofenoxicarbonilo, 3,5-dimetilfenoxicarbonilo, p-clorofenoxicarbonilo, 2-cloro-4-nitrofenoxi-carbonilo y similares); grupos protectores de carbonilo (por ejemplo, grupos acetal y cetal, tales como dimetilacetal, 1,3-dioxolano y similares, grupos acilal y grupos ditiano, tales como 1,3-ditianos, 1,3-ditiolano, y similares); grupos protectores de ácido carboxílico (por ejemplo, grupos éster, tales como éster metílico, éster bencílico, éster t-butílico, ortoésteres y similares; grupos sililo, tales como trimetilsililo, así como cualquiera de los descritos en la presente memoria; y grupos oxazolina); y grupos protectores de fosfato (por ejemplo, grupos éster opcionalmente sustituidos, tales como éster metílico, éster isopropílico, éster 2-cianoetílico, éster alílico, éster t-butílico, éster bencílico, éster fluorenilmetilo, éster 2-(trimetilsilil)etilo, éster 2-(metilsulfonil)etilo, éster 2,2,2-tricloroetilo, éster 3',5'-dimetoxibenzoína, éster p-hidroxifenacilo, y similares).

55 Por "proximidad" o "en proximidad" a un extremo de un oligonucleótido se entiende cerca o más cerca del extremo término indicado que el otro extremo restante. Por ejemplo, una fracción o grupo en la proximidad del extremo 3' de un oligonucleótido está cerca o más cerca del extremo 3' que del extremo 5'. En realizaciones particulares, una fracción o grupo en la proximidad del extremo 3' de un oligonucleótido está a uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, quince o más nucleótidos del extremo 3'. En otras realizaciones, una fracción o grupo en la proximidad del extremo 5' de un oligonucleótido está a uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, quince o más nucleótidos del extremo 5'.

Por "purificación" se entiende la eliminación de cualquier producto sin reaccionar o cualquier agente presente en una mezcla de reacción que pueda reducir la actividad de un agente químico o biológico para ser utilizado en una etapa sucesiva. La purificación puede incluir una o más entre separación cromatográfica, separación electroforética y precipitación del producto o reactivo sin reaccionar para ser eliminado.

5 Por "grupo correactivo reversible" se entiende un grupo reactivo que participa en una reacción reversible. Ejemplos de grupos reactivos no limitantes incluyen grupos fotorreactivos, donde la exposición a una radiación de absorción particular da como resultado un enlace entre los grupos fotorreactivos y la exposición a una radiación de absorción particular diferente da como resultado la escisión del enlace formado (por ejemplo, un grupo cianovinilcarbazol, un grupo cianovinilo y un grupo acrilamida). Otro ejemplo de grupo reactivo no limitante incluye grupos reactivos redox, donde tales grupos pueden reducirse u oxidarse reversiblemente (por ejemplo, un grupo tiol).

15 Por "andamio" se entiende una fracción química que muestra uno o más nodos de diversidad en una geometría especial particular. Los nodos de diversidad normalmente se unen típicamente al andamio durante la síntesis de la biblioteca, pero en algunos casos se puede unir un nodo de diversidad al andamio antes de la síntesis de la biblioteca (por ejemplo, adición de uno o más bloques de construcción y/o una o más etiquetas). En algunas realizaciones, el andamio forma un derivado de manera que puede desprotegerse ortogonalmente durante la síntesis de la biblioteca y posteriormente reacciona con diferentes nodos de diversidad.

20 Por fármaco de "molécula pequeña" o candidato a fármaco de "molécula pequeña" se entiende una molécula que tiene un peso molecular por debajo de aproximadamente 1.000 Dalton. Las moléculas pequeñas pueden ser orgánicas o inorgánicas, aisladas (por ejemplo, a partir de bibliotecas de compuestos o fuentes naturales) u obtenidas por formación de derivados de compuestos conocidos.

25 Por "identidad sustancial" o "sustancialmente idéntica" se entiende una secuencia de polipéptido o polinucleótido que tiene la misma secuencia del polipéptido o polinucleótido, respectivamente, como secuencia de referencia, o tiene un porcentaje especificado de residuos de aminoácidos o nucleótidos, respectivamente, que son iguales en la ubicación correspondiente dentro de una secuencia de referencia cuando las dos secuencias están alineadas de manera óptima. Por ejemplo, una secuencia de aminoácidos que es "sustancialmente idéntica" a una secuencia de referencia tiene al menos 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% o 100% de identidad con la secuencia de aminoácidos de referencia. Para los polipéptidos, la longitud de las secuencias de comparación generalmente será de al menos 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 aminoácidos contiguos, más preferiblemente al menos 25, 50, 75, 90, 100, 150, 200, 250, 300 o 350 aminoácidos contiguos, y lo más preferiblemente la secuencia de aminoácidos de longitud completa. Para ácidos nucleicos, la longitud de las secuencias de comparación generalmente será de al menos 5 nucleótidos contiguos, preferiblemente al menos 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25 nucleótidos contiguos, y más preferiblemente la secuencia de nucleótidos de longitud completa. La identidad de secuencia puede medirse usando un software de análisis de secuencia en la configuración predeterminada (por ejemplo, el paquete de software de análisis de secuencia del Genetics Computer Group, Centro de Biotecnología de la Universidad de Wisconsin, 1710 University Avenue, Madison, WI 53705). Tal software puede hacer coincidir secuencias similares al asignar grados de homología a varias sustituciones, eliminaciones y otras modificaciones.

40 Por "etiqueta" o "etiqueta de oligonucleótido" se entiende una porción de oligonucleótido de la biblioteca, al menos parte de la cual codifica información. Los ejemplos no limitantes de tal información incluyen la adición (por ejemplo, mediante una reacción de unión) de un componente (es decir, un andamio o un bloque de construcción, como en una etiqueta de andamio o una etiqueta de bloque de construcción, respectivamente), la cabeza de dominio en la biblioteca, la identidad de la biblioteca (es decir, como en una etiqueta de identidad), el uso de la biblioteca (es decir, como en una etiqueta de uso) y/o el origen de un miembro de la biblioteca (es decir, como en una etiqueta de origen).

45 Por "cola de dominio" se entiende una porción de oligonucleótido de la biblioteca que está unida al complejo después de la adición de todas las etiquetas precedentes y codifica la identidad de la biblioteca, el uso de la biblioteca y/o el origen de un miembro de la biblioteca.

Por "cebador" se entiende un oligonucleótido que es capaz de hibridar con una plantilla de oligonucleótido y luego extenderse mediante una polimerasa de una forma dependiente de la plantilla.

50 Por "cebador de relevo" se entiende un oligonucleótido que es capaz de hibridar con una plantilla de oligonucleótido que contiene, en la región de la plantilla con la que se hibrida el cebador, al menos un enlace internucleotídico que reduce la capacidad de una polimerasa para leer o translocarse a través de ese enlace. Tras la hibridación, uno o más cebadores de relevo permiten la extensión por una polimerasa de una forma dependiente de la plantilla.

55 Por "recombinación", como se usa en la presente memoria, se entiende la generación de un producto de polimerasa como resultado de al menos dos eventos de hibridación distintos.

Otras características y ventajas de la invención serán evidentes a partir de la siguiente descripción detallada y las reivindicaciones.

#### Breve descripción de los dibujos

5 Las Figuras 1A-1B muestran complejos a modo de ejemplo. La Figura 1A muestra un complejo (en la dirección 5' a 3') que tiene una entidad química (estrella), una secuencia constante fija (por ejemplo, una cabeza de dominio), seguida de tres secuencias de codificación variables (por ejemplo, tres etiquetas) y otra secuencia constante fija (por ejemplo, una cola de dominio), donde el enlace entre las secuencias constantes fijas y las secuencias de codificación variables son legibles. El complejo puede extenderse más usando una polimerasa con un cebador (hibridado con el extremo 3' del complejo), opcionalmente amplificado con PCR, y secuenciado. La Figura 1B muestra un complejo que tiene enlaces ilegibles. El complejo (en la dirección 5' a 3') tiene una entidad química (estrella), una secuencia constante fija (por ejemplo, una cola de dominio), seguida de tres secuencias de codificación variables (por ejemplo, tres etiquetas) y otra secuencia constante fija (por ejemplo, una cola de dominio), donde el enlace entre las secuencias constantes fijas y las secuencias de codificación variables son ilegibles. Cada una de las secuencias de codificación variables incluye un conector 5' y un conector 3' que son secuencias constantes fijas. Debido a que los enlaces son ilegibles, los cebadores de relevo se usan para abarcar los enlaces y para permitir la extensión con una polimerasa, formando así fragmentos de oligonucleótidos. A continuación, los fragmentos se ligaron usando ligasa, opcionalmente se amplificaron con PCR, y se secuenciaron. El proceso de lectura (por ejemplo, extensión opcional, ligación, amplificación opcional y secuenciación) se puede opcionalmente realizar a través de ese enlace después de la selección. Alternativamente, la extensión opcional y la parte de ligación del proceso de lectura pueden realizarse antes de la selección.

La Figura 2 muestra un complejo ilustrativo que tiene un enlace ilegible formado por psoraleno. El esquema proporciona (de la dirección 5' a 3') una secuencia de codificación variable, un conector 3', un conector 5' y otra secuencia de codificación variable. El conector 3' y el conector 5' se hibridan para formar un dúplex, y el psoraleno (una fracción fotorreactiva de intercalación) forma un enlace ilegible entre los conectores 5' y 3'. El cebador de relevo se hibrida con los conectores 5' y 3' y abarca el enlace ilegible. Las porciones legibles del complejo (por ejemplo, la secuencia de codificación variable) se extienden con una polimerasa para formar fragmentos de oligonucleótidos. Los fragmentos y los cebadores de relevo se ligan utilizando ligasa, opcionalmente amplificada con PCR, y secuenciada. El proceso de lectura (por ejemplo, extensión, ligación, amplificación opcional y secuenciación) puede realizarse opcionalmente por completo después de la selección. Alternativamente, la parte de extensión y ligación del proceso de lectura se puede realizar antes de la selección.

La Figura 3 muestra un complejo ilustrativo que tiene enlaces ilegibles formados por un oligonucleótido de entrecruzamiento. En el esquema superior, el complejo incluye (de la dirección 5' a 3') una entidad química (estrella), una secuencia constante fija (por ejemplo, una cabeza de dominio), un oligonucleótido de entrecruzamiento que está entrecruzado a través de un grupo correactivo reversible (por ejemplo, grupos cianovinilcarbazol entrecruzados a aproximadamente 366 nm, enlaces cruzados marcados con X), y seguido por una secuencia de codificación variable que tiene secuencias constantes fijas en los extremos 5' y 3' (por ejemplo, una etiqueta que tiene conectores 5' y 3'). La combinación del oligonucleótido de entrecruzamiento y la secuencia de codificación variable se repiten tres veces más. Luego, sigue el oligonucleótido de entrecruzamiento final y otra secuencia constante fija (por ejemplo, una cola de dominio). Los enlaces cruzados se marcan con X. A continuación, el complejo se hace reaccionar con un proceso químico y/o un proceso enzimático (por ejemplo, una ligasa y opcionalmente una quinasa) para ligar las etiquetas. A continuación, los oligonucleótidos de entrecruzamiento se liberan para formar una plantilla (por ejemplo, usando absorción a aproximadamente 312 nm). Finalmente, la plantilla se amplifica opcionalmente con PCR y se secuencia. El proceso de lectura (por ejemplo, etapa de quinasa opcional, ligación, amplificación opcional y secuenciación) se puede opcionalmente realizar a través de ese enlace después de la selección. Alternativamente, la etapa de quinasa opcional y la etapa de ligación del proceso de lectura pueden realizarse antes de la selección.

La Figura 4 muestra una reacción reversible ilustrativa con grupos correactivos reversibles (un par de un grupo cianovinilcarbazol y una timidina) y el uso de estos grupos permite la formación de enlaces cruzados entre el oligonucleótido de entrecruzamiento y la secuencia constante fija de la etiqueta de oligonucleótido. Los grupos correactivos reversibles (marcados con X y X') están presentes en el conector 3', en el conector 5' y en el oligonucleótido de entrecruzamiento.

La Figura 5 muestra el oligonucleótido CNVKJ, que contiene dos modificaciones de cianovinilcarbazol, CNVK2\_P\_TagB, CNVK\_2\_NP\_TagB y CNVK2\_TagA.

La Figura 6 muestra los resultados de la electroforesis de los productos de entrecruzamiento fotoquímico de los oligonucleótidos CNVKJ, CNVK2\_P\_TagB y CNVK2\_TagA.

55 La Figura 7 muestra la LCMS de CNVK2\_P\_TagB purificado y CNVK2\_TagA.

- La Figura 8 muestra el LCMS de la recuperación de la Double\_CNVK\_template.
- La Figura 9 muestra la LCMS del tratamiento de un complejo de CNVK2\_P\_TagB, CNVK2\_TagA y CNVKJ con ADN ligasa de T4.
- 5 La Figura 10 muestra los resultados de LCMS de la reacción de fosforilación y ligación del par de etiquetas modelo no fosforiladas entrecruzadas fotoquímicamente CNVK2\_NP\_TagB con CNVK2\_TagA y CNVKJ.
- La Figura 11 muestra el oligonucleótido 5PSO2\_A9\_TA, modificado con C2-psoraleno en el extremo 5' y el oligonucleótido PSO\_HP\_A9\_TCT.
- 10 La Figura 12 muestra la LCMS del entrecruzamiento fotoquímico de oligonucleótidos utilizando psoraleno dentro de un tallo para modelar un evento de marcación del oligonucleótido 5PSO2\_A9\_TA, modificado con C2-psoraleno en el extremo 5' y el oligonucleótido PSO\_HP\_A9\_TCT.
- La Figura 13 muestra los oligonucleótidos Tag1\_PsoCVU y SplintC\_PsoC2 y el oligonucleótido 5PSOC2\_A9\_GA (modificado en el extremo 5' con C2-psoraleno).
- La Figura 14 muestra un esquema de fotoligación de psoraleno.
- La Figura 15 muestra un gel del transcurso del tiempo de fotoligación de psoraleno.
- 15 La Figura 16 muestra la LCMS del conjugado fotoligado de 5PsoC2\_A9\_GA y Tag1\_PsoCVU.
- La Figura 17 muestra un oligonucleótido biotinilado en 5' 5Bio\_Tag\_PsoCVU y un oligonucleótido con psoraleno en 5' 5PsoC2\_A9\_GA junto con un oligonucleótido de soporte SplintC\_PsoC2.
- La Figura 18 muestra la LCMS de un conjugado de oligonucleótido ligado fotoquímicamente con psoraleno-timidina (Bio\_Pso).
- 20 La Figura 19 muestra la LCMS de ADNc generado a partir de Bio\_Pso usando extensión y ligación de cebador terminal y no terminal.
- La Figura 20 muestra la LCMS de control sin ligación de ADNc generado a partir de Bio\_Pso usando extensión de cebador terminal y no terminal.
- 25 La Figura 21 muestra los oligonucleótidos CVU\_G y CVU\_A y los oligonucleótidos Tag1\_PsoCVU, SplintC\_CVU y SplintA\_CVU.
- La Figura 22 muestra la estructura de la unión de ligación fotoquímica formada entre una 5'-5-(Carboxi)vinil-2'-desoxiuridina irradiada y una 3'-timidina.
- La Figura 23 muestra un gel del transcurso del tiempo de fotoligación de CVU a 365 nm.
- La Figura 24 muestra la LCMS del producto de ligación fotoquímico de CVU\_G con Tag1\_PsoCVU.
- 30 La Figura 25 muestra la LCMS de un conjugado de oligonucleótido ligado fotoquímicamente con carboxivinil uridina-timidina (Bio\_CVU).
- La Figura 26 muestra la LCMS de la plantilla Bio\_CVU menos un único nucleótido dA.
- La Figura 27 muestra los oligonucleótidos S\_IDTN<sub>3</sub> y S\_IDTAlquino.
- La Figura 28 muestra el análisis de LCMS del conjugado gelificado de S\_IDTN<sub>3</sub> y S\_IDTAlquino.
- 35 La Figura 29 muestra oligonucleótidos TKR\_Central y S\_IDTN<sub>3</sub> y oligo TKR\_DBCO\_S modificado con 5'-dibenzociclooctino (DBCO).
- La Figura 30 muestra el transcurso del tiempo por LCMS de la reacción click libre de cobre entre TKR\_DBCO\_S y TKR\_Central.
- La Figura 31 muestra la LCMS del conjugado 2cl\_s.

La Figura 32 muestra Conjugado\_Click\_S y Conjugado\_Click\_L.

La Figura 33 muestra la generación de ADNc mediada por recombinación dependiente del soporte /amplificación por PCR del Conjugado\_Click\_L y Conjugado\_Click\_S por Deep Vent (exo-)polimerasa en presencia y ausencia de ePsoporte. 24 ciclos, concentración del conjugado de partida 200 pM. 1-marcador; 2-sin plantilla; 3-sin plantilla + ePsoporte; 4-Conjugado\_Click\_S; 5- Conjugado\_Click\_S + ePsoporte; 6-Conjugado\_Click\_L; 5-Conjugado\_Click\_L + ePsoporte.

La Figura 34 muestra el mecanismo propuesto de generación de ADNc/amplificación por PCR mediada por recombinación dependiente del soporte.

La Figura 35 muestra la generación de ADNc mediada por recombinación que depende del soporte/amplificación por PCR de Conjugado\_Click\_S y el Conjugado\_Click\_L mediante exo-polimerasa Deep Vent tanto separadamente como mezcladas juntas en solución libre. Todas las reacciones complementadas con ePsoporte. 1-marcador; 2-sin plantilla; 3- Conjugado\_Click\_S; 4-Conjugado\_Click\_L; 5-Conjugado\_Click\_S y Conjugado\_Click\_L.

La Figura 36 muestra el posible mecanismo de la recombinación de los productos de generación de ADNc/PCR de mediación dependiente del soporte en solución libre con el consiguiente arrastre (pérdida) de asociaciones de etiquetas (información codificada).

La Figura 37 muestra la generación de cDNA mediada por recombinación/amplificación por PCR de la mezcla de Conjugado\_Click\_S y Conjugado\_Click\_L por (exo-) polimerasa Deep Vent en solución libre y en compartimentos acuosos emulsionados con las condiciones siguientes: 30 segundos a 94°C, 30 segundos a 52°C y 60 segundos a 72°C durante 35 ciclos, y concentración del conjugado inicial de 5 pM; 1- marcador; 2- solución libre, 3- compartimentos acuosos emulsionados.

La Figura 38 muestra los conjugados TKR\_S y TKR\_L y sus trazas por LCMS.

La Figura 39 muestra la generación de ADNc mediada por recombinación dependiente de la repetición de (exo-) polimerasa Deep Vent/amplificación por PCR de TKR\_S y TKR\_L. 1- Marcador, 2- sin plantilla, 3- TKR\_S, 4- TKR\_L, 5- TKR L y TKR\_S. Condiciones: 5 µL de regulador Thermopol 10x; 2,5 µL de una solución de 10 µM de cada cebador directo e inverso; 2,5 µL de mezcla 10 mM de dNTP (NEB); 40 µM de concentración final de cualquier conjugado o su mezcla 1:1, 2,5 µL de (exo-) polimerasa Deep Vent y agua hasta 50 µL. La reacción se cicló 21 veces de la siguiente manera: 30 segundos a 94°C, 30 segundos a 52°C y 60 segundos a 72°C.

La Figura 40 muestra datos de secuencia derivados del producto de amplificación clonado de la generación de ADNc mediada por recombinación dependiente de repetición/amplificación por PCR de TKR\_S y TKR\_L mezclados en solución libre.

La Figura 41 muestra los conjugados TKR\_2\_click\_S y TKR\_2\_click\_L y sus trazas por LCMS.

La Figura 42 muestra la generación de ADNc mediada por recombinación dependiente de la repetición de (exo-) polimerasa Deep Vent/amplificación por PCR de 2\_click\_S y 2\_click\_L en solución libre. 1- Marcador; 2- Sin control de plantilla; 3- 2\_click\_S; 4- 2\_click\_L; 5- Mezcla 1:1 de 2\_click\_S y 2\_click\_L. Condiciones: 5 µl de regulador Thermopol 10x; 2,5 µL de una solución 10 µM de cada cebador directo e inverso; 2,5 µL de mezcla dNTP 10 mM (NEB); 40 µM de concentración final de cualquier conjugado o su mezcla 1:1, 2,5 µL de (exo-) polimerasa Deep Vent y agua hasta 50 µL. La reacción se cicló 22 veces de la siguiente manera: 30 segundos a 94°C, 30 segundos a 52°C y 60 segundos a 72°C.

La Figura 43 muestra los datos de secuencia derivados del producto de amplificación clonado del producto de la generación de cDNA mediada por recombinación dependiente de la repetición/producto de amplificación por PCR de la mezcla de TKR\_2\_click\_S y TKR\_2\_click\_L en solución libre.

#### Descripción detallada

Se han desarrollado complejos que tienen al menos un enlace (por ejemplo, un enlace químico) para el que una polimerasa tiene una capacidad reducida para leer o translocarse a través de ese enlace. La capacidad de etiquetar bibliotecas sin estar restringida por la legibilidad de una polimerasa amplía enormemente el intervalo de métodos de etiquetado que se pueden usar. Las posibles ventajas no limitativas incluyen la disminución de la masa de las etiquetas y/o miembros de la biblioteca; mayor solubilidad de los miembros de la biblioteca en diversas condiciones de reacción (por ejemplo, en condiciones acuosas y/u orgánicas adecuadas tanto para la síntesis de la entidad química como de la etiqueta); mayor estabilidad (por ejemplo, en condiciones de reacción acuosas y/u orgánicas, donde tales etiquetas y enlaces de oligonucleótidos pueden ser más resistentes a condiciones de reacción particulares, en comparación con la etiqueta sin tales uniones); costo reducido (por ejemplo, reducir el uso de

reactivos enzimáticos caros); mayor facilidad de uso (por ejemplo, el uso de entrecruzamiento de cianovinilcarbazol se produce en un amplio intervalo de pH de 5,5-9,5, que permite el entrecruzamiento en condiciones de reacción adecuadas para formar la porción de entidad química del complejo); menor número de etapas de reacción y uso reducido de reactivo o reactivos (por ejemplo, uso reducido de una etapa de intercambio de regulador, tal como durante las reacciones de división combinatoria, donde miles de alícuotas pequeñas individuales deben procesarse independientemente; una etapa de precipitación, o una etapa de modificación del pH); mayor fidelidad (por ejemplo, cuando se usan métodos mediados por hibridación para reducir la frecuencia de ocurrencia de eventos de etiquetado erróneo); y una mayor compatibilidad con las condiciones de reacción para formar la entidad química (por ejemplo, permite el uso de funciones ortogonales, donde se pueden utilizar procesos químicos de etiquetado únicos (por ejemplo, el uso de radiación UV) que no se producirán durante la etapa o las etapas de división posterior/mezcla, lo que también reducirá la frecuencia de eventos de etiquetado erróneo).

Debido a que dichos enlaces pueden tener una legibilidad reducida por una polimerasa, también se han desarrollado métodos para permitir la secuenciación de tales complejos. Estos métodos incluyen el uso de cebadores de relevo para abarcar los enlaces y/o el uso de oligonucleótidos entrecruzables liberables. Estos complejos y métodos pueden usarse para crear diversas bibliotecas de entidades químicas seleccionables estableciendo una relación codificada entre etiquetas particulares y reacciones químicas particulares o bloques de construcción. Para identificar una o más entidades químicas, las etiquetas de oligonucleótidos pueden amplificarse, clonarse, secuenciarse y correlacionarse usando la relación establecida. Los métodos para crear y etiquetar bibliotecas de estos complejos se describen en detalle a continuación.

## 20 Complejos

Esta invención presenta un complejo que incluye una entidad química, una o más etiquetas, y una cabeza de dominio operativamente asociada con la entidad química y una o más etiquetas. Al menos uno de los enlaces entre la cabeza de dominio y la etiqueta o entre dos etiquetas es un enlace para el cual una polimerasa tiene una capacidad reducida para leer o translocarse a través de ese enlace. Las entidades químicas, las cabezas de dominio, las etiquetas, los enlaces y los espaciadores bifuncionales se describen con más detalle a continuación.

### Entidades químicas

Las entidades o miembros químicos (por ejemplo, moléculas pequeñas o péptidos) de la invención pueden incluir uno o más bloques de construcción y opcionalmente incluir uno o más andamios.

El andamio S puede ser un solo átomo o un andamio molecular. Ejemplos de andamios de átomos individuales incluyen un átomo de carbono, un átomo de boro, un átomo de nitrógeno o un átomo de fósforo, etc. Los andamios poliatómicos ilustrativos incluyen un grupo cicloalquilo, un grupo cicloalqueno, un grupo heterocicloalquilo, un grupo heterocicloalqueno, un grupo arilo o un grupo heteroarilo. Las realizaciones particulares de un andamio de heteroarilo incluyen una triazina, tal como 1,3,5-triazina, 1,2,3-triazina o 1,2,4-triazina; una pirimidina; una pirazina; una piridazina; un furano; un pirrol; una pirrolina; una pirrolidina; un oxazol; un pirazol; un isoxazol; un pirano; una piridina; un indol; un indazol; o una purina.

El andamio S se puede unir operativamente a la etiqueta mediante cualquier método útil. En un ejemplo, S es una triazina que está vinculada directamente a la cabeza de dominio. Para obtener este andamio ilustrativo, se hace reaccionar la triclorotriazina (es decir, un precursor clorado de la triazina que tiene tres cloros) con un grupo nucleofílico de la cabeza de dominio. Usando este método, S tiene tres posiciones que tienen cloro que están disponibles para sustitución, donde dos posiciones son nodos de diversidad disponibles y una posición se une a la cabeza de dominio. A continuación, se agrega el bloque de construcción  $A_n$  a un nodo de diversidad del andamio, y la etiqueta  $A_n$  que codifica para el bloque de construcción  $A_n$  ("etiqueta  $A_n$ ") se liga a la cabeza de dominio, donde estas dos etapas se pueden realizar en cualquier orden. Luego, el bloque de construcción  $B_n$  se agrega al nodo de diversidad restante, y la etiqueta  $B_n$  que codifica para el bloque de construcción  $B_n$  se ligó al extremo de la etiqueta  $A_n$ . En otro ejemplo, S es una triazina que está operativamente ligada a una etiqueta, donde la triclorotriazina se hace reaccionar con un grupo nucleofílico (por ejemplo, un grupo amino) de un enlazador PEG, alifático o aromático de una etiqueta. Se pueden agregar bloques de construcción y etiquetas asociadas, como se describió anteriormente.

En otro ejemplo más, S es una triazina que está operativamente unida al bloque de construcción  $A_n$ . Para obtener este andamio, bloque de construcción  $A_n$  que tiene dos nodos de diversidad (por ejemplo, un grupo electrófilo y un grupo nucleofílico, tal como un Fmoc-aminoácido) se hace reaccionar con el grupo nucleofílico de un enlazador (por ejemplo, el grupo terminal de un enlazador PEG, alifático o aromático, que está unido a la cabeza de dominio). Luego, la triclorotriazina se hace reaccionar con un grupo nucleofílico del bloque de construcción  $A_n$ . Usando este método, las tres posiciones de cloro de S se usan como nodos de diversidad para los bloques de construcción. Como se describe aquí, se pueden agregar bloques de construcción y etiquetas adicionales, y se pueden agregar andamios adicionales  $S_n$ .

Los bloques de construcción  $A_n$  incluyen, por ejemplo, aminoácidos (por ejemplo, aminoácidos alfa, beta, gamma, delta y épsilon, así como derivados de aminoácidos naturales y no naturales), reactivos químicos reactivos (por ejemplo, cadenas de azida o alquino) con una amina, o un reactivo de tiol, o combinaciones de los mismos. La elección del bloque de construcción  $A_n$  depende, por ejemplo, de la naturaleza del grupo reactivo utilizado en el enlazador, la naturaleza de una fracción de andamio y el disolvente utilizado para la síntesis química.

Los ejemplos de bloques de construcción  $B_n$  y  $C_n$  incluyen cualquier unidad estructural útil de una entidad química, tal como grupos aromáticos opcionalmente sustituidos (por ejemplo, fenilo o bencilo opcionalmente sustituido), grupos heterocíclicos opcionalmente sustituidos (por ejemplo, quinolinilo, isoquinolinilo, indolilo, isoindolilo, azaindolilo, bencimidazolilo, azabencimidazolilo, benzisoxazolilo, piridinilo, piperidilo o pirrolidinilo opcionalmente sustituidos), grupos alquilo opcionalmente sustituidos (por ejemplo, grupos alquilo  $C_{1-6}$  lineales o ramificados opcionalmente sustituidos o grupos aminoalquilo  $C_{1-6}$  opcionalmente sustituidos) o grupos carbocíclico opcionalmente sustituidos (por ejemplo, ciclopropilo, ciclohexilo o ciclohexenilo opcionalmente sustituidos). Los bloques de construcción  $B_n$  y  $C_n$  particularmente útiles incluyen aquellos con uno o más grupos reactivos, tales como un grupo opcionalmente sustituido (por ejemplo, cualquiera descrito aquí) que tiene uno o sustituyentes opcionales que son grupos reactivos o pueden modificarse químicamente para formar grupos reactivos. Los ejemplos de grupos reactivos incluyen uno o más de amina ( $-NR_2$ , donde cada R es, independientemente, H o un alquilo  $C_{1-6}$  opcionalmente sustituido), hidroxilo, alcoxi ( $-OR$ , donde R es un alquilo  $C_{1-6}$  opcionalmente sustituido, tal como metoxi), carboxi ( $-COOH$ ), amida o sustituyentes reactivos químicos. Se puede introducir un sitio de restricción, por ejemplo, en la etiqueta  $B_n$  o  $C_n$ , donde se puede identificar un complejo realizando PCR y digestión de restricción con una de las enzimas de restricción correspondientes.

#### Cabeza de dominio

En la biblioteca, la cabeza de dominio conecta operativamente cada entidad química con su etiqueta de oligonucleótido de codificación. Generalmente, la cabeza de dominio es un oligonucleótido de partida que tiene dos grupos funcionales que pueden formar derivados adicionalmente, donde el primer grupo funcional une operativamente la entidad química (o un componente de la misma) a la cabeza de dominio y el segundo grupo funcional une operativamente una o más etiquetas a la cabeza de dominio. Un espaciador bifuncional puede usarse opcionalmente como una fracción espaciadora entre la cabeza de dominio y la entidad química.

Los grupos funcionales de la cabeza de dominio se pueden usar para formar un enlace covalente con un componente de la entidad química y otro enlace covalente con una etiqueta. El componente puede ser cualquier parte de la molécula pequeña, tal como un andamio con nodos de diversidad o un bloque de construcción. Alternativamente, la cabeza de dominio puede formar derivados para proporcionar un espaciador (es decir, una fracción espaciadora que separa la cabeza de dominio de la molécula pequeña que se formará en la biblioteca) que termina en un grupo funcional (por ejemplo, un grupo hidroxilo, amina, carboxilo, sulfhidrilo, alquino, azido, o fosfato), que se usa para formar el enlace covalente con un componente de la entidad química. El espaciador se puede unir al extremo 5', en una de las posiciones internas, o al extremo 3' de la cabeza de dominio. Cuando el espaciador está unido a una de las posiciones internas, el espaciador puede unirse operativamente a una base que forma derivado (por ejemplo, la posición C5 de uridina) o colocarse internamente dentro del oligonucleótido utilizando técnicas estándar conocidas en la técnica. Ejemplos de espaciadores se describen en el presente documento.

La cabeza de dominio puede tener cualquier estructura útil. La cabeza de dominio puede ser, por ejemplo, de 1 a 100 nucleótidos de longitud, preferiblemente de 5 a 20 nucleótidos de longitud, y lo más preferiblemente de 5 a 15 nucleótidos de longitud. La cabeza de dominio puede ser de cadena sencilla o de cadena doble y puede consistir en nucleótidos naturales o modificados, como se describe en este documento. Por ejemplo, la fracción química puede estar operativamente enlazada al extremo 3' o al extremo 5' de la cabeza de dominio. En realizaciones particulares, la cabeza de dominio incluye una estructura en horquilla formada por bases complementarias dentro de la secuencia. Por ejemplo, la fracción química puede estar operativamente enlazada a la posición interna, el extremo 3' o el extremo 5' de la cabeza de dominio.

Generalmente, la cabeza de dominio incluye una secuencia no autocomplementaria en el extremo 5' o 3' que permite unir una etiqueta de oligonucleótido mediante polimerización, ligación enzimática o reacción química. La cabeza de dominio puede permitir la ligación de etiquetas de oligonucleótidos y las etapas de purificación y fosforilación opcionales. Después de agregar la última etiqueta, se puede agregar una secuencia de adaptador adicional al extremo 5' de la última etiqueta. Las secuencias adaptadoras a modo de ejemplo incluyen una secuencia de unión al cebador o una secuencia que tiene una etiqueta (por ejemplo, biotina). En los casos en que se usan muchos bloques de construcción y etiquetas correspondientes (por ejemplo, 100), se puede emplear una estrategia de mezcla y división durante la etapa de síntesis de oligonucleótidos para crear el número necesario de etiquetas. Tales estrategias de mezcla y división para la síntesis de ADN son conocidas en la técnica. Los miembros de la biblioteca resultante pueden amplificarse mediante PCR después de la selección de entidades de unión frente a un objetivo u objetivos de interés.

La cabeza de dominio o el complejo pueden incluir opcionalmente una o más secuencias de unión al cebador. Por ejemplo, la cabeza de dominio tiene una secuencia en la región de bucle de la horquilla que sirve como región de unión del cebador para la amplificación, donde la región de unión del cebador tiene una temperatura de fusión más alta para su cebador complementario (por ejemplo, que puede incluir regiones identificadoras flanqueantes) que para una secuencia en la cabeza de dominio. En otras realizaciones, el complejo incluye dos secuencias de unión al cebador (por ejemplo, para permitir una reacción de PCR) en cualquier lado de una o más etiquetas que codifican uno o más bloques de construcción. Alternativamente, la cabeza de dominio puede contener una secuencia de unión del cebador en el extremo 5' o 3'. En otras realizaciones, la cabeza de dominio es una horquilla, y la región de bucle forma un sitio de unión del cebador o el sitio de unión del cebador se introduce a través de la hibridación de un oligonucleótido con la cabeza de dominio en el lado 3' del bucle. Un oligonucleótido cebador, que contiene una región homóloga al extremo 3' de la cabeza de dominio y que porta una región de unión al cebador en su extremo 5' (por ejemplo, para permitir una reacción de PCR) puede hibridarse con la cabeza de dominio y puede contener una etiqueta que codifica un bloque de construcción o la adición de un bloque de construcción. El oligonucleótido cebador puede contener información adicional, tal como una región de nucleótidos aleatorios, por ejemplo, de 2 a 16 nucleótidos de longitud, que se incluye para el análisis bioinformático.

La cabeza de dominio puede incluir opcionalmente una estructura de horquilla, donde esta estructura se puede lograr mediante cualquier método útil. Por ejemplo, la cabeza de dominio puede incluir bases complementarias que forman compañeros de emparejamiento de bases intermoleculares, tales como emparejamiento de bases de ADN de Watson-Crick (por ejemplo, adenina-timina y guanina-citosina) y/o emparejamiento de base inestable (por ejemplo, guanina-uracilo, inosina-uracilo, inosina-adenina e inosina-citosina). En otro ejemplo, la cabeza de dominio puede incluir nucleótidos modificados o sustituidos que pueden formar formaciones dúplex de mayor afinidad en comparación con nucleótidos no modificados, siendo conocidos en la técnica tales nucleótidos modificados o sustituidos. En otro ejemplo más, la cabeza de dominio incluye una o más bases entrecruzadas para formar la estructura en horquilla. Por ejemplo, las bases dentro de una única cadena o bases en diferentes cadenas dobles se pueden entrecruzar, por ejemplo, usando psoraleno.

La cabeza de dominio o complejo puede incluir opcionalmente una o más etiquetas que permitan la detección. Por ejemplo, la cabeza de dominio, una o más etiquetas de oligonucleótidos y/o una o más secuencias de cebadores pueden incluir un isótopo, un agente de radioimagenología, un marcador, un trazador, una etiqueta fluorescente (por ejemplo, rodamina o fluoresceína), una etiqueta quimioluminiscente, un punto cuántico y una molécula informadora (por ejemplo, biotina o una etiqueta his).

En otras realizaciones, la cabeza de dominio o etiqueta puede modificarse para soportar la solubilidad en condiciones semiacuosas, acuosas reducidas o no acuosas (por ejemplo, orgánicas). Las bases de nucleótidos de la cabeza de dominio o etiqueta pueden hacerse más hidrófobas modificando, por ejemplo, las posiciones C5 de las bases T o C con cadenas alifáticas sin alterar significativamente su capacidad para formar enlaces de hidrógeno con sus bases complementarias. Los nucleótidos modificados o sustituidos a modo de ejemplo son 5'-dimetoxitritil-N-4-diisobutilaminometiliden-5-(1-propinil)-2'-desoxicidina, 3'-[(2-cianoetil)-(N,N-diisopropil)]-fosforamidita; 5'-dimetoxitritil-5-(1-propinil)-2'-desoxiuridina, 3'-[(2-cianoetil)-(N,N-diisopropil)]-fosforamidita; 5'-dimetoxitritil-5-fluoro-2'-desoxiuridina, 3'-[(2-cianoetil)-(N,N-diisopropil)]-fosforamidita; y 5'-dimetoxitritil-5-(piren-1-il-etinil)-2'-desoxiuridina, o 3'-[(2-cianoetil)-(N,N-diisopropil)]-fosforamidita.

Además, el oligonucleótido de la cabeza de dominio se puede intercalar con modificaciones que promueven la solubilidad en disolventes orgánicos. Por ejemplo, la fosforamidita de azobenceno puede introducir una fracción hidrófoba en el diseño de la cabeza de dominio. Tales inserciones de amiditas hidrófobas en la cabeza de dominio pueden ocurrir en cualquier parte de la molécula. Sin embargo, la inserción no puede interferir con el etiquetado posterior usando etiquetas de ADN adicionales durante la síntesis de la biblioteca o la PCR subsiguiente una vez que se completa una selección o análisis de microarreglos, si se usa para la deconvolución de la etiqueta. Tales adiciones al diseño de la cabeza de dominio aquí descrita harán que la cabeza de dominio sea soluble, por ejemplo, en 15%, 25%, 30%, 50%, 75%, 90%, 95%, 98%, 99% o 100% de disolvente orgánico. De este modo, la adición de residuos hidrófobos en el diseño de la cabeza de dominio permite una solubilidad mejorada en condiciones semiacuosas o no acuosas (por ejemplo, orgánicas), al tiempo que hace que la cabeza de dominio sea competente para el etiquetado de oligonucleótidos. Además, las etiquetas de ADN que se introducen posteriormente en la biblioteca también pueden modificarse en la posición C5 de las bases T o C de manera que también hagan que la biblioteca sea más hidrófoba y soluble en disolventes orgánicos para las siguientes etapas de la síntesis de la biblioteca.

En realizaciones particulares, la cabeza de dominio y la primera etiqueta pueden ser la misma entidad, es decir, se puede construir una pluralidad de entidades de etiqueta de la cabeza de dominio que comparten partes comunes (por ejemplo, una región de unión del cebador) y todas difieren en otra parte (por ejemplo, la región de codificación). Estas pueden utilizarse en la etapa de "división" y agruparse después de que haya ocurrido la codificación.

En realizaciones particulares, la cabeza de dominio puede codificar información, por ejemplo, incluyendo una secuencia que codifica la primera etapa de división o divisiones o una secuencia que codifica la identidad de la biblioteca, tal como mediante el uso de una secuencia particular relacionada con una biblioteca específica.

Etiquetas de oligonucleótidos

5 Las etiquetas de oligonucleótidos descritas en este documento (por ejemplo, una etiqueta o una porción de una cabeza de dominio o una porción de una cola de dominio) pueden usarse para codificar cualquier información útil, tal como una molécula, una porción de una entidad química, la adición de un componente (por ejemplo, un andamio o un bloque de construcción), una cabeza de dominio en la biblioteca, la identidad de la biblioteca, el uso de uno o más miembros de la biblioteca (por ejemplo, el uso de los miembros en una alícuota de una biblioteca), y/o el origen de un miembro de la biblioteca (por ejemplo, mediante el uso de una secuencia de origen).

15 Se puede usar cualquier secuencia en un oligonucleótido para codificar cualquier información. De este modo, una secuencia de oligonucleótidos puede servir a más de un propósito, tal como codificar dos o más tipos de información o proporcionar un oligonucleótido de partida que también codifica para uno o más tipos de información. Por ejemplo, la primera etiqueta puede codificar para la adición de un primer bloque de construcción, así como para la identificación de la biblioteca. En otro ejemplo, se puede usar una cabeza de dominio para proporcionar un oligonucleótido de partida que une operativamente una entidad química a una etiqueta, donde la cabeza de dominio incluye adicionalmente una secuencia que codifica la identidad de la biblioteca (es decir, la secuencia de identificación de la biblioteca). En consecuencia, cualquiera de la información descrita en este documento puede codificarse en etiquetas de oligonucleótidos separadas o puede combinarse y codificarse en la misma secuencia de oligonucleótidos (por ejemplo, una etiqueta de oligonucleótido, tal como una etiqueta, o una cabeza de dominio).

Una secuencia de bloque de construcción codifica la identidad de un bloque de construcción y/o el tipo de reacción de unión conducida con un bloque de construcción. Esta secuencia de bloque de construcción se incluye en una etiqueta, donde la etiqueta puede incluir opcionalmente uno o más tipos de secuencia descritos a continuación (por ejemplo, una secuencia de identificación de biblioteca, una secuencia de uso y/o una secuencia de origen).

25 Una secuencia de identificación de la biblioteca codifica la identidad de una biblioteca particular. Con el fin de permitir la mezcla de dos o más bibliotecas, un miembro de la biblioteca puede contener una o más secuencias que identifican la biblioteca, como en una etiqueta identificadora de la biblioteca (es decir, un oligonucleótido que incluye una secuencia identificadora de la biblioteca), en una etiqueta ligada, en una parte de la secuencia de la cabeza de dominio, o en una secuencia de la cola de dominio. Estas secuencias de identificación de bibliotecas se pueden usar para deducir relaciones de codificación, donde la secuencia de la etiqueta se traduce y se correlaciona con información de historial químico (síntesis). De acuerdo con esto, estas secuencias que identifican la biblioteca permiten la mezcla de dos o más bibliotecas juntas para selección, amplificación, purificación, secuenciación, etc.

35 Una secuencia de uso codifica el historial (es decir, uso) de uno o más miembros de la biblioteca en una alícuota individual de una biblioteca. Por ejemplo, alícuotas separadas pueden tratarse con diferentes condiciones de reacción, bloques de construcción y/o etapas de selección. En particular, esta secuencia puede usarse para identificar tales alícuotas y deducir su historial (uso) y de ese modo permitir la mezcla de alícuotas de la misma biblioteca con diferentes historias (usos) (por ejemplo, experimentos de selección distintos) para los fines de la mezcla juntas de muestras para selección, amplificación, purificación, secuenciación, etc. Estas secuencias de uso se pueden incluir en una cabeza de dominio, una cola de dominio, una etiqueta, una etiqueta de uso (es decir, un oligonucleótido que incluye una secuencia de uso) o cualquier otra etiqueta descrita en el presente documento (por ejemplo, una etiqueta de identificación de biblioteca o una etiqueta de origen).

45 Una secuencia de origen es una secuencia de oligonucleótidos degenerada (aleatoria, generada estocásticamente) de cualquier longitud útil (por ejemplo, aproximadamente seis oligonucleótidos) que codifica el origen del miembro de la biblioteca. Esta secuencia sirve para subdividir estocásticamente los miembros de la biblioteca que son idénticos en todos los aspectos en entidades distinguibles por información de secuencia, de tal manera que las observaciones de productos de amplificación derivados de plantillas únicas de progenitores (por ejemplo, miembros seleccionados de la biblioteca) pueden distinguirse de las observaciones de múltiples productos de amplificación derivados de la misma plantilla progenitora (por ejemplo, un miembro seleccionado de la biblioteca). Por ejemplo, después de la formación de la biblioteca y antes de la etapa de selección, cada miembro de la biblioteca puede incluir una secuencia de origen diferente, tal como en una etiqueta de origen. Después de la selección, los miembros de la biblioteca seleccionados pueden amplificarse para producir productos de amplificación, y la porción del miembro de la biblioteca que se espera que incluya la secuencia de origen (por ejemplo, en la etiqueta de origen) puede ser observada y comparada con la secuencia de origen en cada una de los otros miembros de la biblioteca. Como las secuencias de origen son degeneradas, cada producto de amplificación de cada miembro de la biblioteca debe tener una secuencia de origen diferente. Sin embargo, una observación de la misma secuencia de origen en el producto de amplificación podría indicar múltiples amplicones derivados de la misma molécula plantilla. Cuando se desea determinar las estadísticas y la demografía de la población de etiquetas de codificación antes de la amplificación, en oposición a la posterior amplificación, se puede usar la etiqueta de origen. Estas secuencias de origen pueden

incluirse en una cabeza de dominio, una cola de dominio, una etiqueta, una etiqueta de origen (es decir, un oligonucleótido que incluye una secuencia de origen) o cualquier otra etiqueta descrita en la presente memoria (por ejemplo, una etiqueta identificadora de biblioteca o una etiqueta de uso).

5 Cualquiera de los tipos de secuencias que se describen en este documento puede incluirse en la cabeza de dominio. Por ejemplo, la cabeza de dominio puede incluir una o más de una secuencia de bloque de construcción, una secuencia de identificación de biblioteca, una secuencia de uso o una secuencia de origen.

Cualquiera de estas secuencias descritas aquí se puede incluir en una cola de dominio. Por ejemplo, la cola de dominio puede incluir una o más entre una secuencia de identificación de la biblioteca, una secuencia de uso o una secuencia de origen.

10 Cualquiera de las etiquetas descritas en este documento puede incluir un conector en o cerca del extremo 5' o 3' que tiene una secuencia fija. Los conectores facilitan la formación de enlaces (por ejemplo, enlaces químicos) proporcionando un grupo reactivo (por ejemplo, un grupo químico reactivo o un grupo fotorreactivo) o proporcionando un sitio para un agente que permite un enlace (por ejemplo, un agente de una fracción de intercalación o un grupo reactivo reversible en el o los conectores u oligonucleótido de entrecruzamiento). Cada  
 15 conector 5' puede ser el mismo o diferente, y cada conector 3' puede ser el mismo o diferente. En un complejo ilustrativo no limitante que tiene más de una etiqueta, cada etiqueta puede incluir un conector 5' y un conector 3', donde cada conector 5' tiene la misma secuencia y cada conector 3' tiene la misma secuencia (por ejemplo, cuando la secuencia del conector 5' puede ser la misma o diferente de la secuencia del conector 3'). El conector proporciona una secuencia que se puede usar para uno o más enlaces. Para permitir la unión de un cebador de relevo o para  
 20 hibridar un oligonucleótido de entrecruzamiento, el conector puede incluir uno o más grupos funcionales que permiten un enlace (por ejemplo, un enlace para el cual una polimerasa tiene capacidad reducida para leer o translocarse a través de ese enlace, tal como un enlace químico).

25 Estas secuencias pueden incluir cualquier modificación descrita aquí para oligonucleótidos, tal como una o más modificaciones que promueven la solubilidad en disolventes orgánicos (por ejemplo, cualquiera aquí descrito, tal como para la cabeza de dominio), que proporcionan un análogo del enlace fosfodiéster natural (por ejemplo, un análogo de fosforotioato), o que proporciona uno o más oligonucleótidos no naturales (por ejemplo, nucleótidos sustituidos en 2', tales como nucleótidos 2'-O-metilados y nucleótidos 2'-fluoro, o cualquiera de los descritos en este documento).

30 Estas secuencias pueden incluir cualquier característica descrita aquí para oligonucleótidos. Por ejemplo, estas secuencias se pueden incluir en una etiqueta que tiene menos de 20 nucleótidos (por ejemplo, como se describe en la presente memoria). En otros ejemplos, las etiquetas que incluyen una o más de estas secuencias tienen aproximadamente la misma masa (por ejemplo, cada etiqueta tiene una masa que es aproximadamente +/- 10% de la masa promedio entre un conjunto específico de etiquetas que codifican una variable específica); carecen de una  
 35 región de unión al cebador (por ejemplo, constante); carecen de una región constante; o tienen una región constante de longitud reducida (por ejemplo, una longitud inferior a 30 nucleótidos, inferior a 25 nucleótidos, inferior a 20 nucleótidos, inferior a 19 nucleótidos, inferior a 18 nucleótidos, inferior a 17 nucleótidos, inferior a 16 nucleótidos, inferior a 15 nucleótidos, inferior a 14 nucleótidos, inferior a 13 nucleótidos, inferior a 12 nucleótidos, inferior a 11 nucleótidos, inferior a 10 nucleótidos, inferior a 9 nucleótidos, inferior a 8 nucleótidos o inferior a 7 nucleótidos).

40 Las estrategias de secuenciación para bibliotecas y oligonucleótidos de esta longitud pueden incluir opcionalmente estrategias de concatenación o catenación para aumentar la fidelidad de lectura o la profundidad de la secuencia, respectivamente. En particular, la selección de bibliotecas codificadas que carecen de regiones de unión al cebador se ha descrito en la literatura para SELEX, tal como se describe en Jarosch et al., *Nucleic Acids Res.* 34: e86 (2006). Por ejemplo, un miembro de la biblioteca se puede modificar (por ejemplo, después de una etapa de selección) para  
 45 incluir una primera secuencia de adaptador en el extremo 5' del complejo y una segunda secuencia de adaptador en el extremo 3' del complejo, donde la primera la secuencia es sustancialmente complementaria a la segunda secuencia y da como resultado la formación de un dúplex. Para mejorar aún más el rendimiento, se añaden dos nucleótidos colgantes fijos (por ejemplo, CC) al extremo 5'. En realizaciones particulares, la primera secuencia adaptadora es 5'-GTGCTGC-3' (SEQ ID NO: 1), y la segunda secuencia adaptadora es 5'-GCAGCACCC-3' (SEQ ID NO: 2).

## 50 Enlaces

Los enlaces de la invención están presentes entre oligonucleótidos que codifican información (por ejemplo, tal como entre la cabeza de dominio y una etiqueta, entre dos etiquetas, o entre una etiqueta y una cola de dominio), donde tales enlaces incluyen cualquier enlace para el cual una polimerasa tiene una capacidad reducida para leer o  
 55 translocarse a través de ese enlace. Los enlaces a modo de ejemplo incluyen enlaces químicos que incluyen uno o más de un grupo químico reactivo, un grupo fotorreactivo, una fracción de intercalación, un oligonucleótido de entrecruzamiento, o un grupo correactivo reversible.

Se puede analizar un enlace para determinar si una polimerasa tiene una capacidad reducida para leer o translocarse a través de ese enlace. Esta capacidad puede analizarse mediante cualquier método útil, tal como cromatografía líquida-espectrometría de masas, análisis de RT-PCR, datos demográficos de secuencias y/o análisis de PCR.

5 En realizaciones particulares, la ligación química incluye el uso de uno o más pares químicos reactivos para proporcionar un enlace. Los ejemplos de pares químicos reactivos son un par que incluye un grupo alquínico opcionalmente sustituido y un grupo azido opcionalmente sustituido para formar un triazol mediante una reacción de cicloadición 1,3-dipolar de Huisgen; un dieno opcionalmente sustituido que tiene un sistema de electrones  $4\pi$  (por ejemplo, un compuesto 1,3-insaturado opcionalmente sustituido, tal como 1,3-butadieno, 1-metoxi-3-trimetilsililoxi-1,3-butadieno, ciclopentadieno, ciclohexadieno o furano opcionalmente sustituidos) y un dienófilo opcionalmente sustituido o un heterodienófilo opcionalmente sustituido que tiene un sistema de electrones  $2\pi$  (por ejemplo, un grupo alquénico opcionalmente sustituido o un grupo alquínico opcionalmente sustituido) para formar un cicloalquénico a través de una reacción de Diels-Alder; un nucleofílico (por ejemplo, una amina opcionalmente sustituida o un tiol opcionalmente sustituido) con un electrofílico heterocíclico deformado (por ejemplo, epóxido, aziridina, ion aziridinio o ion episulfonio opcionalmente sustituidos) para formar un heteroalquilo a través de una reacción de apertura de anillo; un grupo fosforotioato con un grupo yodo, tal como en una ligación de soporte de un oligonucleótido que contiene dT con yodo en posición 5' con un oligonucleótido 3'-fosforotioato; un grupo amino opcionalmente sustituido con un grupo aldehído o un grupo cetona, tal como una reacción de un oligonucleótido modificado con aldehído en posición 3', que se puede obtener opcionalmente oxidando un oligonucleótido modificado con glicerilo en posición 3' comercialmente disponible, con oligonucleótido con amino en posición 5' (es decir, en una reacción de aminación reductiva) o un oligonucleótido 5'-hidrazido; un par de un grupo amino opcionalmente sustituido y un grupo de ácido carboxílico o un grupo tiol (por ejemplo, con o sin el uso de succinimidil trans-4-(maleimidilmetil)ciclohexano-1-carboxilato (SMCC) o 1-etil-3-(3-dimetilaminopropil)carbodiimida (EDAC), un par de una hidracina opcionalmente sustituida y un grupo aldehído o una cetona, un par de una hidroxilamina opcionalmente sustituida y un grupo aldehído o una cetona, o un par de un nucleofílico y un haluro de alquilo opcionalmente sustituido.

Los complejos de platino, los agentes alquilantes o los nucleótidos modificados con furano también se pueden usar como un grupo químico reactivo para formar enlaces intercatenarios o intracatenarios. Dichos agentes se pueden usar entre dos oligonucleótidos y opcionalmente pueden estar presentes en el oligonucleótido de entrecruzamiento.

30 Los ejemplos de complejos de platino no limitantes incluyen cisplatino (cis-diaminalcloroplatino (II), por ejemplo, para formar enlaces GG intracatenarios), transplatino (trans-diaminadcloroplatino (II), por ejemplo, para formar enlaces intercatenario GXG, donde X puede ser cualquier nucleótido), carboplatino, picolplatino (ZD0473), ormaplatino u oxaliplatino para formar, por ejemplo, enlaces GC, CG, AG o GG. Cualquiera de estos enlaces pueden ser enlaces inter o intracatenarios.

35 Los agentes alquilantes ilustrativos no limitantes incluyen mostaza nitrogenada (mecloretamina, por ejemplo, para formar enlaces GG), clorambucilo, melfalán, ciclofosfamida, formas profármaco de ciclofosfamida (por ejemplo, 4-hidroperoxiciclofosfamida e ifosfamida), 1,3-bis(2-cloroetil)-1-nitrosourea (BCNU, carmustina), una aziridina (por ejemplo, mitomicina C, trietilenmelamina o trietilenotiofosforamida (tio-tepa) para formar enlaces GG o AG), hexametilmelamina, un sulfonato de alquilo (por ejemplo, busulfano a formar enlaces GG), o una nitrosourea (por ejemplo, 2-cloroetilnitrosourea para formar enlaces GG o CG, tales como carmustina (BCNU), clorozotocina, lomustina (CCNU) y semustina (metil-CCNU)). Cualquiera de estos enlaces puede ser enlaces inter o intracatenarios.

45 Los nucleótidos modificados con furano también pueden usarse para formar enlaces. Tras la oxidación in situ (por ejemplo, con N-bromosuccinimida (NBS)), la fracción de furano forma un derivado reactivo oxo-enal que reacciona con una base complementaria para formar un enlace intercatenario. En algunas realizaciones, los nucleótidos modificados con furano forman enlaces con un nucleótido A o C complementario. Los ejemplos de nucleótidos modificados con furano, no limitantes, incluyen cualquier nucleótido modificado con 2'-(furan-2-il)propanoilamino; o nucleótidos modificados acíclicos del ácido nucleico 2-(furan-2-il)etilglicol.

50 Los grupos fotorreactivos también se pueden usar como un grupo reactivo. Los ejemplos de grupos fotorreactivos no limitativos incluyen una fracción de intercalación, un derivado de psoraleno (por ejemplo, psoraleno, HMT-psoraleno u 8-metoxipsoraleno), un grupo cianovinilcarbazol opcionalmente sustituido, un grupo vinilcarbazol opcionalmente sustituido, un grupo cianovinilo opcionalmente sustituido, un grupo acrilamida opcionalmente sustituido, un grupo diazirina opcionalmente sustituido, una benzofenona opcionalmente sustituida (por ejemplo, succinimidil éster de ácido 4-benzoilbenzoico o isotiocianato de benzofenona), un grupo 5-(carboxi)vinil-uridina opcionalmente sustituido (por ejemplo, 5-(carboxi)vinil-2'-desoxiuridina), o un grupo azida opcionalmente sustituido (por ejemplo, una aril azida o una aril azida halogenada, tal como succinimidil éster de ácido 4-azido-2,3,5,6-tetrafluorobenzoico (ATFB)) .

Las fracciones de intercalación también se pueden usar como un grupo reactivo. Ejemplos de fracciones de intercalación no limitantes incluyen un derivado de psoraleno, un derivado de alcaloide (por ejemplo, berberina, palmatina, coralina, sanguinarina (por ejemplo, formas de iminio o alcanolamina de los mismos) o aristolactama- $\beta$ -

D-glucósido), un catión de etidio (por ejemplo, bromuro de etidio), un derivado de acridina (por ejemplo, proflavina, acriflavina o amsacrina), un derivado de antraciclina (por ejemplo, doxorubicina, epirubicina, daunorrubicina (daunomicina), idarrubicina y aclarrubicina) o talidomida.

5 Para un oligonucleótido de entrecruzamiento, cualquier grupo reactivo útil (por ejemplo, descrito en este documento) puede usarse para formar enlaces intercatenarios o intracatenarios. Los ejemplos de grupos reactivos incluyen un grupo químico reactivo, un grupo fotorreactivo, una fracción de intercalación y un grupo correactivo reversible. Los agentes de entrecruzamiento para su uso con oligonucleótidos de entrecruzamientos incluyen, sin limitación, agentes alquilantes (por ejemplo, como se describe en la presente memoria), cisplatino (cis-diaminodicloroplatino (II)), trans-diaminodicloroplatino (II), psoraleno, HMT-psoraleno, 8-metoxipsoraleno, nucleótidos modificados con furano, 2-fluoro-deoxiinosina (2-F-dI), 5-bromo-deoxicitosina (5-Br-dC), 5-bromo desoxiuridina (5-Br-dU), 5-yodo-desoxicitosina (5-I-dC), 5-yodo-desoxiuridina (5-I-dU), succinimidil trans-4-(maleimidilmetil)ciclohexano-1-carboxilato, SMCC, EDAC o succinimidil acetiltoacetato (SATA).

15 Los oligonucleótidos también se pueden modificar para que contengan fracciones tiol que se pueden hacer reaccionar con una variedad de grupos reactivos con tiol tales como maleimididas, halógenos, yodoacetamidas y, por lo tanto, se pueden usar para entrecruzar dos oligonucleótidos. Los grupos tiol se pueden unir al extremo 5' o al extremo 3' de un oligonucleótido.

20 Para el entrecruzamiento intercatenario entre oligonucleótidos dúplex en una posición de pirimidina (por ejemplo, timidina), se puede elegir la fracción de psoraleno de intercalación fotorreactiva. El psoraleno se intercala en el dúplex y forma enlaces cruzados intercatenarios covalentes con pirimidinas, preferentemente en sitios 5'-TpA, tras la irradiación con luz ultravioleta (aproximadamente 254 nm). La fracción de psoraleno puede unirse covalentemente a un oligonucleótido modificado (por ejemplo, mediante una cadena de alcano, tal como un grupo alquilo C<sub>1-10</sub>, o un grupo polietilenglicol, tal como -(CH<sub>2</sub>CH<sub>2</sub>O)<sub>n</sub>CH<sub>2</sub>CH<sub>2</sub>-, donde n es un número entero de 1 a 50). También se pueden usar ejemplos de derivados de psoraleno, donde los derivados no limitantes incluyen 4'-(hidroxietoximetil)-4,5',8-trimetilpsoraleno (HMT-psoraleno) y 8-metoxipsoraleno.

25 Varias porciones del oligonucleótido de entrecruzamiento se pueden modificar para introducir un enlace. Por ejemplo, los fosforotioatos terminales en oligonucleótidos también se pueden usar para unir dos oligonucleótidos adyacentes. Los uracilos/citosinas halogenados también pueden usarse como modificaciones de entrecruzamiento en el oligonucleótido. Por ejemplo, los oligonucleótidos modificados con 2-fluoro-desoxiinosina (2-F-dI) se pueden hacer reaccionar con diaminas o tiopropilaminas que contienen disulfuros para formar enlaces disulfuro.

30 Como se describe a continuación, los grupos correactivos reversibles incluyen los seleccionados de un grupo cianovinilcarbazol, un grupo cianovinilo, un grupo acrilamida, un grupo tiol o sulfoniletil tioéteres. Un grupo de cianovinilcarbazol (CNV) opcionalmente sustituido también puede usarse en oligonucleótidos para formar enlaces cruzados con una base de pirimidina (por ejemplo, citosina, timina y uracilo, así como bases modificadas de los mismos) en cadenas complementarias. Los grupos CNV promueven la cicloadición [2+2] con la base de pirimidina adyacente tras la irradiación a 366 nm, lo que da como resultado un entrecruzamiento intercatenario. La irradiación a 312 nm invierte el entrecruzamiento y, por lo tanto, proporciona un método para el entrecruzamiento reversible de cadenas de oligonucleótidos. Un grupo CNV no limitante es 3-cianovinilcarbozaol, que puede incluirse como un nucleótido de carboxivinilcarbazol (por ejemplo, tal como 3-carboxivinilcarbazol-1'-β-desoxirribósido-5'-trifosfato).

40 El grupo CNV puede modificarse para reemplazar el grupo ciano reactivo con otro grupo reactivo para proporcionar un grupo vinilcarbazol opcionalmente sustituido. Los ejemplos de grupos reactivos no limitantes para un grupo vinilcarbazol incluyen un grupo amida de -CONR<sub>N1</sub>R<sub>N2</sub>, en el que cada R<sub>N1</sub> y R<sub>N2</sub> puede ser igual o diferente y es independientemente H y alquilo C<sub>1-6</sub>, por ejemplo, -CONH<sub>2</sub>; un grupo carboxilo de -CO<sub>2</sub>H; o un grupo alcocarbonilo C<sub>2-7</sub> (por ejemplo, metoxicarbonilo). Además, el grupo reactivo se puede ubicar en el carbono alfa o beta del grupo vinilo. Los ejemplos de grupos vinilcarbazol incluyen un grupo cianovinilcarbazol, tal como se describe en este documento; un grupo amidovinilcarbazol (por ejemplo, un nucleótido de amidovinilcarbazol, tal como 3-amidovinilcarbazol-1'-β-desoxirribósido-5'-trifosfato); un grupo carboxivinilcarbazol (por ejemplo, un nucleótido de carboxivinilcarbazol, tal como 3-carboxivinilcarbazol-1'-β-desoxirribósido-5'-trifosfato); y un grupo alcocarbonilvinilcarbazol C<sub>2-7</sub> (por ejemplo, un nucleótido de alcocarbonilvinilcarbazol, tal como 3-metoxicarbonilvinilcarbazol-1'-β-desoxirribósido-5'-trifosfato). Se proporcionan grupos y nucleótidos de vinilcarbazol opcionalmente sustituidos adicionales que tienen tales grupos en las fórmulas químicas de la patente de Estados Unidos No. 7.972.792 y de Yoshimura y Fujimoto, Org. Letón. 10: 3227-3230 (2008).

55 Otros grupos reactivos reversibles incluyen un grupo tiol y otro grupo tiol para formar un disulfuro, así como un grupo tiol y un grupo vinilsulfona para formar sulfoniletil tioéteres. Los grupos tiol-tiol pueden incluir opcionalmente un enlace formado por una reacción con bis-((N-yodoacetil)piperazinil)sulfonarodamina. Otros grupos reactivos reversibles (por ejemplo, tales como algunos grupos fotorreactivos) incluyen grupos benzofenona opcionalmente sustituidos. Un ejemplo no limitante es benzofenona uracilo (BPU), que se puede usar para la formación selectiva en el sitio y en la secuencia de un entrecruzamiento intercatenario de dúplex de oligonucleótidos que contienen BPU.

Este enlace cruzado puede revertirse con el calentamiento, proporcionando un método para el entrecruzamiento reversible de dos cadenas de oligonucleótidos.

En otras realizaciones, la ligación química incluye la introducción de un análogo del enlace de fosfodiéster, por ejemplo, para el análisis y la secuenciación de la PCR posterior a la selección. Ejemplos de análogos de un fosfodiéster incluyen un enlace fosforotioato (por ejemplo, como el introducido mediante el uso de un grupo fosforotioato y un grupo saliente, tal como un grupo yodo), un enlace de fosforamida o un enlace fosforoditioato (por ejemplo, como el introducido por el uso de un grupo fosforoditioato y un grupo saliente, tal como un grupo yodo).

Para cualquiera de los grupos descritos en este documento (por ejemplo, un grupo químico reactivo, un grupo fotorreactivo, una fracción de intercalación, un oligonucleótido de entrecruzamiento, o un grupo correactivo reversible), el grupo se puede incorporar en o en la proximidad del extremo de un oligonucleótido o entre los extremos 5' y 3'. Además, uno o más grupos pueden estar presentes en cada oligonucleótido. Cuando se requieren pares de grupos reactivos, entonces los oligonucleótidos se pueden diseñar para facilitar una reacción entre el par de grupos. En el ejemplo no limitante de un grupo cianovinilcarbazol que reacciona conjuntamente con una base de pirimidina, el primer oligonucleótido puede diseñarse para incluir el grupo cianovinilcarbazol en o cerca del extremo 5'. En este ejemplo, se puede diseñar un segundo oligonucleótido para que sea complementario al primer oligonucleótido y para incluir la base de pirimidina correactiva en una posición que se alinea con el grupo cianovinilcarbazol cuando se hibridan el primer y el segundo oligonucleótido. Cualquiera de los grupos de la presente invención y cualquiera de los oligonucleótidos que tienen uno o más grupos se pueden diseñar para facilitar la reacción entre los grupos para formar uno o más enlaces.

## 20 Espaciadores bifuncionales

El espaciador bifuncional entre la cabeza de dominio y la entidad química puede variarse para proporcionar una fracción espaciadora apropiada y/o para aumentar la solubilidad de la cabeza de dominio en disolvente orgánico. Existe una amplia variedad de espaciadores disponibles comercialmente que pueden acoplarse a la cabeza de dominio con la biblioteca de moléculas pequeñas. El espaciador consiste típicamente en cadenas lineales o ramificadas y puede incluir un alquilo C<sub>1-10</sub>, un heteroalquilo de 1 a 10 átomos, un alqueno C<sub>2-10</sub>, un alquino C<sub>2-10</sub>, un ariilo C<sub>5-10</sub>, un sistema cíclico o policíclico de 3 a 20 átomos, un fosfodiéster, un péptido, un oligosacárido, un oligonucleótido, un oligómero, un polímero o un poli alquilglicol (por ejemplo, un polietilenglicol, tal como - (CH<sub>2</sub>CH<sub>2</sub>O)<sub>n</sub>CH<sub>2</sub>CH<sub>2</sub>-, donde n es un número entero de 1 a 50), o combinaciones de los mismos

El espaciador bifuncional puede proporcionar una fracción espaciadora apropiada entre la cabeza de dominio y una entidad química de la biblioteca. En ciertas realizaciones, el espaciador bifuncional incluye tres partes. La parte 1 puede ser un grupo reactivo, que forma un enlace covalente con ADN, tal como, por ejemplo, un ácido carboxílico, preferiblemente activado por un éster de N-hidroxi succinimida (NHS) para reaccionar con un grupo amino en el ADN (por ejemplo, dT modificado con amino), una amidita para modificar el extremo 5' o 3' de una cabeza de dominio de cadena sencilla (logrado mediante química de oligonucleótidos estándar), pares químicos reactivos (por ejemplo, cicloadición de azido-alquino en presencia de catalizador de Cu (I), o cualquiera de los descritos en este documento), o grupos tiol reactivos. La parte 2 también puede ser un grupo reactivo, que forma un enlace covalente con la entidad química, ya sea un bloque de construcción A<sub>n</sub> o un andamio. Tal grupo reactivo podría ser, por ejemplo, una amina, un tiol, una azida o un alquino. La parte 3 puede ser una fracción espaciadora químicamente inerte de longitud variable, introducida entre la parte 1 y la 2. Tal fracción espaciadora puede ser una cadena de unidades de etilenglicol (por ejemplo, PEG de diferentes longitudes), un alcanano, un alqueno, una cadena de polieno, o una cadena peptídica. El espaciador puede contener ramificaciones o insertos con fracciones hidrofóbicas (tales como, por ejemplo, anillos de benceno) para mejorar la solubilidad de la cabeza de dominio en disolventes orgánicos, así como fracciones fluorescentes (por ejemplo, fluoresceína o Cy-3) usadas para fines de detección de bibliotecas. Los residuos hidrófobos en el diseño de la cabeza de dominio se pueden variar con el diseño del espaciador para facilitar la síntesis de la biblioteca en disolventes orgánicos. Por ejemplo, la combinación de la cabeza de dominio y el espaciador está diseñada para tener residuos apropiados en los que el coeficiente de octanol: agua (P<sub>oct</sub>) es de, por ejemplo, 1,0 a 2,5.

Los espaciadores se pueden seleccionar empíricamente para un diseño de biblioteca de molécula pequeña dado, de modo que la biblioteca se puede sintetizar en disolvente orgánico, por ejemplo, en 15%, 25%, 30%, 50%, 75%, 90%, 95 %, 98%, 99% o 100% de disolvente orgánico. El espaciador puede variarse usando reacciones modelo antes de la síntesis de la biblioteca para seleccionar la longitud de cadena apropiada que solubiliza la cabeza de dominio en un disolvente orgánico. Los ejemplos de espaciadores incluyen aquellos que tienen una longitud de cadena de alquilo aumentada, unidades de polietilenglicol aumentadas, especies ramificadas con cargas positivas (para neutralizar las cargas de fosfato negativas en la cabeza de dominio), o cantidades aumentadas de hidrofobicidad (por ejemplo, adición de estructuras de anillo de benceno).

Los ejemplos de espaciadores disponibles comercialmente incluyen espaciadores aminocarboxílicos, tales como aquellos que son péptidos (por ejemplo, Z-Gly-Gly-Gly-Osu (éster N-alfa-benciloxicarbonil-(glicina)<sub>3</sub>-N-succinimidílico) o Z-Gly-Gly-Gly-Gly-Gly-Gly-Osu (éster N-alfa-benciloxicarbonil-(glicina)<sub>6</sub>-N-succinimidílico, SEQ ID

NO: 3)), PEG (por ejemplo, Fmoc-aminoPEG2000-NHS o amino-PEG (12-24)-NHS), o cadenas ácidas de alcano (por ejemplo, ácido Boc- $\epsilon$ -aminocaproico-Osu); espaciadores de pares químicos reactivos, tales como los pares químicos reactivos descritos en este documento en combinación con una fracción peptídica (por ejemplo, azidohomoalanina-Gly-Gly-Gly-OSu (SEQ ID NO: 4) o propargilglicina-Gly-Gly-Gly-OSu (SEQ ID NO: 5)), PEG (por ejemplo, azido-PEG-NHS), o una fracción de cadena ácida de alcano (por ejemplo, ácido 5-azidopentanoico, (S)-2-(azidometil)-1-Boc-pirrolidina, 4-azidoanilina o éster de N-hidroxisuccinimida del ácido 4-azido-butan-1-oico); espaciadores reactivos con tiol, tales como los que son PEG (por ejemplo, SM(PEG)<sub>n</sub>NHS-PEG-maleimida), cadenas de alcano (por ejemplo, ácido 3-(piridin-2-ildiisulfanil)- propiónico-Osu o sulfosuccinimidil 6-(3'-[2-piridilditio]propionamido)hexanoato)); y amiditas para la síntesis de oligonucleótidos, tales como modificadores de amino (por ejemplo, 6-(trifluoroacetilamino)-hexil-(2-cianoetil)-(N,N-diisopropil)-fosforamidita), modificadores de tiol (por ejemplo, S-tritil-6-mercaptohexilo-1-[(2-cianoetil)-(N,N-diisopropil)] - fosforamidita, o modificadores de pares químicos reactivos (por ejemplo, 6-hexin-1-il-(2-cianoetil)-(N,N-diisopropil)-fosforamidita, 3-dimetoxitritiloxi-2-(3-(3-propargiloxipropanamido)propanamido)propil-1-O-succinoilo, alquilamino de cadena larga CPG, o éster N-hidroxisuccinimida de ácido 4-azido-butan-1-oico)). Se conocen espaciadores adicionales en la técnica, y los que se pueden usar durante la síntesis de la biblioteca incluyen, pero no están limitados a, 5'-O-dimetoxitritil-1',2'-didesoxirribosa-3'-[(2-cianoetil)-(N,N-diisopropil)]-fosforamidita; 9-O-dimetoxitritil-trietilenglicol, 1-[(2-cianoetil)-(N,N-diisopropil)]-fosforamidita; 3-(4,4'-dimetoxitritiloxi)propil-1-[(2-cianoetil)-(N,N-diisopropil)]-fosforamidita; y 18-O-dimetoxitritil hexaetilenglicol, 1-[(2-cianoetil)-(N,N-diisopropil)]-fosforamidita. Cualquiera de los espaciadores de la presente memoria se puede agregar en tándem entre sí en diferentes combinaciones para generar espaciadores de diferentes longitudes deseadas.

Los espaciadores también pueden estar ramificados, donde los espaciadores ramificados son bien conocidos en la técnica y los ejemplos pueden consistir en dobladores simétricos o asimétricos o un triplicador simétrico. Véase, por ejemplo, Newcome et al., *Dendritic Molecules: Concepts, Synthesis, Perspectives*, VCH Publishers (1996); Boussif et al., *Proc. Natl. Acad. Sci. USA* 92: 7297-7301 (1995); y Jansen et al., *Science* 266: 1226 (1994).

#### 25 Ligación enzimática y técnicas de ligación química

Se pueden usar diversas técnicas de ligación para añadir andamios, bloques de construcción, espaciadores, enlaces, etiquetas y/o la cabeza de dominio para producir un complejo. En consecuencia, cualquiera de las etapas de unión descritas en el presente documento puede incluir cualquier técnica de ligación útil, tal como ligación enzimática y/o ligación química. Estas etapas de unión pueden incluir la adición de una o más etiquetas a la cabeza de dominio o complejo; la adición de un espaciador a la cabeza de dominio; y la adición de uno o más andamios o bloques de construcción a la cabeza de dominio o complejo. En realizaciones particulares, las técnicas de ligación usadas para cualquier oligonucleótido proporcionan un producto resultante que puede transcribirse y/o transcribirse inversamente para permitir la decodificación de la biblioteca o para la polimerización dependiente de la plantilla con una o más ADN o ARN polimerasas.

35 En general, la ligación enzimática produce un oligonucleótido que tiene un enlace fosfodiéster nativo que puede transcribirse y/o transcribirse en forma inversa. Los métodos ilustrativos de ligación de enzimas se proporcionan en la presente memoria e incluyen el uso de una o más ARN o ADN ligasas, tales como ARN ligasa T4, ADN ligasa T4, ADNes ligasa CircLigase<sup>MR</sup>, ADNes ligasa CircLigase<sup>TM</sup> II y ADNes ligasa ThermoPhage<sup>MR</sup> (Prokaryote Ltd., Reykjavik, Islandia).

40 La ligación química también puede usarse para producir oligonucleótidos capaces de transcribirse o transcribirse de forma inversa. Puede ser necesario probar la eficacia de una técnica de ligación química para proporcionar oligonucleótidos capaces de transcribirse o transcribirse de forma inversa. Esta eficacia se puede analizar mediante cualquier método útil, tal como cromatografía líquida-espectrometría de masas, análisis de RT-PCR y/o análisis de PCR. En realizaciones particulares, la ligación química incluye el uso de uno o más pares químicos reactivos para proporcionar una fracción espaciadora que puede transcribirse o transcribirse de forma inversa. En particular, las reacciones adecuadas para pares químicos reactivos son candidatos preferidos para el proceso de ligación (Kolb et al., *Angew. Chem. Int. Ed.*, 40: 2004-2021 (2001); Van der Eycken et al., *Peine QSAR Sci.*, 26: 1115-1326 (2007)). En una realización, los oligonucleótidos ligados contienen un enlace que las polimerasas tienen una capacidad reducida para leer o translocarse, por ejemplo, un enlace "ilegible".

#### 50 Condiciones de reacción para promover la ligación enzimática o la ligación química

Los métodos descritos en este documento pueden incluir una o más condiciones de reacción que promueven la ligación enzimática o química entre la cabeza de dominio y una etiqueta o entre dos etiquetas. Estas condiciones de reacción incluyen usar nucleótidos modificados dentro de la etiqueta, como se describe en la presente memoria; usar etiquetas de donantes y etiquetas de aceptores que tienen diferentes longitudes y variar la concentración de las etiquetas; usando diferentes tipos de ligasas, así como combinaciones de las mismas (por ejemplo, ADN ligasa CircLigase<sup>MR</sup> y/o ARN ligasa T4) y variando su concentración; usando polietilenglicoles (PEG) que tienen diferentes pesos moleculares y variando su concentración; uso de agentes de agolpamiento que no sean PEG (por ejemplo, betaína o albúmina de suero bovino); variando la temperatura y la duración de la ligación; variando la concentración

de diversos agentes, incluyendo ATP,  $\text{Co}(\text{NH}_3)_6\text{Cl}_3$  y pirofosfato inorgánico de levadura; utilizando etiquetas oligonucleotídicas fosforiladas enzimática o químicamente; usando etiquetas protegidas en 3'; y usando etiquetas preadeniladas. Estas condiciones de reacción también incluyen ligaciones químicas.

5 La cabeza de dominio y/o las etiquetas pueden incluir uno o más nucleótidos modificados o sustituidos. En realizaciones preferidas, la cabeza de dominio y/o las etiquetas incluyen uno o más nucleótidos modificados o sustituidos que promueven la ligación enzimática, tales como nucleótidos 2'-O-metilo (por ejemplo, 2'-O-metilguanina o 2'-O-metiluracilo), nucleótidos 2'-fluoro o cualquier otro nucleótido modificado que se utilice como sustrato para la ligación. Alternativamente, la cabeza de dominio y/o las etiquetas se modifican para incluir uno o más grupos químicamente reactivos para soportar la ligación química (por ejemplo, un grupo alquilil opcionalmente sustituido y un grupo azido opcionalmente sustituido). Opcionalmente, los oligonucleótidos marcadores tienen derivados en ambos extremos con grupos químicamente reactivos y, opcionalmente, uno de estos extremos está protegido, de manera que los grupos pueden abordarse independientemente y las reacciones secundarias pueden reducirse (por ejemplo, reacciones secundarias de polimerización reducida).

15 La ligación enzimática puede incluir una o más ligasas. Ejemplos de ligasas incluyen ADNes ligasa CircLigase<sup>MR</sup> (EPICENTRE Biotechnologies, Madison, WI), ADNes ligasa CircLigase<sup>MR</sup> II (también de EPICENTRE Biotechnologies), ADNes ligasa ThermoPhageMR (Prokaryote Ltd., Reykjavik, Islandia), ARN ligasa T4, y ADN ligasa T4. En realizaciones preferidas, la ligación incluye el uso de una ARN ligasa o una combinación de una ARN ligasa y una ADN ligasa. La ligación puede incluir además uno o más cationes multivalentes solubles, tales como  $\text{Co}(\text{NH}_3)_6\text{Cl}_3$ , en combinación con una o más ligasas.

20 Antes o después de la etapa de ligación, el complejo se puede purificar por tres razones. En primer lugar, el complejo se puede purificar para eliminar cabezas de dominio o etiquetas sin reaccionar que pueden dar lugar a reacciones cruzadas e introducir "ruido" en el proceso de codificación. En segundo lugar, el complejo se puede purificar para eliminar cualquier reactivo o material de partida sin reaccionar que pueda inhibir o disminuir la actividad de ligación de una ligasa. Por ejemplo, el fosfato puede dar como resultado una actividad de ligación reducida. En tercer lugar, las entidades que se introducen en una etapa química o de ligación pueden necesitar ser eliminadas para permitir la siguiente etapa química o de ligación. Los métodos de purificación del complejo se describen en este documento.

30 La ligación enzimática y química puede incluir polietilenglicol que tiene un peso molecular promedio de más de 300 Dalton (por ejemplo, más de 600 Dalton, 3.000 Dalton, 4.000 Dalton o 4.500 Dalton). En realizaciones particulares, el polietilenglicol tiene un peso molecular promedio de aproximadamente 3.000 Daltons a 9.000 Daltons (por ejemplo, de 3.000 Daltons a 8.000 Daltons, de 3.000 Daltons a 7.000 Daltons, de 3.000 Daltons a 6.000 Daltons, y de 3.000 Daltons a 5.000 Daltons). En realizaciones preferidas, el polietilenglicol tiene un peso molecular promedio de aproximadamente 3.000 Dalton a aproximadamente 6.000 Dalton (por ejemplo, de 3.300 Daltons a 4.500 Daltons, de 3.300 Daltons a 5.000 Daltons, de 3.300 Daltons a 5.500 Dalton, de 3.300 Dalton a 6.000 Daltons, de 3.500 Daltons a 4.500 Daltons, de 3.500 Daltons a 5.000 Daltons, de 3.500 Daltons a 5.500 Daltons, y de 3.500 Daltons a 6.000 Daltons, tal como 4.600 Daltons). El polietilenglicol puede estar presente en cualquier cantidad útil, tal como desde aproximadamente el 25% (p/v) hasta aproximadamente el 35% (p/v), tal como el 30% (p/v).

40 En una realización preferida de esta invención, las etiquetas se instalan por ligación de un oligonucleótido de cadena sencilla con un oligonucleótido de cadena sencilla usando el protocolo de ligación descrito a continuación: Cabeza de dominio: 25  $\mu\text{M}$  (extremo 5': 5'-monofosfo/2'-OMe G, nucleótidos intervinientes: 2'-desoxi, y extremo 3': bloqueado en 2'/bloqueado en 3'); Etiqueta: 25  $\mu\text{M}$  (extremo 5': 2'-OMe/5'-OH G, nucleótidos intervinientes: 2'-desoxi, y extremo 3': 3'-OH/2'-OMe);  $\text{Co}(\text{NH}_3)_6\text{Cl}_3$ : 1 mM; PEG 4600: 30% (p/v); ARN ligasa T4 (Promega): 1,5 unidades/ $\mu\text{l}$ ; pirofosfatasa inorgánica de levadura: 0,0025 unidades/ $\mu\text{l}$ ; Tris: 50 mM;  $\text{MgCl}_2$ : 10 mM; ATP: 1 mM; pH: 7,5; y agua: hasta completar.

45 En realizaciones adicionales, el protocolo incluye la incubación a 37°C durante 20 horas. Para los propósitos de la construcción real de la biblioteca, se puede usar una mayor concentración de cabeza de dominio, etiquetas y/o ligasa, y tales modificaciones a estas concentraciones serían evidentes para los expertos en la técnica. Las metodologías de ligación química pueden incluir cualquier método químico que permita la asociación operativa de dos etiquetas de codificación, independientemente de si una polimerasa es capaz de leer o translocarse a través del enlace asociado operativamente o no.

#### Métodos para determinar la secuencia de nucleótidos de un complejo

55 Esta invención presenta un método para determinar la secuencia de nucleótidos de un complejo, de manera que se pueden establecer relaciones de codificación entre la secuencia de la etiqueta y las unidades estructurales (o bloques de construcción) de la entidad química. En particular, la identidad y/o el historial de una entidad química pueden inferirse a partir de la secuencia de bases en el oligonucleótido. Usando este método, una biblioteca que incluye diversas entidades o miembros químicos (por ejemplo, moléculas pequeñas o péptidos) puede abordarse con una secuencia de etiqueta particular.

Las Figuras 1-4 proporcionan varios ejemplos de métodos para determinar la secuencia de nucleótidos de un complejo. Estas figuras proporcionan métodos ilustrativos para abarcar oligonucleótidos entrecruzables de unión y/o liberación ilegibles, que dan como resultado la formación de una plantilla de oligonucleótido.

5 La Figura 1b proporciona un método ilustrativo que usa cebadores de relevo. En particular, los cebadores de relevo se hibridan con las porciones de secuencia constante fija (por ejemplo, los conectores 5' y 3'), donde cada secuencia para los cebadores de relevo puede ser la misma o diferente. Los cebadores de relevo abarcan el enlace ilegible, y la extensión continúa a lo largo de las partes legibles del complejo (es decir, la secuencia de codificación, por ejemplo, una o más etiquetas) para producir fragmentos de oligonucleótidos. Luego, los cebadores y fragmentos de relevo se pueden ligar para formar una plantilla, que se puede amplificar y secuenciar opcionalmente.

10 Las Figuras 2-4 proporcionan enlaces ilegibles a modo de ejemplo. La Figura 2 proporciona un enlace formado por psoraleno, que reacciona dentro de un dúplex formado por conectores 5' y 3' adyacentes. La Figura 3 proporciona un enlace formado por un oligonucleótido entrecruzado, que se hibrida con los conectores 5' y 3' adyacentes.

15 Cualquiera de los enlaces descritos en este documento puede ser reversible o irreversible. Los enlaces reversibles incluyen enlaces fotorreactivos (por ejemplo, un grupo cianovinilcarbozol y timidina, como en la Figura 4) y enlaces redox. En este documento se describen enlaces adicionales.

En una realización alternativa, un enlace "ilegible" se puede reparar enzimáticamente para generar un enlace legible o al menos translocable. Los procesos de reparación enzimáticos son bien conocidos por los expertos en la técnica e incluyen, pero no se limitan a, mecanismos de reparación de dímeros de pirimidina (por ejemplo, timidina) (por ejemplo, usando una fotoliasa o una glicosilasa (por ejemplo, dímero de pirimidina glicosilasa (PDG))), mecanismos de reparación de escisión de bases (por ejemplo, usando una glicosilasa, una endonucleasaapurínica/apirimidínica (AP), una endonucleasa Flap o una polimerasa poli ADP ribosa (por ejemplo, endonucleasaapurínica/apirimidínica (AP) humana, APE 1; proteína endonucleasa III (Nth); endonucleasa IV; endonucleasa V; formamidopirimidina [fapy]-ADN glicosilasa (Fpg); 8-oxoguanina glucosilasa humana 1 (isoforma  $\alpha$ ) (hOGG1); endonucleasa VIII humana tipo 1 (hNEIL1); uracilo-ADN glicosilasa (UDG), uracilo ADN glicosilasa monofuncional selectiva de cadena sencilla (SMUG1) y ADN glucosilasa de alquiladenina humana (hAAG), que pueden combinarse opcionalmente con una o más endonucleasas, ADN o ARN polimerasas, y/o ligasas para la reparación), mecanismos de reparación de la metilación (por ejemplo, uso de una metilguanina metiltransferasa), mecanismos de reparación de AP (por ejemplo, usando una endonucleasaapurínica/apirimidínica (AP) (por ejemplo, APE 1; endonucleasa III; endonucleasa IV; endonucleasa V; Fpg; hOGG1; y hNEIL1), que pueden combinarse opcionalmente con una o más endonucleasas, ADN o ARN polimerasas, y/o ligasas para la reparación), mecanismos de reparación de la escisión de nucleótidos (por ejemplo, utilizando proteínas de complementación cruzada de reparación de escisión o nucleasas de escisión, que pueden combinarse opcionalmente con una o más endonucleasas, ADN o ARN polimerasas, y/o ligasas para la reparación), y mecanismos de reparación de falta de coincidencia (por ejemplo, usando una endonucleasa (por ejemplo, endonucleasa I T7, MutS, MutH y MutL), que se pueden combinar opcionalmente con una o más exonucleasas, endonucleasas, helicasas, ADN o ARN polimerasas, y/o ligasas para la reparación). Se encuentran disponibles mezclas de enzimas comerciales para proporcionar fácilmente estos tipos de mecanismos de reparación, por ejemplo, mezcla de reparación PreCR® (New England Biolabs Inc., Ipswich MA), que incluye Taq ADN Ligasa, Endonucleasa IV, Bst ADN Polimerasa, Fpg, Uracil-ADN Glicosilasa (UDG), T4 PDG (T4 Endonucleasa V) y Endonucleasa VIII.

40 Métodos para etiquetar bibliotecas codificadas

Esta invención presenta un método para asociar operativamente etiquetas de oligonucleótidos con entidades químicas, de manera que se pueden establecer relaciones de codificación entre la secuencia de la etiqueta y las unidades estructurales (o bloques de construcción) de la entidad química. En particular, la identidad y/o el historial de una entidad química pueden inferirse a partir de la secuencia de bases en el oligonucleótido. Usando este método, una biblioteca que incluye diversas entidades o miembros químicos (por ejemplo, moléculas pequeñas o péptidos) puede codificarse con una secuencia de etiqueta particular.

En general, estos métodos incluyen el uso de una cabeza de dominio, que tiene al menos un grupo funcional que puede elaborarse químicamente y al menos un grupo funcional al que se puede unir (o ligar) un oligonucleótido de cadena sencilla. La unión puede realizarse por cualquier medio útil, tal como mediante unión enzimática (por ejemplo, ligación con una o más de una ARN ligasa y/o una ADN ligasa) o mediante unión química (por ejemplo, mediante una reacción de sustitución entre dos grupos funcionales, tales como un nucleofílico y un grupo saliente).

Para crear numerosas entidades químicas dentro de la biblioteca, una solución que contiene la cabeza de dominio se puede dividir en alícuotas múltiples y luego colocar en una multiplicidad de compartimentos físicamente separados, tales como los pozos de una placa de múltiples pozos. En general, este es la etapa de "división". Dentro de cada compartimento o pozo, se realizan sucesivas etapas de reacción química y ligación con una etiqueta monocatenaria dentro de cada alícuota. La relación entre las condiciones de reacción química y la secuencia de la etiqueta monocatenaria se registran. Las etapas de reacción y ligación pueden realizarse en cualquier orden. Luego,

las alícuotas que reaccionaron y se ligaron se combinan o "agrupan" y, opcionalmente, se puede realizar la purificación en este punto. Estas etapas de división y agrupación se pueden repetir opcionalmente.

A continuación, la biblioteca puede probarse y/o seleccionarse para una característica o función particular, como se describe en la presente memoria. Por ejemplo, la mezcla de entidades químicas marcadas se puede separar en al menos dos poblaciones, donde la primera población se une a un objetivo biológico particular y la segunda población no (por ejemplo, mediante selección negativa o selección positiva). La primera población puede capturarse selectivamente (por ejemplo, eluyendo en una columna que proporciona el objetivo de interés o incubando la alícuota con el objetivo de interés) y, opcionalmente, analizarse o probarse más a fondo, tal como con lavado opcional, purificación, selección negativa, selección positiva o etapas de separación.

Finalmente, las historias químicas de uno o más miembros (o entidades químicas) dentro de la población seleccionada pueden determinarse mediante la secuencia del oligonucleótido unido operativamente. Tras correlacionar la secuencia con el bloque de construcción particular, este método puede identificar los miembros individuales de la biblioteca con la característica seleccionada (por ejemplo, una mayor tendencia a unirse a la proteína objetivo y de ese modo provocar un efecto terapéutico). Para pruebas y optimización adicionales, los compuestos terapéuticos candidatos pueden prepararse entonces sintetizando los miembros de la biblioteca identificados con o sin sus etiquetas asociadas de oligonucleótidos.

Los métodos descritos en este documento pueden incluir cualquier cantidad de etapas opcionales para diversificar la biblioteca o interrogar a los miembros de la biblioteca. Para cualquier método de etiquetado descrito en este documento, se puede agregar un número sucesivo de "n" etiquetas con un número "n" adicional de etapas de ligación, separación y/o fosforilación. Las etapas opcionales a modo de ejemplo incluyen la restricción de oligonucleótidos de codificación asociados a miembros de la biblioteca usando una o más endonucleasas de restricción; reparación de los oligonucleótidos codificantes asociados, por ejemplo, con cualquier enzima reparadora, tal como las descritas aquí; ligación de una o más secuencias adaptadoras a uno o ambos extremos para oligonucleótidos codificadores asociados a miembros de la biblioteca, por ejemplo, tal como una o más secuencias adaptadoras para proporcionar una secuencia de cebado para la amplificación y secuenciación o para proporcionar una etiqueta, tal como biotina, para la inmovilización de la secuencia; transcripción inversa o transcripción, opcionalmente seguida de transcripción inversa, de las etiquetas ensambladas en el complejo usando una transcriptasa inversa, transcriptasa u otra polimerasa dependiente de la plantilla; amplificación de las etiquetas ensambladas en el complejo usando, por ejemplo, PCR; generación de aislados clonales de una o más poblaciones de etiquetas ensambladas en el complejo, por ejemplo, mediante el uso de transformación bacteriana, formación de emulsión, dilución, técnicas de captura de superficie, etc.; amplificación de aislados clonales de una o más poblaciones de etiqueta ensamblada en el complejo, por ejemplo, usando aislados clonales como plantillas para la polimerización de nucleótidos dependiente de la plantilla; y determinación de la secuencia de aislados clonales de una o más poblaciones de etiquetas ensambladas en el complejo, por ejemplo, usando aislados clonales como plantillas para la polimerización dependiente de plantilla con nucleótidos marcados en forma fluorescente. Métodos adicionales para amplificar y secuenciar las etiquetas de oligonucleótidos se describen en este documento.

Estos métodos se pueden usar para identificar y descubrir cualquier cantidad de entidades químicas con una característica o función particular, por ejemplo, en una etapa de selección. La característica o función deseada se puede usar como la base para dividir la biblioteca en al menos dos partes con el enriquecimiento concomitante de al menos uno de los miembros o miembros relacionados en la biblioteca con la función deseada. En realizaciones particulares, el método comprende identificar un miembro de la biblioteca similar a un fármaco pequeño que se une o inactiva una proteína de interés terapéutico. En otra realización, se diseña una secuencia de reacciones químicas, y se elige un conjunto de bloques de construcción de modo que la reacción de los bloques de construcción elegidos bajo las condiciones químicas definidas generará una pluralidad combinatoria de moléculas (o una biblioteca de moléculas), donde una o más moléculas pueden tener utilidad como agente terapéutico para una proteína particular. Por ejemplo, las reacciones químicas y los bloques de construcción se eligen para crear una biblioteca que tenga grupos estructurales comúnmente presentes en inhibidores de quinasas. En cualquiera de estos casos, las etiquetas de oligonucleótidos codifican la historia química del miembro de la biblioteca y, en cada caso, una colección de posibilidades químicas puede representarse mediante cualquier combinación particular de etiqueta.

En una realización, la biblioteca de entidades químicas, o una parte de las mismas, se pone en contacto con un objetivo biológico en condiciones adecuadas para que al menos un miembro de la biblioteca se una al objetivo, seguido de la eliminación de miembros de la biblioteca que no se unen al objetivo, y analizar las una o más etiquetas de oligonucleótidos asociadas con el objetivo. Este método puede incluir opcionalmente amplificar las etiquetas mediante métodos conocidos en la técnica. Los ejemplos objetivos biológicos incluyen enzimas (por ejemplo, quinasas, fosfatasa, metilasas, desmetilasas, proteasas y enzimas de reparación del ADN), proteínas implicadas en interacciones proteína:proteína (por ejemplo, ligandos para receptores), receptores objetivo (por ejemplo, GPCR y RTK), canales de iones, bacterias, virus, parásitos, ADN, ARN, priones e hidratos de carbono.

En otra realización, las entidades químicas que se unen a un objetivo no están sujetas a amplificación, sino que se analizan directamente. Los ejemplos de métodos de análisis incluyen análisis de microarreglos, que incluyen análisis

de cristal fotónico de resonancia evanescente; métodos basados en perlas para deconvolucionar etiquetas (por ejemplo, usando etiquetas de his); análisis del biosensor de cristal fotónico libre de etiquetas (por ejemplo, un lector BIND® de SRU Biosystems, Inc., Woburn, MA); o enfoques basados en hibridación (por ejemplo, usando matrices de oligonucleótidos inmovilizados complementarios a las secuencias presentes en la biblioteca de etiquetas).

5 Además, los pares químicos reactivos (o grupos funcionales) se pueden incluir fácilmente en los esquemas de síntesis de oligonucleótidos en fase sólida y apoyarán la ligación química eficiente de los oligonucleótidos. Además, los oligonucleótidos ligados resultantes pueden actuar como plantillas para la polimerización dependiente de la plantilla con una o más polimerasas. Por consiguiente, cualquiera de las etapas de unión descritas en este documento para etiquetar bibliotecas codificadas puede modificarse para incluir una o más entre técnicas de ligación  
10 enzimática y/o de ligación química. Los ejemplos de técnicas de ligación incluyen ligación enzimática, tal como el uso de una o más ARN ligasas y/o ADN ligasas; y ligación química, tal como el uso de pares químicos reactivos (por ejemplo, un par que incluye grupos funcionales alquínico y azido opcionalmente sustituidos).

Además, una o más bibliotecas se pueden combinar en una etapa de división y mezcla. Con el fin de permitir la  
15 mezcla de dos o más bibliotecas, el miembro de la biblioteca puede contener una o más secuencias que identifican la biblioteca, tal como en una etiqueta identificadora de la biblioteca, en una etiqueta ligada, o como parte de la secuencia de la cabeza de dominio, como se describe aquí.

Métodos que tienen masa reducida

Gran parte de la motivación para las estrategias de codificación de cadena sencilla surge de la masa reducida de una etiqueta de cadena sencilla en comparación con una etiqueta de doble cadena. La masa reducida  
20 potencialmente confiere varios beneficios, incluyendo una mayor solubilidad, menor costo, mayor reactividad, mayor accesibilidad al objetivo, disminución del radio hidrodinámico, mayor precisión de las evaluaciones analíticas, etc. Además de utilizar una metodología de etiquetado de cadena sencilla, se pueden lograr reducciones adicionales de la masa incluyendo el uso de uno o más de los siguientes: una o más etiquetas que tienen una longitud reducida, conjuntos de etiquetas de masa constante, una cabeza de dominio de codificación, uno o más miembros de una  
25 biblioteca que carecen de una región de unión del cebador y/o una región constante, uno o más miembros de una biblioteca que tiene una región constante reducida, o cualquier otra metodología descrita en este documento.

Para minimizar la masa de los miembros en la biblioteca, se puede reducir la longitud de una o más etiquetas, tal como hasta una longitud lo más corta posible para codificar cada tamaño dividido. En particular, las etiquetas pueden tener menos de 20 nucleótidos (por ejemplo, menos de 19 nucleótidos, menos de 18 nucleótidos, menos de  
30 17 nucleótidos, menos de 16 nucleótidos, menos de 15 nucleótidos, menos de 14 nucleótidos, menos de 13 nucleótidos, menos de 12 nucleótidos, menos de 11 nucleótidos, menos de 10 nucleótidos, menos de 9 nucleótidos, menos de 8 nucleótidos o menos de 7 nucleótidos). Como se describe a continuación en los Ejemplos, se pueden usar etiquetas más cortas (por ejemplo, aproximadamente 10 nucleótidos o más cortas) para la ligación de etiquetas.

También se pueden usar estrategias de masa constante, lo que podría ayudar en el análisis durante la síntesis de  
35 bibliotecas. Además, los conjuntos de etiquetas de masa constante podrían permitir el reconocimiento de todas las ocurrencias de errores individuales (por ejemplo, errores que surgen de una lectura errónea de una secuencia o de la unión química o enzimática de una etiqueta) y la mayoría de las ocurrencias de errores múltiples. La relación entre la longitud de un conjunto de etiquetas de cadena sencilla de masa constante y la capacidad de codificación (por ejemplo, longitudes mínimas para soportar tamaños de división de bloque de construcción específicos o identidades de biblioteca, etc.) se describe a continuación en la Tabla 1. Por consiguiente, el uso de conjuntos de etiquetas de  
40 masa constante podría usarse para proporcionar una capacidad de codificación beneficiosa, mientras se mantiene el reconocimiento de errores durante la formación de la biblioteca.

Tabla 1

Longitud	Base #1	Base #2	Base #3	Base #4	Combinaciones
1	1	0	0	0	1
2	1	1	0	0	2
3	1	1	1	0	6
4	1	1	1	1	24

5	2	1	1	1	60
6	2	2	1	1	180
7	2	2	2	1	630
8	2	2	2	2	2.520
9	3	2	2	2	7.560
10	3	3	2	2	25.200
11	3	3	3	2	92.400
12	3	3	3	3	369.600
13	4	3	3	3	1.201.200
14	4	4	3	3	4.204.200
15	4	4	4	3	15.765.750
16	4	4	4	4	63.063.000
17	5	4	4	4	214.414.200
18	5	5	4	4	771.891.120
19	5	5	5	4	2.933.186.256
20	5	5	5	5	11.732.745.024

5 Para minimizar la masa en la biblioteca, la cabeza de dominio se puede usar no solo para unir la fracción química y una etiqueta sino también para codificar la identidad de una biblioteca particular o para una etapa particular. Por ejemplo, la cabeza de dominio puede codificar información, por ejemplo, una pluralidad de cabezas de dominio que codifican la primera o primeras divisiones o la identidad de la biblioteca, tal como mediante el uso de una secuencia particular relacionada con una biblioteca específica.

10 Además, las regiones de unión del cebador (por ejemplo, constantes) de la biblioteca de entidades químicas codificadas por ADN pueden excluirse durante la etapa o etapas de selección. Entonces, estas regiones se pueden agregar después de la selección, por ejemplo, mediante ligación de cadena sencilla. Un ejemplo de estrategia incluiría proporcionar una entidad química en el extremo 5' de un oligonucleótido codificante, seleccionar una entidad química particular con base en cualquier característica o función particular útil, y ligar un oligonucleótido de la cola de dominio al extremo 3' del oligonucleótido codificante que incluye una secuencia de unión del cebador y puede contener opcionalmente una o más etiquetas, por ejemplo una etiqueta de "uso", una etiqueta de "origen", etc., como se describe en este documento. Esta secuencia de unión al cebador podría usarse luego para iniciar la polimerización dependiente de la plantilla para generar ADNc (o ARNc) que sea complementario al miembro de la biblioteca seleccionado. El ADNc o ARNc se ligaría entonces en su extremo 3' a un oligonucleótido que contiene una secuencia de unión al cebador y, ahora que la información de codificación está flanqueada en ambos lados por secuencias de unión al cebador, el oligonucleótido puede secuenciarse y/o amplificarse usando enfoques establecidos, tal como cualquiera descrito aquí.

20 La masa se puede minimizar adicionalmente omitiendo o reduciendo el tamaño de una o más secuencias constantes que separan las etiquetas de codificación. La ligación de cadena sencilla no requiere una relación complementaria entre los extremos a ligar o entre estos extremos y un soporte. Por lo tanto, no se requiere una secuencia fija para soportar la ligación enzimática. Las regiones fijas cortas entre etiquetas pueden ser útiles para el análisis informático de etiquetas u otros procesos de deconvolución *in silico*.

## Métodos para codificar entidades químicas dentro de una biblioteca

- Los métodos de la invención se pueden usar para sintetizar una biblioteca que tiene un número diverso de entidades químicas que están codificadas por etiquetas de oligonucleótidos. Ejemplos de bloques de construcción y etiquetas de ADN codificantes se encuentran en la publicación de la solicitud de patente de Estados Unidos No. 2007/0224607, los documentos WO2010094036, WO2009077173, WO2007062664, WO2011127933 y WO2005058479. Cada entidad química está formada por uno o más bloques de construcción y, opcionalmente, un andamio. El andamio sirve para proporcionar uno o más nodos de diversidad en una geometría particular (por ejemplo, una triazina para proporcionar tres nodos dispuestos espacialmente alrededor de un anillo de heteroarilo o una geometría lineal).
- Los bloques de construcción y sus etiquetas de codificación se pueden añadir directa o indirectamente (por ejemplo, a través de un espaciador) a la cabeza de dominio para formar un complejo. Cuando la cabeza de dominio incluye un espaciador, el bloque de construcción o el andamio se agrega al extremo del espaciador. Cuando el espaciador está ausente, el bloque de construcción se puede agregar directamente a la cabeza de dominio o el propio bloque de construcción puede incluir un espaciador que reacciona con un grupo funcional de la cabeza de dominio. Ejemplos de espaciadores y cabezas de dominio se describen en el presente documento.

- El andamio se puede agregar de cualquier manera útil. Por ejemplo, el andamio se puede agregar al extremo del espaciador o a la cabeza de dominio, y se pueden agregar bloques de construcción sucesivos a los nodos de diversidad disponibles del andamio. En otro ejemplo, el bloque de construcción  $A_n$  se agrega primero al espaciador o a la cabeza de dominio, y luego el nodo de diversidad del andamio  $S$  se hace reaccionar con un grupo funcional en el bloque de construcción  $A_n$ . Las etiquetas de oligonucleótidos que codifican un andamio particular se pueden agregar opcionalmente a la cabeza de dominio o al complejo. Por ejemplo, se agrega  $S_n$  al complejo en  $n$  recipientes de reacción, donde  $n$  es un número entero superior a uno, y la etiqueta  $S_n$  (es decir, la etiqueta  $S_1, S_2, \dots, S_{n-1}, S_n$ ) está unida al grupo funcional del complejo.

- Los bloques de construcción se pueden agregar en múltiples etapas de síntesis. Por ejemplo, una alícuota de la cabeza de dominio, que tiene opcionalmente un espaciador unido, se separa en  $n$  recipientes de reacción, donde  $n$  es un número entero de dos o más. En la primera etapa, se agrega el bloque de construcción  $A_n$  a cada recipiente de reacción  $n$  (es decir, el bloque de construcción  $A_1, A_2, \dots, A_{n-1}, A_n$  se agrega al recipiente de reacción 1, 2, ...  $n-1, n$ ), donde  $n$  es un número entero y cada bloque de construcción es único. En la segunda etapa, se agrega el andamio  $S$  a cada recipiente de reacción para formar un complejo  $A_n-S$ . Opcionalmente, el andamio  $S_n$  puede agregarse a cada recipiente de reacción para formar un complejo  $A_n-S_n$ , donde  $n$  es un número entero de más de dos, y cada andamio  $S_n$  puede ser único. En el tercer paso, se construye el bloque  $B_n$  a cada  $n$  recipiente de reacción que contiene el complejo  $A_n-S$  (es decir, el bloque de construcción  $B_1, B_2, \dots, B_{n-1}, B_n$  se agrega al recipiente de reacción 1, 2, ...  $n-1, n$  que contiene el complejo  $A_1-S, A_2-S, \dots, A_{n-1}-S, A_n-S$ ), donde cada bloque de construcción  $B_n$  es único. En etapas adicionales, puede agregarse el bloque de construcción  $C_n$  a cada recipiente de reacción  $n$  que contiene el complejo  $B_n-A_n-S$  (es decir, el bloque de construcción  $C_1, C_2, \dots, C_{n-1}, C_n$  se agrega al recipiente de reacción 1, 2, ...  $n-1, n$  que contiene el complejo  $B_1-A_1-S \dots B_n-A_n-S$ ), donde cada bloque de construcción  $C_n$  es único. La biblioteca resultante tendrá un número  $n^3$  de complejos que tienen  $n^3$  etiquetas. De esta manera, se pueden usar etapas de síntesis adicionales para unir bloques de construcción adicionales para diversificar aún más la biblioteca.

- Después de formar la biblioteca, los complejos resultantes pueden purificarse opcionalmente y someterse a una reacción de polimerización o ligación, por ejemplo, a una cola de dominio. Esta estrategia general se puede ampliar para incluir nodos de diversidad adicionales y bloques de construcción (por ejemplo, D, E, F, etc.). Por ejemplo, el primer nodo de diversidad se hace reaccionar con bloques de construcción y/o  $S$  y se codifica mediante una etiqueta de oligonucleótido. A continuación, se hacen reaccionar bloques de construcción adicionales con el complejo resultante, y el nodo de diversidad posterior forma derivados mediante bloques de construcción adicionales, que se codifica mediante el cebador utilizado para la reacción de polimerización o ligación.

- Para formar una biblioteca codificada, se añaden etiquetas de oligonucleótidos al complejo después o antes de cada etapa de síntesis. Por ejemplo, antes o después de la adición del bloque de construcción  $A_n$  a cada recipiente de reacción, el marcador  $A_n$  se une al grupo funcional de la cabeza de dominio (es decir, la etiqueta  $A_1, A_2, \dots, A_{n-1}, A_n$  se agrega al recipiente de reacción 1, 2, ...  $n-1, n$  que contiene la cabeza de dominio). Cada etiqueta  $A_n$  tiene una secuencia distinta que se correlaciona con cada bloque de construcción único  $A_n$ , y determina la secuencia de la etiqueta  $A_n$  que proporciona la estructura química del bloque de construcción  $A_n$ . De esta manera, se usan etiquetas adicionales para codificar bloques de construcción adicionales o andamios adicionales.

- Además, la última etiqueta añadida al complejo puede incluir una secuencia de unión al cebador o proporcionar un grupo funcional para permitir la unión (por ejemplo, mediante ligación) de una secuencia de unión al cebador. La secuencia de unión al cebador se puede usar para amplificar y/o secuenciar las etiquetas de oligonucleótidos del complejo. Los ejemplos de métodos para amplificar y secuenciar incluyen la reacción en cadena de la polimerasa

(PCR), la amplificación de cadena lineal (LCR), la amplificación por círculo rodante (RCA) o cualquier otro método conocido en la técnica para amplificar o determinar secuencias de ácidos nucleicos.

Usando estos métodos, se pueden formar grandes bibliotecas que tienen una gran cantidad de entidades químicas codificadas. Por ejemplo, una cabeza de dominio se hace reaccionar con un espaciador y un bloque de construcción  $A_n$ , que incluye 1.000 variantes diferentes (es decir,  $n = 1.000$ ). Para cada bloque de construcción  $A_n$ , se liga una etiqueta de ADN  $A_n$  o un cebador extendido a la cabeza de dominio. Estas reacciones se pueden realizar en una placa de 1.000 pozos o en placas de 10 x 100 pozos. Todas las reacciones pueden agruparse, opcionalmente purificarse, y dividirse en un segundo conjunto de placas. A continuación, se puede realizar el mismo procedimiento con el bloque de construcción  $B_n$ , que también incluye 1.000 variantes diferentes. Una etiqueta de ADN  $B_n$  puede ser ligada al complejo  $A_n$ -cabeza de dominio, y todas las reacciones pueden agruparse. La biblioteca resultante incluye 1.000 x 1.000 combinaciones de  $A_n$  x  $B_n$  (es decir, 1.000.000 de compuestos) etiquetados por 1.000.000 de combinaciones diferentes de etiquetas. El mismo enfoque puede extenderse para agregar los bloques de construcción  $C_n$ ,  $D_n$ ,  $E_n$ , etc. La biblioteca generada se puede usar luego para identificar compuestos que se unen al objetivo. La estructura de las entidades químicas que se unen a la biblioteca se puede evaluar opcionalmente mediante PCR y la secuenciación de las etiquetas de ADN para identificar los compuestos que se enriquecieron.

Este método se puede modificar para evitar el etiquetado después de la adición de cada bloque de construcción o para evitar el agrupamiento (o la mezcla). Por ejemplo, el método puede modificarse añadiendo el bloque de construcción idéntico  $A_n$  a  $n$  recipientes de reacción, donde  $n$  es un número entero de más de uno, y añadiendo el bloque de construcción  $B_1$  idéntico a cada pozo de reacción. Aquí,  $B_1$  es idéntico para cada entidad química, y, por lo tanto, no es necesaria una etiqueta de oligonucleótido que codifique este componente. Después de agregar un bloque de construcción, los complejos se pueden agrupar o no agrupar. Por ejemplo, la biblioteca no se agrupa después de la etapa final de adición del bloque de construcción, y las agrupaciones se seleccionan individualmente para identificar el o los compuestos que se unen a un objetivo. Para evitar agrupar todas las reacciones después de la síntesis, se puede usar, por ejemplo, un lector BIND® (de SRU Biosystems, Inc.), por ejemplo, para monitorizar la unión en una superficie del sensor en formato de alto rendimiento (por ejemplo, placas de 384 pozos y placas de 1.536 pozos). Por ejemplo, el bloque de construcción  $A_n$  se puede codificar con la etiqueta  $A_n$  del ADN, y el bloque de construcción  $B_n$  se puede codificar por su posición dentro de la placa del pozo. Los compuestos candidatos pueden identificarse luego utilizando un ensayo de unión (por ejemplo, usando un Biosensor BIND®, también disponible a través de SRU Biosystems, Inc., o usando un ensayo ELISA) y analizando las etiquetas  $A_n$  por secuenciación, análisis de microarreglos y/o análisis de restricción de la digestión. Este análisis permite la identificación de combinaciones de bloques de construcción  $A_n$  y  $B_n$  que producen las moléculas deseadas.

El método de amplificación puede incluir opcionalmente formar una emulsión de agua en aceite para crear una pluralidad de microrreactores acuosos. Las condiciones de reacción (por ejemplo, concentración de complejo y tamaño de microrreactores) pueden ajustarse para proporcionar, en promedio, un microrreactor que tiene al menos un miembro de una biblioteca de compuestos. Cada microrreactor también puede contener el objetivo, una única perla capaz de unirse a un complejo o una porción del complejo (por ejemplo, una o más etiquetas) y/o unirse al objetivo, y una solución de reacción de amplificación que tiene uno o más reactivos necesarios para realizar la amplificación del ácido nucleico. Después de amplificar la etiqueta en los microrreactores, las copias amplificadas de la etiqueta se unirán a las perlas en los microrreactores, y las perlas recubiertas se pueden identificar por cualquier método útil.

Una vez que se han identificado los bloques de construcción de la primera biblioteca que se unen al objetivo de interés, se puede preparar una segunda biblioteca de forma iterativa. Por ejemplo, se pueden agregar uno o dos nodos adicionales de diversidad, y se crea y se muestra la segunda biblioteca, como se describe aquí. Este proceso se puede repetir tantas veces como sea necesario para crear moléculas con las propiedades moleculares y farmacéuticas deseadas.

Se pueden usar diversas técnicas de ligación para agregar el andamio, los bloques de construcción, los espaciadores, los enlaces y las etiquetas. En consecuencia, cualquiera de las etapas de unión descritas en el presente documento puede incluir cualquier técnica o técnicas de ligación útiles. Los ejemplos de técnicas de ligación incluyen la ligación enzimática, tal como el uso de una o más ARN ligasas y/o ADN ligasas, como se describe en la presente memoria; y ligación química, tal como el uso de pares químicos reactivos, como se describe en el presente documento.

### Ejemplo 1

#### Estrategia general para complejos

Los complejos de la invención se pueden formar para incluir uno o más enlaces que una polimerasa tiene capacidad reducida para leer o translocar a través del enlace. Por ejemplo, se pueden formar complejos para incluir (de la dirección 5' a 3') una entidad química (estrella), una secuencia constante fija (por ejemplo, una cabeza de dominio), seguido de tres secuencias de codificación variable (por ejemplo, tres etiquetas), y otra secuencia constante fija (por

ejemplo, una cola de dominio), donde el enlace entre las secuencias constantes fijas y las secuencias de codificación variables puede ser ilegible (Figura 1B).

Los enlaces ilegibles pueden formarse por cualquier fracción o grupo funcional útil. Un ejemplo no limitante es psoraleno (Figura 2), que puede formar un enlace ilegible entre los conectores 5' y 3'.

5 Se pueden formar enlaces ilegibles mediante un oligonucleótido de entrecruzamiento con un grupo correactivo reversible de un grupo cianovinilcarbazol. El complejo puede incluir (de la dirección 5' a 3') una entidad química (estrella), una secuencia constante fija (por ejemplo, una cabeza de dominio), un oligonucleótido de entrecruzamiento que está entrecruzado a través de un grupo correactivo reversible de un grupo cianovinilcarbazol (entrecruzado a aproximadamente 366 nm, marcado con X), y seguido por una secuencia de codificación variable  
10 que tiene secuencias constantes fijas en los extremos 5' y 3' (por ejemplo, una etiqueta que tiene conectores 5' y 3') (Figura 3). Un ejemplo de una reacción reversible con grupos correactivos reversibles (un par de un grupo cianovinilcarbazol y una timidina), que puede estar presente en el conector 3', el conector 5' y el oligonucleótido de entrecruzamiento (Figura 4). Otros enlaces ilegibles pueden volverse deconvolucionables mediante el uso de una enzima o enzimas de reparación.

15 Como parte del proceso de lectura o deconvolución, el complejo puede hacerse reaccionar con un proceso químico y/o un proceso enzimático (por ejemplo, una ligasa y opcionalmente una quinasa) para ligar las etiquetas. Cuando la unión entre etiquetas adyacentes es una muesca, entonces la ligasa puede usarse para ligar las etiquetas. Opcionalmente, se puede usar una quinasa para convertir grupos 5'-hidroxilo terminales en grupos fosfato. Cuando la unión entre etiquetas adyacentes es un espacio que requiere extensión con uno o más nucleótidos, entonces se  
20 puede usar una polimerasa para extenderse a través del espacio antes de la ligación de las etiquetas (que incluye opcionalmente quinasa). Opcionalmente, la polimerasa tiene una actividad de desplazamiento de cadena reducida. A continuación, los oligonucleótidos de entrecruzamientos se liberan para formar una plantilla (por ejemplo, usando irradiación a aproximadamente 312 nm). Finalmente, la plantilla se amplifica opcionalmente con PCR y se secuencia.

25 Para cualquiera de estos complejos, cada una de las secuencias de codificación variable puede incluir opcionalmente un conector 5' y un conector 3' que son secuencias constantes fijas. Debido a que los enlaces son ilegibles, los cebadores de relevo pueden usarse para abarcar los enlaces y para permitir la extensión con una polimerasa, formando así fragmentos de oligonucleótidos. Los fragmentos pueden luego ligarse usando ligasa, opcionalmente amplificarse con PCR, y secuenciarse.

30 Alternativamente, las secuencias codificantes en cualquier lado de un enlace ilegible pueden copiarse en una sola secuencia de ADNc usando un proceso mediado por recombinación en el que se favorece que ocurran eventos de recombinación para secuencias derivadas de la misma secuencia de plantilla ancestral. Este mecanismo conserva la información de asociación de la etiqueta y hace que estas asociaciones sean deducibles de los datos de secuencias derivadas. Un posible método para generar enlaces ilegibles que pueden convertirse en plantillas para el ADNc propenso a la recombinación es establecer repeticiones de regiones homólogas a cada lado del enlace.

## 35 Ejemplo 2

Entrecruzamiento fotoquímico de oligonucleótidos utilizando cianovinil carbazol para modelar un evento de etiquetado

40 El oligonucleótido CNVKJ, que contiene dos modificaciones de cianovinil carbazol, se obtuvo de Nihon Techno Service (Japón) (Figura 5). Los oligonucleótidos CNVK2\_P\_TagB, CNVK2\_NP\_TagB y CNVK2\_TagA se obtuvieron de Integrated DNA Technologies (IDT) (IA). Cada oligonucleótido se disolvió en agua a una concentración de 1 mM

45 Las condiciones de entrecruzamiento fotoquímico fueron las siguientes: Oligonucleótidos CNVKJ, CNVK2\_P\_TagB y CNVK2\_TagA se mezclaron a una relación equimolar a 100 uM cada uno en regulador de fosfato 500 mM a pH 7,0 en alícuotas de 20 µL o 50 µL en tubos de microcentrífuga de polipropileno de 1,5 mL de color natural (Fisher Scientific, 02-682-550). A continuación, se irradió cada reacción usando una lámpara UVL-21 (4 vatios) (95-0018-02) (Ultra Violet Products, CA) a 365 nm durante 2 horas a 4°C. Los productos de las reacciones se analizaron luego por electroforesis en un PAAG desnaturante al 10%, así como por LCMS (Thermo Fisher Scientific) (Figura 6). Más del 70% del material de partida se convirtió en el conjugado de oligonucleótido entrecruzado fotoquímicamente con un PM observado de 16.894 Da (calculado 16.890 Da). Cuando se realizó la misma reacción de entrecruzamiento en regulador de fosfato 500 mM a pH 5,5 o regulador de borato 500 mM a pH 9,5, se observaron resultados y  
50 rendimientos similares.

El conjugado de oligonucleotídico fotoquímicamente entrecruzado de CNVK2\_P\_TagB, CNVK2\_TagA y CNVKJ se purificó en un PAAG desnaturante al 10% y no se entrecruzó calentando a 80°C en alícuotas de 50 µL en tubos de microcentrífuga de polipropileno de 1,5 ml de color natural (Fisher Scientific, 02-682-550) con irradiación (a 80°C) usando una lámpara Spectroline E-Series EB-160C (6 vatios) (Spectronics Corp., NY) a 312 nm durante 1 hora. Los

productos fueron analizados por LCMS. La reacción produce cuantitativamente etiquetas de modelo de oligonucleótido de partida inalterado CNVK2\_P\_TagB con PM observado 6.578 Da (calculado 6.579 Da) y CNVK2\_TagA PM observado 5.558 Da (calculado 5.558 Da) (Figura 7).

5 Para estudiar la disociación de múltiples oligonucleótidos entrecruzados que contienen CNVK, se adaptó el protocolo anterior para preparar un complejo con dos oligos CNVK, CNVKJ y Splint\_14\_CNVK, (Splint\_14\_CNVK es CGAXCGTGTCA XCG, donde X es CNVK, sintetizado por Biosearch Technologies, CA), entrecruzado fotoquímicamente a un solo oligonucleótido, Double\_CNVK\_Template (AAAAAAGTCGTGACACGTCGGAAAAAAAAAAAAACGGTGACACG GTCGAAAAAA, IDT (IA)). Este complejo se analizó mediante LCMS: PM observado 25.079 Da (calculado 25.055 Da). Luego se disolvió en agua hasta una concentración de 100 µM y se irradió con UV 312 nm durante 1 hora a 80°C, nuevamente como se describió anteriormente. Los productos se analizaron mediante LCMS para encontrar la recuperación de la plantilla Double\_CNVK inicial: PM observado 16.492 Da (calculado 16.492 Da) (Figura 8).

### Ejemplo 3

15 Ligación y desentrecruzamiento fotoquímico de oligonucleótidos entrecruzados fotoquímicamente con cianovinilcarbazol para modelar la recuperación de la información de secuencia a partir de etiquetas fotoquímicamente entrecruzadas

Se preparó un complejo de CNVK2\_P\_TagB, CNVK2\_TagA y CNVKJ, se entrecruzó fotoquímicamente y se purificó en gel como se describe en el Ejemplo 2.

20 El par de etiquetas de modelo fotoquímicamente entrecruzadas se disolvió en 1x regulador de ADN ligasa T4 (NEB, MA) y se incubó con ADN ligasa T4 (NEB, MA) durante 1 hora a 37°C. El producto de esta reacción se entrecruzó luego como se describe en el Ejemplo 2 y se analizó mediante electroforesis en gel desnaturizante y LCMS (Figura 9). El producto de la reacción de ligación se observó con un alto rendimiento, se observó un PM de 12.119 Da (calculado 12.119 Da).

25 También se preparó el par correspondiente de etiquetas de modelo no fosforiladas fotoquímicamente entrecruzado CNVK2\_NP\_TagB con CNVK2\_TagA y CNVKJ, como se describió anteriormente. El análisis por LCMS de este complejo reveló un PM observado de 16.503 Da (calculado 16.499 Da, Figura 10). El complejo entrecruzado fotoquímicamente se disolvió en 1x regulador de ligasa T4 (NEB) y se incubó con una mezcla de polinucleótido quinasa T4 y ADN ligasa T4 (NEB) durante 1 hora a 37°C. El producto de esta reacción de ligación por fosforilación acoplada se entrecruzó después fotoquímicamente mediante irradiación con UV a 312 nm durante 1 hora a 80°C como se describe en el Ejemplo 2. El producto de esta reacción de fosforilación-ligación se observó mediante electroforesis en gel desnaturizante y LCMS, revelando un PM observado de 11.894 Da (PM calculado 11.896 Da (fosforilación adicional en el extremo 5' del producto ligado)) (Figura 10).

### Ejemplo 4

35 Entrecruzamiento fotoquímico de oligonucleótidos utilizando psoraleno dentro de un tallo para modelar un evento de marcado

40 El oligonucleótido 5PSO2\_A9\_TA, modificado con C2-Psoraleno en el extremo 5' se obtuvo de Biosearch, CA (Figura 11). El oligonucleótido PSO\_HP\_A9\_TCT se obtuvo de IDT (IA) (Figura 11). Ambos oligonucleótidos se disolvieron a una concentración de 0,5 mM cada uno en regulador de fosfato 500 mM a pH 7,0 y se hibridaron calentando a 95°C y posteriormente enfriando a temperatura ambiente. La reacción se irradió luego con luz UV a 365 nm a 4°C usando una lámpara UV compacta UVL-21 (UVP) durante 30 minutos. Los productos fueron analizados por LCMS. Se observó una rápida formación del producto fotoquímicamente entrecruzado con un PM observado de 17.643 Da (PM calculado 17.643,4 Da, véase la Figura 12).

### Ejemplo 5

Ligación fotoquímica de oligonucleótidos utilizando psoraleno para modelar un evento de etiquetado

45 Los oligonucleótidos Tag1\_PsoCVU y SplintC\_PsoC2 se obtuvieron a través de IDT (IA) (Figura 13). El oligonucleótido 5PSOC2\_A9\_GA (modificado en el extremo 5' con psoraleno C2) se obtuvo de Biosearch (CA) (Figura 13).

50 El conjugado de oligonucleotídico ligado fotoquímicamente de Tag1\_PsoCVU y 5PSOC2\_A9\_GA se preparó de la siguiente manera. La fotoligación se logró mediante la reacción fotoquímica de la fracción 5'-psoralenopirona de un oligonucleótido con la timidina 3' de un segundo oligonucleótido en presencia de un tercer oligonucleótido diseñado para hibridar con el primer y el segundo oligonucleótidos y ubicar conjuntamente sus extremos reactivos (Figura 14).

- Las condiciones de ligación fotoquímica fueron las siguientes: los oligonucleótidos 5PsoC2\_A9\_GA (1 mM final) y Tag1\_PsoCVU (1,1 mM final) y SplintC\_PsoC2 (1,1 mM final) se combinaron en una reacción de fotoligación de 10  $\mu$ L en regulador de fosfato 500 mM a pH 7,0 en tubos de microcentrífuga de polipropileno de 1,5 mL de color natural (Fisher Scientific, 02-682-550). La mezcla de reacción se calentó primero a 95°C seguido de enfriamiento lento a temperatura ambiente. La reacción se irradió luego con luz UV a 365 nm a 4°C usando una lámpara UV compacta UVL-21 (UVP). Se tomaron alícuotas de 1  $\mu$ L en el transcurso del tiempo y se analizaron en un gel de poliacrilamida desnaturalizante de Urea TBE/8M al 15% (Figura 15). El gel se visualizó y se fotografió usando el sombreado UV de una placa fluorescente de TLC. El conjugado fotoligado de 5PsoC2\_A9\_GA y Tag1\_PsoCVU se formó con alto rendimiento. Después de 24 horas, el rendimiento es aproximadamente del 90% (Figura 15).
- La muestra de 24 horas se analizó mediante LCMS (Figura 16). El pico máximo de LC correspondió al conjugado fotoligado de 5PsoC2\_A9\_GA y Tag1\_PsoCVU con un PM observado de 15.252.4 Da (PM calculado 15.243,1 Da).

### Ejemplo 6

- Generación de cDNA a partir de etiquetas de modelo conjugadas con psoraleno ligadas fotoquímicamente no legibles con polimerasa, utilizando extensión del cebador terminal, extensión de cebador no terminal y ligación para modelar la lectura de un solo evento de etiquetado

- Se generó un conjugado de oligonucleótido ligado fotoquímicamente con psoraleno-timidina (Bio\_Pso) usando un método similar al descrito en el Ejemplo 5 con los siguientes dos oligonucleótidos de etiqueta modelo: un oligonucleótido 5'-biotinilado 5Bio\_Tag\_PsoCVU y un oligonucleótido 5'-Psoraleno 5PsoC2\_A9\_GA junto con un oligonucleótido de soporte SplintC\_PsoC2, los tres oligonucleótidos se obtuvieron a través de IDT (IA) y se muestran en la Figura 17. Después de 24 horas de la reacción, se purificó Bio\_Pso en un PAAG desnaturalizante con TBE-Urea al 10% y se analizó mediante LCMS, revelando un producto con un PM observado de 15.834,0 Da (PM calculado 15.812,7 Da) (Figura 18).

- El conjugado de oligonucleótido ligado fotoquímicamente de psoraleno-timidina Bio\_Pso se usó como plantilla para generar ADNc mediante la extensión enzimática tanto de un cebador oligonucleotídico terminal (FAMprimer) como de un cebador oligonucleotídico no terminal (Phos-SplintC\_PsoC2) con ADN polimerasa T4 (NEB, MA). Se generó Phos-SplintC\_PsoC2 mediante la fosforilación 5' de SplintC\_PsoC2 usando T4 PNK (NEB, MA). La generación de ADNc se realizó en una reacción de 100  $\mu$ L que contenía 10  $\mu$ M de FAMprimer, Phos-SplintC\_PsoC2 y Bio\_Pso en 1x regulador ADN ligasa T4 (NEB, MA), complementado con 1 mM de cada dNTP y 10  $\mu$ L de ADN polimerasa T4 (NEB, MA). La mezcla de reacción se incubó a 37°C durante 1 hora y luego se complementó con ATP 0,5 mM y 5  $\mu$ L de ADN ligasa T4 (2000  $\mu$ /L, NEB, MA) y luego se incubó durante 1 hora más a 37°C.

- La mezcla de reacción se incubó luego con 200  $\mu$ L de DynaBeads M280 recubiertas con estreptavidina (Invitrogen, prelavadas con PBS durante 1 hora a temperatura ambiente), las perlas se lavaron con 1 mL de PBS y el producto se eluyó con 35  $\mu$ L de NaOH 100 mM. El eluyente se neutralizó luego mediante la adición de 5  $\mu$ L de Tris HCl 1 M pH 7,0 y se analizó mediante LCMS.

- El análisis por LCMS del producto resultante, con detección a 495 nm, mostró que aproximadamente el 50% del FAMprimer se había extendido a la secuencia complementaria de longitud completa a la plantilla Bio\_Pso menos un solo nucleótido dA (PM observado de 15.170 Da, PM de longitud completa esperado 15.458 Da). La mayor parte del resto del FAMprimer se extendió hasta la unión de ligación fotoquímica (PM observado 11.154 Da, PM calculado 11.154 Da) (Figura 19).
- Cuando la generación de ADNc, descrita anteriormente, se realizó sin la adición de la enzima ligasa, no se observó la secuencia complementaria de longitud completa (menos un solo dA) (Figura 20).

### Ejemplo 7

Ligación fotoquímica de oligonucleótidos utilizando carboxivinil uridina para modelar un evento de etiquetado

- Los oligonucleótidos CVU\_G y CVU\_A se sintetizaron por Trilink (CA) (Figura 21). Ambos oligonucleótidos están modificados en 5' con una 5-(Carboxi)vinil-2'-desoxiuridina (Figura 21). Los oligonucleótidos Tag1\_PsoCVU, SplintC\_CVU y SplintA\_CVU se obtuvieron a partir de ADN IDT (IA) (Figura 21).

La estructura de la unión de ligación fotoquímica formada entre una 5'-5-(Carboxi)vinil-2'-desoxiuridina irradiada y una 3'-timidina se muestra en la Figura 22.

- Se generó una mezcla de reacción de ligación fotoquímica de 10  $\mu$ L con CVU\_G 1 mM, Tag1\_PsoCVU 1,1 mM y SplintC\_CVU 1,1 mM disueltos en regulador de fosfato de sodio 500 mM pH 7,0 en tubos de microcentrífuga de polipropileno de 1,5 mL de color natural (Fisher Scientific, 02-682-550). La mezcla de reacción se calentó a 95°C

seguido de enfriamiento lento a temperatura ambiente. La reacción se irradió a continuación con luz UV a 365 nm durante 24 horas a 4°C usando una lámpara UV compacta UVL-21 (UVP). Se tomaron alícuotas de 1 µL en el transcurso del tiempo y se sometieron a electroforesis en un gel de poliacrilamida desnaturante con Urea al 15% de TBE/8M (Figura 23). El gel se visualizó y fotografió mediante sombreado UV de una placa fluorescente de TLC.

5 Después de 24 horas de irradiación, más del 90% del oligonucleótido CVU\_G había sido ligado fotoquímicamente.

La muestra de 24 horas se analizó mediante LCMS (Figura 24). El pico principal corresponde al producto de ligación fotoquímica de CVU\_G con Tag1\_PsoCVU. El PM observado fue 15.260.6 Da (PM calculado 15.238,7 Da).

### Ejemplo 8

10 Generación de ADNc a partir de etiquetas modelo conjugadas con carboxivinil uridina ligadas fotoquímicamente no legibles por polimerasa, que utilizan extensión del cebador terminal, extensión del cebador no terminal y ligación para modelar la lectura de un único evento de etiquetado

15 Se generó un conjugado de oligonucleótido ligado fotoquímicamente a carboxivinil uridina-timidina (Bio\_CVU) usando un método similar al descrito en el Ejemplo 7 con los siguientes dos oligonucleótidos de etiqueta modelo: Un oligonucleótido 5'-5-carboxivinil Uridina 5'-CVU\_A y un oligonucleótido 5'-biotinilado 5Bio\_Tag\_PsoCVU junto con un oligonucleótido de soporte SplintA\_CVU, los tres oligonucleótidos se obtuvieron a través de IDT (IA) y se muestran en las Figuras 21 y 25. El producto de reacción se purificó en gel como se describe en el Ejemplo 6 y el análisis por LCMS reveló un producto con un PM observado de 15.811,8 Da (PM calculado 15.792,3 Da) (Figura 25).

20 Se usó el conjugado Bio\_CVU de oligonucleótido ligado fotoquímicamente con carboximetil uridina-timidina como plantilla para generar ADNc mediante la extensión enzimática de un cebador oligonucleotídico terminal (FAMprimer) y un cebador oligonucleotídico no terminal (Phos-SplintA\_CVU) con ADN polimerasa T4 (NEB, MA). Se generó Phos-SplintA\_CVU por la 5'-fosforilación de SplintA\_CVU utilizando T4 PNK (NEB, MA). La generación de ADNc se realizó en una reacción de 100 µL que contenía 10 µM de FAMprimer, Phos-SplintA\_CVU y Bio\_CVU en 1x regulador ADN ligasa T4 (NEB, MA), complementado con 1 mM de cada dNTP y 10 µL de ADN polimerasa T4 (NEB, MA). La mezcla de reacción se incubó a 37°C durante 1 hora y luego se complementó con ATP 0,5 mM y 5 µL de ADN ligasa T4 (2000 µ/L, NEB, MA) y luego se incubó durante 1 hora más a 37°C.

25 La mezcla de reacción se incubó luego con 200 µL de DynaBeads M280 recubiertas con estreptavidina (Invitrogen, prelavadas con PBS durante 1 hora a temperatura ambiente), las perlas se lavaron con 1 mL de PBS y el producto se eluyó con 35 µL de NaOH 100 mM. El eluyente se neutralizó luego mediante la adición de 5 µL de Tris HCl 1 M pH 7,0 y se analizó mediante LCMS.

30 El análisis por LCMS del producto resultante, con detección a 495 nm, mostró que aproximadamente el 80% del FAMprimer se había extendido a la secuencia complementaria de longitud completa a la plantilla Bio\_CVU menos un único nucleótido dA (PM observado 15.176 Da, PM calculado de longitud completa 15.458 Da). La mayor parte del resto del FAMprimer se extendió hasta tres nucleótidos antes de la unión de ligación fotoquímica (PM observado 10.560 Da, PM calculado 10.559 Da) (Figura 26).

35 Cuando la generación de ADNc, descrita anteriormente, se realizó sin adición de la enzima ligasa, no se observó la secuencia complementaria de longitud completa (menos un único dA).

### Ejemplo 9

Ligación de oligonucleótidos utilizando una conjugación click no legible con polimerasa para modelar un evento de etiquetado

40 Los oligonucleótidos S\_IDTN<sub>3</sub> y S\_ID\_Talkyne se obtuvieron a partir de IDT (IA) (Fig. 27).

La modificación 3'-azida del oligonucleótido S\_IDTN<sub>3</sub> es una modificación patentada de IDT (modificación de azida 3' IDT/3AzidaN/(PM 350)).

Los oligonucleótidos que contienen la modificación /3AzideN/ se purificaron por HPLC.

Los oligonucleótidos que contienen la modificación Hexinilo se desalaron y se usaron sin purificación adicional.

45 La conjugación click de estas etiquetas modelo se logró mediante una cicloadición de azido alquino 3+1 catalizada por Cu (I) de la siguiente manera:

Las soluciones madre se prepararon de la siguiente manera: Cu(OAc)<sub>2</sub>·H<sub>2</sub>O (FW 181,6), 10 mM en DMF; ascorbato de sodio (FW 198,1), 20 mM en H<sub>2</sub>O y Tris-(benciltriazolilmetil)amina (TBTA) (FW 530,6), 10 mM en DMF. A un tubo Eppendorf se le añadieron 10 µL pH 7,0 de regulador de fosfato (solución acuosa 500 mM), 10 µL de S\_IDTalkyne (solución acuosa 1 mM, 10 nmol, 1 equivalente) y 10 µL de S-IDTN<sub>3</sub> (solución acuosa 1 mM, 10 nmol, 1 equivalente molar). A esta solución se le añadieron 5 µL de Cu\_Premezcla (2 equivalentes molares de Cu(OAc)<sub>2</sub>, 4 equivalentes molares de ascorbato de sodio, 1 equivalente molar de TBTA). La reacción se incubó a temperatura ambiente durante la noche. El producto se analizó mediante LCMS (Thermo) (Figura 28) y se encontró que el PM era 26.304,8 Da (calculado: 26.311,2 Da). El conjugado se purificó adicionalmente mediante electroforesis usando 10% de urea PAAG/8M. Este producto de conjugación se llamó Conjugado\_Click\_S.

## 10 Ejemplo 10

Ligación de oligonucleótidos usando un par de conjugaciones click no legibles con polimerasa ortogonal para modelar un par de eventos de etiquetado ortogonal

Se obtuvieron los oligos TKR\_Central y S\_IDTN<sub>3</sub> de IDT (Figura 29). El oligo TKR\_DBCO\_S modificado con 5'-dibenzo-ciclooctino (DBCO) se obtuvo a través de Biosearch Technologies (CA) (Figura 29).

15 En primer lugar, se realizó una conjugación click con TKR\_Central y TKR\_DBCO\_S usando una reacción click sin cobre basada en ciclooctina. Se mezclaron cantidades molares iguales de cada oligonucleótido (10 µL de soluciones acuosas 1 mM) en regulador de fosfato de sodio 0,2 M a pH 7,0. El progreso de la reacción se controló mediante LCMS. La reacción se completó en 30 a 60 minutos, produciendo más del 90% del conjugado puro. MS indicó un PM de 25.336 Da (calculado 25.341,8 Da) (Figura 30).

20 En segundo lugar, se realizó una conjugación click entre el producto de conjugación de TKR\_Central y TKR\_DBCO\_S con S\_IDTN<sub>3</sub> usando el procedimiento click catalizado por Cu (I) descrito en el Ejemplo 9. El conjugado final resultante, 2cl\_S se purificó en un gel de poliacrilamida de urea TBE/8M gel al 10% y se analizó mediante LCMS (Thermo) revelando el peso molecular esperado para 2cl\_S 38578,1 Da (calculado 38.560,4 Da) (Figura 31).

## 25 Ejemplo 11

Amplificación por PCR de ADNc derivado de etiquetas modelo conjugadas por click no legibles por polimerasa utilizando recombinación mediada por cebador no terminal en solución libre y en gotitas acuosas emulsionadas para modelar la lectura de información de secuencia de etiquetas conjugadas químicamente

30 Se preparó Conjugado\_Click\_S (Figura 32), mediante la conjugación de S\_IDTN<sub>3</sub> y S\_IDTalkyne, como se describe en el Ejemplo 9.

Se preparó Conjugado\_Click\_L (Figura 32) por la conjugación de L\_IDTN<sub>3</sub>: TGCGGTCTAACTGTCTAGGCACTTGTT CGTTTGCCAGTGTGAGGAATGAACAGG/3AzidaN/ y L\_IDTalkyne: /5hexinil/ TACCGAATTGCCTTCCTCGTACAGTT CTAAGGCGCTTGACACCACTTCAATCGGGTGTCTATG CTT (IDT, IA) de manera similar al protocolo de conjugación descrito en el Ejemplo 10. LCMS después de la purificación en gel confirmó el PM como 37.582,6 Da (calculado 37.584,4 Da).

Se obtuvieron cebadores de oligodesoxinucleótidos, que incluyen al cebador directo: TGCGGTCTAACTGTCTA, al cebador inverso: AAG CATAGCACCCGATT y ePsplint: GGCAATTCGGTACCTGTTTCATTCC de (IDT, IA).

40 Los conjugados Conjugado\_Click\_L y Conjugado\_Click\_S se diluyeron cada uno hasta 1 nM y se usaron como plantillas para la generación de ADNc mediada por recombinación dependiente de soporte/amplificación por PCR usando Deep Vent (exo-) ADN polimerasa (NEB, MA). Se preparó una reacción de Dina Vent (exo-) de 50 µL como sigue: 5 µL de regulador Thermopool 10x; 2,5 µL de una solución 10 µM de cada cebador directo e inverso, 1 µL de cebador de ePsplint 100 nM; 2,5 µL de mezcla dNTP 10 mM (NEB); 5 µL de una dilución 2 nM de cualquier conjugado o de su mezcla 1:1; 2,5 µL de Deep Vent (exo-) y agua hasta 50 µL. La reacción se cicló 24 veces de la siguiente manera: 30 segundos a 94°C, 30 segundos a 52°C y 60 segundos a 72°C.

45 Cuando los conjugados Conjugado\_Click\_L o Conjugado\_Click\_S se amplifican por separado en solución libre en presencia de ambos cebadores terminales y el cebador de soporte "ePsplint" y luego se evalúan por electroforesis, cada uno da lugar a un único producto con una movilidad que se correlaciona con la longitud de la plantilla (carriles 5 y 7 en la Figura 33). Amplificaciones similares en ausencia del cebador de soporte "ePsplint" no produjeron ningún producto (carriles 4 y 6 en la Figura 33). Esto sugiere que la generación de ADNc mediada por recombinación dependiente de soporte/amplificación por PCR de ambos conjugados por (exo-) polimerasa Deep Vent depende de la presencia del cebador ePsplint (Figura 33). Un posible mecanismo para esta recombinación y amplificación se presenta en la Figura 34.

5 Cuando ambos conjugados Conjugado\_Click\_L y Conjugado\_Click\_S se mezclan y amplifican juntos en solución libre en presencia de ambos cebadores terminales y el cebador de soporte "ePsplint" y luego se evalúan mediante electroforesis, se genera una escalera de productos de movilidad diferente con movilidades que se correlacionan a las longitudes de cada una de las plantillas, así como con las movilidades que se correlacionan con las longitudes intermedias entre cada una de las plantillas (carril 5 en la Figura 35). Un posible mecanismo para esta recombinación y amplificación se presenta en la Figura 36.

10 Cuando ambos conjugados Conjugado\_Click\_L y Conjugado\_Click\_S se mezclan y se amplifican juntos en una emulsión de agua en aceite en presencia de ambos cebadores terminales y el cebador de soporte "ePsplint" a baja concentración de conjugados en relación con las gotitas de la emulsión, y luego se evalúan mediante electroforesis, se observan un par de productos con movilidades que se correlacionan con las longitudes de cada una de las plantillas (carril 3 en la Figura 37). Las gotitas de la emulsión actúan como microrreactores individuales, dentro de los cuales se amplifican los conjugados individuales aislados. La fase de aceite/tensioactivo se preparó mezclando 4,5% (vol/vol) de Span 80 (Fluka, cat # 85548), 0,4% (vol/vol) de Tween 80 (Sigma, cat # S-8074) y 0,05% (vol/vol) de Triton X-100 (Sigma, cat # T-9284) en aceite mineral ligero (Sigma, cat # M-3516). La fase acuosa (50 µL) se preparó mezclando 5 µL de regulador Thermopol 10x; 2,5 µL de una solución de cada cebador de 10 µM, 1 µL de cebador ePsplint 100 nM; 2,5 µL de mezcla dNTP 10 mM (NEB); 2,5 µL DVexo- (NEB), 2,5 µL de BSA (NEB) y agua hasta 50 µL. Se añadió una mezcla 1:1 de los conjugados Conjugado\_Click\_L y Conjugado\_Click\_S a la reacción hasta una concentración final de 5 pM. A 50 µL de la fase acuosa, se le añadieron luego 150 µL de la fase no acuosa y se realizó la emulsión mediante agitación tipo vórtice durante 5 minutos. El proceso de recombinación/amplificación se realizó de la siguiente manera: 30 segundos a 94°C, 30 segundos a 52°C y 60 segundos a 72°C durante 32 ciclos.

25 Después de la generación de ADNc mediada por recombinación dependiente de soporte/amplificación por PCR, la emulsión se rompió mediante microcentrifugación a 21.000 g durante 20 minutos, luego se eliminó la fase no acuosa y se extrajeron en cloroformo la fase acuosa y la interfase. La fase acuosa se evaluó mediante electroforesis nativa (minigeles de agarosa al 4%, Invitrogen CA).

Los resultados de la generación de ADNc mediada por recombinación dependiente del soporte de emulsión compartimentada/amplificación por PCR muestran que el grado de formación de los productos de PCR mezclados está significativamente disminuido o inexistente (carril 3, Figura 37).

### Ejemplo 12

30 Amplificación por PCR de ADNc derivado de etiquetas modelo conjugadas por click no legibles con polimerasa usando recombinación mediada por homología repetida en solución libre para modelar la lectura de un solo evento de etiquetado

35 Dos conjugados, TKR\_S y TKR\_L (figura 38), que contienen cada uno un único enlace de conjugación click no legible con polimerasa formados entre un oligonucleótido modificado con 3'-azido y un oligonucleótido modificado con 5'-hexinilo usando cicloadición catalizada con Cu (I) y el protocolo descrito en el Ejemplo 9. TKR\_S y TKR-L contienen cada uno una secuencia GGAATGAACAGG de 12 mer duplicada que flanquea el enlace no legible con polimerasa. TKR\_S se preparó por conjugación de los siguientes dos oligonucleótidos de ADN IDT, TKR\_S\_N<sub>3</sub>: TGCGGTCTAACTGTCTAGTTCACCTTCTCCGGAATGAACAGG/3AzidoN/(también denominado como S\_IDTN<sub>3</sub> en el Ejemplo 9) y TKR\_S\_Hex: /5'-hexinilo/GGAATGAACAGGGCCTGATTGAAGCAATCGGGTGCTATGCTT, mientras que TKR\_L se preparó por conjugación de dos oligonucleótidos de ADN IDT TKR\_L\_N<sub>3</sub>: TGCGGTCTAACTGTCTAGGCACTTGTTCGTTGCCAGTGTGATGGACACCACTTGGAAATGAACAGG/3AzidaN/ y TKR\_L\_Hex: /5'Hexinil/GGAATGAACAGGGGAATGAACAGGTTCTCGTACAGTTCTAAGGCGCTCAATCGGGTGCTA TGCTT. Ambos conjugados se purificaron con PAGE en condiciones de desnaturalización y se analizaron mediante LCMS para revelar pesos moleculares para TKR\_S: 26.447,5 Da (calculado 26.449,2 Da), y para TKR\_L 37.582,6 Da (calculado 37.584,4).

50 Los dos conjugados se diluyeron cada uno hasta 1 nM y se utilizaron para una generación de ADNc mediada por recombinación dependiente de la repetición/amplificación por PCR con (exo-) ADN polimerasa Deep Vent (NEB) en solución libre con el cebador directo TGCGGTCTAACTGTCTA y el cebador inverso AAGCATAGCACCCGATT (IDT).

55 La generación de ADNc mediada por recombinación dependiente de la repetición/reacción de amplificación por PCR se realizó usando (exo-) ADN polimerasa Deep Vent (NEB). 5 µL de regulador Thermopol 10x; 2,5 µL de una solución 10 µM de cada cebador directo e inverso; 2,5 µL de mezcla dNTP 10 mM (NEB); 40 pM de concentración final de cualquier conjugado o su mezcla 1:1; 2,5 µL de (exo-) polimerasa Deep Vent y agua hasta un volumen final de 50 µL. La reacción se cicló térmicamente 22 veces de la siguiente manera: 30 segundos a 94°C, 30 segundos a 52°C y 60 segundos a 72°C.

Los productos se visualizaron en un minigel de agarosa al 4% (Invitrogen) (Figura 39).

5 La generación de ADNc mediada por recombinación dependiente de la repetición de (exo-) polimerasa Deep Vent/reacción de amplificación por PCR produjo una gran cantidad de amplicón derivado de cada conjugado amplificado por separado (carriles 3 y 4, Figura 39). La amplificación de la mezcla de los conjugados bajo las mismas condiciones no dio como resultado la producción de ningún producto de amplificación de longitud intermedia (carril 5, Figura 39). Para confirmar aún más la falta de mezcla (pérdida) de asociaciones de etiquetas (información codificada) en estos productos de amplificación, se clonaron y secuenciaron los productos de amplificación derivados de la mezcla coamplificada de TKR\_S y TKR\_L. Estos datos de secuencia confirmaron la falta completa de mezcla de los pares de etiquetas modelo en los productos de amplificación derivados de los conjugados coamplificados (Figura 40).

### Ejemplo 13

Amplificación por PCR de ADNc derivado de etiquetas conjugadas con doble click ilegibles usando recombinación mediada por homología repetida en solución libre para modelar la lectura de un par de eventos de etiquetado

15 Dos conjugados, 2\_cl\_S y 2\_cl\_L (Figura 41), cada uno con dos enlaces de conjugación click ilegibles, cada uno flanqueado por diferentes secuencias de repetición de 12 mer (GAAGGCGTTATG-click-GAAGGCGTTATG) y (GGAATGAACAGG-click-GGAATGAACAGG), respectivamente, se prepararon en un procedimiento de dos etapas similar al descrito en el Ejemplo 10.

20 En primer lugar, se realizó una conjugación utilizando una reacción click libre de cobre basada en ciclooctina entre TKR\_central: /5'hexinilo/GGAATGAACAGGGTAAGCTGGAGTGAAGGCGTTATG/3azidaN/(IDT) y, o bien TKR\_DBCO\_S: (DBCO) GAAGGCGTTATGTCCGTAATCTTGCAATCGGGTGCTATGCTT o TKR\_DBCO\_L (DBCO) GAAGGCGTTATGTGATATCCGTGGTGTGAGTTCCAATCGGGTGCTATGCTT. Ambos oligonucleótidos que contienen DBCO se obtuvieron de Biosearch (CA), DBCO es la modificación de 5'-dibenzo-ciclooctileno para la cual está disponible el reactivo de fosforamidita a través de Glen Research (VA)). Los productos de estas conjugaciones se denominaron TKR\_S\_N3 y TKR\_L\_N3 respectivamente.

25 En segundo lugar, se realizó una conjugación entre los productos de las primeras conjugaciones, TKR\_S\_N3 y TKR\_L\_N3, usando una reacción click catalizada por Cu (I). Los conjugados resultantes, TKR\_2\_click\_S y TKR\_2\_click\_L (Figura 41) se purificaron usando electroforesis en un gel de poliacrilamida y urea TBE/8M al 10% y se analizaron por LCMS (Thermo) revelando los pesos moleculares esperados para TKR\_2\_click\_S 38.578,1 Da (calculado 38.560,4 Da), y para TKR\_2\_click\_L 49,913.3 Da (calculado 49.916,7 Da) (Figura 41).

30 Los dos conjugados se diluyeron cada uno hasta 1 nM y se usaron para la generación de ADNc mediada por recombinación dependiente de la repetición/amplificación por PCR con (exo-)polimerasa Deep Vent (NEB). Para la amplificación, el cebador directo TGCGGTCTAACTGTCTA y el cebador inverso AAGCATAGCACCCGATT (ADN IDT) se usaron 5 µL de regulador Thermopol 10x; 2,5 µL de una solución de 10 µM de cada cebador directo e inverso; 2,5 µL de mezcla dNTP 10 mM (NEB); 40 pM de concentración final ya sea del conjugado o su mezcla 1:1, 35 2,5 µL de (exo-) polimerasa Deep Vent y agua hasta 50 µL. La reacción se cicló térmicamente 22 veces de la siguiente manera: 30 segundos a 94°C, 30 segundos a 52°C y 60 segundos a 72°C. Los productos de amplificación se visualizaron en minigeles de agarosa al 4% (Invitrogen).

40 La generación de ADNc mediada por recombinación dependiente de la repetición de (exo-)polimerasa Deep Vent/reacción de amplificación por PCR produjo grandes cantidades de amplicón derivado de cada conjugado amplificado por separado (carriles 3 y 4, Figura 42). La amplificación de la mezcla de los conjugados en las mismas condiciones no dio como resultado la producción de ningún producto de amplificación de longitud intermedia (carril 5, Figura 42). Para confirmar aún más la falta de mezcla (pérdida) de asociaciones de etiquetas (información codificada) en estos productos de amplificación, se clonaron y secuenciaron los productos de amplificación derivados de la mezcla coamplificada de TKR\_2\_click\_S y TKR\_2\_click\_L. Estos datos de secuencia confirmaron la ausencia completa de mezcla de los pares de etiquetas modelo en los productos de amplificación derivados de los conjugados coamplificados (Figura 43).

## REIVINDICACIONES

1. Un complejo que comprende:

(i) una entidad química que comprende una o más andamios o uno o más bloques de construcción;

5 (ii) una primera etiqueta de oligonucleótido que codifica la identidad de al menos uno de dichos uno o más andamios o bloques de construcción; y

(iii) una cabeza de dominio que tiene un primer grupo funcional y un segundo grupo funcional,

10 en el que dicho primer grupo funcional se asocia operativamente con dicha entidad química y dicho segundo grupo funcional se asocia operativamente con dicha primera etiqueta a través de un primer enlace, en el que dicho primer enlace es un oligonucleótido de entrecruzamiento que comprende un cianovinilcarbazol que abarca la unión entre dos etiquetas adyacentes y que se asocia operativamente con dichas dos etiquetas adyacentes a través de uno o más grupos correactivos reversibles.

2. Un complejo que comprende:

(i) una entidad química que comprende uno o más andamios o uno o más bloques de construcción;

15 (ii) n número de etiquetas de oligonucleótidos que tienen n-1 enlaces, en los que n es un número entero entre 1 y 10, y en el que cada uno de dichos enlaces está entre dos etiquetas adyacentes y cada etiqueta codifica la identidad de al menos uno de dichos uno o más andamios o bloques de construcción; y

(iii) una cabeza de dominio que tiene un primer grupo funcional operativamente asociado con dicha entidad química y un segundo grupo funcional operativamente asociado con al menos uno de dicho n número de etiquetas a través de un primer enlace,

20 en el que al menos uno de dichos primer enlace y n-1 enlaces es un oligonucleótido de entrecruzamiento que comprende un cianovinilcarbazol que abarca la unión entre dos etiquetas adyacentes y que se asocia operativamente con dichas dos etiquetas adyacentes a través de uno o más grupos correactivos reversibles.

3. El complejo de la reivindicación 1 o 2, en el que una o más etiquetas comprenden un conector 5' en el extremo 5' de dichas una o más etiquetas y un conector 3' en el extremo 3' de dichas una o más etiquetas.

25 4. El complejo de una cualquiera de las reivindicaciones 1 a 3, en el que al menos un grupo fotorreactivo está presente en un conector 5' en o cerca del extremo 5' de dicha etiqueta y/o en un conector 3' en o cerca del terminal 3' de dicha etiqueta, incluso más preferiblemente en el que la secuencia de al menos uno de dicho conector 5' es complementario a la secuencia del conector 3' adyacente o es idéntica o suficientemente similar para permitir la hibridación con el mismo oligonucleótido complementario en cualquier lado de la unión en eventos de hibridación separados.

30

5. El complejo de cualquiera de las reivindicaciones 1 a 4, en el que dicha entidad química está operativamente asociada a dicha cabeza de dominio a través de un espaciador bifuncional, preferiblemente en el que dicha entidad química está unida covalentemente a dicha cabeza de dominio.

35 6. El complejo de cualquiera de las reivindicaciones 1 a 5, en el que dicha cabeza de dominio comprende un oligonucleótido seleccionado del grupo que consiste en un oligonucleótido de doble cadena, un oligonucleótido de cadena sencilla, o un oligonucleótido de horquilla, preferiblemente en el que dicha cabeza de dominio comprende una región de unión al cebador.

40 7. El complejo de cualquiera de las reivindicaciones 1 a 6, que comprende además una o más primera o primeras etiquetas identificadoras de bibliotecas, etiqueta o etiquetas de uso y/o etiqueta o etiquetas de origen, preferiblemente en el que dicho complejo comprende entre 2 a 20 etiquetas, más preferiblemente en el que cada una de dichas etiquetas comprende de 5 a 75 nucleótidos, lo más preferiblemente en el que cada una de dichas etiquetas dentro de un conjunto de etiquetas individuales comprende aproximadamente la misma masa.

45 8. El complejo de cualquiera de las reivindicaciones 1 a 7, en el que dicho complejo comprende ARN, ADN, ADN modificado y/o ARN modificado, preferiblemente en el que dicho ADN modificado o ARN modificado es PNA, LNA, GNA, TNA o una mezcla de los mismos dentro del mismo oligonucleótido.

9. El complejo de cualquiera de las reivindicaciones 1 a 8, que comprende adicionalmente una cola de dominio.

10. Una biblioteca que comprende uno o más complejos de cualquiera de las reivindicaciones 1 a 9, preferiblemente en la que dicha biblioteca comprende una pluralidad de cabezas de dominio, más preferiblemente en la que cada cabeza de dominio de dicha pluralidad de cabezas de dominio comprende una región de secuencia idéntica y una región de codificación diferente incluso más preferiblemente en la que dicha región de secuencia idéntica es una región de unión al cebador, incluso más preferiblemente en la que dicha región de codificación diferente es una primera etiqueta que codifica para el uso de la biblioteca, origen de la biblioteca, identidad de la biblioteca, historia de la biblioteca, un enlace, un espaciador o adición de un primer componente; o una secuencia de oligonucleótidos que facilita las tecnologías de hibridación, amplificación, clonación o secuenciación, incluso más preferiblemente en la que dicha biblioteca comprende entre aproximadamente  $10^2$  a  $10^{20}$  complejos, más preferiblemente en la que cada complejo es diferente.

11. Un método para etiquetar una primera biblioteca que comprende una entidad química codificada, comprendiendo dicho método:

(a) proporcionar una cabeza de dominio que tiene un primer grupo funcional y un segundo grupo funcional;

(b) unir dicho primer grupo funcional de dicha cabeza de dominio a un primer componente de dicha entidad química, en el que dicha cabeza de dominio está conectada directamente a dicho primer componente o dicha cabeza de dominio está conectado indirectamente a dicho primer componente mediante un espaciador bifuncional; y

(c) unir dicho segundo grupo funcional de dicha cabeza de dominio a una primera etiqueta de oligonucleótido a través de un primer enlace para formar un complejo;

en el que dichas etapas (b) y (c) pueden realizarse en cualquier orden y en el que dicha primera etiqueta codifica la reacción de unión de dicha etapa (b), proporcionando de ese modo una biblioteca etiquetada, preferiblemente en el que dicho método comprende adicionalmente

(d) unión de un número  $n_c$  de componentes adicionales de dicha entidad química, en la que  $n_c$  es un número entero entre 1 y 10; y

(e) unión de un número  $n_l$  de etiquetas de oligonucleótidos adicionales que tienen  $n_l$  enlaces para formar un complejo, en la que  $n_l$  es un entero entre 1 y 10 y en la que cada uno de dichos enlaces está entre dos etiquetas adyacentes y cada etiqueta codifica la identidad de al menos uno de dichos componentes;

en el que al menos uno de dicho primer enlace y dichos  $n-1$  enlaces y son un oligonucleótido de entrecruzamiento que comprende un cianovinilcarbazol que abarca la unión entre dos etiquetas adyacentes y que se asocia operativamente con dichas dos etiquetas adyacentes a través de uno o más grupos correactivos reversibles; y

en el que dichas etapas (d) y (e) pueden realizarse en cualquier orden y en el que cada etiqueta adicional codifica la reacción de unión de cada componente adicional de dicha etapa (d).

12. Un método para detectar una pluralidad de entidades químicas, comprendiendo dicho método:

(a) poner en contacto un objetivo con un complejo de cualquiera de las reivindicaciones 1 a 9 y/o una biblioteca de la reivindicación 10; y

(b) seleccionar uno o más complejos que tienen una característica predeterminada para dicho objetivo, en comparación con un control, detectando de ese modo dicha entidad química, preferiblemente en el que dicha característica predeterminada comprende una mayor unión para dicho objetivo, en comparación con un control, más preferiblemente en el que antes de la etapa (a), se realizan una o más operaciones, seleccionándose dichas operaciones de la lista que consiste en: hibridar uno o más cebadores de relevo con dicho complejo, en la que dichos uno o más cebadores de relevo abarcan dicho primer enlace y/o  $n-1$  enlaces; extender dichos cebadores de relevo usando una polimerasa para producir fragmentos de oligonucleótidos; y ligar dichos fragmentos de oligonucleótidos para producir una plantilla.

13. Un método para determinar la secuencia de nucleótidos de un complejo de cualquiera de las reivindicaciones 1 a 9, comprendiendo dicho método:

(a) hibridación de uno o más cebadores de relevo con dicho complejo, en el que dichos uno o más cebadores de relevo abarcan dicho primer enlace y/o  $n-1$  enlaces;

(b) extender dichos cebadores de relevo usando una polimerasa para producir fragmentos de oligonucleótidos;

(c) ligar dichos fragmentos de oligonucleótidos para producir una plantilla;

(d) amplificar opcionalmente dicha plantilla mediante reacción en cadena de la polimerasa para producir una mezcla amplificada; y

(e) secuenciar dicha mezcla opcionalmente amplificada para determinar la secuencia de dicho complejo.

5 14. Un método para determinar la secuencia de nucleótidos de un complejo de cualquiera de las reivindicaciones 1 a 9, comprendiendo dicho método:

(a) proporcionar dicho complejo que comprende el enlace químico, en el que dicho enlace químico es un oligonucleótido de entrecruzamiento que abarca la unión entre dos etiquetas adyacentes y que se asocia operativamente con dichas dos etiquetas adyacentes a través de uno o más grupos correactivos reversibles;

10 (b) liberar dicho oligonucleótido de entrecruzamiento para producir una plantilla;

(c) amplificar opcionalmente dicha plantilla mediante reacción en cadena de la polimerasa para producir una mezcla amplificada; y

(d) secuenciar dicha mezcla opcionalmente amplificada para determinar la secuencia de dicho complejo, en el que el método incluye ligar las etiquetas usando un proceso químico o un proceso enzimático antes de la etapa (b).

15

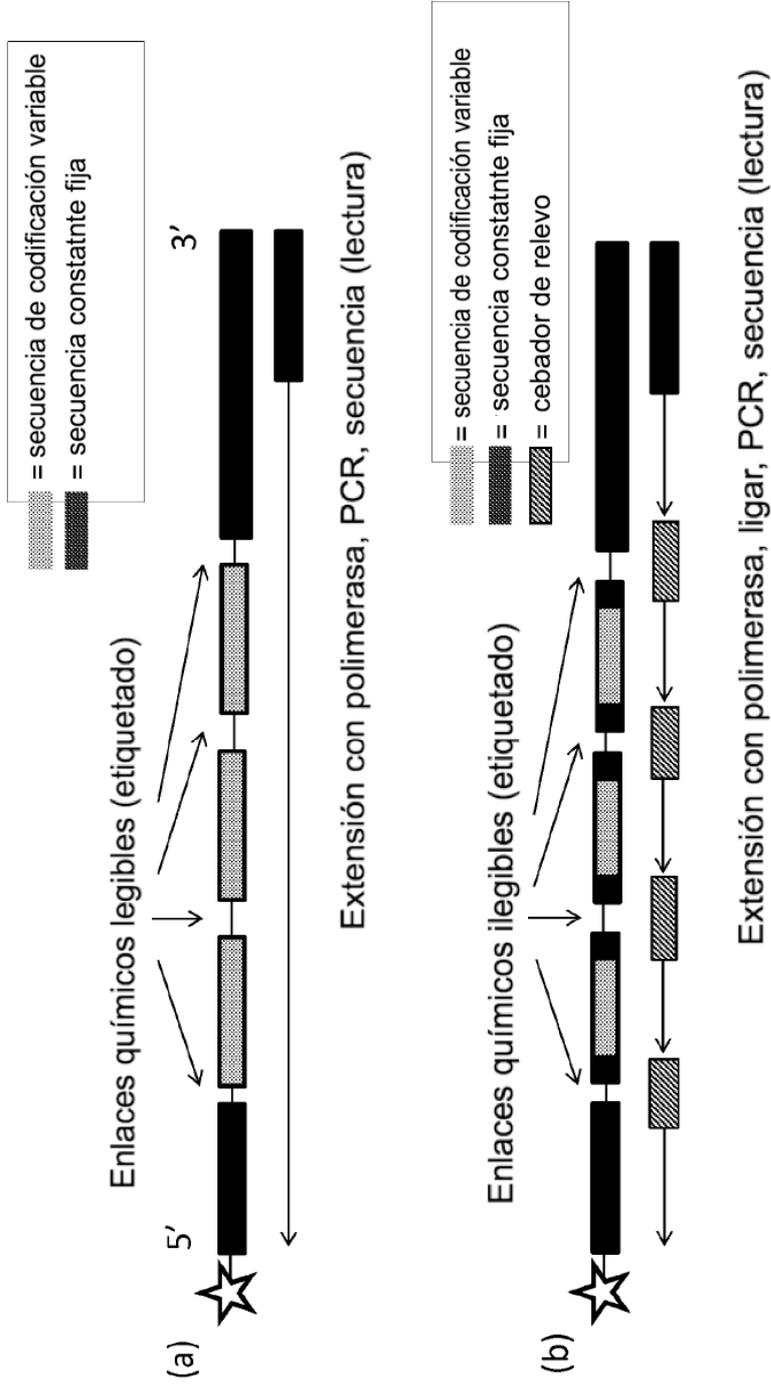


Figura 1



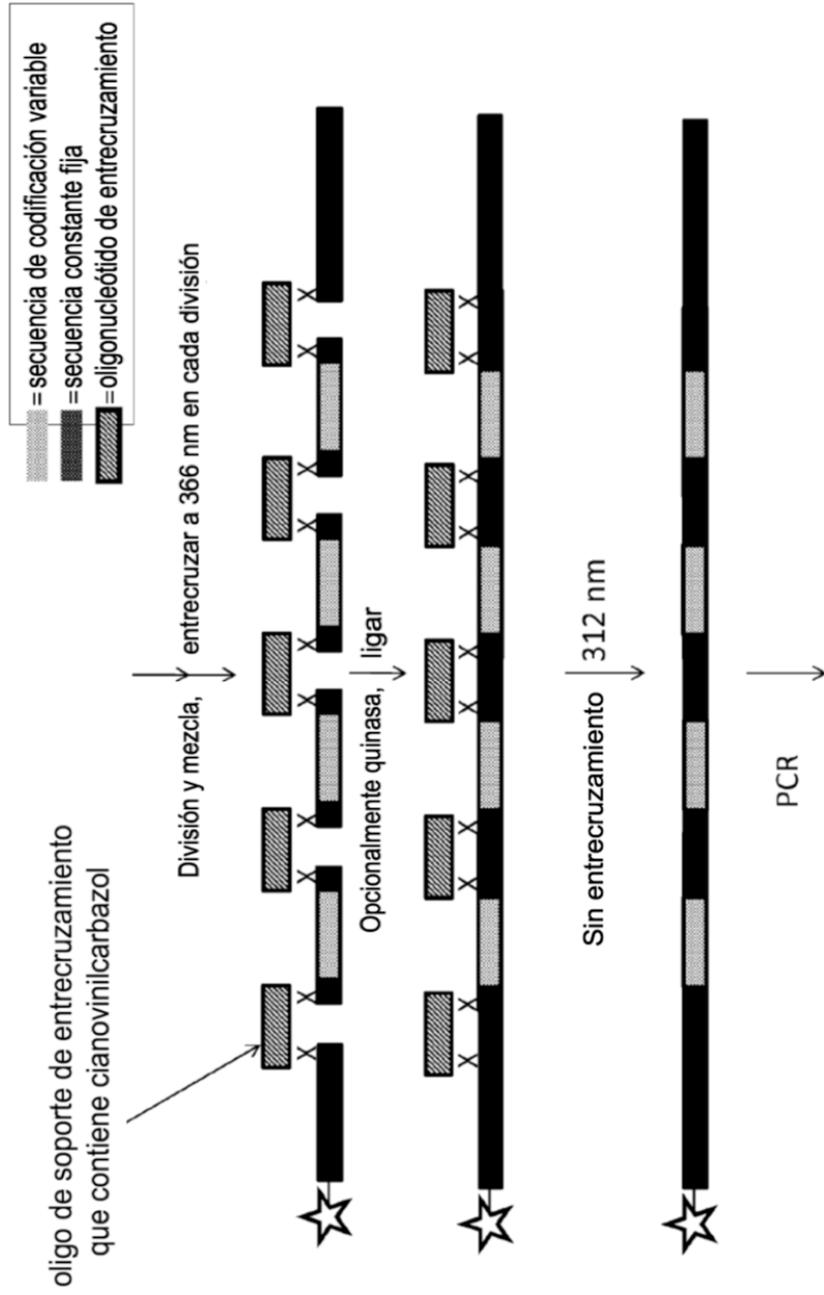


Figura 3



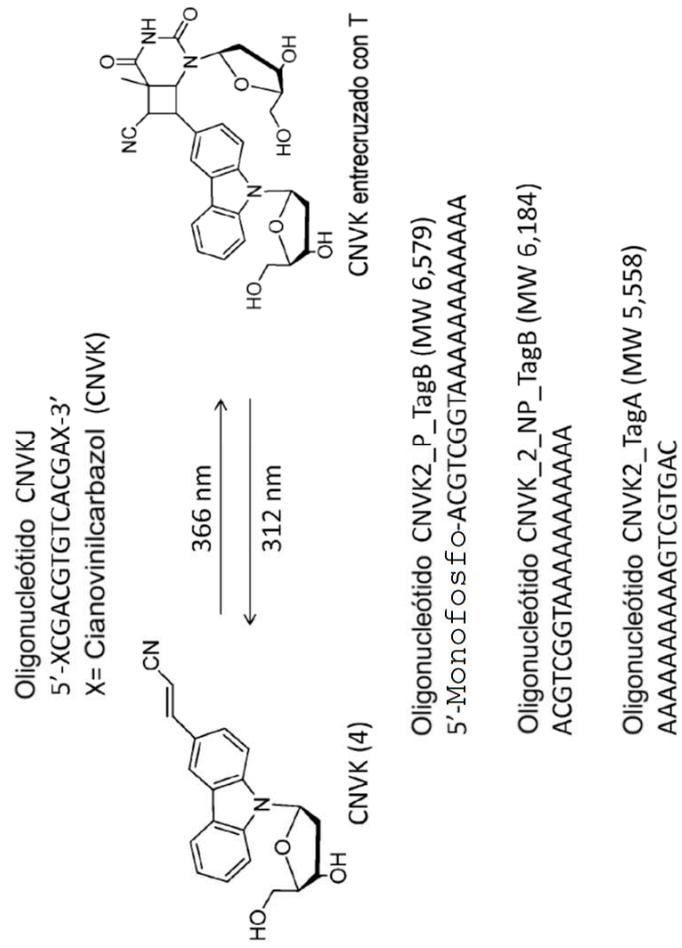
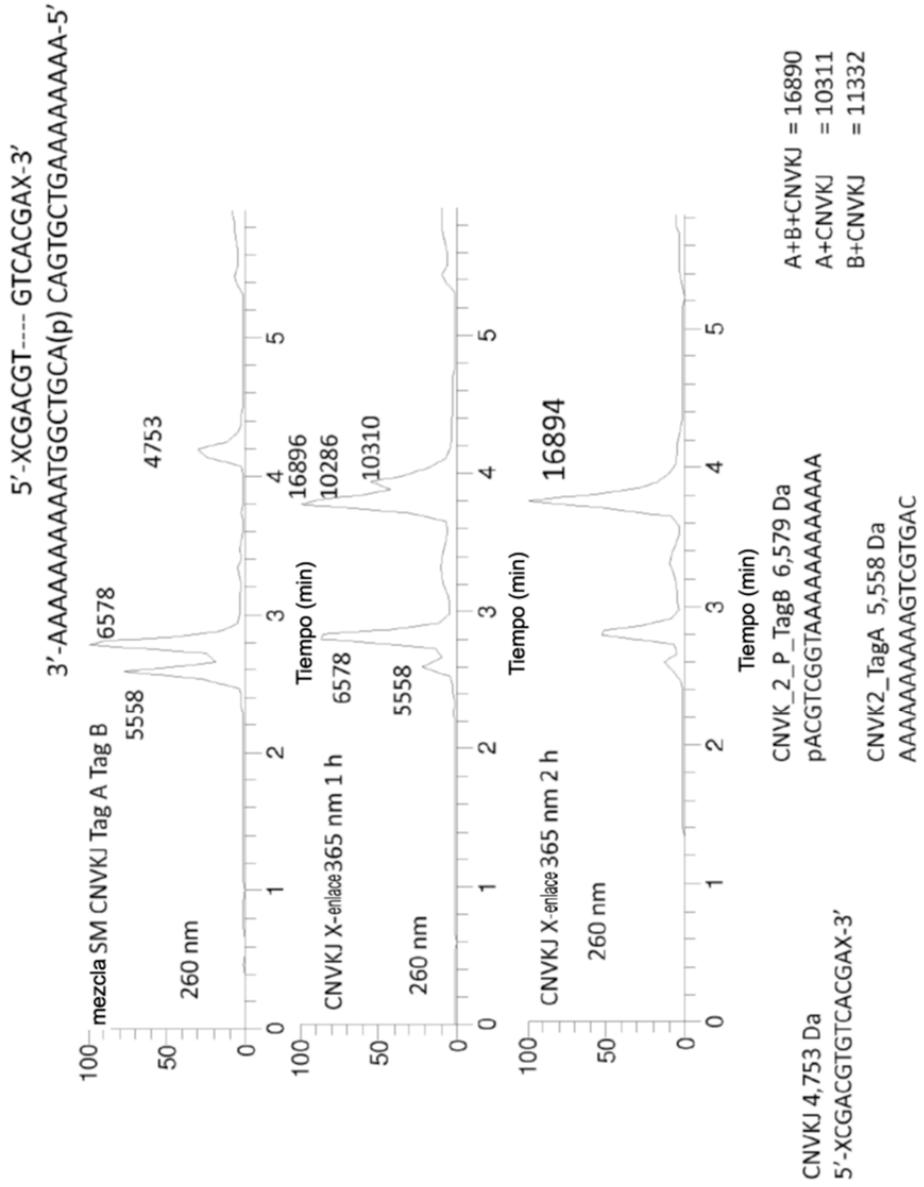
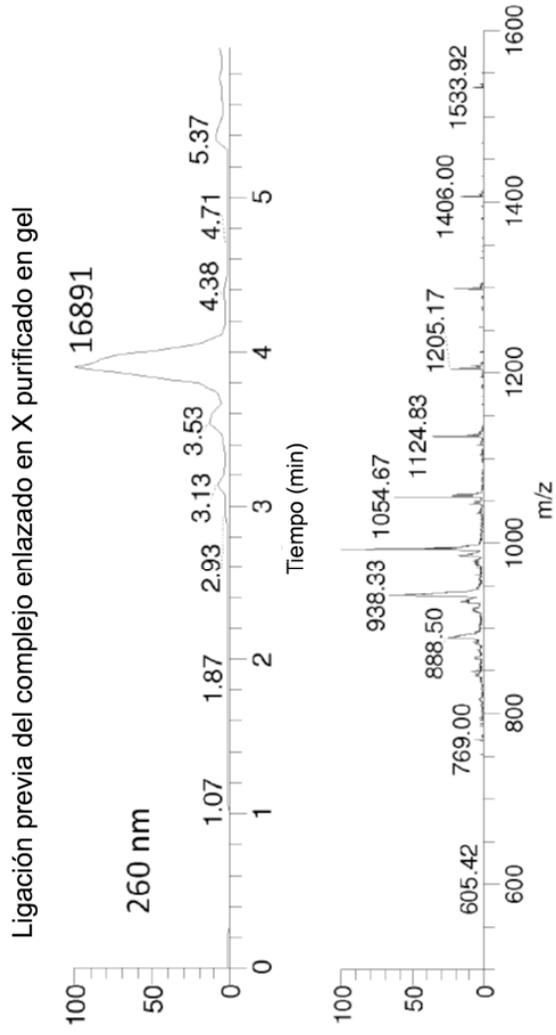


Figura 5



**Figura 6**



Ligación previa del complejo enlazado en X purificado en gel no enlazado en X a 312 nm por 1,5 h a tem. ambiente

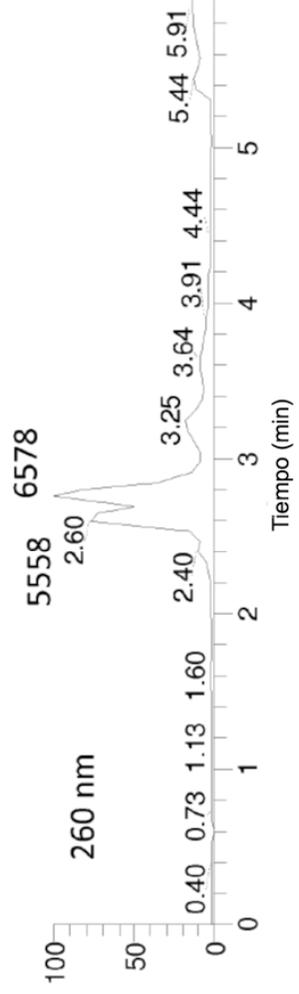


Figura 7

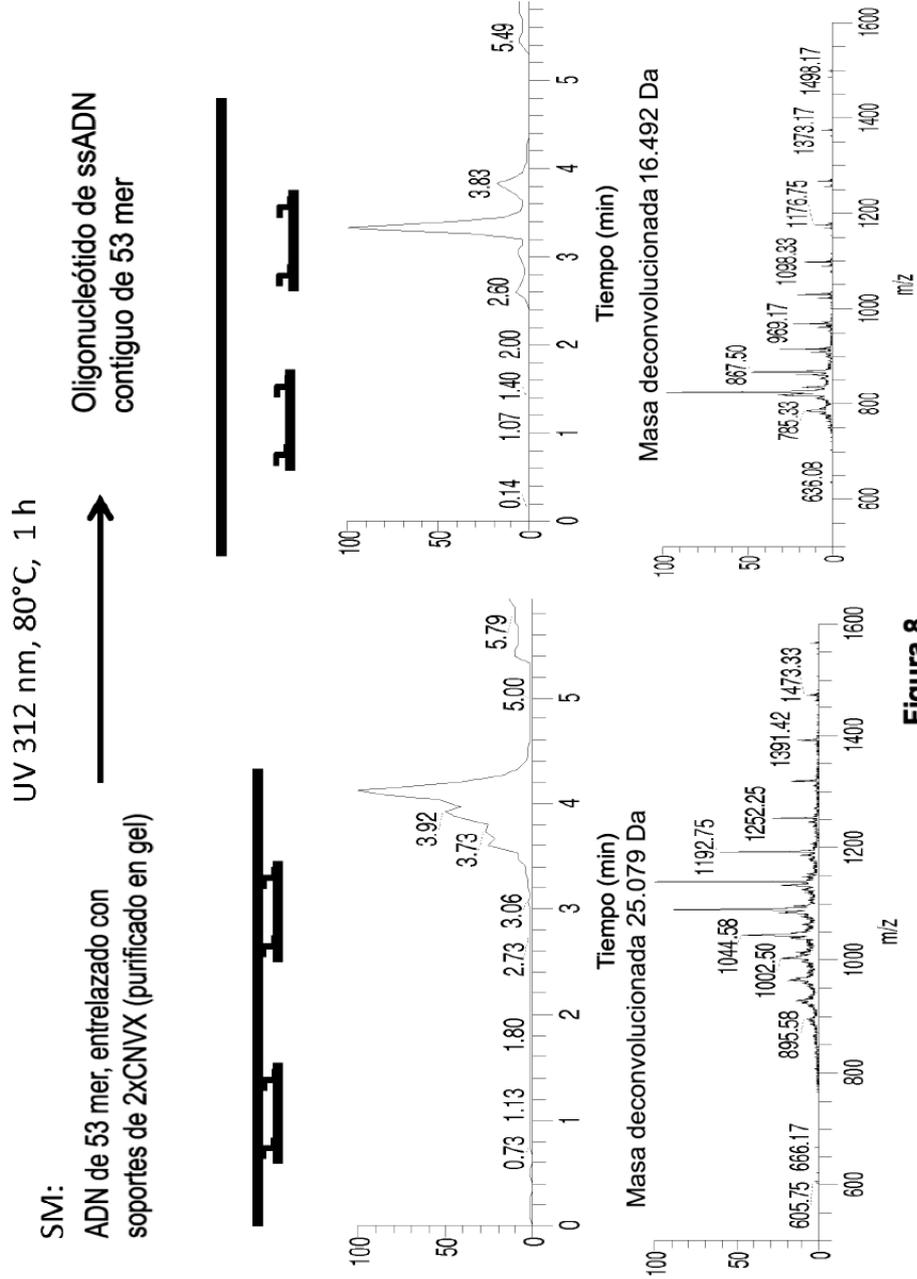


Figura 8

Ligación de ADN T4 del complejo enlazado a X  
seguido por desenlace de X durante 1 hora a 312 nm a 80°C

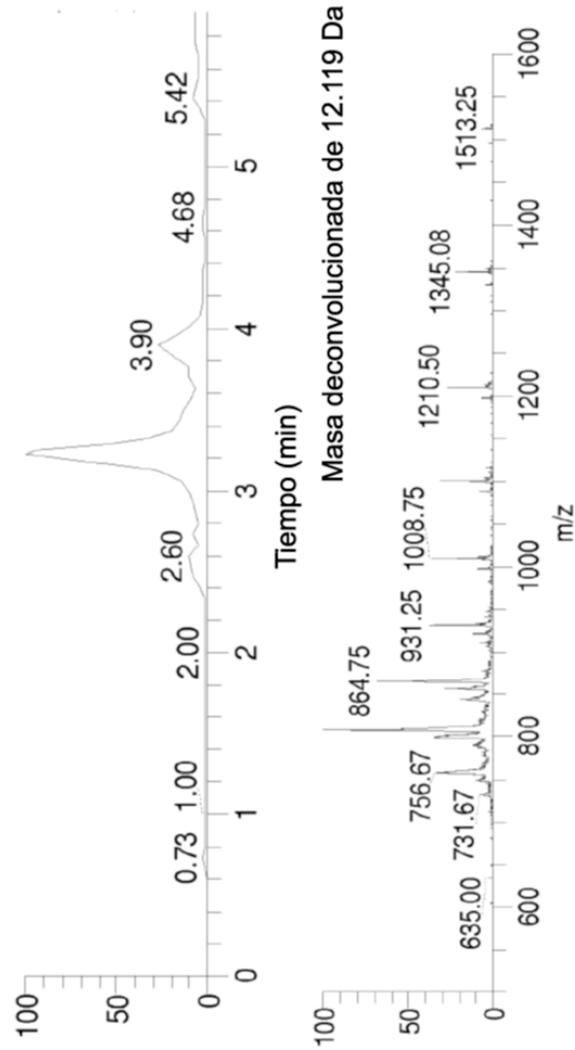


Figura 9

SM:  
Tag A + Tag B + complejo de soporte

Producto:  
Tag A ligado a Tag B

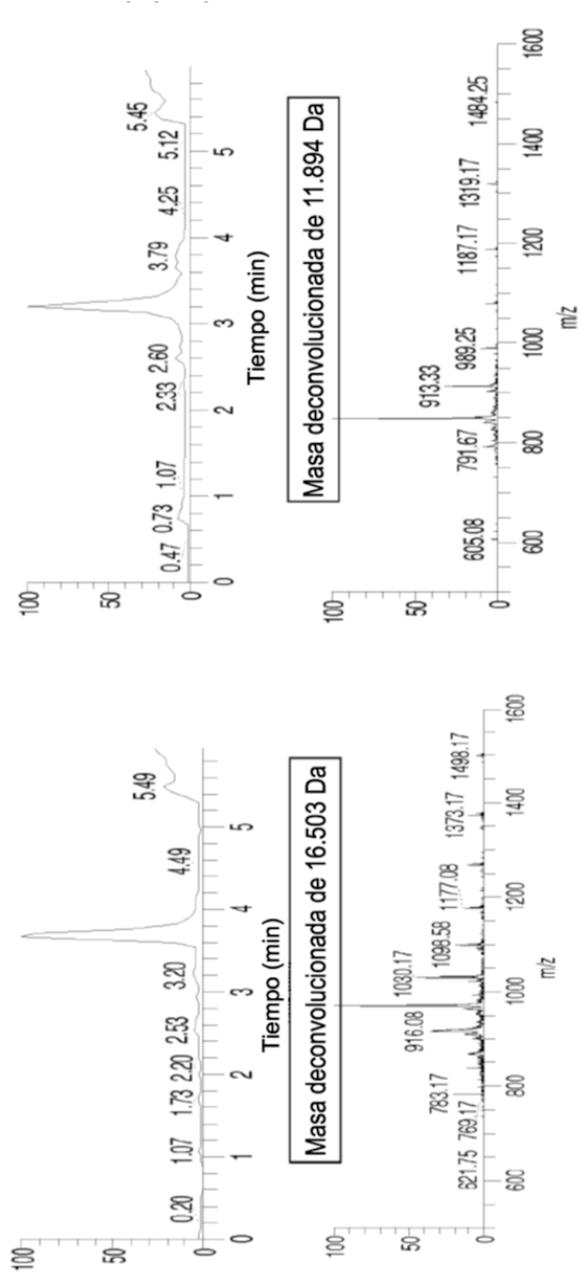


Figura 10

5PSO2\_A9\_TA (PM 11.556,3 Da)  
5'-Psoraleno (enlazado C2)-TAGCGGATGCACAAAAAAGGGGAGCT TGCGTACTG-3'

PSO\_HP\_A9\_TCT (PM 6.087,1 Da)  
5'-AAAAAA AAAGCATCCGCTCT -3'

**Figura 11**

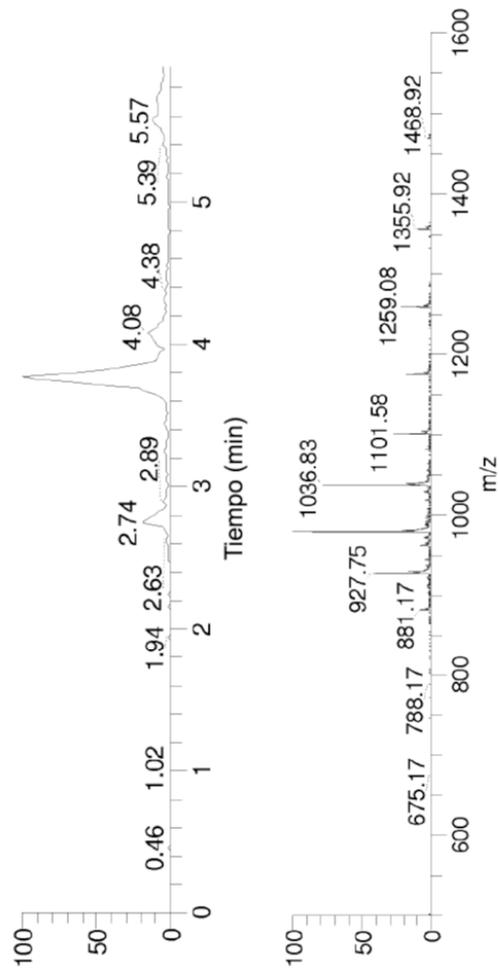
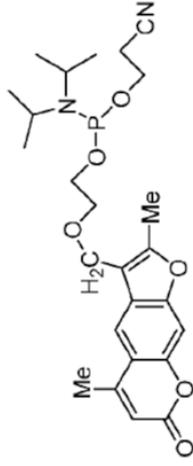


Figura 12



**Psoraleno (C2) Fosforamidita**

Tag1\_PsoCVU (PM 3.662,5 Da)

5'-AAAAACTCAGGT-3'

SplintC\_PsoC2 (PM 4.529,0 Da)

5'-TCCGCTCGACCTGAG-3'

5PsoC2\_A9\_GA (PM 11.580,6 Da)

5'-Psoraleno (C2)-GAGCGGATGCAAAAAAGGGAGCTTGCGTACTG-3'

**Figura 13**

5'-NNNNNNNNNT PNNNNNNNNNNNNNNNNNNNNNNNN-3'  
3'-NNNAGNNNNNNNN-5'

P = 5'-Psoraleno

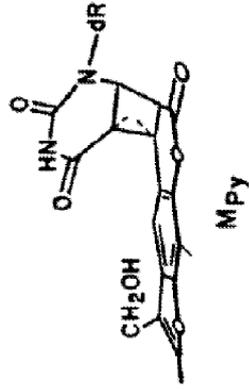


Figura 14

Transcurso del tiempo de la fotoligación de Psoraleno

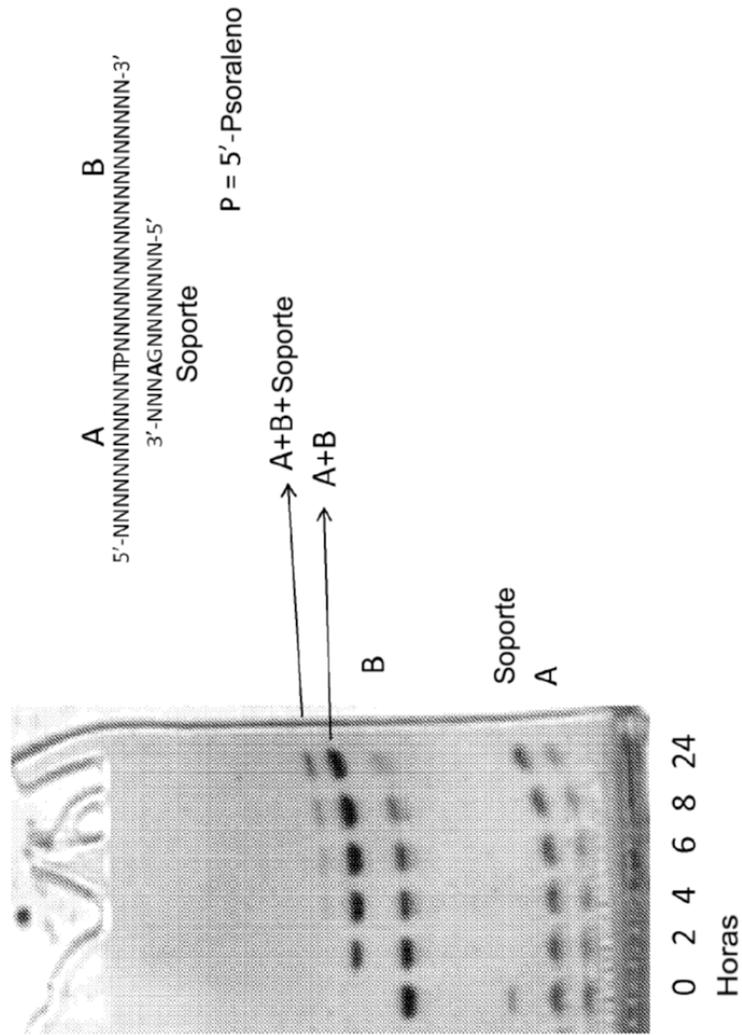


Figura 15

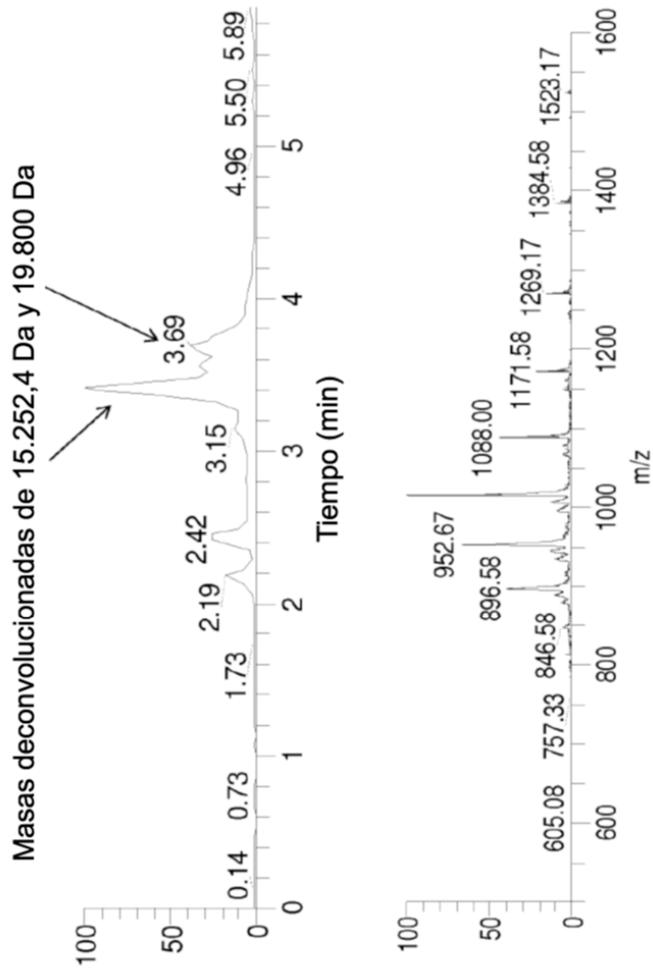


Figura 16

5Bio\_Tag\_PsoCVU (PM 4.232,1 Da) /5BioTEG/ es una modificación  
 5'-/5BioTEG/AAAAACTCAGGT-3' de biotina patentada

5PsoC2\_A9\_GA (PM 11.580,6 Da)

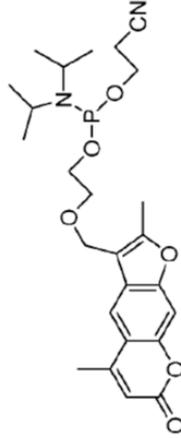
5'-Psoralen (C2)-GAGCGGATGCAAAAAAAGGCGAGCTTGCGTACTG-3'

SplintC\_PsoC2 (PM 4.529,0 Da)

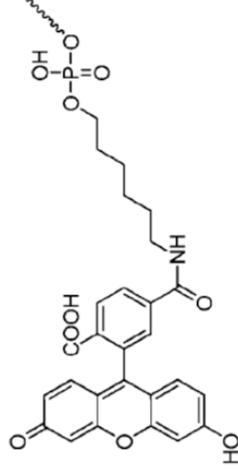
5'-TCCGCTCGACCTGAG-3' FAMprimer (PM 5.099,5 Da)  
 5'-/56-FAM/AGTACGCAAGCTCGC-3'

Phos-SplintC\_PsoC2

5'-Monofosfato-TCCGCTCGACCTGAG-3'



**Psoraleno (C2) Fosforamidita**



**/56-FAM/ Fosforamidita**

**Figura 17**

LCMS de conjugado fotoquímico purificado en gel\_Bio\_Pso

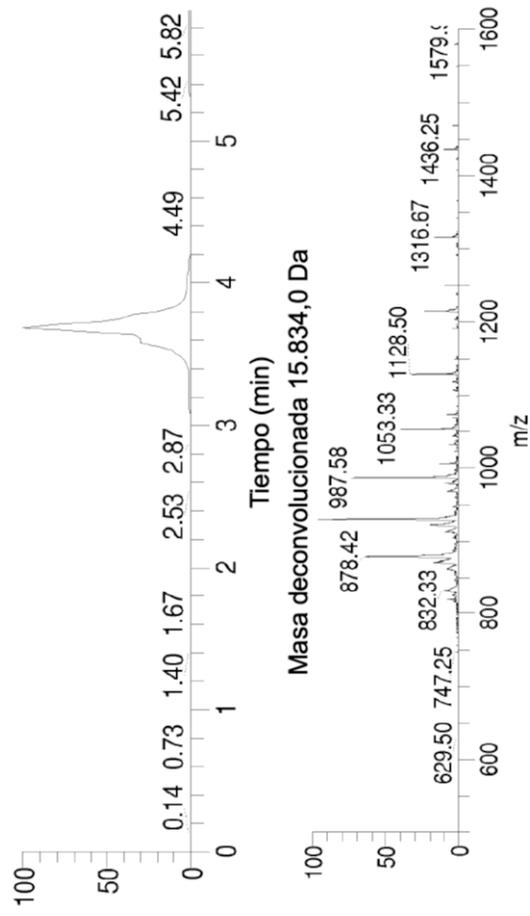


Figura 18

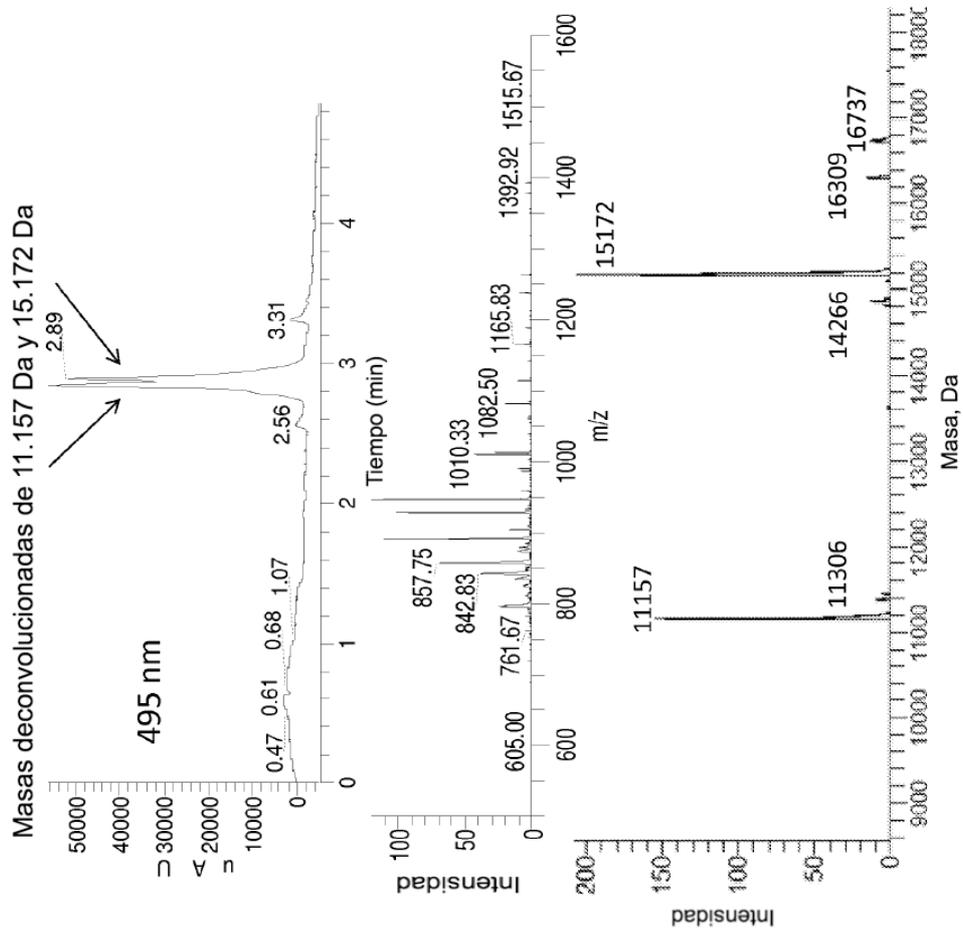


Figura 19

LCMS de control sin ligación de ADNc generado a partir de Bio\_Pso usando extensión de cebador terminal y no terminal

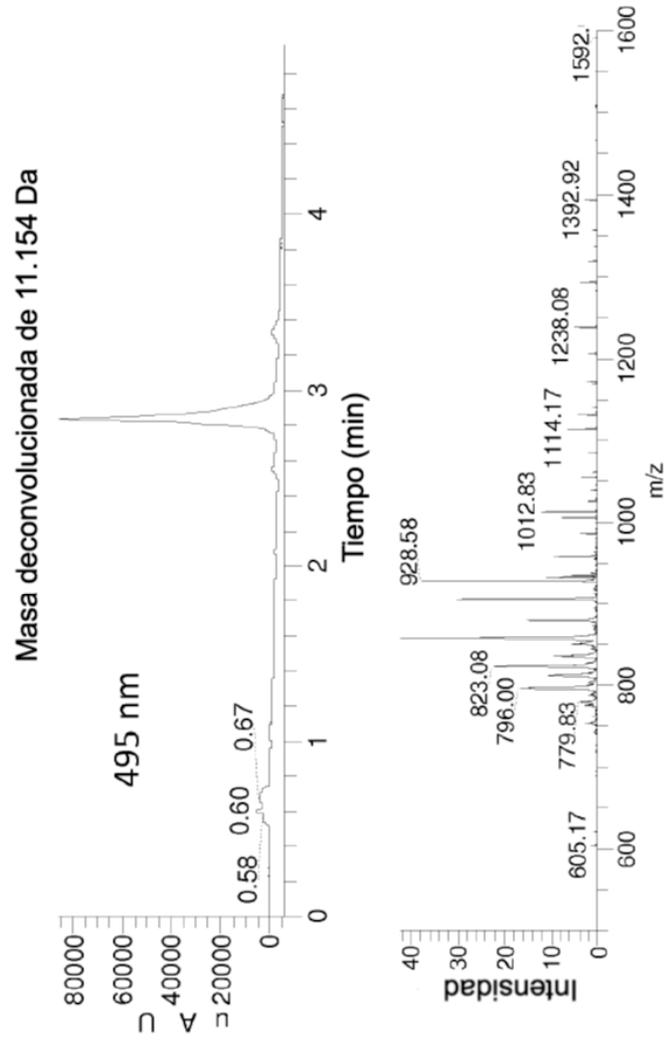
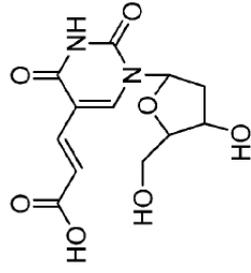


Figura 20



**5-(Carboxi)vinil-2'-deoxiuridina**

SplintC\_CVU  
5'-TCCGCTCAACCTGAG-3'

SplintA\_CVU  
5'-TCCGCTTAAACCTGAG-3'

Tag1\_PsoCVU (PM 3.662,5 Da)  
5'-AAAAACTCAGGT-3'

CVU\_G (PM 11.576,2 Da)  
5'-(5-(Carboxi)vinil-2'-deoxiuridina)GAGCGGATGCAAAAAAAGCGGAGCTTGCGTACTG -3'

CVU\_A (PM 11.560,2 Da)  
5'-(5-(Carboxi)vinil-2'-deoxiuridina)AAGCGGATGCAAAAAAAGCGGAGCTTGCGTACTG -3'

**Figura 21**

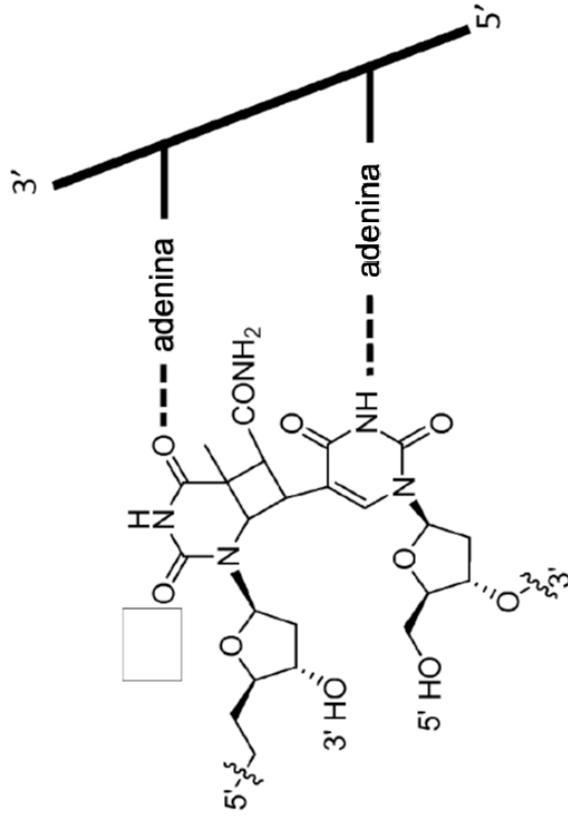


Figura 22

Tiempo transcurrido de la fotoligación de CVU a 365 nm

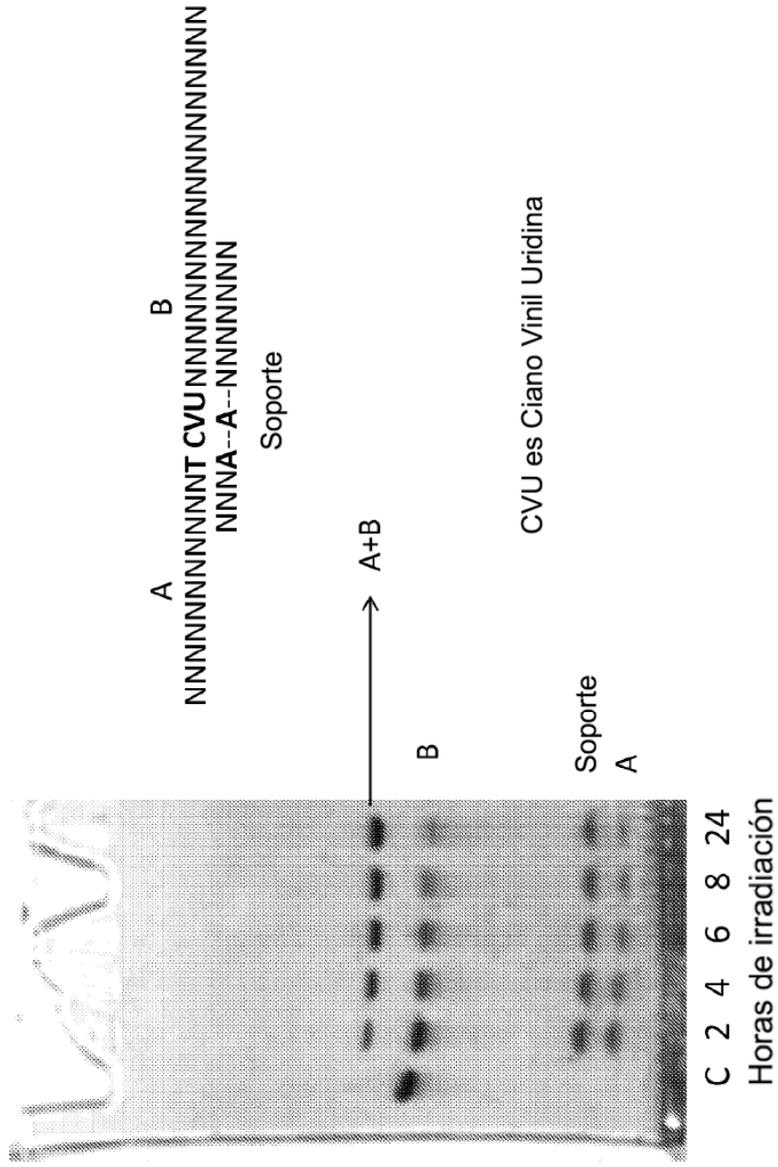


Figura 23

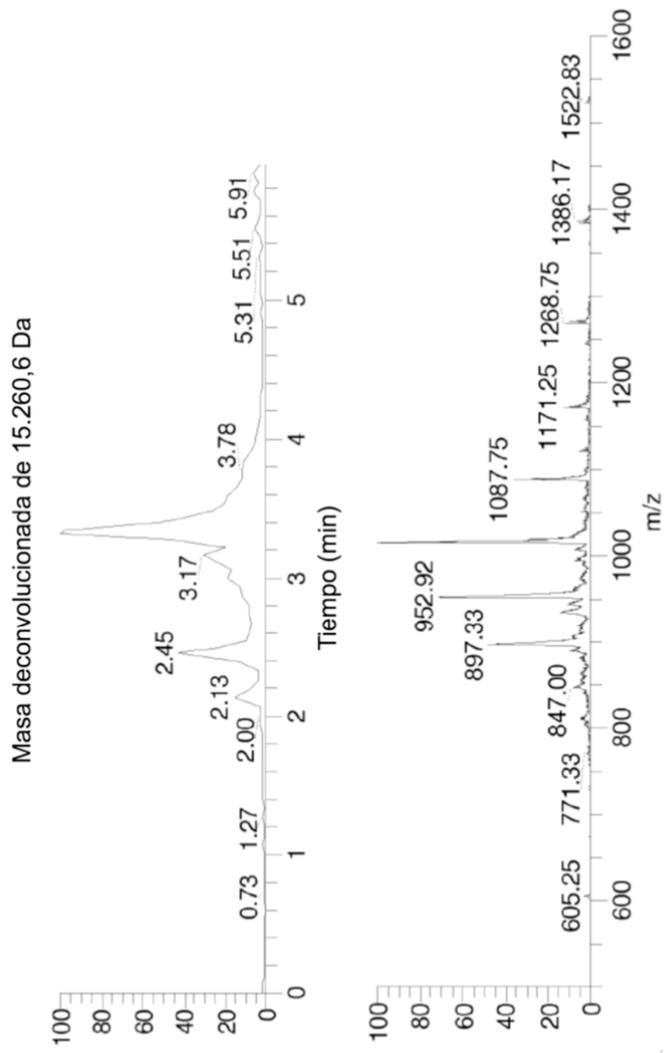


Figura 24

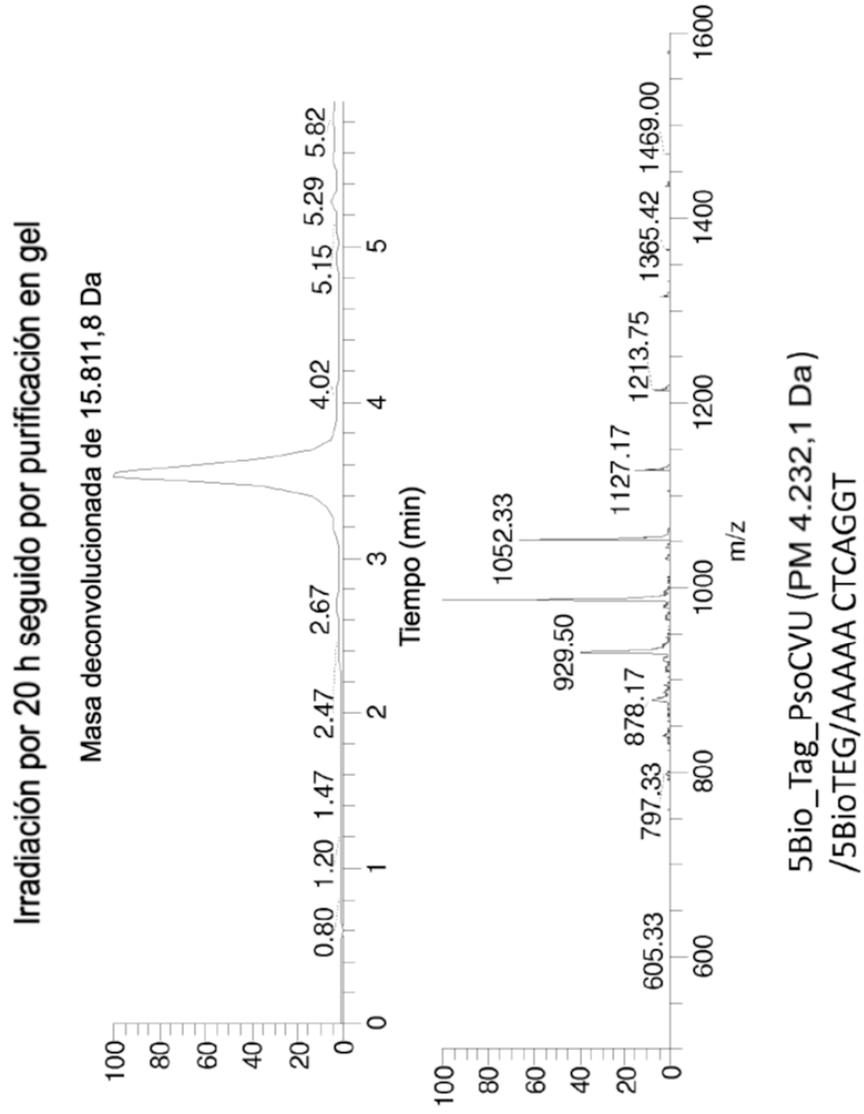
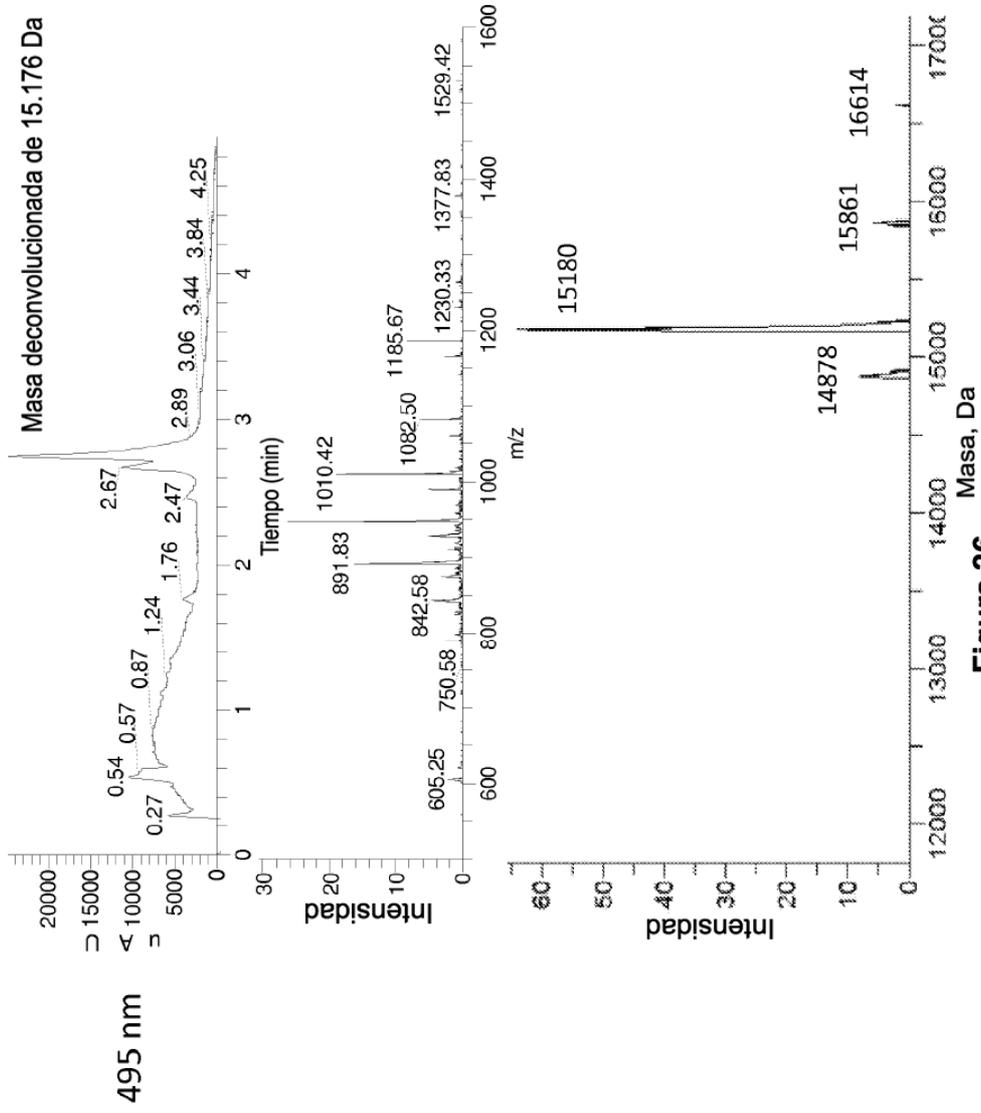


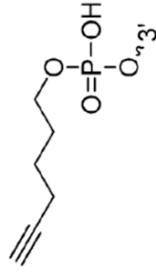
Figura 25



**Figura 26**

S\_IDTN<sub>3</sub>  
TGCGGTCTAACTGTCTAG TTCACCTTCTCCGGAATGAACACAGG/3AzidaN /

S\_IDTalkyne  
/5Hexinilo/TACCGAATTGCCGCCCTGATTGAAGCAATCGGGTGCTATGCTT



Modificación 5Hexinilo

Figura 27

Análisis por LC/MS del conjugado purificado en gel de S\_IDTN<sub>3</sub> y S\_IDTalkyne.

Masa deconvolucionada observada de 26.304,8 Da (calculada 26.311,2 Da)

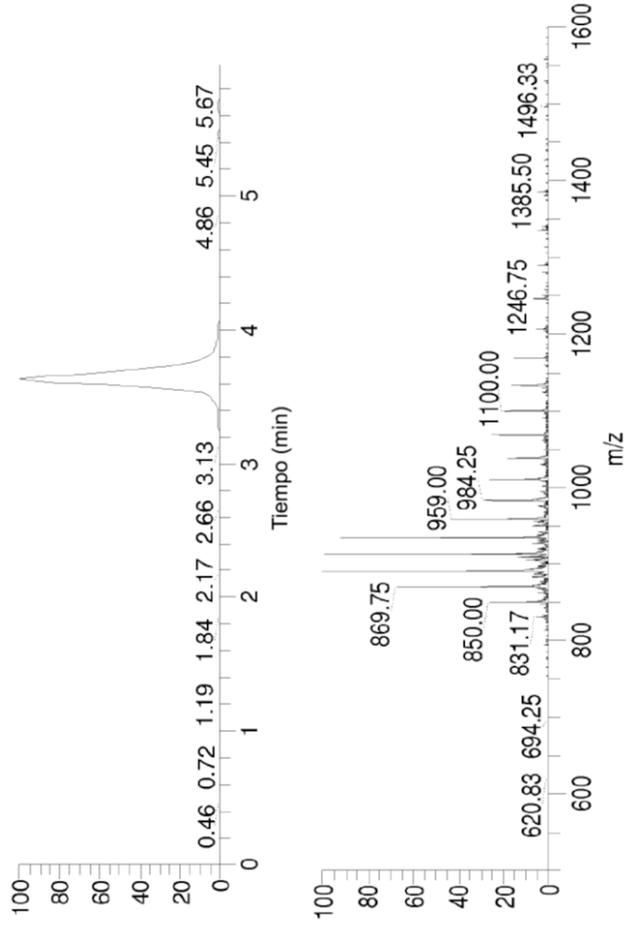


Figura 28



Masa deconvolucionada observada de 25.336 Da (calculada 25.341,8 Da)

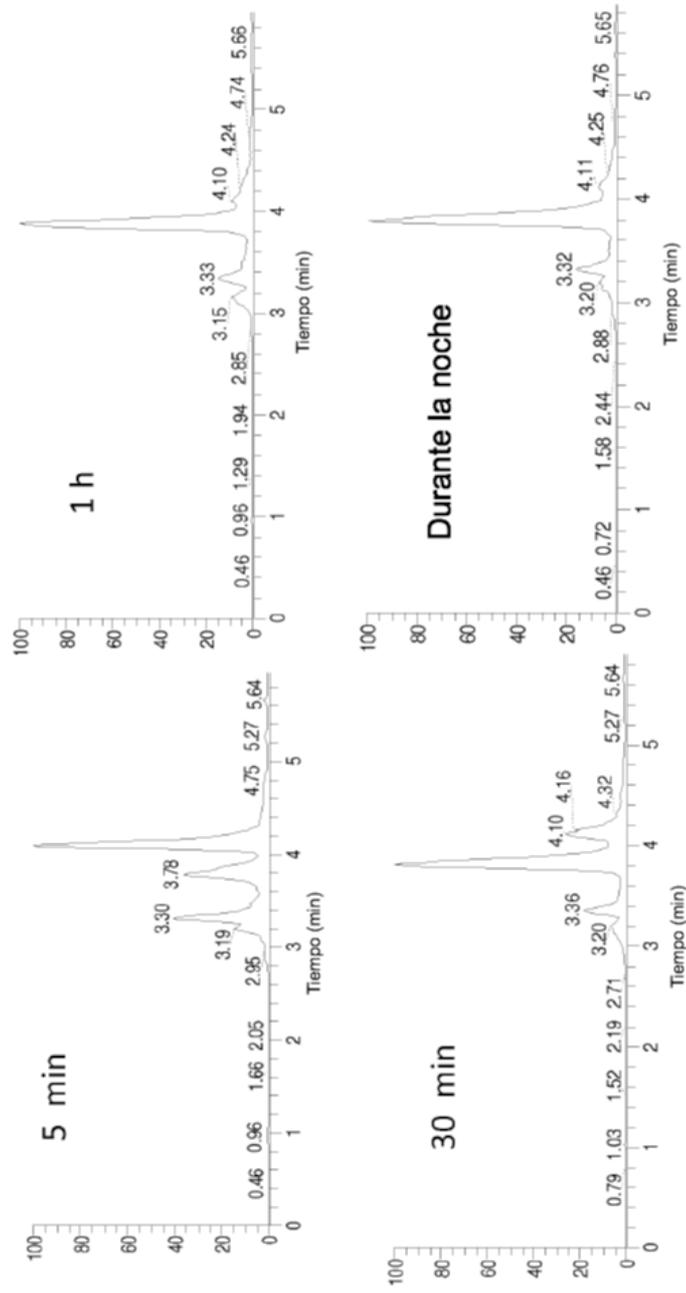


Figura 30

Masa deconvulsionada observada de 38.578,1 Da (calculada 38.560,4 Da)

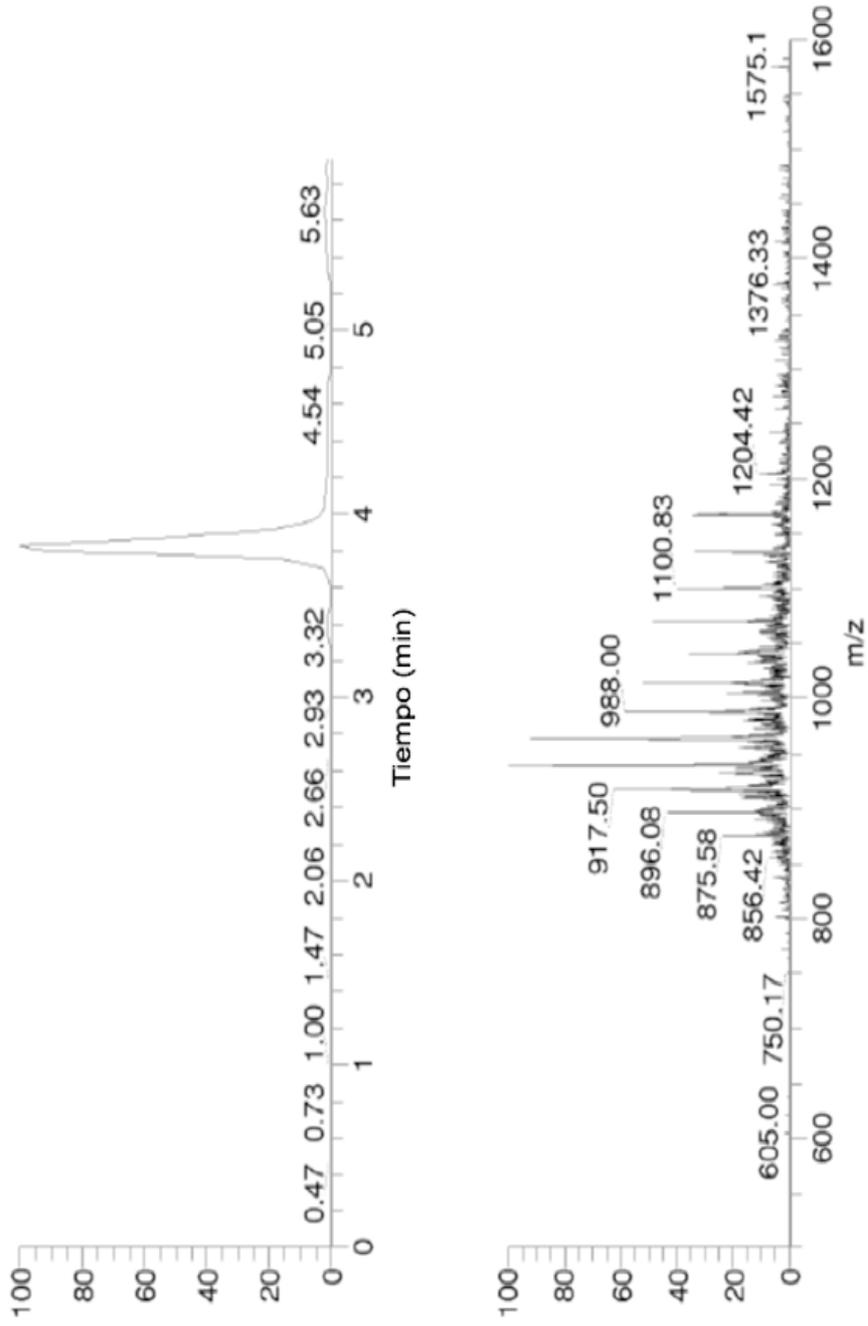
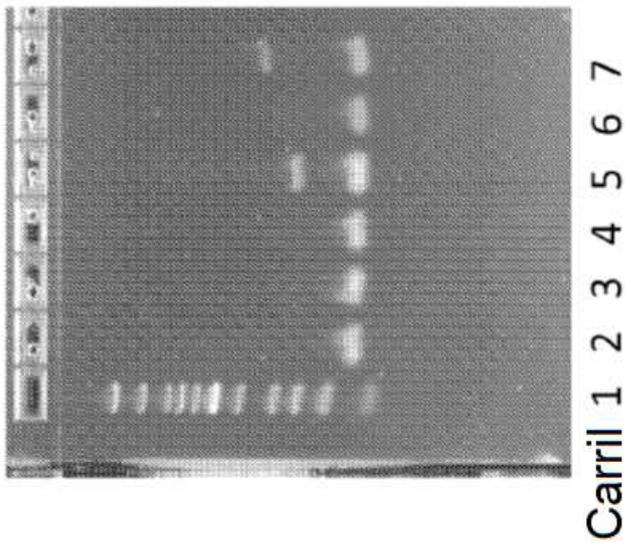


Figura 31

Conjugado\_Click\_S  
TGCGGTCTAACTGCTAGTTCACCTTCTCCGGAATGAACAGG(click)TACCGAATTGCCGCCTGATTGAAG  
-CAATCGGGTGCTATGCTT

Conjugado\_Click\_L  
TGCGGTCTAACTGCTAGGCACCTTGTTTCGTTTGCCAGTGTGAGGAATGAACAGG(click)TACCGAATTGC  
CTTCCTCG-TACAGTTCTAAGGGCGCTTGGACACCCACTTCAATCGGGTGCTATGCTT

Figura 32



**Figura 33**

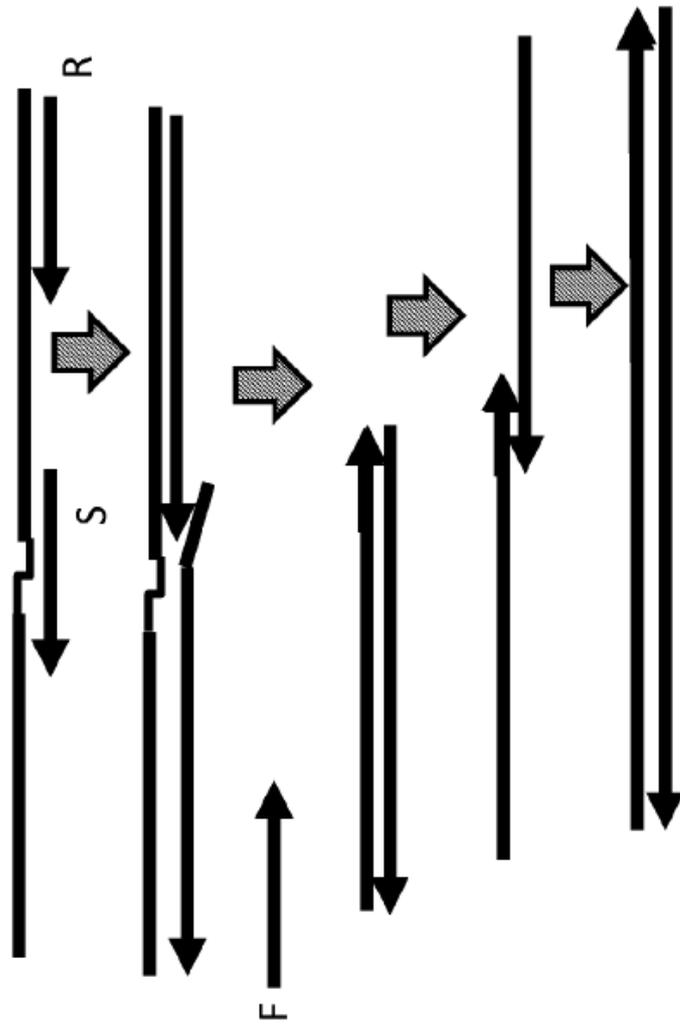
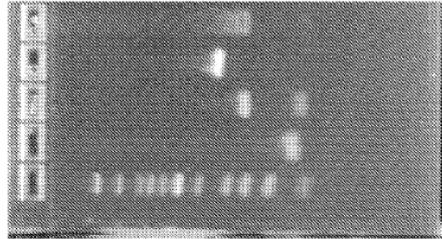


Figura 34



Carril 1 2 3 4 5

**Figura 35**

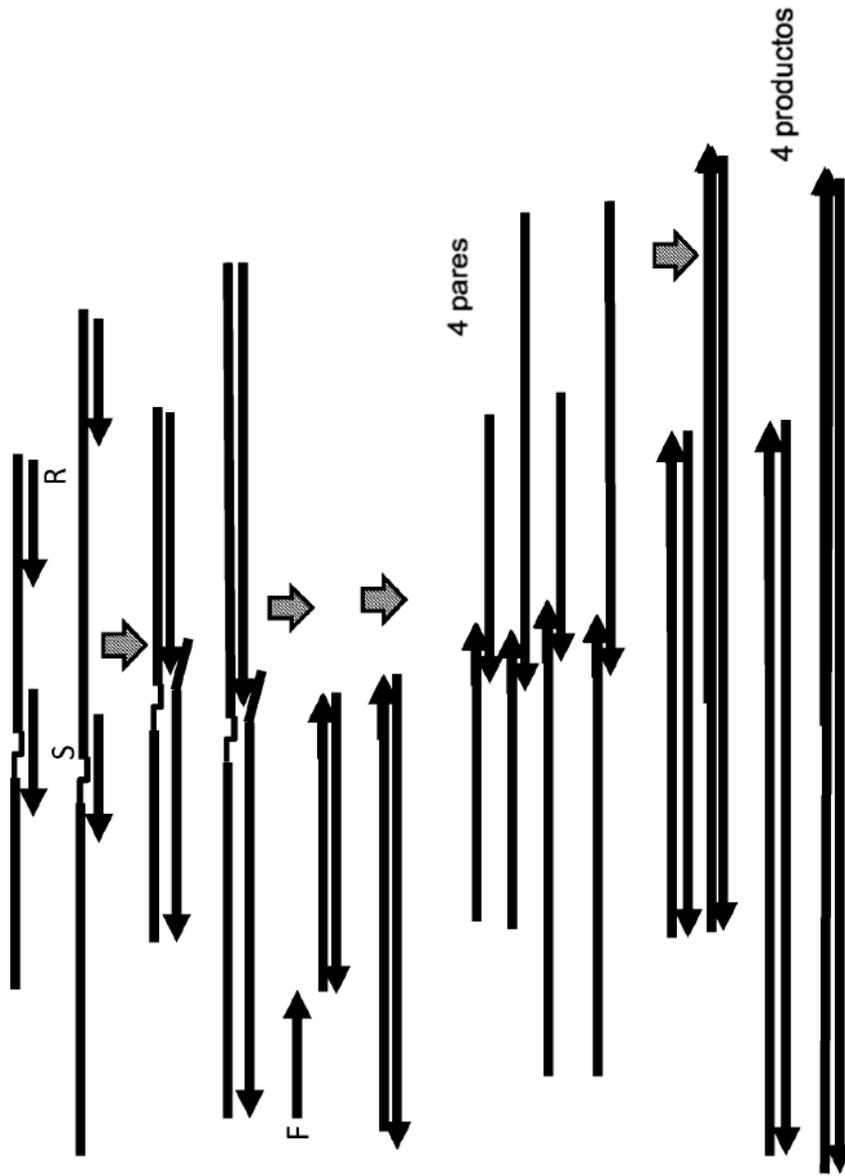
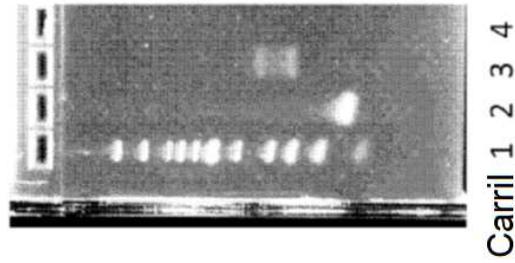


Figura 36

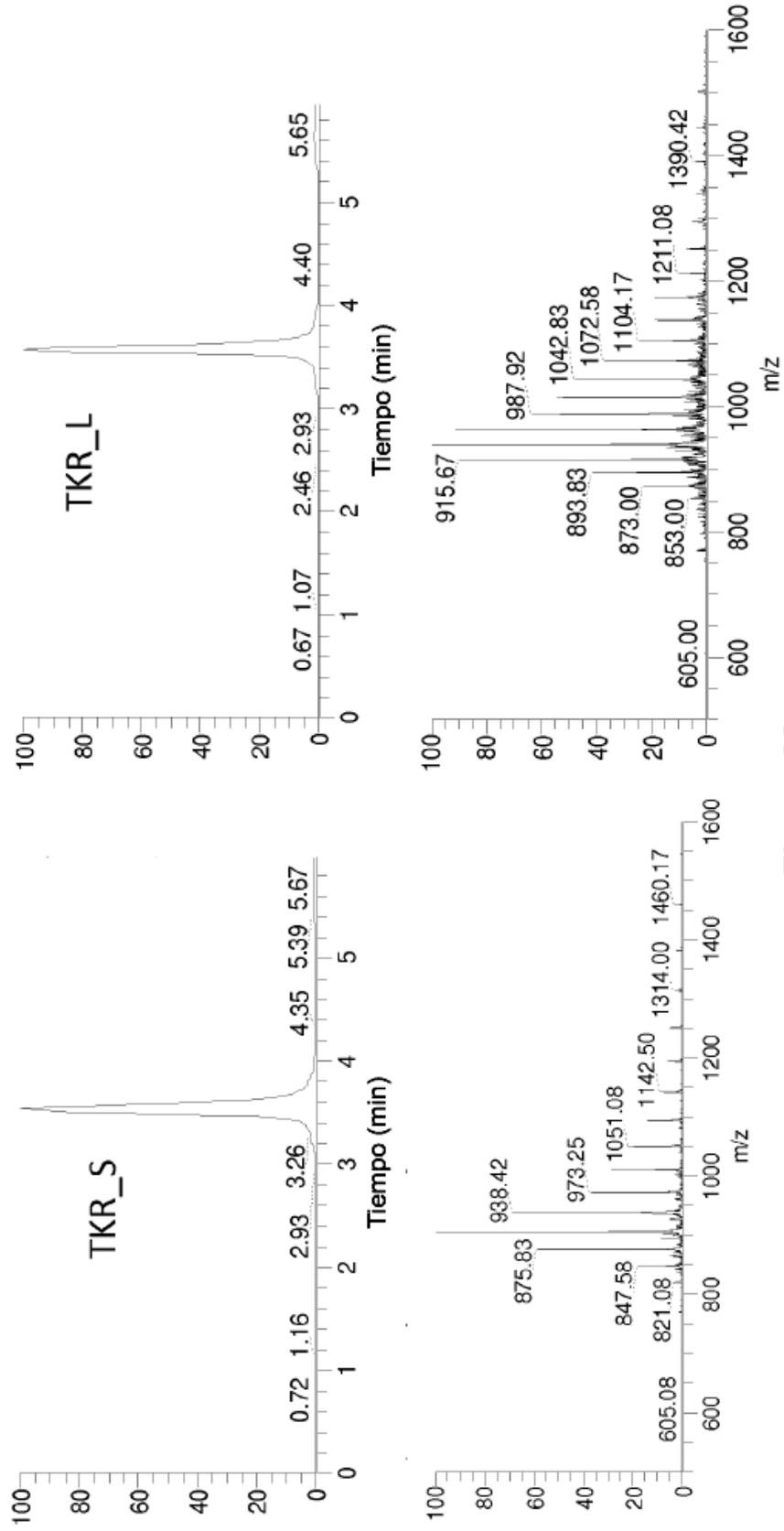


32 ciclos, concentración de partida 5 pM  
1- Marcador  
2 - Solución PCR  
3 - Emulsión PCR

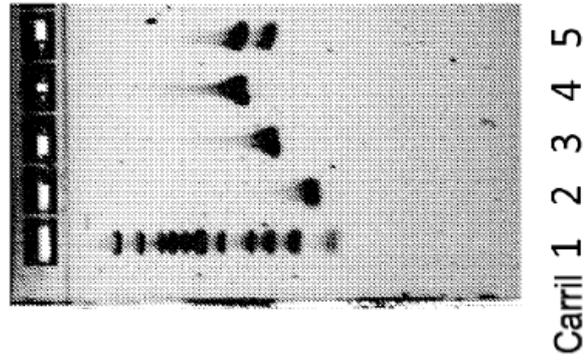
Figura 37

TKR\_S:  
 TCGGGTCTAACTGTCTAGTTCACCTTCTCCGGGAATGAAACAGG(click)GGAATGAAACAGGGCCCTGATTGAAGCAATCGGGTGCTATGCTT

TKR\_L:  
 TGCGGTCTAACTGTCTAGGCACTTGTTGCGTTTGCCAGTGTGATGGACACCCTTGAATGAAACAGG(click)GGAATGAAACAGGGGAATGAACAG  
 GTTCCTCGTACAGTTCTAAGGGCTCAATCGGGTGCTATGCTT



**Figura 38**



**Figura 39**

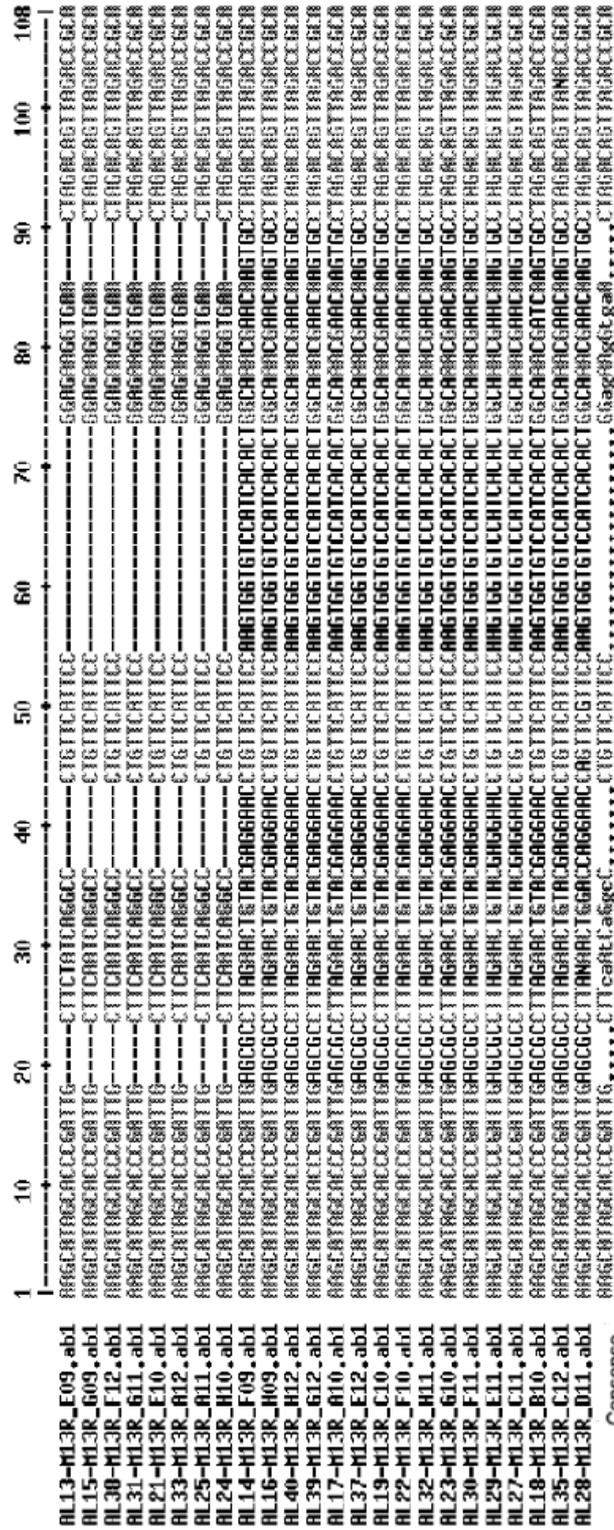
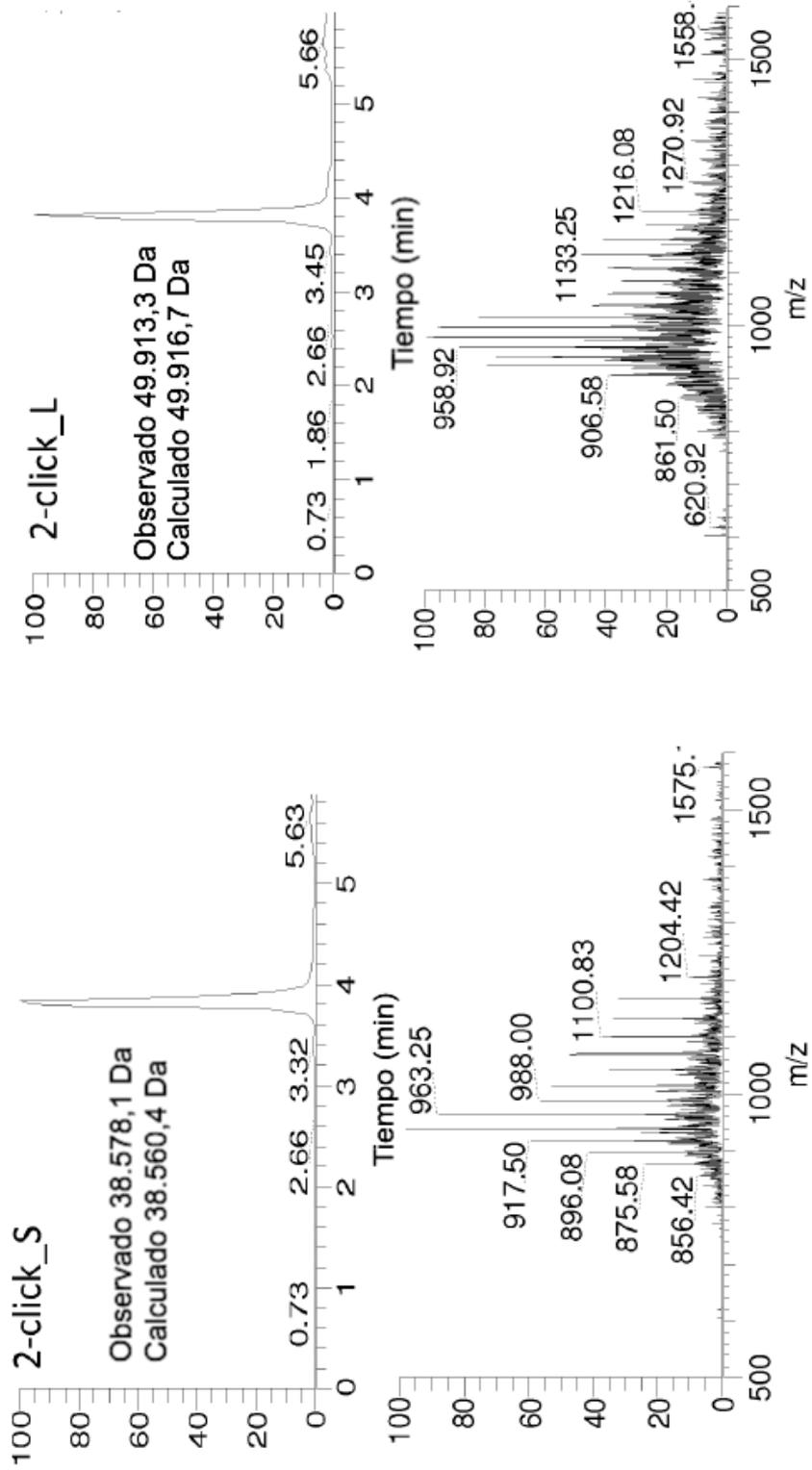


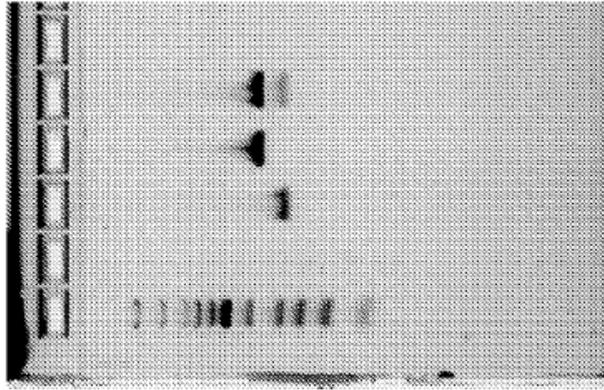
Figura 40

TKR\_2\_click\_S  
 TGCGGTCTAACTGTCTAGTTCACCTTCTCCGGAAATGAACAGG(CLICK)GGAATGAACAGGGTAAGCTGGAGTGAAGGGTTATG(DBCO)  
 GAAGGCGTTATGTCGACTCTTGCAATCGGGTGCTATGCTT

TKR\_2\_click\_L  
 TGCGGTCTAACTGTCTAGGCACCTTGTTCGTTGCCAGTGTGATGGACACCCTTGGAAATGAACAGG(CLICK)GGAATGAACAGGGTAAGC  
 TGGAGTGAAGGCGTTATG(DBCO)GAAGGCGTTATGTGATATCCGTGGTGTCTGAGTCCCAATCGGGTGCTATGCTT



**Figura 41**



Carril 1 2 3 4 5

**Figura 42**

