

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 852 073**

51 Int. Cl.:

**C12N 15/10** (2006.01)

**C40B 50/14** (2006.01)

**C40B 50/16** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.09.2012 E 18158771 (8)**

97 Fecha y número de publicación de la concesión europea: **04.11.2020 EP 3351631**

54 Título: **Métodos para etiquetar bibliotecas codificadas por ADN**

30 Prioridad:

**07.09.2011 US 201161531820 P**

**20.09.2011 US 201161536929 P**

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

**10.09.2021**

73 Titular/es:

**X-CHEM, INC. (100.0%)**

**100 Beaver Street, Suite 101**

**Waltham, Massachusetts 02453, US**

72 Inventor/es:

**KEEFE, ANTHONY D.;**

**WAGNER, RICHARD W.;**

**LITOVCHICK, ALEXANDER;**

**CLARK, MATTHEW;**

**CUOZZO, JOHN W.;**

**CENTRELLA, PAOLO A.;**

**ZHANG, YING y**

**HUPP, CHRISTOPHER D.**

74 Agente/Representante:

**IZQUIERDO BLANCO, María Alicia**

ES 2 852 073 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

Métodos para etiquetar bibliotecas codificadas por ADN

## 5 Antecedentes de la invención

**[0001]** En general, esta invención se refiere a bibliotecas codificadas en el ADN de los compuestos y procedimientos de uso y la creación de tales bibliotecas. La invención también se refiere a composiciones para su uso en tales bibliotecas.

**[0002]** 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 se pueden cribar e interrogar rápidamente. Para aumentar aún más la complejidad, se pueden programar y automatizar varios pasos del proceso de descubrimiento. Estos pasos incluyen el uso de síntesis de múltiples pasos, escindir y agrupar para agregar bloques de construcción a andamios atómicos o poliatómicos y el uso de enzimas y/o ligación química para agregar etiquetas de ADN que codifiquen tanto los pasos sintéticos como los componentes básicos.

**[0003]** A pesar de estos beneficios, numerosos problemas pueden surgir cuando las bibliotecas muy grandes o complejas se deben sintetizar y someter a deconvolución. A medida que aumenta el tamaño de la biblioteca, pueden ser necesarios métodos mejorados para proporcionar altos rendimientos de ligación de etiquetas. Para crear bibliotecas en diversas condiciones de reacción, serían beneficiosas construcciones de nucleótidos ligados estables, 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 ser reconocida por polimerasas dependientes de ADN o ARN, de modo que la población de etiquetas. La demografía se puede determinar mediante polimerización dependiente de la plantilla y determinación de secuencia. Pueden surgir dificultades al crear una biblioteca que tenga todos estos atributos beneficiosos. Por consiguiente, existe la necesidad de métodos mejorados y más robustos de cribado e identificación de pequeños compuestos en bibliotecas codificadas por ADN. El documento WO 2007/062664 A2 describe un método basado en plantilla de ADN para generar una biblioteca codificada por ADN. El documento WO 2006/135786 A2 describe un método para sintetizar una biblioteca codificada por ADN utilizando un enlazador químico trifuncional que une dos cadenas distintas de oligonucleótidos. El documento WO 2005/058479 A2 describe un método para sintetizar una biblioteca codificada por ADN usando un enlazador químico trifuncional que une dos cadenas distintas de oligonucleótidos. El documento WO 2010/094036 A1 describe un método para generar una biblioteca codificada por ADN utilizando un oligonucleótido en horquilla. El documento WO 2004/039825 A2 describe un método para sintetizar una biblioteca codificada por ADN usando oligonucleótidos de codón y anti-codón y síntesis de división y mezcla.

## Resumen de la invención

**[0004]** La presente invención presenta métodos como se define en las reivindicaciones para la creación de bibliotecas, en donde el método incluye una o más condiciones que mejoran la ligación de cadena simple de las etiquetas. También se describen composiciones para su uso en la creación de bibliotecas. Las condiciones ejemplares incluyen el uso de una o más bases sustituidas en 2' dentro de las etiquetas, tales como 2'-O-metilo o 2'-fluoro; el uso de etiquetas de longitud particular; el uso de una o más enzimas; opcionalmente, la inclusión de capacidades de reconocimiento de errores en el diseño de la etiqueta; y/o el uso de uno o más agentes durante la ligación.

**[0005]** En consecuencia, la invención proporciona un método de etiquetado de una primera biblioteca que comprende una entidad química codificada por oligonucleótidos, comprendiendo el método: (i) proporcionar un casco que consiste en un oligonucleótido monocatenario que tiene un primer grupo funcional y un segundo grupo funcional, en donde dicho casco comprende al menos un nucleótido sustituido en 2'; (ii) unir dicho primer grupo funcional de dicho casco a un primer componente de dicha entidad química, en donde dicho casco está conectado directamente a dicho primer componente o dicho casco está conectado indirectamente a dicho primer componente mediante un conector bifuncional (p. ej., un polietileno enlazador de glicol o  $-(CH_2CH_2O)_nCH_2CH_2-$ , donde n es un número entero de 1 a 50); y (iii) unir dicho segundo grupo funcional de dicho casco a una primera etiqueta de bloque de construcción que consiste en un oligonucleótido monocatenario para formar un complejo, en donde dicha primera etiqueta de bloque de construcción comprende un nucleótido sustituido en 2' en ambos extremos 5' y 3', donde los pasos (ii) y (iii) pueden realizarse en cualquier orden y donde dicha primera etiqueta de bloque de construcción codifica la reacción de unión del paso (ii), y donde el paso (iii) comprende una ligasa ARN y/o ADN ligasa para unir dicho casco a dicho primer bloque de construcción, proporcionando así una biblioteca etiquetada.

**[0006]** En algunas realizaciones, la pieza de cabeza incluye un nucleótido 2' sustituido en uno o más del extremo 5', del extremo 3', o la posición interna de la pieza de cabeza. En realizaciones particulares, el casco incluye el nucleótido sustituido en 2' y el segundo grupo funcional en el extremo 5' o en el extremo 3'.

**[0007]** La primera etiqueta de bloque de construcción incluye al menos dos (p. ej., al menos tres, cuatro, cinco, o más) nucleótidos 2' sustituidos. La primera etiqueta de bloque de construcción incluye un nucleótido sustituido en 2' en el extremo 5' y el extremo 3' de la primera etiqueta de bloque de construcción (p. ej., un nucleótido 2'-O-metilo o un

nucleótido 2'-fluoro en ambos de los extremos 5' y 3'). En realizaciones particulares, la primera etiqueta de bloque de construcción incluye un nucleótido sustituido en 2' en la posición interna de la primera etiqueta de bloque de construcción (p. ej., un nucleótido de 2'-O-metilo o un nucleótido 2'-fluoro en ambos extremos 5' y 3'). En algunas realizaciones, la primera etiqueta de bloque de construcción incluye un grupo protector en el extremo 3' o en el extremo 5'.

**[0008]** En cualquiera de las realizaciones descritas en el presente documento, el nucleótido sustituido en 2' es un nucleótido 2'-O-metilo (p. ej., 2'-guanina O-metilo o 2'-O-metilo uracilo) o un 2'-nucleótido fluorado (p. ej., 2'-fluoro guanina o 2'-fluoro uracilo).

**[0009]** En cualquiera de las realizaciones anteriores, la etapa (ii) puede incluir unir, enlazar, o asociarse operativamente la pieza de cabeza directamente al primer componente (p. ej., un andamio o un primer bloque de construcción). En otras realizaciones más, el paso (ii) incluye unir el casco indirectamente al primer componente (p. ej., un andamio o un primer bloque de construcción) a través de un enlazador bifuncional (p. ej., el método incluye atar el casco con el primer grupo funcional del enlazador y unir el primer componente con el segundo grupo funcional del enlazador).

**[0010]** En cualquiera de las realizaciones anteriores, el método puede incluir además la unión (iv) a una segunda etiqueta de bloque de construcción para el extremo 5' o 3' del complejo; y (v) unir un segundo componente (p. ej., un primer bloque de construcción o un segundo bloque de construcción) de la biblioteca química al primer componente, donde las etapas (iv) y (v) se pueden realizar en cualquier orden. En algunas realizaciones, la segunda etiqueta de bloque de construcción codifica la reacción de unión del paso (v). En otras realizaciones, el paso (iv) puede incluir unir la segunda etiqueta de bloque de construcción al extremo 5' del complejo; el complejo incluye un grupo fosfato en el extremo 5'; y la segunda etiqueta de bloque de construcción incluye un grupo hidroxilo en los extremos 3' y 5'. En otras realizaciones, la etapa (iv) puede incluir además purificar el complejo y hacer reaccionar el complejo con una polinucleótido quinasa para formar un grupo fosfato en el extremo 5' antes de unir la segunda etiqueta de bloque de construcción. En otras realizaciones, el paso (iv) puede incluir unir la segunda etiqueta de bloque de construcción al extremo 3' del complejo; el complejo incluye un grupo protector en el extremo 3'; y la segunda etiqueta de bloque de construcción incluye un grupo fosfato en el extremo 5' y un grupo protector en el extremo 3'. En otras realizaciones más, el paso (iv) puede incluir además hacer reaccionar el complejo con un agente hidrolizante para liberar el grupo protector del complejo antes de unir la segunda etiqueta de bloque de construcción al complejo.

**[0011]** En realizaciones adicionales, la segunda etiqueta de bloque de construcción incluye un nucleótido sustituido en 2' (p. ej., un nucleótido 2'-O-metilo o un nucleótido 2'-fluoro) a uno o más del extremo 5', el extremo 3', o la posición interna de la segunda etiqueta de bloque de construcción (p. ej., un nucleótido 2'-O-metilo y/o un nucleótido 2'-fluoro en ambos extremos 5' y 3').

**[0012]** En algunas realizaciones, el paso (iv) puede incluir el uso de una ligasa de ARN (p. ej., ligasa T4 ARN) y/o una ligasa ADN (p. ej., una ligasa ssADN) para enlazar la segunda etiqueta de bloque de construcción al complejo (p. ej., puede incluir el uso tanto de ligasa ARN como de ligasa ADN).

**[0013]** En otras realizaciones, el paso (iii) puede incluir el uso de una ligasa ARN (p. ej., ligasa T4 ARN) y/o un ADN ligasa (p. ej., ligasa ssADN) para enlazar la pieza de cabeza a la primera etiqueta de bloque de construcción (por ejemplo, puede incluir el uso de ligasa ARN y ligasa ADN).

**[0014]** En realizaciones adicionales, el paso (iii) y/o el paso (iv), si está presente, puede incluir el uso de polietilenglicol y/o uno o más cationes multivalentes solubles (p. ej., cloruro de magnesio, cloruro de manganeso (II), o cloruro de hexamina cobalto (III)). En algunas realizaciones, el polietilenglicol está en una cantidad de aproximadamente el 25% (p/v) a aproximadamente el 35% (p/v) (p. ej., de aproximadamente el 25% (p/v) a aproximadamente el 30% (p/v), de aproximadamente 30% (p/v) a aproximadamente 35% (p/v), o aproximadamente 30% (p/v)). En otras realizaciones, el polietilenglicol tiene un peso molecular medio de aproximadamente 3.000 a aproximadamente 5.500 Dalton (p. ej., aproximadamente 4.600 Dalton). En otras realizaciones, el uno o más cationes multivalentes solubles están en una cantidad de aproximadamente 0,05 mM a aproximadamente 10,5 mM (p. ej., de 0,05 mM a 0,5 mM, de 0,05 mM a 0,75 mM, de 0,05 mM a 1,0 mM, de 0,05 mM a 1,5 mM, de 0,05 mM a 2,0 mM, de 0,05 mM a 3,0 mM, de 0,05 mM a 4,0 mM, de 0,05 mM a 5,0 mM, de 0,05 mM a 6,0 mM, de 0,05 mM a 7,0 mM, de 0,05 mM a 8,0 mM, de 0,05 mM a 9,0 mM, de 0,05 mM a 10,0 mM, de 0,1 mM a 0,5 mM, de 0,1 mM a 0,75 mM, de 0,1 mM a 1,0 mM, de 0,1 mM a 1,5 mM, de 0,1 mM a 2,0 mM, de 0,1 mM a 3,0 mM, de 0,1 mM a 4,0 mM, de 0,1 mM a 5,0 mM, de 0,1 mM a 6,0 mM, de 0,1 mM a 7,0 mM, de 0,1 mM a 8,0 mM, de 0,1 mM a 9,0 mM, de 0,1 mM a 10,0 mM, de 0,1 mM a 10,5 mM, de 0,5 mM a 0,75 mM, de 0,5 mM a 1,0 mM, de 0,5 mM a 1,5 mM, de 0,5 mM a 2,0 mM, de 0,5 mM a 3,0 mM, de 0,5 mM a 4,0 mM, de 0,5 mM a 5,0 mM, de 0,5 mM a 6,0 mM, de 0,5 mM a 7,0 mM, de 0,5 mM a 8,0 mM, de 0,5 mM a 9,0 mM, de 0,5 mM a 10,0 mM, de 0,5 mM a 10,5 mM, de 0,75 mM a 1,0 mM, de 0,75 mM a 1,5 mM, de 0,75 mM a 2,0 mM, de 0,75 mM a 3,0 mM, de 0,75 mM a 4,0 mM, de 0,75 mM a 5,0 mM, de 0,75 mM a 6,0 mM, de 0,75 mM a 7,0 mM, de 0,75 mM a 8,0 mM, de 0,75 mM a 9,0 mM, de 0,75 mM a 10,0 mM, de 0,75 mM a 10,5 mM, de 1,0 mM a 1,5 mM, de 1,0 mM a 2,0 mM, de 1,0 mM a 3,0 mM, de 1,0 mM a 4,0 mM, de 1,0 mM a 5,0 mM, de 1,0 mM a 6,0 mM, de 1,0 mM a 7,0 mM, de 1,0 mM a 8,0 mM, de 1,0 mM a 9,0 mM, de 1,0 mM a 10,0 mM, de 1,0 mM a 10,5 mM, de 1,5 mM a 2,0 mM, de 1,5 mM a 3,0 mM, de 1,5 mM a 4,0 mM, de 1,5 mM a 5,0 mM, de 1,5 mM a 6,0 mM, de 1,5 mM a

7,0 mM, de 1,5 mM a 8,0 mM, de 1,5 mM a 9,0 mM, de 1,5 mM a 10,0 mM, de 1,5 mM a 10,5 mM, de 2,0 mM a 3,0 mM, de 2,0 mM a 4,0 mM, de 2,0 mM a 5,0 mM, de 2,0 mM a 6,0 mM, de 2,0 mM a 7,0 mM, de 2,0 mM a 8,0 mM, de 2,0 mM a 9, 0 mM, de 2,0 mM a 10,0 mM y de 2,0 mM a 10,5 mM). En algunas realizaciones, uno o más cationes multivalentes están en una cantidad de aproximadamente 1 mM (p. ej., de 0,5 mM a 1,5 mM). En una realización particular, el catión multivalente está en forma de cloruro de cobalto (III) de hexamina.

**[0015]** En otras realizaciones, el método incluye además separar el complejo de cualquier etiqueta no reaccionada o pieza de cabeza sin reaccionar antes de una cualquiera de las etapas de unión (ii)-(v). En otras realizaciones, el método incluye además purificar el complejo antes de cualquiera de las etapas de unión (ii)-(v). En otras realizaciones, el método incluye además unir uno o más componentes adicionales (p. ej., un andamio o un primer bloque de construcción) y una o más etiquetas de bloque de construcción adicionales, en cualquier orden y después de cualquiera de los pasos de unión (ii)-(v).

**[0016]** La invención también presenta un método como se define en las reivindicaciones, que comprende, además, etiquetar una primera biblioteca que incluye una entidad química codificada por oligonucleótido, incluyendo el método: (i) proporcionar una pieza de cabeza que tiene un primer grupo funcional y un segundo grupo funcional, donde el casco incluye un nucleótido sustituido en 2' en el extremo 5', opcionalmente uno o más nucleótidos en la posición interna del casco, y un grupo protector en la posición 2' y/o la posición 3' en el extremo 3'; (ii) unir el primer grupo funcional del casco a un primer componente de la entidad química, donde el casco está conectado directamente al primer componente o el casco está conectado indirectamente al primer componente mediante un conector bifuncional; y (iii) unir el segundo grupo funcional del casco a una primera etiqueta de bloque de construcción, donde la primera etiqueta de bloque de construcción incluye un nucleótido sustituido en 2' y un grupo hidroxilo en el extremo 5', opcionalmente uno o más nucleótidos en la posición interna de la etiqueta y un nucleótido sustituido en 2' y un grupo hidroxilo en el extremo 3'; donde los pasos (ii) y (iii) se pueden realizar en cualquier orden y donde la primera etiqueta de bloque de construcción codifica la reacción de unión de el paso (ii), proporcionando así una biblioteca etiquetada.

**[0017]** En algunas realizaciones, el nucleótido sustituido en 2' es un nucleótido 2'-O-metilo (p. ej., 2'-O-metilo guanina) o un nucleótido 2'-fluoro (guanina por ejemplo, 2'-fluoro). En otras realizaciones, uno o más nucleótidos en la posición interna del casco son 2'-desoxinucleótidos. En otras realizaciones más, el enlazador bifuncional es un enlazador de polietilenglicol (p. ej.,  $-(CH_2CH_2O)_nCH_2CH_2-$ , donde n es un número entero de 1 a 50).

**[0018]** En otras realizaciones, uno o más nucleótidos (p. ej., uno o más 2'-desoxinucleótidos) están presentes en la posición interna de la pieza de cabeza o la etiqueta.

**[0019]** En algunas realizaciones, el paso (iii) puede incluir el uso de uno o más cationes solubles multivalentes (p. ej., cloruro de magnesio, cloruro de manganeso (II), o cloruro de hexamina cobalto (III), polietilenglicol (p. ej., que tiene un peso molecular medio de aproximadamente 4.600 Daltons) y ligasa ARN (p. ej., ligasa T4 ARN).

**[0020]** En otro aspecto, la invención presenta métodos como se define en las reivindicaciones para identificar y/o descubrir una entidad química, incluyendo el método el marcado de una primera biblioteca que incluye una entidad química de oligonucleótidos codificados (p. ej., incluyendo los pasos (i) a (iii) y opcionalmente incluir los pasos (iv) a (v)) y seleccionar una característica o función particular (p. ej., seleccionar para unirse a una proteína diana que incluye exponer la entidad química codificada por oligonucleótidos o entidad química a la proteína diana y seleccionar la una o más entidades químicas codificadas por oligonucleótidos o entidades químicas que se unen a la proteína diana (p. ej., usando cromatografía de exclusión por tamaño)). También se describe un complejo que incluye un casco y una etiqueta de bloque de construcción, donde la etiqueta incluye 5 a 20 nucleótidos, un nucleótido sustituido en 2' en el extremo 5' y un nucleótido sustituido en 2' en el extremo 3'. En algunas realizaciones, el nucleótido sustituido en 2' en el extremo 5' y/o en el extremo 3' es un nucleótido 2'-O-metilo (p. ej., 2'-O-metilo guanina o 2'-O-metilo uracilo) o un 2'-fluoro-nucleótido (p. ej., 2'-fluoroguanina o 2'-fluorouracilo). En realizaciones particulares, el casco incluye una estructura de horquilla. En algunas realizaciones, el casco incluye un nucleótido sustituido en 2' en uno o más del extremo 5', el extremo 3' o la posición interna del casco. En otras realizaciones, el casco incluye además un extremo 5' preadenilado. En otras realizaciones más, el casco incluye de 5 a 20 nucleótidos.

**[0021]** En cualquiera de las realizaciones anteriores, la pieza de cabeza, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción, o una o más etiquetas de bloque de construcción más adicionales, si está presente, incluye un extremo 5' preadenilado.

**[0022]** En cualquiera de las realizaciones anteriores, el método incluye además la unión de una o más (p. ej., una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, o diez) etiquetas de bloque de construcción adicional al complejo y unir uno o más (p. ej., uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve o diez) componentes adicionales (p. ej., andamios o bloques de construcción) al complejo, donde la etiqueta de uno o más bloques de construcción adicionales codifica uno o más componentes adicionales o codifica la reacción de unión de uno o más componentes adicionales, proporcionando así una biblioteca etiquetada.

**[0023]** En cualquiera de las realizaciones anteriores, el nucleótido sustituido en 2' es un nucleótido 2'-O-metilo, tales como 2'-O-metilo guanina, 2'-O-metilo uracilo, 2'-O-metilo adenosina, 2'-O-metilo timidina, 2'-O-metilo inosina, 2'-O-



metilo citidina o 2'-O-metilo diamino purina. Alternativamente, en cualquiera de las realizaciones anteriores, el nucleótido sustituido en 2' es un nucleótido 2'-fluoro, tal como 2'-fluoroguanina, 2'-fluoro uracilo, 2'-fluoro adenosina, 2'-fluoro timidina, 2' -fluoro-inosina, 2'-fluoro-citidina o 2'-fluoro-diamino-purina.

5 **[0024]** En cualquiera de las realizaciones anteriores, la ligasa ARN es ligasa T4 ARN y/o la ligasa ADN es una ligasa ssADN.

10 **[0025]** En cualquiera de las realizaciones anteriores, el método incluye una pluralidad de piezas de cabeza. En algunas realizaciones de este método, cada casco de la pluralidad de cascos incluye una región de secuencia idéntica y una región de codificación diferente. En realizaciones particulares, la región de secuencia idéntica es una región de unión a cebador. En otras realizaciones, la región de codificación diferente es una etiqueta de bloque de construcción inicial que codifica para el casco o para una adición de un componente inicial.

15 **[0026]** En cualquiera de las realizaciones anteriores, la unión en al menos uno de los pasos (ii)-(iv), si está presente, incluye la enzima de ligación y/o ligación química. En algunas realizaciones, la ligación enzimática incluye el uso de una ligasa ARN (p. ej., ligasa T4 ARN) o una ligasa ADN (p. ej., ligasa ssADN). En otras realizaciones, la ligación enzimática incluye el uso de una ligasa ARN (p. ej., ligasa T4 ARN) y una ligasa ADN (p. ej., ligasa ssADN). En algunas realizaciones, la ligación química incluye el uso de uno o más pares químicamente co-reactivos (p. ej., un par que incluye un grupo alquínico opcionalmente sustituido con un grupo azido opcionalmente sustituido; un par que incluye un dieno opcionalmente sustituido que tiene un sistema de 4 electrones  $\pi$  (por ejemplo, un compuesto 1,3-insaturado opcionalmente sustituido, tal como 1,3-butadieno opcionalmente sustituido, 1-metoxi-3-trimetilsililoxi-1,3-butadieno, ciclopentadieno, ciclohexadieno o furano) con un dienófilo opcionalmente sustituido o un heterodienófilo sustituido que tiene un sistema de 2 electrones  $\pi$  (p. ej., un grupo alquénico opcionalmente sustituido o un grupo alquínico opcionalmente sustituido); un par que incluye un nucleófilo (p. ej., una amina opcionalmente sustituida o un tiol opcionalmente sustituido) con un electrófilo heterocíclico deformado (p. ej., epóxido, aziridina, ion aziridinio o ion episulfonio opcionalmente sustituidos); un par que incluye un grupo fosforotioato con un grupo yodo (p. ej., un grupo fosforotioato en el extremo 3' y un grupo yodo en el extremo 5'); o un par que incluye un grupo aldehído con un grupo amino (p. ej., un grupo amino primario o un grupo amino secundario, incluyendo un grupo hidrazido)). En realizaciones particulares, el par químicamente co-reactivo produce un espaciador resultante que tiene una longitud de aproximadamente 4 a aproximadamente 24 átomos (p. ej., de aproximadamente 4 a aproximadamente 10 átomos). En otras realizaciones, la ligación química incluye el uso de un grupo fosforotioato (p. ej., en el extremo 3') y un grupo yodo (p. ej., en el extremo 5'). En realizaciones adicionales, la ligación química incluye un oligonucleótido de férula en la reacción de unión. En algunas realizaciones, la ligación química incluye el uso de un grupo fosforotioato (p. ej., en el extremo 3' del casco, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción, la una o más etiquetas de bloque de construcción adicionales, la etiqueta de identificación de biblioteca, la etiqueta de uso y/o la etiqueta de origen, si está presente), un grupo de yodo (p. ej., en el extremo 5' del casco, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción, la una o más etiquetas de bloques de construcción adicionales, la etiqueta de identificación de la biblioteca, la etiqueta de uso y/o la etiqueta de origen, si está presente) y un oligonucleótido de férula en la reacción de unión, donde el uso evita el uso de uno o más grupos protectores. En otras realizaciones, la ligación química de múltiples marcadores comprende el uso alternativo de pares ortogonales químicamente co-reactivos (p. ej., dos o más pares químicamente co-reactivos que se describen en el presente documento) para ligar marcadores sucesivos.

45 **[0027]** En cualquiera de las realizaciones anteriores, la pieza de cabeza puede incluir una estructura de una sola cadena (p. ej., de horquilla).

50 **[0028]** En cualquiera de las realizaciones anteriores, la pieza de cabeza, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción, la una o más etiquetas de bloque de construcción más adicionales, la etiqueta de biblioteca de identificación, la etiqueta de uso, y/o la etiqueta de origen, si está presente, incluye una secuencia que es sustancialmente idéntica (p. ej., al menos 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, o 100% idéntica) a cualquier secuencia descrita en este documento (p. ej., la secuencia en cualquiera de las SEQ ID NO: 6-21, 26, 27 o 29-31), o una secuencia que es complementaria a una secuencia que es sustancialmente idéntica (p. ej., al menos 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% o 100% idéntica) a cualquier secuencia descrita en el presente documento (p. ej., la secuencia en cualquiera de las SEQ ID NO: 6-21, 26, 27 o 29-31). En realizaciones particulares, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción, la una o más etiquetas de bloque de construcción adicionales, la etiqueta de identificación de biblioteca, la etiqueta de uso y/o la etiqueta de origen, si está presente, incluye además una secuencia que es sustancialmente idéntica (p. ej., al menos 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% o 100% idéntica) a la secuencia de SEQ ID NO:1 o SEQ ID NO: 2.

60 **[0029]** En cualquiera de las realizaciones anteriores, los métodos o complejos incluyen sólo las moléculas de cadena sencilla, en donde la pieza de cabeza, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción, y/o una o más etiquetas de bloques de construcción adicionales son monocatenarias. En algunas realizaciones, una o más de las moléculas monocatenarias tienen una estructura de horquilla. En realizaciones particulares, la pieza de cabeza incluye una estructura de horquilla y una o más etiquetas de bloques de construcción no incluyen una estructura de horquilla.

**[0030]** En cualquiera de las realizaciones anteriores, el procedimiento comprende además una o más etapas opcionales de diversificar la biblioteca o para 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 biblioteca similar a un fármaco que se une o inactiva una proteína de interés terapéutico. En otras realizaciones, el método comprende además poner en contacto a un miembro de la biblioteca con una diana biológica en condiciones adecuadas para que al menos un miembro de la biblioteca se una a la diana, eliminando uno o más miembros de la biblioteca que no se unen a la diana, y analizar las una o más etiquetas oligonucleotídicas asociadas con ellas.

**[0031]** Como se describe en el presente documento, el uso de moléculas de cadena simple (p. ej., incluyendo moléculas de horquilla) podría tener numerosos beneficios. Por consiguiente, en cualquiera de las realizaciones descritas en el presente documento, los métodos y complejos incluyen un casco, una o más etiquetas de bloques de construcción, un complejo, una entidad química, una molécula o cualquier miembro de una biblioteca etiquetada que tenga una masa disminuida, una solubilidad aumentada (p. ej., en un solvente orgánico), menor costo, mayor reactividad, mayor accesibilidad a la diana, menor radio hidrodinámico y/o mayor precisión de las evaluaciones analíticas, en comparación con un método que incluye una o más moléculas de doble hebra (p. ej., una pieza de cabeza de doble hebra o una etiqueta de bloque de construcción de doble hebra). En algunas realizaciones, cada una de las etiquetas de bloque de construcción (p. ej., la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción y/o una o más etiquetas de bloque de construcción adicionales, si están presentes) tiene aproximadamente la misma masa (p. ej., cada etiqueta de bloque de construcción tiene una masa que es aproximadamente +/- 10% de la masa promedio entre dos o más etiquetas de bloques de construcción). En realizaciones particulares, la etiqueta de bloque de construcción tiene una masa reducida (p. ej., menos de aproximadamente 15.000 Dalton, aproximadamente 14.000 Dalton, aproximadamente 13.000 Dalton, aproximadamente 12.000 Dalton, aproximadamente 11.000 Dalton, aproximadamente 10.000 Dalton, aproximadamente 9.000 Dalton, aproximadamente 8.000 Dalton, aproximadamente 7.500 Daltons, aproximadamente 7.000 Daltons, aproximadamente 6.000 Daltons, aproximadamente 6.500 Daltons, aproximadamente 5.000 Daltons, aproximadamente 5.500 Daltons, aproximadamente 4.000 Daltons, aproximadamente 4.500 Daltons o aproximadamente 3.000 Daltons) en comparación con una etiqueta de doble hebra (p. ej., una etiqueta de doble hebra que tiene una masa de aproximadamente 15.000 Dalton, aproximadamente 14.000 Dalton, aproximadamente 13.000 Dalton o aproximadamente 12.000 Dalton). En otras realizaciones, la etiqueta de bloque de construcción tiene una longitud reducida en comparación con una etiqueta de doble hebra (p. ej., una etiqueta de doble hebra que tiene una longitud de menos de aproximadamente 20 nucleótidos, menos de aproximadamente 19 nucleótidos, menos de aproximadamente 18 nucleótidos, menos de 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 de bloques de construcción o miembros de la biblioteca carecen de una región de unión de cebadores y/o una región constante (p. ej., durante un paso de selección, como la selección mediante cromatografía de exclusión por tamaños). En algunas realizaciones, una o más etiquetas de bloques de construcción o miembros de la biblioteca tienen una región constante reducida (p. ej., una longitud menor de aproximadamente 30 nucleótidos, menos de aproximadamente 25 nucleótidos, menos de aproximadamente 20 nucleótidos, menos de aproximadamente 19 nucleótidos, menos de aproximadamente 18 nucleótidos, menos de 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 otras realizaciones, los métodos incluyen un casco que codifica una molécula, una porción de una entidad química, una reacción de unión (p. ej., ligación química o enzimática) de un paso, o la identidad de una biblioteca, donde el casco codificador elimina la necesidad de una etiqueta de bloque de construcción adicional para codificar dicha información.

**[0032]** En cualquiera de las realizaciones anteriores, un oligonucleótido (p. ej., la pieza de cabeza, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción, y/o una o más etiquetas de bloque de construcción más adicionales, si está presente) codifica para la identidad de la biblioteca. En algunas realizaciones, el oligonucleótido (p. ej., pieza de cabeza, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción y/o una o más etiquetas de bloque de construcción adicionales, si están presentes) incluye una primera secuencia de identificación de biblioteca, donde la secuencia codifica para la identidad de la primera biblioteca. En realizaciones particulares, el oligonucleótido es una primera etiqueta de identificación de biblioteca. En algunas realizaciones, el método incluye proporcionar una primera etiqueta de identificación de biblioteca, donde la etiqueta incluye una secuencia que codifica 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 una segunda biblioteca.

**[0033]** En cualquiera de las realizaciones anteriores, un oligonucleótido (p. ej., una pieza de cabeza y/o uno o más bloques de construcción) codifica para el uso del miembro de la biblioteca (p. ej., el uso en una etapa de selección o

una etapa de unión, como se describe aquí). En algunas realizaciones, el oligonucleótido (p. ej., la pieza de cabeza, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción y/o una o más etiquetas de bloque de construcción 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 uno o más pasos (p. ej., un paso de selección y/o un paso de unión). En realizaciones particulares, el oligonucleótido es una etiqueta de uso que incluye una secuencia de uso. En algunas realizaciones, un oligonucleótido (p. ej., un casco y/o uno o más bloques de construcción) codifica el origen del miembro de la biblioteca (p. ej., en una parte particular de la biblioteca). En algunas realizaciones, el oligonucleótido (p. ej., la pieza de cabeza, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción y/o una o más etiquetas de bloque de construcción adicionales, si están presentes) incluye una secuencia de origen (p. ej., una secuencia degenerada aleatoria 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 realizaciones particulares, el oligonucleótido es una etiqueta de origen que incluye una secuencia de origen. En algunas realizaciones, el método incluye además unir, enlazar o asociar operativamente una etiqueta de uso y/o una etiqueta de origen al complejo.

**[0034]** En cualquiera de las realizaciones anteriores, los métodos, las composiciones y complejos incluyen, opcionalmente, un tubo de desagüe, donde el tubo de desagüe incluye una o más de una secuencia de biblioteca de identificación, una secuencia de uso, o una secuencia de origen, tal como se describe aquí. En realizaciones particulares, los métodos incluyen además unir, enlazar o asociar operativamente el cordal (p. ej., incluyendo una o más de una secuencia de identificación de biblioteca, una secuencia de uso o una secuencia de origen) al complejo.

**[0035]** En cualquiera de las realizaciones anteriores, los métodos, composiciones y complejos, o porciones de los mismos (p. ej., la pieza de cabeza, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción, y/o la una o más etiquetas de bloque de construcción más adicionales, si está presente), incluye un grupo fosfato modificado (p. ej., un fosforotioato o un enlace 5'-N-fosforamidita) entre el nucleótido terminal en el extremo 3' y el nucleótido adyacente al nucleótido terminal. En realizaciones particulares, el grupo fosfato modificado minimiza la mezcla durante la ligación enzimática entre dos oligonucleótidos (p. ej., minimiza la inclusión de un nucleótido adicional o la escisión de un nucleótido en el producto final o complejo, en comparación con las secuencias de dos oligonucleótidos que se van a ligar, tal como como entre un casco y una etiqueta de bloque de construcción o entre una primera etiqueta de bloque de construcción y una segunda etiqueta de bloque de construcción), en comparación con la ligación entre dos oligonucleótidos (p. ej., un casco y una etiqueta de bloque de construcción o una primera etiqueta de bloque de construcción y una segunda etiqueta de bloque de construcción) que carece del grupo fosfato modificado. En algunas realizaciones, el complejo puede incluir un grupo fosforotioato o triazol.

**[0036]** En cualquiera de las realizaciones anteriores, los métodos, composiciones y complejos, o porciones de los mismos (p. ej., la pieza de cabeza, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción, y/o la una o más etiquetas de bloque de construcción más adicionales, si está presente), incluye una modificación que apoya la solubilidad en condiciones semi, reducidas o no acuosas (p. ej., orgánicas). En algunas realizaciones, el enlazador bifuncional, la pieza de cabeza o una o más etiquetas de bloques de construcción se modifican 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 de anillo hidrófoba. En algunas realizaciones, la modificación incluye uno o más nucleótidos modificados que tienen un resto hidrófobo (p. ej., modificado en las posiciones C5 de las bases T o C con cadenas alifáticas, como en 5'-dimetoxitritilo-N4-diisobutilaminometiliden-5-(1-propinilo)-2'-desoxicitidina, 3'-[(2-cianoetilo)-(N,N-diisopropilo)]-fosforamidita; 5'-dimetoxitritilo-5-(1-propinilo)-2'-desoxiuridina, 3'-[(2-cianoetilo)-(N,N-diisopropilo)]-fosforamidita; 5'-dimetoxitritilo-5-fluoro-2'-desoxiuridina, 3'-[(2-cianoetilo)-(N,N-diisopropilo)]-fosforamidita; y 5'-dimetoxitritilo-5-(piren-1-ilo-etinilo)-2'-desoxiuridina, o 3'-[(2-cianoetilo)-(N,N-diisopropilo)]-fosforamidita) o una inserción que tiene un resto hidrófobo (p. ej., un azobenceno). En algunas realizaciones, el miembro de la biblioteca tiene un octanol: agua coeficiente de aproximadamente 1,0 a aproximadamente 2,5 (p. ej., 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 de aproximadamente 2,0 a aproximadamente 2,5).

**[0037]** En cualquiera de las realizaciones anteriores, la pieza de cabeza, el tubo de desagüe, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción, la una o más etiquetas de bloque de construcción más adicionales, la etiqueta de biblioteca de identificación, la etiqueta de uso, y/o la etiqueta de origen, si está presente, puede incluir de 5 a 20 nucleótidos (p. ej., 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 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 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 y de 19 a 20 nucleótidos). En realizaciones particulares, la pieza de cabeza, la primera etiqueta de bloque de construcción, la segunda etiqueta de bloque de construcción, la una o más etiquetas de bloque de construcción adicionales, la etiqueta de identificación de biblioteca, la etiqueta de uso y/o la etiqueta de origen, si está presente, tienen un longitud de menos de 20 nucleótidos (p. ej., 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 de 10 nucleótidos, menos de 9 nucleótidos, menos de 8 nucleótidos o menos de 7 nucleótidos).

**[0038]** En realizaciones particulares, la primera etiqueta de bloque de construcción y la segunda etiqueta de bloque de construcción incluyen el mismo número de nucleótidos. En otras realizaciones, la primera etiqueta de bloque de construcción o la segunda etiqueta de bloque de construcción incluye más de 8 nucleótidos (p. ej., más de 9 nucleótidos, más de 10 nucleótidos, más de 11 nucleótidos, más de 12 nucleótidos, más de 13 nucleótidos, más de 14 nucleótidos y más de 15 nucleótidos). En algunas realizaciones, la primera etiqueta de bloque de construcción es una etiqueta donante (p. ej., como se define aquí) que tiene de 8 a 20 nucleótidos (p. ej., 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 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, y de 19 a 20 nucleótidos).

#### *Definiciones*

**[0039]** Por "nucleótidos sustituidos 2'" se entiende una base de nucleótidos que tiene una sustitución en la posición 2' de la ribosa en la base.

**[0040]** Por "aproximadamente" se quiere decir +/- 10% del valor recitado.

**[0041]** Por "bifuncional" se entiende tener dos grupos reactivos que permiten la unión de dos restos químicos. Por ejemplo, un enlazador bifuncional es un enlazador, como se describe en el presente documento, que tiene dos grupos reactivos que permiten la unión de una pieza de cabeza y una entidad química.

**[0042]** 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/u otras fuerzas intermoleculares. La unión se puede efectuar por cualquier medio útil, como mediante unión enzimática (p. ej., ligación enzimática) o mediante unión química (p. ej., ligación química).

**[0043]** Por "bloque de construcción" se entiende una unidad estructural de una entidad química, donde la unidad está directamente vinculada a otras unidades estructurales químicas o indirectamente enlazadas a través del andamio. Cuando la entidad química es polimérica u oligomérica, los componentes básicos 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. Generalmente, 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, los pasos químicos o de unión codificados pueden incluir varios componentes químicos (p. ej., reacciones de condensación de múltiples componentes o procesos de múltiples pasos). Los grupos reactivos en dos bloques de construcción diferentes deben ser complementarios, es decir, capaces de reaccionar juntos para formar un enlace covalente o no covalente.

10 **[0044]** Por "etiqueta de bloque de construcción" se entiende una porción de oligonucleótido de la biblioteca que codifica la adición (p. ej., por una reacción de unión) de un componente (es decir, un andamio o un bloque de construcción), la pieza de cabeza en la biblioteca, la identidad de la biblioteca, el uso de la biblioteca y/o el origen de un miembro de la biblioteca. Por "etiqueta aceptora" se entiende una etiqueta de bloque de construcción que tiene una entidad reactiva (p. ej., un grupo hidroxilo en el extremo 3' en el caso de la ligación enzimática). Por "etiqueta donante" se entiende  
 15 una etiqueta de bloque de construcción que tiene una entidad capaz de reaccionar con la entidad reactiva en la etiqueta aceptora (p. ej., un grupo fosforilo en el extremo 5' en el caso de la ligación enzimática).

**[0045]** Por "entidad química" se entiende un compuesto que comprende uno o más bloques de construcción y, opcionalmente, un andamio. La entidad química puede ser cualquier fármaco peptídico o de molécula pequeña o candidato a fármaco diseñado o construido para tener una o más características deseadas, p. ej., capacidad para  
 20 unirse a una diana biológica, 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 determinadas realizaciones, la entidad química puede reaccionar además como una entidad bifuncional o trifuncional (o mayor).

25 **[0046]** Por "par químicamente co-reactivo" se quiere decir un par de grupos reactivos que participa en una reacción modular con un alto rendimiento y una alta ganancia termodinámica, produciendo así un espaciador. Las reacciones ejemplares y los pares químicamente reactivos incluyen 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 4 electrones  $\pi$  y un dienófilo  
 30 opcionalmente sustituido o un heterodienófilo opcionalmente sustituido que tiene un sistema de 2 electrones  $\pi$ ; una reacción de apertura de anillo con un nucleófilo y un electrófilo heterocíclico tensado; una reacción de ligación en férula con un grupo fosforotioato y un grupo yodo; y una reacción de aminación reductora con un grupo aldehído y un grupo amino, como se describe en el presente documento.

35 **[0047]** Por "complejo" o "complejo ligado" se quiere decir una pieza de cabeza que está asociada 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 conector bifuncional entre la entidad química y la pieza de cabeza.

40 **[0048]** Por "componente" de una entidad química se entiende o bien un andamio o un bloque de construcción.

**[0049]** Por "nodo diversidad" se quiere decir un grupo funcional en una posición en el andamio o el bloque de construcción que permite la adición de otro bloque de construcción.

45 **[0050]** Por "pieza de cabeza" se entiende un oligonucleótido de partida para la síntesis de la biblioteca que está unida operativamente a un componente de una entidad química y a una etiqueta de bloque de construcción. Opcionalmente, un enlazador bifuncional conecta la pieza de cabeza al componente.

50 **[0051]** Por "biblioteca" se entiende una colección de moléculas o entidades químicas. Opcionalmente, las moléculas o entidades químicas se unen a uno o más oligonucleótidos que codifican las moléculas o partes de la entidad química.

**[0052]** Por "enlazador" se entiende una entidad de conexión química que enlaza la pieza de cabeza a una entidad química.

55 **[0053]** Por "catión multivalente" se quiere decir 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. Los cationes multivalentes ejemplares incluyen los de metales alcalinotérreos (p. ej., magnesio) y metales de transición (p. ej., manganeso (II) o cobalto (III)), y aquellos que están opcionalmente unidos a uno o más aniones y/o uno o más ligandos univalentes o polidentados, tales como cloruro, amina y/o etilendiamina.

60 **[0054]** Por "oligonucleótido" se entiende un polímero de nucleótidos que tienen 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 por qué tener bases contiguas, pero se puede intercalar con restos enlazadores. El  
 65 polímero oligonucleotídico puede incluir bases naturales (p. ej., adenosina, timidina, guanosa, citidina, uridina, desoxiadenosina, desoxitimidina, desoxiguanosina, desoxicitidina, inosina o purina de diamino), análogos de bases

(p. ej., 2-aminoadenosina, 2-tiotimidina, pirrolo-pirimidina, 3-metilo adenosina, C5-propinilcitolina, C5-propiniluridina, C5-bromouridina, C5-fluorouridina, C5-yodouridina, C5-metilcitolina, 7-desazaadenosina, 7-desazaguanosina, 8-oxoadenosina, 8-oxoguanosina, O(6)-metilguanina y 2-tiocitidina), nucleótidos modificados (p. ej., nucleótidos sustituidos en 2', como bases 2'-O-metiladas y bases 2'-fluoro), bases intercaladas, azúcares modificados (p. ej., 2'-fluororibosa, ribosa, 2'-desoxirribosa, arabinosa y hexosa) y/o grupos fosfato modificados (p. ej., fosforotioatos y enlaces 5'-N-fosforamidita). En este documento se describen otras bases modificadas. Por "oligonucleótido aceptor" se entiende un oligonucleótido que tiene una entidad reactiva (p. ej., un grupo hidroxilo en el extremo 3' en el caso de la ligación enzimática o un grupo azido opcionalmente sustituido en el caso de la ligación química). Por "oligonucleótido donante" se entiende un oligonucleótido que tiene una entidad capaz de reaccionar con la entidad reactiva en el oligonucleótido aceptor (p. ej., un grupo fosforilo en el extremo 5' en el caso de la ligación enzimática o un grupo alquililo opcionalmente sustituido en el caso de la ligación química).

**[0055]** Por "unido operativamente" o "asociado operativamente" se quiere decir que dos o más estructuras químicas están directamente o indirectamente vinculadas de una manera tal como para permanecer unidas a través de las diversas manipulaciones a las que se espera que se sometan. Normalmente, la entidad química y la pieza de cabeza están operativamente enlazadas de una manera indirecta (p. ej., covalentemente mediante un enlazador apropiado). Por ejemplo, el enlazador puede ser un resto bifuncional con un sitio de unión para la entidad química y un sitio de unión para la pieza de cabeza. Además, la entidad química y la etiqueta oligonucleotídica pueden unirse operativamente directa o indirectamente (p. ej., covalentemente a través de un enlazador apropiado).

**[0056]** Por "grupo protector" se entiende un grupo destinado a proteger el extremo 3' o extremo 5' de un oligonucleótido frente a reacciones indeseables durante uno o más pasos de unión de etiquetado de una biblioteca codificada por ADN. Los grupos protectores comúnmente usados se describen en Greene, Protective Groups in Organic Synthesis, 4ª edición (John Wiley & Sons, Nueva York, 2007). Los ejemplos de grupos protectores 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 (p. ej., éster de O-( $\alpha$ -metoxietilo), O-isovalerilo éster y O-levulinilo éster), grupos tritilo (p. ej., dimetoxitritilo y monometoxitritilo), grupos xantenilo (p. ej., 9-fenilxanten-9-ilo y 9-(p-metoxifenilo) xanten-9-ilo), grupos acilo (por ejemplo, fenoxiacetilo y acetilo) y grupos sililo (p. ej., t-butildimetilsililo).

**[0057]** Por "purificar" se entiende la eliminación de cualquier producto que no ha reaccionado o cualquier agente presente en una mezcla de reacción que pueda reducir la actividad de un agente químico o biológico a ser utilizado en una etapa sucesiva. La purificación puede incluir una o más de separación cromatográfica, separación electroforética y precipitación del producto o reactivo sin reaccionar a ser eliminado.

**[0058]** Por "andamio" se entiende un resto químico que muestra uno o más nodos de la diversidad en una geometría especial particular. Los nodos de diversidad se adjuntan típicamente al andamio durante la síntesis de la biblioteca, pero en algunos casos se puede adjuntar un nodo de diversidad al andamio antes de la síntesis de la biblioteca (p. ej., adición de uno o más bloques de construcción y/o una o más etiquetas). En algunas realizaciones, el andamio se deriva de manera que se puede desproteger ortogonalmente durante la síntesis de la biblioteca y posteriormente reaccionar con diferentes nodos de diversidad.

**[0059]** Por fármaco de "molécula pequeña" o candidato de fármaco "molécula pequeña" se entiende una molécula que tiene un peso molecular por debajo de aproximadamente 1.000 Daltons. Las moléculas pequeñas pueden ser orgánicas o inorgánicas, aisladas (p. ej., de bibliotecas de compuestos o fuentes naturales) u obtenidas por derivatización de compuestos conocidos.

**[0060]** Por "identidad sustancial" o "sustancialmente idéntica" se quiere decir una secuencia de polipéptido o polinucleótido que tiene la misma secuencia de polipéptido o polinucleótido, respectivamente, como una 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 será generalmente 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 los ácidos nucleicos, la longitud de las secuencias de comparación será generalmente 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 se puede medir usando software de análisis de secuencia en el ajuste predeterminado (p. ej., Sequence Analysis Software Package del Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). Dicho software puede hacer coincidir secuencias similares asignando grados de homología a diversas sustituciones, delecciones y otras modificaciones.

**[0061]** Por "pieza de cola" se entiende una porción de oligonucleótido de la biblioteca que se une al complejo después de la adición de todas las etiquetas de bloque de construcción y codifica para la identidad de la biblioteca, el uso de la biblioteca, y/o la origen de un miembro de la biblioteca.

- 5 **[0062]** Otras características y ventajas de la invención serán evidentes a partir de la siguiente descripción detallada y las reivindicaciones.

#### Breve descripción de los dibujos

10 **[0063]**

**La Figura 1** muestra un método ejemplar para la síntesis general de bibliotecas químicas utilizando etiquetas de ADN monocatenarias que se unen secuencialmente mediante ligación enzimática y/o química. "BB" se refiere al bloque de construcción.

15 **Las Figuras 2A-2B** muestran métodos ejemplares para el etiquetado de bibliotecas de ADN monocatenario usando ligación enzimática. La Figura 2A muestra un método ejemplar para etiquetar bibliotecas usando ligación enzimática monocatenaria con un oligonucleótido 5'-monofosfato (5'-P) protegido (reinstalado), donde los recuadros grises se refieren a nucleótidos 2'-OMe, "X" se refiere a un grupo protector o un componente de una entidad química, y "PNK" se refiere a polinucleótido quinasa. La Figura 2B muestra un método ejemplar para etiquetar bibliotecas usando ligación monocatenaria con un oligonucleótido 3'-OH protegido, donde las cajas negras unidas a -O- se refieren a un grupo protector del terminal 3'-OH y "LC" se refiere a la separación cromatográfica de líquido del grupo protector.

20 **La Figura 3** muestra un método ejemplar para etiquetar bibliotecas usando ligación monocatenaria con un oligonucleótido 5' preadenilado (etiquetado como "5'-App") (pieza de cabeza) con un extremo 3' que está bloqueado, p. ej., por una entidad química (etiquetado "X-3'"). Este método se puede utilizar para ligar una etiqueta oligonucleotídica fosforilada en 5' (etiquetada como "Etiqueta A") a la pieza de cabeza y etiquetas adicionales que tengan un extremo 3'-OH (etiquetado como "Etiqueta B" y "Etiqueta C") al complejo en la presencia de ATP.

25 **Las Figuras 4A-4E** muestran complejos ejemplares, cada uno con una pieza de cabeza, un enlazador y una molécula pequeña que incluye un andamio ("S") y nodos de diversidad A, B y C. Los recuadros de color gris oscuro se refieren a nucleótidos 2'-OMe, y las líneas de puntos se refieren a la presencia de una o más bases complementarias. Las Figuras 4A-4B son esquemas de complejos que tienen un cabezal oligonucleotídico lineal monocatenario, donde el conector y la molécula pequeña están conectados al extremo 3' (Figura 4A) o al extremo 5' (Figura 4B) del cabezal. Las figuras 4C-4D son esquemas de complejos que tienen un cabezal oligonucleotídico en horquilla monocatenario, donde el conector y la molécula pequeña están conectados a la posición interna (Figura 4C) o al extremo 3' (Figura 4D) del cabezal. La Figura 4E muestra un método ejemplar para marcar bibliotecas que tienen una pieza de cabeza de oligonucleótidos en horquilla, donde la estrella se refiere a un resto químico e" y "en el extremo 3' se refiere a un grupo protector. Las etiquetas de oligonucleótidos están marcadas de 1 a 4 y la secuencia adaptadora es la línea negra en el extremo 5'.

30 **Las Figuras 5A-5C** muestran la ligación de oligonucleótidos por ligasa T4 ARN o CircLigase™ ligasa ssADN. La Figura 5A es un esquema de la reacción de ligación enzimática. El oligonucleótido donante está fosforilado en 5' y lleva un marcador de fluoresceína 3', imitando una pieza de cabeza con una biblioteca química en el extremo 3'. El oligonucleótido aceptor no está fosforilado. La Figura 5B muestra el análisis de electroforesis en gel de una reacción de ligación en un gel de urea 8 M/poliacrilamida al 15% (PAAG). "SM" se refiere a un donante marcado con fluorescencia, "Producto" se refiere a un producto de ligación y "Donante adenilado" se refiere a un donante 5'App, como se describió anteriormente. La Figura 5C muestra la ligación de alto rendimiento lograda para la ligasa T4 ARN a altas concentraciones de enzima y oligonucleótidos.

35 **Las Figuras 6A-6B** representan la optimización del peso molecular de PEG (Figura 6A) y la concentración (Figura 6B) para lograr el rendimiento máximo de ligación mediante ligasa T4 ARN. Las condiciones de reacción son las descritas anteriormente para las Figuras 5A-5C. La Figura 6A es un gráfico que cuantifica el análisis electroforético de una reacción de ligación con etiquetas de donante y aceptor de 15 mer MNA/ADN después de la incubación durante 5 horas o 20 horas con PEG al 25% (p/v) que tiene un peso molecular de 300 a 20.000 (20K). La Figura 6B muestra el efecto de la concentración sobre la ligación después de la incubación durante 18-20 horas en presencia de 5% a 45% (p/v) de PEG4600.

40 **Las Figuras 7A-7B** muestran una correlación entre la eficacia de la ligación por CircLigase™ (Figura 7A) y la ligasa ARN T4 (Figura 7B) y la longitud de los oligonucleótidos donadores o aceptores. La Figura 7A representa un gráfico que cuantifica el efecto de la longitud del aceptor sobre el rendimiento de la ligación en la reacción de ligación de CircLigase™. La Figura 7B representa un gráfico y una tabla que cuantifican el efecto de la longitud de nucleótidos de las etiquetas de MNA/ADN aceptor y donante sobre la ligación monocatenaria con ligasa T4 ARN. Estos datos representan un promedio de dos experimentos independientes obtenidos por densitometría de geles fluorescentes a una excitación de 450 nm.

45 **Las Figuras 8A-8B** son espectros de CL-EM para una etiqueta de MNA/ADN antes y después de la fosforilación. Los datos se muestran para la etiqueta de 15 mer 5'-HO-mUAC GTA TAC GAC TGmG-OH-3' (SEQ ID NO: 13) (a 250 μm) antes (Figura 8A) y después (Figura 8B) de la reacción con la polinucleótido quinasa T4 (50 unidades por 5 nmoles de etiqueta).

**La Figura 9** muestra un gel electroforético para la ligación secuencial monocatenaria de etiquetas A-C. El extremo 3' incluía fluoresceína para representar un compuesto de biblioteca (o entidad química), y el asterisco (\*) indica purificación del producto ligado (o complejo) antes de la fosforilación.

**Las Figuras 10A-10B** muestran esquemas de una reacción de "par químicamente co-reactivo" entre oligonucleótidos donantes y aceptores que dan como resultado un espaciador "corto" de 5 átomos (Figura 10A) y un espaciador "largo" de 24 átomos (Figura 10B).

**Las Figuras 11A-11E** muestran los resultados de la transcripción inversa (TI) y el análisis de PCR de las plantillas de ADN de 75 mer que contienen un solo espaciador corto o largo, como se muestra en las Figuras 10A-10B. La Figura 11A es un esquema de la reacción de TI. Los espectros CLEM de la TI se registraron a 260 nm y 650 nm para la plantilla de ADN de 75 mer de control (Figura 11B), la plantilla de ADN de 75 mer que contenía un espaciador de 5 átomos ("corto") (Figura 11C) y el ADN de plantilla de 75 mer que contiene un solo espaciador de 24 átomos ("largo") (Figura 11D). La Figura 11E muestra el análisis de TI-PCR de la plantilla de ADN de control de 75 mer ("templ75"), una plantilla de ADN de 75 mer con un espaciador de 5 átomos ("clic corto") y una plantilla de ADN de 75 mer con un espaciador de 24 átomos ("clic largo").

**Las Figuras 12A-12G** muestran los resultados de una reacción de ligación química entre un oligonucleótido de ADN modificado con yodo en 5' y un oligonucleótido de ADN de 3'-fosforotioato en presencia o ausencia de un oligonucleótido de férula complementario. La Figura 12A muestra un ejemplo esquemático de la reacción. El oligonucleótido de 5'-yodo está marcado con 6-FAM en el extremo 3', mientras que el oligonucleótido de fosforotioato 3' está marcado con Cy5 en el extremo 5'. La Figura 12B muestra un análisis de electroforesis en gel de las reacciones de ligación en presencia (+spl) o ausencia (-spl) de una férula complementaria. CCy5 y CFL indican bandas visibles de Cy5 y material de partida marcado con fluoresceína, respectivamente. La Figura 12C muestra un curso de tiempo de la reacción de ligación ferulizada en las condiciones anteriores, que se cuantificó usando Cy5 (635 nm) y detección de fluoresceína (450 nm). La Figura 12D muestra el análisis CL-EM de la ligación de CFL y CCy5 en ausencia (superior, a 260 nm, 495 nm y 650 nm) y presencia (inferior, a 260 nm, 495 nm y 650 nm) de una férula, donde las reacciones de ligación se incubaron durante siete días. La Figura 12E muestra el análisis CL-EM de la ligación de CFL y CCy5 en ausencia de una férula (a 260 nm, 495 nm y 650 nm), donde las reacciones de ligación se incubaron durante ocho días. La Figura 12F muestra el análisis por EM de la reacción del oligonucleótido CFL con piperidina, en donde se pretendía que esta reacción desplazara al yodo. Las condiciones de reacción incluyeron oligonucleótidos a 100 µM, piperidina a 40 mM (400 equivalentes) en tampón borato 100 mM, pH 9,5, durante 20 horas a temperatura ambiente (izquierda); y oligonucleótidos a 400 µM, piperidina a 2 M (4.000 equivalentes) en tampón borato 200 mM, pH 9,5, durante 2 horas a 65°C (derecha). La Figura 12G muestra el análisis de EM de una reacción de ligación ferulizada de oligonucleótidos CFL y CCy5 a 50 µM realizada en presencia de 400 equivalentes de piperidina en tampón borato 100 mM, pH 9,5, durante 20 horas a temperatura ambiente.

**Las Figuras 13A-13C** muestran el uso de oligonucleótidos modificados para minimizar la mezcla. La Figura 13A muestra un análisis CL-EM de una reacción de ligación monocatenaria de una pieza de cabeza de ssHP fosforilado en 5' (3.636 Da) y una etiqueta (etiqueta 15; 2.469 Da) que tiene nucleótidos 2'-O metilo. El análisis de CL-EM mostró tres picos: pico 1 para la etiqueta (2.469 Da); pico 2 para la pieza de cabeza adenilada (3.965 Da); y el pico 3 que tiene dos (en algunos casos tres) subpicos que contienen productos con pesos moleculares de 6.089 Da (producto de ligación esperado); 5.769 Da (esperado 6.089 Da - 320 Da); y 6.409 Da (esperado 6.089 Da + 320 Da). Esta diferencia de masa de 320 Da corresponde exactamente a la eliminación o adición de un nucleótido 2'-O-Me C adicional. Las Figuras 13B-1 a 13B-3 muestran un mecanismo propuesto no limitante de la mezcla de nucleótidos, donde aproximadamente el 90% de la reacción proporciona el producto de ligación esperado (normal) y aproximadamente el 10% de la reacción proporciona productos de ligación aberrantes ("Producto - 1 nt" y "Producto + 1 nt"). La Figura 13C muestra un análisis CL-EM de la ligación de la pieza de cabeza HP-PS con la etiqueta 15. La pieza de cabeza HP-PS tiene la secuencia de la pieza de cabeza ssHP pero incluye un enlace de fosforotioato en el extremo 5'. El análisis LC mostró tres picos: pico 1 para la etiqueta (2.469), pico 2 para la pieza de cabeza adenilada (3.984) y pico 3 para un solo producto de ligación (6.107) sin casi ninguna mezcla de nucleótidos observada. Las trazas de +/- 320 picos probablemente corresponden a la conversión oxidativa del enlace fosforotioato en un enlace fosfodiéster nativo o se deben a una sulfuración incompleta.

**La Figura 14** es un gráfico que muestra la separación de los miembros de la biblioteca usando cromatografía de exclusión por tamaño, donde los miembros de la biblioteca unidos a la diana (izquierda en el gráfico) eluyen en un tiempo más corto que los miembros de la biblioteca no unidos (a la derecha en el gráfico).

**La Figura 15A** es un esquema ejemplar que muestra la ligación química de etiquetas de ADN codificantes usando una única química que no depende de la férula, p. ej., 5'-azido/3'-alquinilo. Los grupos reactivos están presentes en los extremos 3' y 5' de cada etiqueta (Etiqueta A, B y C), y uno de los grupos reactivos en cada extremo (p. ej., el extremo 3') está protegido para evitar la ciclación, polimerización o ligación de ciclo incorrecto de las etiquetas. El ciclo de ligación de etiquetas incluye ligación química, seguida de desprotección del grupo funcional restante para hacer que la entidad ligada en crecimiento sea competente para el siguiente ciclo de ligación. Cada ciclo también incluye la adición de uno o más bloques de construcción (BBA, BBB y BBC, que están codificados por etiqueta A, B y C, respectivamente). El proceso de ligación química puede incluir opcionalmente la adición de un cordal.

**La Figura 15B** es un esquema ejemplar que muestra la ligación química de etiquetas de ADN codificantes usando una única química que depende de la férula. La naturaleza dependiente de la plantilla de este enfoque



reduce la frecuencia de aparición de polimerización de la etiqueta, ciclación de la etiqueta, así como de eventos de error. De manera similar a la Figura 15A, este esquema incluye etiquetas (Etiqueta A, B y C) y uno o más bloques de construcción codificados por etiquetas (BBA, BBB y BBC).

**La Figura 15C** es un esquema de ejemplo que muestra el uso de una sucesión de etiquetas ligadas químicamente como plantilla para polimerización dependiente de plantilla, generando ADNc que es competente para amplificación por PCR y secuenciación, así como el uso de una polimerasa dependiente de plantilla capaz de leer a través de las uniones ligadas químicamente.

**La Figura 16A** es un esquema de ejemplo que muestra la ligación química de etiquetas de ADN codificantes usando etiquetas de alquino protegidas por TIPS y química de "clic". Cada ciclo de síntesis de la biblioteca incluye la ligación química catalizada por Cu(I) de la etiqueta protegida con TIPS al alquino desprotegido del ciclo anterior. Después de la ligación, el grupo TIPS se elimina (desprotege), activando así el alquino para la siguiente etapa de ligación química.

**La Figura 16B** muestra la estructura de CPG de DMT-succinil-3'-O-TIPS-propargil uridina que se usa para iniciar la síntesis en fase sólida de oligonucleótidos que llevan 3'-O-TIPS-propargil uridina en el extremo 3'.

**La Figura 16C** es un esquema ejemplar que muestra el uso de una sucesión de etiquetas ligadas químicamente "clic" como plantilla para la polimerización dependiente de plantilla, generando ADNc que es competente para la amplificación y secuenciación por PCR, así como el uso de una polimerasa dependiente de plantilla capaz de lectura a través de las uniones ligadas químicamente "clic".

**Las Figuras 17A-17C** muestran la síntesis de las plantillas Y55 e Y185 de "un solo clic" biotiniladas en 5'. La figura 17A proporciona un esquema ejemplar. La Figura 17B y la Figura 17C muestran el análisis CL-EM de Y55 e Y185, respectivamente.

**Las Figuras 18A-18C** proporcionan un ensayo ejemplar para la "lectura completa" de una plantilla de "un solo clic". La Figura 18A muestra un esquema en donde el cebador marcado con FAM se hibrida con la plantilla biotinilada y se incuban con la polimerasa dependiente de la plantilla, de acuerdo con las condiciones recomendadas por el fabricante. Posteriormente, los complejos se incuban con cuentas de estreptavidina, se lavan, se eluyen con NaOH y luego se neutralizan. Después de la neutralización, las muestras se analizan mediante CL-EM. La Figura 18B y la Figura 18C muestran datos de CL-EM de la copia del fragmento de Klenow de las plantillas Y55 e Y185, respectivamente.

**Las Figuras 19A-19D** proporcionan la síntesis de la plantilla YDC de "doble clic" 5'-biotinilada y la YTC de "plantilla de triple clic" usando una etiqueta de alquino protegida con TIPS. Las Figuras 19A y 19B muestran esquemas ejemplares para esta síntesis. Las Figuras 19C y 19D muestran el análisis CL-EM de las plantillas YDC y YTC respectivamente.

**Las Figuras 20A-20C** proporcionan un ensayo de "lectura completa" de clic ejemplar utilizando plantillas de "doble clic" y "triple clic". La Figura 20A es un esquema, donde el cebador marcado con FAM se hibrida con la plantilla biotinilada y se incuban con el fragmento Klenow de la polimerasa de ADN I de *E. coli* de acuerdo con las condiciones de reacción recomendadas por el fabricante. Los complejos se incuban con cuentas de estreptavidina, se lavan, se eluyen con NaOH y se neutralizan. Después de la neutralización, las muestras se analizan mediante CL-EM. Las Figuras 20B y 20C muestran datos de CL-EM de la copia del fragmento de Klenow de las plantillas YDC y YTC, respectivamente.

**La Figura 21** es un gráfico que muestra la eficacia de la "lectura completa" del clic utilizando plantillas de "clic único", "doble clic" y "triple clic" en comparación con una plantilla de ADN de control "sin clic". Estos datos se obtuvieron utilizando el ensayo de "lectura completa" descrito en el presente documento, y los rendimientos se midieron mediante análisis CL-EM en comparación con un patrón interno.

**Las Figuras 22A-22C** proporcionan esquemas ejemplares de ligación química con química ortogonal. La Figura 22A es un esquema de la estrategia de ligación química para etiquetas que codifican ADN que (i) utiliza dos químicas ortogonales sucesivas para (ii) estrategias de lectura completa disponibles. Cada etiqueta contiene dos grupos reactivos ortogonales, indicados por símbolos diferentes para el extremo 5' y el extremo 3' de cada etiqueta. En cada ciclo sucesivo de ligación química, se utiliza una química ortogonal. Esta estrategia reduce la frecuencia de ocurrencia de eventos de error y también puede usarse sin la protección de los grupos terminales reactivos. La Figura 22B es un esquema de la polimerización dependiente de la plantilla "lectura completa" de una plantilla generada por la ligación química ortogonal de etiquetas de ADN ortogonales para generar ADNc a partir del cual se puede deducir la secuencia de las etiquetas. La Figura 22C es la misma que la Figura 22B pero incluye un cordal autocebante, que se puede convertir en bicatenario mediante digestión de restricción para facilitar la separación de la cadena durante la amplificación por PCR.

**La Figura 23** es un esquema de ejemplo que muestra la estrategia de ligación química para etiquetas que codifican ADN que utiliza dos químicas ortogonales sucesivas específicas. Cada etiqueta contiene grupos reactivos al clic y fosforotioato/yodoreactivos. Las etiquetas que llevan grupos reactivos ortogonales en sus extremos 3' y 5' no pueden polimerizar y tienen una frecuencia reducida de ocurrencia de eventos de error. Sin desear estar limitado, este enfoque puede eliminar la necesidad de la protección TIPS del 3'-alquino. En el ciclo A, la etiqueta 5'-yodo/3'-alquino se liga mediante ligación dependiente de la férula a la pieza de cabeza 3'-fosforotioato, dejando un alquino 3' reactivo para el siguiente ciclo de ligación química a una etiqueta de 5'-azido/3'-fosforotioato. Los ciclos de ligación ortogonal se pueden repetir tantas veces como se desee.

**Las Figuras 24A-24B** muestran la protección y el uso de grupos 3'-fosforotioato/5'-yodo en etiquetas de ADN. La Figura 24A muestra un esquema ejemplar para usar grupos protectores (PG) para estas etiquetas. La Figura 24B muestra un esquema de ejemplo para el uso de etiquetas 3'-fosforotioato/5'-yodo para ligar

químicamente la sucesión de etiquetas de ADN codificantes que codifican una biblioteca química instalada covalentemente en el extremo 5'.

**Las Figuras 25A-25B** muestran la protección y el uso de grupos 3'-fosforotioato en etiquetas de ADN. La Figura 25A muestra el esquema de protección de estos grupos. La Figura 25B muestra el esquema para el uso de etiquetas de 3'-fosforotioato/5'-azido y 3'-propargil/5'-yodo para ligar químicamente una sucesión de etiquetas de ADN codificantes ortogonales que codifican una biblioteca química instalada covalentemente sobre el extremo 5'.

## Descripción detallada

**[0064]** La invención presenta métodos como se define en las reivindicaciones, que comprende el uso de la ligación de una sola hebra para instalar etiquetas de oligonucleótidos en complejos de entidad-oligonucleótidos químicos. Este método se puede utilizar para crear diversas bibliotecas de entidades químicas seleccionables mediante el establecimiento de una relación codificada entre etiquetas particulares y reacciones químicas o bloques de construcción particulares. Para identificar una o más entidades químicas, las etiquetas de oligonucleótidos se pueden amplificar, clonar, secuenciar y correlacionar utilizando la relación establecida. En particular, se identificaron las condiciones de reacción que promueven la ligación de etiquetas de una sola hebra. Estas condiciones incluyen el uso de uno o más nucleótidos sustituidos en 2' (p. ej., nucleótidos de 2'-O-metilo o nucleótidos 2'-flúor) dentro de las etiquetas, el uso de etiquetas de longitud particular (p. ej., entre 5 y 15 nucleótidos), el uso de una o más enzimas (p. ej., ligasa ARN y/o ligasa ADN), y/o el uso de uno o más agentes durante la ligación (p. ej., polietilenglicol y/o un catión multivalente soluble, como  $\text{Co}(\text{NH}_3)_6\text{Cl}_3$ ). Estos métodos incluyen adicionalmente métodos para unir oligonucleótidos químicamente, de modo que la secuencia del producto oligonucleotídico unido puede utilizarse como plantilla para una reacción de polimerasa dependiente de la plantilla. Los métodos para crear y etiquetar bibliotecas de estos complejos se describen en detalle a continuación.

## Métodos para etiquetar bibliotecas codificadas

**[0065]** Esta invención presenta un método como se define en las reivindicaciones, para unir de manera operativa etiquetas oligonucleotídicas con entidades químicas, de manera que se puedan 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 la historia de una entidad química pueden inferirse de la secuencia de bases en el oligonucleótido. Usando este método, una biblioteca que incluye diversas entidades o miembros químicos (p. ej., pequeñas moléculas o péptidos) puede direccionarse con una secuencia de etiqueta particular.

**[0066]** Generalmente, estos métodos incluyen el uso de una pieza de cabeza, que tiene al menos un grupo funcional que puede ser elaborado químicamente y al menos un grupo funcional al que un oligonucleótido monocatenario puede unirse (o ligarse). La unión se puede efectuar por cualquier medio útil, como mediante unión enzimática (p. ej., unión con una o más de una ligasa ARN y/o una ligasa ADN) o mediante unión química (p. ej., mediante una reacción de sustitución entre dos grupos funcionales, como como nucleófilo y grupo saliente), en donde el paso (iii) comprende una ligasa ARN y/o ligasa ADN para unir dicha pieza de cabeza a dicho primer bloque de construcción, proporcionando así una biblioteca etiquetada.

**[0067]** Para crear numerosas entidades químicas dentro de la biblioteca, una solución que contiene la pieza de cabeza se puede dividir en múltiples alícuotas y luego se coloca en una multiplicidad de compartimentos físicamente separados, tales como los pocillos de una placa de múltiples pocillos. Generalmente, este es el paso "dividido". Dentro de cada compartimento o pocillo, se realizan sucesivas etapas de ligación y reacción química con una etiqueta monocatenaria dentro de cada alícuota. Se registra la relación entre las condiciones de la reacción química y la secuencia de la etiqueta monocatenaria. Los pasos de reacción y ligación pueden realizarse en cualquier orden. Luego, las alícuotas reaccionadas y ligadas se combinan o "juntan" y, opcionalmente, se puede realizar la purificación en este punto. Estos pasos de división y agrupación se pueden repetir opcionalmente.

**[0068]** A continuación, la biblioteca puede ser probada y/o seleccionada para una característica o función particular, como se describe en el presente documento. Por ejemplo, la mezcla de entidades químicas etiquetadas se puede separar en al menos dos poblaciones, donde la primera población se une a una diana biológica particular y la segunda población no. A continuación, la primera población puede capturarse selectivamente (p. ej., eluyendo en una columna que proporciona la diana de interés o incubando la alícuota con la diana de interés) y, opcionalmente, analizada o probada adicionalmente, como con lavado opcional, purificación, selección negativa, selección positiva, o pasos de separación.

**[0069]** Por último, las historias químicas de uno o más miembros (o entidades químicas) dentro de la población seleccionada se pueden determinar por 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 (p. ej., una mayor tendencia a unirse a la proteína diana y de ese modo provocar un efecto terapéutico). Para pruebas y optimización adicionales, los compuestos terapéuticos candidatos pueden prepararse luego sintetizando los miembros de la biblioteca identificados con o sin sus etiquetas oligonucleotídicas asociadas.

**[0070]** Las Figuras 1-3 proporcionan varios ejemplos de métodos para etiquetar las bibliotecas usando la ligación de una sola hebra con una pieza de cabeza, en donde las etiquetas se pueden ligar en el extremo 5' o el extremo 3' de la pieza de cabeza. Para controlar el orden en donde se ligan las etiquetas y reducir las reacciones secundarias, estos métodos aseguran que sólo estén presentes durante la ligación un extremo 5' reactivo y un extremo 3' reactivo. Además, estos métodos ejemplares utilizan nucleótidos sustituidos en 2' (p. ej., nucleótidos 2'-desoxi/2'-O-metilo mixtos) en las etiquetas, y estas etiquetas actúan como plantillas para una polimerasa dependiente de ADN o ARN capaz de polimerizar nucleótidos de una manera dependiente de la plantilla. Sin desear estar limitado por la teoría, el uso de uno o más nucleótidos sustituidos en 2' (p. ej., nucleótidos 2'-O-metilo y/o nucleótidos 2'-fluoro) dentro de una etiqueta podría promover la ligación por ligasa ARN de manera más estrecha similar al ARN, conservando al mismo tiempo la robustez física y química del medio de grabación, así como la capacidad de extraer información de secuencia mediante polimerización dependiente de la plantilla.

**[0071]** La Figura 1 proporciona un método ejemplar para reducir reacciones secundarias, donde el complejo ligado y las etiquetas son diseñadas para evitar reacciones no deseadas entre grupos 3'-OH reactivo y 5'-monofosfato ("5'-P"). En particular, este esquema describe el enfoque del ciclo de fosforilación-ligación. Durante la ligación, solo están disponibles un grupo 3'-OH (en la etiqueta) y un grupo 5'-P (en la pieza de cabeza) y, por lo tanto, solo es posible un evento de ligación. Después de los pasos de ligación y purificación, se forma un grupo 5'-OH en el complejo, y este grupo se puede convertir en un 5'-P para añadir etiquetas oligonucleotídicas posteriores. El extremo 3' del complejo está bloqueado por X, que puede ser un grupo protector o un componente de una entidad química (p. ej., que incluye opcionalmente un enlazador que actúa como espaciador entre la entidad química y la pieza de cabeza).

**[0072]** Como se muestra en la Figura 1, el método de ejemplo incluye la ligación de la etiqueta de bloque de construcción 1 ("etiqueta 1") para el extremo 5' de la pieza de cabeza, creando de este modo un complejo, y realizar ligaciones sucesivas para el extremo 5' del complejo. El extremo 5' reactivo es un grupo fosfato en el complejo y el extremo 3' reactivo es un grupo hidroxilo en las etiquetas. Después de la adición de cada etiqueta, el complejo ligado se separa de la pieza de cabeza y las etiquetas que no han reaccionado y no ligado y de otros reactivos (p. ej., fosfato, cobalto u otros reactivos presentes durante el paso de ligación). La separación puede realizarse mediante cualquier método útil (p. ej., mediante separación cromatográfica o electroforética de productos ligados y no ligados o mediante precipitación de un reactivo). Luego, el complejo ligado se expone a un agente (p. ej., una polinucleótido quinasa o un agente de fosforilación químico) para formar un grupo fosfato en el extremo 5' del complejo. Los pasos de separación y fosforilación se pueden realizar en cualquier orden. En particular, si se usa una quinasa en el paso de fosforilación, la quinasa debe inactivarse o eliminarse antes de la adición de las etiquetas posteriores que también pueden contener un grupo 5'-OH, o cualquier reactivo que pueda inhibir la quinasa debe eliminarse de la mezcla de reacción antes del paso de fosforilación.

**[0073]** En otra realización, el método incluye la unión de etiquetas sucesivas desde el extremo 3' del complejo ligado anterior. En este método, el complejo ligado carece de un grupo 3'-OH reactivo inmediatamente después del paso de ligación, pero contiene un grupo que puede convertirse en un grupo 3'-OH (p. ej., mediante la liberación de un grupo protector). La Figura 2A proporciona un esquema que muestra un método ejemplar para marcar el extremo 3' de un complejo, y la Figura 2B proporciona un esquema de reacción ejemplar para un extremo 3' protegido que contiene un grupo 3'-OH convertible tras la liberación del grupo protector 3' enlazado. Como se muestra en la Figura 2A, la etiqueta 1 del bloque de construcción ("etiqueta 1") tiene un grupo protegido en 3'. En el primer paso, el método ejemplar incluye la ligación de la etiqueta al extremo 3' de la pieza de cabeza, creando así un complejo. Se realizan ligaciones sucesivas al extremo 3' del complejo. El extremo 5' reactivo es un grupo fosfato en la etiqueta y el extremo 3' reactivo es un grupo hidroxilo en el complejo. Después de la adición de cada etiqueta, el complejo ligado se desprotege (p. ej., mediante la adición de un agente hidrolizante) para liberar el grupo protector 3'.

**[0074]** En aún otra realización, el método incluye las etiquetas sucesivas de unión mediante el uso de un oligonucleótido 5' preadenilado (5'-App) y una ligasa (p. ej., ligasa T4 ARN). En presencia de ATP, la ligasa T4 ARN utilizará el cofactor de ATP para formar un intermedio adenilado antes de la ligación. En ausencia de ATP, la ligasa T4 ARN solo ligará oligonucleótidos preadenilados y no se producirán posibles reacciones secundarias con oligonucleótidos 5'-P. Por lo tanto, la ligación monocatenaria con reacciones secundarias reducidas se puede realizar con un oligonucleótido 5'-App sintetizado químicamente en presencia de una etiqueta 5'-monofosforilada, donde el oligonucleótido 5'-App se puede ligar a una pieza de cabeza antes del etiquetado o para un complejo formado después de múltiples rondas de marcado.

**[0075]** La figura 3 proporciona un esquema que muestra un método ejemplar para etiquetar el extremo 5' de una pieza de cabeza preadenilada. La adenilación del nucleótido donante en el grupo 5'-fosfato es el primer paso en la reacción de ligación, y esta reacción generalmente requiere una molécula de ATP. En el segundo paso, el grupo 3'-OH del oligonucleótido aceptor reacciona con el donante adenilado y forma un enlace diéster entre dos oligonucleótidos, liberando así una molécula de AMP. El grupo 5'-fosfato químicamente adenilado del oligonucleótido donante imita un producto de la primera etapa de la reacción de ligación y puede ligarse al segundo oligonucleótido en ausencia de ATP. En el siguiente esquema, se liga una pieza de cabeza 5'-App al grupo 3'-OH de una etiqueta oligonucleotídica fosforilada en 5' (etiquetada "Etiqueta A"). Debido a la presencia del extremo 5' adenilado del oligonucleótido, la ligación puede ocurrir en ausencia de ATP. En estas condiciones, el grupo 5'-fosfato de la etiqueta A no sirve como

donante de ligación. La etiqueta B del bloque de construcción se puede ligar proporcionando un nucleótido que tiene un extremo 3'-OH (marcado como "Etiqueta B") en presencia de ATP, y se pueden incluir etiquetas adicionales (etiquetadas como "Etiqueta C").

**[0076]** En la Figura 3, el extremo 3' de la pieza de cabeza puede ser bloqueado con cualquier grupo protector (p. ej., un grupo protector irreversible, tal como DDN, o un grupo protector reversible). En el primer paso, el método incluye la ligación de la etiqueta al extremo 5' de la pieza de cabeza en ausencia de ATP, creando así un complejo. Se realizan ligaciones sucesivas al extremo 5' del complejo en presencia de ATP. Este método puede modificarse para realizar ligaciones sucesivas al extremo 3' de un complejo. Por ejemplo, el método puede incluir el uso de una etiqueta preadenilada en 5' y una pieza de cabeza que tiene un extremo 3'-OH reactivo. Este método puede además requerir bloquear el extremo 3' de la etiqueta para evitar reacciones cruzadas entre las etiquetas, tales como el método descrito anteriormente y en la Figura 2.

**[0077]** El método general proporcionado en la Figura 3 puede ser modificado mediante la sustitución de la imprimación con una pieza de cabeza. En este caso, la pieza de cabeza tiene que ser adenilada químicamente en el extremo 5' y la etiqueta A se fosforila en el extremo 5'. La unión de esta etiqueta A fosforilada a la pieza de cabeza adenilada se produce en las mismas condiciones estándar, descritas aquí, pero omitiendo el ATP. Usando esta condición de ligación, se puede prevenir la ligación del extremo 5' fosforilado. En el siguiente paso, la ligación de la etiqueta B requiere que esta etiqueta tenga un grupo hidroxilo libre en el extremo 5' (es decir, no fosforilado). Se pueden realizar reacciones de ligación sucesivas en presencia de ATP, seguido de fosforilación del extremo 5' del oligonucleótido resultante si se desea una extensión adicional de las etiquetas (p. ej., etiqueta C en la Figura 3).

**[0078]** Los métodos descritos en este documento pueden incluir cualquier número de pasos opcionales para diversificar la biblioteca o interrogar a los miembros de la biblioteca. Para cualquier método de marcado descrito en este documento (p. ej., como en las Figuras 1-3), se puede añadir un número "n" sucesivo de etiquetas con un número "n" adicional de etapas de ligación, separación y/o fosforilación. Los pasos opcionales ejemplares incluyen la restricción de los miembros de la biblioteca usando una o más endonucleasas de restricción; ligación de una o más secuencias adaptadoras a uno o ambos extremos de la biblioteca, p. ej., tal como una o más secuencias adaptadoras para proporcionar una secuencia de cebado para amplificación y secuenciación o para proporcionar un marcador, tal como biotina, para inmovilizar 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 plantilla; amplificación de las etiquetas ensambladas en el complejo usando, p. ej., PCR; generación de aislados clonales de una o más poblaciones de etiquetas ensambladas en el complejo, p. ej., 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, p. ej., 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, p. ej., usando aislados clonales como plantillas para la polimerización dependiente de plantilla con nucleótidos marcados con fluorescencia. En el presente documento se describen métodos adicionales para amplificar y secuenciar las etiquetas oligonucleotídicas.

**[0079]** Estos métodos se pueden utilizar para identificar y descubrir cualquier número de entidades químicas con una característica particular o función, p. ej., en una etapa de selección. La característica o función deseada puede usarse como 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 pequeño miembro de biblioteca similar a un fármaco 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 genere 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 componentes básicos se eligen para crear una biblioteca que tenga grupos estructurales comúnmente presentes en los inhibidores de quinasas. En cualquiera de estos casos, las etiquetas codifican la historia química del miembro de la biblioteca y, en cada caso, una colección de posibilidades químicas puede estar representada por cualquier combinación de etiquetas particular.

**[0080]** En una realización, la biblioteca de entidades químicas, o una porción de la misma, se pone en contacto con una diana biológica en condiciones adecuadas para al menos un miembro de la biblioteca que se une a la diana, seguido de la eliminación de miembros de la biblioteca que no se unen a la diana y analizan una o más etiquetas oligonucleotídicas asociadas con ellas. Este método puede incluir opcionalmente amplificar las etiquetas mediante métodos conocidos en la técnica. Los ejemplos de dianas biológicas incluyen enzimas (p. ej., quinasas, fosfatasa, metilasas, demetilasas, proteasas y enzimas de reparación de ADN), proteínas implicadas en interacciones proteína:proteína (p. ej., ligandos para receptores), dianas de receptor (p. ej., GPCR y RTK), canales de iones, bacterias, virus, parásitos, ADN, ARN, priones y carbohidratos.

**[0081]** En otra realización, las entidades químicas que se unen a una diana no se someten a amplificación sino que se analizan directamente. Los métodos de análisis ejemplares incluyen análisis de micromatrices, incluyendo análisis

de cristales fotónicos de resonancia evanescente; métodos basados en cuentas para deconvolución de etiquetas (p. ej., usando etiquetas his); análisis de biosensor de cristal fotónico sin etiquetas (p. ej., un lector BIND® de SRU Biosystems, Inc., Woburn, MA); o enfoques basados en hibridación (p. ej., mediante el uso de matrices de oligonucleótidos inmovilizados complementarios a las secuencias presentes en la biblioteca de etiquetas).

**[0082]** Además, los pares químicamente co-reactivos (o grupos funcionales) se pueden incluir fácilmente en los esquemas de oligonucleótidos en fase sólida síntesis y apoyarán la ligación química eficiente de 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 los pasos de unión descritos en el presente documento para marcar bibliotecas codificadas puede modificarse para incluir una o más técnicas de ligación enzimática y/o ligación química. Las técnicas de ligación ejemplares incluyen ligación enzimática, tal como el uso de una o más ligasas de ARN y/o ligasas de ADN; y ligación química, tal como el uso de pares químicamente reactivos (p. ej., un par que incluye grupos funcionales alquínico y azido opcionalmente sustituidos).

**[0083]** Además, una o más bibliotecas se pueden combinar en una etapa de división y mezcla. Para permitir la mezcla de dos o más bibliotecas, el miembro de la biblioteca puede contener una o más secuencias de identificación de la biblioteca, como en una etiqueta de identificación de la biblioteca, en una etiqueta de bloque de construcción ligada o como parte de la secuencia de la pieza de cabeza, como se describe aquí.

#### *Métodos que tienen masa reducida*

**[0084]** La mayor parte de la motivación para las estrategias de codificación de cadena simple surge de la masa reducida de una etiqueta de una sola cadena, en comparación con una etiqueta de doble cadena. La masa reducida potencialmente confiere varios beneficios que incluyen mayor solubilidad, menor costo, mayor reactividad, mayor accesibilidad a la diana, menor radio hidrodinámico, mayor precisión de las evaluaciones analíticas, etc. Además de utilizar una metodología de marcado de una sola hebra, reducciones adicionales de masa se pueden lograr al incluir 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 pieza de cabeza de codificación, uno o más miembros de una biblioteca que carecen de una región de unión a cebador y/o una región constante, uno o más miembros de una biblioteca que tienen una región constante reducida, o cualquier otra metodología descrita en el presente documento.

**[0085]** Para reducir al mínimo la masa de los miembros de la biblioteca, la longitud de una o más etiquetas de bloque de construcción puede ser reducida, tal como a una longitud que es tan corta como sea posible para codificar cada tamaño de división. En particular, las etiquetas pueden tener menos de 20 nucleótidos (p. ej., 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 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 (p. ej., aproximadamente 10 nucleótidos o más cortas) para la ligación de etiquetas.

**[0086]** Estrategias de masa constante también se pueden utilizar, lo que podría ayudar en el análisis durante la síntesis de la biblioteca. Además, los conjuntos de etiquetas de masa constante podrían permitir el reconocimiento de todas las ocurrencias de errores individuales (p. ej., errores que surgen de una mala lectura de una secuencia o de la ligación 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 monocatenarias de masa constante y la capacidad de codificación (p. ej., longitudes mínimas para admitir tamaños de división de bloques de construcción específicos o identidades de biblioteca, etc.) se describe a continuación en la Tabla 1. En consecuencia, el empleo de conjuntos de etiquetas de masa constantes 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 nº 1	Base nº 2	Base nº 3	Base nº 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

(Continuación)

Longitud	Base nº 1	Base nº 2	Base nº 3	Base nº 4	Combinaciones
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

**[0087]** Para minimizar la masa en la biblioteca, la pieza de cabeza se puede utilizar no sólo para enlazar el resto químico y una etiqueta sino también para codificar la identidad de una biblioteca particular o para un paso particular. Por ejemplo, la pieza de cabeza puede codificar información, p. ej., una pluralidad de piezas de cabeza que codifican la(s) primera(s) división(es) o la identidad de la biblioteca, p. ej., utilizando una secuencia particular relacionada con una biblioteca específica.

**[0088]** Además, las regiones de unión (p. ej., constante) de cebadores a partir de la biblioteca de entidades químicas codificadas-ADN pueden ser excluidas durante el paso de selección. Entonces, estas regiones pueden añadirse después de la selección mediante, p. ej., ligación monocatenaria. Una estrategia ejemplar incluiría proporcionar una entidad química en el extremo 5' de un oligonucleótido codificante, seleccionar una entidad química particular en función de cualquier característica o función particular útil y ligar un oligonucleótido de cordal al extremo 3' del oligonucleótido codificante que incluye una secuencia de unión al cebador y puede contener opcionalmente una o más etiquetas, p. ej., una etiqueta de "uso", una etiqueta de "origen", etc., como se describe en el presente documento. Esta secuencia de unión del cebador podría usarse luego para iniciar la polimerización dependiente de la plantilla para generar ADNc (o ARNc) que es complementario al miembro de la biblioteca seleccionado. El cADN o cARN luego se ligaría en su extremo 3' a un oligonucleótido que contiene una secuencia de unión de cebador y, ahora que la información codificante está flanqueada en ambos lados por secuencias de unión de cebador, el oligonucleótido se puede secuenciar y/o amplificar usando enfoques establecidos, como cualquiera de los descritos en este documento.

**[0089]** La masa se puede minimizar además omitiendo o reduciendo el tamaño de una o más secuencias constantes que separan las etiquetas de codificación. La ligación monocatenaria no requiere ninguna relación complementaria entre los extremos a ligar o entre estos extremos y una férula. 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 sintáctico de etiquetas u otros procesos de deconvolución *in silico*.

#### Etiquetas de oligonucleótidos

**[0090]** Las etiquetas de oligonucleótidos descritas en este documento (p. ej., una etiqueta de bloque de construcción o una parte de una pieza de cabeza) se puede utilizar para codificar cualquier información útil, tal como una molécula, una parte de una entidad química, la adición de un componente (p. ej., un andamio o un bloque de construcción), una pieza de cabeza en la biblioteca, la identidad de la biblioteca, el uso de uno o más miembros de la biblioteca (p. ej., el uso de los miembros en una alícuota de una biblioteca) y/o el origen de un miembro de la biblioteca (p. ej., mediante el uso de una secuencia de origen).

**[0091]** Cualquier secuencia en un oligonucleótido puede usarse para codificar cualquier información. Por tanto, una secuencia de oligonucleótidos puede servir para más de un propósito, como codificar dos o más tipos de información o proporcionar un oligonucleótido de partida que también codifica uno o más tipos de información. Por ejemplo, la primera etiqueta de bloque de construcción 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 pieza de cabeza para proporcionar un oligonucleótido de partida que une operativamente una entidad química a una etiqueta de bloque de construcción, donde la pieza de cabeza incluye adicionalmente una secuencia que codifica la identidad de la biblioteca (es decir, la secuencia de identificación de la biblioteca). Por consiguiente, cualquier información descrita en el presente documento puede codificarse en etiquetas oligonucleotídicas separadas o puede combinarse y codificarse en la misma secuencia oligonucleotídica (p. ej., una etiqueta oligonucleotídica, como una etiqueta de bloque de construcción o una pieza de cabeza).

**[0092]** Una secuencia de bloque de construcción codifica la identidad de un bloque de construcción y/o el tipo de reacción de unión realizado con un bloque de construcción. Esta secuencia de bloques de construcción está incluida en una etiqueta de bloques de construcción, donde la etiqueta puede incluir opcionalmente uno o más tipos de secuencia que se describen a continuación (p. ej., una secuencia de identificación de biblioteca, una secuencia de uso y/o una secuencia de origen).

**[0093]** Una biblioteca de identificación codifica secuencias para la identidad de una biblioteca en particular. Para permitir la mezcla de dos o más bibliotecas, un miembro de la biblioteca puede contener una o más secuencias de identificación de la biblioteca, como en una etiqueta de identificación de la biblioteca (es decir, un oligonucleótido que incluye una secuencia de identificación de la biblioteca), en una etiqueta de bloque de construcción ligado, en una parte de la secuencia de la pieza de cabeza o en una secuencia del cordal. 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 la información del historial químico (síntesis). En consecuencia, estas secuencias de la biblioteca de identificación permiten la mezcla de dos o más bibliotecas juntas para la selección, amplificación, purificación, secuenciación, etc.

**[0094]** Una secuencia de uso codifica la historia (es decir, el uso) de uno o más miembros de la biblioteca en una alícuota individual de una biblioteca. Por ejemplo, se pueden tratar alícuotas separadas 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, por lo tanto, permitir la mezcla de alícuotas de la misma biblioteca con diferentes historias (usos) (p. ej., distintos experimentos de selección) para los fines de la mezcla de muestras juntas para selección, amplificación, purificación, secuenciación, etc. Estas secuencias de uso se pueden incluir en una pieza de cabeza, un cordal, una etiqueta de bloque de construcción, una etiqueta de uso (es decir, un oligonucleótido que incluye una secuencia de uso) o cualquier otra etiqueta descrita en el presente documento (p. ej., una etiqueta de identificación de biblioteca o una etiqueta de origen).

**[0095]** Una secuencia de origen es una secuencia de oligonucleótido degenerada (aleatoria) de cualquier longitud útil (p. ej., aproximadamente seis oligonucleótidos) que codifica para el origen del elemento de biblioteca. Esta secuencia sirve para subdividir estocásticamente los miembros de la biblioteca que son idénticos en todos los aspectos en entidades distinguibles por la información de la secuencia, de modo que las observaciones de los productos de amplificación derivados de plantillas progenitoras únicas (p. ej., miembros de la biblioteca seleccionados) se pueden distinguir de las observaciones de múltiples productos de amplificación derivados de la misma plantilla progenitora (p. ej., un miembro de la biblioteca seleccionado). Por ejemplo, después de la formación de la biblioteca y antes del paso de selección, cada miembro de la biblioteca puede incluir una secuencia de origen diferente, como en una etiqueta de origen. Después de la selección, los miembros de la biblioteca seleccionados se pueden amplificar para producir productos de amplificación, y la parte del miembro de la biblioteca que se espera que incluya la secuencia de origen (p. ej., en la etiqueta de origen) se puede observar y comparar con la secuencia de origen en cada uno de los otros miembros de biblioteca. Como las secuencias de origen están 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 una fuente de error, como un error de amplificación o un error de ciclación en la secuencia que produce secuencias repetidas, y el punto de partida o la fuente de estos errores se puede rastrear mediante la observación de la secuencia de origen en cada paso (p. ej., en cada paso de selección o paso de amplificación) de utilización de la biblioteca. Estas secuencias de origen se pueden incluir en una pieza de cabeza, un cordal, una etiqueta de bloque de construcción, una etiqueta de origen (es decir, un oligonucleótido que incluye una secuencia de origen) o cualquier otra etiqueta descrita en este documento (p. ej., una etiqueta de identificación de biblioteca o una etiqueta de uso).

**[0096]** Cualquiera de los tipos de secuencias descritas en este documento se pueden incluir en la pieza de cabeza. Por ejemplo, la pieza de cabeza puede incluir una o más de una secuencia de bloques de construcción, una secuencia de identificación de biblioteca, una secuencia de uso o una secuencia de origen.

**[0097]** Cualquiera de estas secuencias descritas en el presente documento se pueden incluir en un tubo de desagüe. Por ejemplo, el cordal puede incluir una o más de una secuencia de identificación de biblioteca, una secuencia de uso o una secuencia de origen.

**[0098]** Estas secuencias pueden incluir cualquier modificación descrita en el presente documento para los oligonucleótidos, tales como una o más modificaciones que promueven la solubilidad en disolventes orgánicos (p. ej., cualquiera de los descritos en el presente documento, tales como por la pieza de cabeza), que proporcionan un análogo del enlace fosfodiéster natural (p. ej., un análogo de fosforotioato), o que proporcionan uno o más oligonucleótidos no naturales (p. ej., nucleótidos sustituidos en 2', tales como nucleótidos 2'-O-metilados y nucleótidos 2'-fluorados, o cualquiera de los descritos en el presente documento).

**[0099]** Estas secuencias pueden incluir cualesquiera características descritas en este documento para los oligonucleótidos. Por ejemplo, estas secuencias pueden incluirse en una etiqueta de menos de 20 nucleótidos (p. ej., como se describe en este documento). En otros ejemplos, las etiquetas que incluyen una o más de estas secuencias tienen aproximadamente la misma masa (p. ej., cada etiqueta tiene una masa que es aproximadamente +/- 10% de la masa promedio entre dos o más etiquetas); carecen de una región de unión a cebadores (p. ej., constante); carece de una región constante; o tienen una región constante de longitud reducida (p. ej., una longitud de menos de 30 nucleótidos, menos de 25 nucleótidos, menos de 20 nucleótidos, 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 de 10 nucleótidos, menos de 9 nucleótidos, menos de 8 nucleótidos o menos de 7 nucleótidos).

**[0100]** Las estrategias de secuenciación para bibliotecas y oligonucleótidos de esta longitud pueden incluir opcionalmente estrategias de concatenación o de cadena para aumentar la fidelidad de lectura o la profundidad de secuenciación, respectivamente. En particular, la selección de bibliotecas codificadas que carecen de regiones de unión a cebadores se ha descrito en la bibliografía 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 (p. ej., después de un paso de selección) para incluir una primera secuencia adaptadora en el extremo 5' del complejo y una segunda secuencia adaptadora en el extremo 3' del complejo, donde la primera secuencia es sustancialmente complementaria a la segunda secuencia y da como resultado la formación de un dúplex. Para mejorar aún más el rendimiento, se agregan dos nucleótidos colgantes fijos (p. ej., 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).

#### **Pieza de cabeza**

**[0101]** En la biblioteca, la pieza de cabeza enlaza operativamente cada entidad química a su etiqueta oligonucleotídica codificante. Generalmente, la pieza de cabeza es un oligonucleótido de partida que tiene dos grupos funcionales que pueden derivatizarse adicionalmente, donde el primer grupo funcional une operativamente la entidad química (o un componente de la misma) a la pieza de cabeza y el segundo grupo funcional enlaza operativamente una o más etiquetas a la fuente. Opcionalmente, se puede utilizar un enlazador como espaciador entre la pieza de cabeza y la entidad química.

**[0102]** Los grupos funcionales de la pieza de cabeza pueden utilizarse para formar un enlace covalente con un componente de la entidad química y un enlace covalente con una etiqueta. El componente puede ser cualquier parte de la molécula pequeña, como un andamio con nodos de diversidad o un bloque de construcción. Alternativamente, la pieza de cabeza puede derivatizarse para proporcionar un enlazador (es decir, un espaciador que separa la pieza de cabeza de la molécula pequeña que se formará en la biblioteca) que termina en un grupo funcional (p. ej., un grupo hidroxilo, amina, carboxilo, sulfhidrilo, alquínilo, azido o fosfato), que se utiliza para formar el enlace covalente con un componente de la entidad química. El enlazador se puede conectar al extremo 5', en una de las posiciones internas, o al extremo 3' de la pieza de cabeza. Cuando el enlazador se une a una de las posiciones internas, el enlazador puede enlazarse operativamente a una base derivatizada (p. ej., la posición C5 de la uridina) o colocarse internamente dentro del oligonucleótido usando técnicas estándar conocidas en la técnica. En este documento se describen ejemplos de enlazadores. La pieza de cabeza puede tener cualquier estructura útil. La pieza de cabeza puede tener, p. ej., 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 pieza de cabeza puede ser monocatenaria o bicatenaria y puede consistir en nucleótidos naturales o modificados, como se describe en el presente documento. En las Figuras 4A-4D se describen realizaciones ejemplares particulares de la pieza de cabeza. Por ejemplo, el resto químico se puede unir operativamente al extremo 3' (Figura 4A) o al extremo 5' (Figura 4B) de la pieza de cabeza. En realizaciones particulares, la pieza de cabeza incluye una estructura de horquilla formada por bases complementarias dentro de la secuencia. Por ejemplo, el resto químico se puede unir operativamente a la posición interna (Figura 4C), al extremo 3' (Figura 4D) o al extremo 5' de la pieza de cabeza.

**[0103]** Generalmente, la pieza de cabeza incluye una secuencia no complementaria en el extremo 5' o 3' que permite unir una etiqueta oligonucleotídica por polimerización, ligación enzimática o reacción química. En la Figura 4E, la pieza de cabeza ejemplar permite la ligación de etiquetas oligonucleotídicas (marcadas 1-4), y el método incluye etapas de purificación y fosforilación. Después de la adición de la etiqueta 4, se puede agregar una secuencia adaptadora adicional al extremo 5' de la etiqueta 4. Ejemplos de secuencias adaptadoras incluyen una secuencia de unión de cebador o una secuencia que tiene una etiqueta (p. ej., biotina). En los casos en los que se utilizan muchos bloques de construcción y etiquetas correspondientes (p. ej., 100 etiquetas), se puede emplear una estrategia de mezcla y división durante el paso 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 resultantes pueden amplificarse mediante PCR después de la selección de entidades de unión frente a una diana o diana de interés.

**[0104]** La pieza de cabeza o el complejo puede incluir opcionalmente uno o más cebadores de secuencias de unión. Por ejemplo, la pieza de cabeza tiene una secuencia en la región del bucle de la horquilla que sirve como una 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 (p. ej., que puede incluir regiones de identificación flanqueantes) que para una secuencia en la pieza de cabeza. En otras realizaciones, el complejo incluye dos secuencias de unión de cebadores (p. ej., para permitir una reacción de PCR) en cualquier lado de una o más etiquetas que codifican uno o más bloques de construcción. Alternativamente, la pieza de cabeza puede contener una secuencia de unión de cebador en el extremo 5' o 3'. En otras realizaciones, la pieza de cabeza es una horquilla y la región del bucle forma un sitio de unión del cebador o el sitio de enlace del cebador se introduce mediante la hibridación de un oligonucleótido a la pieza de cabeza en el lado 3' del bucle. Un oligonucleótido cebador, que contiene una región homóloga al extremo 3' de la pieza de cabeza y que lleva una región de unión al cebador en su extremo 5' (p. ej., para permitir una reacción de PCR) se puede hibridar con la pieza de cabeza 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 aleatorizados, p. ej., de 2 a 16 nucleótidos de longitud, que se incluye para el análisis bioinformático.

**[0105]** La pieza de cabeza puede incluir opcionalmente una estructura de horquilla, donde esta estructura se puede conseguir por cualquier método útil. Por ejemplo, la pieza de cabeza puede incluir bases complementarias que forman parejas de emparejamiento de bases intermoleculares, como por emparejamiento de bases de ADN Watson-Crick (p. ej., adenina-timina y guanina-citosina) y/o por emparejamiento de bases oscilantes (p. ej., guanina-uracilo, inosina-uracilo, inosina-adenina e inosina-citosina). En otro ejemplo, la pieza de cabeza puede incluir nucleótidos modificados o sustituidos que pueden formar formaciones dúplex de mayor afinidad en comparación con los nucleótidos no modificados, siendo estos nucleótidos modificados o sustituidos conocidos en la técnica. En otro ejemplo más, la pieza de cabeza incluye una o más bases reticuladas para formar la estructura de horquilla. Por ejemplo, las bases dentro de una sola hebra o las bases en diferentes hebras dobles pueden reticularse, p. ej., usando psoraleno.

**[0106]** La pieza de cabeza o complejo pueden incluir opcionalmente una o más etiquetas que permiten la detección. Por ejemplo, la pieza de cabeza, una o más etiquetas de oligonucleótidos y/o una o más secuencias de cebadores pueden incluir un isótopo, un agente de radiografía, un marcador, un trazador, una etiqueta fluorescente (p. ej., rodamina o fluoresceína), una etiqueta quimioluminiscente, un punto cuántico y una molécula informadora (p. ej., biotina o una etiqueta his).

**[0107]** En otras realizaciones, la pieza de cabeza o etiqueta se puede modificar para soportar la solubilidad en condiciones semi, reducidas o no acuosas (p. ej., orgánicas). Las bases de nucleótidos de la pieza de cabeza o etiqueta pueden volverse más hidrófobas modificando, p. ej., las posiciones C5 de las bases T o C con cadenas alifáticas sin alterar significativamente su capacidad de enlace de hidrógeno con sus bases complementarias. Ejemplos de nucleótidos modificados o sustituidos son 5'-dimetoxitritil-N4-diisobutilaminometiliden-5-(1-propinilo)-2'-desoxicitidina, 3'-[(2-cianoetilo)-(N,N-diisopropilo)]-fosforamidita; 5'-dimetoxitritilo-5-(1-propinilo)-2'-desoxiuridina, 3'-[(2-cianoetilo)-(N,N-diisopropilo)]-fosforamidita; 5'-dimetoxitritilo-5-fluoro-2'-desoxiuridina, 3'-[(2-cianoetilo)-(N,N-diisopropilo)]-fosforamidita; y 5'-dimetoxitritilo-5-(piren-1-ilo-etinilo)-2'-desoxiuridina, o 3'-[(2-cianoetilo)-(N,N-diisopropilo)]-fosforamidita.

**[0108]** Además, el oligonucleótido de pieza de cabeza se puede intercalar con modificaciones que promueven la solubilidad en disolventes orgánicos. Por ejemplo, la fosforamidita de azobenceno puede introducir un resto hidrófobo en el diseño de la pieza de cabeza. Tales inserciones de amiditas hidrófobas en la pieza de cabeza pueden ocurrir en cualquier parte de la molécula. Sin embargo, la inserción no puede interferir con el etiquetado posterior utilizando etiquetas de ADN adicionales durante la síntesis de la biblioteca o la posterior PCR una vez que se completa la selección o el análisis de microarrays, si se usa para la deconvolución de etiquetas. Tales adiciones al diseño de la pieza de cabeza descrita en este documento harían que la pieza de cabeza fuera soluble en, p. ej., 15%, 25%, 30%, 50%, 75%, 90%, 95%, 98%, 99% o 100% de disolvente orgánico. Por tanto, la adición de residuos hidrófobos en el diseño de la pieza de cabeza permite una solubilidad mejorada en condiciones semi o no acuosas (p. ej., orgánicas), mientras que hace que la pieza de cabeza sea competente para el marcado de oligonucleótidos. Además, las etiquetas de ADN que se introducen posteriormente en la biblioteca también se pueden modificar en la posición C5 de las bases T o C de modo que también hagan que la biblioteca sea más hidrófoba y soluble en disolventes orgánicos para los pasos posteriores de la síntesis de la biblioteca.

**[0109]** En realizaciones particulares, la pieza de cabeza y la primera etiqueta de bloque de construcción pueden ser la misma entidad, es decir, se puede construir una pluralidad de entidades de etiqueta de pieza de cabeza que comparten partes comunes (p. ej., una región de unión de cebador) y todas difieren en otra parte (p. ej., región de codificación). Estos pueden utilizarse en el paso de "división" y agruparse después de que se haya producido el evento que están codificando.

**[0110]** En realizaciones particulares, la pieza de cabeza puede codificar la información, p. ej., mediante la inclusión de una secuencia que codifica el primer paso de división(es) o una secuencia que codifica la identidad de la biblioteca, tales como mediante el uso de una secuencia particular relacionada a una biblioteca específica.

## **Ligación enzimática y ligación química técnica**

**[0111]** Diversas técnicas de ligación pueden utilizarse para añadir andamios, bloques de construcción, enlazadores, etiquetas de bloque de construcción, y/o la pieza de cabeza para producir un complejo. Por consiguiente, cualquiera de los pasos de unión descritos en este documento puede incluir cualquier técnica de ligación útil, tal como ligación enzimática y/o ligación química. Estos pasos de unión pueden incluir la adición de una o más etiquetas de bloques de construcción a la pieza de cabeza o complejo; la adición de un enlazador a la pieza de cabeza; y la adición de uno o más andamios o bloques de construcción a la pieza de cabeza o complejo. En realizaciones particulares, las técnicas de ligación utilizadas para cualquier oligonucleótido proporcionan un producto resultante que se puede transcribir y/o transcribir de forma inversa para permitir la decodificación de la biblioteca o la polimerización dependiente de pieza de cabeza con una o más polimerasas de ADN o ARN.

[0112] Generalmente, la enzima de ligación produce un oligonucleótido que tiene un enlace fosfodiéster nativo que puede ser transcrito y/o transcrito a la inversa. En este documento se proporcionan métodos ejemplares de ligación enzimática e incluyen el uso de una o más ligasas ARN o ADN, como ligasa T4 ARN, ligasa T4 ADN, CircLigase™ ligasa ssADN, CircLigase™ II ligasa ssADN y ThermoPhage™ ligasa ssADN (Prokaryote Ltd., Reykjavik, Islandia).

[0113] La ligación química también se puede utilizar para producir oligonucleótidos capaces de transcribirse o transcripción inversa. Un beneficio de la ligación química es que la síntesis en fase sólida de tales oligonucleótidos se puede optimizar para soportar un rendimiento de ligación eficiente. Sin embargo, 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 probar mediante cualquier método útil, como cromatografía líquida-espectrometría de masas, análisis de TI-PCR y/o análisis de PCR. En el Ejemplo 5 se proporcionan ejemplos de estos métodos.

[0114] En realizaciones particulares, la ligación química incluye el uso de uno o más pares químicamente co-reactivos para proporcionar un espaciador que se puede transcribir o transcribir de forma inversa. En particular, las reacciones adecuadas para pares químicamente co-reactivos son candidatos preferidos para el proceso de ciclación (Kolb et al., Angew. Chem. Int. Ed., 40: 2004-2021 (2001); Van der Eycken et al., QSAR Comb. Sci., 26: 1115-1326 (2007)). Ejemplos de pares químicamente co-reactivos son un par que incluye un grupo alquínico opcionalmente sustituido y un grupo azido opcionalmente sustituido para formar un espaciador de triazol mediante una reacción de cicloadición 1,3-dipolar de Huisgen; un dieno opcionalmente sustituido que tiene un sistema de 4 electrones  $\pi$  (p. ej., un compuesto 1,3-insaturado opcionalmente sustituido, tal como 1,3-butadieno, 1-metoxi-3-trimetilsililo-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  $2\pi$  (p. ej., un grupo alquénico opcionalmente sustituido o un grupo alquínico opcionalmente sustituido) para formar un espaciador de cicloalquénico mediante una reacción de Diels-Alder; un nucleófilo (p. ej., una amina opcionalmente sustituida o un tiol opcionalmente sustituido) con un electrófilo heterocíclico deformado (p. ej., epóxido, aziridina, ion aziridinio o ion episulfonio opcionalmente sustituidos) para formar un espaciador heteroalquilo mediante una reacción de apertura de anillo; un grupo fosforotioato con un grupo yodo, tal como en una ligación ferulizada de un oligonucleótido que contiene 5'-yodo dT con un oligonucleótido 3'-fosforotioato; y un grupo aldehído y un grupo amino, tal como una reacción de un oligonucleótido modificado con aldehído en 3', que puede obtenerse opcionalmente oxidando un oligonucleótido modificado con glicerilo en 3' disponible comercialmente, con oligonucleótido 5'-amino (es decir, en una reacción de aminación reductora) o un oligonucleótido 5'-hidrazido.

[0115] En otras realizaciones, la ligación química incluye la introducción de un análogo del enlace fosfodiéster, p. ej., para análisis y secuenciación de post-selección de PCR. Ejemplos de análogos de un fosfodiéster incluyen un enlace fosforotioato (p. ej., introducido mediante el uso de un grupo fosforotioato y un grupo saliente, como un grupo yodo), un enlace fosforamida o un enlace fosforoditioato (p. ej., introducido mediante el uso de un grupo fosforoditioato y un grupo saliente, como un grupo yodo).

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

[0116] La invención también presenta una o más condiciones de reacción tal como se definen en las reivindicaciones, que promueven ligación enzimática o química entre la pieza de cabeza y una etiqueta o entre dos etiquetas. Estas condiciones de reacción incluyen el uso de nucleótidos modificados dentro de la etiqueta, como se describe en el presente documento; usar etiquetas donantes y etiquetas aceptoras que tienen diferentes longitudes y varían la concentración de las etiquetas; usando diferentes tipos de ligasas, así como combinaciones de las mismas (p. ej., CircLigase™ ligasa ADN y/o ligasa T4 ARN), y variando su concentración; utilizando polietilenglicoles (PEG) que tienen diferentes pesos moleculares y varían su concentración; uso de agentes de apiñamiento no PEG (p. ej., betaína o albúmina de suero bovino); variando la temperatura y la duración de la ligación; variar la concentración de varios agentes, incluidos ATP,  $\text{Co}(\text{NH}_3)_6\text{Cl}_3$  y pirofosfato inorgánico de levadura; utilizando etiquetas oligonucleotídicas fosforiladas enzimática o químicamente; utilizando etiquetas protegidas en 3'; y el uso de etiquetas preadeniladas. Estas condiciones de reacción también incluyen ligaciones químicas.

[0117] La pieza de cabeza y/o etiquetas pueden incluir uno o más nucleótidos modificados o sustituidos. En realizaciones preferidas, la pieza de cabeza y/o etiquetas incluyen uno o más nucleótidos modificados o sustituidos que promueven la ligación enzimática, tales como 2'-O-metilo-nucleótidos (p. ej., 2'-O-metilo guanina o 2'-O-metilo uracilo), 2'-fluoro-nucleótidos o cualquier otro nucleótido modificado que se utilice como sustrato para la ligación. Alternativamente, la pieza de cabeza y/o las etiquetas se modifican para incluir uno o más grupos químicamente reactivos para soportar la ligación química (p. ej., un grupo alquínico opcionalmente sustituido y un grupo azido opcionalmente sustituido). Opcionalmente, los oligonucleótidos de etiqueta se funcionalizan en ambos extremos con grupos químicamente reactivos y, opcionalmente, uno de estos extremos está protegido, de modo que los grupos se pueden abordar de forma independiente y las reacciones secundarias pueden reducirse (p. ej., reacciones secundarias de polimerización reducidas).

[0118] La ligación enzimática puede incluir una o más ligasas. Ejemplos de ligasas incluyen CircLigase™ ligasa ssADN (EPICENTRE Biotechnologies, Madison, WI), CircLigase™ II ligasa ssADN (también de EPICENTRE Biotechnologies), ThermoPhage™ ligasa ssADN (Prokaryote Ltd., Reykjavik, Islandia), ligasa T4 ARN y ligasa T4

ADN. En realizaciones preferidas, la ligación incluye el uso de una ligasa ARN o una combinación de una ligasa ARN y una ligasa ADN. 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.

**[0119]** Antes o después del paso de ligación, el complejo se puede purificar por tres razones. En primer lugar, el complejo se puede purificar para eliminar la pieza de cabeza o las etiquetas que no han reaccionado que pueden dar lugar a reacciones cruzadas e introducir "ruido" en el proceso de codificación. En segundo lugar, el complejo puede purificarse 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 reducir la actividad de ligación. En tercer lugar, es posible que sea necesario eliminar las entidades que se introducen en una etapa química o de ligación para permitir la posterior etapa química o de ligación. En este documento se describen métodos para purificar el complejo.

**[0120]** La ligación enzimática y química puede incluir polietilenglicol que tiene un peso molecular promedio de más de 300 Daltons (p. ej., más de 600 Daltons, 3.000 Daltons, 4.000 Daltons, o 4.500 daltons). En realizaciones particulares, el polietilenglicol tiene un peso molecular promedio de aproximadamente 3.000 Dalton a 9.000 Dalton (p. ej., de 3000 Dalton a 8.000 Dalton, de 3.000 Dalton a 7.000 Dalton, de 3.000 Dalton a 6.000 Dalton, y de 3.000 Daltons a 5.000 Daltons). En realizaciones preferidas, el polietilenglicol tiene un peso molecular medio de aproximadamente 3.000 Dalton a aproximadamente 6.000 Dalton (p. ej., de 3.300 Dalton a 4.500 Dalton, de 3.300 Dalton a 5.000 Dalton, de 3.300 Dalton a 5.500 Dalton, de 3.300 Dalton a 6.000 Dalton, de 3.500 Dalton a 4.500 Dalton, de 3.500 Dalton a 5.000 Dalton, de 3.500 Dalton a 5.500 Dalton, y de 3.500 Dalton a 6.000 Dalton, como 4.600 Dalton). 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).

**[0121]** En una realización preferida de esta invención, las etiquetas de bloque de construcción se instalan por ligación de un oligonucleótido monocatenario a un oligonucleótido de cadena sencilla utilizando el protocolo de ligación se indica a continuación:

Pieza de cabeza:	25 $\mu\text{m}$ (extremo 5': 5'-monofosfo/2'-OMe G, nucleótidos intermedios: 2'-desoxi y extremo 3': bloqueado 2'/bloqueado 3')
Etiqueta de bloque de construcción:	25 $\mu\text{m}$ (extremo 5': 2'-OMe/5'-OH G, nucleótidos intermedios: 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)
Ligasa T4 ARN (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

**[0122]** En realizaciones adicionales, el protocolo incluye incubación a 37°C durante 20 horas. Para los propósitos de la construcción de la biblioteca real, se puede usar una mayor concentración de pieza de cabeza, etiquetas y/o ligasa, y tales modificaciones a estas concentraciones serán evidentes para los expertos en la técnica.

#### Métodos para la codificación de entidades químicas dentro de una biblioteca

**[0123]** Los métodos de la invención se pueden usar para sintetizar una biblioteca que tiene un número diverso de entidades químicas que son codificadas por las etiquetas de oligonucleótidos. En la publicación de solicitud de patente de EE.UU. n° 2007/0224607 se encuentran ejemplos de bloques de construcción y etiquetas de ADN codificantes.

**[0124]** 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 (p. ej., una triazina para proporcionar tres nodos dispuestos espacialmente alrededor de un anillo heteroarilo o una geometría lineal).

**[0125]** Los bloques de construcción y sus etiquetas que codifican pueden añadirse directamente o indirectamente (p. ej., a través de un enlazador) a la pieza de cabeza para formar un complejo. Cuando el cabezal incluye un enlazador, el bloque de construcción o andamio se agrega al final del enlazador. Cuando el enlazador está ausente, el bloque de construcción se puede agregar directamente a la pieza de cabeza o el bloque de construcción en sí puede incluir un enlazador que reacciona con un grupo funcional de la pieza de cabeza. En este documento se describen conectores y piezas de cabeza ejemplares.

**[0126]** El andamio se puede añadir de cualquier forma útil. Por ejemplo, el andamio se puede agregar al final del enlazador o la pieza de cabeza, 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 enlazador o a la pieza de cabeza, 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 añadir opcionalmente a la pieza de cabeza o al complejo. Por ejemplo,  $S_n$  se agrega al complejo en  $n$  recipientes de reacción, donde  $n$  es un número entero más que uno, y la etiqueta  $S_n$  (es decir, la etiqueta  $S_1, S_2, \dots, S_{n-1}, S_n$ ) es ligada al grupo funcional del complejo.

**[0127]** Bloques de construcción se pueden agregar en múltiples pasos sintéticos. Por ejemplo, una parte alícuota de la pieza de cabeza, que tiene opcionalmente un conector adjunto, se separa en  $n$  recipientes de reacción, donde  $n$  es un número entero de dos o más. En el primer paso, el bloque de construcción  $A_n$  se agrega a cada  $n$  recipiente de reacción (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  $A_n$  es único. En el segundo paso, se agrega el andamio  $S$  a cada recipiente de reacción para formar un complejo  $A_n-S$ . Opcionalmente, el andamio  $S_n$  se puede añadir a cada recipiente de reacción a partir de 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, el bloque de construcción  $B_n$  es para 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,  $\dots, n-1, n$  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 pasos posteriores, el bloque de construcción  $C_n$  se puede agregar 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 a recipiente de reacción 1, 2,  $\dots, n-1, n$  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  $n^3$  números de complejos que tienen  $n^3$  etiquetas. De esta manera, se pueden usar pasos sintéticos adicionales para unir bloques de construcción adicionales para diversificar aún más la biblioteca.

**[0128]** Después de la formación de la biblioteca, los complejos resultantes pueden opcionalmente ser purificados y sometidos a una polimerización de reacción o de ligación utilizando uno o más cebadores. Esta estrategia general se puede ampliar para incluir nodos de diversidad y bloques de construcción adicionales (p. ej., 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 oligonucleotídica. Luego, se hacen reaccionar bloques de construcción adicionales con el complejo resultante, y el nodo de diversidad subsiguiente se deriva mediante bloques de construcción adicionales, que están codificados por el cebador utilizado para la reacción de polimerización o ligación.

**[0129]** Para formar una biblioteca codificada, se agregan etiquetas de oligonucleótidos al complejo después o antes de cada paso sintético. Por ejemplo, antes o después de la adición del bloque de construcción  $A_n$  a cada recipiente de reacción, la etiqueta  $A_n$  está unida al grupo funcional de la pieza de cabeza (es decir, la etiqueta  $A_1, A_2, \dots, A_{n-1}, A_n$  se añade al recipiente de reacción 1, 2,  $\dots, n-1, n$  que contiene la pieza de cabeza). Cada etiqueta  $A_n$  tiene una secuencia distinta que se correlaciona con cada bloque de construcción único  $A_n$ , y la determinación de la secuencia de la etiqueta  $A_n$  proporciona la estructura química del bloque de construcción  $A_n$ . De esta manera, se utilizan etiquetas adicionales para codificar bloques de construcción adicionales o andamios adicionales.

**[0130]** Por otra parte, la última etiqueta que se añade al complejo puede incluir una secuencia de cebador o proporcionar un grupo funcional para permitir la unión (p. ej., por ligación) de una secuencia de cebador. La secuencia del cebador se puede usar para amplificar y/o secuenciar las etiquetas de oligonucleótidos del complejo. Los métodos ejemplares para amplificar y secuenciar incluyen la reacción en cadena de la polimerasa (PCR), amplificación de cadena lineal (LCR), amplificación por círculo rodante (RCA) o cualquier otro método conocido en la técnica para amplificar o determinar secuencias de ácido nucleico.

**[0131]** Usando estos métodos, se pueden formar grandes bibliotecas que tienen un gran número de entidades químicas codificadas. Por ejemplo, una pieza de cabeza se reacciona con un enlazador 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 se extiende un cebador a la pieza de cabeza. Estas reacciones se pueden realizar en una placa de 1.000 pocillos o en placas de 10 x 100 pocillos. 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. Se puede ligar una etiqueta de ADN  $B_n$  al complejo  $A_n$ -pieza de cabeza y se pueden combinar todas las reacciones. La biblioteca resultante incluye 1.000 x 1.000 combinaciones de  $A_n \times B_n$  (es decir, 1.000.000 de compuestos) etiquetados por 1.000.000 de combinaciones diferentes de etiquetas. El mismo enfoque puede extenderse para agregar bloques de construcción  $C_n, D_n, E_n$ , etc. La biblioteca generada puede usarse luego para identificar compuestos que se unen a diana. La estructura de las entidades químicas que se unen a la biblioteca se puede evaluar opcionalmente mediante PCR y secuenciación de las etiquetas de ADN para identificar los compuestos que se enriquecieron.

**[0132]** Este método se puede modificar para evitar el marcado tras la adición de cada bloque de construcción o para evitar la puesta en común (o mezcla). Por ejemplo, el método se puede modificar agregando el bloque de construcción  $A_n$  a  $n$  recipientes de reacción, donde  $n$  es un número entero de más de uno, y agregando 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 se necesita una etiqueta de oligonucleótido que codifique este bloque de construcción. Después de agregar un bloque de construcción, los complejos pueden agruparse o no agruparse. Por ejemplo, la biblioteca no se agrupa después del paso final de la

adición de bloques de construcción, y las agrupaciones se analizan individualmente para identificar los compuestos que se unen a una diana. Para evitar agrupar todas las reacciones después de la síntesis, se puede utilizar un lector BIND® (de SRU Biosystems, Inc.), p. ej., para controlar la unión en la superficie del sensor en un formato de alto rendimiento (p. ej., placas de 384 pocillos y placas de 1.536 pocillos). Por ejemplo, el bloque de construcción  $A_n$  puede estar codificado con la etiqueta de ADN  $A_n$ , y el bloque de construcción  $B_n$  puede estar codificado por su posición dentro de la placa de pocillos. Los compuestos candidatos se pueden identificar mediante un ensayo de unión (p. ej., utilizando un Biosensor BIND®, también disponible por SRU Biosystems, Inc., o utilizando un ensayo ELISA) y analizando las etiquetas  $A_n$  mediante secuenciación, análisis de microarray y/o análisis de digestión de restricción. Este análisis permite la identificación de combinaciones de componentes básicos  $A_n$  y  $B_n$  que producen las moléculas deseadas.

**[0133]** 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 (p. ej., concentración de complejo y tamaño de microrreactores) se pueden ajustar para proporcionar, en promedio, un microrreactor que tiene al menos un miembro de una biblioteca de compuestos. Cada microrreactor también puede contener la diana, una sola cuenta capaz de unirse a un complejo o una porción del complejo (p. ej., una o más etiquetas) y/o unirse a la diana, y una solución de reacción de amplificación que tiene uno o más reactivos necesarios para realizar amplificación de ácido nucleico. Después de amplificar la etiqueta en los microrreactores, las copias amplificadas de la etiqueta se unirán a las cuentas en los microrreactores, y las cuentas recubiertas se pueden identificar mediante cualquier método útil.

**[0134]** Una vez que los bloques de construcción de la primera biblioteca que se unen a la diana de interés han sido identificados, una segunda biblioteca se puede preparar de forma iterativa. Por ejemplo, se pueden agregar uno o dos nodos adicionales de diversidad, y se crea y muestrea la segunda biblioteca, como se describe en este documento. Este proceso puede repetirse tantas veces como sea necesario para crear moléculas con las propiedades moleculares y farmacéuticas deseadas.

**[0135]** Diversas técnicas de ligación pueden utilizarse para añadir el andamio, bloques de construcción, enlazadores, y la construcción de etiquetas de bloque. Por consiguiente, cualquiera de los pasos de unión descritos en el presente documento puede incluir cualquier técnica o técnicas de unión útiles. Las técnicas de ligación ejemplares incluyen ligación enzimática, tal como el uso de una o más ligasas de ARN y/o ligasas de ADN, como se describe en el presente documento; y ligación química, tal como el uso de pares químicamente co-reativos, como se describe en este documento.

#### Andamios y bloques de construcción

**[0136]** El andamio S puede ser un solo átomo o un andamio molecular. Los andamios de un solo átomo ejemplares incluyen un átomo de carbono, un átomo de boro, un átomo de nitrógeno o un átomo de fósforo, etc. Los andamios poliatómicos ejemplares incluyen un grupo cicloalquilo, un grupo cicloalqueno, un grupo heterocicloalquilo, un grupo heterocicloalqueno, un grupo aroli o un grupo heteroaroli. Las realizaciones particulares de un andamio de heteroaroli 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.

**[0137]** 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 pieza de cabeza. Para obtener este andamio ejemplar, se hace reaccionar triclorotriazina (es decir, un precursor clorado de triazina que tiene tres cloros) con un grupo nucleófilo de la pieza de cabeza. 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 está unida a la pieza de cabeza. A continuación, el bloque de construcción  $A_n$  se agrega a un nodo de diversidad del andamio, y la codificación de la etiqueta  $A_n$  para el bloque de construcción  $A_n$  ("etiqueta  $A_n$ ") se liga a la pieza de cabeza, donde estos dos pasos se pueden realizar en cualquier orden. Luego, el bloque de construcción  $B_n$  se agrega al nodo de diversidad restante, y la codificación de la etiqueta  $B_n$  para el bloque de construcción  $B_n$  se liga al final de la etiqueta  $A_n$ . En otro ejemplo, S es una triazina que está operativamente enlazada al enlazador de una etiqueta, donde la triclorotriazina reacciona con un grupo nucleofílico (p. ej., un grupo amino) de un enlazador PEG, alifático o aromático de una marca. Se pueden agregar bloques de construcción y etiquetas asociadas, como se describe anteriormente.

**[0138]** En otro ejemplo, S es una triazina que está unida operativamente a la construcción de bloque  $A_n$ . Para obtener este andamio, el bloque de construcción  $A_n$  que tiene dos nodos de diversidad (p. ej., un grupo electrófilo y un grupo nucleófilo, como un Fmoc-aminoácido) se hace reaccionar con el grupo nucleófilo de un enlazador (p. ej., el grupo terminal de un PEG, alifático, o enlazador aromático, que se adjunta a una pieza de cabeza). Luego, la triclorotriazina se hace reaccionar con un grupo nucleofílico del bloque de construcción  $A_n$ . Con este método, las tres posiciones de cloro de S se utilizan como nodos de diversidad para los bloques de construcción. Como se describe en este documento, se pueden agregar bloques de construcción y etiquetas adicionales, y se pueden agregar andamios adicionales  $S_n$ .

**[0139]** Los bloques de construcción ejemplares  $A_n$  incluyen, p. ej., aminoácidos (p. ej., aminoácidos alfa, beta, gamma, delta y épsilon, así como derivados de aminoácidos naturales y no naturales), reactivos químicamente co-reativos (p. ej., cadenas de azida o alquino) con una amina, o un reactivo tiol, o combinaciones de los mismos. La elección del bloque de construcción  $A_n$  depende, p. ej., de la naturaleza del grupo reactivo utilizado en el enlazador, la naturaleza de un resto de andamio y el disolvente utilizado para la síntesis química.

**[0140]** Los bloques de construcción ejemplares  $B_n$ 's y  $C_n$ 's incluyen cualquier unidad estructural útil de una entidad química, tal como grupos aromáticos opcionalmente sustituidos (p. ej., fenilo o bencilo opcionalmente sustituidos), grupos heterociclilo opcionalmente sustituidos (p. ej., quinolinilo, isoquinolinilo, indolilo, isoindolilo, azaindolilo, bencimidazolilo, azabencimidazolilo, bencisoxazolilo, piridinilo, piperidilo, o pirrolidinilo opcionalmente sustituidos), grupos alquilo opcionalmente sustituido (p. ej., grupos  $C_{1-6}$  alquilo lineal o ramificado opcionalmente sustituidos o grupos  $C_{1-6}$  aminoalquilo opcionalmente sustituidos), o grupos carbociclilo opcionalmente sustituidos (p. ej., ciclopropilo, ciclohexilo o ciclohexenilo opcionalmente sustituidos). Particularmente bloque de construcción útil  $B_n$ 's y  $C_n$ 's incluyen los que tienen uno o más grupos reactivos, tal como un grupo opcionalmente sustituido (p. ej., cualquiera de los descritos en este documento) que tiene uno o más sustituyentes opcionales que son grupos reactivos o se pueden modificar 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  $C_{1-6}$  alquilo opcionalmente sustituido), hidroxilo, alcoxi ( $-OR$ , donde R es un  $C_{1-6}$  alquilo opcionalmente sustituido, como metoxi), carboxi ( $-COOH$ ), amida o sustituyentes químicamente co-reativos. Se puede introducir un sitio de restricción, p. ej., en la etiqueta  $B_n$  o  $C_n$ , donde se puede identificar un complejo realizando PCR y digestión de restricción con una de las correspondientes enzimas de restricción.

### Enlazadores

**[0141]** El enlazador bifuncional entre la pieza de cabeza y la entidad química se pueden variar para proporcionar un espaciador apropiado y/o para aumentar la solubilidad de la pieza de cabeza en disolvente orgánico. Está disponible comercialmente una amplia variedad de enlazadores que pueden acoplar la pieza de cabeza con la biblioteca de moléculas pequeñas. El enlazador consiste típicamente en cadenas lineales o ramificadas y puede incluir un  $C_{1-10}$  alquilo, un heteroalquilo de 1 a 10 átomos, un  $C_{2-10}$  alquenoilo, un  $C_{2-10}$  alquínilo,  $C_{5-10}$  arilo, un sistema cíclico o policíclico de 3 a 20 átomos, un fosodiéster, un péptido, un oligosacárido, un oligonucleótido, un oligómero, un polímero o un polialquilglicol (p. ej., un polietilenglicol, como  $-(CH_2CH_2O)_nCH_2CH_2-$ , donde n es un número entero de 1 a 50), o combinaciones de los mismos.

**[0142]** El enlazador bifuncional puede proporcionar un espaciador apropiado entre la pieza de cabeza y una entidad química de la biblioteca. En ciertas realizaciones, el enlazador bifuncional incluye tres partes. La parte 1 puede ser un grupo reactivo, que forma un enlace covalente con el ADN, como, p. ej., un ácido carboxílico, preferiblemente activado por un éster de N-hidroxisuccinimida (NHS) para reaccionar con un grupo amino en el ADN (p. ej., dT modificado por amino), una amidita para modificar el extremo 5' o 3' de una pieza de cabeza monocatenaria (lograda mediante la química de oligonucleótidos estándar), pares químicamente co-reativos (p. ej., cicloadición de azidoalquino en presencia de catalizador de Cu(I), o cualquiera descrito en el presente documento), o grupos reactivos con tiol. La parte 2 también puede ser un grupo reactivo, que forma un enlace covalente con la entidad química, ya sea el bloque de construcción  $A_n$  o un andamio. Tal grupo reactivo podría ser, p. ej., una amina, un tiol, una azida o un alquino. La parte 3 puede ser un espaciador químicamente inerte de longitud variable, introducido entre las partes 1 y 2. Dicho espaciador puede ser una cadena de unidades de etilenglicol (p. ej., PEG de diferentes longitudes), un alcano, un alqueno, una cadena de polienos o una cadena de péptidos. El enlazador puede contener ramificaciones o inserciones con restos hidrófobos (tales como, p. ej., anillos de benceno) para mejorar la solubilidad de la pieza de cabeza en disolventes orgánicos, así como restos fluorescentes (p. ej., fluoresceína o Cy-3) usados con fines de detección de bibliotecas. Los residuos hidrófobos en el diseño de la pieza de cabeza pueden variar con el diseño del enlazador para facilitar la síntesis de bibliotecas en disolventes orgánicos. Por ejemplo, la combinación de pieza de cabeza y conector está diseñada para tener residuos apropiados en los que el coeficiente octanol: agua ( $P_{oct}$ ) es de, p. ej., 1,0 a 2,5.

**[0143]** Los enlazadores se pueden seleccionar empíricamente para un diseño de biblioteca de moléculas pequeñas dado, de modo que la biblioteca se pueda sintetizar en un disolvente orgánico, p. ej., en 15%, 25%, 30%, 50%, 75%, 90%, 95%, 98%, 99% o 100% de disolvente orgánico. El enlazador se puede variar usando reacciones modelo antes de la síntesis de la biblioteca para seleccionar la longitud de cadena apropiada que solubiliza la pieza de cabeza en un disolvente orgánico. Los enlazadores ejemplares incluyen aquellos que tienen mayor longitud de cadena de alquilo, mayores unidades de polietilenglicol, especies ramificadas con cargas positivas (para neutralizar las cargas negativas de fosfato en la pieza de cabeza), o mayores cantidades de hidrofobicidad (p. ej., adición de estructuras de anillo de benceno).

**[0144]** Los ejemplos de enlazadores disponibles comercialmente incluyen enlazadores aminocarboxílicos, como los que son péptidos (p. ej., Z-Gly-Gly-Gly-Osu (N-alfa-benciloxycarbonil-(glicina)<sub>3</sub>-N-succinimidil éster) o Z-Gly-Gly-Gly-Gly-Gly-Gly-Osu (N-alfa-benciloxycarbonil-(glicina)<sub>6</sub>-N-succinimidil éster, SEQ ID NO: 3), PEG (p. ej., Fmoc-aminoPEG2000-NHS o amino-PEG (12-24)-NHS), o cadenas de ácido alcano (p. ej., ácido Boc-ε-aminocaproico-Osu); enlazadores de pares químicamente co-reativos, tales como los pares químicamente co-reativos descritos aquí en combinación con un resto peptídico (p. ej., azidohomoalanina-Gly-Gly-Gly-OSu (SEQ ID NO: 4) o propargilglicina-Gly-

Gly-Gly-OSu (SEQ ID NO: 5), PEG (p. ej., Azido-PEG-NHS) o un resto de cadena de ácido alcano (p. ej., ácido 5-azidopentanoico, (S)-2-(azidometilo)-1-Boc-pirrolidina, 4-azidoanilina o éster de N-hidroxisuccinimida del ácido 4-azido-butan-1-oico); enlazadores reactivos con tiol, como los que son PEG (p. ej., SM(PEG)<sub>n</sub> NHS-PEG-maleimida), cadenas de alcanos (p. ej., ácido 3-(piridin-2-ildisulfanil)-propiónico-Osu o sulfosuccinimidil 6-(3'-[2-piridilditio]-propionamido)hexanoato); y amiditas para la síntesis de oligonucleótidos, tales como modificadores de amino (p. ej., 6-(trifluoroacetilamino)-hexil-(2-cianoetilo)-(N,N-diisopropilo)-fosforamidita), modificadores de tiol (p. ej., S-tritilo-6-mercaptohexil-1-[(2-cianoetilo)-(N,N-diisopropilo)]-fosforamidita, o modificadores de pares químicamente co-activos (p. ej., 6-hexil-1-ilo-(2-cianoetilo)-(N,N-diisopropilo)fosforamidita, 3-dimetoxitritilo-2-(3-(3-propargiloxipropanamido)propanamido)propil-1-O-succinoilo, alquilamino CPG de cadena larga o éster de N-hidroxisuccinimida del ácido 4-azido-butan-1-oico)). Se conocen enlazadores adicionales en la técnica, y los que se pueden usar durante la síntesis de bibliotecas incluyen, pero no se limitan a 5'-O-dimetoxitritilo-1',2'-didesoxirribosa-3'-[(2-cianoetilo)-(N,N-diisopropilo)]-fosforamidita; 9-odimetoxitritilo-trietilenglicol, 1-[(2-cianoetilo)-(N,N-diisopropilo)]-fosforamidita; 3-(4,4'-dimetoxitritilo)propil-1-[(2-cianoetilo)-(N,N-diisopropilo)]-fosforamidita; y 18-O-dimetoxitritil hexaetilenglicol, 1-[(2-cianoetilo)-(N,N-diisopropilo)]-fosforamidita. Cualquiera de los enlazadores de la presente puede añadirse en tándem entre sí en diferentes combinaciones para generar enlazadores de diferentes longitudes deseadas.

[0145] Los enlazadores también pueden ser ramificados, donde los enlazadores ramificados son bien conocidos en la técnica y los ejemplos pueden consistir en duplicadores simétricos o asimétricos o un agudo simétrico. Véase, p. ej., Newcome et al., *Dendritic Molecules: Concepts, Synthesis, Perspectives*, VCH Publishers (1996); Boussif et al., *Proc. Natl. Acad. Sci. EE.UU.* 92: 7297 - 7301 (1995); y Jansen et al., *Science* 266: 1226 (1994).

### Ejemplo 1

#### Estrategia general para mejorar la ligación de una sola cadena de etiquetas de ADN

[0146] Se exploraron diversas condiciones de reacción para mejorar la ligación de una sola cadena de etiquetas para formar una biblioteca codificada. Estas condiciones de reacción incluyen el uso de nucleótidos modificados dentro de la etiqueta (p. ej., el uso de uno o más nucleótidos que tienen un grupo 2'-OMe para formar una etiqueta de MNA/ADN, donde "MNA" se refiere a un oligonucleótido que tiene al menos un 2'-O-metilnucleótido); usar etiquetas donantes y etiquetasceptoras que tienen diferentes longitudes y varían la concentración de las etiquetas; utilizando diferentes tipos de ligasas, así como combinaciones de las mismas (p. ej., CircLigase™ ligasa ssADN y/o ligasa T4 ARN), y variando su concentración; purificar el complejo eliminando los materiales de partida que no han reaccionado; utilizando polietilenglicoles (PEG) que tienen diferentes pesos moleculares y varían su concentración; variar la temperatura y la duración de la reacción, como la ligación; variar la concentración de varios agentes, incluidos ATP, Co(NH<sub>3</sub>)<sub>6</sub>Cl<sub>3</sub> y pirofosfato inorgánico de levadura; utilizando etiquetas oligonucleotídicas fosforiladas enzimática o químicamente; utilizando etiquetas protegidas en 3'; y usando etiquetas adeniladas químicamente en 5'.

[0147] Después de un análisis exhaustivo de las diferentes condiciones, combinaciones óptimas de parámetros que presentó hasta un 90% de eficiencia de ligación (p. ej., la Figura 5C), según lo determinado por la fracción de producto final ligado al reactivo de partida no ligado ("ligado en fracción"), fueron encontrados. En la Figura 5A se muestra un esquema de la reacción de ligación usando ligasa, y en la Figura 5B se muestra una electroforesis en gel de poliacrilamida desnaturalizante típica. El oligonucleótido donante se marcó en el extremo 3' y se pudo detectar en un gel escaneando a 450 nm de excitación en un Storm™ 800 PhosphorImager. El gel representa un donante no ligado (o material de partida) y el producto ligado. En particular, el donante adenilado puede resolverse y distinguirse del material de partida en este gel.

[0148] La Tabla 2 proporciona las eficiencias de ligación medidas en función de la composición del oligonucleótido (es decir, oligonucleótidos con todos los nucleótidos de ADN frente a oligonucleótidos con al menos un nucleótido 2'-O-metilo, marcado como "MNA") y el tipo de ligasa (es decir, ligasa ARN frente a ligasa de ADNss). Estos experimentos de ligación incluyeron las siguientes etiquetas: un donante totalmente de ADN que tiene la secuencia de 5'-P-GCT GTG CAG GTA GAG TGC-6-FAM-3' (SEQ ID NO: 6); un donante de 5'-MNA-ADN que tiene la secuencia de 5'-P-mGCT GTG CAG GTA GAG TGC-6-FAM-3' (SEQ ID NO: 7); un donante todo-MNA que tiene la secuencia de 5'-P-mGmUmG mCmAmG mGmUmA mGmAmG mUmGmC-6-FAM-3' (SEQ ID NO: 8); un aceptor de ADN-3'MNA que tiene la secuencia de 5'-HO-TAC GTA TAC GAC TGG-OH-3' (SEQ ID NO: 9); un aceptor de todo ADN que tiene la secuencia de 5'-HO-GCA GAC TAC GTA TAC GAC TGG-OH-3' (SEQ ID NO: 10); y un aceptor todo-MNA que tiene la secuencia de 5'-HO-mUmAmC mGmUmA mUmAmC mGmAmC mUmGmG-OH-3' (SEQ ID NO: 11), donde "m" indica una base 2'-OMe, "P" indica un nucleótido fosforilado y "FAM" indica fluoresceína.

[0149] Eficiencias de ligación se calculan a partir de los datos de densitometría en gel como la relación entre la intensidad del producto de ligación y la suma de la intensidad a partir del producto de ligación y el material de partida no ligado. Las condiciones de reacción para la ligasa T4 ARN incluyeron lo siguiente: 5 µm de cada uno de los oligonucleótidos donantes y aceptores (15-18 nucleótidos (nts) de longitud) en una solución tampón que contiene Tris HCl 50 mM, MgCl<sub>2</sub> 10 mM, cloruro de cobalto de hexamina 1 mM, ATP 1 mM, PEG4600 al 25% y 5 unidades de ligasa ARN T4 (unidades nuevas NEB) a pH 7,5. Las reacciones se incubaron a 37°C durante 16 horas. Las condiciones de reacción para CircLigase™ incluyeron lo siguiente: 5 µm de cada uno de los oligonucleótidos donantes y aceptores

(longitud 15 o 18 nts) incubados en un tampón que contiene MOPS 50 mM (pH 7,5), KCl 10 mM, MgCl<sub>2</sub> 5 mM, 1 mM TDT, 0,05 mM de ATP, 2,5 mM MnCl<sub>2</sub>, y 25% (p/v) PEG 8000 con 20 unidades de CircLigase™ (Epicentre) a 50°C durante 16 horas. Las reacciones se resolvieron en urea 8M/PAAG al 15%, seguido de densitometría usando excitación a 450 nm.

Tabla 2

Donante	Aceptor	T4 ligasa ARN	CircLigase™
Todo ADN	Todo ADN	9%	89%
Todo ADN	Todo-MNA	14%	68%
Todo ADN	ADN-3'MNA	46%	85%
Todo-MNA	Todo ADN	11%	84%
Todo-MNA	Todo-MNA	20%	29%
Todo-MNA	ADN-3'MNA	32%	73%
5'-MNA-ADN	Todo ADN	29%	90%
5'-MNA-ADN	Todo-MNA	16%	46%
5'-MNA-ADN	ADN-3'MNA	69%	81%

**[0150]** En general, CircLigase™ produce mayores rendimientos de ligación que la ligasa T4 ARN (Tabla 2). Cuando tanto el donante como el aceptor eran oligonucleótidos híbridos de ADN/MNA, se logró una ligación eficaz con la ligasa T4 ARN.

**[0151]** La Figura 5C muestra la ligación de alto rendimiento conseguida para la ligasa T4 ARN a concentraciones elevadas de enzima y oligonucleótidos. Las condiciones de reacción incluyeron lo siguiente: 250 µm de cada oligonucleótido donante y aceptor en un tampón que contiene Tris HCl 50 mM, MgCl<sub>2</sub> 10 mM, cloruro de cobalto de hexamina 1 mM, ATP 2,5 mM, PEG4600 al 30% (p/v), pH 7,5, diferentes cantidades de ligasa T4 ARN a 40 unidades/µL (NEB- nuevas unidades), y 0,1 unidad de pirofosfatasa inorgánica de levadura. Las reacciones se incubaron a 37°C durante 5 y 20 horas y se resolvieron en urea 8M/PAAG al 15%, seguido de densitometría usando excitación a 450 nm.

**[0152]** En general, estos datos sugieren que la ligación enzimática se puede optimizar mediante la inclusión de uno o más 2'-nucleótidos modificados y/o mediante el uso de una ligasa ARN o ADN. A continuación se analizan más detalles para varias otras condiciones probadas, como PEG o longitud de etiqueta, que pueden contribuir a la eficacia de la ligación.

## Ejemplo 2

### Efecto de la PEG en la ligación de una sola cadena

**[0153]** Para determinar el efecto del peso molecular de PEG (MW) en la ligación, las etiquetas de una sola cadena se ligaron con 25% (p/v) de PEG que tiene un PM de 300 a 20.000 Daltons. Como se muestra en la Figura 6A, se observó un 80% o más de ligación para PEG que tenía un PM de 3.350, 4.000, 6.000, 8.000 y 20.000. Estos experimentos de ligación incluyeron las siguientes etiquetas: un donante de 15 mer que tiene la secuencia de 5'-P-mGTG CAG GTA GAG TGC-6-FAM-3' (SEQ ID NO: 12) y un aceptor de 15 mer que tiene la secuencia de 5'-HO-mUAC GTA TAC GAC TGmG-OH-3' (SEQ ID NO: 13). Estas etiquetas de oligonucleótidos eran secuencias de ADN con una o dos bases de ARN 2'-O-metilo (2'-OMe) de extremo (p. ej., 2'-OMe-U (mU) o 2'-OMeG (mG)).

**[0154]** También se realizaron experimentos para determinar el efecto de la concentración de PEG. Se ligaron etiquetas monocatenarias con diversas concentraciones de PEG que tenían un PM de 4.600 Daltons (PEG4600). Como se muestra en la Figura 6B, se observó un 70% o más de ligación, en promedio, para 25% (p/v) a 35% (p/v) de PEG4600.

## Ejemplo 3

### Efecto de la longitud de la etiqueta en la ligación de una sola cadena

**[0155]** Para determinar el efecto de la longitud de etiqueta en las etiquetas de ligación, aceptores y donantes de diversas longitudes se construyeron. Para los experimentos de CircLigase™, se utilizó un donante de 15 mer con la secuencia 5'-P-mGTG CAG GTA GAG TGC-6-FAM-3' (SEQ ID NO: 12) y se combinó con 10, 12, 14, 16 y 18 mer oligonucleótidos aceptores de ADN. Para los experimentos de ligasa T4 ARN, las etiquetas incluían una o más bases 2'-OMe (designadas como etiquetas MNA/ADN). La Tabla 3 proporciona la secuencia para las tres etiquetas donantes (15mer, 8mer y 5mer) y las tres etiquetasceptoras (15mer, 8mer y 5mer).



Tabla 3

Etiqueta de oligonucleótido	Secuencia*
Donante 15mer	5'-P-mGTG CAG GTA GAG TGC-6-FAM-3' (SEQ ID NO: 12)
Aceptor 15mer	5'-HO-mUAC GTA TAC GAC TGmG-OH-3' (SEQ ID NO: 13)
Donante 8mer	5'-P-mGT GAG TGC-6-FAM-3' (SEQ ID NO: 14)
Aceptor 8mer	5'-HO-C A GAC TGmG-OH-3' (SEQ ID NO: 15)
Donante 5mer	5'-P-mGT GAC-6-FAM-3' (SEQ ID NO: 16)
Aceptor 5mer	5'-HO-mAC TGmG-OH-3' (SEQ ID NO: 17)
* "m" indica una base 2'-OMe, "P" indica un nucleótido fosforilado y "FAM" indica fluoresceína.	

**[0156]** El grado de ligación se analizó por densitometría de geles electroforéticos (Figuras 7A-7B). Los resultados de las reacciones CircLigase™ indican una fuerte dependencia del rendimiento de la ligación de la longitud del oligonucleótido aceptor (Figura 7A). El rendimiento de ligación más alto se observó con un aceptor de 18 mer (62%), mientras que el rendimiento de ligación con un aceptor de 10 mer fue inferior al 10%. Los resultados de las reacciones de la ligasa ARN T4 indican que la combinación de un aceptor de 8 mer con un donante de 8 mer proporcionó el mayor rendimiento y que las combinaciones que tenían un donante de 15 mer con cualquiera de los aceptores probados proporcionaron rendimientos superiores al 75% (Figura 7B). Si una biblioteca incluye etiquetas más cortas (es decir, aproximadamente 10 mer o menos), entonces puede preferirse la ligasa T4 ARN para la ligación de etiquetas. En otros casos, la ligación se puede optimizar aún más utilizando CircLigase™ o la combinación de ligasa T4 ARN y CircLigase™.

#### Ejemplo 4

##### Efecto de la purificación sobre la ligación de una sola cadena

**[0157]** Para determinar el efecto de la purificación en la ligación, las etiquetas de una sola cadena se ligaron para imitar el proceso sintético de biblioteca. Para estos experimentos, las etiquetas incluían etiquetas de donante de 15 mer y de aceptor de 15 mer, como se proporciona anteriormente en la Tabla 3. La entidad química se unió al extremo 3' de la biblioteca, donde la entidad química era fluoresceína en este ejemplo para ayudar en la visualización. Como se muestra en la Figura 9 (derecha), las etiquetas sucesivas se ligaron al grupo 5'-OH del complejo después de la fosforilación por T4 PNK.

**[0158]** Los experimentos también se llevaron a cabo por la purificación del producto ligado (es decir, el complejo) antes de la reacción PNK, donde los agentes particulares útiles en la reacción de ligación (p. ej., fosfato, cobalto, y/o las etiquetas que no han reaccionado) pueden inhibir la reacción de fosforilación con PNK o reducir el rendimiento de la ligación. Como se muestra en la Figura 9 (izquierda), la purificación del complejo (es decir, precipitación mínima) antes de la reacción de PNK aumentó la ligación (ver datos marcados con \*, que indican purificación). Las Figuras 8A-8B muestran los espectros de CL-EM para una etiqueta de MNA/ADN de 15 mer antes y después de la fosforilación. La presencia o ausencia de DTT no tuvo ningún efecto sobre la fosforilación.

#### Ejemplo 5

##### Ligación de par químicamente co-reactiva y la transcripción inversa de las uniones

**[0159]** Los métodos descritos en el presente documento puede incluir además técnicas de ligación de par químicamente co-reactiva, así como de técnicas de ligación enzimática. En consecuencia, como ejemplo de ligación química, un par ejemplar químicamente co-reactivo (es decir, un par alquino y un azido en una reacción de cicloadición) en dos variantes: se utilizó un par químicamente co-reactivo corto y un par químicamente co-reactivo largo.

##### Materiales

**[0160]** En una primera variante, se utilizó un par químicamente co-reactivo corto (Figura 10A). El par incluía (i) un oligonucleótido que tiene la secuencia 5'-GCG TGA ACA TGC ATC TCC CGT ATG CGT ACA GTC CAT T/propargilG/-3' ("5extremo3propargil", SEQ ID NO: 18) y (ii) un oligonucleótido que tiene la secuencia 5'-/azidoT/ATA GCG CGA TAT ACA CAC TGG CGA GCT TGC GTA CTG-3' ("3extremo5azido", SEQ ID NO: 19). Este par de oligonucleótidos fue preparado por TriLink BioTechnologies, Inc. (San Diego, CA). Estos oligonucleótidos se diseñaron para producir un espaciador corto entre dos oligonucleótidos tras la ligación, donde el enlazador tendría 5 átomos de largo (contando desde la posición C3' del oligonucleótido 5extremo3propargil hasta la posición C5' del oligonucleótido azido 3extremo5). Además, el oligonucleótido 5'-azido (3extremo5azido) se preparó convirtiendo el grupo yodo en el oligonucleótido 5'-yodo correspondiente en un grupo azido.

**[0161]** En una segunda variante, se utilizó un par químicamente co-reactivo largo (Figura 10B). El par incluía (i) un oligonucleótido que tiene la secuencia 5'-GCG TGA ACA TGC ATC TCC CGT ATG CGT ACA GTC CAT TG/espaciador7-azida/-3' ("5extremo3azida", SEQ ID NO: 20) y (ii) un oligonucleótido que tiene la secuencia 5'-/hexiniloTA GCG CGA TAT ACA CAC TGG CGA GCT TGC GTA CTG-3' ("3extremo5hexinilo", SEQ ID NO: 21). Este par de oligonucleótidos fue preparado por Integrated ADN Technologies, Inc. (IDT ADN, San Diego, CA y Coralville,

IA). El oligonucleótido 5extremo3azida se preparó haciendo reaccionar un éster de azidobutirato N-hidroxisuccinimida con un 3'-amino-modificador C7 (2-dimetoxitritil oximetil-6-fluorenilmetoxycarbonilamino-hexano-1-succiniloalquilamino de cadena larga), que se introdujo durante la síntesis de un alquilamino de cadena larga), que se introdujo. Este par fue diseñado para producir un espaciador de 24 átomos de longitud entre los oligonucleótidos (contando desde la posición C3' del oligonucleótido 5extremo3azida hasta la posición C5' del oligonucleótido 3extremo5hexinilo).

**[0162]** Para la transcripción inversa (como se muestra en el esquema de la Figura 11A), los cebadores y las plantillas incluían lo siguiente: un cebador de transcripción inversa que tiene la secuencia de 5'-Cy5/CAG TAC GCA AGC TCG-3' ("Cy5s\_cebador15", SEQ ID NO: 22); una plantilla de control que tiene la secuencia de 5'-GCG TGA ACA TGC ATC TCC CGT ATG CGT ACA GTC CAT TGT ATA GCG CGA TAT ACA CAC TGG CGA GCT TGC GTA CTG-3' ("templ75", SEQ ID NO: 23); un cebador 5'-PCR que tiene la secuencia de 5'-GCG TGA ACA TGC ATC TCC-3' (SEQ ID NO: 24); y un cebador 3'-PCR que tiene la secuencia de 5'-CAG TAC GCA AGC TCG CC-3' (SEQ ID NO: 25), donde estas secuencias se obtuvieron a partir de ADN de IDT. Se usó un cebador de ADN marcado con Cy5 para los experimentos para permitir la detección separada de los productos de transcripción inversa por LC.

#### *Condiciones experimentales*

**[0163]** Para las ligaciones de par químicamente co-reactivas, soluciones 1 mM de pares químicamente co-reactivos, tales como 5extremo3propargil + 3extremo5azido (corto) o 5extremo3azida + 3extremo5hexinil (largo), se incubaron durante 12 horas en presencia de 100 equivalentes de ligando TBTA (tris-[(1-bencilo-1H-1,2,3-triazol-4-ilo)metilo]amina) y 50 equivalentes de CuBr en una mezcla de agua/acetato de dimetilo. Después de la reacción, se añadió un exceso de EDTA y las mezclas de reacción se desalaron usando columnas de desalación Zeba Spin (Invitrogen Corp., Carlsbad, CA) y luego se precipitó con etanol. Para las reacciones de transcripción inversa, las plantillas se purificaron en un gel de poliacrilamida al 15% que contenía urea 8M.

**[0164]** Se realizó cromatografía líquida-espectrometría de masas (CL-EM) en un Thermo Scientific LCQ Fleet usando una columna ACE 3 C18-300 (50 x 2,1 mm) y un gradiente de 5 minutos de 5-35% de tampón B usando tampón A (hexafluoroisopropanol al 1% (HFIP), diisopropiletilamina al 0,1% (DIEA), EDTA 10 µm en agua) y tampón B (HFIP al 0,075%, DIEA al 0,0375%, EDTA 10 µm, acetonitrilo al 65%/agua al 35%). La CL se controló a 260 nm y 650 nm. La EM se detectó en modo negativo y la desconvolución de los picos de masa se realizó utilizando el software ProMass.

**[0165]** Las reacciones de transcripción inversa se realizaron utilizando ThermoScript™ RT (Invitrogen Corp.), según el protocolo del fabricante, a 50°C durante 1-2 horas. Los resultados fueron analizados por CL-EM y por PCR. La PCR se realizó usando Platinum® SuperMix y se resolvió en E-Gels de agarosa al 4% (ambos de Invitrogen Corp.). Se realizaron once y dieciocho ciclos de PCR con o sin una reacción de TI previa. La plantilla de 75 mer no se transcribió de forma inversa y se utilizó directamente para la amplificación por PCR.

#### *Resultados y discusión*

**[0166]** En ambas ligaciones que forman un espaciador corto y un espaciador largo, rendimientos de la reacción eran altos, cerca de cuantitativos, tal como se analizó por CL-EM. Por consiguiente, la ligación química proporciona una técnica de alto rendimiento para unir o asociar operativamente una pieza de cabeza a una o más etiquetas de bloques de construcción.

**[0167]** Para que una estrategia de ligación química viable produzca bibliotecas de ADN codificada, el complejo resultante debe ser capaz de experimentar PCR o TI-PCR para otras aplicaciones de secuenciación. Si bien la PCR y la TI-PCR pueden no ser un problema con las etiquetas ligadas enzimáticamente, como se describió anteriormente, los enlazadores químicos no naturales pueden ser difíciles de procesar por las polimerasas de ARN o ADN. Los datos proporcionados en Figuras 11B-11E sugieren que los oligonucleótidos que tienen un espaciador de longitudes particulares se pueden transcribir y/o transcribir de forma inversa.

**[0168]** En el caso de un enlazador de par químicamente co-reactivo que resulta en un oligonucleótido enlazado por triazol, una dependencia se observó en la longitud del enlazador. Para el par corto químicamente co-reactivo, la plantilla resultante se transcribió de forma inversa y se analizó mediante CL-EM. El análisis de CL reveló tres picos de absorción principales a 2,79 min, 3,47 min y 3,62 min para 260 nm, donde los picos a 3,47 min y 3,62 min también proporcionaron picos de absorción a 650 nm. Análisis de EM del pico a los 3,47 min. mostró solo la presencia de la plantilla 23097,3 (calculado 23098,8), y el pico a los 3,62 min. contenía una plantilla (23098,0) y un cebador completamente extendido (23670,8, calculado: 23671,6) en una proporción de aproximadamente 1,7:1, lo que sugiere un rendimiento del 50-60% para esta reacción de TI (Figura 11C). A modo de comparación, la transcripción inversa (TI) del control que tiene una plantilla de ADN produjo el cebador extendido (pico 23068,9) en una cantidad aproximadamente equivalente a la plantilla (23078,7), lo que sugiere un rendimiento cercano al 100% (Figura 11B).

**[0169]** Para el par largo químicamente co-reactivo, LC de la reacción de TI mostró dos picos de absorción a 2,77 min y 3,43 min para 260 nm, donde el pico a 3,43 min también proporcionó picos de absorción a 650 nm, es decir, contenía un material etiquetado con Cy5, que es el producto de TI esperado. Análisis de EM del pico a los 3,43 min. reveló la plantilla (observado 23526,6, calculado: 23534,1), así como el cebador Cy5 extendido al enlazador (11569,1). No se

observó producto de longitud completa por CL-EM, lo que indica que la reacción de TI no se produjo en una cantidad medible (Figura 11D).

**[0170]** TI-PCR se realizó con las plantillas descritas anteriormente y reveló que sólo el enlazador corto produjo producto de transcripción inversa, aunque a 5-10 menor eficiencia (Figura 11E). Se estimó que la eficiencia de la TI era aproximadamente 2 veces menor que la plantilla (templ75). Por ejemplo, el producto de PCR de la plantilla ligada corta es aproximadamente 2 veces menor después de TI y aproximadamente 5-10 veces menor sin TI, en comparación con el producto de PCR de la plantilla 75 de todo el ADN (templ75). Por consiguiente, estos datos proporcionan soporte para el uso de la ligación química para producir un complejo que se puede transcribir y/o transcribir de forma inversa, y se pueden usar piezas de cabeza y/o etiquetas ligados químicamente en cualquiera de los pasos de unión descritos en este documento para producir bibliotecas codificadas.

## Ejemplo 6

### Ligación de oligonucleótidos de 3'-fosforotioato con oligonucleótidos de 5'-yodo

**[0171]** Para determinar la flexibilidad de los métodos descritos en el presente documento, se determinó la eficacia de ligación de oligonucleótidos que tienen otras modificaciones. En particular, los análogos del enlace fosfodiéster natural (p. ej., un análogo de fosforotioato) podrían proporcionar un resto alternativo para el análisis y secuenciación por PCR posterior a la selección.

**[0172]** Los siguientes oligonucleótidos fueron sintetizados por TriLink Biotechnologies, Inc. (San Diego, CA): (i) 5'-/Cy5/CGA TAT ACA CAC TGG CGA GCT/tiofosfato/-3' ("CCy5, SEQ ID NO: 26), (ii) 5'-IododT/GC GTA CTG AGC/6-FAM/-3' ("CFL," SEQ ID NO: 27), como se muestra en la Figura 12A, y (iii) un oligonucleótido de férula que tiene la secuencia de CAG TAC GCA AGC TCG CC ("spl", SEQ ID NO: 28). Las reacciones de ligación se realizaron con 100 µm de cada reactivo oligonucleótido en un tampón que contiene 50 mM de Tris HCl (pH 7,0), NaCl 100 mM, y MgCl<sub>2</sub> 10 mM ("tampón de ligación") a temperatura ambiente. Las reacciones de ligación se complementaron con cualquiera de los siguientes: 100 µm del oligonucleótido de férula, 10 mM de Co(NH<sub>3</sub>)<sub>6</sub>Cl<sub>3</sub>, 40% (p/v) de PEG4000 o 80% (p/v) de PEG300. Se dejó progresar la reacción durante hasta 48 horas. Los productos de ligación se analizaron por CL-EM usando detección a 260 nm, 495 nm y 650 nm, así como por un gel de urea 8 M/poliacrilamida al 15% (PAAG) que se escaneó adicionalmente a 450 y 635 nm de excitación en un Storm™ 800 PhosphorImager.

**[0173]** En ausencia del oligonucleótido de férula, no se observó la ligación (Figura 12B, carriles marcados como "spl"). En presencia del oligonucleótido de férula, se produjo la ligación y alcanzó aproximadamente el 60% de la fracción ligada después de 48 horas (Figuras 12B-12C). CL-EM reveló varios picos en el cromatograma, con un pico a los 3,00 min absorbiendo a 260 nm, 495 nm y 650 nm. La EM de este pico mostró principalmente el producto de ligación a 11539,6 Da (calculado 11540) con menos del 10% de oligonucleótido CCy5 a 7329,8 Da (calculado 7329,1). Se detectaron niveles bajos de ligación en presencia de PEG y hexamina cobalto, donde la hexamina cobalto provocó la precipitación del oligonucleótido marcado con Cy5. Estos datos sugieren que los cascos y/o etiquetas que tienen grupos fosfato modificados (p. ej., enlaces fosfodiéster modificados, tales como enlaces fosforotioato) se pueden usar en cualquiera de los pasos de unión descritos en este documento para producir bibliotecas codificadas.

**[0174]** Con el fin de estudio adicional la reacción de ligación yodo-fosforotioato, la ligación de 5'-I dT-oligo-3'-FAM (CFL) se realizó y 5'-Cy5-oligo-3' PS (CCy5) en ausencia y presencia de una férula en diferentes condiciones de reacción.

**[0175]** En un primer conjunto de condiciones, los experimentos de ligación se llevaron a cabo con una incubación de siete a ocho días. Estos experimentos se realizaron en el mismo tampón de ligación anterior con 50 µm de cada oligonucleótido y se incubaron durante una semana a temperatura ambiente. La Figura 12D muestra el análisis CL-EM de la ligación de CFL y CCy5 en ausencia (arriba) y presencia (abajo) de una férula (control positivo), donde las reacciones de ligación se incubaron durante siete días. Se registraron tres trazas de LC para cada reacción a 260 nm (para detectar todos los ácidos nucleicos), a 495 nm (para detectar el oligonucleótido CFL y el producto de ligación) y a 650 nm (para detectar el oligonucleótido CCy5 y el producto de ligación).

**[0176]** En ausencia de la férula, no se produjo la ligación, y sólo materiales de partida CFL (4339 Da) y CCy5 (7329 Da) se detectó (Figura 12D, arriba). Cuando el oligonucleótido de férula estuvo presente durante siete días, se observó un pico característico en el canal de 495 nm con un tiempo de retención de 2,98 min, que corresponde al producto ligado (11542 Da) (Figura 12D, parte inferior). Este pico se solapaba con el del oligonucleótido CCy5 observado en el canal de 650 nm y, por tanto, no se podía distinguir de CCy5 a 650 nm.

**[0177]** La Figura 12E muestra el análisis CL-EM de CFL y CCy5 en ausencia de una férula, donde las reacciones de ligación se incubaron durante ocho días a 400 µm de cada oligonucleótido. No se detectó producto de ligación. El pico 1 (a 495 nm) contenía material de partida CFL (4339 Da), así como trazas de la pérdida de producto de yodo (4211 Da) y un producto de degradación desconocido (4271 Da, posiblemente desplazamiento de etil mercaptano). El pico 2 (a 650 nm) contenía material de partida CCy5 (7329 Da) y oligonucleótido CCy5 oxidado (7317 Da). El pico 3 (a 650 nm) contenía CCy5 dimerizado (14663 Da).

**[0178]** En un segundo conjunto de condiciones, las reacciones de desplazamiento de yodo se llevaron a cabo en presencia de piperidina y a un pH mayor que 7,0. La Figura 12F muestra el análisis de MS para una reacción de oligonucleótido CFL con piperidina, donde se pretendía que esta reacción desplazara el yodo terminal presente en CFL. Una condición de reacción incluyó oligonucleótidos a 100  $\mu$ m, piperidina a 40 mM (400 equivalentes) en tampón borato 100 mM, pH 9,5, durante 20 horas a temperatura ambiente (datos mostrados en el panel izquierdo de la Figura 12F); y otra condición de reacción incluyó oligonucleótidos a 400  $\mu$ m, piperidina a 2 M (4.000 equivalentes) en tampón borato 200 mM, pH 9,5, durante 2 horas a 65°C (datos mostrados en el panel derecho de la Figura 12F).

**[0179]** En las condiciones de reacción incluyendo 40 mM de piperidina (Figura 12F, izquierda), no se observó desplazamiento de piperidina, y se detectó una pequeña cantidad de producto de hidrólisis (4229 Da). Además, se observaron trazas de la pérdida de yodo (4211 Da) y producto de degradación desconocido (4271 Da). En la condición de reacción que incluía 2 M de piperidina (Figura 12F, derecha), se observó el desplazamiento de piperidina del yodo (4296 Da) y la cantidad de material de partida disminuyó sustancialmente (4339 Da). Además, también se observaron picos correspondientes a la hidrólisis de yodo (por desplazamiento de OH) o impureza (4229 Da) y pérdida de yodo (4214 Da). Estos datos muestran que la presencia de una amina (p. ej., como parte de la síntesis de la biblioteca química) no afectará negativamente a la porción de oligonucleótidos de los miembros de la biblioteca y/o interferirá con esta estrategia de ligación.

**[0180]** En un tercer conjunto de condiciones, reacciones de ligación de férula se llevaron a cabo en presencia de piperidina y a un pH mayor que 7,0. La Figura 12G muestra una reacción de ligación en férula de oligonucleótidos CFL y CCy5 a 50  $\mu$ m realizada en presencia de 400 equivalentes de piperidina en tampón borato 100 mM, pH 9,5, durante 20 horas a temperatura ambiente. El pico característico detectado en la traza de LC (a 495 nm) contenía predominantemente el producto de ligación a 11541,3 Da (calculado 11540 Da). En base a estos resultados, se puede concluir que la piperidina no altera la ligación enzimática y que la presencia de otras aminas (p. ej., como parte de la síntesis de bibliotecas químicas) probablemente no interferirá con esta estrategia de ligación.

**[0181]** Considerados en conjunto, estos datos indican que esta estrategia de ligación se puede realizar bajo diferentes condiciones de reacción que son adecuadas para una amplia gama de transformaciones químicas, incluyendo tiempos de incubación extendidos, condiciones de pH elevado y/o presencia de una o más aminas. Por tanto, los presentes métodos pueden ser útiles para desarrollar miembros de bibliotecas con diversas condiciones de reacción y excluir la necesidad de intercambio de tampón, tal como precipitación u otros métodos intensivos en recursos.

## Ejemplo 7

### Minimización de arrastrar los pies con nucleótidos modificados

**[0182]** Durante la ligación enzimática de una sola hebra con ligasa T4 ARN, puede ocurrir la reordenación baja a moderada de nucleótido terminal. La mezcla aleatoria puede resultar en la inclusión o escisión de un nucleótido, donde el producto o complejo final incluye o excluye un nucleótido en comparación con la secuencia ligada esperada (es decir, una secuencia que tiene la secuencia completa para los oligonucleótidos aceptores y donantes).

**[0183]** Aunque se pueden tolerar niveles bajos de reorganización, la reordenación se puede minimizar incluyendo un grupo fosfato modificado. En particular, el grupo fosfato modificado es un enlace fosforotioato entre el nucleótido terminal en el extremo 3' de un oligonucleótido aceptor y el nucleótido adyacente al nucleótido terminal. Mediante el uso de un enlace fosforotioato de este tipo, la mezcla se redujo en gran medida. Solo se detectó una mezcla residual mediante espectrometría de masas, donde la mezcla probablemente surgió debido a la conversión incompleta del enlace fosfodiéster nativo en el enlace fosforotioato o a bajos niveles de oxidación del enlace fosforotioato seguido de conversión en el enlace fosfodiéster nativo. En conjunto estos datos y los datos de ligación en el Ejemplo 6, uno o más grupos fosfato modificados (p. ej., un fosforotioato o un enlace 5'-N-fosforamidita) podrían incluirse en cualquier secuencia de oligonucleótidos descrita en este documento (p. ej., entre el nucleótido terminal en el extremo 3' de una pieza de cabeza, un complejo, una etiqueta de bloque de construcción o cualquier etiqueta descrita en el presente documento, y el nucleótido adyacente al nucleótido terminal) para minimizar la mezcla durante la ligación de una sola hebra.

**[0184]** Una pieza de cabeza de hebra única (PSSS, 3636 Da) se fosforiló en el extremo 5' y modificado con un enlazador hexilamina en el extremo 3' para proporcionar la secuencia de 5'-P-mCGAGTCACGTC/Aminohex/-3' (SEQ ID NO: 29). La pieza de cabeza se ligó a una etiqueta (etiqueta 15, XTAGSS000015, 2469 Da) que tiene la secuencia de 5'-mCAGTGTCmA-3' (SEQ ID NO: 30), donde mC y mA indican nucleótidos de metilo 2'-O. El análisis CL-EM (Figura 13A) reveló que el pico del producto de ligación contenía hasta tres especies, que se separaron parcialmente por LC y tenían los siguientes pesos moleculares: 6089 Da (esperado), 5769 Da (-320 Da del esperado) y 6409 Da (+320 Da del esperado). Esta diferencia de masa de 320 Da corresponde exactamente a la eliminación o adición de un nucleótido O-Me C adicional ("reordenación terminal de nucleótidos").

**[0185]** Los experimentos con otros nucleótidos O-Me terminales, así como con nucleótidos 2'-fluoro terminales, confirmaron que la mezcla se produce probablemente por escisión del nucleótido de extremo 5' del oligonucleótido

donante, probablemente después de la adenilación del último. Se desconoce el mecanismo de este evento. Sin estar limitado por el mecanismo, la Figura 13B ilustra un posible esquema para la reordenación de nucleótidos durante la reacción de ligasa T4 ARN entre una pieza de cabeza y una etiqueta, donde un experto en la técnica entendería que esta reacción podría ocurrir entre cualquier oligonucleótido donante y aceptor (p. ej., entre dos etiquetas, donde una etiqueta es el oligonucleótido donante y la otra etiqueta es el oligonucleótido aceptor).

**[0186]** En general, la mayoría de la reacción de ligación con ligasa T4 ARN (T4Rnl1) proporciona el producto de ligación esperado (normal) que tiene la secuencia combinada de los oligonucleótidos tanto de donante como aceptor (Figura 13B-1, la reacción de la izquierda). Una pequeña minoría de la reacción proporciona productos de ligación aberrantes (Figura 13B-1, reacción a la derecha), donde estos productos aberrantes incluyen aquellos que tienen la eliminación o adición de un nucleótido terminal ("Producto -1 nt" y "Producto +1 nt", respectivamente, en la Figura 13B-2).

**[0187]** Sin estar limitado por el mecanismo, la escisión del oligonucleótido donante ("pieza de cabeza" o "HP" en la Figura 13B-1) puede ocurrir al reaccionar con el grupo 3'-OH del aceptor ("etiqueta"), proporcionando así un donante 5'-fosforilado que carece de un nucleótido ("HP-1 nt") y un nucleótido adenilado con un grupo 3'-OH accesible ("1 nt"). La Figura 13B-2 muestra dos esquemas de ejemplo para la reacción entre la pieza de cabeza (HP), etiqueta, HP-1 nt y 1 nt. Para proporcionar un producto con un nucleótido terminal escindido (Figura 13B-2, izquierda), el donante 5' fosforilado que carece de un nucleótido (HP-1 nt) actúa como sustrato para el evento de ligación. Esta pieza de cabeza de HP-1 nt se vuelve a adenilar mediante ligasa T4 ARN (para proporcionar "HP-1 nt adenilado" en la Figura 13B-2) y se liga a la etiqueta, lo que da como resultado un producto de ligación menos un nucleótido ("Producto-1 nt"). Para proporcionar un producto con un nucleótido terminal adicional (Figura 13B-2, izquierda), el nucleótido adenilado (1 nt) probablemente sirve como sustrato para la ligación a la etiqueta, produciendo así un oligonucleótido que tiene un nucleótido más largo que el aceptor ("Etiqueta +1 nt"). Este oligonucleótido de etiqueta +1 nt probablemente sirve como aceptor para la pieza de cabeza inalterada, donde esta reacción proporciona un producto de ligación que tiene un nucleótido adicional ("Producto +1 nt"). Se realizaron análisis de CL-EM de "Producto", "Producto-1 nt" y "Producto +1 nt" (Figura 13B-3). Cuando una etiqueta aberrante y una pieza de cabeza aberrante (es decir, Etiqueta +1 nt y HP-1 nt, respectivamente) se recombinan, entonces el producto de ligación resultante es indistinguible del producto esperado.

**[0188]** Para estudiar adicionalmente el mecanismo de reordenación de nucleótido terminal, se preparó una pieza de cabeza (HP-PS) que tiene la secuencia de 5'-mC\*GAGTCACGTC/Aminohex/-3' (SEQ ID NO: 31). La pieza de cabeza HP-PS tiene la misma secuencia que ssHP pero contiene una modificación, a saber, el primer enlace fosfodiéster entre el nucleótido 5' terminal mC y el siguiente G se sintetizó como un enlace fosforotioato (un oxígeno fosfato no puente se sustituyó por un azufre). El análisis de CL-EM de la ligación de HP-PS a la etiqueta 15 reveló que la reordenación estaba casi completamente inhibida (Figura 13C). Las trazas de +/- 320 picos probablemente corresponden a la conversión oxidativa del enlace fosforotioato en enlaces fosfodiéster nativos o sulfuración incompleta.

## Ejemplo 8

### Cromatografía de exclusión de miembros de la biblioteca

**[0189]** Las bibliotecas de entidades químicas que se generan utilizando oligonucleótidos monocatenarios cortos como que elementos codificadores son muy adecuadas para el enriquecimiento de los ligantes a través de cromatografía de exclusión por tamaño (SEC). SEC es una técnica cromatográfica que separa moléculas en función del tamaño, donde las moléculas más grandes que tienen un peso molecular más alto fluyen a través de la columna más rápida que las moléculas más pequeñas que tienen un peso molecular más bajo.

**[0190]** Los complejos de proteínas y los miembros de la biblioteca de ssADN se pueden separar fácilmente de los miembros de la biblioteca no unidos usando SEC. La Figura 14 es una traza ultravioleta de un experimento SEC en donde una pequeña molécula unida covalentemente a ssADN corto (un rango de oligonucleótidos con longitudes definidas en el rango de 20-50 mer) se mezcló con una proteína diana que se sabe que se une a la molécula pequeña. Los picos que eluyen primero de la columna, en el intervalo de tiempo de 11 a 13 minutos, representan miembros de la biblioteca asociados a la diana. Los picos posteriores, que eluyen entre 14 y 17 minutos, representan miembros de la biblioteca no unidos. La proporción de proteína diana a molécula de biblioteca fue 2:1, por lo que aproximadamente el 50% de las moléculas de biblioteca deberían asociarse con la proteína en la fracción de elución temprana, como se observa en la Figura 14. Las bibliotecas con regiones codificantes de oligonucleótidos bicatenarios más grandes no pueden seleccionarse utilizando este método ya que los miembros de la biblioteca no vinculados co-migran con los miembros de la biblioteca vinculados en SEC. Por lo tanto, las bibliotecas de moléculas pequeñas unidas a la codificación de oligonucleótidos monocatenarios en el rango de longitud de 20 a 50 mer permiten el uso de una poderosa técnica de separación que tiene el potencial de aumentar significativamente la relación señal-ruido requerida para la selección efectiva de ligantes de moléculas pequeñas a una o más dianas, p. ej., nuevas dianas proteicas que están opcionalmente sin marcar y/o proteínas de tipo silvestre. En particular, estos enfoques permiten identificar entidades químicas que se unen a la diana en bibliotecas codificadas generadas combinatoriamente sin la necesidad de marcar o inmovilizar la diana (p. ej., una diana de proteína).

**Ejemplo 9****Codificación con las etiquetas de ADN ligadas químicamente utilizando la misma química para cada etapa de ligación**

[0191] Etiquetas de ADN codificadoras se pueden ligar enzimáticamente o químicamente. En la Figura 15A se ilustra un enfoque general para la ligación química de etiquetas de ADN. Cada etiqueta tiene grupos reactivos co-complementarios en sus extremos 5' y 3'. Para evitar la polimerización o ciclación de las etiquetas, (i) protección de uno o ambos grupos reactivos (Figura 15A), p. ej., en el caso de alquinos 3' protegidos con TIPS, o (ii) química de ligación dependiente de la férula (Figura 15B), p. ej., en el caso de la ligación de 5'-yodo/3'-fosforotioato. Para (i), las etiquetas no ligadas se pueden eliminar o tapar después de cada ciclo de biblioteca para evitar el error o la polimerización de la etiqueta desprotegida. Este paso puede ser opcional para (ii), pero aún puede incluirse. Las reacciones de extensión del cebador, utilizando enzimas polimerasas que son capaces de leer a través de uniones ligadas químicamente, también se pueden realizar para demostrar que las etiquetas ligadas son legibles y, por lo tanto, la información codificada se puede recuperar mediante la amplificación y secuenciación posteriores a la selección (Figura 15C).

[0192] Una estrategia de etiquetado de biblioteca que implementa la ligación de las etiquetas usando cicloadición azida/alquino catalizada por "química de clic" (Cu(I)) se muestra en la Figura 16A. La implementación de esta estrategia se basa en la capacidad de ligación sucesiva precisa de las etiquetas, evitando errores y polimerizaciones de etiquetas, así como en la capacidad de copiar el ADN ligado químicamente en ADN natural amplificable (ADNc) para la amplificación y secuenciación posteriores a la selección (Figura 16C).

[0193] Para lograr nucleótidos 3' propargilo precisos protegidos por (TIPS) de triisopropilsililo de ligación de etiqueta, (sintetizados a partir de propargilo U en la forma de una matriz CPG utilizado para la síntesis de oligonucleótidos) se utilizó (Figura 16B). El grupo protector TIPS se puede eliminar específicamente mediante tratamiento con fluoruro de tetrabutilamonio (TBAF) en DMF a 60°C durante 1-4 horas. Como resultado, la ligación durante la síntesis de la biblioteca incluye un nucleótido 5'-azido/3'-TIPS-propargilo (Etiqueta A) que reacciona con el 3'-propargilo de la pieza de cabeza mediante una reacción de clic. Después de la purificación, el ciclo anterior se trata con TBAF para eliminar los TIPS y generar el alquino reactivo que a su vez reacciona con la etiqueta del siguiente ciclo. El procedimiento se repite durante tantos ciclos como sea necesario para producir 2, 3 o 4 o más etiquetas de codificación instaladas sucesivamente (Figura 16A).

**Materiales y métodos**

[0194] Oligos: Los siguientes oligos fueron sintetizados por Trilink Biotechnologies, San Diego CA: ss-HP-alquino: 5'-NH<sub>2</sub>-TCG AAT GAC TCC GAT AT (3'-Propargil G)-3' (SEQ ID NO: 32); ss-azido-TP: 5'-azido dT ATA GCG CGA TAT ACA CAC TGG CGA GCT TGC GTA CTG -3' (SEQ ID NO: 33); y B-azido: 5' azido dT ACA CAC TGG CGA GCT TGC GTA CTG -3' (SEQ ID NO: 34).

[0195] Etiqueta clic-TIPS: 5'-azido dT AT GCG TAC AGT CC (propargil U-TIPS)-3' (SEQ ID NO: 35) y 5'-Dimetoxitritil 2'-succinil 3'-O-(triisopropil sililo) Las cpg de propargiluridina fueron sintetizadas por Prime Organics, Woburn MA.

[0196] Los siguientes oligos se sintetizaron mediante tecnologías de ADN IDT, Coralville, IA: FAM-clic-cebador: (5'-6-FAM) CAG TAC GCA AGC TCG CC -3' (SEQ ID NO: 36) y Cy5-clic-cebador: (5'-Cy5) CAG TAC GCA AGC TCG CC -3' (SEQ ID NO: 37).

[0197] ADN55-control: 5'-Biotin-TEG//ispC3//ispC3/-TCGAATGACTCCGATATGT ATA GCG CGA TAT ACA CAC TGG CGA GCT TGC GTA CTG -3' (SEQ ID NO: 38).

[0198] rADN55-control: 5'-Bio-TEG//ispC3//ispC3/-TCGAATGACTCCGATAT(riboG)T ATA GCG CGA TAT ACA CAC TGG CGA GCT TGC GTA CTG -3' (SEQ ID NO: 39)

[0199] Síntesis de las plantillas: En los siguientes ejemplos, la frase "etiquetas ligadas químicamente", o secuencias de control relacionadas con ellas, se denominan "plantillas" porque el paso posterior ("lectura") las utiliza como plantillas para la polimerización dependiente de la plantilla.

[0200] Ligación de etiquetas: A una solución de 1 equivalente (1 mM) de PSSS-alquino y 1 equivalente (1 mM) de ss-azidoTP en mM pH tampón 7,0 de fosfato de 500, se añadió una solución de pre-mezclado 2 eq de acetato de Cu (II) (hasta una concentración final de 2 mM), 4 eq de ascorbato de sodio (hasta una concentración final de 4 mM), 1 eq de TBTA (hasta una concentración final de 1 mM) en DMF/agua. La mezcla se incubó a temperatura ambiente durante la noche. Después de la confirmación por CL-EM de la finalización de la reacción, la reacción se precipitó usando sal/etanol.

**[0201]** Plantillas de "clic único" Y55 y Y185 se sintetizaron mediante la reacción de ss-HP-alquino con ss-azido-TP y B-azido, respectivamente. Las plantillas de doble y triple clic (YDC y YTC) se sintetizaron mediante ligación de clic de ss-HPalquino con etiqueta clic-TIPS, seguida de desprotección de TIPS usando TBAF (fluoruro de tetrabutilamonio) en DMF a 60°C durante una hora, seguido de ligación de clic con ss-azido TP. Para la plantilla de triple clic (YTC), la ligación y la desprotección de etiqueta clic TIPS se repitieron dos veces.

**[0202]** Se hicieron reaccionar las plantillas con biotina (EG)<sub>4</sub>-NHS y se desalaron (Figura 17A). Los productos finales se purificaron mediante RP HPLC y/o en un gel de poli(acrilamida) al 15-20%/urea 8M y se analizaron por CL-EM.

**[0203]** Enzimas: Las siguientes polimerasas de ADN con sus tampones de reacción se adquirieron de New England Biolabs: fragmento Klenow de la polimerasa de ADN I de E. coli, fragmento Klenow (exo-), polimerasa de ADN I de E. coli, Therminator™, 9°N™, Superscript III™.

**[0204]** Las Dynabeads® M280 de estreptavidina magnética se adquirieron de Invitrogen.

**[0205]** Evaluación de polimerización dependiente de plantilla: Cada plantilla (5 µm) se incubó con 1 equivalente de cualquiera de Cy5 o FAM Clic-cebador en 40 a 50 µL del correspondiente tampón 1x de reacción y cada enzima, utilizando la reacción en condiciones de acuerdo con las pautas del fabricante durante 1 hora. Ciertas reacciones (como SSII o transcripciones SSIII) se complementaron adicionalmente con 1 mM MnCl<sub>2</sub>. El producto de la reacción se cargó en 125 µl de cuentas SA prelavadas durante 30 minutos con agitación. A continuación, se recogieron las cuentas y se descartó el flujo a través. Las cuentas se lavaron con 1 ml de solución salina tamponada con Tris (pH 7,0) y se eluyeron con 35 µl de NaOH 100 mM. El eluido se neutralizó inmediatamente añadiendo 10 µl de 1 M Tris HCl, pH 7,0. Los productos se analizaron mediante CL-EM.

## Resultados y discusión

**[0206]** Preparación de plantilla: Cada plantilla, Y55, Y185 (Figuras 17B y 17C), YDC y YTC (Figura 19) se sintetizaron y se purificaron hasta más de 85% de pureza (la principal impureza es plantilla no biotinilada). CL-EM reveló los siguientes PM para las plantillas: Y55 17.624 (calculado 17.619) Da; YDC 22.228 (calculado 22.228) Da; y YTC 26.832 (calculado 26.837) Da.

**[0207]** Las plantillas Y55 y Y185 de clic único (Figuras 17B y 17C) se sintetizaron a partir de oligonucleótidos que llevan sólo una funcionalidad química modular (alquino o azida). La eficacia de la reacción de clic (ligación química) fue superior al 90% en una reacción durante la noche utilizando catalizador de Cu(I) generado *in situ*.

**[0208]** Las plantillas YDC y YTC (Figuras 19A-19D) sirven para demostrar ligaciones químicas sucesivas. Tanto YDC como YTC utilizan etiquetas individuales que contienen simultáneamente funcionalidades de alquino protegidas con azido y con TIPS. La plantilla YTC demuestra tres ciclos sucesivos de marcado que pueden usarse para codificar tres pasos de generación de bibliotecas químicas.

**[0209]** Se probaron todas las plantillas anteriores para determinar la extensión del cebador a través y más allá de los enlaces de ligación de clic para demostrar que las etiquetas ligadas son legibles y, por lo tanto, que la información codificada es recuperable.

**[0210]** *Polimerización dependiente de la plantilla usando la plantilla Y55 de "un solo clic"*: se probó un gran conjunto de polimerasas para leer a través de un enlace de clic de triazol (Figura 18A). Los experimentos iniciales se realizaron utilizando Cy5-clic-cebador. En experimentos posteriores se usó FAM-clic-cebador. El fluoróforo no tuvo ningún efecto sobre la copia de la plantilla, es decir, los resultados fueron equivalentes usando cualquier cebador. Como plantilla de control se utilizaron el control ADN55 y el control rADN55 (para probar el efecto de un solo ribonucleótido en la plantilla, ya que propargil-G se utilizó para la ligación de clic es un derivado de ribonucleótido).

**[0211]** Productos de longitud completa que se esperan de las tres plantillas tienen el mismo peso molecular, que es 17446 (cebador FAM) (Figura 18B) o 17.443 (cebador Cy5). También se observó una pequeña cantidad del producto que corresponde a la extensión del cebador hasta, pero deteniéndose en, el enlace de ligación de clic (11880 Da) para algunas polimerasas.

**[0212]** Un conjunto de polimerasas que pueden producir grado sustancial de lectura a través de la articulación de clic (producción de ADNc de longitud completa) se descubrieron y se tabularon a continuación.

Rendimientos de ADNc de longitud completa de más del 50%
Fragmento de Klenow de la polimerasa de ADN de <i>E. coli</i> I
Fragmento de Klenow (exo-)
Polimerasa de ADN de <i>E. coli</i> I
Therminator™
9°N™
Superscript III™ suplementado con 1 mM de MnCl <sub>2</sub>

- 5
- 10 **[0213]** Los rendimientos más altos se lograron utilizando fragmento Klenow (más del 80% de lectura a través de un solo cruce de clic) con incubación a 37°C (Figura 18B). Se observó un rendimiento algo menor usando la polimerasa de ADN I de *E. coli*. Se lograron rendimientos del 50% con Therminator™ y polimerasas 9°N™, así como con el fragmento Klenow exo-.
- 15 **[0214]** La transcriptasa inversa SuperScript III™ produjo aproximadamente 50% de rendimiento de cADN cuando el tampón se suplementó con 1 mM MnCl<sub>2</sub>. Sin embargo, el manganeso provocó la incorporación errónea de nucleótidos que se observó por MS, es decir, se redujo la fidelidad de polimerización.
- 20 **[0215]** *Polimerización dependiente de la plantilla usando la plantilla Y185 de "un solo clic"*: La plantilla Y185 presenta el mismo sitio de unión del cebador que todas las plantillas utilizadas en este ejemplo, excepto, debido a una cola B-azido diferente, la distancia entre el último nucleótido del sitio de unión del cebador al enlace de clic es de 8 nucleótidos, en comparación con 20 nucleótidos en Y55 y todas las demás plantillas. La plantilla se usó para probar si la transcripción de un enlace clic todavía era posible cuando la enzima estaba en conformación de iniciación-alargamiento temprano. Klenow fue capaz de copiar la plantilla Y185 con una eficacia similar a la de Y55, lo que abre
- 25 la posibilidad de reducir la longitud de las etiquetas codificantes ligadas por clic (Figura 18C).
- [0216]** *Polimerización dependiente de la plantilla usando plantillas YDC y YTC ligadas con clic doble y triple*: Después de establecer que el fragmento de Klenow era la enzima más eficiente para leer los enlaces de ligación con clic en las condiciones de ensayo empleadas, también se generaron las ADNc usando plantillas YDC y YTC (Figuras 20A-20C).
- 30 Las reacciones de extensión del cebador con las plantillas YDC y YTC produjeron productos completos. Otros productos observados, que componían alrededor del 10-15% de la producción total de la reacción, correspondían al cebador parcialmente extendido, estancado en cada unión de clic, como por ejemplo, 11880 Da y 16236 Da. Los rendimientos se midieron mediante análisis CL-EM en presencia del estándar interno y fueron alrededor del 80-90% por unión (es decir, alrededor del 85% para 1 clic, 55% para 2 clics y 50% para plantillas de 3 clics, ver Figura 21).
- 35 **[0217]** El producto de transcripción de YDC carecía 1 dA de nucleótidos (calculado 22110, observado 27197 Da; - 313 dA Figura 20B) y el producto de la transcripción YTC carecía de 2 nucleótidos dA (calculado 26773, observado 26147; - 626 2xdA) (Figura 20C). Esto se correlaciona con el número de nucleótidos de propargil U en la plantilla. Sin desear estar limitado por el mecanismo, se puede plantear la hipótesis de que Klenow se saltó esas U en el contexto de la unión T-triazol-U. Por el contrario, el nucleótido propargil G en la unión del primer clic se copió correctamente.
- 40

## Ejemplo 10

### Uso de etiquetas de 3'-fosforotioato/5'-yodo para ligar químicamente una sucesión de etiquetas de ADN codificantes que codifican una biblioteca química instalada covalentemente en el extremo 5'

45

- [0218]** *Protección de 3'-fosforotioato en la etiqueta*: como se muestra en la Figura 24A, se disolvió una etiqueta de 5'-yodo-3'-fosforotioato (1 eq.) en agua para dar una concentración final de 5 mM. Posteriormente, se añadió vinil metilo sulfona (20 eq.) y la reacción se incubó a temperatura ambiente durante la noche. Una vez completada la reacción, el
- 50 producto se precipitó con etanol.

#### Síntesis de la biblioteca (Figura 24B)

- [0219]** *Ciclo A*: Para cada pocillo en la división se añadió pieza de cabeza de ADN monocatenario (1 eq., solución 1 mM en 500 mM tampón de pH 9,5 de borato), una etiqueta protegida de ciclo A (1,5 eq.) y férula (1,2 eq.). La ligación química se incubó a temperatura ambiente durante la noche. A cada pocillo (en la división) se le añadió un aminoácido Fmoc (100 eq.), seguido de cloruro de 4-(4,6-dimetoxi-1,3,5-triazin-2-ilo)-4-metilmorfolinio (100 eq.). La reacción química se incubó a temperatura ambiente durante la noche. Una vez completada, se reunieron todos los pocillos y los productos se precipitaron usando etanol. El grupo del ciclo A se purificó usando CL y se liofilizó hasta sequedad,
- 60 y luego se disolvió en agua para dar una concentración final de 1 mM y se añadió piperidina (10% v/v) para realizar la desprotección de la etiqueta del ciclo A (60°C, 2 h). El producto desprotegido se precipitó de nuevo con etanol.

- [0220]** *Ciclo B*: El grupo de ciclo A desprotegido se disolvió en 500 mM, pH 9,5, borato de tampón para dar una concentración 1 mM y luego dividido en pocillos de reacción separados (1 eq. de producto de ciclo A en cada pocillo).
- 65 A cada pocillo se le añadió una etiqueta protegida del ciclo B (1,5 eq.) y una férula (1,2 eq.). La ligación química se incubó a temperatura ambiente durante la noche. A cada pocillo (en la división) se le añadió una mezcla de un ácido



de formilo (100 eq.), diisopropil carbodiimida (100 eq.) y 1-hidroxi-7-aza-benzotriazol (100 eq.). La reacción química se incubó a temperatura ambiente durante la noche. Una vez completada, se reunieron todos los pocillos y los productos se precipitaron usando etanol. El grupo del ciclo B se purificó usando CL y se liofilizó hasta sequedad, y luego se disolvió en agua para dar una concentración final de 1 mM y se añadió piperidina (10% v/v) para realizar la desprotección de la etiqueta del ciclo B (60°C, 2 h). El producto desprotegido se precipitó de nuevo con etanol.

**[0221]** *Ciclo C:* El grupo de ciclo B desprotegido se disolvió en tampón de fosfato 500 mM pH 5,5 para dar un 1 mM de concentración y luego se dividió en pocillos de reacción separados (1 eq de producto de ciclo B en cada pocillo). A cada pocillo se le añadió una etiqueta C de ciclo (1,5 eq.) y una férula (1,2 eq.). La ligación química se incubó a temperatura ambiente durante la noche. A cada pocillo (en la división) se le añadió una amina (80 eq.) y cianoborohidruro de sodio (80 eq.). La reacción química se incubó a 60°C durante 16 h. Una vez completada, se reunieron todos los pocillos y los productos se precipitaron usando etanol. El conjunto del ciclo C se purificó usando CL y se liofilizó hasta sequedad.

## Ejemplo 11

### Codificación con las etiquetas de ADN ligadas químicamente usando un par de químicas ortogonales para cada etapa de ligación de etiqueta sucesiva

**[0222]** Otro enfoque para la generación de etiquetas de ADN que codifican ligados químicamente es el uso de un par de ortogonales químicas para ligaciones sucesivas (Figura 22A). Las etiquetas que llevan grupos reactivos ortogonales en sus extremos no polimerizarán ni ciclarán las etiquetas, y la naturaleza ortogonal de los pasos de ligación sucesivos reducirá la frecuencia de los eventos de confusión. Tales enfoques requieren (i) tener al menos dos químicas ortogonales disponibles para la conjugación de oligonucleótidos, y (ii) una estrategia de lectura completa disponible para cada una de las uniones así creadas (Figuras 22B y 22C). Este enfoque también puede obviar la necesidad del uso de grupos de protección o etapas de protección, simplificando así el proceso de ligación de etiquetas.

**[0223]** *Estrategia de ligación química ortogonal que utiliza ligación de 5'-Azido/3'-Alquinilo y 5'-Yodo/3'-Fosforotioato para etapas sucesivas:* Un ejemplo del uso de ligación de etiquetas de dos químicas ortogonales es la combinación de ligaciones 5'-azido/3'-alquinilo y 5'-yodo/3'-fosforotioato. La Figura 23 muestra un esquema ejemplar de la síntesis de una estrategia de etiquetado de ligación química ortogonal de 3 ciclos usando estas químicas de ligación sucesivas. Las Figuras 25A-25B muestran un ejemplo del uso de etiquetas 3'-fosforotioato/5'-azido y 3'-propargil/5'-yodo para ligar químicamente una sucesión de etiquetas de ADN codificantes ortogonales que codifican una biblioteca química instalada covalentemente sobre el extremo 5'.

**[0224]** *Protección de 3'-fosforotioato en las etiquetas:* Como se muestra en la Figura 25A, una etiqueta 5'-azido-3'-fosforotioato (1 eq.) Se disolvió en agua para dar una concentración final de 5 mM. Posteriormente, se añadió vinil metilo sulfona (20 eq.) y la reacción se incubó a temperatura ambiente durante la noche. Una vez completada la reacción, el producto se precipitó con etanol.

### Síntesis de la biblioteca (Figura 25B)

**[0225]** *Ciclo A:* Para cada pocillo en la división se añadió la pieza de cabeza de ADN de hebra única (1 eq., solución 1 mM en 500 mM tampón pH 9,5 de borato), una etiqueta de ciclo A (1,5 eq) y férula (1,2 eq.). La ligación química se incubó a temperatura ambiente durante la noche. A cada pocillo (en la división) se le añadió un aminoácido Fmoc (100 eq.), seguido de cloruro de 4-(4,6-dimetoxi-1,3,5-triazin-2-ilo)-4-metilmorfolinio (100 eq.). La reacción química se incubó a temperatura ambiente durante la noche. Una vez completada, se reunieron todos los pocillos y los productos se precipitaron usando etanol. El grupo del ciclo A se purificó usando CL y se liofilizó hasta sequedad. La desprotección de Fmoc se realizó en el grupo del ciclo A tratando la piscina (1 mM en agua) con piperidina (10% v/v) durante 2 h a temperatura ambiente. El producto desprotegido se precipitó de nuevo con etanol.

**[0226]** *Ciclo B:* El ciclo A purificado se disolvió en tampón fosfato 500 mM, pH 7,0 para dar una concentración de 1 mM y luego se dividió en pocillos de reacción separados (1 eq. de producto del ciclo A en cada pocillo). A cada pocillo se le añadió una etiqueta protegida del ciclo B (1,2 eq.), acetato de cobre (II) (2 eq.), ascorbato de sodio (4 eq.) y tris-(benciltriazolilmetilo)amina (1 eq.). La ligación química se incubó a temperatura ambiente durante la noche. Una vez completados, los productos se precipitaron (en la división) usando etanol y luego se diluyeron a una concentración de 1 mM usando tampón borato 500 mM, pH 9,5. A cada pocillo (en la división) se le añadió una mezcla de un ácido de formilo (100 eq.), diisopropil carbodiimida (100 eq.) y 1-hidroxi-7-aza-benzotriazol (100 eq.). La reacción química se incubó a temperatura ambiente durante la noche. Una vez completada, se reunieron todos los pocillos y los productos se precipitaron usando etanol. El conjunto del ciclo B se disolvió luego en agua para dar una concentración final 1 mM, y se añadió piperidina (10% v/v) para realizar la desprotección de la etiqueta del ciclo B (temperatura ambiente, 18 h). El producto desprotegido se precipitó de nuevo con etanol. El conjunto del ciclo B desprotegido se purificó usando CL y se liofilizó a sequedad.

**[0227]** *Ciclo C:* El grupo de ciclo B purificado se disolvió en 500 mM, tampón fosfato pH 5,5 para dar una concentración 1 mM y luego dividido en pocillos de reacción separados (1 eq de producto de ciclo B en cada pocillo). A cada pocillo

se le añadió una etiqueta C de ciclo (1,5 eq.) y una férula (1,2 eq.). La ligación química se incubó a temperatura ambiente durante la noche. A cada pocillo (en la división) se le añadió una amina (80 eq.) y cianoborohidruro de sodio (80 eq.). La reacción química se incubó a 60°C durante 16 h. Una vez completada, se reunieron todos los pocillos y los productos se precipitaron usando etanol. El conjunto del ciclo C se purificó usando CL y se liofilizó hasta sequedad.

5

## LISTA DE SECUENCIAS

**[0228]**

10

&lt;110&gt; X-CHEM, INC.

&lt;120&gt; MÉTODOS PARA MARCAR BIBLIOTECAS CODIFICADAS POR ADN

&lt;130&gt; C69049PCEPT1

15

&lt;150&gt; 61/536,929

&lt;151&gt; 2011-09-20

&lt;150&gt; 61/531,820

20

&lt;151&gt; 2011-09-07

&lt;160&gt; 39

&lt;170&gt; Patente en versión 3,5

25

&lt;210&gt; 1

&lt;211&gt; 7

&lt;212&gt; ADN

&lt;213&gt; Secuencia artificial

30

&lt;220&gt;

&lt;223&gt; Constructo sintético

&lt;400&gt; 1

35

gtgctgc 7

&lt;210&gt; 2

&lt;211&gt; 9

&lt;212&gt; ADN

40

&lt;213&gt; Secuencia artificial

&lt;220&gt;

&lt;223&gt; Constructo sintético

45

&lt;400&gt; 2

gcagcaccc 9

&lt;210&gt; 3

&lt;211&gt; 8

&lt;212&gt; PRT

50

&lt;213&gt; Secuencia artificial

&lt;220&gt;

&lt;223&gt; Constructo sintético

55

&lt;220&gt;

&lt;221&gt; MISC\_FEATURE

&lt;222&gt; (1)..(1)

&lt;223&gt; Xaa es N-alfa-benciloxycarbonilo

60

&lt;220&gt;

&lt;221&gt; MISC\_FEATURE

&lt;222&gt; (8)..(8)

&lt;223&gt; Xaa es OSu (éster N-succinimidílico)

65

&lt;400&gt; 3

Xaa Gly Gly Gly Gly Gly Gly Xaa  
1 5

5	<210> 4 <211> 5 <212> PRT <213> Secuencia artificial
10	<220> <223> Constructo sintético
15	<220> <221> MISC_FEATURE <222> (1)..(1) <223> Xaa es azidohomoalanina
20	<220> <221> MISC_FEATURE <222> (5)..(5) <223> Xaa es OSu (éster N-succinimidílico)
	<400> 4

25 Xaa Gly Gly Gly Xaa  
1 5

30      <210> 5  
          <211> 5  
          <212> PRT  
          <213> Secuencia artificial

35                    <220>  
                     <223> Constructo sintético

```

40      <220>
      <221> MISC_FEATURE
      <222> (1)..(1)
      <223> Xaa es propargilglicina

```

```

45      <220>
      <221> MISC_FEATURE
      <222> (5)..(5)
      <223> Xaa es OSu (éster de N-succinimidilo)
      <400> 5

```

50 Xaa Gly Gly Gly Xaa  
1 5

55

<210>	6
<211>	18
<212>	ADN
<213>	Secuencia artificial

<220>  
<223> Constructo sintético

```

64         <220>
        <221> misc_feature
        <222> (1)..(1)
        <223> Grupo fosfato extremo 5'
65
        <220>

```

<221> misc\_feature  
 <222> (18)..(18)  
 <223> 6-FAM extremo 3'  
 5 <400> 6  
 gctgtgcagg tagagtgc 18  
 <210> 7  
 <211> 18  
 10 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> Constructo sintético  
 15 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> grupo fosfato extremo 5'  
 20 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> 2'-OMe se adjunta a g  
 25 <220>  
 <221> misc\_feature  
 <222> (18)..(18)  
 <223> 6-FAM extremo 3'  
 30 <400> 7  
 gctgtgcagg tagagtgc 18  
 <210> 8  
 <211> 15  
 35 <212> ARN  
 <213> Secuencia artificial  
 <220> 5  
 <223> Constructo sintético  
 40 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> Grupo fosfato extremo 5'  
 45 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> 2'-OMe se adjunta a g  
 50 <220>  
 <221> misc\_feature  
 <222> (2)..(2)  
 <223> 2'-OMe se adjunta a u  
 55 <220>  
 <221> misc\_feature  
 <222> (3)..(3)  
 <223> 2' - OMe se adjunta a g  
 60 <220>  
 <221> misc\_feature  
 <222> (4)..(4)  
 <223> 2'-OMe se adjunta a c  
 65

5  
 <220>  
 <221> misc\_feature  
 <222> (5)..(5)  
 <223> 2'-OMe se adjunta a un

10  
 <220>  
 <221> misc\_feature  
 <222> (6)..(6)  
 <223> 2'-OMe se adjunta a g

15  
 <220>  
 <221> misc\_feature  
 <222> (7)..(7)  
 <223> 2'-OMe se adjunta a g

20  
 <220>  
 <221> misc\_feature  
 <222> (8)..(8)  
 <223> 2'-OMe se adjunta a u

25  
 <220>  
 <221> misc\_feature  
 <222> (9)..(9)  
 <223> 2'-OMe se adjunta a un

30  
 <220>  
 <221> misc\_feature  
 <222> (10)..(10)  
 <223> 2'-OMe se adjunta a g

35  
 <220> 5  
 <221> misc\_feature  
 <222> (11)..(11)  
 <223> 2'-OMe se adjunta a a

40  
 <220>  
 <221> misc\_feature  
 <222> (12)..(12)  
 <223> 2'-OMe se adjunta a g

45  
 <220>  
 <221> misc\_feature  
 <222> (13)..(13)  
 <223> 2'-O Me se adjunta a u

50  
 <220>  
 <221> misc\_feature  
 <222> (14)..(14)  
 <223> 2'-OMe se adjunta a g

55  
 <220>  
 <221> misc\_feature  
 <222> (15)..(15)  
 <223> 2'-OMe se adjunta a c

60  
 <220>  
 <221> misc\_feature  
 <222> (15)..(15)  
 <223> 6-FAM extremo 3'

65  
 <400> 8  
 gugcagguag agugc 15

<210> 9  
 <211> 15  
 <212> ADN

<213> Secuencia artificial  
 <220>  
 <223> Constructo sintético  
 5  
 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> HO- extremo 5'  
 10  
 <220>  
 <221> misc\_feature  
 <222> (15)..(15)  
 <223> 2'-OMe se adjunta a g  
 15  
 <220>  
 <221> misc\_feature  
 <222> (15)..(15)  
 <223> -OH extremo 3'  
 20  
 <400> 9  
 tacgtatagc actgg 15  
 25  
 <210> 10  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 30  
 <220>  
 <223> Constructo sintético  
 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> HO- extremo 5'  
 35  
 <220>  
 <221> misc\_feature  
 <222> (21)..(21)  
 <223> -OH extremo 3'  
 40  
 <400> 10  
 gcagactagc tatacgactg g 21  
 45  
 <210> 11  
 <211> 15  
 <212> ARN  
 <213> Secuencia artificial  
 50  
 <220>  
 <223> Constructo sintético  
 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> HO- extremo 5'  
 55  
 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> 2'-OMe se adjunta a u  
 60  
 <220>  
 <221> misc\_feature  
 <222> (2)..(2)  
 <223> 2'-OMe se adjunta a a  
 65

	<220>
	<221> misc_feature
	<222> (3)..(3)
5	<223> 2'-OMe se adjunta a c
	<220>
	<221> misc_feature
	<222> (4)..(4)
10	<223> 2'-OMe se adjunta a g
	<220>
	<221> misc_feature
	<222> (5)..(5)
15	<223> 2'-OMe se adjunta a u
	<220>
	<221> misc_feature
	<222> (6)..(6)
20	<223> 2'-OMe se adjunta a un
	<220>
	<221> misc_feature
	<222> (7)..(7)
25	<223> 2'-OMe se adjunta a u
	<220>
	<221> misc_feature
	<222> (8)..(8)
30	<223> 2'-OMe se adjunta a un
	<220>
	<221> misc_feature
	<222> (9)..(9)
35	<223> 2'-OMe se adjunta a c
	<220>
	<221> misc_feature
	<222> (10)..(10)
40	<223> 2'-OMe se adjunta a g
	<220>
	<221> misc_feature
	<222> (11)..(11)
45	<223> 2'-OMe se adjunta a a
	<220>
	<221> misc_feature
	<222> (12)..(12)
50	<223> 2'-OMe se adjunta a c
	<220>
	<221> misc_feature
	<222> (13)..(13)
55	<223> 2'-OMe se adjunta a u
	<220>
	<221> misc_feature
	<222> (14)..(14)
60	<223> 2'-OMe se adjunta a g
	<220>
	<221> misc_feature
	<222> (15)..(15)
65	<223> 2'-OMe se adjunta a g

5  
 <220>  
 <221> misc\_feature  
 <222> (15)..(15)  
 <223> -OH extremo 3'

10  
 <400> 11  
 uacguauacg acugg 15

15  
 <210> 12  
 <211> 15  
 <212> ADN  
 <213> Secuencia artificial

20  
 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> grupo fosfato extremo 5'

25  
 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> 2'-OMe se adjunta a g

30  
 <220>  
 <221> misc\_feature  
 <222> (15)..(15)  
 <223> 6-FAM extremo 3'

35  
 <400> 12  
 gtcaggtag agtgc 15

40  
 <210> 13  
 <211> 15  
 <212> ADN  
 <213> Secuencia artificial

45  
 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> HO- extremo 5'

50  
 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> 2'-OMe se adjunta a u

55  
 <220>  
 <221> misc\_feature  
 <222> (15)..(15)  
 <223> 2'-OMe se adjunta a g

60  
 <220>  
 <221> misc\_feature  
 <222> (15)..(15)  
 <223> -OH extremo 3'

65  
 <400> 13  
 uacgtatacg actgg 15

<210> 14



	<211> 8
	<212> ADN
	<213> Secuencia artificial
5	<220>
	<223> Constructo sintético
	<220>
10	<221> misc_feature
	<222> (1)..(1)
	<223> Grupo fosfato extremo 5'
	<220>
15	<221> misc_feature
	<222> (1)..(1)
	<223> 2'-OMe se adjunta a g
	<220>
20	<221> misc_feature
	<222> (8)..(8)
	<223> 6-FAM extremo 3'
25	<400> 14
	gtgagtgc 8
	<210> 15
	<211> 8
	<212> ADN
30	<213> Secuencia artificial
	<220>
	<223> Constructo sintético
35	<220>
	<221> misc_feature
	<222> (1)..(1)
	<223> HO- extremo 5'
40	<220>
	<221> misc_feature
	<222> (8)..(8)
	<223> 2'-OMe se adjunta a g
45	<220>
	<221> misc_feature
	<222> (8)..(8)
	<223> -OH extremo 3'
50	<400> 15
	cagactgg 8
	<210> 16
	<211> 5
	<212> ADN
55	<213> Secuencia artificial
	<220>
	<223> Constructo sintético 5
60	<220>
	<221> misc_feature
	<222> (1)..(1)
	<223> grupo fosfato extremo 5'
65	<220>
	<221> misc_feature

	<222> (1)..(1)
	<223> 2'-OMe se adjunta a g
5	<220> <221> misc_feature <222> (5)..(5) <223> 6-FAM extremo 3'
10	<400> 16 gtgac 5
15	<210> 17 <211> 5 <212> ADN <213> Secuencia artificial
20	<220> <223> Constructo sintético
25	<220> <221> misc_feature <222> (1)..(1) <223> HO- extremo 5'
30	<220> <221> misc_feature <222> (1)..(1) <223> 2'-OMe se adjunta a a
35	<220> <221> misc_feature <222> (5)..(5) <223> 2'-OMe se adjunta a g
40	<220> <221> misc_feature <222> (5)..(5) <223> -OH extremo 3'
45	<400> 17 actgg 5
50	<210> 18 <211> 38 <212> ADN <213> Secuencia artificial
55	<220> <223> Constructo sintético
60	<220> 5 <221> misc_feature <222> (38)..(38) <223> g es propargilG
65	<400> 18 gcgtgaacat gcatctcccg tatgcgtaca gtccattg 38
	<210> 19 <211> 37 <212> ADN <213> Secuencia artificial
	<220> <223> Constructo sintético

<220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> t es azidoT  
 5  
 <400> 19  
 tatagcgga tatacacact ggcgagcttg cgtactg 37  
 10  
 <210> 20  
 <211> 38  
 <212> ADN  
 <213> Secuencia artificial  
 15  
 <220>  
 <223> Constructo sintético  
 20  
 <220>  
 <221> misc\_feature  
 <222> (38)..(38)  
 <223> Extremo 3' incluye espaciador7-azida  
 <400> 20  
 gcgtgaacat gcatctcccg tatgcgtaca gtccattg 38  
 25  
 <210> 21  
 <211> 35  
 <212> ADN  
 <213> Secuencia artificial  
 30  
 <220>  
 <223> Constructo sintético  
 35  
 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> hexinilo extremo 5'  
 <400> 21  
 tagcgcgata tacacactgg cgagcttgcg tactg 35  
 40  
 <210> 22  
 <211> 15  
 <212> ADN 5  
 <213> Secuencia artificial  
 45  
 <220>  
 <223> Constructo sintético  
 50  
 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> Cy5 extremo 5'  
 <400> 22  
 cagtacgcaa gctcg 15  
 55  
 <210> 23  
 <211> 75  
 <212> ADN  
 <213> Secuencia artificial  
 60  
 <220>  
 <223> Constructo sintético  
 65  
 <400> 23

	gcgtgaacat gcatctccc tatgcgtaca gtccattgta tagcgcgata tacacactgg	60
5	cgagcttgcg tactg	75
10	<210> 24 <211> 18 <212> ADN <213> Secuencia artificial	
15	<220> <223> Constructo sintético	
20	<400> 24 gcgtgaacat gcatctcc 18 <210> 25 <211> 17 <212> ADN <213> Secuencia artificial	
25	<220> <223> Constructo sintético	
30	<400> 25 cagtacgcaa gctcgcc 17 <210> 26 <211> 21 <212> ADN <213> Secuencia artificial	
35	<220> <223> Constructo sintético	
40	<220> <221> misc_feature <222> (1)..(1) <223> Cy5 extremo 5'	
45	<220> <221> misc_feature <222> (21)..(21) <223> Tiofosfato extremo 3'	
50	<400> 26 cgatatacac actggcgagc t 21 <210> 27 <211> 12 <212> ADN <213> Secuencia artificial	
55	<220> <223> Constructo sintético	
60	<220> <221> misc_feature <222> (1)..(1) <223> t es lododT	
65	<220> <221> misc_feature <222> (12)..(12) <223> 6-FAM extremo 3'	

	<400> 27 tgcgtactga gc 12
5	<210> 28 <211> 17 <212> ADN <213> Secuencia artificial
10	<220> <223> Constructo sintético
15	<400> 28 cagtacgcaa gctcgcc 17
20	<210> 29 <211> 11 <212> ADN <213> Secuencia artificial
25	<220> <221> misc_feature <222> (1)..(1) <223> Grupo fosfato extremo 5'
30	<220> <221> misc_feature <222> (1)..(1) <223> 2'-OMe se adjunta a c
35	<220> <221> misc_feature <222> (11)..(11) <223> aminohex extremo 3'
40	<400> 29 cgagtcacgt c 11
45	<210> 30 <211> 8 <212> ADN <213> Secuencia artificial
50	<220> <223> Constructo sintético
55	<220> <221> misc_feature <222> (1)..(1) <223> 2'-OMe se adjunta a c
60	<220> <221> misc_feature <222> (8)..(8) <223> 2'-OMe se adjunta a a
65	<400> 30 cagtgta 8
	<210> 31 <211> 11 <212> ADN <213> Secuencia artificial

<220>  
 <223> Constructo sintético

5

<220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> Grupo fosfato extremo 5'

10

<220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> 2' -OMe se adjunta a c

15

<220>  
 <221> misc\_feature  
 <222> (2)..(2)  
 <223> g se sintetiza como un enlace de fosforotioato

20

<220> 5  
 <221> misc\_feature  
 <222> (11)..(11)  
 <223> aminohex extremo 3'

25

<400> 31  
 cgagtcacgt c 11

30

<210> 32  
 <211> 18  
 <212> ADN  
 <213> Secuencia artificial

35

<220>  
 <223> Constructo sintético

40

<220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> amidación

45

<220>  
 <221> misc\_feature  
 <222> (18)..(18)  
 <223> g es propargilG

50

<400> 32  
 tcgaatgact ccgatatg 18

55

<210> 33  
 <211> 37  
 <212> ADN  
 <213> Secuencia artificial

60

<220>  
 <223> Constructo sintético

65

<220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> t es azido dT

<400> 33  
 tatagcgcca tatacacact ggcgagcttg cgtactg 37

<210> 34  
 <211> 25

<212> ADN  
 <213> Secuencia artificial  
  
 5        <220>  
          <223> Constructo sintético  
  
          <220>  
          <221> misc\_feature  
 10       <222> (1)..(1)  
          <223> t es azido dT  
  
          <400> 34  
          tacacactgg cgagcttgcg tactg 25  
  
 15       <210> 35  
          <211> 15  
          <212> ADN  
          <213> Secuencia artificial  
  
 20       <220>  
          <223> Constructo sintético  
  
          <220>  
          <221> misc\_feature  
 25       <222> (1)..(1)  
          <223> t es azido dT  
  
          <220>  
          <221> misc\_feature  
 30       <222> (15)..(15)  
          <223> u es propargil U-TIPS  
  
          <400> 35  
          tatgcgtaca gtccu 15  
  
 35       <210> 36  
          <211> 17  
          <212> ADN  
          <213> Secuencia artificial  
  
 40       <220>  
          <223> Constructo sintético  
  
          <220>  
          <221> misc\_feature  
 45       <222> (1)..(1)  
          <223> 6-FAM extremo 5'  
  
          <400> 36  
 50       cagtacgcaa gctcgcc 17  
  
          <210> 37  
          <211> 17  
          <212> ADN  
 55       <213> Secuencia artificial  
  
          <220>  
          <223> Constructo sintético  
  
 60       <220>  
          <221> misc\_feature  
          <222> (1)..(1)  
          <223> Cy5 extremo 5'  
  
 65       <400> 37  
          cagtacgcaa gctcgcc 17

5           <210> 38  
             <211> 55  
             <212> ADN 5  
             <213> Secuencia artificial

            <220>  
             <223> Constructo sintético

10           <220>  
             <221> misc\_feature  
             <222> (1)..(1)  
             <223> Biotina extremo 5' TEG//ispC3//ispC3

15           <400> 38  
             tcgaatgact ccgatatgta tagcgcgata tacacactgg cgagcttgcg tactg 55

20           <210> 39  
             <211> 55  
             <212> ADN  
             <213> Secuencia artificial

25           <220>  
             <223> Constructo sintético

30           <220>  
             <221> misc\_feature  
             <222> (1)..(1)  
             <223> 5' terminal Bio-TEG//ispC3//ispC3

35           <220>  
             <221> misc\_feature  
             <222> (18)..(18)  
             <223> g es riboG

40           <400> 39  
             tcgaatgact ccgatatgta tagcgcgata tacacactgg cgagcttgcg tactg 55

45

50

55

60

65



## REIVINDICACIONES

1. Un método de etiquetado de una primera biblioteca que comprende una entidad química codificada por oligonucleótidos, comprendiendo dicho método:

- (i) proporcionar una pieza de cabeza que consiste en un oligonucleótido monocatenario que tiene un primer grupo funcional y un segundo grupo funcional, en donde dicha pieza de cabeza comprende al menos un nucleótido sustituido en 2';
- (ii) unir dicho primer grupo funcional de dicha pieza de cabeza a un primer componente de dicha entidad química, en donde dicha pieza de cabeza está conectada directamente a dicho primer componente o dicha pieza de cabeza está conectada indirectamente a dicho primer componente mediante un conector bifuncional;
- y
- (iii) unir dicho segundo grupo funcional de dicha pieza de cabeza a una primera etiqueta de bloque de construcción que consiste en un oligonucleótido monocatenario para formar un complejo, en donde dicha primera etiqueta de bloque de construcción comprende un nucleótido sustituido en 2' en los extremos tanto 5' como 3',

en donde dichos pasos (ii) y (iii) se pueden realizar en cualquier orden y en donde dicha primera etiqueta de bloque de construcción codifica la reacción de unión de dicho paso (ii), y en donde el paso (iii) comprende una ligasa ARN y/o ligasa ADN para unir dicha pieza de cabeza a dicho primer bloque de construcción, proporcionando así una biblioteca etiquetada.

2. El método de la reivindicación 1, en el que cada nucleótido sustituido en 2' es 2'-O-metilnucleótido o 2'-fluoro-nucleótido.

3. El método de la reivindicación 1 o 2, en el que cada nucleótido sustituido en 2' es 2'-O-metilo guanina, 2'-O-metilo uracilo, 2'-O-metilo adenosina, 2'-O-metilo timidina, 2'-O-metilo inosina, 2'-O-metilo citidina, 2'-O-metilo diamino purina, 2'-fluoro guanina, 2'-fluoro uracilo, 2'-fluoro adenosina, 2'-fluoro timidina, 2'-fluoro-inosina, 2'-fluorocitidina o 2'-5 fluoro diamino purina.

4. El método de una cualquiera de las reivindicaciones 1-3, en donde el paso (ii) comprende unir dicha pieza de cabeza directamente a dicho primer componente.

5. El método de una cualquiera de las reivindicaciones 1-3, en donde el paso (ii) comprende unir dicha pieza de cabeza indirectamente a dicho primer componente mediante un conector bifuncional.

6. El método de una cualquiera de las reivindicaciones 1-5, que comprende además

- (iv) unir una segunda etiqueta de bloque de construcción que consiste en un oligonucleótido monocatenario al extremo 5' o al extremo 3' de dicho complejo; y
- (v) unir un segundo componente de dicha biblioteca química a dicho primer componente, en donde dichas etapas (iv) y (v) pueden realizarse en cualquier orden.

7. El método de la reivindicación 6, en el que dicha segunda etiqueta de bloque de construcción comprende un nucleótido sustituido en 2' en uno o más del extremo 5', el extremo 3' o la posición interna de dicha segunda etiqueta de bloque de construcción.

8. El método de cualquiera de las reivindicaciones 1-7, en el que dicho paso (iii) y/o paso (iv), si está presente, comprende una ligasa ARN y/o una ligasa ADN para unir dicha segunda etiqueta de bloque de construcción a dicho complejo. y/o comprende polietilenglicol y/o uno o más cationes multivalentes solubles.

9. El método de una cualquiera de las reivindicaciones 1-8, en el que dicho método comprende además separar dicho complejo de cualquier etiqueta sin reaccionar o pieza de cabeza sin reaccionar antes de cualquiera de los pasos de unión (ii)-(v) y/o purificar dicho complejo antes de cualquiera de los pasos de unión (ii)-(v).

10. El método de una cualquiera de las reivindicaciones 1-9, en el que dicho método comprende además unir una o más etiquetas de bloques de construcción adicionales a dicho complejo y unir uno o más componentes adicionales a dicho complejo.

11. El método de una cualquiera de las reivindicaciones 1-10, en el que dicha pieza de cabeza comprende una estructura de horquilla.

12. El método de una cualquiera de las reivindicaciones 1-11, en el que dicha pieza de cabeza, dicha primera etiqueta de bloque de construcción, dicha segunda etiqueta de bloque de construcción y/o dicha una o más etiquetas de bloque de construcción adicionales, si están presentes, comprende además una primera secuencia de identificación de biblioteca, secuencia de uso y/o secuencia de origen.

13. El método de una cualquiera de las reivindicaciones 1-12, en el que dicho método comprende además unir una primera etiqueta de identificación de biblioteca, una secuencia de uso, una secuencia de origen y/o un cordal a dicho complejo.

5 14. El método de una cualquiera de las reivindicaciones 1-13, en el que dicho método comprende una pluralidad de piezas de cabeza.

10 15. El método de las reivindicaciones 1-14, en el que el paso (iii) comprende una ligasa ARN, preferiblemente en donde dicha ligasa ARN es ligasa T4 ARN.

15

20

25

30

35

40

45

50

55

60

65

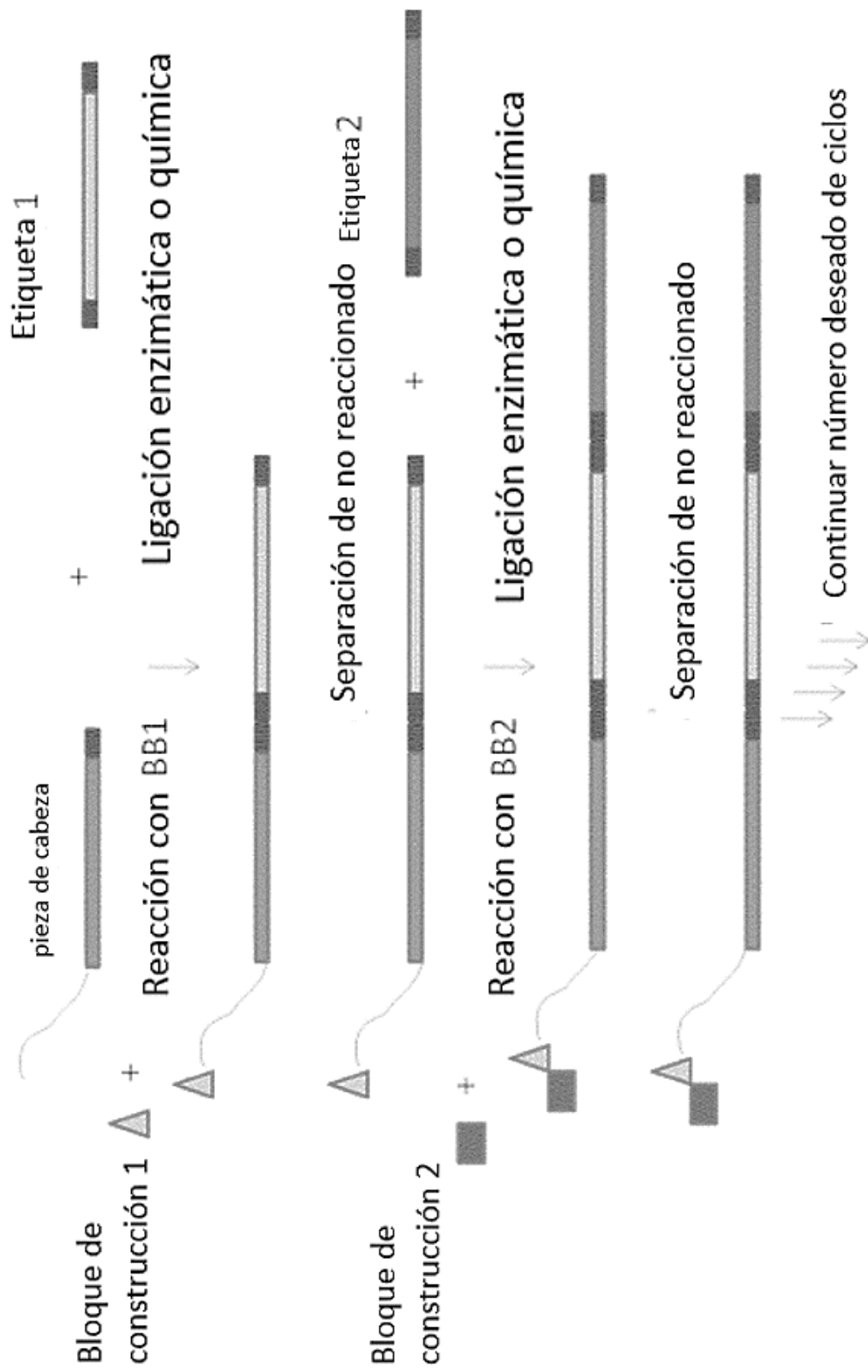
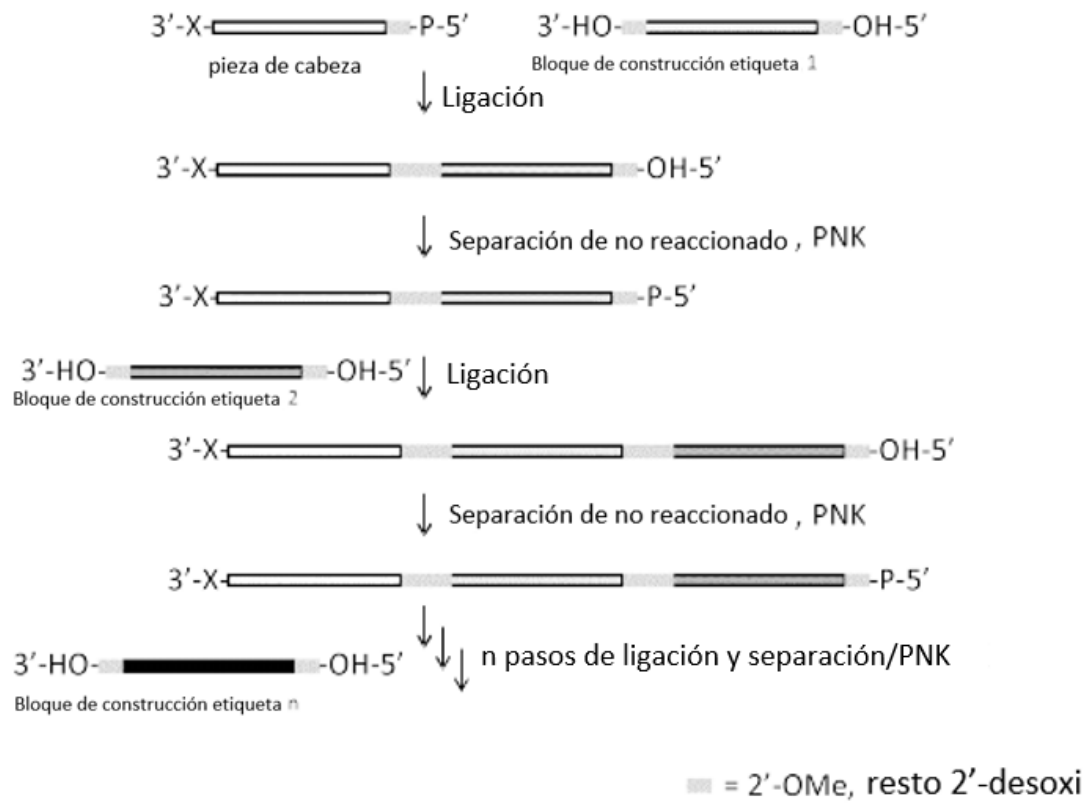


Figura 1



**Figura 2A**

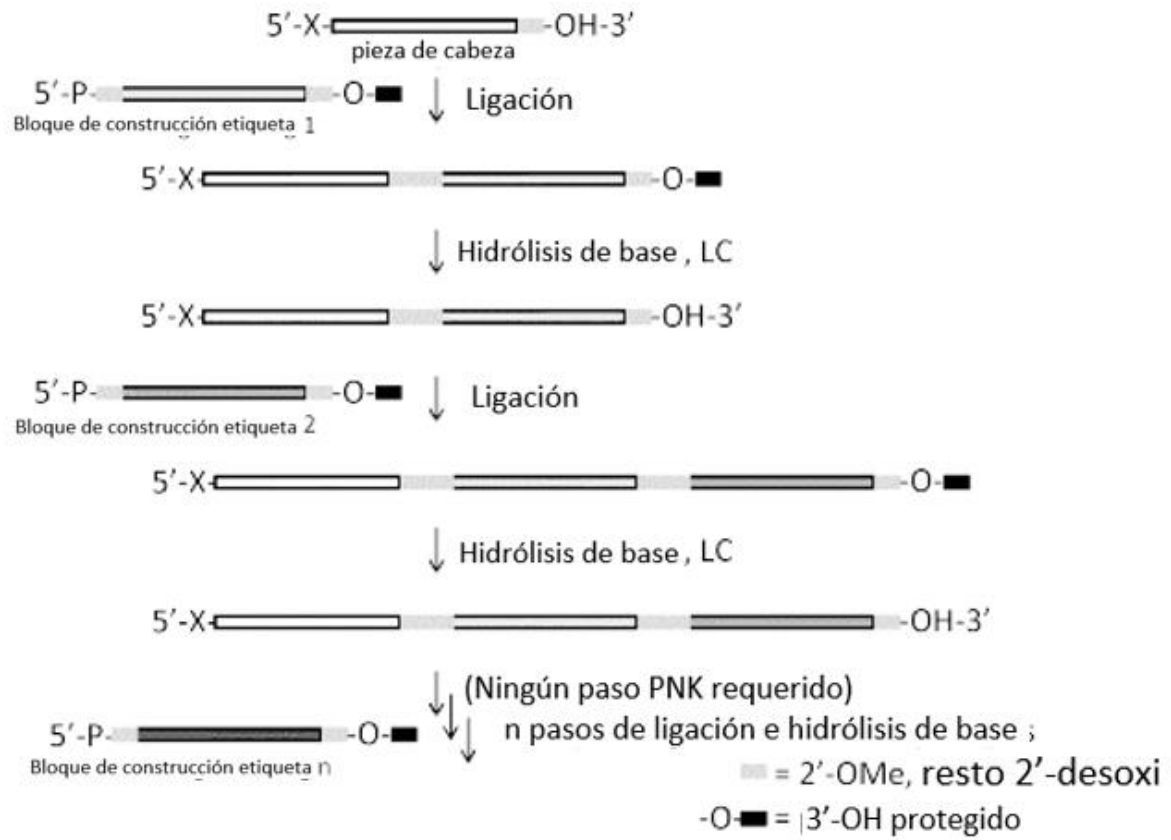
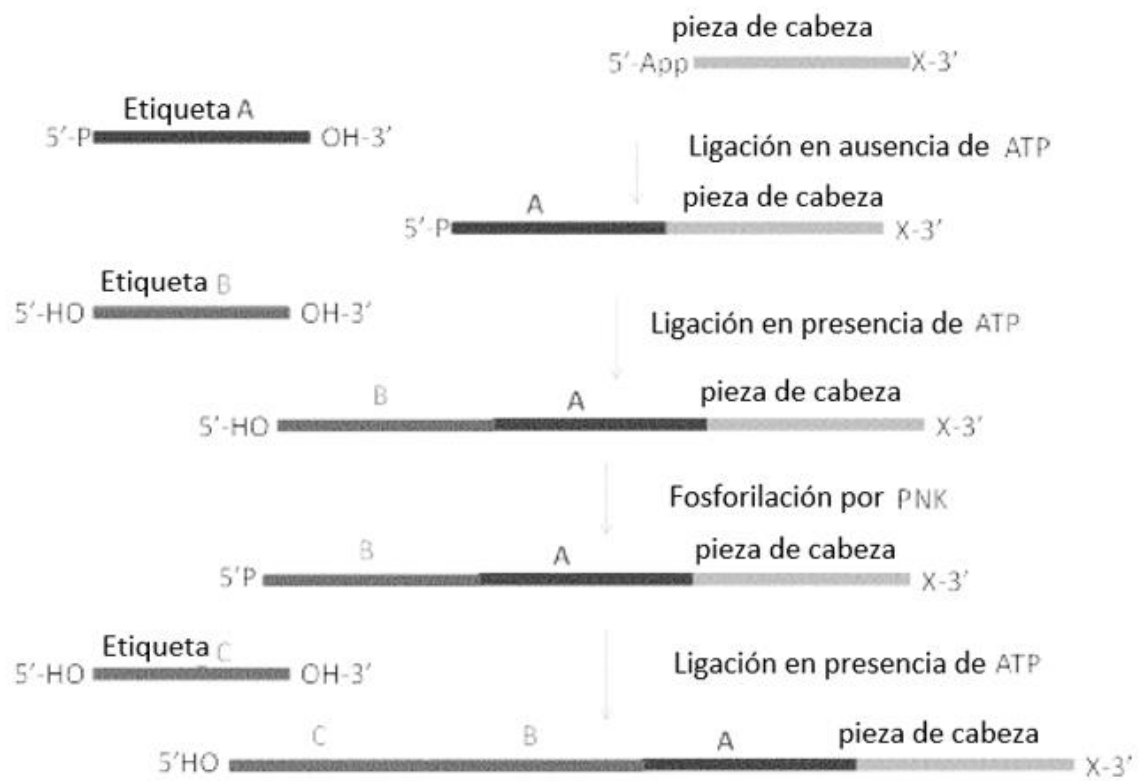
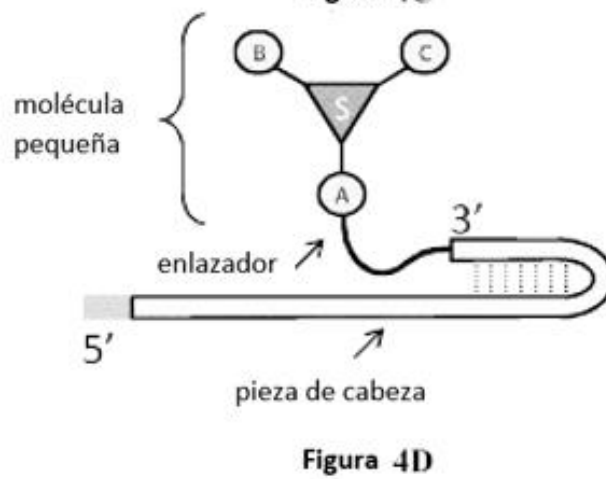
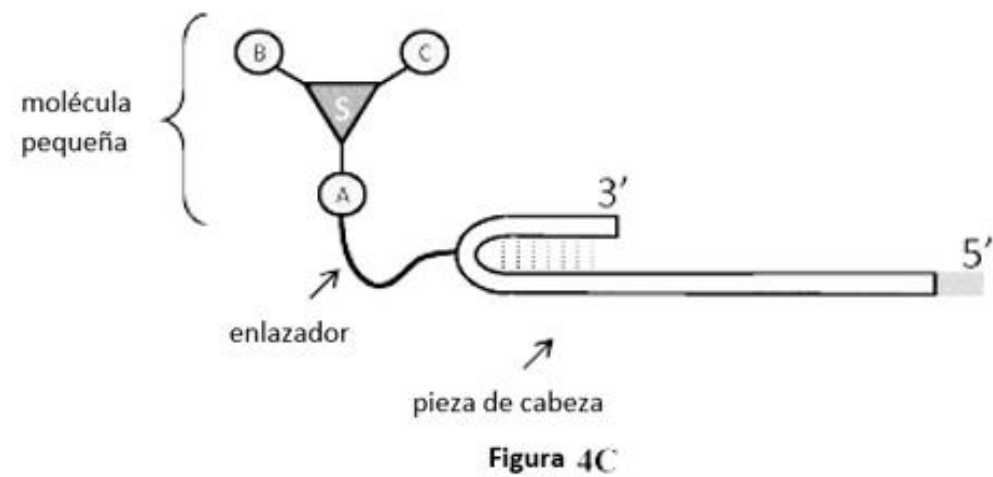
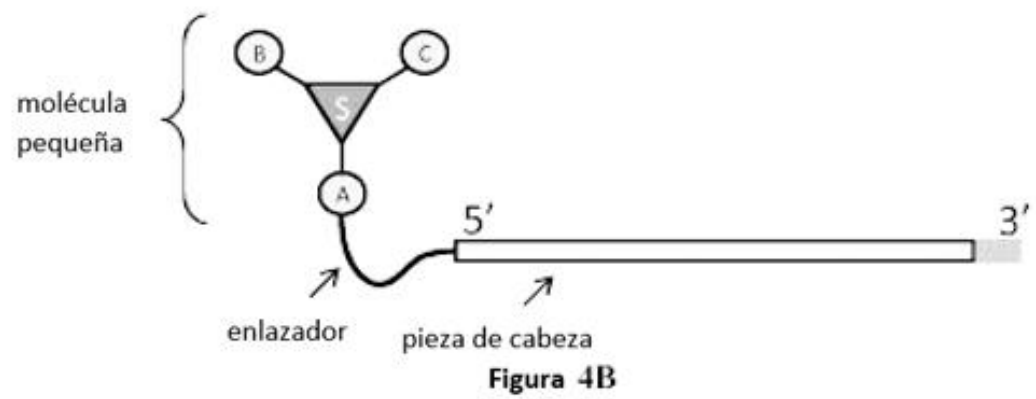
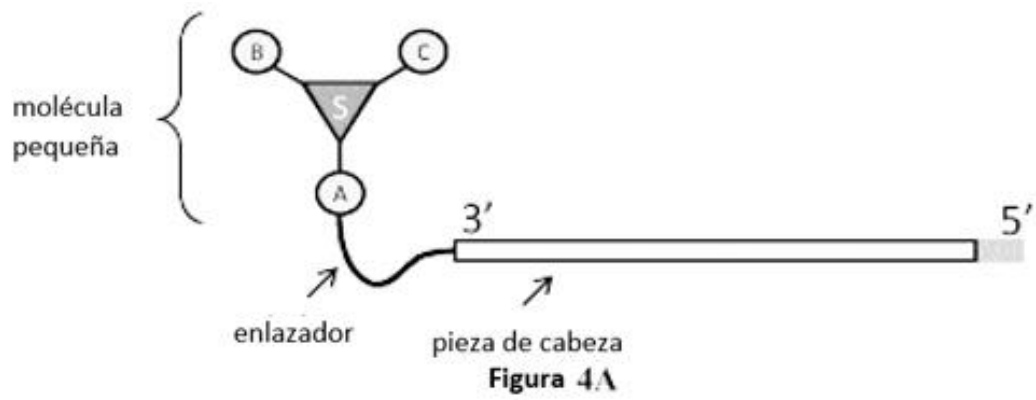


Figura 2B



**Figura 3**



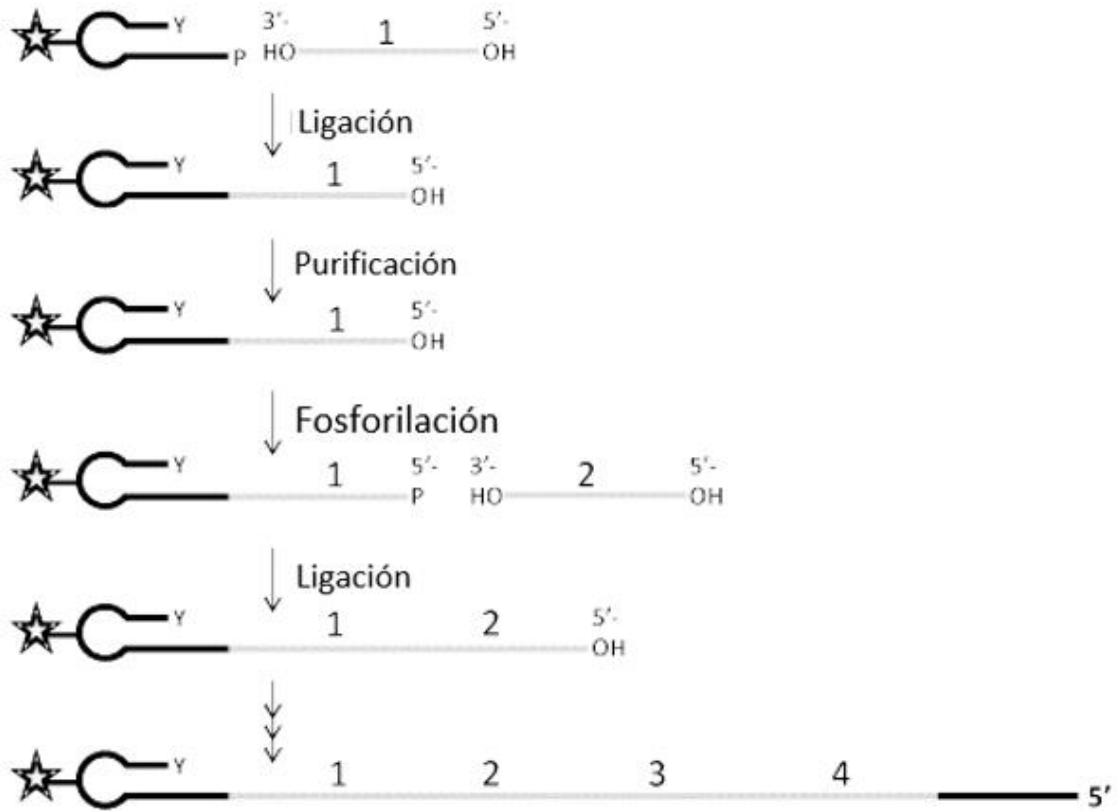


Figura 4E



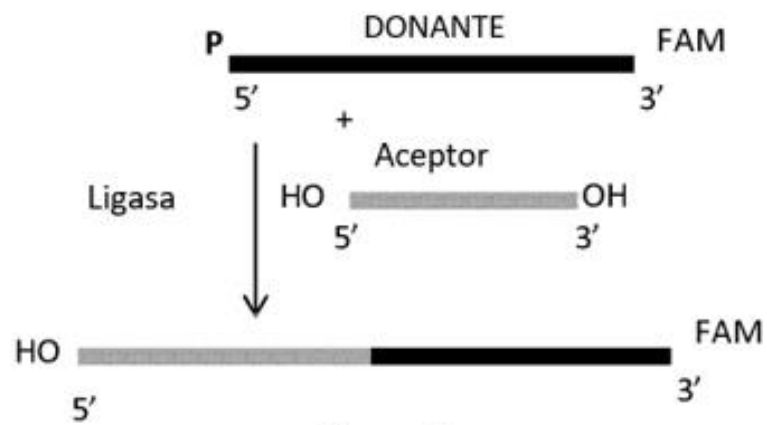


Figura 5A

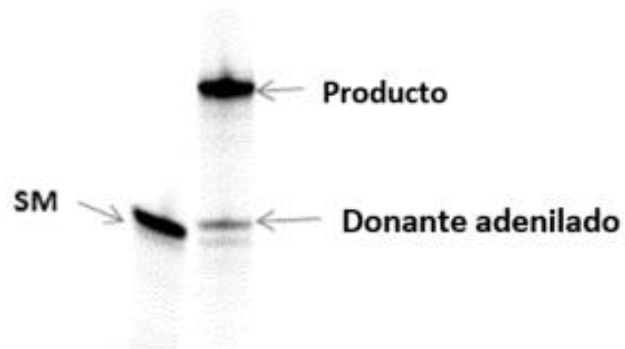


Figura 5B

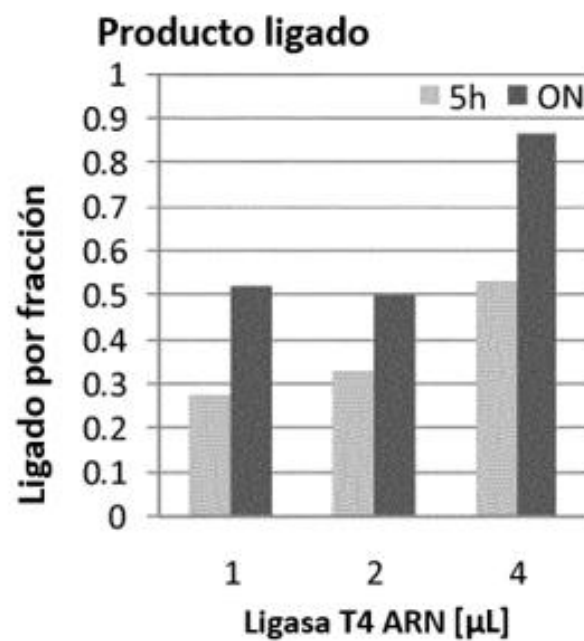
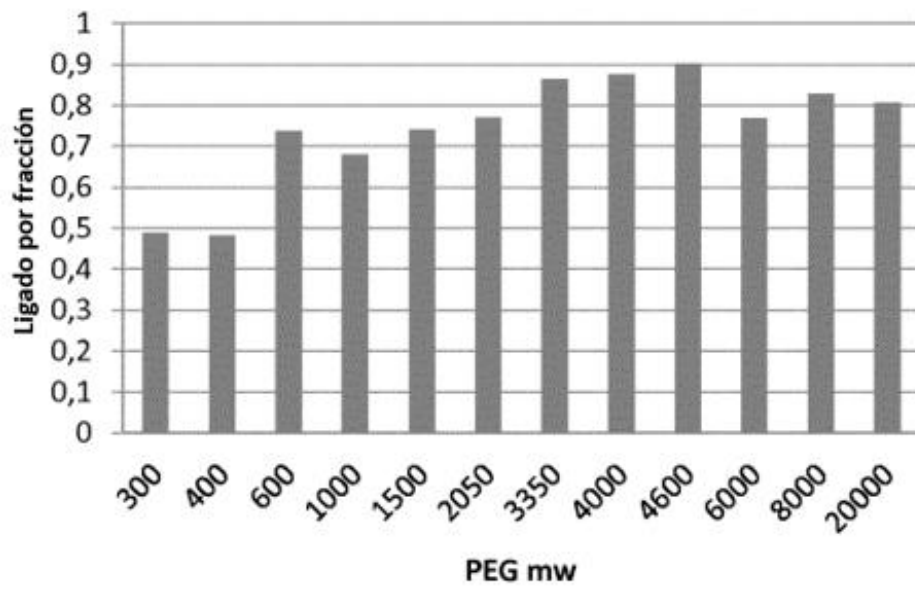
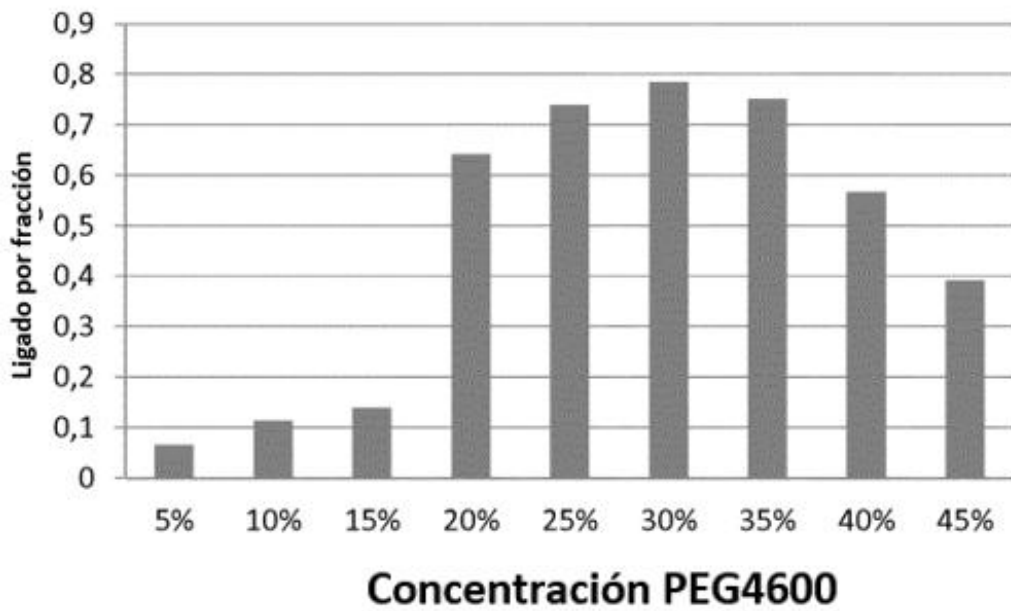


Figura 5C



**Figura 6A**



**Figura 6B**

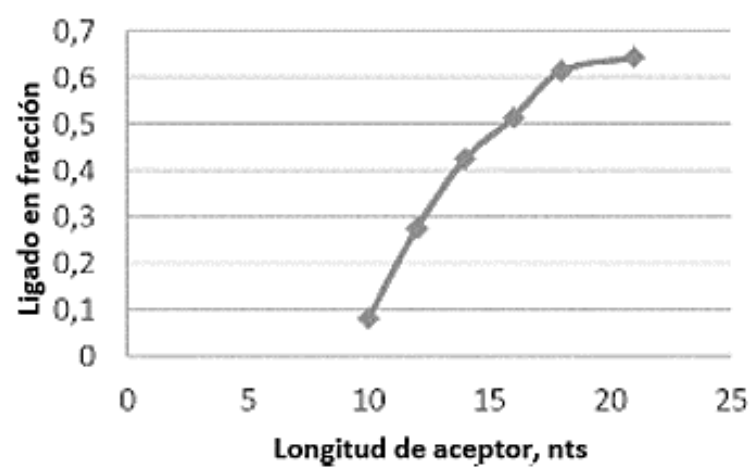
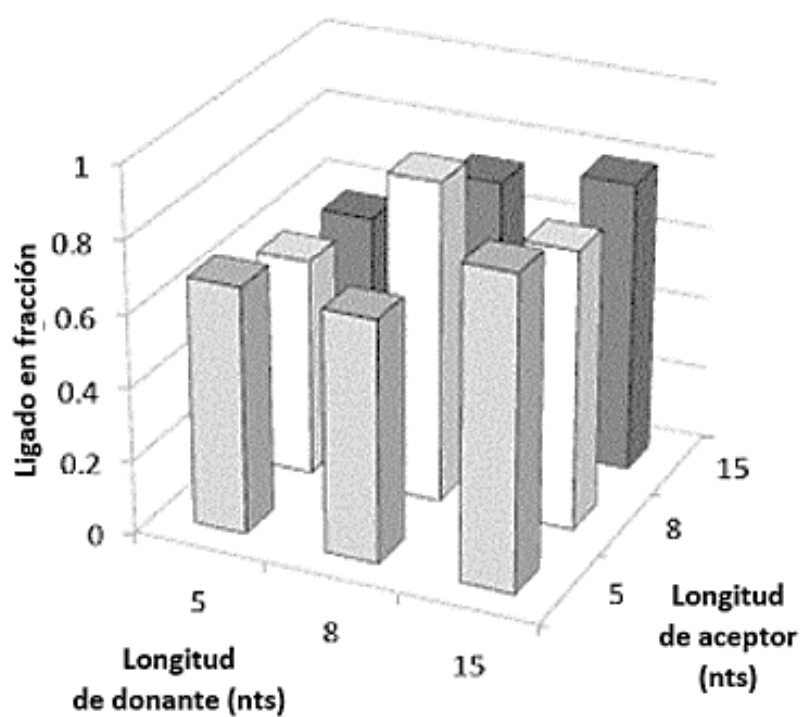


Figura 7A



Longitud de donante \ Longitud de aceptor	Longitud de donante		
	5	8	15
5	0.68	0.66	0.86
8	0.60	0.88	0.76
15	0.57	0.74	0.80

Figura 7B

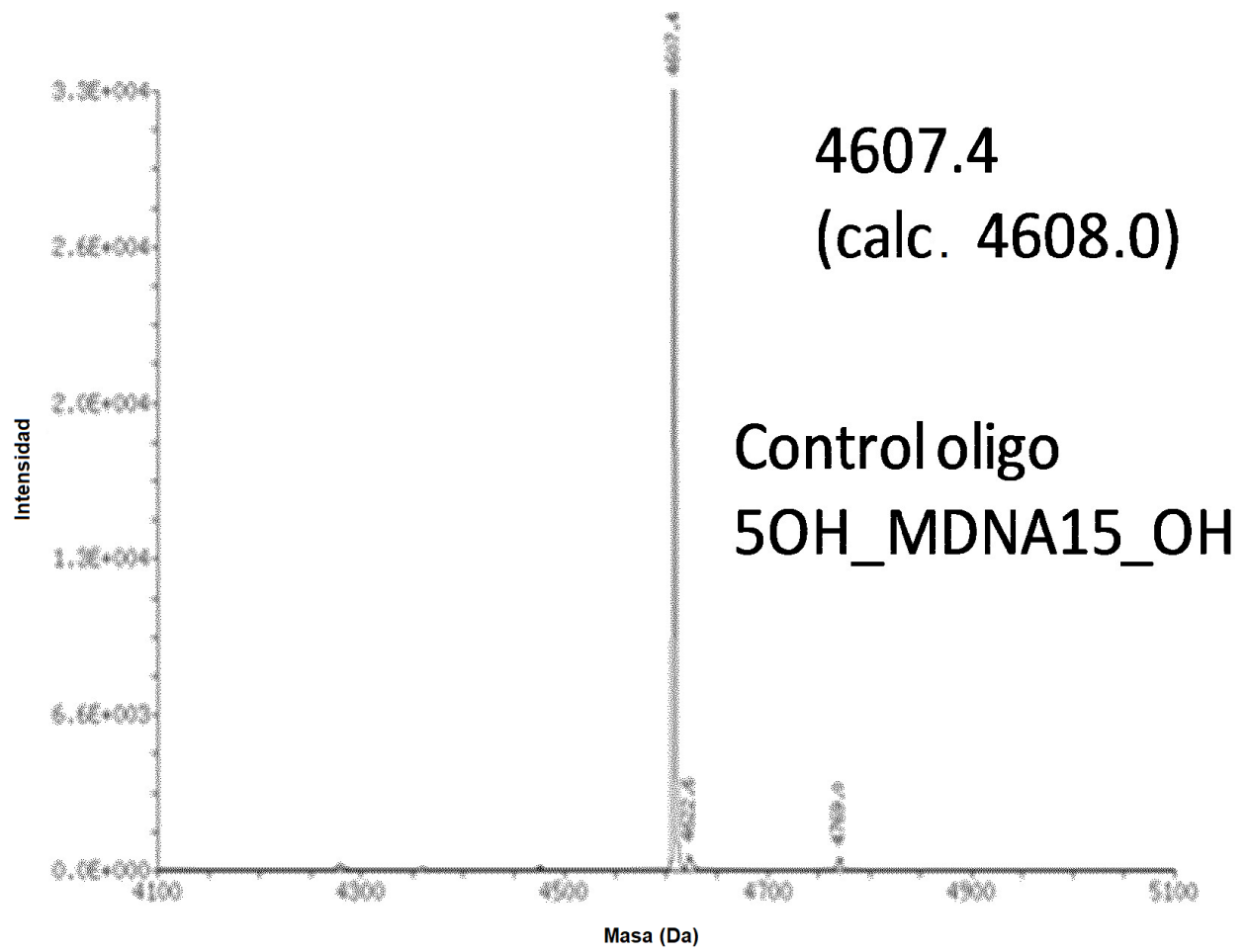


Figura 8A

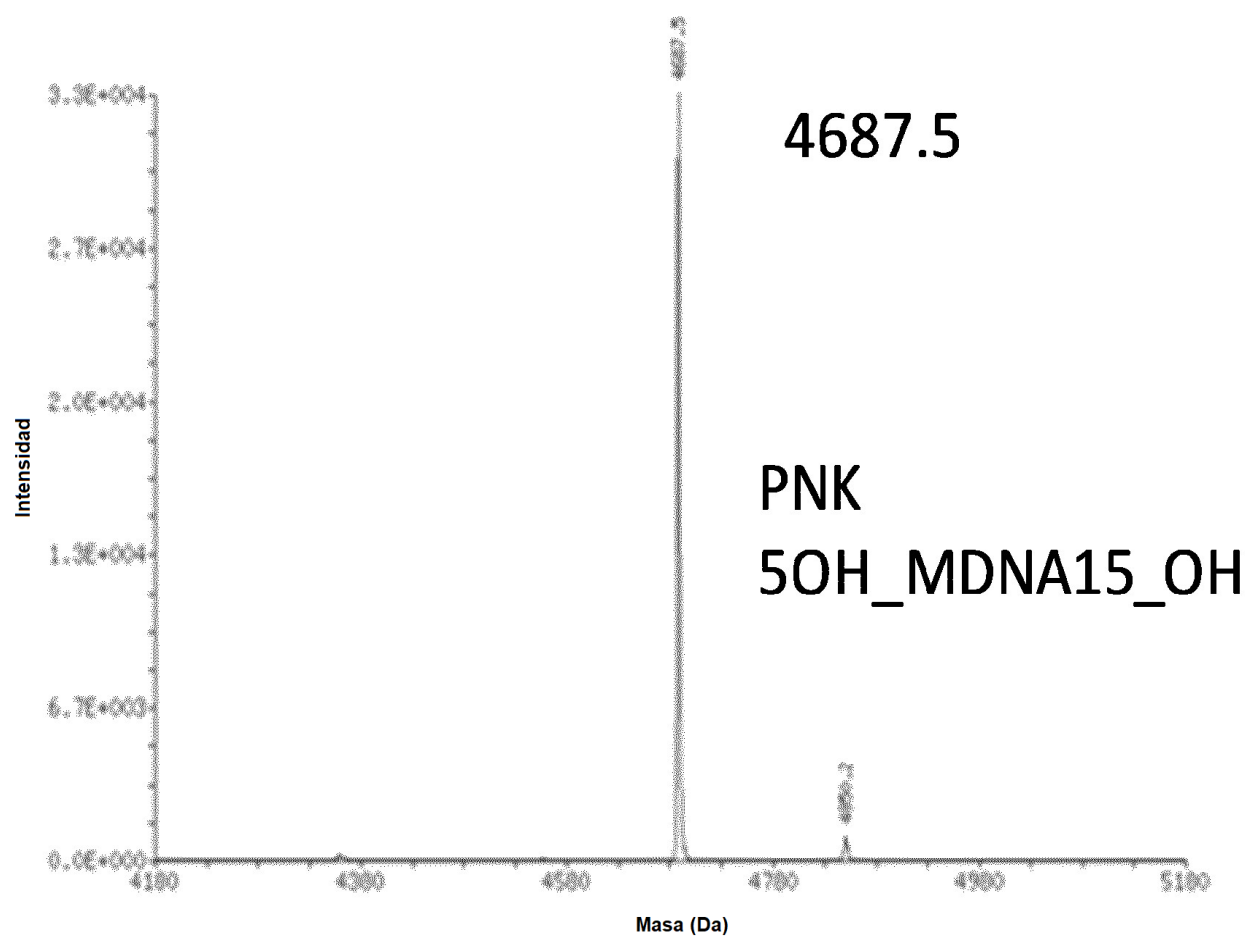
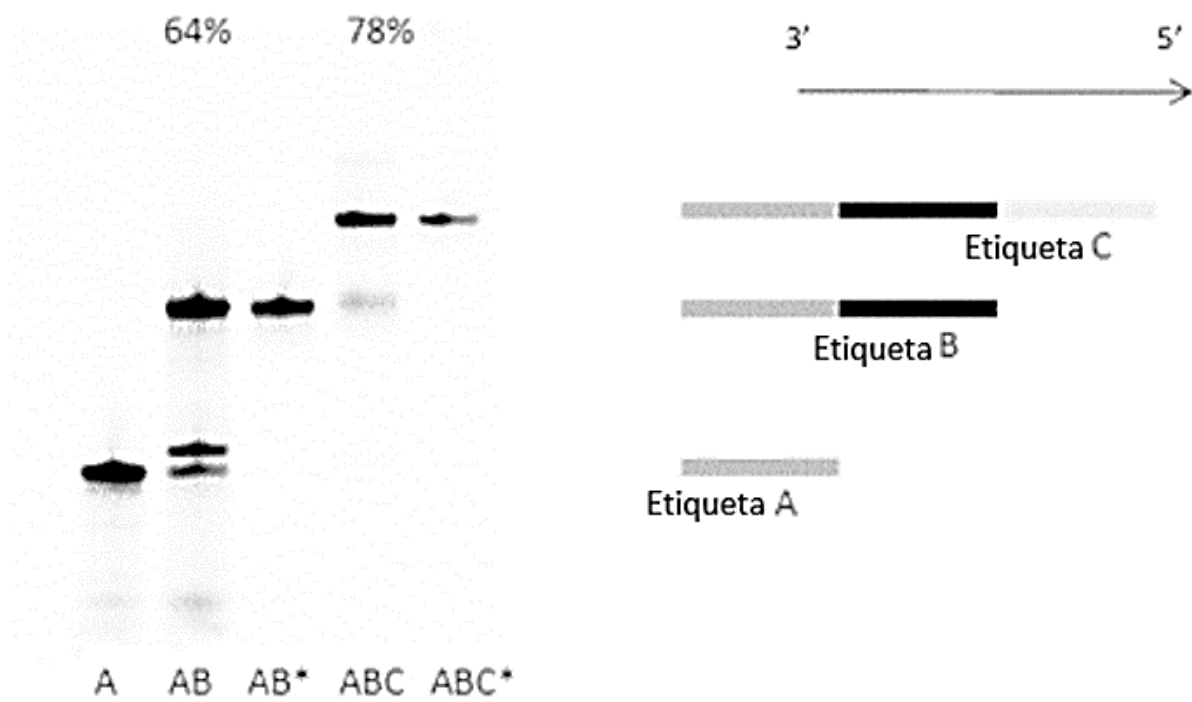


Figura 8B



\*- purificado por gel

**Figura 9**

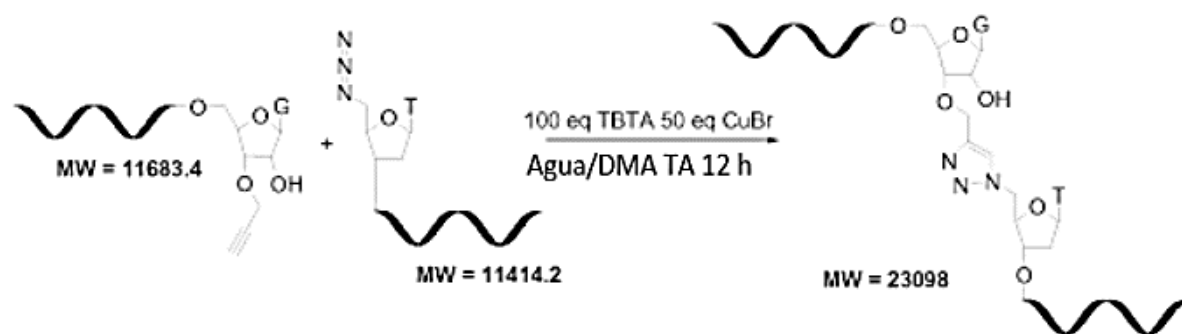


Figura 10A

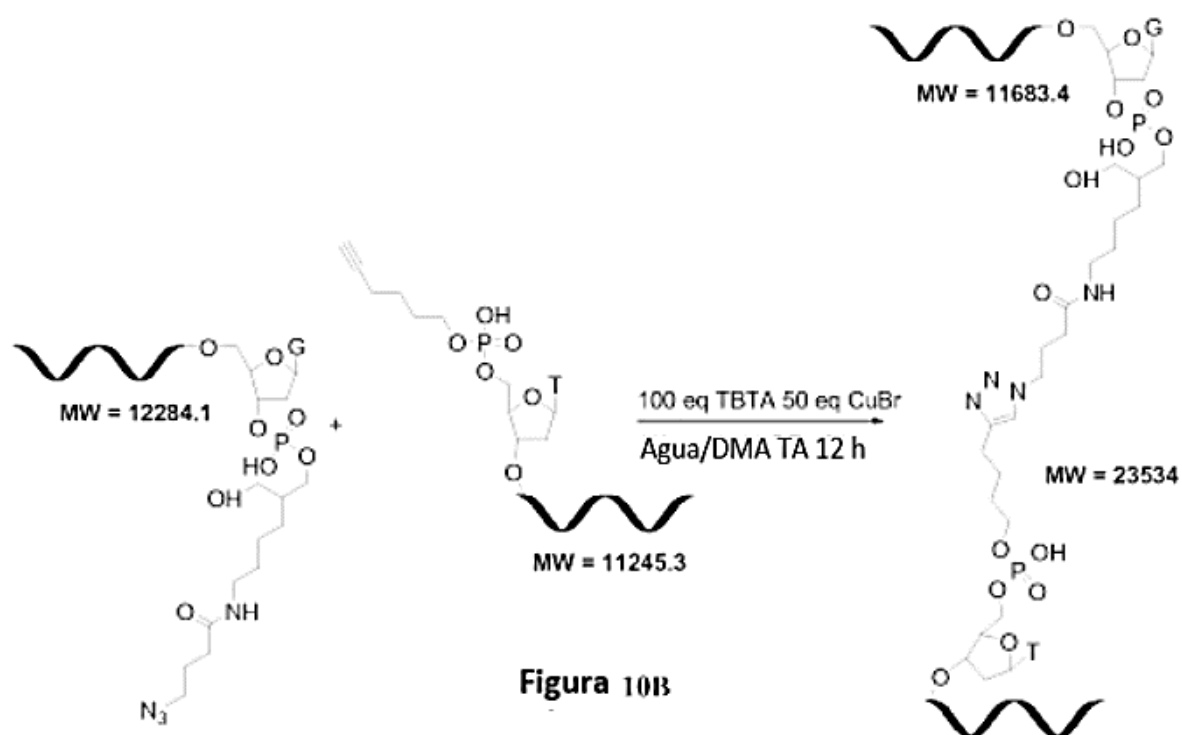


Figura 10B





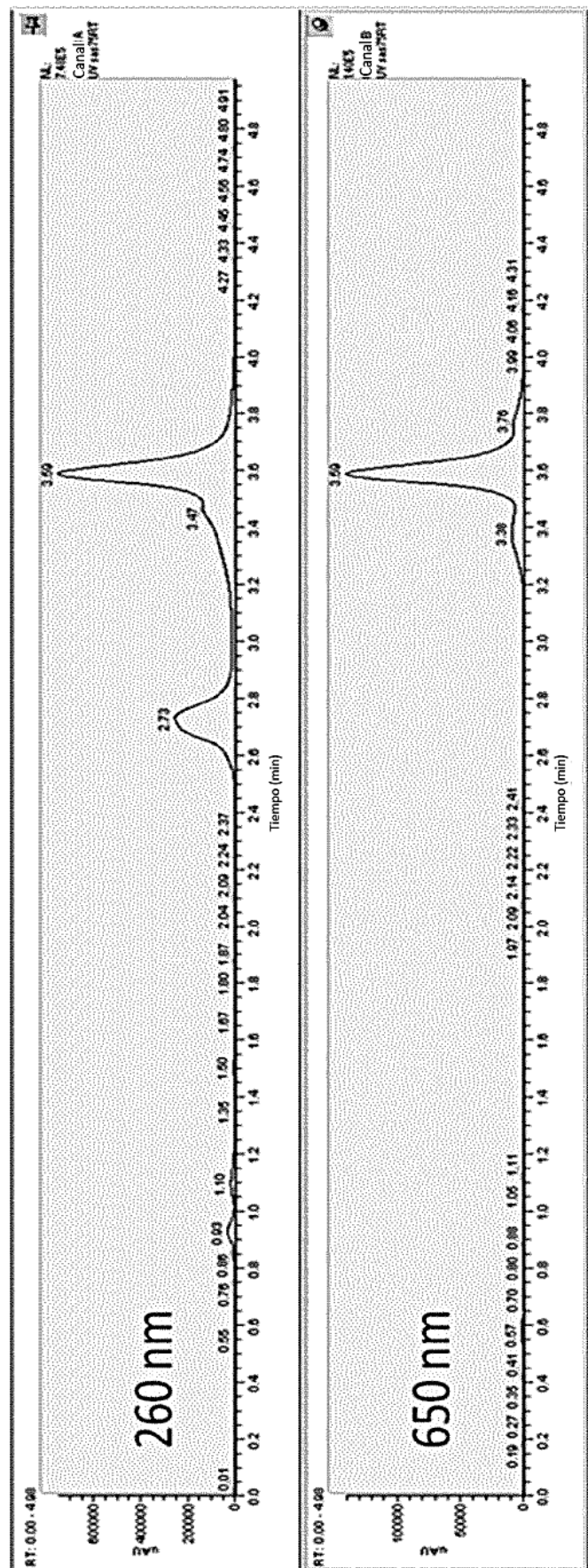


Figura 11B-1

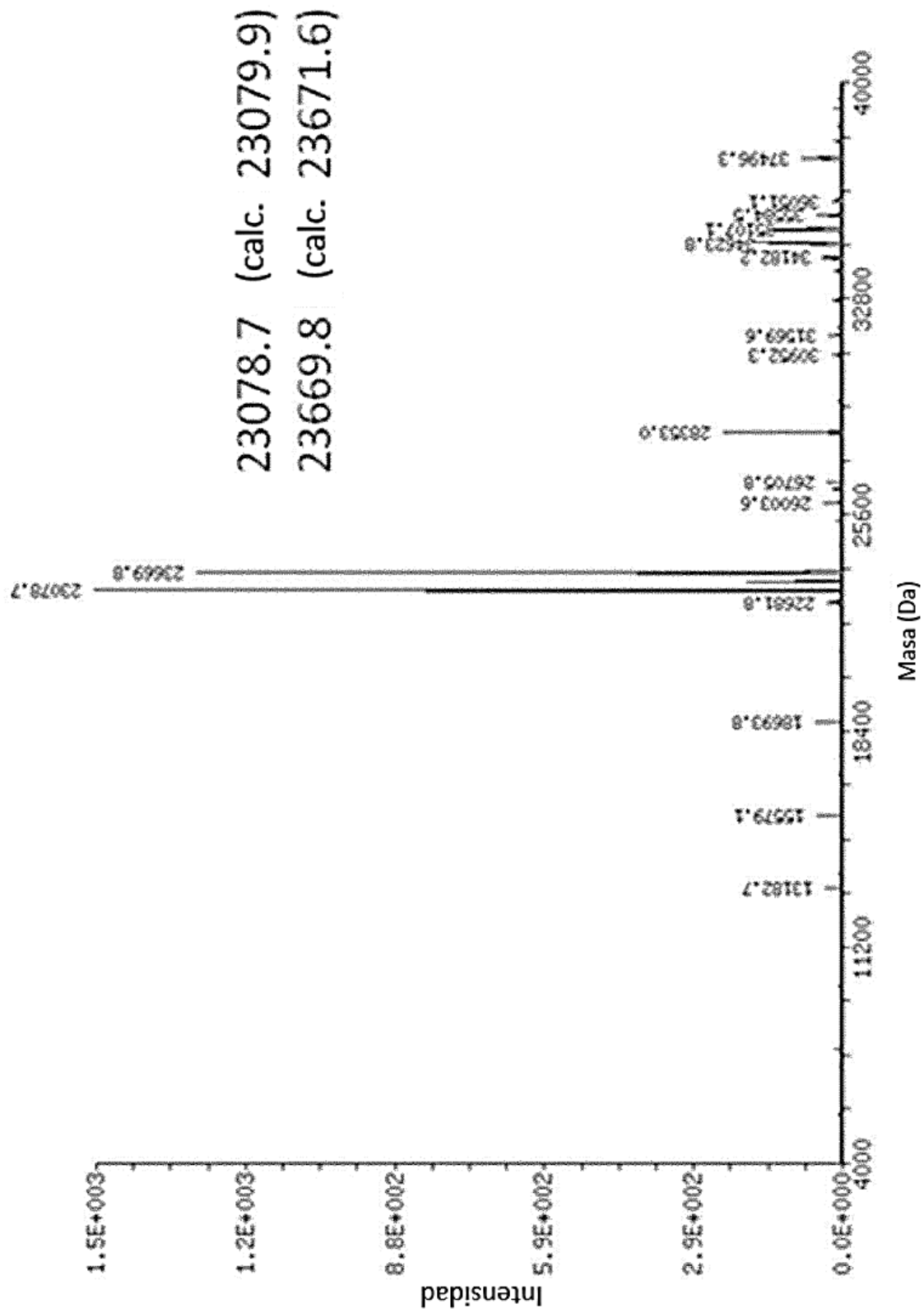


Figura 11B-2

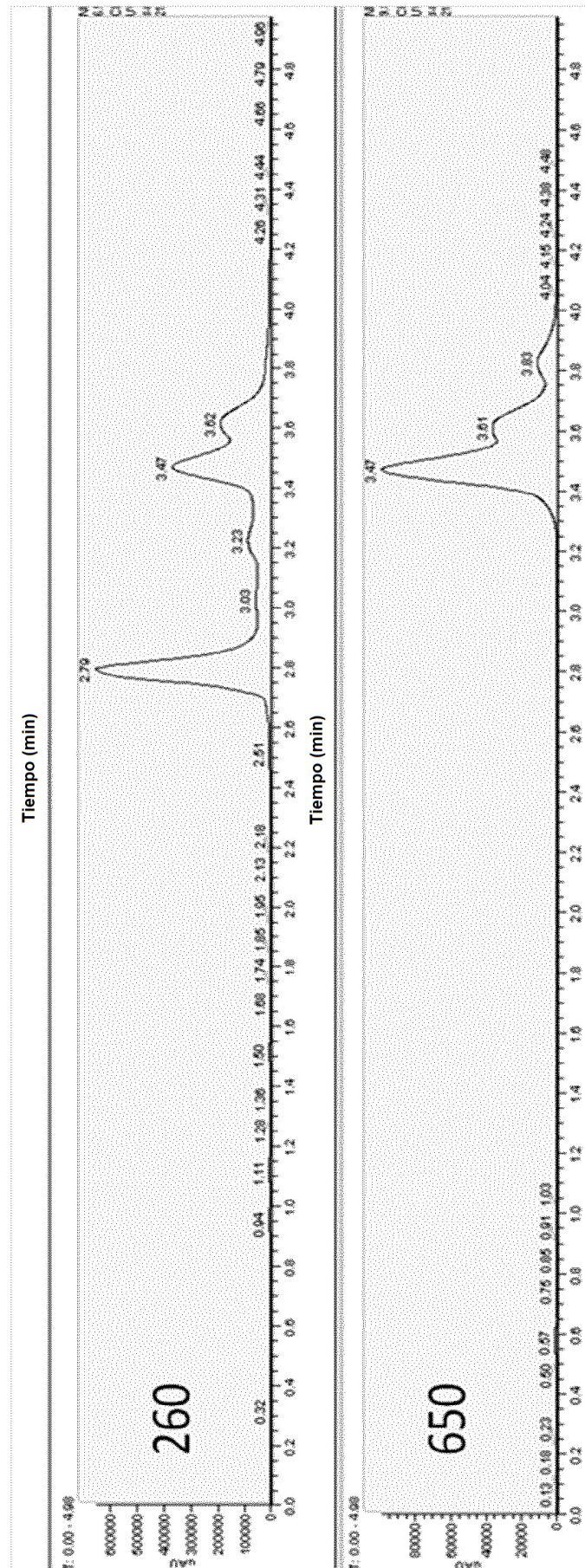


Figura 11C-1

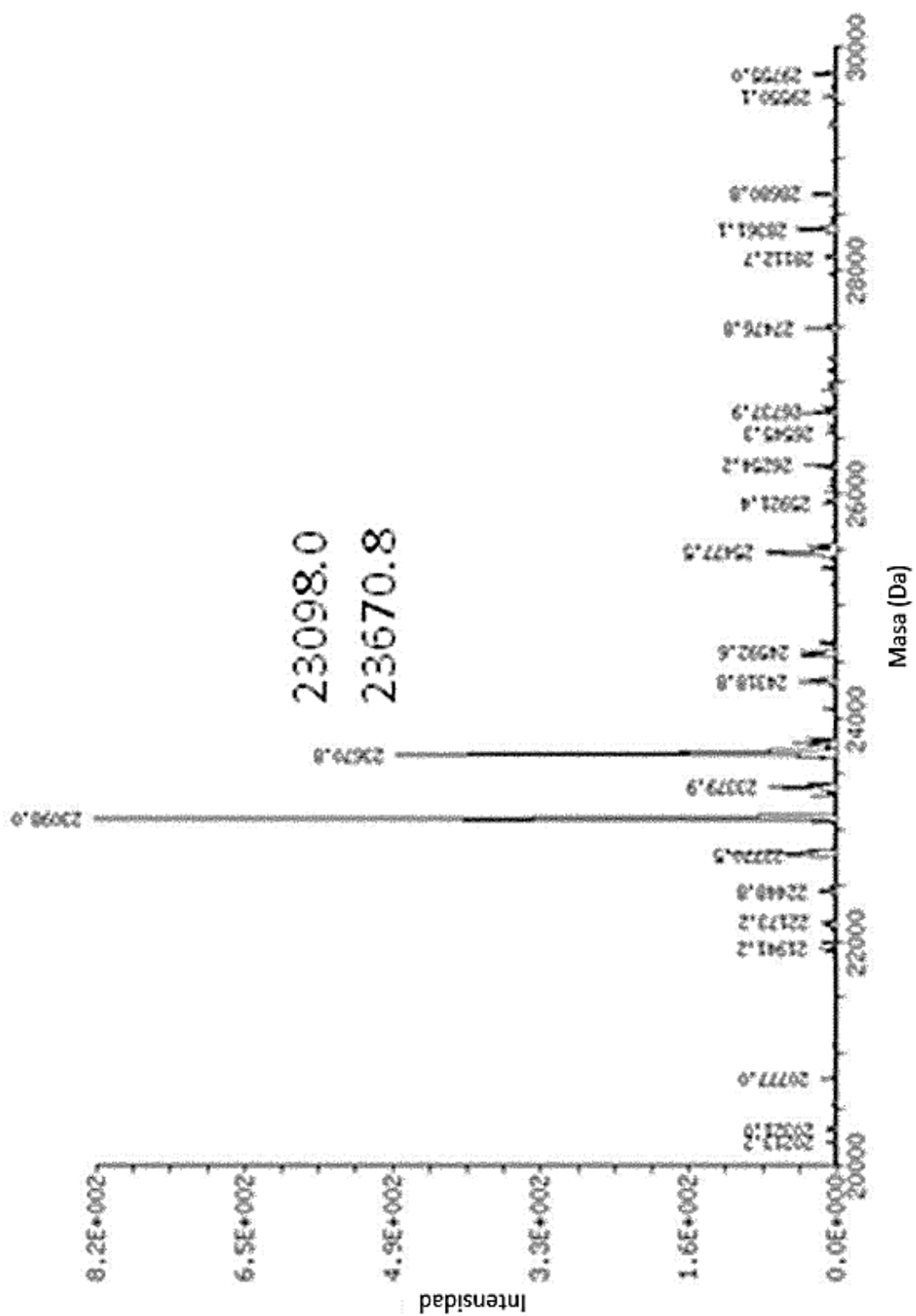


Figura 11C-2



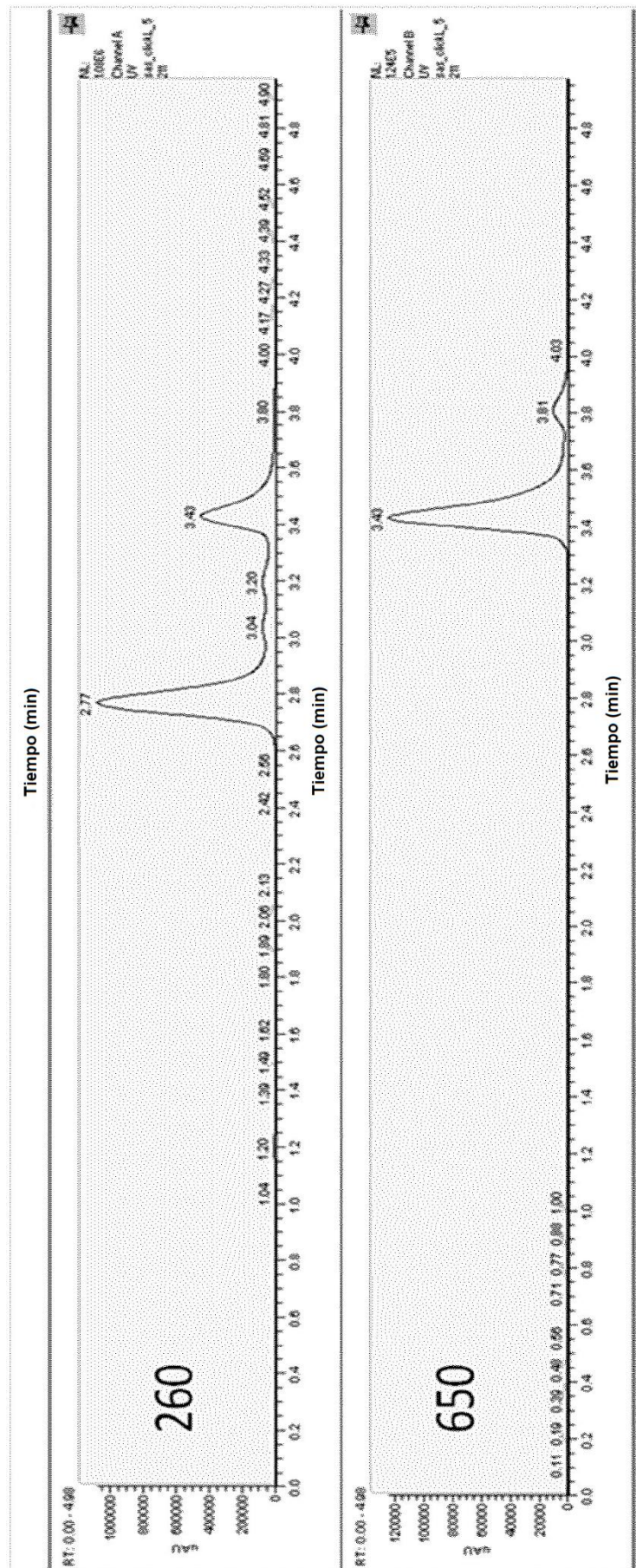


Figura 11D-1

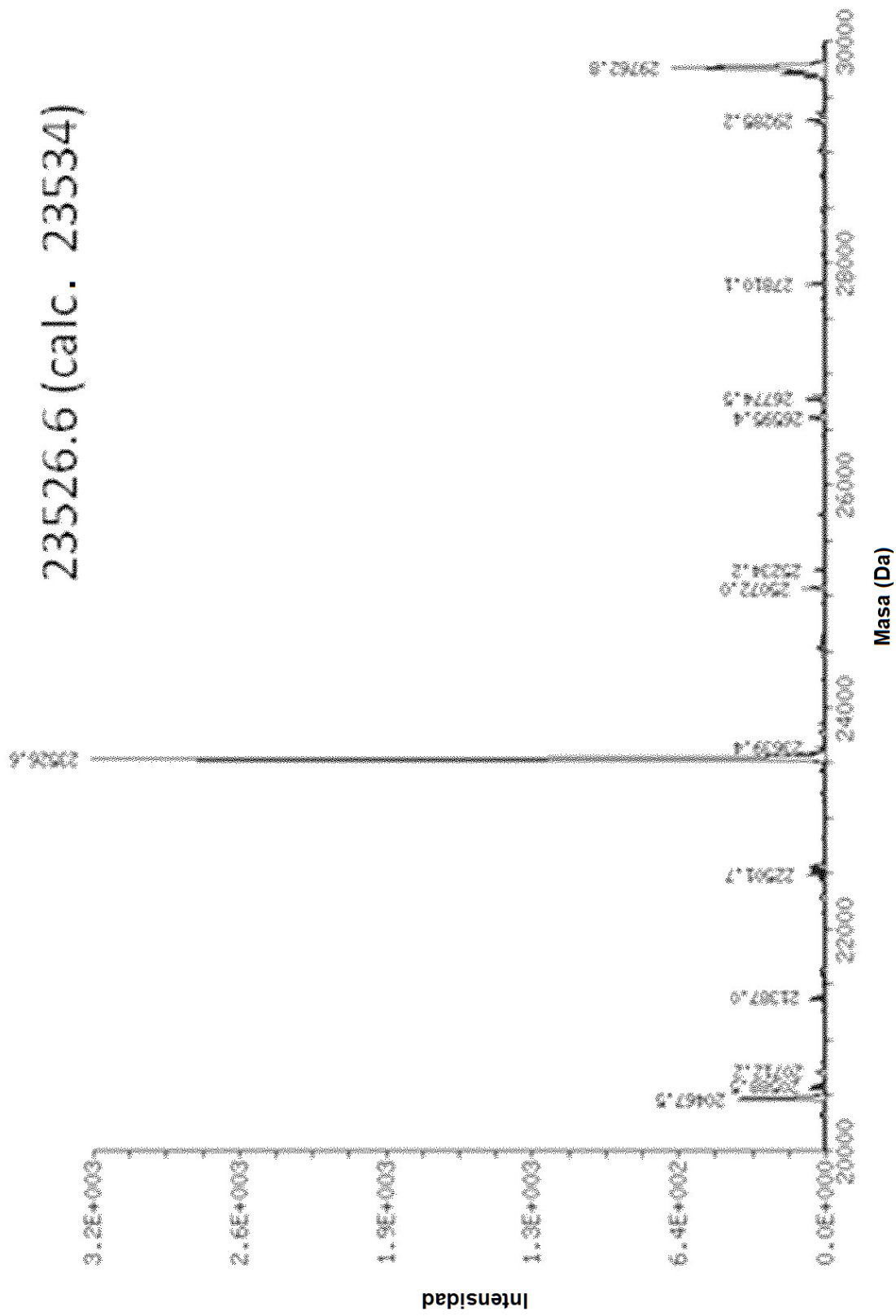
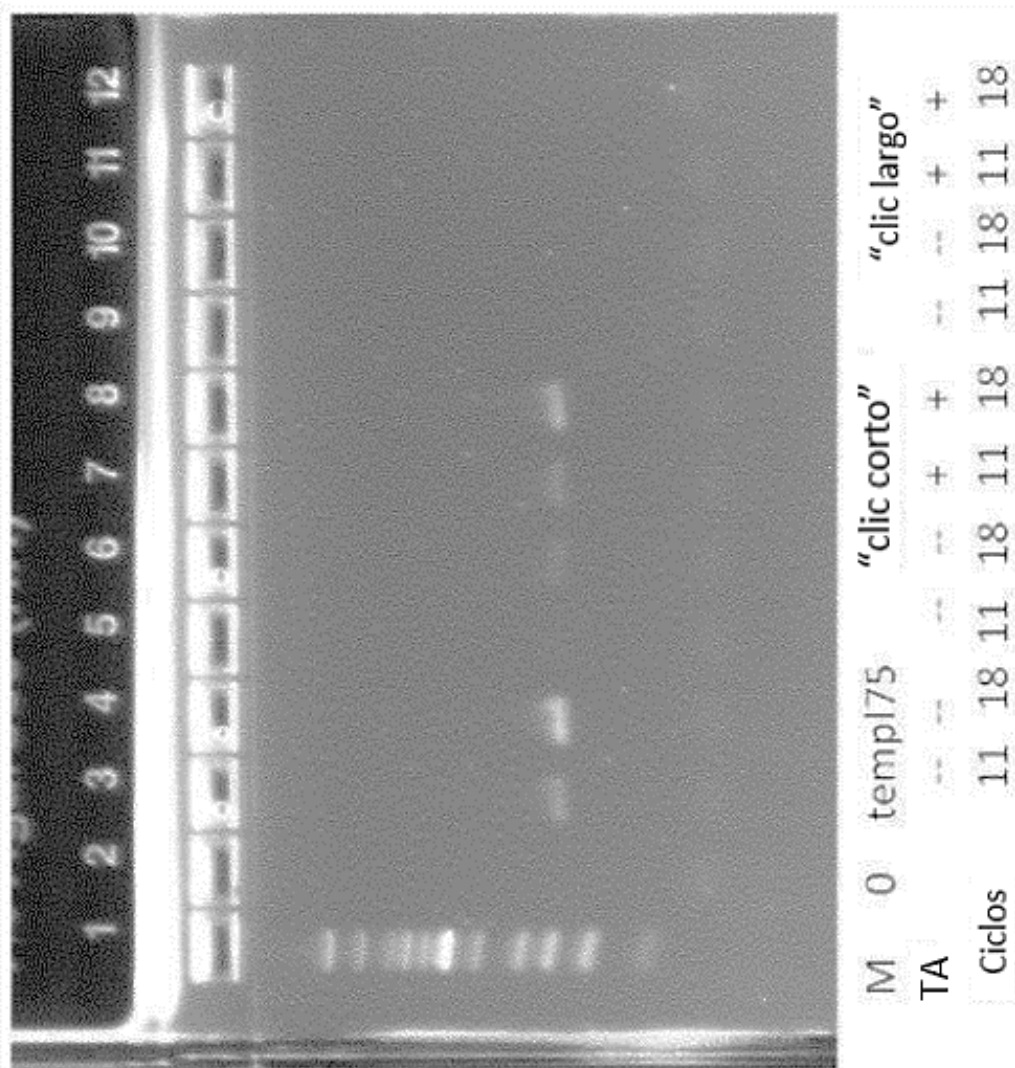


Figura 11D-2



**Figura 11E**



Figura 12A

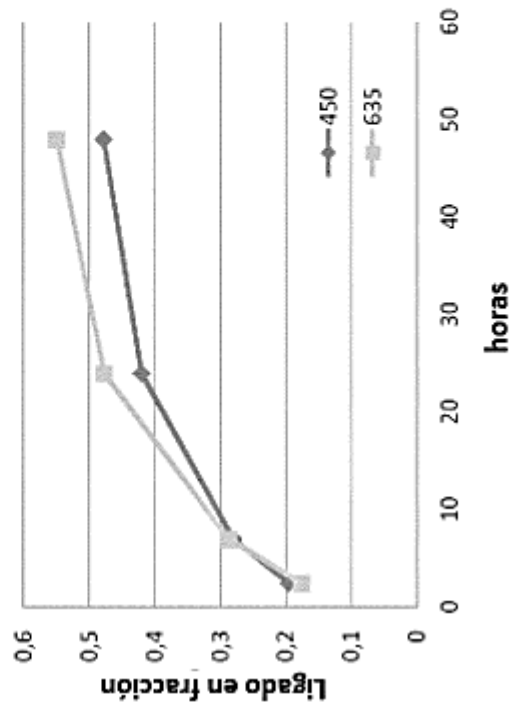


Figura 12C

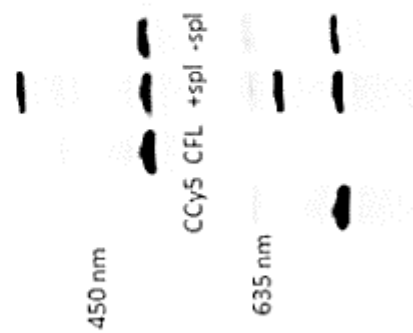
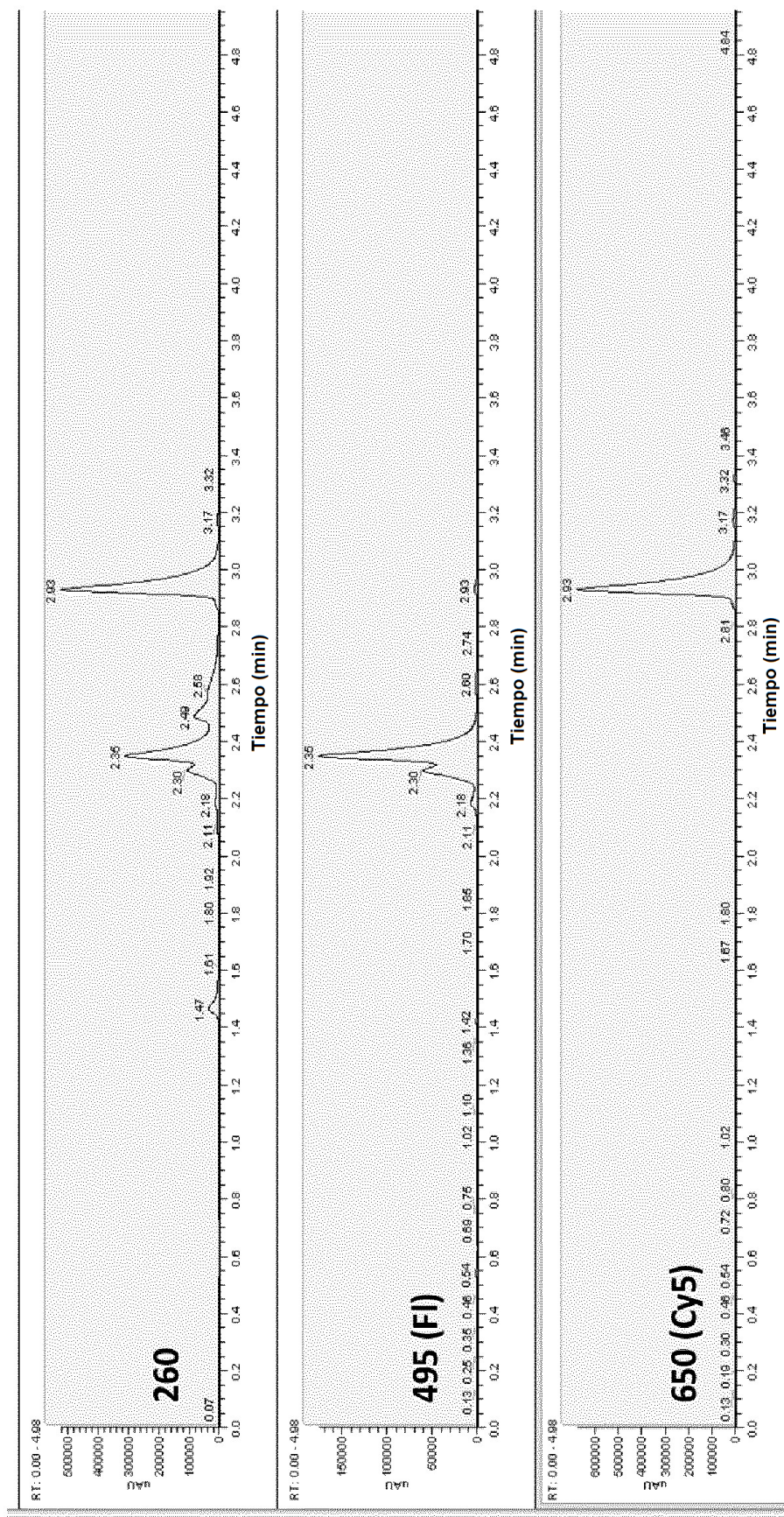


Figura 12B



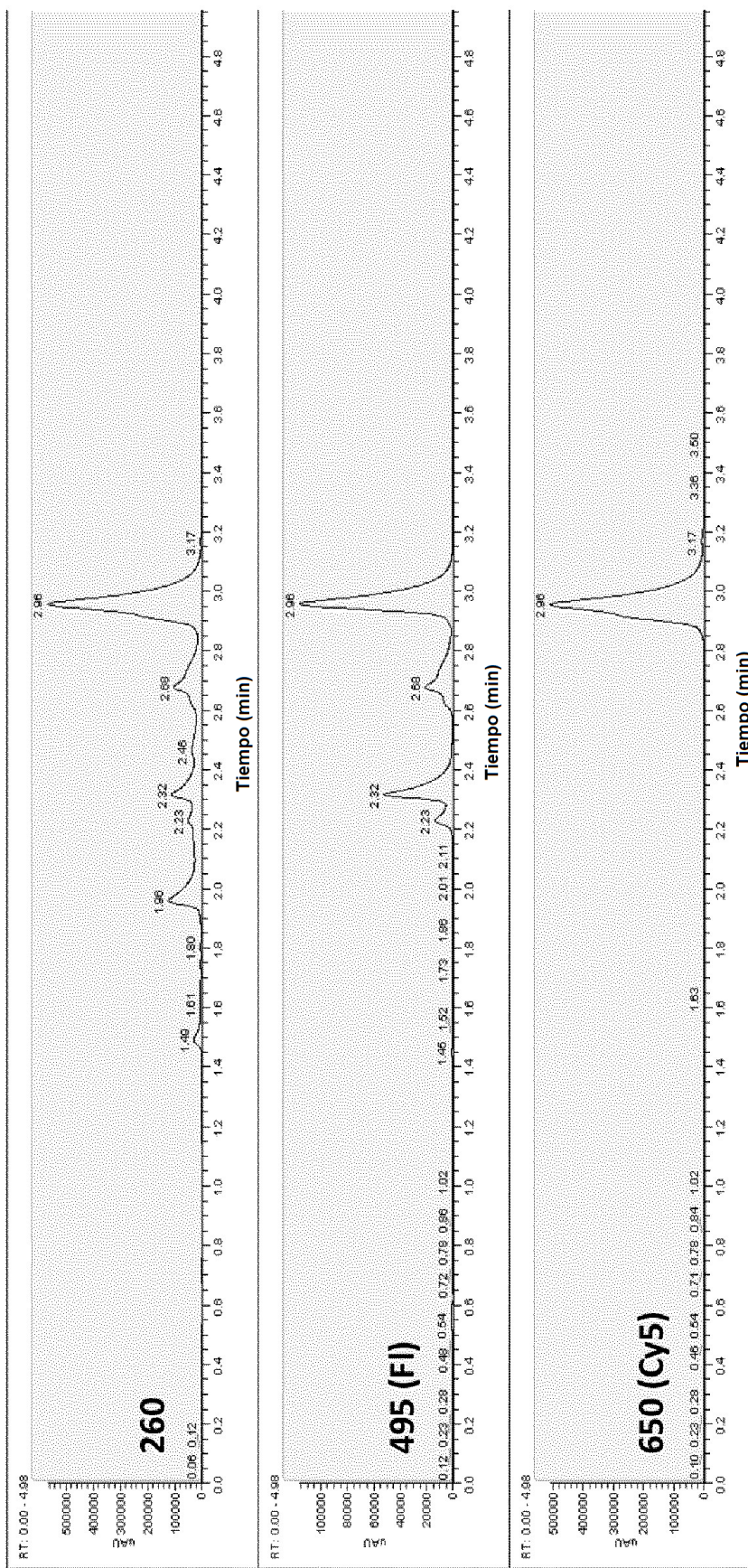
Sin férula: incubación durante 7 días con 50  $\mu$ M oligonucleótidos



Solo materiales de partida: lodo dt\_12\_Fam(4339) y Cy5\_21\_TioFosfato (7329)

Figura 12D-1

**Férula: Incubación durante 7 días con 50  $\mu$ M oligonucleótidos**



**Producto (11542)**

**Figura 12D-2**



Sin férula: incubación durante 8 días con 400 uM oligonucleótidos

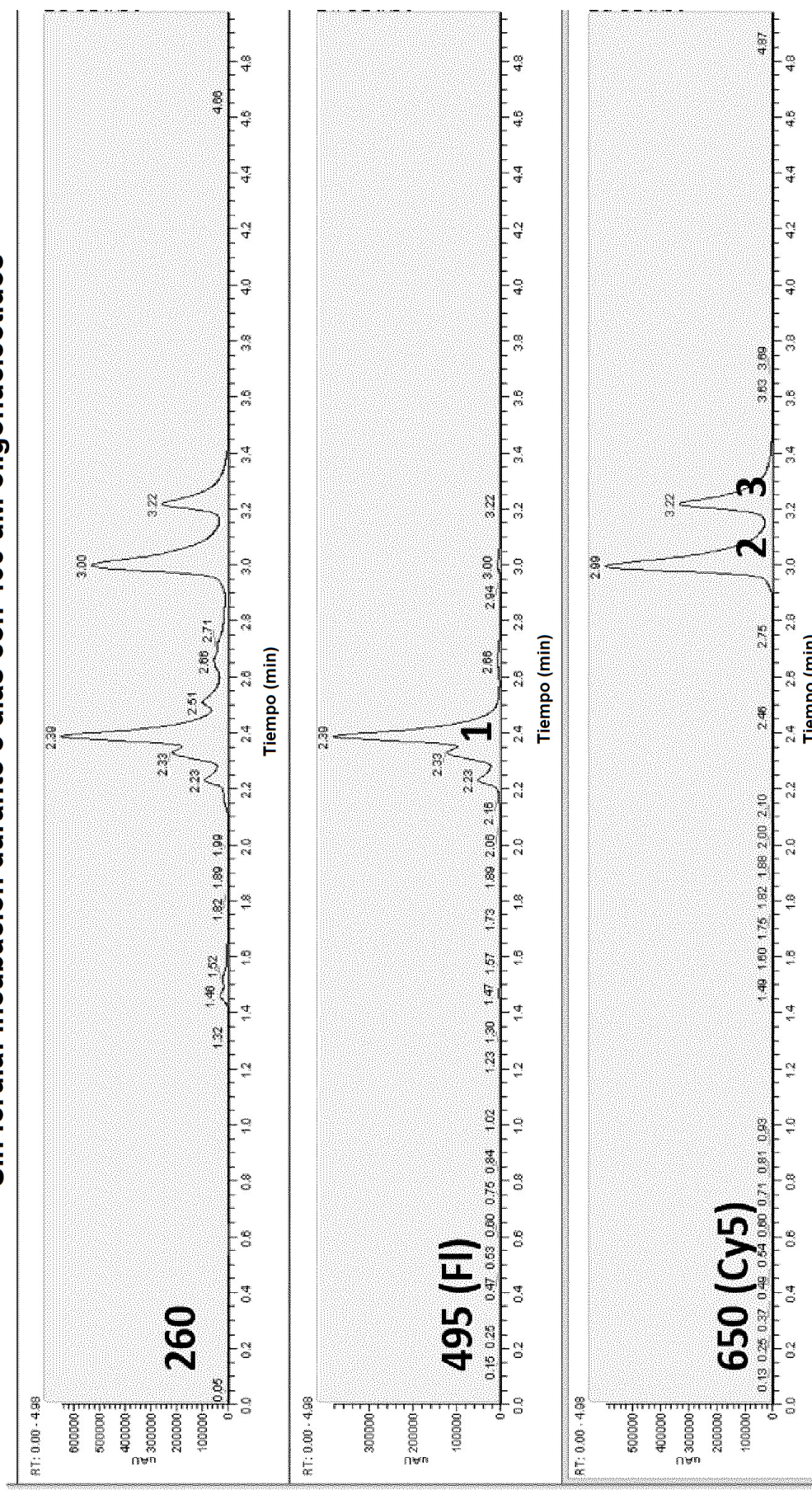
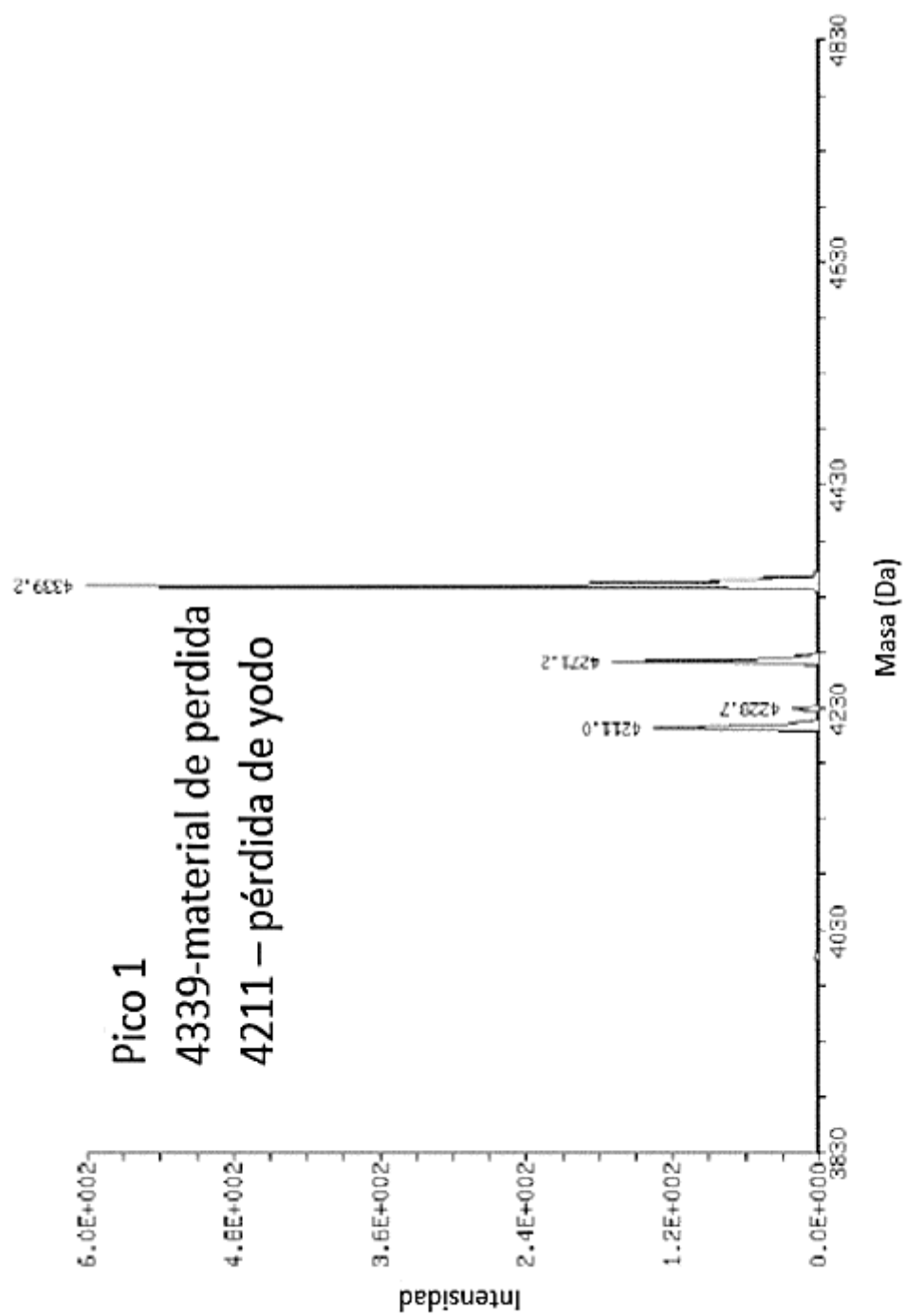


Figura 12E-1



**Figura 12E-2**

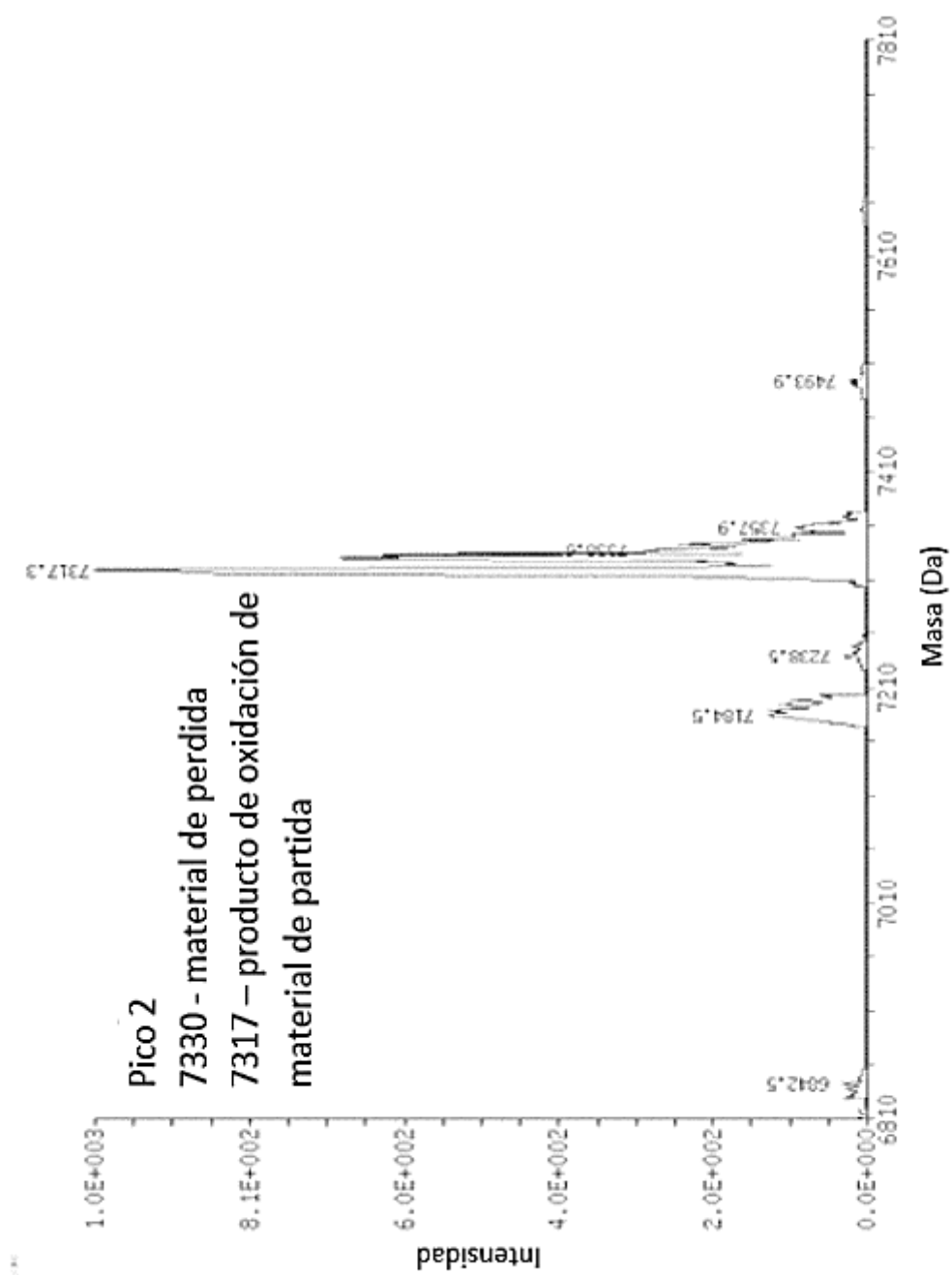


Figura 12F-3

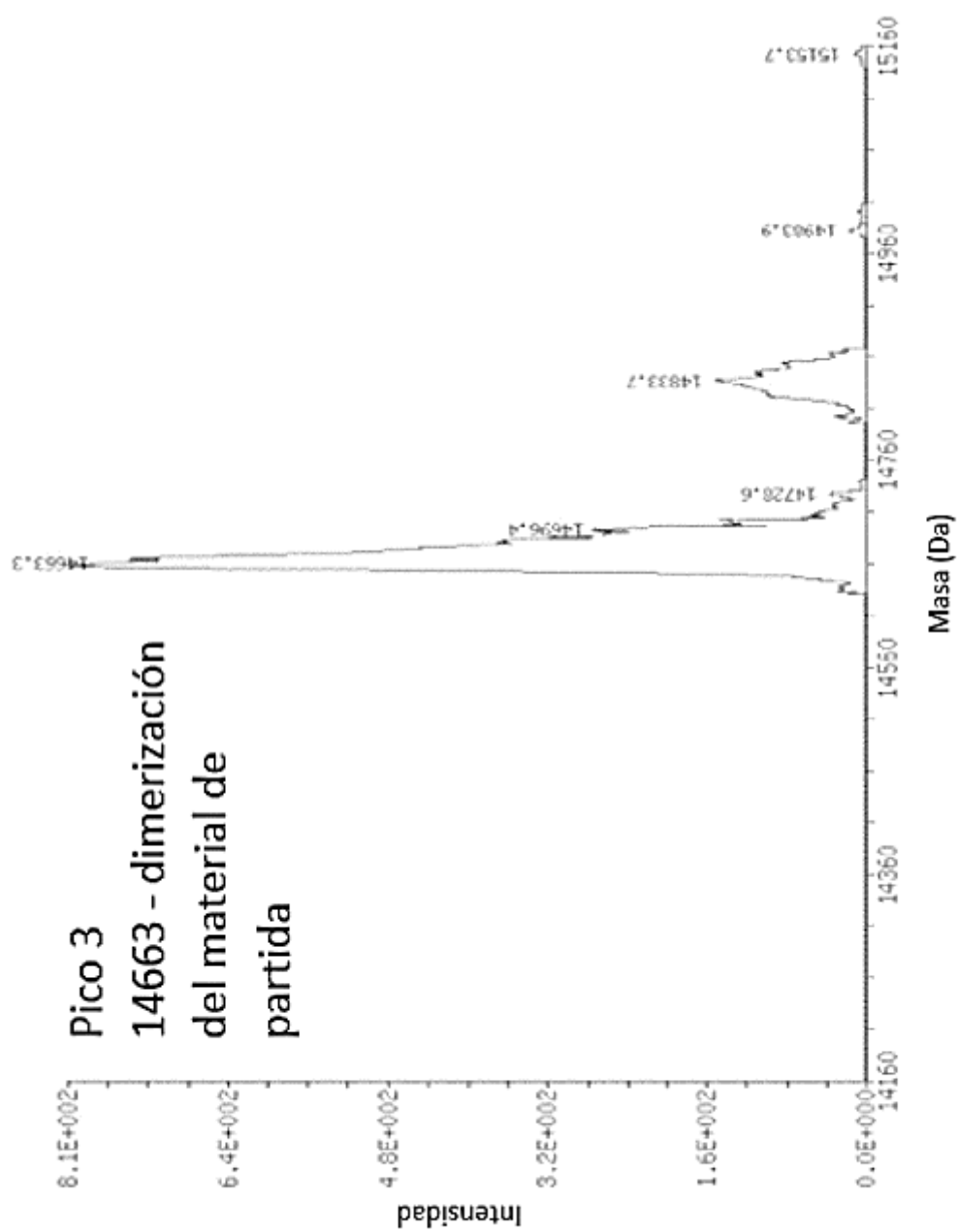
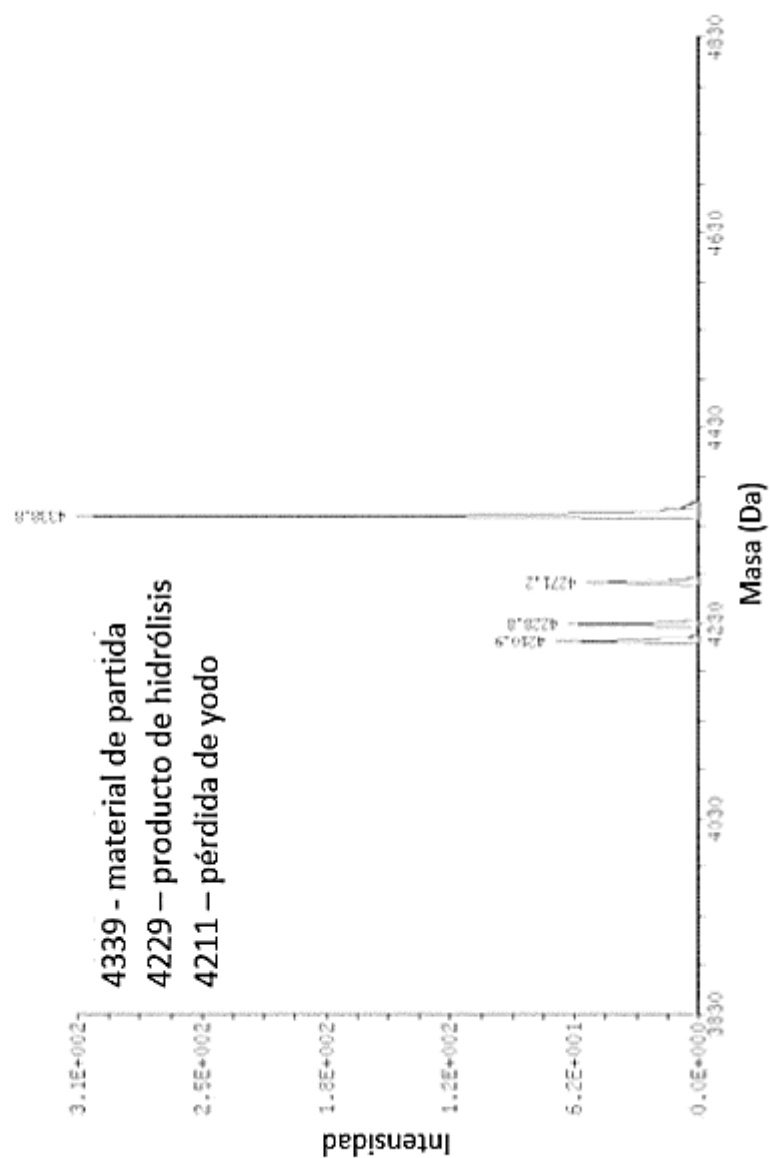


Figura 12E-4

**Oligonucleótido CFL en presencia de piperidina**  
**400 eq. de piperidina durante la noche a TA**



**Figura 12F-1**

# 400 eq. de piperidina durante 2 h a 65°C

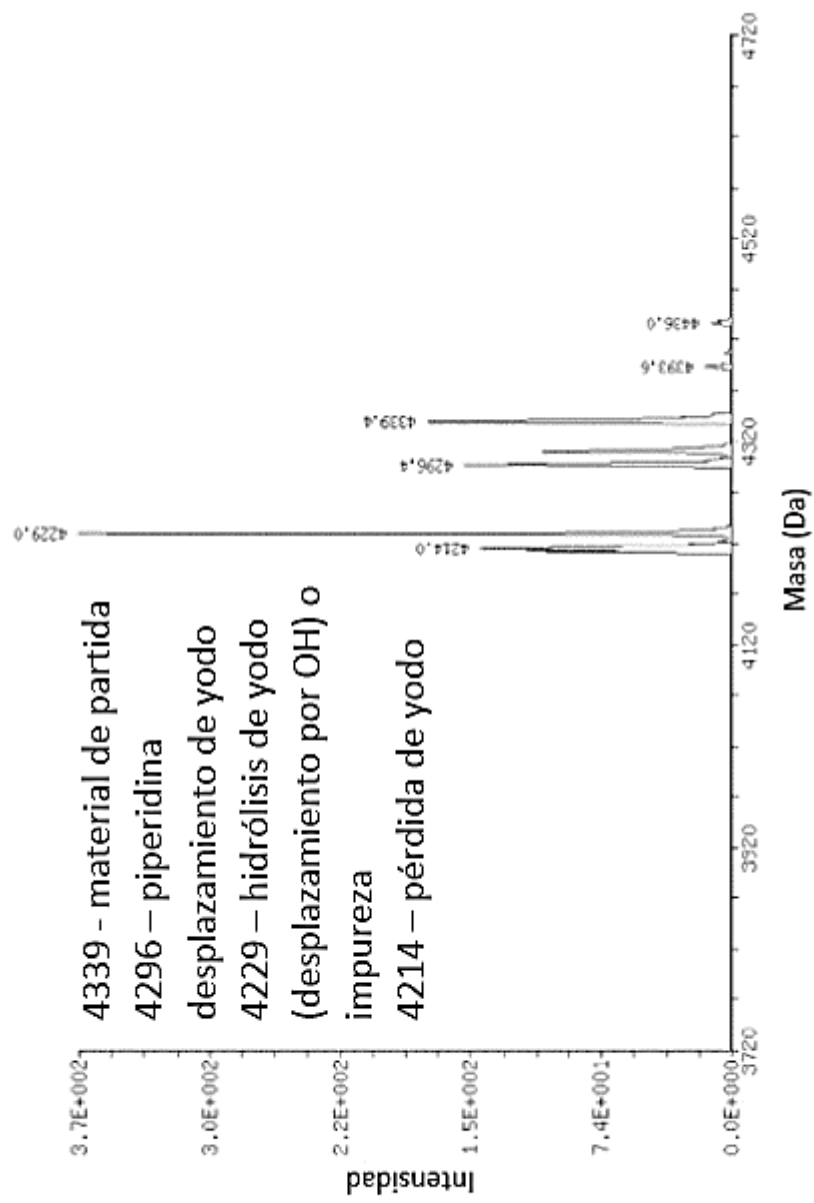


Figura 12F-2



# Ligación de férula: Oligonucleótidos CFL y CCy5 en la presencia de piperidina

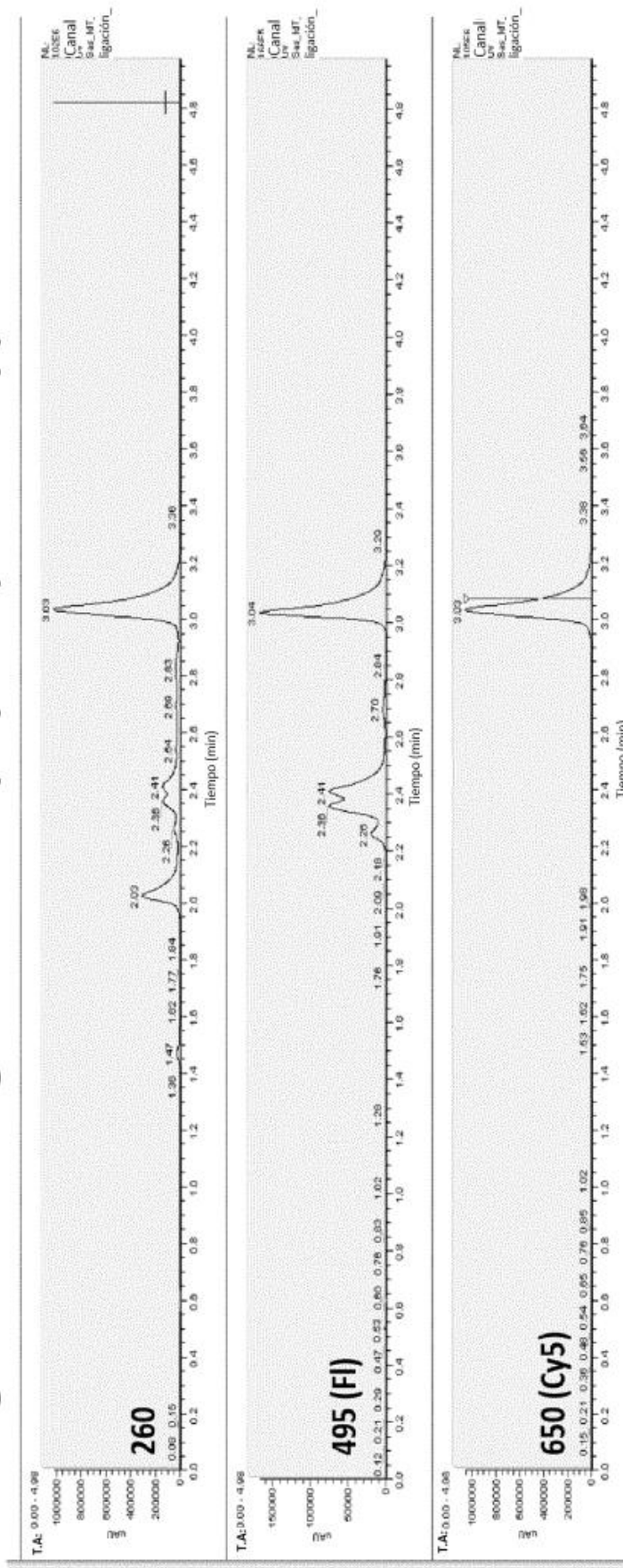


Figura 12G-1

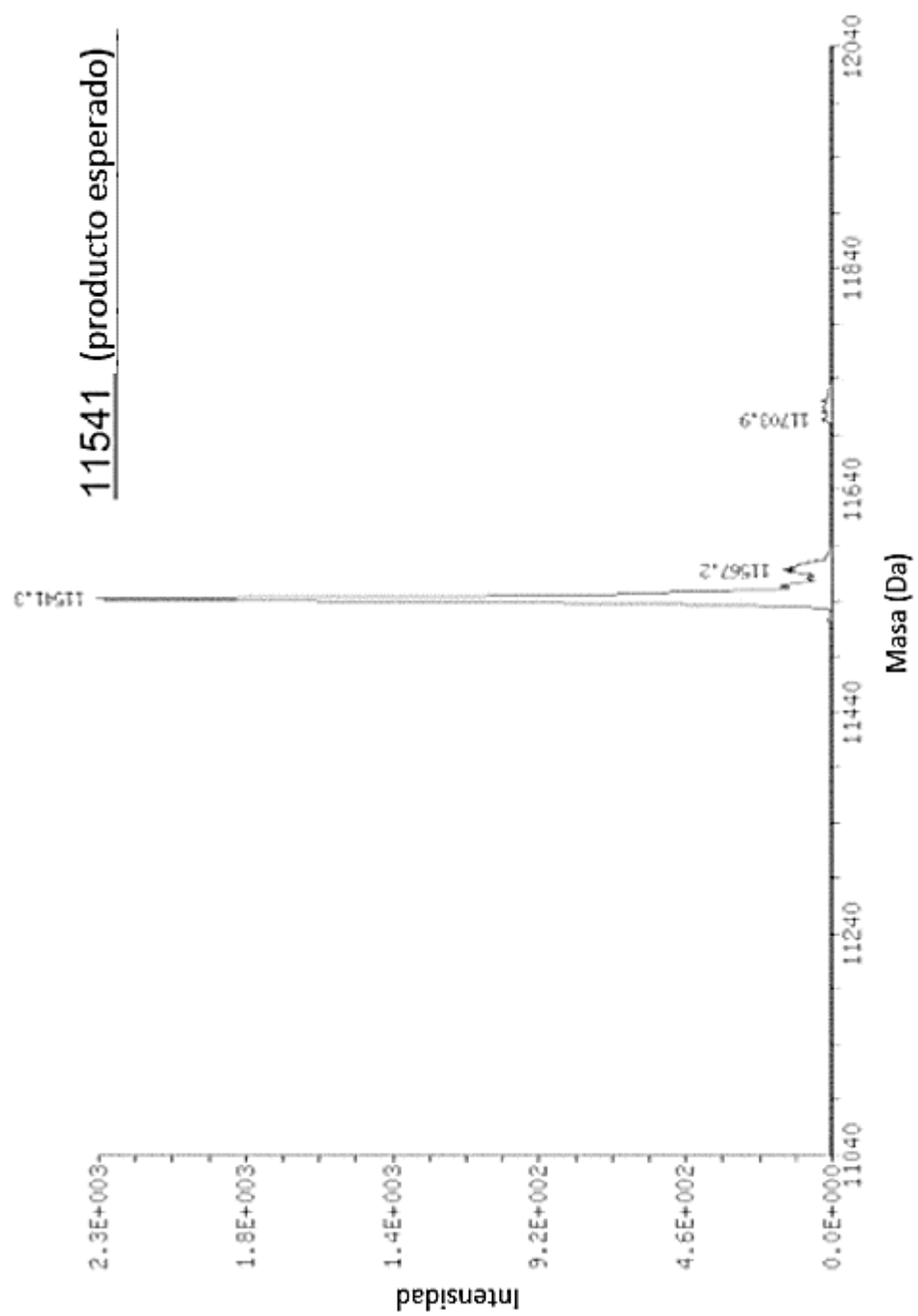
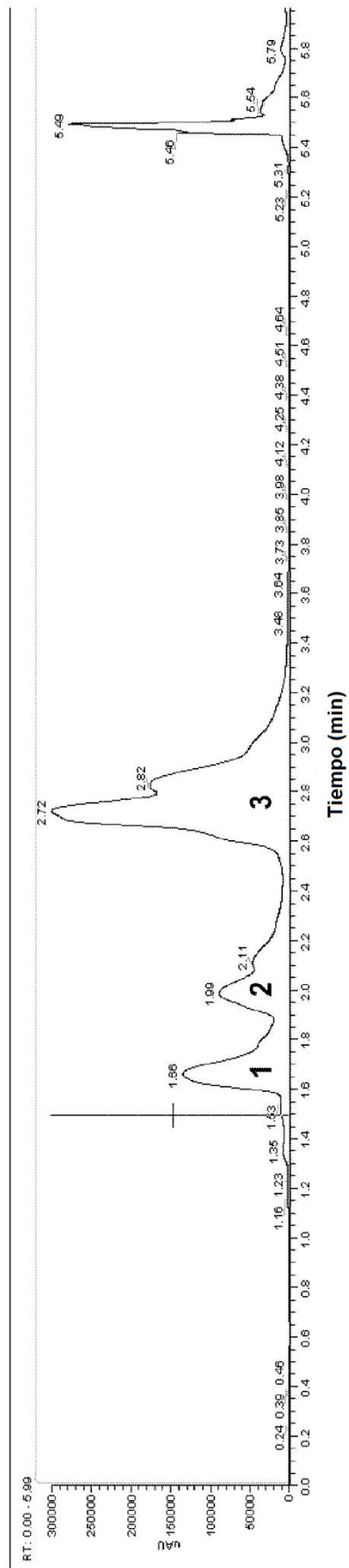
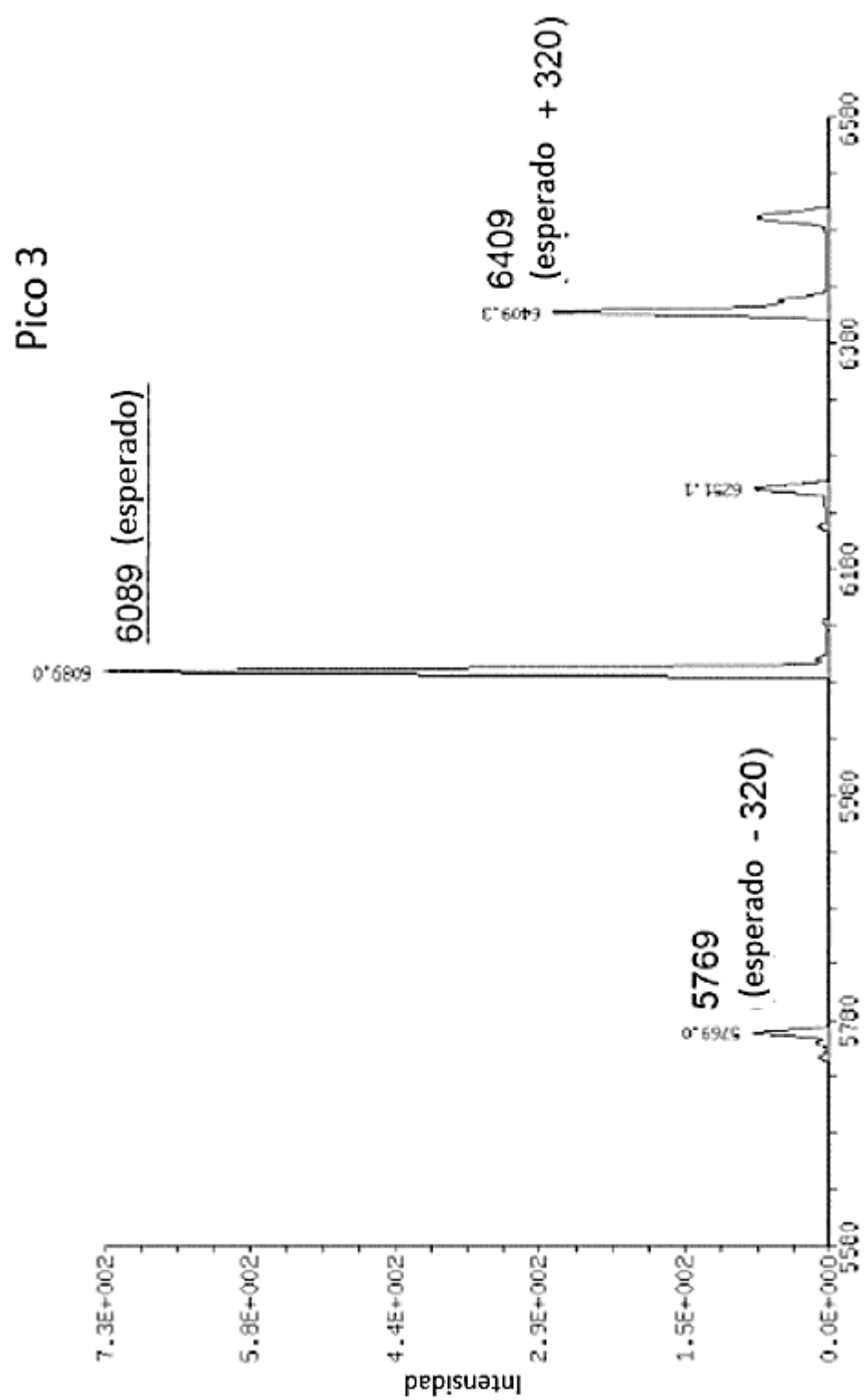


Figura 12G-2



Pico 1 - etiqueta, pico 2 - HP adenilado)

Figura 13A-1



**Figura 13A-2**

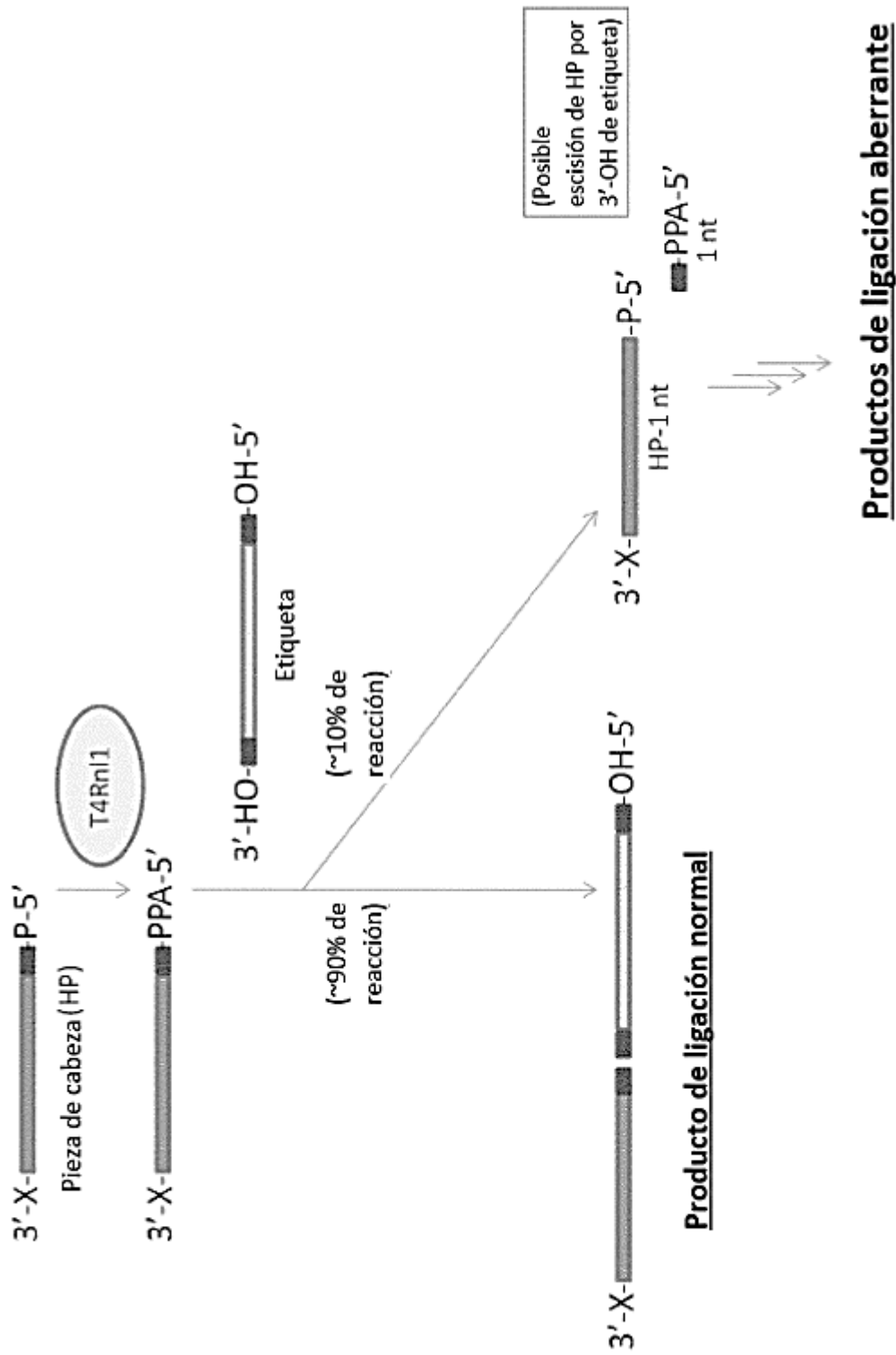


Figura 13B-1

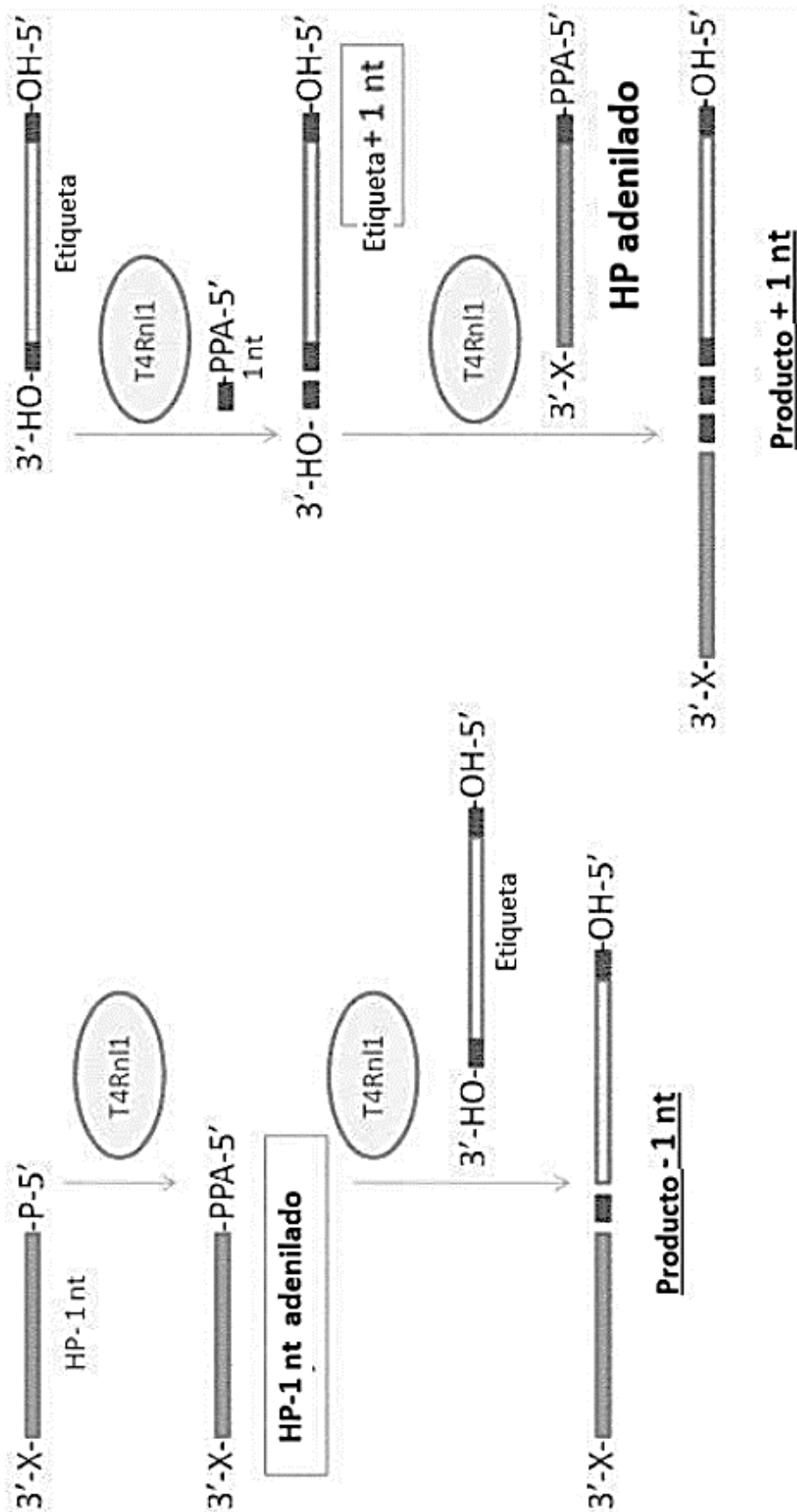


Figura 13B-2

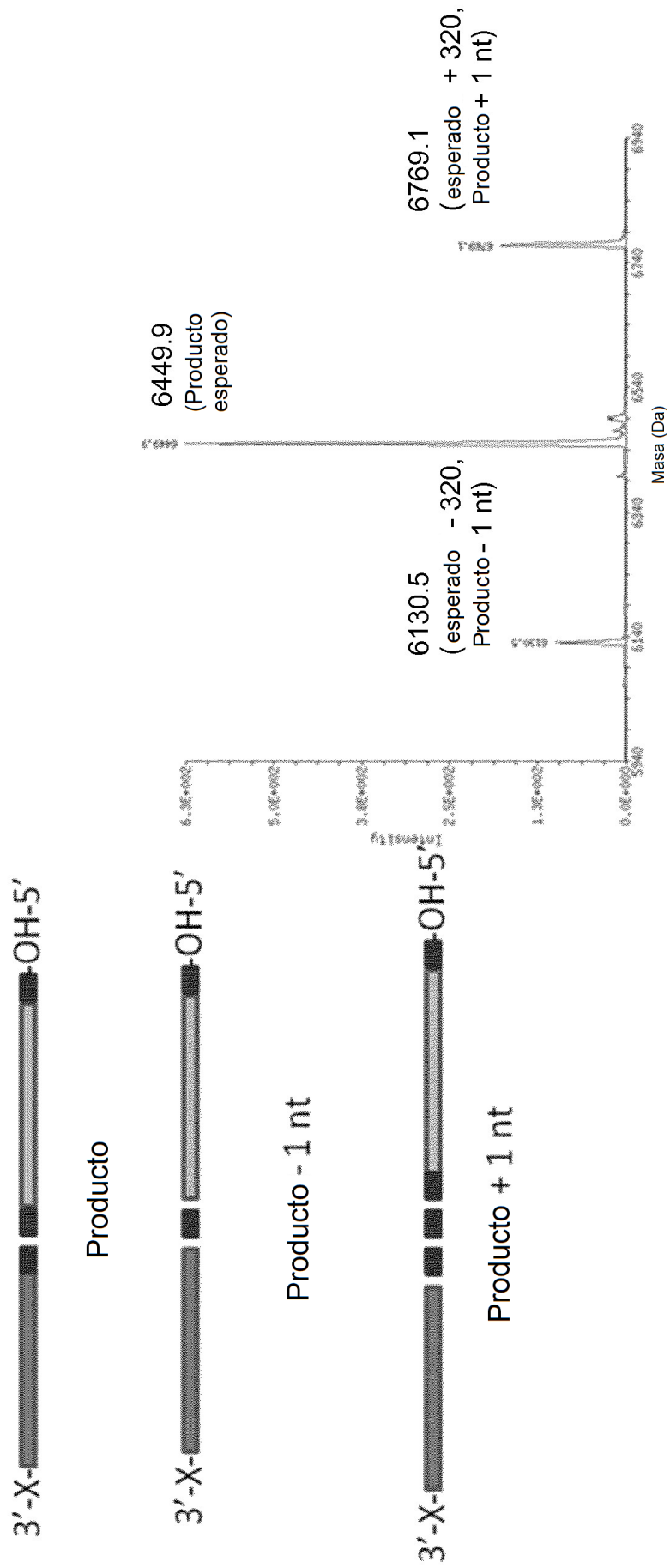
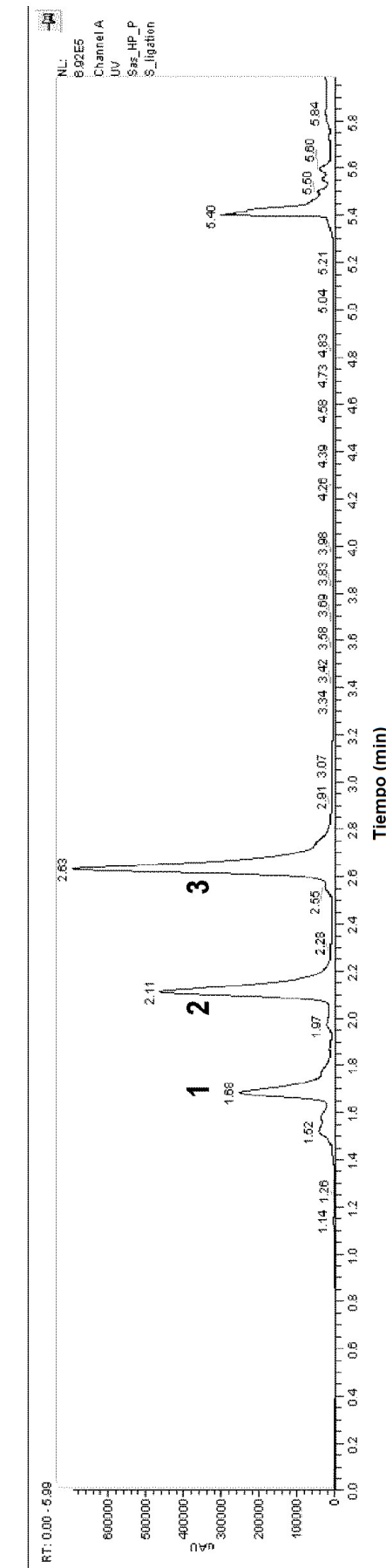


Figura 13B-3



(Pico 1- etiqueta, pico 2- HP adenilado)

Figura 13C-1



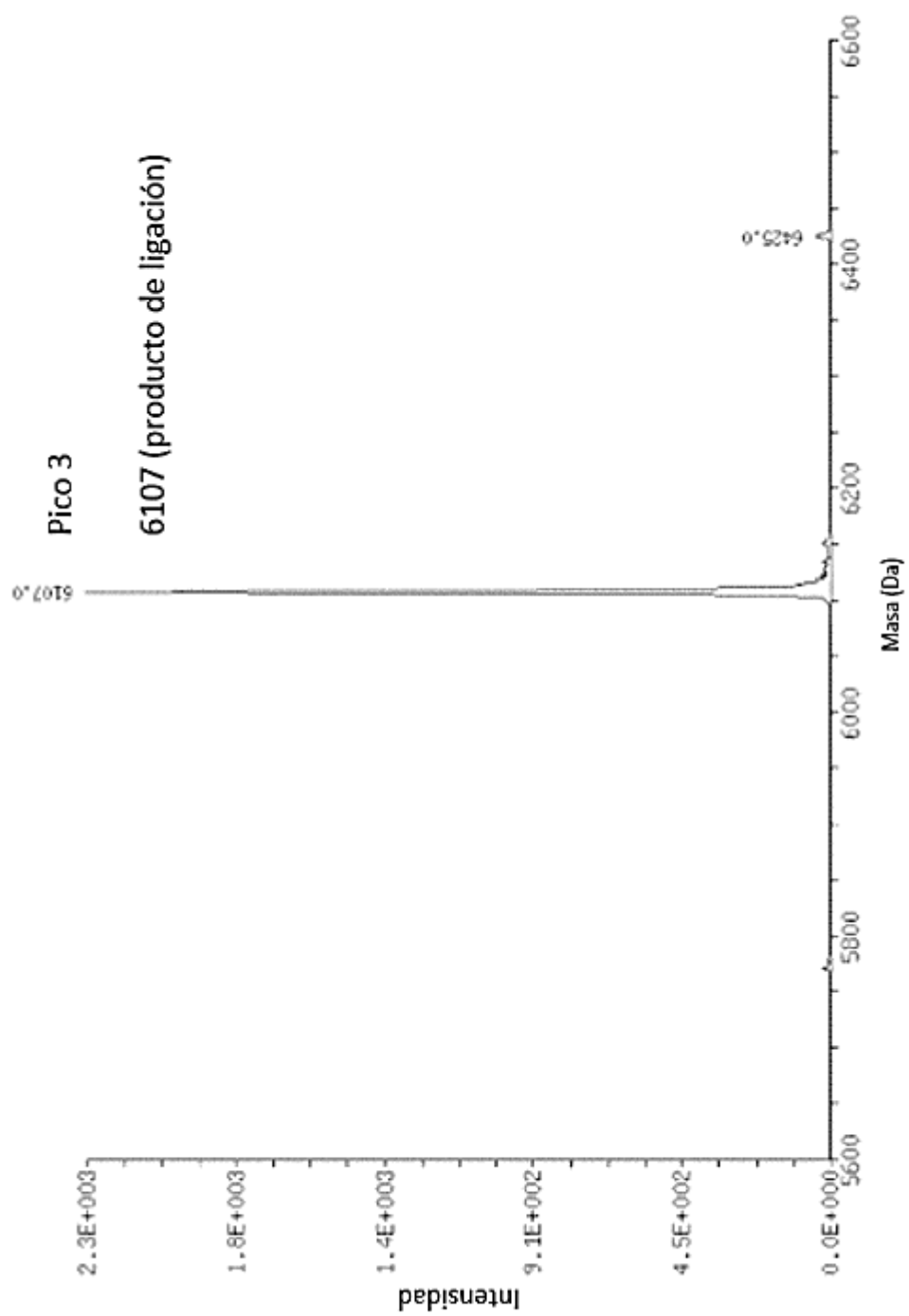


Figura 13C-2

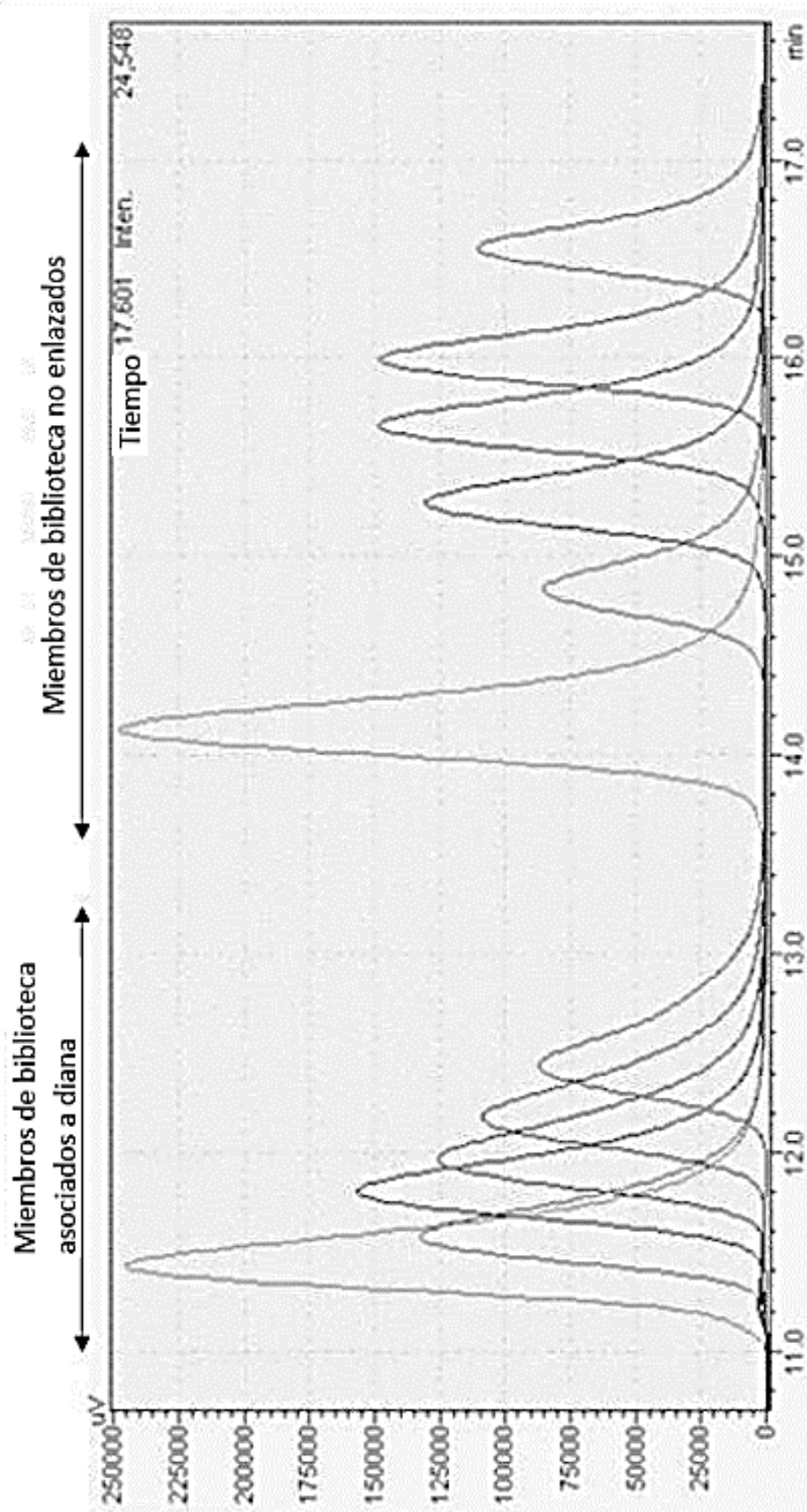


Figura 14

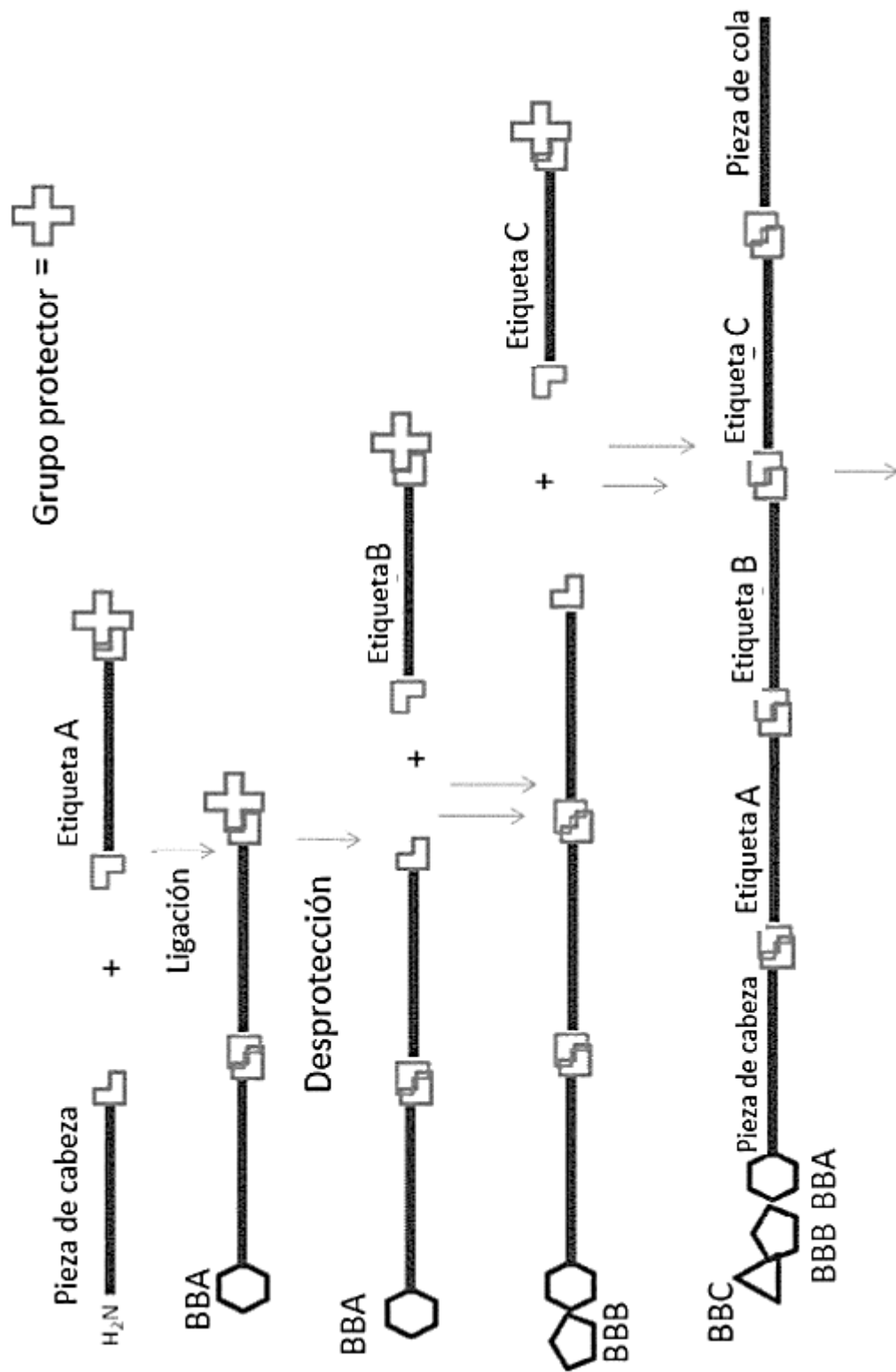


Figura 15A

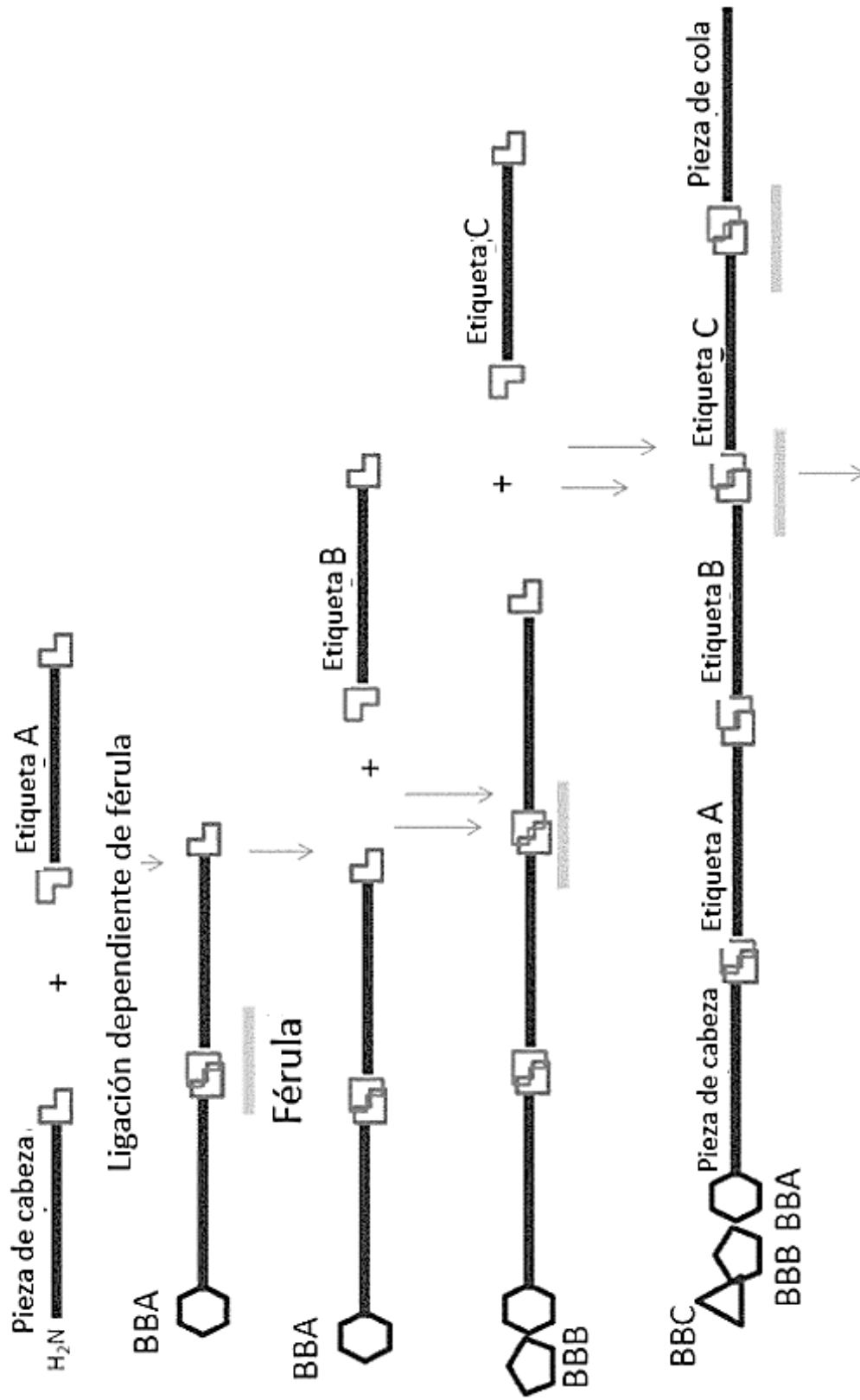


Figura 15B

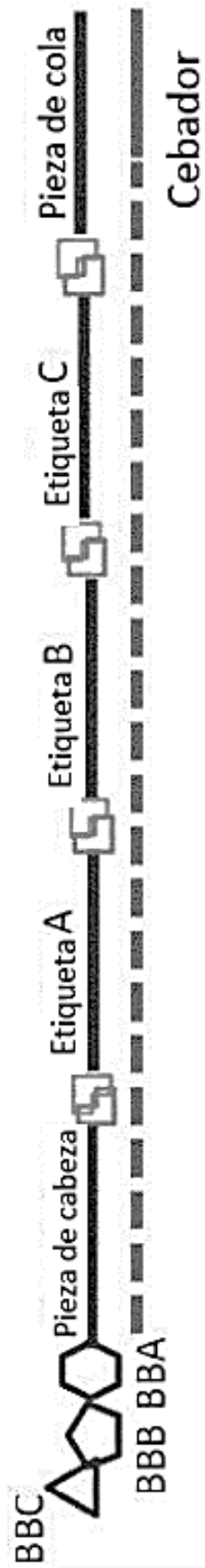


Figura 15C

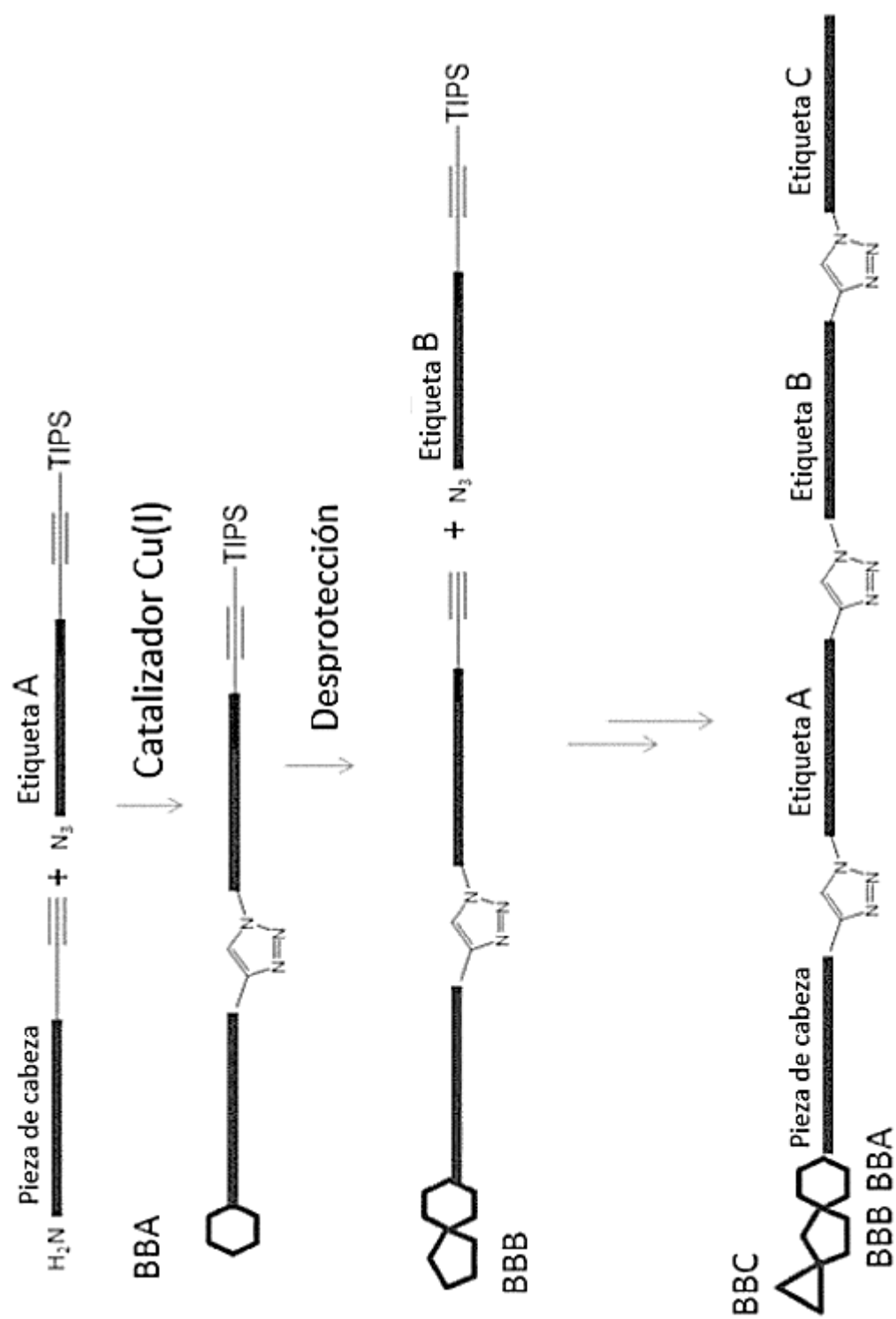
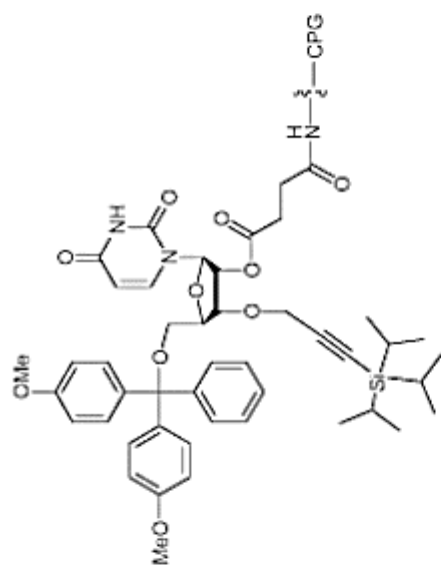
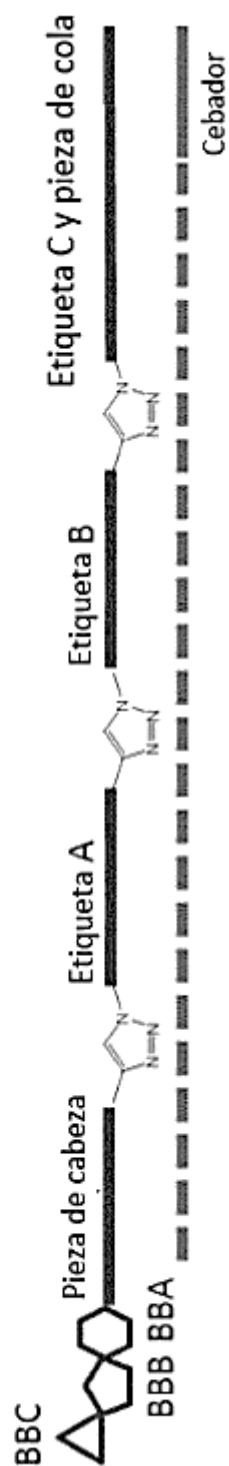


Figura 16A



**Figura 16B**



Extensión de cebador por polimerasa de ADN capaz de lectura por juntas ligadas químicamente

**Figura 16C**

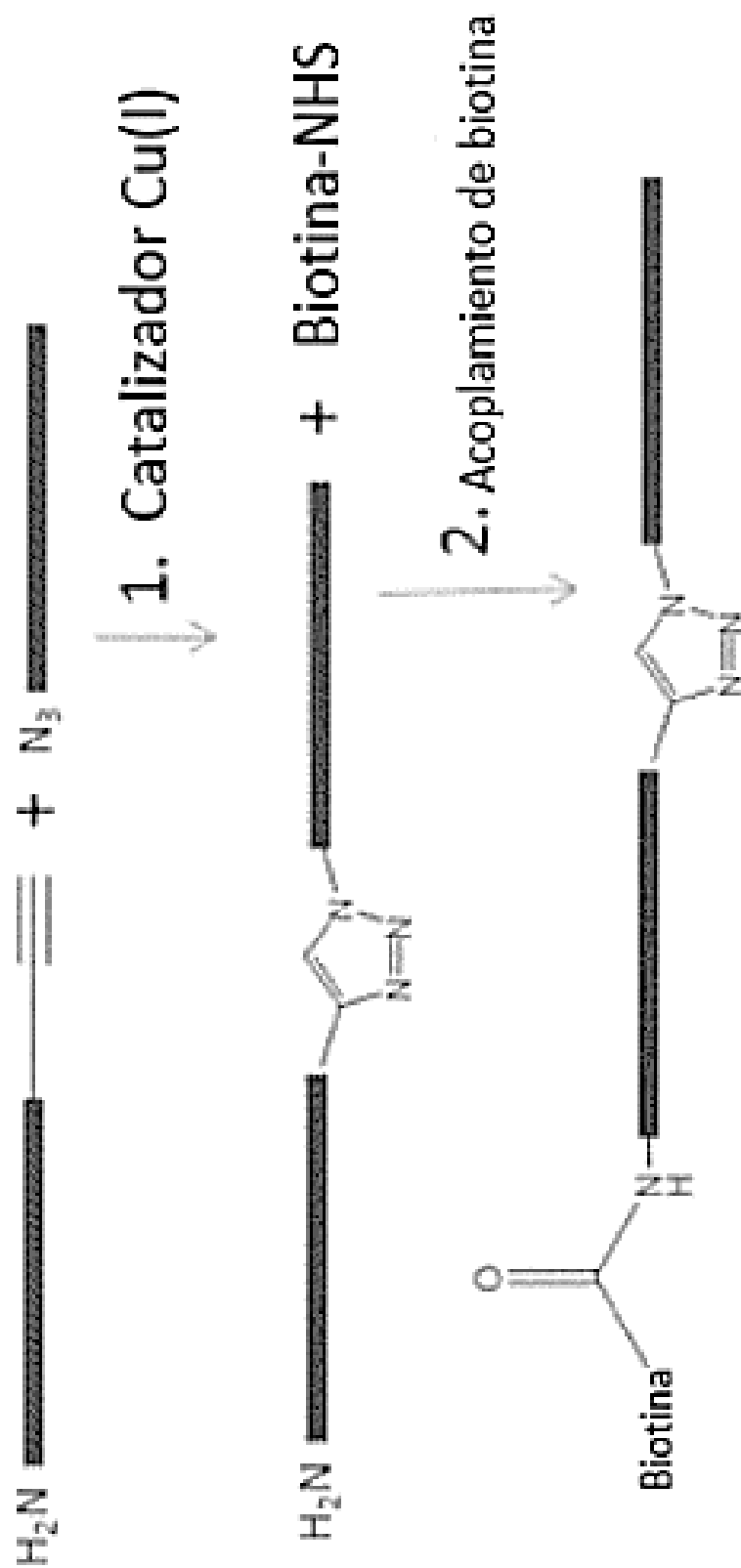


Figura 17A



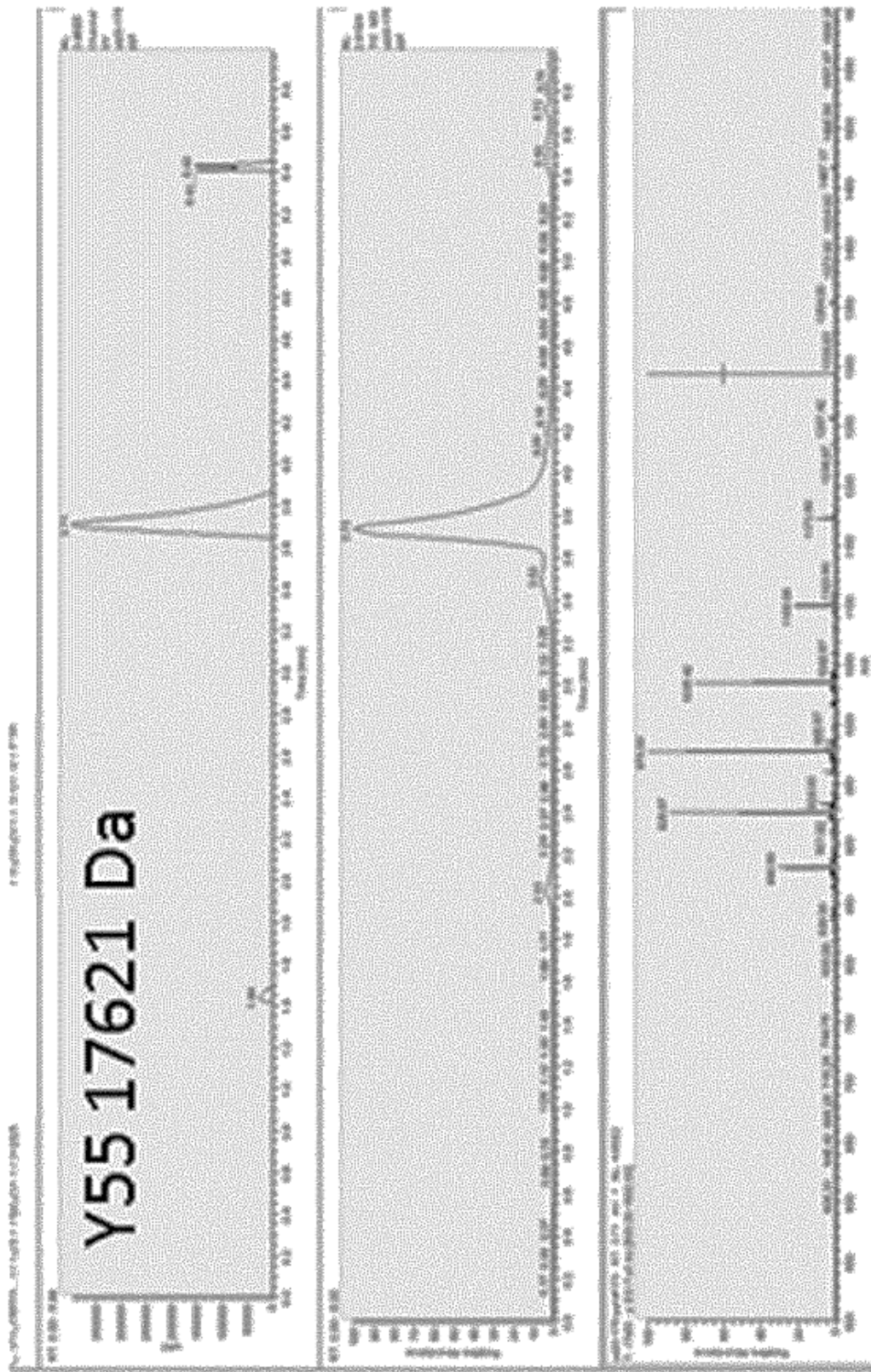


Figura 17B



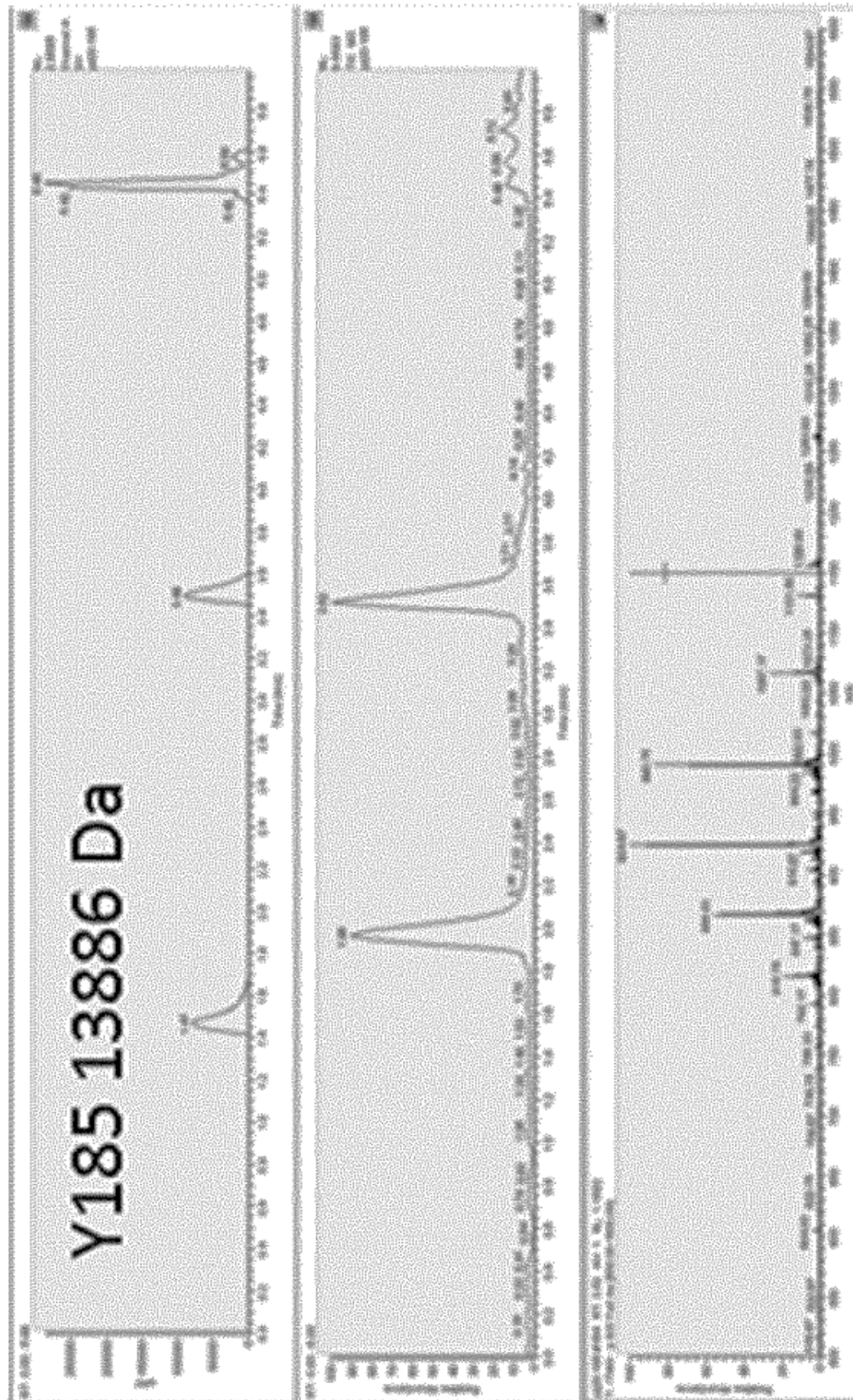


Figure 17C

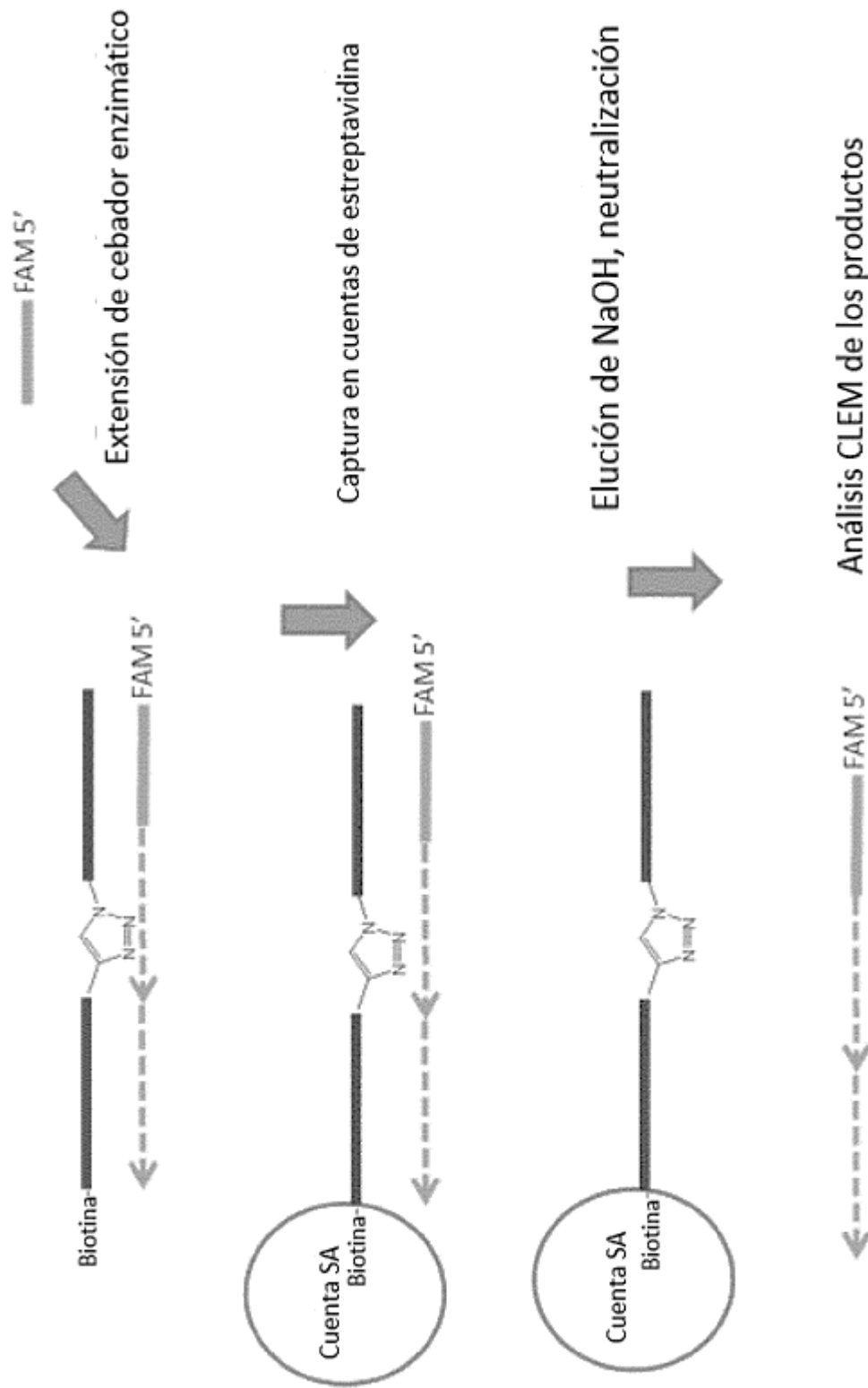


Figura 18A





# Producto Klenow Y185 13763

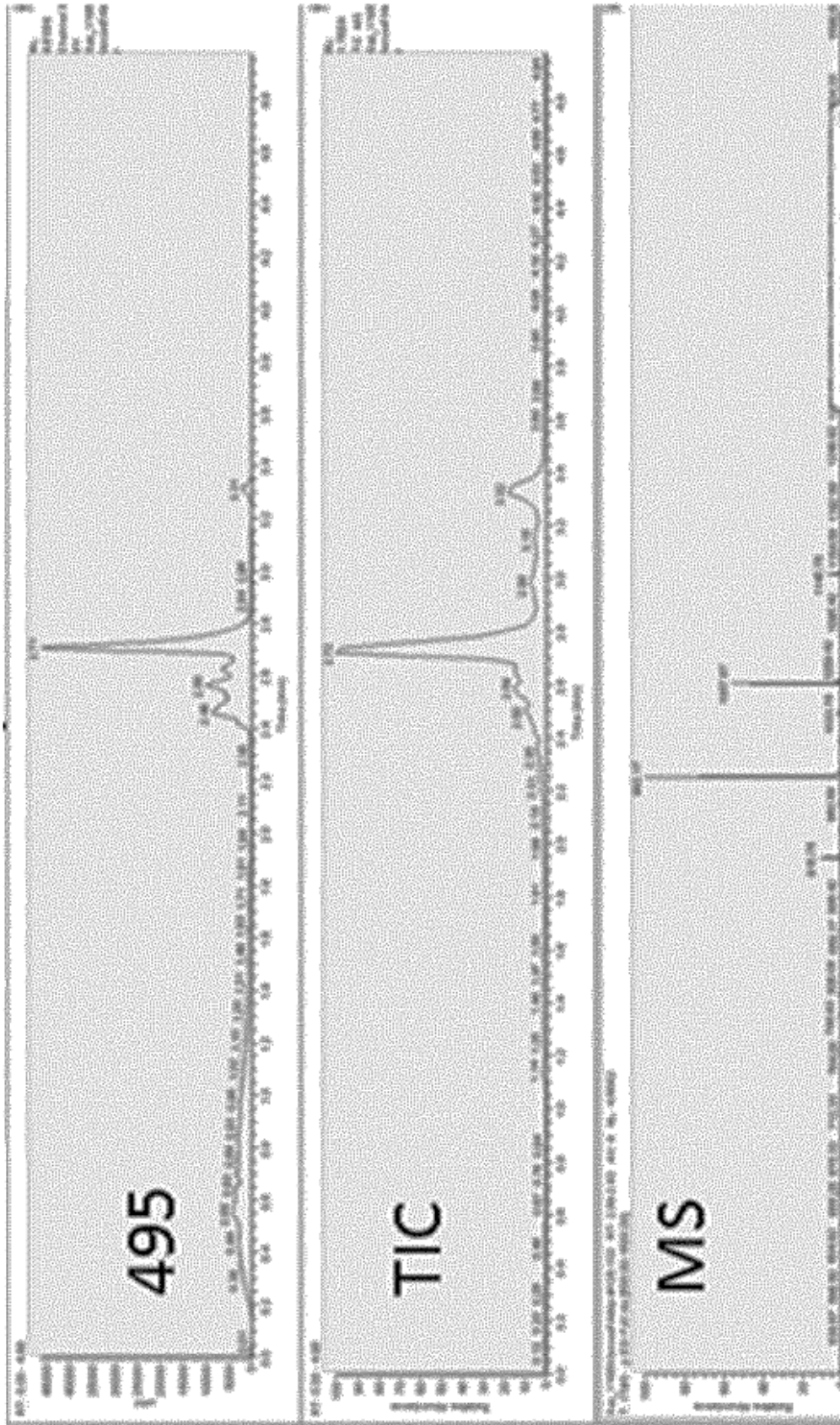


Figura 18C

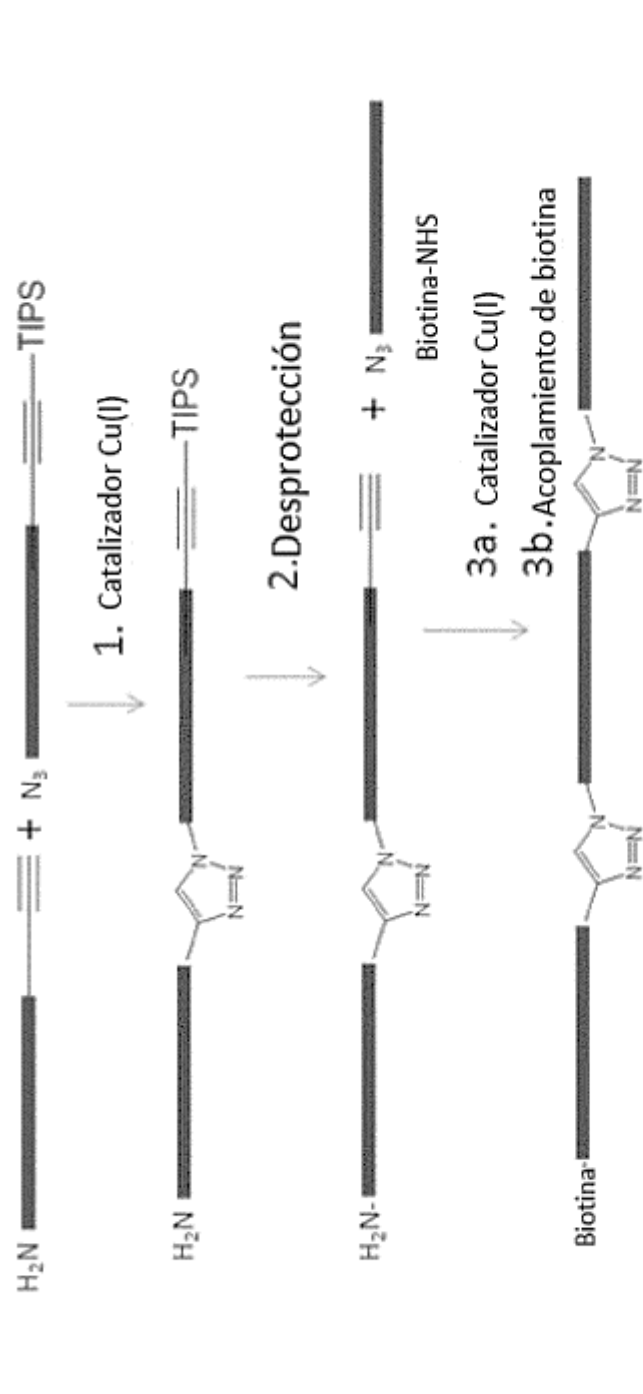


Figura 19A

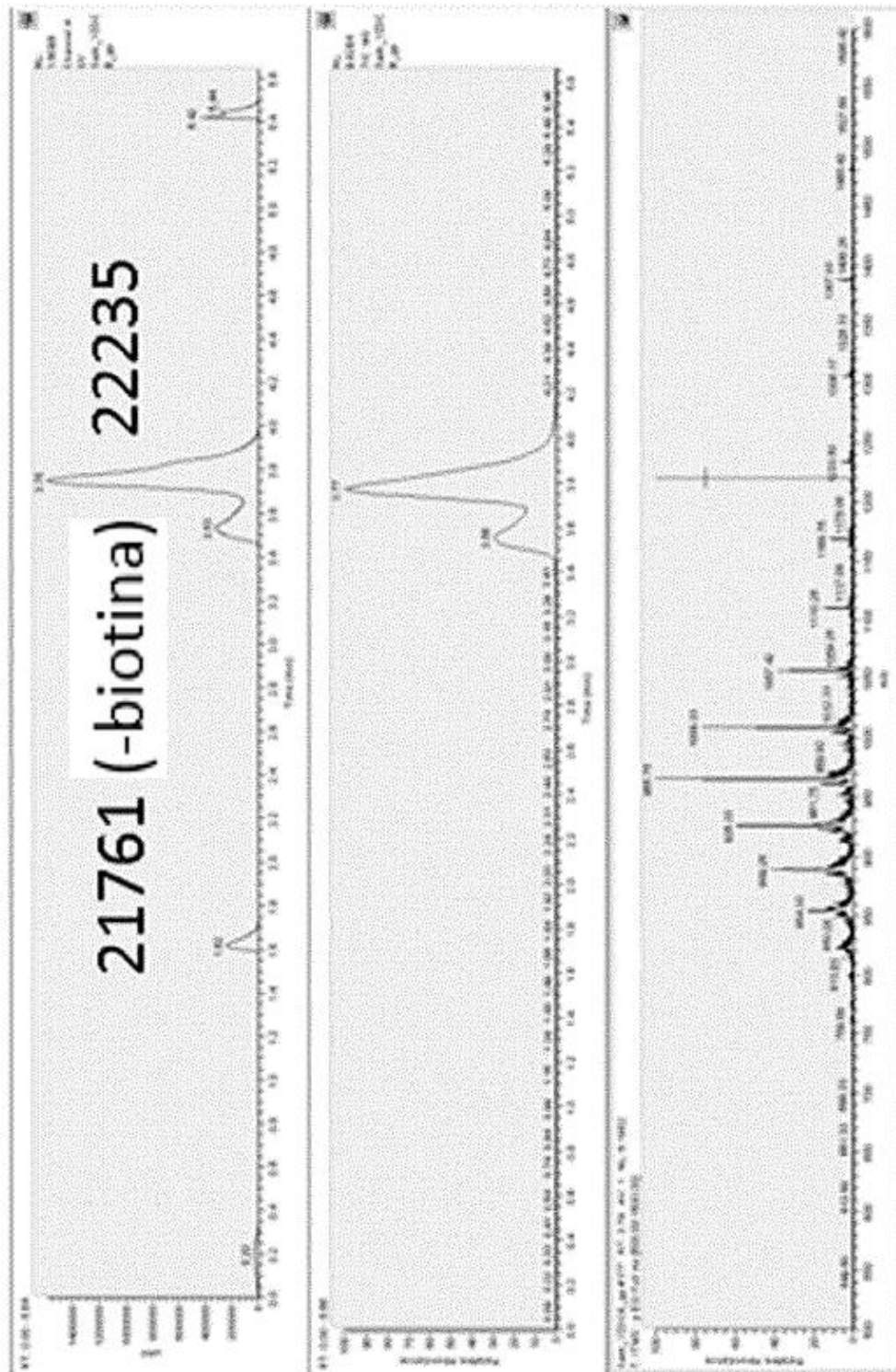
Pasos 2 y 3a repetidos dos veces



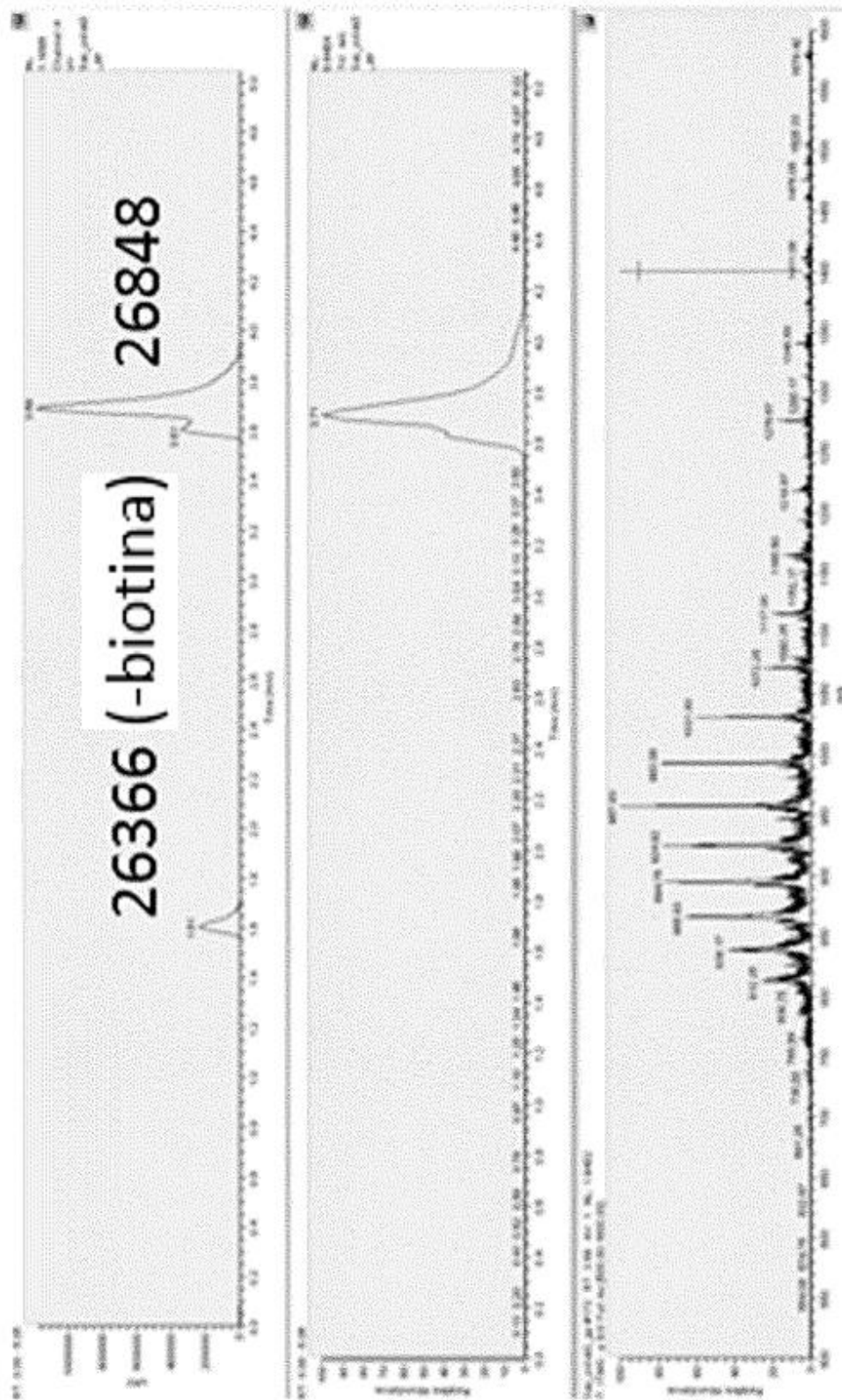
Figura 19B



# Plantilla YDC



# Plantilla YDC





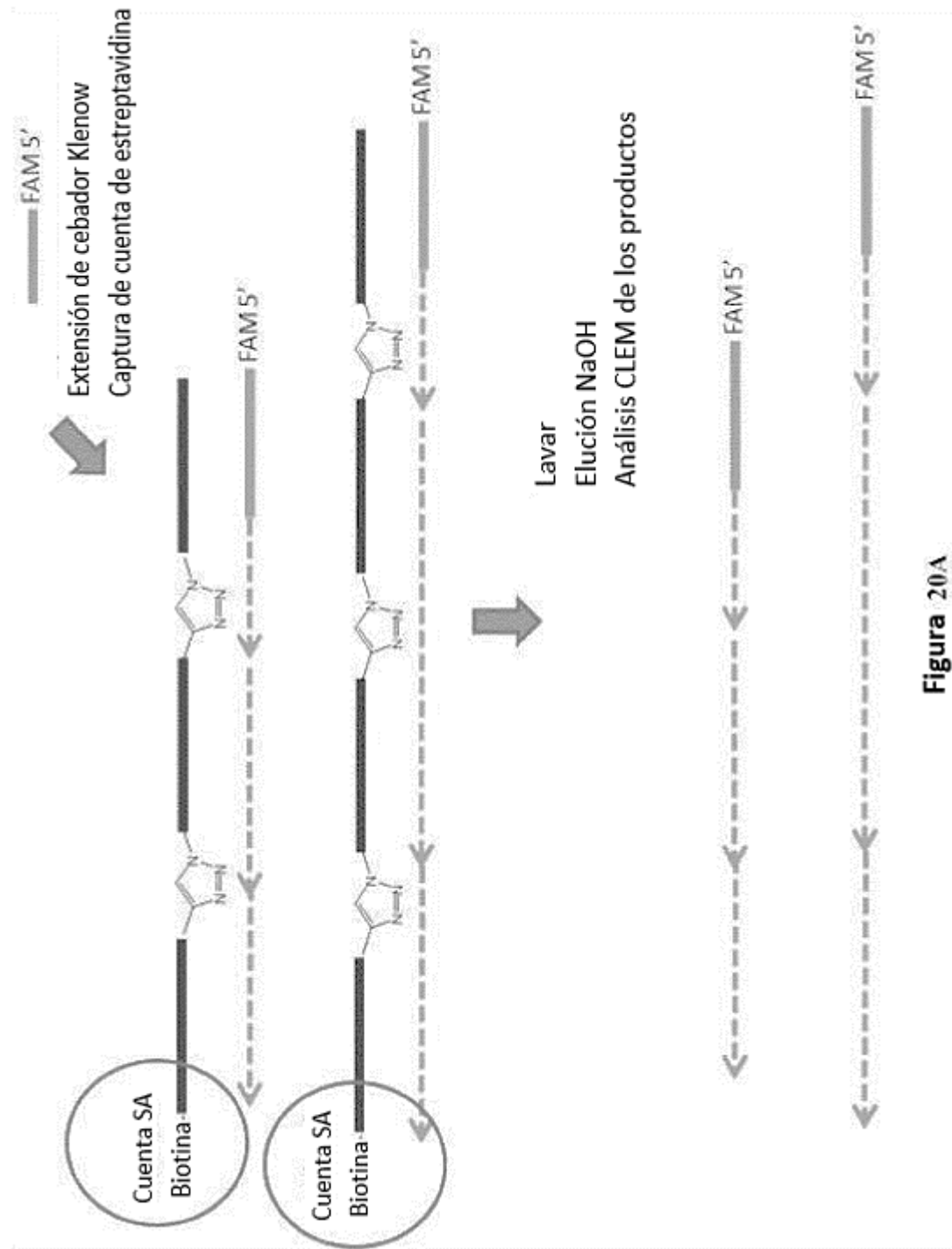


Figura 20A

# Producto Klenow YDC 27197

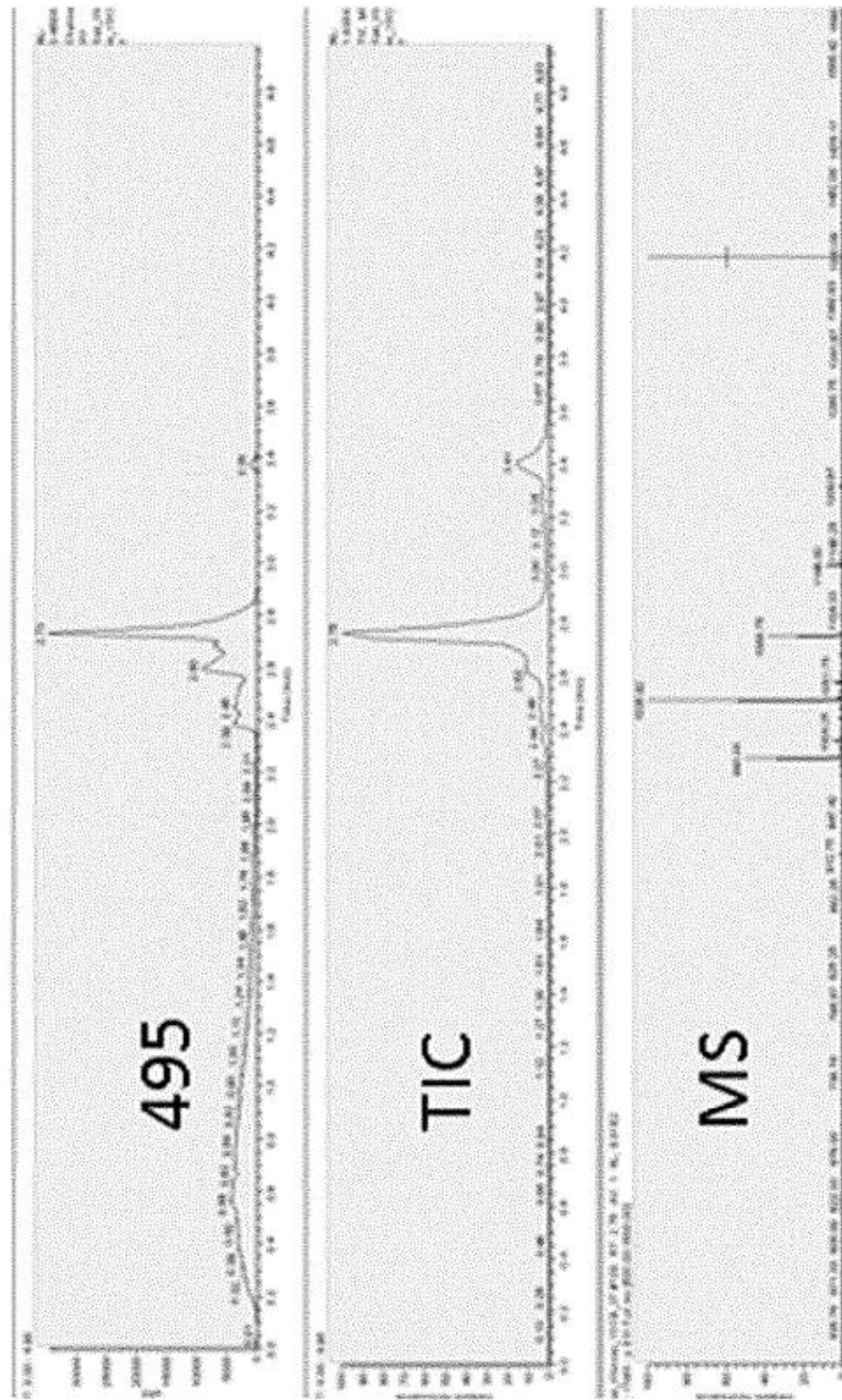


Figura 20B



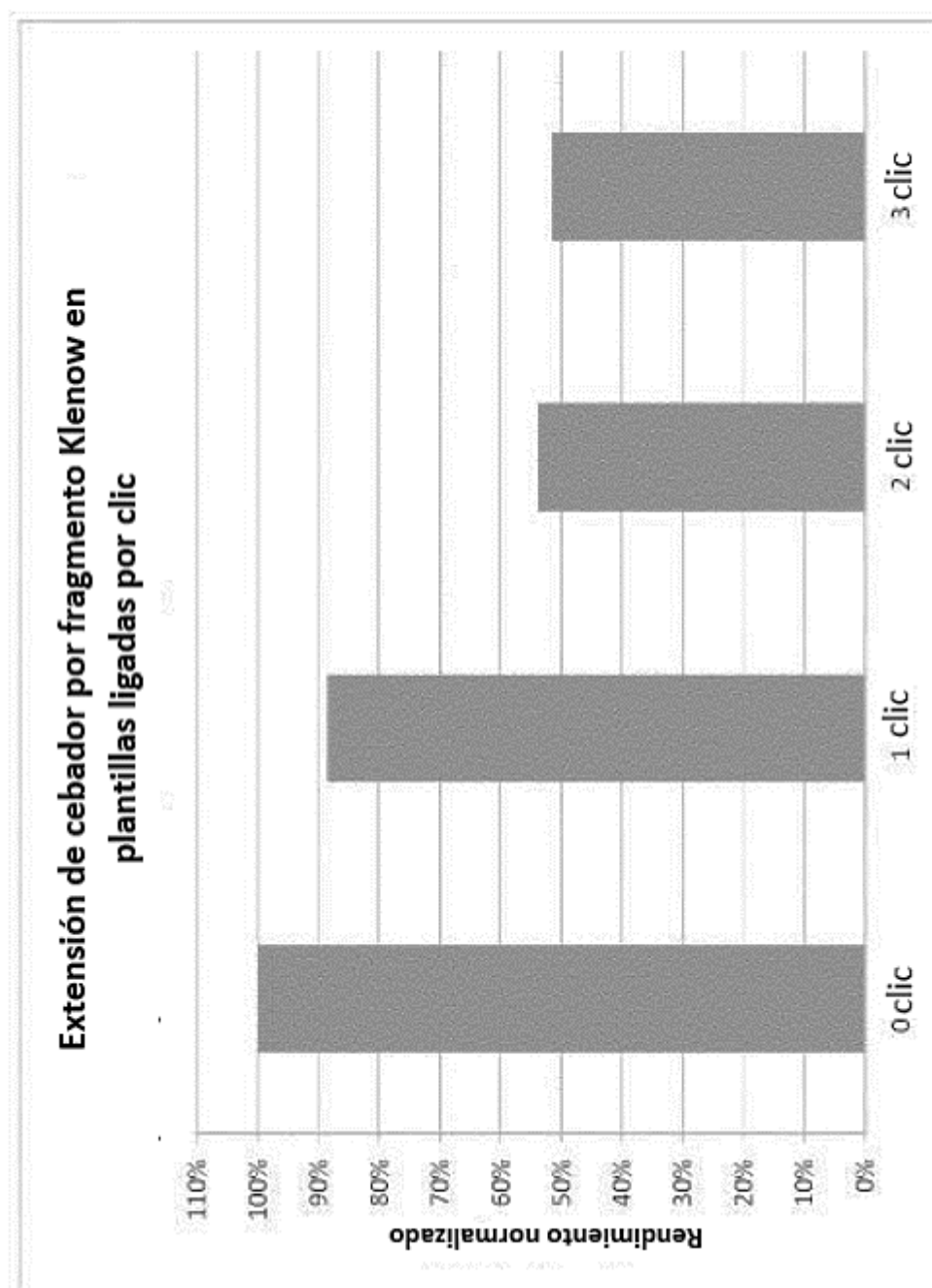


Figura 21

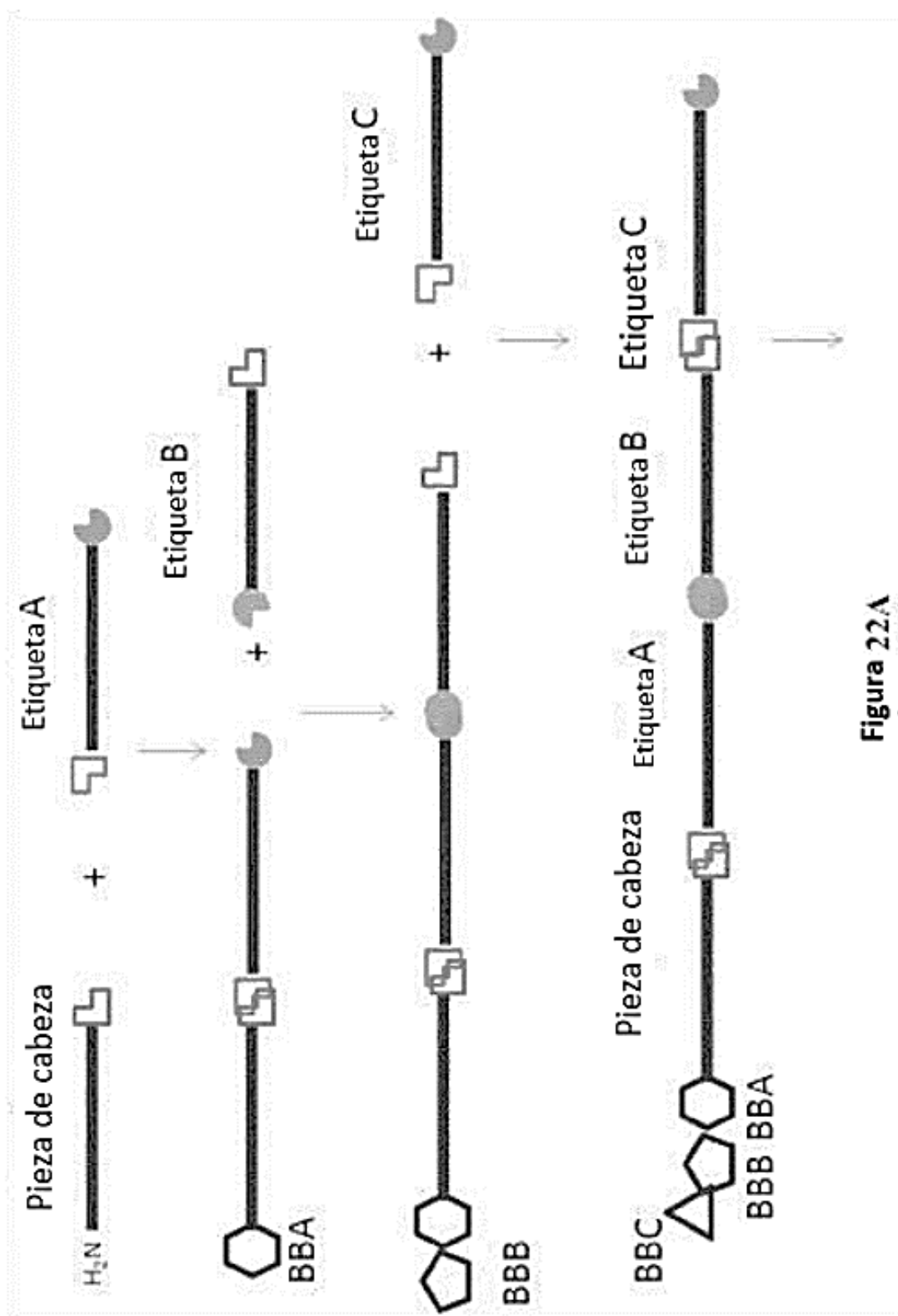


Figura 22A

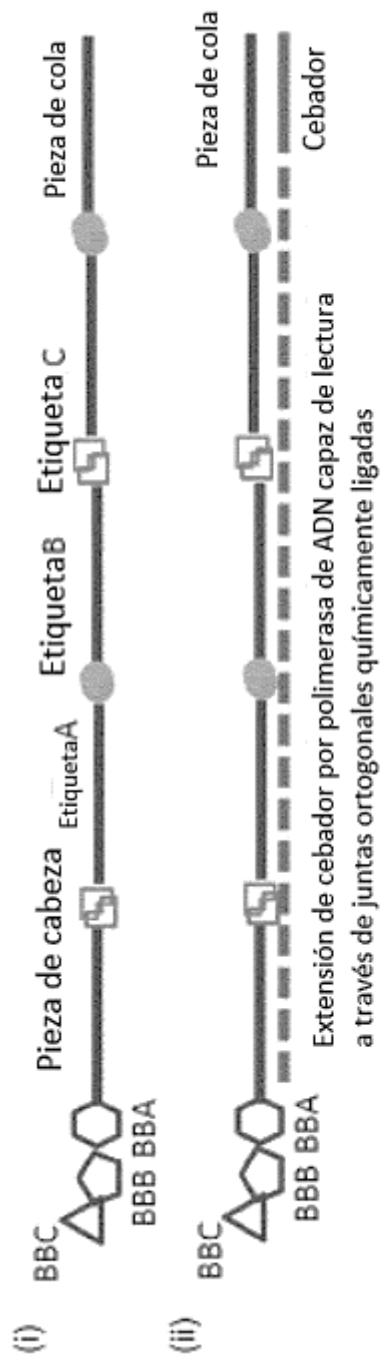


Figura 22B

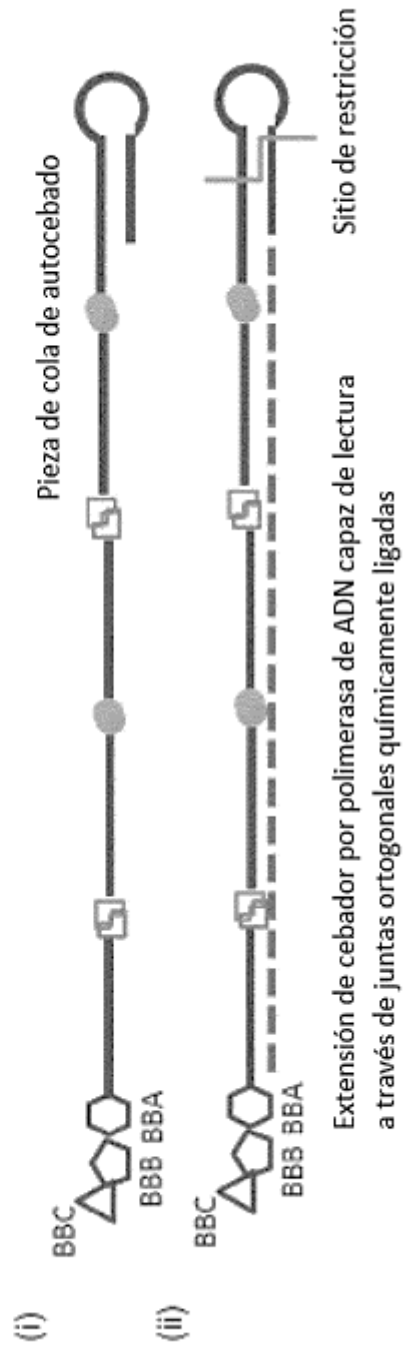


Figura 22C

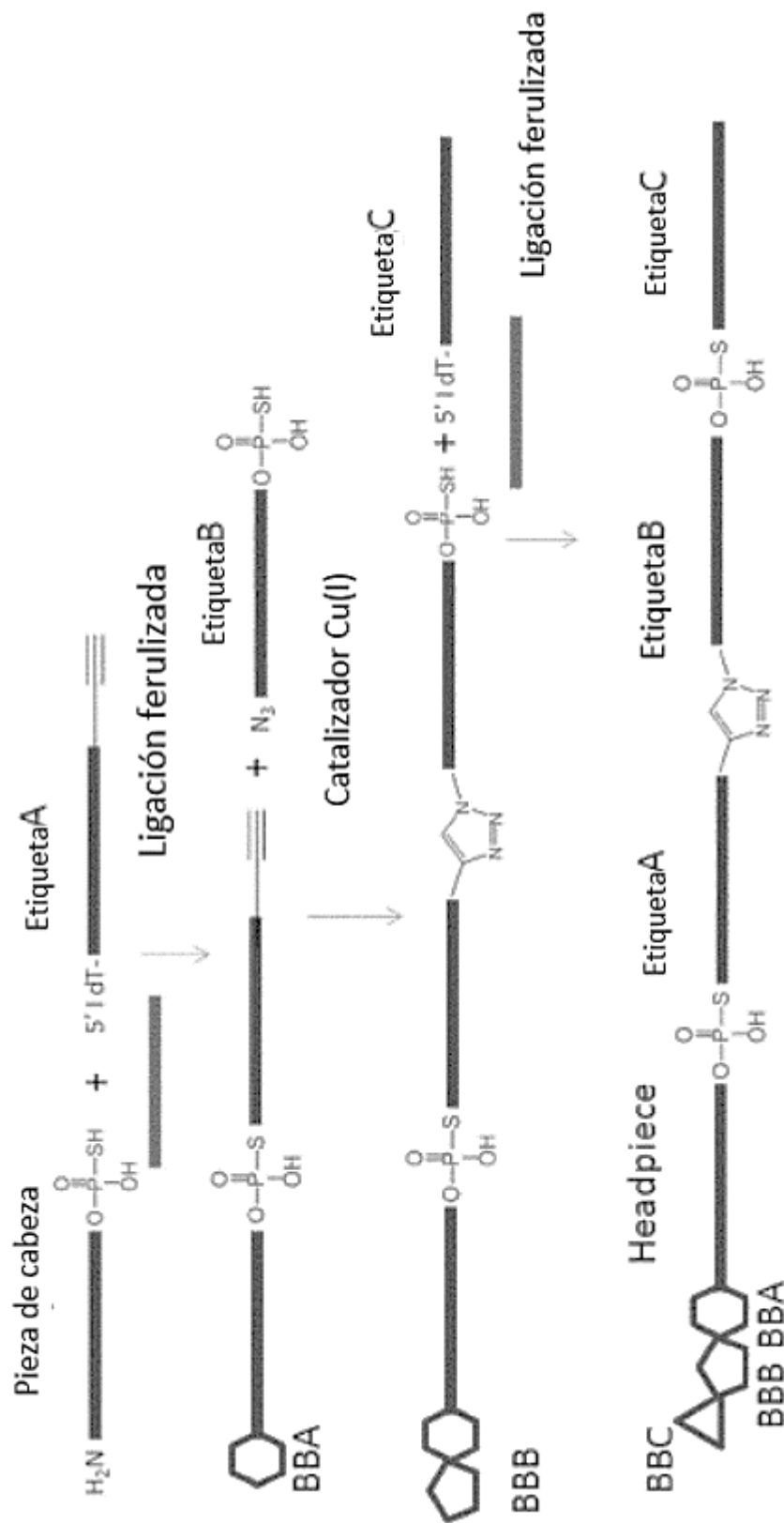


Figura 23



Figura 24A



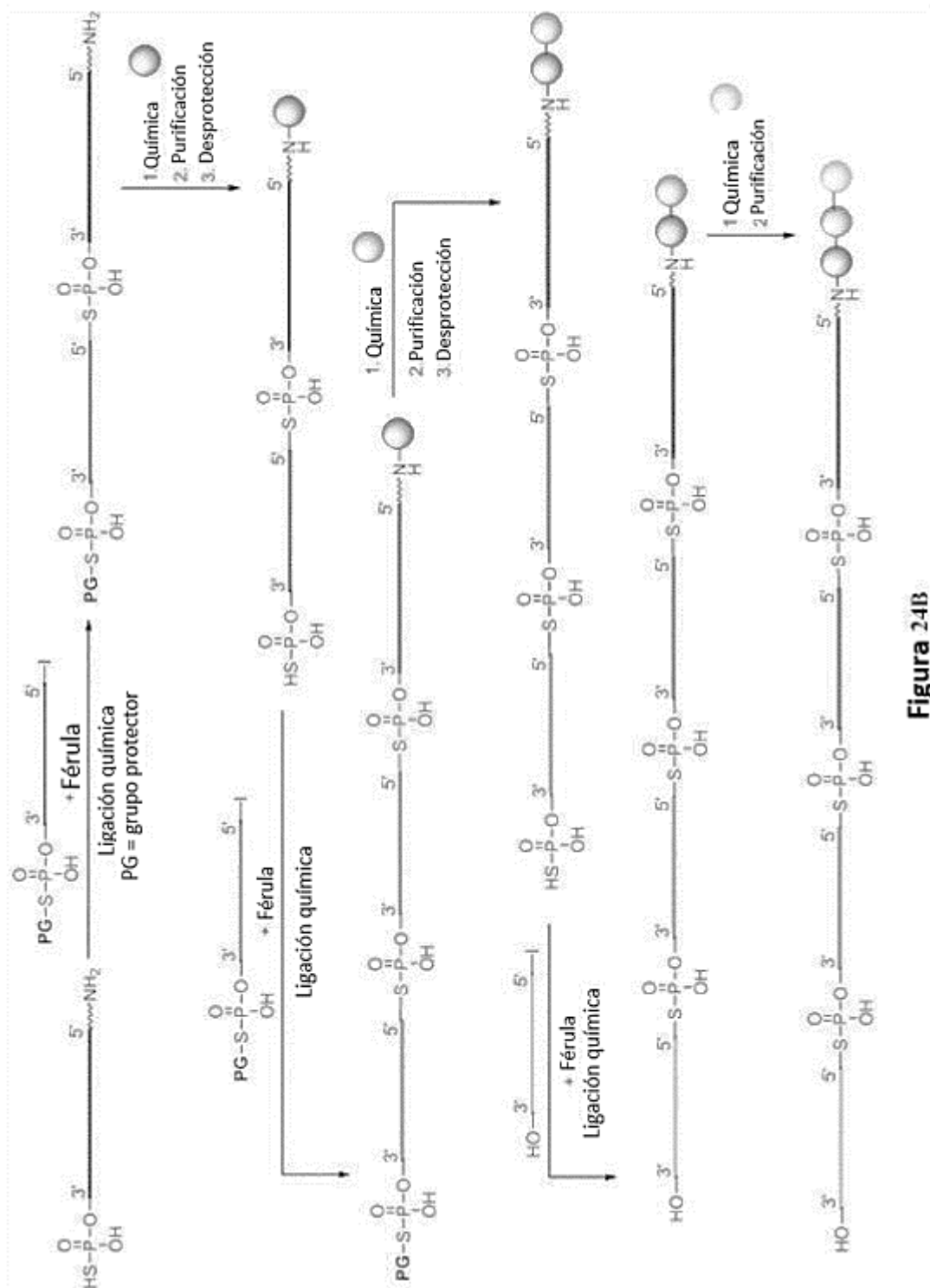


Figura 24B

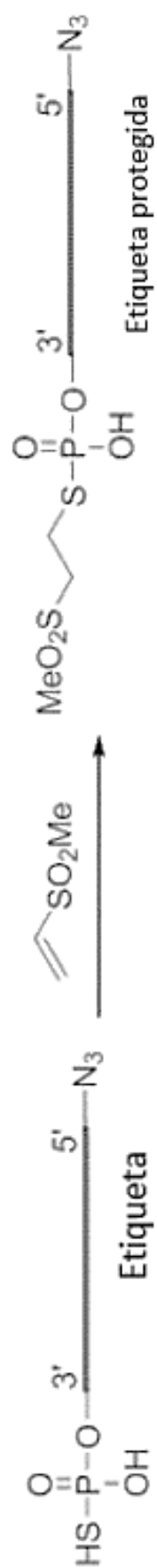


Figura 25A

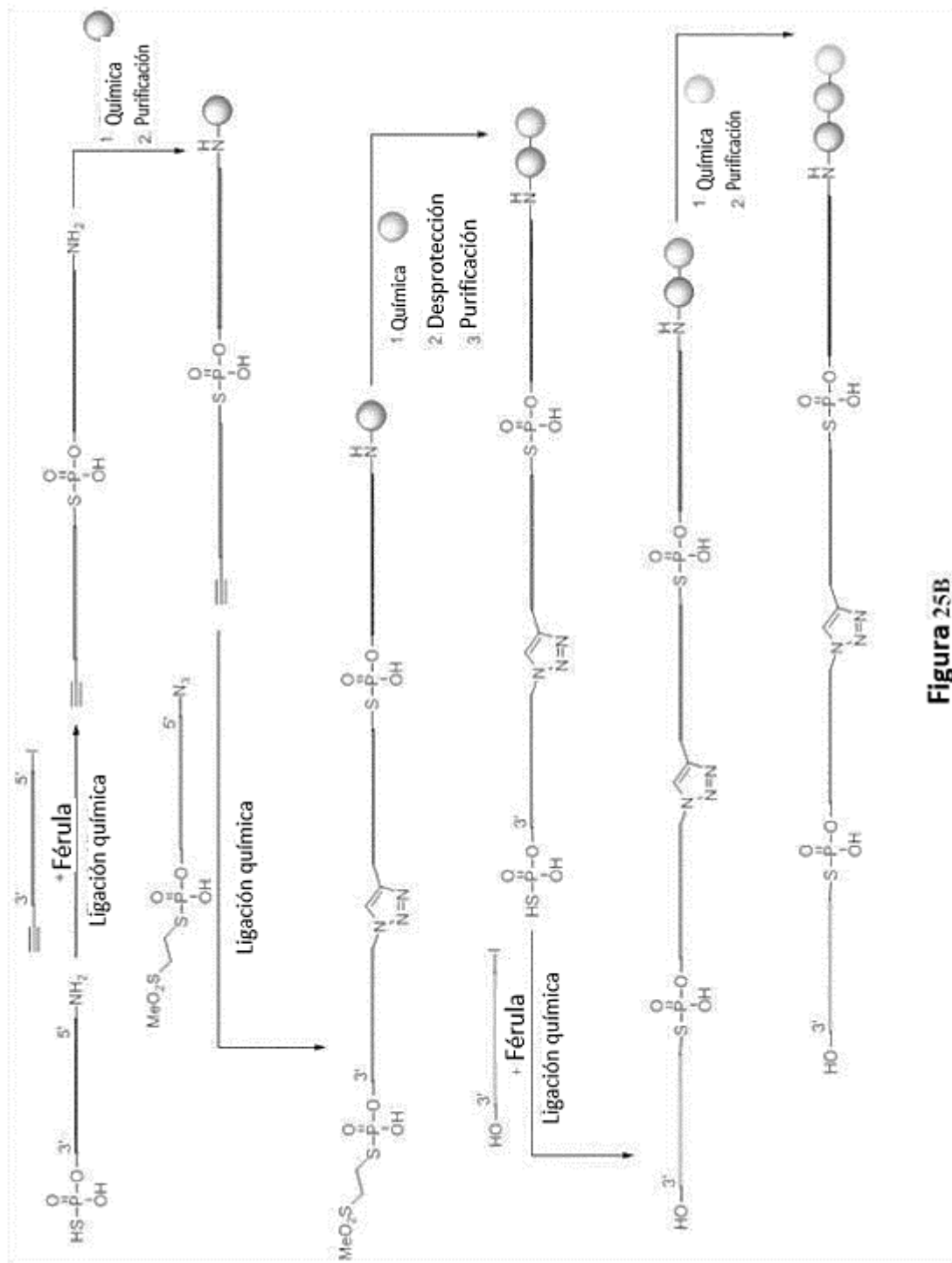


Figura 25B