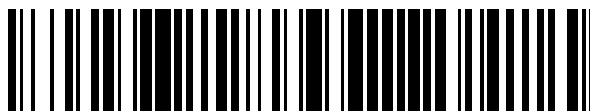


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 737 957**

51 Int. Cl.:

**C12N 15/10** (2006.01)

**C12N 15/66** (2006.01)

**C12P 19/34** (2006.01)

**C40B 50/06** (2006.01)

**G06F 19/10** (2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **23.08.2012 E 15159715 (0)**

97 Fecha y número de publicación de la concesión europea: **24.04.2019 EP 2944693**

54 Título: **Composiciones y métodos para el ensamblaje de alta fidelidad de ácidos nucleicos**

30 Prioridad:

**26.08.2011 US 201161527922 P**  
**09.09.2011 US 201161532825 P**

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

**17.01.2020**

73 Titular/es:

**GEN9, INC. (100.0%)**  
**27 Drydock Avenue, 8th Floor**  
**Boston, MA 02210, US**

72 Inventor/es:

**JACOBSON, JOSEPH;**  
**SCHINDLER, DANIEL y**  
**LAWTON, SCOTT S.**

74 Agente/Representante:

**VALLEJO LÓPEZ, Juan Pedro**

**ES 2 737 957 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

Composiciones y métodos para el ensamblaje de alta fidelidad de ácidos nucleicos

5 **Campo de la invención**

Los métodos y las composiciones de la invención se refieren al ensamblaje de ácidos nucleicos y, en particular, a reacciones de ensamblaje de ácidos nucleicos múltiple de alta fidelidad.

10 **Antecedentes**

Los ácidos nucleicos recombinantes y sintéticos tienen muchas aplicaciones en investigación, industria, agricultura y medicina. Los ácidos nucleicos recombinantes y sintéticos pueden usarse para expresar y obtener grandes cantidades de polipéptidos, incluyendo enzimas, anticuerpos, factores de crecimiento, receptores y otros polipéptidos que pueden usarse para diversos fines médicos, industriales o agrícolas. Los ácidos nucleicos recombinantes y sintéticos también pueden usarse para producir organismos modificados genéticamente incluyendo bacterias modificadas, levaduras, mamíferos, plantas y otros organismos. Los organismos modificados genéticamente pueden usarse en investigación (por ejemplo, como modelos animales de enfermedades, como herramientas para comprender procesos biológicos, etc.), en industria (por ejemplo, como organismos hospedadores para la expresión de proteínas, como biorreactores para generar productos industriales, como herramientas para recuperación ambiental, para aislar o modificar compuestos naturales con aplicaciones industriales, etc.), en agricultura (por ejemplo, cultivos modificados con rendimiento aumentado o resistencia aumentada a enfermedades o estrés ambiental, etc.) y para otras aplicaciones. Los ácidos nucleicos recombinantes y sintéticos también pueden usarse como composiciones terapéuticas (por ejemplo, para modificar la expresión génica, para terapia génica, etc.) o como herramientas de diagnóstico (por ejemplo, como sondas para patologías, etc.).

Se han desarrollado numerosas técnicas para modificar ácidos nucleicos existentes (por ejemplo, ácidos nucleicos de origen natural) para generar ácidos nucleicos recombinantes. Por ejemplo, pueden usarse combinaciones de amplificación, mutagénesis, digestión con nucleasas, ligadura, clonación de ácidos nucleicos y otras técnicas para producir muchos ácidos nucleicos recombinantes diferentes. Los polinucleótidos sintetizados químicamente se usan con frecuencia como cebadores o adaptadores para la amplificación, la mutagénesis y la clonación de ácidos nucleicos.

También se están desarrollando técnicas para el ensamblaje de ácidos nucleicos de novo mediante las cuales se preparan ácidos nucleicos (por ejemplo, se sintetizan químicamente) y se ensamblan para producir ácidos nucleicos objetivo más largos de interés. Por ejemplo, se están desarrollando diferentes técnicas de ensamblaje múltiple para ensamblar oligonucleótidos en ácidos nucleicos sintéticos más grandes que pueden usarse en investigación, industria, agricultura y/o medicina. Kosuri et al, *Nature Biotech* (2010) 28 12 1295 publica la síntesis génica escalable mediante la amplificación selectiva de grupos de ADN a partir de microchips. Richmond et al, *Nucl Acid Res* 2004 32 17 5011-5018 publica la síntesis génica de alto rendimiento mediante la amplificación y el ensamblaje de ADN eluido de chip. Los documentos WO2011/085075 y WO2011/066186 publican la síntesis de ácidos nucleicos, por ejemplo, sobre un soporte sólido. Weber et al, *Plos One* (2011) 6 2 e16765 publica un sistema de clonación modular jerárquico que permite la creación de construcciones multigénicas eucarióticas. Engler et al, *Plos One* (2008) 3 11 e3647 publica un método de clonación en un solo recipiente y en una sola etapa. El documento WO2008/027558 publica métodos para ensamblar moléculas de ácido nucleico usando activación iterativa de uno o más rasgos codificados por vectores para ensamblar progresivamente un inserto de ácido nucleico más largo. El documento WO2010/070295 publica un método para el ensamblaje de secuencias de ácido polinucleico. Sin embargo, una limitación de las técnicas de ensamblaje actualmente disponibles es la tasa de error relativamente alta. Como tal, se necesitan métodos de alta fidelidad y bajo coste.

50 **Sumario de la invención**

Los aspectos de la invención se exponen en las reivindicaciones. La divulgación se refiere a métodos de producción de un ácido nucleico objetivo. El método, de acuerdo con alguna divulgación del presente documento, incluye: (1) proporcionar una pluralidad de fragmentos de ácido nucleico bicatenario de extremos romos que tienen una secuencia de reconocimiento de enzimas de restricción en ambos extremos de cada uno de la pluralidad de fragmentos de ácido nucleico bicatenario de extremos romos; (2) producir una pluralidad de fragmentos de ácido nucleico bicatenario de extremos cohesivos mediante digestión enzimática de la pluralidad de fragmentos de ácido nucleico bicatenario de extremos romos en la proximidad de la secuencia de reconocimiento de enzimas de restricción, en los que cada uno de la pluralidad de fragmentos de ácido nucleico bicatenario de extremos cohesivos tiene dos salientes diferentes y no complementarios; (3) ligar la pluralidad de fragmentos de ácido nucleico bicatenario de extremos cohesivos con una ligasa, en los que un primer saliente de un primer fragmento de ácido nucleico bicatenario de extremos cohesivos es únicamente complementario a un segundo saliente de un segundo fragmento de ácido nucleico bicatenario de extremos cohesivos; y (4) formar una disposición lineal de la pluralidad de fragmentos de ácido nucleico bicatenario de extremos cohesivos, en la que la disposición única comprende el ácido nucleico objetivo. En determinada divulgación del presente documento, la pluralidad de fragmentos de ácido

nucleico bicatenario de extremos romos puede proporcionarse liberando una pluralidad de oligonucleótidos sintetizados sobre un soporte sólido y sintetizando cadenas complementarias de la pluralidad de oligonucleótidos usando una reacción basada en polimerasa.

5 En otra divulgación del presente documento, se proporciona un método para diseñar una pluralidad de ácidos nucleicos de partida que se han de ensamblar en un ácido nucleico objetivo. El método, de acuerdo con alguna divulgación del presente documento, puede incluir: (1) obtener una secuencia objetivo de un ácido nucleico objetivo; (2) seleccionar una pluralidad de subsecuencias en la misma de manera que cada dos subsecuencias adyacentes se solapen entre sí por N bases; (3) almacenar las secuencias de N-bases solapadas resultantes en una memoria; (4) comparar las secuencias de N-bases solapadas entre sí para garantizar que difieran entre sí en al menos una base; y (5) repetir las etapas (2) a (4) hasta que se obtenga una pluralidad de ácidos nucleicos de partida satisfactorios, en las que cualquiera de los dos ácidos nucleicos de partida adyacentes se solapen entre sí mediante N bases.

15 Otra divulgación más en el presente documento se refiere a una pluralidad de ácidos nucleicos de partida que se han de ensamblar en un ácido nucleico objetivo, diseñado de acuerdo con los métodos que se describen en el presente documento. En determinada divulgación del presente documento, la pluralidad de ácidos nucleicos de partida puede incluir cada uno adicionalmente un sitio de unión de cebador universal diseñado para amplificar la pluralidad de ácidos nucleicos de partida a partir de los mismos. La pluralidad de ácidos nucleicos de partida también puede incluir cada uno adicionalmente una secuencia de reconocimiento de enzimas de restricción diseñada por ingeniería genética.

25 En otra divulgación más en el presente documento, se proporciona un sistema para ensamblar un ácido nucleico objetivo. El sistema incluye: (1) un soporte sólido para sintetizar la pluralidad de ácidos nucleicos de partida que se describen en el presente documento, en el que cada ácido nucleico de partida comprende adicionalmente un sitio de unión de cebador universal diseñado por ingeniería genética y una secuencia de reconocimiento de enzimas de restricción diseñada por ingeniería genética; (2) una unidad de reacción de polimerasa para sintetizar cadenas complementarias de la pluralidad de ácidos nucleicos de partida una reacción basada en polimerasa usando un cebador universal complementario al sitio de unión del cebador universal, produciendo de este modo una pluralidad de fragmentos de ácido nucleico bicatenario de extremos romos; (3) una unidad de digestión para producir una pluralidad de fragmentos de ácido nucleico bicatenario de extremos cohesivos a través de la digestión enzimática de la pluralidad de fragmentos de ácido nucleico bicatenario de extremos romos en la proximidad de la secuencia de reconocimiento de enzimas de restricción, en la que la pluralidad de fragmentos de ácido nucleico bicatenario de extremos cohesivos tienen dos salientes diferentes y no complementarios; y (4) una unidad de ligadura para ligar la pluralidad de fragmentos de ácido nucleico bicatenario de extremos cohesivos con una ligasa, en la que un primer saliente de un primer fragmento de ácido nucleico bicatenario de extremos cohesivos es únicamente complementario a un segundo saliente de un segundo fragmento de ácido nucleico bicatenario de extremos cohesivos.

40 Una divulgación adicional en el presente documento proporciona un producto de programa informático para diseñar una pluralidad de ácidos nucleicos de partida que se han de ensamblar en un ácido nucleico objetivo, dicho programa reside en un medio de almacenamiento de soporte físico legible por ordenador y tiene una pluralidad de instrucciones que, cuando son ejecutadas por un procesador, provocan que el procesador realice operaciones que comprenden: (1) obtener una secuencia objetivo de un ácido nucleico objetivo; (2) seleccionar una pluralidad de subsecuencias en la misma de manera que cada dos subsecuencias adyacentes se solapen entre sí por N bases; (3) almacenar las secuencias de N-bases solapadas resultantes en una memoria; (4) comparar las secuencias de N-bases solapadas entre sí para garantizar que difieran entre sí en al menos una base; y (5) repetir las etapas (2) a (4) hasta que se obtenga una pluralidad de ácidos nucleicos de partida satisfactorios, en las que cualquiera de los dos ácidos nucleicos de partida adyacentes se solapen entre sí mediante N bases.

## 50 Breve descripción de las figuras

La FIG. 1 ilustra un diseño de ejemplo de oligonucleótidos para una reacción de ensamblaje de oligonucleótidos múltiple.

55 La FIG. 2 ilustra la posición relativa de los cebadores utilizados para someter a ensayo productos de la reacción de ensamblaje múltiple.

La FIG. 3 ilustra una realización de una reacción de ensamblaje de oligonucleótidos por pares.

La FIG. 4 ilustra realizaciones de una reacción de ensamblaje de oligonucleótidos múltiple.

60 La FIG. 5 ilustra un ensayo basado en PCR de los productos de la reacción de ensamblaje de oligonucleótidos múltiple de la FIG. 4.

La FIG. 6 ilustra la confirmación de secuenciación de los productos de la reacción de ensamblaje de oligonucleótidos múltiple de la FIG. 4.

Las FIG. 7A y 7B ilustran realizaciones de un ensayo de ligadura de desapareamiento por pares.

La FIG. 8 ilustra productos de ensamblaje alternativos basados en el diseño de la FIG. 1.

Las FIG. 9A y 9B ilustran dos estrategias de diseño para secuencias que flanquean fragmentos de ensamblaje.

65 La FIG. 10A y 10B ilustran dos estrategias de ensamblaje de compensación.

**Descripción detallada de la invención**

Las divulgaciones se refieren a métodos y composiciones para unir covalentemente una pluralidad de fragmentos de ácido nucleico para producir un producto de ácido nucleico más largo en una sola etapa de ensamblaje. La divulgación del presente documento puede usarse para ensamblar grandes cantidades de fragmentos de ácido nucleico eficientemente y/o para reducir el número de etapas necesarias para generar grandes productos de ácido nucleico, reduciendo al mismo tiempo la tasa de error de ensamblaje. La divulgación del presente documento puede incorporarse a los procedimientos de ensamblaje de ácidos nucleicos para aumentar la fidelidad, el rendimiento y/o la eficiencia del ensamblaje, disminuir el coste y/o reducir el tiempo de ensamblaje. Alguna divulgación del presente documento puede automatizarse y/o implementarse en un contexto de ensamblaje de alto rendimiento para facilitar la producción paralela de muchos productos de ácido nucleico objetivo diferentes.

Ensamblaje de oligonucleótidos múltiple

Un fragmento de ácido nucleico predeterminado puede ensamblarse a partir de una pluralidad de diferentes ácidos nucleicos de partida (por ejemplo, oligonucleótidos) en una reacción de ensamblaje múltiple (por ejemplo, una reacción mediada por enzimas múltiple, una reacción de ensamblaje químico múltiple o una combinación de las mismas). Determinados aspectos de las reacciones de ensamblaje de ácidos nucleicos múltiple se ilustran mediante la siguiente descripción de determinadas realizaciones de reacciones de ensamblaje de oligonucleótidos múltiple. Debe apreciarse que la descripción de las reacciones de ensamblaje en el contexto de oligonucleótidos no pretende ser limitante. Las reacciones de ensamblaje que se describen en el presente documento pueden realizarse usando ácidos nucleicos de partida obtenidos de una o más fuentes diferentes (por ejemplo, polinucleótidos sintéticos o naturales, productos de amplificación de ácidos nucleicos, productos de degradación de ácidos nucleicos, oligonucleótidos, etc.). Los ácidos nucleicos de partida pueden denominarse ácidos nucleicos de ensamblaje (por ejemplo, oligonucleótidos de ensamblaje). Como se usa en el presente documento, un ácido nucleico de ensamblaje tiene una secuencia que está diseñada para incorporarse en el producto de ácido nucleico generado durante el proceso de ensamblaje. Sin embargo, debe apreciarse que la descripción de las reacciones de ensamblaje en el contexto de ácidos nucleicos bicatenario no pretende ser limitante. En alguna divulgación del presente documento, uno o más de los ácidos nucleicos de partida ilustrados en las figuras y que se describen en el presente documento pueden proporcionarse como ácidos nucleicos monocatenarios. En consecuencia, debe apreciarse que cuando las figuras y la descripción ilustran el ensamblaje de ácidos nucleicos bicatenarios de extremos cohesivos, se contempla la presencia de uno o más ácidos nucleicos monocatenarios.

Como se usa en el presente documento, un oligonucleótido puede ser una molécula de ácido nucleico que comprenda al menos dos restos de nucleótidos unidos covalentemente. En alguna divulgación del presente documento, un oligonucleótido puede tener entre 10 y 1.000 nucleótidos de longitud. Por ejemplo, un oligonucleótido puede tener entre 10 y 500 nucleótidos de longitud o entre 500 y 1.000 nucleótidos de longitud. En alguna divulgación del presente documento, un oligonucleótido puede tener entre aproximadamente 20 y aproximadamente 300 nucleótidos de longitud (por ejemplo, de aproximadamente 30 a 250, de 40 a 220, de 50 a 200, de 60 a 180, o de aproximadamente 65 o aproximadamente 150 nucleótidos de longitud), entre aproximadamente 100 y aproximadamente 200, entre aproximadamente 200 y aproximadamente 300 nucleótidos, entre aproximadamente 300 y aproximadamente 400, o entre aproximadamente 400 y aproximadamente 500 nucleótidos de longitud. Sin embargo, pueden usarse oligonucleótidos más cortos o más largos. Un oligonucleótido puede ser un ácido nucleico monocatenario. Sin embargo, en alguna divulgación del presente documento puede usarse un oligonucleótido bicatenario como se describe en el presente documento. En determinada divulgación del presente documento, un oligonucleótido puede sintetizarse químicamente como se describe con más detalle a continuación. En alguna divulgación del presente documento, un ácido nucleico de entrada (por ejemplo, un oligonucleótido sintético) puede amplificarse antes de su uso. El producto resultante puede ser bicatenario.

En determinada divulgación del presente documento, cada oligonucleótido puede diseñarse para que tenga una secuencia que sea idéntica a una porción diferente de la secuencia de un ácido nucleico objetivo predeterminado que se ha de ensamblar. En consecuencia, en alguna divulgación del presente documento, cada oligonucleótido puede tener una secuencia que sea idéntica a una porción de una de las dos cadenas de un ácido nucleico objetivo bicatenario. Para una mayor claridad, las dos cadenas complementarias de un ácido nucleico bicatenario se denominan en el presente documento las cadenas positiva (P) y negativa (N). Esta designación no pretende implicar que las cadenas sean cadenas sentido y antisentido de una secuencia codificante. Se refieren solo a las dos cadenas complementarias de un ácido nucleico (por ejemplo, un ácido nucleico objetivo, un fragmento de ácido nucleico intermedio, etc.) independientemente de la secuencia o función del ácido nucleico. En consecuencia, en alguna divulgación del presente documento, una cadena P puede ser una cadena sentido de una secuencia codificante, mientras que en otra divulgación del presente documento una cadena P puede ser una cadena antisentido de una secuencia codificante. Debe apreciarse que la referencia a ácidos nucleicos complementarios o regiones de ácidos nucleicos complementarios en el presente documento se refiere a ácidos nucleicos o regiones de los mismos que tienen secuencias que son complementos inversos entre sí de manera que pueden hibridarse de una manera antiparalela típica del ADN natural.

De acuerdo con una divulgación del presente documento, un ácido nucleico objetivo puede ser la cadena P, la

cadena N o un ácido nucleico bicatenario que comprenda las cadenas tanto P como N. Debe apreciarse que diferentes oligonucleótidos pueden diseñarse para que tengan diferentes longitudes. En alguna divulgación del presente documento, uno o más oligonucleótidos diferentes pueden tener regiones de secuencia solapantes (por ejemplo, regiones 5' solapadas y/o regiones 3' solapadas). Las regiones de secuencia solapantes pueden ser idénticas (es decir, correspondientes a la misma cadena del fragmento de ácido nucleico) o complementarias (es decir, correspondientes a cadenas complementarias del fragmento de ácido nucleico). La pluralidad de oligonucleótidos puede incluir uno o más pares de oligonucleótidos con regiones de secuencia idénticas solapantes, uno o más pares de oligonucleótidos con regiones de secuencia complementarias solapantes o una combinación de los mismos. Las secuencias solapantes pueden ser de cualquier longitud adecuada. Por ejemplo, las secuencias solapantes pueden abarcar toda la longitud de uno o más ácidos nucleicos utilizados en una reacción de ensamblaje. Las secuencias solapantes pueden ser de entre aproximadamente 2 y aproximadamente 50 (por ejemplo, de entre 3 y 20, entre 3 y 10, entre 3 y 8 o 4, 5, 6, 7, 8, 9, etc. nucleótidos de longitud). Sin embargo, pueden usarse longitudes solapantes más cortas, más largas o intermedias. Debe apreciarse que los solapamientos entre diferentes ácidos nucleicos de entrada utilizados en una reacción de ensamblaje pueden tener longitudes y/o secuencias diferentes. Por ejemplo, las secuencias solapantes pueden ser diferentes entre sí en al menos un nucleótido, 2 nucleótidos, 3 nucleótidos o más. Suponiendo que las secuencias solapantes se diferencien entre sí en  $x$  nucleótidos, entonces pueden ensamblarse entre sí hasta  $(4^x+1)$  fragmentos de diferentes ácidos nucleicos de entrada en una sola reacción.

En una reacción de ensamblaje de oligonucleótidos múltiple diseñada para generar un fragmento de ácido nucleico predeterminado, las secuencias combinadas de los oligonucleótidos diferentes en la reacción pueden abarcar la secuencia de todo el fragmento de ácido nucleico en la cadena positiva, en la cadena negativa, en ambas cadenas o en una combinación de porciones de la cadena positiva y porciones de la cadena negativa. La pluralidad de oligonucleótidos diferentes puede proporcionar secuencias positivas, secuencias negativas o una combinación de secuencias positivas y negativas correspondientes a la secuencia completa del fragmento de ácido nucleico que se ha de ensamblar. En alguna divulgación del presente documento, la pluralidad de oligonucleótidos puede incluir uno o más oligonucleótidos que tengan secuencias idénticas a una o más porciones de la secuencia positiva y uno o más oligonucleótidos que tengan secuencias que sean idénticas a una o más porciones de la secuencia negativa del fragmento de ácido nucleico. Uno o más pares de oligonucleótidos diferentes pueden incluir secuencias que sean idénticas a porciones solapantes de la secuencia de fragmento de ácido nucleico predeterminada como se describe en el presente documento (por ejemplo, porciones de secuencias solapantes de la misma cadena o de una cadena complementarias del fragmento de ácido nucleico). En alguna divulgación del presente documento, la pluralidad de oligonucleótidos incluye un conjunto de oligonucleótidos que tienen secuencias que se combinan para abarcar toda la secuencia positiva y un conjunto de oligonucleótidos que tienen secuencias que se combinan para abarcar la secuencia negativa completa del fragmento de ácido nucleico predeterminado. Sin embargo, en determinada divulgación del presente documento, la pluralidad de oligonucleótidos puede incluir uno o más oligonucleótidos con secuencias que sean idénticas a las porciones de secuencia en una sola cadena (ya sea la cadena positiva o negativa) del fragmento de ácido nucleico, pero ningún oligonucleótido con secuencias que sean complementarias a esas porciones de secuencia. En una realización, una pluralidad de oligonucleótidos incluye solo oligonucleótidos que tienen secuencias idénticas a porciones de la secuencia positiva del fragmento de ácido nucleico predeterminado. En una realización, una pluralidad de oligonucleótidos incluye solo oligonucleótidos que tienen secuencias idénticas a porciones de la secuencia negativa del fragmento de ácido nucleico predeterminado. Estos oligonucleótidos pueden ensamblarse por ligadura secuencial o en una reacción basada en la prolongación (por ejemplo, si se añade a la reacción un oligonucleótido que tiene una región 3' que es complementaria a uno de la pluralidad de oligonucleótidos).

En un aspecto, un fragmento de ácido nucleico puede ensamblarse en una reacción de ensamblaje mediada por ligasa a partir de una pluralidad de oligonucleótidos que se combinan y se ligan en una o más rondas de ligaduras mediadas por ligasa. Las técnicas de ensamblaje basadas en ligasa pueden implicar una o más enzimas ligasa adecuadas que puedan catalizar la unión covalente de los extremos adyacentes de los ácidos nucleicos 3' y 5' (por ejemplo, un fosfato 5' y un hidroxilo 3' de uno o más ácidos nucleicos hibridados en un ácido nucleico molde complementario de manera que el extremo 3' esté inmediatamente adyacente al extremo 5'). En consecuencia, una ligasa puede catalizar una reacción de ligadura entre el fosfato 5' de un primer ácido nucleico y el hidroxilo 3' de un segundo ácido nucleico si el primer y el segundo ácido nucleico se hibridan uno al lado del otro en un ácido nucleico molde). Una ligasa puede obtenerse de fuentes recombinantes o naturales. En alguna divulgación del presente documento, pueden usarse una o más ligasas de baja temperatura (por ejemplo, de temperatura ambiente o inferior) (por ejemplo, ADN ligasa T3, ADN ligasa T4, ADN ligasa T7 y/o ADN ligasa de *E. coli*). Una ligasa de temperatura más baja puede ser útil para salientes más cortos (por ejemplo, salientes de aproximadamente 3, de aproximadamente 4, de aproximadamente 5 o de aproximadamente 6 bases) que pueden no ser estables a temperaturas más altas. Una ligasa también puede ser una ligasa termoestable. En alguna divulgación del presente documento, puede usarse una ligasa termoestable de un organismo termófilo. Los ejemplos de ligasas de ADN termoestables incluyen, pero sin limitación: ADN ligasa Tth (de *Thermus thermophilus*, disponible en, por ejemplo, Eurogentec y GeneCraft); ADN ligasa Pfu (una ligasa hipertermófila de *Pyrococcus furiosus*); ligasa Taq (de *Thermus aquaticus*), cualquier otra ligasa termoestable adecuada o cualquier combinación de las mismas.

La divulgación del presente documento puede usarse para potenciar diferentes tipos de reacciones de ensamblaje

de ácidos nucleicos (por ejemplo, reacciones de ensamblaje de ácidos nucleicos múltiple). La divulgación del presente documento puede usarse en combinación con una o más reacciones de ensamblaje que se describen en, por ejemplo, Carr et al., 2004, *Nucleic Acids Research*, Vol. 32, n.º 20, el62 (9 páginas); Richmond et al., 2004, *Nucleic Acids Research*, Vol. 32, n.º 17, págs. 5011-5018; Caruthers et al., 1972, *J. Mol. Biol.* 72, 475-492; Hecker et al., 1998, *Biotechniques* 24:256-260; Kodumal et al., 2004, *PNAS* Vol. 101, N.º 44, págs. 15573-15578; Tian et al., 2004, *Nature*, vol. 432, págs. 1050-1054; y las Patentes de los EE.UU. N.º 6.008.031 y 5.922.539, cuyas divulgaciones se incorporan en el presente documento por referencia. Determinada divulgación del presente documento de reacciones de ensamblaje de ácidos nucleicos múltiple para generar un fragmento de ácido nucleico predeterminado se ilustra con referencia a las FIG. 1-10. Debe apreciarse que los métodos de síntesis y ensamblaje que se describen en el presente documento (incluyendo, por ejemplo, la síntesis de oligonucleótidos, el ensamblaje por etapas, el ensamblaje de ácidos nucleicos múltiple, el ensamblaje jerárquico de fragmentos de ácido nucleico o cualquier combinación de los mismos) pueden realizarse en cualquier formato adecuado, incluyendo en un tubo de reacción, en una placa de múltiples pocillos, sobre una superficie, en una columna, en un dispositivo microfluído (por ejemplo, un tubo microfluído), un tubo capilar, etc. Por ejemplo, en alguna divulgación del presente documento, el ácido nucleico objetivo puede ensamblarse mediante "ensamblaje recursivo" o "ensamblaje jerárquico". En esta realización, el ácido nucleico objetivo se divide en primer lugar en dos o más fragmentos de ácido nucleico solapantes (o fragmentos de subensamblaje). Cada fragmento de ácido nucleico se subdivide después en dos o más fragmentos de ácido nucleico más pequeños que solapantes.

## 20 Oligonucleótidos sintéticos

Los oligonucleótidos pueden sintetizarse usando cualquier técnica adecuada. Por ejemplo, los oligonucleótidos pueden sintetizarse en una columna u otro soporte (por ejemplo, un chip). Los ejemplos de técnicas de síntesis basadas en chip incluyen técnicas utilizadas en dispositivos de síntesis o métodos disponibles en CombiMatrix, Agilent, Affymetrix u otras fuentes. Un oligonucleótido sintético puede tener cualquier tamaño adecuado, por ejemplo, de entre 10 y 1.000 nucleótidos de longitud (por ejemplo, de entre 10 y 200, 200 y 500, 500 y 1.000 nucleótidos de longitud o cualquier combinación de los mismos). Una reacción de ensamblaje puede incluir una pluralidad de oligonucleótidos, cada uno de los cuales puede tener independientemente entre 10 y 300 nucleótidos de longitud (por ejemplo, entre 20 y 250, entre 30 y 200, de 50 a 150, de 50 a 100 o cualquier número intermedio de nucleótidos). Sin embargo, pueden usarse uno o más oligonucleótidos más cortos o más largos en determinada divulgación del presente documento.

Como se usa en el presente documento, los términos "soporte" y "sustrato" se usan indistintamente y se refieren a un material insoluble en disolvente, poroso o no poroso, sobre el que se sintetizan o inmovilizan polímeros tales como ácidos nucleicos. Como se usa en el presente documento, "poroso" significa que el material contiene poros que tienen diámetros sustancialmente uniformes (por ejemplo, en el intervalo de nm). Los materiales porosos pueden incluir, pero sin limitación, papel, filtros sintéticos y similares. En materiales porosos de este tipo, la reacción puede tener lugar dentro de los poros. El soporte puede tener una cualquiera de una serie de formas, tales como pasador, tira, placa, disco, varilla, bandas, estructura cilíndrica, partícula, incluyendo perla, nanopartículas y similares. El soporte puede tener anchos variables.

El soporte puede ser hidrófilo o capaz de volverse hidrófilo. El soporte puede incluir polvos inorgánicos tales como sílice, sulfato de magnesio y alúmina; materiales poliméricos naturales, en particular, materiales celulósicos y materiales derivados de celulosa, tales como papeles que contienen fibra, por ejemplo, papel de filtro, papel cromatográfico, etc.; polímeros de origen natural sintéticos o modificados, tales como nitrocelulosa, acetato de celulosa, poli(cloruro de vinilo), poliacrilamida, dextrano reticulado, agarosa, poliacrilato, polietileno, polipropileno, poli(4-metilbuteno), poliestireno, polimetacrilato, poli(terefalato de etileno), nailon, poli(butirato de vinilo), membrana de difluoruro de polivinilideno (PVDF), vidrio, vidrio de poros controlados, vidrio de poros controlados magnéticamente, cerámica, metales y similares; ya sea utilizados por sí solos o junto con otros materiales.

En alguna divulgación del presente documento, los oligonucleótidos se sintetizan en un formato de matriz. Por ejemplo, los oligonucleótidos monocatenarios se sintetizan in situ sobre un soporte común en el que cada oligonucleótido se sintetiza en un elemento (o mancha) separado o diferenciado sobre el sustrato. En una divulgación preferida del presente documento, los oligonucleótidos monocatenarios se unen a la superficie del soporte o elemento. Como se usa en el presente documento, el término "matriz" se refiere a una disposición de elementos diferenciados para almacenar, enviar, amplificar y liberar oligonucleótidos u oligonucleótidos complementarios para reacciones adicionales. En una realización preferida, el soporte o matriz es direccionable: el soporte incluye dos o más elementos direccionables diferenciados en una ubicación predeterminada particular (es decir, una "dirección") sobre el soporte. Por tanto, cada molécula de oligonucleótido de la matriz se ubica en una ubicación conocida y definida sobre el soporte. La secuencia de cada oligonucleótido puede determinarse a partir de su posición sobre el soporte. Además, los soportes o matrices direccionables permiten el control directo de volúmenes aislados individuales, tales como gotitas. El tamaño del elemento definido puede elegirse para que permita la formación de una gotita microvolumétrica sobre el elemento, manteniéndose cada una de las gotitas separadas entre sí. Como se describe en el presente documento, los elementos normalmente, pero no necesariamente, están separados por espacios interelementos para garantizar que las gotitas entre dos elementos adyacentes no se fusionen. Los interelementos normalmente no llevarán ningún oligonucleótido sobre su superficie y

se corresponderán con un espacio inerte. En alguna divulgación del presente documento, los elementos y los interelementos pueden diferir en sus propiedades de hidrofilia o hidrofobia. En alguna divulgación del presente documento, los elementos y los interelementos pueden comprender un modificador como se describe en el presente documento.

5 Las matrices pueden construirse, pedirse por encargo o adquirirse en un proveedor comercial (por ejemplo, CombiMatrix, Agilent, Affymetrix, Nimblegen). Los oligonucleótidos se fijan, se aplican puntualmente, se inmovilizan, se unen a la superficie, se soportan o se sintetizan sobre los elementos diferenciados de la superficie o matriz. Los oligonucleótidos pueden fijarse covalentemente a la superficie o depositarse sobre la superficie. En la técnica se conocen bien diversos métodos de construcción, por ejemplo, sintetizadores de matriz sin máscara, métodos dirigidos por luz que utilizan máscaras, métodos de canal de flujo, métodos de aplicación puntual, etc.

15 En alguna divulgación del presente documento, los oligonucleótidos de construcción y/o selección pueden sintetizarse sobre un soporte sólido usando un sintetizador de matriz sin máscara (MAS). Se describen sintetizadores de matriz sin máscara, por ejemplo, en la solicitud PCT N.º WO 99/42813 y en la Patente de los EE.UU. N.º 6.375.903 correspondiente. Se conocen otros ejemplos de instrumentos sin máscara que pueden fabricar una micromatriz de ADN a petición del cliente en la que cada uno de los elementos de la matriz tiene una molécula de ADN monocatenaria de secuencia deseada.

20 Otros métodos para sintetizar oligonucleótidos de construcción y/o selección incluyen, por ejemplo, métodos dirigidos por luz que utilizan máscaras, métodos de canal de flujo, métodos de aplicación puntual, métodos basados en pasadores y métodos que utilizan múltiples soportes.

25 Se describen métodos dirigidos por luz que utilizan máscaras (por ejemplo, métodos VLSIPS™) para la síntesis de oligonucleótidos, por ejemplo, en las Patentes de los EE.UU. N.º 5.143.854; 5.510.270 y 5.527.681. Estos métodos implican activar regiones predefinidas de un soporte sólido y después poner en contacto el soporte con una solución de monómeros preseleccionada. Las regiones seleccionadas pueden activarse mediante la irradiación con una fuente de luz a través de una máscara de manera muy parecida a las técnicas fotolitográficas utilizadas en la fabricación de circuitos integrados. Otras regiones del soporte permanecen inactivas porque la iluminación es bloqueada por la máscara y permanecen protegidas químicamente. De este modo, un patrón de luz define qué regiones del soporte reaccionan con un monómero dado. Activando repetidamente diferentes conjuntos de regiones predefinidas y poniendo en contacto diferentes soluciones de monómeros con el soporte, se produce una serie diversa de polímeros sobre el soporte. Pueden usar opcionalmente otras etapas, tales como lavar la solución de monómeros sin reaccionar del soporte. Otros métodos aplicables incluyen técnicas mecánicas tales como las que se describen en la Patente de los EE.UU. N.º 5.384.261.

35 Se describen métodos adicionales aplicables a la síntesis de oligonucleótidos de construcción y/o selección en sobre solo soporte, por ejemplo, en la Patente de los EE.UU. N.º 5.384.261. Por ejemplo, pueden entregarse reactivos al soporte ya sea (1) fluyendo dentro de un canal definido sobre regiones predefinidas o (2) "aplicándolos puntualmente" en regiones predefinidas. También pueden emplearse otros enfoques, así como combinaciones de aplicación puntual y flujo. En cada caso, determinadas regiones activadas del soporte se separan mecánicamente de otras regiones cuando las soluciones de monómeros se entregan a los diversos sitios de reacción. Los métodos de canal de flujo implican, por ejemplo, sistemas microfluidos para controlar la síntesis de oligonucleótidos sobre un soporte sólido. Por ejemplo, pueden sintetizarse diversas secuencias de polímeros en regiones seleccionadas de un soporte sólido mediante la formación de canales de flujo sobre una superficie del soporte a través de los cuales fluyen reactivos apropiados o en los que se colocan reactivos apropiados. Los métodos de aplicación puntual para la preparación de oligonucleótidos sobre un soporte sólido incluyen la entrega de reactivos en cantidades relativamente pequeñas depositándolos directamente en regiones seleccionadas. En algunas etapas, toda la superficie de soporte puede pulverizarse o recubrirse con una solución, si es más eficiente hacerlo. Pueden depositarse gota a gota alícuotas de las soluciones de monómeros medidas con precisión mediante un dispensador que se mueve de una región a otra.

40 Se describen métodos basados en pasadores para la síntesis de oligonucleótidos sobre un soporte sólido, por ejemplo, en la Patente de los EE.UU. N.º 5.288.514. Los métodos basados en pasadores utilizan un soporte que tiene una pluralidad de pasadores u otras prolongaciones. Cada uno de los pasadores se inserta simultáneamente en recipientes de reactivos individuales en una bandeja. Una matriz de 96 pasadores se utiliza habitualmente con una bandeja de 96 recipientes, tal como una placa de microtitulación de 96 pocillos. Cada bandeja se llena con un reactivo particular para el acoplamiento en una reacción química particular sobre un pasador individual. En consecuencia, las bandejas con frecuencia contendrán diferentes reactivos. Puesto que las reacciones químicas se han optimizado de manera que cada una de las reacciones pueda realizarse con un conjunto relativamente similar de condiciones de reacción, es posible realizar múltiples etapas de acoplamiento químico simultáneamente.

45 Otras micromatrices y métodos adecuados para sintetizar oligonucleótidos incluyen los que se describen en las Patentes de los EE.UU. N.º 7.323.320 y 7.563.600, cuyas divulgaciones enteras se incorporan en su totalidad por referencia en el presente documento. En un ejemplo, los oligonucleótidos sintetizados a partir de los mismos se escinden químicamente, enzimáticamente o físicamente, o se liberan de otro modo de las micromatrices para una

amplificación, digestión con enzimas de restricción y/o ensamblaje adicionales.

En otra divulgación del presente documento, una pluralidad de oligonucleótidos puede sintetizarse o inmovilizarse (por ejemplo, fijarse) sobre múltiples soportes, tales como perlas. Un ejemplo es un método de síntesis basado en perlas que se describe, por ejemplo, en las Patentes de los EE.UU. N.º 5.770.358; 5.639.603; y 5.541.061. Para la síntesis de moléculas tales como oligonucleótidos sobre perlas, se suspende una gran cantidad de perlas en un vehículo adecuado (tal como agua) en un recipiente. Las perlas están provistas de moléculas espaciadoras opcionales que tienen un sitio activo con el que forma complejo, opcionalmente, un grupo protector. En cada etapa de la síntesis, las perlas se dividen para acoplarse en una pluralidad de recipientes. Una vez que las cadenas de oligonucleótidos nacientes se desprotegen, se añade una solución de monómero diferente a cada recipiente, de manera que, en todas las perlas de un recipiente dado, se produce la misma reacción de adición de nucleótidos. Después, las perlas se lavan para retirar el exceso de reactivos, se agrupan en un solo recipiente, se mezclan y se redistribuyen en otra pluralidad de recipientes en preparación para la siguiente ronda de síntesis. Debe observarse que, en virtud del gran número de perlas utilizadas al principio, habrá, de forma similar, un gran número de perlas dispersadas al azar en el recipiente, cada una de las cuales tendrá una secuencia oligonucleotídica única sintetizada sobre una superficie de la misma después de numerosas rondas de adición aleatoria de bases. Una perla individual puede marcarse con una secuencia que sea única para el oligonucleótido bicatenario sobre la misma, para permitir la identificación durante el uso.

En otra divulgación más del presente documento, una pluralidad de oligonucleótidos puede fijarse o sintetizarse sobre nanopartículas. Las nanopartículas incluyen, pero sin limitación, materiales coloidales de metal (por ejemplo, oro, plata, cobre y platino), semiconductores (por ejemplo, CdSe, CdS y CdS recubierto con ZnS) y magnéticos (por ejemplo, ferromagnetita). Se conocen en la técnica métodos para fijar oligonucleótidos a las nanopartículas. En otra realización, las nanopartículas se fijan al sustrato. Las nanopartículas con o sin oligonucleótidos inmovilizados pueden fijarse a sustratos como se desvela en, por ejemplo, Grabar et al., *Analyt. Chem.*, 67, 73-743 (1995); Bethell et al., *J. Electroanal. Chem.*, 409, 137 (1996); Bar et al., *Langmuir*, 12, 1172 (1996); Colvin et al., *J. Am. Chem. Soc.*, 114, 5221 (1992). Las nanopartículas desnudas pueden fijarse en primer lugar al sustrato y los oligonucleótidos pueden fijarse a las nanopartículas inmovilizadas.

Las secuencias oligonucleotídicas y/o polinucleotídicas presintetizadas pueden fijarse a un soporte o sintetizarse in situ usando métodos dirigidos por luz, canal de flujo y métodos de aplicación puntual, métodos de inyección de tinta, métodos basados en pasadores y métodos basados en perlas establecidos en las siguientes referencias: McGall et al. (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93:13555; *Synthetic DNA Arrays In Genetic Engineering*, Vol. 20:111, Plenum Press (1998); Duggan et al. (1999) *Nat. Genet.* S21:10; *Microarrays: Making Them and Using Them In Microarray Bioinformatics*, Cambridge University Press, 2003; la Publicaciones de Solicitud de Patente de los EE.UU. N.º 2003/0068633 y 2002/0081582; las Patentes de los EE.UU. N.º 6.833.450, 6.830.890, 6.824.866, 6.800.439, 6.375.903 y 5.700.637; y las publicaciones PCT N.º WO 04/031399, WO 04/031351, WO 04/029586, WO 03/100012, WO 03/066212, WO 03/065038, WO 03/064699, WO 03/064027, WO 03/064026, WO 03/046223, WO 03/040410 y WO 02/24597; cuyas divulgaciones se incorporan por referencia en el presente documento en su totalidad para todos los fines. En alguna divulgación del presente documento, los oligonucleótidos presintetizados se fijan a un soporte o se sintetizan usando una metodología de aplicación puntual en la que las soluciones de monómeros se depositan gota a gota mediante un dispensador que se mueve de una región a otra (por ejemplo, chorro de tinta). En alguna divulgación del presente documento, los oligonucleótidos se aplican puntualmente sobre un soporte usando, por ejemplo, un dispensador accionado por onda mecánica.

Una preparación de un oligonucleótido modificado por ingeniería genética para que tenga una determinada secuencia puede incluir moléculas de oligonucleótidos que tengan la secuencia diseñada además de moléculas de oligonucleótidos que contengan errores (por ejemplo, que difieran de la secuencia diseñada al menos en una posición). Un error de secuencia puede incluir una o más supresiones, adiciones, sustituciones (por ejemplo, transversión o transición), inversiones, duplicaciones de nucleótidos o cualquier combinación de dos o más de las mismas. Los errores de oligonucleótidos pueden generarse durante la síntesis de oligonucleótidos. Diferentes técnicas de síntesis pueden ser propensas a diferentes perfiles y frecuencias de error. En alguna divulgación del presente documento, las tasas de error pueden variar de 1/10 a 1/200 errores por base, dependiendo del protocolo de síntesis que se use. Sin embargo, en alguna divulgación del presente documento, pueden conseguirse menores tasas de error. Además, los tipos de errores pueden depender de las técnicas de síntesis que se usen. Por ejemplo, en alguna divulgación del presente documento, la síntesis de oligonucleótidos basada en chips puede dar como resultado relativamente más supresiones que las técnicas de síntesis basadas en columnas.

En alguna divulgación del presente documento, una o más preparaciones de oligonucleótidos pueden someterse a un proceso de reducción de errores o de filtración de errores para retirar (o reducir el número o la frecuencia de) oligonucleótidos que contienen errores. Dicho proceso puede usarse para aumentar el número de oligonucleótidos sin errores en las preparaciones de oligonucleótidos. Los métodos para realizar la reducción de errores o la filtración de errores pueden incluir, por ejemplo, la hibridación con un oligonucleótido de selección, la unión a un agente de unión de desapareamiento o a una proteína de unión de desapareamiento o combinaciones de los mismos.

En alguna divulgación del presente documento, puede usarse una técnica de hibridación en la que una preparación



de oligonucleótidos (es decir, oligonucleótidos de construcción) se hibride en condiciones rigurosas, una o más veces, con una preparación de oligonucleótidos inmovilizados (es decir, oligonucleótidos de selección) diseñados para que tengan una secuencia complementaria. La expresión "oligonucleótido de selección", como se usa en el presente documento, se refiere a un oligonucleótido monocatenario que es complementario al menos a una parte de un oligonucleótido de construcción (o el complemento del oligonucleótido de construcción). Los oligonucleótidos de selección pueden usarse para retirar copias de un oligonucleótido de construcción que contiene errores de secuenciación (por ejemplo, una desviación de la secuencia deseada) de un grupo de oligonucleótidos de construcción. En alguna divulgación del presente documento, un oligonucleótido de selección puede inmovilizarse terminalmente sobre un sustrato. En otra divulgación más del presente documento, los oligonucleótidos de selección pueden estar en solución. En una realización, los oligonucleótidos de selección pueden ser oligonucleótidos sintéticos que se han sintetizado en paralelo sobre un sustrato como se desvela en el presente documento.

Los oligonucleótidos de construcción que no se unen o que forman dúplex inestables pueden retirarse con el fin de retirar selectivamente o específicamente oligonucleótidos que contienen errores que desestabilizarían la hibridación en las condiciones utilizadas. Debe apreciarse que este proceso puede no retirar todos los oligonucleótidos que contienen errores puesto que algunos oligonucleótidos que contienen errores aún pueden unirse a los oligonucleótidos de selección inmovilizados con suficiente afinidad a través de este proceso de selección. Por ejemplo, los oligonucleótidos que contienen errores pueden diferir de los oligonucleótidos de selección en una o dos bases y aún pueden unirse a los oligonucleótidos de selección en las condiciones de reacción del proceso de selección.

En alguna divulgación del presente documento, puede incluirse una proteína de unión a ácido nucleico o recombinasa (por ejemplo, RecA) en una o más de las etapas de procesamiento de oligonucleótidos para mejorar la selección de oligonucleótidos sin errores. Por ejemplo, promoviendo preferentemente la hibridación de oligonucleótidos que sean completamente complementarios a los oligonucleótidos inmovilizados, puede reducirse la cantidad de oligonucleótidos que contienen errores que estén unidos. Como resultado, el procedimiento de procesamiento de oligonucleótidos que se describe en el presente documento puede retirar más oligonucleótidos que contienen errores y generar una preparación de oligonucleótidos que tenga una frecuencia de error más baja (por ejemplo, con una tasa de error inferior a 1/50, inferior a 1/100, inferior a 1/200, inferior a 1/300, inferior a 1/400, inferior a 1/500, inferior a 1/1.000 o inferior a 1/2.000 errores por base).

En alguna divulgación del presente documento, la corrección de errores puede incluirse entre cada repetición del proceso y al final del proceso de síntesis para aumentar la población relativa de polinucleótidos sintetizados sin desviación de las secuencias deseadas. Dicha corrección de errores puede incluir la secuenciación directa y/o la aplicación de corrección de errores basada en enzimas correctoras, tales como las nucleasas correctoras de errores (por ejemplo, CEL I), corrección de errores basada en la unión de MutS o homólogos de MutS u otras proteínas de unión de desapareamiento (véase, por ejemplo, la Solicitud Internacional N.º PCT/US2010/057405), otros medios de corrección de errores conocidos en la técnica o cualquier combinación de los mismos. En una realización de ejemplo, CEL I puede añadirse a los dúplex de oligonucleótidos en el medio fluido. CEL I es una endonucleasa específica de desapareamiento que escinde todos los tipos de desapareamientos, tales como polimorfismos de un solo nucleótido, inserciones pequeñas o supresiones. La adición de la endonucleasa da como resultado la escisión de los oligonucleótidos bicatenarios en el sitio o región del desapareamiento.

Debe apreciarse que no se incluyen una o más proteínas de unión a ácidos nucleicos o recombinasas preferentemente en una técnica de optimización de fidelidad posterior a la síntesis (por ejemplo, una técnica de detección que usa MutS o un homólogo de MutS), porque el procedimiento de optimización implica la retirada de ácidos nucleicos que contienen errores a través de la producción y retirada de heterodúplex. En consecuencia, cualesquier proteínas de unión a ácido nucleico o recombinasas (por ejemplo, RecA) que se incluyeron en las etapas de síntesis se retiran preferentemente (por ejemplo, mediante inactivación, purificación en columna u otra técnica adecuada) después de la síntesis y antes de la optimización de la fidelidad.

En determinada divulgación del presente documento, puede ser útil incluir uno o más oligonucleótidos modificados. Un oligonucleótido puede modificarse incorporando una base modificada (por ejemplo, un análogo de nucleótido) durante la síntesis, modificando el oligonucleótido después de la síntesis o cualquier combinación de los mismos. Los ejemplos de modificaciones incluyen, pero sin limitación, uno o más de los siguientes: bases universales tales como nitro indoles, dP y dK, inosina, uracilo; bases halogenadas tales como BrdU; bases marcadas con fluorescencia; marcadores no radiactivos tales como biotina (como derivado de dT) y digoxigenina (DIG); 2,4-dinitrofenilo (DNP); nucleótidos radiactivos; modificación posterior al acoplamiento tal como dR-NH<sub>2</sub> (desoxirribosa-NEb); acridina (6-cloro-2-metoxiacridina); y fosforamidas espaciadoras que se usan durante la síntesis para añadir un "brazo" espaciador a la secuencia, tales como C3, C8 (octanodiol), C9, C12, HEG (hexaetilenglicol) y C18.

#### Oligonucleótidos amplificadores

Los oligonucleótidos pueden proporcionarse o sintetizarse como productos sintéticos monocatenarios. En alguna divulgación del presente documento, los oligonucleótidos también pueden proporcionarse o sintetizarse como preparaciones bicatenarias que incluyen una cadena complementaria hibridada. Los oligonucleótidos pueden ser

moléculas de ADN, ARN, ANP o cualquier combinación de los mismos. Se puede producir un oligonucleótido bicatenario amplificando un oligonucleótido sintético monocatenario u otro molde adecuado (por ejemplo, una secuencia en una preparación de ácido nucleico tal como un vector de ácido nucleico o ácido nucleico genómico). En consecuencia, puede proporcionarse una pluralidad de oligonucleótidos diseñados para que tengan las características de secuencia que se describen en el presente documento como una pluralidad de oligonucleótidos monocatenarios que tengan esas características o también pueden proporcionarse junto con oligonucleótidos complementarios. En alguna divulgación del presente documento, un oligonucleótido puede estar fosforilado (por ejemplo, con un fosfato 5'). En alguna divulgación del presente documento, un oligonucleótido puede no estar fosforilado.

En alguna divulgación del presente documento, un oligonucleótido puede amplificarse usando un par de cebadores apropiado con un cebador correspondiente a cada extremo del oligonucleótido (por ejemplo, uno que sea complementario al extremo 3' del oligonucleótido y uno que sea idéntico al extremo 5' del oligonucleótido). En alguna divulgación del presente documento, un oligonucleótido puede diseñarse para que contenga una secuencia de ensamblaje central (diseñada para que se incorpore en el ácido nucleico objetivo) flanqueada por una secuencia de amplificación 5' (por ejemplo, una secuencia universal 5') y/o una secuencia de amplificación 3' (por ejemplo, una secuencia universal 3'). Pueden usarse cebadores de amplificación (por ejemplo, de entre 10 y 50 nucleótidos de longitud, de entre 15 y 45 nucleótidos de longitud, de aproximadamente 25 nucleótidos de longitud, etc.) correspondientes a las secuencias de amplificación flanqueantes para amplificar el oligonucleótido (por ejemplo, un cebador puede ser complementario a la secuencia de amplificación 3' y un cebador pueden tener la misma secuencia que la secuencia de amplificación 5'). Las secuencias de amplificación pueden retirarse después del oligonucleótido amplificado usando cualquier técnica adecuada para producir un oligonucleótido que contenga solo la secuencia de ensamblaje.

En alguna divulgación del presente documento, una pluralidad de oligonucleótidos diferentes (por ejemplo, aproximadamente 5, 10, 50, 100 o más) con diferentes secuencias de ensamblaje centrales pueden tener secuencias de amplificación 5' idénticas y/o secuencias de amplificación 3' idénticas. Todos estos oligonucleótidos pueden amplificarse en la misma reacción usando los mismos cebadores de amplificación.

Una pluralidad de oligonucleótidos utilizados en una reacción de ensamblaje puede contener preparaciones de oligonucleótidos sintéticos, oligonucleótidos monocatenarios, oligonucleótidos bicatenarios, productos de amplificación, oligonucleótidos que se procesan para retirar (o reducir la frecuencia de) variantes que contienen errores, etc. o cualquier combinación de dos o más de los mismos. En algunos aspectos, pueden usarse productos de amplificación bicatenarios como oligonucleótidos de ensamblaje y pueden añadirse a una reacción de ensamblaje como se describe en el presente documento. En alguna divulgación del presente documento, el oligonucleótido puede amplificarse mientras está todavía fijado al soporte. En alguna divulgación del presente documento, el oligonucleótido puede retirarse o escindirse del soporte antes de la amplificación o después de la amplificación.

En alguna divulgación del presente documento, un oligonucleótido sintético puede incluir una secuencia de ensamblaje central flanqueada por secuencias de amplificación 5' y 3'. La secuencia de ensamblaje central se diseña para su incorporación en un ácido nucleico objetivo ensamblado o subensamblaje objetivo. Las secuencias flanqueantes se diseñan para la amplificación y no se pretende que se incorporen en el ácido nucleico ensamblado. Las secuencias de amplificación flanqueantes pueden usarse como secuencias cebadoras universales para amplificar una pluralidad de oligonucleótidos de ensamblaje diferentes que comparten las mismas secuencias de amplificación pero que tienen secuencias de ensamblaje centrales diferentes. En alguna divulgación del presente documento, las secuencias flanqueantes se retiran después de la amplificación para producir un oligonucleótido que contiene solo la secuencia de ensamblaje.

En determinada divulgación del presente documento, los productos de amplificación bicatenarios pueden someterse a digestión con enzimas de restricción para retirar las secuencias flanqueantes. Para ese fin, las secuencias flanqueantes pueden diseñarse para incluir uno o más sitios de restricción o sitios de reconocimiento de enzimas de restricción. El sitio de restricción puede estar presente en el extremo 5' o 3' de la secuencia de amplificación siempre que el sitio de escisión esté entre la secuencia flanqueante que se retirará y la secuencia de ensamblaje central. El sitio de restricción puede incluirse en la secuencia de amplificación (es decir, el sitio de unión del cebador). El sitio de restricción también puede estar fuera de la secuencia de amplificación.

Después de la digestión con enzimas de restricción, las secuencias flanqueantes escindidas pueden separarse y retirarse usando cualquier técnica adecuada. En alguna divulgación del presente documento, las secuencias flanqueantes escindidas pueden ser fragmentos de menos de aproximadamente 40, aproximadamente 35, aproximadamente 30, aproximadamente 25, aproximadamente 20 o aproximadamente 15 bases de longitud. Como tal, pueden usarse técnicas de separación dependientes del tamaño conocidas en la técnica, tales como afinidad diferencial a la sílice, filtración por tamaño, precipitación diferencial con PEG (polietilenglicol) o CTAB (bromuro de cetiltrimetilamonio) o cualquier combinación de los mismos, de manera de separar las secuencias flanqueantes escindidas de las secuencias de ensamblaje centrales que pueden diseñarse para que tengan un tamaño más largo que las secuencias flanqueantes.

En alguna divulgación del presente documento, los cebadores de amplificación pueden estar biotinilados. Por tanto, los productos de amplificación resultantes también están biotinilados en ambos extremos. Tras la digestión con enzimas de restricción, las secuencias flanqueantes escindidas que tienen los cebadores biotinilados conservan los marcadores de biotina, mientras que las secuencias de ensamblaje centrales no están biotiniladas. De este modo, las secuencias flanqueantes escindidas pueden purificarse por afinidad y retirarse usando estreptavidina (por ejemplo, unida a una perla, columna u otra superficie). En alguna divulgación del presente documento, los cebadores de amplificación también pueden diseñarse para incluir determinadas características de secuencia (por ejemplo, sitios de restricción) que pueden usarse para retirar las regiones cebadoras después de la amplificación con el fin de producir un fragmento de ensamblaje bicatenario que incluya la secuencia de ensamblaje sin las secuencias de amplificación flanqueantes.

#### Salientes monocatenarios

La divulgación del presente documento implica ácidos nucleicos bicatenarios con salientes monocatenarios. Se pueden generar salientes usando cualquier técnica adecuada. En alguna divulgación del presente documento, un fragmento de ácido nucleico bicatenario (por ejemplo, un fragmento ensamblado en un ensamblaje múltiple) puede digerirse con una enzima de restricción apropiada para generar un saliente monocatenario terminal. En alguna divulgación del presente documento, pueden digerirse fragmentos que estén diseñados para ser adyacentes entre sí en un producto ensamblado con la misma enzima para exponer salientes complementarios. También pueden usarse diferentes enzimas que generen salientes complementarios.

En alguna divulgación del presente documento, pueden generarse salientes usando una enzima de restricción de tipo IIS. Las enzimas de restricción de tipo IIS son enzimas que se unen a un ácido nucleico bicatenario en un sitio, denominado sitio de reconocimiento y hacen un único corte bicatenario fuera del sitio de reconocimiento. El corte bicatenario, denominado sitio de escisión, generalmente se encuentra a 0-20 bases de distancia del sitio de reconocimiento. El sitio de reconocimiento generalmente tiene entre 4 y 8 pb de longitud. Todas las enzimas de restricción de tipo IIS presentan al menos un reconocimiento asimétrico parcial. Reconocimiento asimétrico significa que las secuencias de reconocimiento 5'→3' son diferentes para cada cadena del ácido nucleico. La actividad enzimática también muestra polaridad, lo que significa que los sitios de escisión están ubicados en un solo lado del sitio de reconocimiento. De este modo, generalmente solo hay un corte bicatenario correspondiente a cada sitio de reconocimiento. La escisión generalmente produce salientes monocatenarios de 1-6 nucleótidos, con extremos 5' o 3', aunque algunas enzimas producen extremos romos. Cualquiera de los cortes es útil en el contexto de la divulgación, aunque en algunos casos se producen los que producen salientes monocatenarios. Hasta ahora, se han identificado aproximadamente 80 enzimas de tipo IIS. Los ejemplos adecuados incluyen pero sin limitación BstF5 I, BtsC I, BsrD I, Bts I, Alw I, Bcc I, BsmA I, Ear I, Mly I (romo), Ple I, Bmr I, Bsa I, BsmB I, BspQ I, Fau I, Mnl I, Sap I, Bbs I, BciV I, Hph I, Mbo II, BfuA I, BspCN I, BspM I, SfaN I, Hga I, BseR I, Bbv I, Eci I, Fok I, BceA I, BsmF I, BtgZ I, BpuE I, Bsg I, Mme I, BseG I, Bse3D I, BseM I, AclW I, Alw26 I, Bst6 I, BstMA I, EamI 104 I, Ksp632 I, Pps I, Sch I (romo), Bfi I, Bso31 I, BspTN I, Eco31 I, Esp3 I, Smu I, Bfu I, Bpi I, BpuA I, BstV2 I, AsuHP I, Acc36 I, Lwe I, Aar I, BseM II, TspDT I, TspGW I, BseX I, BstVI I, Eco571, Eco57M I, Gsu I y Beg I. En alguna divulgación del presente documento pueden usarse Bsa I, BsmB I, BspQ I, BtgZ I, BsmF I, Fok I, Bbv I, cualquier variante de los mismos o cualquier combinación de los mismos. Dichas enzimas e información con respecto a sus sitios de reconocimiento y escisión están disponibles de proveedores comerciales tales como New England Biolabs.

En alguna divulgación del presente documento, cada uno de una pluralidad de fragmentos de ácido nucleico diseñados para el ensamblaje puede tener un sitio de restricción de tipo IIS en cada extremo. Los sitios de restricción de tipo IIS pueden orientarse de manera que los sitios de escisión sean internos con respecto a las secuencias de reconocimiento. Como resultado, la digestión enzimática expone una secuencia interna (por ejemplo, un saliente dentro de una secuencia interna) y retira las secuencias de reconocimiento de los extremos. En consecuencia, pueden usarse los mismos sitios de tipo IIS de tipo para ambos extremos de todos los fragmentos de ácido nucleico que se preparan para el ensamblaje. Sin embargo, también pueden usarse diferentes sitios de tipo IIS. Dos fragmentos que se diseñan para que sean adyacentes en un producto ensamblado pueden incluir cada uno una secuencia terminal solapante idéntica y un sitio de tipo IIS flanqueante que se ubica de manera apropiada para exponer salientes complementarios dentro de la secuencia solapante tras la digestión con enzimas de restricción. En consecuencia, puede generarse una pluralidad de fragmentos de ácido nucleico con diferentes salientes complementarios. El sitio de restricción en cada extremo de un fragmento de ácido nucleico puede ubicarse de manera que la digestión con la enzima de tipo IIS apropiada retire el sitio de restricción y exponga una región monocatenaria que sea complementaria a una región monocatenaria en un fragmento de ácido nucleico que se diseña para que sea adyacente en el producto de ácido nucleico ensamblado. En determinada divulgación del presente documento, las enzimas de restricción pueden seleccionarse de manera que los fragmentos de ácido nucleico de ensamblaje estén libres de los sitios de restricción correspondientes.

Como se ha analizado anteriormente, los sitios de restricción pueden colocarse dentro o fuera, 5' o 3' con respecto a la secuencia de amplificación. Como ilustra la Figura 9A, los sitios de restricción (que se muestran en **negrita**) pueden incluirse dentro de la secuencia de amplificación (que se muestra en  *cursiva*) y distales con respecto al fragmento de ensamblaje central (**negro**). A modo de ejemplo, los sitios BtgZI y BsmFI se usan en cualquiera de los extremos del fragmento de ensamblaje bicatenario y sus respectivos sitios de escisión se indican mediante flechas.

BtgZI y BsmFI se escinden a 10 nucleótidos/14 nucleótidos de distancia de sus sitios de reconocimiento. También pueden usarse otras enzimas de restricción que se escinden a una distancia corta (por ejemplo, 5-25, 10-20 o aproximadamente 15 nucleótidos) del sitio de reconocimiento. Como alternativa, como ilustra la Figura 9B, los sitios de restricción (que se muestran en negrita) pueden estar fuera de la secuencia de amplificación (que se muestra en cursiva) y proximales con respecto al fragmento de ensamblaje central (fuente normal). Los sitios BsaI se usan en ambos extremos del fragmento de ensamblaje bicatenario como ejemplo, cuyos sitios de escisión también se indican mediante flechas. Como puede observarse a partir de las Figuras 9A y 9B, cuando los sitios de restricción se colocan distales con respecto al fragmento de ensamblaje central y se incluyen en la secuencia de amplificación, la longitud total del ácido nucleico de partida es más corta que cuando los sitios de restricción se colocan cerca del fragmento de ensamblaje central y no se incluyen en la secuencia de amplificación. Por tanto, la primera estrategia (Figura 9A) puede ser más rentable y menos propensa a errores para sintetizar ácidos nucleicos de partida más cortos (por ejemplo, sobre un chip). La primera estrategia también usa cebadores universales más cortos (para amplificar los fragmentos) y, por tanto, reduce los costes adicionalmente. Después de la digestión con enzimas de restricción, los trozos terminales que se han de retirar de los fragmentos de ensamblaje central también son más cortos y, por tanto, son más fáciles, más baratos y más rápidos de retirar en la primera estrategia que en la segunda.

Las digestiones enzimáticas de ADN con enzimas de restricción de tipo IIS u otras enzimas de restricción específicas de sitio normalmente generan un saliente de cuatro a seis nucleótidos. Se muestra inesperadamente en la presente divulgación, que estos extremos cohesivos cortos son suficientes para ligar múltiples fragmentos de ácido nucleico que contengan extremos complementarios para formar el ácido nucleico objetivo. Convencionalmente, para garantizar la eficiencia, una reacción de ligadura generalmente implica dos fragmentos, ya que la eficiencia de la ligadura disminuye significativamente con tres o más fragmentos. Asimismo, los métodos convencionales requieren extremos cohesivos más largos para mejorar la especificidad, ya que con frecuencia se producen desapareamientos. Además, para seleccionar el producto de ligadura correcto, se requiere un proceso de clonación y detección que requiera mucho trabajo y tiempo.

La presente divulgación proporciona, entre otras cosas: (1) una ligadura satisfactoria de múltiples fragmentos (por ejemplo, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8 o más) en una sola reacción (por ejemplo, grupo único); (2) una reacción de ligadura rápida y económica (por ejemplo, 30 minutos a temperatura ambiente); (3) una alta especificidad que discrimina desapareamientos; y (4) una etapa de PCR rápida para seleccionar el producto correcto, sin necesidad de clonación y detección. Otra ventaja de la presente divulgación es la capacidad de usar directamente oligonucleótidos sintéticos de chips o micromatrices disponibles en el mercado para construir cualquier ácido nucleico objetivo de interés, que pueden ser de cualquier secuencia y/o cualquier longitud (por ejemplo, al menos 500 pb, al menos 1 kb, al menos 2 kb, al menos 5 kb, al menos 10 kb o más). Dichos oligonucleótidos sintéticos pueden ser sustancialmente del mismo tamaño (por ejemplo, aproximadamente 50 bases, aproximadamente 100 bases, aproximadamente 200 bases, aproximadamente 300 bases o más) y, por tanto, son fáciles de manipular.

En un ejemplo, asumiendo que cada oligonucleótido o fragmento sobre el chip tiene una carga útil de 100 nucleótidos y los fragmentos tienen salientes de 4 bases, si el número de fragmentos es  $n$ , entonces la longitud del producto de ligadura =  $(n \cdot 100) - (4 \cdot (n - 1))$ , con  $(n - 1)$  uniones de ligadura. Debe tenerse en cuenta que, para garantizar la especificidad de la ligadura, los salientes pueden seleccionarse o diseñarse para que sean únicos para cada sitio de ligadura; es decir, cada par de salientes complementarios para dos fragmentos diseñados para que sean adyacentes en un producto ensamblado debe ser único y diferir de cualquier otro par de salientes complementarios en al menos un nucleótido.

Otra estrategia (ensamblaje de compensación) para exponer los extremos cohesivos se ilustra en la Figura 10A. A partir de un chip, puede sintetizarse una pluralidad de oligonucleótidos (por ejemplo,  $A_1$ - $A_{10}$ ). Los oligonucleótidos pueden diseñarse para que tengan secuencias de ensamblaje centrales que, cuando se ensamblan correctamente, forman el ácido nucleico objetivo 5'- $A_1$ - $A_3$ - $A_5$ - $A_7$ - $A_9$ -3' (siendo la cadena inversa 3'- $A_2$ - $A_4$ - $A_6$ - $A_8$ - $A_{10}$ -5'). Es decir, dos oligonucleótidos adyacentes  $A_n$  and  $A_{n+1}$  pueden diseñarse para solaparse. Como se usa en el presente documento, oligonucleótidos adyacentes se refiere a oligonucleótidos en los que un primer oligonucleótido está en el extremo 5' o en el extremo 3' de un segundo oligonucleótido a lo largo de la secuencia de ácido nucleico lineal. En alguna divulgación del presente documento, los oligonucleótidos adyacentes pueden ser contiguos. Como se usa en el presente documento, oligonucleótidos contiguos se refiere a dos oligonucleótidos en los que el primer oligonucleótido termina en la posición fijada arbitrariamente en -1 y el segundo fragmento comienza en la posición fijada arbitrariamente en 0 a lo largo de la secuencia de ácido nucleico lineal. Las secuencias de ensamblaje central pueden tener cualquier longitud deseable, tal como de aproximadamente 50-500 nucleótidos, aproximadamente 60-300 nucleótidos, aproximadamente 70-200 nucleótidos, o más cortos o más largos. La pluralidad de oligonucleótidos puede tener una longitud uniforme para facilitar su manipulación. A modo de ejemplo, los oligonucleótidos sintetizados también pueden incluir secuencias de amplificación en cada extremo, que pueden tener sitios de restricción incorporados. Las secuencias de amplificación pueden ser de aproximadamente 10-30 nucleótidos, aproximadamente 15-25 nucleótidos, o más cortos o más largos. La Figura 10A muestra secuencias de ensamblaje centrales de 70-meros y oligonucleótidos totales de 120-meros. Los oligonucleótidos sintetizados pueden eluirse, escindirse, o liberarse del chip de otro modo, y someterse a amplificación por PCR usando el par de cebadores  $A_L$  y  $A_R$ . Los productos amplificados pueden escindirse (por ejemplo, con una enzima de restricción) para retirar las

secuencias de amplificación (cabezas de flecha) y las secuencias de ensamblaje bicatenarias centrales de 70-meros pueden purificarse a partir de los mismos. Estas secuencias de ensamblaje bicatenarias pueden fundirse (por ejemplo, a 95 °C) y volver a hibridarse (por ejemplo, a 65 °C) en una sola etapa de reorganización. Después de reorganizar los oligonucleótidos monocatenarios, el 25 % de los productos serán productos de ensamblaje de compensación (por ejemplo, A<sub>1</sub>/A<sub>2</sub>, A<sub>2</sub>/A<sub>3</sub>, A<sub>3</sub>/A<sub>4</sub>, A<sub>4</sub>/A<sub>5</sub>, etc.) que tengan extremos cohesivos. Estos extremos cohesivos pueden ensamblarse entre sí (en etapas o en una sola reacción jerárquicamente) usando una ligasa, formando de este modo el ácido nucleico objetivo 5'-A<sub>1</sub>-A<sub>3</sub>-A<sub>5</sub>-A<sub>7</sub>-A<sub>9</sub>-3' (siendo la cadena inversa 3'-A<sub>2</sub>-A<sub>4</sub>-A<sub>6</sub>-A<sub>8</sub>-A<sub>10</sub>-5'). Debe apreciarse que los oligonucleótidos también pueden diseñarse de manera que el ácido nucleico objetivo sea 5'-A<sub>1</sub>...A<sub>3</sub>...A<sub>5</sub>...A<sub>7</sub>...A<sub>9</sub>-3' (es decir, se permiten huecos entre A<sub>n</sub> y A<sub>n+2</sub>, que puede rellenarse usando una secuencia A<sub>n+1</sub> como molde). Para ese fin, pueden usarse una polimerasa y dNTP para prolongar y llenar los huecos antes de la ligadura.

Una segunda estrategia de ensamblaje de compensación se ilustra en la Figura 10B, donde puede usarse una sola etapa de combinada de ensamblaje-(prolongación)-ligadura, a diferencia de dos etapas separadas (es decir, etapa de ensamblaje y etapa de ligadura). Por ejemplo, después de la etapa de reorganización (por ejemplo, fusión a 95 °C y rehibridación a 65 °C), los oligonucleótidos de análisis sin huecos pueden ligarse para formar un producto de longitud completa o un producto de subensamblaje. Si hay presentes huecos en el análisis, los oligonucleótidos pueden incubarse en presencia de una polimerasa y dNTP para rellenar los huecos por prolongación de la cadena antes de la ligadura. En alguna divulgación del presente documento, el análisis con espacios puede someterse simultáneamente a prolongación de cadena por polimerasa y ligadura. Como se usa en el presente documento, el término "subensamblaje" se refiere a una molécula de ácido nucleico que se ha ensamblado a partir de un conjunto de oligonucleótidos de construcción. Preferentemente, un subensamblaje es al menos aproximadamente 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 20 veces, 50 veces, 100 veces o más, más largo que los oligonucleótidos de construcción.

También pueden usarse otros métodos para generar extremos cohesivos. Por ejemplo, puede usarse un método basado en polimerasa (por ejemplo, ADN polimerasa T4) para sintetizar extremos cohesivos deseables. Independientemente del método de generación de salientes específicos (por ejemplo, salientes complementarios para ácidos nucleicos diseñados para que sean adyacentes en un producto de ácido nucleico ensamblado), pueden diseñarse y/o producirse salientes de diferentes longitudes. En alguna divulgación del presente documento, pueden usarse salientes monocatenarios largos (3' o 5') para promover la especificidad y/o el ensamblaje eficiente. Por ejemplo, un saliente monocatenario 3' o 5' puede tener una longitud superior a 8 bases, por ejemplo, 8-14, 14-20, 20-25, 25-50, 50-100, 100-500 o más bases de longitud.

### 35 Ensamblaje de alta fidelidad

De acuerdo con la divulgación del presente documento, puede ensamblarse una pluralidad de fragmentos de ácido nucleico en un único procedimiento en el que la pluralidad de fragmentos se mezcla en condiciones que promueven el ensamblaje covalente de los fragmentos para generar un ácido nucleico más largo específico. De acuerdo con una divulgación del presente documento, una pluralidad de fragmentos de ácido nucleico puede ensamblarse covalentemente *in vitro* usando una ligasa. En alguna divulgación del presente documento, pueden ensamblarse 5 o más (por ejemplo, 10 o más, 15 o más, de 15 a 20, de 20 a 25, de 25 a 30, de 30 a 35, de 35 a 40, de 40 a 45, de 45 a 50, 50 o más, etc.) fragmentos de ácido nucleico diferentes. Sin embargo, debe apreciarse que puede ensamblarse cualquier número de ácidos nucleicos (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, etc.) usando técnicas de ensamblaje adecuadas. Cada fragmento de ácido nucleico que se ensambla puede tener entre aproximadamente 100 nucleótidos de longitud y aproximadamente 1.000 nucleótidos de longitud (por ejemplo, aproximadamente 200, aproximadamente 300, aproximadamente 400, aproximadamente 500, aproximadamente 600, aproximadamente 700, aproximadamente 800, aproximadamente 900). Sin embargo, pueden ensamblarse fragmentos de ácido nucleico más largos (por ejemplo, de aproximadamente 2.500 o más nucleótidos, de aproximadamente 5.000 o más nucleótidos, de aproximadamente 7.500 o más nucleótidos, de aproximadamente 10.000 o más nucleótidos, etc.) o más cortos usando una técnica de ensamblaje (por ejemplo, ensamblaje aleatorio en un vector plasmídico). Debe apreciarse que el tamaño de cada fragmento de ácido nucleico puede ser independiente del tamaño de otros fragmentos de ácido nucleico añadidos a un ensamblaje. Sin embargo, en alguna divulgación del presente documento, cada fragmento de ácido nucleico puede tener aproximadamente el mismo tamaño o longitud (por ejemplo, entre aproximadamente 100 nucleótidos de longitud y aproximadamente 400 nucleótidos de longitud). Por ejemplo, la longitud de los oligonucleótidos puede tener una mediana de la longitud de entre aproximadamente 100 nucleótidos de longitud y aproximadamente 400 nucleótidos de longitud y varían en aproximadamente, +/- 1 nucleótidos, +/- 4 nucleótidos, +/- 10 nucleótidos. Debe apreciarse que la longitud de un fragmento de ácido nucleico bicatenario puede indicarse mediante el número de pares de bases. Como se usa en el presente documento, un fragmento de ácido nucleico denominado de "x" nucleótidos de longitud corresponde a "x" pares de bases de longitud cuando se usa en el contexto de un fragmento de ácido nucleico bicatenario. En alguna divulgación del presente documento, uno o más ácidos nucleicos que se ensamblan en una reacción (por ejemplo, 1-5, 5-10, 10-15, 15-20, etc.) pueden tener codones optimizados y/o no ser de origen natural. En alguna divulgación del presente documento, todos los ácidos nucleicos que se ensamblan en una reacción tienen codones optimizados y/o no son de origen natural.

En alguna divulgación del presente documento, los fragmentos de ácido nucleico que se ensamblan se diseñan para que tengan secuencias complementarias solapantes. En alguna divulgación del presente documento, los fragmentos de ácido nucleico son fragmentos de ácido nucleico bicatenarios con salientes monocatenarios 3' y/o 5'. Estos salientes pueden ser extremos cohesivos que pueden hibridarse con extremos cohesivos complementarios en diferentes fragmentos de ácido nucleico. De acuerdo con una divulgación del presente documento, la presencia de secuencias complementarias (y, en particular, de extremos cohesivos complementarios) en dos fragmentos de ácido nucleico promueve su ensamblaje covalente. En alguna divulgación del presente documento, se ensambla una pluralidad de fragmentos de ácido nucleico con diferentes extremos cohesivos monocatenarios complementarios solapantes y su orden en el producto de ácido nucleico ensamblado se determina mediante la identidad de los extremos cohesivos en cada fragmento. Por ejemplo, los fragmentos de ácido nucleico pueden diseñarse de manera que un primer ácido nucleico tenga un primer extremo cohesivo que sea complementario a un primer extremo cohesivo de un segundo ácido nucleico y un segundo extremo cohesivo que sea complementario a un primer extremo cohesivo de un tercer ácido nucleico. Un segundo extremo cohesivo del segundo ácido nucleico puede ser complementario a un primer extremo cohesivo de un cuarto ácido nucleico. Un segundo extremo cohesivo del tercer ácido nucleico puede ser complementario a un primer extremo cohesivo de un quinto ácido nucleico. Y así sucesivamente hasta el ácido nucleico final. De acuerdo con una divulgación del presente documento, esta técnica puede usarse para generar una disposición lineal que contenga fragmentos de ácido nucleico ensamblados en un orden lineal predeterminado (por ejemplo, primero, segundo, tercero, hasta el final).

En determinada divulgación del presente documento, las regiones complementarias solapantes entre fragmentos de ácido nucleico adyacentes se diseñan (o se seleccionan) para que sean lo suficientemente diferentes para promover (por ejemplo, favorecer termodinámicamente) el ensamblaje de una alineación única de fragmentos de ácido nucleico (por ejemplo, una alineación seleccionada o diseñada de fragmentos). Sorprendentemente, en condiciones de ligadura adecuadas, la diferencia en tan poco como un nucleótido proporciona suficiente poder de discriminación entre el apareamiento perfecto (100 % de extremos cohesivos complementarios) y el desapareamiento (menos del 100 % de extremos cohesivos complementarios). Como tal, los salientes de 4 bases pueden permitir que hasta  $(4^4+1)=257$  fragmentos diferentes se ligen con una alta especificidad y fidelidad.

Debe apreciarse que pueden usarse regiones solapantes de diferentes longitudes. En alguna divulgación del presente documento, pueden usarse extremos cohesivos más largos cuando se ensamblan números más altos de fragmentos de ácido nucleico. Los extremos cohesivos más largos pueden proporcionar más flexibilidad para diseñar o seleccionar secuencias suficientemente distintas para discriminar entre la hibridación de extremos cohesivos correcta (por ejemplo, que implica extremos cohesivos diseñados para hibridarse entre sí) y la hibridación de extremos cohesivos incorrecta (por ejemplo, entre extremos cohesivos no complementarios).

Para conseguir dicho ensamblaje de alta fidelidad, pueden usarse una o más ligasas adecuadas. Una ligasa puede obtenerse de fuentes recombinantes o naturales. En alguna divulgación del presente documento, puede usarse ADN ligasa T3, ADN ligasa T4, ADN ligasa T7 y/o ADN Ligasa de *E. coli*. Estas ligasas pueden usarse a una temperatura relativamente baja (por ejemplo, temperatura ambiente) y son particularmente útiles para salientes relativamente cortos (por ejemplo, salientes de aproximadamente 3, de aproximadamente 4, de aproximadamente 5 o de aproximadamente 6 bases). En determinadas reacciones de ligadura (por ejemplo, 30 min de incubación a temperatura ambiente), la ADN ligasa T7 puede ser más eficiente para la ligadura de múltiples vías que las otras ligasas. También puede usarse una ligasa termoestable, tal como una o más de entre ADN ligasa Tth; ADN ligasa Pfu; ligasa Taq, cualquier otra ligasa termoestable adecuada o cualquier combinación de las mismas.

En alguna divulgación del presente documento, pueden diseñarse o seleccionarse dos o más pares de extremos cohesivos complementarios entre diferentes fragmentos de ácido nucleico para que tengan secuencias idénticas o similares con el fin de promover el ensamblaje de productos que contengan una disposición (y/o número) relativamente aleatoria de los fragmentos que tienen extremos cohesivos similares o idénticos. Esto puede ser útil para generar bibliotecas de productos de ácido nucleico con diferentes disposiciones de secuencia y/o diferentes números de copias de determinadas regiones de secuencia interna.

Debe apreciarse que la variación en la concentración de fragmentos individuales que se han de ensamblar podría dar como resultado el ensamblaje de construcciones intermedias incompletas. Por ejemplo, en el ensamblaje de la secuencia de ácido nucleico objetivo (ABCDEF) usando oligonucleótidos A, B, C, D, E, F, teniendo cada uno de los cuales el extremo saliente cohesivo apropiado, si la concentración de los fragmentos individuales no es equimolar (por ejemplo, si la concentración de A, B y C es superior a la concentración de D, E y F), pueden formarse especies de terminación (tales como AB y BC) dando como resultado una mezcla de productos intermedios no ligados. Para evitar la formación de construcciones intermedias incompletas, el ácido nucleico objetivo puede ensamblarse a partir de al menos dos grupos de fragmentos individuales (por ejemplo, grupo 1: A, C, E y Grupo 2: B, D, F). En alguna divulgación del presente documento, cada uno de los dos grupos comprende una pluralidad de fragmentos de ácido nucleico, teniendo cada fragmento de ácido nucleico del primer grupo un extremo terminal complementario a un extremo terminal de un fragmento de ácido nucleico en el segundo grupo. En alguna divulgación del presente documento, los al menos dos grupos pueden formarse dividiendo la población de oligonucleótidos en los al menos dos grupos y amplificando los oligonucleótidos en cada grupo por separado. En otra divulgación del presente documento, los al menos dos grupos pueden formarse liberando (por ejemplo, eluyendo, escindiendo o amplificando)

oligonucleótidos de una primera matriz de oligonucleótidos en un primer grupo y liberando los oligonucleótidos de una segunda matriz de oligonucleótidos en un segundo grupo. En otra realización más, los al menos dos grupos diferentes pueden formarse amplificando secuencias de oligonucleótidos usando al menos dos conjuntos diferentes de marcadores de amplificación como se describe en el presente documento. A modo de ejemplo, el segundo grupo que comprende los oligonucleótidos B, D y F puede diluirse tal como la concentración molar de los oligonucleótidos B, D y F presentes en el segundo grupo es más baja que la concentración molar de los oligonucleótidos A, C y E presentes en el primer grupo. Por ejemplo, la concentración molar de los oligonucleótidos en el segundo grupo puede ser aproximadamente dos veces, 10 veces, 20 veces, 50 veces, 100 veces, o más, más baja que la concentración molar de los oligonucleótidos en el primer grupo. Después de mezclar y ligar los dos grupos, el producto resultante comprende el ácido nucleico objetivo que tiene la secuencia predeterminada y puede separarse del exceso de oligonucleótidos que forman el primer grupo. En determinada divulgación del presente documento, puede ser deseable formar grupos de dímeros de oligonucleótidos que tengan diferentes concentraciones molares. Por ejemplo, el ensamblaje de las secuencias de ácido nucleico objetivo ABCDEFGH puede realizarse usando al menos dos grupos diferentes, comprendiendo el primer grupo los oligonucleótidos A, B, E. Comprendiendo F y el segundo grupo los oligonucleótidos C, D, G, H. El segundo grupo puede diluirse de manera que la concentración molar de los oligonucleótidos C, D, G, H sea inferior (por ejemplo, 10 veces o 100 veces) a la concentración molar de los oligonucleótidos A, B, E, F. Los oligonucleótidos que tienen los extremos salientes cohesivos apropiados pueden ligarse para formar los productos intermedios AB y EF en el primer grupo y CD y GH en el segundo grupo. Puesto que la concentración molar de C, D, G, H es inferior a la concentración molar de A, B, E, F, la concentración molar de CD y GH es inferior a la concentración molar de AB y EF. Después de mezclar los productos intermedios AB, CD, EF, GH en condiciones de ligadura, el producto resultante que comprende el ácido nucleico objetivo que tiene la secuencia predeterminada puede separarse del exceso de dímeros AB y EF.

En alguna divulgación del presente documento, los fragmentos de ácido nucleico se mezclan y se incuban con una ligasa. Debe apreciarse que la incubación en condiciones que promueven la hibridación específica de los extremos cohesivos puede aumentar la frecuencia de ensamblaje (por ejemplo, ensamblaje correcto). En alguna divulgación del presente documento, los diferentes extremos cohesivos se diseñan para que tengan temperaturas de fusión similares (por ejemplo, dentro de aproximadamente 5 °C entre sí) de manera que se promueva la hibridación correcta de todos los fragmentos en las mismas condiciones. La hibridación correcta puede promoverse a una temperatura diferente dependiendo de la longitud de los extremos cohesivos que se usan. En alguna divulgación del presente documento, pueden usarse extremos cohesivos de entre aproximadamente 4 y aproximadamente 30 nucleótidos de longitud (por ejemplo, extremos cohesivos de aproximadamente 5, aproximadamente 10, aproximadamente 15, aproximadamente 20, aproximadamente 25 o aproximadamente 30 nucleótidos de longitud). Las temperaturas de incubación pueden oscilar entre aproximadamente 20 °C y aproximadamente 50 °C (incluyendo, por ejemplo, la temperatura ambiente). Sin embargo, pueden usarse temperaturas más altas o más bajas. La duración de la incubación puede optimizarse basándose en la longitud de los salientes, la complejidad de los salientes y el número de diferentes ácidos nucleicos (y, por tanto, el número de diferentes salientes) que se mezclan entre sí. El tiempo de incubación también puede depender de la temperatura de hibridación y la presencia o ausencia de otros agentes en la mezcla. Por ejemplo, puede añadirse una proteína de unión a ácido nucleico y/o una recombinasa (por ejemplo, RecA, por ejemplo, una proteína RecA termoestable).

El complejo resultante de ácidos nucleicos puede someterse a una reacción en cadena de la polimerasa, en presencia de un par de cebadores específicos de secuencia objetivo, para amplificar y seleccionar el producto de ligadura correcto (es decir, el ácido nucleico objetivo). Como alternativa, el complejo resultante de ácidos nucleicos puede ligarse en un vector adecuado y transformarse en una célula hospedadora para una detección adicional de colonias.

#### Análisis de secuencia y diseño y selección de fragmentos

Una divulgación del presente documento puede incluir el análisis de la secuencia de un ácido nucleico objetivo y el diseño de una estrategia de ensamblaje basada en la identificación de regiones, dentro de la secuencia de ácido nucleico objetivo, que pueden usarse para generar extremos cohesivos apropiados (por ejemplo, salientes monocatenarios). Estas regiones pueden usarse para definir los extremos de fragmentos de ácido nucleico que pueden ensamblarse (por ejemplo, en una reacción) para generar el ácido nucleico objetivo. Los fragmentos de ácido nucleico pueden entonces proporcionarse o fabricarse (por ejemplo, en una reacción de ensamblaje múltiple). Los fragmentos de ácido nucleico pueden seleccionarse de manera que tengan un tamaño relativamente uniforme para que sean fáciles de manipular (por ejemplo, purificación).

De acuerdo con alguna divulgación del presente documento, la secuencia de ácido nucleico puede diseñarse y/o analizarse de manera asistida por ordenador para generar un conjunto de oligonucleótidos bicatenarios o monocatenarios analizados. Como se usa en el presente documento, el término "analizado" significa que una secuencia de ácido nucleico objetivo se ha delineado, por ejemplo, de manera asistida por ordenador, tal como para identificar una serie de secuencias de oligonucleótidos adyacentes. Los oligonucleótidos adyacentes o fragmentos de ácido nucleico se solapan preferentemente en un número apropiado de nucleótidos para facilitar el ensamblaje de acuerdo con los métodos de la divulgación. Las secuencias de oligonucleótidos pueden sintetizarse individualmente y ensamblarse usando los métodos de la divulgación.

En alguna divulgación del presente documento, una secuencia de ácido nucleico objetivo puede analizarse para identificar regiones que contengan al menos un nucleótido diferente en una cadena del ácido nucleico objetivo. Estas regiones pueden usarse para generar extremos cohesivos. Debe apreciarse que la longitud de un extremo cohesivo es preferentemente suficiente para proporcionar especificidad. Por ejemplo, los extremos cohesivos pueden ser lo suficientemente largos para que tengan secuencias lo suficientemente diferentes (por ejemplo, al menos diferencias de 1 base) para prevenir o reducir el emparejamiento incorrecto entre extremos cohesivos similares. Sin embargo, su longitud preferentemente no es lo suficientemente larga como para estabilizar los pares erróneos entre secuencias cohesivas similares. En alguna divulgación del presente documento, puede usarse una longitud de aproximadamente 3 a aproximadamente 10 bases. Sin embargo, puede seleccionarse cualquier longitud adecuada para una región que se ha de usar para generar un saliente cohesivo. La importancia de la especificidad puede depender del número de fragmentos diferentes que se ensamblan simultáneamente. Además, la longitud adecuada requerida para evitar la estabilización de regiones emparejadas erróneamente puede depender de las condiciones utilizadas para la hibridación de diferentes extremos cohesivos.

En alguna divulgación del presente documento, pueden seleccionarse regiones alternas si están separadas por distancias que definen fragmentos con longitudes adecuadas para el diseño del ensamblaje. En alguna divulgación del presente documento, las regiones alternas pueden estar separadas por aproximadamente 100 a aproximadamente 500 bases. Sin embargo, puede seleccionarse cualquier distancia adecuada más corta o más larga. Por ejemplo, las regiones cohesivas pueden estar separadas por aproximadamente 200 a aproximadamente 1.000 bases. Debe apreciarse que pueden estar disponibles diferentes patrones de regiones alternas dependiendo de varios factores (por ejemplo, dependiendo de la secuencia del ácido nucleico objetivo, la longitud elegida de los extremos cohesivos y la longitud del fragmento deseado). En alguna divulgación del presente documento, si hay varias opciones disponibles, las regiones pueden seleccionarse para maximizar las diferencias de secuencia entre diferentes extremos cohesivos.

La selección de las regiones cohesivas define los fragmentos que se ensamblarán para generar el ácido nucleico objetivo. En consecuencia, el tamaño del fragmento puede tener una longitud de entre aproximadamente 100 y aproximadamente 500 pares de bases, entre aproximadamente 200 y aproximadamente 1.000 bases, o más corto o más largo dependiendo del ácido nucleico objetivo. Los fragmentos pueden generarse u obtenerse usando cualquier técnica adecuada. En alguna divulgación del presente documento, cada fragmento puede ensamblarse (por ejemplo, en una reacción de ensamblaje de dúplex múltiple) de manera que esté flanqueado por regiones bicatenarias que pueden usarse para generar las regiones monocatenarias cohesivas.

En alguna divulgación del presente documento, los métodos para permitir el ensamblaje de un polinucleótido objetivo se basan en la información de la secuencia del ácido nucleico objetivo. En alguna divulgación del presente documento, puede usarse un software de ordenador para analizar la secuencia objetivo (por ejemplo,  $A_1.A_n$ ) dividiéndola en un conjunto de oligonucleótidos solapantes ( $A_1, A_2, A_3, \dots A_n$ ) de longitud especificada. Los oligonucleótidos  $A_1, A_2, A_3, \dots A_n$  pueden sintetizarse a partir de un chip o micromatriz. En alguna divulgación del presente documento, las secuencias de oligonucleótidos pueden diseñarse para incluir: secuencia de cebador de amplificación, sitio de reconocimiento para una enzima de restricción, tal como una enzima de restricción de tipo IIS, relleno, carga útil, relleno, complemento inverso del sitio de reconocimiento para una enzima de restricción (igual o diferente), complemento inverso de una secuencia de cebador de amplificación diferente. La carga útil puede ser un subconjunto solapante del gen objetivo (o cualquier secuencia de ácido nucleico arbitraria). La carga útil puede rellenarse, si se desea, con  $m$  nucleótidos  $M$  ( $M_m$ ) para permitir la generación de extremos cohesivos únicamente complementarios después de la escisión con la enzima o enzimas de restricción. Los cebadores permiten la amplificación. Los sitios de reconocimiento para la enzima o enzimas de restricción permiten que los cebadores se escindan de la carga útil.

En determinada divulgación del presente documento, es ventajoso usar el mismo sitio de reconocimiento en secuencias objetivo múltiples. Sin embargo, debe tenerse en cuenta que, si una secuencia objetivo ya contiene el sitio de reconocimiento, entonces se cortará el oligonucleótido que contiene ese sitio de reconocimiento (en un análisis de izquierda a derecha o de derecha a izquierda), evitando el ensamblaje correcto. En alguna divulgación del presente documento, si la secuencia objetivo solo contiene una única aparición del sitio de reconocimiento, el problema puede resolverse iniciando el análisis dentro del sitio y analizando un conjunto de oligonucleótidos a la izquierda y el otro conjunto a la derecha del sitio de reconocimiento. Puesto que el sitio se dividirá entre 2 oligonucleótidos, no existirá como una secuencia intacta y, por tanto, no se reconocerá ni se cortará. Si hay una longitud o un intervalo de longitudes de oligonucleótido deseados, el último oligonucleótido en cada lado del análisis puede rellenarse con un número apropiado de nucleótidos  $M$  ( $M_m$ ).

Este enfoque puede prolongarse a más de una aparición de un sitio de reconocimiento si esos sitios de restricción aparecen dentro de un múltiplo entero del intervalo de longitud permitido para una carga útil. Como ejemplo del caso más simple (e ignorando cualquier solapamiento deseado para los propósitos de este ejemplo), si alguna porción de los 2 sitios de restricción está separada exactamente 100 pb para un tamaño de carga útil de 100 pb deseado, el análisis desde cualquiera de ellos separará automáticamente el otro. Si la carga útil puede variar de 90-110 pb, entonces pueden acomodarse un par de sitios de restricción dentro de este intervalo de distancia. Con este mismo



intervalo de carga útil, también podría dividirse un par a distancias más largas: 180-220 pb, 270-330 pb, etc.

5 Cuando se analiza una secuencia objetivo en oligonucleótidos, la longitud del último oligonucleótido (o el último en cada dirección si se analiza desde el interior) puede caer fuera del intervalo deseado de longitudes de oligonucleótido. El último oligonucleótido puede rellenarse a la longitud deseada. Sin embargo, esto puede ocurrir con el coste de producir pares de bases adicionales que de otra manera no serían útiles, especialmente cuando se ensamblan una gran cantidad de secuencias objetivo. En alguna divulgación del presente documento, una solución a este problema es concatenar cada secuencia objetivo en un solo pseudo-objetivo largo (con secuencias de cebadores opcionales entre las secuencias objetivo reales) y después dividirlo en fragmentos solapantes más pequeños y de la longitud deseada (por ejemplo, por escisión o amplificación mediante PCR). El cálculo de la longitud de un fragmento se presenta a continuación:

$$\text{longitud} = (\text{trozos} * \text{longitud\_de\_oligonucleótido\_máx}) - (\text{uniones} * \text{solapamiento})$$

15 donde uniones = trozos - 1

Por ejemplo:

$$20 \quad \text{longitud } 484 = (\text{trozos } 5 * \text{longitud\_de\_oligonucleótido\_máx } 100) - (\text{uniones } 4 * \text{solapamiento } 4)$$

$$\text{longitud } 504 = (\text{trozos } 5 * \text{longitud\_de\_oligonucleótido\_máx } 104) - (\text{uniones } 4 * \text{solapamiento } 4)$$

25 Si algunas de las secuencias objetivo contienen un sitio de restricción, entonces en algunos casos, el orden en que se concatenan las secuencias objetivo puede elegirse de manera de tener el sitio de restricción en una unión (y dentro del intervalo de longitudes de oligonucleótido deseado). En el caso general, puede añadirse relleno adicional solo al subconjunto de secuencias objetivo que contienen el sitio de restricción, obteniendo aun así el beneficio completo de eliminar el relleno en la mayoría de las secuencias objetivo.

30 Los ejemplos de la presente divulgación muestran que determinadas enzimas ligasa en determinadas condiciones distinguen correctamente 2 oligonucleótidos con salientes que tengan la misma última base y bases diferentes de la segunda a la última. En alguna divulgación del presente documento, puede ser deseable diseñar los oligonucleótidos de manera que la última base en cada saliente sea única. A, C, G, T únicos en el extremo (4 uniones) permiten la ligadura de hasta 5 trozos, que es un número comercialmente útil para ensamblar. También se contemplan números mayores de trozos de ligadura en la presente divulgación, como se ejemplifica a continuación:

- 35 - últimas 2 bases únicas:  $4^2 = 16$  uniones, hasta 17 trozos
- últimas 3 bases únicas:  $4^3 = 64$  uniones, hasta 65 trozos
- últimas 4 bases únicas:  $4^4 = 256$  uniones, hasta 257 trozos

40 Aspectos de la divulgación se relacionan con algoritmos para analizar la secuencia de ácido nucleico objetivo de entrada. En alguna divulgación del presente documento, pueden usarse algoritmos para garantizar que la última base (o las últimas 2, 3 o 4 bases) de la pluralidad de oligonucleótidos sea única. Por ejemplo, pueden usarse algoritmos de la divulgación para definir una pluralidad de oligonucleótidos analizados que juntos comprenden la secuencia objetivo (de origen natural, no natural o cualquier secuencia de ácido nucleico arbitraria, teniendo los oligonucleótidos aproximadamente la misma longitud y con un solapamiento de 4 bases siendo la última base (o las últimas 2, 3 o 4 bases) única. En alguna divulgación más del presente documento, los oligonucleótidos pueden definirse de manera que del segundo al último o del tercero al último, etc. o combinaciones de los mismos sean únicos.

50 En alguna divulgación del presente documento, un primer algoritmo comprende las siguientes etapas de diseño o descomposición:

- 55 Etapa 1: es moverse en la cantidad objetivo, por ejemplo, 100 pb,
- Etapa 2: almacenar las 1-4 bases pertinentes en un conjunto (por ejemplo, en una memoria),
- Etapa 3: retroceder en el solapamiento (4 pb),
- Etapa 4: moverse nuevamente. Para este segundo movimiento y cada movimiento posterior de 100 pb, si ya existen las 1-4 bases pertinentes en el conjunto, entonces desplazarse 1 base a la vez hasta encontrar una secuencia de 1-4 bases que aún no esté en el conjunto.
- Etapa 5: añadir la nueva secuencia de 1-4 bases al conjunto,
- 60 Etapa 6: después repetir. Si se alcanza el número deseado de trozos antes de alcanzar al final de la secuencia de ADN, entonces comenzar de nuevo con un nuevo conjunto, retroceder en un solapamiento apropiado para el ensamblaje de fragmentos (que puede o no ser un método diferente al ensamblaje de oligonucleótidos en un fragmento).

65 Un experto en la materia observará que el desplazamiento de 1 base podría variar en dirección, por ejemplo, siempre a la izquierda (más corto) si la longitud nominal es la longitud máxima deseada, siempre a la derecha (más

largo) si la longitud nominal es la longitud mínima deseada, o alguna combinación de los mismos. Para centrarse alrededor de la longitud nominal, el desplazamiento podría alternar, por ejemplo, posiciones de verificación en el siguiente orden: -1, +1, -2, +2, etc. El desplazamiento también podría ponderarse para preferir, por ejemplo, que sea más corto, pero permitir que sea más largo, por ejemplo, -1, -2, +1, -3, -4, +2, etc.

5 Este algoritmo puede limitarse al diseño de determinadas secuencias objetivo, ya que el desplazamiento requerido puede ser grande puesto que los grados de libertad se reducen con cada adición posterior al conjunto. Por ejemplo, el primer extremo puede ser una "A", pero el último extremo puede no tener una "A" en varias bases, lo que hace que el último oligonucleótido sea muy corto o muy largo, lo que puede ser indeseable. Una solución a este problema es almacenar una matriz de datos para cada unión, después elegir el menor número de oligonucleótidos para desplazar o la menor distancia de desplazamiento total entre todos los oligonucleótidos, o alguna combinación de los mismos.

15 Las estadísticas de la frecuencia con la que aparecerá cualquier secuencia corta dada (por ejemplo, para un sitio de restricción) en una secuencia aleatoria de 1.000 pb son las siguientes. Por ejemplo, si se usa un sitio de restricción de 6 pb que no se analiza a partir de la mitad de una secuencia objetivo, el 22 % de las secuencias no se podría construir con ese sitio de restricción. Con el mismo sitio de 6 pb y el análisis a partir de la mitad, solo el 3 % de las secuencias que contienen 2 sitios no pudo construirse (o requeriría un análisis adicional). Más en particular:

20 - Si la aparición una sola vez de un sitio de restricción impidió la construcción:

Con la cantidad 1 de longitud 5 pb, el 62 % tendrá al menos 1 sitio Con la cantidad 1 de longitud 6 pb, el 22 % tendrá al menos 1 sitio Con la cantidad 1 de longitud 7 pb, el 6 % tendrá al menos 1 sitio

25 - Si el análisis desde el interior permite 2 apariciones:

Con la cantidad 1 de longitud 5 pb, el 25 % tendrá al menos 2 sitios  
 Con la cantidad 1 de longitud 6 pb, el 3 % tendrá al menos 2 sitios  
 Con la cantidad 1 de longitud 7 pb, <1 % tendrá al menos 2 sitios (aproximadamente el 0,2 %)

30 - Si se usa más de una enzima de restricción (y el sitio correspondiente) y si se permite una sola aparición:

Con la cantidad 2 de longitud 5 pb, el 38 % tendrá al menos 1 sitio  
 Con la cantidad 2 de longitud 6 pb, el 5 % tendrá al menos 1 sitio  
 Con la longitud de 7 pb y la longitud 6 pb, el 1 % tendrá al menos 1 sitio  
 Con la cantidad 3 de longitud 5 pb, el 24 % tendrá al menos 1 sitio  
 Con la cantidad 3 de longitud 6 pb, el 1 % tendrá al menos 1 sitio

40 - Si hay más de una enzima de restricción, permitiendo 2 apariciones:

Con la cantidad 2 de longitud 5 pb, el 6 % tendrá al menos 2 sitios  
 Con la cantidad 2 de longitud 6 pb, <1 % tendrá al menos 2 sitios (aproximadamente el 0,06 %)  
 Con la cantidad 3 de longitud 5 pb, el 2 % tendrá al menos 2 sitios.

#### 45 Aplicaciones

La divulgación del presente documento puede ser útil para una gama de aplicaciones que implican la producción y/o el uso de ácidos nucleicos sintéticos. Como se describe en el presente documento, la divulgación proporciona métodos para ensamblar ácidos nucleicos sintéticos con mayor eficiencia. Los ácidos nucleicos ensamblados resultantes pueden amplificarse *in vitro* (por ejemplo, usando PCR, LCR o cualquier técnica de amplificación adecuada), amplificarse *in vivo* (por ejemplo, a través de clonación en un vector adecuado), aislarse y/o purificarse. Un ácido nucleico ensamblado (solo o clonado en un vector) puede transformarse en una célula hospedadora (por ejemplo, una célula hospedadora procariota, eucariota, de insecto, de mamífero u otra). En alguna divulgación del presente documento, la célula hospedadora puede usarse para propagar el ácido nucleico. En determinada divulgación del presente documento, el ácido nucleico puede integrarse en el genoma de la célula hospedadora. En alguna divulgación del presente documento, el ácido nucleico puede reemplazar una región de ácido nucleico correspondiente en el genoma de la célula (por ejemplo, mediante recombinación homóloga). En consecuencia, los ácidos nucleicos pueden usarse para producir organismos recombinantes. En alguna divulgación del presente documento, un ácido nucleico objetivo puede ser un genoma completo o fragmentos grandes de un genoma que se usan para reemplazar todo o parte del genoma de un organismo hospedador. Los organismos recombinantes también pueden usarse para diversas aplicaciones de investigación, industriales, agrícolas y/o médicas.

65 Muchas de las técnicas que se describen en el presente documento pueden usarse juntas, aplicando técnicas de ensamblaje adecuadas en uno o más puntos para producir moléculas de ácido nucleico largas. Por ejemplo, puede usarse ensamblaje basado en ligasa para ensamblar dúplex de oligonucleótidos y fragmentos de ácido nucleico de menos de 100 a más de 10.000 pares de bases de longitud (por ejemplo, de 100 a 500 pares, de 500 a 1.000

meros, de 1.000 a 5.000 meros, de 5.000 meros a 10.000 meros, 25.000 meros, 50.000 meros, 75.000 meros, 100.000 meros, etc.). En una realización de ejemplo, los métodos que se describen en el presente documento pueden usarse durante el ensamblaje de un genoma completo (o un fragmento grande del mismo, por ejemplo, aproximadamente el 10 %, el 20 %, el 30 %, el 40 %, el 50 %, el 60 %, el 70 %, el 80 %, el 90 % o más) de un organismo (por ejemplo, vírico, bacteriano, levadura u otro organismo procariótico o eucariótico), que opcionalmente incorporan modificaciones específicas en la secuencia en una o más ubicaciones deseadas.

Cualquiera de los productos de ácido nucleico (por ejemplo, incluyendo los ácidos nucleicos que se amplifican, se clonan, se purifican, se aíslan, etc.) puede envasarse en cualquier formato adecuado (por ejemplo, en un tampón estable, liofilizado, etc.) para su almacenamiento y/o envío (por ejemplo, para su envío a un centro de distribución o a un cliente). De forma similar, cualquiera de las células hospedadoras (por ejemplo, células transformadas con un vector o que tienen un genoma modificado) puede prepararse en un tampón adecuado para su almacenamiento y/o transporte (por ejemplo, para su distribución a un cliente). En alguna divulgación del presente documento, las células pueden estar congeladas. Sin embargo, también pueden usarse otras preparaciones celulares estables.

Las células hospedadoras pueden crecer y expandirse en cultivo. Las células hospedadoras pueden usarse para expresar uno o más ARN o polipéptidos de interés (por ejemplo, proteínas terapéuticas, industriales, agrícolas y/o médicas). Los polipéptidos expresados pueden ser polipéptidos naturales o polipéptidos no naturales. Los polipéptidos pueden aislarse o purificarse para su uso posterior.

En consecuencia, las moléculas de ácido nucleico generadas usando métodos de la divulgación pueden incorporarse en un vector. El vector puede ser un vector de clonación o un vector de expresión. En alguna divulgación del presente documento, el vector puede ser un vector vírico. Un vector vírico puede comprender secuencias de ácido nucleico capaces de infectar células diana. De forma similar, en alguna divulgación del presente documento, puede usarse un vector de expresión procariótico unido operativamente a un sistema promotor apropiado para transformar células diana. En otra divulgación del presente documento, un vector eucariótico unido operativamente a un sistema promotor apropiado puede usarse para transfectar células o tejidos diana.

La transcripción y/o traducción de las construcciones que se describen en el presente documento puede realizarse *in vitro* (es decir, usando sistemas sin células) o *in vivo* (es decir, expresadas en células). En alguna divulgación del presente documento, pueden prepararse lisados celulares. En determinada divulgación del presente documento, pueden aislarse o purificarse polipéptidos o ARN expresados. Los ácidos nucleicos de la divulgación también pueden usarse para añadir marcadores de detección y/o purificación a polipéptidos expresados o fragmentos de los mismos. Los ejemplos de fusión basada en polipéptidos/marcador incluyen, pero sin limitación, hexa-histidina (His<sup>6</sup>) Myc y HA y otros polipéptidos con utilidad, tales como GFP GST, MBP, quitina y similares. En alguna divulgación del presente documento, los polipéptidos pueden comprender uno o más restos de aminoácidos no naturales.

En alguna divulgación del presente documento, pueden prepararse anticuerpos contra polipéptidos o uno o más fragmentos de los mismos codificados por uno o más ácidos nucleicos sintéticos. En determinada divulgación del presente documento, los ácidos nucleicos sintéticos pueden proporcionarse como bibliotecas para la detección en investigación y desarrollo (por ejemplo, para identificar proteínas o péptidos terapéuticos potenciales, para identificar posibles dianas proteicas para el desarrollo de fármacos, etc.) En alguna divulgación del presente documento, puede usarse un ácido núcleo sintético como agente terapéutico (por ejemplo, para terapia génica o para regulación génica). Por ejemplo, un ácido nucleico sintético puede administrarse a un paciente en una cantidad suficiente para expresar una cantidad terapéutica de una proteína. En otra divulgación del presente documento, un ácido nucleico sintético puede administrarse a un paciente en una cantidad suficiente para regular (por ejemplo, regular negativamente) la expresión de un gen.

Debe apreciarse que los diferentes actos o divulgaciones del presente documento que se describen en el presente documento pueden realizarse independientemente y pueden realizarse en diferentes emplazamientos en los Estados Unidos o fuera de los Estados Unidos. Por ejemplo, cada uno de los actos de recibir un orden para un ácido nucleico objetivo, analizar una secuencia de ácido nucleico objetivo, diseñar uno o más ácidos nucleicos de partida (por ejemplo, oligonucleótidos), sintetizando uno o más ácidos nucleicos de partida, purificar uno o más ácidos nucleicos de partida, ensamblar uno o más ácidos nucleicos de partida, aislar uno o más ácidos nucleicos ensamblados, confirmar la secuencia de uno o más ácidos nucleicos ensamblados, manipular uno o más ácidos nucleicos ensamblados (por ejemplo, amplificar, clonar, insertar en un genoma hospedador, etc.) y cualesquier otros actos o parte de estos actos puede realizarse independientemente en un emplazamiento o en sitios diferentes dentro de los Estados Unidos o fuera de los Estados Unidos. En alguna divulgación del presente documento, un procedimiento de ensamblaje puede implicar una combinación de actos que se realizan en un sitio (en los Estados Unidos o fuera de los Estados Unidos) y actos que se realizan en uno o más sitios lejanos (dentro de los Estados Unidos o fuera de los Estados Unidos).

#### Aplicaciones automatizadas

Los aspectos de los métodos y dispositivos que se proporcionan en el presente documento pueden incluir la automatización de uno o más actos que se describen en el presente documento. En alguna divulgación del presente

documento, una o más etapas de una reacción de amplificación y/o ensamblaje pueden automatizarse usando uno o más dispositivos de manipulación de muestras automatizadas (por ejemplo, uno o más dispositivos de manipulación de líquidos o fluidos automatizados). Los dispositivos y procedimientos automatizados pueden usarse para entregar reactivos de reacción, incluyendo uno o más de los siguientes: ácidos nucleicos de partida, tampones, enzimas (por ejemplo, una o más ligasas y/o polimerasas), nucleótidos, sales y cualesquier otros agentes adecuados, tales como agentes estabilizantes. También pueden usarse dispositivos y procedimientos automatizados para controlar las condiciones de reacción. Por ejemplo, puede usarse un termociclador automático para controlar las temperaturas de reacción y cualesquier ciclos de temperatura que se puedan usar. En alguna divulgación del presente documento, un láser de barrido puede automatizarse para proporcionar una o más temperaturas de reacción o ciclos de temperatura adecuados para incubar polinucleótidos. De forma similar, puede automatizarse el análisis posterior de los productos polinucleotídicos ensamblados. Por ejemplo, la secuenciación puede automatizarse usando un dispositivo de secuenciación y protocolos de secuenciación automatizados. Las etapas adicionales (por ejemplo, amplificación, clonar, etc.) también pueden automatizarse usando uno o más dispositivos apropiados y protocolos relacionados. Debe apreciarse que uno o más de los dispositivos o componentes de dispositivo que se describen en el presente documento pueden combinarse en un sistema (por ejemplo, un sistema robótico) o en un microentorno (por ejemplo, una cámara de reacción microfluida). Pueden transferirse mezclas de reacción de ensamblaje (por ejemplo, muestras de reacción líquidas) de un componente del sistema a otro usando dispositivos y procedimientos automatizados (por ejemplo, la manipulación y/o transferencia robóticas de muestras y/o recipientes de muestras, incluyendo dispositivos de pipeteo automatizados, micro-sistemas, etc.). El sistema y cualesquier componentes del mismo pueden controlarse mediante un sistema de control.

En consecuencia, las etapas y/o aspectos del método de los dispositivos que se proporcionan en el presente documento pueden automatizarse usando, por ejemplo, un sistema informático (por ejemplo, un sistema controlado por ordenador). Un sistema informático en el que pueden implementarse aspectos de la tecnología que se proporciona en el presente documento puede incluir un ordenador para cualquier tipo de procesamiento (por ejemplo, análisis de secuencias y/o control de dispositivos automatizados como se describe en el presente documento). Sin embargo, debe apreciarse que determinadas etapas de procesamiento pueden ser proporcionadas por uno o más de los dispositivos automatizados que forman parte del sistema de ensamblaje. En alguna divulgación del presente documento, un sistema informático puede incluir dos o más ordenadores. Por ejemplo, un ordenador puede acoplarse, través de una red, a un segundo ordenador. Un ordenador puede realizar análisis de secuencia. El segundo ordenador puede controlar uno o más de los dispositivos de síntesis y ensamblaje automatizados del sistema. En otros aspectos, pueden incluirse ordenadores adicionales en la red para controlar uno o más de los actos de análisis o procesamiento. Cada ordenador puede incluir una memoria y un procesador. Los ordenadores pueden tomar cualquier forma, ya que los aspectos de la tecnología que se proporciona en el presente documento no se limitan a implementarse en ninguna plataforma informática en particular. De forma similar, la red puede tomar cualquier forma, incluyendo una red privada o una red pública (por ejemplo, Internet). Pueden asociarse dispositivos de pantalla a uno o más de los dispositivos y ordenadores. Como alternativa, o además, un dispositivo de pantalla puede ubicarse en un sitio lejano y conectarse para mostrar el resultado de un análisis de acuerdo con la tecnología que se proporciona en el presente documento. Las conexiones entre los diferentes componentes del sistema pueden ser por cable, fibra óptica, transmisión inalámbrica, transmisión satelital, cualquier otra transmisión adecuada o cualquier combinación de dos o más de las anteriores.

Cada uno de los diferentes aspectos, realizaciones o actos de la tecnología que se proporciona en el presente documento puede automatizarse e implementarse independientemente en cualquiera de numerosas maneras. Por ejemplo, cada aspecto, realización o acto puede implementarse independientemente usando hardware, software o una combinación de los mismos. Cuando se implementa en el software, el código del software puede ejecutarse en cualquier procesador o colección de procesadores adecuados, ya sea que se proporcionen en un solo ordenador o se distribuyan entre varios ordenadores. Debe apreciarse que cualquier componente o colección de componentes que realicen las funciones que se han descrito anteriormente pueden considerarse genéricamente como uno o más controladores que controlan las funciones analizadas anteriormente. El uno o más controladores pueden implementarse de varias maneras, tal como con hardware dedicado o con hardware con fines generales (por ejemplo, uno o más procesadores) que se programa usando microcódigo o software para realizar las funciones mencionadas anteriormente.

A este respecto, debe apreciarse que una implementación de la divulgación de la tecnología que se proporciona en el presente documento comprende al menos un medio legible por ordenador (por ejemplo, una memoria de ordenador, un disquete, un disco compacto, una cinta, etc.) codificado con un programa de ordenador (es decir, una pluralidad de instrucciones), que, cuando se ejecuta en un procesador, realiza una o más de las funciones de la tecnología descrita anteriormente en el presente documento. El medio legible por ordenador puede ser transportable de manera que el programa almacenado en él se pueda cargar en cualquier recurso del sistema informático para implementar una o más funciones de la tecnología que se proporciona en el presente documento. Asimismo, debe apreciarse que la referencia a un programa de ordenador que, cuando se ejecuta, realiza las funciones mencionadas anteriormente, no se limita a un programa de aplicación que se ejecuta en un ordenador central. Más bien, la expresión programa de ordenador se usa en el presente documento en un sentido genérico para hacer referencia a cualquier tipo de código informático (por ejemplo, software o microcódigo) que pueda emplearse para programar un procesador para implementar los aspectos analizados anteriormente de la tecnología que se proporciona en el

presente documento.

Debería apreciarse que, de acuerdo con la divulgación de la tecnología que se proporciona en el presente documento, en la que los procesos se almacenan en un medio legible por ordenador, los procesos implementados por ordenador pueden, durante el curso de su ejecución, recibir información manualmente (por ejemplo, de un usuario).

En consecuencia, el control a nivel de sistema general de los dispositivos o componentes de ensamblaje que se describen en el presente documento puede realizarse mediante un controlador del sistema que puede proporcionar señales de control a los sintetizadores de ácido nucleico asociados, dispositivos de manipulación de líquidos, termocicladores, dispositivos de secuenciación, componentes robóticos asociados, así como otros sistemas adecuados para realizar la entrada/salida deseada u otras funciones de control. De este modo, el controlador del sistema junto con los controladores de cualquier dispositivo juntos forman un controlador que controla el funcionamiento de un sistema de ensamblaje de ácidos nucleicos. El controlador puede incluir un sistema de procesamiento de datos con fines generales, que puede ser un ordenador con fines generales o una red de ordenadores con fines generales y otros dispositivos asociados, incluyendo dispositivos de comunicaciones, módems y/u otros circuitos o componentes para realizar la entrada/salida deseada u otras funciones. El controlador también puede implementarse, al menos en parte, como un único circuito integrado con fines especiales (por ejemplo, ASIC) o una matriz de ASIC, teniendo cada uno una sección de procesador central o central para el control general, a nivel de sistema y secciones separadas dedicadas para realizar diversos cálculos, funciones y otros procesos específicos diferentes con el control de la sección del procesador central. El controlador también puede implementarse usando una pluralidad de circuitos o dispositivos electrónicos integrados programables u otros circuitos o dispositivos dedicados separados, por ejemplo, circuitos electrónicos o circuitos lógicos cableados, tales como circuitos de elementos diferenciados o dispositivos lógicos programables. El controlador también puede incluir cualquier otro componente o dispositivo, tales como dispositivos de entrada/salida del usuario (monitores, pantallas, impresoras, un teclado, un dispositivo señalador de usuario, pantalla táctil u otra interfaz de usuario, etc.), dispositivos de almacenamiento de datos, motores de accionamiento, conexiones, controladores de válvulas, dispositivos robóticos, bombas de vacío y otras, sensores de presión, detectores, fuentes de alimentación, fuentes de pulsos, dispositivos de comunicación u otros circuitos o componentes electrónicos, etc. El controlador también puede controlar el funcionamiento de otras partes de un sistema, tales como el procesamiento automatizado de pedidos de clientes, el control de calidad, el envasado, el transporte, la facturación, etc., para realizar otras funciones adecuadas conocidas en la técnica, pero no se desvelan en detalle en el presente documento.

Pueden usarse diversos aspectos de la presente invención solos, en combinación o en diversas disposiciones no analizadas específicamente en las realizaciones descritas anteriormente y, por tanto, no se limitan en su aplicación a los detalles y la disposición de los componentes expuestos en la descripción anterior o ilustrados en los dibujos. Por ejemplo, pueden combinarse aspectos que se describen en una realización de cualquier manera con los aspectos que se describen en otras realizaciones.

El uso de términos ordinales tales como "primero", "segundo", "tercero", etc., en las reivindicaciones para modificar un elemento de una reivindicación no denota por sí mismo ninguna prioridad, precedencia u orden de un elemento reivindicado respecto de otro, o el orden temporal en el que se realizan los actos de un método, sino que se utilizan simplemente como etiquetas para distinguir un determinado elemento que tiene un determinado nombre de otro elemento que tiene el mismo nombre (salvo por el uso del término ordinal) para distinguir los elementos de la reivindicación.

Además, la fraseología y la terminología utilizadas en el presente documento son con fines de descripción y no deben considerarse limitantes. El uso de "que incluye", "que comprende", o "que tiene", "que contiene", "que implica" y variaciones de los mismos en el presente documento, pretenden incluir los elementos que se enumeran a continuación en el presente documento y equivalentes de los mismos así como elementos adicionales.

### Ejemplos

La Figura 1 muestra la secuencia de una secuencia bicatenaria elegida arbitrariamente de aproximadamente 836 pb de longitud. Los fragmentos de 60 pb se seleccionaron y se marcaron de 1 a 28 (los fragmentos 1-14 están en la cadena positiva; los fragmentos 15-28 en la cadena negativa). Estos fragmentos de 60 pb se encargaron a IDT (Integrated DNA Technologies, Coralville, Iowa) ("oligonucleótidos IDT"), con las siguientes secuencias flanqueantes:

```
GTCACTACCGCTATCATGGCGGTC . . . . . GAGACCAGGAGACAGGACCGACCAAA
CAGTGATGGCGATAGTACCGCAGAG . . . . . CTCTGGTCCTCTGTCTGGCTGGTTT
```

Está subrayado el sitio de reconocimiento de Bsal-HF, que produce un saliente de 4 bases:

```
5'...GGTCTC(N)1▼...3'
3'...CCAGAG(N)5▲...5'
```

Los sitios de reconocimiento Bsal-HF están flanqueados por cebadores universales que son útiles para la amplificación de estos fragmentos.

- 5 También se diseñaron cebadores de PCR A-E (flechas discontinuas en la Figura 1) para amplificar el producto de ligadura correcto. La Figura 2 muestra la posición relativa de los cebadores ("oligoA" a "oligoE") como puntas de flecha, así como el tamaño previsto de los productos de PCR correspondientes.

Los oligonucleótidos IDT bicatenarios se sometieron a digestión con Bsal-HF, en las siguientes condiciones:

- 10
- 1X NEBuffer 4
  - Complementado con Albúmina de suero bovino 100 µg/ml
  - Incubar a 37 °C.
- 15 Se purificaron oligonucleótidos bicatenarios digeridos que tenían extremos cohesivos (oligonucleótidos 1-28) por electroforesis en un gel al 4 %. Después se sometieron diversas combinaciones de oligonucleótidos 1-28 purificados a reacciones de ligadura. Se sometieron a ensayo varias ligasas, temperaturas y tiempos de incubación diferentes para determinar las condiciones de ligadura óptimas. Las ligasas sometidas a ensayo incluyen:
- 20 ADN ligasa T4  
ADN ligasa T4 + sal 300 mM (para menor actividad, mayor especificidad)  
ADN ligasa T3  
ADN ligasa T7  
ADN ligasa Pfu
- 25 ADN ligasa Taq  
ADN ligasa E. coli

Se muestran resultados de ejemplo realizados a temperatura ambiente durante 30 minutos en las Figuras 3-5. La Figura 3 muestra los resultados de electroforesis de la ligadura por pares (de dos oligonucleótidos), de izquierda a derecha del gel: escalera, sin ligasa, ADN ligasa T4, ADN ligasa T4 + sal, ADN ligasa T3, ADN ligasa T7. Las bandas de la parte inferior a la parte superior del gel corresponden a: oligonucleótidos libres, producto ligado correcto, un producto ligado y medio, dímero del producto ligado. La ADN ligasa T7 produjo el producto ligado más correcto y, por tanto, apareció como la más eficiente en estas condiciones experimentales, manteniéndose igual las otras cosas.

La Figura 4 muestra los resultados de ligadura de los oligonucleótidos 1-10 (calles 1-6) y los oligonucleótidos 11-14 (calles 7-10), con las diferentes ligasas indicadas en la parte superior del gel. Se observaron múltiples bandas, lo que indica la presencia de diferentes productos de ligadura. Sin embargo, tras la amplificación por PCR usando oligonucleótidos A y B como cebadores, se observó una banda fuerte a aproximadamente 300 pb. Debido a que el producto de PCR previsto de los oligonucleótidos A y B es de 337 pb (véase la Figura 2), esta banda corresponde al producto de ligadura correcto que comprende los oligonucleótidos 1-6 (véase la Figura 1). La banda se cortó del gel, se purificó y se secuenció. Los resultados de la secuenciación se muestran en la Figura 6, lo que confirma el 100 % de fidelidad del producto de ligadura en comparación con la secuencia esperada. La ADN ligasa Taq no produjo ningún producto de ligadura, probablemente debido a la baja temperatura de reacción (temperatura ambiente), ya que la ADN ligasa Taq solo es activa a temperaturas elevadas (45 °C-65 °C).

Se desarrolló un ensayo de desapareamiento por pares para someter a ensayo la especificidad de diversas ligasas. Se diseñó un par de oligonucleótidos con salientes de 4 bases, donde la secuencia de apareamiento perfecto ("P", del inglés *perfect match*) es GGTG y la secuencia de desapareamiento ("M", del inglés *mismatch*) es GCTG, que difiere de la secuencia correcta en un nucleótido. Como se muestra en las Figuras 7A y 7B, pueden observarse dos bandas principales, correspondiendo la banda inferior a oligonucleótidos no ligados (según lo indicado por los controles sin ligasa) y correspondiendo la banda superior al producto ligado. ADN ligasa T4 + sal, ADN ligasa T3, ADN ligasa T7 y ADN ligasa de E. coli produjeron una banda fuerte que corresponde al producto ligado cuando se usan los salientes de apareamiento perfecto. Por el contrario, cuando se usaron salientes de desapareamiento, la mayor parte del producto eran oligonucleótidos no ligados. Estos experimentos muestran que en estas condiciones de reacción, ADN ligasa T4 + sal, ADN ligasa T3, ADN ligasa T7 y ADN ligasa de E. coli demostraron una alta especificidad y una discriminación de desapareamiento de un nucleótido de diferencia.

Además del producto de ligadura que tiene los oligonucleótidos 1-6 mostrados anteriormente, también se produjeron otros productos de ligadura, incluyendo productos más largos. Un producto parecía tener los oligonucleótidos 1-6 ligados al oligonucleótido 14. Esto se debe al hecho de que los oligonucleótidos 7 y 14 tenían el mismo extremo cohesivo (GTTC, recuadros en la Figura 8).

## REIVINDICACIONES

1. Un método de producción de un ácido nucleico objetivo que tiene una secuencia predefinida, comprendiendo el método:

5 proporcionar una pluralidad de fragmentos de ácido nucleico bicatenario de extremos romos que tienen una secuencia de reconocimiento de enzimas de restricción en cada extremo; en donde la pluralidad de fragmentos de ácido nucleico bicatenario de extremos romos se generan a partir de una pluralidad de oligonucleótidos monocatenarios inmovilizados sobre un soporte sólido y se amplifican usando un cebador de amplificación que se une a una secuencia de amplificación ubicada en el extremo 5' y/o el extremo 3', en donde el cebador de amplificación es un cebador universal para todos los oligonucleótidos monocatenarios;

10 producir una pluralidad de fragmentos de ácido nucleico bicatenario de extremos cohesivos, que juntos comprenden la secuencia de ácido nucleico objetivo, a través de la digestión enzimática de la pluralidad de fragmentos de ácido nucleico bicatenario de extremos romos en o cerca de la secuencia de reconocimiento de enzimas de restricción, en donde la pluralidad de fragmentos de ácido nucleico bicatenario de extremos cohesivos tienen cada uno dos salientes diferentes y no complementarios, teniendo los salientes de 3 a 10 bases de longitud;

15 hibridar la pluralidad de fragmentos de ácido nucleico bicatenario de extremos cohesivos en los salientes, en donde un primer saliente de un primer fragmento de ácido nucleico bicatenario de extremos cohesivos se diseña para que comprenda una región que sea únicamente complementaria a una región en un segundo saliente de un segundo fragmento de ácido nucleico bicatenario de extremos cohesivos; y

20 ligar la pluralidad de fragmentos de ácido nucleico bicatenario de extremos cohesivos con una ligasa, formando de este modo el ácido nucleico objetivo que tiene una secuencia predefinida.

25 2. El método de la reivindicación 1, en el que el cebador de amplificación tiene un marcador de afinidad para facilitar la retirada por afinidad de productos de digestión enzimática indeseables.

3. El método de la reivindicación 2, en el que el marcador de afinidad es biotina.

30 4. El método de una cualquiera de las reivindicaciones 1-3, en el que en la etapa de proporcionar, la pluralidad de ácidos nucleicos bicatenarios de extremos romos comprenden al menos 3, 4, 5, 6, 7, 8, 10, 15 o 20 fragmentos de ácido nucleico bicatenario de extremos romos diferentes y en donde, preferentemente, cada uno de la pluralidad de fragmentos de ácido nucleico bicatenario de extremos romos tiene al menos 50, 100, 200 o 300 bases de longitud.

35 5. El método de una cualquiera de las reivindicaciones 1-4, en el que en la etapa de proporcionar, la pluralidad de fragmentos de ácido nucleico bicatenario de extremos romos comprende la misma secuencia de reconocimiento de enzimas de restricción o en el que la pluralidad de fragmentos de ácido nucleico bicatenario de extremos romos comprenden al menos dos secuencias de reconocimiento de enzimas de restricción diferentes reconocibles por dos enzimas de restricción diferentes que se seleccionan para producir salientes que tengan el mismo número de bases.

40 6. El método de una cualquiera de las reivindicaciones 1-5, en el que en la etapa de proporcionar, la secuencia de reconocimiento de enzimas de restricción es susceptible de ser reconocida por una enzima de restricción de tipo IIs y en donde, preferentemente, la enzima de restricción de tipo IIs se selecciona entre BstF5 I, BtsC I, BsrD I, Bts I, Alw I, Bcc I, BsmA I, Ear I, Ple I, Bmr I, Bsa I, BsmB I, BspQ I, Fau I, Mnl I, Sap I, Bbs I, BciV I, Hph I, Mbo II, BfuA I, BspCN I, BspM I, SfaN I, Hga I, BseR I, Bbv I, Eci I, Fok I, BceA I, BsmF I, BtgZ I, BpuE I, Bsg I, Mme I, BseG I, Bse3D I, BseM I, AclW I, Alw26 I, Bst6 I, BstMA I, EamI 104 I, Ksp632 I, Pps I, Bfi I, Bso31 I, BspTN I, Eco31 I, Esp3 I, Smu I, Bfu I, Bpi I, BpuA I, BstV2 I, AsuHP I, Acc36 I, Lwe I, Aar I, BseM II, TspDT I, TspGW I, BseX I, BstVI I, Eco571, Eco57M I, Gsu I y Bcg I, cualquier variante de los mismos o cualquier combinación de los mismos.

50 7. El método de una cualquiera de las reivindicaciones 1-6, en el que en la etapa de producción, la pluralidad de fragmentos de ácido nucleico bicatenario de extremos cohesivos se diseñan de manera que un extremo cohesivo seleccionado en un fragmento de ácido nucleico bicatenario de extremos cohesivos seleccionado es únicamente complementario al siguiente extremo cohesivo en un fragmento de ácido nucleico bicatenario de extremos cohesivos adyacente.

55 8. El método de una cualquiera de las reivindicaciones 1-7, en el que en la etapa de producción, los salientes son salientes 5', salientes 3' o salientes 5' y 3',

60 en donde, preferentemente, los salientes tienen al menos 3, 4, 5, 6, 7 u 8 bases de longitud, y en donde, preferentemente, los salientes difieren entre sí en al menos 1, 2, 3 o 4 bases.

9. El método de una cualquiera de las reivindicaciones 1-8, que comprende adicionalmente, antes de la etapa de hibridación, purificar la pluralidad de fragmentos de ácido nucleico bicatenario de extremos cohesivos para retirar productos de digestión enzimática indeseables y en el que preferentemente, los productos de digestión enzimática indeseables incluyen fragmentos de menos de aproximadamente 40, aproximadamente 35, aproximadamente 30, aproximadamente 25, aproximadamente 20 o aproximadamente 15 bases de longitud.

10. El método de una cualquiera de las reivindicaciones 1-9, en el que en la etapa de ligadura, la ligasa es la ADN ligasa T3, ADN ligasa T4, ADN ligasa T7, ADN ligasa de *E. coli*, cualquier variante de los mismos o cualquier combinación de los mismos.
- 5
11. El método de una cualquiera de las reivindicaciones 1-10, que comprende adicionalmente someter la pluralidad hibridada de fragmentos de ácido nucleico bicatenario de extremos cohesivos a una reacción de prolongación de cadena antes o simultáneamente durante la etapa de ligadura.
- 10
12. El método de una cualquiera de las reivindicaciones 1-11, que comprende adicionalmente amplificar el ácido nucleico objetivo usando un par de cebadores específicos para el ácido nucleico objetivo y una polimerasa.
13. El método de una cualquiera de las reivindicaciones 1-12, que comprende adicionalmente confirmar la secuencia del ácido nucleico objetivo.
- 15
14. El método de una cualquiera de las reivindicaciones 1-13, en el que el ácido nucleico objetivo es lineal o se incorpora en un vector.
- 20
15. El método de una cualquiera de las reivindicaciones 1-14 en el que en la etapa de ligadura, la pluralidad de fragmentos de ácido nucleico se ligan en un único grupo o en el que la pluralidad de fragmentos de ácido nucleico están en al menos dos grupos, teniendo cada fragmento de ácido nucleico del primer grupo un extremo terminal complementario a un fragmento de ácido nucleico del segundo grupo.



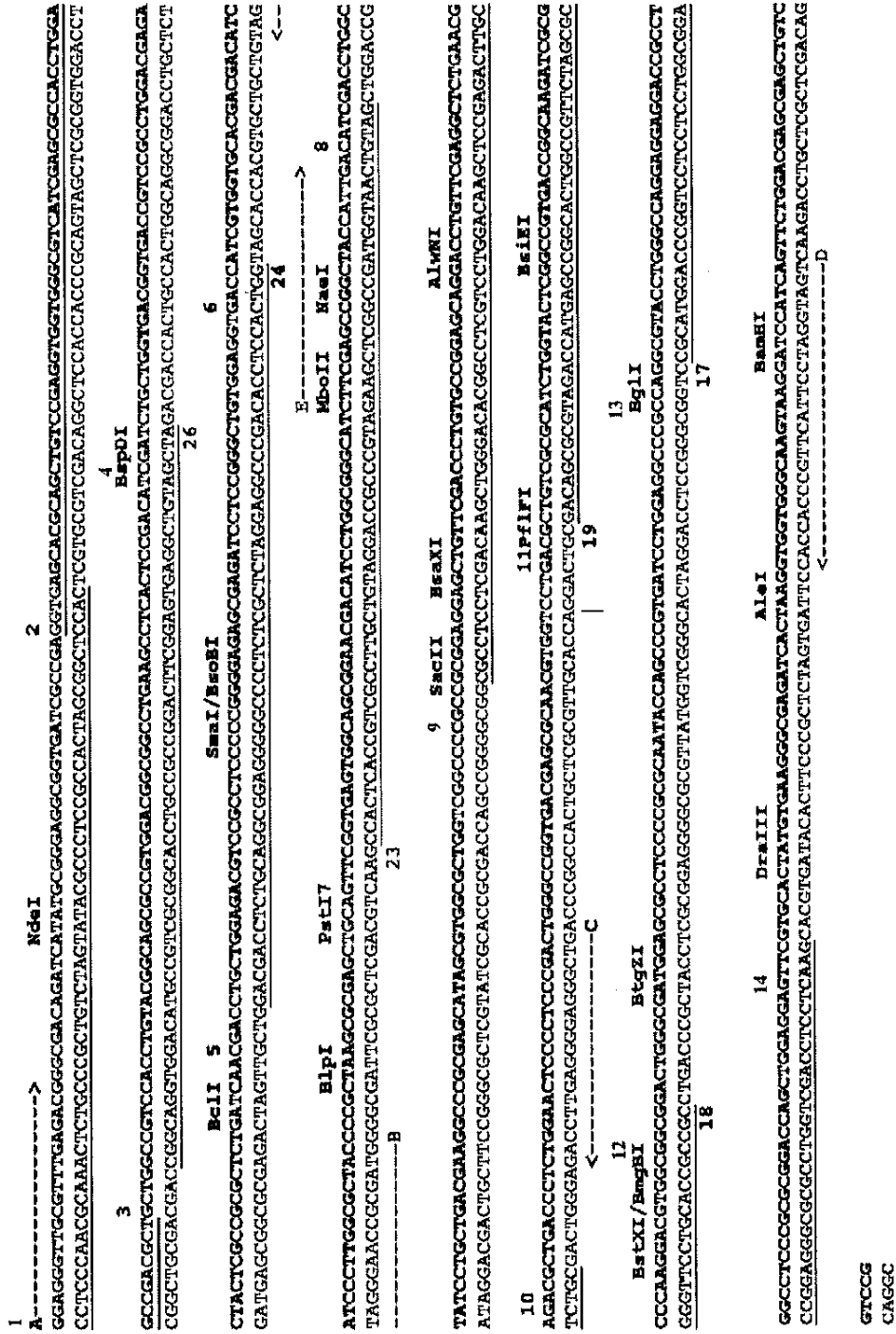


Figura 1

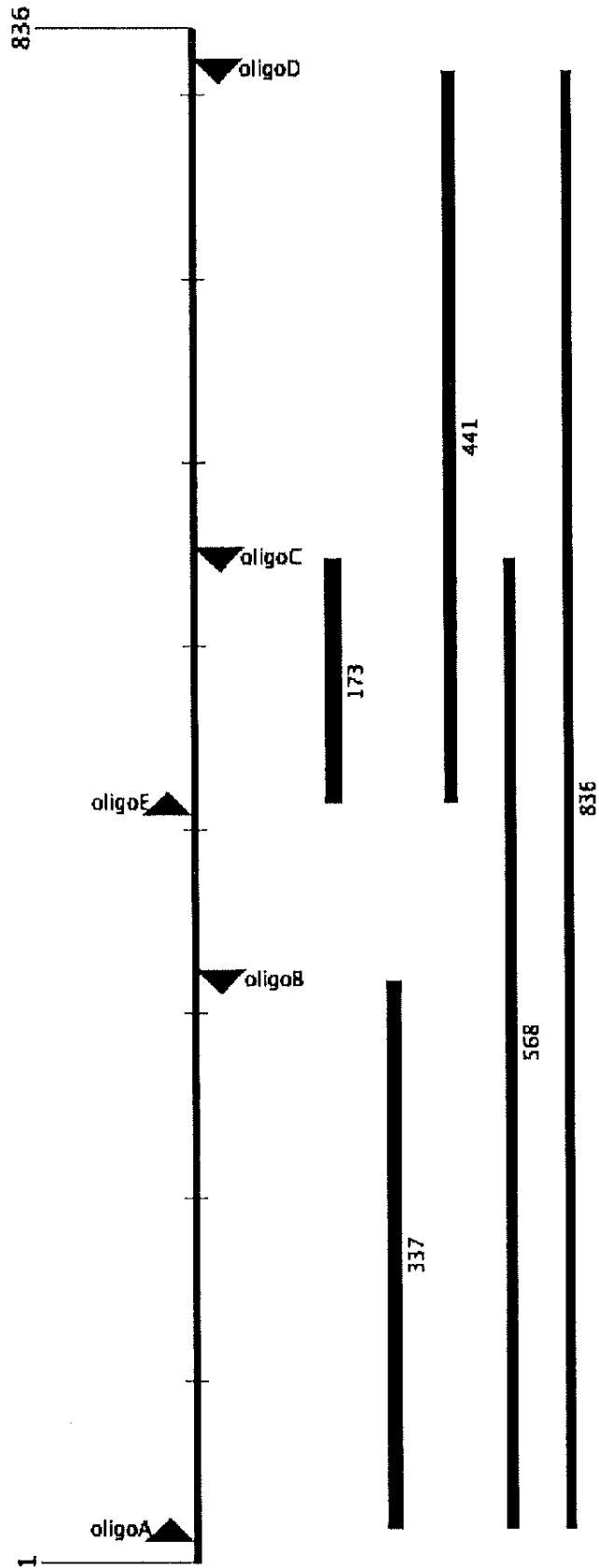


Figura 2

### Ligadura por pares

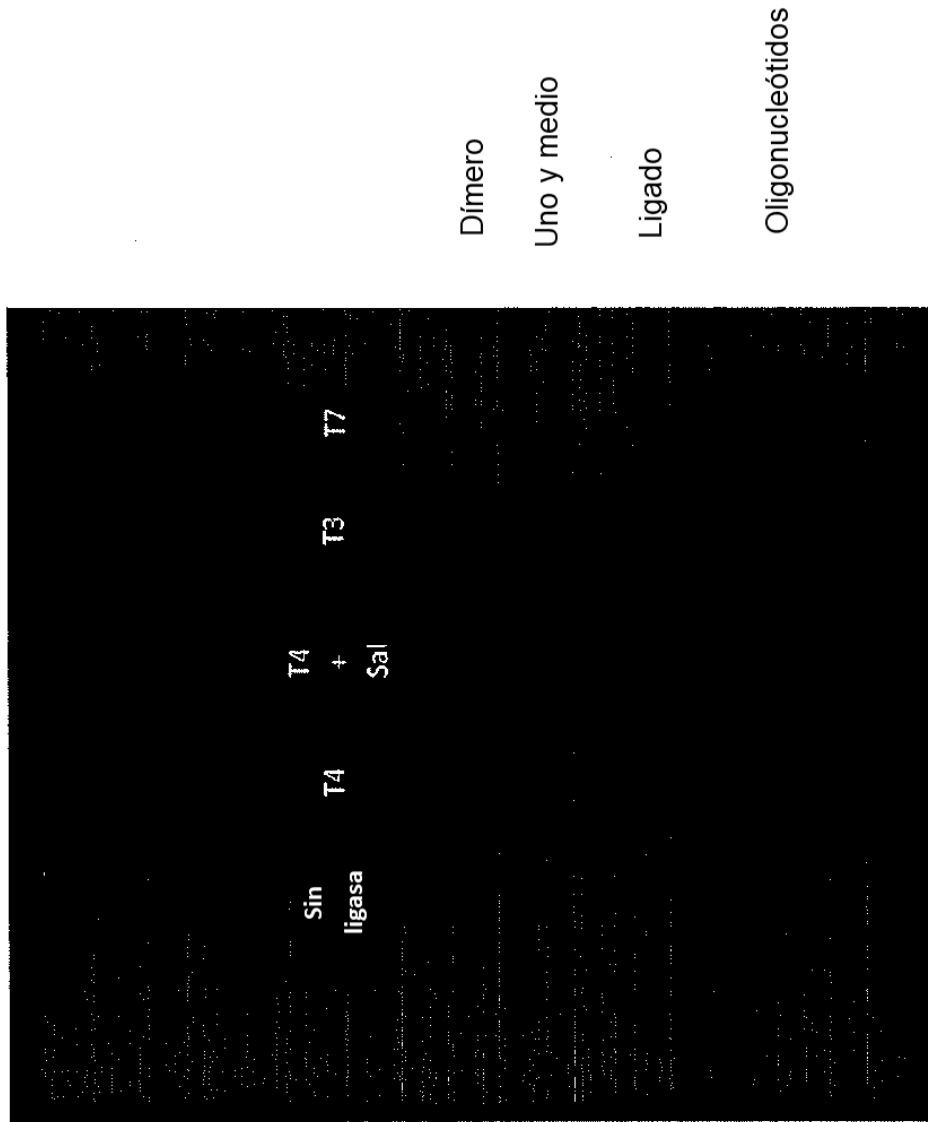
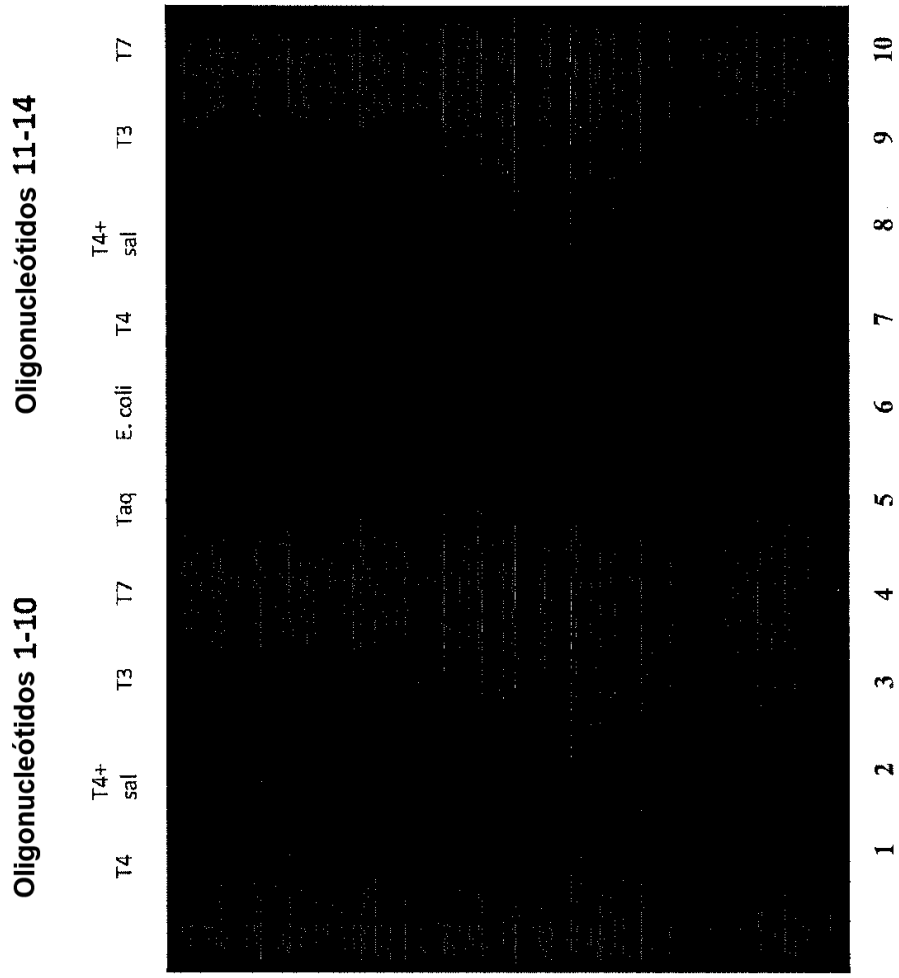


Figura 3



**Figura 4**

PCR de oligonucleótidos 1-10 ligados con oligonucleótidos A y B

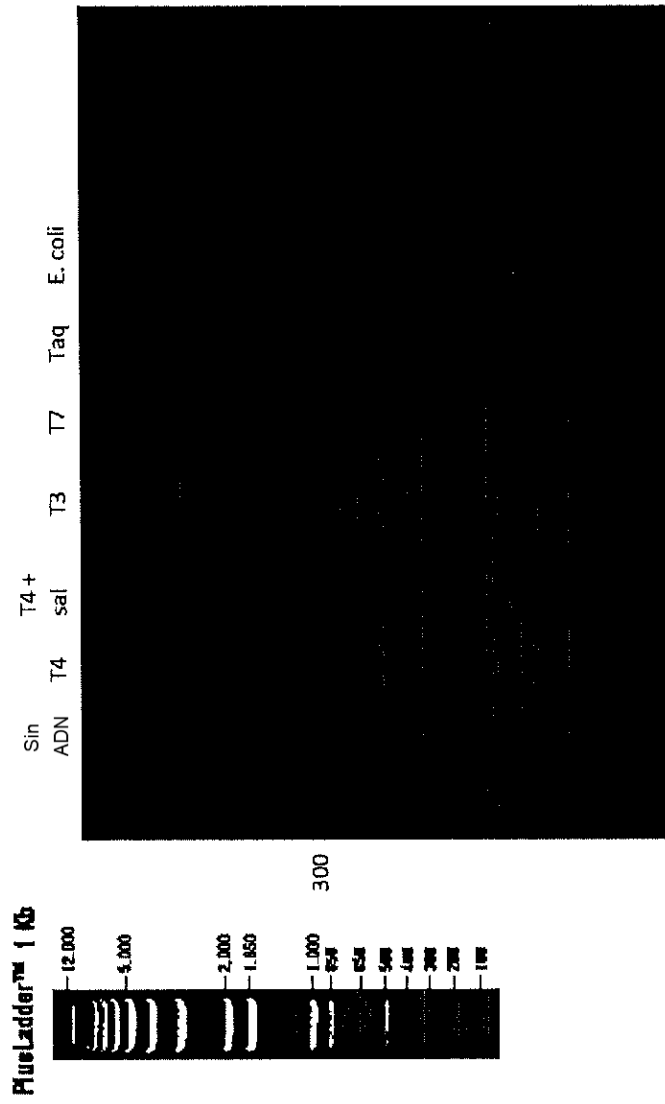


Figura 5

**Secuencia de producto de ligadura en comparación con el esperado**

```

>IcII42826 2
Longitud=550

Puntuación = 623 bits (337), Esperado = 0,0
Identidades = 337/337 (100%), Huecos = 0/337 (0%)
Cadena = Más/Más

Consulta 1  GGAGGGTTGCGTTTGAGACGGGGACAGATCATATATGCGGGAGGGCGGTGATCGCCGAGGTG 60
            |||
Sujeto 102  GGAGGGTTGCGTTTGAGACGGGGACAGATCATATATGCGGGAGGGCGGTGATCGCCGAGGTG 161

Consulta 61  AGCACGCAGCTGTCCGAGGTGGTGGGCGTCAFCGAGCGCCACCTGGAGCCGACGCTGCTG 120
            |||
Sujeto 162  AGCACGCAGCTGTCCGAGGTGGTGGGCGTCAFCGAGCGCCACCTGGAGCCGACGCTGCTG 221

Consulta 121 GCGTCCACCTGTACGGCAGCGCGTGGACGGCGCCTGAAGCCTCACTCCGACATCGAT 180
            |||
Sujeto 222  GCGTCCACCTGTACGGCAGCGCGTGGACGGCGCCTGAAGCCTCACTCCGACATCGAT 281

Consulta 181 CTGCTGGTGACGGTGACCGTCCGCTGGACGAGACTACTCGCCGCGCTCTGATCAACGAC 240
            |||
Sujeto 282  CTGCTGGTGACGGTGACCGTCCGCTGGACGAGACTACTCGCCGCGCTCTGATCAACGAC 341

Consulta 241 CTGCTGGAGACGTCGCGCTCCCGGGGAGAGCGAGATCCTCCGGGCTGTGGAGGTGACC 300
            |||
Sujeto 342  CTGCTGGAGACGTCGCGCTCCCGGGGAGAGCGAGATCCTCCGGGCTGTGGAGGTGACC 401

Consulta 301 ATCGTGGTGCACGACGACATCATCCCTTGGCGGTACC 337
            |||
Sujeto 402  ATCGTGGTGCACGACGACATCATCCCTTGGCGGTACC 438
    
```

**Figura 6**

Ensayo de ligadura de desapareamiento por pares

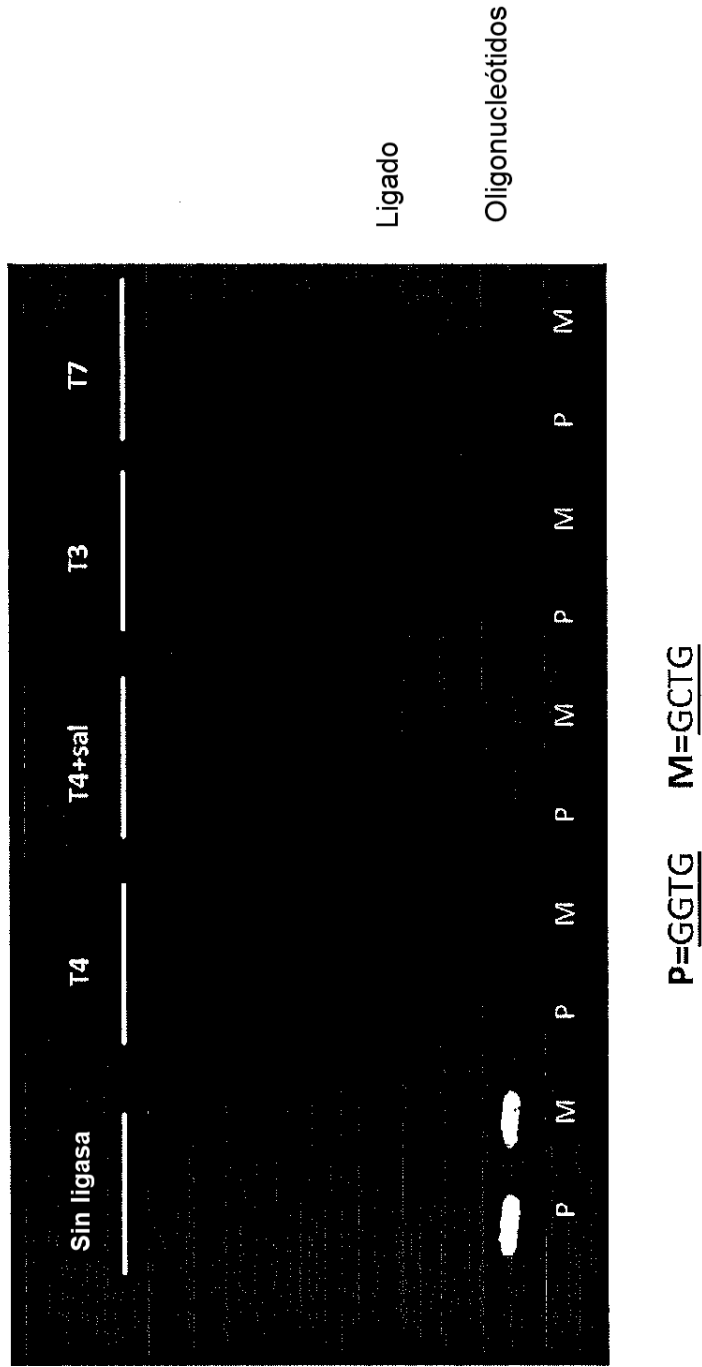


Figura 7A

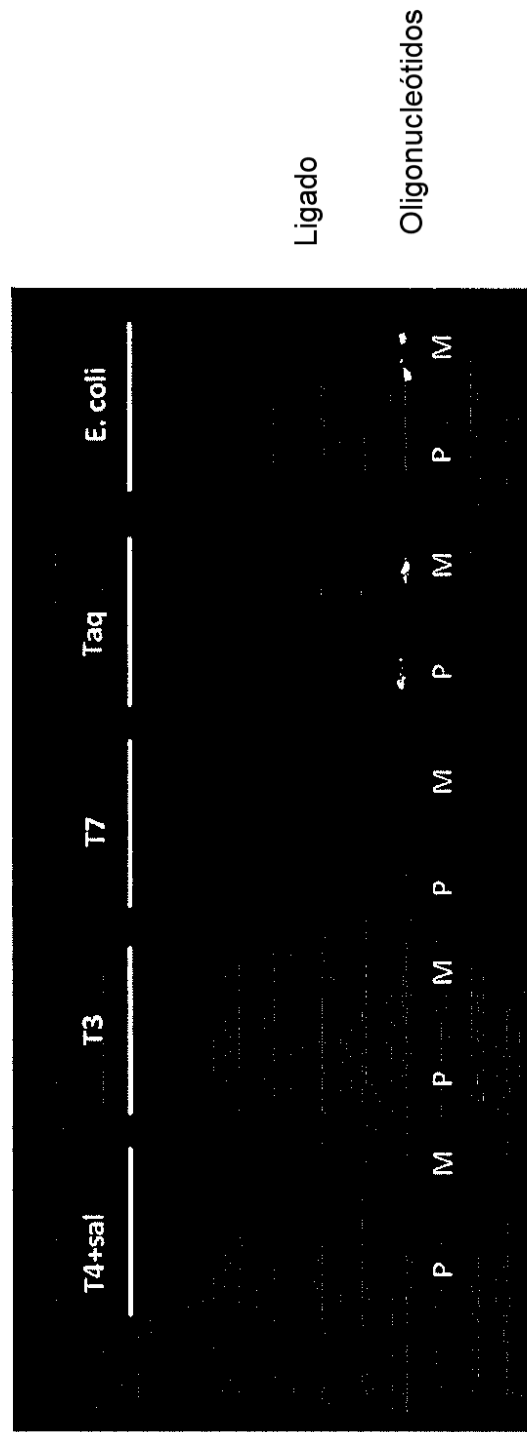


Figura 7B





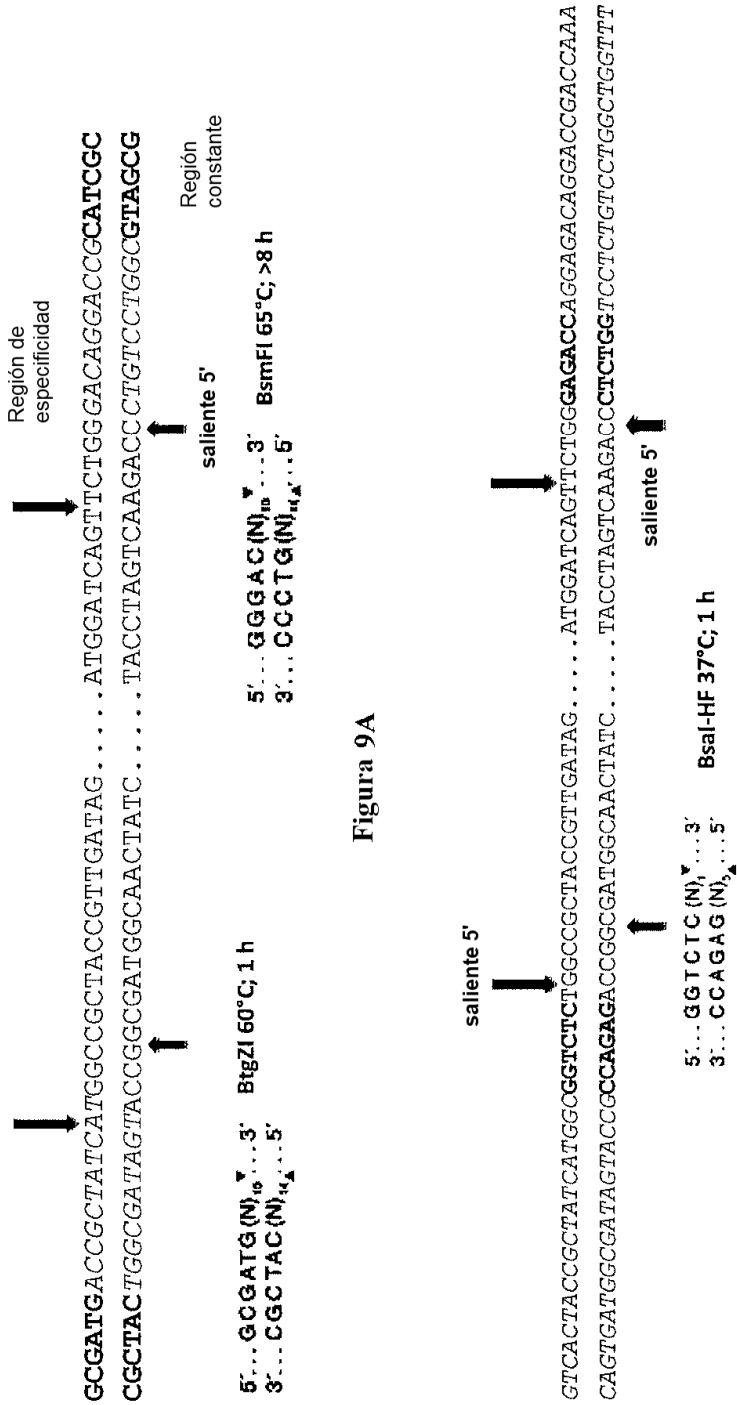


Figura 9A

Figura 9B

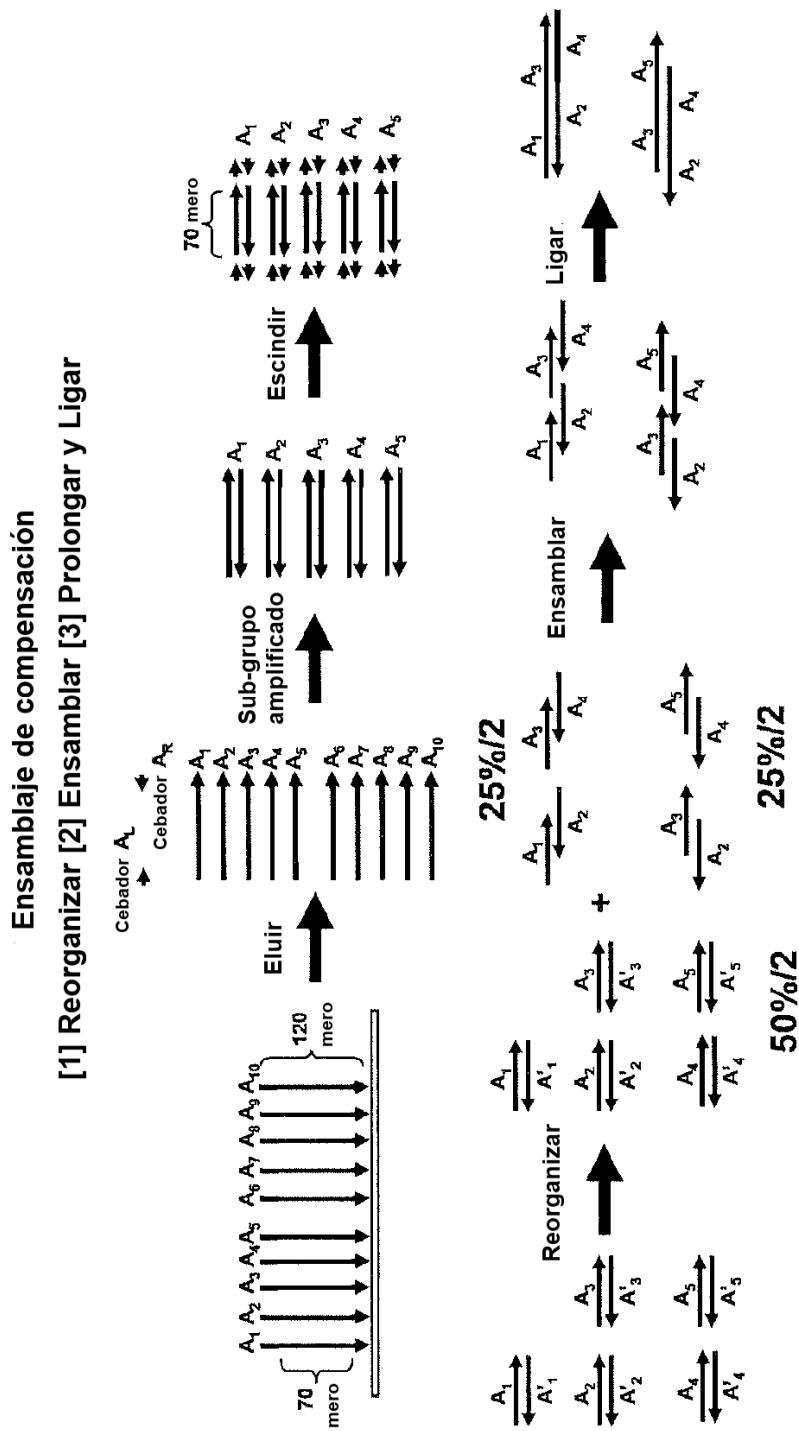


Figura 10A

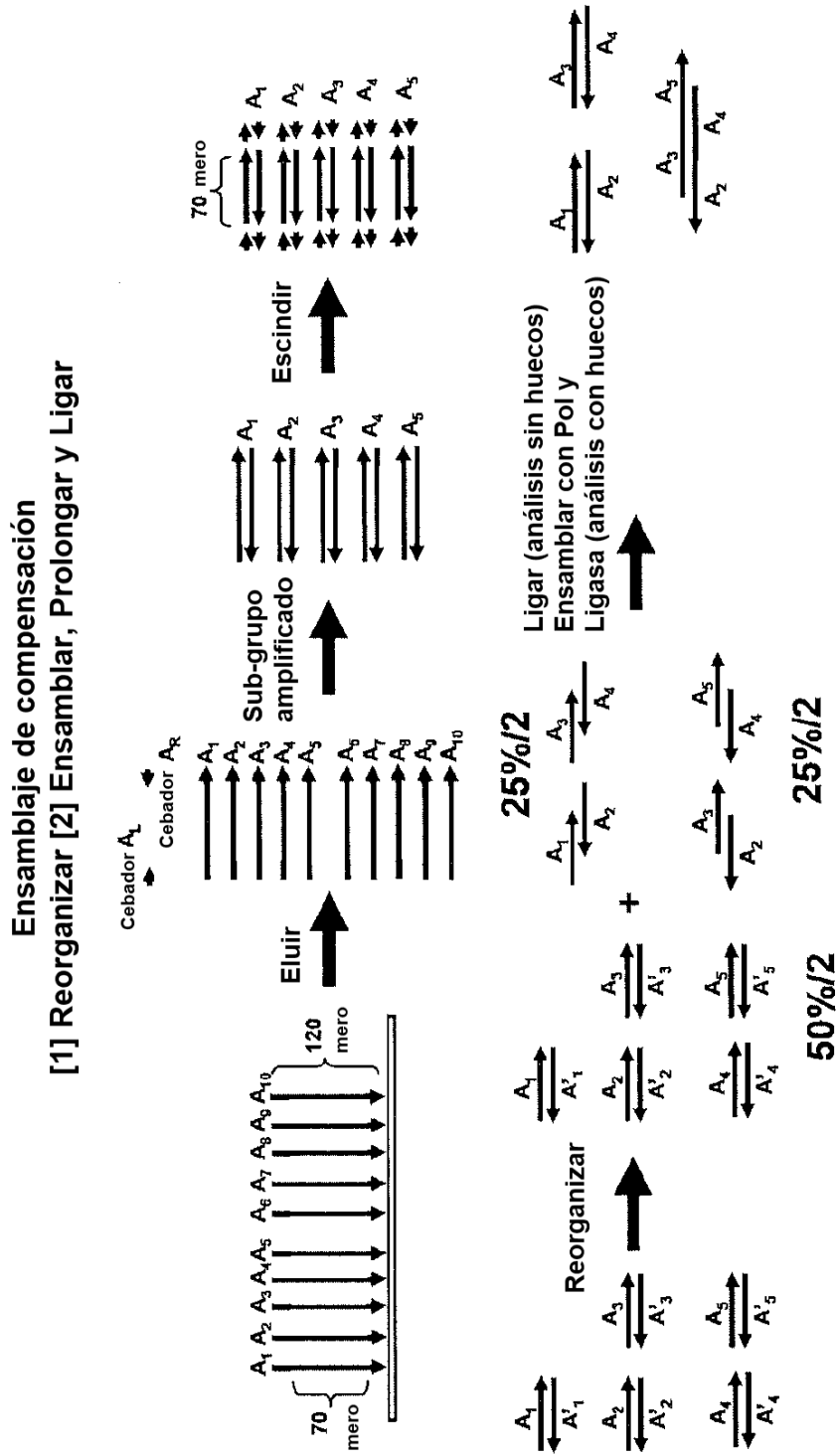


Figura 10B