(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 805 874**

(51) Int. Cl.:

C12N 15/10 (2006.01)
C12Q 1/68 (2008.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **17.01.2012** **E 17185075 (3)**

(97) Fecha y número de publicación de la concesión europea: **22.04.2020** **EP 3263703**

(54) Título: **Ligación enzimática de ácidos nucleicos**

(30) Prioridad:

17.01.2011 US 201161433502 P
17.01.2011 US 201161433488 P
11.04.2011 US 201161474205 P
11.04.2011 US 201161474168 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
15.02.2021

(73) Titular/es:

LIFE TECHNOLOGIES CORPORATION (100.0%)
5823 Newton Drive
Carlsbad, CA 92008, US

(72) Inventor/es:

HENDRICKS, STEPHEN

(74) Agente/Representante:

SÁNCHEZ SILVA, Jesús Eladio

ES 2 805 874 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

Ligación enzimática de ácidos nucleicos

Antecedentes

Las ligasas de ADN pueden unir polinucleótidos juntos, por ejemplo, mediante la catálisis de la formación de un enlace fosfodiéster en roturas monocatenarias o bicatenarias en el ADN dúplex. Las ligasas pueden ser sensibles al grado de hibridación entre las cadenas de ácido nucleico opuestas en un dúplex. Por ejemplo, la ligación exitosa puede ocurrir con menos frecuencia (o nada en absoluto) cuando una cadena a unir a una cadena adyacente está en un dúplex y no es complementaria a su cadena opuesta en el dúplex. En algunos casos, una no coincidencia de un solo nucleótido entre cadenas en un dúplex puede perjudicar o prevenir significativamente la ligación. La capacidad de las ligasas para la discriminación en base a la hibridación, que incluye la discriminación de un solo nucleótido, ha llevado al desarrollo de técnicas de detección mediadas por ligasas (*por ejemplo*, Landegren, U., Bioessays, 15(11):761-765 (1993), y Barany, PNAS USA, 88(1):189-193 (1991)). La amplificación de la señal lineal en base a la ligasa que se conoce como LDR (*es decir,* reacción de detección de ligasa), que se combina con la PCR (*es decir,* la reacción en cadena de la polimerasa) en base a la amplificación objetivo específica de genes, ha demostrado ser una herramienta útil en la detección de mutaciones genéticas de cáncer y enfermedades. Las técnicas de la PCR/LDR típicamente se basan en dos propiedades de una ligasa de ADN: (i) la especificidad y (ii) la termoestabilidad.

Resumen

Esta aplicación se refiere a reactivos y métodos de ligación. Entre otras cosas, se proporcionan métodos y reactivos para ligar ácidos nucleicos a otros ácidos nucleicos, que incluye la ligación de dos polinucleótidos. Cualquiera o ambos polinucleótidos pueden ser monocatenarios o bicatenarios. Uno o ambos polinucleótidos pueden ser un oligonucleótido corto. En algunas modalidades, los oligonucleótidos y/o los polinucleótidos de diferentes longitudes pueden unirse entre sí. La ligación puede ser enzimática, y las ligaciones pueden ser dependientes del molde o independientes del molde. Los ácidos nucleicos que se ligan o involucran en la reacción de ligación pueden marcarse o no marcarse e inmovilizarse o en solución.

Algunas modalidades implican la ligación independiente del molde de polinucleótidos bicatenarios o monocatenarios (por ejemplo, oligonucleótidos). Los polinucleótidos monocatenarios opcionalmente no se hibridan con otro polinucleótido. En algunas modalidades, un oligonucleótido puede hibridarse o asociarse de otra forma con todo o una porción de un saliente u otra región monocatenaria de un ácido nucleico dúplex y ligarse a un extremo libre de una cadena del dúplex u otro oligonucleótido que se hibride o asocie de otra forma con un saliente u otra porción monocatenaria del dúplex.

Algunas modalidades implican la ligación de polinucleótidos monocatenarios (por ejemplo, oligonucleótidos). Opcionalmente, los polinucleótidos monocatenarios se hibridan hibridados adyacentes o cerca uno del otro en otro polinucleótido único. En algunas modalidades, un oligonucleótido puede hibridarse o asociarse de otra forma con todo o una porción de un saliente u otra región monocatenaria de un ácido nucleico dúplex y ligarse a un extremo libre de una cadena del dúplex u otro oligonucleótido que se hibride o asocie de otra forma con un saliente u otra porción monocatenaria del dúplex.

En algunos aspectos, se proporcionan métodos y reactivos para hibridar o asociar de otra forma un primer oligonucleótido y un segundo oligonucleótido a un tercer oligonucleótido o a un polinucleótido tal que los extremos terminales del primer y segundo oligonucleótidos sean adyacentes o cercanos entre sí. Tal hibridación o asociación puede ocurrir de forma secuencial, simultánea o sustancialmente simultánea. El extremo terminal del primer oligonucleótido puede ligarse al extremo terminal adyacente o cercano del segundo oligonucleótido.

Las no coincidencias de bases de nucleótidos entre un primer y/o segundo oligonucleótido y un tercer oligonucleótido o polinucleótido pueden afectar la eficiencia de la ligación. Por ejemplo, las no coincidencias en la posición terminal de uno o ambos oligonucleótidos primero y segundo pueden afectar la eficiencia de la ligación, al reducir la probabilidad de una ligación exitosa o al excluir la ligación por completo. Las no coincidencias en otras posiciones o en múltiples posiciones también pueden afectar la eficiencia de la ligación, al reducir la probabilidad de una ligación exitosa o al excluir la ligación por completo.

También se proporcionan métodos y reactivos para realizar múltiples ligaciones secuencialmente, en paralelo, o tanto secuencialmente como en paralelo.

Opcionalmente, uno o más del cebador, sonda o molde se marcan. Por ejemplo, la sonda puede marcarse. De lo contrario, el cebador o molde se marcan.

En algunas modalidades, se proporcionan métodos de ligación que proporcionan información sobre la secuencia de un ácido nucleico. Por ejemplo, en algunos aspectos puede realizarse una ligación en presencia de múltiples oligonucleótidos que son al menos parcialmente complementarios a una región objetivo en un molde. Pueden usarse

sondas de oligonucleótidos que hibridan o se asocian de otra forma con un molde adyacente o cerca de un extremo terminal de un cebador o sonda que se hibrida o se asocia de otra forma con un molde. Pueden usarse múltiples sondas de oligonucleótidos, cada una al menos parcialmente complementaria a una región de un molde para determinar la información de secuencia de una manera dependiente del molde, como se conoce en la técnica. Por ejemplo, la ligación de oligonucleótidos se usa para determinar la información de la secuencia de ácido nucleico en el Sistema SOLiD (Life Technologies-Applied Biosystems, Carlsbad, CA). De acuerdo con algunas modalidades, la información de secuencia se determina mediante la ligación de sondas oligonucleotídicas a un cebador oligonucleotídico de una manera dependiente de la secuencia del molde, por ejemplo, mediante el uso de una SFL. Las sondas de oligonucleótidos pueden ser un conjunto de múltiples sondas que tienen diferentes secuencias y marcadores distintivos, y el cebador y las sondas pueden tener longitudes de no más de 8, 7, 6, 5, 4, 3 o 2 nucleótidos. Los marcadores pueden proporcionar información sobre la secuencia de la sonda.

En algunas modalidades, la ligación puede realizarse mediante una "ligasa de huella pequeña" (en la presente "SFL") que puede ligar polinucleótidos cortos. Las SFL pueden usarse en cada una de las modalidades de ligaciones que se discuten anteriormente y en el resto de esta descripción, así como en otras modalidades de ligaciones que se conocen por un experto en la técnica. Por ejemplo, en algunas modalidades, una SFL puede ligar los extremos terminales de un primer oligonucleótido y un segundo oligonucleótido. El primer oligonucleótido puede ser un cebador y el segundo oligonucleótido puede ser una sonda, cada uno que se hibrida o se asocia de otra forma con una porción de un tercer oligonucleótido o un polinucleótido.

En algunas modalidades, la SFL puede ligar oligonucleótidos que tienen 8, 7, 6, 5, 4, 3 o 2 nucleótidos de longitud a un polinucleótido. La ligación de tales oligonucleótidos puede ser a oligonucleótidos de la misma longitud o de diferente longitud o a un polinucleótido. Por ejemplo, un oligonucleótido de 2 o 3 nucleótidos de longitud puede ligarse a un oligonucleótido de 2, 3, 4, 5, 6, 7, 8 o más nucleótidos de longitud o a oligonucleótidos más largos o a un polinucleótido.

También se proporciona un kit que comprende una ligasa de huella pequeña ("SFL") o una variante funcional o fragmento o derivado de esta. Las SFL ejemplares se identifican en la presente y se proporcionan sus secuencias. Opcionalmente, el kit también puede incluir una o más sondas de oligonucleótidos de menos de 12 nucleótidos de longitud (por ejemplo, no más de 8, 6, 5, 4, 3 o 2 nucleótidos de longitud). Opcionalmente, el kit incluye la ligasa CV y una o más sondas de oligonucleótidos de menos de 6 nucleótidos de longitud.

Breve Descripción de los Dibujos

Figura 1. (A) Sustrato de ligación de perlas covalente PI extendido. Se muestra un sustrato de ligación que se diseña para imitar las condiciones que se usan en la secuenciación SOLiD™. La Figura 1 (A) representa el ADN unido covalentemente a una perla (tal como una perla magnética, por ejemplo, de 0,1 a 1 µm de diámetro). El ADN de esta perla se enriqueció por la PCR mediante el uso de la secuencia PI de 40 nucleótidos. Para la región GCGGATGTACGGTACAGCAG, se alinea un "cebador" complementario de 20-mer que tiene un 5' PO_4 que reacciona con el grupo 3'OH de la sonda SOLiD, así como un marcador fluorescente 3' para la detección mediante electroforesis capilar. Para la secuencia CGAATAGA saliente, las sondas complementarias pueden hibridarse para demostrar la reacción de ligación. (B) Sustrato de ligación de perlas covalente PI. Se muestra un sustrato de ligación similar al de la Figura 1 (A), sin embargo, la reacción en este caso ocurrirá mucho más cerca de la superficie de la perla. Las sondas se hibridan con la región saliente AGTCGGTGAT, donde los residuos que se subrayan son opuestos al triplete de inosina de las sondas que tienen la estructura Colorante-5 'III (s) -xy-NNN 3', como se describe con más detalle en la presente.

Figura 2: Secuencias de ligasas ejemplares. La identificación de GenBank y el organismo fuente se proporcionan donde corresponda. También se proporcionan tres ligasas artificiales. La DLX difiere de la ligasa de ADN Hin en 1 aminoácido, que se designa como la letra en negrita que se subraya. La DLXd difiere de la ligasa de ADN Hin en 2 aminoácidos, que se designan como letras en negrita que se subraya. La DLXd2 es 22 aminoácidos más corta que la ligasa de ADN Hin y difiere de la ligasa de ADN Hin en 2 aminoácidos, que se designan como letras en negrita que se subraya.

Figura 3: Ligación de oligonucleótidos cortos con varias ligasas. Ligación de oligonucleótidos cortos con (A) la ligasa CV, (B) la ligasa DLXd y (C) la ligasa MnM. Las reacciones de ligación directa se realizaron en las siguientes condiciones: ligasa 2,0 µM, oligo corto 2-5 µM, cebador/molde 2,0 nM (unido a perlas magnéticas) y se procedió durante 20 minutos a 15 °C. La eficiencia de la ligación se calculó como la relación de áreas de pico determinadas por CE donde se usó un cebador marcado con FAM. Eficiencia = ligado/(ligado + no ligado).

Figura 4: Ligación de oligonucleótidos cortos con varias ligasas. Ligación de oligonucleótidos cortos con (A) la ligasa CV, (B) la ligasa DLXd y (C) la ligasa MnM. Las reacciones de ligación inversa se realizaron en las siguientes condiciones: ligasa 2,0 µM, oligo corto 2-5 µM, cebador/molde 2,0 nM (unido a perlas magnéticas) y se procedió durante 20 minutos a 15 °C. La eficiencia de la ligación se calculó como la relación de áreas de pico determinadas por CE donde se usó un cebador marcado con FAM. Eficiencia = ligado/(ligado + no ligado).

Figura 5: Ligación de 2-mers. La reacción de ligación en la dirección directa se realizó en las siguientes condiciones:

DLXd 2,0 µM, dinucleótido 123 µM (5'-CG-3 '), cebador/molde 2,0 nM (unido a perlas magnéticas) y continuó durante 20 minutos a 15 °C. La eficiencia de la ligación se calculó como la relación de áreas de pico determinadas por CE donde se usó un cebador marcado con FAM. Eficiencia = ligado/(ligado + no ligado).

Definiciones

"Degenerar", con respecto a una posición en un polinucleótido que es uno de una población de polinucleótidos, significa que la identidad de la base del nucleósido que ocupa esa posición varía entre los diferentes miembros de la población. Una población de polinucleótidos en este contexto es opcionalmente una mezcla de polinucleótidos dentro de una única fase continua (por ejemplo, un fluido). La "posición" puede designarse mediante un valor numérico que se asigna a uno o más nucleótidos en un polinucleótido, generalmente con respecto al extremo 5' o 3'. Por ejemplo, al nucleótido terminal en el extremo 3' de una sonda de extensión puede asignársele la posición 1. Así, en un grupo de sondas de extensión de estructura 3'-XXXNXXXX-5', el N está en la posición 4. Se dice que una posición es k-degenerada si puede ocuparse por nucleósidos que tienen cualquiera de k diferentes identidades. Por ejemplo, una posición que puede ocuparse por nucleósidos que comprenden cualquiera de las 4 bases diferentes es degenerada 4 veces.

En líneas similares, debe entenderse que una declaración de que se produjo un resultado (por ejemplo, la ligación, la unión) tiene la intención de indicar que el resultado se produjo a un nivel significativo o sustancial o un nivel mejorado en comparación con cuando no se produjo. Por ejemplo. Se dice que la ligación no se produjo si no es significativa, insustancial o muy reducida (por ejemplo, reducida al menos en un 80 %, 90 %, 95 % o 99 % en comparación con cuando ocurre la ligación (por ejemplo, en las condiciones que se describen en el último párrafo).

Los términos "micropartículas", "perlas", "microperlas", etc., se refieren a partículas (opcional pero no necesariamente de forma esférica) que tienen una longitud de sección transversal más pequeña (por ejemplo, un diámetro) de 50 micrómetros o menos, preferentemente 10 micrómetros o menos, 3 micrómetros o menos, aproximadamente 1 micrómetro o menos, aproximadamente 0,5 micrómetros o menos, por ejemplo, aproximadamente 0,1, 0,2, 0,3 o 0,4 micrómetros, o menos (por ejemplo, menos de 1 nanómetro, aproximadamente 1-10 nanómetros, aproximadamente 10- 100 nanómetros, o alrededor de 100-500 nanómetros). Las micropartículas (por ejemplo, Dynabeads de Dynal, Oslo, Noruega) pueden hacerse de una variedad de materiales inorgánicos u orgánicos que incluyen, pero no se limitan a, vidrio (por ejemplo, vidrio de poro controlado), sílice, circonia, poliestireno reticulado, poliacrilato, polimetilmetacrilato, dióxido de titanio, látex, poliestireno, etc. La magnetización puede facilitar la recolección y concentración de los reactivos que se unen a las micropartículas (por ejemplo, polinucleótidos o ligasas) después de la amplificación, y facilita los pasos adicionales (por ejemplo, lavados, eliminación de reactivos, etc.). En ciertas modalidades, puede usarse una población de micropartículas que tienen diferentes tamaños de formas y/o colores. Las micropartículas pueden codificarse opcionalmente, por ejemplo, con puntos cuánticos de modo que cada micropartícula pueda identificarse individual o exclusivamente.

El término "secuencia" se refiere a la información de secuencia sobre un polinucleótido o polipéptido o cualquier porción del polinucleótido o polipéptido que tiene dos o más unidades (nucleótidos o aminoácidos) de largo. El término también puede usarse como una referencia al polinucleótido o molécula de polipéptido en sí o una porción relevante de este. La información de la secuencia de polinucleótidos se refiere a la sucesión de bases de nucleótidos en el polinucleótido, y en un polipéptido se relaciona con la sucesión de cadenas laterales de aminoácidos en el polipéptido o una porción de este. Por ejemplo, si el polinucleótido contiene bases adenina, guanina, citosina, timina o uracilo, la secuencia de polinucleótidos puede representarse por una sucesión correspondiente de letras A, G, C, T o U), por ejemplo, una molécula de ADN o ARN. Las secuencias que se muestran en la presente se presentan en una orientación 5'→3' a menos que se indique de otra forma.

"Dúplex perfectamente apareado" en referencia a sondas y polinucleótidos molde significa que uno forma una estructura bicatenaria con el otro tal que cada nucleósido en la estructura bicatenaria sufre un apareamiento de bases de Watson-Crick con un nucleósido en el otro. El término también comprende el apareamiento de análogos de nucleósidos, tales como la desoxinosina, nucleósidos con bases de 2-aminopurina, y similares, que pueden emplearse para reducir la degeneración de las sondas, ya sea que tal apareamiento implique o no la formación de enlaces de hidrógeno.

El término "polimorfismo" recibe su significado ordinario en la técnica y se refiere a una diferencia en la secuencia del genoma entre individuos de la misma especie. Un "polimorfismo de un solo nucleótido" (SNP) se refiere a un polimorfismo en una sola posición.

"Sondas", "oligonucleótidos" o "cebadores" pretenden ser términos intercambiables en la presente, de modo que cualquiera de estos pueda tomarse como referencia a otro. Estos son polinucleótidos que no necesariamente se limitan a cualquier longitud. Donde se desee, estos pueden tener menos de 100 nucleótidos de largo, a veces menos de 30 nucleótidos de largo, por ejemplo, menos de 20 nucleótidos, opcionalmente menos de 12 nucleótidos, por ejemplo, menos de ocho nucleótidos de longitud. En algunos casos, estos tienen 2, 3, 4, 5, 6, 7, 8 o más nucleótidos de longitud. En algunos casos, estos tienen 3 o 4 nucleótidos de longitud.

Un "polinucleótido", también llamado "ácido nucleico", es un polímero lineal de dos o más nucleótidos que se unen por enlaces internucleosídicos covalentes, o un fragmento variante o funcional de este. Una secuencia de letras, como "ATGCCTG", pretende representar una secuencia de polinucleótidos en el orden 5'→73' de izquierda a derecha a menos que se especifique de otra forma. En ejemplos naturales de estos, el enlace internucleosídico es típicamente un enlace fosfodiéster. Sin embargo, otros ejemplos opcionalmente comprenden otros enlaces internucleosídicos, tales como enlaces de fosforotiolato y pueden comprender o no un grupo fosfato. En otros casos, el polinucleótido puede contener cadenas principales no nucleotídicas, por ejemplo, poliamida (por ejemplo, Ácidos nucleicos peptídicos (PNA)) y polimorfolino y otros polímeros de ácido nucleico específicos de secuencia sintética, siempre que los polímeros contengan nucleobases en una configuración que permita el apareamiento de bases y apilamiento de bases, como el que se encuentra en el ADN y el ARN.

Como se usa en la presente, "polinucleótido", "oligonucleótido", "sonda", "cebador", "molde", "ácido nucleico" y similares se pueden tomar para referirse a poblaciones o grupos de moléculas individuales que son sustancialmente idénticas a través de su longitud total o en una parte relevante de interés. Por ejemplo, el término "molde" puede indicar una pluralidad de moléculas de molde que son sustancialmente idénticas, etc. En el caso de polinucleótidos que están degenerados en una o más posiciones, se apreciará que el polinucleótido degenerado comprende una pluralidad de moléculas de polinucleótidos, que tienen secuencias que son sustancialmente idénticas solo en las posiciones no degeneradas y difieren en la secuencia en las posiciones degeneradas. Por lo tanto, la referencia a "un" polinucleótido (por ejemplo, "un" cebador, sonda, oligonucleótido, molde, etc.) puede significar una población de moléculas de polinucleótidos que son sustancialmente idénticas en al menos una parte de interés, de modo que la naturaleza plural de la población no necesita indicarse explícitamente, pero si así se desea puede hacerlo. Estos términos también pretenden proporcionar un soporte adecuado para una declaración que especifica explícitamente una sola molécula de polinucleótido. Se entenderá que los miembros de una población no necesitan ser 100 % idénticos, por ejemplo, puede ocurrir un cierto número de "errores" durante el curso de la síntesis. Preferentemente, al menos el 90 %, al menos el 95 %, al menos el 99 % o más de los miembros de una población son sustancialmente idénticos. Preferentemente, el porcentaje de identidad de al menos el 95 % o más preferentemente al menos el 99 % de los miembros de la población a una molécula de ácido nucleico de referencia es al menos 98 %, 99 %, 99,9 % o más. El porcentaje de identidad puede calcularse mediante la comparación de dos secuencias que se alinean de manera óptima, mediante la determinación del número de posiciones en las que se produce una base de ácido nucleico idéntica (por ejemplo, A, T, C, G, U o I) en ambas secuencias para obtener el número de posiciones pareadas, mediante la división del número de posiciones pareadas por el número total de posiciones, y mediante la multiplicación del resultado por 100 para obtener el porcentaje de identidad de secuencia. Se apreciará que, en ciertos casos, una molécula de ácido nucleico tal como un molde, sonda, cebador, etc., puede ser una porción de una molécula de ácido nucleico más grande que también contiene una porción que no sirve a una función de molde, sonda o cebador. En ese caso, los miembros individuales de una población no necesitan ser sustancialmente idénticos con respecto a esa porción.

Los nucleótidos de un polinucleótido pueden tener cualquier combinación de bases, que incluye las que se mencionan en la presente, por ejemplo, uracilo, adenina, timina, citosina, guanina, inosina, xatanina, hipoxatanina, isocitosina, isoguanina, etc. Opcionalmente, el polinucleótido es un ADN que tiene las bases de nucleótidos A, C, T y/o G. Opcionalmente, el polinucleótido es un ARN que tiene las bases de nucleótidos A, C, T y/o U.

Los polinucleótidos incluyen los ADN bicatenario y monocatenario, así como los ARN bicatenario y monocatenario, los híbridos de ADN:ARN, los péptidos-ácidos nucleicos (PNA) e híbridos entre PNA y ADN o ARN, y también incluyen tipos de modificaciones que se conocen, por ejemplo, marcadores que se conocen en la técnica, metilación, "caperuzas", sustitución de uno o más de los nucleótidos naturales con un análogo, modificaciones internucleotídicas tales como, por ejemplo, aquellos con enlaces no cargados (por ejemplo, metilfosfonatos, fosfotriésteres, fosforamidatos, carbamatos, etc.), con enlaces cargados negativamente (por ejemplo, fosforotioatos, fosforoditioatos, etc.), y con enlaces cargados positivamente (por ejemplo, aminoalquilfosforamidatos, aminoalquilfosfotriésteres), aquellos que contienen restos colgantes, tales como, por ejemplo, proteínas (que incluye nucleasas, toxinas, anticuerpos, péptidos señal, poli-L-lisina, etc.), aquellos con intercaladores (por ejemplo, acridina, psoraleno, etc.), aquellos que contienen quelantes (por ejemplo, metales, metales radiactivos, boron, metales oxidativos, etc.), aquellos que contienen alquiladores, aquellos con enlaces modificados (por ejemplo, ácidos nucleicos anoméricos alfa, etc.), así como formas no modificadas del polinucleótido u oligonucleótido. Los polinucleótidos pueden unirse opcionalmente a uno o más restos no nucleotídicos tales como marcadores y otras moléculas pequeñas, moléculas grandes tales como proteínas, lípidos, azúcares y soportes sólidos o semisólidos, por ejemplo a través del extremo 5'o 3'. Los marcadores incluyen cualquier resto que es detectable mediante el uso de un método de detección de elección, y por lo tanto hace que el nucleótido o polinucleótido que se une sea detectable de manera similar mediante el uso de un método de detección de elección. Opcionalmente, el marcador emite una radiación electromagnética que es ópticamente detectable o visible. En algunos casos, el nucleótido o polinucleótido no se une a un marcador, y la presencia del nucleótido o polinucleótido se detecta directamente, y/o se detecta la generación de subproductos de la ligación, como PPi o NMN. Opcionalmente, la presencia del nucleótido, polinucleótido o subproducto se detecta mediante un transistor de efecto de campo químico, por ejemplo, donde la carga en el electrodo de puerta se genera mediante un proceso químico. Opcionalmente, el transistor de efecto de campo químico es un transistor de efecto de campo sensible a iones.

Cuando se marcan dos o más reactivos, los marcadores se distinguen preferentemente entre sí con el método de detección que se elige. Por ejemplo, los marcadores pueden ser espectralmente resolubles, es decir, distinguibles en función de sus características espectrales, particularmente la longitud de onda de emisión de fluorescencia, en condiciones de funcionamiento. En otros casos, el marcador puede comprender un compuesto generador de señal (SGC). Un SGC es opcionalmente una sustancia que es detectable en un ensayo de elección, o capaz de reaccionar para formar una entidad química o física (es decir, un producto de reacción) que es detectable en un ensayo de elección. Ejemplos representativos de productos de reacción incluyen precipitados, señales fluorescentes, compuestos que tienen un color y similares. Los SGC representativos incluyen, por ejemplo, compuestos bioluminiscentes (por ejemplo, luciferasa), fluoróforos (por ejemplo, a continuación), compuestos bioluminiscentes y quimioluminiscentes, radioisótopos (por ejemplo, ^{131}I, ^{125}I, ^{14}C, ^{3}H, ^{35}S, ^{32}P y similares), enzimas (por ejemplo, a continuación), proteínas de unión (por ejemplo, biotina, avidina, estreptavidina y similares), partículas magnéticas, compuestos químicamente reactivos (por ejemplo, manchas coloreadas), oligonucleótidos marcados; sondas moleculares (por ejemplo, CY3, Research Organics, Inc.) y similares. Los fluoróforos representativos incluyen isotiocianato de fluoresceína, succinil fluoresceína, rodamina B, lisamina, 9,10-difenilantraceno, perileno, rubreno, pireno y los derivados fluorescentes de estos tales como isocianato, isotiocianato, cloruro de ácido o cloruro de sulfonilo, umbeliferona, quelatos de tierras raras tales como lantánidos tales como Europio (Eu) y similares. Los compuestos generadores de señales también incluyen SGC cuyos productos son detectables por longitudes de onda fluorescentes y quimioluminiscentes, por ejemplo, colorantes de secuenciación, luciferasa, metales emisores de fluorescencia tales como ^{152}Eu u otros de la serie de los lantánidos; compuestos tales como luminol, isoluminol, sales de acridinio y similares; compuestos bioluminiscentes tales como luciferina; proteínas fluorescentes (por ejemplo, GFP o variantes de estas); y similares. Los sujetos SGC son opcionalmente detectables mediante el uso de un método visual u óptico; preferentemente, con un método susceptible de automatización tal como un método espectrofotométrico, un método de fluorescencia, un método quimioluminiscente, un método nanométrico eléctrico que implica, por ejemplo, un cambio en la conductancia, impedancia, resistencia y similares y un método de campo magnético. Algunos SGC son opcionalmente detectables a simple vista o con un aparato de detección de señal cuando se encuentra en una concentración adecuada.

Un "nucleótido" se refiere a un nucleótido, nucleósido o análogo de este. Opcionalmente, el nucleótido es un glucósido N o C de una base de purina o pirimidina. (por ejemplo, un desoxirribonucleósido que contiene 2-desoxi-D-ribosa o ribonucleósido que contiene D-ribosa). Los ejemplos de otros análogos incluyen, sin limitación, fosforotioatos, fosforamidatos, metilfosfonatos, quiral-metilfosfonatos, 2-O-metil ribonucleótidos.

Las bases de nucleótidos o nucleobases generalmente tienen un anillo o anillos aromáticos parentales sustituidos o no sustituidos. En ciertas modalidades, el anillo o anillos aromáticos contienen al menos un átomo de nitrógeno. En ciertas modalidades, la base de nucleótidos es capaz de formar enlaces de hidrógeno de Watson-Crick y/o Hoogsteen con una base de nucleótidos apropiadamente complementaria. Los ejemplos de bases de nucleótidos y análogos de estos incluyen, pero no se limitan a, purinas tales como 2-aminopurina, 2,6-diaminopurina, adenina (A), etenoadenina, N6-Δ2-isopenteniladenina (6iA), N6-Δ2-isopentenil-2-metiltioadenina (2ms6iA), N6-metiladenina, guanina (G), isoguanina, N2-dimetilguanina (dmG), 7-metilguanina (7 mG), 2-tiopirimidina, 6-tioguanina (6sG), hipoxantina y O6-metilguanina; 7-deaza-purinas tales como 7-deazaadenina (7-deaza-A) y 7-deazaguanina (7-deaza-G); pirimidinas como la citosina (C), 5-propinilocitosina, isocitosina, timina (T), 4-tiotimina (4sT), 5,6-dihidrotimina, O4-metiltilamina, uracilo (U), 4-tiouracilo (4sU) y 5, 6-dihidrouracilo (dihidrouracilo; D); indoles tales como nitroindol y 4-metilindol; pirrol tales como nitropirrol; nebularina; base (Y); etc. En ciertas modalidades, las bases de nucleótidos son bases de nucleótidos universales. Pueden encontrarse bases de nucleótidos ejemplares adicionales, por ejemplo, en Fasman, 1989, Practical Handbook of Biochemistry and Molecular Biology, págs. 385-394, CRC Press, Boca Raton, Florida, y las referencias que se citan allí. Un nucleósido es generalmente un compuesto que tiene una base de nucleótidos que se une covalentemente al carbono C-1'de un azúcar pentosa. En ciertas modalidades, el enlace es a través de un anillo heteroaromático de nitrógeno. Los azúcares de pentosa típicos incluyen, pero no se limitan a, aquellas pentosas en las que uno o más de los átomos de carbono están cada uno independientemente sustituido con uno o más de estos o diferentes grupos -R, -OR, -NRR o halógeno, donde cada R es independientemente hidrógeno, alquilo (C1-C6) o arilo (C5-C14). El azúcar pentosa puede estar saturado o insaturado. Los ejemplos de azúcares pentosa y análogos de estos incluyen, pero no se limitan a, ribosa, 2'-desoxirribosa, 2'-(C1-C6) alcoxirribosa, 2'-(C5-C14) ariloxirribosa, 2',3'-dideoxirribosa, 2',3'-didehidroribosa, 2'-desoxi-3'-haloribosa, 2'-desoxi-3'-fluororibosa, 2'-desoxi-3'-clororibosa, 2'-desoxi-3'-aminoribosa, 2'-desoxi-3'-(C1-C6) alquilribosa, 2'-desoxi-3'-(C1-C6) alcoxirribosa y 2'-desoxi-3'-(C5-C14) ariloxirribosa. Uno o más de los carbonos de pentosa de un nucleósido pueden sustituirse con un éster de fosfato, como se describe en la patente de EE.UU. Núm. 7255994. En ciertas modalidades, los nucleósidos son aquellos en los que la base de nucleótidos es una purina, una 7-deazapurina, una pirimidina, una base de nucleótidos universal, una base de nucleótidos específica o un análogo de esta. Los análogos de nucleótidos incluyen derivados en los que el azúcar pentosa y/o la base de nucleótidos y/o uno o más de los ésteres de fosfato de un nucleósido pueden reemplazarse con su análogo respectivo. Los análogos ejemplares de azúcar pentosa y análogo de base de nucleótidos se describen anteriormente. Ejemplos de análogos de éster de fosfato incluyen, pero no se limitan a, alquilfosfonatos, metilfosfonatos, fosforamidatos, fosfotriésteres, fosforotioatos, fosforoditioatos, fosforoselenoatos, fosforodiselenoatos, fosforoanilotioatos, fosforoanilidatos, fosforoamidatos, boronofosfatos, etc., y pueden incluir los contraiones que se asocian. Otros análogos de nucleótidos son los monómeros de análogos de nucleótidos que pueden polimerizarse en análogos de polinucleótidos en los que el esqueleto de fosfato del ADN/ARN y/o el esqueleto de éster de fosfato del azúcar se reemplaza con un tipo diferente de enlace. Los ejemplos de análogos de

polinucleótidos incluyen, pero no se limitan a, ácidos nucleicos peptídicos, en los que el esqueleto de fosfato del azúcar del polinucleótido se reemplaza por un esqueleto peptídico.

Los enlaces internucleosídicos pueden ser un enlace fosfodiéster, aunque pueden usarse otros enlaces (por ejemplo, enlaces escindibles que pueden escindirse sustancialmente en condiciones en las que los enlaces fosfodiéster no se cortan sustancialmente). Por ejemplo, un enlace que contiene un sitio sensible a la endonucleasa AP, por ejemplo, un residuo abásico, un residuo que contiene una base dañada que es un sustrato para su eliminación por una glicosilasa de ADN, u otro residuo o enlace que es un sustrato para la escisión por una AP endonucleasa, o un nucleósido disacárido.

El término adjetivo "hibridado" se refiere opcionalmente a dos polinucleótidos que se unen entre sí por dos o más pares de bases adyacentes secuencialmente. El término "hibridación" se refiere al proceso por el cual los polinucleótidos se hibridan entre sí. Dos polinucleótidos monocatenarios pueden considerarse "complementarios" si, cuando se hibridan juntos, el polinucleótido más largo forma un saliente monocatenario y el polinucleótido más corto puede ligarse eficientemente a un tercer polinucleótido adyacente que forma un dúplex perfectamente compatible con el monocatenario saliente. Cuando el saliente monocatenario tiene menos de ocho nucleótidos, puede alargarse arbitrariamente a ocho nucleótidos mediante la agregación de una combinación aleatoria de nucleótidos al saliente.

De manera similar, los residuos de nucleótidos pueden considerarse como complementarios si cuando ambos están apareados entre sí en dos polinucleótidos hibridados, cualquiera de los nucleótidos puede ligarse en una reacción de ligación dirigida por un molde cuando se sitúa como el nucleótido terminal en su polinucleótido. Los nucleótidos que se incorporan eficientemente por las polimerasas de ADN opuestas entre sí durante la replicación del ADN en condiciones fisiológicas también se consideran complementarios. En una modalidad, los nucleótidos complementarios pueden formar pares de bases entre sí, tales como los pares de bases A-T/U y G-C que se forman mediante enlaces de hidrógeno de tipo Watson-Crick específicos entre las nucleobases de nucleótidos y/o las posiciones de polinucleótidos antiparalelas entre sí. La complementariedad de otros pares de bases artificiales puede basarse en otros tipos de enlaces de hidrógeno y/o hidrofobicidad de bases y/o complementariedad de forma entre bases.

En casos apropiados, los polinucleótidos pueden considerarse como complementarios cuando pueden experimentar un apareamiento de bases acumulativo en dos o más posiciones correspondientes individuales en orientación antiparalela, como en un dúplex hibridado. Opcionalmente puede haber una complementariedad "completa" o "total" entre una primera y segunda secuencia de polinucleótidos donde cada nucleótido en la primera secuencia de polinucleótidos puede experimentar una interacción de apareamiento de bases estabilizadoras con un nucleótido en la posición antiparalela correspondiente en el segundo polinucleótido. La complementariedad "parcial" describe secuencias de polinucleótidos en las que al menos el 20 %, pero menos del 100 %, de los residuos de un polinucleótido son complementarios a los residuos en el otro polinucleótido. Una "no coincidencia" está presente en cualquier posición en la que los dos nucleótidos opuestos no son complementarios. En algunos ensayos de ligación, un polinucleótido puede experimentar una ligación sustancial dependiente del molde, incluso cuando tiene una o más no coincidencia con su molde hibridado. Opcionalmente, el polinucleótido no tiene más de 4, 3 o 2 no coincidencias, por ejemplo, 0 o 1 no coincidencia, con su molde. En algunos ensayos, el polinucleótido no se somete a una ligación sustancial dependiente del molde a menos que sea al menos un 60 % complementario, por ejemplo, al menos aproximadamente 70 %, 80 %, 85 %, 90 %, 95 %, 99 % o 100 % complementario a su molde.

Como se usa en la presente, una "muestra biológica" se refiere a una muestra de tejido o líquido aislado de un individuo, que incluye, pero no se limita a, por ejemplo, plasma, suero, líquido cefalorraquídeo, semen, líquido linfático, las secciones externas de la piel, tractos respiratorios, intestinales y genitourinarios, lágrimas, saliva, leche, células sanguíneas, tumores, órganos y también muestras de constituyentes de cultivo celular in vitro (que incluyen, pero no se limitan a, medio condicionado resultante del crecimiento de células en medio de cultivo celular, putativamente células infectadas con virus, células recombinantes y componentes celulares).

La identidad de secuencia (que también se llama homología) se refiere a la similitud en la secuencia de dos o más secuencias (por ejemplo, secuencias de nucleótidos o polipéptidos). En el contexto de dos o más secuencias homólogas, el porcentaje de identidad u homología de las secuencias o subsecuencias de estas indica el porcentaje de todas las unidades monoméricas (por ejemplo, nucleótidos o aminoácidos) que son iguales (es decir, aproximadamente 70 % de identidad, preferentemente 75 %, 80 %, 85 %, 90 %, 95 % o 99 % de identidad). El porcentaje de identidad puede estar sobre una región que se especifica, cuando se compara y se alinea para una correspondencia máxima sobre una ventana de comparación, o una región que se designa y se mide con un algoritmo de comparación de secuencia BLAST o BLAST 2.0 con los parámetros predeterminados que se describen a continuación, o mediante la alineación manual e inspección visual. Se dice que las secuencias son "sustancialmente idénticas" cuando hay al menos un 90 % de identidad a nivel de aminoácidos o a nivel de nucleótidos. Preferentemente, la identidad existe sobre una región que tiene al menos aproximadamente 2, 3, 4, 5, 6, 7, 8, 10, 12, 15, 20, 25, 50 o 100 residuos de longitud, o en toda la longitud de al menos una secuencia que se compara. Un algoritmo preferente para determinar el porcentaje de identidad de secuencia y la similitud de secuencia son los algoritmos BLAST y BLAST 2.0, que se describen en Altschul y otros, Nuc. Acidos Res. 25: 3389-3402 (1977). Otros métodos incluyen los algoritmos de Smith & Waterman, Adv. Appl. Math. 2: 482 (1981), y Needleman & Wunsch, J. Mol. Biol. 48: 443 (1970), etc. Otra indicación de que dos secuencias de ácido nucleico son sustancialmente idénticas es que las dos moléculas

o sus complementos se hibridan entre sí en condiciones rigurosas.

En las reivindicaciones, cualquier verbo activo (o su gerundio) se destina a indicar la acción real o intencionada correspondiente, incluso si no se produce una acción real. Por ejemplo, el verbo "hibridar" y el gerundio "hibridando" y similares se refieren a la hibridación real o al intento de hibridación al poner en contacto secuencias de ácido nucleico en condiciones adecuadas para la hibridación, incluso si no se produce una hibridación real. De manera similar, "detectar" y "detección" cuando se usa en las reivindicaciones se refiere a la detección real o al intento de detección, incluso si no se detecta realmente ningún objetivo.

La "hibridación inespecífica" se usa para referirse a cualquier hibridación no intencionada o insignificante, por ejemplo, la hibridación a una secuencia polinucleotídica no intencionada distinta de la secuencia polinucleotídica objetivo que se desea. La secuencia de polinucleótidos no intencionada puede estar en el mismo polinucleótido o en uno diferente del objetivo que se desea. En algunos casos, la única hibridación que se prevé puede ser el apareamiento de bases de Watson-Crick entre dos polinucleótidos. Otros tipos de apareamientos de bases que se prevén pueden incluir el apareamiento de bases entre los análogos correspondientes de tales nucleótidos o entre isocitidina e isoguanina. En algunos casos donde la hibridación solo se pretende entre bases complementarias, cualquier enlace entre bases no complementarias se considera hibridación no específica.

En referencia a la ligación de dos polinucleótidos, el extremo terminal "proximal" de cualquiera de los polinucleótidos es el extremo terminal que se pretende ligar al otro polinucleótido. En general, este es el extremo terminal que se pone en contacto con el sitio activo de la ligasa, o el extremo terminal que finalmente se liga al otro polinucleótido, mientras que el extremo terminal opuesto es el extremo terminal "distal". El residuo de nucleótido terminal en el extremo terminal proximal puede denominarse nucleótido proximal, y la posición de nucleótido proximal que opcionalmente se designa como posición 1, o -1 en dependencia del lado del sitio de ligación al que nos referimos, la penúltima posición de nucleótido como posición 2 o -2, etc. Con referencia a dos polinucleótidos que se hibridan adyacentemente, el extremo terminal proximal es generalmente el extremo terminal de un polinucleótido que está más cerca del otro polinucleótido. En algunos casos no limitantes de ligación dependiente del molde, los extremos terminales proximales de ambos polinucleótidos se hibridan adyacentemente entre sí.

"Soporte", como se usa en la presente, se refiere a una estructura o matriz sobre o en la cual los reactivos de ligación, por ejemplo, las moléculas de ácido nucleico, micropartículas y similares, pueden inmovilizarse de manera que se evite de manera significativa o total que se difundan libremente o se muevan con respeto a otro. Los reactivos pueden, por ejemplo, ponerse en contacto con el soporte, y opcionalmente unirse de forma covalente o no covalente o incrustarse parcial/completamente

Una "base universal", como se usa en la presente, es una base que es complementaria a más de otra base. Las bases completamente universales pueden aparearse con cualquiera de las bases que se encuentran típicamente en los ácidos nucleicos naturales. La base no necesita ser igualmente capaz de aparearse con cada una de las bases naturales. Alternativamente, la base universal puede aparearse solo o selectivamente con dos o más bases, pero no todas las bases. Opcionalmente, la base universal se aparea solo o selectivamente con purinas, o alternativamente con pirimidinas. Si se desea, pueden incluirse dos o más bases universales en una posición particular en una sonda. Se conocen varias bases universales en la técnica que incluyen, pero no se limitan a, inosina, hipoxantina, 3-nitropirrol, 4-nitroindol, 5-nitroindol, 4-nitrobencimidazol, 5-nitroindazol, 8-aza-7-deazaadenina, 6H, 8H-3,4-dihidropirimido[4,5-c][1,2]oxazin-7-ona, 2-amino-6-metoxiaminopurina, etc. La hipoxantina es una base completamente universal preferente. Los nucleósidos que comprenden la hipoxantina incluyen, pero no se limitan a, inosina, isoinosina, 2'-desoxiinosina y 7-deaza-2'-desoxiinosina, 2-aza-2'deoxiinosina.

"Purificado" generalmente se refiere al aislamiento de una sustancia (compuesto, polinucleótido, proteína, polipéptido, composición de polipéptidos) tal que la sustancia comprende un porcentaje significativo, como una proporción mayor de la que se encuentra naturalmente (por ejemplo, mayor que el 2 %, mayor que el 5 %, mayor que el 10 %, mayor que el 20 %, mayor que el 50 % o más, a veces más del 90 %, 95 % o 99 %) de la muestra en la que reside. En ciertas modalidades, un componente sustancialmente purificado comprende al menos un 50 %, 80 % -85 % o 90-95 % de la muestra. Las técnicas para purificar polinucleótidos y polipéptidos de interés se conocen bien en la técnica e incluyen, por ejemplo, cromatografía de intercambio iónico, cromatografía de afinidad y sedimentación de acuerdo con la densidad. "Aislado" se refiere al material retirado de su entorno original (por ejemplo, el entorno natural si es natural) y, por lo tanto, se altera "por la mano del hombre" de su estado natural. Por ejemplo, un polinucleótido aislado puede ser parte de un vector o una composición de materia, o puede contenerse dentro de una célula, y aún estar "aislado" porque ese vector, composición de materia o célula particular no es el entorno original o natural del polinucleótido.

Las ligasas ejemplares comprenden un polipéptido. Los términos "polipéptido", "péptido" y "proteína" se usan de manera intercambiable en la presente para referirse a un polímero de residuos de aminoácidos, o cualquier variante o fragmento funcional de este. Los términos se aplican a los polímeros de aminoácidos en los que uno o más residuos de aminoácidos son un mimético químico artificial de un aminoácido natural correspondiente, así como a los polímeros de aminoácidos naturales y los polímeros de aminoácidos no naturales.

El término "aminoácido" incluye aminoácidos naturales y sintéticos, así como análogos de aminoácidos y miméticos

de aminoácidos que funcionan de manera similar a los aminoácidos naturales. Los aminoácidos naturales son aquellos que se codifican por el código genético, así como aquellos aminoácidos que se modifican posteriormente, por ejemplo, hidroxiprolina, gamma-carboxiglutamato y O-fosfoserina. Los análogos de aminoácidos se refieren a compuestos que tienen la misma estructura química básica que un aminoácido natural, es decir, un carbono que está unido a un hidrógeno, un grupo carboxilo, un grupo amino y un grupo R, por ejemplo, homoserina, norleucina, metionina sulfóxido, metionina metil sulfonio. Tales análogos tienen grupos R modificados (por ejemplo, norleucina) o esqueletos de péptidos modificados, pero conservan la misma estructura química básica que un aminoácido natural. Los miméticos de aminoácidos se refieren a compuestos químicos que tienen una estructura que es diferente de la estructura química general de un aminoácido, pero que funciona de manera similar a un aminoácido natural.

Las variantes o derivados de una secuencia de nucleótidos o secuencia de polipéptidos dada son variantes que se modifican opcionalmente de forma conservadora. Con respecto a las secuencias de ácido nucleico particulares, las variantes que se modifican conservadoramente se refieren a aquellos ácidos nucleicos que codifican secuencias de aminoácidos idénticas o esencialmente idénticas, o donde el ácido nucleico no codifica una secuencia de aminoácidos, a secuencias esencialmente idénticas.

En cuanto a las secuencias de aminoácidos, un experto reconocerá que las sustituciones, eliminaciones o adiciones individuales a una secuencia de ácido nucleico, péptido, polipéptido o proteína que altera, agrega o elimina un solo aminoácido o un pequeño porcentaje de aminoácidos en la secuencia que se codifica es una "variante conservadoramente modificada" donde la alteración resulta en la sustitución de un aminoácido con un aminoácido químicamente similar. Las tablas de sustitución conservadoras que proporcionan aminoácidos funcionalmente similares se conocen bien en la técnica. Tales variantes conservadoramente modificadas son adicionales y no excluyen variantes polimórficas, homólogos entre especies y alelos de la descripción. (véase, por ejemplo, Creighton, Proteins (1984)).

La identidad de secuencia (que también se llama homología) se refiere a la similitud en la secuencia de dos o más secuencias (por ejemplo, secuencias de nucleótidos o polipéptidos). En el contexto de dos o más secuencias homólogas, el porcentaje de identidad u homología de las secuencias o subsecuencias de estas indica el porcentaje de todas las unidades monoméricas (por ejemplo, nucleótidos o aminoácidos) que son iguales (es decir, aproximadamente 70 % de identidad, preferentemente 75 %, 80 %, 85 %, 90 %, 95 % o 99 % de identidad). El porcentaje de identidad puede estar sobre una región que se especifica, cuando se compara y se alinea para una correspondencia máxima sobre una ventana de comparación, o una región que se designa y se mide con un algoritmo de comparación de secuencia BLAST o BLAST 2.0 con los parámetros predeterminados que se describen a continuación, o mediante la alineación manual e inspección visual. Se dice que las secuencias son "sustancialmente idénticas" cuando hay al menos un 90 % de identidad a nivel de aminoácidos o a nivel de nucleótidos. Esta definición también se refiere al complemento de una secuencia de prueba. Preferentemente, la identidad existe sobre una región que tiene al menos aproximadamente 25, 50 o 100 residuos de longitud, o en toda la longitud de al menos una secuencia comparada. Un algoritmo preferente para determinar el porcentaje de identidad de secuencia y la similitud de secuencia son los algoritmos BLAST y BLAST 2.0, que se describen en Altschul y otros, Nuc. Acidos Res. 25: 3389-3402 (1977). Otros métodos incluyen los algoritmos de Smith & Waterman, Adv. Appl. Math. 2: 482 (1981), y Needleman & Wunsch, J. Mol. Biol. 48: 443 (1970), etc. Otra indicación de que dos secuencias de ácido nucleico son sustancialmente idénticas es que las dos moléculas o sus complementos se hibridan específicamente entre sí en condiciones rigurosas, como las que se describen en la presente. Los ácidos nucleicos que no se hibridan específicamente entre sí en condiciones rigurosas siguen siendo sustancialmente idénticos si los polipéptidos que codifican son sustancialmente idénticos.

Dos polinucleótidos se hibridan selectivamente (o específicamente) entre sí si se unen de manera significativa o detectable entre sí en condiciones de hibridación rigurosas cuando están presentes en una mezcla de polinucleótidos complejos, como el ADN celular total o de la biblioteca. Para la hibridación selectiva o específica, una señal positiva es al menos dos veces el fondo, preferentemente 10 veces la hibridación de fondo. Opcionalmente, las condiciones rigurosas se seleccionan para que sean aproximadamente 5-10 ºC más bajas que el punto de fusión térmica para la secuencia específica a un pH de fuerza iónica que se define. Las condiciones rigurosas son opcionalmente en las que la concentración de sal es menor que aproximadamente 1,0 M de iones de sodio, típicamente de aproximadamente 0,01 a 1,0 M de concentración de iones de sodio (u otras sales) a un pH de 7,0 a 8,3 y la temperatura es de al menos aproximadamente 30 ºC para sondas cortas (por ejemplo, de 10 a 50 nucleótidos) y al menos aproximadamente 60 ºC para sondas largas (por ejemplo, más de 50 nucleótidos). También pueden lograrse condiciones rigurosas con la adición de agentes desestabilizadores como la formamida. Las condiciones de hibridación rigurosas ejemplares pueden ser las siguientes: 50 % de formamida, 5 x SSC y 1 % de SDS, la incubación a 42 °C, o 5 x SSC, 1 % de SDS, la incubación a 65 °C, con lavado en 0,2 x SSC y SDS al 0,1 % a 65 ºC.

Descripción detallada

Entre otra información novedosa y sorprendente se presenta en la presente, la nueva y sorprendente ligación enzimática de polinucleótidos cortos se presenta en la presente.

1) Ligaciones

La ligación en la presente se refiere a la formación enzimática de un enlace covalente entre los extremos terminales de dos o más cadenas de polinucleótidos. La "ligación" implica la formación de una unión o enlace covalente entre los extremos terminales 5'y 3' de dos o más ácidos nucleicos, por ejemplo, oligonucleótidos y/o polinucleótidos, opcionalmente en una reacción que es dirigida por un molde. La naturaleza de la unión o enlace puede variar ampliamente y la ligación preferentemente se logra enzimáticamente. La naturaleza de la unión o enlace puede variar ampliamente. Las ligaciones ilustrativas no limitantes se realizan enzimáticamente para formar un enlace fosfodiéster entre un nucleótido 5' terminal de una cadena polinucleotídica con un nucleótido 3' terminal de un polinucleótido.

La ligación puede ser uno o más de los siguientes tipos de ligación que se describen en la presente. Un primer tipo de ligación enzimática implica la formación de un enlace covalente entre un primer extremo terminal de una primera cadena de polinucleótidos y un segundo extremo terminal diferente de una segunda cadena de polinucleótidos. El primer y el segundo extremo terminal de polinucleótidos pueden estar en diferentes cadenas de polinucleótidos, o ambos pueden estar en la misma cadena de polinucleótidos (lo que resulta en la circularización). Opcionalmente, la primera y la segunda cadena de polinucleótidos no se hibridan con un tercer polinucleótido. Opcionalmente, los extremos terminales de la primera y segunda cadena de polinucleótidos se unen independientemente de sus secuencias (por ejemplo, la ligación de extremo romo o unión de extremo no homólogo). En otra variación, pueden ligarse dos polinucleótidos bicatenarios con porciones sobresalientes monocatenarias que son complementarias entre sí (por ejemplo, la ligación con extremo cohesivo). Un tercer tipo de ligación (ligación dependiente de molde) se describe a continuación.

En cualquiera de los métodos que se describen en la presente, las cadenas de polinucleótidos pueden estar en formato monocatenario o se hibridan con cadenas complementarias en formato bicatenario. En la ligación de extremo romo, ambas cadenas de polinucleótidos a unir se hibridan con dos cadenas complementarias diferentes tal que no exista saliente.

Se proporcionan métodos para ligar dos polinucleótidos. Un ejemplo de método de ligación logra la ligación entre un primer extremo terminal de una primera secuencia de polinucleótidos y un segundo extremo terminal de una segunda secuencia de polinucleótidos. Para facilitar la referencia, la primera secuencia de polinucleótidos se llama "sonda de inicialización" o "cebador", el segundo polinucleótido se llama "sonda de extensión" o "sonda". Una tercera secuencia de polinucleótidos que está opcionalmente presente (por ejemplo, en la ligación dependiente del molde) se denomina molde o secuencia objetivo. Cualquiera o más de las secuencias del cebador (sonda de inicialización), sonda (sonda de extensión) y/o molde pueden ubicarse en la misma cadena de polinucleótidos, o en diferentes cadenas de polinucleótidos. En un sistema de una cadena, el cebador, la sonda y el molde están todos en la misma cadena de polinucleótidos. En un ejemplo de un sistema de dos cadenas, la sonda o el cebador (pero no ambos) están en la misma cadena que el molde (por ejemplo, la hibridación entre la secuencia del molde y las secuencias del cebador o sonda forma una estructura de tallo-bucle u horquilla). En otro ejemplo de un sistema de dos cadenas, la sonda o el cebador están ambos en la misma cadena. En un sistema de tres cadenas, el molde, el cebador y la sonda están en cadenas de polinucleótidos separadas. Cualquier método de ligación que se describe en la presente puede realizarse en un sistema de una, dos o tres cadenas.

Opcionalmente, la ligación convierte una cadena polinucleotídica lineal en una cadena polinucleotídica circular (por ejemplo, en un sistema monocatenario a bicatenario). Opcionalmente, la ligación reduce el número de cadenas de polinucleótidos en uno (por ejemplo, en un sistema de dos o tres cadenas).

La ligación opcionalmente crea un enlace entre un nucleótido terminal de la sonda con el nucleótido terminal del cebador. Opcionalmente, el extremo terminal proximal del cebador y/o de la sonda está ligado. El nucleótido terminal del cebador puede ser el nucleótido 5' terminal y el nucleótido terminal de la sonda de extensión puede ser el nucleótido 3' terminal. Alternativamente, el nucleótido terminal del cebador puede ser el nucleótido 3' terminal y el nucleótido terminal de la sonda de extensión puede ser el nucleótido 5' terminal. El extremo 5' terminal de un polinucleótido, por ejemplo, tiene el quinto carbono en el anillo de azúcar de la desoxirribosa o ribosa en su terminal, opcionalmente con un grupo fosfato que se une a él, donde el grupo fosfato es capaz de formar un enlace fosfodiéster con un nucleótido 3' terminal. El nucleótido 3' terminal opcionalmente tiene un grupo 3'-hidroxilo que es capaz de formar un enlace fosfodiéster con un nucleótido 5' terminal. La ligación opcionalmente resulta en la formación de un enlace de fosfodiéster.

Cualquiera de los polinucleótidos, sin importar cómo se designen (por ejemplo, como "sondas" o "cebadores" o "moldes") puede ser de cualquier secuencia, cualquier longitud, en cualquier forma y de cualquier fuente. Los polinucleótidos pueden comprender una secuencia natural o ser altamente homólogos a una secuencia natural y/o derivarse de una secuencia natural. La secuencia natural puede ser cualquier porción de un gen, una secuencia reguladora, ADN genómico o fragmento, ADNc, ARN que incluye ARNm y ARNr, u otros. Los polinucleótidos pueden comprender opcionalmente cualquier secuencia artificial también. El polinucleótido puede derivarse u obtenerse de una muestra tal como una muestra de diagnóstico. El polinucleótido puede ser un producto secundario de una reacción, por ejemplo, un producto de ligación de una reacción de ligación o un ensayo como los que se describen en la presente, una sonda que se extiende de una reacción de la PCR o un producto de amplificación de la PCR ("amplicón"), el producto de una reacción de escisión invasiva, etc. El polinucleótido puede tener un 5' fosfato, o alternativamente

puede carecer de un 5' fosfato.

El producto de ligación de cualquiera de las reacciones puede someterse opcionalmente a reacciones adicionales de ligación y/o no ligación. Por ejemplo, el producto de ligación puede usarse como cebador (sonda de inicialización) o sonda de extensión y/o molde en una ligación posterior. También, por ejemplo, puede usarse como molde y/o cebador para una reacción de extensión de polimerasa, como en la PCR. La sonda, el cebador, el molde y/o el producto de ligación pueden someterse opcionalmente a una o más modificaciones antes o después de la ligación. Por ejemplo, la sonda, el cebador, el molde y/o el producto de ligación pueden escindirse enzimática o químicamente (por ejemplo, en enlaces de escisión), pueden tratarse con exo o endonucleasas, quinasas, fosfatasas, etc. Los extremos de un producto bicatenario pueden ser de punta roma o rellenada, con caperuza o adenilada, etc.

Opcionalmente, la sonda no tiene más de 20 nucleótidos consecutivos de largo, por ejemplo, no más de 15, 12, 10, 8, 7 nucleótidos consecutivos, preferentemente no más de 6, 5, 4 3 o 2 nucleótidos consecutivos. Opcionalmente, la sonda tiene al menos 2, 3, 4, 5, 6 o 7 nucleótidos de longitud. En algunos casos, la sonda es de cualquiera de las longitudes mínimas que se especifican (por ejemplo, 2, 3, 4, 5, 6 o 7 nucleótidos de longitud) y no tiene más de 20, 15, 12, 10, 8, 7 nucleótidos, preferentemente no más de 6, 5, 4 3 o 2 nucleótidos. En algunos ejemplos, la sonda es un 2-mer, 3-mer, 4-mer, 5-mer, 6-mer, 7-mer, 8-mer, 9-mer, 10-mer, 11-mer, 12- mer, 13-mer, 14-mer, 15-mer o 20-mer, o cualquier combinación de tales oligonucleótidos.

Opcionalmente, el cebador no tiene más de 20 nucleótidos consecutivos de largo, por ejemplo, no más de 15, 12, 10, 8, 7 nucleótidos consecutivos, preferentemente no más de 6, 5, 4, 3 o 2 nucleótidos consecutivos. En algunos ejemplos, el cebador es uno o más 2-mer, 3-mer, 4-mer, 5-mer, 6-mer, 7-mer, 8-mer, 9-mer, 10-mer, 11-mer, 12-mer, 13-mer, 14-mer, 15-mer o 20-mer, o cualquier combinación de tales oligonucleótidos.

Debe entenderse que la sonda o cebador (o molde, si está presente) puede "mezclarse" o "componerse", es decir, que comprende una mezcla de uno o más polinucleótidos de diferentes secuencias.

Opcionalmente, la ligación se realiza mediante el uso de la ligasa CV en combinación con una o más sondas de al menos 3 nucleótidos de longitud y no más de 6 o 5 nucleótidos de longitud. Opcionalmente, la una o más sondas tienen al menos 3 nucleótidos de longitud y no más de 4 nucleótidos de longitud. En otros ejemplos, la una o más sondas tienen al menos 4 nucleótidos de longitud y no más de 5 nucleótidos de longitud. Alternativamente, todas las sondas de una o más sondas pueden ser de 3-mers. De otra forma, todas las sondas de una o más sondas pueden ser de 4-mers.

Cuando la ligación se realiza mediante el uso de las ligasas DLX, DLXd, DLXd2, opcionalmente la ligación se realiza con una o más sondas que tienen al menos 2 o 3 nucleótidos de longitud y no más de 6 o 5 nucleótidos de longitud. Opcionalmente, la una o más sondas tienen al menos 3 nucleótidos de longitud y no más de 4 nucleótidos de longitud. En otros ejemplos, la una o más sondas tienen al menos 4 nucleótidos de longitud y no más de 5 nucleótidos de longitud. Alternativamente, todas las sondas de una o más sondas pueden ser de 3-mers. De otra forma, todas las sondas de una o más sondas pueden ser de 4-mers.

La ligación puede ser una ligación dependiente del molde. En la ligación dependiente de molde, la ligación entre una secuencia de cebador y una secuencia de sonda ocurre tras la hibridación de al menos una porción de una o ambas secuencias a una secuencia de molde. En algunos casos, ambas sondas deben hibridarse con el molde para que se produzca una ligación significativa. En un ejemplo típico, la ligación dependiente del molde no puede tener lugar a menos que ambos polinucleótidos se hibridan con la secuencia del molde. La porción del cebador o sonda que se hibrida con la secuencia objetivo generalmente tiene al menos dos nucleótidos de longitud. La porción que se hibrida opcionalmente no tiene más de 20 nucleótidos consecutivos de largo, por ejemplo no más de 15, 12, 10, 8, 7 nucleótidos consecutivos, preferentemente no más de 6, 5, 4, 3 o 2 nucleótidos consecutivos. La porción que se hibrida es opcionalmente una porción terminal del primer o segundo polinucleótido (por ejemplo, una porción que incluye el nucleótido 5' o 3' terminal). Por ejemplo, la porción que se hibrida puede consistir en los 2, 3, 4, 5, 6, 7, 8, 10, 15 o 20 nucleótidos terminales del extremo 5' o 3'.

Opcionalmente, la ligación ocurre cuando no hay presentes no coincidencias dentro de una o más porciones que se hibridan. En otros casos, la ligación ocurre cuando uno, dos o tres no coincidencias pueden estar presentes dentro de una o más porciones que se hibridan. En algunos casos, la ligación no ocurre cuando el nucleótido terminal y/o el segundo nucleótido terminal y/o el tercer nucleótido más terminal no coinciden. Como se mencionó, los nucleótidos terminales pueden ser los nucleótidos 5' o 3' terminales del polinucleótido.

Opcionalmente, el molde, si está presente, no tiene más de 11 nucleótidos de longitud, por ejemplo, no más de 10, 9, 8, 7, 6, 5 o 4 nucleótidos. Opcionalmente, el molde es uno o más N-mers, donde N es 4, 5, 6, 7, 8, 9, 10 u 11.

Opcionalmente, la ligación dependiente del molde de un ácido nucleico comprende: a) proporcionar un primer oligonucleótido que tiene menos de 6 nucleótidos; b) proporcionar un segundo oligonucleótido; c) acercar los extremos terminales 3' de uno de los oligonucleótidos primero y segundo con los extremos terminales 5' del otro oligonucleótido; y d) ligar el primer y el segundo oligonucleótidos. Opcionalmente, el primer oligonucleótido tiene una longitud de 5

nucleótidos. Opcionalmente, el primer oligonucleótido tiene una longitud de 4 nucleótidos. Opcionalmente, el primer oligonucleótido tiene una longitud de 3 nucleótidos. Opcionalmente, la ligación se realiza mediante el uso de una ligasa de huella pequeña (SFL). Opcionalmente, el segundo oligonucleótido incluye una secuencia complementaria a una porción de un ácido nucleico molde. Opcionalmente, el segundo oligonucleótido se hibrida con el ácido nucleico molde en la región de complementariedad. Opcionalmente, el primer oligonucleótido tiene una secuencia complementaria al ácido nucleico molde. Opcionalmente, el primer oligonucleótido se hibrida con el ácido nucleico molde en la región de complementariedad, y en donde un extremo terminal del primer oligonucleótido es adyacente a un extremo terminal del segundo oligonucleótido.

En algunas variaciones (por ejemplo, en la "ligación de muesca" o la ligación "dependiente del molde"), tanto el cebador como la sonda deben hibridarse adyacentes entre sí en el molde para que se produzca la ligación. Opcionalmente, la sonda y el cebador se hibridan adyacente y pueden ligarse solo cuando un nucleótido terminal del cebador se hibrida con un primer nucleótido del molde y un nucleótido terminal de la sonda de extensión se hibrida con un segundo nucleótido del molde, donde el primero y segundo nucleótidos en el molde no se separan por un nucleótido intermedio del molde. En otras modalidades, los nucleótidos intermedios pueden estar presentes entre el primer y segundo nucleótidos en el molde (opcionalmente, algunos nucleótidos, por ejemplo, no más de 1, 2, 3, 5, 10 o 15 nucleótidos). En tales modalidades, puede realizarse una etapa de "relleno de espacios" para extender el extremo 3' terminal de la sonda o la sonda antes de que pueda ligarse al extremo 5' terminal de la otra.

Opcionalmente, al menos una de las sondas, el molde (si la ligación depende del molde) y/o el cebador se inmoviliza mientras que otro de estos tres se marca. Por ejemplo, en la secuenciación de la ligación, el molde y o el cebador pueden inmovilizarse y la sonda puede marcarse.

Una sonda puede tener, por ejemplo, N residuos de nucleótidos de longitud, donde N es de 2 a 10, por ejemplo, 2, 3 o 4. N también puede ser menor que 6, por ejemplo, si el extremo terminal proximal de la sonda es su extremo 3' terminal.

Opcionalmente, la ligación es una ligación "hacia adelante" (es decir, la ligación del extremo 3' terminal de la sonda al extremo 5' terminal del cebador). Alternativamente, puede lograrse la ligación "inversa", donde el extremo 5' terminal de la sonda se liga al extremo 3' terminal del cebador.

Una sonda también puede tener una longitud N y puede comprender una porción proximal que se hibrida perfectamente con el molde y tiene una longitud de L nucleótidos, donde el nucleótido L+1 de la sonda no coincide con el molde. L puede ser, por ejemplo, de 2 a 8, y además puede ser menor que 6 si el extremo terminal proximal de la sonda es su extremo 3' terminal. En otras modalidades, L puede ser 2, 3, 4, 5, 6, 7 u 8.

La ligación puede repetirse al menos una vez, para cualquier número deseado de ciclos. Opcionalmente, cualquier producto de ligación de una reacción de ligación anterior se usa como cebador de una próxima ligación. Opcionalmente, todas las ligaciones extienden el producto de ligación en la misma dirección. Por ejemplo, todas las reacciones de ligación pueden ser ligaciones directas o todas pueden ser ligaciones inversas. En otras modalidades, algunas ligaciones pueden ser directas y algunas inversas. Opcionalmente, cualquier cebador que no se liga se vuelve no ligable antes de iniciar la siguiente ligación. Si así se desea (por ejemplo, en la secuenciación de la ligación), el método comprende además detectar si la sonda se ligó al cebador antes de repetir la siguiente reacción de ligación.

Cuando así se desee, cualquier producto de ligación de la reacción de ligación anterior puede usarse como molde de la siguiente reacción de ligación, por ejemplo, en una reacción en cadena de ligasa.

Opcionalmente, el extremo 5' terminal de la sonda de menos de 6 nucleótidos de largo se liga al extremo 3' terminal del cebador por una SFL como la ligasa CV. Por ejemplo, la sonda tiene 2, 3 o 4 nucleótidos de longitud. Si así se desea, el cebador también es un oligonucleótido corto. Por ejemplo, el cebador puede tener menos de 6 nucleótidos, por ejemplo, 3 o 4 nucleótidos de largo.

La ligación debe producir una cantidad significativa o detectable de producto de ligación. Opcionalmente, la eficiencia de la ligación es al menos del 5 %, a veces del 10 %, 20 %, 30 %, 50 %, 60 %, 70 %, 75 %, 80 %, 85 %, 90 % o 95 %. La eficiencia de la ligación (en términos porcentuales) puede considerarse en alguna modalidad como la porción porcentual del reactivo de ligación que se liga para formar el producto de ligación al final de la reacción de ligación. La eficiencia se determina opcionalmente después de que una reacción de ligación alcanzó el equilibrio, de modo que aumentar el tiempo de ligación no resultará en un aumento sustancial del producto de ligación que se forma. Generalmente, puede decirse que una reacción de ligación alcanzó el equilibrio después de 20 minutos. Por ejemplo, puede decirse que una reacción de ligación en la que se liga el 90 % del cebador o la sonda tiene una eficacia del 90 %. El reactivo de ligación que se usa para medir la eficiencia de la ligación es opcionalmente cualquier reactivo que esté en una concentración más baja que los demás. Opcionalmente, los otros reactivos y condiciones no son limitantes para la eficacia de la ligación (por ejemplo, otros reactivos están presentes en exceso o en una concentración que es al menos igual o mayor que el reactivo limitante).

En algunas modalidades, la SFL puede ligar una sonda corta que es más corta que N nucleótidos al menos un X %

tan eficientemente como la SFL puede ligar el N-mer correspondiente. N es opcionalmente 4, 5, 6, 7, 8, 10, 12, 15 o 20. En algunas modalidades, N es 6 o 7. X es opcionalmente al menos 30 %, 50 %, 60 %, 70 %, 80 % o 90 %. En algunas modalidades de la ligación dependiente del molde, la longitud de la sonda corta es de Y residuos, y los residuos proximales Y del N-mer más largo correspondiente son idénticos a la sonda, y los residuos NY distales del N-mer más largo correspondiente son perfectamente complementarios al molde. Y es, por ejemplo, 2, 3, 4, 5 o 6.

En algunas modalidades, la SFL puede ligar una sonda corta que es más corta que N nucleótidos cuando la sonda corta se conjuga con un colorante aproximadamente tan eficientemente como la SFL puede ligar la sonda no conjugada. N es opcionalmente 4, 5, 6, 7, 8, 10, 12, 15 o 20. En algunas modalidades, N es 6 o 7. En otras modalidades, la SFL puede ligar la sonda corta conjugada a un colorante al menos 30 %, 50 %, 60 %, 70 %, 80 % o 90 % tan eficientemente como la SFL puede ligar la sonda no conjugada. Los colorantes ejemplares incluyen al Cy5.

Se entiende que la eficiencia de la ligación (ya sea que se exprese en términos absolutos o relativos) puede aumentar o disminuir en dependencia de las condiciones de reacción exactas que se usan. Opcionalmente, la eficiencia de la ligación se mide cuando el cebador o la sonda están a una concentración de aproximadamente 10^{-9} a 10^{-4} M, por ejemplo a aproximadamente 10^{-8}, 10^{-7}, 10^{-6} o 10^{-5} M. En ciertas aplicaciones, por ejemplo, aplicaciones de secuenciación, una concentración de trabajo que se usa a menudo en el intervalo de 10^{-8} a 10^{-5}, por ejemplo, 10^{-7} o 10^{-6} M de sonda. Cuando se usa una mezcla de sondas, la concentración de la sonda es la concentración de solo aquellas sondas que pueden ligarse en esa reacción de ligación particular (por ejemplo, la(s) sonda(s) complementaria(s) al molde). Por lo tanto, otras sondas que no son capaces de participar en la reacción de ligación opcionalmente no se consideran al calcular la concentración de la sonda de interés. Opcionalmente, la concentración de la sonda está entre 1 picomolar y 1 milimolar, por ejemplo, aproximadamente 0,01-100 µM, por ejemplo, 1-10 µM, por ejemplo, 1-5 µM. En el caso de 2-mers, la concentración se incrementa opcionalmente a 10-1000 µM, por ejemplo aproximadamente 100 µM. Opcionalmente, la concentración de ligasa está entre 1 picomolar y 1 milimolar, por ejemplo, 1-2 micromolar. Opcionalmente, el ensayo de ligación se realiza a 15-35 °C, por ejemplo 15 °C, 20 °C, 25 °C o 30 °C. Cuando se involucran dos o más reactivos y se especifica la concentración de un reactivo particular, los otros reactivos están opcionalmente en exceso y/o no a una concentración que es limitante para la ligación.

La ligación puede realizarse en condiciones in vitro que se han determinado experimentalmente como adecuadas u óptimas para la actividad de la ligasa. Preferentemente, las condiciones de reacción que se eligen son (i) sustancialmente similares a las condiciones in vivo o fisiológicas en las que una forma natural de la ligasa que se usa es activa, o (ii) se ha determinado experimentalmente que resulta en una eficiencia de ligación que es comparable o mejor que la eficiencia que se obtiene mediante el uso de las de tipo (i). Si en la presente se especifican condiciones de ligación in vitro ejemplares para una SFL particular, entonces las condiciones de reacción sustancialmente similares son generalmente apropiadas para esa ligasa particular. En otras modalidades, las condiciones son tales que el ensayo de ligación de referencia produce una ligación significativa o detectable en 30 minutos, en 10 minutos, en 1 minuto o en diez segundos. Otro ejemplo no limitante de una eficiencia significativa o detectable de la ligación genera en el intervalo de 100 pM de producto de ligación, opcionalmente aproximadamente 1000 pM o 10.000 pM.

Opcionalmente, la eficiencia relativa puede expresarse como el porcentaje de eficiencia relativa, que puede calcularse como A/B x 100, donde A es el porcentaje de reactivo de prueba (por ejemplo, la sonda) ligado en un ensayo de prueba y B es el porcentaje del reactivo de referencia (por ejemplo, la sonda de referencia) que se liga en un ensayo de referencia. Cuando la eficiencia relativa de la ligación se especifica en términos comparativos o relativos en comparación con un ensayo de ligación de referencia, está implícito que todos los demás reactivos y condiciones (por ejemplo, la temperatura, la concentración de todos los reactivos, el pH, la concentración de iones necesarios como Mg++ y Mn++, la concentración de cofactores necesarios tales como NAD y/o ATP, las sales, los tampones, las concentraciones molares de todos los reactivos, que incluyen la enzima, el molde, la sonda, el cebador, los oligonucleótidos, etc.) se mantienen idénticos. Por ejemplo, una condición de que una SFL puede ligar una sonda corta (por ejemplo, de menos de 6 nucleótidos) al menos un X % de manera tan eficiente como la SFL puede ligar un octanucleótido correspondiente, puede significar que los dos ensayos de ligación diferentes contienen todos los reactivos excepto para las sondas (por ejemplo, el cebador, el molde, las enzimas y cualquier otro reactivo) y todas las condiciones de reacción (por ejemplo, la temperatura, las concentraciones de reactivo, las concentraciones de cualquier otro reactivo, etc.) se mantienen iguales para fines prácticos.

Opcionalmente, la ligasa tiene una mejor eficiencia de ligación que la ligasa de ADN T4, por ejemplo, en cualquier método que se describe en la presente. En una modalidad, la eficiencia de la ligación es mayor que la de la ligasa de ADN T4 para la misma mezcla de sonda(s), cebador(s) y molde(s), en cualquier método de ligación y condiciones elegidas, que incluye cualquiera de las que se describen en la presente. La eficiencia de la ligación es, por ejemplo, al menos un 5 % mayor que la ligasa T4, opcionalmente al menos un 10 %, 15 % o 20 % mayor. El aumento en la eficiencia debe ser estadísticamente confiable y significativo, por ejemplo, con un intervalo de confianza de al menos 95 %, 99 %, 99,9 %, 99,99 % o 99,999999 %. En un ejemplo, la SFL muestra una mayor eficiencia que la ligasa de ADN T4 cuando se une una sonda de 8 nucleótidos o menos a un cebador en la dirección hacia adelante o hacia atrás. Las condiciones de ensayo de la ligasa, el molde, la sonda, el cebador y/o la ligación son, por ejemplo, las que se describen en la presente.

La ligasa opcionalmente tiene buena fidelidad de ligación. La fidelidad de la ligación puede evaluarse como el

porcentaje de eventos de ligación incorrectos para una combinación dada de ligasa, molde, cebador y sonda. Un evento de ligación incorrecto es aquel en el que la sonda o el cebador no son perfectamente complementarios al molde, o el producto de ligación no es perfectamente complementario al molde. En una modalidad, la fidelidad de la ligación es mayor que la de la ligasa de ADN de T4 para la misma mezcla de sonda(s), cebador(s) y molde(s), en cualquier método de ligación y condiciones elegidas, que incluye cualquiera de las que se describen en la presente. La fidelidad de la ligación es, por ejemplo, al menos 5 % mayor que la ligasa T4, opcionalmente al menos 10 %, 15 % o 20 % mayor. El aumento en la fidelidad debe ser estadísticamente confiable y significativo, por ejemplo, con un intervalo de confianza de al menos 95 %, 99 %, 99,9 %, 99,99 % o 99,999999 %. En un ejemplo, la SFL muestra una mayor fidelidad que la ligasa de ADN de T4 cuando se une una sonda de 8 nucleótidos o menos a un cebador en la dirección hacia adelante o hacia atrás. Las condiciones de ensayo de la ligasa, el molde, la sonda, el cebador y/o la ligación son, por ejemplo, las que se describen en la presente.

2) Ligasas de huella pequeña

Opcionalmente, la ligación enzimática de polinucleótidos se logra mediante una ligasa de huella pequeña (SFL). Por ejemplo, cualquier método de ligación que se proporciona en la presente puede usar una ligasa que se muestra en la Tabla 1A, 1B o 1C. Una SFL es una ligasa que puede ligar polinucleótidos cortos. Como se usa en la presente, el término "ligasa" pretende incluir cualquier fragmento o variante o derivado de esa ligasa. El fragmento o variante o derivado posee opcionalmente una o más actividades funcionales de una ligasa. Una SFL opcionalmente comprende un polipéptido que tiene una o más de las siguientes actividades: (1) ataque nucleofílico en ATP o NAD^+ que resulta en la liberación de PPi o NMN y la formación de un intermediario covalente de ligasa-adenilato; (2) transferir el adenilato al extremo 5' terminal de la cadena de ADN que termina en fosfato 5' para formar adenilato de ADN (por ejemplo, el oxígeno 5'-fosfato de la cadena de ADN ataca el fósforo de la ligasa-adenilato); y (3) formación de un enlace covalente que se une a los extremos del polinucleótido y liberación de AMP (por ejemplo, por el ataque del 3'-OH sobre el ADN-adenilato). Opcionalmente, la SFL puede mediar una o más de las siguientes transformaciones de enlace: de fosfoanhidruro (ATP) a fosforamidato (ligasa-adenilato); desde fosforamidato (ligasa-adenilato) a fosfoanhidruro (ADN-adenilato); o de fosfoanhidruro (ADN-adenilato) a fosfodiéster (ADN sellado). Por lo tanto, las SFL ejemplares pueden comprender una secuencia de polipéptidos que es homóloga o una variante de una secuencia de SFL que se conoce o cualquier parte de esta. Las SFL ejemplares tienen opcionalmente una identidad de secuencia de aminoácidos de al menos 70 %, opcionalmente al menos 85 %, opcionalmente al menos 90, 95 %, 97 % o 99 %, con una ligasa conocida o SFL conocida.

Los ejemplos representativos de SFL incluyen la ligasa CV, DLX, DLXd, DLXd2 y la ligasa MnM. Una SFL preferente es la ligasa del virus Chlorella. Se identifican algunas ligasas ejemplares y sus números GI o de acceso se proporcionan en la Tabla 1A a continuación:

Tabla 1A

PRK08224		
Organismo	**Nombre de la proteína**	**Acceso**
B. Acidobacterias		
Candidatus Koribacter versatilis Ellin345 Candidatus Koribacter (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_592504
Candidatus Solibacter usitatus Ellin6076 Candidatus Solibacter (1 proteínas)		
	Ligasa de ADN dependiente de ATP	YP_826317
C. Actinobacterias		
Acidothermus cellulolyticus 11BAcidothermus (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_873134
Actinosynnema mirum DSM 43827 Actinosynnema (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_003099374
Arthrobacter aurescens TClArthrobacter (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_949544
Arthrobacter chlorophenolicus A6Arthrobacter (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_002489901

Arthrobacter sp. FB24Arthrobacter (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_833558
C. Actinobacterias		
Beutenbergia cavernae DSM 12333 Beutenbergia (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_002880505
Catenulispora acidiphila DSM 44928Catenulispora (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_003116519
Catenulispora acidiphila DSM 44928Catenulispora (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_003116565
Frankia alni ACN14aFrankia (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_712338
Frankia sp. EANlpecFrankia (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_001509433
Kineococcus radiotolerans SRS30216Kineococcus (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_001360406
Kytococcus sedentarius DSM 20547 Kytococcus (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_003149340
Mycobacterium abscessus ATCC 19977 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_001701033
Mycobacterium avium 104 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_879648
Mycobacterium avium subsp. paratuberculosis K-10 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	NP_959275
Mycobacterium gilvum PYR-GCK Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_001132524
Mycobacterium gilvum PYR-GCK Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_001132543
Mycobacterium marinum MMycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_001853525
Mycobacterium smegmatis str. MC2 155 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_890520
Mycobacterium smegmatis str. MC2 155 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_890521
Mycobacterium sp. JLS Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_001073538

Mycobacterium sp. JLS Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_001073546
Mycobacterium sp. JLS Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_001073574
Mycobacterium ulcerans Agy99 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_907815
Mycobacterium vanbaalenii PYR-1 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_956315
Mycobacterium vanbaalenii PYR-1 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_956321
Mycobacterium tuberculosis H37RaMycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_001285120
Mycobacterium tuberculosis H37Rv Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	NP_218248
Mycobacterium bovis AF2122/97 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	NP_857396
Mycobacterium bovis BCG str. Pasteur 1173P2 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_979871
Mycobacterium bovis BCG str. Tokyo 172 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_002646832
Mycobacterium tuberculosis CDC1551 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	NP_338389
Mycobacterium tuberculosis F11 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_001289690
Mycobacterium tuberculosis KZN 1435 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_003033779
Mycobacterium sp. KMS Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_940984
Mycobacterium sp. MCS Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_642076
Mycobacterium sp. KMS Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_941006
Mycobacterium sp. MCS Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_642099
Nakamurella multipartita DSM 44233 Nakamurella (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_003200226

Nocardia farcinica IFM 10152 Nocardia (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_118771
Nocardioides sp. JS614 Nocardioides (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_922117
Rhodococcus erythropolis PR4 Rhodococcus (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_002768423
Rhodococcus jostii RHA1 Rhodococcus (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_705046
Rhodococcus opacus B4 Rodococo (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_002782360
Saccharopolyspora erythraea NRRL 2338Saccharopolyspora (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_001104098
Salinispora arenicola CNS-205 Salinispora (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_001536378
Salinispora tropica CNB-440 Salinispora (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_001158390
Streptomyces avermitilis MA-4680 Streptomyces (3 proteínas)	Ligasa de ADN dependiente de ATP	NP_822873
Streptomyces coelicolor A3 (2) Streptomyces (3 proteínas)	Ligasa de ADN dependiente de ATP	NP_630780
Streptomyces griseus subsp. griseus NBRC 13350 Streptomyces (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_001822536
F. Chlamydiae / Verrucomicrobia		
Opitutus terrae PB90-1Opitutus (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_001821013
N.Alphaproteobacteria		
Bradyrhizobium japonicum USDA 110Bradyrhizobium (3 proteínas)	Ligasa de ADN dependiente de ATP	NP_771853
Bradyrhizobium sp. BTAilBradyrhizobium (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_001236299
Bradyrhizobium sp. ORS278Bradyrhizobium (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_001202307
Chelativorans sp. BNC1Chelativorans (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_675242
Mesorhizobium loti MAFF303099 Mesorhizobium (2 proteínas)	Ligasa de ADN dependiente de ATP	NP_108245

Mesorhizobium loti MAFF303099 (plásmido) Mesorhizobium (2 proteínas)	Ligasa de ADN dependiente de ATP	NP_109531
Methylocella silvestris BL2 Methylocella (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_002363964
Nitrobacter hamburgensis X14 Nitrobacter (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_579055
Phenylobacterium zucineum HLK1 Phenylobacterium (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_002131547
Rhizobium leguminosarum bv. trifolii WSM2304 (plásmido) Rhizobium (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_002279148
Rhizobium sp. NGR234 (plásmido) Rhizobium (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_002823307
	Ligasa de ADN dependiente de ATP	
Sinorhizobium medicae WSM419 (plásmido) Sinorhizobium (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_001312861
Sinorhizobium meliloti 1021 (plásmido) Sinorhizobium (2 proteínas)	Ligasa de ADN dependiente de ATP	NP_436551
P.Deltaproteobacteria		
Anaeromyxobacter dehalogenans 2CP-1 Anaeromyxobacter (5 proteínas)	Ligasa de ADN dependiente de ATP	YP_002491286
P.Deltaproteobacteria		
Anaeromyxobacter dehalogenans 2CP-CAnaeromyxobacter (5 proteínas)	Ligasa de ADN dependiente de ATP	YP_464028
Anaeromyxobacter sp. Fw109-5 Anaeromyxobacter (5 proteínas)	Ligasa de ADN dependiente de ATP	YP_001378773
Anaeromyxobacter sp. Fw109-5 Anaeromyxobacter (5 proteínas)	Ligasa de ADN dependiente de ATP	YP_001381200
Anaeromyxobacter sp. KAnaeromyxobacter (5 proteínas)	Ligasa de ADN dependiente de ATP	YP_002133229
PRK09125		
Organismo	**Nombre de la proteína**	**Acceso**
O.Betaproteobacteria		
Acidovorax sp. JS42Acidovorax (1 proteínas)	Ligasa de ADN	YP_986978

Aromatoleum aromaticum EbN1 Aromatoleum (1 proteínas)	Ligasa de ADN	YP_161050
Azoarcus sp. BH72Azoarcus (1 proteínas)	Ligasa de ADN	YP_934633
Candidatus Accumulibacter phosphatis clade IIA str. UW-1 Candidatus Accumulibacter (1 proteínas)	Ligasa de ADN	YP_003169249
Dechloromonas aromatica RCB Dechloromonas (1 proteínas)	Ligasa de ADN	YP_284461
Diaphorobacter sp. TPSYDiaphorobacter (1 proteínas)	Ligasa de ADN	YP_002553689
Herminiimonas arsenicoxydansHerminiimonas (1 proteínas)	Ligasa de ADN	YP_001100009
Leptothrix cholodnii SP-6 Leptothrix (1 proteínas)	Ligasa de ADN	YP_001791742
Methylibium petroleiphilum PM1Methylibium (1 proteínas)	Ligasa de ADN	YP_001020556
Neisseria gonorrhoeae FA 1090 Neisseria (7 proteínas)	Ligasa de ADN	YP_209054
Neisseria gonorrhoeae NCCP11945Neisseria (7 proteínas)	Ligasa de ADN	YP_002002827
PRK09125		
Organismo	Nombre de la proteína	**Acceso**
O.Betaproteobacteria		
Neisseria meningitidis 053442 Neisseria (7 proteínas)	Ligasa de ADN	YP_001598310
Neisseria meningitidis FAM18 Neisseria (7 proteínas)	Ligasa de ADN	YP_975951
Neisseria meningitidis MC58Neisseria (7 proteínas)	Ligasa de ADN	NP_275038
Neisseria meningitidis Z2491 Neisseria (7 proteínas)	Ligasa de ADN	YP_002341892
Neisseria meningitidis alphal4 Neisseria (7 proteínas)	Ligasa de ADN	YP_003082363
Polaromonas naphthalenivorans CJ2 Polaromonas (2 proteínas)	Ligasa de ADN	YP_982249
Polaromonas sp. JS666 Polaromonas (2 proteínas)	Ligasa de ADN	YP_549233
Rhodoferax ferrireducens T118 Rhodoferax (1 proteínas)	Ligasa de ADN	YP_522700
Thauera sp. MZ1TThauera (1 proteínas)	Ligasa de ADN	YP_002353773

Thiobacillus denitrificans ATCC 25259 Thiobacillus (1 proteínas)	Ligasa de ADN	YP_314570
Variovorax paradoxus S110Variovorax (1 proteínas)	Ligasa de ADN	YP_002944627
Verminephrobacter eiseniae EF01-2 Verminephrobacter (1 proteínas)	Ligasa de ADN	YP_998235
P.Deltaproteobacteria		
Desulfobacterium autotrophicum HRM2Desulfobacterium (1 proteínas)	LigA2	YP_002604477
Myxococcus xanthus DK 1622 Myxococcus (1 proteínas)	Ligasa de ADN	YP_628883
QE psilonproteobacteria		
Arcobacter butzleri RM4018 Arcobacter (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_001489632
Campylobacter concisus 13826 Campylobacter (10 proteínas)	Ligasa de ADN dependiente de ATP	YP_001467307
Campylobacter curvus 525.92 Campylobacter (10 proteínas)	Ligasa de ADN dependiente de ATP	YP_001407695
Campylobacter fetus subsp. feto 82-40 Campylobacter (10 proteínas)	Ligasa de ADN dependiente de ATP	YP_892536
Campylobacter hominis ATCC BAA-381Campylobacter (10 proteínas)	Ligasa de ADN dependiente de ATP	YP_001406356
Campylobacter jejuni RM1221 Campylobacter (10 proteínas)	Ligasa de ADN dependiente de ATP	YP 179811
Campylobacter jejuni subsp. doylei 269.97 Campylobacter (10 proteínas)	Ligasa de ADN dependiente de ATP	YP_001398949
Campylobacter jejuni subsp. jejuni 81-176 Campylobacter (10 proteínas)	Ligasa de ADN dependiente de ATP	YP_001001312
Campylobacter jejuni subsp. jejuni 81116Campylobacter (10 proteínas)	Ligasa de ADN dependiente de ATP	YP_001483144
Campylobacter jejuni subsp. jejuni NCTC 11168 Campylobacter (10 proteínas)	Ligasa de ADN dependiente de ATP	YP_002345037
Campylobacter lari RM2100Campylobacter (10 proteínas)	Ligasa de ADN dependiente de ATP	YP_002574655

Sulfurimonas denitrificans DSM 1251Sulfurimonas (1 proteínas)	Ligasa de ADN	YP_393098
R.Gammaproteobacteria		
Actinobacillus succinogenes 130ZActinobacillus (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_001344487
Aggregatibacter actinomycetemcomitans D11 S-1Aggregatibacter (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_003256304
Aggregatibacter aphrophilus NJ8700 Agregatibacter (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_003007537
Alcanivorax borkumensis SK2Alcanivorax (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_694422
Aliivibrio salmonicida LFI1238 Aliivibrio (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_002262821
Vibrio fischeri ES1 14Aliivibrio (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_204833
Vibrio fischeri MJ11Aliivibrio (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_002156265
Alteromonas macleodii 'Ecotipo profundo' Alteromonas (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_002127707
Mannheimia succiniciproducens MBEL55EBasfia (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_088131
Colwellia psychrerythraea 34HColwellia (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_271053
Haemophilus influenzae 86-028NPHaemophilus (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_248841
Haemophilus influenzae PittEEHaemophilus (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_001290961
Haemophilus influenzae PittGGHaemophilus (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_001293088
Haemophilus somnus 129PTHistophilus (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_719536
Haemophilus somnus 2336Histophilus (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_001783642
Idiomarina loihiensis L2TRIdiomarina (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_156435
Marinobacter aquaeolei VT8 Marinobacter (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_960951

Marinomonas sp. MWYL1Marinomonas (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_001341226
Photobacterium profundum SS9Photobacterium (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_132765
Pseudoalteromonas atlantica T6c Pseudoalteromonas (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_659659
Pseudoalteromonas haloplanktis TAC125 Pseudoalteromonas (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_340675
Psychromonas ingrahamii 37Psychromonas (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_942593
Shewanella amazonensis SB2BShewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_927870
Shewanella baltica OS155 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_001050227
Shewanella baltica OS185 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_001366044
Shewanella baltica OS195 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_001554317
Shewanella baltica OS223 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_002358357
Shewanella frigidimarina NCIMB 400 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_750174
Shewanella halifaxensis HAW-EB4 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_001673966
Shewanella loihica PV-4 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_001093713
Shewanella oneidensis MR-1 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	NP_717802
Shewanella pealeana ATCC 700345 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_001502366
Shewanella piezotolerans WP3 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_002312387
Shewanella sediminis HAW-EB3 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_001474374
Shewanella sp. ANA-3 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_870036

Shewanella sp. MR-4 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_734321
Shewanella sp. MR-7 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_738313
Shewanella woodyi ATCC 51908 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_001760369
Shewanella putrefaciens CN-32 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_001183272
Shewanella sp. W3-18-1 Shewanella (18 proteínas)	Ligasa de ADN dependiente de ATP	YP_963655
Thiomicrospira crunogena XCL-2Thiomicrospira (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_391938
Vibrio cholerae MJ-1236 Vibrio (9 proteínas)	Ligasa de ADN dependiente de ATP	YP_002878565
Vibrio parahaemolyticus RIMD 2210633Vibrio (9 proteínas)	Ligasa de ADN	NP_797856
Vibrio splendidus LGP32Vibrio (9 proteínas)	Ligasa de ADN	YP_002417130
Vibrio vulnificus CMCP6Vibrio (9 proteínas)	Ligasa de ADN	NP_761477
Vibrio vulnificus YJ016 Vibrio (9 proteínas)	Ligasa de ADN	NP_934427
Vibrio cholerae M66-2Vibrio (9 proteínas)	Ligasa de ADN	YP_002810248
Vibrio cholerae O1 biovar El Tor str. N16961Vibrio (9 proteínas)	Ligasa de ADN	NP_231182
Vibrio cholerae O395Vibrio (9 proteínas)	Ligasa de ADN	YP_001217094
Vibrio cholerae O395Vibrio (9 proteínas)	Ligasa de ADN	YP_002819900
PHA0454		
Organismo	**Nombre de la proteína**	**Acceso**
b. Virus		
Virus similares al fago 13aT7 de enterobacterias (16 proteínas)	Ligasa de ADN dependiente de ATP	YP_002003941

Virus similares al fago BA14T7 de enterobacterias (16 proteínas)	Ligasa de ADN dependiente de ATP	YP_002003458
Virus similares al fago EcoDS1T7 de enterobacterias (16 proteínas)	Ligasa de ADN dependiente de ATP	YP_002003747
Virus similares al fago K1FT7 de enterobacterias (16	Ligasa de ADN dependiente de ATP	YP_338096
Virus similares al fago T3T7 de enterobacterias (16 proteínas)	Ligasa de ADN dependiente de ATP	NP_523305
Virus similares al fago T7T7 de enterobacterias (16 proteínas)	Ligasa de ADN dependiente de ATP	NP_041963
Virus similares al fago K11T7 de Klebsiella (16 proteínas)	Ligasa de ADN dependiente de ATP	YP_002003797
Virus similares al fago KvplT7 de Kluyvera (16 proteínas)	Ligasa de ADN	YP_002308390
Virus similares al fago MmP1T7 de Morganella (16 proteínas)	Ligasa de ADN dependiente de ATP	YP_002048633
Virus similares al fago gh-1T7 de Pseudomonas (16 proteínas)	Ligasa de ADN dependiente de ATP	NP_813751
Virus similares al fago phiSG-JL2T7 de Salmonella (16 proteínas)	Ligasa de ADN dependiente de ATP	YP_001949754
Virus similares al Vibriofago VP4T7 (16 proteínas)	Ligasa de ADN dependiente de ATP	YP_249578
Virus similares al fago phiAl 122T7 de Yersinia pestis (16 proteínas)	Ligasa de ADN dependiente de ATP	NP_848267
Virus similares al fago BerlinT7 de Yersinia (16 proteínas)	Ligasa de ADN dependiente de ATP	YP_918989
Virus similares al fago Yepe2T7 de Yersinia (16 proteínas)	Ligasa de ADN dependiente de ATP	YP_002003318
PHA0454		
Organismo	**Nombre de la proteína**	**Acceso**
b. Virus		
Virus similares al fago phiYeO3-12T7 de Yersinia (16 proteínas)	Ligasa de ADN dependiente de ATP	NP_052075
Virus similares al fago LKAlphiKMV de enterobacterias (7 proteínas)	Ligasa de ADN dependiente de ATP	YP_001522868

Virus similares al fago phiKMVphiKMV de enterobacterias (7 proteínas)	Ligasa de ADN dependiente de ATP	NP_877456
Virus similares al fago LKD16phiKMV de Pseudomonas (7 proteínas)	Ligasa de ADN dependiente de ATP	YP_001522807
Virus similares al fago LUZl9phiKMV de Pseudomonas (7 proteínas)	Ligasa de ADN dependiente de ATP	YP_001671961
Virus similares al fago PT2phiKMV de Pseudomonas (7 proteínas)	Ligasa de ADN dependiente de ATP	YP_002117800
Virus similares al fago PT5phiKMV de Pseudomonas (7 proteínas)	Ligasa de ADN dependiente de ATP	YP_002117741
Virus similares al fago phikF77phiKMV de Pseudomonas (7 proteínas)	Ligasa de ADN dependiente de ATP	YP_002727838
CLSZ2445448		
Organismo	**Nombre de la proteína**	**Acceso**
a.Eucariota		
Paramecium tetraurelia cepa d4-2 Paramecium (5 proteínas)	Ligasa de ADN	XP_001347270
Paramecium tetraurelia cepa d4-2 Paramecium (5 proteínas)	proteína hipotética	XP_001422985
Paramecium tetraurelia cepa d4-2 Paramecium (5 proteínas)	proteína hipotética	XP_001431968
Paramecium tetraurelia cepa d4-2 Paramecium (5 proteínas)	proteína hipotética	XP_001435874
CLSZ2445448		
Organismo	**Nombre de la proteína**	**Acceso**
a.Eucariota		
Paramecium tetraurelia cepa d4-2 Paramecium (5 proteínas)	proteína hipotética	XP_001460273
Tetrahymena thermophila Tetrahymena (1 proteínas)	Ligasa de ADN dependiente de ATP	XP_001011861
CLSP2344013		
Organismo	**Nombre de la proteína**	**Acceso**
b. Virus		

Enterobacterias fago Felix 01 Myoviridae sin clasificar (4 proteínas)	Ligasa de ADN putativa	NP 944942
Enterobacterias fago WV8 Myoviridae sin clasificar (4 proteínas)	proteína hipotética	YP_002922879
Erwinia fago phiEa21-4 Myoviridae sin clasificar (4 proteínas)	ligasa de ADN putativa	YP_002456101
Escherichia fago rv5 Myoviridae sin clasificar (4 proteínas)	Ligasa de ADN dependiente de ATP	YP_002003586
PRK07636		
Organismo	**Nombre de la proteína**	**Acceso**
J.Firmicutes		
Bacilo clausii KSM-K16 Bacilo	Ligasa de ADN dependiente de ATP	YP_176304
Bacillus subtilis subsp. subtilis str. 168Bacillus	Ligasa de ADN dependiente de ATP	NP_389932
Bacillus licheniformis ATCC 14580 Bacilo	Ligasa de ADN dependiente de ATP	YP_078721
Bacillus licheniformis ATCC 14580 Bacilo	Ligasa de ADN dependiente de ATP	YP_091132
Geobacillus sp. Y412MC10Geobacillus	Ligasa de ADN dependiente de ATP	YP_003240778
Paenibacillus sp. JDR-2 Paenibacillus	Ligasa de ADN dependiente de ATP	YP_003009892
CLSK2551528		
Organismo	**Nombre de la proteína**	**Acceso**
J.Firmicutes		
Geobacillus sp. Y412MC10Geobacillus (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_003245332
Paenibacillus sp. JDR-2 Paenibacillus (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_003013681
CLSK2532515		
Organismo	**Nombre de la proteína**	**Acceso**
B. Acidobacterias		
Candidatus Solibacter usitatus Ellin6076 Candidatus Solibacter (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_829024

E.Bacteroides / Chlorobi		
Flavobacteriaceae bacteria 3519-10 Flavobacteriaceae sin clasificar (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_003095681
CLSK2470953		
Organismo	**Nombre de la proteína**	**Acceso**
C. Actinobacterias		
Arthrobacter chlorophenolicus A6 (plásmido) Arthrobacter (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_002478051
Arthrobacter chlorophenolicus A6 (plásmido) Arthrobacter (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_002478427
Nocardioides sp. JS614 Nocardioides (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_923463
CLSK2469924		
Organismo	**Nombre de la proteína**	**Acceso**
J.Firmicutes		
Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446 Alliciclobacilo (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_003185050
Brevibacillus brevis NBRC 100599 Brevibacillus (1 proteínas)	ligasa de ADN dependiente de ATP putativa	YP_002773127
CLSK2340991		
Organismo	**Nombre de la proteína**	**Acceso**
N.Alphaproteobacteria		
Phenylobacterium zucineum HLK1 (plásmido) Phenylobacterium (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_002128561
Phenylobacterium zucineum HLK1 (plásmido) Phenylobacterium (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_002128631
CLSK2333706		
Organismo	**Nombre de la proteína**	**Acceso**
J.Firmicutes		
Candidatus Desulforudis audaxviator MP104CCandidatus Desulforudis (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_001716762

Natranaerobius thermophilus JW / NM-WN-LF Natranaerobius (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_001916325
CLSK2303611		
Organismo	**Nombre de la proteína**	**Acceso**
C. Actinobacterias		
Streptomyces coelicolor A3 (2) Streptomyces (2 proteínas)	Ligasa de ADN dependiente de ATP	NP 631399
Streptomyces griseus subsp. Griseus NBRC 13350 Streptomyces (2 proteínas)	ligasa de ADN dependiente de ATP putativa	YP_001828202
CLSK962101		
Organismo	**Nombre de la proteína**	**Acceso**
C. Actinobacterias		
Nocardioides sp. JS614 Nocardioides (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_922436
Salinispora arenicola CNS-205 Salinispora (2 proteínas)	Región de ligasa LigD de ADN polimerasa	YP_001539124
Salinispora tropica CNB-440 Salinispora (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_001160776
CLSK915249		
Organismo	**Nombre de la proteína**	**Acceso**
C. Actinobacterias Véase CLSK2303611 arriba		
Streptomyces avermitilis MA-4680 (plásmido) Streptomyces (2 proteínas)	ligasa de ADN dependiente de ATP putativa	NP_828839
Streptomyces sp. HK1 (plásmido) Streptomyces (2 proteínas)	ligasa de ADN dependiente de ATP putativa	YP_001661618
CLSK899085		
Organismo	**Nombre de la proteína**	**Acceso**
O.Betaproteobacteria		
Burkholderia cenocepacia HI2424 (plásmido) Burkholderia (3 proteínas)	Ligasa de ADN dependiente de ATP	YP_840498
Burkholderia cenocepacia J2315 (plásmido) Burkholderia (3 proteínas)	ligasa putativa	YP_002235530

Burkholderia multivorans ATCC 17616 (plásmido) Burkholderia (3 proteínas)	Polimerasa de ADN LigD subunidad ligasa	YP_001573706
R.Gammaproteobacteria		
Pseudomonas putida (plásmido) Pseudomonas (1 proteínas)	ligasa putativa	NP_542805
CLSK862724		
Organismo	**Nombre de la proteína**	**Acceso**
A.Archaea		
Archaeoglobus fulgidus DSM 4304 Archaeoglobus (1 proteínas)	Ligasa de ADN, putativa	NP_070553
J.Firmicutes		
Desulfotomaculum reduce MI-1 Desulfotomaculum (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_001113345
Moorella thermoacetica ATCC 39073 Moorella (1 proteínas)	Ligasa de ADN dependiente de ATP, central	YP_430340
Pelotomaculum thermopropionicum SIPelotomaculum (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_001211793
Thermoanaerobacter pseudethanolicus ATCC 33223 Termoanaerobacter (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_001664477
Thermoanaerobacter sp. X514 Termoanaerobacter (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_001662589
CLSK820690		
Organismo	**Nombre de la proteína**	**Acceso**
A.Archaea		
arqueón metanogénico no cultivado RC-Muestras ambientales (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_686457
C. Actinobacterias		
Mycobacterium avium 104 Mycobacterium (2 proteínas)	Polimerasa de ADN LigD subunidad ligasa	YP_882332
Mycobacterium avium subsp. paratuberculosis K-10 Mycobacterium (2 proteínas)	proteína hipotética	NP_960263

Saccharopolyspora erythraea NRRL 2338Saccharopolyspora (1 proteínas)	Ligasa de ADN, dependiente de ATP	YP_001107793
N.Alphaproteobacteria		
N.Alphaproteobacteria		
Bradyrhizobium japonicum USDA 110Bradyrhizobium (2 proteínas)	Ligasa de ADN	NP_774671
Bradyrhizobium sp. BTAilBradyrhizobium (2 proteínas)	ligasa de ADN dependiente de ATP putativa	YP_001243518
CLSK808255		
Organismo	**Nombre de la proteína**	**Acceso**
N.Alphaproteobacteria		
Sinorhizobium medicae WSM419Sinorhizobium (2 proteínas)	Región de ligasa LigD de ADN polimerasa	YP_001326990
Sinorhizobium meliloti 1021 (plásmido) Sinorhizobium (2 proteínas)	proteína ligasa de ADN dependiente de ATP putativa	NP_437750
CLSK806855		
Organismo	**Nombre de la proteína**	**Acceso**
N.Alphaproteobacteria		
Agrobacterium tumefaciens str. C58 (plásmido) Agrobacterium (3 proteínas)	Ligasa de ADN dependiente de ATP	NP_395985
Agrobacterium tumefaciens str. C58 (plásmido) Agrobacterium (3 proteínas)	Ligasa de ADN dependiente de ATP	NP_396032
Agrobacterium tumefaciens str. C58 (plásmido) Agrobacterium (3 proteínas)	Ligasa de ADN dependiente de ATP	NP_396609
Rhizobium etli CFN 42 (plásmido) Rhizobium (10 proteínas)	proteína de la Ligasa de ADN (ATP) putativa	YP_472413
Rhizobium etli CIAT 652 Rhizobium (10 proteínas)	probable proteína ligasa de ADN dependiente de ATP	YP_001977317
Rhizobium etli CIAT 652 (plásmido) Rhizobium (10 proteínas)	proteína ligasa de ADN dependiente de ATP putativa	YP_001985803
Rhizobium leguminosarum bv. trifolii WSM1325 (plásmido) Rhizobium (10 proteínas)	LigD de ADN polimerasa, proteína del dominio ligasa	YP_002973496

Rhizobium leguminosarum bv. trifolii WSM1325 (plásmido) Rhizobium (10 proteínas)	LigD de ADN polimerasa, proteína del dominio ligasa	YP_002984974
CLSK806855		
Organismo	**Nombre de la proteína**	**Acceso**
N.Alphaproteobacteria		
Rhizobium leguminosarum bv. trifolii WSM1325 (plásmido) Rhizobium (10 proteínas)	LigD de ADN polimerasa, proteína del dominio ligasa	YP_002984992
Rhizobium leguminosarum bv. trifolii WSM2304 Rhizobium (10 proteínas)	LigD de ADN polimerasa, proteína del dominio ligasa	YP_002281897
Rhizobium leguminosarum bv. trifolii WSM2304 (plásmido) Rhizobium (10 proteínas)	LigD de ADN polimerasa, proteína del dominio ligasa	YP_002278005
Rhizobium leguminosarum bv. viciae 3841 (plásmido) Rhizobium (10 proteínas)	ligasa de ADN putativa	YP_764723
Rhizobium leguminosarum bv. viciae 3841 (plásmido) Rhizobium (10 proteínas)	ligasa de ADN putativa	YP_771149
Sinorhizobium meliloti (plásmido) Sinorhizobium (2 proteínas)	ligasa de ADN dependiente de ATP putativa	YP_001965531
Sinorhizobium meliloti 1021 (plásmido) Sinorhizobium (2 proteínas)	Ligasa de ADN dependiente de ATP	NP_435470
CLSK762775		
Organismo	**Nombre de la proteína**	**Acceso**
N.Alphaproteobacteria		
Rhodococcus jostii RHA1 (plásmido) Rhodococcus (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_708952
Rhodococcus opacus B4 (plásmido) Rhodococcus (2 proteínas)	ligasa de ADN dependiente de ATP putativa	YP_002776886
CLSK761995		
Organismo	**Nombre de la proteína**	**Acceso**
N.Alphaproteobacteria		
Nitrobacter hamburgensis X14 Nitrobacter (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_579015
Rhodopseudomonas palustris BisB5 Rodopseudomonas (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_569297
CLSK761995		

Organismo	Nombre de la proteína	Acceso
N.Alphaproteobacteria		
Rhodopseudomonas palustris TIE-1 Rhodopseudomonas (2 proteínas)	LigD de ADN polimerasa, proteína del dominio ligasa	YP_001991309
CLSK523944		
Organismo	**Nombre de la proteína**	**Acceso**
R.Gammaproteobacteria		
Pseudomonas fluorescens (plásmido) Pseudomonas (3 proteínas)	ligasa de ADN dependiente de ATP putativa	YP_002887417
Pseudomonas putida (plásmido) Pseudomonas (3 proteínas)	fragmento de ligasa putativa	NP_863069
Pseudomonas sp. ND6 (plásmido) Pseudomonas (3	Ligasa de ADN dependiente de ATP	NP_943185
CLSK390680		
Organismo	**Nombre de la proteína**	**Acceso**
R.Gammaproteobacteria		
Mesorhizobium loti MAFF303099 Mesorhizobium (3 proteínas)	Ligasa de ADN dependiente de ATP	NP_108227
Mesorhizobium loti MAFF303099 Mesorhizobium (3 proteínas)	proteína hipotética	NP_108282
Mesorhizobium loti MAFF303099 (plásmido) Mesorhizobium (3 proteínas)	Proteína similar a Ligasa de ADN	NP_109396

Un subconjunto de ligasas de interés se encuentra en la Tabla 1B a continuación.

Tabla 1B

PRK08224		
B. Acidobacterias		
Bacterias; Grupo de fibrobacteres / acidobacterias; Acidobacterias; Acidobacterias no clasificadas; Candidatus Koribacter; Candidatus Koribacter versatilis		
Candidatus Solibacter usitatus Ellin6076 Candidatus Solibacter (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_826317
C. Actinobacterias		
Bacterias; Actinobacterias; Actinobacterias (clase); Actinobacteridae; Actinomicetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium marinum		
Mycobacterium gilvum PYR-GCK Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_001132524
C. Actinobacterias		
Mycobacterium vanbaalenii PYR-1 Mycobacterium (26 proteínas)	Ligasa de ADN dependiente de ATP	YP_956315

Mycobacterium sp. MCS Mycobacterium (26 proteínas)	ADN dependiente de ATP Ligasa	YP_642076
F. Chlamydiae / Verrucomicrobia		
Bacterias; Grupo Chlamydiae / Verrucomicrobia; Verrucomicrobia; Opitutae Opitutales; Opitutaceae; Opituto; Opitutus terrae		
Opitutus terrae PB90-1Opitutus (1 proteínas)	ADN dependiente de ATP Ligasa	YP_001821013
PRK09125		
Organismo	**Nombre de la proteína**	**Acceso**
O.Betaproteobacteria		
Neisseria meningitidis Z2491 Neisseria (7 proteínas)	Ligasa de ADN	YP_002341892
Thiobacillus denitrificans ATCC 25259 Thiobacillus (1 proteínas)	Ligasa de ADN	YP 314570
Variovorax paradoxus S110Variovorax (1 proteínas)	Ligasa de ADN	YP_002944627
Verminephrobacter eiseniae EF01-2 Verminephrobacter (1 proteínas)	Ligasa de ADN	YP 998235
	-	-
P.Deltaproteobacteria		
Desulfobacterium autotrophicum HRM2Desulfobacterium (1 proteínas)	LigA2	YP_002604477
Myxococcus xanthus DK 1622 Myxococcus (1 proteínas)	Ligasa de ADN	YP 628883
	-	-
Q.Epsilonproteobacteria		
Campylobacter jejuni subsp. jejuni NCTC 11168 Campylobacter (10 proteínas)	ADN dependiente de ATP ligasa	YP_002345037
Sulfurimonas denitrificans DSM 1251 Sulfurimonas (1 proteínas)	Ligasa de ADN	YP_393098
	-	-
R.Gammaproteobacteria		
Aggregatibacter aphrophilus NJ8700 Agregatibacter (2 proteínas)	ADN dependiente de ATP ligasa	YP_003007537
Haemophilus influenzae PittEEHaemophilus (3 proteínas)	ADN dependiente de ATP ligasa	YP_001290961
Shewanella baltica OS195 Shewanella (18 proteínas)	ADN dependiente de ATP ligasa	YP_001554317
Shewanella loihica PV-4 Shewanella (18 proteínas)	ADN dependiente de ATP ligasa	YP_001093713
Vibrio cholerae M66-2Vibrio (9 proteínas)	Ligasa de ADN	YP_002810248
R.Gammaproteobacteria		

	-	-
PHA0454		
Organismo	**Nombre de la proteína**	**Acceso**
b.Virus		
Virus; virus dsDNA, sin etapa de ARN; Caudovirales; Podoviridae; Autographivirinae; virus similares a phiKMV		
Virus similares al fago LKD16phiKMV de Pseudomonas (7 proteínas)	Ligasa de ADN dependiente de ATP	YP_001522807
CLSZ2445448		
Organismo	**Nombre de la proteína**	**Acceso**
a.Eucariota		
Eucariota; Alveolata; Ciliophora; Intramacronucleata; Oligohymenophorea; Peniculida; Parameciidae; Paramecio; Paramecium tetraurelia		
Paramecium tetraurelia cepa d4-2 Paramecium (5 proteínas)	Ligasa de ADN	XP_001347270
PRK07636		
Organismo	**Nombre de la proteína**	**Acceso**
J.Firmicutes		
Bacterias; Firmicutes; Bacilos; Bacillales Bacillaceae; Bacilo; Bacillus clausii		
Bacillus subtilis subsp. subtilis str. 168Bacillus	Una Ligasa de ADN dependiente de TP	NP 389932
Bacterias; Firmicutes; Bacilos; Bacillales Bacillaceae; Geobacillus		
Geobacillus sp. Y412MC10Geobacillus	Ligasa de ADN dependiente de ATP	YP_003240778
CLSK2551528		
Organismo	**Nombre de la proteína**	**Acceso**
J.Firmicutes		
Bacterias; Firmicutes; Bacilos; Bacillales Bacillaceae; Geobacillus		
Geobacillus sp. Y412MC10Geobacillus (1 **proteínas)**	Ligasa de ADN dependiente de ATP	YP_003245332
	-	-
CLSK2470953		
Organismo	**Nombre de la proteína**	**Acceso**
C. Actinobacterias		
Bacterias; Actinobacterias; Actinobacterias (clase); Actinobacteridae; Actinomicetales; Micrococcineae; Micrococcaceae; Artrobacter; Arthrobacter chlorophenolicus		

CLSK2470953		
Organismo	**Nombre de la proteína**	**Acceso**
C. Actinobacterias		
Arthrobacter chlorophenolicus A6 (plásmido) Arthrobacter (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_002478427
CLSK2469924		
Organismo	**Nombre de la proteína**	**Acceso**
J.Firmicutes		
Bacterias; Firmicutes; Bacilos; Bacillales Alicyclobacillaceae; Alicyclobacillus; Alicyclobacillus acidocaldarius; Alicyclobacillus acidocaldarius subsp. acidocaldarius		
Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446Alicyclobacillus	Ligasa de ADN dependiente de ATP	YP_003185050
CLSK2340991		
Organismo	**Nombre de la proteína**	**Acceso**
N.Alphaproteobacteria		
Bacterias; Proteobacterias; Alfaproteobacterias; Caulobacterales; Caulobacteraceae; Fenilobacteria; Phenylobacterium zucineum		
Phenylobacterium zucineum HLK1 (plásmido) Phenylobacterium (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_002128631
	-	-
CLSK2333706		
Organismo	**Nombre de la proteína**	**Acceso**
J.Firmicutes		
Bacterias; Firmicutes; Clostridios; Clostridiales; Peptococcaceae; Candidatus Desulforudis; Candidatus Desulforudis audaxviator		
Candidatus Desulforudis audaxviator MP104CCandidatus Desulforudis (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_001716762
	-	-
CLSK962101		
Organismo	**Nombre de la proteína**	**Acceso**
C. Actinobacterias		
Bacterias; Actinobacterias; Actinobacterias (clase); Actinobacteridae; Actinomicetales; Micromonosporineae; Micromonosporaceae; Salinispora; Salinispora arenicola		
Salinispora arenicola CNS-205 Salinispora (2 proteínas)	Región de ligasa LigD de ADN polimerasa	YP_001539124

Salinispora tropica CNB-440 Salinispora (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_001160776
	-	-
CLSK915249		
Organismo	**Nombre de la proteína**	**Acceso**
C. Actinobacterias Véase CLSK2303611 arriba		
Bacterias; Actinobacterias; Actinobacterias (clase); Actinobacteridae; Actinomicetales; Streptomycineae; Streptomycetaceae; Streptomyces; Streptomyces coelicolor		
Streptomyces avermitilis MA-4680 (plásmido) Streptomyces (2 proteínas)	ligasa de ADN dependiente de ATP putativa	NP 828839
Streptomyces sp. HK1 (plásmido) Streptomyces (2 proteínas)	ligasa de ADN dependiente de ATP putativa	YP_001661618
CLSK862724		
Organismo	**Nombre de la proteína**	**Acceso**
A.Archaea		
Arquea; Euryarchaeota; Archaeoglobi; Arqueoglobales; Archaeoglobaceae; Archaeoglobus; Archaeoglobus fulgidus		
Archaeoglobus fulgidus DSM 4304 Archaeoglobus (1 proteínas)	Ligasa de ADN, putativa	NP 070553
J.Firmicutes		
Pelotomaculum thermopropionicum SIPelotomaculum (1 proteínas)	Ligasa de ADN dependiente de ATP	YP_001211793
Thermoanaerobacter pseudethanolicus ATCC 33223 Termoanaerobacter (2 proteínas)	Ligasa de ADN dependiente de ATP	YP_001664477
CLSK820690		
Organismo	**Nombre de la proteína**	**Acceso**
A.Archaea		
Arquea; Euryarchaeota; muestras ambientales		
arqueón metanogénico no cultivado RC-Muestras ambientales (1 proteínas)	Ligasa de ADN dependiente de ATP	YP 686457
N.Alphaproteobacteria		
Bacterias; Proteobacterias; Alfaproteobacterias; Rizobiales; Bradyrhizobiaceae; Bradyrhizobium; Bradyrhizobium japonicum		

Bradyrhizobium japonicum USDA 110Bradyrhizobium (2 proteínas)	Ligasa de ADN	NP 774671
Bradyrhizobium sp. BTAilBradyrhizobium (2 proteínas)	ligasa de ADN dependiente de ATP putativa	YP_001243518
CLSK808255		
Organismo	**Nombre de la proteína**	**Acceso**
N.Alphaproteobacteria		
Bacterias; Proteobacterias; Alfaproteobacterias; Rizobiales; Rhizobiaceae; Grupo Sinorhizobium / Ensifer; Sinorhizobio; Sinorhizobium medicae		
Sinorhizobium medicae WSM419Sinorhizobium (2 proteínas)	Región de ligasa LigD de ADN polimerasa	YP_001326990
Sinorhizobium meliloti 1021 (plásmido) Sinorhizobium (2 proteínas)	proteína ligasa de ADN dependiente de ATP putativa	NP 437750
CLSK806855		
Organismo	**Nombre de la proteína**	**Acceso**
N.Alphaproteobacteria		
Bacterias; Proteobacterias; Alfaproteobacterias; Rizobiales; Rhizobiaceae; Grupo Rhizobium/Agrobacterium; Agrobacterium; Agrobacterium tumefaciens		
Agrobacterium tumefaciens str. C58 (plásmido) Agrobacterium (3 proteínas)	Ligasa de ADN dependiente de ATP	NP 396032
Rhizobium leguminosarum bv. trifolii WSM1325 (plásmido) Rhizobium (10 proteínas)	LigD de ADN polimerasa, proteína del dominio ligasa	YP_002973496
Rhizobium leguminosarum bv. trifolii WSM2304 (plásmido) Rhizobium (10 proteínas)	LigD de ADN polimerasa, proteína del dominio ligasa	YP_002278005
CLSK390680		
Organismo	**Nombre de la proteína**	**Acceso**
N.Alphaproteobacteria		
Bacterias; Proteobacterias; Alfaproteobacterias; Rizobiales; Phyllobacteriaceae; Mesorhizobio; Mesorhizobium loti		
Mesorhizobium loti MAFF303099 Mesorhizobium (3 proteínas)	proteína hipotética	NP_108282

Se identifican algunas ligasas ejemplares y sus números GI o de acceso se proporcionan en la Tabla 1C a continuación:

Tabla 1c

Ligasa de ADN CV, identificación de GenBank AAC96909.1, del virus de Paramecium bursaria Chlorella 1: MAITKPLLAATLENIEDVQFPCLATPKIDGIRSVKQTQMLSRTFKPIRNSVMNRLL TELLPEGSDGEISIEGATFQDTTSAVMTGHKMYNAKFSYYWFDYVTDDPLKKYI DRVEDMKNYITVHPHILEHAQVKIIPLIPVEINNITELLQYERDVLSKGFEGVMIRK PDGKYKFGRSTLKEGILLKMKQFKDAEATIISMTALFKNTNTKTKDNFGYSKRST HKSGKVEEDVMGSIEVDYDGWFSIGTGFDADQRRDFWQNKESYIGKMVKFKY FEMGSKDCPRFPVFIGIRHEEDR
Ligasa de ADN MnM, identificación de GenBank YP_333052.1, de Burkholderia pseudomallei 1710b (secuencia equivalente a ABA50091)
MSGVPYGFKPNLAATLTKPELIKFPVWASPKIDGIRCVFFGGVAYSRSLKPIPNPV VQEFAKAYANLLEGLDGELTVGSPTDANCMQNSMAVMSKAAAPDFTFHVFDW FHPAQAHIEFWQRSDWEDRIVQFYDRYPEVDIRAAPQVLCTSLAHLDTNEARW LADGYEGMMIRDHCGRYKFGRSTEREGGLVKVKRFTDAEAIVIGFEEEMHNANE AKRDATGRTERSTSKAGLHGKGTLGALWKNERGIVFNIGTGFTAAQRADYWA NHPSLFGKMVKFKHFDHGTVDAPRHPVFIGFRHPEDM
Ligasa de ADN Hin, identificación de GenBank P44121, de Haemophilus influenza MKFYRTLLLFFASSFAFANSDLMLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQ LLTRQGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEIS **T** ITKSFKGDGWEKLKLYV FDVPDAEGNLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGE GVWRNPNAPYERKRSSQILKLKTAR **G** EECTVIAHHKGKGQFENVMGALTCKN HRGEFKIGSGFNLNERENPPPIGSVITYKYRGITNSGKPRFATYWREKK
Ligasa de ADN DLX, ligasa artificial derivada de la ligasa de ADN Hin de Haemophilus influenza: MKFYRTLLLFFASSFAFANSDLMLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQ LLTRQGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEIS **S** ITKSFKGDGWEKLKLYV FDVPDAEGNLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGE GVWRNPNAPYERKRSSQILKLKTARGEECTVIAHHKGKGQFENVMGALTCKNH RGEFKIGSGFNLNERENPPPIGSVITYKYRGITNSGKPRFATYWREKK
Ligasa de ADN DLXd, ligasa artificial derivada de la ligasa Hin D de Haemophilus influenza: MKFYRTLLLFFASSFAFANSDLMLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQ LLTRQGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEIS **S** ITKSFKGDGWEKLKLYV FDVPDAEGNLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGE GVWRNPNAPYERKRSSQILKLKTAR **D** EECTVIAHHKGKGQFENVMGALTCKNH RGEFKIGSGFNLNERENPPPIGSVITYKYRGITNSGKPRFATYWREKK
Ligasa de ADN DLXd2 (Gammaproteobacteria, Haemophilus influenza) (modificado) MLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTRQGQRLSPPAYFIKDFPPFA IDGELFSERNHFEEIS**S**ITKSFKGDGWEKLKLYVFDVPDAEGNLFERLAKLKAHLL EHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGEGVWRNPNAPYERKRSSQILKL KTAR**D**EECTVIAHHKGKGQFENVMGALTCKNHRGEFKIGSGFNLNERENPPPIGS VITYKYRGITNSGKPRFATYWREKK

La SFL es en un aspecto una enzima que puede mediar la formación de un enlace covalente entre dos extremos terminales polinucleotídicos, por ejemplo, un extremo 3'-OH terminal y un extremo 5'-PO4 terminal se unen para formar un enlace fosfodiéster. En algunos casos, la ligación de ADN implica uno o más de tres pasos secuenciales de transferencia de nucleotidilo, que se analizan a continuación. Los tres pasos químicos dependen de un cofactor de cationes divalente. En un aspecto, la SFL es una ligasa dependiente de ATP o una ligasa dependiente de NAD^+.

Opcionalmente, la SFL es la ligasa de ADN del virus Chlorella (ChVLig). Ho, y otros, J Virol, 71 (3): 1931-19374 (1997) o fragmento funcional o variante de esta. Por ejemplo, la SFL puede comprender uno o más dominios característicos de una ligasa, por ejemplo, un dominio nucleotidiltransferasa N-terminal (NTasa) y/o un dominio OB C-terminal. El dominio OB opcionalmente comprende un barril beta antiparalelo de cinco cadenas más una hélice alfa. Dentro del dominio NTasa hay un bolsillo de unión a adenilato compuesto por los seis motivos peptídicos que definen la familia de enzimas covalentes de la enzima NTasa de las ligasas de polinucleótidos. Opcionalmente, el dominio NTasa puede comprender uno o más de los motivos de aminoácidos de ligasa I, Ia, III, IIIa, IV, V y VI, preferentemente los seis motivos. El motivo I (por ejemplo, KxDGxR o un motivo "KXDG") contiene opcionalmente una lisina. Las secuencias ejemplares para cada motivo en la ligasa CV son ATPKIDGIR (motivo I), SRT (motivo Ia), EGSDGEIS (motivo III), YWFDY (motivo IIIa), EGVMIR (motivo IV), LLKMK (motivo V). El motivo 1 contiene preferentemente un residuo de lisina. Otros ejemplos de motivos incluyen CELKLDGLA, VEHKVDGLS, CEPKLDGLA, CELKLDGVA, AEIKYDGVR, CEYKYDGQR, VDYKYDGER, FEIKYDGAR, FEGKWDGYR, AREKIHGTN, ACEKVHGTN, ILTK. Los ejemplos de motivos Ia incluyen TRG, SRT, SRR, SRN, SRS, KRT, KRS, SKG y TRG. Los ejemplos del motivo III incluyen

LEVRGEVF, VEVRGECY, LEVRGEVY, LEARGEAF, FMLDGELM, EGSDGEIS, FILDTEAV, FIIEGEIV, AIVEGELV, VVLDGEAV, YQVFGEFA, LVLNGELF, FTANFEFV y LILVGEMA. Los ejemplos del motivo IIIa incluyen FCYGV, FLYTV, TFYAL, ICHGL, NAYGI, FVYGL, KLYAI, YWFDY, YAFDI, FLFDL, NLFDV, WAFDL, YVFDI, FAFDI, ILLNA Y FLFDV. Los ejemplos del motivo IV incluyen DGVVIK, DGIVIK, DGVVVK, DGTVLK, EGLIVK, EGVMIR, EGLMVK, EGVMVK, EGLMAK, EGVIAK, EGYVLK, EGVVIR, EGYVAV y EGIIMK. Los ejemplos del motivo V incluyen AVAFK, AIAYK, ALAYK, AIAYK, WWKMK, LLKMK, WLKLK, WIKLK, WLKIK, WVKDK, AIKCK, IIKLR, HFKIK e IVKYV. La SFL opcionalmente comprende los seis motivos. Opcionalmente, los seis motivos se encuentran juntos en una ligasa natural, como una SFL que se identifica en la presente. Opcionalmente, la SFL no es una enzima que forma caperuza en el ARN.

La ligasa opcionalmente comprende cualquier porción funcional de una SFL. La ligasa puede ser homóloga a una SFL o a cualquier porción funcional de esta, por ejemplo, más del 75 %, 85 %, 90 %, 95 % o 99 % homóloga a nivel de aminoácidos.

Opcionalmente, cualquiera de los métodos de ligación que se describen en la presente se realiza mediante una SFL que no es la ligasa CV o un fragmento funcional o derivado de esta. Opcionalmente, la SFL no es la ligasa de ADN T4 o un fragmento funcional o variante de esta. Opcionalmente, la SFL es menos del 70 %, 80 % o 90 % idéntica a la ligasa de ADN T4. En algunos ejemplos, la SFL no es una ligasa que se conoce por la fecha efectiva de presentación de esta solicitud para ligar eficientemente oligonucleótidos de menos de 6 nucleótidos de longitud. En algunos ejemplos, la SFL no es una ligasa que se conoce antes del 11 de enero de 2011, para ligar eficientemente oligonucleótidos de menos de 6 nucleótidos de longitud.

En un ensayo típico, una sonda que tiene una longitud de 2, 3, 4, 5 o 6 nucleótidos se liga de manera dependiente del molde a un cebador mediante el uso de una SFL. Opcionalmente, el extremo 3' terminal de la sonda se liga al extremo 5' terminal del cebador, o viceversa. Opcionalmente, la SFL es la ligasa CV. La eficiencia de la ligación es, por ejemplo, más del 5 %, 10 %, 20 %, 30 % o 50 %. En otros casos, una sonda que tiene 6, 7, 8, 9, 10, 11 o 12 nucleótidos de longitud se liga de manera dependiente del molde a un cebador mediante el uso de una SFL. Opcionalmente, el extremo 5' terminal de la sonda se liga al extremo 3' terminal del cebador, o viceversa. Opcionalmente, la SFL es la ligasa CV. La eficiencia de la ligación es, por ejemplo, más del 80 %, 90 % o 95 %.

En cualquier ligación que se describe en la presente, el producto de la ligación se detecta opcionalmente. La ligación puede detectarse por cualquier método conocido o método que se describe en la presente. Por ejemplo, el cebador y/o el molde pueden inmovilizarse en un soporte y etiquetar la sonda. El marcador de la sonda inmovilizada ligada puede detectarse después de que la sonda no ligada se haya lavado. En otras modalidades, se marca una o más de la sonda, cebador o molde. Cualquiera de estos reactivos puede inmovilizarse opcionalmente.

En cualquier ligación en la presente, la ligación puede repetirse durante un número deseado de ciclos, por ejemplo, cualquier reactivo que se someta a un primer ciclo de ligación se usa como reactivo en un siguiente ciclo de ligación. Por ejemplo, el cebador, la sonda o el molde de un primer ciclo de ligación pueden usarse como cebador, sonda o molde en el siguiente ciclo de ligación. En algunas modalidades (por ejemplo, en la secuenciación de ligación), el cebador de un primer ciclo se usa como cebador en el siguiente ciclo. En otros casos, el cebador de un primer ciclo puede usarse como sonda o molde en el siguiente, la sonda de un primer ciclo puede usarse como cebador o molde, o el molde de un primer ciclo puede usarse como cebador o sonda siguiente.

El siguiente ciclo puede diseñarse de manera que el reactivo ligado y no ligado de un primer ciclo pueda actuar como reactivo en el siguiente ciclo. Por ejemplo, el cebador ligado y no ligado de un primer ciclo de ligación puede usarse como cebador en el siguiente ciclo. Alternativamente, el siguiente ciclo puede diseñarse de modo que solo un producto de ligación del primer ciclo pueda actuar como un reactivo de ligación en el siguiente ciclo, y los reactivos que permanecen sin ligar en el primer ciclo no actuarán como reactivos en el siguiente ciclo. En algunos ejemplos, los reactivos no ligados del ciclo anterior se "protegen con caperuza" o se vuelven no ligables antes de que se ejecute el siguiente ciclo de ligación.

Cualquiera o más ligasas y/o métodos de ligación que se abarcan por la invención pueden usarse opcionalmente en uno o más formatos de ensayo de ligación, y/o se realizan en el contexto de una o más aplicaciones de ligación específicas. Los ejemplos no limitantes de formatos de ensayo incluyen: el ensayo de ligación de oligonucleótidos (OLA), una reacción en cadena de la ligasa (LCR), una reacción de detección de ligasa (LDR) y los ensayos que se combinan como el OLA junto con la reacción en cadena de la polimerasa (PCR), por ejemplo, OLA- PCR y PCR-OLA, la reacción en cadena combinada (CCR; una combinación de PCR y LCR) y PCR-LDR (véase, por ejemplo, Landegren y otros, Science 241: 1077-80, 1988; Barany, Proc. Natl. Acad Sci. 88: 189-93, 1991; Grossman y otros, Nucl. Acidos Res. 22 (21): 4527-34, 1994; Bi y Stambrook, Nucl. Acidos Res. 25 (14): 2949-51, 1997; Zirvi y otros, Nucl. Acids Res., 27 (24): e40, 1999; Pat. Nº 4.988.617; y las publicaciones PCT Núms. WO 97/31256 y WO 01/92579. Los ejemplos no limitantes de aplicaciones específicas incluyen: la amplificación del molde, la detección y/o cuantificación de la presencia de un ácido nucleico particular, por ejemplo, en una muestra de diagnóstico, la secuenciación de ligación, el análisis de polimorfismo de un solo nucleótido (SNP), la genotipificación de SNP, la detección de mutaciones, la identificación de genes de copia única, la detección de secuencias repetidas de microsatélites y el mapeo de aductos de ADN, entre otras cosas. Véase también Whitely y otros, Patente de EE.UU. Núm. 4.883.750; Letsinger y otros,

Patente de EE.UU. Núm. 5.476.930; Fung y otros, Patente de EE.UU. Núm. 5.593.826; Kool, patente de EE.UU. Núm. 5.426.180; Landegren y otros, Patente de EE.UU. Núm. 5.871.921; Xu y Kool, Nucleic Acids Research, 27: 875-881 (1999); Higgins y otros, Methods in Enzymology, 68: 50-71 (1979); Engler y otros, The Enzymes, 15-3-29 (1982); y Namsaraev, publicación de patente estadounidense 2004/0110213.

En una modalidad, el método de ligación comprende o consiste en un ensayo de ligación por proximidad (PLA). Los PLA generalmente implican al menos tres pasos. El primer paso es típicamente la unión de las sondas primera y segunda (por ejemplo, las sondas de anticuerpos) a un ligando (por ejemplo, una proteína de interés) de modo que las sondas estén muy cerca de otra. Cada una de las sondas contiene típicamente un oligonucleótido. Los oligonucleótidos se acercan entre sí con la unión de las sondas y, en el segundo paso, se unen entre sí (por ejemplo, el evento de ligación). Los oligonucleótidos ligados pueden amplificarse y detectarse para determinar la presencia del ligando con una muestra (por ejemplo, una muestra biológica).

Un ensayo de PLA ejemplar comprende las etapas de determinar la presencia o ausencia de una proteína objetivo en una muestra, que comprende (a) poner en contacto la proteína objetivo con al menos un primer y segundo agente de unión, teniendo cada agente de unión especificidad de unión para la proteína y siendo unido a al menos un polinucleótido, (b) ligar los oligonucleótidos en el primer y segundo agente de unión entre sí mediante el uso de una ligasa para producir un ácido nucleico objetivo y amplificar el ácido nucleico objetivo; (c) detectar si el ácido nucleico objetivo que se amplificó está presente o no, y (d) concluir que la proteína objetivo está presente en la muestra si se detecta una cantidad significativa de ácido nucleico objetivo que se amplificó, y/o concluir que la proteína objetivo está ausente de la muestra si no se detecta una cantidad significativa del ácido nucleico objetivo que se amplificó. Opcionalmente, la etapa (d) comprende de manera alternativa o adicional medir o evaluar de otro modo la cantidad de ácido nucleico objetivo que se amplificó, y tomar la cantidad de ácido nucleico que se amplificó como un indicador de la cantidad de proteína objetivo presente en la muestra.

En cualquiera de los métodos que se describen en la presente, las sondas de ligación de polinucleótidos pueden usarse para la detección de un ácido nucleico objetivo. La sonda de ligación es opcionalmente capaz de unirse a un ácido nucleico objetivo de secuencia complementaria a través de uno o más tipos de enlaces químicos, generalmente a través del apareamiento de bases complementarias, generalmente a través de la formación de enlaces de hidrógeno. Como se usa en la presente, una sonda puede incluir las bases naturales (es decir, A, G, C o T) o modificadas (7-deazaguanosina, inosina, etc.). Además, las bases en una sonda pueden unirse por un enlace que no sea un enlace fosfodiéster, siempre que no interfiera con la hibridación. Así, por ejemplo, las sondas pueden ser ácidos nucleicos peptídicos en los que las bases constituyentes se unen por enlaces peptídicos en lugar de enlaces fosfodiéster. Un experto en la técnica entenderá que las sondas pueden unirse a secuencias objetivo que carecen de completa complementariedad con la secuencia de la sonda, en dependencia de la rigurosidad de las condiciones de hibridación. Las sondas se marcan preferentemente directamente como con isótopos, cromóforos, lumóforos, cromógenos, o se marcan indirectamente como con biotina a la que puede unirse posteriormente un complejo de estreptavidina. Al analizar la presencia o ausencia de la sonda, puede detectarse la presencia o ausencia de la secuencia o subsecuencia seleccionada.

El ensayo de ligación de oligonucleótidos (OLA) es un método conveniente y altamente riguroso que permite la distinción entre variantes de secuencia de ADN que se conocen (Landegren, 1988). El análisis múltiple de loci altamente polimórficos es útil para la identificación de individuos, por ejemplo, para pruebas de paternidad y en ciencias forenses, compatibilidad de donantes y receptores de trasplantes de órganos, el diagnóstico de enfermedades genéticas, el pronóstico y asesoramiento prenatal, y otras pruebas genéticas que dependen sobre la discriminación de diferencias de base única en una multiplicidad de loci (Delahunty, 1996). Por ejemplo, diferentes ensayos en los que dos quimeras de ADN-PNA, una quimera de secuencia de tipo salvaje (WT) y una quimera de secuencia mutante, tienen diferentes colorantes fluorescentes. Solo cuando la secuencia mutante está presente en la muestra objetivo, la quimera de secuencia mutante se ligará a la segunda sonda (oligo) alineada adyacente si el par de bases mutantes está en el sitio de ligación.

a) Determinación de secuencia

En algunas modalidades, pueden realizarse ciclos sucesivos de ligación, donde el producto de ligación de un ciclo anterior se usa como cebador, sonda y/o molde en un ciclo posterior.

Un uso ejemplar de tal ligación repetitiva es la secuenciación de la ligación. Un molde para secuenciar opcionalmente contiene una región de unión al cebador y una región polinucleotídica de secuencia desconocida. Un cebador con un extremo terminal ligable (por ejemplo, un grupo 3' OH libre o un grupo fosfato 5') se alinea a la región de unión del cebador. Una sonda de extensión se hibrida con el molde adyacente al cebador. El nucleótido proximal de la sonda forma un par de bases complementarias con un nucleótido desconocido en el molde. La sonda de extensión se liga al cebador, opcionalmente con una SFL, lo que resulta en un dúplex extendido. Después de la ligación, el marcador adherido a la sonda de extensión se detecta opcionalmente. El proceso se repite durante un número deseado de ciclos, mediante el uso del producto de ligación de un ciclo como cebador del siguiente ciclo. Opcionalmente, la sonda de extensión tiene un extremo terminal distal no ligable y luego se corta para proporcionar un dúplex extendido ligable.

En un ejemplo, el extremo 3' terminal del cebador se liga al extremo 5' terminal de la sonda de extensión. El paso opcional de escisión puede hacerse, por ejemplo, en un enlace de fosforotiolato mediante el uso de $AgNO_3$ u otra sal que proporcione iones Ag^+, y dejar un grupo fosfato en el extremo 3' terminal del dúplex extendido. El tratamiento con fosfatasa se usa para generar un extremo terminal de sonda extensible en el dúplex extendido.

En la codificación de una base, el marcador corresponde a la identidad del nucleótido X. Por lo tanto, el nucleótido Y se identifica como el nucleótido complementario al nucleótido X. En otras codificaciones, el marcador no tiene una correspondencia biunívoca con la identidad de ningún nucleótido particular en el molde.

También se proporcionan familias de sondas para su uso en diversos métodos de ligación en la presente, por ejemplo, la secuenciación. Las familias de sondas se caracterizan opcionalmente porque cada familia de sondas comprende una pluralidad de sondas de oligonucleótidos marcadas de secuencia diferente y, en cada posición en la secuencia, una familia de sondas comprende al menos 2 sondas que tienen bases diferentes en esa posición. Las sondas en cada familia de sondas comprenden el mismo marcador. Preferentemente, las sondas comprenden un enlace internucleósido escindible. El enlace de tijera puede ubicarse en cualquier lugar de la sonda. Preferentemente, las sondas tienen un resto que no es extensible por la ligasa en un extremo terminal. Preferentemente, las sondas se marcan en una posición entre el enlace escindible y el resto que no es extensible por la ligasa, de modo que la escisión del enlace escindible después de la ligación de una sonda a un extremo terminal de sonda extensible resulta en una porción no marcada que se liga al extensible terminal de la sonda y una porción marcada que ya no está unida a la porción no etiquetada.

En codificaciones de múltiples bases, las sondas en cada familia de sondas comprenden preferentemente al menos j nucleósidos X, en donde j es al menos 2, y en donde cada X es al menos 2 veces degenerado entre las sondas en la familia de sondas. Las sondas en cada familia de sondas comprenden además al menos k nucleósidos N, en donde k es al menos 2, y en donde N representa cualquier nucleósido. En general, j + k es igual o menor que 100, típicamente menor o igual que 30. Los nucleósidos X pueden ubicarse en cualquier lugar de la sonda. Los nucleósidos X no necesitan ubicarse en posiciones contiguas. De este modo, los nucleósidos N no necesitan ubicarse en posiciones contiguas. En otras palabras, los nucleósidos X y N pueden intercalarse. Sin embargo, puede considerarse que los nucleósidos X tienen una secuencia 5'→ 3', al entender que los nucleósidos no necesitan ser contiguos. Por ejemplo, se consideraría que los nucleósidos X en una sonda de estructura $X_ANX_GNNX_CN$ tienen la secuencia AGC. De manera similar, puede considerarse que los nucleósidos N tienen una secuencia.

Los nucleósidos X pueden ser idénticos o diferentes, pero no se seleccionan independientemente, es decir, la identidad de cada X se limita por la identidad de uno o más nucleósidos X en la sonda. Así, en general, solo ciertas combinaciones de nucleósidos X están presentes en cualquier sonda particular y dentro de las sondas en cualquier familia de sonda particular. En otras palabras, en cada sonda, las secuencias de nucleósidos X solo pueden representar un subconjunto de todas las secuencias posibles de longitud j. Así, la identidad de uno o más nucleótidos en X limita las identidades posibles para uno o más de los otros nucleósidos.

Los nucleósidos N se seleccionan preferentemente independientemente y pueden ser A, G, C o T (u, opcionalmente, un nucleósido reductor de la degeneración). Preferentemente, la secuencia de nucleósidos N representa todas las secuencias posibles de longitud k, excepto que uno o más N puede ser un nucleósido reductor de la degeneración. Por lo tanto, las sondas contienen dos porciones, de las cuales la porción que consiste en nucleósidos N se denomina porción no restringida y la porción que consiste en nucleósidos X se llama porción restringida. Como se describió anteriormente, las porciones no necesitan ser nucleósidos contiguos. Las sondas que contienen una porción restringida y una porción no restringida se denominan en la presente sondas parcialmente restringidas. Preferentemente, uno o más nucleósidos en la porción restringida están en el extremo terminal proximal de las sondas, es decir, en el extremo terminal que contiene el nucleósido que se liga al extremo terminal de la sonda extensible, que puede ser el extremo 5' o 3' terminal de la sonda de oligonucleótidos en diferentes modalidades de la descripción.

Dado que la porción restringida de cualquier sonda oligonucleotídica solo puede tener ciertas secuencias, conocer la identidad de uno o más de los nucleósidos en la porción restringida de una sonda, ya sea por sí misma u opcionalmente en combinación con la identidad del marcador en la sonda, proporciona información sobre uno o más de los otros nucleótidos en la porción restringida. La información puede o no ser suficiente para identificar con precisión uno o más de los otros nucleósidos, pero es suficiente para eliminar una o más combinaciones y/o permutaciones posibles de nucleótidos en la porción restringida. Opcionalmente, la información no es suficiente para eliminar cualquier identidad posible de un nucleótido individual en la porción restringida. En ciertas modalidades preferentes, conocer la identidad de un nucleósido en la porción restringida de una sonda es suficiente para identificar con precisión cada uno de los otros nucleósidos en la porción restringida, es decir, para determinar la identidad y el orden de los nucleósidos que comprenden la porción restringida.

Como en los métodos de secuenciación de codificación de una base que se describen anteriormente, el nucleósido más proximal en una sonda de extensión que es complementaria al molde se liga a un extremo terminal extensible de un oligonucleótido inicializador (en el primer ciclo de extensión, ligación y detección) y a un extremo terminal extensible de una sonda de oligonucleótidos que se extiende en ciclos posteriores de extensión, ligación y detección. La detección del marcador que se asocia determina el nombre de la familia de sondas a la que pertenece la sonda recién ligada.

Dado que cada posición en la porción restringida de la sonda es degenerada al menos 2 veces, el nombre de la familia de la sonda no identifica en sí misma ningún nucleótido en la porción restringida. Sin embargo, dado que la secuencia de la porción restringida es una de un subconjunto de todas las secuencias posibles de longitud j, la identificación de la familia de sondas elimina ciertas posibilidades para la secuencia de la porción restringida. La porción restringida de la sonda constituye su porción determinante de secuencia. Por lo tanto, eliminar una o más posibilidades para la identidad de uno o más nucleósidos en la porción restringida de la sonda al identificar la familia de sondas a la que pertenece elimina una o más posibilidades para la identidad de un nucleótido en el molde a la que la sonda de extensión hibrida. En modalidades preferentes, las sondas parcialmente restringidas comprenden un enlace escindible entre cualquiera de los dos nucleósidos.

En ciertas modalidades, las sondas parcialmente restringidas tienen la estructura general $(X)_j(N)_k$, en la que X representa un nucleósido, $(X)_j$ es degenerado al menos 2 veces en cada posición, de modo que X puede ser cualquiera de al menos 2 nucleósidos que tienen diferentes especificidades de apareamiento de bases, N representa cualquier nucleósido, j es al menos 2, k está entre 1 y 100, y al menos un N o X que no sea X en el extremo terminal de la sonda comprende un resto detectable. Preferentemente $(N)_k$ es independientemente degenerado 4 veces en cada posición de modo que, en cada sonda, $(N)_k$ representa todas las secuencias posibles de longitud k, excepto que una o más posiciones en $(N)_k$ pueden ocuparse por un nucleótido reductor de la degeneración. Los nucleósidos en $(X)_j$ pueden ser idénticos o diferentes, pero no se seleccionan independientemente. En otras palabras, en cada sonda, $(X)_j$ solo puede representar un subconjunto de todas las secuencias posibles de longitud j. Por lo tanto, la identidad de uno o más nucleótidos en $(X)_j$ limita las identidades posibles para uno o más de los otros nucleósidos. Por lo tanto, las sondas contienen dos porciones, de las cuales $(N)_k$ es la porción no restringida y $(X)_j$ es la porción restringida.

En ciertas modalidades preferentes, las sondas parcialmente restringidas tienen la estructura 5'-$(X)_j(N)_kN_B$*-3' o 3'-$(X)_j(N)_kN_B$*-5', en donde N representa cualquier nucleósido, N_B representa un resto que no es extensible por la ligasa, * representa un resto detectable, $(X)_j$ es una parte restringida de la sonda que tiene al menos 2 degeneraciones en cada posición, los nucleósidos en $(X)_j$ pueden ser idénticos o diferentes, pero no se seleccionan independientemente, al menos un enlace internucleosídico es un enlace escindible, j es al menos 2 y k está entre 1 y 100, con la condición de que un resto detectable pueda estar presente en cualquier nucleósido N o X que no sea X en el extremo terminal de la sonda en lugar de, o además de, N_B. El enlace escindible puede estar entre dos nucleósidos en $(X)_j$, entre el nucleótido más distal en $(X)_j$ y el nucleósido más próximo en $(N)_k$, entre los nucleósidos dentro de $(N)_k$, o entre el nucleósido terminal en $(N)_k$ y N_B. Preferentemente, el enlace escindible es un enlace fosforotiolato.

Una pluralidad de familias de sondas se denomina "colección" de familias de sondas. Las sondas en cada familia de sondas en una colección de familias de sondas se marcan con un marcador que se distingue de los marcadores que se usan para marcar otras familias de sondas en la colección. Cada familia de sondas tiene preferentemente su propio conjunto definido de secuencias, que opcionalmente no se superponen con ninguna otra familia de sondas. Preferentemente, la combinación de conjuntos de secuencias definidas para familias de sondas en una colección de familias de sondas incluye todas las secuencias posibles de la longitud de la porción que determina la secuencia. Preferentemente, una colección de familias de sondas comprende o consiste en 4 familias de sondas que se marcan distintivamente.

Preferentemente, las porciones que determinan la secuencia de las sondas en cada familia de sondas tienen la misma longitud, y preferentemente las porciones que determinan la secuencia de las familias de sondas en una colección de familias de sondas son de la misma longitud. Preferentemente, la porción que determina la secuenciación es una porción restringida que tiene al menos 2 nucleósidos de longitud.

En algunos casos, una serie de ciclos de ligación a partir de un cebador particular permite la determinación parcial de una secuencia, es decir, la identificación de nucleótidos individuales separados entre sí en un molde. Opcionalmente, para reunir información más completa, se realiza una pluralidad de reacciones en las que cada reacción usa un oligonucleótido de inicialización diferente i. Los oligonucleótidos de inicialización se unen a diferentes porciones de la región de unión. Preferentemente, los oligonucleótidos de inicialización se unen en posiciones tales como los extremos terminales extensibles de los diferentes oligonucleótidos de inicialización están compensados por 1 o más nucleótidos entre sí cuando se hibridan con el molde. Por ejemplo, como se muestra en la Figura 3, se realizan las reacciones de secuenciación 1...N. Los oligonucleótidos de inicialización i_1... i_n tienen la misma longitud y se unen de tal manera que sus nucleótidos terminales se hibridan con posiciones adyacentes sucesivas en el molde. Las sondas de extensión e_1...e_n se unen así en regiones adyacentes sucesivas del molde y se unen a los extremos terminales extensibles de los oligonucleótidos de inicialización. El nucleótido terminal de la sonda e_n ligada a i_n es complementario al nucleótido de la región polinucleotídica, es decir, el primer polinucleótido desconocido en el molde. En el segundo ciclo de extensión, ligación y detección, el nucleótido terminal de la sonda e_{12} es complementario al nucleótido de la región polinucleotídica, es decir, el segundo nucleótido de secuencia desconocida. De la misma forma, los nucleótidos terminales de las sondas de extensión ligadas al dúplex que se inicializaron con los oligonucleótidos de inicialización i_2, i_3, i_4, etc., serán complementarios al tercer, cuarto y quinto nucleótidos de secuencia desconocida. Se apreciará que los oligonucleótidos de inicialización pueden unirse a regiones progresivamente más alejadas de la región polinucleotídica en lugar de estar progresivamente más cerca de ella.

La función espaciadora de los nucleótidos no terminales de las sondas de extensión permite la adquisición de

información de secuencia en posiciones en el molde que se eliminan considerablemente de la posición en la que se une el oligonucleótido de inicialización sin requerir que se realice un número correspondientemente grande de ciclos en cualquier molde dado.

a. Formación de caperuza

En cualquiera de los métodos de ligación repetitiva en la presente, es posible que menos que todas las sondas con extremos terminales extensibles participen en una reacción de ligación exitosa en cada ciclo de extensión, ligación y escisión. Se apreciará que, si tales sondas participaran en ciclos sucesivos, la precisión de cada paso de identificación de nucleótidos podría disminuir progresivamente. En ciertas modalidades, se incluye una etapa de formación de caperuza para evitar que los extremos terminales extensibles que no experimenten ligación participen en ciclos futuros. Cuando se secuencia en la dirección 5'→3' mediante el uso de sondas de extensión que contienen un enlace de fosforotiolato 3'-OPS-5', la formación de caperuza puede realizarse al extender los extremos extensibles no ligados con una polimerasa de ADN y un resto no extensible, por ejemplo, una cadena-nucleótido terminal tal como un didesoxinucleótido o un nucleótido con un resto de bloqueo unido, por ejemplo, después de la etapa de ligación o detección. Cuando se secuencia en la dirección 3'→5' mediante el uso de sondas de extensión que contienen un enlace de fosforotiolato 3'-SPO-5', puede realizarse la formación de caperuza, por ejemplo, al tratar el molde con una fosfatasa, por ejemplo, después de la ligación o detección. También pueden usarse otros métodos de formación de caperuza.

Se contempla que cualquier ensayo o método de ligación en la presente puede ser altamente multiplexado de tal manera que pueda realizarse un gran número de ensayos (por ejemplo, más de 1.000) en paralelo, por ejemplo, simultáneamente.

Cualquier ensayo de ligación multiplexada puede realizarse opcionalmente en sustratos sólidos donde uno o más reactivos de ligación (por ejemplo, la ligasa, el molde, la sonda y/o el cebador) pueden inmovilizarse en un soporte sólido o superficie. Opcionalmente, los cebadores, la sonda y/o el molde pueden unirse a diferentes porciones de un sustrato sólido en forma de matriz. Opcionalmente, el molde, el cebador o la sonda pueden unirse covalentemente al sustrato sólido.

Opcionalmente, uno o más reactivos de ligación se marcan (por ejemplo, el molde, la sonda y/o el cebador) para que los productos de ligación puedan discriminarse entre sí. Alternativamente, los productos de ligación pueden distinguirse en función de: (i) el tamaño mediante electroforesis o cromatografía y/o (ii) marcas detectables (Grossman, 1994). Por ejemplo, los ensayos de ligación multiplexados pueden realizarse en una sola muestra en un solo recipiente.

Cualquiera de los reactivos en la presente (por ejemplo, el cebador, la sonda y/o el molde) puede inmovilizarse opcionalmente en una fase sólida. La fase sólida opcionalmente comprende una superficie a la que uno o más reactivos pueden unirse electrostáticamente, hidrofóbicamente o covalentemente. Las fases sólidas representativas incluyen, por ejemplo, nylon 6; nylon 66; poliestireno; perlas de látex; perlas magnéticas; perlas de vidrio; polietileno; polipropileno; polibutileno; copolímeros de butadienoestireno; caucho silastic; poliésteres; poliamidas; celulosa y derivados; acrilatos; metacrilatos; polivinilo; cloruro de vinilo; cloruro de polivinilo; fluoruro de polivinilo; copolímeros de poliestireno; gel de sílice; obleas de sílice de vidrio; agarosa; dextranos; liposomas; proteínas insolubles metales; y nitrocelulosa. Pueden formarse fases sólidas representativas en artículos apropiados tales como perlas (por ejemplo, micropartículas), tubos, tiras, discos, papeles de filtro, placas y similares.

En una modalidad, cualquiera de los siguientes puede unirse a un soporte sólido (por ejemplo, una perla): la SFL, uno o más reactivos polinucleotídicos (por ejemplo, una sonda de inicialización, una sonda de extensión, un ADN o molde objetivo, un polinucleótido 3' terminal o 5' terminal).

Si se desea, pueden acoplarse uno o más marcadores a cualquier reactivo de interés, por ejemplo, la SFL o un anticuerpo que se une específicamente a la SFL, uno o más polinucleótidos (por ejemplo, el molde, el polinucleótido 5' terminal, el polinucleótido 3' terminal, el molde, la sonda de inicialización, la sonda de extensión, etc.

Composiciones

También se proporciona un polipéptido que comprende o que consiste esencialmente en la secuencia de aminoácidos de una SFL que se proporciona en la presente, generalmente que se proporciona como números de acceso de GenBank en las Tablas, o un fragmento o variante de esta. El fragmento o variante de esta puede ligar preferentemente oligonucleótidos cortos con una eficiencia comparable a su enzima original. Opcionalmente, las variantes o fragmentos de las SFL son al menos 70 %, 80, 90 %, 95 %, 99 % idénticos a la SFL original. En un ejemplo, la SFL es una ligasa derivada de la ligasa de ADN Hin (por ejemplo, la DLX, DLXd o DLXd2) o cualquier fragmento o variante de esta que aún retenga uno o más residuos mutantes que se muestran en la Figura 2, y/o tiene uno o más aminoácidos del extremo C-terminal eliminados, por ejemplo, 22 aminoácidos del extremo C-terminal. Por ejemplo, la ligasa de ADN Hin mutante es al menos un 70 % idéntica a la secuencia de la ligasa Hin D que se proporciona en la Figura 2 o en el Núm. de acceso de GenBank P44121, cuya ligasa comprende una mutación de aminoácidos en la posición 193 de la secuencia de la ligasa Hin D que se proporciona en la Figura 2 o en el Núm. De acceso GenBank P44121.

Opcionalmente, la mutación de aminoácidos consiste en cambiar la glicina en la posición 193 a ácido aspártico o ácido glutámico.

También se proporciona un ácido nucleico que codifica las nuevas SFL o fragmentos o variantes que se describen en la presente. También se proporcionan genes de la nueva SFL que comprenden secuencias que dirigen la expresión de la SFL. También se proporcionan vectores de ácido nucleico que comprenden un ácido nucleico que codifica la nueva SFL y una célula huésped que comprende tal vector. También se proporciona un anticuerpo que puede unirse específicamente a la nueva SFL pero no a su progenitor natural.

Opcionalmente, la nueva SFL o ácido nucleico o vector o célula huésped o anticuerpo se purifica, se aísla o se recombina.

También se proporciona un método para preparar una cualquiera o más ligasas que se describen en la presente que comprende: expresar la SFL en una célula huésped, por ejemplo, mediante el cultivo de dichas células huésped en condiciones tales que se exprese la ligasa; y opcionalmente recuperar o purificar dicha ligasa.

Kits

También se proporcionan kits que comprenden componentes para realizar las reacciones de ligación de acuerdo con la descripción. Los kits pueden contener una SFL o una variante funcional o fragmento de esta. Los kits pueden incluir alternativa o adicionalmente una o más sondas de oligonucleótidos de menos de 12 nucleótidos de longitud (por ejemplo, menos de 8, 7, 6, 5, 4, 3 o 2 nucleótidos de longitud) y/u oligonucleótidos cebadores. En algunas modalidades, un kit puede incluir la ligasa CV y una o más sondas de oligonucleótidos de menos de 6 nucleótidos de longitud. Una o más sondas, por ejemplo, comprenden un 5'-fosfato y/o un marcador. El marcador opcionalmente se adhiere al extremo 5' terminal, o alternativamente al extremo 3' terminal. Opcionalmente, una o más sondas son escindibles.

Los kits opcionalmente proporcionan uno o más cebadores, sondas y/ moldes que se describen en la presente. Los cebadores y/o sondas pueden tener una o más características o características que se describen en la presente. Por ejemplo, las sondas pueden comprender un enlace escindible (por ejemplo, un fosforotiolato). Los kits pueden contener un reactivo de escisión adecuado para escindir enlaces de fosfororotiolato, por ejemplo, AgNO3 y tampones apropiados en los que realizar la escisión. Las sondas pueden comprender un residuo desencadenante tal como un nucleósido que contiene una base dañada o un residuo abásico. Los kits pueden contener un reactivo de escisión adecuado para escindir un enlace entre un nucleósido y un residuo abásico adyacente y/o un reactivo adecuado para eliminar una base dañada de un polinucleótido, por ejemplo, una glicosilasa de ADN. Ciertos kits contienen sondas de oligonucleótidos que comprenden un nucleótido disacárido y contienen peryodato como reactivo de escisión. En ciertas modalidades, los kits contienen una colección de familias o colecciones de sondas de oligonucleótidos marcadas distintivamente las que se describen en la presente. Opcionalmente, el kit no comprende una polimerasa. Opcionalmente, el kit no comprende una enzima que no sea una ligasa, fosfatasa o quinasa. Opcionalmente, el kit no comprende una enzima que no sea una SFL que tenga una o más actividades que se mencionan en la presente. Los kits pueden incluir reactivos de ligación (por ejemplo, ligasa, tampones, etc.) e instrucciones para practicar la modalidad particular de la descripción. Pueden incluirse tampones apropiados para las otras enzimas que pueden usarse, por ejemplo, fosfatasa, polimerasas. En algunos casos, estos tampones pueden ser idénticos. Los kits también pueden incluir un soporte, por ejemplo, perlas magnéticas, que están pre-unidas a cebadores, cebadores o moldes o se derivatizan para poder unirse a tales moléculas. Las perlas pueden funcionalizarse con un cebador para realizar la amplificación por la PCR. Otros componentes opcionales incluyen soluciones de lavado; vectores para insertar moldes para amplificación por la PCR; reactivos de la PCR tales como cebadores de amplificación, polimerasa termoestable, nucleótidos; reactivos para preparar una emulsión; reactivos para preparar un gel, etc. Los kits también pueden comprender una pluralidad de sondas marcadas distintivamente, que pueden marcarse de tal manera que la identidad del marcador proporcione información sobre la secuencia de la sonda. Los kits también pueden comprender otras composiciones o reactivos adicionales o para su uso en el protocolo de determinación de secuenciación de ligación. Opcionalmente, la identidad del marcador proporciona la identidad exacta de un nucleótido que ocupa una posición degenerada en la sonda. Por ejemplo, la identidad del marcador opcionalmente elimina una o más combinaciones o permutaciones de nucleótidos en dos o más posiciones degeneradas. Las sondas, por ejemplo, se degeneran en dos o más posiciones restringidas, donde la identidad de un nucleótido en una posición restringida elimina al menos una posible identidad de un nucleótido en otra posición restringida. Opcionalmente, conocer la identidad de un residuo restringido en esa sonda elimina además al menos una identidad posible para otro residuo restringido en la sonda.

A menos que sea de otra forma evidente por el contexto, cualquier característica puede reclamarse en combinación con cualquier otra, o puede reclamarse como no presente en combinación con otra característica. Una característica puede ser, por ejemplo, cualquier variación, etapa, característica, propiedad, composición, método, etapa, grado, nivel, componente, material, sustancia, elemento, modo, variable, aspecto, medida, cantidad o modalidad.

Muchas características las que se describen en la presente se destinan a ser opcionales. Si alguna característica no se indica explícitamente como necesaria, debe considerarse como opcional. Los ejemplos no limitantes de lenguaje que indican que una característica es opcional incluyen términos como "variación", "dónde", "mientras", "cuándo", "opcionalmente", "incluir", "preferente", "especial", "recomendado", "aconsejable", "particular", "debería", "alternativa",

"típica", "representativa", "variada", "y similar", "puede", "podría", "ejemplo", "modalidad" o "en un aspecto" o "si' o cualquier combinación y/o variación de tales términos.

Los ejemplos no limitantes de términos exclusivos o negativos incluyen "solo", "únicamente", "que consiste en", "que consiste esencialmente en", "solo", "sin", "en ausencia de (por ejemplo, otros artículos del mismo tipo, estructura y/o función) "excluyente", "no", "no", "no se puede", o cualquier combinación y/o variación de tal lenguaje.

De este modo, los referentes como "un", "uno/una", "dicho/dicha" o "el/la/los" se destinan a apoyar referentes explícitamente únicos o plurales donde así se desee. Ejemplos no limitantes de referentes plurales incluyen "al menos uno", "uno o más", "más de uno", "dos o más", "una multiplicidad", "una pluralidad", "cualquier combinación de", "cualquier permutación de" "cualquiera o más de", etc.

Los siguientes ejemplos se proporcionan con fines ilustrativos y no pretenden limitar la invención.

Ejemplos

Ejemplo 1

Este ejemplo describe la ligación de sondas cortas.

Los reactivos oligonucleotídicos fueron los siguientes. Los oligos cortos en la ligación directa (es decir, la ligación del extremo 3' terminal de la sonda al extremo 5' terminal del cebador) fueron: 8-mer: 5'-CTCATTCG-3'; 6-mer: 5'-TCATTCG-3' 5-mer: 5'-ATTCG-3'; 4-mer: 5'-TTCG-3'; 3-mer: 5'-TCG-3'. Los oligos cortos que se usaron para la ligación inversa (es decir, la ligación del extremo 5' terminal de la sonda al extremo 3' terminal del cebador) fueron los mismos que los oligos de ligación directa, excepto que tienen un fosfato 5'. Todos los oligonucleótidos se obtuvieron de Integrated DNA Technologies. El cebador directo fue: PO4-5'-CTGCTGTACCGTACATCCGC-3'-6FAM. El cebador inverso fue: 5'-FAM-CTGCCCCGGGTTCCTCATTCTCT-3'

Preparación del sustrato de la ligasa: Se obtuvieron perlas magnéticas MyOne marcadas con ácido carboxílico de Invitrogen. El oligonucleótido de 41 bases marcado con amino 5' (CCA CTA CGC CTC CGC TTT CCT CTC TAT GGG CAG TCG GTG AT) se acopló a perlas MyOne mediante el uso de la química de acoplamiento de amina estándar como se enseña en Nakajima y otros, Bioconjugate Chem. 1995, 6 (1), 123-130 (Figura lb). Se usó la PCR para extender este oligonucleótido mediante el uso de (CTG CCC CGG GTT CCT CAT TCT CTA TTC GCT GCT GTA CCG TAC ATC CGC CTT GGC CGT ACA GCA GAT CAC CGA CTG CCC ATA GAG AGG). El siguiente sustrato de ligación de 105 bases se unió a perlas MyOne (CCA CTA CGC CTC CGC TTT CCT CTC TAT GGG CAG TCG GTG ATC TGC TGT ACG GCC AAG GCG GAT GTA CGG TAC AGC AGC GAA TAG AGA ATG AGG AAC CCG GGG CAG, Figura la). La cantidad de oligo de 105 bases que se ligó a la perla se midió por hibridación de un oligo "cebador" marcado con fluorescencia, con el marcador fluoresceína 3'(6FAM) para la detección por electroforesis capilar. La cantidad de fluorescencia desnaturalizada del molde se comparó con una curva estándar.

Las reacciones de ligación se realizaron en tampón que contenía Tris-HCl 50 mM, MgCL2 10 mM, ATP 1 mM, DTT 1 mM y PEG 8000 al 5 % p/v a pH 7,5 (tampón de ligasa). El oligonucleótido cebador se alineó al molde en tampón de ligasa mediante el calentamiento a 85 ºC durante 10 minutos seguido de enfriamiento lento a temperatura ambiente. En el caso de la ligación directa, el cebador estaba fosforilado en 5' y un 6FAM marcado en 3'. En el caso de la ligación inversa, el cebador estaba marcado 5' con 6FAM. Las reacciones de ligación se realizaron en una placa de 96 pocillos mediante el uso del molde/cebador 2 nM, ligasa 2,0 µM, oligonucleótido corto 2-5 µM en un volumen total de 50 µL. Las reacciones de ligación continuaron durante 20 minutos a 15 °C. Las reacciones se detuvieron mediante el uso de SDS al 1 % v/v seguido de separación magnética de las perlas con el molde de la solución. Las perlas se lavaron tres veces en 100 µl de SDS al 1 % antes de lavar con 100 µl de formamida al 100 %. Después se analizaron 5 µl del sobrenadante que contenía el cebador desnaturalizado, marcado con FAM y producto de la ligación por electroforesis capilar. La eficiencia de la ligación se calculó como la proporción de áreas de pico determinadas por CE, donde se separaron los cebadores marcados, ligados y ligados con FAM, es decir, Eficiencia = ligado /(ligado + no ligado).

En las Figuras 3-5 se muestran una serie de ligaciones de sondas de diferentes longitudes con una variedad de SFL. En cada panel, la reacción de ligación se traza en función de la longitud de la sonda. Estos datos demuestran claramente que las ligasas de huella pequeña pueden ligar sustratos de sonda tan cortos como dos nucleótidos. Mediante el uso de la SFL, el tamaño de un conjunto de sondas de secuenciación (por ejemplo, para la secuenciación de la ligación en SOLiD™) puede reducirse del conjunto actual de 1024 8-mers a tan solo 16 (en el caso de los di-mers), 64 (en el caso de trímeros), o 256 en el caso de tetrámeros.

En otro experimento, se comparó la eficiencia de la ligación de la ligasa SF mediante el uso del mismo sistema experimental que el anterior con la ligasa de ADN T4, mediante el uso de las sondas de secuenciación de ligación de 8-mers que se conjugaron o no con colorantes. La SFL es la eficiencia de ligación global más alta tanto para la sonda oscura (persecución) como para la sonda de secuenciación marcada. Esto es cierto en la concentración de sonda de 200 nM, que se usa actualmente en SOLiD. Específicamente, se usa una mezcla de 1024 sondas en SOLiD, donde cada sonda individual de las 1024 sondas está presente en una concentración de trabajo final de 200 nM. Las otras

concentraciones fueron 2000 nM de enzima y 1 nM de molde de perla, donde la ligación se realizó a 15 °C durante 30 minutos. Los resultados se muestran en la Tabla 2.

Tabla 2

Máxima eficiencia de ligación de sondas 8-mer por varias enzimas ligasas		
Ligasa	**Oscuro**	**Cy 5**
SFL	0,91 ± 0,01	0,82 ± 0,01
T4	0,67 ± 0,02	0,62 ± 0,02

Ejemplo 2

Se determinó que la fidelidad de una ligasa de huella pequeña (ligasa CV) es sorprendentemente mejor que la ligasa de ADN T4 en la secuenciación de ADN mediante ligación en el sistema de secuenciación SOLiD. Un molde de ADN de 35 nucleótidos se unió covalentemente a las perlas, de modo que cada perla magnética tenía aproximadamente 100.000 moléculas de ADN unidas covalentemente en promedio. Aproximadamente 40 millones de perlas se sometieron a una reacción de secuenciación de ligación a 20 °C. La reacción de secuenciación ligó las "sondas" oligonucleotídicas de 8-mer que se marcaron con un colorante fluorescente en un cebador, y se realizaron un total de siete eventos consecutivos de ligación o sondeo en cada molécula molde. El número aproximado de eventos de ligación fue, por lo tanto, 100 000 (es decir, número de moléculas de molde por perla) x 40 000 000 (es decir, número de cuentas) x 7 (número de eventos de ligación por molde) = $2,8 \times 10^{13}$.

Los resultados se dan en la Tabla 3 a continuación. 0MM = Porcentaje de todos los productos de ligación (que son el resultado de siete eventos de ligación consecutivos) sin no coincidencias; 3MM = Porcentaje de todos los productos de ligación (que son el resultado de siete eventos de ligación consecutivos) con menos de 3 no coincidencias.

Tabla 3

Ligasa	**0MM**	**3MM**
T4	20 %	57 %
SF	27 %	65 %

La secuenciación por datos de ligación demuestra que el 20 % del tiempo el producto de ligación es perfectamente complementario al molde y se forma por siete eventos de ligación consecutivos que son correctos. A partir de la ecuación puede calcularse una estimación de la fidelidad de la ligación (que se expresa como el porcentaje de probabilidad de que un evento de ligación sea correcto o "p"): tasa de 0MM = $(p/100)^n$ donde p es el porcentaje de fidelidad de la ligación y n es el número de eventos de ligación por producto de ligación final (aquí, siete). Por lo tanto, para la ligasa de ADN T4, la fidelidad de la ligación es del 79,5 % y para la SFL la fidelidad de la ligación es del 83 %.

Listado de secuencias

<110> LIFE TECHNOLOGIES CORPORATION

<120> LIGACIÓN ENZIMÁTICA DE ÁCIDOS NUCLEICOS

<130> LT00422PCT

<140> PCT/US2012/021465
<141> 2012-01-17

<150> 61/474,205
<151> 2011-04-11

<150> 61/474,168
<151> 2011-04-11

<150> 61/433,502
<151> 2011-01-17

<150> 61/433,488
<151> 2011-01-17

<160> 87

<170> PatentIn version3.5

<210> 1
<211> 20
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<223> /nota="Descripción de la Secuencia Artificial: Oligonucléotido sintético"

<400> 1
gcggatgtac ggtacagcag 20

<210> 2
<211> 10
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<223> /nota="Descripción de la Secuencia Artificial: Oligonucléotido sintético"

<400> 2
agtcggtgat 10

<210> 3
<211> 298
<212> PRT
<213> Paramecium bursaria Chlorella virus 1

<400> 3

```
Met Ala Ile Thr Lys Pro Leu Leu Ala Ala Thr Leu Glu Asn Ile Glu
1               5                   10                  15
Asp Val Gln Phe Pro Cys Leu Ala Thr Pro Lys Ile Asp Gly Ile Arg
            20                  25                  30

Ser Val Lys Gln Thr Gln Met Leu Ser Arg Thr Phe Lys Pro Ile Arg
        35                  40                  45

Asn Ser Val Met Asn Arg Leu Leu Thr Glu Leu Leu Pro Glu Gly Ser
    50                  55                  60

Asp Gly Glu Ile Ser Ile Glu Gly Ala Thr Phe Gln Asp Thr Thr Ser
65                  70                  75                  80

Ala Val Met Thr Gly His Lys Met Tyr Asn Ala Lys Phe Ser Tyr Tyr
                85                  90                  95

Trp Phe Asp Tyr Val Thr Asp Asp Pro Leu Lys Lys Tyr Ile Asp Arg
            100                 105                 110

Val Glu Asp Met Lys Asn Tyr Ile Thr Val His Pro His Ile Leu Glu
        115                 120                 125

His Ala Gln Val Lys Ile Ile Pro Leu Ile Pro Val Glu Ile Asn Asn
    130                 135                 140
```

Ile Thr Glu Leu Leu Gln Tyr Glu Arg Asp Val Leu Ser Lys Gly Phe
145 150 155 160

Glu Gly Val Met Ile Arg Lys Pro Asp Gly Lys Tyr Lys Phe Gly Arg
165 170 175

Ser Thr Leu Lys Glu Gly Ile Leu Leu Lys Met Lys Gln Phe Lys Asp
180 185 190

Ala Glu Ala Thr Ile Ile Ser Met Thr Ala Leu Phe Lys Asn Thr Asn
195 200 205

Thr Lys Thr Lys Asp Asn Phe Gly Tyr Ser Lys Arg Ser Thr His Lys
210 215 220

Ser Gly Lys Val Glu Glu Asp Val Met Gly Ser Ile Glu Val Asp Tyr
225 230 235 240

Asp Gly Val Val Phe Ser Ile Gly Thr Gly Phe Asp Ala Asp Gln Arg
245 250 255

Arg Asp Phe Trp Gln Asn Lys Glu Ser Tyr Ile Gly Lys Met Val Lys
260 265 270

Phe Lys Tyr Phe Glu Met Gly Ser Lys Asp Cys Pro Arg Phe Pro Val
275 280 285

Phe Ile Gly Ile Arg His Glu Glu Asp Arg
290 295

<210> 4
<211> 306
<212> PRT
<213> Burkholderia pseudomallei

<400> 4

Met Ser Gly Val Pro Tyr Gly Phe Lys Pro Asn Leu Ala Ala Thr Leu
1 5 10 15

Thr Lys Pro Glu Leu Ile Lys Phe Pro Val Trp Ala Ser Pro Lys Ile
20 25 30

Asp Gly Ile Arg Cys Val Phe Phe Gly Gly Val Ala Tyr Ser Arg Ser
35 40 45

Leu Lys Pro Ile Pro Asn Pro Val Val Gln Glu Phe Ala Lys Ala Tyr
50 55 60

Ala Asn Leu Leu Glu Gly Leu Asp Gly Glu Leu Thr Val Gly Ser Pro
65 70 75 80

Thr Asp Ala Asn Cys Met Gln Asn Ser Met Ala Val Met Ser Lys Ala
85 90 95

Ala Ala Pro Asp Phe Thr Phe His Val Phe Asp Trp Phe His Pro Ala
100 105 110

Gln Ala His Ile Glu Phe Trp Gln Arg Ser Asp Val Val Glu Asp Arg
115 120 125

Ile Val Gln Phe Tyr Asp Arg Tyr Pro Glu Val Asp Ile Arg Ala Ala
130 135 140

Pro Gln Val Leu Cys Thr Ser Leu Ala His Leu Asp Thr Asn Glu Ala
145 150 155 160

Arg Trp Leu Ala Asp Gly Tyr Glu Gly Met Met Ile Arg Asp His Cys
165 170 175

Gly Arg Tyr Lys Phe Gly Arg Ser Thr Glu Arg Glu Gly Gly Leu Val
180 185 190

Lys Val Lys Arg Phe Thr Asp Ala Glu Ala Ile Val Ile Gly Phe Glu
195 200 205

Glu Glu Met His Asn Ala Asn Glu Ala Lys Arg Asp Ala Thr Gly Arg
210 215 220

Thr Glu Arg Ser Thr Ser Lys Ala Gly Leu His Gly Lys Gly Thr Leu
225 230 235 240

Gly Ala Leu Val Val Lys Asn Glu Arg Gly Ile Val Phe Asn Ile Gly
245 250 255

Thr Gly Phe Thr Ala Ala Gln Arg Ala Asp Tyr Trp Ala Asn His Pro
260 265 270

Ser Leu Phe Gly Lys Met Val Lys Phe Lys His Phe Asp His Gly Thr
275 280 285

Val Asp Ala Pro Arg His Pro Val Phe Ile Gly Phe Arg His Pro Glu
290 295 300

Asp Met
305

<210> 5
<211> 268
<212> PRT
<213> Haemophilus influenza

<400> 5

Met Lys Phe Tyr Arg Thr Leu Leu Leu Phe Phe Ala Ser Ser Phe Ala
1 5 10 15

Phe Ala Asn Ser Asp Leu Met Leu Leu His Thr Tyr Asn Asn Gln Pro
20 25 30

Ile Glu Gly Trp Val Met Ser Glu Lys Leu Asp Gly Val Arg Gly Tyr
35 40 45

Trp Asn Gly Lys Gln Leu Leu Thr Arg Gln Gly Gln Arg Leu Ser Pro
50 55 60

Pro Ala Tyr Phe Ile Lys Asp Phe Pro Pro Phe Ala Ile Asp Gly Glu
65 70 75 80

Leu Phe Ser Glu Arg Asn His Phe Glu Glu Ile Ser Thr Ile Thr Lys
85 90 95

Ser Phe Lys Gly Asp Gly Trp Glu Lys Leu Lys Leu Tyr Val Phe Asp
100 105 110

Val Pro Asp Ala Glu Gly Asn Leu Phe Glu Arg Leu Ala Lys Leu Lys
115 120 125

Ala His Leu Leu Glu His Pro Thr Thr Tyr Ile Glu Ile Ile Glu Gln
130 135 140

Ile Pro Val Lys Asp Lys Thr His Leu Tyr Gln Phe Leu Ala Gln Val
145 150 155 160

Glu Asn Leu Gln Gly Glu Gly Val Val Val Arg Asn Pro Asn Ala Pro
165 170 175

Tyr Glu Arg Lys Arg Ser Ser Gln Ile Leu Lys Leu Lys Thr Ala Arg
180 185 190

Gly Glu Glu Cys Thr Val Ile Ala His His Lys Gly Lys Gly Gln Phe
195 200 205

Glu Asn Val Met Gly Ala Leu Thr Cys Lys Asn His Arg Gly Glu Phe
210 215 220

Lys Ile Gly Ser Gly Phe Asn Leu Asn Glu Arg Glu Asn Pro Pro Pro
225 230 235 240

Ile Gly Ser Val Ile Thr Tyr Lys Tyr Arg Gly Ile Thr Asn Ser Gly
245 250 255

Lys Pro Arg Phe Ala Thr Tyr Trp Arg Glu Lys Lys
260 265

<210> 6
<211> 268
<212> PRT
<213> Secuencia Artificial

<220>
<221> fuente
<223> /nota="Descripción de la Secuencia Artificial: Oligonucléotido sintético"

<400> 6

Met Lys Phe Tyr Arg Thr Leu Leu Leu Phe Phe Ala Ser Ser Phe Ala
1 5 10 15

Phe Ala Asn Ser Asp Leu Met Leu Leu His Thr Tyr Asn Asn Gln Pro
20 25 30

Ile Glu Gly Trp Val Met Ser Glu Lys Leu Asp Gly Val Arg Gly Tyr
35 40 45

Trp Asn Gly Lys Gln Leu Leu Thr Arg Gln Gly Gln Arg Leu Ser Pro
50 55 60

Pro Ala Tyr Phe Ile Lys Asp Phe Pro Pro Phe Ala Ile Asp Gly Glu
65 70 75 80

Leu Phe Ser Glu Arg Asn His Phe Glu Glu Ile Ser Ser Ile Thr Lys
85 90 95

Ser Phe Lys Gly Asp Gly Trp Glu Lys Leu Lys Leu Tyr Val Phe Asp
100 105 110

Val Pro Asp Ala Glu Gly Asn Leu Phe Glu Arg Leu Ala Lys Leu Lys
115 120 125

Ala His Leu Leu Glu His Pro Thr Thr Tyr Ile Glu Ile Ile Glu Gln
130 135 140

Ile Pro Val Lys Asp Lys Thr His Leu Tyr Gln Phe Leu Ala Gln Val
145 150 155 160

Glu Asn Leu Gln Gly Glu Gly Val Val Val Arg Asn Pro Asn Ala Pro
165 170 175

Tyr Glu Arg Lys Arg Ser Ser Gln Ile Leu Lys Leu Lys Thr Ala Arg
180 185 190

Gly Glu Glu Cys Thr Val Ile Ala His His Lys Gly Lys Gly Gln Phe
195 200 205

Glu Asn Val Met Gly Ala Leu Thr Cys Lys Asn His Arg Gly Glu Phe
210 215 220

Lys Ile Gly Ser Gly Phe Asn Leu Asn Glu Arg Glu Asn Pro Pro Pro
225 230 235 240

Ile Gly Ser Val Ile Thr Tyr Lys Tyr Arg Gly Ile Thr Asn Ser Gly
245 250 255

Lys Pro Arg Phe Ala Thr Tyr Trp Arg Glu Lys Lys
260 265

<211> 268
<212> PRT
<213> Secuencia Artificial

<220>
<221> fuente
<223> /nota="Descripción de la Secuencia Artificial: Oligonucléotido sintético"

<400> 7

Met Lys Phe Tyr Arg Thr Leu Leu Leu Phe Phe Ala Ser Ser Phe Ala
1 5 10 15

Phe Ala Asn Ser Asp Leu Met Leu Leu His Thr Tyr Asn Asn Gln Pro
20 25 30

Ile Glu Gly Trp Val Met Ser Glu Lys Leu Asp Gly Val Arg Gly Tyr
35 40 45

Trp Asn Gly Lys Gln Leu Leu Thr Arg Gln Gly Gln Arg Leu Ser Pro
50 55 60

Pro Ala Tyr Phe Ile Lys Asp Phe Pro Pro Phe Ala Ile Asp Gly Glu
65 70 75 80

Leu Phe Ser Glu Arg Asn His Phe Glu Glu Ile Ser Ser Ile Thr Lys
85 90 95

Ser Phe Lys Gly Asp Gly Trp Glu Lys Leu Lys Leu Tyr Val Phe Asp
100 105 110

Val Pro Asp Ala Glu Gly Asn Leu Phe Glu Arg Leu Ala Lys Leu Lys
115 120 125

Ala His Leu Leu Glu His Pro Thr Thr Tyr Ile Glu Ile Ile Glu Gln
130 135 140

Ile Pro Val Lys Asp Lys Thr His Leu Tyr Gln Phe Leu Ala Gln Val
145 150 155 160

Glu Asn Leu Gln Gly Glu Gly Val Val Val Arg Asn Pro Asn Ala Pro
165 170 175

Tyr Glu Arg Lys Arg Ser Ser Gln Ile Leu Lys Leu Lys Thr Ala Arg
180 185 190

Asp Glu Glu Cys Thr Val Ile Ala His His Lys Gly Lys Gly Gln Phe
195 200 205

Glu Asn Val Met Gly Ala Leu Thr Cys Lys Asn His Arg Gly Glu Phe
210 215 220

Lys Ile Gly Ser Gly Phe Asn Leu Asn Glu Arg Glu Asn Pro Pro Pro
225 230 235 240

Ile Gly Ser Val Ile Thr Tyr Lys Tyr Arg Gly Ile Thr Asn Ser Gly
245 250 255

Lys Pro Arg Phe Ala Thr Tyr Trp Arg Glu Lys Lys
260 265

<210> 8
<211> 246
<212> PRT
<213> Secuencia Artificial

<220>
<221> fuente
<223> /nota="Descripción de la Secuencia Artificial: Oligonucléotido sintético"

<400> 8

```
Met Leu Leu His Thr Tyr Asn Asn Gln Pro Ile Glu Gly Trp Val Met
1               5                   10                  15

Ser Glu Lys Leu Asp Gly Val Arg Gly Tyr Trp Asn Gly Lys Gln Leu
            20                  25                  30

Leu Thr Arg Gln Gly Gln Arg Leu Ser Pro Pro Ala Tyr Phe Ile Lys
        35                  40                  45

Asp Phe Pro Pro Phe Ala Ile Asp Gly Glu Leu Phe Ser Glu Arg Asn
    50                  55                  60

His Phe Glu Glu Ile Ser Ser Ile Thr Lys Ser Phe Lys Gly Asp Gly
65                  70                  75                  80

Trp Glu Lys Leu Lys Leu Tyr Val Phe Asp Val Pro Asp Ala Glu Gly
                85                  90                  95

Asn Leu Phe Glu Arg Leu Ala Lys Leu Lys Ala His Leu Leu Glu His
            100                 105                 110

Pro Thr Thr Tyr Ile Glu Ile Ile Glu Gln Ile Pro Val Lys Asp Lys
        115                 120                 125

Thr His Leu Tyr Gln Phe Leu Ala Gln Val Glu Asn Leu Gln Gly Glu
    130                 135                 140

Gly Val Val Val Arg Asn Pro Asn Ala Pro Tyr Glu Arg Lys Arg Ser
145                 150                 155                 160

Ser Gln Ile Leu Lys Leu Lys Thr Ala Arg Asp Glu Glu Cys Thr Val
                165                 170                 175

Ile Ala His His Lys Gly Lys Gly Gln Phe Glu Asn Val Met Gly Ala
            180                 185                 190

Leu Thr Cys Lys Asn His Arg Gly Glu Phe Lys Ile Gly Ser Gly Phe
        195                 200                 205

Asn Leu Asn Glu Arg Glu Asn Pro Pro Pro Ile Gly Ser Val Ile Thr
    210                 215                 220

Tyr Lys Tyr Arg Gly Ile Thr Asn Ser Gly Lys Pro Arg Phe Ala Thr
225                 230                 235                 240

Tyr Trp Arg Glu Lys Lys
                245
```

<210> 9
<211> 6

<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<220>
<221> MOD_RES
<222> (2)..(2)
<223> Cualquier aminoácido

<220>
<221> MOD_RES
<222> (5)..(5)
<223> Cualquier aminoácido

<400> 9

```
Lys Xaa Asp Gly Xaa Arg
1               5
```

<210> 10
<211> 9
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 10

```
Ala Thr Pro Lys Ile Asp Gly Ile Arg
1               5
```

<210> 11
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 11

```
Glu Gly Ser Asp Gly Glu Ile Ser
1               5
```

<210> 12
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 12

```
Tyr Trp Phe Asp Tyr
1               5
```

<210> 13
<211> 6

<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 13

Glu Gly Val Met Ile Arg
1 5

<210> 14
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 14

Leu Leu Lys Met Lys
1 5

<210> 15
<211> 9
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 15

Cys Glu Leu Lys Leu Asp Gly Leu Ala
1 5

<210> 16
<211> 9
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 16

Val Glu His Lys Val Asp Gly Leu Ser
1 5

<210> 17
<211> 9
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 17

Cys Glu Pro Lys Leu Asp Gly Leu Ala
1 5

<210> 18
<211> 9
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 18

Cys Glu Leu Lys Leu Asp Gly Val Ala
1 5

<210> 19
<211> 9
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 19

Ala Glu Ile Lys Tyr Asp Gly Val Arg
1 5

<210> 20
<211> 9
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 20

Cys Glu Tyr Lys Tyr Asp Gly Gln Arg
1 5

<210> 21
<211> 9
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 21

Val Asp Tyr Lys Tyr Asp Gly Glu Arg
1 5

<210> 22
<211> 9
<212> PRT
<213> Desconocido

<220>

<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 22

Phe Glu Ile Lys Tyr Asp Gly Ala Arg
1 5

<210> 23
<211> 9
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 23

Phe Glu Gly Lys Trp Asp Gly Tyr Arg
1 5

<210> 24
<211> 9
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 24

Ala Arg Glu Lys Ile His Gly Thr Asn
1 5

<210> 25
<211> 9
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 25

Ala Cys Glu Lys Val His Gly Thr Asn
1 5

<210> 26
<211> 9
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 26

Ile Leu Thr Lys Glu Asp Gly Ser Leu
1 5

<210> 27
<211> 9

<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 27

Val Glu Glu Lys Val Asp Gly Tyr Asn
1 5

<210> 28
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 28

Leu Glu Val Arg Gly Glu Val Phe
1 5

<210> 29
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 29

Val Glu Val Arg Gly Glu Cys Tyr
1 5

<210> 30
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 30

Leu Glu Val Arg Gly Glu Val Tyr
1 5

<210> 31
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 31

Leu Glu Ala Arg Gly Glu Ala Phe
1 5

<210> 32
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 32

Phe Met Leu Asp Gly Glu Leu Met
1 5

<210> 33
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 33

Phe Ile Leu Asp Thr Glu Ala Val
1 5

<210> 34
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 34

Phe Ile Ile Glu Gly Glu Ile Val
1 5

<210> 35
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 35

Ala Ile Val Glu Gly Glu Leu Val
1 5

<210> 36
<211> 8
<212> PRT
<213> Desconocido

<220>

<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 36

Val Val Leu Asp Gly Glu Ala Val
1 5

<210> 37
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 37

Tyr Gln Val Phe Gly Glu Phe Ala
1 5

<210> 38
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 38

Leu Val Leu Asn Gly Glu Leu Phe
1 5

<210> 39
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 39

Phe Thr Ala Asn Phe Glu Phe Val
1 5

<210> 40
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 40

Leu Ile Leu Val Gly Glu Met Ala
1 5

<210> 41
<211> 5
<212> PRT

<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 41

Phe Cys Tyr Gly Val
1 5

<210> 42
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 42

Phe Leu Tyr Thr Val
1 5

<210> 43
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 43

Thr Phe Tyr Ala Leu
1 5

<210> 44
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 44

Ile Cys His Gly Leu
1 5

<210> 45
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 45

Asn Ala Tyr Gly Ile
1 5

<210> 46
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 46

Phe Val Tyr Gly Leu
1 5

<210> 47
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 47

Lys Leu Tyr Ala Ile
1 5

<210> 48
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 48

Tyr Ala Phe Asp Ile
1 5

<210> 49
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 49

Phe Leu Phe Asp Leu
1 5

<210> 50
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 50

Asn Leu Phe Asp Val
1 5

<210> 51
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 51

Trp Ala Phe Asp Leu
1 5

<210> 52
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 52

Tyr Val Phe Asp Ile
1 5

<210> 53
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 53

Phe Ala Phe Asp Ile
1 5

<210> 54
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 54

Ile Leu Leu Asn Ala
1 5

<210> 55
<211> 5
<212> PRT
<213> Desconocido

<220>

<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 55

Phe Leu Phe Asp Val
1 5

<210> 56
<211> 6
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 56

Asp Gly Val Val Ile Lys
1 5

<210> 57
<211> 6
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 57

Asp Gly Ile Val Ile Lys
1 5

<210> 58
<211> 6
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 58

Asp Gly Val Val Val Lys
1 5

<210> 59
<211> 6
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 59

Asp Gly Thr Val Leu Lys
1 5

<210> 60
<211> 6

<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 60

Glu Gly Leu Ile Val Lys
1 5

<210> 61
<211> 6
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 61

Glu Gly Leu Met Val Lys
1 5

<210> 62
<211> 6
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 62

Glu Gly Val Met Val Lys
1 5

<210> 63
<211> 6
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 63

Glu Gly Leu Met Ala Lys
1 5

<210> 64
<211> 6
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 64

Glu Gly Val Ile Ala Lys
1 5

<210> 65
<211> 6
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 65

Glu Gly Tyr Val Leu Lys
1 5

<210> 66
<211> 6
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 66

Glu Gly Val Val Ile Arg
1 5

<210> 67
<211> 6
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 67

Glu Gly Tyr Val Ala Val
1 5

<210> 68
<211> 6
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 68

Glu Gly Ile Ile Met Lys
1 5

<210> 69
<211> 5
<212> PRT
<213> Desconocido

<220>

<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 69

Ala Val Ala Phe Lys
1 5

<210> 70
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 70

Ala Ile Ala Tyr Lys
1 5

<210> 71
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 71

Ala Leu Ala Tyr Lys
1 5

<210> 72
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 72

Trp Trp Lys Met Lys
1 5

<210> 73
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 73

Trp Leu Lys Leu Lys
1 5

<210> 74
<211> 5
<212> PRT

<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 74

Trp Ile Lys Leu Lys
1 5

<210> 75
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 75

Trp Leu Lys Ile Lys
1 5

<210> 76
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 76

Trp Val Lys Asp Lys
1 5

<210> 77
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 77

Ala Ile Lys Cys Lys
1 5

<210> 78
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 78

Ile Ile Lys Leu Arg
1 5

<210> 79
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 79

His Phe Lys Ile Lys
1 5

<210> 80
<211> 5
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción del Desconocido: Péptido de motivo ligasa"

<400> 80

Ile Val Lys Tyr Val
1 5

<210> 81
<211> 20
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<223> /nota="Descripción de la Secuencia Artificial: Cebador sintético"

<400> 81
ctgctgtacc gtacatccgc 20

<210> 82
<211> 23
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<223> /nota="Descripción de la Secuencia Artificial: Cebador sintético"

<400> 82
ctgccccggg ttcctcattc tct 23

<210> 83
<211> 41
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<223> /nota="Descripción de la Secuencia Artificial: Oligonuclétido Sintético"

<400> 83
ccactacgcc tccgctttcc tctctatggg cagtcggtga t 41

<210> 84
<211> 87

<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<223> /nota="Descripción de la Secuencia Artificial: Oligonucléotido sintético"

<400> 84

ctgccccggg ttcctcattc tctattcgct gctgtaccgt acatccgcct tggccgtaca 60

gcagatcacc gactgcccat agagagg 87

<210> 85
<211> 105
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<223> /nota="Descripción de la Secuencia Artificial: Oligonucléotido sintético"

<400> 85

ccactacgcc tccgctttcc tctctatggg cagtcggtga tctgctgtac ggccaaggcg 60

gatgtacggt acagcagcga atagagaatg aggaacccgg ggcag 105

<210> 86
<211> 28
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<223> /nota="Descripción de la Secuencia Artificial: Oligonucléotido sintético"

<400> 86
gcggatgtac ggtacagcag cgaataga 28

<210> 87
<211> 14
<212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<223> /nota="Descripción de la Secuencia Artificial: Oligonucléotido sintético"

<400> 87
gcccatagag agga 14

REIVINDICACIONES

1. Una ligasa de ADN Hin mutante que tiene al menos 70 %, al menos 80 % o al menos 90 %, o al menos 95 % o al menos 99 % de identidad con la secuencia de la ligasa de ADN Hin que se proporciona en la Tabla 1C, cuya ligasa comprende dos mutaciones de aminoácidos en las posiciones 93 y 193 de la secuencia de la ligasa de ADN Hin proporcionada en la Tabla 1C que consisten en cambiar la glicina en la posición 193 a ácido aspártico o ácido glutámico y cambiar la treonina en la posición 93 a serina.

2. Una molécula de ácido nucleico que codifica la ligasa de cualquiera de las reivindicaciones anteriores.

3. Un método de ligación que comprende ligar enzimáticamente un primer oligonucleótido y un segundo oligonucleótido, en donde:
el primer oligonucleótido tiene N residuos de nucleótidos de longitud, donde N es de 2 a 7, preferentemente en donde N es 2 o 3, y en donde la ligación se realiza mediante una ligasa de ADN Hin mutante de la reivindicación 1.

4. El método de la reivindicación 3, en donde el primer y el segundo oligonucleótidos son monocatenarios.

5. El método de la reivindicación 3 o 4, en donde el primer oligonucleótido y el segundo oligonucleótido son polinucleótidos diferentes y no se hibridan con el mismo molde.

6. El método de la reivindicación 3, en donde se realiza una ligación del primer y segundo oligonucleótidos antes de la amplificación en un ensayo de ligación por proximidad.

7. El método de la reivindicación 3 o 4, que comprende ligar un extremo terminal proximal del primer oligonucleótido a un extremo terminal proximal del segundo oligonucleótido, en donde el primer y el segundo oligonucleótidos se hibridan a un molde de manera que sus extremos terminales proximales sean adyacentes entre sí.

8. Un método de ligación dependiente del molde que comprende ligar enzimáticamente un primer oligonucleótido y un segundo oligonucleótido, en donde:

   a) el primer y segundo oligonucleótidos se hibridan ambos con el molde de manera que sus extremos terminales proximales sean adyacentes entre sí;
   b) el primer oligonucleótido es de longitud N y comprende una porción proximal que tiene L nucleótidos de largo; y

en donde la ligación se realiza por una ligasa de ADN Hin mutante de la reivindicación 1.

9. El método de cualquiera de las reivindicaciones anteriores, en donde el segundo oligonucleótido no tiene más de 6 nucleótidos de longitud.

10. El método de cualquiera de las reivindicaciones anteriores, en donde el extremo terminal proximal del primer oligonucleótido es su extremo 3' terminal y el extremo terminal proximal del segundo oligonucleótido es su extremo 5' terminal o en donde el extremo terminal proximal del primer oligonucleótido es su extremo 5' terminal y el extremo terminal proximal del segundo oligonucleótido es su extremo 3' terminal.

11. El método de cualquiera de las reivindicaciones anteriores, en donde el primer oligonucleótido se marca y/o en donde el segundo oligonucleótido o molde se marca, preferentemente en donde al menos dos del primer oligonucleótido, el segundo oligonucleótido y el molde se marcan.

12. El método de cualquiera de las reivindicaciones anteriores, en donde al menos uno del primer oligonucleótido, el molde y/o el segundo oligonucleótido se inmoviliza.

13. El método de cualquiera de las reivindicaciones anteriores, en donde la ligación se repite al menos una vez, opcionalmente en donde el método comprende una reacción en cadena de la ligasa.

14. El método de la reivindicación 13, en donde cualquier producto de ligación de una reacción de ligación anterior se usa como un tercer oligonucleótido de una siguiente ligación, opcionalmente en donde cualquier primer y/o segundo oligonucleótido no ligado se vuelve no ligable antes de iniciar la siguiente ligación y/o que comprende además detectar si el primer oligonucleótido se ligó al segundo oligonucleótido antes de repetir la ligación.

15. El método de la reivindicación 12, en donde la ligación se realiza en presencia de múltiples oligonucleótidos que son al menos parcialmente complementarios a la misma región objetivo en el molde adyacente al extremo terminal proximal del segundo oligonucleótido.

16. Un kit que comprende una ligasa de ADN Hin mutante de la reivindicación 1 o la reivindicación 2 y un

oligonucleótido que tiene una porción que está ligada que tiene menos de N nucleótidos de longitud, donde N es de 2 a 7.

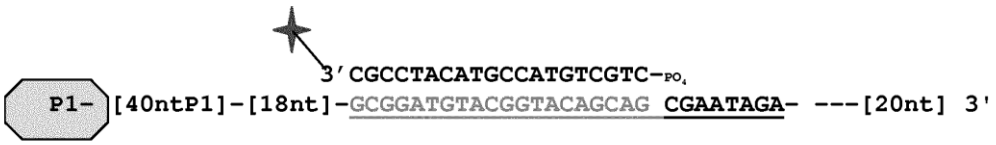
17. El método de una cualquiera de las reivindicaciones 3 a 15, en donde el primer oligonucleótido es una sonda y el segundo oligonucleótido es un cebador.

18. El método de la reivindicación 8, en donde la porción proximal que tiene L nucleótidos de largo (i) se hibrida perfectamente con el molde y en donde (ii) el nucleótido L+1 no coincide con el molde.

19. El método de la reivindicación 18, en donde L es de 2 a 8, y además es menor que 6 si el extremo terminal proximal del primer oligonucleótido es su extremo 3' terminal,

FIGURA 1

(A)


3′CGCCTACATGCCATGTCGTC–PO4
P1–
[40ntP1]–[18nt]–GCGGATGTACGGTACAGCAG CGAATAGA– ---[20nt] 3'

(B)

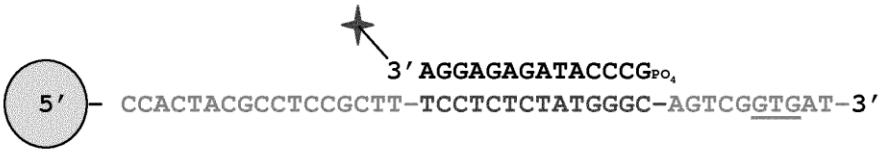

3′AGGAGAGATACCCGPO4
5′
– CCACTACGCCTCCGCTT–TCCTCTCTATGGGC–AGTCGGTGAT–3′

FIGURA 2

Ligasa de ADN CV, identificación de GenBank AAC96909.1, del virus de Paramecium bursaria Chlorella 1:

MAITKPLLAATLENIEDVQFPCLATPKIDGIRSVKQTQMLSRTFKPIRNSVMNRLLTELL
PEGSDGEISIEGATFQDTTSAVMTGHKMYNAKFSYYWFDYVTDDPLKKYIDRVEDMK
NYITVHPHILEHAQVKIIPLIPVEINNITELLQYERDVLSKGFEGVMIRKPDGKYKFGRST
LKEGILLKMKQFKDAEATIISMTALFKNTNTKTKDNFGYSKRSTHKSGKVEEDVMGSIE
VDYDGVVFSIGTGFDADQRRDFWQNKESYIGKMVKFKYFEMGSKDCPRFPVFIGIRHE
EDR

Ligasa de ADN MnM, identificación de GenBank YP_333052.1, de Burkholderia pseudomallei 1710b (secuencia equivalente a ABA50091)

MSGVPYGFKPNLAATLTKPELIKFPVWASPKIDGIRCVFFGGVAYSRSLKPIPNPVVQEF
AKAYANLLEGLDGELTVGSPTDANCMQNSMAVMSKAAAPDFTFHVFDWFHPAQAHI
EFWQRSDVVEDRIVQFYDRYPEVDIRAAPQVLCTSLAHLDTNEARWLADGYEGMMIR
DHCGRYKFGRSTEREGGLVKVKRFTDAEAIVIGFEEEMHNANEAKRDATGRTERSTSK
AGLHGKGTLGALVVKNERGIVFNIGTGFTAAQRADYWANHPSLFGKMVKFKHFDHGT
VDAPRHPVFIGFRHPEDM

Ligasa de ADN Hin, identificación de GenBank P44121, de Haemophilus influenza

MKFYRTLLLFFASSFAFANSDLMLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTR
QGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEIS**T**ITKSFKGDGWEKLKLYVFDVPDAE
GNLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGEGVVVRNPNAP
YERKRSSQILKLKTAR**G**EECTVIAHHKGKGQFENVMGALTCKNHRGEFKIGSGFNLNE
RENPPPIGSVITYKYRGITNSGKPRFATYWREKK

Ligasa de ADN DLX, ligasa artificial derivada de la ligasa de ADN Hin de Haemophilus influenza:

MKFYRTLLLFFASSFAFANSDLMLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTR
QGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEIS**S**ITKSFKGDGWEKLKLYVFDVPDAEG
NLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGEGVVVRNPNAPY
ERKRSSQILKLKTARGEECTVIAHHKGKGQFENVMGALTCKNHRGEFKIGSGFNLNER
ENPPPIGSVITYKYRGITNSGKPRFATYWREKK

Ligasa de ADN DLXd, ligasa artificial derivada de la ligasa Hin D de Haemophilus influenza:

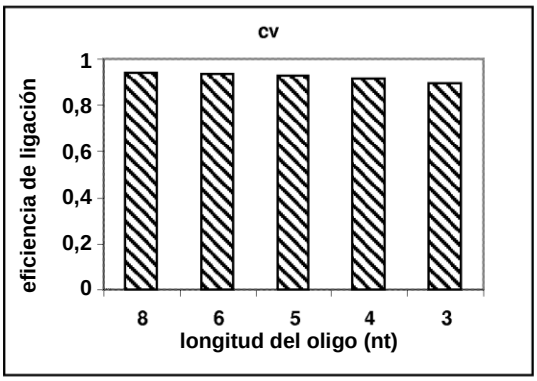
MKFYRTLLLFFASSFAFANSDLMLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTR
QGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEIS**S**ITKSFKGDGWEKLKLYVFDVPDAEG
NLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGEGVVVRNPNAPY
ERKRSSQILKLKTAR**D**EECTVIAHHKGKGQFENVMGALTCKNHRGEFKIGSGFNLNER
ENPPPIGSVITYKYRGITNSGKPRFATYWREKK

Ligasa de ADN DLXd2 (Gammaproteobacteria, Haemophilus influenza) (modificado)

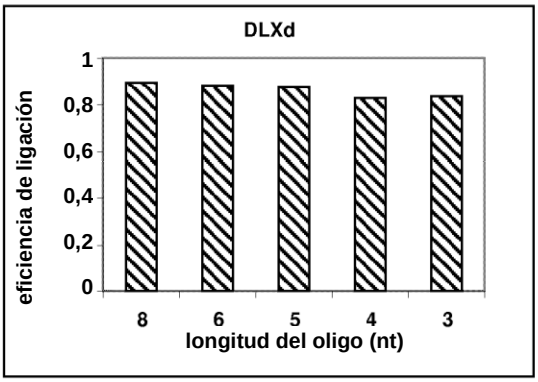
MLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTRQGQRLSPPAYFIKDFPPFAIDGE
LFSERNHFEEIS**S**ITKSFKGDGWEKLKLYVFDVPDAEGNLFERLAKLKAHLLEHPTTYIE
IIEQIPVKDKTHLYQFLAQVENLQGEGVVVRNPNAPYERKRSSQILKLKTAR**D**EECTVIA
HHKGKGQFENVMGALTCKNHRGEFKIGSGFNLNERENPPPIGSVITYKYRGITNSGKPR
FATYWREKK

FIGURA 3

(A)


cv
eficiencia de ligación
1
0,8
0,6
0,4
0,2
0
8
6
5
4
3
longitud del oligo (nt)

(B)


DLXd
eficiencia de ligación
1
0,8
0,6
0,4
0,2
0
8
6
5
4
3
longitud del oligo (nt)

(C)

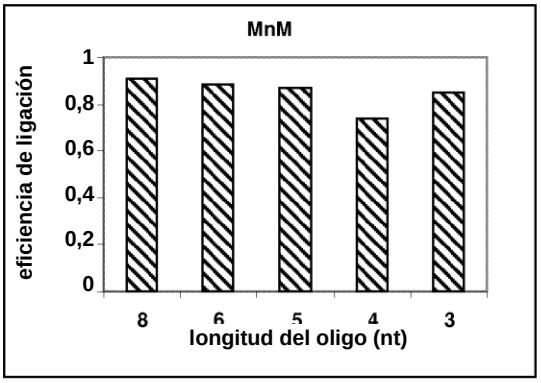
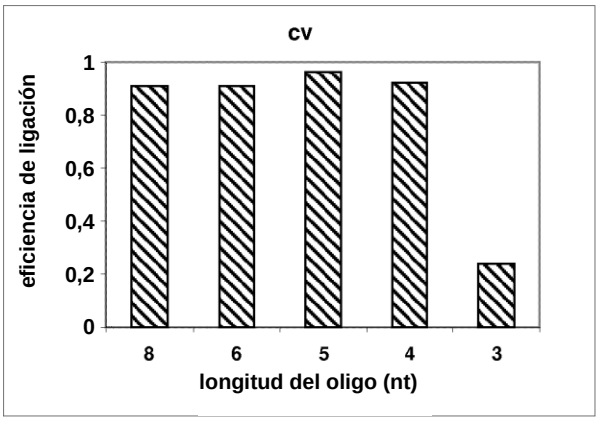
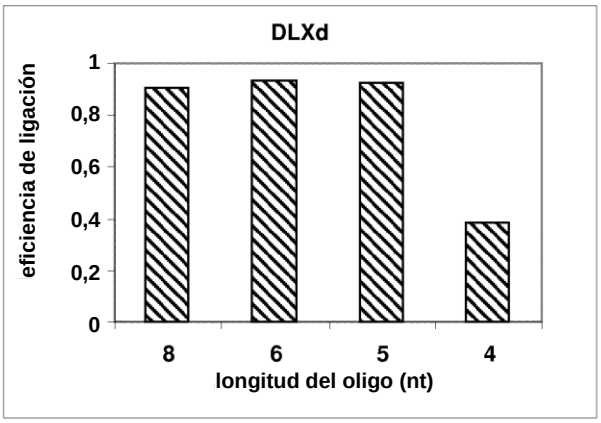

MnM
eficiencia de ligación
1
0,8
0,6
0,4
0,2
0
8
6
5
4
3
longitud del oligo (nt)

FIGURA 4

(A)


cv
eficiencia de ligación
1
0,8
0,6
0,4
0,2
0
8
6
5
4
3
longitud del oligo (nt)

(B)


DLXd
eficiencia de ligación
1
0,8
0,6
0,4
0,2
0
8
6
5
4
longitud del oligo (nt)

(C)

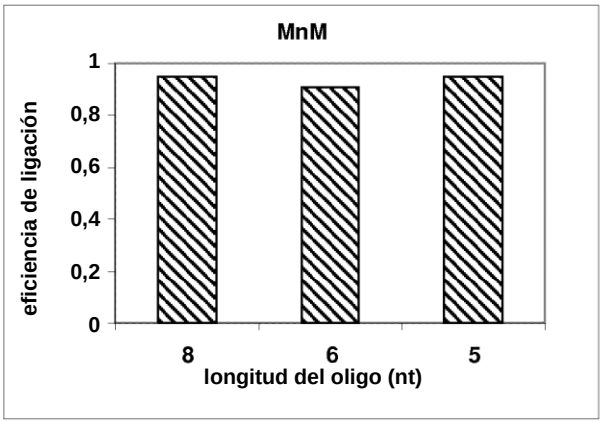
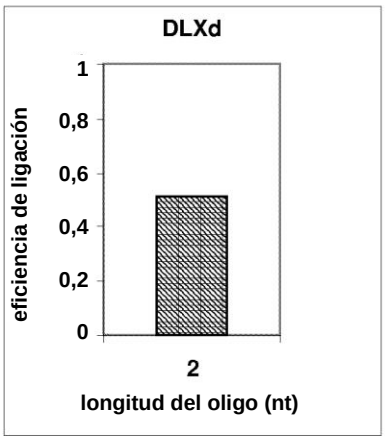

MnM
eficiencia de ligación
1
0,8
0,6
0,4
0,2
0
8
6
5
longitud del oligo (nt)

FIGURA 5


DLXd
eficiencia de ligación
1
0,8
0,6
0,4
0,2
0
2
longitud del oligo (nt)