

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 645 953**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **17.01.2012 PCT/US2012/021465**

87 Fecha y número de publicación internacional: **26.07.2012 WO12099832**

96 Fecha de presentación y número de la solicitud europea: **17.01.2012 E 12701312 (6)**

97 Fecha y número de publicación de la concesión europea: **09.08.2017 EP 2665815**

54 Título: **Ligamiento enzimático 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:
11.12.2017

73 Titular/es:

LIFE TECHNOLOGIES CORPORATION (100.0%)

**5823 Newton Drive
Carlsbad, CA 92008, US**

72 Inventor/es:

HENDRICKS, STEPHEN

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 645 953 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

Ligamiento enzimático de ácidos nucleicos

Esta solicitud reivindica la prioridad de las siguientes aplicaciones provisionales de EE. UU.: No. de solicitud 61/474.205 presentada el 11 de abril, 2011; No. de solicitud 61/474.168 presentada el 1 de abril de 2011; No. de solicitud 61/433.502 presentada el 17 de enero de 2011; y No. de solicitud 61/433.488 presentada el 17 de enero de 2011.

Antecedentes de la invención

Las ADN ligasas pueden unir polinucleótidos entre sí, por ejemplo, catalizando la formación de un enlace fosfodiéster en roturas monocatenarias o bicatenarias en ADN dúplex. Las ligasas pueden ser sensibles al grado de hibridación entre las cadenas de ácido nucleico opuestas en un dúplex. Por ejemplo, el ligamiento exitoso puede ocurrir con menos frecuencia (o no hacerlo) cuando una cadena a ligar a una cadena adyacente está en un dúplex y no es complementaria a su cadena opuesta en el dúplex. En algunos casos, una falta de coincidencia de un solo nucleótido entre las cadenas en un dúplex puede afectar significativamente o impedir el ligamiento. La capacidad de las ligasas para la discriminación basada en la hibridación, incluida la discriminación de un solo nucleótido, ha conducido al desarrollo de técnicas de detección mediadas por ligasa (p. ej., Landegren, U., *Bioessays*, 15 (11): 761-765 (1993), y Barany, *PNAS USA*, 88 (1): 189 - 193 (1991)). Se ha demostrado que la amplificación de la señal lineal basada en ligasa conocida como LDR (es decir, la reacción de detección de ligasa), combinada con la amplificación de dianas específicas del gen basada en PCR (es decir, la reacción en cadena de la polimerasa), es una herramienta útil en la detección de mutaciones genéticas en el cáncer y enfermedades. Las técnicas de PCR/LDR típicamente se basan en dos propiedades de una ADN ligasa: (i) especificidad, y (ii) termoestabilidad. Odell et al., *Nucleic Acids Research*, 2003, vol. 31, No. 17, páginas 5090 a 5100, describen el uso de CV ligasa para unir oligonucleótidos pequeños, pero no en un ligamiento repetitivo. La eficiencia para una sonda 10mer, 8mer y 6mer fue entre 75 y 80%, mientras que la 4mer no se ligó eficientemente. El documento WO2007121489 describe la secuenciación por ligamiento, es decir, el ligamiento repetitivo, usando ligasa T4 y oligonucleótidos de 6 nucleótidos de longitud o mayores.

Sumario

Esta solicitud se relaciona con reactivos y métodos de ligamiento. Entre otras cosas, se proporcionan métodos y reactivos para ligar ácidos nucleicos a otros ácidos nucleicos, incluyendo el ligamiento de dos polinucleótidos. La invención proporciona un método de ligamiento repetitiva, que comprende ligar enzimáticamente un oligonucleótido ("sonda") y un polinucleótido ("cebador") usando una ADN ligasa del Virus Chlorella (CV ligasa) o fragmento funcional de la misma, en la que: la sonda contiene N restos nucleotídicos de longitud, donde N es de 2 a 7. Cualquiera o ambos polinucleótidos pueden ser monocatenarios o bicatenarios. Uno o ambos polinucleótidos pueden ser un oligonucleótido corto. En algunas realizaciones, los oligonucleótidos y/o polinucleótidos de diferentes longitudes se pueden ligar entre sí. El ligamiento puede ser enzimático y los ligamientos pueden ser dependientes de la plantilla o independientes de la plantilla. Los ácidos nucleicos que están ligados o implicados en la reacción de ligamiento pueden marcarse o no marcarse e inmovilizarse o estar en disolución.

Algunas realizaciones implican el ligamiento independiente de plantilla de polinucleótidos bicatenarios o monocatenarios (por ejemplo, oligonucleótidos). Los polinucleótidos monocatenarios opcionalmente no se hibridan con otros polinucleótidos. En algunas realizaciones, un oligonucleótido se puede hibridar o asociar de otro modo con todo o parte de un saliente u otra región monocatenaria de un ácido nucleico dúplex y ligarse a un extremo libre de una cadena del dúplex o a otro oligonucleótido que sea hibridado o asociado de otra manera con un saliente u otra porción monocatenaria del dúplex.

Algunas realizaciones implican el ligamiento de polinucleótidos monocatenarios (por ejemplo, oligonucleótidos). Opcionalmente, los polinucleótidos monocatenarios se hibridan adyacentes o cercanos entre sí en otro único polinucleótido. En algunas realizaciones, un oligonucleótido se puede hibridar o asociar de otro modo con todo o parte de un saliente u otra región monocatenaria de un ácido nucleico dúplex y ligarse a un extremo libre de una cadena del dúplex o a otro oligonucleótido que es hibridado o asociado de otra manera con un saliente u otra porción monocatenaria del dúplex.

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

Los desapareamientos de bases de nucleótidos entre un primer y/o segundo oligonucleótidos y un tercer oligonucleótido o polinucleótido pueden afectar la eficacia del ligamiento. Por ejemplo, los desapareamientos en la posición terminal de uno o ambos oligonucleótidos primero y segundo pueden afectar la eficacia de ligamiento, reduciendo la probabilidad de un ligamiento exitoso o el ligamiento excluido por completo. Los desapareamientos en

otras posiciones o en múltiples posiciones también pueden afectar a la eficacia del ligamiento, reduciendo la probabilidad de un ligamiento exitoso o impidiendo por completo el ligamiento.

También se proporcionan métodos y reactivos para realizar ligamientos múltiples secuencialmente, en paralelo, o ambas secuencialmente y en paralelo.

- 5 Opcionalmente, uno o más de los iniciadores, sondas o plantillas están marcados. Por ejemplo, la sonda puede marcarse. De lo contrario, el cebador o plantilla está marcado.

En algunas realizaciones, se proporcionan métodos de ligamiento que proporcionan información sobre la secuencia de un ácido nucleico. Por ejemplo, en algunos aspectos, se puede realizar un ligamiento en presencia de múltiples oligonucleótidos que son al menos parcialmente complementarios a una región diana en una plantilla. Se pueden usar sondas de oligonucleótidos que se hibridan o se asocian de otro modo con una plantilla adyacente o cercana a un extremo de un cebador o sonda que se hibrida o se asocia de otro modo con una plantilla. Pueden usarse múltiples sondas de oligonucleótidos, cada una al menos parcialmente complementaria a una región de una plantilla, para determinar la información de secuencia de una manera dependiente de plantilla como se conoce en la técnica. Por ejemplo, el ligamiento de oligonucleótidos se usa para determinar la información de la secuencia de ácidos nucleicos en el sistema SOLiD (Life Technologies-Applied Biosystems, Carlsbad, CA). De acuerdo con algunas realizaciones, la información de la secuencia se determina ligando las sondas oligonucleotídicas a un cebador oligonucleotídico de una manera dependiente de la secuencia de la plantilla, por ejemplo, usando un 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 ó 2 nucleótidos. Los marcadores pueden proporcionar información sobre la secuencia de la sonda.

En algunas realizaciones, el ligamiento puede realizarse mediante una "ligasa de huella pequeña" (en este documento "SFL") que puede ligar polinucleótidos cortos. Los SFL se pueden usar en cada una de las realizaciones de ligamientos discutidas anteriormente y en el resto de esta descripción, así como en otras realizaciones de ligamientos es una ADN ligasa del virus Chlorella (CV ligasa) o fragmento funcional de la misma. Por ejemplo, en algunas realizaciones, un SFL puede ligar los extremos de un primer oligonucleótido y un segundo oligonucleótido. El primer oligonucleótido puede ser un cebador y el segundo oligonucleótido puede ser una sonda, hibridándose cada uno de ellos o asociándose de otro modo con una porción de un tercer oligonucleótido o un polinucleótido.

En algunas realizaciones, el SFL puede ligar oligonucleótidos que tienen 8, 7, 6, 5, 4, 3 ó 2 nucleótidos de longitud a un polinucleótido. El ligamiento 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 ó 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 proporcionan kits que comprenden una ADN ligasa del virus Chlorella (CV ligasa) o un fragmento funcional de la misma. Opcionalmente, el kit también puede incluir una o más sondas oligonucleotídicas de menos de 12 nucleótidos de longitud (por ejemplo, no más de 8, 6, 5, 4, 3 ó 2 nucleótidos de longitud). Opcionalmente, el kit incluye una CV ligasa y una o más sondas de oligonucleótidos de menos de 6 nucleótidos de longitud.

Breve descripción de los dibujos

Figura 1. (A) Substrato de ligamiento de perlas covalentes de P1 extendido. Se muestra un sustrato de ligamiento diseñado para imitar las condiciones utilizadas en la secuenciación SOLiD™. La Figura 1(A) representa ADN unido covalentemente a una perla (tal como una perla magnética, por ejemplo, de 0,1 a 1 µm de diámetro). El ADN en esta perla se enriqueció mediante PCR usando la secuencia P1 de 40 nucleótidos. Para la región GCGGATGTACGGTACAGCAG, se templea un "cebador" complementario de 20 mer que tiene un 5' PO₄ 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 saliente CGAATAGA, las sondas complementarias pueden hibridarse para demostrar la reacción de ligamiento. (B) sustrato de ligamiento de perlas covalentes P1. Se muestra un sustrato de ligamiento 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 subrayados son opuestos al triplete de inosina de las sondas que tienen la estructura Dye-5' III (s)-xy-NNN3', como se describe con más detalle en el presente documento.

Figura 2: Secuencias de ejemplos de ligasas. La identificación GenBank y el organismo fuente se proporcionan cuando corresponde. Tres ligasas artificiales también se proporcionan. DLX difiere de Hin DNA Ligasa en 1 aminoácido, designado como la letra en negrita subrayada. DLXd difiere de Hin DNA Ligasa en 2 aminoácidos, designado como las letras en negrita subrayadas. DLXd2 es 22 aminoácidos más corto que Hin DNA Ligasa y se diferencia de Hin DNA Ligasa en 2 aminoácidos, designado como las letras en negrita subrayadas.

Figura 3: Ligamiento de oligonucleótidos cortos con diversas ligasas. Ligamiento de oligonucleótidos cortos con (A) CV ligasa, (B) DLXd ligasa y (C) MnM ligasa. Las reacciones de ligamiento directo se realizaron en las siguientes condiciones: ligasa de 2,0 µM, oligo corto de 2-5 µM, cebador/plantilla 2,0 nM (unido a perlas magnéticas) y se

continuó durante 20 minutos a 15°C. La eficacia de ligamiento se calculó como la relación de áreas de pico determinadas por CE en la que se utilizó un cebador marcado con FAM. Eficiencia = ligado/(ligado + no ligado).

Figura 4: Ligamiento de oligonucleótidos cortos con diversas ligasas. Ligamiento de oligonucleótidos cortos con (A) CV ligasa, (B) DLXd ligasa y (C) MnM ligasa. Las reacciones de ligamiento inverso se realizaron en las siguientes condiciones: ligasa de 2,0 µM, oligo corto de 2-5 µM, cebador/plantilla 2,0 nM (unido a perlas magnéticas) y se continuó durante 20 minutos a 15°C. La eficacia de ligamiento se calculó como la relación de áreas de pico determinada por CE en la que se utilizó un cebador marcado con FAM. Eficiencia = ligado/(ligado + no ligado).

Figura 5: Ligamiento de 2-mers. La reacción de ligamiento en la dirección directa se realizó en las siguientes condiciones: 2,0 µl de DLXd, 123 µM de dinucleótido (5'-CG-3'), cebador/plantilla 2,0 nM (unido a perlas magnéticas) y se continuó durante 20 minutos a 15°C. La eficacia del ligamiento se calculó como la relación de las áreas de pico determinadas por CE, en la que se utilizó un cebador marcado con FAM. Eficiencia = ligado/(ligado + no ligado).

Definiciones

"Degenerado", 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 por un valor numérico asignado 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 se le puede asignar la posición 1. Por lo tanto, en un conjunto de sondas de extensión de la estructura 3'- XXXNXXXX-5', el N está en la posición 4. Se dice que una posición está k-veces degenerada si puede ser ocupada por nucleósidos que tienen cualquiera de las k identidades diferentes. Por ejemplo, una posición que puede ser ocupada por nucleósidos que comprenden cualquiera de 4 bases diferentes está 4 veces degenerada.

En líneas similares, debe entenderse que una afirmación de que se ha producido un resultado (por ejemplo, un ligamiento, unión) pretende indicar que el resultado se ha producido a un nivel significativo o sustancial o a un nivel mejorado en comparación con cuando no se ha producido. Por ejemplo, se dice que el ligamiento no ha ocurrido si no es significativamente, insustancialmente o muy reducida (p. ej., reducida en al menos un 80%, 90%, 95% o 99% en comparación con cuando ocurre el ligamiento (p. ej., bajo las condiciones descritas en el último párrafo).

Los términos "micropartícula", "perlas", "microperlas", etc., se refieren a partículas (opcionalmente pero no necesariamente de forma esférica) que tienen una longitud de sección transversal mínima (por ejemplo, diámetro) de 50 micras o menos, preferiblemente 10 micras o menos, 3 micras o menos, aproximadamente 1 micra o menos, aproximadamente 0,5 micras o menos, por ejemplo, aproximadamente 0,1, 0,2, 0,3, ó 0,4 micras, o más pequeñas (p. ej., menos de 1 nanómetro, aproximadamente 1-10 nanómetros, aproximadamente 10-100 nanómetros, o aproximadamente 100-500 nanómetros). Las micropartículas (p. ej., Dynabeads de Dynal, Oslo, Noruega) pueden estar hechas de una variedad de materiales inorgánicos u orgánicos que incluyen, pero no se limitan a, vidrio (por ejemplo, vidrio de poro controlado), sílice, zirconia, poliestireno reticulado, poliacrilato, polimetilmetacrilato, dióxido de titanio, látex, poliestireno, etc. La magnetización puede facilitar la recolección y concentración de los reactivos unidos a micropartículas (por ejemplo, polinucleótidos o ligasas) después de la amplificación, y facilitar pasos adicionales (por ejemplo, lavados, eliminación de reactivos, etc.). En ciertas realizaciones de la invención, se puede usar una población de micropartículas que tienen diferentes tamaños de formas y/o colores. Las micropartículas se pueden codificar opcionalmente, por ejemplo, con puntos cuánticos de manera que cada micropartícula se pueda identificar 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 longitud. El término también puede usarse como una referencia al propio polinucleótido o molécula polipeptídica o a una parte relevante de la misma. La información de la secuencia de polinucleótidos se refiere a la sucesión de bases de nucleótidos en el polinucleótido, y en un polipéptido se refiere a la sucesión de cadenas laterales de aminoácidos en el polipéptido o parte del mismo. 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 este documento se presentan en una orientación de 5' a 3' a menos que se indique lo contrario.

"Dúplex perfectamente combinado" en referencia a las sondas y los polinucleótidos de plantilla significa que uno forma una estructura bicatenaria con el otro de manera que cada nucleósido en la estructura bicatenaria sufre un emparejamiento de bases Watson-Crick con un nucleósido en el otro. La expresión también comprende el emparejamiento de análogos de nucleósidos, tales como desoxiinosina, nucleósidos con bases de 2-aminopurina, y similares, que pueden emplearse para reducir la degeneración de las sondas, ya sea que dicho emparejamiento implique o no la formación de enlaces de hidrógeno.

El término "polimorfismo" tiene 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.

- 5 "Sondas", "oligonucleótidos" o "cebadores" pretenden ser términos intercambiables en el presente documento, de modo que cualquiera de estos puede tomarse como una referencia del otro. Estos son polinucleótidos no necesariamente limitados a cualquier longitud. Donde así se desee, estos pueden tener menos de 100 nucleótidos de longitud, a veces menos de 30 nucleótidos de longitud, 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 ó 4 nucleótidos de longitud.
- 10 Un "polinucleótido", también llamado "ácido nucleico", es un polímero lineal de dos o más nucleótidos unidos por enlaces internucleosídicos covalentes, o una variante o fragmento funcional de los mismos. Una secuencia de letras, tal como "ATGCCTG", pretende representar una secuencia de polinucleótidos en el orden de 5' a 3' de izquierda a derecha a menos que se especifique lo contrario. En ejemplos que ocurren naturalmente de estos, el enlace internucleosídico es típicamente un enlace fosfodiéster. Sin embargo, otros ejemplos opcionalmente comprenden otros enlaces internucleosídicos, tales como enlaces fosforotiolato y pueden o no comprender 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 ácidos nucleicos específicos de secuencia sintéticos, que proporcionan que los polímeros contengan nucleobases en una configuración que permita el emparejamiento de bases y el apilamiento de bases, tal como se encuentra en el ADN y el ARN.
- 15 Como se usa en este documento, "polinucleótido", "oligonucleótido", "sonda", "cebador", "plantilla", "ácido nucleico" y similares se pueden tomar para referirse a una población o conjunto de moléculas individuales que son sustancialmente idénticas en toda su longitud o en una porción relevante de interés. Por ejemplo, el término "plantilla" puede indicar una pluralidad de moléculas plantilla 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 la(s) posición(es) no degenerada(s) y difieren en la secuencia en las posiciones degeneradas. Por lo tanto, la referencia "un" polinucleótido (p. ej., "un" cebador, sonda, oligonucleótido, plantilla, etc.) puede tomarse para indicar una población de moléculas de polinucleótidos que son sustancialmente idénticas respecto a al menos una porción de interés, de modo que la naturaleza plural de la población no necesita ser explícitamente indicada, pero puede, si así lo desea. Estos términos también pretenden proporcionar un soporte adecuado para una reivindicación que especifique explícitamente una sola molécula de polinucleótido propia. 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. Preferiblemente, al menos el 90%, al menos el 95%, al menos el 99% o más de los miembros de una población son sustancialmente idénticos. Preferiblemente, el porcentaje de identidad de al menos el 95% o más preferiblemente, al menos el 99% de los miembros de la población con respecto a una molécula de ácido nucleico de referencia es al menos el 98%, 99%, 99,9% o mayor. El porcentaje de identidad puede calcularse comparando dos secuencias óptimamente alineadas, determinando el número de posiciones en las que se encuentra la base de ácido nucleico idéntica (p. ej., A, T, C, G, U o I) en ambas secuencias para obtener el número de posiciones coincidentes, dividiendo el número de posiciones coincidentes por el número total de posiciones, y multiplicando el resultado por 100 para obtener el porcentaje de identidad de secuencia. Se apreciará que, en ciertos casos, una molécula de ácido nucleico, tal como una plantilla, sonda, cebador, etc., puede ser una porción de una molécula de ácido nucleico más grande que también contiene una porción que no sirve para una función de plantilla, sonda o cebador. En ese caso, los miembros individuales de una población no necesitan ser sustancialmente idénticos con respecto a esa parte.
- 20 Los nucleótidos de un polinucleótido pueden tener cualquier combinación de bases, incluidas las mencionadas en este documento, por ejemplo, uracilo, adenina, timina, citosina, guanina, inosina, xantina, hipoxantina, 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.
- 25 Los polinucleótidos incluyen ADN bicatenario y monocatenario, así como ARN de doble cadena y monocatenario, híbridos de ADN:ARN, ácidos nucleicos peptídicos (ANP) e híbridos entre ANP y ADN o ARN, y también incluyen tipos conocidos de modificaciones, por ejemplo, marcadores que son conocidos en la técnica, metilación, "tapones", sustitución de uno o más de los nucleótidos naturales con modificaciones análogas e internucleotídicas tales como, por ejemplo, aquellas con enlaces no cargados (p. ej., metilfosfonatos, fosfotriésteres, fosforamidatos, carbamatos, etc.), con enlaces cargados negativamente (por ejemplo, fosforotioatos, fosforoditioatos, etc.) y con enlaces cargados positivamente (por ejemplo, aminoalquilfosforamidatos, aminoalquilfosfotriésteres), que contienen restos colgantes, tales como, por ejemplo, proteínas (incluyendo nucleasas, toxinas, anticuerpos, péptidos señal, poli-L-lisina, etc.), aquellos con intercaladores (p. ej., acridina, psoraleno, etc.), aquellos que contienen quelantes (p. ej., metales, metales radioactivos, boro, metales oxidativos, etc.), aquellos que contienen alquilantes, aquellos con enlaces modificados (por ejemplo, ácidos nucleicos anoméricos alfa, etc.), así como formas no modificadas del polinucleótido u oligonucleótido. Los polinucleótidos se pueden unir 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
- 30
- 35
- 40
- 45
- 50
- 55
- 60

- cualquier resto que sea detectable usando un método de detección de elección, y de este modo hace que el nucleótido o polinucleótido unido sea detectable de forma similar usando 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 está unido a un marcador, y la presencia del nucleótido o polinucleótido se detecta directamente, y/o se detecta la generación de subproductos de ligamiento tales 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 compuerta se genera mediante un proceso químico. Opcionalmente, el transistor de efecto de campo químico es un transistor de efecto de campo sensible a los iones.
- 10 Cuando dos o más reactivos están marcados, los marcadores son preferiblemente distinguibles entre sí con el método de detección de elección. Por ejemplo, los marcadores pueden ser espectralmente resolubles, es decir, distinguibles sobre la base de sus características espectrales, particularmente la longitud de onda de emisión de fluorescencia, en condiciones de operación. En otros casos, el marcador puede comprender un compuesto generador de señal (SGC). Un SGC es opcionalmente una sustancia que en sí misma 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. Los ejemplos representativos de productos de reacción incluyen precipitados, señales fluorescentes, compuestos que tienen un color y similares. El SGC representativo incluye, por ejemplo, compuestos bioluminiscentes (por ejemplo, luciferasa), fluoróforos (por ejemplo, más abajo), compuestos bioluminiscentes y quimioluminiscentes, radioisótopos (por ejemplo, ¹³¹I, ¹²⁵I, ¹⁴C, ³H, ³⁵S, ³²P y similares), enzimas (por ejemplo, más abajo), proteínas de unión (por ejemplo, biotina, avidina, estreptavidina y similares), partículas magnéticas, compuestos químicamente reactivos (por ejemplo, tintes de color), oligonucleótidos marcados; sondas moleculares (por ejemplo, CY3, Research Organics, Inc.) y similares. Fluoróforos representativos incluyen isotiocianato de fluoresceína, succinil fluoresceína, rodamina B, lisamina, 9,10-difenil-antraceno, perileno, rubreno, pireno y derivados fluorescentes de los mismos tales como isocianato, isotiocianato, cloruro de ácido o cloruro de sulfonilo, umbeliferona, quelatos de tierras raras de lantánidos tales como Europio (Eu) y similares. Los compuestos generadores de señal también incluyen SGC cuyos productos son detectables por longitudes de onda fluorescentes y quimioluminiscentes, por ejemplo, tintes de secuenciación, luciferasa, metales que emiten fluorescencia tales como ¹⁵²Eu u otros de la serie de lantánidos; compuestos tales como luminol, isoluminol, sales de acridinio y similares; compuestos bioluminiscentes tales como luciferina; proteínas fluorescentes (por ejemplo, GFP o variantes de las mismas); y similares. Los SGC objetos son opcionalmente detectables usando un método visual u óptico; preferiblemente, con un método susceptible a la automatización, tal como un método espectrofotométrico, un método de fluorescencia, un método quimioluminiscente, un método nanométrico eléctrico que implique, 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 encuentran en una concentración apropiada.

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

- 40 Las bases de nucleótidos o nucleobases tienen habitualmente un anillo o anillos aromáticos primarios sustituidos o no sustituidos. En ciertas realizaciones, el anillo o anillos aromáticos contienen al menos un átomo de nitrógeno. En ciertas realizaciones, la base de nucleótidos es capaz de formar enlaces de hidrógeno Watson-Crick y/o Hoogsteen con una base de nucleótidos apropiadamente complementaria. Las bases de nucleótidos a modo de ejemplo y análogos de las mismas incluyen, pero no se limitan a, purinas tales como 2-aminopurina, 2,6-diaminopurina, adenina (A), etenoadenina, N6- Δ 2-isopenteniladenina (6iA), N6- Δ 2-isopentenil-2-metiltoadenina (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 tales como citosina (C), 5-propinilcitosina, isocitosina, timina (T), 4-tiotimina (4sT), 5,6-dihidrotimina, O4-metilamina, uracilo (U), 4-tiouracilo (4sU) y 5,6-dihidrouracilo (dihidrouracilo; D); indoles tales como nitroindol y 4-metilindol; pirroles tales como nitropirrol; nebularina; base (Y); etc. En ciertas realizaciones, las bases de nucleótidos son bases de nucleótidos universales. Se pueden encontrar bases de nucleótidos ejemplares adicionales, por ejemplo, en Fasman, 1989, Practical Handbook of Biochemistry and Molecular Biology, págs. 385-394, CRC Press, Boca Raton, Fla., y las referencias citadas en este documento. Un nucleósido es usualmente un compuesto que tiene una base de nucleótidos unida covalentemente al carbono C-1' de un azúcar pentosa. En ciertas realizaciones, el enlace es a través de un nitrógeno de anillo heteroaromático. Los azúcares pentosos típicos incluyen, pero no se limitan a, aquellas pentosas en las que uno o más de los átomos de carbono están sustituidos independientemente con uno o más de los mismos 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 pentosas y análogos de los mismos incluyen, pero no se limitan a, ribosa, 2'-desoxirribosa, 2'-alcoxirribosa (C1-C6), 2'-ariloxirribosa (C5-C14), 2',3'-didesoxirribosa, 2',3'-dideshidrorribosa, 2'-desoxi-3'-halorribosa, 2'-desoxi-3'-fluororribosa, 2'-desoxi-3'-clororribosa, 2'-desoxi-3'-aminorribosa, 2'-desoxi-3'-alquilribosa (C1-C6), 2'-desoxi-3'-alcoxirribosa (C1-C6) y 2'-desoxi-3'-ariloxirribosa (C5-C14). Uno o más de los carbonos de pentosa de un nucleósido pueden estar sustituidos con un éster de fosfato,

como se describe en la patente de EE.UU. No 7255994. En ciertas realizaciones, los nucleósidos son aquellos en los que la base de nucleótidos es una purina, una 7-desazapurina, una pirimidina, una base de nucleótidos universal, una base de nucleótidos específica, o un análogo de los mismos. Los análogos de nucleótidos incluyen derivados en los que el azúcar de 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. Ejemplos de análogos de azúcar de pentosa y análogo de bases de nucleótidos se han descrito anteriormente. Ejemplos de análogos de ésteres de fosfato incluyen, pero no se limitan a, alquilfosfonatos, metilfosfonatos, fosforamidatos, fosforiranos, fosforotioatos, fosforoditioatos, fosforoselenoatos, fosforodiselenoatos, fosforoanilotoatos, fosforoanilidatos, fosforoamidatos, boronofosfatos, etc., y pueden incluir contraiones asociados. Otros análogos de nucleótidos son monómeros análogos de nucleótidos que pueden polimerizarse en análogos de polinucleótidos en los que el éster de fosfato de ADN/ARN y/o la cadena principal del éster de fosfato de 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 la cadena principal de fosfato de azúcar del polinucleótido se reemplaza por una estructura peptídica.

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 están sustancialmente escindidos). 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 ADN glicosilasa u otro residuo o enlace que es un sustrato para la escisión por una AP endonucleasa, o un disacárido nucleósido.

El término adjetivo "hibridado" se refiere opcionalmente a dos polinucleótidos que están unidos entre sí mediante dos o más emparejamientos de bases secuencialmente adyacentes. El término "hibridación" se refiere al proceso mediante el cual los polinucleótidos se hibridan entre sí. Se pueden considerar dos polinucleótidos monocatenarios como "complementarios" si cuando se hibridan juntos, el polinucleótido más largo forma un saliente monocatenario y el polinucleótido más corto se puede ligar de manera eficaz a un tercer polinucleótido adyacente que forma un dúplex perfectamente coincidente con el saliente monocatenario. Cuando el saliente monocatenario es inferior a ocho nucleótidos, se puede alargar arbitrariamente hasta ocho nucleótidos agregando una combinación aleatoria de nucleótidos al saliente.

De forma similar, los residuos de nucleótidos pueden considerarse complementarios si cuando ambos están emparejados por bases entre sí dentro de dos polinucleótidos hibridados, cualquiera de los dos nucleótidos puede ligarse en una reacción de ligamiento dirigido por una plantilla cuando se sitúa como el nucleótido terminal en su polinucleótido. Los nucleótidos que se incorporan eficazmente mediante ADN polimerasas opuestas entre sí durante la replicación del ADN en condiciones fisiológicas también se consideran complementarios. En una realización, los nucleótidos complementarios pueden formar pares de bases entre sí, tales como los pares de bases A-T/U y G-C formados mediante enlaces de hidrógeno de tipo Watson-Crick específicos entre las nucleobases de nucleótidos y/o 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 la hidrofobicidad de bases y/o la forma de complementariedad entre bases.

En casos apropiados, los polinucleótidos se pueden considerar complementarios cuando pueden experimentar apareamiento de bases acumulativo en dos o más posiciones individuales correspondientes en orientación antiparalela, como en un dúplex hibridado. Opcionalmente puede haber complementariedad "completa" o "total" entre una primera y una segunda secuencia de polinucleótidos donde cada nucleótido en la primera secuencia de polinucleótidos puede experimentar una interacción de emparejamiento de bases estabilizadora 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. Un "desapareamiento" está presente en cualquier posición en la que los dos nucleótidos opuestos no son complementarios. En algunos ensayos de ligamiento, un polinucleótido puede someterse a una unión sustancial dependiente de plantilla incluso cuando tiene uno o más desapareamientos en su plantilla hibridada. Opcionalmente, el polinucleótido no tiene más de 4, 3 ó 2 desapareamientos, por ejemplo, 0 ó 1 desapareamientos, con su plantilla. En algunos ensayos, el polinucleótido no sufrirá ningún ligamiento sustancial dependiente de plantilla a menos que sea al menos 60% complementario, por ejemplo, al menos aproximadamente 70%, 80%, 85%, 90%, 95%, 99% o 100% complementario a su plantilla.

Como se usa en este documento, una "muestra biológica" se refiere a una muestra de tejido o fluido aislada de un individuo, que incluye, pero no se limita a, por ejemplo, plasma, suero, fluido espinal, semen, fluido linfático, las secciones externas de la piel, tractos respiratorios, intestinal y genitourinario, 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 que resulta del crecimiento de células en el medio de cultivo celular, células infectadas por virus putativamente, células recombinantes y componentes celulares).

La identidad de secuencia (también llamada 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 las mismas indica el porcentaje de todas las unidades monoméricas (por ejemplo, nucleótidos o aminoácidos) que son iguales (es decir,

aproximadamente 70% de identidad, preferiblemente 75 %, 80%, 85%, 90%, 95% o 99% de identidad). El porcentaje de identidad puede estar sobre una región especificada, cuando se compara y alinea para una correspondencia máxima sobre una ventana de comparación, o región designada medida usando un BLAST o algoritmos BLAST 2.0 de comparación de secuencia con parámetros predeterminados descritos a continuación, o mediante alineación manual e inspección visual. Se dice que las secuencias son "sustancialmente idénticas" cuando hay al menos un 90% de identidad en el nivel de aminoácido o en el nivel de nucleótido. Preferiblemente, la identidad existe en una región que tiene al menos aproximadamente 2, 3, 4, 5, 6, 7, 8, 10, 12, 15, 20, 25, 50 ó 100 residuos de longitud, o en toda la longitud de al menos una secuencia comparada. Un algoritmo preferido 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 col., Nuc. Acids 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) está destinado a indicar la acción real o de intento correspondiente, incluso si no se produce una acción real. Por ejemplo, el verbo "hibridar" y la forma de gerundio "hibridando" y similares se refieren a la hibridación real o a la hibridación intentada poniendo en contacto las secuencias de ácido nucleico en condiciones adecuadas para la hibridación, incluso si no se produce una hibridación real. De forma similar, "detectanto" y "detección", cuando se usan en las reivindicaciones, se refieren 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 involuntaria o no intencional, por ejemplo, la hibridación a una secuencia de polinucleótidos no deseada distinta de la secuencia de polinucleótidos diana deseada. La secuencia de polinucleótidos no activada puede estar en el mismo o diferente polinucleótido del objetivo pretendido. En algunos casos, la única hibridación prevista puede ser la del emparejamiento de bases Watson-Crick entre dos polinucleótidos. Otros tipos de apareamientos de bases pretendidos pueden incluir el emparejamiento de bases entre análogos correspondientes de tales nucleótidos o entre iso-citidina e iso-guanina. En algunos casos en los que la hibridación solo se pretende entre bases complementarias, se considera que cualquier unión entre bases no complementarias es una hibridación no específica.

En referencia al ligamiento de dos polinucleótidos, el extremo "proximal" de cualquier polinucleótido es el extremo que se pretende que esté ligado al otro polinucleótido. Este es generalmente el extremo que se pone en contacto con el sitio activo de la ligasa, o el extremo que eventualmente se liga al otro polinucleótido, mientras que el extremo opuesto es el término "distal". El residuo de nucleótido terminal en el extremo proximal se puede denominar nucleótido proximal, y la posición de nucleótido proximal opcionalmente designada como posición 1, o -1 dependiendo del lado del sitio de ligamiento al que nos referimos, la posición de penúltimo nucleótido como posición 2 o -2, etc. Con referencia a dos polinucleótidos hibridados adyacentes, el extremo próximo es generalmente el extremo de un polinucleótido que está más cerca del otro polinucleótido. En algunos casos no limitantes de ligamiento dependiente de plantilla, los extremos proximales de ambos polinucleótidos se hibridan adyacentes entre sí.

"Soporte", como se usa en el presente documento, se refiere a una estructura o matriz en la que los reactivos de ligamiento, p. ej., la moléculas de ácido nucleico, micropartículas y similares pueden inmovilizarse de manera que se evite de manera significativa o completa 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 covalentemente o no covalentemente o integrarse parcialmente/completamente.

Una "base universal", como se usa en este documento, es una base que es complementaria a más de una base diferente. Las bases completamente universales pueden emparejarse con cualquiera de las bases que se encuentran típicamente en los ácidos nucleicos naturales. La base no necesita ser igualmente capaz de emparejarse con cada una de las bases naturales. Alternativamente, la base universal puede emparejarse solo o selectivamente con dos o más bases pero no con todas las bases. Opcionalmente, la base universal se empareja solo o selectivamente con purinas, o alternativamente con pirimidinas. Si así se desea, se pueden incluir dos o más bases universales en una posición particular en una sonda. En la técnica se conocen varias bases universales 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 preferida completamente universal. Los nucleósidos que comprenden hipoxantina incluyen, entre otros, inosina, isoinosina, 2'- desoxi-inosina y 7-deaza-2'-desoxiinosina, 2-aza-2'-desoxinosina.

"Purificado" generalmente se refiere al aislamiento de una sustancia (compuesto, polinucleótido, proteína, polipéptido, composición polipeptídica) tal que la sustancia comprende un porcentaje significativo, tal como una proporción mayor que la que se encuentra naturalmente (por ejemplo, mayor que 2%, mayor que 5%, mayor que 10%, mayor que 20%, mayor que 50% o más, a veces mayor que 90%, 95% o 99%) de la muestra en la que reside. En ciertas realizaciones, un componente sustancialmente purificado comprende al menos 50%, 80% - 85%, o 90 - 95% de la muestra. Técnicas para purificar polinucleótidos y polipéptidos de interés son bien conocidas 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 eliminado de su entorno original (por ejemplo, el entorno natural si se produce de forma natural) y, por lo tanto, es alterado "por la mano del hombre" de su estado natural. Por ejemplo, un polinucleótido aislado podría ser parte de un vector o una composición de materia, o podría estar contenido dentro de una célula, y aún estar "aislado" porque ese vector, composición de materia o celda particular no es el entorno original o natural donde aparece del polinucleótido.

Ejemplos de ligasas comprenden un polipéptido. Los términos "polipéptido", "péptido" y "proteína" se usan indistintamente en este documento para referirse a un polímero de residuos de aminoácidos, o cualquier variante o fragmento funcional del mismo. Los términos se aplican a 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 polímeros de aminoácidos naturales y 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 una manera similar a los aminoácidos naturales. Los aminoácidos naturales son aquellos codificados por el código genético, así como aquellos aminoácidos que se modifican posteriormente, p. ej. hidroxiprolina, γ -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 (p. ej., norleucina) o esqueletos peptídicos modificados, pero conservan la misma estructura química básica que un aminoácido de origen 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 funcionan de manera similar a un aminoácido de origen natural.

Las variantes o derivados de una secuencia de nucleótidos o secuencia polipeptídica dada son opcionalmente variaciones modificadas conservativamente. Con respecto a secuencias particulares de ácidos nucleicos, las variantes modificadas conservativamente se refieren a aquellos ácidos nucleicos que codifican secuencias de aminoácidos idénticas o esencialmente idénticas, o cuando el ácido nucleico no codifica una secuencia de aminoácidos, a secuencias esencialmente idénticas.

En cuanto a las secuencias de aminoácidos, cualquier experto reconocerá que las sustituciones, deleciones o adiciones individuales a una secuencia de ácido nucleico, péptido, polipéptido o proteína que altera, agrega o elimina un único aminoácido o un pequeño porcentaje de aminoácidos en la secuencia codificada es una "variante conservativamente modificada" donde la alteración resulta en la sustitución de un aminoácido con un aminoácido químicamente similar. Las tablas de sustitución conservativa que proporcionan aminoácidos funcionalmente similares son bien conocidas en la técnica. Tales variantes conservativamente modificadas son además, y no excluyen, variantes polimórficas, homólogos interespecies y alelos de la invención (véase, por ejemplo, Creighton, Proteins (1984)). La identidad de secuencia (también llamada 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 las mismas indica el porcentaje de todas las unidades monoméricas (por ejemplo, nucleótidos o aminoácidos) que son iguales (es decir, identidad de aproximadamente 70%, preferiblemente 75%, 80%, 85%, 90%, 95% o 99% de identidad). El porcentaje de identidad puede estar sobre una región especificada, cuando se compara y alinea para la correspondencia máxima sobre una ventana de comparación, o región designada medida usando un BLAST o algoritmos BLAST 2.0 de comparación de secuencia con parámetros predeterminados descritos a continuación, o mediante alineación manual e inspección visual. Se dice que las secuencias son "sustancialmente idénticas" cuando hay al menos un 90% de identidad en el nivel de aminoácido o en el nivel de nucleótido. Esta definición también se refiere al complemento de una secuencia de prueba. Preferiblemente, la identidad existe en una región que tiene al menos aproximadamente 25, 50 ó 100 residuos de longitud, o en toda la longitud de al menos una secuencia comparada. Un algoritmo preferido 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 col., Nuc. Acids 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, tales como las descritas en este documento. Los ácidos nucleicos que no se hibridan específicamente entre sí en condiciones rigurosas son aun 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, tal como ADN celular total o de biblioteca. Para la hibridación selectiva o específica, una señal positiva es al menos dos veces el fondo, preferiblemente 10 veces la hibridación de fondo. Opcionalmente, las condiciones rigurosas se seleccionan para que sean aproximadamente 5-10°C más bajas que el punto de fusión térmico para la secuencia específica a un pH de fuerza iónica definido. Las condiciones rigurosas son opcionalmente en las que la concentración de sal es menor que aproximadamente 1,0 M de ion sodio, típicamente aproximadamente de 0,01 a 1,0 M de concentración de ión sodio (u otras sales) a pH de 7,0 a 8,3 y la temperatura es de al menos aproximadamente 30°C para sondas cortas (por ejemplo, de 10 a 50 nucleótidos) y al menos

aproximadamente 60°C para sondas largas (por ejemplo, más de 50 nucleótidos). También se pueden lograr condiciones restrictivas con la adición de agentes desestabilizadores tales como formamida. Las condiciones de hibridación rigurosas ejemplares pueden ser las siguientes: formamida al 50%, 5 x SSC, y 1% SDS, incubando a 42°C, o, 5 x SSC, 1% de SDS, incubando a 65°C, con lavado en 0.2 x veces SSC y 0,1% de SDS a 65°C.

5 Descripción detallada

Entre otra información novedosa y sorprendente presentada en este documento, se presenta en este documento la novedosa y sorprendente ligamiento enzimático de polinucleótidos cortos.

1) Ligamientos

10 El ligamiento en el presente documento se refiere a la formación enzimática de un enlace covalente entre los extremos de dos o más cadenas de polinucleótidos. El "ligamiento" implica la formación de un enlace covalente o enlace entre los extremos 5' y 3' de dos o más ácidos nucleicos, p. ej. oligonucleótidos y/o polinucleótidos, opcionalmente en una reacción impulsada por plantilla. La naturaleza del enlace o ligamiento puede variar ampliamente y el ligamiento se logra preferiblemente enzimáticamente. La naturaleza del enlace o ligamiento puede variar ampliamente. Los ligamientos ejemplares no limitativos se llevan a cabo enzimáticamente para formar un
15 enlace fosfodiéster entre un nucleótido 5' terminal de una cadena polinucleotídica con un nucleótido 3' terminal de un polinucleótido.

20 El ligamiento puede ser uno o más de los siguientes tipos de ligamiento descritos en este documento. Un primer tipo de ligamiento enzimático implica la formación de un enlace covalente entre un primer término de una primera cadena de polinucleótidos y un segundo término diferente de una segunda cadena de polinucleótidos. El primer y segundo extremos de polinucleótidos pueden estar en diferentes cadenas de polinucleótidos, o ambos pueden estar en la misma cadena de polinucleótidos (lo que resulta en circularización). Opcionalmente, la primera y la segunda cadena de polinucleótidos no están ambas hibridadas con un tercer polinucleótido. Opcionalmente, los extremos de la primera y la segunda cadena de polinucleótidos se unen independientemente de sus secuencias (por ejemplo, ligamiento de extremos romos, o unión final no homóloga). En otra variación, se pueden ligar dos polinucleótidos
25 bicatenarios con porciones sobresalientes monocatenarias que son complementarias entre sí (por ejemplo, ligamiento del extremo cohesivo). Un tercer tipo de ligamiento (ligamiento dependiente de plantilla) se describe más adelante a continuación.

30 En cualquiera de los métodos descritos en este documento, las cadenas de polinucleótidos pueden estar en formato monocatenario o hibridar con cadenas complementarias en formato bicatenario. En el ligamiento de extremos romos, ambas cadenas de polinucleótidos a ligar se hibridan con dos cadenas complementarias diferentes de forma tal que no existe el saliente.

35 Se proporcionan métodos para ligar dos polinucleótidos. Un método de ligamiento ejemplar consigue el ligamiento entre un primer término de una primera secuencia de polinucleótidos y un segundo término de una segunda secuencia de polinucleótidos. Para facilitar la referencia, la primera secuencia de polinucleótidos se denomina la "sonda de inicialización" o "cebador", el segundo polinucleótido se denomina "sonda de extensión" o "sonda". Una tercera secuencia de polinucleótidos que está presente opcionalmente (por ejemplo, en el ligamiento dependiente de plantilla) se denomina secuencia plantilla o diana. Una cualquiera o más de las secuencias del cebador (sonda de inicialización), sonda (sonda de extensión) y/o plantilla se pueden localizar en la misma cadena de polinucleótidos, o en diferentes cadenas de polinucleótidos. En un sistema de una cadena, el cebador, la sonda y la plantilla están
40 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 la plantilla (p. ej., la hibridación entre la secuencia de plantilla y las secuencias de cebador o sonda forma una estructura de bucle de vástago 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, la plantilla, el cebador y la sonda están todos en cadenas de polinucleótidos separadas. Cualquier método de
45 ligamiento descrito en este documento puede realizarse en un sistema de una cadena, dos cadenas o tres cadenas.

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

50 El ligamiento crea opcionalmente un enlace entre un nucleótido terminal de la sonda con el nucleótido terminal del cebador. Opcionalmente, el extremo proximal del cebador y/o 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' de un polinucleótido, por ejemplo, tiene el quinto carbono en el anillo de azúcar de la desoxirribosa o ribosa en su extremo, opcionalmente
55 con un grupo fosfato unido a él, donde el grupo fosfato es capaz de formar un enlace fosfodiéster con un nucleótido terminal 3'. El nucleótido terminal 3' tiene opcionalmente un grupo hidroxilo 3' que es capaz de formar un enlace fosfodiéster con un nucleótido terminal 5'. El ligamiento da como resultado opcionalmente la formación de un enlace fosfodiéster.

Cualquiera de los polinucleótidos, independientemente de cómo se designe (por ejemplo, como "sondas" o "cebadores" o "plantillas") puede ser de cualquier secuencia, cualquier longitud, en cualquier forma y de cualquier fuente. Los polinucleótidos pueden comprender una secuencia de origen natural o ser altamente homólogos a una secuencia de origen natural, y/o derivarse de una secuencia natural. La secuencia natural puede ser cualquier porción de un gen, una secuencia reguladora, ADN o fragmento genómico, ADNc, ARN que incluye ARNm y ARNr u otros. Los polinucleótidos pueden comprender opcionalmente cualquier secuencia artificial también. El polinucleótido puede derivarse u obtenerse a partir de una muestra tal como una muestra de diagnóstico. El polinucleótido puede ser un producto secundario de una reacción, por ejemplo, un producto de ligamiento de una reacción de ligamiento o ensayo tal como los descritos en este documento, una sonda extendida de una reacción de PCR o producto de amplificación por PCR ("amplicón"), el producto de una reacción de escisión invasiva, etc. El polinucleótido puede tener un fosfato 5', o puede carecer alternativamente de un fosfato 5'.

El producto de ligamiento de cualquiera de las reacciones se somete a otras reacciones de ligamiento y/o no ligamiento a su vez. Por ejemplo, el producto de ligamiento puede usarse como cebador (sonda de inicialización) o sonda de extensión y/o plantilla en un ligamiento posterior. También, por ejemplo, se puede usar como plantilla y/o cebador para una reacción de extensión de polimerasa, tal como en PCR. La sonda, el cebador, la plantilla y/o el producto de ligamiento opcionalmente pueden someterse a una o más modificaciones antes o después del ligamiento. Por ejemplo, la sonda, el cebador, la plantilla y/o el producto de ligamiento se pueden escindir enzimática o químicamente (por ejemplo en enlaces escindibles), tratarse con exo- o endonucleasas, quinasas, fosfatasa, etc. Los extremos de un producto bicatenario pueden ser romos o rellenos, cerrados o adenilados, etc.

La sonda tiene N restos de nucleótidos de longitud, donde N es de 2 a 7, por ejemplo, 7 nucleótidos consecutivos, preferiblemente no más de 6, 5, 4, 3 ó 2 nucleótidos consecutivos. Opcionalmente, la sonda tiene al menos 2, 3, 4, 5, 6 ó 7 nucleótidos de longitud. En algunos casos, la sonda tiene cualquiera de las longitudes mínimas especificadas (por ejemplo, 2, 3, 4, 5, 6 ó 7 nucleótidos de longitud) y no tiene más de 7 nucleótidos, preferiblemente no más de 6, 5, 4, 3 ó 2 nucleótidos. En algunos ejemplos, la sonda es un 2-mer, 3-mer, 4-mer, 5-mer, 6-mer, 7-mer, o cualquier combinación de tales oligonucleótidos.

Opcionalmente, la sonda no tiene más de 20 nucleótidos consecutivos de longitud, por ejemplo, no más de 15, 12, 10, 8, 7 nucleótidos consecutivos, preferiblemente no más de 6, 5, 4, 3 ó 2 nucleótidos consecutivos. En algunos ejemplos, la sonda tiene 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.

Se debe entender que la sonda o el cebador (o la plantilla, si está presente) puede "mezclarse" o "componerse", es decir, que comprende una mezcla de uno o más polinucleótidos de diferentes secuencias.

Opcionalmente, el ligamiento se realiza usando CV ligasa en combinación con una o más sondas que tienen al menos 3 nucleótidos de longitud y no más de 6 ó 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 tener 3 meros. De lo contrario, todas las sondas de una o más sondas pueden tener 4 meros.

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

Opcionalmente, el ligamiento se produce cuando no hay falta de coincidencia dentro de una o más porciones hibridadas. En otros casos, el ligamiento ocurre cuando pueden estar presentes uno, dos o tres desapareamientos dentro de una o más porciones hibridadas. En algunos casos, el ligamiento no ocurre cuando el nucleótido terminal y/o el segundo nucleótido terminal más terminal y/o el tercer nucleótido terminal más largo no coinciden. Como se mencionó, los nucleótidos terminales pueden ser los nucleótidos terminales 5' o 3' del polinucleótido.

Opcionalmente, la plantilla, si está presente, no tiene más de 11 nucleótidos de longitud, por ejemplo, no más de 10, 9, 8, 7, 6, 5 ó 4 nucleótidos. Opcionalmente, la plantilla tiene uno o más N-meros, donde N es 4, 5, 6, 7, 8, 9, 10 ó 11.

Opcionalmente, el ligamiento dependiente de plantilla de un ácido nucleico comprende: a) proporcionar un primer oligonucleótido que tiene menos de 6 nucleótidos; b) proporcionar un segundo oligonucleótido; c) poner en

proximidad los extremos 3' de uno de los oligonucleótidos primero y segundo con los extremos 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, el ligamiento se realiza usando una ligasa de huella pequeña (SFL). Opcionalmente, el segundo oligonucleótido incluye una secuencia complementaria a una porción de un ácido nucleico plantilla. Opcionalmente, el segundo oligonucleótido se hibrida con el ácido nucleico plantilla en la región de complementariedad. Opcionalmente, el primer oligonucleótido tiene una secuencia complementaria con el ácido nucleico plantilla. Opcionalmente, el primer oligonucleótido se hibrida con el ácido nucleico plantilla en la región de complementariedad, y en el que un extremo del primer oligonucleótido es adyacente a un extremo del segundo oligonucleótido.

En algunas variaciones, (por ejemplo, ligamiento de "mella" o "ligamiento dependiente de plantilla"), tanto el cebador como la sonda deben hibridarse adyacentes entre sí en la plantilla para que se produzca el ligamiento. Opcionalmente, la sonda y el cebador se hibridizan de forma adyacente y se pueden ligar solo cuando un nucleótido terminal del cebador se hibrida con un primer nucleótido de la plantilla y un nucleótido terminal de la sonda de extensión se hibrida con un segundo nucleótido de la plantilla, donde el primer y segundo nucleótidos en la plantilla no están separados por un nucleótido intermedio de la plantilla. En otras realizaciones, los nucleótidos intermedios pueden estar presentes entre el primer y el segundo nucleótidos en la plantilla (opcionalmente unos pocos nucleótidos, por ejemplo, no más de 1, 2, 3, 5, 10 ó 15 nucleótidos). En tales realizaciones, se puede realizar una etapa de "llenar huecos" para extender el extremo 3' de la sonda o sondar antes de que se pueda ligar al extremo 5' de la otra.

Opcionalmente, al menos uno de la sonda, la plantilla (si el ligamiento depende de la plantilla) y/o el cebador está inmovilizado mientras que otro de estos tres está marcado. Por ejemplo, en la secuenciación de ligamiento, la plantilla y/o el cebador pueden inmovilizarse y la sonda puede marcarse.

Una sonda puede ser, por ejemplo, restos de N nucleótidos de longitud, donde N es de 2 a 7, por ejemplo, 2, 3 ó 4. N también puede ser menor que 6, por ejemplo, si el extremo proximal de la sonda es su extremo 3'.

Opcionalmente, el ligamiento es un ligamiento "directo" (es decir, un ligamiento del extremo 3' de la sonda al extremo 5' del cebador). Alternativamente, se puede lograr el ligamiento "inverso", donde el extremo 5' de la sonda se liga al extremo 3' del cebador.

Una sonda también puede ser de longitud N, y puede comprender una parte proximal que está perfectamente hibridada con la plantilla y tiene L nucleótidos largos, donde el L + I nucleótidos de la sonda no coincide con la plantilla. L puede ser, por ejemplo, de 2 a 8, y además puede ser menor que 6 si el extremo proximal de la sonda es su extremo 3'. En otras realizaciones, L puede ser 2, 3, 4, 5, 6, 7 u 8.

El ligamiento puede repetirse al menos una vez, para cualquier número de ciclos deseado. Opcionalmente, se usa cualquier producto de ligamiento de una reacción de ligamiento previa como iniciador de un próximo ligamiento. Opcionalmente, todos los ligamientos extienden el producto de ligamiento en la misma dirección. Por ejemplo, todas las reacciones de ligamiento pueden ser ligamientos directos o todas pueden ser ligamientos inversos. En otras realizaciones, algunos ligamientos pueden ser hacia adelante y algunos hacia atrás. Opcionalmente, cualquier cebador sin ligar se vuelve ineligible antes de iniciar el siguiente ligamiento. Si así se desea (por ejemplo, en la secuenciación de ligamiento), el método comprende además detectar si la sonda se ha ligado al cebador antes de repetir la siguiente reacción de ligamiento.

Cuando así se desee, se puede usar cualquier producto de ligamiento de la reacción de ligamiento previa como plantilla de la siguiente reacción de ligamiento, por ejemplo, en una reacción en cadena de la ligasa.

Opcionalmente, el extremo 5' de la sonda de menos de 6 nucleótidos de longitud se liga al extremo 3' del cebador mediante una CV ligasa. Por ejemplo, la sonda tiene 2, 3 ó 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 ó 4 nucleótidos de longitud.

El ligamiento debe producir una cantidad significativa o detectable del producto de ligamiento. Opcionalmente, la eficacia del ligamiento es al menos del 5%, a veces del 10%, 20%, 30%, 50%, 60%, 70%, 75%, 80%, 85%, 90% o 95%. La eficacia del ligamiento (en términos porcentuales) puede considerarse en alguna realización como la porción en porcentaje de reactivo de ligamiento que se liga para formar un producto de ligamiento al final de la reacción de ligamiento. La eficacia se determina opcionalmente después de que una reacción de ligamiento haya alcanzado el equilibrio de modo que aumentar el tiempo de ligamiento no dará como resultado un aumento sustancial del producto de ligamiento formado. Generalmente, se puede decir que una reacción de ligamiento ha alcanzado el equilibrio después de 20 minutos. Por ejemplo, se puede decir que una reacción de ligamiento en la que se liga el 90% del cebador o la sonda ha avanzado al 90% de eficacia. El reactivo de ligamiento utilizado para medir la eficacia del ligamiento es opcionalmente el reactivo que se encuentre en menor concentración que los demás. Opcionalmente, los otros reactivos y condiciones son no limitantes para la eficacia de ligamiento (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 realizaciones, el SFL puede ligar una sonda corta que es más corta que N nucleótidos, al menos X% con la misma eficacia que el SFL puede ligar el N-mer correspondiente. N es opcionalmente 4, 5, 6, 7, 8, 10, 12, 15 ó 20.

5 En algunas realizaciones, N es 6 ó 7. X es opcionalmente al menos 30%, 50%, 60%, 70%, 80% o 90%. En algunas realizaciones de ligamiento dependiente de plantilla, la longitud de la sonda corta tiene Y residuos, y los residuos proximales Y del N-mer más largo correspondiente son idénticos a la sonda, y los residuos N-Y distales del N-mer más largo correspondiente son perfectamente complementarios a la plantilla. Y es, por ejemplo, 2, 3, 4, 5 ó 6.

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

15 Se entiende que la eficacia de ligamiento (ya sea expresada en términos absolutos o relativos) puede aumentar o disminuir dependiendo de las condiciones de reacción exactas utilizadas. Opcionalmente, la eficacia de ligamiento se mide cuando el cebador o la sonda está en 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, a menudo se usa una concentración de trabajo en el rango 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
20 pueden ligarse en esa reacción de ligamiento particular (por ejemplo, sonda(s) complementaria a la plantilla). Por lo tanto, otras sondas que no son capaces de participar en la reacción de ligamiento opcionalmente no se consideran cuando se calcula la concentración de la sonda de interés. Opcionalmente, la concentración de la sonda está entre 1 picomolar y 1 milimolar, por ejemplo aproximadamente de 0,01-100 μ M, por ejemplo, 1-10 μ M, por ejemplo, 1-5 μ M. En el caso de 2-meros, la concentración se incrementa opcionalmente a 10-1000 μ M, por ejemplo aproximadamente
25 100 μ M. Opcionalmente, la concentración de ligasa está entre 1 picomolar y 1 milimolar, por ejemplo, 1 - 2 micromolar. Opcionalmente, el ensayo de ligamiento se realiza a 15-35°C, por ejemplo 15°C, 20°C, 25°C o 30°C. Cuando están implicados 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 sea limitante para el ligamiento.

30 El ligamiento puede realizarse en condiciones in vitro que se han determinado experimentalmente que son adecuadas u óptimas para la actividad de la ligasa. Preferiblemente, las condiciones de reacción de elección son (i) sustancialmente similares a las condiciones in vivo o fisiológicas en las que está activa una forma natural de la ligasa que se usa, o (ii) se ha determinado experimentalmente que da como resultado una eficacia de ligamiento que es comparable o mejor que la eficiencia obtenida usando condiciones del tipo (i). Si se especifican en este documento ejemplos de condiciones de ligamiento in vitro para un SFL particular, entonces las condiciones de reacción
35 sustancialmente similares son generalmente apropiadas para esa ligasa particular. En otras realizaciones, las condiciones son tales que el ensayo de ligamiento de referencia produce un ligamiento significativo o detectable en 30 minutos, en 10 minutos, en 1 minuto o en diez segundos. Otro ejemplo no limitante de una eficacia de ligamiento significativa o detectable se genera en el intervalo de 100 pM del producto de ligamiento, opcionalmente aproximadamente 1000 pM o 10.000 pM.

40 Opcionalmente, la eficiencia relativa puede expresarse como porcentaje relativo de eficiencia, que puede calcularse como $A/B \times 100$, 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 eficacia relativa del ligamiento se especifica en términos comparativos o relativos en comparación con un ensayo de ligamiento de referencia, está implícito que todos los demás reactivos y condiciones
45 (p. ej., la temperatura, concentración de todos los reactivos, pH, concentración de iones necesarios, tales como Mg^{++} y Mn^{++} , concentración de cofactores requeridos tales como NAD y/o ATP, sales, tampones, concentraciones molares de todos los reactivos, incluyendo enzimas, plantillas, sondas, cebadores, oligonucleótidos, etc.) se mantienen idénticos de otra forma. Por ejemplo, una condición de que un SFL puede ligar una sonda corta (por ejemplo, menos de 6 nucleótidos) al menos un X% con la misma eficacia que el SFL puede ligar un octanucleótido
50 correspondiente, puede significar que los dos ensayos de ligamiento diferentes son todos reactivos, excepto para las sondas (p. ej., cebador, plantilla, enzimas y cualquier otro reactivo) y todas las condiciones de reacción (p. ej., temperatura, concentraciones de reactivo, concentraciones de cualquier otro reactivo, etc.) se mantienen iguales para propósitos prácticos.

55 Opcionalmente, la ligasa tiene una eficacia de ligamiento mejor que la ADN ligasa de T4, por ejemplo en cualquier método descrito en este documento. En una realización, la eficacia de ligamiento es más alta que la de la T4 ADN ligasa para la misma mezcla de sonda(s), cebador(es) y plantilla(s), en cualquier método de ligamiento y condiciones elegidas, incluyendo cualquiera descrita en este documento. La eficacia de ligamiento es, por ejemplo, al menos 5% mayor que la T4 ligasa, opcionalmente al menos 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%,
60 99,99% o 99,99999%. En un ejemplo, el SFL muestra una eficacia mayor que la T4 ADN ligasa cuando se liga una

sonda de 8 nucleótidos o menos a un cebador en la dirección directa o inversa. Las condiciones del ensayo de la ligasa, plantilla, sonda, cebador y/o ligamiento son, por ejemplo, cualquiera de las descritas en este documento.

La ligasa tiene opcionalmente una buena fidelidad de ligamiento. La fidelidad de ligamiento puede evaluarse como el porcentaje de eventos de ligamiento incorrectos para una combinación dada de ligasa, plantilla, cebador y sonda. Un evento de ligamiento incorrecto es aquel en el que la sonda o el cebador no se complementan perfectamente con la plantilla, o el producto de ligamiento no se complementa perfectamente con la plantilla. En una realización, la fidelidad de ligamiento es más alta que la de la T4 ADN ligasa para la misma mezcla de sonda(s), cebador(es) y plantilla(s), en cualquier método de ligamiento y condiciones elegidas, incluyendo cualquiera de las descritas en este documento. La fidelidad de ligamiento es, por ejemplo, al menos 5% más alta que la T4 ligasa, 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 99,99999%. En un ejemplo, el SFL muestra una fidelidad mayor que la T4 ADN ligasa cuando se liga una sonda de 8 nucleótidos o menos a un cebador en dirección directa o inversa. Las condiciones del ensayo de ligasa, plantilla, sonda, cebador y/o ligamiento son, por ejemplo, cualquiera de las descritas en este documento.

2) Ligasas para huella pequeña

El ligamiento enzimático de polinucleótidos se logra mediante una ligasa de huella pequeña (SFL), ADN ligasa del virus Chlorella (CV ligasa) o un fragmento funcional o variante del mismo. Otras ligasas se muestran en la Tabla 1A, 1B o 1C. Una SFL es una ligasa que puede ligar polinucleótidos cortos. Como se usa en el presente documento, 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. Un SFL comprende opcionalmente un polipéptido que tiene una cualquiera o más de las siguientes actividades: (1) ataque nucleofílico sobre ATP o NAD⁺ que da como resultado la liberación de PPi o NMN y la formación de un intermedio de adenilato de ligasa covalente; (2) transferir el adenilato al extremo 5' de la cadena de ADN terminada en 5'-fosfato 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 une los extremos del polinucleótido y la liberación de AMP (p. ej., mediante el ataque del 3'-OH sobre la ADN-adenilato). Opcionalmente, el SFL puede mediar una o más de las siguientes transformaciones de enlace: desde fosfoanhídrido (ATP) a fosforamidato (ligasa-adenilato); desde fosforamidato (ligasa-adenilato) a fosfoanhídrido (ADN-adenilato); o de fosfoanhídrido (ADN-adenilato) a fosfodiéster (ADN sellado). Por lo tanto, los SFL ejemplares pueden comprender una secuencia polipeptídica que es homóloga a una variante de una secuencia SFL conocida o cualquier porción de la misma. Los 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 CV ligasa, DLX, DLXd, DLXd2 y MnM ligasa. El SFL usado en los métodos y kits de esta invención es la ligasa del virus Chlorella. Se identifican algunas ligasas ejemplares y sus IG o números de acceso se proporcionan en la Tabla 1A a continuación:

Tabla 1A

PRK08224		
Organismo	Nombre de la proteína	Entrada
B. Acidobacteria		
Candidatus Koribacter versatilis Ellin345Candidatus Koribacter (1 proteínas)	ADN Ligasa dependiente de ATP	YP_592504
Candidatus Solibacter usitatus		
Ellin6076Candidatus Solibacter (1 proteína)	ADN Ligasa dependiente de ATP	YP_826317
C. Actinobacteria		
Acidothermus cellulolyticus 11BAcidothermus (1 proteína)	ADN Ligasa dependiente de ATP	YP_873134
Actinosynnema mirum DSM 43827Actinosynnema (1	ADN Ligasa dependiente de	YP_003099374

ES 2 645 953 T3

proteína)	ATP	
Arthrobacter aurescens TCIArthrobacter (3 proteínas)	ADN Ligasa dependiente de ATP	YP_949544
Arthrobacter chlorophenolicus A6Arthrobacter (3 proteínas)	ADN Ligasa dependiente de ATP	YP_002489901
Arthrobacter sp. FB24Arthrobacter (3 proteínas)	ADN Ligasa dependiente de ATP	YP_833558
Beutenbergia cavernae DSM 12333 Beutenbergia (1 proteína)	ADN Ligasa dependiente de ATP	YP_002880505
Catenulispora acidiphila DSM 44928Catenulispora (2 proteínas)	ADN Ligasa dependiente de ATP	YP_003116519
Catenulispora acidiphila DSM 44928Catenulispora (2 proteínas)	ADN dependiente de ATP	YP_003116565
Frankia alni ACN14aFrankia (2 proteínas)	ADN Ligasa dependiente de ATP	YP_712338
Frankia sp. EAN1pecFrankia (2 proteínas)	ADN Ligasa dependiente de ATP	YP_001509433
Kineococcus radiotolerans SRS30216Kineococcus (1 proteína)	ADN Ligasa dependiente de ATP	YP_001360406
Kytococcus sedentarius DSM 20547Kytococcus (1 proteína)	ADN Ligasa dependiente de ATP	YP_003149340
Mycobacterium abscessus ATCC 19977Mycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_001701033
Mycobacterium avium 104Mycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_879648
Mycobacterium avium subsp. paratuberculosis K-10Mycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	NP_959275
Mycobacterium gilvum PYR-GCKMycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_001132524

ES 2 645 953 T3

Mycobacterium gilvum PYR-GCKMycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_001132543
Mycobacterium marinum MMycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_001853525
Mycobacterium smegmatis str. MC2 155Mycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_890520
Mycobacterium smegmatis str. MC2 155Mycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_890521
Mycobacterium sp. JLSMycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_001073538
Mycobacterium sp. JLSMycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_001073546
Mycobacterium sp. JLSMycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_001073574
Mycobacterium ulcerans Agy99Mycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_907815
Mycobacterium vanbaalenii PYR-1 Mycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_956315
Mycobacterium vanbaalenii PYR-1 Mycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_956321
Mycobacterium tuberculosis H37RaMycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_001285120
Mycobacterium tuberculosis H37RvMycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	NP_218248
Mycobacterium bovis AF2122/97Mycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	NP_857396
Mycobacterium bovis BCG str. Pasteur 1173P2Mycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_979871
Mycobacterium bovis BCG str. Tokio 172Micobacteria (26 proteínas)	ADN Ligasa dependiente de ATP	YP_002646832

ES 2 645 953 T3

proteínas)	ATP	
Mycobacterium tuberculosis CDC1551Mycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	NP_338389
Mycobacterium tuberculosis F11Mycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_001289690
Mycobacterium tuberculosis KZN 1435Mycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_003033779
Mycobacterium sp. KMSMycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_940984
Mycobacterium sp. MCSMycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_642076
Mycobacterium sp. KMSMycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_941006
Mycobacterium sp. MCSMycobacterium (26 proteínas)	ADN Ligasa dependiente de ATP	YP_642099
Nakamurella multipartita DSM 44233Nakamurella (1 proteína)	ADN Ligasa dependiente de ATP	YP_003200226
Nocardia farcinica IFM 10152Nocardia (1 proteína)	ADN Ligasa dependiente de ATP	YP_118771
Nocardioides sp. JS614Nocardioides (1 proteína)	ADN Ligasa dependiente de ATP	YP_922117
Rhodococcus erythropolis PR4Rhodococcus (3 proteínas)	ADN Ligasa dependiente de ATP	YP_002768423
Rhodococcus jostii RHA1 Rhodococcus (3 proteínas)	ADN Ligasa dependiente de ATP	YP_705046
Rhodococcus opacus B4Rhodococcus (3 proteínas)	ADN Ligasa dependiente de ATP	YP_002782360
Saccharopolyspora erythraea NRRL 2338Saccharopolyspora (1 proteína)	ADN Ligasa dependiente de ATP	YP_001104098

ES 2 645 953 T3

Salinispora arenicola CNS-205Salinispora (2 proteínas)	ADN Ligasa dependiente de ATP	YP_001536378
Salinispora tropica CNB-440Salinispora (2 proteínas)	ADN Ligasa dependiente de ATP	YP_001158390
Streptomyces avermitilis MA-4680Streptomyces (3 proteínas)	ADN Ligasa dependiente de ATP	NP_822873
Streptomyces coelicolor A3(2)Streptomyces (3 proteínas)	ADN Ligasa dependiente de ATP	NP_630780
Streptomyces griseus subsp. griseus NBRC 13350Streptomyces (3 proteínas)	ADN Ligasa dependiente de ATP	YP_001822536
F. Chlamydiae/Verrucomicrobia		
Opitutus terrae PB90-1Opitutus (1 proteína)	ADN Ligasa dependiente de ATP	YP_001821013
N.Alphaproteobacteria		
Bradyrhizobium japonicum USDA 110Bradyrhizobium (3 proteínas)	ADN Ligasa dependiente de ATP	NP_771853
Bradyrhizobium sp. BTAiIiBradyrhizobium (3 proteínas)	ADN Ligasa dependiente de ATP	YP_001236299
Bradyrhizobium sp. ORS278Bradyrhizobium (3 proteínas)	ADN Ligasa dependiente de ATP	YP_001202307
Chelativorans sp. BNCIChelativorans (1 proteína)	ADN Ligasa dependiente de ATP	YP_675242
Mesorhizobium loti MAFF303099Mesorhizobium (2 proteínas)	ADN Ligasa dependiente de ATP	NP_108245
Mesorhizobium loti MAFF303099 (plásmido) Mesorhizobium (2 proteínas)	ADN Ligasa dependiente de ATP	NP_109531
Methylocella silvestris BL2Methylocella (1 proteína)	ADN Ligasa dependiente de	YP_002363964

ES 2 645 953 T3

	ATP	
Nitrobacter hamburgensis X14Nitrobacter (1 proteína)	ADN Ligasa dependiente de ATP	YP_579055
Phenylobacterium zucineum HLKIPhenylobacterium (1 proteína)	ADN Ligasa dependiente de ATP	YP_002131547
Rhizobium leguminosarum bv. trifolii WSM2304 (plásmido)Rhizobium (2 proteínas)	ADN Ligasa dependiente de ATP	YP_002279148
Rhizobium sp. NGR234 (plásmido) Rhizobium (2 proteínas)	ADN Ligasa dependiente de ATP	YP_002823307
ADN Ligasa dependiente de ATP		
Sinorhizobium medicae WSM419 (plásmido)Sinorhizobium (2 proteínas)	ADN Ligasa dependiente de ATP	YP_001312861
Sinorhizobium meliloti 1021 (plásmido) Sinorhizobium (2 proteínas)	ADN Ligasa dependiente de ATP	NP_436551
P.Deltaproteobacteria		
Anaeromyxobacter dehalogenans 2CP-1 Anaeromyxobacter (5 proteínas)	ADN Ligasa dependiente de ATP	YP_002491286
Anaeromyxobacter dehalogenans 2CP-CAnaeromyxobacter (5 proteínas)	ADN Ligasa dependiente de ATP	YP_464028
Anaeromyxobacter sp. Fw109-5Anaeromyxobacter (5 proteínas)	ADN Ligasa dependiente de ATP	YP_001378773
Anaeromyxobacter sp. Fw109-5Anaeromyxobacter (5 proteínas)	ADN Ligasa dependiente de ATP	YP_001381200
Anaeromyxobacter sp. KAnaeromyxobacter (5 proteínas)	ADN Ligasa dependiente de ATP	YP_002133229
PRK09125		
Organismo	Nombre de la proteína	Entrada

ES 2 645 953 T3

O.Betaproteobacteria		
Acidovorax sp. JS42Acidovorax (1 proteína)	ADN ligasa	YP_986978
Aromatoleum aromaticum EbNIAromatoleum (1 proteína)	ADN ligasa	YP_161050
Azoarcus sp. BH72Azoarcus (1 proteína)	ADN ligasa	YP_934633
Candidatus Accumulibacter phosphatis clade IIA str. UW-1 Candidatus Accumulibacter (1 proteína)	ADN ligasa	YP_003169249
Dechloromonas aromatica RCBDechloromonas (1 proteína)	ADN ligasa	YP_284461
Diaphorobacter sp. TPSYDiaphorobacter (1 proteína)	ADN ligasa	YP_002553689
Herminiimonas arsenicoxydansHerminiimonas (1 proteína)	ADN ligasa	YP_001100009
Leptothrix cholodnii SP-6Leptothrix (1 proteína)	ADN ligasa	YP_001791742
Methylibium petroleiphilum PMIMethylibium (1 proteína)	ADN ligasa	YP_001020556
Neisseria gonorrhoeae FA 1090 Neisseria (7 proteínas)	ADN ligasa	YP_209054
Neisseria gonorrhoeae NCCPI 1945 Neisseria (7 proteínas)	ADN ligasa	YP_002002827
Neisseria meningitidis 053442 Neisseria (7 proteínas)	ADN ligasa	YP_001598310
Neisseria meningitidis FAM18 Neisseria (7 proteínas)	ADN ligasa	YP_975951
Neisseria meningitidis MC58 Neisseria (7 proteínas)	ADN ligasa	NP_275038
Neisseria meningitidis Z2491 Neisseria (7 proteínas)	ADN ligasa	YP_002341892
Neisseria meningitidis alpha14Neisseria (7 proteínas)	ADN ligasa	YP_003082363
Polaromonas naphthalenivorans CJ2Polaromonas (2 proteínas)	ADN ligasa	YP_982249
Polaromonas sp. JS666Polaromonas (2 proteínas)	ADN ligasa	YP_549233

ES 2 645 953 T3

Rhodospirillum rubrum T118Rhodospirillum (1 proteína)	ADN ligasa	YP_522700
Thauera sp. MZTThauera (1 proteína)	ADN ligasa	YP_002353773
Thiobacillus denitrificans ATCC 25259Thiobacillus (1 proteína)	ADN ligasa	YP_314570
Variovorax paradoxus S110Variovorax (1 proteína)	ADN ligasa	YP_002944627
Verminephrobacter eiseniae EF01-2Verminephrobacter (1 proteína)	ADN ligasa	YP_998235
P.Deltaproteobacteria		
Desulfobacterium autotrophicum HRM2Desulfobacterium (1 proteína)	LigA2	YP_002604477
Myxococcus xanthus DK 1622Myxococcus (1 proteína)	ADN ligasa	YP_628883
Q.Epsilonproteobacteria		
Arcobacter butzleri RM4018Arcobacter (1 proteína)	ATP dependiente de la ADN ligasa	YP_001489632
Campylobacter concisus 13826Campylobacter (10 proteínas)	ADN-ligasa dependiente de ATP	YP_001467307
Campylobacter curvus 525.92 (10 proteínas)	ATP dependiente de la ADN ligasa	YP_001407695
Campylobacter fetus subsp. fetus 82-40Campylobacter (10 proteínas)	ADN ligasa dependiente de ATP	YP_892536
Campylobacter hominis ATCC BAA-Campylobacter (10 proteínas)	ADN ligasa dependiente de ATP	YP_001406356381
Campylobacter jejuni RM 1221 Campylobacter (10 proteínas)	ADN-ligasa dependiente de ATP	YP_179811
Campylobacter jejuni subsp. doylei 269.97Campylobacter (10 proteínas)	ADN ligasa dependiente de ATP	YP_001398949

ES 2 645 953 T3

Campylobacter jejuni subsp. jejuni 81-176Campylobacter (10 proteínas)	ADN ligasa dependiente de ATP	YP_001001312
Campylobacter jejuni subsp. Jejuni 81116Campylobacter (10 proteínas)	ADN ligasa dependiente de ATP	YP_001483144
Campylobacter jejuni subsp. jejuni NCTC 11168Campylobacter (10 proteínas)	ADN-ligasa dependiente de ATP	YP_002345037
Campylobacter lari RM2100Campylobacter (10 proteínas)	ADN-ligasa dependiente de ATP	YP_002574655
Sulfurimonas denitrificans DSM 1251Sulfurimonas (1 proteína)	ADN ligasa	YP_393098
R. Gammaproteobacteria		
Actinobacillus succinogenes 130ZActinobacillus (1 proteína)	ADN ligasa dependiente de ATP	YP_001344487
Aggregatibacter actinomycetemcomitans D11S-1Aggregatibacter (2 proteínas)	ADN-ligasa dependiente de ATP	YP_003256304
Aggregatibacter aphrophilus NJ8700Aggregatibacter (2 proteínas)	ADN ligasa dependiente de ATP	YP_003007537
Alcanivorax borkumensis SK2 Alcanivorax (1 proteína)	ADN ligasa dependiente de ATP	YP_694422
Aliivibrio salmonicida LF11238Aliivibrio (3 proteínas)	ADN ligasa dependiente de ATP	YP_002262821
Vibrio fischeri ES114Aliivibrio (3 proteínas)	ADN ligasa dependiente de ATP	YP_204833
Vibrio fischeri MJ11 Aliivibrio (3 proteínas)	ADN ligasa dependiente de ATP	YP_002156265
Alteromonas macleodii 'Deep ecotype'Alteromonas (1 proteína)	ADN ligasa dependiente de ATP	YP_002127707
Mannheimia succiniciproducens MBEL55EBasfia (1 proteína)	ADN ligasa dependiente de ATP	YP_088131

ES 2 645 953 T3

Colwellia psychrerythraea 34HColwellia (1 proteína)	ADN ligasa dependiente de ATP	YP_271053
Haemophilus influenzae 86-028NPHaemophilus (3 proteínas)	ADN ligasa dependiente de ATP	YP_248841
Haemophilus influenzae PittEEHaemophilus (3 proteínas)	ADN ligasa dependiente de ATP	YP_001290961
Haemophilus influenzae PittGGHaemophilus (3 proteínas)	ADN-ligasa dependiente de ATP	YP_001293088
Haemophilus somnus 129PTHistophilus (2 proteínas)	ADN ligasa dependiente de ATP	YP_719536
Haemophilus somnus 2336Histophilus (2 proteínas)	ADN ligasa dependiente de ATP	YP_001783642
Idiomarina loihiensis L2TRIdiomarina (1 proteína)	ADN ligasa dependiente de ATP	YP_156435
Marinobacter aquaeolei VT8Marinobacter (1 proteína)	ADN ligasa dependiente de ATP	YP_960951
Marinomonas sp. MWYLIIMarinomonas (1 proteína)	ADN ligasa dependiente de ATP	YP_001341226
Photobacterium profundum SS9Photobacterium (1 proteína)	ADN ligasa dependiente de ATP	YP_132765
Pseudoalteromonas atlantica T6cPseudoalteromonas (2 proteínas)	ADN-ligasa dependiente de ATP	YP_659659
Pseudoalteromonas haloplanktis TAC125Pseudoalteromonas (2 proteínas)	ADN-ligasa dependiente de ATP	YP_340675
Psychromonas ingrahamii 37Psychromonas (1 proteína)	ADN ligasa dependiente de ATP	YP_942593
Shewanella amazonensis SB2B Shewanella (18 proteínas)	ADN ligasa dependiente de ATP	YP_927870

ES 2 645 953 T3

Shewanella baltica OS155Shewanella (18 proteínas)	ADN ligasa dependiente de ATP	YP_001050227
Shewanella baltica OS185Shewanella (18 proteínas)	ADN ligasa dependiente de ATP	YP_001366044
Shewanella baltica OS195Shewanella (18 proteínas)	ADN ligasa dependiente de ATP	YP_001554317
Shewanella baltica OS223Shewanella (18 proteínas)	ADN ligasa dependiente de ATP	YP_002358357
Shewanella frigidimarina NCIMB 400Shewanella (18 proteínas)	ADN ligasa dependiente de ATP	YP_750174
Shewanella halifaxensis HAW-EB4Shewanella (18 proteínas)	ADN-ligasa dependiente de ATP	YP_001673966
Shewanella loihica PV-4Shewanella (18 proteínas)	ADN ligasa dependiente de ATP	YP_001093713
Shewanella oneidensis MR-1 Shewanella (18 proteínas)	dependientes de ADN ATP	NP_717802
Shewanella pealeana ATCC 700345 Shewanella (18 proteínas)	ADN-ligasa dependiente de ATP	YP_001502366
Shewanella piezotolerans WP3 Shewanella (18 proteínas)	ADN ligasa dependiente de ATP	YP_002312387
Shewanella sediminis HAW-EB3 Shewanella (18 proteínas)	ADN-ligasa dependiente de ATP	YP_001474374
Shewanella sp. ANA-3 Shewanella (18 proteínas)	ADN ligasa dependiente de ATP	YP_870036
Shewanella sp. MR-4 Shewanella (18 proteínas)	ADN ligasa dependiente de ATP	YP_734321
Shewanella sp. MR-7 Shewanella (18 proteínas)	ADN ligasa dependiente de ATP	YP_738313

ES 2 645 953 T3

Shewanella woodyi ATCC 51908Shewanella (18 proteínas)	ADN-ligasa dependiente de ATP	YP_001760369
Shewanella putrefaciens CN-32Shewanella (18 proteínas)	ADN ligasa dependiente de ATP	YP_001183272
Shewanella sp. W3-18-1 Shewanella (18 proteínas)	ADN ligasa dependiente de ATP	YP_963655
Thiomicrospira crunogena XCL-2Thiomicrospira (1 proteína)	ADN ligasa dependiente de ATP	YP_391938
Vibrio cholerae MJ-1236Vibrio (9 proteínas)	ADN ligasa dependiente de ATP	YP_002878565
Vibrio parahaemolyticus RIMD 2210633Vibrio (9 proteínas)	ADN ligasa	NP_797856
Vibrio splendidus LGP32Vibrio (9 proteínas)	ADN ligasa	YP_002417130
Vibrio vulnificus CMCP6Vibrio (9 proteínas)	ADN ligasa	NP_761477
Vibrio vulnificus YJ016Vibrio (9 proteínas)	ADN ligasa	NP_934427
Vibrio cholerae M66-2Vibrio (9 proteínas)	ADN ligasa	YP_002810248
Vibrio cholerae O1 biovar El Tor str. N16961 Vibrio (9 proteínas)	ADN ligasa	NP_231182
Vibrio cholerae O395 Vibrio (9 proteínas)	ADN ligasa	YP_001217094
Vibrio cholerae O395 Vibrio (9 proteínas)	ADN ligasa	YP_002819900
PHA0454		
Organismo	Nombre de la proteína	Entrada
b. Virus		
Enterobacteria fago 13a AT7-tipo virus (16 proteínas)	ADN ligasa dependiente de	YP_002003941

ES 2 645 953 T3

	ATP	
Enterobacteria fago BA14T7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	YP_002003458
Enterobacteria fago EcoDSIT7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	YP_002003747
Enterobacteria fago KIFT7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	YP_338096
Enterobacteria fago T3T7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	NP_523305
Enterobacteria fago T7T7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	NP_041963
Klebsiella fago K11T7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	YP_002003797
Fago de Kluyvera KvpIT7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	YP_002308390
Fago de Morganella MmPIT7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	YP_002048633
Fago de Pseudomonas gh-1T7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	NP_813751
Fago de Salmonella phiSG-JL2T7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	YP_001949754
Vibriophage VP4T7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	YP_249578
Fago de Yersinia pestis phiA1122T7-tipo virus (16 proteínas)	ADN-ligasa dependiente de ATP	NP_848267
Fago de Yersinia BerlinT7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	YP_918989
Fago de Yersinia Yepe2T7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	YP_002003318

ES 2 645 953 T3

Fago de Yersinia phiYe03-12T7-tipo virus (16 proteínas)	ADN ligasa dependiente de ATP	NP_052075
Fago de Enterobacteria LKA1phiKMV-tipo virus (7 proteínas)	ADN-ligasa dependiente de ATP	YP_001522868
Fago de Enterobacteria phiKMVphiKMV-tipo virus (7 proteínas)	ADN ligasa dependiente de ATP	NP_877456
Fago de Pseudomonas LKD16phiKMV-tipo virus (7 proteínas)	ADN-ligasa dependiente de ATP	YP_001522807
Fago de Pseudomonas LUZ19phiKMV-tipo virus (7 proteínas)	ADN-ligasa dependiente de ATP	YP_001671961
Fago de Pseudomonas PT2phiKMV-tipo virus (7 proteínas)	ADN ligasa dependiente de ATP	YP_002117800
Fago de Pseudomonas PT5phiKMV-tipo virus (7 proteínas)	ADN ligasa dependiente de ATP	YP_002117741
Fago de Pseudomonas phikF77phiKMV-tipo virus (7 proteínas)	ADN-ligasa dependiente de ATP	YP_002727838
CLSZ2445448		
Organismo	Nombre de la proteína	Entrada
a. Eukaryota		
Cepa de Paramecium tetraurelia d4-2Paramecium (5 proteínas)	ADN ligasa	XP_001347270
Cepa de Paramecium tetraurelia d4-2Paramecium (5 proteínas)	proteína hipotética	XP_001422985
Cepa de Paramecium tetraurelia d4-2Paramecium (5 proteínas)	proteína hipotética	XP_001431968
Cepa de Paramecium tetraurelia d4-2Paramecium (5 proteínas)	proteína hipotética	XP_001435874

ES 2 645 953 T3

Cepa de Paramecium tetraurelia d4-2Paramecium (5 proteínas)	proteína hipotética	XP_001460273
Tetrahymena thermophilaTetrahymena (1 proteína)	ADN ligasa dependiente de ATP	XP_001011861
CLSP2344013		
Organismo	Nombre de la proteína	Entrada
b. Virus		
Fago de Enterobacteria Felix 01 no clasificada Myoviridae (4 proteínas)	supuesta ADN ligasa	NP_944942
Fago de Enterobacteria WV8 no clasificada Myoviridae (4 proteínas)	proteína hipotética	YP_002922879
Fago de Erwinia phiEa21-4 no clasificado Myoviridae (4 proteínas)	supuesta ADN ligasa	YP_002456101
Fago de Escherichia rv5 no clasificada Myoviridae (4 proteínas)	ADN-ligasa dependiente de ATP	YP_002003586
PRK07636		
Organismo	Nombre de la proteína	Entrada
J.Firmicutes		
Bacillus clausii KSM-K16Bacillus	ADN-ligasa dependiente de ATP	YP_176304
Cepa de Bacillus subtilis subsp. subtilis 168Bacillus	ADN-ligasa dependiente de ATP	NP_389932
Bacillus licheniformis ATCC 14580Bacillus	ADN-ligasa dependiente de ATP	YP_078721
Bacillus licheniformis ATCC 14580Bacillus	ADN-ligasa dependiente de ATP	YP_091132

ES 2 645 953 T3

Geobacillus sp. Y412MC10 Geobacillus	ADN ligasa dependiente de ATP	YP_003240778
Paenibacillus sp. JDR-2Paenibacillus	ADN-ligasa dependiente de ATP	YP_003009892
CLSK2551528		
Organismo	Nombre de la proteína	Entrada
J.Firmicutes		
Geobacillus sp. Y412MC10Geobacillus (1 proteína)	ADN ligasa dependiente de ATP	YP_003245332
Paenibacillus sp. JDR-2Paenibacillus (1 proteína)	ADN ligasa dependiente de ATP	YP_003013681
CLSK2532515		
Organismo	Nombre de la proteína	Entrada
B. Acidobacteria		
Candidatus Solibacter usitatus Ellin6076Candidatus Solibacter (1 proteína)	ADN ligasa dependiente de ATP	YP_829024
E.Bacteroides/Chlorobi		
Flavobacteriaceae bacterium 3519-10 no clasificado Flavobacteriaceae (1 proteína)	ADN ligasa dependiente de ATP	YP_003095681
CLSK2470953		
Organismo	Nombre de la proteína	Entrada
C. Actinobacteria		
Arthrobacter chlorophenolicus A6 (plásmido) Arthrobacter (2 proteínas)	ADN ligasa dependiente de ATP	YP_002478051
Arthrobacter chlorophenolicus A6 (plásmido) Arthrobacter (2 proteínas)	ADN ligasa dependiente de ATP	YP_002478427

ES 2 645 953 T3

Nocardioides sp. JS614 Nocardioides (1 proteína)	ADN ligasa dependiente de ATP	YP_923463
CLSK2469924		
Organismo	Nombre de la proteína	Entrada
J.Firmicutes		
Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446Alicyclobacillus (1 proteína)	ADN ligasa dependiente de ATP	YP_003185050
Brevibacillus brevis NBRC 100599Brevibacillus (1 proteína)	supuesta ADN ligasa dependiente de ATP	YP_002773127
CLSK2340991		
Organismo	Nombre de la proteína	Entrada
N. Alphaproteobacteria		
Phenylobacterium zucineum HLK1 (plásmido) Phenylobacterium (2 proteínas)	ADN ligasa dependiente de ATP	YP_002128561
Phenylobacterium zucineum HLK1 (plásmido) Phenylobacterium (2 proteínas)	ADN ligasa dependiente de ATP	YP_002128631
CLSK2333706		
Organismo	Nombre de la proteína	Entrada
J.Firmicutes		
Candidatus Desulforudis audaxviator MP104CCandidatus Desulforudis (1 proteína)	ADN ligasa dependiente de ATP	YP_001716762
Natranaerobius thermophilus JW/NM-WN-LFNatranaerobius (1 proteína)	ADN-ligasa dependiente de ATP	YP_001916325
CLSK2303611		
Organismo	Nombre de la proteína	Entrada
C.Actinobacteria		
Streptomyces coelicolor A3(2) Streptomyces (2 proteínas)	ADN ligasa dependiente de ATP	NP_631399

ES 2 645 953 T3

Streptomyces griseus subsp. griseus NBRC 13350 Streptomyces (2 proteínas)	supuesta ADN ligasa dependiente de ATP	YP_001828202
CLSK962101		
Organismo	Nombre de la proteína	Entrada
C.Actinobacteria		
Nocardioides sp. JS614 Nocardioides (1 proteína)	ADN ligasa dependiente de ATP	YP_922436
Salinispora arenicola CNS-205 Salinispora (2 proteínas)	región de ADN polimerasa LigD ligasa	YP_001539124
Salinispora tropica CNB-440 Salinispora (2 proteínas)	ADN ligasa dependiente de ATP	YP_001160776
CLSK915249		
Organismo	Nombre de la proteína	Entrada
C.Actinobacteria Véase CLSK2303611 más arriba		
Streptomyces avermitilis MA -4680 (plásmido) Streptomyces (2 proteínas)	supuesta ADN ligasa dependiente de ATP	NP_828839
Streptomyces sp. HK1 (plásmido) Streptomyces (2 proteínas)	supuesta ADN ligasa dependiente de ATP	YP_001661618
CLSK899085		
Organismo	Nombre de la proteína	Entrada
O.Betaproteobacteria		
Burkholderia cenocepacia HI2424 (plásmido) Burkholderia (3 proteínas)	ADN ligasa dependiente de ATP	YP_840498
Burkholderia cenocepacia J2315 (plásmido) Burkholderia (3 proteínas)	supuesta ligasa	YP_002235530
Burkholderia multivorans ATCC 17616 (plásmido) Burkholderia (3 proteínas)	subunidad de ADN polimerasa ligasa LigD	YP_001573706
R.Gammaproteobacteria		

ES 2 645 953 T3

Pseudomonas putida (plásmido) Pseudomonas (1 proteína)	supuesta ligasa	NP_542805
CLSK862724		
Organismo	Nombre de la proteína	Entrada
A. Archaea		
Archaeoglobus fulgidus DSM 4304 Archaeoglobus (1 proteína)	ADN ligasa, supuesta	NP_070553
J.Firmicutes		
Desulfotomaculum reduce MI-1 Desulfotomaculum (1 proteína)	ADN ligasa dependiente de ATP	YP_001113345
Moorella thermoacetica ATCC 39073Moorella (1 proteína)	ADN-ligasa dependiente de ATP, central	YP_430340
Pelotomaculum thermopropionicum SIPelotomaculum (1 proteína)	ADN-ligasa dependiente de ATP	YP_001211793
Thermoanaerobacter pseudethanolicus ATCC 33223Thermoanaerobacter (2 proteínas)	ADN-ligasa dependiente de ATP	YP_001664477
Thermoanaerobacter sp. X514Thermoanaerobacter (2 proteínas)	ADN ligasa dependiente de ATP	YP_001662589
CLSK820690		
Organismo	Nombre de la proteína	Entrada
A. Archaea		
Muestras de archaeon metanógena RC no cultivada-lenvironmental (1 proteína)	ADN-ligasa dependiente de ATP	YP_686457
C.Actinobacteria		
Mycobacterium avium 104Mycobacterium (2 proteínas)	Subunidad de ADN polimerasa ligasa LigD	YP_882332
Mycobacterium avium subsp. paratuberculosis K-10Mycobacterium (2 proteínas)	proteína hipotética	NP_960263

ES 2 645 953 T3

Saccharopolyspora erythraea NRRL 2338Saccharopolyspora (1 proteínas)	ADN ligasa, ATP dependiente	YP_001107793
N.Alphaproteobacteria		
Bradyrhizobium japonicum USDA 110Bradyrhizobium (2 proteínas)	ADN ligasa	NP_774671
Bradyrhizobium sp. BTAi1Bradyrhizobium (2 proteínas)	supuesta ADN ligasa dependiente de ATP	YP_001243518
CLSK808255		
Organismo	Nombre de la proteína	Entrada
N.Alphaproteobacteria		
Sinorhizobium medicae WSM419Sinorhizobium (2 proteínas)	región ADN polimerasa ligase LigD	YP_001326990
Sinorhizobium meliloti 1021 (plásmido) Sinorhizobium (2 proteínas)	supuesta proteína ADN ligasa dependiente de ATP	NP_437750
CLSK806855		
Organismo	Nombre de la proteína	Entrada
N.Alphaproteobacteria		
Cepa de Agrobacterium tumefaciens C58 (plásmido) Agrobacterium (3 proteínas)	ADN-ligasa dependiente de ATP	NP_395985
Cepa de Agrobacterium tumefaciens C58 (plásmido) Agrobacterium (3 proteínas)	ADN ligasa dependiente de ATP	NP_396032
Cepa de Agrobacterium tumefaciens C58 (plásmido) Agrobacterium (3 proteínas)	ADN ligasa dependiente de ATP	NP_396609
Rhizobium etli CFN 42 (plásmido) Rhizobium (10 proteínas)	supuesta proteína de ADN ligasa (ATP)	YP_472413
Rhizobium etli CIAT 652Rhizobium (10 proteínas)	probable proteína ADN ligasa dependiente de ATP	YP_001977317

ES 2 645 953 T3

Rhizobium etli CIAT 652 (plásmido) Rhizobium (10 proteínas)	supuesta proteína de ADN ligasa dependiente de ATP	YP_001985803
Rhizobium leguminosarum bv. trifolii WSM1325 (plásmido) Rhizobium (10 proteínas)	proteína del dominio ligasa ADN polimerasa LigD	YP_002973496
Rhizobium leguminosarum bv. trifolii WSM1325 (plásmido) Rhizobium (10 proteínas)	proteína del dominio ligasa LigD, ADN polimerasa	YP_002984974
Rhizobium leguminosarum bv. trifolii WSM1325 (plásmido) Rhizobium (10 proteínas)	proteína del dominio de ligasa, ADN polimerasa LigD	YP_002984992
Rhizobium leguminosarum bv. trifolii WSM2304Rhizobium (10 proteínas)	proteína del dominio ligasa, ADN polimerasa LigD	YP_002281897
Rhizobium leguminosarum bv. trifolii WSM2304 (plásmido) Rhizobium (10 proteínas)	proteína del dominio de ligasa, ADN polimerasa LigD	YP_002278005
Rhizobium leguminosarum bv. viciae 3841 (plásmido) Rhizobium (10 proteínas)	supuesta ADN ligasa	YP_764723
Rhizobium leguminosarum bv. viciae 3841 (plásmido) Rhizobium (10 proteínas)	supuesta ADN ligasa	YP_771149
Sinorhizobium meliloti (plásmido) Sinorhizobium (2 proteínas)	supuesta ADN ligasa dependiente de ATP	YP_001965531
Sinorhizobium meliloti 1021 (plásmido) Sinorhizobium (2 proteínas)	ADN ligasa dependiente de ATP	NP_435470
CLSK762775		
Organismo	Nombre de la proteína	Entrada
N.Alphaproteobacteria		
Rhodococcus jostii RHA1 (plásmido) Rhodococcus (2 proteínas)	ADN-ligasa dependiente de ATP	YP_708952
Rhodococcus opacus B4 (plásmido) Rhodococcus (2 proteínas)	supuesta ADN ligasa dependiente de ATP	YP_002776886
CLSK761995		
Organismo	Nombre de la proteína	Entrada
N.Alphaproteobacteria		

Nitrobacter hamburgensis X14Nitrobacter (1 proteína)	ADN ligasa dependiente de ATP	YP_579015
Rhodopseudomonas palustris BisB5 Rodopseudomonas (2 proteínas)	ADN ligasa dependiente de ATP	YP_569297
Rhodopseudomonas palustris TIE-1 Rhodopseudomonas (2 proteínas)	proteína del dominio ligasa, ADN polimerasa LigD	YP_001991309
CLSK523944		
Organismo	Nombre de la proteína	Entrada
R.Gammaproteobacteria		
Pseudomonas fluorescens (plásmido) Pseudomonas (3 proteínas)	supuesta ADN ligasa dependiente de ATP	YP_002887417
Pseudomonas putida (plásmido) Pseudomonas (3 proteínas)	supuesto fragmento de ligasa	NP_863069
Pseudomonas sp. ND6 (plásmido) Pseudomonas (3 proteínas)	ADN-ligasa dependiente de ATP	NP_943185
CLSK390680		
Organismo	Nombre de la proteína	Entrada
R.Gammaproteobacteria		
Mesorhizobium loti MAFF303099Mesorhizobium (3 proteínas)	ADN ligasa dependiente de ATP	NP_108227
Mesorhizobium loti MAFF303099Mesorhizobium (3 proteínas)	Proteína hipotética	NP_108282
Mesorhizobium loti MAFF303099 (plásmido) Mesorhizobium (3 proteínas)	Proteína de tipo ADN ligasa	NP_109396

La CV ligasa se proporciona a continuación:

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

MAITKPLLAATLENIEDVQFPCLATPKIDGIRSVKQTQMLSRTFKPIRNSVMNRL
 TELLPEGSDGEISIEGATFQDTSVMTGHKMYNAKFSYYWFDYVTDDPLKKYI
 DRVEDMKNYITVHPHILEHAQVKIIPVEINNITELLQYERDVLSKGFEGVMIRK
 PDGKYKFGFRSTLKEGILLKMKQFKDAEATIISMTALFKNTNTKTKDNFGYSKRST
 HKSGKVEEDVMGSIEVDYDGVVFSIGTGFDADQRRDFWQNKESYIGKMKVFKY
 FEMGSKDCPRFPVFIGIRHEEDR

5 El SFL es, en un aspecto, una enzima que puede mediar la formación de un enlace covalente entre dos extremos polinucleotídicos, por ejemplo, un extremo 3'-OH y un extremo 5'-PO4 se unen para formar un enlace fosfodiéster.

En algunos casos, el ligamiento de ADN implica uno o más de tres pasos de transferencia de nucleótido secuencial, discutidos a continuación. Los tres pasos químicos dependen de un cofactor de catión divalente. En un aspecto, el SFL es una ligasa dependiente de ATP o una ligasa dependiente de NAD⁺.

5 El SFL es ADN ligasa del virus de Chlorella (ChVLIg). Ho, y col., J Virol, 71 (3): 1931-19374 (1997) o un fragmento funcional del mismo. Por ejemplo, el SFL puede comprender uno o más dominios característicos de una ligasa, por ejemplo, un dominio de nucleótidiltransferasa N-terminal (NTasa) y/o un dominio OB C-terminal. El dominio OB comprende opcionalmente un beta-barril antiparalelo de cinco cadenas más una alfa-hélice. Dentro del dominio de la NTasa hay un bolsillo de unión a adenilato compuesta por los seis motivos peptídicos que definen la familia de enzimas polinucleótidas de la NTasa covalente. Opcionalmente, el dominio de la NTasa puede comprender uno o más de los motivos de aminoácidos de la ligasa I, Ia, III, IIIa, IV, V y VI, preferiblemente los seis motivos. El motivo I (por ejemplo, KxDGxR o un motivo "KXDG") contiene opcionalmente una lisina. Las secuencias ejemplares para cada motivo en CV ligasa son ATPKIDGIR (motivo I), SRT (motivo Ia), EGSDGEIS (motivo III), YWFDY (motivo IIIa), EGVMIR (motivo IV), LLKMK (motivo V). El motivo 1 contiene preferiblemente un residuo de lisina. Otros ejemplos de motivo I incluyen CELKLDGLA, VEHKVDGLS, CEPKLDGLA, CELKLDGVA, AEIKYDQVR, CEYKYDQQR, VDYKYDGER, FEIKYDQAR, FEGKWDGYR, AREKIHGTN, ACEKVHGTN, ILTKEDGSL y VEEKVDGY. Ejemplos del motivo Ia incluyen TRG, SRT, SRR, SRN, SRS, KRT, KRS, SKG y TRG. Ejemplos del motivo III incluyen LEVRGEVF, VEVRGECY, LEVRGEVY, LEARGEAF, FMLDGELM, EGSDGEIS, FILDTEAV, FIIIEGIV, AIVEGELV, VVLDGEAV, YQVFGGEFA, 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. Ejemplos del motivo IV incluyen DGVIK, DGIVK, DGVVVK, DGTVLK, EGLIVK, EGVMIR, EGLMVK, EGVMVK, EGLMAK, EGVIK, EGYVLK, EGVVIR, EGYVAV y EGIIMK. Ejemplos del motivo V incluyen AVAFK, AIAYK, ALAYK, AIAYK, WWKMK, LLKMK, WLKLV, WIKLV, WIKLV, WVKDK, AIKCK, IIKLR, HFKIK y IVKYV. El SFL opcionalmente comprende los seis motivos. Opcionalmente, los seis motivos se encuentran juntos en una ligasa de origen natural, tal como una SFL identificada en este documento. Opcionalmente, la SFL no es una enzima que bloquee el ARN.

Opcionalmente, cualquiera de los métodos de ligamiento descritos en este documento se realiza mediante un SFL que es un fragmento funcional o derivado de la CV ligasa.

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

35 En cualquier ligamiento descrito en este documento, el producto de ligamiento se detecta opcionalmente. El ligamiento puede detectarse mediante cualquier método conocido o método descrito en este documento. Por ejemplo, el cebador y/o la plantilla pueden inmovilizarse en un soporte y marcarse con una sonda. El marcador en la sonda inmovilizada ligada se puede detectar después de que la sonda no ligada se haya eliminado por lavado. En otras realizaciones, se marca una o más de las sondas, cebadores o plantillas. Cualquiera o más de estos reactivos se pueden inmovilizar opcionalmente.

40 En cualquier ligamiento en el presente documento, el ligamiento puede repetirse durante un número deseado de ciclos, por ejemplo, cualquier reactivo que se haya sometido a un primer ciclo de ligamiento se usa como reactivo en un siguiente ciclo de ligamiento. Por ejemplo, el cebador, la sonda o plantilla de un primer ciclo de ligamiento puede usarse como cebador, sonda o plantilla en un próximo ciclo de ligamiento. En algunas realizaciones (por ejemplo, secuenciación de ligamiento), 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 una sonda o plantilla en el siguiente, la sonda de un primer ciclo puede usarse como cebador o plantilla, o la plantilla de un primer ciclo puede usarse como cebador o sonda del siguiente.

50 El siguiente ciclo puede diseñarse de manera que tanto el reactivo ligado como el no ligado de un primer ciclo puedan actuar como reactivo en el siguiente ciclo. Por ejemplo, el cebador ligado y no ligado de un primer ciclo de ligamiento puede usarse como cebador en el siguiente ciclo. Alternativamente, el siguiente ciclo puede diseñarse de modo que solo un producto de ligamiento del primer ciclo pueda actuar como reactivo de ligamiento en el ciclo siguiente, 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 están "tapados" o no se pueden volver a unir antes de que se ejecute el siguiente ciclo de ligamiento.

55 Cualquiera o más ligasas y/o métodos de ligamiento abarcados por la invención pueden usarse opcionalmente en uno o más formatos de ensayo de ligamiento, y/o se realizan en el contexto de una o más aplicaciones de ligamiento específicas. Los ejemplos no limitantes de formatos de ensayo incluyen: ensayo de ligamiento de oligonucleótidos (OLA), una reacción en cadena de ligasa (LCR), una reacción de detección de ligasa (LDR) y ensayos de combinación tales como OLA acoplado con la reacción en cadena de la polimerasa (PCR), por ejemplo, OLA-PCR y

60

PCR-OLA, la reacción en cadena combinada (CCR, una combinación de PCR y LCR) y PCR-LDR (véase, por ejemplo, Landegren y col., *Science* 241: 1077-80, 1988; Barany, *Proc. Natl. Acad. Sci.* 88:189-93, 1991; Grossman et al., *Nucl. Acids Res.* 22(21): 4527 - 34, 1994; Bi y Stambrook, *Nucl. Acids Res.* 25(14): 2949 - 51, 1997; Zirvi y col., *Nucl. Acids Res.*, 27 (24): e40, 1999; Pat. de EE.UU. No. 4.988.617; y Publicaciones PCT Nos. WO 97/31256 y WO 01/92579. Ejemplos no limitativos de aplicaciones específicas incluyen: amplificación de plantilla, detección y/o cuantificación de la presencia de un ácido nucleico particular, por ejemplo, en una muestra de diagnóstico, secuenciación de ligamiento, análisis de polimorfismo de nucleótido simple (SNP), genotipificación de SNP, detección de mutaciones, identificación de genes de copia única, detección de secuencias repetidas de microsatélite y mapeo de aducto de ADN, entre otras cosas. Véase, también, Whitely et al., Pat. U.S. No. 4.883.750; Letsinger et al., Pat. U.S. No. 5.476.930; Fung et al., Pat. U.S. No. 5.593.826; Kool, U.S., Pat. No. 5.426.180; Landegren et al., Pat. U.S. No. 5.871.921; Xu y Kool, *Nucleic Acids Research*, 27: 875 - 881 (1999); Higgins et al., *Methods in Enzymology*, 68: 50 - 71 (1979); Engler et al., *The Enzymes*, 15 - 3 - 29 (1982); y Namsaraev, publicación de patente de los Estados Unidos 2004/0110213.

En una realización, el método de ligamiento comprende o consiste en un ensayo de ligamiento de proximidad (PLA). Los PLA normalmente implican al menos tres pasos. La primera etapa es típicamente la unión de la primera y la segunda sondas (por ejemplo, sondas de anticuerpos) a un ligando (por ejemplo, una proteína de interés) de manera que las sondas están muy cerca entre sí. Cada una de las sondas contiene típicamente un oligonucleótido. Los oligonucleótidos se acercan entre sí con la unión de las sondas y, en la segunda etapa, se ligan entre sí (por ejemplo, el evento de ligamiento). Los oligonucleótidos ligados pueden luego 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 diana en una muestra, que comprende (a) poner en contacto la proteína diana con al menos un primer y segundo agente de unión, teniendo cada agente de unión especificidad de unión por la proteína y ser unido a al menos un polinucleótido, (b) ligar los oligonucleótidos en el primer y el segundo agente de unión entre sí usando una ligasa para producir un ácido nucleico diana y amplificar el ácido nucleico diana; (c) detectar si está presente o no el ácido nucleico diana amplificado, y (d) concluir que la proteína diana está presente en la muestra si se detecta una cantidad significativa de ácido nucleico diana amplificado, y/o concluir que la proteína diana está ausente de la muestra si no se detecta una cantidad significativa de ácido nucleico diana amplificado. Opcionalmente, la etapa (d) comprende alternativamente o adicionalmente medir o evaluar de otro modo la cantidad de ácido nucleico diana amplificado, y tomar la cantidad de ácido nucleico amplificado como un indicador de la cantidad de proteína diana presente en la muestra.

En cualquiera de los métodos descritos en este documento, las sondas de ligamiento de polinucleótidos pueden usarse para la detección de un ácido nucleico diana. La sonda de ligamiento es opcionalmente capaz de unirse a un ácido nucleico diana de secuencia complementaria a través de uno o más tipos de enlaces químicos, habitualmente a través de emparejamiento de bases complementarias, habitualmente a través de la formación de enlaces de hidrógeno. Como se usa en este documento, una sonda puede incluir bases naturales (es decir, A, G, C o T) o modificadas (7-desazaguanosina, inosina, etc.). Además, las bases en una sonda se pueden unir mediante un enlace distinto de un enlace fosfodiéster, siempre que no interfiera con la hibridación. De este modo, por ejemplo, las sondas pueden ser ácidos nucleicos peptídicos en los que las bases constituyentes se unen mediante enlaces peptídicos en lugar de enlaces fosfodiéster. Cualquier experto en la técnica entenderá que las sondas pueden unirse a secuencias diana que carecen de complementariedad completa con la secuencia de sonda dependiendo de la rigurosidad de las condiciones de hibridación. Las sondas preferiblemente se marcan directamente como con isótopos, cromóforos, luminóforos, cromógenos o se marcan indirectamente, tal como con biotina, a la que más adelante se puede unir un complejo de estreptavidina. Al ensayar la presencia o ausencia de la sonda, se puede detectar la presencia o ausencia de la secuencia seleccionada o subsecuencia.

El ensayo de ligamiento de oligonucleótidos (OLA) es un método conveniente y altamente riguroso que permite la distinción entre variantes de secuencia de ADN conocidas (Landegren, 1988). El análisis multiplex de loci altamente polimórficos es útil para la identificación de individuos, por ejemplo, para pruebas de paternidad y en ciencia forense, emparejamiento entre donante y receptor de trasplante de órgano, diagnóstico de enfermedad genética, pronóstico y asesoramiento prenatal, y otras pruebas genéticas que dependen de la discriminación de las diferencias de una sola base en una multiplicidad de loci (Delahunty, 1996). Por ejemplo, diferentes ensayos en los que dos quimeras de PNA-ADN, una quimera de secuencia de tipo salvaje (WT) y una quimera de secuencia mutante, portan diferentes colorantes fluorescentes. Solo cuando la secuencia mutante está presente en la muestra diana, la quimera de la secuencia mutante se ligará a la adyacente segunda sonda hibridada (oligo) si el par de bases mutantes está en el sitio de ligamiento.

a) Determinación de la secuencia

En algunas realizaciones, se pueden llevar a cabo ciclos sucesivos de ligamiento, donde el producto de ligamiento de un ciclo anterior se usa como cebador, sonda y/o plantilla en un ciclo sucesivo.

Un uso ejemplar de tal ligamiento repetitivo es la secuenciación de ligamiento. Una plantilla a secuenciar contiene opcionalmente una región de unión a cebador y una región de polinucleótido de secuencia desconocida. Un cebador

con un extremo ligable (por ejemplo, un grupo 3' OH libre o un grupo 5' fosfato) se hibrida con la región de unión del cebador. Una sonda de extensión se hibrida con la plantilla adyacente al cebador. El nucleótido proximal de la sonda forma un par de bases complementarias con un nucleótido desconocido en la plantilla. La sonda de extensión se liga luego al cebador, opcionalmente con un SFL, dando como resultado un dúplex extendido. Después del ligamiento, el marcador unido a la sonda de extensión se detecta opcionalmente. El proceso se repite durante un número deseado de ciclos, utilizando el producto de ligamiento de un ciclo como el cebador del siguiente ciclo. Opcionalmente, la sonda de extensión tiene un extremo distal no ligable, y luego se escinde para proporcionar un dúplex extendido ligable.

En un ejemplo, el extremo 3' del cebador se liga al extremo 5' de la sonda de extensión. La etapa opcional de escisión puede realizarse, por ejemplo, en un enlace fosforotiolato usando AgNO_3 u otra sal que proporcione iones Ag^+ , dejando un grupo fosfato en el extremo 3' del dúplex extendido. El tratamiento con fosfatasa se usa para generar un extremo de sonda extensible en el dúplex extendido.

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

También se proporcionan familias de sondas para su uso en diversos métodos de ligamiento en este documento, por ejemplo, secuenciación. Las familias de sondas se caracterizan opcionalmente porque cada familia de sondas comprende una pluralidad de sondas oligonucleotídicas marcadas de diferente secuencia 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. Preferiblemente, las sondas comprenden un enlace internucleosídico escindible. El enlace divisible se puede ubicar en cualquier lugar de la sonda. Preferiblemente, las sondas tienen un resto que no es extensible por ligasa en un extremo. Preferiblemente, las sondas se marcan en una posición entre el enlace escindible y el resto que no es extensible por ligasa, de manera que la escisión del enlace escindible después de la unión de una sonda a un extremo de sonda extensible da como resultado una porción no marcada ligada al extensible terminal de la sonda y una parte marcada que ya no está unida a la parte no marcada.

En codificaciones de bases múltiples, las sondas en cada familia de sondas comprenden preferiblemente 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 inferior a 100, generalmente inferior o igual a 30. Los nucleósidos X pueden ubicarse en cualquier parte de la sonda. Los nucleósidos X no necesitan ubicarse en posiciones contiguas. De manera similar, los nucleósidos N no necesitan ubicarse en posiciones contiguas. En otras palabras, los nucleósidos X y N pueden ser intercalados. Sin embargo, se puede considerar que los nucleósidos X tienen una secuencia de 5' a 3', en el entendimiento de que los nucleósidos no necesitan ser contiguos. Por ejemplo, se consideraría que los nucleósidos X en una sonda de estructura $X_A N X_G N N X_C N$ tienen la secuencia AGC. Del mismo modo, se puede considerar 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 está restringida por la identidad de uno o más nucleósidos diferentes X en la sonda. Por lo tanto, en general, solo ciertas combinaciones de nucleósidos X están presentes en cualquier sonda particular y dentro de las sondas en cualquier familia de sondas 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 . Por lo tanto, la identidad de uno o más nucleótidos en X limita las posibles identidades para uno o más de los otros nucleósidos.

Los nucleósidos N preferiblemente se seleccionan independientemente y pueden ser A, G, C o T (u, opcionalmente, un nucleósido reductor de la degeneración). Preferiblemente, la secuencia de nucleósidos N representa todas las secuencias posibles de longitud k , excepto que uno o más N pueden ser un nucleósido reductor de la degeneración. Las sondas contienen así dos porciones, de las cuales la porción que consiste en los nucleósidos N se denomina porción no limitada y la porción que consiste en los nucleósidos X se denomina porción restringida. Como se describió anteriormente, las porciones no necesitan ser nucleósidos contiguos. Las sondas que contienen una parte restringida y una parte no restringida se denominan en este documento sondas parcialmente restringidas. Preferiblemente, uno o más nucleósidos en la porción restringida están en el extremo proximal de las sondas, es decir, en el extremo que contiene el nucleósido que se ligará al extremo de la sonda extensible, que puede ser el extremo 5' o 3' de la sonda de oligonucleótidos en diferentes realizaciones de la invención.

Dado que la porción restringida de cualquier sonda de oligonucleótido 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 parte restringida. La información puede ser o no suficiente para identificar con precisión uno o más de los otros nucleósidos, pero será suficiente para eliminar una o más combinaciones de nucleótidos posibles y/o permutaciones en la parte restringida. Opcionalmente, la información no es suficiente para eliminar cualquier posible identidad de cualquier nucleótido individual en la porción restringida. En ciertas

realizaciones preferidas, 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, determinar la identidad y el orden de los nucleósidos que comprenden la porción restringida.

5 Como en los métodos de secuenciación de codificación de una sola base descritos anteriormente, el nucleósido más proximal en una sonda de extensión que es complementario a la plantilla se liga a un extremo extensible de un oligonucleótido inicializante (en el primer ciclo de extensión, ligamiento y detección) y a un extremo extensible de una sonda de oligonucleótidos extendida en ciclos posteriores de extensión, ligamiento y detección. La detección del marcador asociado 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 al menos 2 veces degenerada, el nombre de la familia de sondas no identifica en sí mismo ningún nucleótido en la porción restringida. Sin embargo, dado que la secuencia de la porción restringida es uno 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 parte restringida de la sonda constituye su porción de determinación 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 identificando la familia de sondas a la que pertenece elimina una o más posibilidades para la identidad de un nucleótido en la plantilla a la que la sonda de extensión se hibrida. En realizaciones preferidas de la invención, las sondas parcialmente restringidas comprenden un enlace escindible entre dos nucleósidos cualesquiera.

20 En ciertas realizaciones, las sondas parcialmente restringidas tienen la estructura general $(X)_j(N)_k$, en la que X representa un nucleósido, $(X)_j$ es al menos 2 veces degenerado en cada posición de modo que X puede ser cualquiera de al menos 2 nucleósidos que tienen diferentes especificidades de emparejamiento de bases, N representa cualquier nucleósido, j es al menos 2, k está entre 1 y 100, y al menos un N o X distinto de X en el extremo de la sonda comprende un resto detectable. Preferiblemente $(N)_k$ es independientemente 4 veces degenerado 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 estar ocupadas por un nucleótido de reducción de la degeneración. Los nucleósidos en $(X)_j$ pueden ser idénticos o diferentes, pero no se seleccionan de forma independiente. 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. Las sondas contienen así dos porciones, de las cuales $(N)_k$ es la porción no restringida y $(X)_j$ es la porción restringida.

30 En ciertas realizaciones preferidas de la invención, las sondas parcialmente restringidas tienen la estructura $5'-(X)_j(N)_kN_B^*-3'$ o $3'-(X)_j(N)_kN_B^*-5'$, donde N representa cualquier nucleósido, N_B representa un resto que no es extensible por ligasa, $*$ representa un resto detectable, $(X)_j$ es una porción restringida de la sonda que es al menos 2 veces degenerada 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 otro que la X en el 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 nucleótido más proximal en $(N)_k$, entre nucleósidos dentro de $(N)_k$, o entre el nucleósido terminal en $(N)_k$ y N_B . Preferiblemente, el enlace escindible es un enlace fosforotiolato.

40 Una pluralidad de familias de sondas se conoce como una "colección" de familias de sondas. Las sondas en cada familia de sondas en una colección de familias de sondas están marcadas con un marcador que se distingue de los marcadores utilizados para marcar otras familias de sondas en la colección. Cada familia de sondas preferiblemente tiene su propio conjunto definido de secuencias, que opcionalmente no se solapan con ninguna otra familia de sondas. Preferiblemente, 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 determinante de la secuencia. Preferiblemente, una colección de familias de sondas comprende o consiste en 4 familias de sondas marcadas de manera distinguible.

50 Preferiblemente, las porciones que determinan la secuencia de las sondas en cada familia de sondas tienen la misma longitud, y preferiblemente las porciones determinantes de la secuencia de las familias de sondas en una colección de familias de sondas tienen la misma longitud. Preferiblemente, la porción de determinación de la secuenciación es una porción restringida que tiene al menos 2 nucleósidos de longitud.

En algunos casos, una serie de ciclos de ligamiento a partir de un cebador particular permite la determinación parcial de una secuencia, es decir, la identificación de nucleótidos individuales espaciados entre sí en una plantilla. Opcionalmente, para reunir información más completa, se realiza una pluralidad de reacciones en las que cada reacción utiliza un oligonucleótido de inicialización i diferente. Los oligonucleótidos de inicialización i se unen a diferentes porciones de la región de unión. Preferiblemente, los oligonucleótidos de inicialización se unen en posiciones tales que los extremos extensibles de los diferentes oligonucleótidos de inicialización se compensan con 1 o más nucleótidos entre sí cuando se hibridan con la plantilla. Por ejemplo, como se muestra en la Fig. 3, se realizan las reacciones de secuenciación $1...N$. Los oligonucleótidos iniciadores $i_1...i_n$ tienen la misma longitud y se unen de forma tal que sus nucleótidos terminales se hibridan con posiciones adyacentes sucesivas en la plantilla. Las sondas de extensión $e_1...e_n$ se unen así sucesivamente a las regiones adyacentes sucesivas de la plantilla y se

ligan a los extremos 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 del polinucleótido, es decir, el primer polinucleótido desconocido en la plantilla. En el segundo ciclo de extensión, el ligamiento y la detección, el nucleótido terminal de la sonda e_{12} es complementario al nucleótido de la región del polinucleótido, es decir, el segundo nucleótido de secuencia desconocida. Del mismo modo, los nucleótidos terminales de las sondas de extensión ligadas a los dúplex inicializados con los oligonucleótidos iniciadores i_2, i_3, i_4 y así sucesivamente, serán complementarios al tercer, cuarto y quinto nucleótidos de la secuencia desconocida. Se apreciará que los oligonucleótidos de inicialización pueden unirse a regiones progresivamente más alejadas de la región del polinucleótido en lugar de progresivamente más cerca de ella.

10 La función espaciadora de los nucleótidos no terminales de las sondas de extensión permite la adquisición de información de secuencia en posiciones en la plantilla que se eliminan considerablemente de la posición en la que el oligonucleótido iniciador se une sin requerir que se realice un número correspondientemente grande de ciclos en cualquier plantilla dada.

a. Taponar

15 En cualquiera de uno o más métodos de ligamiento repetitivo de la presente invención, es posible que menos de todas las sondas con extremos extensibles participen en una reacción de ligamiento exitoso en cada ciclo de extensión, ligamiento y escisión. Se apreciará que si tales sondas participaban en ciclos sucesivos, la precisión de cada paso de identificación de nucleótidos podría disminuir progresivamente. En ciertas realizaciones de la invención, se incluye una etapa de taponación para evitar que los extremos extensibles no sufran un ligamiento de la participación en ciclos futuros. Cuando se secuencian en la dirección 5' a 3' usando sondas de extensión que contienen un enlace 3'-O-P-S-5' fosforotiolato, se puede realizar el taponamiento extendiendo los extremos extensibles no ligados con una ADN polimerasa y un resto no extensible, por ejemplo, un nucleótido que termina la cadena tal como un didesoxinucleótido o un nucleótido con un resto bloqueante unido, por ejemplo, después de la etapa de ligamiento o detección. Cuando se secuencian en la dirección 3' a 5' usando sondas de extensión que contienen un enlace 3'-S-P-O-5' fosforotiolato, se puede realizar el taponamiento, por ejemplo, tratando la plantilla con una fosfatasa, por ejemplo, después del ligamiento o la detección. También se pueden usar otros métodos de taponamiento.

Se contempla que cualquier ensayo o método de ligamiento en la presente memoria pueda ser altamente multiplexado de modo que se pueda realizar un gran número de ensayos (por ejemplo, mayores de 1.000) en paralelo, por ejemplo, simultáneamente.

Cualquier ensayo de ligamiento multiplexado puede realizarse opcionalmente en sustratos sólidos en los que uno o más reactivos de ligamiento (por ejemplo, la ligasa, plantilla, sonda y/o cebador) pueden inmovilizarse sobre un soporte o superficie sólida. Opcionalmente, los cebadores, la sonda y/o la plantilla se pueden unir a diferentes porciones de un sustrato sólido en forma de una matriz. Opcionalmente, la plantilla, el cebador o la sonda se pueden unir covalentemente al sustrato sólido.

Opcionalmente, uno o más reactivos de ligamiento están marcados (por ejemplo, plantilla, sonda y/o cebador) de manera que los productos de ligamiento pueden discriminarse entre sí. Alternativamente, los productos de ligamiento se pueden distinguir en función de: (i) el tamaño usando electroforesis o cromatografía y/o (ii) marcadores detectables (Grossman, 1994). Por ejemplo, los ensayos de ligamiento multiplexados pueden realizarse en una única muestra en un solo vaso.

Cualquiera de los reactivos de la presente invención (por ejemplo, cebador, sonda y/o plantilla) se puede inmovilizar opcionalmente en una fase sólida. La fase sólida comprende opcionalmente una superficie a la que se pueden unir uno o más reactivos electrostáticamente, hidrofólicamente 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 silástico; poliésteres; poliamidas; celulosa y derivados; acrilatos; metacrilatos; polivinilo; cloruro de vinilo; cloruro de polivinilo; poli(fluoruro de vinilo); copolímeros de poliestireno; gel de sílice; vidrio de obleas de sílice; agarosa; dextranos; liposomas; metales proteicos insolubles; y, nitrocelulosa. Las fases sólidas representativas se pueden formar en artículos apropiados tales como perlas (por ejemplo, micropartículas), tubos, tiras, discos, papeles de filtro, placas y similares.

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

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

Composiciones

También se proporciona un polipéptido que comprende o consiste esencialmente en la secuencia de aminoácidos de un SFL proporcionado en este documento, generalmente proporcionado como números de entrada del GenBank en las tablas, o un fragmento o variante de la misma. El fragmento o variante de la misma preferiblemente puede ligar oligonucleótidos cortos con una eficacia comparable a su enzima original. Opcionalmente, las variantes o fragmentos de los SFL son al menos un 70%, 80, 90%, 95%, 99% idénticos al SFL original. En un ejemplo, el SFL es una ligasa derivada de la ADN-ligasa Hin (por ejemplo, DLX, DLXd o DLXd2) o cualquier fragmento o variante de la misma que aún retiene uno o más residuos mutantes mostrados en la Figura 2, y/o tiene uno o más aminoácidos C-terminados eliminados, por ejemplo, 22 aminoácidos C-terminales eliminados. Por ejemplo, la ADN-ligasa Hin mutante es al menos 70% idéntica a la secuencia de Hin D ligasa proporcionada en la Figura 2 o en el Número de entrada del GenBank P44121, cuya ligasa comprende una mutación de aminoácido en la posición 193 de la secuencia Hin D ligasa proporcionada en la Figura 2 o en el número de entrada de GenBank P44121. Opcionalmente, la mutación de aminoácido consiste en cambiar la glicina en la posición 193 por ácido aspártico o ácido glutámico.

También se proporciona un ácido nucleico que codifica los nuevos SFL o fragmentos o variantes descritos en este documento. También se proporcionan genes del nuevo SFL que comprende las secuencias que dirigen la expresión de SFL. También se proporcionan vectores de ácido nucleico que comprenden un ácido nucleico que codifica el nuevo SFL y una célula hospedadora que comprende dicho vector. También se proporciona un anticuerpo que se puede unir específicamente al nuevo SFL pero no a su progenitor natural.

Opcionalmente, el nuevo 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 o más ligasas descritas en la presente memoria que comprende: expresar el SFL en una célula hospedadora, por ejemplo, cultivando dichas células hospedadoras en condiciones tales que se expresa la ligasa; y opcionalmente recuperar o purificar dicha ligasa.

Kits

También se proporcionan kits que comprenden componentes para realizar reacciones de ligamiento de acuerdo con la descripción. Los kits de la invención comprenden una ADN ligasa del virus Chlorella (CV ligasa) o un fragmento funcional de la misma, un oligonucleótido ("sonda") y un polinucleótido ("cebador"), en donde: la sonda tiene N residuos de nucleótidos de longitud, donde N es de 2 a 7. En algunas realizaciones, un kit puede incluir CV ligasa 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 está unido opcionalmente al extremo 5', o alternativamente al extremo 3'. Opcionalmente, una o más sondas son escindibles.

Los kits opcionalmente proporcionan uno o más cebadores, sondas y/o plantillas descritos en este documento. Los cebadores y/o sondas pueden tener uno o más rasgos o características descritos en este documento. Por ejemplo, las sondas pueden comprender un enlace escindible (por ejemplo, fosforotiolato). Los kits pueden contener un reactivo de escisión adecuado para escindir enlaces fosforotiolato, por ejemplo, AgNO_3 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 ADN glicosilasa. Ciertos kits contienen sondas de oligonucleótidos que comprenden un nucleótido de disacárido y contienen peryodato como reactivo de escisión. En ciertas realizaciones, los kits contienen una colección de familias o colecciones de sondas de oligonucleótidos marcadas de manera distinguible descritas en este documento. 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 un SFL que tiene una o más actividades mencionadas en este documento. Los kits pueden incluir reactivos de ligamiento (por ejemplo, ligasa, tampones, etc.) e instrucciones para practicar la realización particular de la invención. Se pueden incluir 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, p. ej. perlas magnéticas, que están preunidas a cebadores, cebadores o plantillas o se derivatizan para poder unirse a tales moléculas. Las perlas pueden ser funcionalizadas con un cebador para realizar la amplificación por PCR. Otros componentes opcionales incluyen soluciones de lavado; vectores para insertar plantillas para la amplificación por PCR; reactivos de 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 de manera distinguible, que pueden marcarse de manera que la identidad del marcador proporcione información sobre la secuencia de la sonda. Los kits también pueden comprender composiciones o reactivos diferentes o adicionales para usar en el protocolo de determinación de secuenciación de ligamiento. 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 son, por ejemplo, degeneradas 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 evidente por el contexto, cualquier característica se puede reivindicar en combinación con cualquier otra, o se puede reivindicar que no está presente en combinación con otra característica. Una característica puede ser, por ejemplo, cualquier variación, paso, característica, propiedad, composición, método, paso, grado, nivel, componente, material, sustancia, elemento, modo, variable, aspecto, medida, cantidad o realización.

Muchas características descritas en este documento son opcionales. Si alguna característica no se indica explícitamente como necesaria, se debe considerar como opcional. Ejemplos no limitativos de lenguaje que indican que una característica es opcional incluyen términos tales como "variación", "dónde", "mientras", "cuándo", "opcionalmente", "incluir", "preferido", "especial", "recomendado", "aconsejable", "particular", "debería", "alternativa", "típico", "representativo", "varios", "tales como", "similares", "puede", "ejemplo", "realización" o "en un aspecto" o "si" o cualquier combinación y/o variación de tales términos.

Cualquier indicación de que una característica sea opcional está destinada a proporcionar un soporte adecuado para las reivindicaciones que incluyen un lenguaje cerrado o exclusivo o negativo (por ejemplo, bajo 35 U.S.C. 112 o Art. 83 y 84 de EPC). El lenguaje exclusivo excluye específicamente la función recitada particular de incluir cualquier tema adicional. Por ejemplo, si se indica que A puede ser un medicamento X, dicho lenguaje tiene la intención de proporcionar soporte para una condición que especifique explícitamente que A está compuesto solo por X, o que A no incluye ningún otro medicamento además de X. El término "negativo" excluye explícitamente la característica opcional propia del alcance de las reivindicaciones. Por ejemplo, si se indica que el elemento A puede ser X, dicho término está destinado a proporcionar soporte para una condición que especifique explícitamente que A no es X.

Ejemplos no limitantes de términos exclusivos o negativos incluyen "solo", "exclusivamente", "que consiste en", "que consiste esencialmente en", "solo", "sin", "en ausencia de (por ejemplo, otros elementos del mismo tipo, estructura y/o función)" "excluyendo", "no", "ni", "no puede", o cualquier combinación y/o variación de dichas expresiones.

De forma similar, los referentes tales como "un", "uno/a", "dcho/a" o "el/la" están destinados a admitir a los referentes explícitamente únicos o plurales cuando así lo deseen. Los ejemplos no limitantes de referentes en plural incluyen "al menos uno", "uno o más", "más de uno", "dos o más", "una multiplicidad", "una pluralidad", "cualquier combinación de", "cualquier permutación de", "uno cualquiera o más", etc.

La cita de cualquier publicación es para su descripción antes de la fecha de presentación y no debe interpretarse como una admisión de que la presente invención no tiene derecho a anteceder a dicha publicación en virtud de la invención anterior.

Los siguientes ejemplos se proporcionan con fines ilustrativos y no están destinados a limitar la invención.

Ejemplos

Ejemplo 1

Este ejemplo describe el ligamiento de sondas cortas.

Los reactivos oligonucleotídicos fueron los siguientes. Los oligos cortos en el ligamiento hacia adelante (es decir, el ligamiento del extremo 3' de la sonda al extremo 5' del cebador) fueron: 8mer: 5'-CTCATTCG-3'; 6mer: 5'-TCATTCG-3' 5mer: 5'-ATTTCG-3'; 4mer: 5'-TTTCG-3'; 3mer: 5'-TCG-3'. Los oligos cortos usados para el ligamiento inverso (es decir, el ligamiento del extremo 5' de la sonda al extremo 3' del cebador) eran los mismos que los oligos de ligamiento directo 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-CTGCCCGGGTTCCTCATTCTCT-3'.

Preparación del sustrato de ligasa: las perlas magnéticas de MyOne marcadas con ácido carboxílico se obtuvieron de Invitrogen. El oligonucleótido de base 41 marcado con amino en 5' (CCA CTA CGC, CTC, CGC, TTT, CCT, CTC, TAT, GGG, CAG, TCG, GTG, AT) se acopló a perlas MyOne usando química de acoplamiento de amina estándar como se describe en Nakajima y col., Bioconjugate Chem. 1995, 6(1), 123-130 (Figura 1b). Se usó PCR para extender este oligonucleótido usando (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 ligamiento de 105 bases estaba ahora unido 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 GAA TAG AGA ATG AGG AAC CCG GGG CAG, Figura 1 a). Se midió la cantidad de oligo-unido de 105 bases a la perla mediante hibridación de un oligo "cebador" marcado fluorescentemente, con marcador de fluoresceína 3' (6FAM) para la detección por electroforesis capilar. La cantidad de fluorescencia desnaturalizada de la plantilla se comparó con una curva estándar.

Las reacciones de ligamiento 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 de cebador se hibridó a la plantilla en tampón de ligasa calentando a 85°C durante 10 minutos seguido de enfriamiento lento a temperatura ambiente. En el caso del ligamiento directo, el cebador se fosforiló en 5' y se marcó 3' 6FAM. En el caso del ligamiento inverso, el

5 cebador fue marcado en 5' con 6FAM. Las reacciones de ligamiento se realizaron en una placa de 96 pocillos usando plantilla/cebador 2 nM, ligasa de 2,0 μM, oligonucleótido corto de 2-5 μM en un volumen total de 50 μM. Las reacciones de ligamiento continuaron durante 20 minutos a 15°C. Las reacciones se detuvieron usando SDS al 1% v/v seguido de separación magnética de las perlas con plantilla de la solución. Las perlas se lavaron tres veces en 100 μl de SDS al 1% antes del lavado con 100 μl de formamida al 100%. Luego se analizaron por electroforesis capilar 5 μl del sobrenadante que contenía el iniciador y el producto de ligamiento desnaturalizados marcados con FAM. La eficacia de ligamiento se calculó como la relación de las áreas de los picos determinada por CE, donde el cebador FAM ligado, no ligado y marcado se separó, es decir, Eficiencia = ligado/(ligado + no ligado).

10 En las Figuras 3-5 se muestran una serie de ligamientos de sondas de diferentes longitudes con una variedad de SFL. En cada panel, la reacción de ligamiento se representa 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. Al utilizar SFL, el tamaño de un conjunto de sonda de secuenciación (por ejemplo, para la secuenciación de ligamiento en SOLiD™) se puede reducir del conjunto actual de 1024 8-mers a tan solo 16 (en el caso de los dí-meros), 64 (en el caso de los trímeros), o 256 en el caso de los tetrámeros.

15 En otro experimento, se comparó la eficacia de ligamiento de la SF ligasa usando el mismo sistema experimental que antes para la T4 ADN ligasa, usando sondas de secuenciación de ligamiento de 8-meros que se conjugaron o no se conjugaron con colorantes. SFL es la eficacia de ligamiento más alta para la sonda oscura (búsqueda) y la sonda de secuenciación marcada. Esto es cierto a la concentración de sonda 200 nM, que actualmente se utiliza en SOLiD. Específicamente, se usa una mezcla de 1024 sondas en SOLiD, donde cada sonda individual de las sondas 20 1024 está presente a una concentración de trabajo final de 200 nM. Las otras concentraciones fueron 2000 nM de enzima y 1 nM de plantilla de perlas, donde el ligamiento se realizó a 15°C durante 30 minutos. Los resultados se muestran en la Tabla 2.

Tabla 2

Máxima eficiencia de ligamiento de las sondas de 8meros mediante diversas enzimas ligasas		
Ligasa	Oscuro	Cy 5
SFL	0,91 ± 0,01	0,82 ± 0,01
T4	0,67 ± 0,02	0,62 ± 0,02

25 Ejemplo 2

Se determinó que la fidelidad de una ligasa de huella pequeña (ligasa CV) era sorprendentemente mejor que la T4 ADN ligasa en la secuenciación de ADN por ligamiento en el sistema de secuenciación SOLiD. Una plantilla de ADN de 35 nucleótidos se unió covalentemente a perlas, de manera 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 ligamiento a 20°C. La reacción de secuenciación ligó "sondas" de oligonucleótidos de 8-meros marcadas con colorante fluorescente sobre un cebador, y se realizaron un total de siete sucesos consecutivos de ligamiento o sondeo en cada molécula plantilla. El número aproximado de eventos de ligamiento fue, por lo tanto, de 100.000 (es decir, el número de moléculas patrón por perla) x 40.000.000 (es decir, número de perlas) x 7 (número de eventos de ligamiento por plantilla) = 2,8 x 10¹³.

35 Los resultados se dan en la Tabla 3 a continuación. OMM = Porcentaje de todos los productos de ligamiento (que son el resultado de siete eventos consecutivos de ligamiento) sin desapareamientos; 3MM = Porcentaje de todos los productos de ligamiento (que son el resultado de siete eventos consecutivos de ligamiento) con menos de 3 desapareamientos.

Tabla 3

Ligasa	OMM	3MM
T4	20%	57%
SF	27%	65%

40 Los datos de la secuenciación por ligamiento demuestran que el 20% del tiempo el producto de ligamiento era perfectamente complementario a la plantilla y se formaba mediante siete sucesos de ligamiento consecutivos que eran todos correctos. Se puede calcular una estimación de la fidelidad de ligamiento (expresada como la probabilidad porcentual de que un evento de ligamiento sea correcto, o "p") a partir de la ecuación: tasa OMM = (p/100)ⁿ donde p es el porcentaje de fidelidad de ligamiento y n es el número de eventos de ligamiento por producto 45

de ligamiento final (en este documento, siete). Así, para la T4 ADN ligasa, la fidelidad del ligamiento es del 79,5% y para el SFL la fidelidad del ligamiento es del 83%.

Reivindicaciones

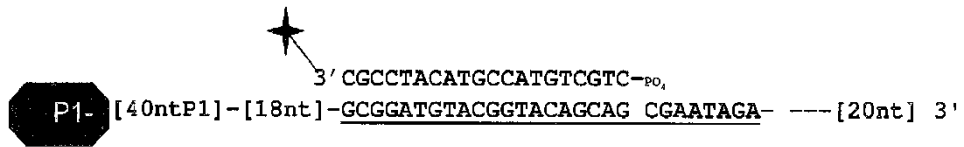
1. Un método de ligamiento repetitivo, que comprende ligar enzimáticamente un oligonucleótido ("sonda") y un polinucleótido ("cebador") usando una ADN ligasa del Virus Chlorella (CV ligasa) o fragmento funcional de la misma, en la que: la sonda contiene N restos nucleotídicos en longitud, donde N es de 2 a 7.
- 5 2. Un kit que comprende:
 - a. una ADN ligasa del virus Chlorella (CV ligasa) o fragmento funcional de la misma;
 - b. un cebador de polinucleótido monocatenario; y
 - c. una sonda de oligonucleótidos en la que la sonda tiene N residuos de nucleótidos de longitud, y N es de 2 a 7.
3. El método de la reivindicación 1 o el kit de la reivindicación 2, en el que N es 2 ó 3.
- 10 4. El método o kit de cualquiera de las reivindicaciones precedentes, en el que la sonda y el cebador son monocatenarios.
5. El método de una cualquiera de las reivindicaciones precedentes, en el que la sonda y el cebador no se hibridan con el mismo polinucleótido ("plantilla") si son polinucleótidos diferentes.
6. El método de la reivindicación 1, en el que el ligamiento repetitivo se usa para la secuenciación de ligamiento.
- 15 7. El método de una cualquiera de las reivindicaciones 1 ó 3, que comprende ligar un extremo proximal de la sonda a un extremo proximal del cebador, en el que la sonda y el cebador están hibridados a un polinucleótido ("plantilla") de modo que sus extremos proximales están de modo adyacente hibridados entre sí.
8. Un método de ligamiento dependiente de plantilla que comprende ligar enzimáticamente un oligonucleótido ("sonda") y un polinucleótido ("cebador") usando una ADN ligasa de Virus de Chlorella (CV ligasa) o fragmento funcional de la misma, en la que:
 - a) la sonda y el cebador se hibridan ambos a la plantilla de manera que sus extremos proximales son adyacentes entre sí;
 - b) la sonda tiene una longitud N, y comprende una parte proximal, donde la porción proximal está (i) perfectamente hibridada con la plantilla y tiene (ii) L nucleótidos de longitud, donde el nucleótido L + 1º de la sonda no coincide con la plantilla; y
 - 25 c) L es de 2 a 8, y además es menor que 6 si el extremo proximal de la sonda es su extremo 3'.
9. El método de las reivindicaciones 1 u 8, en el que la eficacia del ligamiento es más del 80%, 90% o 95%.
10. El método de cualquiera de las reivindicaciones precedentes, en el que la eficacia de ligamiento es al menos 5% más alta que la T4 ligasa.
- 30 11. El método o kit de una cualquiera de las reivindicaciones precedentes, en el que el cebador no tiene más de 6 nucleótidos de longitud, por ejemplo, no más de 5, 4, 3 ó 2 nucleótidos y/o

en donde la plantilla, si está presente, no tiene más de 11 nucleótidos de longitud, por ejemplo, no más de 10, 9, 8, 7, 6, 5, 4, 3 ó 2 nucleótidos.
- 35 12. El método de una cualquiera de las reivindicaciones precedentes, en el que el extremo proximal de la sonda es su extremo 3' y el extremo proximal del cebador es su extremo 5' o en donde el extremo proximal de la sonda es su extremo 5' y el extremo proximal del cebador es su extremo 3'.
13. El método de la reivindicación 1 o el kit de la reivindicación 2, en el que N tiene al menos 3 nucleótidos de longitud y no más de 6 ó 5 nucleótidos de longitud.
- 40 14. El método o kit de una cualquiera de las reivindicaciones precedentes, en el que la sonda está marcada y/o en la que el cebador o plantilla está marcado.
15. El método o kit de cualquiera de las reivindicaciones precedentes, en donde al menos uno de la sonda, la plantilla y/o el cebador están inmovilizados.
16. El método de la reivindicación 8, en el que el ligamiento se repite al menos una vez, opcionalmente en donde el método comprende una reacción en cadena de ligasa.
- 45 17. El método de la reivindicación 1 o la reivindicación 16, en el que se llevan a cabo ciclos sucesivos de ligamiento, en donde el producto de un ciclo anterior se usa como cebador, sonda y/o plantilla en un ciclo sucesivo.

18. El método de la reivindicación 17 o el kit de la reivindicación 2, en el que la sonda tiene un extremo distal no ligable, y luego se escinde después del ligamiento, para proporcionar un dúplex prolongado ligable para el siguiente ciclo.

FIGURA 1

(A)



(B)

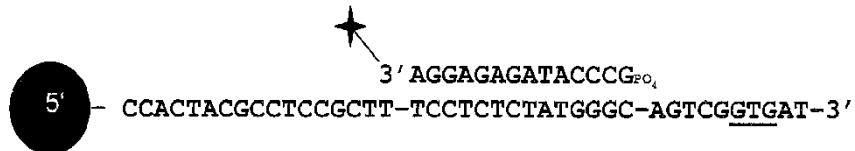


FIGURA 2

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

MAITKPLLAATLENIEDVQFPCLATPKIDGIRSVKQTQMLSRTEFKPIRNSVMNRLLELL
 PEGSDGEISIEGATFQDTTSAVMTGHKMYNAKFSYYWFDYVTD DPLKKYIDRVEDMK
 NYITVHPHILEHAQVKIIPVEINNITELLQYERDVLSKGFEGVMIRKPDGKYKFRST
 LKEGILLKMKQFKDAEATIISMALFKNTNTKTKDNFGYSKRSTHKS GKVEEDVMGSIE
 VDYDGVVFSIGTGFADQRRDFWQNKESYIGKMKVFKYFEMGSKDCPRFPVFIGIRHE
 EDR

MnM ADN Ligasa. GenBank ID YP_333052.1. de Burkholderia pseudomallei 1710b
 (secuencia equivalente a ABA 50091)

MSGVPYGFKNLAATLTKPELIKFPVWASPKIDGIRCVFFGGVAYSRS LKPIP NPVVQEF
 AKAYANLLEGLD GELTVGSPTDANCMQNSMAVMSKAAAPDFTFHVFDWFHPAQAH
 EFWQRSDVVEDRIVQFYDRYPEVDIRAAPQVLCTSLAHLDTNEARWLADGYEGMMIR
 DHCGRYKFRSTEREGGLVKVCRFTDAEAVIGFEEEMHNANEAKRDATGRTERSTSK
 AGLHGKGT LGALVVKNERGIVFNIGTGFTAAQRADYWANHPSLFGKMKVFKHFDHGT
 VDAPRHPVFIGFRHPEDM

Hin ADN Ligasa. GenBank ID P44121, de Haemophilus influenza

MKFYRLLLLFFASSFANSDLM LHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTR
 QGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEISITKSFKGDGWEK LKLYVFDVPDAE
 GNLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQEGGVVVRNPAP
 YERKRSSQILKLKTARGEECTVIAHHKGGKGFENVMGALTCKNHRGEFKIGSGFNLNE
 RENPPPIGSVITYKYRGITNSGKPRFATYWREKK

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

MKFYRLLLLFFASSFANSDLM LHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTR
 QGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEISITKSFKGDGWEK LKLYVFDVPDAEG
 NLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQEGGVVVRNPAPY
 ERKRSSQILKLKTARGEECTVIAHHKGGKGFENVMGALTCKNHRGEFKIGSGFNLNER
 ENPPPIGSVITYKYRGITNSGKPRFATYWREKK

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

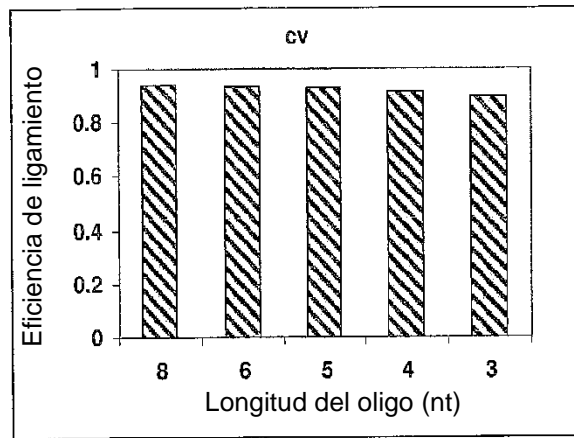
MKFYRLLLLFFASSFANSDLM LHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTR
 QGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEISITKSFKGDGWEK LKLYVFDVPDAEG
 NLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQEGGVVVRNPAPY
 ERKRSSQILKLKTARDEECTVIAHHKGGKGFENVMGALTCKNHRGEFKIGSGFNLNER
 ENPPPIGSVITYKYRGITNSGKPRFATYWREKK

DLXd2 ADN Ligasa (Gammaproteobacteria, Haemophilus influenza) (modificada)

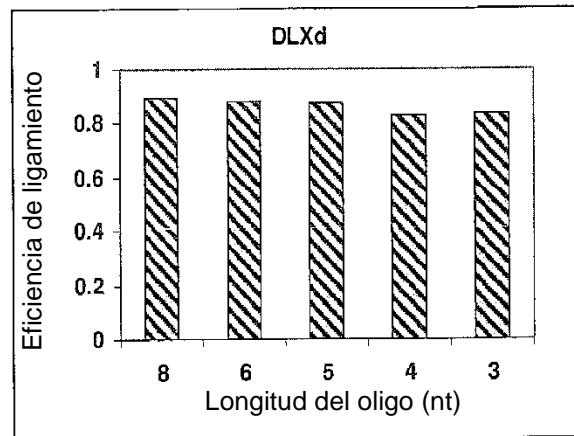
MLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTRQGQRLSPPAYFIKDFPPFAIDGE
 LFSERNHFEEISITKSFKGDGWEK LKLYVFDVPDAEGNLFERLAKLKAHLLEHPTTYIE
 IIEQIPVKDKTHLYQFLAQVENLQEGGVVVRNPAPYERKRSSQILKLKTARDEECTVIA
 HHKGGKGFENVMGALTCKNHRGEFKIGSGFNLNERENPPPIGSVITYKYRGITNSGKPR
 FATYWREKK

FIGURA 3

(A)



(B)



(C)

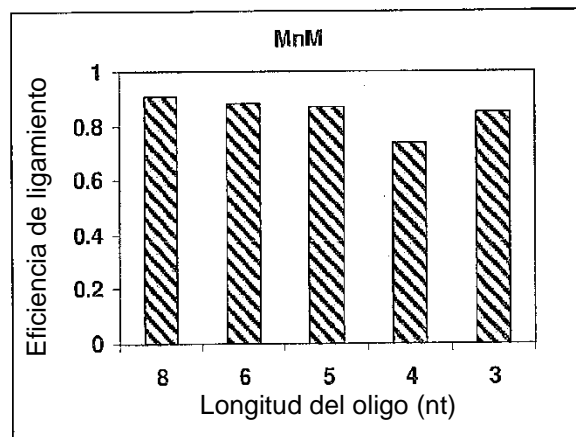
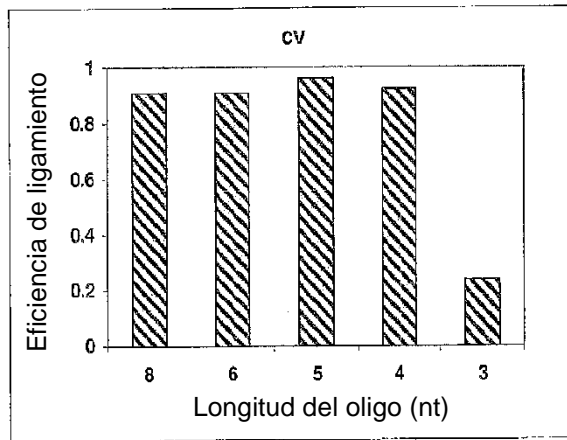
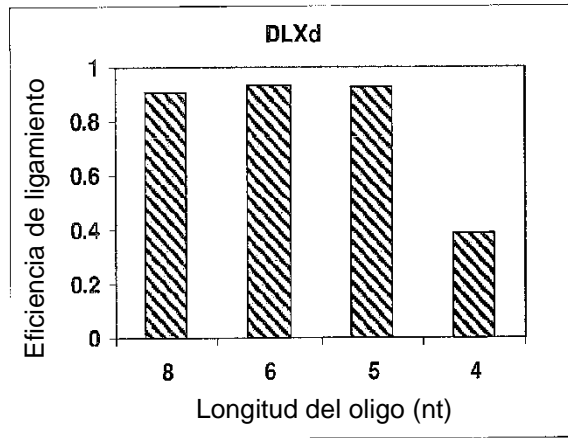


FIGURA 4

(A)



(B)



(C)

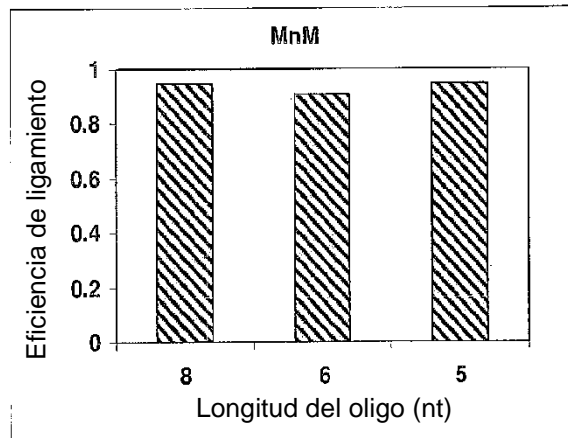


FIGURA 5

