

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 378 790**

51 Int. Cl.:
C12N 15/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **09004736 .6**
- 96 Fecha de presentación: **31.03.2009**
- 97 Número de publicación de la solicitud: **2110435**
- 97 Fecha de publicación de la solicitud: **21.10.2009**

54 Título: **Procedimiento para la preparación de una genoteca de ácido nucleico**

30 Prioridad:
31.03.2008 EP 08006472

45 Fecha de publicación de la mención BOPI:
17.04.2012

45 Fecha de la publicación del folleto de la patente:
17.04.2012

73 Titular/es:
**SLONING BIOTECHNOLOGY GMBH
ZEPPELINSTRASSE 4
82178 PUCHHEIM, DE**

72 Inventor/es:
**Van den Brulle, Jan;
Fuhrmann, Markus y
Strohner, Ralf**

74 Agente/Representante:
Carpintero López, Mario

ES 2 378 790 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

Procedimiento para la preparación de una genoteca de ácido nucleico

La presente invención se refiere a un procedimiento para preparar una genoteca de ácido nucleico, que comprende una pluralidad de elementos y una genoteca de ácido nucleico obtenible de este modo.

5 La generación *de novo* de moléculas de ácido nucleico se está usando cada vez más en la investigación biofarmacéutica para sustituir los procedimientos de clonación, a menudo bastante complejos, necesarios para producir construcciones de ADN deseadas con propiedades optimizadas, por ejemplo, un alto nivel de expresión de proteína en sistemas *in vivo* o *in vitro* adecuados. Existen una diversidad de procedimientos conocidos para sintetizar dichas moléculas de ADN. Prácticamente, todos estos procedimientos dependen de la síntesis, apareamiento y posterior ligación de oligonucleótidos monocatenarios sintéticos para ensamblar moléculas de ADN bicatenario de mayor tamaño, que consisten típicamente en de más de cien a varios miles de pares de bases. Sin embargo, la eficacia de estos procedimientos está limitada por varios factores: (i) la calidad de los oligonucleótidos usados, (ii) el tamaño de la construcción deseada y (iii) la proporción de secuencias "difíciles", por ejemplo, aquellas con regiones auto-complementarias, alto contenido de GC, tétradas de G, pliegues de ADN o bloques de secuencia repetitiva. Los propios componentes básicos de oligonucleótidos están contaminados con diversos productos de terminación y deleciones internas. Son especialmente problemáticos los productos n-1 (oligonucleótidos que contienen deleciones internas de un nucleótido que se producen como resultado de reacciones de protección terminal incompletas), que difícilmente pueden separarse del oligonucleótido de longitud completa deseado. Como muchos oligonucleótidos tienen que ensamblarse para generar un gen completo, la probabilidad de obtener un clon sin errores, es decir, que no incorpore ni siquiera un oligonucleótido defectuoso con un cambio de base o una deleción interna, se aproxima al 0%. Por ejemplo, si un gen se ensamblase a partir de cincuenta oligonucleótidos teniendo cada uno una pureza del 90%, la probabilidad de crear un producto sin errores sería de aproximadamente $0,9^{50} = 0,005$. Generalmente, deben emplearse procedimientos de corrección de errores pesados para obtener una construcción sin errores del 100%. En muchos casos, los productos de síntesis defectuosos no pueden tolerarse porque errores en la secuencia codificante pueden causar la generación de productos de transcripción o traducción acortados debido, por ejemplo, a un desplazamiento de la fase de lectura abierta. Mientras que los primeros dos problemas pueden aliviarse mediante el uso de oligonucleótidos de muy alta pureza, la formación de estructuras secundarias no deseadas que pueden causar deleciones en el producto de síntesis puede en muchos casos suprimirse únicamente si se permiten alteraciones en la secuencia de ADN.

30 En la técnica anterior se conocen una diversidad de procedimientos para producir ADN sintético. Casi 30 años atrás, el trabajo pionero de Khorana y colaboradores (Sekiya T, Brown EL, Belagaje R, Fritz HJ, Gait MJ, Lees RG, Ryan MJ, Khorana HG, Norris KE. (1979), J Biol Chem. 254 (13): 5781-6, y 5787-801) demostró la síntesis *de novo* completa de un gen de ARNt supresor por ligación de parejas de oligonucleótidos apareados. En este procedimiento y procedimientos relacionados, oligonucleótidos monocatenarios complementarios que comprenden la secuencia de ADN deseada completa se aparean en parejas para dar fragmentos bicatenarios, que se alinean en el orden correcto en virtud de salientes monocatenarios complementarios (Stabinsky, Patente de Estados Unidos 4.652.639). Los fragmentos resultantes se ligan después de forma secuencial o en una reacción de un tubo (Jayaraman, Patente de Estados Unidos 5.132.215) enzimática o químicamente. Después de la purificación y/o clonación estos fragmentos génicos pueden unirse entre sí para formar construcciones de ADN de mayor tamaño. En la denominada "síntesis de casete", cada pareja de oligonucleótidos apareados se clona por separado en un vector plasmídico antes de unir los fragmentos usando endonucleasas de restricción (Richards y col., Patente de Estados Unidos 5.093.251).

45 Como alternativa, pueden ensamblarse construcciones de ADN a partir de oligonucleótidos parcialmente apareados, que después de la hibridación contienen huecos monocatenarios que deben rellenarse por ADN polimerasas; este procedimiento se denomina comúnmente procedimiento de "rellenado de huecos". De acuerdo con este procedimiento, una diversidad de oligonucleótidos parcialmente solapantes se sintetizan, se purifican y posteriormente se hibridan, habitualmente en parejas o en subgrupos. Después de la síntesis de las cadenas opuestas respectivas usando una ADN polimerasa, los fragmentos individuales se ligan entre sí. Los productos de ligación bicatenarios generados de esta forma pueden clonarse como fragmentos parciales o amplificarse en una reacción en cadena de la polimerasa (PCR) con cebadores oligonucleotídicos terminales. Sin embargo, este procedimiento está plagado de frecuentes acontecimientos de cebado erróneo y deleciones internas debido a la formación de estructuras secundarias.

55 Ambos procedimientos son de un uso limitado ya que al aumentar la longitud de la molécula de ácido nucleico a sintetizar aumenta la probabilidad de que uno o varios oligonucleótidos con una secuencia incorrecta se incorporen en el producto final. Dichos errores se copian después en la reacción de ADN polimerasa. Además, también pueden introducirse errores de secuencia durante la reacción de PCR.

60 Se describe una combinación de los procedimientos anteriores en la Patente de Estados Unidos 6.472.184, en la que una serie de oligonucleótidos enlazables que representan regiones contiguas en una cadena de la secuencia diana se hibridan con oligonucleótidos no enlazables que son complementarios a los extremos 3' o 5' de los oligonucleótidos enlazables que se van a conectar. Este procedimiento es relativamente simple y sencillo pero

también está plagado de los problemas comunes compartidos por todos los procedimientos que usan oligonucleótidos monocatenarios como componentes básicos: la formación de estructuras secundarias no deseadas y la incorporación de oligonucleótidos n-x, conduciendo ambas a deleciones internas.

5 Aparte de estos procedimientos convencionales, existen procedimientos adicionales conocidos en la técnica para la producción de moléculas de ADN sintéticas. La Solicitud de Patente Internacional WO 98/15567 y la Patente de Estados Unidos 6.110.668 muestran un procedimiento dirigido por molde de acoplamiento de oligonucleótidos para dar construcciones de ADN sintéticas por ligación de una pluralidad de oligonucleótidos que son al menos parcialmente complementarios al ADN de molde monocatenario, y los extremos de dichos oligonucleótidos se ligan en el orden correcto en etapas de apareamiento y desnaturalización sucesivas. Sin embargo, una condición previa para la aplicación de este procedimiento es la existencia previa de un ADN de molde adecuado, excluyendo su uso en la síntesis *de novo*.

15 La Solicitud de Patente Internacional WO 99/47536 desvela un procedimiento de síntesis génica en fase sólida en el que oligonucleótidos monocatenarios se ligan secuencialmente a una molécula iniciadora inmovilizada en una orientación definida. Una desventaja de este procedimiento es que son necesarias muchas etapas para sintetizar genes de mayor tamaño, dando como resultado un rendimiento reducido y un enriquecimiento de secuencias defectuosas. Además, este procedimiento es difícil de automatizar, lo cual es un requisito indispensable para una síntesis rápida normalizada.

20 La Solicitud de Patente Internacional WO 00/75368 desvela una síntesis en fase sólida combinatoria de ácidos nucleicos usando una genoteca de oligonucleótidos bicatenarios como componentes básicos normalizados. El uso de componentes básicos normalizados hace que sea innecesario sintetizar un nuevo conjunto de oligonucleótidos para cada nueva síntesis. Estos oligonucleótidos de genoteca bicatenarios comparten generalmente una estructura global idéntica y, por lo tanto, evitan los problemas de síntesis comunes causados por la formación de estructuras secundarias alternativas de los componentes básicos de oligonucleótidos, tales como la introducción de deleciones. En una versión preferida, contienen un bucle terminal, un tallo bicatenario y un saliente monocatenario corto. Hay dos clases diferentes de oligonucleótidos de genoteca, que están caracterizados por la presencia de diferentes sitios de reconocimiento para enzimas de restricción de tipo IIS dentro de su secuencia, y la presencia o ausencia o el tipo de una modificación interna. Los nucleótidos en el saliente y la región directamente adyacente forman la porción variable que contribuye realmente al ácido nucleico a sintetizar; la secuencia restante es generalmente idéntica en todos los oligonucleótidos que pertenecen a la misma clase.

30 Para construir un ácido nucleico bicatenario, su secuencia se descompone primero en fragmentos de menor tamaño (habitualmente de entre 6 y 30 pares de bases cada uno). Estos denominados bloques de alargamiento se sintetizan después en reacciones en paralelo. En una reacción de este tipo, dos oligonucleótidos de genoteca bicatenarios, uno de cada clase, se ligan por emparejamiento de salientes monocatenarios. Los productos de ligación de los mismos se escinden posteriormente con la enzima de restricción de tipo IIS, que es específica para el oligonucleótido que dona los nucleótidos. El efecto neto de dicho ciclo de ligación/restricción es la adición de un pequeño número de pares de bases (típicamente de entre uno y cinco) al oligonucleótido de partida. Este procedimiento se repite después hasta que se completa la síntesis del bloque de alargamiento deseado.

40 En una segunda fase de reacción, la denominada transposición, esos bloques de alargamiento que son adyacentes en el ácido nucleico a sintetizar se ligan de una forma por parejas después de que cada bloque se haya escindido con una enzima de restricción de tipo IIS diferente. Repitiendo este procedimiento varias veces la longitud de los intermedios de transposición se duplica en cada etapa, mientras que el número de reacciones se reduce a la mitad. Por lo tanto, una molécula de ácido nucleico definida puede generarse en muy pocos ciclos. La ventaja de este procedimiento reside en el ensamblaje combinatorio por parejas de los fragmentos de la molécula de ácido nucleico a sintetizar, de una forma independiente de secuencia. De este modo puede generarse cualquier bloque de alargamiento deseado a partir de una genoteca de ácido nucleico normalizada con un número definido de elementos.

El número de los elementos de dicha genoteca depende de la longitud de los salientes generados por la enzima de restricción de tipo IIS individual, así como del número de nucleótidos que se añaden a los oligonucleótidos en crecimiento en cada ciclo de alargamiento.

50 Este procedimiento ofrece varias ventajas: puede automatizarse completamente puesto que no existe la necesidad de sintetizar y purificar nuevos oligonucleótidos para construir grandes genes o fragmentos de ADN, los componentes básicos se preparan a gran escala y pueden usarse para ensamblar muchas construcciones diferentes hasta que se agote el suministro, reduciendo de este modo el coste de oligonucleótidos en de uno a dos órdenes de magnitud.

55 En diversos campos tales como la enzimología y la terapia basada en anticuerpos, sin embargo, existe la necesidad de optimizar las proteínas respectivas, es decir, enzimas y anticuerpos, respectivamente. Típicamente, dicha optimización se realiza alterando la secuencia de aminoácidos de dichas proteínas. En el caso de que dichas proteínas sean productos de traducción de una molécula de ácido nucleico, dicha molécula de ácido nucleico codificante puede estar sujeta a dicha optimización en términos de alteración de la secuencia de ácido nucleico. En

la técnica se describen diversas estrategias para manipular la secuencia de aminoácidos de dicha proteína y la secuencia de ácido nucleico, respectivamente, que codifica dicha proteína. Sin embargo, la mutación de la secuencia de tipo silvestre típicamente no es suficiente para proporcionar una proteína optimizada. En su lugar, es necesario un conjunto de secuencias mutadas que proporcione un conjunto de secuencias de aminoácidos mutadas y, por lo tanto, proteínas, que también se denominan muteínas, para identificar la muteína o muteínas que tengan la característica o características deseadas. Esta estrategia más heurística es todavía prometedora y ampliamente perseguida aunque cada vez están disponibles más datos sobre la estructura cristalina de las moléculas diana de dichas proteínas. No obstante, el código de acuerdo con el cual se pliega una secuencia de aminoácidos y la secuencia de aminoácidos necesaria para generar una estructura tridimensional distinta de la proteína, están todavía en una fase temprana.

En la técnica anterior, dichas muteínas se generan por mutagénesis. Aparte de la mutagénesis aleatoria, también se realiza una mutagénesis específica de sitio. Aunque la mutagénesis aleatoria, por ejemplo, por PCR propensa a errores, es un medio potente que proporciona al menos una diversidad de muteínas partiendo de una secuencia de aminoácidos básica, que es típicamente la secuencia de aminoácidos de tipo silvestre, el resultado de la mutagénesis aleatoria es poco predecible. Además el uso de procedimientos de mutagénesis específica de sitio que dirigen un nivel controlado de aleatorización en posiciones específicas no es necesariamente adecuado para proporcionar una diversidad bien definida de dichas muteínas (revisado en Neylon (2004), NAR, 1448-59). Al menos, la generación de dicha diversidad es laboriosa. Indudablemente, la generación de una genoteca de muteínas que tengan mutaciones distintas y bien caracterizadas en comparación con la secuencia de tipo silvestre o de partida es más deseable, puesto que tendrá un impacto directo sobre el esfuerzo necesario para identificar clones con características deseadas en un procedimiento de exploración posterior. Además, es deseable controlar la composición de dicha genoteca en términos de la representación relativa de secuencias individuales. Dependiendo del uso o aplicación deseada, dicha genoteca puede tener una distribución uniforme de las diversas secuencias o una distribución sesgada, es decir, una en la que secuencias individuales están presentes con una frecuencia aumentada o disminuida.

El procedimiento para sintetizar un ácido nucleico como se desvela en la Solicitud de Patente Internacional WO 00/75368 puede, por supuesto, ser útil también en la generación de las diversas secuencias de ácido nucleico que posteriormente se traducen en las muteínas correspondientes. Sin embargo, la genoteca respectiva tiene entonces que generarse por combinación de las secuencias de ácido nucleico o secuencias de aminoácidos individuales. Dicho procedimiento, aunque proporciona las secuencias de ácido nucleico y secuencias de aminoácidos específicas, respectivamente, es costoso, reduce el rendimiento global del procedimiento y requiere bastantes etapas de reacción. Además, cuando se combinan las diversas secuencias de aminoácidos y secuencias de ácido nucleico, respectivamente, de nuevo puede producirse un sesgo no deseado de secuencias individuales.

Por consiguiente, existe la necesidad en la técnica de proporcionar un procedimiento para preparar una genoteca de ácido nucleico que comprenda una pluralidad de elementos, por lo que los elementos difieren en al menos una posición que puede ser un solo nucleótido, pero puede ser, en el caso de un ácido nucleico que codifica una secuencia de aminoácidos, un triplete de nucleótidos que codifican un aminoácido, por lo que el aminoácido así codificado es diferente entre al menos algunos de los elementos. Más específicamente, existe la necesidad en la técnica de proporcionar un procedimiento para preparar dicha clase de genoteca, por lo que la genoteca de ácido nucleico tiene una distribución distinta de los diversos elementos individuales de dicha genoteca de ácido nucleico en términos de frecuencia.

Éste y otros problemas se resuelven por la materia objeto de las reivindicaciones independientes. Las realizaciones preferidas pueden tomarse a partir de las reivindicaciones dependientes.

Más específicamente, en un primer aspecto el problema subyacente a la presente solicitud se resuelve mediante un procedimiento para preparar una genoteca de ácido nucleico que comprende una pluralidad de elementos, por lo que cada elemento de dicha genoteca de ácido nucleico comprende una primera extensión de nucleótidos, una segunda extensión de nucleótidos y una tercera extensión de nucleótidos, por lo que la secuencia de la primera extensión de nucleótidos y de la tercera extensión de nucleótidos es idéntica en cada elemento de la genoteca de ácido nucleico y los elementos de la genoteca de ácido nucleico difieren en la secuencia de la segunda extensión de nucleótidos, por lo que el procedimiento comprende las etapas siguientes:

a) proporcionar un primer oligonucleótido al menos parcialmente bicatenario que tenga una estructura bicatenaria que tenga un saliente monocatenario, por lo que el saliente monocatenario proporciona una secuencia de nucleótidos correspondiente a la secuencia de la primera o tercera extensión de nucleótidos de los elementos de la genoteca de ácido nucleico, por lo que el oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una primera enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento,

b) proporcionar una primera genoteca de oligonucleótidos que comprende varios miembros, por lo que cada miembro es un segundo oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria, por lo que el segundo oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una segunda enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento, y un saliente

monocatenario, por lo que el saliente monocatenario o parte del mismo es el mismo para los miembros de la primera genoteca de oligonucleótidos y por lo que dicho saliente monocatenario es esencialmente complementario al saliente monocatenario del primer oligonucleótido al menos parcialmente bicatenario, por lo que la estructura bicatenaria comprende una extensión de nucleótidos por lo que los miembros de la primera genoteca de oligonucleótidos difieren en la secuencia de dicha extensión de nucleótidos y la secuencia de dicha extensión de nucleótidos de los miembros de la primera genoteca de oligonucleótidos corresponden a la secuencia, o parte de la misma, de la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico,

c) combinar el primer oligonucleótido al menos parcialmente bicatenario y la primera genoteca de oligonucleótidos y ligar una molécula del primer oligonucleótido al menos parcialmente bicatenario con una molécula de cada miembro de la primera genoteca de oligonucleótidos mediante sus salientes monocatenarios, después de lo cual se forma un producto de primera ligación,

d) cortar el producto de primera ligación con dicha segunda enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico del miembro de la primera genoteca de oligonucleótidos, por lo que dicha escisión proporciona una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados y segundos oligonucleótidos al menos parcialmente bicatenarios acortados,

e) eliminar los segundos oligonucleótidos al menos parcialmente bicatenarios acortados,

f) proporcionar una segunda genoteca de oligonucleótidos que comprende varios miembros, por lo que cada miembro es un tercer oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria, por lo que el tercer oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una tercera enzima de restricción de tipo IIS, que corta fuera de su sitio de reconocimiento, y un saliente monocatenario, por lo que el saliente monocatenario o parte del mismo es diferente para los miembros de la segunda genoteca de oligonucleótidos, y por lo que dichos salientes monocatenarios o parte de los mismos del tercer oligonucleótido son esencialmente complementarios a los salientes monocatenarios de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados, y por lo que los miembros de la segunda genoteca de oligonucleótidos comprenden una extensión de nucleótidos en la estructura bicatenaria que es idéntica en todos los miembros y corresponde a la secuencia, o parte de la misma, de la tercera o primera extensión de nucleótidos de los elementos de la genoteca de ácido nucleico,

g) combinar la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados y la segunda genoteca de oligonucleótidos y ligar cada molécula de la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados con una molécula de cada uno de los miembros de la segunda genoteca de oligonucleótidos, por lo que el saliente monocatenario del primero oligonucleótido al menos parcialmente bicatenario alargado es esencialmente complementario al saliente del tercer oligonucleótido al menos parcialmente bicatenarios, después de lo cual se forma un producto de segunda ligación,

h) cortar el producto de segunda ligación con la tercera de enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico de los terceros oligonucleótidos al menos parcialmente bicatenarios, proporcionando una pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados y terceros oligonucleótidos al menos parcialmente bicatenarios acortados,

i) eliminar los terceros oligonucleótidos al menos parcialmente bicatenarios acortados,

j) proporcionar un cuarto oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria que tiene un saliente monocatenario, por lo que el saliente monocatenario proporciona una secuencia de nucleótidos esencialmente complementaria al saliente monocatenario del segundo oligonucleótido al menos parcialmente bicatenario alargado, por lo que el oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una cuarta enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento,

k) combinar el cuarto oligonucleótido al menos parcialmente bicatenario y la pluralidad de los segundos oligonucleótidos al menos parcialmente bicatenarios alargados y ligar una molécula del cuarto oligonucleótido y cada molécula de la pluralidad de los segundos oligonucleótidos al menos parcialmente bicatenarios alargados formando un producto de tercera ligación;

l) cortar el producto de tercera ligación con la cuarta enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico del cuarto o cuartos oligonucleótidos al menos parcialmente bicatenarios, proporcionando una pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados y cuarto oligonucleótido al menos parcialmente bicatenario acortado, y

m) eliminar el cuarto oligonucleótido al menos parcialmente bicatenario acortado.

En una realización del primer aspecto el procedimiento comprende además las etapas de:

n) repetir las etapas j) a m) como etapas ja) a ma), por lo que el cuarto oligonucleótido al menos parcialmente

bicatenario de la etapa j) es un oligonucleótido al menos parcialmente bicatenario adicional en la etapa ja), el producto de tercera ligación en la etapa k) es un producto de ligación adicional en la etapa ka), la pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados de la etapa l) es una pluralidad de oligonucleótidos al menos parcialmente bicatenarios alargados adicionales en la etapa la), y el cuarto oligonucleótido al menos parcialmente bicatenario acortado de la etapa l) es un oligonucleótido al menos parcialmente bicatenario acortado adicional en la etapa la).

En una realización del primer aspecto después de la etapa de ligación ka), el producto de ligación adicional se escinde, por lo que la escisión se produce en la secuencia de ácido nucleico del oligonucleótido al menos parcialmente bicatenario alargado tradicional.

En un segundo aspecto, el problema subyacente a la presente solicitud se resuelve mediante un procedimiento para preparar una genoteca de ácido nucleico que comprende una pluralidad de elementos, por lo que cada elemento de dicha genoteca de ácido nucleico comprende una primera extensión de nucleótidos, una segunda extensión de nucleótidos y una tercera extensión de nucleótidos, por lo que la secuencia de la primera extensión de nucleótidos y de la tercera extensión de nucleótidos es idéntica en cada elemento de la genoteca de ácido nucleico y los elementos de la genoteca de ácido nucleico difieren en la secuencia de la segunda extensión de nucleótidos, por lo que el procedimiento comprende las etapas siguientes:

a) proporcionar un primer oligonucleótido al menos parcialmente bicatenario que tenga una estructura bicatenaria que tenga un saliente monocatenario, por lo que el saliente monocatenario proporciona una secuencia de nucleótidos correspondiente a la secuencia de la primera o tercera extensión de nucleótidos de los elementos de la genoteca de ácido nucleico, por lo que el oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una primera enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento,

b) proporcionar una primera genoteca de oligonucleótidos que comprende varios miembros, por lo que cada miembro es un segundo oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria, por lo que el oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una segunda enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento, y un saliente monocatenario, por lo que el saliente monocatenario o parte del mismo es el mismo para los miembros de la primera genoteca de oligonucleótidos, y por lo que dicho saliente monocatenario es esencialmente complementario al saliente monocatenario del primer oligonucleótido al menos parcialmente bicatenario, por lo que la estructura bicatenaria comprende una extensión de nucleótidos por lo que los miembros de la primera genoteca de oligonucleótidos difieren en la secuencia de dicha extensión de nucleótidos y la secuencia de dicha extensión de nucleótidos de los miembros de la primera genoteca de oligonucleótidos corresponde a la secuencia, o parte de la misma, de la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico,

c) combinar el primer oligonucleótido al menos parcialmente bicatenario y la primera genoteca de oligonucleótidos y ligar una molécula del primer oligonucleótido al menos parcialmente bicatenario con una molécula de cada miembro de la primera genoteca de oligonucleótidos mediante sus salientes monocatenarios, después de lo cual se forma un producto de primera ligación,

d) cortar el producto de primera ligación con dicha segunda enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico del miembro de la primera genoteca de oligonucleótidos, por lo que dicha escisión proporciona una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados y segundos oligonucleótidos al menos parcialmente bicatenarios acortados, por lo que dicha pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados consiste en diversas especies moleculares que difieren en la secuencia del saliente monocatenario,

e) eliminar los segundos oligonucleótidos al menos parcialmente bicatenarios acortados,

f) proporcionar una segunda genoteca de oligonucleótidos que comprende varios subgrupos de oligonucleótidos y comprendiendo cada subgrupo varios miembros, por lo que cada miembro es un tercer oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria, por lo que el tercer oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una tercera enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento y un saliente monocatenario, por lo que el saliente monocatenario o parte del mismo es diferente para los diversos subgrupos de oligonucleótidos de la segunda genoteca de oligonucleótidos y por lo que el saliente monocatenario, o parte del mismo, de cada subgrupo de la segunda genoteca de oligonucleótidos es esencialmente complementario al saliente monocatenario de cada especie molecular de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados, y por lo que los miembros de la segunda genoteca de oligonucleótidos comprenden una extensión de nucleótidos en la estructura bicatenaria que es diferente en los miembros de cada subgrupo de la segunda genoteca de oligonucleótidos y corresponde a la secuencia, o parte de la misma, de la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico,

g) combinar la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados y la

segunda genoteca de oligonucleótidos y ligar cada molécula de la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados con cada molécula de los miembros de la segunda genoteca de oligonucleótidos, por lo que se ligan aquellas moléculas cuyos salientes son esencialmente complementarios entre sí, después de lo cual se forman productos de segunda ligación,

5 h) cortar los productos de segunda ligación con la tercera enzima de restricción de tipo II S, por lo que la escisión se produce en la secuencia de ácido nucleico del tercer oligonucleótido al menos parcialmente bicatenario, proporcionando una pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados y terceros oligonucleótidos al menos parcialmente bicatenarios acortados, por lo que dicha pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados consiste en diversos subgrupos
10 que difieren en sus salientes monocatenarios,

i) eliminar los terceros oligonucleótidos al menos parcialmente bicatenarios acortados,

15 j) proporcionar una tercera genoteca de oligonucleótidos que comprende varios miembros, por lo que cada miembro es un cuarto oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria que tiene un saliente monocatenario, por lo que el saliente monocatenario proporciona una secuencia de nucleótidos esencialmente complementaria a los salientes monocatenarios de los segundos oligonucleótidos al menos parcialmente bicatenarios alargados y una extensión, preferentemente adyacente a dicho saliente, que es idéntica en los diversos miembros y proporciona la tercera o primera extensión de nucleótidos de los elementos de la genoteca de ácido nucleico, por lo que los miembros de la tercera genoteca de oligonucleótidos difieren en sus salientes monocatenarios y los salientes corresponden a la segunda extensión, o parte de la
20 misma, de la genoteca de ácido nucleico, por lo que los oligonucleótidos comprenden un sitio de reconocimiento, o parte del mismo, para una cuarta enzima de restricción de tipo IIS, que corta fuera de su sitio de reconocimiento,

25 k) combinar la pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados y la tercera genoteca de oligonucleótidos y ligar cada molécula de los miembros de la tercera genoteca de oligonucleótidos con cada molécula de la pluralidad de los segundos oligonucleótidos al menos parcialmente bicatenarios alargados, por lo que se ligan aquellas moléculas cuyo salientes son esencialmente complementarios entre sí, después de lo cual se forma un producto de tercera ligación;

30 l) cortar el producto de tercera ligación con la cuarta enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico del miembro de la tercera genoteca de oligonucleótidos, proporcionando una pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados y cuartos oligonucleótidos al menos parcialmente bicatenarios acortados,

m) opcionalmente eliminar los cuartos oligonucleótidos al menos parcialmente bicatenarios acortados,

35 n) proporcionar un quinto oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria que tiene un saliente monocatenario, por lo que el saliente monocatenario proporciona una secuencia de nucleótidos esencialmente complementaria al saliente monocatenario de la pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados, por lo que el quinto oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una quinta enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento,

40 o) combinar la pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados y la cuarta genoteca de oligonucleótidos y ligar cada molécula de la pluralidad del tercer oligonucleótido al menos parcialmente bicatenario alargado con una molécula del quinto oligonucleótido al menos bicatenario, después de lo cual se forma un producto de cuarta ligación;

45 p) cortar el producto de cuarta ligación con la quinta enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico del quinto oligonucleótido al menos parcialmente bicatenario, proporcionando una pluralidad de cuartos oligonucleótidos al menos parcialmente bicatenarios alargados y el quinto oligonucleótido al menos parcialmente bicatenario acortado,

q) opcionalmente eliminar el quinto oligonucleótido al menos parcialmente bicatenario acortado.

En una realización del segundo aspecto, el procedimiento comprende además las etapas de:

50 r) repetir las etapas n) a q) como etapas na) a qa), por lo que el quinto oligonucleótido al menos parcialmente bicatenario de la etapa n) es un oligonucleótido al menos bicatenario adicional en la etapa na), el producto de cuarta ligación en la etapa o) es un producto de ligación adicional en la etapa oa), la pluralidad de cuartos oligonucleótidos al menos parcialmente bicatenarios alargados de la etapa p) es una pluralidad de oligonucleótidos al menos parcialmente bicatenarios alargados adicionales en la etapa pa) y el quinto oligonucleótido al menos parcialmente bicatenario acortado de la etapa p) es un oligonucleótido al menos
55 parcialmente bicatenario acortado adicional en la etapa pa).

- En una realización del segundo aspecto después de la etapa de ligación oa), el producto de ligación adicional se escinde por lo que la escisión se produce en la secuencia de ácido nucleico del oligonucleótido al menos parcialmente bicatenario alargado adicional.
- 5 En una realización del primer y del segundo aspecto, la primera enzima de restricción de tipo IIS se selecciona del grupo que comprende Bpil, BbsI, Esp3I, BsmBI, Eco31I, BsaI, BfuAI, FokI, BseRI, BbvI y BsgI.
- En una realización del primer y del segundo aspecto, la segunda, tercera, cuarta, quinta y cualquier enzima de restricción de tipo IIS adicional se selecciona cada una e individualmente del grupo que consiste Earl, Eam1104I, SapI, Lgul, Ksp632I y BspQI.
- 10 En una realización del primer y del segundo aspecto, el ácido nucleico es un ácido nucleico monocatenario y la longitud de la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico consiste en dos nucleótidos, tres nucleótidos, cuatro nucleótidos, cinco nucleótidos, seis nucleótidos, siete nucleótidos, ocho nucleótidos o múltiplos de tres nucleótidos.
- 15 En una realización del primer y del segundo aspecto, en la que el ácido nucleico es un ácido nucleico bicatenario y la longitud de la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico consiste en dos nucleótidos, tres nucleótidos, cuatro nucleótidos, cinco nucleótidos, seis nucleótidos, siete nucleótidos, ocho nucleótidos o múltiplos de tres nucleótidos.
- En una realización del primer y del segundo aspecto, la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico está codificando uno o varios aminoácidos.
- 20 En una realización del primer y del segundo aspecto, el segundo oligonucleótido al menos parcialmente bicatenario, el tercer oligonucleótido al menos parcialmente bicatenario, el cuarto oligonucleótido al menos parcialmente bicatenario, el quinto y el oligonucleótido al menos parcialmente bicatenario adicional comprenden una modificación, por lo que dicha modificación permite la inmovilización del oligonucleótido en una superficie.
- En una realización del primer y del segundo aspecto, la modificación se selecciona del grupo que comprende un resto de biotina, un resto de digoxigenina, un resto de isotiocianato de fluoresceína, un compuesto amino o un éster de succinilo.
- 25 En una realización del primer y del segundo aspecto, la genoteca de ácido nucleico es una genoteca de moléculas de ácido nucleico que codifican un ácido nucleico funcional, por lo que dicho ácido nucleico se selecciona del grupo que comprende aptámeros, promotores, ribozimas y moléculas de mediación de iARN.
- 30 En una realización del primer y del segundo aspecto, la genoteca de ácido nucleico es una genoteca de moléculas de ácido nucleico que codifican un péptido, polipéptido o proteína.
- En una realización del primer y del segundo aspecto, el péptido, polipéptido y proteína se seleccionan del grupo que comprende aptámeros peptídicos, enzimas, enzimas de restricción, dominios de unión a ADN, vacunas, anticuerpos, proteínas farmacéuticamente activas, anticalinas, DARPinas, nanocuerpos y AdNectinas.
- 35 En una realización del primer y del segundo aspecto, la genoteca de ácido nucleico es una genoteca de moléculas de ácido nucleico que codifican un polipéptido, por lo que el polipéptido es un armazón que comprende una o varias regiones constantes, por lo que una o varias regiones variables y una o varias de las regiones variables están codificadas por la secuencia de la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico.
- 40 En una realización del primer y del segundo aspecto, la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico es la única diferencia entre dichos elementos.
- En un tercer aspecto el problema subyacente a la presente solicitud se resuelve mediante una genoteca de ácido nucleico que puede obtenerse mediante un procedimiento de acuerdo con el primer y/o segundo aspecto.
- En una realización del tercer aspecto los miembros de la genoteca difieren en una posición de nucleótido, por lo que dicha posición de nucleótido comprende al menos 1, 2, 3, 4, 5 y 6 nucleótidos.
- 45 En un cuarto aspecto el problema subyacente a la presente solicitud se resuelve mediante un procedimiento para la fabricación de una genoteca de moléculas de ácido nucleico, por lo que la genoteca de moléculas de ácido nucleico comprende una pluralidad de elementos, por lo que los elementos comprenden al menos una extensión constante y al menos una extensión variable, y por lo que la al menos una extensión constante comprende una secuencia de nucleótidos que es la misma en todos los elementos, y la al menos una extensión variable comprende una secuencia de nucleótidos que es diferente en todos los elementos, por lo que el procedimiento comprende las etapas
- 50 siguientes:
- a) proporcionar un primer oligonucleótido, por lo que el primer oligonucleótido es un oligonucleótido al menos parcialmente bicatenario que tiene un saliente monocatenario y que comprende un sitio de reconocimiento para

una primera enzima de restricción de tipo IIS que corta fuera de su secuencia de reconocimiento, o una genoteca de ácido nucleico que comprende una pluralidad de elementos, por lo que la genoteca de ácido nucleico es la genoteca de ácido nucleico de acuerdo con el primer, segundo y/o tercer aspecto y los elementos de la genoteca de ácido nucleico tienen un saliente monocatenario, y los elementos de la genoteca de ácido nucleico comprenden un sitio de reconocimiento para una primera enzima de restricción de tipo IIS que corta fuera de su secuencia de reconocimiento;

b) proporcionar un segundo oligonucleótido, por lo que dicho segundo oligonucleótido es una genoteca de ácido nucleico que comprende una pluralidad de elementos, por lo que la genoteca de ácido nucleico es la genoteca de ácido nucleico de acuerdo con cualquiera del primer, segundo y/o tercer aspecto y los elementos de la genoteca de ácido nucleico tienen un saliente monocatenario que es al menos parcialmente complementario al saliente monocatenario del oligonucleótido proporcionado en la etapa a), los elementos de la genoteca de ácido nucleico comprenden un sitio de reconocimiento para una segunda enzima de restricción de tipo IIS que corta fuera de su secuencia de reconocimiento, por lo que la segunda enzima de restricción de tipo IIS de los elementos de la genoteca de ácido nucleico es diferente de la primera enzima de restricción de tipo IIS del oligonucleótido proporcionado en la etapa a);

c) combinar los oligonucleótidos de la etapa a) y de la etapa b) y ligar una molécula del oligonucleótido de la etapa a) con cada molécula de cada elemento de la genoteca de ácido nucleico de la etapa b) mediante sus salientes monocatenarios con la condición de que el oligonucleótido de la etapa a) sea diferente de la genoteca de ácido nucleico de acuerdo con cualquiera del primer, segundo y/o tercer aspecto; o

combinar los oligonucleótidos de la etapa a) y de la etapa b) y ligar cada molécula de cada elemento de la genoteca de ácido nucleico de la etapa a) con cada molécula de cada elemento de la genoteca de ácido nucleico de la etapa b) mediante sus salientes monocatenarios con la condición de que el oligonucleótido de la etapa a) sea una genoteca de ácido nucleico de acuerdo con cualquiera del primer, segundo y/o tercer aspecto;

d) opcionalmente eliminar reactivos y enzimas sin reaccionar,

e) escindir el producto de ligación de la etapa c) con una enzima de restricción de tipo IIS que corte fuera de su secuencia de reconocimiento, por lo que la escisión se produce en el oligonucleótido de la etapa a) o de la etapa b) proporcionando una genoteca de moléculas de ácido nucleico prolongadas; y

f) opcionalmente separar las moléculas de ácido nucleico prolongadas de la mezcla de reacción.

En una realización del cuarto aspecto las etapas a) a f) se repiten con la condición de que el oligonucleótido de la etapa a) sea un cuarto oligonucleótido que sea una genoteca de moléculas de ácido nucleico prolongadas proporcionada en la etapa e) o un cuarto oligonucleótido al menos parcialmente bicatenario que tiene un saliente monocatenario y que comprende un sitio de reconocimiento para una cuarta enzima de restricción de tipo IIS que corta fuera de su secuencia de reconocimiento;

el oligonucleótido de la etapa b) es un tercer oligonucleótido que es una genoteca de moléculas de ácido nucleico prolongadas proporcionada en la etapa e),

por lo que el tercer y cuarto oligonucleótido comprenden cada uno un saliente monocatenario y por lo que ambos salientes son al menos parcialmente complementarios entre sí;

cuando se combinan los oligonucleótidos de la etapa a) y la etapa b) y se liga una molécula del oligonucleótido de la etapa a) con cada molécula de cada elemento de la genoteca de ácido nucleico de la etapa b) mediante sus salientes monocatenarios, el oligonucleótido de la etapa a) es diferente de la genoteca de moléculas de ácido nucleico prolongadas de la etapa b); o

cuando se combinan los oligonucleótidos de la etapa a) y la etapa b) y se liga cada molécula de cada elemento de la genoteca de ácido nucleico de la etapa a) con cada molécula de cada elemento de la genoteca de ácido nucleico de la etapa b) mediante sus salientes monocatenarios, el oligonucleótido de la etapa a) es la genoteca de moléculas de ácido nucleico prolongadas de la etapa b); y

en la que en la etapa e) se obtiene una genoteca de moléculas de ácido nucleico prolongadas adicional.

En una realización del cuarto aspecto la extensión variable o parte de la misma se proporciona por la segunda extensión de los elementos de la genoteca de ácido nucleico de acuerdo con cualquiera del primer, segundo y/o tercer aspecto. En una realización del cuarto aspecto, la extensión constante o parte de la misma se proporciona mediante la primera extensión y/o la tercera extensión de los elementos de la genoteca de ácido nucleico como se describe en cualquiera del primer, segundo y/o tercer aspecto.

En una realización del primer, segundo y/o tercer aspecto la proporción de los diversos elementos y miembros de las genotecas usados en cualquier procedimiento de este tipo es equimolar.

En una realización del primer, segundo y/o tercer aspecto la proporción de los diversos elementos y miembros de las genotecas usados en cualquier procedimiento de este tipo no es equimolar o está sesgada.

Los presentes inventores han descubierto sorprendentemente que es posible usar el procedimiento descrito en la Solicitud de Patente Internacional WO 00/75368 para preparar una genoteca de ácido nucleico que comprenda una pluralidad de elementos, por lo que los diversos elementos o moléculas de ácido nucleico difieren, de una forma

controlada, en una o varias posiciones de nucleótidos distintas. Preferentemente, la genoteca tiene una composición definida, es decir, los elementos individuales de la genoteca, es decir, las diversas secuencias de ácido nucleico están contenidas en la genoteca con una frecuencia definida preferentemente controlada. Se entenderá que la secuencia de ácido nucleico individual que también se denomina en el presente documento elemento de la genoteca de ácido nucleico, como especies de ácido nucleico, está como tal al menos una vez, preferentemente varias veces presente en la genoteca.

Hasta ahora, los presentes inventores se han alejado del concepto básico subyacente al procedimiento descrito en la Solicitud de Patente Internacional WO 00/75368 que proporciona explícitamente una sola secuencia de ácido nucleico definida y, por lo tanto, una secuencia de aminoácidos definida si dicha secuencia de ácido nucleico es una secuencia de ácido nucleico codificante, siendo preferentemente el caso. Dada la meticulosidad de dicho procedimiento, no era evidente para los presentes inventores usar dicho procedimiento y modificarlo para proporcionar una genoteca de ácido nucleico como se define en el presente documento. Es el mérito de los presentes inventores que han modificado el procedimiento que se describe en la Solicitud de Patente Internacional WO 00/75368 de una forma para preparar una genoteca de ácido nucleico de acuerdo con la presente invención, reflejándose las ventajas intrínsecas al procedimiento del documento WO 00/75368 en el procedimiento de la presente invención en la medida en que las secuencias individuales, es decir, el elemento individual de la genoteca, se proporcionen con la meticulosidad necesaria mientras que al mismo tiempo se proporcione una composición de la genoteca que esté bien definida y realmente controlada en términos de la frecuencia con la que los elementos individuales y, más específicamente, las especies de ácido nucleico individuales, están presentes en dicha genoteca. Hasta ahora, a pesar de proporcionar una diversidad de secuencia que se refleja como una genoteca de ácido nucleico que comprende una pluralidad de elementos, el procedimiento de acuerdo con la presente invención proporciona las ventajas intrínsecas a la Solicitud de Patente Internacional WO 00/75368 en términos de calidad, rendimiento y fiabilidad de las secuencias de ácido nucleico a sintetizar.

El procedimiento de acuerdo con la presente invención se usa ventajosamente para el diseño de proteínas y, más específicamente, la optimización de proteínas por selección de una especie proteica optimizada de una diversidad de muteínas. El campo de uso de dichas genotecas y muteínas, respectivamente, reside en enzimas industriales en el campo de la alimentación, los piensos, los biocombustibles y similares, y en la fabricación de proteínas terapéuticas como anticuerpos, nanocuerpos, armazones tales como anticalinas, DARPinas y AdnECTINAs (revisado en Binz y col. (2005), Nature Biotechnology, 1257-68) y aptámeros peptídicos como, por ejemplo, se describen en Crawford M y col., Briefings in Functional Genomics and Proteomics, Vol. 2, N° 1, 72-79, abril de 2003. Un campo de uso adicional es la optimización de vacunas, la optimización y producción de agentes biofarmacéuticos, la optimización de enzimas de restricción, la fabricación de dominios de unión a ADN específicos y la optimización de rutas metabólicas. Sin embargo, el experto en la materia reconocerá que la aplicación de los procedimientos de acuerdo con la invención no se limita a los mismos.

Aunque es evidente que la presente invención es particularmente ventajosa para crear una genoteca de secuencias de ácido nucleico que codifiquen una diversidad de proteínas, por lo que dicha diversidad de proteínas difieren en una o varias posiciones de aminoácidos en comparación con un material de partida que es preferentemente la secuencia de tipo silvestre, se incluye en la presente invención que la genoteca de ácido nucleico preparada de acuerdo con la presente invención pueda usarse para ácidos nucleicos no codificantes. Preferentemente, dichos ácidos nucleicos no codificantes son aptámeros como, por ejemplo, se describe en la Patente Europea EP 0 533 838, promotores, ribozimas, construcciones de ARN monocatenario y bicatenario e iARN.

Se reconocerá que los productos de reacción del procedimiento de acuerdo con la presente invención son típicamente productos de alargamiento, particularmente productos de alargamiento en el sentido de la Solicitud de Patente Internacional WO 00/75368, que pueden usarse adicionalmente, particularmente usarse en la segunda fase de reacción, es decir, la denominada transposición, del procedimiento de acuerdo con la Solicitud de Patente Internacional WO 00/75368. Preferentemente, el producto de reacción proporcionado por el procedimiento de acuerdo con la presente invención es un elemento de una pareja de moléculas de ácido nucleico que se usa de una forma por parejas y dichas moléculas de ácido nucleico difieren al menos en un saliente monocatenario que permite la ligación de dichas dos moléculas.

Cualquiera de los oligonucleótidos al menos parcialmente bicatenarios usados en relación con los procedimientos de acuerdo con la presente solicitud, si no se indica explícitamente lo contrario en el presente documento, puede estar formado por un solo oligonucleótido, por lo que la extensión bicatenaria del mismo está formada por una parte del oligonucleótido plegada hacia atrás sobre otra parte del oligonucleótido, o puede estar formada por dos o varios oligonucleótidos. En el último caso, es suficiente que la extensión bicatenaria de dicho oligonucleótido se forme, en su mayor parte con independencia de cómo estén implicados muchos oligonucleótidos individuales en la reacción. Sin embargo, se prefiere que dicho oligonucleótido esté formado por dos oligonucleótidos que estén apareados entre sí, preferentemente a través de emparejamiento de bases de Watson-Crick. También dentro del alcance de la presente invención se incluye que dicho oligonucleótido esté formado por un oligonucleótido ramificado. De nuevo, dicho nucleótido ramificado es adecuado para el uso en el procedimiento de acuerdo con el primer aspecto de la presente invención con tal de que cumpla los criterios que se definen en la etapa a) de dicho procedimiento. El oligonucleótido ramificado puede generarse por unión de un oligonucleótido al menos parcialmente bicatenario y un oligonucleótido monocatenario, o a partir de dos o varios oligonucleótidos monocatenarios cuyas bases se

emparejan para formar dicho oligonucleótido al menos parcialmente bicatenario que tiene una protuberancia o saliente monocatenario. El punto de ramificación es un nucleótido, que se localiza preferentemente en la región de bucle, y que tiene un resto químico que permite el acoplamiento químico de un nucleótido u oligonucleótido adicional con el mismo. Dichos restos químicos que permiten la ramificación de un oligonucleótido o ácido nucleico, respectivamente, son conocidos por los expertos en la materia. Son ejemplos de dichos restos químicos engarces amino o carboxi que pueden activarse con EDC (1-etil-3-(3-dimetilaminopropil)carbodiimida) y NHS (N-hidroxisulfosuccinimida) descritos por Hoare y Koshland jr. (Hoare D G, Koshland D E, Jr. (1967), J. Biol. Chem. 242: 2447-2453) y Johnsson y col. (Johnsson B, Lofas S, Lindquist G. (1991), Anal Biochem. 198: 268-77), respectivamente. Dicho primer oligonucleótido de la etapa a) también se denomina en el presente documento "iniciador izquierdo", mientras que el primer oligonucleótido análogo de la etapa b) también se apoda "iniciador derecho".

Acerca de la longitud de los diversos salientes que se describen en el presente documento, dicha longitud se selecciona preferentemente del grupo que comprende un saliente de 1 nucleótido, un saliente de 2 nucleótidos, un saliente de 3 nucleótidos, un saliente de 4 nucleótidos o un saliente de 5 nucleótidos. Como norma general, la longitud del saliente se determina por la enzima o enzimas de restricción usadas en la fase de alargamiento, es decir, la primera etapa del procedimiento para la fabricación de un ácido nucleico. Los nucleótidos del saliente monocatenario más los nucleótidos adyacentes distales del sitio de escisión de la enzima de restricción específica para el segundo oligonucleótido al menos parcialmente bicatenario componen los nucleótidos que se transfieren en cada etapa de alargamiento. El número de nucleótidos transferidos determina a su vez el tamaño de la genoteca de oligonucleótidos necesario para construir cada secuencia posible. Por ejemplo, si se transfieren seis nucleótidos en cada etapa, entonces será necesaria una genoteca de $4^6 = 4096$ oligonucleótidos diferentes para representar todas las permutaciones de hexámeros posibles. Se reconocerá que un saliente particularmente preferido en relación con los procedimientos de la presente invención consiste en 3 nucleótidos, por lo que dichos 3 nucleótidos codifican un aminoácido.

En una realización, la modificación de los oligonucleótidos usados en relación con el procedimiento de acuerdo con la presente invención es un elemento de una pareja de interacción específica tal como biotina y avidina, estreptavidina, extravidina y cualquier mutante o derivado de la misma, incluyendo sitios de unión a biotina artificiales. Las parejas de interacción específicas adicionales incluyen, sin embargo no se limitan a, FITC-anticuerpos anti-FITC (Serke S, Pachmann K (1988), J Immunol Methods. 112(2): 207-11) y digoxigenina-anticuerpos anti-digoxigenina (Kessler C, Holtke HJ, Seibl R, Burg J, Muhlegger K. (1990), Biol Chem Hoppe Seyler. 371(10): 917-27)).

Un experto en la materia también reconocerá que dicha modificación puede ser un nucleótido o secuencia de nucleótidos que permita una inmovilización específica en una superficie.

En relación con las enzimas de restricción de tipo IIS, las siguientes combinaciones son, en principio, adecuadas para llevar a cabo la presente invención.

Un grupo de enzimas de restricción de tipo IIS que se usa preferentemente como primera enzima de restricción de tipo IIS en relación con los procedimientos de la presente invención comprende las enzimas siguientes: Bpil, BbsI, Esp3I, BsmBI, Eco31I, BsaI, BfuAI, FokI, BseRI, BbvI y BsgI.

Otro grupo de enzimas de restricción de tipo IIS que se usa preferentemente como segunda, tercera, cuarta, quinta y una enzima de restricción de tipo IIS adicional en relación con los procedimientos de la presente invención comprende las enzimas siguientes: EarI, Ears1104I, SapI, LglI, Ksp632I y BspQI. En relación con este grupo se prefiere particularmente la enzima de restricción 1104I.

La presente invención se ilustra ahora adicionalmente por referencia a las figuras adjuntas, de las que pueden tomarse características, realizaciones y ventajas adicionales.

La Fig. 1 muestra una ilustración del procedimiento para la preparación de una genoteca de ácido nucleico de acuerdo con el primero aspecto de la presente invención, por lo que los miembros de la genoteca de ácido nucleico difieren en una posición de la secuencia y dicha posición proporciona un aminoácido diferente en la traducción;

la Fig. 2 muestra una ilustración del procedimiento para la preparación de una genoteca de ácido nucleico de acuerdo con el segundo aspecto de la presente invención, por lo que los miembros de la genoteca de ácido nucleico difieren en una posición de la secuencia y dicha posición comprende dos tripletes consecutivos de nucleótidos, codificando cada uno un aminoácido;

la Fig. 3 muestra una ilustración del procedimiento para la preparación de una genoteca de ácido nucleico de acuerdo con el primer aspecto de la presente invención, por lo que los miembros de la genoteca de ácido nucleico difieren en una posición de la secuencia y dicha posición proporciona diferentes aminoácidos tras la traducción, por lo que la posición variable es una posición terminal;

la Fig. 4 muestra una ilustración del procedimiento para la preparación de una genoteca de ácido nucleico de

acuerdo con el segundo aspecto de la presente invención, por lo que los miembros de la genoteca de ácido nucleico difieren en dos oposiciones de la secuencia y dichas dos posiciones proporcionan dos aminoácidos diferentes tras la traducción, por lo que las posiciones variables son dos posiciones terminales consecutivas;

- 5 la Fig. 5 es un histograma que indica las frecuencias de codones obtenidas en la síntesis de una genoteca de ácido nucleico usando el procedimiento de acuerdo con el primer aspecto de la presente solicitud;
- la Fig. 6 es un histograma que indica las frecuencias de codones en dos posiciones consecutivas obtenidas en la síntesis de una genoteca de ácido nucleico usando el procedimiento de acuerdo con el segundo aspecto de la presente solicitud;
- 10 la Fig. 7 es un histograma que indica las frecuencias de codones en una posición variable con una distribución igual de los codones obtenidos en la síntesis de una genoteca de ácido nucleico usando el procedimiento de acuerdo con el primer aspecto de la presente solicitud; y
- la Fig. 8 es un histograma que indica las frecuencias de codones en una posición variable con una distribución desigual de los codones obtenidos en la síntesis de una genoteca de ácido nucleico usando el procedimiento de acuerdo con el primer aspecto de la presente solicitud.
- 15

Un experto en la materia reconocerá que lo siguiente es una descripción de los procedimientos de acuerdo con la presente invención para preparar una genoteca de ácido nucleico que comprende una pluralidad de elementos, por lo que sólo se describen las etapas esenciales en más detalle. Los detalles adicionales en la realización del procedimiento de acuerdo con la presente invención son evidentes para un experto en la materia a la luz de la divulgación de la presente solicitud y los principios de reacción subyacentes que se basan en el procedimiento definido en la Solicitud de Patente Internacional WO 00/75368.

20

Procedimiento para preparar una genoteca de ácido nucleico como se ilustra en la Fig. 1.

El procedimiento de la presente invención se describe en relación con la síntesis de una secuencia diana que se representa en la Fig. 1. Dicha secuencia diana comprende cuatro tripletes, por lo que leyendo de izquierda a derecha, que es típicamente en la dirección 5' -> 3', el segundo triplete se varía tanto a nivel de ácido nucleico como también, y en particular, a nivel de aminoácido. El primer, tercer y cuarto tripletes son los mismos para los diversos elementos de la genoteca de ácido nucleico a preparar al igual que lo son preferentemente los otros nucleótidos no representados específicamente y preferentemente los tripletes y, por lo tanto, las secuencias de aminoácidos a sintetizar de acuerdo con el procedimiento de acuerdo con la presente invención.

25

La genoteca de ácido nucleico de acuerdo con la presente invención comprende una pluralidad de elementos, por lo que cada elemento de dicha genoteca de ácido nucleico comprende una primera extensión de nucleótidos que corresponde al primer triplete, una segunda extensión de nucleótidos que corresponde al segundo triplete y una tercera extensión de nucleótidos que corresponde al tercer y cuarto triplete. La secuencia la primera extensión de nucleótidos y de la tercera extensión de nucleótidos es idéntica en cada elemento de la genoteca de ácido nucleico y los elementos de la genoteca de ácido nucleico difieren en la secuencia de la segunda extensión de nucleótidos. En la secuencia diana, la primera extensión es por lo tanto una extensión que comprende CAC, y la tercera extensión es una extensión que comprende los dos tripletes GAG y ACA. La segunda extensión es, por consiguiente, la que comprende ATG, TAT o GAG. Como se ha resumido anteriormente, los elementos de la genoteca de ácido nucleico difieren al menos con respecto a la segunda extensión. Por consiguiente, en una realización preferida la genoteca de acuerdo con la presente invención comprende tres elementos, comprendiendo el primer elemento la secuencia de CACATGGAGACA, comprendiendo el segundo elemento la secuencia de CACTATGAGACA y comprendiendo el tercer elemento la secuencia de ácido nucleico CACGAGGAGACA. Sin embargo, se entenderá que la segunda extensión puede comprender tripletes alternativos adicionales, que se suman por lo tanto a la complejidad de la genoteca de ácido nucleico.

30

35

40

En la etapa 1, se proporciona un primer oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria. Dicho oligonucleótido al menos parcialmente bicatenario tiene un saliente monocatenario y también se denomina en el presente documento engarce de cebador de secuenciación o esplínquer (*splinker*). El saliente monocatenario proporciona una secuencia de nucleótidos correspondiente a la secuencia de la primera o la tercera extensión de nucleótidos, o parte de las mismas, de los elementos de la genoteca de ácido nucleico. Si el saliente monocatenario corresponde a una parte de la primera o tercera extensión, dicha correspondencia es preferentemente con la parte o partes más terminales de la primera y tercera extensión, respectivamente. Como se usa preferentemente en el presente documento, la expresión de que secuencias son correspondientes entre sí significa que las secuencias son complementarias entre sí, preferentemente basándose en el emparejamiento de bases de Watson-Crick, las secuencias son complementarias entre sí, preferentemente completamente complementarias, es decir, muestran un ajuste perfecto. En una realización preferida, el oligonucleótido comprende un sitio de reconocimiento o parte del mismo para una primera enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento. En el presente caso, dicho primer oligonucleótido al menos parcialmente bicatenario tiene un saliente monocatenario que comprende GTG y, por lo tanto, es complementario al primer triplete y por lo tanto a

45

50

55

la primera extensión de la secuencia diana. El primer oligonucleótido al menos parcialmente bicatenario consiste en la realización representada en la Fig. 1 (1) de un solo oligonucleótido que está formado por una parte de dicho oligonucleótido individual que se pliega hacia atrás hacia otra parte del oligonucleótido, formando de este modo un bucle. Sin embargo, también se incluye en la presente invención que el oligonucleótido al menos parcialmente bicatenario esté formado por dos oligonucleótidos monocatenarios separados cuyas bases se emparejan entre sí, generando de este modo dicho oligonucleótido al menos parcialmente y el saliente monocatenario descrito anteriormente. En una realización preferida, el extremo de dicho último oligonucleótido al menos parcialmente bicatenario que no está proporcionando el saliente monocatenario necesario para la ligación se bloquea para evitar cualquier ligación o reacción que sea potencialmente posible en las etapas posteriores del procedimiento de acuerdo con la presente invención.

En la etapa 2, se proporciona una primera genoteca de oligonucleótidos que comprende varios miembros. Cada miembro de dicha primera genoteca de oligonucleótidos se denomina también en el presente documento segundo oligonucleótido al menos parcialmente bicatenario y tiene una estructura bicatenaria. La estructura bicatenaria, como en relación con el primer oligonucleótido al menos bicatenario, puede estar formada por dos cadenas sencillas que tengan la modificación adicional que se ha descrito anteriormente. Los miembros de la primera genoteca de oligonucleótidos comprenden un sitio de reconocimiento, o parte del mismo, para una segunda enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento. Además, dicho segundo oligonucleótido al menos parcialmente bicatenario proporciona un saliente monocatenario. Los miembros de la primera genoteca de oligonucleótidos tienen el mismo saliente monocatenario y dicho saliente monocatenario es esencialmente o parcialmente complementario al saliente monocatenario del primer oligonucleótido al menos parcialmente bicatenario. En el presente caso, dicho saliente monocatenario es CAC. Posterior al saliente monocatenario, los segundos oligonucleótidos al menos parcialmente bicatenarios comprenden una estructura bicatenaria que comprende una extensión de nucleótidos, por lo que los miembros de la primera genoteca de oligonucleótidos difieren en la secuencia de dicha extensión de nucleótidos y las secuencias de dicha extensión de nucleótidos corresponden a las secuencias, o parte de las mismas, de la segunda extensión de los elementos de la genoteca de ácido nucleico de acuerdo con la presente invención. Por consiguiente, como se representa en la Fig. 1 (1), se proporciona una primera genoteca de oligonucleótidos en la que la parte bicatenaria de los miembros posterior al saliente monocatenario comprende las diversas secuencias sujetas a la segunda extensión de los nucleótidos de los elementos de la genoteca de ácido nucleico. Dentro de la presente invención se incluye que la estructura bicatenaria comprenda nucleótidos adicionales que, en una realización, puedan o no contener nucleótidos adicionales de la secuencia diana. En el uso particular representado en la Fig. 1, estas secuencias son ATG, TAT y GAG que están formando emparejamiento de bases con sus secuencias complementarias, es decir, TAC, ATA y CTC (en dirección 3' -> 5'), que también puede estar representadas en el presente documento como ATG/TAC, TAT/ATA y GAG/CTC. Los miembros de la primera genoteca de oligonucleótidos difieren por lo tanto en la secuencia de dicha extensión de nucleótidos y la secuencia de dicha extensión de nucleótidos de los miembros de la primera genoteca de nucleótidos corresponde a la secuencia o parte de la misma de la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico. Por consiguiente, la complejidad de la genoteca de ácido nucleico a preparar usando el procedimiento de acuerdo con la presente invención, que en el caso de la Fig. 1 es de tres, es decir, consiste en tres elementos diferentes, está reflejada y preferentemente es en realidad la misma que la complejidad de la primera genoteca de oligonucleótidos, que también consiste en tres miembros diferentes.

Además, los miembros de la primera genoteca de oligonucleótidos comprenden una modificación. En el presente caso dicha modificación permite la inmovilización de los diversos miembros en una superficie, típicamente una superficie de un soporte o vehículo. La inmovilización está específicamente mediada por una interacción entre la modificación y un compañero de interacción de dicha modificación. En el presente caso, la modificación es biotina que está interaccionando con estreptavidina como su compañero de interacción y proporciona un complejo estable que permite la inmovilización de los miembros de la primera genoteca de oligonucleótidos y los productos de reacción de cualquier reacción de ligación usando los mismos. Los miembros, debido a la presencia de la modificación, también se denominan en el presente documento anclajes o moléculas de anclaje.

En la Fig. 1, el sitio de reconocimiento para la enzima de restricción de tipo IIS de los miembros de la primera genoteca de oligonucleótidos se indica como un recuadro en negro que ilustra el hecho de que para esta clase de enzima de restricción, el sitio de reconocimiento se encuentra fuera del sitio de escisión.

En la etapa 3, el primer oligonucleótido al menos parcialmente bicatenario y la primera genoteca de oligonucleótidos y los miembros de la misma, respectivamente, se combinan. Tras dicha combinación, una molécula del primer oligonucleótido al menos parcialmente bicatenario se aparea y posteriormente se liga con una molécula de un miembro de la primera genoteca de oligonucleótidos, preferentemente por medio de una ligasa o actividad de ligasa. Dicha ligación se produce tras la formación de un primer producto de preligación a través de la hibridación de los salientes monocatenarios de las dos clases de moléculas, actuando dichos salientes como "extremos cohesivos". El producto de reacción de la ligación se denomina producto de primera ligación. Debido a que la primera genoteca de oligonucleótidos comprende diversos elementos y especies, respectivamente, la reacción de ligación da como resultado una combinación de productos de primera ligación o una pluralidad de productos de primera ligación. Además, en esta pluralidad de productos de ligación que en realidad constituye una genoteca adicional, los miembros difieren al menos y, preferentemente únicamente en el presente ejemplo, en la secuencia correspondiente al segundo triplete y, por lo tanto, la segunda extensión de la secuencia diana, mientras que típicamente el resto del

producto de primera ligación en términos de secuencia de nucleótidos es preferentemente igual para los diversos elementos de la genoteca de productos de primera ligación así formada. La pluralidad de productos de primera ligación comprende en el presente caso tres productos de ligación diferentes según se determina por el número de los segundos oligonucleótidos al menos parcialmente bicatenarios diferentes que forman la primera genoteca de oligonucleótidos. Debe reconocerse que en el procedimiento de acuerdo con el primer aspecto, el producto de primera ligación comprende diversas especies de moléculas que difieren en su secuencia correspondiente a la segunda extensión de la secuencia diana. La diversidad de secuencia de ácido nucleico de dichas especies se proporciona por la primera genoteca de oligonucleótidos.

En la etapa 4, los productos de primera ligación se cortan con una enzima de restricción de tipo IIS. Dicho sitio de reconocimiento de enzima de restricción de tipo IIS es el proporcionado por los segundos oligonucleótidos al menos parcialmente bicatenarios y se denomina en el presente documento segunda enzima de restricción de tipo IIS. El corte de los productos de primera ligación se realiza en la combinación y la genoteca, respectivamente, que consisten en los productos de primera ligación. Por consiguiente, el corte o escisión de los productos de primera ligación da como resultado una pluralidad de moléculas así cortadas, por lo que el producto de reacción de dicho corte es una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados que consisten en diferentes especies de primeros oligonucleótidos al menos parcialmente bicatenarios alargados y una pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios acortados.

Se reconocerá que la realización de la etapa 3 y 4, respectivamente, puede llevarse a cabo en condiciones de reacción diferentes en términos de inmovilización o no inmovilización de los miembros de la primera genoteca de oligonucleótidos que consisten en el producto de ligación de primera genoteca, respectivamente. En cada caso, las moléculas respectivas proporcionan una modificación que, como se ha resumido anteriormente, permite la inmovilización de las moléculas en un soporte que comprende el compañero de interacción de la modificación. Por consiguiente, cualquier secuencia con referencia a una etapa de inmovilización puede, en principio, realizarse en diversas realizaciones. Por consiguiente, dentro de la presente invención se incluye realizar la reacción de ligación sujeta a la etapa 3 en una fase líquida, es decir, no estando los miembros de la primera genoteca de oligonucleótidos inmovilizados. Posteriormente, puede producirse el corte del producto de primera ligación. En una realización, dicho corte se realiza entonces también en la fase líquida, no estando el producto de primera ligación inmovilizado en un soporte mediante su modificación. Después del corte, el segundo oligonucleótido al menos parcialmente bicatenario acortado se elimina de la mezcla de reacción mediante su modificación por unión al soporte. Como alternativa, después de la ligación, el producto de primera ligación puede inmovilizarse en una superficie y el corte del producto de primera ligación se realiza estando el producto de primera ligación inmovilizado en dicha superficie. En cualquier caso, el segundo oligonucleótido al menos parcialmente bicatenario acortado se elimina en última instancia de la reacción por inmovilización del mismo a través de su modificación a un soporte. Se reconocerá que debido a la modificación que permite una interacción específica con su compañero de interacción tal como biotina con estreptavidina, para los fines de la inmovilización la reacción se transfiere a un recipiente de reacción, constituyendo preferentemente las paredes del mismo un soporte que contiene el compañero de interacción de la modificación, en el presente ejemplo dado que la modificación es biotina, estreptavidina. Debido a la alta afinidad de la biotina y la estreptavidina, la unión de la molécula que contiene la modificación al soporte se produce inmediatamente. En otras palabras, si se va a realizar una reacción tal como la reacción de ligación con una molécula que tiene dicha modificación, el recipiente de reacción o cualquier superficie disponible en dicha reacción no debe comprender un compañero de interacción de la modificación, mientras que en el caso de que se desee la inmovilización, se usa un recipiente de reacción cuya superficie proporcione dicho compañero de interacción de la modificación, mediante el recipiente de reacción o mediante otras superficies prevalentes en dicha reacción.

En otras palabras, en una realización del procedimiento de acuerdo con el primer aspecto, tras la formación del producto de ligación es posible cortar el producto de ligación con la enzima de restricción de tipo IIS respectiva, cuya secuencia de reconocimiento se proporciona por la genoteca de oligonucleótidos respectiva o cualquier oligonucleótido que proporcione una modificación que permita la inmovilización del producto de ligación en un soporte, por lo que el corte proporciona una pluralidad de oligonucleótidos al menos parcialmente bicatenarios alargados y oligonucleótido al menos parcialmente bicatenario acortado, por lo que dicho oligonucleótido al menos parcialmente bicatenario acortado se elimina de la mezcla de reacción por inmovilización en dicho soporte mediante la modificación. En una realización alternativa, tras la formación del producto de ligación, dicho producto de ligación se inmoviliza en un soporte mediante la modificación, y el producto de ligación así inmovilizado se corta mediante una enzima de restricción de tipo IIS cuya secuencia de reconocimiento se proporciona por la genoteca de oligonucleótidos respectiva o cualquier oligonucleótido que proporcione una modificación que permita la inmovilización del producto de ligación en un soporte, por lo que el corte proporciona una pluralidad de oligonucleótidos al menos parcialmente bicatenarios alargados y oligonucleótido al menos parcialmente bicatenario acortado, por lo que dicho oligonucleótido al menos parcialmente bicatenario acortado se une o permanece unido al soporte.

La etapa 5 de la reacción consiste en la eliminación de los segundos oligonucleótidos al menos parcialmente bicatenarios acortados que se realiza preferentemente por inmovilización del segundo oligonucleótido al menos parcialmente bicatenario acortado.

Debido a esta secuencia de reacción, la reacción y preferentemente el sobrenadante de la misma contienen una

pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados. En una etapa adicional, que es típicamente la etapa 6, se proporciona una segunda genoteca de oligonucleótidos que comprende varios miembros. Cada miembro de dicha segunda genoteca de oligonucleótidos es un tercer oligonucleótido al menos parcialmente bicatenario y comprende una estructura bicatenaria. La estructura bicatenaria puede diseñarse tal como se resume en el presente documento en relación con otros oligonucleótidos al menos parcialmente bicatenarios, tal como el primer oligonucleótido al menos parcialmente bicatenario. Más específicamente, el tercer oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una tercera enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento y un saliente monocatenario. El saliente monocatenario o parte del mismo es diferente para los miembros de la segunda genoteca de oligonucleótidos. Dichos salientes monocatenarios o parte de los mismos de la segunda genoteca de oligonucleótidos son esencialmente complementarios a los salientes monocatenarios de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados. Un experto en la materia entenderá que debido al mecanismo subyacente al procedimiento de la presente invención, el saliente de una especie de la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados es complementario a un miembro de la segunda genoteca de oligonucleótidos. También se reconocerá que los salientes de las diversas especies de la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados y el saliente de los diversos miembros de la segunda genoteca de oligonucleótidos son, si se toman en su totalidad, complementarios entre sí, aunque no cada saliente de los diversos primeros oligonucleótidos al menos parcialmente bicatenarios alargados será complementario a cada saliente del miembro de la segunda genoteca de oligonucleótidos. La segunda genoteca de oligonucleótidos proporciona por lo tanto salientes monocatenarios de ATG, TAT y GAG, respectivamente. Por consiguiente, en el presente caso, la segunda genoteca de oligonucleótidos comprende tres miembros. Los miembros de la segunda genoteca de oligonucleótidos comprenden además una extensión de nucleótidos en la estructura bicatenaria que es idéntica en todos los miembros y que corresponde a la secuencia o parte de la misma de la tercera o primera extensión de nucleótidos de los elementos de la genoteca de ácido nucleico. Los miembros de la segunda genoteca de oligonucleótidos comprenden una modificación, por lo que dicha modificación es preferentemente la misma que en relación con la primera genoteca de oligonucleótidos.

En la etapa 7, la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados y la segunda genoteca de oligonucleótidos se combinan. Debido a la complementariedad de los salientes, una molécula de una especie de la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados se liga con una molécula de un miembro de la segunda genoteca de oligonucleótidos. La pareja respectiva de moléculas a ligarse está definida por la complementariedad de los salientes que definen una pareja de reacción. Una vez que los dos oligonucleótidos correspondientes, como se definen por sus salientes monocatenarios, se hibridan o aparean, se produce típicamente la ligación. El producto de ligación de dicha reacción de ligación se denomina producto de segunda ligación. En el presente caso, el producto de segunda ligación comprende un total de tres secuencias diferentes y por lo tanto tres especies diferentes, por lo que las secuencias de dichas especies difieren al menos, y típicamente exclusivamente, en la secuencia correspondiente a la segunda extensión de la secuencia diana. Los productos de segunda ligación forman de nuevo una genoteca.

Las etapas siguientes, es decir, la etapa 8 que es cortar el producto de segunda ligación con la tercera enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico del tercer oligonucleótido al menos parcialmente bicatenario, está diseñada en principio de la misma forma que se ha descrito en relación con el producto de primera ligación. Esto es aplicable particularmente con respecto a la secuencia potencial de unión y corte y la eliminación por inmovilización, respectivamente. Como resultado de la etapa 8 en relación con la cual el producto de segunda ligación se corta con la tercera enzima de restricción de tipo IIS, se proporciona una pluralidad de los segundos oligonucleótidos al menos parcialmente bicatenarios alargados que comprende un total de tres especies diferentes de oligonucleótidos, y terceros oligonucleótidos al menos parcialmente bicatenarios acortados. De nuevo, los terceros oligonucleótidos al menos parcialmente bicatenarios acortados así liberados del producto de segunda ligación se eliminan de la reacción por la modificación y su interacción con un compañero de interacción tal como estreptavidina en relación con que sea biotina la modificación unida a una superficie.

En la etapa 9, se proporciona un cuarto oligonucleótido al menos parcialmente bicatenario. El cuarto oligonucleótido bicatenario tiene un saliente monocatenario y, de otro modo, el diseño de este cuarto oligonucleótido al menos parcialmente bicatenario puede ser, en una realización, similar a las realizaciones descritas en realización con otros oligonucleótidos al menos parcialmente bicatenarios descritos en relación con la Fig 1. El saliente monocatenario del cuarto oligonucleótido al menos parcialmente bicatenario proporciona una secuencia de nucleótidos esencialmente complementaria al saliente monocatenario del segundo oligonucleótido al menos parcialmente bicatenario alargado. El saliente del segundo oligonucleótido al menos parcialmente bicatenario alargado es el mismo y por lo tanto idéntico para las tres especies del segundo oligonucleótido al menos parcialmente bicatenario alargado, ya que dicho saliente corresponde al primer triplete de la tercera extensión. Además, el cuarto oligonucleótido al menos parcialmente bicatenario comprende un sitio de reconocimiento, o parte del mismo, para una cuarta enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento. El saliente monocatenario de dicho cuarto oligonucleótido al menos parcialmente bicatenario es, como se representa en la Fig. 1 (2), GAG y por lo tanto idéntico al primer triplete de la tercera extensión. Dicho cuarto oligonucleótido y la pluralidad de los segundos oligonucleótidos al menos parcialmente bicatenarios alargados se combinan y se ligan. De nuevo, una molécula del cuarto oligonucleótido se aparea con una molécula de segundos oligonucleótidos al menos parcialmente

bicatenarios alargados debido a la complementariedad del saliente de ambos oligonucleótidos. Las moléculas así apareadas se ligan después proporcionando un producto de tercera ligación. Como en relación con otras reacciones, debido a la complementariedad de los salientes, se produce una reacción de uno en uno, es decir, una reacción entre una molécula de cada especie de reactivo contenido en la reacción. El producto de tercera ligación comprende en última instancia la secuencia diana que se representa en la Fig. 1 (1). Dicho producto de tercera ligación puede cortarse por la cuarta enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico del cuarto o cuartos oligonucleótidos al menos parcialmente bicatenarios y proporciona una pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados y cuartos oligonucleótidos al menos parcialmente bicatenarios acortados. De nuevo, el diseño de la secuencia de reacciones adicionales puede variar como se ha resumido en relación con el primer oligonucleótido al menos parcialmente bicatenario y la primera genoteca de oligonucleótidos.

Como se entenderá con respecto a las secuencias representadas específicamente en la Fig. 1, la enzima de restricción de tipo IIS usada para la escisión es una proporcionada por las moléculas de anclaje, es decir, las moléculas que proporcionan la modificación que permite la interacción con un compañero de interacción y, por lo tanto, la inmovilización de la molécula incluyendo cualquier resto que pueda haberse añadido por ligación o eliminado por escisión. La enzima de restricción de tipo IIS específica intrínseca a todas las diversas secuencias de anclaje usadas en relación con el ejemplo descrito en la Fig. 1 es Eam1104I.

Dicho tercer oligonucleótido al menos parcialmente bicatenario alargado puede someterse a etapas de reacción adicionales que también se denominan etapas de alargamiento, como se ha resumido anteriormente.

Los expertos en la materia reconocerán que, como alternativa, la secuencia diana así obtenida y más específicamente el tercer oligonucleótido al menos parcialmente bicatenario alargado pueden usarse después en una etapa de transposición como se describe en relación con la Solicitud de Patente Internacional WO 00/75368.

Procedimiento para preparar una genoteca de ácido nucleico como se ilustra en la Fig. 2

En relación con el procedimiento para la preparación de una genoteca de ácido nucleico, por lo que los elementos de la misma contienen una posición variable que consiste en dos tripletes de nucleótidos posteriores, se aplica típicamente la siguiente secuencia de etapas. La secuencia diana como se representa en la Fig. 2 (1) comprende como primera extensión CAC, como segunda extensión que es variable en secuencia y comprende en el presente ejemplo (en dirección 5' → 3') ATGCTC, ATGGAG, ATGTTT, TATGAG, TATCTC, TATTTT, GAGTTT, GAGCTC o GAGGAG una secuencia de ácido nucleico que por lo tanto es diferente para cada elemento de la genoteca de ácido nucleico, pero en cualquier caso una secuencia definida por una secuencia de nucleótidos conocida más que una extensión variable de secuencia arbitraria, y como una tercera extensión ACA.

La segunda extensión variable comprende en el caso ilustrado una primera mitad de la segunda extensión y una segunda mitad de la segunda extensión, por lo que la primera mitad de la segunda extensión presenta uno de los triplete siguientes: ATA, TAT o GAG, y la segunda mitad de la segunda extensión comprende los diferentes tripletes siguientes: CTC, GAG o TTT.

La particularidad de este aspecto del procedimiento de la presente invención es que la primera mitad de la segunda extensión y la segunda mitad de la segunda extensión se proporcionan o incorporan en el ácido nucleico en etapas de reacción diferentes. Se reconocerá que la longitud de la primera mitad de la segunda extensión y de la segunda mitad de la segunda extensión puede variar, dependiendo de las características de corte de las enzimas de restricción de tipo IIS usadas. Se prefiere particularmente si la longitud de cada una de la primera mitad y la segunda mitad consiste en tres nucleótidos consecutivos que codifican un aminoácido.

En una primera etapa se proporciona un primer oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria. Dicho oligonucleótido al menos parcialmente bicatenario tiene un saliente monocatenario y también se denomina en el presente documento esplínquer. El saliente monocatenario proporciona una secuencia de nucleótidos correspondiente a la secuencia de la primera o la tercera extensión de nucleótidos, o parte de la misma, de los elementos de la genoteca de ácido nucleico. Más específicamente, la secuencia de saliente monocatenario es complementaria a dicha primera o tercera extensión, respectivamente. En una realización preferida, el primer oligonucleótido al menos parcialmente bicatenario comprende un sitio de reconocimiento o parte del mismo para una primera enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento. En el presente caso, dicho primer oligonucleótido al menos parcialmente bicatenario tiene un saliente monocatenario que comprende GTC y que por lo tanto es complementario al primer triplete de la secuencia diana. El oligonucleótido al menos parcialmente bicatenario en la realización representada en la Fig. 2 (1) está formado por una parte del oligonucleótido monocatenario que se pliega hacia atrás hacia otra parte del oligonucleótido, formando de este modo un bucle y la estructura bicatenaria. Sin embargo, también se incluye en la presente invención que el oligonucleótido al menos parcialmente bicatenario esté formado por dos oligonucleótidos monocatenarios separados cuyas bases se emparejan entre sí generando de este modo dicho oligonucleótido al menos parcialmente bicatenario como se describe en el presente documento. El extremo del oligonucleótido monocatenario descrito anteriormente que no está proporcionando el saliente monocatenario necesario para la ligación se bloquea para evitar cualquier ligación o reacción que sea potencialmente posible en las etapas posteriores del procedimiento de acuerdo con la presente

invención.

En la etapa 2, se proporciona una primera genoteca de oligonucleótidos que comprende varios miembros. Cada miembro de dicha primera genoteca de oligonucleótidos también se denomina en el presente documento segundo oligonucleótido al menos parcialmente bicatenario y tiene una estructura bicatenaria. Esta clase de segundo oligonucleótido al menos parcialmente bicatenario también se denomina en el presente documento anclaje o molécula de anclaje. La estructura bicatenaria, como en relación con el primer oligonucleótido al menos parcialmente bicatenario, puede estar formada por dos cadenas sencillas que tienen la modificación que permite la inmovilización de los segundos oligonucleótidos al menos parcialmente bicatenarios como se ha descrito anteriormente, y particularmente en relación con la realización representada en la Fig. 1. Los miembros de la primera genoteca de oligonucleótidos comprenden un sitio de reconocimiento, o parte del mismo, para una segunda enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento. Además, dicho segundo oligonucleótido al menos parcialmente bicatenario proporciona un saliente monocatenario. Los miembros de la primera genoteca de oligonucleótidos tienen un saliente monocatenario o parte del mismo que es el mismo para los miembros de la primera genoteca de oligonucleótidos, y por lo que dicho saliente monocatenario es esencialmente complementario al saliente monocatenario del primer oligonucleótido al menos parcialmente bicatenario. En el presente caso, dicho saliente monocatenario es CAC. Posterior al saliente monocatenario, la estructura bicatenaria comprende una extensión de nucleótidos y, más específicamente, una estructura bicatenaria de elemento de estructura bicatenaria, por lo que los miembros de la primera genoteca de oligonucleótidos difieren en la secuencia de dicha extensión o estructura bicatenaria en los nucleótidos, y las secuencias de dicha extensión de nucleótidos corresponden a las secuencias, o parte de las mismas, de la primera mitad de la segunda extensión de los nucleótidos de los elementos de la genoteca de ácido nucleico de acuerdo con la presente invención. Por consiguiente, como se representa en la Fig. 2 (1), se proporciona una primera genoteca de oligonucleótidos, por lo que los miembros de la misma, es decir, los segundos oligonucleótidos al menos parcialmente bicatenarios, están caracterizados de modo que preferentemente la parte bicatenaria de los miembros posterior al saliente monocatenario comprende las diversas secuencias sujetas a la primera mitad de la segunda extensión de los nucleótidos de los elementos de la genoteca de ácido nucleico. En relación con lo mismo, debe reconocerse que un miembro de la primera genoteca de oligonucleótidos comprende una de las secuencias posibles respectivas de la primera mitad de la segunda extensión como también contenida en un elemento, es decir, el elemento correspondiente de la genoteca de ácido nucleico. Dentro de la presente invención se incluye que la estructura bicatenaria comprenda nucleótidos adicionales que puedan, en una realización, o que no puedan, en una realización diferente, contener nucleótidos adicionales de la secuencia diana. En el presente caso, las secuencias respectivas son ATG, TAT y GAG que forman emparejamiento de bases con las secuencias complementarias, es decir (en dirección 3' -> 5') TAC, ATA y CTC, formando parte de la estructura bicatenaria de los segundos oligonucleótidos al menos parcialmente bicatenarios. Los miembros de la primera genoteca de oligonucleótidos difieren por lo tanto en la secuencia de dicha extensión de nucleótidos y la secuencia de dicha extensión de nucleótidos de los miembros de la primera genoteca de nucleótidos corresponde a la secuencia o parte de la misma de la primera mitad de la segunda extensión de los nucleótidos de los elementos de la genoteca de ácido nucleico.

Además, los miembros de la primera genoteca de oligonucleótidos comprenden una modificación, como se describe en el presente documento, que permite la inmovilización de dichas moléculas en o sobre un soporte que puede ser vehículo o superficie que comprenda un compañero de interacción de dicha modificación. Debido a la presencia dicha modificación, esta clase de oligonucleótidos se denomina también anclaje o molécula de anclaje. La inmovilización está específicamente mediada por una interacción entre la modificación y un compañero de interacción de dicha modificación. En el presente caso, la modificación es biotina que está interaccionando con estreptavidina como su compañero de interacción y proporciona un complejo estable, permitiendo la inmovilización de los miembros de la primera genoteca de oligonucleótidos. En la Fig. 2 el sitio de reconocimiento para la enzima de restricción de tipo IIS de los miembros de la primera genoteca de oligonucleótidos está indicado como un recuadro en negro que ilustra el hecho de que, para esta clase de enzima de restricción, el sitio de reconocimiento se encuentra fuera del sitio de escisión. Todos los miembros de la primera genoteca de oligonucleótidos tienen el mismo sitio de reconocimiento, o parte del mismo, para una enzima de restricción de tipo IIS.

En la etapa 3, el primer oligonucleótido menos parcialmente bicatenario y la primera genoteca de oligonucleótidos y los miembros de la misma, respectivamente, se combinan. Tras dicha combinación, una molécula del primer oligonucleótido al menos parcialmente bicatenario se liga con una molécula de un miembro de la primera genoteca de oligonucleótidos, preferentemente por medio de una ligasa o una actividad de ligasa. Dicha ligación se produce tras la formación de un primer producto de preligación a través de la hibridación de los salientes monocatenarios de las dos clases de moléculas, actuando dichos salientes como "extremos cohesivos". El producto de reacción de la ligación se denomina producto de primera ligación. Debido a la complejidad de la primera genoteca de oligonucleótidos, que es en el presente caso de 3, también los productos de primera ligación comprenden 3 especies que, al menos, difieren en la secuencia correspondiente a la primera mitad de la segunda extensión. Típicamente, el resto de los productos de primera ligación en términos de secuencias de nucleótidos es preferentemente el mismo para los diversos elementos y, por lo tanto, especies de la genoteca de productos de primera ligación así formada.

En la etapa 4, el producto de primera ligación se corta con una enzima de restricción de tipo IIS. Dicho sitio de reconocimiento de enzima de restricción de tipo IIS es el proporcionado por el segundo oligonucleótido al menos parcialmente bicatenario, y se denomina en el presente documento segunda enzima de restricción de tipo IIS. El

corte del producto de primera ligación se realiza en la combinación y la genoteca, respectivamente, que consisten en los productos de primera ligación. Por consiguiente, el corte o escisión del producto de primera ligación da como resultado una diversidad de moléculas así cortadas, por lo que el producto de reacción de dicho corte es una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados y una pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios acortados. Se reconocerá que la pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados consiste en diversas especies moleculares que difieren en la secuencia del saliente monocatenario.

Se reconocerá también que la realización de la etapa 3 y 4, respectivamente, puede llevarse a cabo en condiciones de reacción diferentes en términos de inmovilización o no inmovilización de los miembros de la primera genoteca de oligonucleótidos y el producto de primera ligación, respectivamente. En cada caso, las moléculas respectivas proporcionan una modificación que, como se ha resumido anteriormente, permite la inmovilización. Por consiguiente, cualquier secuencia con respecto a la etapa de inmovilización puede, en principio, realizarse en diversas realizaciones. Por consiguiente, dentro de la presente invención se incluye realizar la reacción de ligación sujeta a la etapa 3 en la fase líquida, es decir, no estando los miembros de la primera genoteca de oligonucleótidos inmovilizados. Posteriormente, se produce el corte del producto de primera ligación. En una realización, dicho corte se realiza entonces también en la fase líquida, no estando el producto de primera ligación inmovilizado en un soporte mediante su modificación. Después del corte el segundo oligonucleótido al menos parcialmente bicatenario acortado se elimina de la mezcla de reacción mediante su modificación por unión al soporte. Como alternativa, después de la ligación, el producto de primera ligación puede inmovilizarse en o sobre un soporte y la etapa de corte del producto de primera ligación se realiza estando el producto de primera ligación inmovilizado en dicho soporte. En cualquier caso, el segundo oligonucleótido al menos parcialmente bicatenario acortado se elimina de la reacción por inmovilización del mismo a través de su modificación en un soporte. Se reconocerá que debido a la modificación que permite una interacción específica tal como entre biotina y estreptavidina, para los fines de la inmovilización la reacción se transfiere a un recipiente de reacción, preferentemente cuyas paredes tienen un compañero de interacción de la modificación, en el presente ejemplo dado que la modificación es biotina, estreptavidina. Debido a la alta afinidad de la biotina y la estreptavidina, la unión de la molécula que contiene la modificación a la superficie que comprende estreptavidina que actúa como soporte se produce inmediatamente. En otras palabras, si se va a realizar una reacción con una molécula que tiene dicha modificación, el recipiente de reacción o cualquier soporte disponible en dicha reacción no debe comprender un compañero de interacción de la modificación, mientras que en el caso de que se desee una inmovilización, se usa un recipiente de reacción cuya superficie proporciona dicho compañero de interacción de la modificación, mediante el recipiente de reacción o mediante otras superficies o soportes prevalentes en dicha reacción.

La etapa 5 de la reacción consiste en la eliminación de los segundos oligonucleótidos al menos parcialmente bicatenarios acortados que se realiza preferentemente por inmovilización del segundo oligonucleótido al menos parcialmente bicatenario acortado

Debido a esta secuencia de reacción, la reacción y preferentemente el sobrenadante de la misma contienen una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados. Los diversos primeros oligonucleótidos al menos parcialmente bicatenarios alargados definen, por su saliente monocatenario, diversas especies o subgrupos de oligonucleótidos. En el presente caso, la pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados comprende tres especies o subgrupos diferentes. Cada subgrupo comprende a su vez diversos tipos de moléculas que preferentemente difieren con respecto a la extensión de nucleótidos que forma la extensión bicatenaria o parte de la misma, y que corresponde a la segunda extensión de nucleótidos, o parte de la misma, del ácido nucleico diana. Se reconocerá que los diversos subgrupos contienen preferentemente moléculas que, en su totalidad, tienen la misma extensión de nucleótidos que forma la extensión bicatenaria o parte de la misma y que corresponde a la segunda extensión de nucleótidos, o parte de la misma, del ácido nucleico diana. En una etapa adicional, que es típicamente la etapa 6, se proporciona una segunda genoteca de oligonucleótidos que comprende varios miembros. Cada miembro de dicha segunda genoteca de oligonucleótidos es un tercer oligonucleótido al menos parcialmente bicatenario y tiene una estructura bicatenario. Dicha estructura bicatenaria puede diseñarse tal como se ha resumido en el presente documento en relación con otros oligonucleótidos al menos parcialmente bicatenarios, tales como y preferentemente, el primer oligonucleótido al menos parcialmente bicatenario. Más específicamente, el tercer oligonucleótido al menos parcialmente bicatenario comprende un sitio de reconocimiento, o parte del mismo, para una tercera enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento, y un saliente monocatenario. El saliente monocatenario o parte del mismo es diferente para los miembros de la segunda genoteca de oligonucleótidos. En el presente ejemplo, dicho saliente es ATG, TAT o GAG. Dichos salientes monocatenarios o parte de los mismos de los terceros oligonucleótidos al menos parcialmente bicatenarios son esencialmente complementarios al saliente monocatenario de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados. Un experto en la materia reconocerá que solamente una especie de los primeros oligonucleótidos bicatenarios alargados puede reaccionar y en última instancia ligarse con una especie de los terceros oligonucleótidos al menos parcialmente bicatenarios, concretamente aquellos cuyos salientes monocatenarios sean preferentemente, esencialmente o parcialmente complementarios entre sí. Sólo tomados en su totalidad los salientes de las diversas especies de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados y de las diversas especies de los terceros oligonucleótidos al menos parcialmente bicatenarios son complementarios entre sí. Los miembros de la segunda genoteca de oligonucleótidos

comprenden además una extensión de nucleótidos en la estructura bicatenaria que es diferente en los diversos miembros de la segunda genoteca de oligonucleótidos. La diversidad con respecto a esta extensión se obtiene como resultado de las posibilidades combinatorias de la primera mitad de la segunda extensión y la segunda mitad de la segunda extensión de la secuencia diana. En el presente ejemplo, la primera mitad de la segunda extensión y la segunda mitad de la segunda extensión comprenden cada una tres tripletes diferentes, de modo que el número de terceros oligonucleótidos al menos parcialmente bicatenarios debe ser de 3×3 , es decir, nueve moléculas diferentes. Si la primera mitad consistía en tres tripletes diferentes y la segunda mitad consistía en cuatro tripletes diferentes, la complejidad de la segunda genoteca de oligonucleótidos sería de 3×4 , es decir, 12, y dicha genoteca proporcionaría 12 especies de oligonucleótidos diferentes. Por consiguiente, en el presente caso la segunda genoteca de oligonucleótidos consiste en 9 miembros, proporcionando de este modo cualquier combinación que surja de la variabilidad de la segunda extensión de la secuencia diana, más específicamente que surja del hecho de que dicha segunda extensión consiste en una primera y una segunda mitad que contribuyen a la secuencia diana mediante moléculas diferentes, por lo que y cada una de dichas primera y segunda extensión comprende tres formas o secuencias diferentes que dan como resultado las 9 especies diferentes. De nuevo, los miembros de la segunda genoteca de oligonucleótidos comprenden una modificación, por lo que dicha modificación es preferentemente la misma que se describe en relación con la primera genoteca de oligonucleótidos.

En la etapa 7, la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados que consiste en tres especies diferentes, y la segunda genoteca de oligonucleótidos se combinan. Debido a la complementariedad de los salientes como se ha resumido anteriormente, cada molécula de la pluralidad, es decir, cada una de las tres especies diferentes de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados se liga con una molécula de un miembro de la segunda genoteca de oligonucleótidos. La pareja de moléculas respectiva a ligar está definida por la complementariedad de sus salientes que definen una pareja de reacción. Una vez que se hibridan o aparean las dos moléculas correspondientes, la ligación típicamente se produce tras la adición o presencia de una actividad de ligasa respectiva. En dichos productos de ligación, la primera mitad de la segunda extensión y la segunda mitad de la segunda extensión de la secuencia diana ya están contenidas en el producto de ligación y recuadradas en la representación de la Fig. 2 (2). Los nueve productos de ligación diferentes comprenden, como una extensión bicatenaria, ya la primera mitad de la segunda extensión del ácido nucleico diana y la segunda mitad de la segunda extensión del ácido nucleico diana.

En el presente caso, el producto de segunda ligación comprende un total de nueve secuencias diferentes, por lo que las secuencias difieren en el saliente monocatenario introducido en el producto de segunda ligación por los primeros oligonucleótidos al menos parcialmente bicatenarios alargados y, en el presente ejemplo, el primer triplete de la estructura bicatenaria proporcionado por los miembros de la segunda genoteca de oligonucleótidos. En otras realizaciones, diferencias adicionales pueden residir en partes adicionales del producto de ligación, sin embargo, se prefiere que las dos diferencias mencionadas anteriormente en los diversos miembros de la genoteca que consisten en el producto de segunda ligación sean de otro modo las mismas. De nuevo, los diversos miembros de la genoteca que consiste en los productos de segunda ligación tienen un sitio de reconocimiento de enzima de restricción de tipo IIS, o parte del mismo, que corresponde a uno de los segundos oligonucleótidos al menos parcialmente bicatenarios.

Las siguientes etapas, es decir, la etapa 8, consiste en el corte del producto de segunda ligación con la tercera enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico del tercer oligonucleótido al menos parcialmente bicatenario que está, en principio, diseñado y generado, respectivamente, del mismo modo que se ha descrito en relación con el producto de primera ligación. Esto es aplicable particularmente con respecto a la secuencia potencial de unión y corte y la eliminación por inmovilización, respectivamente. Como resultado de la etapa 8 en el transcurso de la cual el producto de segunda ligación se corta con la tercera enzima de restricción de tipo IIS, se proporciona una pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados y terceros oligonucleótidos al menos parcialmente bicatenarios acortados. De nuevo, los terceros oligonucleótidos al menos parcialmente bicatenarios acortados así liberados del producto de segunda ligación se eliminan de la reacción por la modificación y su interacción con un compañero de interacción tal como estreptavidina, en relación con que sea biotina la modificación unida a una superficie.

Los nueve segundos oligonucleótidos al menos parcialmente bicatenarios alargados diferentes así obtenidos se representan en la Fig. 2 (2).

En la etapa 9, se proporciona una tercera genoteca de oligonucleótidos con una diversidad de cuartos oligonucleótidos al menos parcialmente bicatenarios. Dichos cuartos oligonucleótidos al menos parcialmente bicatenarios difieren en sus salientes. De otro modo, el diseño de este cuarto oligonucleótido al menos parcialmente bicatenario puede ser, en una realización, similar a las realizaciones descritas en relación con otros oligonucleótidos al menos parcialmente bicatenarios descritos en relación con las Fig. 1 y 2, respectivamente. El saliente monocatenario proporciona una secuencia de nucleótidos que es esencialmente complementaria a los salientes monocatenarios de los diversos segundos oligonucleótidos al menos parcialmente bicatenarios alargados. El número de las especies respectivas de la tercera genoteca de oligonucleótidos surge de y corresponde a la variabilidad de la segunda mitad de la segunda extensión del ácido nucleico a preparar, y por lo tanto los diversos segundos oligonucleótidos al menos parcialmente bicatenarios alargados. Dichos segundos oligonucleótidos al menos parcialmente bicatenarios alargados tienen, en el presente caso, tres salientes monocatenarios diferentes, si se toman en su conjunto. Más específicamente, estos salientes son AAA, CTC y GAG, siendo complementarios a la

segunda mitad de la segunda extensión del ácido nucleico diana, en concreto, TTT, GAG y CTC. Además, los cuartos oligonucleótidos al menos parcialmente bicatenarios comprenden un sitio de reconocimiento, o parte del mismo, para una cuarta enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento. Dichos cuartos oligonucleótidos al menos parcialmente bicatenarios, es decir, en el presente ejemplo, tres de estos oligonucleótidos diferentes, y la pluralidad de los segundos oligonucleótidos al menos parcialmente bicatenarios alargados, en el presente caso nueve especies diferentes de dichos oligonucleótidos, se combinan entre sí y se ligan por medio de una actividad de ligasa del apareamiento de las dos moléculas por hibridación a través de sus salientes monocatenarios complementarios. De nuevo, en la molécula de los cuartos oligonucleótidos al menos parcialmente bicatenarios y una molécula de los diversos segundos oligonucleótidos al menos parcialmente bicatenarios alargados se aparean debido a la complementariedad de sus salientes. Hasta ahora, uno de dichos tres cuartos oligonucleótidos al menos parcialmente bicatenarios diferentes es, en principio, reactivo con tres de los segundos oligonucleótidos al menos parcialmente bicatenarios alargados, en concreto aquellos cuya cadena sencilla es complementaria a la cadena sencilla del cuarto oligonucleótido al menos parcialmente bicatenario respectivo. Esto, de nuevo, proporciona un total de nueve oligonucleótidos diferentes. Las moléculas así apareadas se ligan después por medio de una actividad de ligasa que proporciona un producto de tercera ligación. Dicho producto de tercera ligación consiste en nueve especies diferentes de productos de ligación, por lo que los productos de ligación contienen tanto la primera como la segunda mitad de la segunda extensión del ácido nucleico diana, estando cada una de dichas secuencias flanqueada en ambos lados por otro grupo, en el presente caso tres nucleótidos. Se reconocerá que la complejidad de las genotecas y oligonucleótidos usados en el procedimiento de la presente solicitud es el resultado inmediato de y corresponde a la variabilidad de la primera mitad y la segunda mitad de la segunda extensión del ácido nucleico a preparar, o del producto de multiplicación de la misma. Como en relación con otras reacciones, debido a la complementariedad de los salientes, se produce una reacción de uno en uno, es decir, una reacción entre una molécula de cada especie de reactivo contenido en la reacción. El producto de tercera ligación comprende por lo tanto en última instancia la secuencia diana como se representa en la Fig. 2 (1). La secuencia variable de la secuencia diana, es decir, la que es diferente en las diversas especies del producto de ligación, está recuadrada en la Fig. 2 (3).

De nuevo, como se describe en relación con el procedimiento de la Fig. 1, dicha molécula puede cortarse con una enzima de restricción de tipo IIS haciendo uso de la secuencia de reconocimiento respectiva proporcionada por el cuarto oligonucleótido al menos parcialmente bicatenario, y someterse a reacciones adicionales usando los mismos esquemas de reacción que se han resumido anteriormente, o puede someterse a la escisión usando una enzima de restricción de tipo IIS que después se proporciona por el tercer oligonucleótido al menos parcialmente bicatenario, permitiendo la liberación de la secuencia diana. En la última alternativa, el producto de tercera ligación puede cortarse por lo tanto para introducirse en una reacción de transposición con otras moléculas, típicamente preparadas de una forma similar para permitir una reacción de transposición como se describe en relación con el documento WO 00/75368.

Procedimiento para preparar una genoteca de ácido nucleico como se ilustra en la Fig. 3

En una alternativa adicional de los procedimientos de acuerdo con la presente invención y, más específicamente, el procedimiento de acuerdo con el primer aspecto, se va a sintetizar una secuencia diana como se representa en la Fig. 3 (1), por lo que la primera posición, es decir, el primer triplete en el caso representado, es variable, y en el presente caso consiste en tres nucleótidos. Dicha posición variable también se denomina primera extensión, mientras que los tripletes posteriores se denominan segunda y tercera extensión, respectivamente. Excepto las cuestiones analizadas en lo siguiente, las etapas y las moléculas usadas en dicho procedimiento corresponden a las descritas en el presente documento en relación con el procedimiento de acuerdo con el primer aspecto de la invención, que se incorpora en el presente documento por referencia para evitar cualquier repetición innecesaria.

En una primera etapa, se proporciona una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios. Dichos oligonucleótidos al menos parcialmente bicatenarios tienen una estructura bicatenaria y un saliente monocatenario. Esta clase de moléculas también se denominan en el presente documento esplínquer. Estos primeros oligonucleótidos al menos parcialmente bicatenarios difieren con respecto a la secuencia del saliente monocatenario. La diversidad de los salientes monocatenarios corresponde a la diversidad de la posición variable, por lo que los salientes de las diversas especies de los primeros oligonucleótidos al menos parcialmente bicatenarios son complementarias a las secuencias de los diversos elementos de la genoteca de ácido nucleico a preparar y, por lo tanto, complementarias a la primera extensión. En el presente ejemplo, el saliente es (en dirección 3' -> 5') TAC, ATA o CTC. En una realización, los primeros oligonucleótidos al menos parcialmente bicatenarios comprenden un sitio de reconocimiento, o parte del mismo, para una primera enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento.

En una segunda etapa, se proporciona una primera genoteca de oligonucleótidos que comprende varios miembros y varias especies, respectivamente. Cada miembro de dicha primera genoteca de oligonucleótidos también se denomina en el presente documento segundo oligonucleótido al menos parcialmente bicatenario y tiene una estructura bicatenaria. La estructura bicatenaria, como en relación con el primer oligonucleótido al menos parcialmente bicatenario, puede estar formada por dos cadenas sencillas que tienen una modificación adicional como se ha descrito anteriormente que, en principio, permite la inmovilización de dicha primera genoteca de oligonucleótidos en un soporte, por lo que dicho soporte comprende o proporciona un compañero de interacción

para dicha modificación. En un ejemplo preferido, la modificación es biotina y el compañero de interacción de dicho soporte es estreptavidina. Los miembros de la primera genoteca de oligonucleótidos comprenden un sitio de reconocimiento, o parte del mismo, para una segunda de enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento. Además, dichos segundos oligonucleótidos al menos parcialmente bicatenarios proporcionan un saliente monocatenario. Los miembros y especies, respectivamente, de la primera genoteca de oligonucleótidos difieren entre sí en la medida en que el saliente monocatenario o parte del mismo sea diferente para los miembros o especies de la primera genoteca de oligonucleótidos, y por lo que dichos salientes monocatenarios son esencialmente o parcialmente complementarios al saliente monocatenario de los primeros oligonucleótidos al menos parcialmente bicatenarios. En el presente caso, dichos salientes monocatenarios son ATG, TAT y GAG, respectivamente. De nuevo, la complejidad de la genoteca se define por la variabilidad de la primera extensión del ácido nucleico a preparar. Por consiguiente, si la complejidad era, como se representa en el ejemplo, de tres debido a las tres secuencias diferentes, la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios y de la primera genoteca de oligonucleótidos será de tres. Posterior al saliente monocatenario, la estructura bicatenaria de los segundos oligonucleótidos al menos parcialmente bicatenarios comprende una extensión de nucleótidos, por lo que los miembros de la primera genoteca de oligonucleótidos tienen la misma secuencia. Inmediatamente después de los tres nucleótidos que comprenden el saliente, los segundos oligonucleótidos al menos parcialmente bicatenarios tienen una extensión que corresponde a la secuencia de la segunda extensión de la secuencia diana. En los segundos oligonucleótidos al menos parcialmente bicatenarios, sin embargo, dicha segunda extensión de la secuencia diana está presente en una forma bicatenaria. La pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios y la primera genoteca de oligonucleótidos y sus miembros, es decir, los segundos oligonucleótidos al menos parcialmente bicatenarios, se representa en la Fig. 3 (1).

Los miembros de la primera genoteca de oligonucleótidos comprenden una modificación. En el presente caso, dicha modificación permite la inmovilización de los diversos miembros de la genoteca en un soporte que típicamente es la superficie de un tubo de reacción o de un pocillo. La inmovilización está específicamente mediada por una interacción por la modificación y el compañero de interacción de dicha modificación que está típicamente presente en el soporte o vehículo. En el presente caso, la modificación es biotina que está interaccionando con estreptavidina ya que su compañero de interacción proporciona un complejo estable que permite la inmovilización de los miembros de la primera genoteca de oligonucleótidos.

Como en relación con los procedimientos representados en la Figs. 1 y 2, el sitio de reconocimiento para la enzima de restricción de tipo IIS de los miembros de la primera genoteca de oligonucleótidos está indicado como un recuadro en negro que ilustra el hecho de que para esta clase de enzima de restricción el sitio de reconocimiento se encuentra fuera del sitio de escisión.

En la etapa 3, los primeros oligonucleótidos al menos parcialmente bicatenarios y los miembros de la primera genoteca de oligonucleótidos se combinan. Tras dicha combinación, una molécula del primer oligonucleótido al menos parcialmente bicatenario se liga con una molécula de los miembros de la primera genoteca de oligonucleótidos, preferentemente por medio de una ligasa o actividad de ligasa. Dicha ligación se produce formando un primer producto de preligación a través de la hibridación de los salientes monocatenarios de las dos clases de moléculas, actuando dichos salientes como "extremos cohesivos". El producto de reacción de la ligación se denomina, en su totalidad, producto de primera ligación como se representa en la Fig. 3 (1). Como se indica en la Fig. 3 (1), dicho producto de primera ligación varía y representa las posiciones variables de la secuencia diana. En otras palabras, el producto de primera ligación forma por sí mismo una genoteca que comprende, en el presente ejemplo, tres especies diferentes de producto de ligación que difieren en la posición que representa la posición variable de la secuencia diana. Dicho producto de ligación comprende además la segunda extensión de la secuencia diana que es, en el presente caso, GAG. Hasta ahora, los miembros de esta combinación de productos de ligación al menos, y preferentemente, difieren únicamente en la secuencia correspondiente al primer triplete de la secuencia diana que es la posición variable, mientras que típicamente el resto del producto de primera ligación en términos de secuencia de nucleótidos es preferentemente el mismo para los diversos elementos de la genoteca de productos de primera ligación así formada.

En la etapa 4, el producto de primera ligación se corta con una enzima de restricción de tipo IIS. Dicha enzima de restricción de tipo IIS es una cuyo sitio de reconocimiento se proporciona por las moléculas de anclaje, es decir, por los segundos oligonucleótidos al menos parcialmente bicatenarios. El corte de los productos de primera ligación se realiza en la combinación y la genoteca, respectivamente, que consisten en los productos de primera ligación. Por consiguiente, el corte o escisión de los productos de primera ligación da como resultado una diversidad de moléculas así cortadas, por lo que el producto de reacción de dicho corte es una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados y una pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios acortados. En cada caso, en el presente ejemplo, se proporcionan tres especies diferentes de las moléculas respectivas.

Como se ha descrito en relación con los procedimientos sujetos a las Figs. 1 y 2, la realización de la etapa 4 puede llevarse a cabo en escenarios diferentes en términos de inmovilización o no inmovilización de los miembros del producto de primera ligación, respectivamente. La divulgación respectiva se incorpora en el presente documento por referencia.

La etapa adicional de la reacción consiste en la eliminación de los segundos oligonucleótidos al menos parcialmente bicatenarios acortados que se realiza preferentemente por inmovilización de los segundos oligonucleótidos al menos parcialmente bicatenarios acortados.

Debido al esquema de reacción, la reacción, preferentemente el sobrenadante de la misma, contiene una pluralidad de oligonucleótidos al menos parcialmente bicatenarios que también se denominan primeros oligonucleótidos al menos parcialmente bicatenarios alargados. Dicha pluralidad de oligonucleótidos consiste en tres especies diferentes que tienen la misma secuencia en el saliente monocatenario, pero difieren en la secuencia de los tres nucleótidos adyacentes que corresponden a la secuencia diana y que están recuadrados en la Fig. 3 (2). En una etapa adicional, que es típicamente la etapa 6, se proporciona un tercer oligonucleótido al menos parcialmente bicatenario. El tercer oligonucleótido al menos parcialmente bicatenario tiene una estructura bicatenaria que puede diseñarse tal como se resume en el presente documento en relación con otros oligonucleótidos al menos parcialmente bicatenarios y, preferentemente, como el primer oligonucleótido al menos parcialmente bicatenario. Más específicamente, dicho tercer oligonucleótido al menos parcialmente bicatenario comprende un sitio de reconocimiento, o parte del mismo, para una tercera enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento, y el saliente monocatenario. El saliente monocatenario o parte del mismo es complementario al saliente monocatenario de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados. Debido a esto, el tercer oligonucleótido al menos parcialmente bicatenario consiste en una especie solamente. El tercer oligonucleótido al menos parcialmente bicatenario comprende por lo tanto una extensión de nucleótidos en la estructura bicatenaria que es idéntica a la tercera extensión de la secuencia diana, es decir, en el presente caso ACA. Este tercer oligonucleótido al menos parcialmente bicatenario que, debido a tener una modificación, permite su inmovilización como se ha descrito anteriormente, y que por lo tanto también puede denominarse molécula de anclaje, se liga después con los tres miembros de la genoteca formando la pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados. De nuevo, el apareamiento y la ligación se realizan de una forma similar a la descrita en relación con otros procedimientos de ligación. Además, dichos procedimientos de ligación pueden variarse de acuerdo con lo que se ha dicho anteriormente. En otras palabras, debido a la complementariedad de los salientes, cada molécula de la pluralidad del primer oligonucleótido al menos parcialmente bicatenario alargado se liga con una sola molécula del tercer oligonucleótido al menos parcialmente bicatenario. El producto de ligación así formado de dicha reacción de ligación también se denomina producto de segunda ligación y se representa en la Fig. 3 (2), estando la extensión variable de la secuencia diana recuadrada. Dicho producto de ligación comprende entonces la secuencia diana total, por lo que la primera posición de la misma muestra la variabilidad de la secuencia diana, representada por las diferentes especies del producto de segunda ligación.

Procedimiento para preparar una genoteca de ácido nucleico como se ilustra en la Fig. 4

En otra realización del procedimiento de acuerdo con la presente invención y, más específicamente, del segundo aspecto del mismo, la secuencia diana es como se representa en la Fig. 4 (1), es decir, comprende dos posiciones variables que forman una primera extensión, más específicamente una primera mitad de la primera extensión y una segunda mitad de la primera extensión, y una segunda extensión. Excepto las cuestiones analizadas en lo siguiente, las etapas y las moléculas usadas en dicho procedimiento corresponden a las descritas en el presente documento en relación con el procedimiento de acuerdo con el segundo aspecto de la presente invención, que se incorpora en el presente documento por referencia para evitar cualquier repetición innecesaria.

En el presente ejemplo, la primera mitad de la primera extensión consiste en uno de los tres tripletes ATG, TAT y GAG, y la segunda mitad de la primera extensión consiste en CTC, GAG o TTT. La segunda extensión consiste en ACA.

En una primera etapa, se proporciona una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios. Dicha pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios mostrados en la Fig. 4 (1) comprende tres especies diferentes de primeros oligonucleótidos al menos parcialmente bicatenarios. Las diferentes especies o miembros de esta clase de genoteca difieren al menos, y preferentemente, en el saliente monocatenario. Los diversos primeros oligonucleótidos al menos parcialmente bicatenarios difieren más específicamente de modo que el saliente monocatenario es (en dirección 3' -> 5') TAC, ATA y CTC, siendo por lo tanto complementario a las diversas secuencias de la secuencia variable de la primera mitad de la primera extensión de la secuencia diana. Aparte de eso, esta clase de primeros oligonucleótidos al menos parcialmente bicatenarios tiene una estructura similar o idéntica a la descrita en relación con las otras realizaciones del procedimiento de la presente invención y, más específicamente, las del segundo aspecto de los procedimientos de acuerdo con la presente invención, más particularmente comprende una secuencia de reconocimiento o parte de la misma de una primera enzima de restricción de tipo IIS.

En la etapa 2, se proporciona una primera genoteca de oligonucleótidos que comprende varios miembros. Cada miembro de dicha primera genoteca de oligonucleótidos también se denomina en el presente documento segundo oligonucleótido al menos parcialmente bicatenario y tiene una estructura bicatenaria. La estructura bicatenaria, como en realización con el primer oligonucleótido al menos parcialmente bicatenario, puede estar formada por dos cadenas sencillas que tengan la modificación adicional que se ha descrito anteriormente. Los miembros de la primera genoteca de oligonucleótidos comprenden un sitio de reconocimiento, o parte del mismo, para una segunda enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento. Además dicho segundo

oligonucleótido al menos parcialmente bicatenario proporciona un saliente monocatenario. Los miembros de la primera genoteca de oligonucleótidos difieren entre sí al menos en la medida en que el saliente monocatenario o parte del mismo sea el mismo para un subgrupo de los miembros de la primera genoteca de oligonucleótidos, y por lo que dicho saliente monocatenario es esencialmente o parcialmente complementario al saliente monocatenario de una especie de los diversos primeros oligonucleótidos al menos parcialmente bicatenarios. En otras palabras, la complejidad de la primera genoteca se determina por la complejidad de la genoteca de ácido nucleico a preparar y puede calcularse multiplicando la variabilidad de la primera mitad de la primera extensión con la variabilidad de la segunda mitad de la primera extensión. Por consiguiente, en su totalidad, todos los oligonucleótidos de la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios tienen un saliente que, de nuevo en su totalidad, será complementario a los salientes de los oligonucleótidos de la primera genoteca de oligonucleótidos. Más específicamente, dicho saliente monocatenario es ATG, correspondiente al saliente monocatenario del primer oligonucleótido al menos parcialmente bicatenario TAC, TAT, correspondiente al saliente monocatenario del primer oligonucleótido al menos parcialmente bicatenario ATA, y GAG correspondiente al saliente monocatenario del primer oligonucleótido al menos parcialmente bicatenario CTC. Hasta ahora, esta primera genoteca de oligonucleótidos proporciona tres especies diferentes en términos de diferentes salientes. Cada una de estas especies comprende tres especies adicionales que están caracterizadas en la medida en que, inmediatamente después de los salientes monocatenarios, la secuencia de acuerdo con la segunda mitad de la primera extensión de la secuencia diana se repite. Dicha secuencia de nucleótidos es parte de la estructura bicatenaria. Hasta ahora, el primer subgrupo de la primera genoteca de oligonucleótidos consiste en ATG CTG/GAG, ATG GAG/CTC y ATG TTT/AAA, la segunda subespecie consiste en TAT CTC/GAG, TAT GAG/CTC y TAT TTT/AAA, la tercera subespecie consiste en GAG CTC/GAG, GAG GAG/CTC y GAG TTT/AAA, por lo que los oligonucleótidos indicados después de la "r" son indicativos de los tres nucleótidos que son complementarios a los tres nucleótidos inmediatamente precedentes. Hasta ahora, la primera genoteca de oligonucleótidos consiste en nueve miembros que se forman por la variabilidad de la primera mitad de la primera extensión, multiplicada por las variantes de la segunda mitad de la primera extensión. En otras palabras, el número de los miembros de la primera genoteca de oligonucleótidos es el producto del número de secuencias diferentes de la primera mitad de la primera extensión multiplicado por el número de diferentes secuencias de la segunda mitad de la primera extensión del ácido nucleico diana.

De forma similar a los miembros de las primeras genotecas de oligonucleótidos usadas en relación con las otras realizaciones de la presente invención, los segundos oligonucleótidos al menos parcialmente bicatenarios comprenden una modificación como se especifica en relación con los mismos.

En la etapa 3, los diversos primeros oligonucleótidos al menos parcialmente bicatenarios y la primera genoteca de oligonucleótidos y los miembros de la misma, respectivamente, se combinan. Tras dicha combinación, una molécula del primer oligonucleótido al menos parcialmente bicatenario se liga con una molécula de un miembro de la primera genoteca de oligonucleótidos, preferentemente por medio de una ligasa o actividad de ligasa. Dicha ligación se produce tras la formación de un primer producto de preligación a través de la hibridación de los salientes monocatenarios correspondientes de las dos clases de moléculas, actuando dichos salientes como extremos cohesivos. El producto de reacción de la ligación se denomina producto de primera ligación. Debido a la diversidad del primer oligonucleótido al menos parcialmente bicatenario y la primera genoteca de oligonucleótidos, la reacción de ligación da como resultado una combinación de productos de primera ligación, en el presente caso de nueve especies diferentes de dichos productos de primera ligación. Dichos productos de ligación difieren en su secuencia en la medida en que contienen la segunda extensión variable que consiste en la primera mitad y la segunda mitad de la primera extensión del ácido nucleico diana que está indicada como un recuadro en la Fig. 4 (2). Aparte de esto, los productos de ligación que en realidad constituyen una genoteca adicional difieren hasta ahora, y preferentemente, en esta secuencia, mientras que típicamente el resto del producto de primera ligación en términos de secuencia de nucleótidos es preferentemente el mismo para los diversos elementos de la genoteca de productos de primera ligación así formada.

En la etapa 4, los diversos productos de primera ligación se cortan con una enzima de restricción de tipo IIS. Dicha restricción de tipo IIS es aquella cuyo sitio de reconocimiento es el proporcionado por los segundos oligonucleótidos al menos parcialmente bicatenarios (que está marcado mediante un recuadro negro en las Figs. 4 (1,2)). El corte del producto de primera ligación se realiza en la reacción y la genoteca, respectivamente, que comprenden el producto o productos de primera ligación. Por consiguiente, el corte o escisión del producto de primera ligación da como resultado una diversidad de moléculas así cortadas, por lo que el producto de reacción de dicho corte es una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados y una pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios acortados. Como en las diversas otras realizaciones descritas en relación con las Figs. 1 a 3, dicha segunda enzima de restricción de tipo IIS es Eam1104I.

De nuevo, se reconocerá que la realización de la etapa 4 puede llevarse a cabo en escenarios diferentes en términos de inmovilización o no inmovilización de los miembros de la primera genoteca de oligonucleótidos como se describe en relación con las otras realizaciones del procedimiento de acuerdo con la invención.

Debido al corte del producto de primera ligación se generan un total de nueve primeros oligonucleótidos al menos parcialmente bicatenarios alargados diferentes. Típicamente, la especie individual de la pluralidad o diversidad así generada de primeros oligonucleótidos al menos parcialmente bicatenarios alargados comprende nueve especies diferentes definidas por las diversas secuencias diana deseadas, por lo que existen tres subgrupos diferentes de

moléculas que están definidas por su saliente monocatenario que es complementario a la segunda mitad de la primera extensión de la secuencia diana. Más específicamente, en el presente ejemplo, los tres subgrupos son aquellos que están definidos por un saliente monocatenario de GAG, uno definido por un saliente monocatenario de CTC y uno definido por un saliente monocatenario de AAA. Cada subgrupo varía a su vez en la medida en que la extensión y, en particular, los tres nucleótidos siguientes o precedentes, respectivamente, al saliente monocatenario y que forman la primera extensión bicatenaria de la estructura bicatenaria corresponden a la primera mitad de la primera extensión de la secuencia diana y, por lo tanto, son ATG/TAC, TAT/ATA y GAG/CTC, respectivamente, para cada uno de dichos subgrupos.

En una etapa siguiente se añade una segunda genoteca de oligonucleótidos que consiste en terceros oligonucleótidos al menos parcialmente bicatenarios como se representa en la Fig. 4 (2). Los miembros de la segunda genoteca de oligonucleótidos difieren con respecto a sus salientes monocatenarios, que corresponden a la segunda mitad de la primera extensión de la secuencia diana, es decir, en el presente caso CTC, GAG y TTT. De otro modo, los miembros son los mismos que se describen en relación con la primera genoteca de oligonucleótidos, por lo que comprenden una secuencia de reconocimiento para una tercera enzima de restricción de tipo IIS, o parte de la misma, y comprende una modificación.

La primera parte de la estructura bicatenaria del tercer oligonucleótido al menos parcialmente bicatenario después del saliente monocatenario es la misma dentro de todos los miembros de la segunda genoteca de oligonucleótidos y corresponde a la segunda extensión de la secuencia diana, que es en el presente caso ACA. Como esta extensión del tercer oligonucleótido al menos parcialmente bicatenario es parte de la estructura bicatenaria de dicho tercer oligonucleótido al menos parcialmente bicatenario, la estructura bicatenaria respectiva en los diversos miembros de la segunda genoteca de oligonucleótidos es por lo tanto ACA/TGT para cada uno y cualquiera de los miembros de dicha segunda genoteca de oligonucleótidos.

Los miembros de la segunda genoteca de oligonucleótidos se ligan después, típicamente después de una etapa de apareamiento, con los diversos primeros oligonucleótidos al menos parcialmente bicatenarios alargados y hasta ahora, las etapas adicionales corresponden al descrito en relación con cada una y cualquiera de las otras realizaciones del procedimiento de acuerdo con la presente invención. El producto de reacción respectivo es el producto de segunda ligación que consiste, según se determina por la complejidad de la genoteca de ácido nucleico a preparar, en nueve especies diferentes.

Los expertos en la materia entenderán que es posible variar la concentración con la que los diversos oligonucleótidos al menos parcialmente bicatenarios se añaden a las reacciones individuales. Aparte de cualquier proporción equimolar respecto al otro, la proporción puede estar sesgada, lo que dará como resultado una sobre o infra representación de la posición variable introducida en la molécula de ácido nucleico por el oligonucleótido u oligonucleótidos respectivos. De acuerdo con lo mismo, la genoteca de ácido nucleico preparada por los procedimientos de acuerdo con la presente invención refleja en términos de prevalencia de los diversos elementos la proporción sesgada de los oligonucleótidos al menos parcialmente bicatenarios usados como moléculas de anclaje.

Se reconocerá que los procedimientos de acuerdo con la presente invención pueden usarse para la preparación de ácidos nucleicos monocatenarios, así como ácidos nucleicos bicatenarios. La preparación de un ácido nucleico monocatenario se basa típicamente en la preparación de un ácido nucleico bicatenario, después de lo cual el ácido nucleico bicatenario se desnaturaliza proporcionando los ácidos nucleicos monocatenarios respectivos. La separación de las dos cadenas sencillas puede efectuarse en el caso de los diversos oligonucleótidos usados en la preparación del ácido nucleico bicatenario, consistiendo cada uno en dos cadenas sencillas que no están unidas covalentemente entre sí. Como alternativa, las cadenas sencillas unidas covalentemente que proporcionan las dobles cadenas pueden separarse por escisión covalente.

También se reconocerá que los procedimientos de acuerdo con la presente invención pueden usarse para la preparación de una genoteca de ácido nucleico, por lo que los elementos de la genoteca de ácido nucleico difieren en la secuencia, particularmente en la segunda extensión, y por lo que dicha extensión consiste en más de seis nucleótidos como se ilustra en el presente documento. Como una cuestión de hecho, la repetición de las etapas individuales, es decir, la repetición de las etapas n) a q) como etapas na) a qa), por lo que el quinto oligonucleótido al menos parcialmente bicatenario de la etapa n) es un oligonucleótido al menos bicatenario adicional en la etapa na), el cuarto producto de ligación en la etapa o) es un producto de ligación adicional en la etapa oa), la pluralidad de cuartos oligonucleótidos al menos parcialmente bicatenarios alargados en la etapa p) es una pluralidad de oligonucleótidos al menos parcialmente bicatenarios alargados adicionales en la etapa pa), y el quinto oligonucleótido al menos parcialmente bicatenario acortado de la etapa p) es un oligonucleótido al menos bicatenario adicional acortado en la etapa pa), no está limitando el número de los nucleótidos que son diferentes en los diversos elementos de la genoteca de ácido nucleico. En su lugar, los procedimientos de acuerdo con la presente invención permiten la preparación de extensiones significativamente más largas, dependiendo del fin para el que se prepare la genoteca de ácido nucleico. En el caso de proporcionar un centro activo optimizado de una enzima o un anticuerpo que tenga un sitio de unión optimizado, la longitud de la extensión que es diferente en los diversos miembros de la genoteca de ácido nucleico corresponde a la longitud de la secuencia de aminoácidos que forma el centro activo de dicha enzima o el sitio de unión de dicho anticuerpo, y el ácido nucleico codificante por lo tanto, respectivamente.

Por último, se reconocerá que la longitud de la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico o cualquier extensión de nucleótidos correspondiente a la misma, preferentemente correspondiente a la misma en términos de función, puede consistir, en principio, en cualquier longitud. Preferentemente, consiste en dos nucleótidos, tres nucleótidos, cuatro nucleótidos, cinco nucleótidos, seis nucleótidos, siete nucleótidos u ocho o múltiplos de tres nucleótidos.

Ejemplo 1: Fabricación de una genoteca de ácido nucleico con una sola posición variable

Como se ha resumido anteriormente, la invención proporciona un procedimiento altamente eficaz para la síntesis de una genoteca de ácido nucleico. Un ejemplo para la generación de una genoteca con una sola posición variable como se representa también en la Fig. 1 se describe en más detalle en lo siguiente. La posición variable abarca 20 codones diferentes (en lugar de tres como se representa en las figuras) que representan todos los aminoácidos posibles, el uso de codones se adaptó a *E. coli*.

1) Primera ligación

Se proporcionaron 20-200 pmol de una molécula esplínquer, es decir, un primer oligonucleótido al menos parcialmente bicatenario con un saliente GTG, en un recipiente de reacción en de 25 a 200 µl de tampón de reacción. Después, se añadieron 20-200 pmol de una mezcla de 20 moléculas de anclaje, es decir, la primera genoteca de oligonucleótidos. Estos anclajes están presentes en la mezcla en cantidades equimolares y todos tienen el mismo saliente monocatenario (CAC, correspondiente a la primera extensión) que es complementario al saliente de la molécula esplínquer, pero difieren en su parte variable adyacente que representa los codones para los 20 aminoácidos diferentes (véase la tabla 1). Después de la adición de 10 a 50 U de ADN-Ligasa de T4 las moléculas se ligaron durante 60 minutos a 25°C proporcionando una pluralidad de productos de primera ligación.

2) Aislamiento de una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados

La pluralidad de productos de primera ligación se inmovilizó mediante su resto de biotina en un recipiente o pocillo revestido con estreptavidina durante 15 minutos a temperatura ambiente. El material no unido se eliminó por lavado tres veces con tampón de lavado (Tris/Acetato 33 mM, Acetato de Magnesio 10 mM, Acetato de Potasio 66 mM, pH 7,9). Después, se añadieron de 10 a 100 U de una enzima de restricción tal como Eam1104I, que es la segunda enzima de restricción de tipo IIS, en de 25 a 200 µl de tampón de reacción, seguido de una incubación de 15 a 90 minutos a 37°C. La pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados liberados de este modo se transfirió a un nuevo recipiente o pocillo.

3) Segunda ligación

Se añadieron 20-200 pmol de una mezcla de 20 moléculas de anclaje, es decir, una segunda genoteca de oligonucleótidos, a la pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados en de 25 a 200 µl de tampón de reacción. Las moléculas de anclaje están presentes en la mezcla en cantidades equimolares y difieren en sus salientes monocatenarios (correspondientes a la segunda extensión de la secuencia diana) que son complementarios a los salientes proporcionados por la pluralidad de primeros oligonucleótidos alargados, pero son todas idénticas en su parte adyacente. Después de la adición de 10 a 50 U de ADN-Ligasa de T4, las moléculas se ligaron durante 60 minutos a 25°C, proporcionando una pluralidad de productos de segunda ligación.

4) Aislamiento de una pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados

La pluralidad de productos de segunda ligación se inmovilizó mediante su resto de biotina en un recipiente o pocillo revestido con estreptavidina durante 15 minutos a temperatura ambiente. El material no unido se eliminó por lavado tres veces con tampón de lavado. Después, se añadieron de 10 a 100 U de una enzima de restricción tal como Eam1104I, que es la segunda enzima de restricción de tipo IIS, en de 25 a 200 µl de tampón de reacción, seguido de una incubación de 15 a 90 minutos a 37°C. La pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados así liberados se transfirió a un nuevo recipiente o pocillo.

5) Tercera ligación

Se añadieron 20-200 pmol de una molécula de anclaje, es decir, un cuarto oligonucleótido al menos parcialmente bicatenario con un saliente (GAG, correspondiente a la primera mitad de la primera extensión) que es complementario al saliente proporcionado por la pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados (CTC), a la pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados en de 25 a 200 µl de tampón de reacción. Después de la adición de 10 a 50 U de ADN-Ligasa de T4, las moléculas se ligaron durante 60 minutos a 25°C proporcionando una pluralidad de productos de tercera ligación.

6) Síntesis de un bloque de alargamiento

Un ciclo de alargamiento que comprende la unión del producto de tercera ligación a la superficie de estreptavidina, la escisión del producto inmovilizado por Eam1104I y la ligación de los oligonucleótidos alargados así liberados con un

anclaje apropiado se repitió tres veces partiendo del producto de tercera ligación de la etapa 5, dando como resultado una pluralidad de productos de sexta ligación, también denominados bloques de alargamiento. Las condiciones de cada etapa de reacción son como se han descrito anteriormente.

7) Síntesis del producto final

- 5 El bloque de alargamiento así sintetizado se usó en la síntesis posterior del producto génico completo mediante el procedimiento de transposición descrito anteriormente (Solicitud de Patente Internacional WO 00/75368).

8) Evaluación de calidad del producto final

- 10 El producto final, es decir, una genoteca de ácido nucleico que tenía una sola posición variable, se clonó en un vector convencional (pPCR-Script, Stratagene) y se secuenciaron 96 clones individuales. Las secuencias de los elementos de la genoteca de ácido nucleico deseada parecían ser idénticas excepto por su posición variable particular. La distribución obtenida de los 20 codones diferentes en la posición variable se muestra en la tabla 1.

Tabla 1:

Saliente	Aminoácido	Nº de secuencias
Total		96
GAA	E	3
ATT	I	3
CTT	L	7
GTG	V	3
TTT	F	7
ATG	M	1
TGC	C	7
GCA	A	9
GGC	G	2
CCT	P	9
ACA	T	3

- 15 Las figuras respectivas junto con el análisis estadístico de las mismas se indican en la Fig. 5, que es un histograma de las frecuencias de codones obtenidas en una sola posición permutada. El valor esperado μ de cada uno de los 20 codones en una muestra de 96 es de 4,8, si se asume una distribución igual. En este caso, se espera que más del 95% de todos los codones se encuentren dentro de un intervalo de dos desviaciones típicas ($\mu \pm 2\sigma$). En este ejemplo, todas las frecuencias observadas están bien dentro de un intervalo de 2 desviaciones típicas de una distribución normal, confirmando la hipótesis de una distribución igual de frecuencias de codones.

Ejemplo 2: Fabricación de una genoteca de ácido nucleico con una extensión de posiciones variables

- 20 El ejemplo 2 es un ejemplo para la generación de una genoteca con una extensión de dos posiciones variables, comprendiendo cada una tres nucleótidos posteriores que forman un triplete que codifica un aminoácido que también se representa en la Fig. 2 y se describe en más detalle en lo siguiente. La posición variable contiene 10 codones diferentes, proporcionando cada uno una complejidad de la genoteca de 100 (10 x 10), más que tres codones diferentes como se ilustra en la Fig. 2. El uso de codones se adaptó a *E. coli*.

25 1) Primera ligación

- 30 Se proporcionaron 20-200 pmol de una molécula esplínquer, es decir, un primer oligonucleótido al menos parcialmente bicatenario con un saliente GTG, en un recipiente de reacción en de 25 a 200 μ l de tampón de reacción. Después, se añadieron 20-200 pmol de una mezcla de 10 moléculas de anclaje, es decir, la primera genoteca de oligonucleótidos. Estos anclajes están presentes en la mezcla en cantidades equimolares y todos tienen el mismo saliente monocatenario (CAC, correspondiente a la primera extensión de la secuencia diana) que es complementario al saliente de la molécula esplínquer, pero difieren en su parte variable adyacente que representa los codones para 10 aminoácidos diferentes (véase la tabla 2). Después de la adición de 10 a 50 U de ADN-Ligasa

de T4, las moléculas se ligaron durante 60 minutos a 25°C, proporcionando una pluralidad de productos de primera ligación.

2) Aislamiento de una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados

5 La pluralidad de productos de primera ligación se inmovilizó mediante su resto de biotina en un recipiente o pocillo revestido con estreptavidina durante 15 minutos a temperatura ambiente. El material no unido se eliminó por lavado tres veces con tampón de lavado (Tris/Acetato 33 mM, Acetato de Magnesio 10 mM, Acetato de Potasio 66 mM, pH 7,9). Después, se añadieron de 10 a 100 U de una enzima de restricción tal como Eam1104I, que es la segunda enzima de restricción de tipo IIS, en de 25 a 200 µl de tampón de reacción, seguido de una incubación de 15 a 90 minutos a 37°C. La pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados así liberados se transfirió a un nuevo recipiente o pocillo.

3) Segunda ligación

15 Se añadieron 20-200 pmol de una mezcla de 100 moléculas de anclaje diferentes (10 tripletes como de la primera mitad de la segunda extensión multiplicados por 10 tripletes como de la segunda mitad de la segunda extensión, por lo que la segunda extensión es la extensión de la genoteca de ácido nucleico a preparar que proporciona las posiciones variables), es decir, la segunda genoteca de oligonucleótidos, a la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados en de 25 a 200 µl de tampón de reacción. Todas las moléculas de anclaje están presentes en la mezcla en cantidades equimolares y difieren en sus salientes monocatenarios (que corresponden a la primera mitad de la segunda extensión de la secuencia diana) que son complementarios a los 10 salientes diferentes proporcionados por la pluralidad de primeros oligonucleótidos alargados, así como en su parte adyacente. En la parte adyacente también se proporcionaron 10 codones diferentes, conduciendo a una mezcla combinatoria de 100 anclajes. Después de la adición de 10 a 50 U de ADN-Ligasa de T4, las moléculas se ligaron durante 60 minutos a 25°C, proporcionando una pluralidad de productos de segunda ligación.

4) Aislamiento de una pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados

25 La pluralidad de productos de segunda ligación se inmovilizó mediante su resto de biotina en un recipiente o pocillo revestido con estreptavidina durante 15 minutos a temperatura ambiente. El material no unido se eliminó por lavado tres veces con tampón de lavado. Después, se añadieron de 10 a 100 U de una enzima de restricción tal como Eam1104I, que es la segunda enzima de restricción de tipo IIS, en de 25 a 200 µl de tampón de reacción, seguido de una incubación de 15 a 90 minutos a 37°C. La pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados así liberados se transfirió a un nuevo recipiente o pocillo.

5) Tercera ligación

35 Se añadieron 20-200 pmol de una mezcla de 10 moléculas de anclaje, es decir, la tercera genoteca de oligonucleótidos, a la pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados en de 25 a 200 µl de tampón de reacción. Las moléculas de anclaje están presentes en la mezcla en cantidades equimolares y difieren en sus salientes monocatenarios (que corresponden a la segunda mitad de la segunda extensión de la secuencia diana) que son complementarios a los salientes proporcionados por la pluralidad de primeros oligonucleótidos alargados, pero son idénticas en su parte adyacente que constituye la tercera extensión de la secuencia diana. Después de la adición de 10 a 50 U de ADN-Ligasa de T4, las moléculas se ligaron durante 40 60 minutos a 25°C proporcionando un producto de tercera ligación que en realidad consiste en una pluralidad de productos de tercera ligación.

6) Aislamiento de una pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados

45 La pluralidad de productos de tercera ligación se inmovilizó mediante su resto de biotina en un recipiente o pocillo revestido con estreptavidina durante 15 minutos a temperatura ambiente. El material no unido se eliminó por lavado tres veces con tampón de lavado. Después, se añadieron de 10 a 100 U de una enzima de restricción tal como Eam1104I, que es la segunda enzima de restricción de tipo IIS, en de 25 a 200 µl de tampón de reacción, seguido de una incubación de 15 a 90 minutos a 37°C. La pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados así liberados se transfirió a un nuevo recipiente o pocillo.

7) Cuarta ligación

50 Se añadieron 20-200 pmol de una molécula de anclaje, es decir, el quinto oligonucleótido al menos parcialmente bicatenario con un saliente (ACA, que corresponde a la tercera extensión de la secuencia diana) que es complementario al saliente proporcionado por la pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados (TGT), a la pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados en de 25 a 200 µl de tampón de reacción. Después de la adición de 10 a 50 U de ADN-Ligasa de T4, las moléculas se ligaron durante 60 minutos a 25°C, proporcionando un cuarto producto de ligación que en realidad

consiste en una pluralidad de cuartos productos de ligación.

8) Síntesis de un bloque de alargamiento

5 Este ciclo de alargamiento que comprende la unión del producto de ligación a la superficie de estreptavidina, la escisión del producto inmovilizado por Eam11041 y la ligación de los oligonucleótidos alargados así liberados con un anclaje apropiado se repitió dos veces, dando como resultado una pluralidad de productos de sexta ligación, también denominados bloques de alargamiento. Las condiciones de cada etapa de reacción eran como se han descrito anteriormente.

9) Síntesis del producto final

10 El bloque de alargamiento así sintetizado se usó en la síntesis posterior del producto génico completo mediante el procedimiento de transposición descrito anteriormente (Solicitud de Patente Internacional WO 00/75368).

10) Evaluación de calidad del producto final

15 El producto final, es decir, una genoteca de ácido nucleico que tiene una extensión variable que comprende dos tripletes, por lo que cada triplete está codificando uno de un total de 10 aminoácidos posibles en la posición respectiva, se clonó en un vector convencional y se secuenciaron 62 clones individuales. Las secuencias parecían ser idénticas excepto por sus dos posiciones variables particulares. La distribución obtenida de los 10 codones diferentes en las posiciones variables se muestra en la Tabla 2.

Tabla 2:

Tripletes	Triplete Variable 1	Triplete Variable 2
AAA	3	5
ATG	9	5
ACT	8	5
CCC	6	7
CCG	10	7
CTG	9	10
GAT	7	2
GCA	4	7
TTA	4	7
TGG	2	7
Total	62	62

20 Las figuras respectivas junto con el análisis estadístico de las mismas se indican en la Fig. 6, que es un histograma de las frecuencias de codones obtenidas en dos tripletes consecutivos. El valor esperado μ de cada uno de los 10 codones en una muestra de 62 es de 6,2, si se asume una distribución igual. Examinando ambos tripletes por igual, todas las frecuencias observadas están bien dentro de un intervalo de 2 desviaciones típicas de una distribución normal. Este resultado confirma la hipótesis de una distribución igual puesto que en este caso más del 95% de las frecuencias deberían estar presentes dentro de este intervalo. La hipótesis se confirma por el resultado de que 13 de las 20 frecuencias (65%) se encuentran dentro de un intervalo de 1 desviación típica de una distribución igual, satisfaciendo de nuevo por lo tanto los requisitos estadísticos de una distribución igual.

Ejemplo 3: Fabricación de una genoteca de ácido nucleico con una sola posición variable con proporciones definidas

30 Un ejemplo para la generación de una genoteca con una sola posición variable por lo que los codones aparecen en proporciones definidas se describe en más detalle en lo siguiente. Hasta ahora, el presente ejemplo ilustra una realización del procedimiento de acuerdo con el primer aspecto de la presente invención. La posición variable contiene 10 codones diferentes. El uso de codones se adaptó a *E. coli*.

1) Primera ligación

En una primera reacción (R1), se proporcionaron 20-200 pmol de una molécula esplínquer, es decir, un primer

oligonucleótido al menos parcialmente bicatenario con un saliente GCC (en dirección 5' -> 3', correspondiente a la primera extensión de la secuencia diana) en un recipiente de reacción en de 25 a 200 µl de tampón de reacción. Después, se añadieron 20-200 pmol de una mezcla de 10 moléculas de anclaje, es decir, la primera genoteca de oligonucleótidos. Estos anclajes tienen todos el mismo saliente monocatenario (GGC; en dirección 5' -> 3') que es complementario al saliente de la molécula esplínquer, pero difieren en su parte variable adyacente que representa los codones para 10 aminoácidos diferentes (véase la tabla 1). Todas las moléculas de anclaje estaban presentes en cantidades equimolares (cada anclaje el 10% de la cantidad total). Después de la adición de 10 a 50 U de ADN-Ligasa de T4, las moléculas se ligaron durante 60 minutos a 25°C, proporcionando una pluralidad de productos de primera ligación.

5 En una reacción paralela (**R2**), de nuevo se proporcionaron 20-200 pmol de una molécula esplínquer, es decir, un primer oligonucleótido al menos parcialmente bicatenario con un saliente GCC (en dirección 5' -> 3', correspondiente a la primera extensión de la secuencia diana) en un recipiente de reacción en de 25 a 200 µl de tampón de reacción. Después, se añadieron 20-200 pmol de una mezcla de 10 moléculas de anclaje, es decir, la primera genoteca de oligonucleótidos. Estos anclajes tienen todos también el mismo saliente monocatenario (GGC; en dirección 5' -> 3') que es complementario al saliente de la molécula esplínquer, pero difieren en su parte variable adyacente que representa los codones para 10 aminoácidos diferentes (véase la tabla 1). Al contrario que en la primera reacción, esta vez una de las moléculas de anclaje (que conduce a un saliente TTT después de la restricción con Eam1104I, 30% de la cantidad total) está presente en exceso respecto a los otros nueve anclajes (7,8% de la cantidad total de cada). Después de la adición de 10 a 50 U de ADN-Ligasa de T4, las moléculas se ligaron durante 60 minutos a 25°C proporcionando una pluralidad de productos de primera ligación.

2) Aislamiento de una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados

La pluralidad de productos de primera ligación se inmovilizó mediante su resto de biotina en un recipiente o pocillo revestido con estreptavidina durante 15 minutos a temperatura ambiente. El material no unido se eliminó por lavado tres veces con tampón de lavado (Tris/Acetato 33 mM, Acetato de Magnesio 10 mM, Acetato de Potasio 66 mM, pH 7,9). Después se añadieron de 10 a 100 U de una enzima de restricción tal como Eam1104I, que es la segunda enzima de restricción de tipo IIS, en de 25 a 200 µl de tampón de reacción, seguido de una incubación de 15 a 90 minutos a 37°C. La prioridad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados así liberados se transfirió a un nuevo recipiente o pocillo.

Esta reacción se procesó en paralelo para tanto **R1** como **R2**.

3) Segunda ligación

Se añadieron 20-200 pmol de una mezcla de 10 moléculas de anclaje, es decir, la segunda genoteca de oligonucleótidos, a la primera reacción (**R1**) que contenía la pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados, en de 25 a 200 µl de tampón de reacción. Las moléculas de anclaje difieren en sus salientes monocatenarios (que corresponden a la segunda extensión de la secuencia diana) que son complementarios a los salientes proporcionados por la pluralidad de primeros oligonucleótidos alargados, pero son todas idénticas en su parte adyacente. Todas las moléculas de anclaje estaban presentes en cantidades equimolares (cada anclaje el 10% de la cantidad total). Después de la adición de 10 a 50 U de ADN-Ligasa de T4, las moléculas se ligaron durante 60 minutos a 25°C proporcionando una pluralidad de productos de segunda ligación.

40 En paralelo, se añadieron 20-200 pmol de una mezcla de 10 moléculas de anclaje, es decir, la segunda genoteca de oligonucleótidos a la segunda reacción (**R2**) que contenía la pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados en de 25 a 200 µl de tampón de reacción. Las moléculas de anclaje difieren en sus salientes monocatenarios (que corresponden a la segunda extensión de la secuencia diana), que son complementarios a los salientes proporcionados por la pluralidad de primeros oligonucleótidos alargados, pero son todas idénticas en su parte adyacente. Una de las moléculas de anclaje contenía un saliente (AAA) que es complementario al saliente proporcionado por el miembro de la pluralidad de los primeros oligonucleótidos alargados, que estaba presente en una cantidad del 30% de la cantidad total en la primera ligación, y este anclaje también estaba presente en una cantidad del 30% de la segunda genoteca de oligonucleótidos. Los otros nueve anclajes estaban presentes en una cantidad del 7,8% cada uno. Después de la adición de 10 a 50 U de ADN-Ligasa de T4, las moléculas se ligaron durante 60 minutos a 25°C proporcionando una pluralidad de productos de segunda ligación.

4) Aislamiento de una pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados

La pluralidad de productos de segunda ligación se inmovilizó mediante su resto de biotina en un recipiente o pocillo revestido con estreptavidina durante 15 minutos a temperatura ambiente. El material no unido se eliminó por lavado tres veces con tampón de lavado. Después, se añadieron de 10 a 100 U de una enzima de restricción tal como Eam1104I, que es la segunda enzima de restricción de tipo IIS, en de 25 a 200 µl de tampón de reacción, seguido de una incubación de 15 a 90 minutos a 37°C. La pluralidad de segundos oligonucleótidos al menos parcialmente

bicatenarios alargados así liberados se transfirió a un nuevo recipiente o pocillo.

La reacción se procesó en paralelo para tanto **R1** como **R2**.

5) Tercera ligación

5 Se añadieron 20-200 pmol de una molécula de anclaje con un saliente que es complementario al saliente proporcionado por la pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados, y que es el cuarto oligonucleótido al menos parcialmente bicatenario, a la pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados en de 25 a 200 µl de tampón de reacción. Después de la adición de 10 a 50 U de ADN-Ligasa de T4, las moléculas se ligaron durante 60 minutos a 25°C proporcionando una pluralidad de productos de tercera ligación.

10 La reacción se procesó en paralelo para tanto **R1** como **R2**.

6) Síntesis de un bloque de alargamiento

15 Este ciclo de alargamiento que comprende la unión de los productos de ligación a la superficie de estreptavidina, la escisión de los productos inmovilizados por Eam1104l y ligación de los oligonucleótidos alargados así liberados con un anclaje apropiado se repitió tres veces, dando como resultado una pluralidad de productos de sexta ligación, también denominados bloques de alargamiento. Las condiciones de cada etapa de reacción eran como se han descrito anteriormente.

Las reacciones se procesaron en paralelo para tanto **R1** como **R2**.

7) Síntesis del producto final

20 Los bloques de alargamiento así sintetizados se usaron en la síntesis posterior de los productos génicos completos, es decir, la genoteca de ácido nucleico a preparar, mediante el procedimiento de transposición descrito anteriormente (Solicitud de Patente Internacional WO 00/75368).

8) Evaluación de calidad del producto final

25 Los productos finales, es decir, las genotecas de ácido nucleico que tienen una sola posición variable, por lo que 10 codones diferentes pueden aparecer en proporciones equimolares (R1) o no equimolares o sesgadas (R2), se clonaron en un vector convencional y se secuenciaron 96 clones individuales de cada. Como se usa preferentemente en el presente documento, el término sesgado se refiere a una mezcla que comprende diversos tipos de moléculas, por lo que uno o varios de los tipos de moléculas están contenidos en una proporción no equimolar en comparación con uno o varios otros de dichos diversos tipos de moléculas. Las secuencias parecían ser idénticas excepto por su posición variable particular. La distribución obtenida de los 10 codones diferentes en la
30 posición variable se muestra en la Tabla 3.

Tabla 3:

Tripletes	R 1	R 2
AAA	12	30
ATG	4	5
ACT	7	9
CCC	9	6
CCG	11	10
CTG	14	6
GAT	9	8
GCA	8	7
TTA	15	4
TGG	7	11
Total	96	96

Las figuras respectivas junto con el análisis estadístico de las mismas se indican en las Figs. 7 y 8, que son histogramas de frecuencias de codones obtenidas en una sola posición variable. La Fig. 7 muestra la aparición de los diferentes codones dentro de la genoteca que tienen la distribución equimolar (R1), la Fig. 8 muestra la aparición de los diferentes codones dentro de la genoteca en la que el triplete AAA está presente en un exceso de tres veces respecto a los otros codones (R2).

En R1, el valor esperado μ de cada uno de los 10 codones en una muestra de 96 es 9,6, si se asume una distribución igual. En este caso, se espera que más del 95% de todos los codones se encuentren dentro de un intervalo de dos desviaciones típicas ($\mu \pm 2\sigma$). En este ejemplo, todas las frecuencias observadas están dentro de un intervalo de 2 desviaciones típicas de una distribución normal, confirmando la hipótesis de una distribución igual de frecuencias de codones. La hipótesis se confirma por el resultado de que 7 de las 10 frecuencias (70%) se encuentran dentro de un intervalo de 1 desviación típica de una distribución igual, satisfaciendo de nuevo por lo tanto los requisitos estadísticos de una distribución igual (más del 65% de las frecuencias dentro de un intervalo de 1 desviación típica).

En R2, el valor esperado μ para el codón AAA en una muestra de 96 es de 28,8, si se asume que la distribución observada sigue las especificaciones aplicadas. En la misma medida, el valor esperado para los otros nueve codones es de 7,5. En el caso de una distribución especificada se espera que más del 95% de todos los codones se encuentren dentro de un intervalo de dos desviaciones típicas ($\mu \pm 2\sigma$). En este ejemplo, todas las frecuencias observadas están dentro de un intervalo de 2 desviaciones típicas lejos del valor esperado respectivo, confirmando la hipótesis de la aparición de la distribución de frecuencias de codones especificada. La hipótesis se confirma por el resultado de que 8 de las 10 frecuencias (80%) se encuentran dentro de un intervalo de 1 desviación típica de la distribución especificada, satisfaciendo de nuevo por lo tanto los requisitos estadísticos de una distribución especificada (más del 65% de las frecuencias dentro de un intervalo de 1 desviación típica).

Las características de la presente invención desveladas en la memoria descriptiva, el listado de secuencias, las reivindicaciones y/o los dibujos pueden tanto por separado como en cualquier combinación de las mismas ser material para realizar la invención en diversas formas de la misma.

LISTADO DE SECUENCIAS

<110> Sloning BioTechnology GMBH

<120> Procedimiento para la preparación de una genoteca de ácido nucleico

<130> S 10040 EP-EP

<150> EP 08 006 472.8

<151> 31-03-2008

<160> 24

<170> PatentIn versión 3.3

<210> 1

<211> 12

<212> ADN

<213> Artificial

<220>

<223> sintético

<400> 1

cacatggaga ca 12

<210> 2

<211> 12

<212> ADN

<213> Artificial

<220>

<223> sintético

<400> 2

cactatgaga ca 2

ES 2 378 790 T3

	<210> 3	
	<211> 12	
	<212> ADN	
	<213> Artificial	
5		
	<220>	
	<223> sintético	
10	<400> 3	
	cacgaggaga ca	12
	<210> 4	
	<211> 12	
	<212> ADN	
	<213> Artificial	
15		
	<220>	
	<223> sintético	
20	<400> 4	
	tgtctcatg tg	12
	<210> 5	
	<211> 12	
	<212> ADN	
	<213> Artificial	
25		
	<220>	
	<223> sintético	
30	<400> 5	
	tgtctcatag tg	12
	<210> 6	
	<211> 12	
	<212> ADN	
	<213> Artificial	
35		
	<220>	
	<223> sintético	
40	<400> 6	
	tgtctcctcg tg	12
	<210> 7	
	<211> 12	
	<212> ADN	
	<213> Artificial	
45		
	<220>	
	<223> sintético	
50	<400> 7	
	caatgctca ca	12
	<210> 8	
	<211> 12	
	<212> ADN	
	<213> Artificial	
55		
	<220>	
	<223> sintético	
60	<400> 8	
	cactatctca ca	12
65		

	<210> 9	
	<211> 12	
	<212> ADN	
5	<213> Artificial	
	<220>	
	<223> sintético	
10	<400> 9 cacgagctca ca	12
	<210> 10	
	<211> 12	
15	<212> ADN	
	<213> Artificial	
	<220>	
	<223> sintético	
20	<400> 10 tgtgagcatg tg	12
	<210> 11	
	<211> 12	
25	<212> ADN	
	<213> Artificial	
	<220>	
	<223> sintético	
30	<400> 11 tgtgagatag tg	12
	<210> 12	
	<211> 12	
35	<212> ADN	
	<213> Artificial	
	<220>	
40	<223> sintético	
	<400> 12 tgtgagctcg tg	12
45	<210> 13	
	<211> 12	
	<212> ADN	
	<213> Artificial	
50	<220>	
	<223> sintético	
	<400> 13 cacatggaga ca	12
55	<210> 14	
	<211> 12	
	<212> ADN	
	<213> Artificial	
60	<220>	
	<223> sintético	
	<400> 14 cactatgaga ca	12
65		

ES 2 378 790 T3

	<210> 15	
	<211> 12	
	<212> ADN	
	<213> Artificial	
5		
	<220>	
	<223> sintético	
10	<400> 15	
	cacgaggaga ca	12
	<210> 16	
	<211> 12	
	<212> ADN	
15	<213> Artificial	
	<220>	
	<223> sintético	
20	<400> 16	
	tgtctcatg tg	12
	<210> 17	
	<211> 12	
25	<212> ADN	
	<213> Artificial	
	<220>	
	<223> sintético	
30	<400> 17	
	tgtctcatag tg	12
	<210> 18	
35	<211> 12	
	<212> ADN	
	<213> Artificial	
	<220>	
40	<223> sintético	
	<400> 18	
	tgtctcctcg tg	12
45	<210> 19	
	<211> 12	
	<212> ADN	
	<213> Artificial	
50	<220>	
	<223> sintético	
	<400> 19	
55	cacatgttta ca	12
	<210> 20	
	<211> 12	
	<212> ADN	
60	<213> Artificial	
	<220>	
	<223> sintético	
65	<400> 20	
	cactatttta ca	12

ES 2 378 790 T3

5
<210> 21
<211> 12
<212> ADN
<213> Artificial

<220>
<223> sintético

10
<400> 21
cacgagtta ca 12

<210> 22
<211> 12
<212> ADN
15
<213> Artificial

<220>
<223> sintético

20
<400> 22
tgtaaactg tg 12

<210> 23
<211> 12
25
<212> ADN
<213> Artificial

<220>
<223> sintético
30

<400> 23
tgtaaactg tg 12

<210> 24
<211> 12
35
<212> ADN
<213> Artificial

<220>
40
<223> sintético

<400> 24
tgtaaactg tg 12

REIVINDICACIONES

1. Un procedimiento para preparar una genoteca de ácido nucleico que comprende una pluralidad de elementos, por lo que cada elemento de dicha genoteca de ácido nucleico comprende una primera extensión de nucleótidos, una segunda extensión de nucleótidos y una tercera extensión de nucleótidos, por lo que la secuencia de la primera extensión de nucleótidos y de la tercera extensión de nucleótidos son idénticas en cada elemento de la genoteca de ácido nucleico y los elementos de la genoteca de ácido nucleico difieren en la secuencia de la segunda extensión de nucleótidos, por lo que el procedimiento comprende las etapas siguientes:

a) proporcionar un primer oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria que tiene un saliente monocatenario, por lo que el saliente monocatenario proporciona una secuencia de nucleótidos correspondiente a la secuencia de la primera o tercera extensión de nucleótidos de los elementos de la genoteca de ácido nucleico, por lo que el oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una primera enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento,

b) proporcionar una primera genoteca de oligonucleótidos que comprende varios miembros, por lo que cada miembro es un segundo oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria, por lo que el segundo oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una segunda enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento, y un saliente monocatenario, por lo que el saliente monocatenario o parte del mismo es el mismo para los miembros de la primera genoteca de oligonucleótidos y por lo que dicho saliente monocatenario es esencialmente complementario al saliente monocatenario del primer oligonucleótido al menos parcialmente bicatenario, por lo que la estructura bicatenaria comprende una extensión de nucleótidos, por lo que los miembros de la primera genoteca de oligonucleótidos difieren en la secuencia de dicha extensión de nucleótidos y la secuencia de dicha extensión de nucleótidos de los miembros de la primera genoteca de oligonucleótidos corresponde a la secuencia, o parte de la misma, de la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico,

c) combinar el primer oligonucleótido al menos parcialmente bicatenario y la primera genoteca de oligonucleótidos y ligar una molécula del primer oligonucleótido al menos parcialmente bicatenario con una molécula de cada miembro de la primera genoteca de oligonucleótidos mediante sus salientes monocatenarios, después de lo cual se forma un producto de primera ligación,

d) cortar el producto de primera ligación con dicha segunda enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico del miembro de la primera genoteca de oligonucleótidos, por lo que dicha escisión proporciona una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados y segundos oligonucleótidos al menos parcialmente bicatenarios acortados,

e) eliminar los segundos oligonucleótidos al menos parcialmente bicatenarios acortados,

f) proporcionar una segunda genoteca de oligonucleótidos que comprende varios miembros, por lo que cada miembro es un tercer oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria, por lo que el tercer oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una tercera enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento y un saliente monocatenario, por lo que el saliente monocatenario o parte del mismo es diferente para los miembros de la segunda genoteca de oligonucleótidos, y por lo que dichos salientes monocatenarios o parte de los mismos del tercer oligonucleótido son esencialmente complementarios a los salientes monocatenarios de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados, y por lo que los miembros de la segunda genoteca de oligonucleótidos comprenden una extensión de oligonucleótidos en la estructura bicatenaria que es idéntica en todos los miembros y corresponde a la secuencia, o parte de la misma, de la tercera o primera extensión de nucleótidos de los elementos de la genoteca de ácido nucleico,

g) combinar la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados y la segunda genoteca de oligonucleótidos y ligar cada molécula de la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados con una molécula de cada uno de los miembros de la segunda genoteca de oligonucleótidos, por lo que el saliente monocatenario del primer oligonucleótido al menos parcialmente bicatenario alargado es esencialmente complementario al saliente del tercer oligonucleótido al menos parcialmente bicatenario, después de lo cual se forma un producto de segunda ligación,

h) cortar el producto de segunda ligación con la tercera de enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico de los terceros oligonucleótidos al menos parcialmente bicatenarios, proporcionando una pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados y terceros oligonucleótidos al menos parcialmente bicatenarios acortados,

i) eliminar los terceros oligonucleótidos al menos parcialmente bicatenarios acortados,

j) proporcionar un cuarto oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria que tiene un saliente monocatenario, por lo que el saliente monocatenario proporciona una secuencia de nucleótidos esencialmente complementaria al saliente monocatenario del segundo oligonucleótido al menos parcialmente bicatenario alargado, por lo que el oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una cuarta enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento,

k) combinar el cuarto oligonucleótido al menos parcialmente bicatenario y la pluralidad de los segundos oligonucleótidos al menos parcialmente bicatenarios alargados y ligar una molécula del cuarto oligonucleótido y cada molécula de la pluralidad de los segundos oligonucleótidos al menos parcialmente bicatenarios alargados formando un producto de tercera ligación;

l) cortar el producto de tercera ligación con la cuarta enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico del cuarto oligonucleótido u oligonucleótidos al menos parcialmente bicatenarios, proporcionando una pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados y cuarto oligonucleótido al menos parcialmente bicatenario acortado, y
 m) eliminar el cuarto oligonucleótido al menos parcialmente bicatenario acortado.

2. El procedimiento de acuerdo con la reivindicación 1, por lo que el procedimiento comprende además las etapas de:

n) repetir las etapas j) a m) como etapas ja) a ma), por lo que el cuarto oligonucleótido al menos parcialmente bicatenario de la etapa j) es un oligonucleótido al menos parcialmente bicatenario adicional en la etapa ja), el producto de tercera ligación en la etapa k) es un producto de ligación adicional en la etapa ka), la pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados en la etapa l) es una pluralidad de oligonucleótidos al menos parcialmente bicatenarios alargados adicionales en la etapa la), y el cuarto oligonucleótido al menos parcialmente bicatenario acortado de la etapa l) es un oligonucleótido al menos parcialmente bicatenario acortado adicional en la etapa la).

3. El procedimiento de acuerdo con la reivindicación 2, por lo que después de la etapa de ligación ka), el producto de ligación adicional se escinde, por lo que la escisión se produce en la secuencia de ácido nucleico del oligonucleótido al menos parcialmente bicatenario alargado adicional.

4. Un procedimiento para preparar una genoteca de ácido nucleico que comprende una pluralidad de elementos, por lo que cada elemento de dicha genoteca de ácido nucleico comprende una primera extensión de nucleótidos, una segunda extensión de nucleótidos y una tercera extensión de nucleótidos, por lo que la secuencia de la primera extensión de nucleótidos y de la tercera extensión de nucleótidos son idénticas en cada elemento de la genoteca de ácido nucleico y los elementos de la genoteca de ácido nucleico difieren en la secuencia de la segunda extensión de nucleótidos, por lo que el procedimiento comprende las etapas siguientes:

a) proporcionar un primer oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria que tiene un saliente monocatenario, por lo que el saliente monocatenario proporciona una secuencia de nucleótidos correspondiente a la secuencia de la primera o tercera extensión de nucleótidos de los elementos de la genoteca de ácido nucleico, por lo que el oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una primera enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento,

b) proporcionar una primera genoteca de oligonucleótidos que comprende varios miembros, por lo que cada miembro es un segundo oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria, por lo que el oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una segunda enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento, y un saliente monocatenario, por lo que el saliente monocatenario o parte del mismo es el mismo para los miembros de la primera genoteca de oligonucleótidos, y por lo que dicho saliente monocatenario es esencialmente complementario al saliente monocatenario del primer oligonucleótido al menos parcialmente bicatenario, por lo que la estructura bicatenaria comprende una extensión de nucleótidos por lo que los miembros de la primera genoteca de oligonucleótidos difieren en la secuencia de dicha extensión de nucleótidos y la secuencia de dicha extensión de nucleótidos de los miembros de la primera genoteca de oligonucleótidos corresponde a la secuencia, o parte de la misma, de la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico,

c) combinar el primer oligonucleótido al menos parcialmente bicatenario y la primera genoteca de oligonucleótidos y ligar una molécula del primer oligonucleótido al menos parcialmente bicatenario con una molécula de cada miembro de la primera genoteca de oligonucleótidos mediante sus salientes monocatenarios, después de lo cual se forma un producto de primera ligación,

d) cortar el producto de primera ligación con dicha segunda enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico del miembro de la primera genoteca de oligonucleótidos, por lo que dicha escisión proporciona una pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados y segundos oligonucleótidos al menos parcialmente bicatenarios acortados, por lo que dicha pluralidad de primeros oligonucleótidos al menos parcialmente bicatenarios alargados consiste en diversas especies moleculares que difieren en la secuencia del saliente monocatenario,

e) eliminar los segundos oligonucleótidos al menos parcialmente bicatenarios acortados,

f) proporcionar una segunda genoteca de oligonucleótidos que comprende varios subgrupos de oligonucleótidos, y comprendiendo cada subgrupo varios miembros, por lo que cada miembro es un tercer oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria, por lo que el tercer oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una tercera enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento y un saliente monocatenario, por lo que el saliente monocatenario o parte del mismo es diferente de los diversos subgrupos de oligonucleótidos de la segunda genoteca de oligonucleótidos, y por lo que el saliente monocatenario, o parte del mismo, de cada subgrupo de la segunda genoteca de oligonucleótidos es esencialmente complementario al saliente monocatenario de cada especie molecular de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados, y por lo que los miembros de la segunda genoteca de oligonucleótidos comprenden una extensión de nucleótidos en la estructura bicatenaria que es diferente en los miembros de cada subgrupo de la segunda genoteca de oligonucleótidos y corresponde a la secuencia, o parte de la misma, de la segunda extensión de nucleótidos de

los elementos de la genoteca de ácido nucleico,

g) combinar la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados y la segunda genoteca de oligonucleótidos y ligar cada molécula de la pluralidad de los primeros oligonucleótidos al menos parcialmente bicatenarios alargados con cada molécula de los miembros de la segunda genoteca de oligonucleótidos, por lo que se ligan aquellas moléculas cuyos salientes son esencialmente complementarios entre sí, después de lo cual se forman productos de segunda ligación,

h) cortar los productos de segunda ligación con la tercera enzima de restricción de tipo II S, por lo que la escisión se produce en la secuencia de ácido nucleico del tercer oligonucleótido al menos parcialmente bicatenario, proporcionando una pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados y terceros oligonucleótidos al menos parcialmente bicatenarios acortados, por lo que dicha pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados consiste en diversos subgrupos que difieren en sus salientes monocatenarios,

i) eliminar los terceros oligonucleótidos al menos parcialmente bicatenarios acortados,

j) proporcionar una tercera genoteca de oligonucleótidos que comprende varios miembros, por lo que cada miembro es un cuarto oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria que tiene un saliente monocatenario, por lo que el saliente monocatenario proporciona una secuencia de nucleótidos esencialmente complementaria a los salientes monocatenarios de los segundos oligonucleótidos al menos parcialmente bicatenarios alargados y una extensión, preferentemente adyacente a dicho saliente, que es idéntica en los diversos miembros y proporciona la tercera o primera extensión de nucleótidos de los elementos de la genoteca de ácido nucleico, por lo que los miembros de la tercera genoteca de oligonucleótidos difieren en sus salientes monocatenarios y los salientes corresponden a la segunda extensión, o parte de la misma, de la genoteca de ácido nucleico, por lo que los oligonucleótidos comprenden un sitio de reconocimiento, o parte del mismo, para una cuarta enzima de restricción de tipo IIS, que corta fuera de su sitio de reconocimiento,

k) combinar la pluralidad de segundos oligonucleótidos al menos parcialmente bicatenarios alargados y la tercera genoteca de oligonucleótidos y ligar cada molécula de los miembros de la tercera genoteca de oligonucleótidos con cada molécula de la pluralidad de los segundos oligonucleótidos al menos parcialmente bicatenarios alargados, por lo que se ligan aquellas moléculas cuyos salientes son esencialmente complementarios entre sí, después de lo cual se forma un producto de tercera ligación;

l) cortar el producto de tercera ligación con la cuarta enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico del miembro de la tercera genoteca de oligonucleótidos, proporcionando una pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados y cuartos oligonucleótidos al menos parcialmente bicatenarios acortados,

m) opcionalmente eliminar los cuartos oligonucleótidos al menos parcialmente bicatenarios acortados,

n) proporcionar un quinto oligonucleótido al menos parcialmente bicatenario que tiene una estructura bicatenaria que tiene un saliente monocatenario, por lo que el saliente monocatenario proporciona una secuencia de nucleótidos esencialmente complementaria al saliente monocatenario de la pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados, por lo que el quinto oligonucleótido comprende un sitio de reconocimiento, o parte del mismo, para una quinta enzima de restricción de tipo IIS que corta fuera de su sitio de reconocimiento,

o) combinar la pluralidad de terceros oligonucleótidos al menos parcialmente bicatenarios alargados y la cuarta genoteca de oligonucleótidos y ligar cada molécula de la pluralidad del tercer oligonucleótido al menos parcialmente bicatenario alargado con una molécula del quinto oligonucleótido al menos bicatenario, después de lo cual se forma un producto de cuarta ligación;

p) cortar el producto de cuarta ligación con la quinta enzima de restricción de tipo IIS, por lo que la escisión se produce en la secuencia de ácido nucleico del quinto oligonucleótido al menos parcialmente bicatenario, proporcionando una pluralidad de cuartos oligonucleótidos al menos parcialmente bicatenarios alargados y el quinto oligonucleótido al menos parcialmente bicatenario acortado,

q) opcionalmente eliminar el quinto oligonucleótido al menos parcialmente bicatenario acortado.

5. El procedimiento de acuerdo con la reivindicación 4, por lo que el procedimiento comprende además las etapas de:

r) repetir las etapas n) a q) como etapas na) a qa), por lo que el quinto oligonucleótido al menos parcialmente bicatenario de la etapa n) es un oligonucleótido al menos bicatenario adicional en la etapa na), el producto de cuarta ligación en la etapa o) es un producto de ligación adicional en la etapa oa), la pluralidad de cuartos oligonucleótidos al menos parcialmente bicatenarios alargados de la etapa p) es una pluralidad de oligonucleótidos al menos parcialmente bicatenarios alargados adicionales en la etapa pa) y el quinto oligonucleótido al menos parcialmente bicatenario acortado de la etapa p) es un oligonucleótido al menos parcialmente bicatenario acortado adicional en la etapa pa).

6. El procedimiento de acuerdo con la reivindicación 5, por lo que después de la etapa de ligación oa) el producto de ligación adicional se escinde, por lo que la escisión se produce en la secuencia de ácido nucleico del oligonucleótido al menos parcialmente bicatenario alargado adicional.

7. El procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 6, en el que la primera enzima de restricción de tipo IIS se selecciona del grupo que comprende Bpil, BbsI, Esp3I, BsmBI, Eco31I, BsaI, BfuAI, FokI,

BseRI, BbvI y BsgI.

8. El procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 7, en el que la segunda, tercera, cuarta, quinta y cualquier enzima de restricción de tipo IIS adicional se selecciona cada una e individualmente del grupo que comprende EarI, Eam1104I, SapI, LguI, Ksp632I y BspQI.
- 5 9. El procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 8, en el que el ácido nucleico es un ácido nucleico monocatenario y la longitud de la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico consiste en dos nucleótidos, tres nucleótidos, cuatro nucleótidos, cinco nucleótidos, seis nucleótidos, siete nucleótidos, ocho nucleótidos o múltiplos de tres nucleótidos.
- 10 10. El procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 9, en el que el ácido nucleico es un ácido nucleico bicatenario y la longitud de la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico consiste en dos nucleótidos, tres nucleótidos, cuatro nucleótidos, cinco nucleótidos, seis nucleótidos, siete nucleótidos, ocho nucleótidos o múltiplos de tres nucleótidos.
11. El procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 10, en el que la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico está codificando uno o varios aminoácidos.
- 15 12. El procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 11, en el que el segundo oligonucleótido al menos parcialmente bicatenario, el tercer oligonucleótido al menos parcialmente bicatenario, el cuarto oligonucleótido al menos parcialmente bicatenario, el quinto y el oligonucleótido al menos parcialmente bicatenario adicional comprenden una modificación, por lo que dicha modificación permite la inmovilización del oligonucleótido en una superficie.
- 20 13. El procedimiento de acuerdo con la reivindicación 12, por lo que la modificación se selecciona del grupo que comprende un resto biotina, un resto de digoxigenina, un resto de isotiocianato de fluoresceína, un compuesto amino o un éster de succinilo.
- 25 14. El procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 11, en el que la genoteca de ácido nucleico es una genoteca de moléculas de ácido nucleico que codifica un ácido nucleico funcional, por lo que dicho ácido nucleico se selecciona del grupo que comprende aptámeros, promotores, ribozimas y moléculas de mediación de iARN.
15. El procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 14, en el que la genoteca de ácido nucleico es una genoteca de moléculas de ácido nucleico que codifican un péptido, polipéptido o proteína.
- 30 16. El procedimiento de acuerdo con la reivindicación 15, en el que el péptido, polipéptido y proteína se seleccionan del grupo que comprende aptámeros peptídicos, enzimas, enzimas de restricción, dominios de unión a ADN, vacunas, anticuerpos, proteínas farmacéuticamente activas, anticualinas, DARPinas, nanocuerpos y AdNectinas.
- 35 17. El procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 13, en el que la genoteca de ácido nucleico es una genoteca de moléculas de ácido nucleico que codifican un polipéptido, por lo que el polipéptido es un armazón que comprende una o varias regiones constantes, por lo que una o varias regiones variables y una o varias de las regiones variables están codificadas por la secuencia de la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico.
18. El procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 17, en el que la segunda extensión de nucleótidos de los elementos de la genoteca de ácido nucleico es la única diferencia entre dichos elementos.
- 40 19. Un procedimiento para la fabricación de una genoteca de moléculas de ácido nucleico, por lo que la genoteca de moléculas de ácido nucleico comprende una pluralidad de elementos, por lo que los elementos comprenden al menos una extensión constante y al menos una extensión variable, y por lo que la al menos una extensión constante comprende una secuencia de nucleótidos que es la misma en todos los elementos, y la al menos una extensión variable comprende una secuencia de nucleótidos que es diferente en todos los elementos, por lo que el procedimiento comprende las etapas siguientes:
- 45 a) proporcionar un primer oligonucleótido, por lo que el primer oligonucleótido es un oligonucleótido al menos parcialmente bicatenario que tiene un saliente monocatenario y comprende un sitio de reconocimiento para una primera enzima de restricción de tipo IIS que corta fuera de su secuencia de reconocimiento, o una genoteca de ácido nucleico que comprende una pluralidad de elementos, por lo que la genoteca de ácido nucleico es la genoteca de ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 18 y los elementos de la
- 50 genoteca de ácido nucleico tienen un saliente monocatenario, y los elementos de la genoteca de ácido nucleico comprenden un sitio de reconocimiento para una primera enzima de restricción de tipo IIS que corta fuera de su secuencia de reconocimiento;
- b) proporcionar un segundo oligonucleótido, por lo que dicho segundo oligonucleótido es una genoteca de ácido nucleico que comprende una pluralidad de elementos, por lo que la genoteca de ácido nucleico es la genoteca de
- 55 ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 18, y los elementos de la genoteca de

- ácido nucleico tienen un saliente monocatenario que es al menos parcialmente complementario al saliente monocatenario del oligonucleótido proporcionado en la etapa a), los elementos de la genoteca de ácido nucleico comprenden un sitio de reconocimiento para una segunda enzima de restricción de tipo IIS que corta fuera de su secuencia de reconocimiento, por lo que la segunda enzima de restricción de tipo IIS de los elementos de la genoteca de ácido nucleico es diferente de la primera enzima de restricción de tipo IIS del oligonucleótido proporcionado en la etapa a);
- 5 c) combinar los oligonucleótidos de la etapa a) y la etapa b) y ligar una molécula del oligonucleótido de la etapa a) con cada molécula de cada elemento de la genoteca de ácido nucleico de la etapa b) mediante sus salientes monocatenarios con la condición de que el oligonucleótido de la etapa a) sea diferente de la genoteca de ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 18; o
- 10 combinar los oligonucleótidos de la etapa a) y la etapa b) y ligar cada molécula de cada elemento de la genoteca de ácido nucleico de la etapa a) con cada molécula de cada elemento de la genoteca de ácido nucleico de la etapa b) mediante sus salientes monocatenarios con la condición de que el oligonucleótido de la etapa a) sea una genoteca de ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 18;
- 15 d) opcionalmente eliminar reactivos y enzimas sin reaccionar,
- e) escindir el producto de ligación de la etapa c) con una enzima de restricción de tipo IIS que corta fuera de su secuencia de reconocimiento, por lo que la escisión se produce en el oligonucleótido de la etapa a) o la etapa b) proporcionando una genoteca de moléculas de ácido nucleico prolongadas; y
- f) opcionalmente separar las moléculas de ácido nucleico prolongadas de la mezcla de reacción.
- 20 20. El procedimiento de acuerdo con la reivindicación 19, por lo que la etapa a) a f) se repiten con la condición de que el oligonucleótido de la etapa a) sea un cuarto oligonucleótido que es una genoteca de moléculas de ácido nucleico prolongadas proporcionada en la etapa e) o un cuarto oligonucleótido al menos parcialmente bicatenario que tiene un saliente monocatenario que comprende un sitio de reconocimiento para una cuarta enzima de restricción de tipo IIS que corta fuera de su secuencia de reconocimiento;
- 25 el oligonucleótido de la etapa b) es un tercer oligonucleótido que es una genoteca de moléculas de ácido nucleico prolongadas proporcionada en la etapa e), por lo que el tercer y cuarto oligonucleótidos comprenden cada uno un saliente monocatenario y por lo que ambos salientes son al menos parcialmente complementarios entre sí; cuando se combinan los oligonucleótidos de la etapa a) y la etapa b) y se liga una molécula del oligonucleótido de la etapa a) con cada molécula de cada elemento de la genoteca de ácido nucleico de la etapa b) mediante sus salientes monocatenarios, el oligonucleótido de la etapa a) es diferente de la genoteca de moléculas de ácido nucleico prolongadas de la etapa b); o cuando se combinan los oligonucleótidos de la etapa a) y la etapa b) y se liga cada molécula de cada elemento de la genoteca de ácido nucleico de la etapa a) con cada molécula de cada elemento de la genoteca de ácido nucleico de la etapa b) mediante sus salientes monocatenarios, el oligonucleótido de la etapa a) es la genoteca de moléculas de ácido nucleico prolongadas de la etapa b); y
- 30 en el que en la etapa e) se obtiene una genoteca adicional de moléculas de ácido nucleico prolongadas.
- 35 21. El procedimiento de acuerdo con cualquiera de las reivindicaciones 19 a 20, en el que la extensión variable o parte de la misma se proporciona por la segunda extensión de los elementos de la genoteca de ácido nucleico como se describe en cualquiera de las reivindicaciones 1 a 20.
- 40 22. El procedimiento de acuerdo con cualquiera de las reivindicaciones 19 a 21, en el que la extensión constante o parte de la misma se proporciona por la primera extensión y/o la tercera extensión de los elementos de la genoteca de ácido nucleico como se describe en cualquiera de las reivindicaciones 1 a 18.
23. El procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 18, en el que la proporción de los diversos elementos y miembros de los genotecas usados en cualquiera de dichos procedimientos es equimolar.
- 45 24. El procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 18, en el que la proporción de los diversos elementos y miembros de las genotecas usados en cualquiera de dichos procedimientos no es equimolar o está sesgada.

Sloning BioTechnology GmbH
S 10040 EP-EP

Fig. 1 (1)

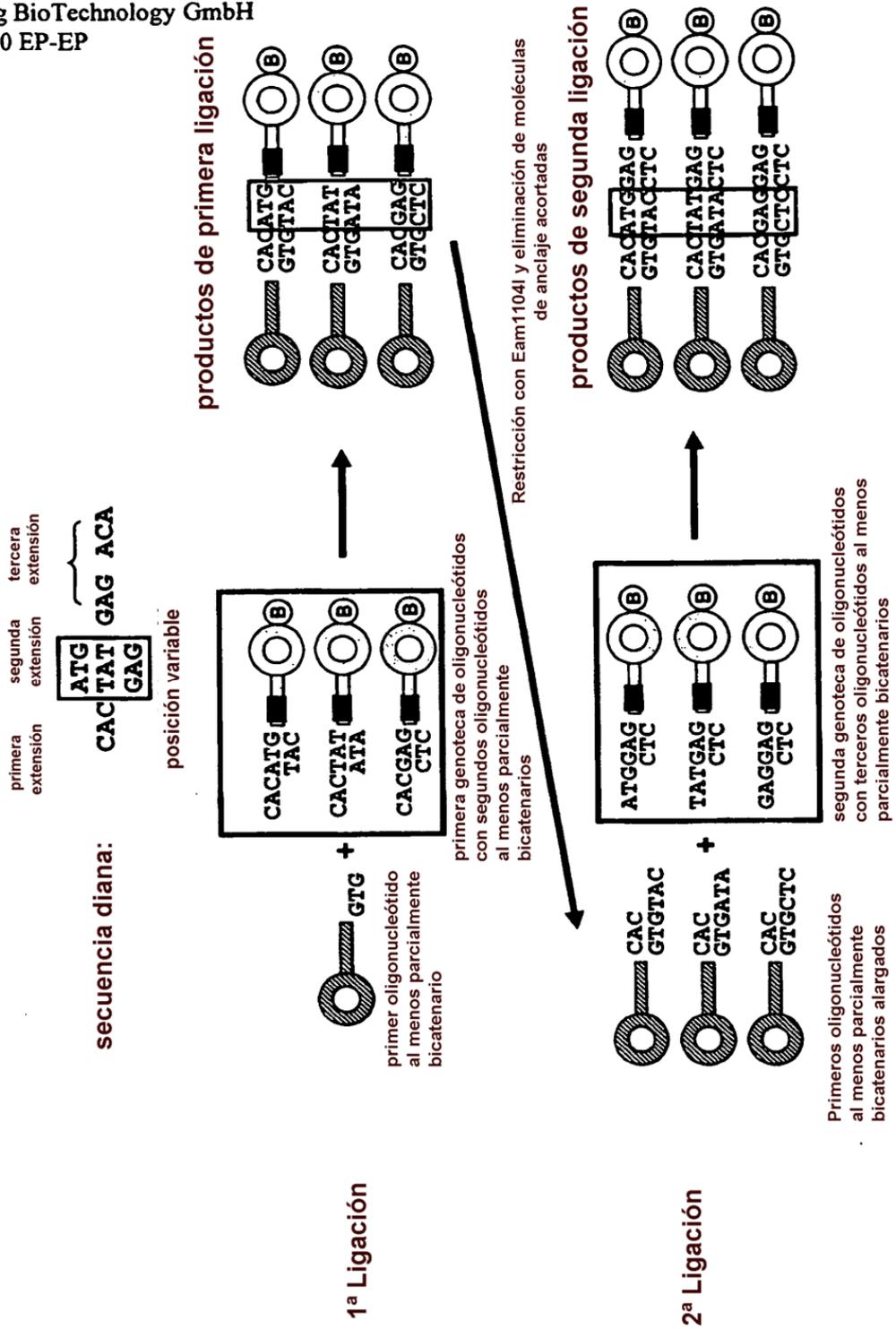


Fig. 1 (2)

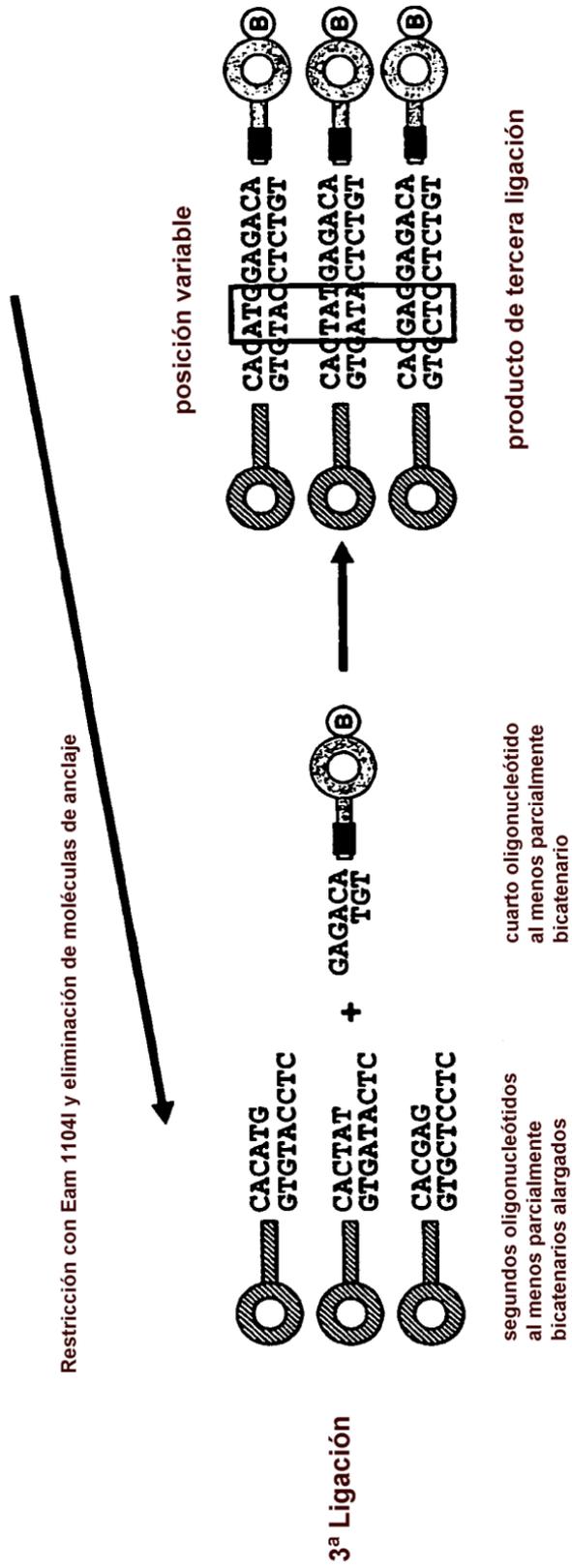


Fig. 2 (1)

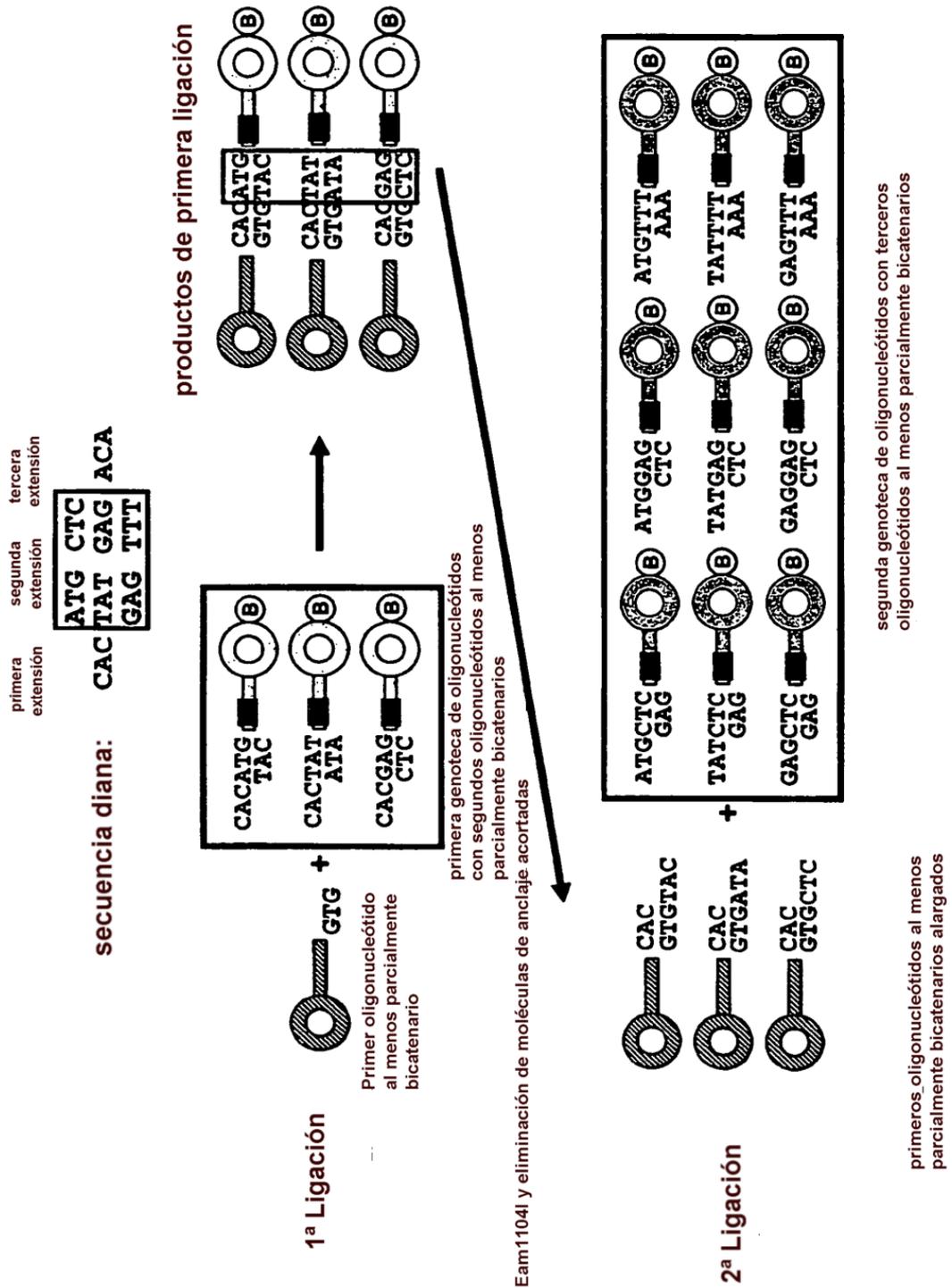


Fig. 2 (2)

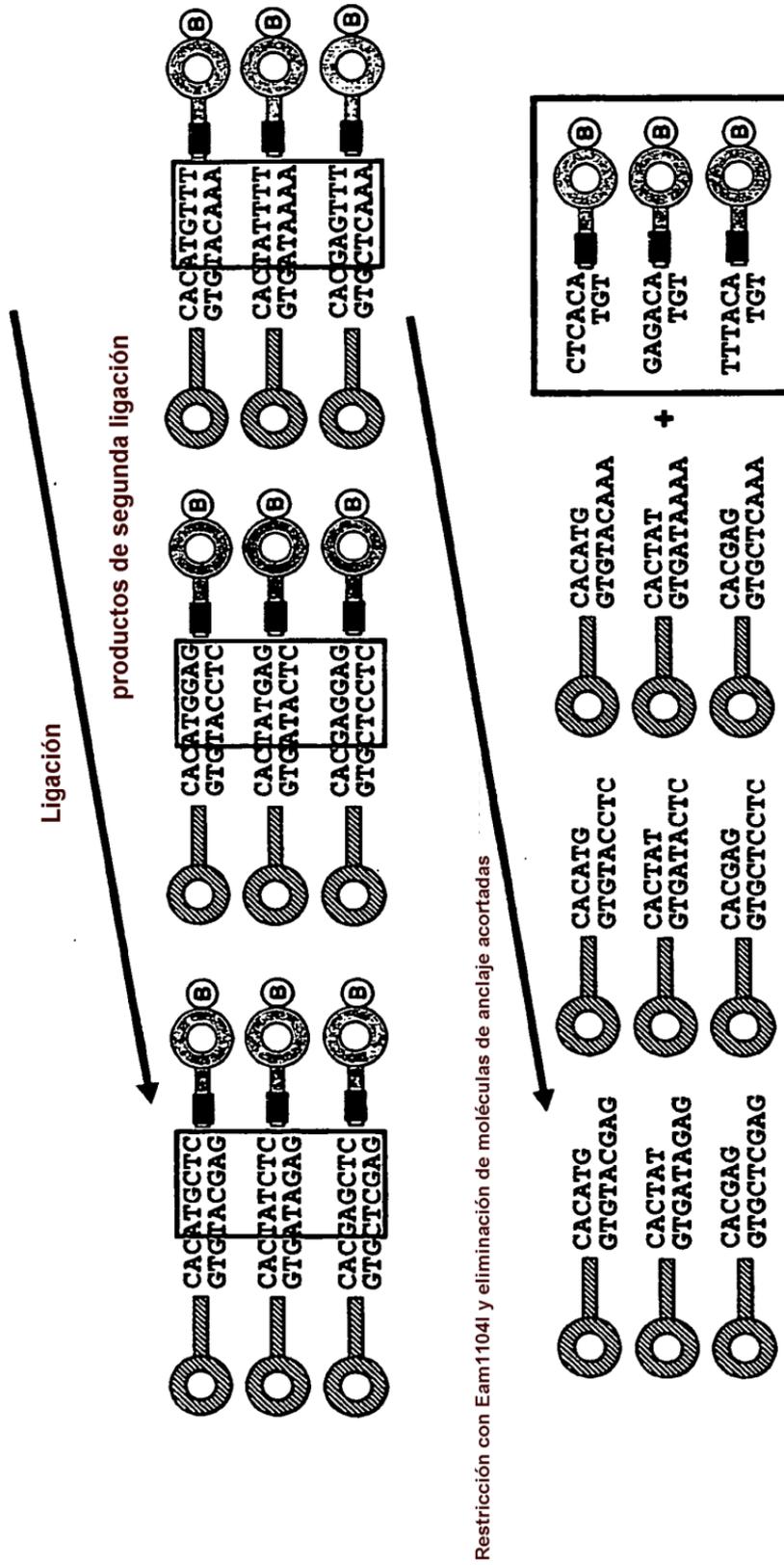


Fig. 2 (3)

Ligación

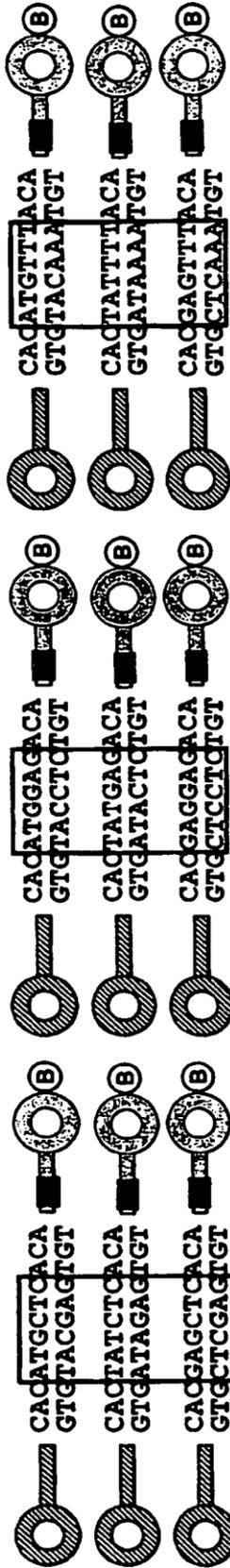


Fig. 3 (1)

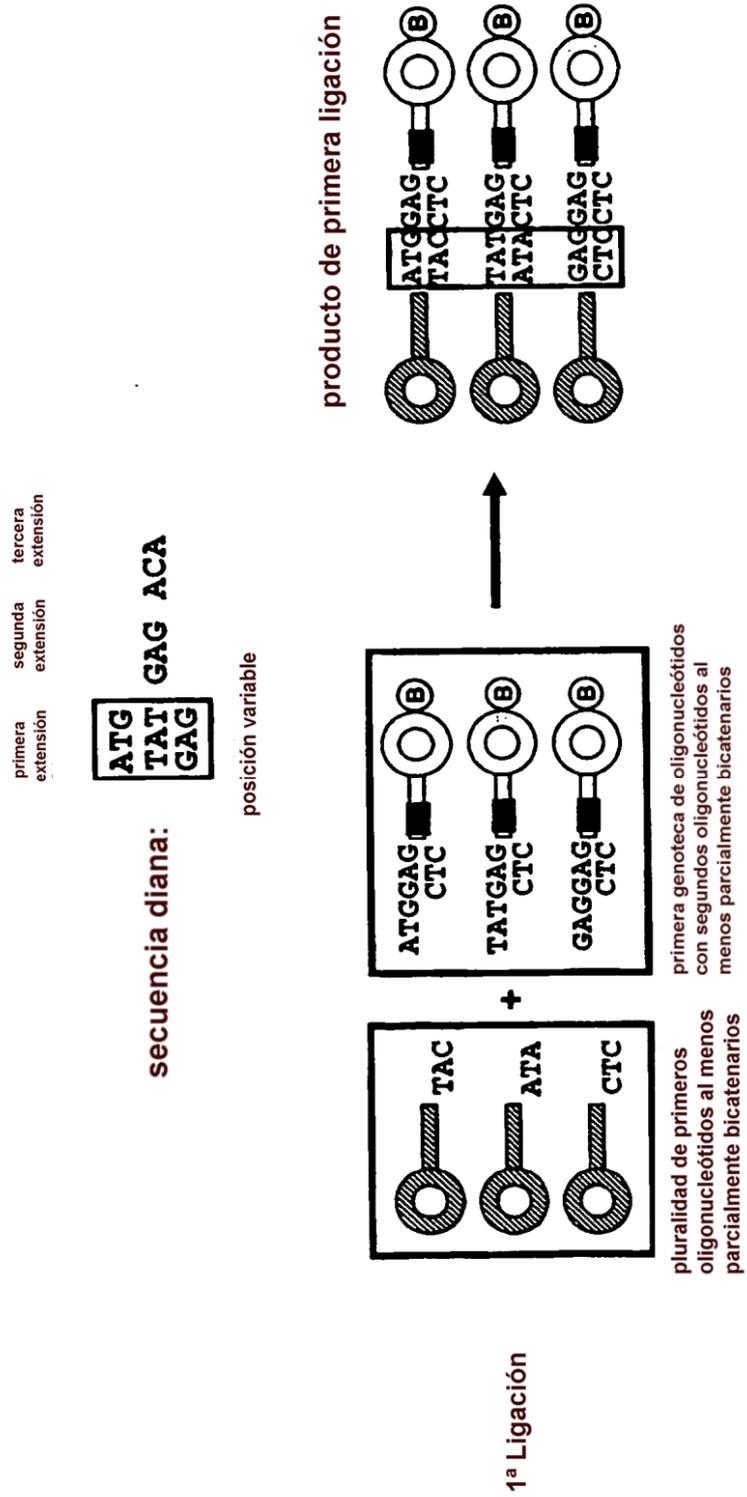


Fig. 3 (2)

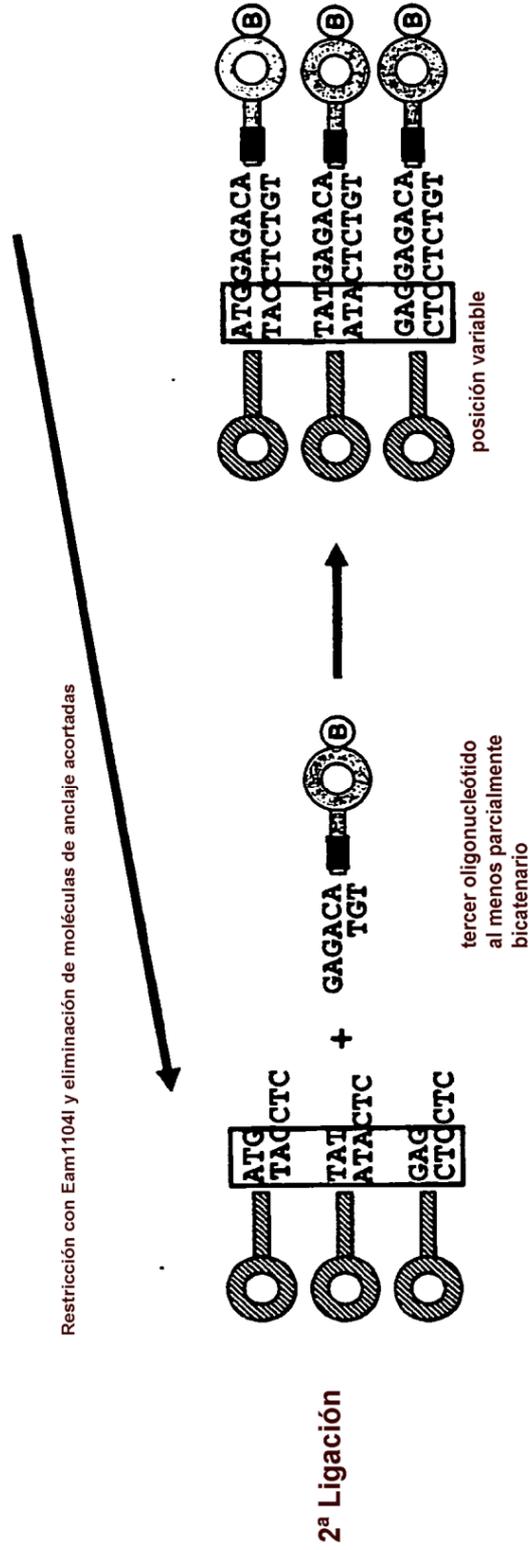


Fig. 4 (1)

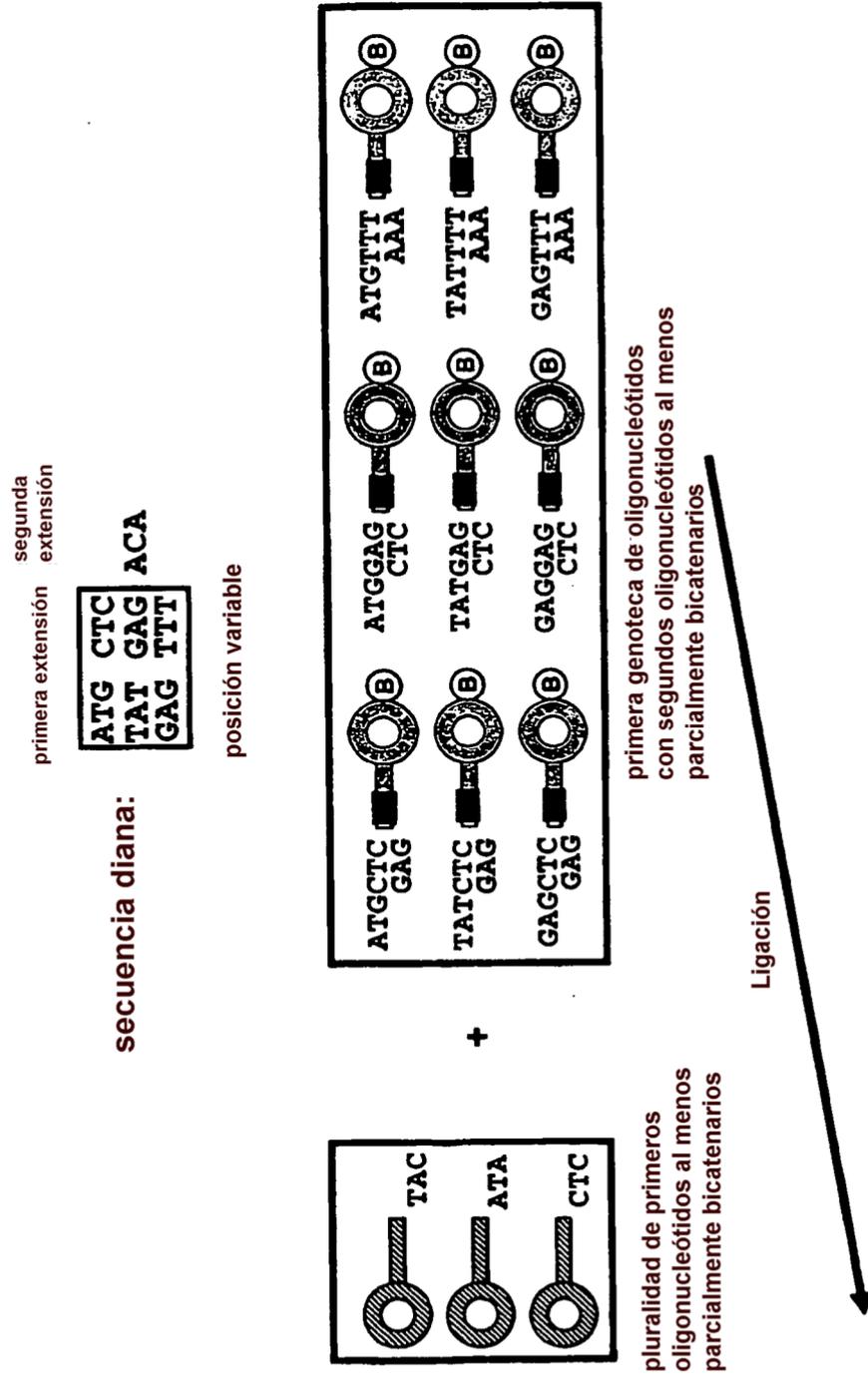


Fig. 4 (2)

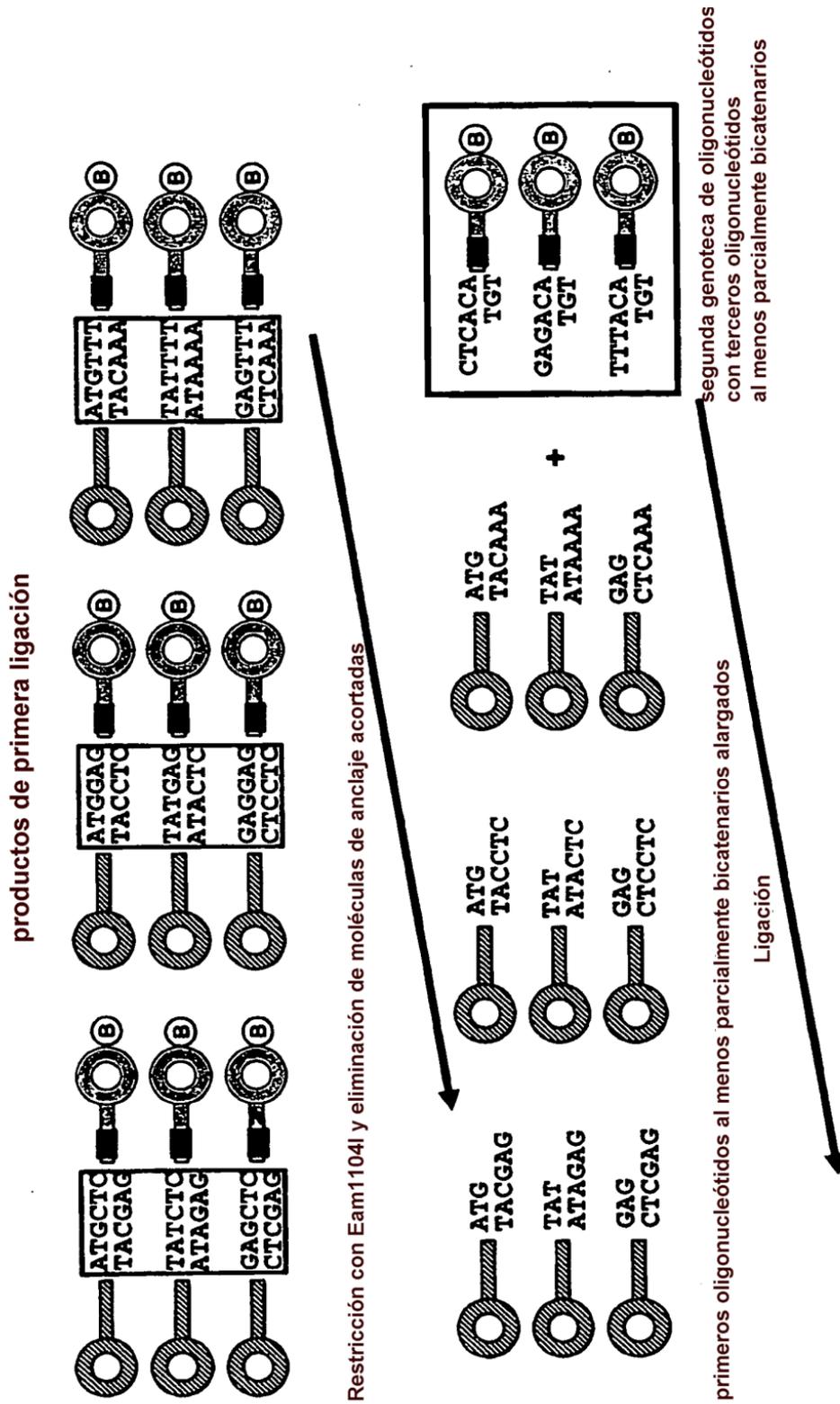
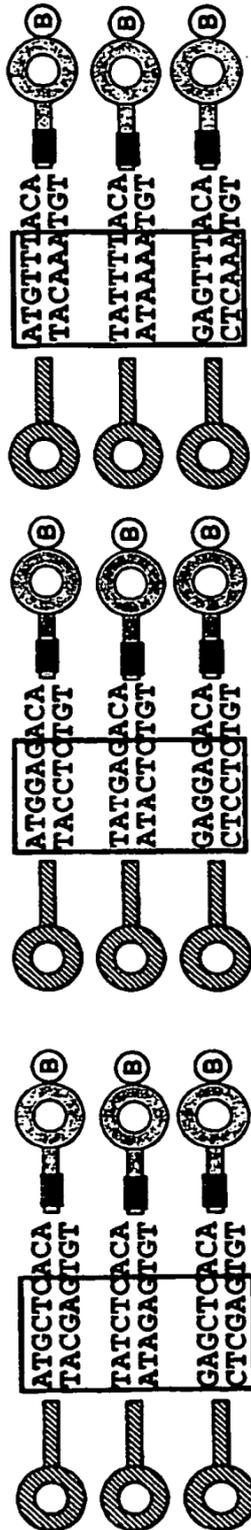


Fig. 4 (3)



productos de segunda ligación

Fig. 5

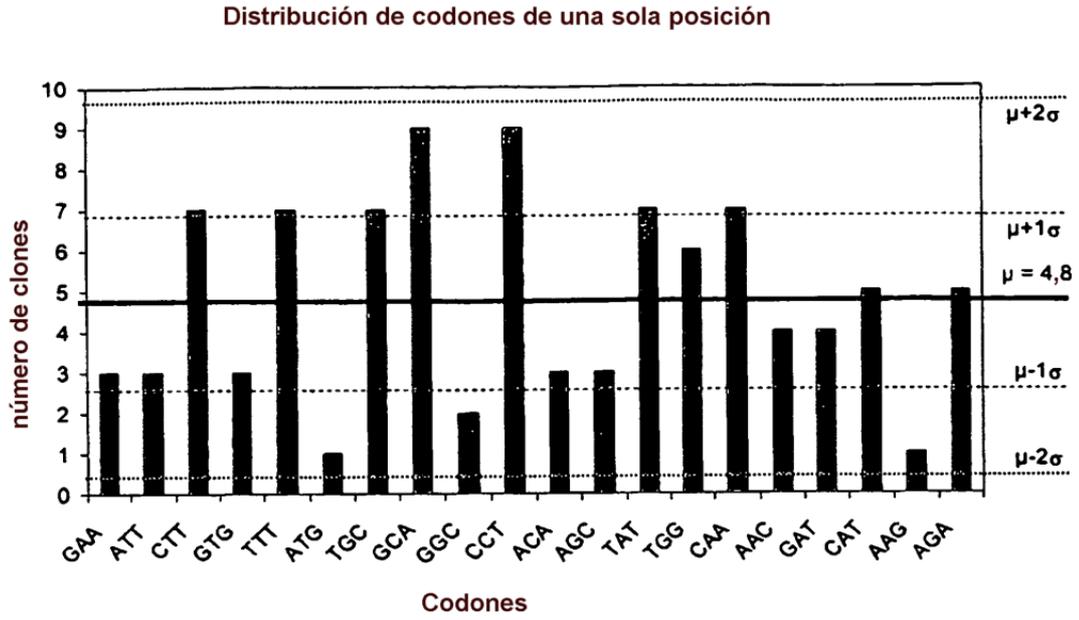


Fig. 6

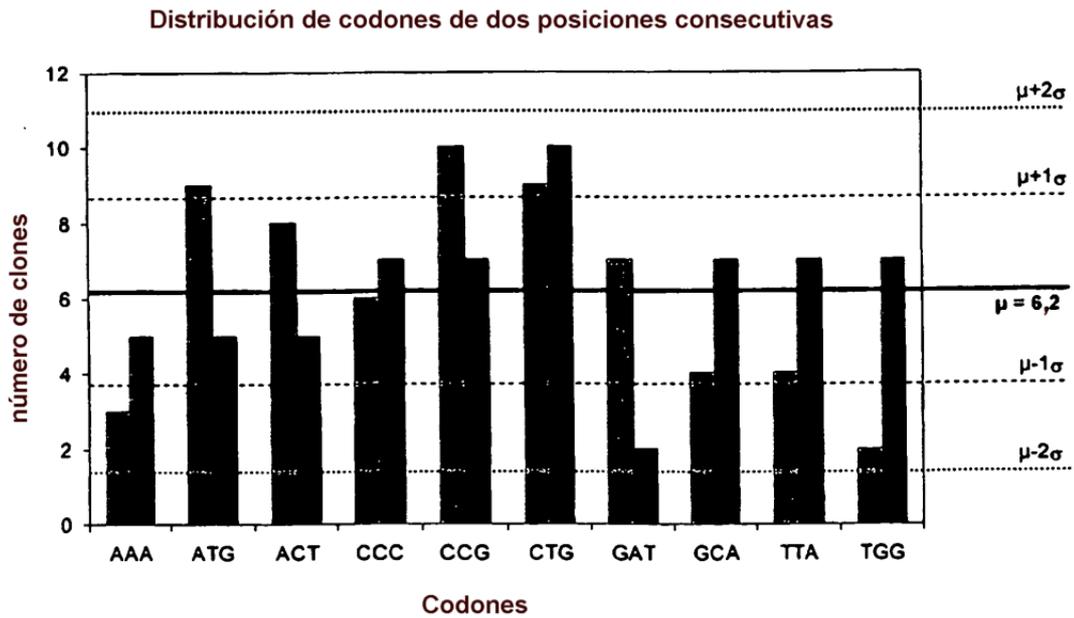


Fig. 7

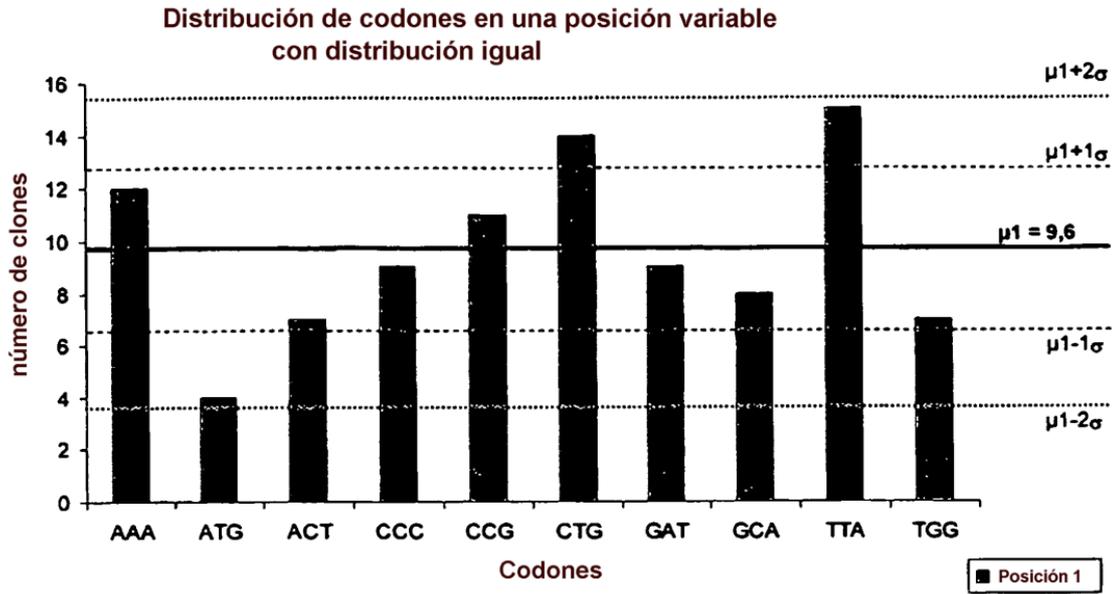


Fig. 8

