



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 324 513**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12N 15/66 (2006.01)

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **00940666 .1**

96 Fecha de presentación : **27.06.2000**

97 Número de publicación de la solicitud: **1196554**

97 Fecha de publicación de la solicitud: **17.04.2002**

54

Título: **Procedimientos de clonación y producción de cadenas de fragmentos con contenido de información legible.**

30

Prioridad: **28.06.1999 NO 20003190**
28.06.1999 NO 20003191

45

Fecha de publicación de la mención BOPI:
10.08.2009

45

Fecha de la publicación del folleto de la patente:
10.08.2009

73

Titular/es: **Complete Genomics AS.**
P.O. Box 64, Blindern
0313 Oslo, NO

72

Inventor/es: **Lexow, Preben**

74

Agente: **Curell Suñol, Marcelino**

ES 2 324 513 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

Procedimientos de clonación y producción de cadenas de fragmentos con contenido de información legible.

5 La presente invención se refiere a nuevos procedimientos de generación de ADN con un contenido de información fácilmente legible.

10 Actualmente los procedimientos de clonación conocidos implican generalmente la utilización de enzimas de restricción que se utilizan para generar fragmentos para la inserción y vectores de escisión para las secuencias correspondientes producidas y por consiguiente terminales complementarias. Generalmente, las enzimas que se utilizan cortan secuencias palindrómicas y de este modo producen prolongaciones idénticas. Diferentes secuencias que son cortadas con las mismas endonucleasas de restricción pueden ligarse a continuación conjuntamente para formar nuevos ácidos nucleicos recombinantes.

15 Mandecki y Bolling; *Gene*. 15 de agosto de 1988; 63(1):101-7 describe un procedimiento para la síntesis de un gen u otro fragmento de ADN con una secuencia definida.

20 Se ha descubierto actualmente de manera sorprendente que al generar fragmentos que comprenden zonas que representan una unidad de información que corresponden a uno o más elementos del código, puede sintetizarse una molécula de ácido nucleico bicatenaria que puede almacenar código alfanumérico. Según un primer aspecto de la invención, un procedimiento para sintetizar una molécula de ácido nucleico bicatenario que contiene información en forma de código alfanumérico para almacenar la información en forma de código alfanumérico comprende:

- 25 i) generar n fragmentos de ácido nucleico bicatenario, en los que por lo menos n-2 fragmentos presentan zonas monocatenarias en ambos terminales y 2 fragmentos presentan zonas monocatenarias en por lo menos un terminal, en los que (n-1) zonas monocatenarias son complementarias con (n-1) otras zonas monocatenarias, produciendo de este modo (n-1) pares complementarios;
- 30 ii) poner en contacto dichos n fragmentos de ácido nucleico bicatenarios simultánea o consecutivamente, para efectuar la unión de dichos pares complementarios de zonas monocatenarias; y
- iii) ligar opcionalmente dichos pares complementarios simultánea o consecutivamente para producir una molécula de ácido nucleico constituida por n fragmentos;

35 en la que dichos fragmentos comprenden zonas que representan una unidad de información correspondiente a uno o más elementos del código.

40 Según un segundo aspecto de la invención, un procedimiento identifica los elementos del código contenidos en una molécula de ácido nucleico preparada según el primer aspecto, en el que una sonda portadora de unos medios de señalización específica para uno o más elementos del código, está unida a la molécula de ácido nucleico y se detecta una señal generada por los medios de señalización, por lo que pueden identificarse uno o más elementos del código.

45 “Zonas monocatenarias” como se hace referencia en la presente memoria son las zonas que sobresalen en el extremo, es decir en el terminal de las moléculas de ácido nucleico. Estas zonas son suficientes para permitir la unión específica de moléculas con zonas monocatenarias complementarias y la ligadura posterior entre estas moléculas. De este modo, las zonas monocatenarias son por lo menos de 1 base de longitud, preferentemente de 3 bases de longitud, pero preferentemente de por lo menos 4 bases, por ejemplo de 4 a 10 bases, por ejemplo de 4, 5 ó 6 bases de longitud. Se contemplan zonas monocatenarias de hasta 20 bases de longitud que permitirán la utilización de fragmentos en el procedimiento de la invención que son de hasta Mb de longitud.

“Unión” tal como se utiliza en la presente memoria se refiere a la etapa de asociación de zonas monocatenarias complementarias (es decir, unión no covalente). La “ligadura” ulterior de las secuencias consigue la unión covalente.

55 “Complementario” tal como se utiliza en la presente memoria se refiere al reconocimiento de bases específicas mediante por ejemplo la complementariedad base-base. Sin embargo, la complementariedad tal como se hace referencia en la presente memoria incluye el emparejamiento de nucleótidos en el emparejamiento de bases Watson-Crick además del emparejamiento de análogos de nucleósido, por ejemplo desoxinosina que son susceptibles de hibridación específica a la base en las moléculas de ácido nucleico y otros análogos que dan como resultado dicha hibridación específica, por ejemplo, APN, ADN y sus análogos. La complementariedad de una zona monocatenaria con otras se considera que es suficiente, cuando, en las condiciones utilizadas, se consigue la fijación específica. De este modo en el caso de zonas monocatenarias largas puede tolerarse alguna falta de especificidad base-base, por ejemplo, mal emparejadas, por ejemplo si una base en un serie de 10 bases no es complementaria. Dichos desemparejamientos ligeros que no afectan a la última fijación ni a la ligadura de las zonas monocatenarias se consideran que son complementarios para los fines de la presente invención. Las zonas monocatenarias pueden conservar partes, en la unión, que continúan siendo monocatenarias, por ejemplo cuando se emplean prolongaciones de diferentes tamaños o las porciones complementarias no comprenden todas las zonas monocatenarias. En dichos casos, tal como se mencionó anteriormente, puede conseguirse proporcionar el enlace de zonas monocatenarias que se considera que son complementarias. En

ES 2 324 513 T3

estos casos, antes de la ligadura, las bases desaparecidas pueden rellenarse en por ejemplo utilizando el fragmento de Klenow, u otras técnicas apropiadas cuando sea necesario.

5 La ligadura entre las moléculas de ácido nucleico se consigue por cualquier técnica apropiada conocida en la materia (véase por ejemplo, Sambrook *et al.*, en “*Molecular Cloning: A Laboratory Manual*”, 2ª ed., Editor Chris Nolan, Cold Spring Harbor Laboratory Press, 1989). Por ejemplo, la ligadura puede conseguirse químicamente o mediante la utilización de ligasas naturales apropiadas o variantes de las mismas. Las ligasas apropiadas que pueden utilizarse incluyen la T4 ADN ligasa, y ligasas termoestables, tales como Pfu, Taq y TTH ADN ligasa. La ligadura puede evitarse o permitirse controlando el estado de fosforilación de las bases terminales por ejemplo mediante la utilización
10 apropiada de cinasas o fosfatasas. Pueden utilizarse también de manera apropiada grandes volúmenes para permitir ligaduras intermoleculares. De este modo, puede utilizarse adaptador elevado para las relaciones vector/inserción para evitar que el vector o la inserción se vuelvan a ligar en su material original.

15 En el procedimiento de la invención, se combinan múltiples fragmentos mediante la selección apropiada de las zonas monocatenarias que aparecen en sus extremos. Esto tiene aplicación en la producción de secuencias específicas con fines biológicos, pero presenta utilidad específica en la producción de cadenas de moléculas de ácido nucleico en las que las unidades que preparan las cadenas cada una significa una unidad de información, es decir, las cadenas pueden utilizarse para almacenar información, tal como se describirá con más detalle a continuación. Tal como se utiliza en la presente memoria “cadenas” se refiere a una disposición en serie de los fragmentos tal como se describe
20 en la presente memoria. Dichas cadenas son preferentemente lineales e incluyen secuencias de fragmentos ramificadas y no ramificadas. De este modo, por ejemplo, pueden utilizarse fragmentos de ADN ramificados para proporcionar cadenas con una disposición ramificada de los fragmentos.

25 Para producir cadenas de moléculas de ácido nucleico con diferentes fragmentos unitarios, es decir cadenas de fragmentos puede utilizarse el procedimiento siguiente. En primer lugar es necesario generar fragmentos que tengan prolongaciones en su extremo, para permitirles que se unan uno con otro. (Los fragmentos 3' y 5' últimos pueden sin embargo tener una prolongación solamente en el extremo que estará acoplado a los fragmentos internos). Tal como se describirá con mayor detalle a continuación, para determinadas aplicaciones, oligonucleótidos apropiados pueden proceder de bancos en los que los miembros presentan variabilidad en por lo menos alguna de sus bases. Si deben
30 producirse bancos en los que los miembros son bicatenarios, se valorará que el número de miembros en dicho banco pueda ser más bien elevado. Esto puede sin embargo reducirse eficazmente utilizando un número más pequeño de bloques de construcción más pequeño.

35 Una estrategia consiste en construir dos oligonucleótidos monocatenarios utilizando técnicas convencionales. Pueden utilizarse oligonucleótidos que tienen una zona de 6 bases que se complementan entre sí y de este modo permiten la hibridación. Como no todas las moléculas están implicadas en la hibridación, las zonas monocatenarias se extienden más allá de la zona de hibridación creando de este modo zonas monocatenarias. Por conveniencia el número de miembros del banco requeridos puede reducirse aún más si aparecen secuencias repetidas con frecuencia en la cadena del fragmento. Esto se describirá con mayor detalle a continuación.
40

Una vez se han creado las unidades (es decir, los fragmentos) de la cadena bicatenaria apropiadas pueden ligarse en la misma solución, proporcionando que las diferentes prolongaciones presentes en las secuencias sean únicas.

45 Por lo tanto, en un primer aspecto, la presente invención proporciona un procedimiento de síntesis de una molécula de ácido nucleico bicatenaria que contiene información en forma de código alfanumérico para almacenar información que comprende por lo menos las etapas siguientes:

- 1) generar n fragmentos de ácido nucleico bicatenarios, en el que por lo menos n-2 fragmentos tienen zonas monocatenarias en ambos terminales y 2 fragmentos tienen zonas monocatenarias en por lo menos un terminal, en el que (n-1) zonas monocatenarias son complementarias con otras (n-1) zonas monocatenarias, produciendo de este modo (n-1) pares complementarios,
50
- 2) poner en contacto dichos n fragmentos de ácido nucleico bicatenario, simultánea o consecutivamente, para efectuar la unión de dichos pares complementarios de zonas monocatenarias,
55
- 3) ligar opcionalmente dichos pares complementarios simultánea o consecutivamente para producir una molécula de ácido nucleico constituida por n fragmentos, en los que dichos fragmentos comprenden zonas que representan una unidad de información correspondiente a uno o más elementos del código.
60

Las expresiones “molécula de ácido nucleico”, “zonas monocatenarias”, “complementario”, “enlace” y “ligadura” se describieron anteriormente en la presente memoria.

65 En la etapa 1) se hace referencia a (n-1) zonas monocatenarias complementarias a “otras” (n-1) zonas monocatenarias. Ésta describe dos familias de zonas monocatenarias, que comprenden conjuntamente 2(n-1) miembros, que forman n-1 pares. Por lo tanto “otras” se refiere a las zonas monocatenarias en la segunda familia que no están presentes en la primera familia.

ES 2 324 513 T3

“Poner en contacto” tal como se utiliza en la presente memoria se refiere a poner juntos los fragmentos bicatenarios en condiciones que sean conductores para la asociación de las zonas monocatenarias complementarias. Dependiendo del procedimiento utilizado, esto puede permitir por último la ligadura de los fragmentos que llevan estas zonas. Sin embargo debe indicarse que los fragmentos pueden estar unidos por otros procedimientos aparte de la ligadura. Por ejemplo puede utilizarse la PCR con cebadores apropiados, por ejemplo pares de cebadores.

La puesta en contacto y/o ligadura simultánea o consecutiva se refiere a la posibilidad de añadir los fragmentos individualmente o en grupos a una cadena en crecimiento o añadir simultáneamente los n fragmentos juntos, en los que la ligadura puede realizarse después de cada adición o una vez se han combinado los n fragmentos. Preferentemente la ligadura se efectúa una vez se han combinado todos los fragmentos.

“Fragmentos” tal como se utiliza en la presente memoria son preferentemente mayores de 6 bases de longitud (en los que dicha longitud se refiere a la longitud de cada oligonucleótido monocatenario que construye el fragmento que puede diferenciarse ligeramente en la longitud de uno a otro), por ejemplo entre 6 y 50 bases, por ejemplo de 8 a 25 bases.

Tal como se hace referencia en la presente memoria, “n” es un número entero de por lo menos 4, por ejemplo, por lo menos 10 ó 100, por ejemplo entre 25 y 200.

Preferentemente, tal como se mencionó anteriormente, los fragmentos se generan mediante la utilización de oligonucleótido monocatenarios para generar moléculas bicatenarias apropiadas.

De particular interés en dichos procedimientos es la producción de cadenas de fragmentos que pueden utilizarse para almacenar información en forma de código a la que puede accederse fácilmente.

Existe actualmente una gran necesidad de almacenar información con diferentes propósitos (por ejemplo programas informáticos, música, películas, bases de datos, etc.). Es por consiguiente imperativo encontrar unos medios de almacenamiento eficaces, que den como resultado el desarrollo de tecnología de CD ROM, DVD, etc. las moléculas de ácido nucleico ofrecen procedimientos muchos más eficaces para almacenar información y presentan varias ventajas sobre procedimientos de almacenamiento actualmente en utilización. Por ejemplo, la capacidad de almacenamiento de las moléculas de ácido nucleico es extensa. En principio, uno tubo de ensayo que contiene moléculas de ADN puede contener tanta información como varios millones de CD ROM o más. El ácido nucleico puede copiarse rápida y eficazmente utilizando sistema naturales que son mejorados en gran medida por técnicas que han sido desarrolladas tales como PCR, LCR, etc. Cuando se almacenan apropiadamente, las moléculas de ácido nucleico pueden conservarse durante periodos sumamente largos. Las herramientas naturales para la manipulación de moléculas nucleicas están ya disponibles para el tratamiento de las moléculas, por ejemplo polimerasas, enzimas de restricción, factores de transcripción, ribosomas, etc. Las moléculas de ácido nucleico pueden también tener propiedades catalíticas.

Además, pueden utilizarse moléculas de ácido nucleico como sistemas seguros ya que pueden construirse de tal manera que no sean copiadas fácilmente, a diferencia de la copia de los sistemas de almacenamiento actuales, por ejemplo los CD, etc., que está aumentando constantemente.

Anteriormente, sin embargo, no fue posible aprovecharse del enorme potencial ofrecido por las moléculas de ácido nucleico debido a la ausencia de algunos procedimientos eficaces para escribir mensajes de ADN o leer mensajes de ADN. El procedimiento descrito anteriormente proporciona procedimientos que superan este problema permitiendo la síntesis rápida de grandes moléculas de ADN y procedimientos para escanear rápida y eficazmente aquellas moléculas para recuperar la información.

La clave para la recuperación eficaz de la información codificada por las moléculas de ácido nucleico producidas según el procedimiento descrito en la presente memoria, es la expansión de la unidad que proporciona información en la molécula. En la naturaleza y en los procedimientos utilizados anteriormente, cada base en la secuencia presenta un contenido de información individual. De hecho se han descrito procedimientos en los que una sola base puede significar más de una sola unidad de información, por ejemplo en código binario, las bases A=“00”, C=“01”, G=“10” y T=“11”. Aunque esto presenta ventajas siempre que cantidades significativas de información puedan estar contenidas en una sola molécula, el sistema adolece de graves inconvenientes ya que requiere procedimientos de escritura y lectura en los que las bases individuales puedan acoplarse y discriminarse.

En un procedimiento preferido de la invención por consiguiente, se proporcionan unidades de información que no son bases individuales, sino que en su lugar son secuencias cortas. Las técnicas descritas anteriormente permiten la producción rápida de dichas cadenas y que pueda accederse fácilmente a la información.

Por lo tanto las unidades que representan información codificada pueden generarse y leerse. Cada unidad de información por consiguiente puede representar un elemento del código, en el que el código puede por ejemplo ser código alfanumérico o una representación más simple tal como un código binario. En cada caso los elementos individuales del código, por ejemplo “a”, “b”, “c”, “1”, “0”, etc. es necesario estén representados por una secuencia individualizada y específica.

ES 2 324 513 T3

Tal como se utiliza en la presente memoria “unidades de información” se refiere a secuencias cortas discretas que representan una sola pieza de información, por ejemplo uno o más elementos (es decir, combinaciones de los mismos) de un código.

5 “Elementos” del código, tal como se mencionó anteriormente se refieren a diferentes miembros que construyen un código tal como un código binario o alfanumérico.

Por lo tanto, según el procedimiento de la invención, los fragmentos que están unidos comprenden zonas que representan una unidad de información correspondiente a uno o más elementos del código. El código es alfanumérico.
10 Especialmente con preferencia el código es binario. De este modo, por ejemplo, considerando un sistema binario de captura de información, si se desea producir cadenas constituidas por “0”, “1” fragmentos, las combinaciones de secuencias apropiadas pueden atribuirse a “0” ó “1”.

Por conveniencia cada uno o más de dichos elementos del código (juntos) presenta la fórmula:
15

$$(X)_a,$$

en la que
20

X es un nucleótido A, T, G, C o un derivado del mismo que permite la unión complementaria y puede ser igual o diferente en cada posición, y

a es un número entero mayor de 2, por ejemplo mayor de 4, por ejemplo de 2 a 20, preferentemente de 4 a 10, por
25 ejemplo de 6 a 8,

en el que $(X)_a$ es diferente de cada uno o más elementos del código.

Especialmente preferentemente, en el caso del código binario, los elementos del código “1” y “0” puede presentar
30 las fórmulas:

$$“0”=(X)_a \quad y \quad “1”=(Y)_b,$$

en la que
35

$(X)_a$ e $(Y)_b$ no son idénticos,

X e Y son cada uno un nucleótido A, T, G, C o un derivado de los mismos que permite la unión complementaria y
40 pueden ser iguales o diferentes en cada posición, y

a y b son números enteros mayores de 2, por ejemplo mayores de 4, por ejemplo de 2 a 20, preferentemente de 4 a
10, por ejemplo de 6 a 8.

45 Tal como se hace referencia en la presente memoria un “derivado” que es susceptible de unión complementaria se refiere a un análogo o variante de nucleótido que es susceptible de unirse a un nucleótido presente en una cadena complementaria, e incluye variantes de nucleótidos naturales o sintéticas específicas, por ejemplo uracilo o nucleótidos metilados o amidados, etc.

50 En la forma más sencilla y preferida, X e Y son los mismos en cada posición, por ejemplo “0” = GGGGGGGG y “1” = AAAAAAAA. Sin embargo, pueden utilizarse secuencias repetidas tales como $[AC]_6A$ o $[GT]_6A$. La secuencia del código puede presentar también una propiedad funcional, por ejemplo, puede ser un elemento de integración tal como AttP1 o AttP2.

55 Se valorará sin embargo que las secuencias descritas anteriormente puedan también indicar más de un solo elemento del código. De este modo por ejemplo la unidad de información puede indicar 2 o más elementos del código, por ejemplo de 2 a 32 elementos, preferentemente de 2 a 4 elementos del código. Si, por ejemplo, se considera el código binario, cada unidad de información puede referirse a “01” ó “00” u “11” ó “10”.

60 En el procedimiento descrito en la presente memoria, las cadenas que comprenden dichas propiedades pueden prepararse de la manera siguiente. Para producir una cadena con, por ejemplo, 8 0/1 fragmentos, ocho fragmentos que empiezan por “0” con diferentes prolongaciones y 8 fragmentos que comienzan por “1” con diferentes prolongaciones se generan como se ilustra en la Figura 1. En este caso los fragmentos “0” estaban constituidos por la secuencia GGGGGGGG, aunque esta podría sustituirse por otras secuencias. Además los fragmentos son sintetizados de modo
65 que tengan prolongaciones únicas de modo que puedan ligarse solamente en una posición. Por lo tanto, los fragmentos para la posición 1 en la cadena se producen de modo que tengan una prolongación que esté complementada por una de las prolongaciones en los fragmentos para la posición 2. De este modo, los fragmentos de la posición 2 se sintetizan de modo que puedan unirse a los fragmentos de la posición 1. Asimismo los fragmentos de la posición 3 pueden

ES 2 324 513 T3

unirse solamente a los fragmentos de la posición 2 en uno de sus terminales y los fragmentos de la posición 4 al otro terminal y así sucesivamente. Estos fragmentos se almacenan por separado. Con el fin de construir una cadena, se hace la selección de una de las dos alternativas para cada posición de modo que se produzca una cadena binaria apropiada.

5 Por lo tanto, en el esquema esbozado anteriormente, para producir un fragmento de cadena que representa una cadena 01001011, los fragmentos "0" de las posiciones 1, 3, 4 y 6 se mezclan con los fragmentos "1" de las posiciones 2, 5, 7 y 8. Si los fragmentos se ligan a continuación añadiendo ligasa o utilizando otros procedimientos de ligadura mencionados anteriormente, se producirá la cadena descrita anteriormente. Como se apreciará, ésta cadena podría también conseguirse utilizando por ejemplo solamente 4 fragmentos si la unidad de información transportase cada
10 fragmento de los elementos del código 2 indicados.

Además es posible combinar fragmentos intermedios de las cadenas, (por ejemplo que contienen por lo menos 4 fragmentos) con otros fragmentos de las cadenas, que proporcionan prolongaciones apropiadas existentes en sus terminales puedan ligarse para formar fragmentos de cadenas compuestos. Por lo tanto, pueden realizarse en paralelo varios ciclos y combinarse los productos. En el procedimiento representado en la Figura 1 los fragmentos del extremo presentan extremos truncados, pero evidentemente, pueden utilizarse fragmentos apropiados que tengan prolongaciones asimismo en los terminales.

Una técnica apropiada para producir 8 cadenas de fragmentos, que contiene cada una 8 fragmentos que pueden ser ligados entonces se ilustra en la Figura 2. Para el fragmento de la cadena 1, se utilizan fragmentos terminales de modo que es posible para la cadena de fragmento completa ligarse al fragmento de la cadena 2 y así sucesivamente. Éstos pueden combinarse a continuación para producir una cadena de 64 fragmentos. Asimismo, dichas cadenas de 8 fragmentos pueden combinarse para producir cadenas de fragmentos que comprenden 512 fragmentos.

25 Como se apreciará, tal como en la producción de cadenas más cortas, la etapa de ligadura, cuando se lleva a cabo, se efectúa convenientemente una vez que se han combinado todas las cadenas del fragmento. Sin embargo, si se desea la etapa de ligadura puede realizarse sucesivamente además de cada cadena de fragmentos posterior.

Para combinar 8 fragmentos binarios por ciclo, se requieren 16 fragmentos de partida diferentes, que representan las diferentes alternativas "0", "1" en cada posición. Para construir una cadena de 64 fragmentos utilizando dos ciclos, es decir para producir 8 cadenas con 8 fragmentos que se ligan a continuación, solamente se requieren $16 + (4 \times 7) = 44$ fragmentos de partida. Por lo tanto, el número de diferentes fragmentos de partida requeridos refleja un aumento casi lineal en contraste con las combinaciones de las cadenas de fragmentos que pueden producirse que aumenta exponencialmente con el número de ciclos. Como consecuencia, pueden producirse fragmentos de cadena muy largos con un número relativamente de fragmentos de partida.

Desde luego, como se mencionó anteriormente, pueden producirse cadenas intermedias más largas o más cortas que 8. Ya que existe un gran número de permutaciones en la zona que sobresale, pueden utilizarse más fragmentos de partida permitiendo de este modo que se construyan fragmentos mayores en un solo ciclo. Por lo tanto, puede reducirse el número de ciclos necesarios para producir cadenas largas.

Las cadenas de fragmentos pequeños producidas según los procedimientos descritos en la presente memoria pueden acoplarse conjuntamente también utilizando variaciones de las técnicas descritas en la presente memoria. Por ejemplo, pueden utilizarse pares de cebadores complementarios para unir varias cadenas tal como se describe en el ejemplo 5. En esta técnica, la ampliación de las cadenas de fragmentos se consigue utilizando diferentes pares de cebadores. El segundo cebador en el par cebador 1 es complementario con el primer cebador en el par cebador 2 y el segundo cebador en el que el par es complementario con el primer cebador en el par cebador 3 y así sucesivamente. Se realizan a continuación reacciones de PCR que producen productos que en forma monocatenaria son capaces de unirse entre sí mediante sus extremos complementarios introducidos por los pares de cebador. Éstos pueden ligarse a continuación.

Alternativamente, las cadenas de fragmentos preparadas por los procedimientos descritos en la presente memoria pueden ampliarse con un cebador que contiene una secuencia de restricción con una nucleasa que escinde fuera su secuencia de reconocimiento. Estos productos de ampliación se digieren a continuación con esta nucleasa para producir prolongaciones no palindrómicas en el extremo de cada cadena de fragmento. Mediante la selección de la secuencia apropiada (por ejemplo en el cebador o los fragmentos que se utilizan) las prolongaciones que se generan permiten a las diferentes cadenas de fragmentos combinarse en orden.

Por lo tanto en un aspecto preferido, la invención proporciona un procedimiento de síntesis de una molécula de ácido nucleico bicatenaria que comprende por lo menos las etapas siguientes:

- 1) generar cadenas de fragmentos según el procedimiento descrito anteriormente en la presente memoria;
- 2) generar opcionalmente zonas monocatenarias en el extremo de dichas cadenas de fragmentos en las que dichas zonas monocatenarias son complementarias con otras zonas monocatenarias en dichas cadenas de fragmentos formando de este modo pares complementarios de zonas monocatenarias,
- 3) poner en contacto dichas cadenas de fragmentos entre sí, simultánea o consecutivamente, para efectuar la unión de dichos pares complementarios de cadenas monocatenarias.

ES 2 324 513 T3

Opcionalmente dichas cadenas se ligan, sin embargo, pueden utilizarse técnicas alternativas para formar la cadena última, por ejemplo puede utilizarse la PCR tal como se describe en la presente memoria.

Preferentemente las cadenas de fragmentos intermedios son entre 4 y 20 fragmentos de longitud, por ejemplo de 5 a 10, y entre 5 y 50 dichas cadenas de fragmentos se combinan por ejemplo entre 10 y 20.

Por conveniencia los fragmentos que deben utilizarse en el procedimiento de la invención están contenidos en bancos. Los procedimientos de producción de los fragmentos que construyen la librería son bien conocidos en la técnica. Por ejemplo, pueden producirse una serie de oligonucleótidos que contienen dos porciones. Una primera porción que formará una prolongación en un extremo y una segunda porción que efectuará la unión a un oligonucleótido complementario y que contienen en esta porción la unidad de información. Al producir las porciones de hibridación comunes y las prolongaciones de las variantes, se crea una serie de oligonucleótidos bicatenarios para uno o más elementos del código (indicados por lo menos por una parte de la porción de hibridación). Esto proporciona un banco para uno (o una combinación de) elementos del código. Pueden crearse bancos diferentes para diferentes elementos del código (o combinaciones de los mismos), mediante alteración apropiada de la unidad de información, es decir, la secuencia en la porción de hibridación.

Por conveniencia para su utilización en la invención, estos diferentes oligonucleótidos bicatenarios están dispuestos en dos redes bidimensionales de modo que en una dimensión las posiciones consecutivas dentro del último fragmento están indicadas y en la segunda dimensión el posible elemento del código (o combinaciones de los mismos) se proporcionan. En el caso más sencillo, en el código binario, en el que "0" y "1" están representados por diferentes secuencias, la primera dimensión comprendería fragmentos para cada posición del fragmento propuesto y la segunda dimensión tendría solamente 2 variantes ("0" y "1"). Esto puede simplemente verse como un solo banco o dos bancos, es decir los bancos "0" ó "1". Una vez producidos estos bancos, las cadenas de fragmentos en cualquier orden deseado de fragmentos pueden producirse fácilmente.

Con el fin de que los miembros del banco se dirijan de manera apropiada a su zona o pocillo correcta (es decir el banco puede estar compuesto por soportes sólidos separados, o un soporte sólido con diferentes direcciones, por ejemplo pocillos o diferentes pocillos que contienen diferentes soluciones), cualquier técnica de clasificación apropiada puede utilizarse. Esta clasificación puede conseguirse en virtud del procedimiento utilizado para la producción de los miembros del banco, o puede conseguirse la clasificación mediante una técnica apropiada, por ejemplo por unión a oligonucleótidos complementarios en la zona del banco oportuna.

Los soportes sólidos apropiados adecuados para acoplar miembros del banco son bien conocidos en la técnica y se describen extensamente en la bibliografía y generalmente hablando, el soporte sólido puede ser cualquiera de los soportes o matrices bien conocidos que se utilizan extensamente en la actualidad o se proponen para la inmovilización, separación, etc. en los procedimientos químicos o bioquímicos. De este modo, por ejemplo, los restos de inmovilización pueden tomar la forma de perlas, partículas, hojas, geles, filtros, membranas, tiras de microfibras, tubos o placas, fibras o capilares, hechos por ejemplo de material polimérico, por ejemplo agarosa, celulosa, aginatos, teflón, látex o poliestireno. Resultan generalmente preferidos los materiales en partículas, por ejemplo perlas. Por conveniencia, el resto de inmovilización puede comprender partículas magnéticas, tales como partículas superparamagnéticas.

En una forma de realización preferida, se utilizan placas o láminas para permitir la fijación de las moléculas en disposición lineal. Las placas pueden comprender también pocillos perpendiculares a la placa en los que pueden acoplarse moléculas. El acoplamiento al soporte sólido puede realizarse directa o indirectamente y la técnica que se utiliza dependerá de si la molécula que debe acoplarse es un oligonucleótido para la fijación del miembro del banco o el propio miembro del banco. Para acoplar los miembros del banco directamente, es decir, no mediante la unión a un oligonucleótido, el acoplamiento puede realizarse por conveniencia indirectamente mediante la utilización de un resto de acoplamiento transportado en las moléculas de ácido nucleico y/o el soporte sólido. Por lo tanto, por ejemplo, puede utilizarse un par de acompañantes para la unión por afinidad, tal como avidina, estreptavidina o biotina, ADN o proteína de unión a ADN (por ejemplo la proteína represora lac I o la secuencia del operador lac a la que se une), anticuerpos (que pueden ser mono- o policlonales) fragmentos de anticuerpo o los epítopos o haptenos de anticuerpos. En estos casos, un acompañante del par de unión se acopla al soporte sólido (o es parte inherente del mismo) y el otro acompañante se acopla a (o es parte inherente de) las moléculas de ácido nucleico. Alternativamente, pueden utilizarse técnicas de acoplamiento directo tales como por ejemplo si se utiliza un filtro, el acoplamiento puede realizarse por reticulación producida por UV. Cuando se acoplan fragmentos de ADN, puede utilizarse también la tendencia natural del ADN a adherirse al vidrio.

Los oligonucleótidos que deben utilizarse para captura de los miembros del banco pueden acoplarse al soporte sólido mediante la utilización de grupos funcionales apropiados en el soporte sólido.

El acoplamiento de grupos funcionales apropiados al soporte sólido puede realizarse por procedimientos bien conocidos en la técnica, que incluyen por ejemplo, el acoplamiento por grupos hidroxilo, carboxilo, aldehído o amino que pueden proporcionarse mediante el tratamiento del soporte sólido para proporcionar recubrimientos de la superficie adecuados. El acoplamiento de grupos funcionales apropiados a las moléculas de ácido nucleico de la invención puede realizarse por ligadura o introducirse durante la síntesis o ampliación, utilizando por ejemplo cebadores que llevan un resto apropiado, tal como biotina o una secuencia específica para captura.

ES 2 324 513 T3

El mezclado apropiado puede conseguirse por automatización. Por ejemplo, en el caso de los fragmentos “0” y “1”, la combinación correcta de estos elementos es la etapa crítica desde el punto de vista del consumo de recursos y de tiempo. Este procedimiento esta descrito con mayor detalle en el Ejemplo 2. En particular, el procedimiento puede minimizarse proporcionando procedimientos de ampliación adecuados (tales como la clonación y/o PCR) se emplean en la ultima etapa. Por lo tanto, pueden utilizarse técnicas que utilizan tecnología tal como la clasificación utilizando citómetros de flujo tal como se describe en la Figura 3C. Dichos procedimientos de clasificación están muy probados y son capaces de clasificar aproximadamente 5 a 30.000 gotitas por segundo para un equipo convencional, pero hasta 300.000 gotitas por segundo para los citómetros más avanzados.

Como se mencionó anteriormente, es posible que cada fragmento pueda indicar más de un solo elemento del código. Si por ejemplo cada fragmento indica 5 elementos del código, utilizando la tecnología existente y un banco de 32x100 componentes del banco, si se conectan 3.200 recipientes al dispositivo de clasificación ilustrado en la Figura 3C debería ser posible escribir varios miles de cadenas con 500 elementos del código por segundo. Evidentemente, un procedimiento que puede generar secuencias de ácido nucleico con tal rapidez ofrece ventajas significativas en los procedimientos conocidos en la técnica.

La molécula de ácido nucleico (es decir, la cadena de fragmentos) producida según el procedimiento descrito anteriormente y las moléculas monocatenarias de la misma comprenden más propiedades de la invención. Estas moléculas pueden, cuando proceda, ser incluidas en un vector.

Una vez producidas, las cadenas de fragmentos, en forma bicatenaria o monocatenaria, pueden utilizarse para almacenar información. En dichos casos se requiere unos medios apropiados de lectura de la información almacenada en estas cadenas. En algunas aplicaciones, las cadenas de fragmentos pueden dirigirse de manera apropiada a zonas específicas, por ejemplo mediante la unión a oligonucleótidos transportados sobre soportes sólidos que son complementarios a las prolongaciones en un terminal de las cadenas de fragmentos. Alternativamente pueden utilizarse el anticuerpo/antígeno apropiado, o los sistemas de reconocimiento de ADN:proteína. Por lo tanto, a continuación puede accederse a la información almacenada en las moléculas dirigidas de esta manera, o en la solución.

El documento WO 00/39333 describe técnicas apropiadas para dirigir y leer la información contenida en moléculas de ácido nucleico. De particular interés a este respecto son las técnicas en las que se detecta la fluorescencia de las sondas que llevan marcadores fluorescentes dirigidos a secuencias específicas. En dichas técnicas, las sondas que llevan marcadores como los descritos anteriormente en la presente memoria, pueden dirigirse a zonas de fragmentos específicas, particularmente a zonas que indican elementos del código. Las señales generadas (directa o indirectamente) por los marcadores pueden detectarse a continuación e identificarse de este modo el elemento del código. Si se utiliza un solo sistema binario solamente se requieren 2 marcadores discretos y puede determinarse su modelo de unión. Alternativamente, si un código más complejo se refleja en las cadenas de fragmentos, igualmente se requieren marcadores más discretos para la detección poco ambiciosa.

Por lo tanto en otro aspecto, la presente invención proporciona un procedimiento de identificación de los elementos del código contenidos en una molécula de ácido nucleico preparada tal como se describió anteriormente en la presente memoria (es decir cadena de fragmentos) en la que una sonda, portadora de unos medios de señalización (por ejemplo un marcador), específica para uno o más elementos del código, está unida a dicha molécula de ácido nucleico y se detecta una señal generada por dichos medios de señalización, por lo que puede identificarse dicho elemento o más elementos del código.

Preferentemente dichos medios de señalización son un marcador tal como se describió anteriormente en la presente memoria.

Una “sonda” tal como se hace referencia en la presente memoria se refiere a una molécula de ácido nucleico apropiada, por ejemplo de secuencias de ADN, ARN o APN, o híbridos de la misma, que puede unirse a una molécula diana de ácido nucleico (que puede ser monocatenaria o bicatenaria) por interacciones específicas, es decir, es específica para elementos del código específicos, por ejemplo mediante unión complementaria a una secuencia específica. Las sondas pueden ser de cualquier longitud conveniente, que permita la unión específica, por ejemplo del orden de 5 a 50 bases, preferentemente de 8 a 20 bases de longitud.

Unos “medios de señalización” tales como se utiliza en la presente memoria se refieren a unos medios para generar una señal directa o indirectamente. Una señal puede ser cualquier propiedad física o química que pueda detectarse, por ejemplo presencia de un producto específico, color, fluorescencia, radiación, magnetismo, paramagnetismo, carga eléctrica, tamaño o volumen. Preferentemente el marcador es un fluoróforo cuya fluorescencia se detecta. En dichos casos para la detección del marcador pueden utilizarse escáneres de fluorescencia y de este modo la identificación de los elementos del código.

Un elemento del código específico o una combinación de elementos puede identificarse mediante la aparición de una señal específica. Evidentemente la posición de cada señal es crucial para determinar la secuencia de los elementos del código. Como consecuencia se utilizarían los procedimientos en los que puede obtenerse la información de la posición (absoluta o relativa). En el documento WO 00/39333 se describen técnicas apropiadas, por ejemplo utilizando moléculas diana que se han acoplado a un soporte sólido en un extremo.

ES 2 324 513 T3

Existen numerosas aplicaciones para las cadenas de fragmentos una vez producidas en nano y pico-tecnología, entre otras por ejemplo por alargamiento de las cadenas de fragmentos mediante una corriente de líquido, electricidad u otra tecnología y utilizándolas como plantillas para nano y pico-estructuras. Pueden utilizarse también productos para marcar productos que pueden identificarse a continuación para crear su identidad. Alternativamente, pueden utilizarse moléculas para almacenar la información, por ejemplo fotografías, texto, música o como almacenamiento de datos en ordenadores de ADN. La producción rápida y las técnicas de lectura hacen posibles dichas aplicaciones por primera vez.

Los ejemplos siguientes solamente en los que las Figuras referidas son las siguientes se dan a título de ilustración:

La Figura 1 presenta la producción de una cadena de fragmentos que utiliza 8 fragmentos de partida “0” y “1” con diferentes prolongaciones;

la Figura 2 presenta la producción de una cadena de 64 fragmentos en la que se producen 8 cadenas que comprenden 8 fragmentos cada una en la que los terminales de las cadenas 1 y 2, y 2 y 3, etc. son complementarias de modo que pueden ligarse;

la Figura 3 presenta 3 técnicas para mezclar los fragmentos “0” y “1” de un banco de fragmentos ordenados para cada posición, en las que A) se seleccionan fragmentos apropiados por aspiración de los pocillos apropiados, B) se liberan los fragmentos apropiados en los pocillos del banco y C) se utiliza un citómetro de flujo para dirigir gotitas apropiadas a la cámara de mezclado;

la Figura 4 presenta la ampliación por PCR de la cadena señal 1-0-1-0-0 utilizando los cebadores SP6 y T7. Banda 1: 1 μ g de un escalón de ADN de 1 kb (Gibco BRL), Banda 2: 10 μ l de ADN de la cadena de fragmentos ampliada por PCR utilizando los cebadores SP6 y T7. Banda 3: igual que la banda 2 excepto para la utilización de los cebadores SP6 y T7-Cy5; y

la Figura 5 presenta la utilización de pares de cebadores durante el procedimiento de ampliación para unir cadenas de fragmentos.

Ejemplo 1

Automatización y miniaturización de la síntesis de la cadena

Este procedimiento describe un proceso rápido para mezclar fragmentos “0” y “1” apropiados con las prolongaciones correctas para producir una cadena específica constituida por los “0” y “1”.

Se producen dos bancos, uno con fragmentos “0” y otro con fragmentos “1”. Como se menciona en la descripción, se generan con prolongaciones que pueden estar ligadas a prolongaciones correspondientes para los fragmentos en posiciones adyacentes. Estos elementos separados se presentan en pocillos separados para formar el banco, de modo que los fragmentos en la posición 1 están presentes en el pocillo 1, los fragmentos en la posición 2 están presentes en el pocillo 2 y así sucesivamente. Los dos bancos proporcionan de este modo alternativas para cada posición. Con el fin de generar la cadena es necesario por consiguiente seleccionar solamente el fragmento correcto “0” o “1” para la posición 1, y a continuación la posición 2, etc. Ya que estos fragmentos, como consecuencia de sus únicas prolongaciones pueden hibridarse solamente con fragmentos para las posiciones adyacentes, es necesario seleccionar solamente los fragmentos correctos, a continuación mezclar y ligar estos fragmentos simultáneamente. En la Figura 3 se presentan diferentes modos de conseguir este efecto que permiten tres alternativas diferentes de mezclado.

En la Figura 3A, por ejemplo para producir la cadena 0-1-0-0-1, se utiliza el aparato para aspirar del banco “0” en las posiciones 1, 3 y 4 y aspirar del banco “1” en las posiciones 2 y 5. Los líquidos que se han aspirado pueden mezclarse a continuación junto con la ligasa y un tampón apropiado. En la alternativa B, cada pocillo en el banco está conectado a un tubo/boquilla que puede cerrarse/abrirse electrónicamente. El líquido de las boquillas se dirige a la cámara de ligadura junto con ligasa y un tampón apropiado. Pueden diferentes cadenas construirse cambiando de manera apropiada el modelo de boquillas que se abren/cierran.

El procedimiento pueden miniaturizarse también, por ejemplo utilizando la tecnología de citómetro de flujo tal como se ilustra en la Figura 3C. En este procedimiento, los componentes del banco se almacenan en recipientes en la parte superior de la “maquina de escritura”. Gotitas de cada recipiente se guían a continuación al pocillo de residuos o de producción dependiendo de la naturaleza de la cadena que debe construirse. El mecanismo de guía es tal como se utiliza en los citómetros de flujo ordinarios, es decir, se cargan las gotitas cuando dejan el recipiente y pueden guiarse electrónicamente en diferentes direcciones.

ES 2 324 513 T3

Ejemplo 2

Bancos que comprenden oligonucleótidos para su utilización en la invención

5 Por conveniencia, el procedimiento de clonación puede realizarse utilizando bancos que contienen oligonucleótidos. Por ejemplo un banco puede contener:

1. Oligonucleótidos con una parte común y 5 bases en el extremo 5' que varían para proporcionar todas las permutaciones posibles, es decir 1.024 variantes.
- 10 2. Oligonucleótidos con una parte común y 4 bases en el extremo 5' que varían para proporcionar todas las permutaciones posibles, es decir 256 variantes.
- 15 3. Oligonucleótidos con una parte común y 5 bases en el extremo 3' que varían para proporcionar todas las permutaciones posibles, es decir 1024 variantes.
4. Oligonucleótidos con una parte común y 6 bases en el extremo 3' que varían para proporcionar todas las permutaciones posibles, es decir 4.096 variantes.

20 En lo expuesto anteriormente, los oligonucleótidos se producen de tal modo que todos los oligonucleótidos "1" son complementarios a los oligonucleótidos "2" en virtud de las bases invariantes, es decir para generar una molécula bicatenaria con prolongaciones de 4/5 bases variantes. Asimismo son complementarios los oligonucleótidos "3" y "4".

25 Los oligonucleótidos combinados de este modo (es decir, con prolongaciones en su extremo de 4 a 6 bases pueden también combinarse con oligonucleótidos bicatenarios complementarios generados también combinando determinados miembros del banco. De esta manera las prolongaciones variables de diferentes longitudes pueden crearse en la molécula resultante, por ejemplo una molécula con una prolongación de 4 bases en ambos extremos 3' y 5'.

30 Puede también proporcionarse oligonucleótidos en el banco que permiten que se unan adaptadores 5' y 3'. De este modo, por ejemplo, pueden proporcionarse oligonucleótidos que presentan la siguiente forma:

5. 5'-AAAA-[comp1]-FFFFF-3'
6. 5'-DDDDD-[comp1]-FFFFF-3'
- 35 7. 5'-AAAA-[comp1]-HHHHHH-3'
8. 5'-DDDDD-[comp1]-HHHHHH-3'
- 40 9. 3'-[comp1*]-5'
10. 5'-BBBB-[comp2]-3'
11. 5'-EEEE-[comp2*]-3'
- 45 12. 5'-[comp3]-GGGG-3'
13. 5'-[comp3*]-IIIII-3'

50 en los que "comp_x" se refiere a una zona que es complementaria de la zona "comp_x*", es decir, "5", "6", "7" u "8" pueden unirse a "9". Además, "comp₂" puede unirse al oligonucleótido 1 anterior, "comp_{2a}" puede unirse al oligonucleótido 2, "comp₃" puede unirse al oligonucleótido 4 y "comp₃*" puede unirse al oligonucleótido "3". Las bases indican que "A" se une a "B", es decir "7" y "10" pueden unirse a sus extremos. Asimismo "D" se une a "E", "F" se une a "G" y "H" se une a "I". (Estas bases cuando están juntas pueden tener un contenido variable, por ejemplo AAAA = GAGA y entonces BBBB = TCTC).

55 Mediante la utilización apropiada de los enlazadores descritos anteriormente pueden combinarse los adaptadores 5' y 3'. Por ejemplo, el oligonucleótido "2" con la prolongación 5' de la base 4 específica puede unirse mediante su zona complementaria a un enlazador "11" oligonucleotídico que a continuación dejará un solapamiento "EEEE". Éste puede estar unido al oligonucleótido "8" por el solapamiento que puede por sí mismo unirse al oligonucleótido "9" mediante su zona complementaria. El solapamiento "HHHHHH" puede unirse al oligonucleótido "13" que puede acoplar un oligonucleótido "4" por la unión a la zona complementaria. Estas diversas permutaciones pueden realizarse que dan como resultado varias longitudes de solapamiento, por ejemplo cualquier combinación de 4, 5 ó 6 solapamientos de bases que pueden estar en la misma o diferentes cadenas.

65

ES 2 324 513 T3

Ejemplo 3

Identificación de conjuntos de oligonucleótidos con prolongaciones de 6 pares de bases con ligaduras mínimas desemparejadas

5

Con el fin de identificar conjuntos de oligonucleótidos con prolongaciones de 6 pares de bases que es improbable que formen ligaduras de desemparejamiento entre sí pueden realizarse las etapas siguientes.

10

1. Crear los 2.048 pares de prolongaciones de 6 bases.

2. Eliminar los 32 pares palindrómicos.

Esto produce un conjunto final de 2.016 pares que sobresalen.

15

Parte 1

Tomar un par como par nº 1 y seleccionar el par siguiente ejecutando el apartado 1.

20

Apartado 1

Algoritmo 1

25

Calcular las (2016-n) tablas de puntuaciones desemparejadas no ponderadas entre los n par(es) ya seleccionados y los (2016-n) pares restantes y hallar entre los últimos par(es) los que la puntuación menor en la tabla es la mayor (véase a continuación para detalles acerca del cálculo de la puntuación). Si existe solamente dicho par, entonces seleccionarlo. Si existen varios pares, a continuación calcular las puntuaciones mal emparejadas ponderadas de las comparaciones de la prolongación que dan la menor puntuación no ponderada y hallar el/los par(es) para los que la menor puntuación ponderada es la mayor. Si existe solamente uno de dichos pares, entonces seleccionarlo. Si existen varios pares, entonces rehacer el procedimiento completo utilizando la segunda puntuación menor no ponderada en la tabla de desemparejamientos, a continuación la tercera menor y así sucesivamente. Si varios pares permanecen unidos después de todas las puntuaciones equivocadas que se han considerado, mantenerlos todos.

30

35

Repetir el algoritmo 1 para cada par seleccionado e iterarlo en el número deseado de posiciones para obtener la(s) cadena(s) de pares que sobresalen. Este procedimiento genera un árbol con un par que sobresale en cada rama. Se calculan las puntuaciones más bajas no ponderadas y mal emparejadas ponderadas de la combinación específica de pares de cada punto. Se interrumpe una serie de reacciones específica (1) cuando se alcanza el número deseado de posiciones, o (2) cuando la combinación de pares es la que ya se ha encontrado al principio, o (3) cuando las puntuaciones mal emparejadas más bajas de esta combinación son menores que las puntuaciones más bajas de la(s) cadena(s) completa(s) ya construidas. El punto (3) asegura que cada nueva cadena completa siempre tiene puntuaciones mal emparejadas más pequeñas que son mayores o por lo menos iguales a las de la(s) cadena(s) construida(s) anteriormente. Obsérvese también que, como resultado de este proceso, todos los pares en una cadena dada son exclusivos y todas las cadenas completas en el árbol son únicas. El proceso total termina cuando se interrumpe la última serie de reacciones que ha de explorarse. Mantener la(s) cadena(s) completa(s) cuyas puntuaciones mal emparejadas más bajas son las mayores.

40

45

Repetir el apartado 1 empezando en cada uno de los 2016 pares como par nº 1 para producir una serie de 2016 cadenas que sobresalen. Encontrar la(s) mejor(es) cadena(s) aplicando el algoritmo 2.

50

Algoritmo 2

55

Para todas las cadenas, calcular las tablas de puntuaciones mal emparejadas no ponderadas entre todos los pares que están presentes en la cadena, y encontrar la(s) cadena(s) para la(s) que la puntuación más baja en la tabla es la mayor (véase a continuación para detalles). Si existe solamente una cadena tal, entonces seleccionarla. Si existen varias cadenas, entonces calcular las puntuaciones mal emparejadas ponderadas de las comparaciones que sobresalen que dan la puntuación más baja no ponderada y encontrar la(s) cadena(s) para la que la puntuación ponderada menor es la mayor. Si existe solamente una de dichas cadenas, entonces seleccionarla. Si existen varias cadenas, entonces rehacer el procedimiento completo utilizando la segunda puntuación más baja no ponderada en la tabla de que sobresalen, a continuación la tercera más baja, y así sucesivamente. Si varias cadenas permanecen enlazadas después de que se han considerado todas las puntuaciones mal emparejadas, entonces conservarlas todas.

60

Esto permite la producción de un conjunto de una o más cadenas que sobresalen.

65

ES 2 324 513 T3

Parte 2

Tomar una cadena y ejecutar el apartado 2.

5 Apartado 2

Algoritmo 3

10 Para esta cadena, hallar el/los par(es) que sobresale(s) que es(son) responsable(s) de las puntuaciones no ponderadas y ponderadas menores en la tabla de puntuaciones desemparejadas entre todos los pares en la cadena. A continuación, crear nuevas cadenas sustituyendo este par con todos los pares que sobresalen restantes que no están presentes en la cadena original (si existen varios pares que han de sustituirse, sustituir un par una vez). A partir del conjunto completo de cadenas recién generadas y de la cadena original, seleccionar una o más cadenas siguiendo el algoritmo 2. Entonces, incluyendo la cadena original en el algoritmo 2 se asegura que las cadenas seleccionadas siempre tienen una puntuación mal emparejada que es mayor o por lo menos igual a la puntuación de la cadena original. La mejora (si existe) puede implicar la puntuación menor o de orden n menor no ponderada, o la puntuación ponderada correspondiente.

20 Repetir el algoritmo 3 para cada cadena seleccionada. Este procedimiento genera un árbol con una cadena en cada rama. Cada nueva cadena que se añade al árbol tiene una puntuación mal emparejada mayor o igual a la puntuación de la cadena hallada en la etapa previa. Se interrumpe una serie de reacciones específicas cuando la cadena seleccionada es la que se ha encontrado ya al principio. Esto asegura que todas las cadenas en el árbol son únicas. El procedimiento completo termina cuando se interrumpe la última serie de reacciones que debe explorarse. Mantener todas las cadenas que están presentes en el árbol.

25 Repetir el apartado 2 (es decir, construir un árbol) partiendo de cada una de las cadenas seleccionadas al final de la parte 1.

30 A partir de la serie completa de cadenas presentes en todos los árboles, seleccionar una o más cadenas siguiendo el algoritmo 2.

Esto produce una serie final de una o más cadenas que sobresalen.

Cálculo de puntuaciones mal emparejadas

35 *Puntuación no ponderada*

La puntuación no ponderada para una ligadura entre dos prolongaciones de 6 bases es el número de incompatibilidades observadas, considerando los tripletes de las primeras 3 bases y las últimas 3 bases por separado. Por ejemplo, la puntuación para la ligadura AAAAAC/TTTGCA es de 0 a 3 y la puntuación para AAAAC/TCAGGG es 2-2. Todas las puntuaciones posibles se ordenan desde la mayor a la menor según el orden siguiente:

mayor: 3-3

45 3-2/2-3

2-2

3-1/1-3

50 2-1/1-2

1-1

3-0/0-3

55 2-0/0-2

menor: 1-0/0-1

60 *Puntuación ponderada*

La puntuación ponderada (WS) para una ligadura se calcula de la manera siguiente:

65
$$WS = 6 - \sum_{i=1}^6 BPS_i$$

ES 2 324 513 T3

en la que PBS_i es la puntuación para el par de bases concreto en el punto i y se da en la tabla siguiente:

AA = 1,0	CA = 0,6	GA = 1,0	TA = 0,0
AC = 0,6	CC = 1,0	GC = 0,0	TC = 0,6
AG = 1,0	CG = 0,0	GG = 0,9	TG = 0,2
AT = 0,0	CT = 0,6	GT = 0,2	TT = 0,6

para el perfecto emparejamiento entre una prolongación y su complemento, WS = 6.

Comparación entre pares y construcción de tablas de puntuaciones

Encontrar el siguiente par que sobresale

Para seleccionar el siguiente par que sobresale, se calculan las tablas de puntuaciones con compatibilidad entre los pares seleccionados en las posiciones anteriores y se calculan todos los pares restantes. Para construir dicha tabla, se comparan todos los pares anteriormente seleccionados con un nuevo par y también cada prolongación se compara con ella misma. De este modo, si se han seleccionado ya n pares, el número de ligaduras consideradas para cada tabla es $4n + 2(n+1) = 6n+2$. Cuando se comparan dos prolongaciones que están en la misma cadena de ADN, una de ellas está invertida.

Se considera el ejemplo siguiente en el que los pares AAAAAC/TTTTTG (1A/1B) y AAACGT/TTTGCA (2A/2B) se han seleccionado previamente y el nuevo par AGTCCC/TCAGGG (3A/3B) se selecciona en la posición siguiente:

La tabla correspondiente es:

Comparación	Prolongación	Ligadura	Puntuación no ponderada	Puntuación ponderada
1 vs 1	1A	AAAAAC	3-3	0,8
	1A	CAAAAA		
	1B	TTTTTG	3-3	3,2
	1B	GTTTTT		
2 vs 2	2A	AAACGT	2-2	2,8
	2A	TGCAAA		
	2B	TTTGCA	2-2	4,4
	2B	ACGTTT		
3 vs 3	3A	AGTCCC	2-2	3,6
	3A	CCCTGA		
	3B	TCAGGG	2-2	3,6
	3B	GGGACT		
1 vs 3	1A	AAAAAC	3-2	2,6
	3A	CCCTGA		
	1A	AAAAAC	2-2	2,4
	3B	TCAGGG		
	1B	TTTTTG	2-2	4,0
	3A	AGTCCC		
	1B	TTTTTG	3-2	4,6
	3B	GGGACT		
2 vs 3	2A	AAACGT	3-2	2,7
	3A	CCCTGA		
	2A	AAACGT	2-2	3,3
	3B	TCAGGG		
	2B	TTTGCA	2-2	3,6
	3A	AGTCCC		
	2B	TTTGCA	3-2	3,4
	3B	GGGACT		

En la presente memoria, la puntuación más baja es 2-2; 2,4 dada por la ligadura entre las prolongaciones 1A y 3B.

ES 2 324 513 T3

Tabla de puntuación para una cadena

5 Para calcular la tabla de las puntuaciones mal emparejadas para una cadena, todos los pares que sobresalen contenidos en la cadena se comparan entre sí y también cada prolongación se compara con ella misma. De este modo, para una cadena de p pares que sobresalen, el número de ligaduras consideradas es $4p(p-1) / 2 + 2p = 2(p^2)$. Como anteriormente, una de las dos prolongaciones se invierte en la comparación cuando ambas son de la misma cadena de ADN.

10 Por ejemplo, se considera la siguiente cadena de 3 pares (es decir, la posición 4): AAAAAC/TTTTTG (1A/1B), AAACGT/TTTGCA (2A/2B), AGTCCC/TCAGGG (3A/3B) en la que 1A está en un fragmento, 1B y 2A están en un segundo fragmento, 2B y 3A están en un tercer fragmento y 3B está en un cuarto fragmento.

La tabla correspondiente es:

Comparación	Prolongación	Ligadura	Puntuación no ponderada	Puntuación ponderada
1 vs 1	1A	AAAAAC	3-3	0,8
	1A	CAAAAA		
	1B	TTTTTG	3-3	3,2
	1B	GTTTTT		
2 vs 2	2A	AAACGT	2-2	2,8
	2A	TGCAAA		
	2B	TTTGCA	2-2	4,4
	2B	ACGTTT		
3 vs 3	3A	AGTCCC	2-2	3,6
	3A	CCCTGA		
	3B	TCAGGG	2-2	3,6
	3B	GGGACT		
1 vs 2	1A	AAAAAC	2-3	1,8
	2A	TGCAAA		
	1A	AAAAAC	0-3	3,8
	2B	TTTGCA		
	1B	TTTTTG	0-3	5,0
	2A	AAACGT		
	1B	TTTTTG	2-3	3,8
	2B	ACGTTT		
1 vs 3	1A	AAAAAC	3-2	2,6
	3A	CCCTGA		
	1A	AAAAAC	2-2	2,4
	3B	TCAGGG		
	1B	TTTTTG	2-2	4,0
	3A	AGTCCC		
	1B	TTTTTG	3-2	4,6
	3B	GGGACT		
2 vs 3	2A	AAACGT	3-2	2,7
	3A	CCCTGA		
	2A	AAACGT	2-2	3,3
	3B	TCAGGG		
	2B	TTTGCA	2-2	3,6
	3A	AGTCCC		
	2B	TTTGCA	3-2	3,4
	3B	GGGACT		

65 En la presente memoria, la puntuación más baja es 0-3; 3,8 proporcionada por la ligadura entre las prolongaciones 1A y 2B.

ES 2 324 513 T3

Resultados obtenidos

TABLA DE PUNTOS DE ROTURA

5 Parte 1

Nº de posiciones	Puntuación no ponderada	Puntuación ponderada	nº de cadenas iguales
3	3-3	1,6	48
4	2-2	4,0	48
9	2-2	2,5	12
10	3-1	3,2	12
14	3-1	2,4	6
15	2-1	4,6	6
33	2-1	3,0	12
34	3-0	4,6	12
90	3-0	3,1	

Parte 2

Nº de posiciones	Puntuación no ponderada	Puntuación ponderada	nº de cadenas iguales
3	3-3	1,6	48
4	3-2	2,2	48
9	2-2	2,5	12
10	3-1	3,2	12
14	3-1	2,4	6
15	3-1	2,0	6
33	2-1	3,0	12
34	3-0	4,6	12
90			

40 Debe apreciarse que la puntuación mal emparejada no ponderada (en la que 9 = 3-3, 8 = 3-2, 7 = 2-2, 6 = 3-1, 5 = 2-1, 4 = 1-1, 3 = 3-0, 2 = 2-0, 1 = 1-0) reduce a medida que aumenta el número de posiciones.

45 *Muestras de las cadenas obtenidas al final de la parte 1 y al final de la parte 2*

3 posiciones (esta cadena se obtiene al final de ambas partes):

50 AACTCG/TTGAGC
TCTCAC/AGAGTG

4 posiciones:

55 parte 1

60 AATTGG/TTAACC
TGCCAC/ACGGTG
ATAGTC/TATCAG

parte 2

65 AATGGG/TTACCC
TCGGAC/AGCCTG
TTAACG/AATTGC

ES 2 324 513 T3

9 posiciones (esta cadena puede obtenerse al final de ambas partes):

5	AATCAC/TTAGTG TGAGGG/ACTCCC TCGGAT/AGCCTA	TACACG/ATGTGC ACATTG/TGTAAG GGCTAG/CCGATC	AGGCTG/TCCGAG TTTAGC/AAATCG
---	---	---	--------------------------------

10 posiciones (esta cadena puede obtenerse al final de ambas partes):

10	AAAACC/TTTTGG TTGGGG/AACCCC TCATAG/AGTATC	AGGCTC/TCCGAG GTCATG/CAGTAC TGCAGT/ACGTCA	TCGATA/AGCTAT ATTCAG/TAAGTC AGAGAT/TCTCTA
----	---	---	---

15 posiciones (esta cadena puede obtenerse al final de ambas partes):

20	ACGTGC/TGCACG TATGAG/ATACTC TGCACG/ACGTGC ATACAC/TATGTG AACTTG/TTGAAC	GTTGGC/CAACCG TTGCGG/AACGCC AGTATC/TCATAG TGACTA/ACTGAT ACTCCG/TGAGGC	TCAGCC/AGTCGG AGAGGG/TCTCCC CACCGC/GTGGCG
----	---	---	---

25 posiciones:

parte 1:

30	AAAACC/TTTTGG TTGGGG/AACCCC TCATAG/AGTATC AGGCTC/TCCGAG GTCATG/CAGTAC	TGCAGT/ACGTCA TCGATA/AGCTAT ATTCAG/TAAGTC AGAGAT/TCTCTA TACTTC/ATGAAG	AAGTAA/TTCATT CCGTCC/GGCAGG TGTAAC/ACATTG ACCGTG/TGGCAC
----	---	---	--

35 posiciones:

40	AAAACC/TTTTGG TTGGGG/AACCCC TCATAG/AGTATC AGGCTC/TCCGAG GACAAG/CTGTTC	TCTGCT/AGACGA TCGATA/AGCTAT ATTCAG/TAAGTC AGAGAT/TCTCTA TACTTC/ATGAAG	AAGTAA/TTCATT CCGTCC/GGCAGG TGTAAC/ACATTG ACCGTG/TGGCAC
----	---	---	--

45 posiciones (esta cadena puede obtenerse al final de ambas partes):

50	AACTAG/TTGATC TGGAGC/ACCTCG TCAAAT/AGTTTA CAGGCC/GTCCGG TATCAC/ATAGTG AGATTC/TCTAAG TTCCGT/AAGGCA CCCACG/GGGTGC ATGCCG/TACGGC TCCGTC/AGGCAG CCACGC/GGTGCG ACTATG/TGATAC TTAGCA/AATCGT	GTAAGG/CATTCC AAACTA/TTTGAT GTCTCC/CAGAGG ACAGCG/TGTGCG CACATC/GTGTAG TGTGTA/ACACAT TAATGC/ATTACG GGTAAG/CCATTC AGTTAT/TCAATA CAACAG/GTTGTC ATCGGC/TAGCCG AATGCT/TTACGA TTGGAG/AACCTC	TCGCCT/AGCGGA TCTCGG/AGAGCC ACCCCC/TGGGGG TTTTCG/AAAAGC AAGTCA/TTCAGT GTTCTC/CAAGAG
----	---	---	--

60 posiciones (esta cadena puede obtenerse al final de ambas partes):

65	AACTCT/TTGAGA TCGAAC/AGCTTG	TTATTG/AATAAG CACAAAG/GTGTTC	CCAATC/GGTTAG ACTTAT/TGAATA
----	--------------------------------	---------------------------------	--------------------------------

ES 2 324 513 T3

5 CAGGGC/GTCCCG TAAAGG/ATTTCC TGTGCG/ACACGC ATGTAG/TACATC TTCCCC/AAGGGG AATCTC/TTAGAG TGGCGT/ACCGCA 10 GGCTGC/CCGACG TTTGAC/AAACTG ACACCG/TGTGGC 15 TGAGGC/ACTCCG	TCCGAT/AGGCTA AGTAGC/TCATCG CCGTCG/GGCAGC TCACTA/AGTGAT GTGACG/CACTGC TGAAAT/ACTTTA AGCATG/TCGTAC ACCGTC/TGGCAG	AAAGAG/TTTCTC TTGATA/AACTAT AAGACC/TTCTGG CAATCC/GTTAGG TCTCGC/AGAGCG AGGGGG/TCCCCC TGCCAG/ACGGTC TACTAC/ATGATG
--	--	--

90 posiciones (esta cadena puede obtenerse al final de la parte 1):

20 AAAAAA/TTTTTT CCGGCC/GGCCGG AGGTAG/TCCATC TCCATC/AGGTAG 25 ATCTGC/TAGACG TAGACG/ATCTGC ACTGTG/TGACAC TGACAC/ACTGTG CATTAC/GTAATG 30 ACCCCA/TGGGGT ATGGTA/TACCAT CGAAGC/GCTTCG ATTACC/TAATGG 35 TAATGG/ATTACC CTCCTC/GAGGAG AGTTGA/TCAACT AATGCT/TTACGA 40 TTACGA/AATGCT AAGCGC/TTCGCG TTCGCG/AAGCGC CCCAAG/GGGTTC 45 GGGTTC/CCCAAG ACATCC/TGTAGG TGTAGG/ACATCC AACTTG/TTGAAC 50 TTGAAC/AACTTG ATAGAC/TATCTG TATCTG/ATAGAC AGACCG/TCTGGC 55 TCCGAG/AGGCTC CGGGGC/GCCCCG	TCTGGC/AGACCG ACGCAG/TGCGTC TGCGTC/ACGCAG AGTCAT/TCAGTA TCAGTA/AGTCAT CAGCCG/GTCGGC GTCGGC/CAGCCG AATTTT/TTAAAG TTAAAG/AATTTT CCAACG/GGTTGC GGTTCG/CCAACG CACCAC/GTGGTG AGAATA/TCTTAT TCTTAT/AGAATA ATCAAT/TAGTTA TAGTTA/ATCAAT ACTTCA/TGAAGT AGCCCC/TCGGGG TCGGGG/AGCCCC ACCATG/TGGTAC TGGTAC/ACCATG AGGGGA/TCCCCT CTAATC/GATTAG CGAGAG/GCTCTC GCTCTC/CGAGAG ACACGT/TGTGCA TGTGCA/ACACGT CCTGTC/GGACAG GGACAG/CCTGTC	AAACGG/TTTGCC TTTGCC/AAACGG AACCAA/TTGGTT CAAAAC/GTTTTG AAGGAA/TTCCCT CGCCGC/GCGGCG AGTGCG/TCACGC TCACGC/AGTGCG ATTTTA/TAAAAT ATCCTA/TAGGAT AGTATC/TCATAG TCATAG/AGTATC ATGTGG/TACACC TACACC/ATGTGG ATGCAC/TACGTG TACGTG/ATGCAC ACTAAC/TGATTG TGATTG/ACTAAC CAGTGC/GTCACG GTCACG/CAGTGC AATAAG/TTATTC TTATTC/AATAAG AGATAT/TCTATA TCTATA/AGATAT AAGTCG/TTCAGC TTCAGC/AAGTCG AATCGA/TTAGCT TTAGCT/AATCGA AGGCTC/TCCGAG
---	---	---

60 Ejemplo 4

Construcción de una cadena de 5 fragmentos que codifica la secuencia binaria 1-0-1-0-0

65 Este experimento demuestra la construcción de una cadena específica de 5 fragmentos utilizando un conjunto de cuatro pares que sobresalen de 6 bases 5' no palindrómicas. El conjunto de cuatro únicos pares que sobresalen se halló utilizando un programa informático tal como se describe en el Ejemplo 5.

ES 2 324 513 T3

Basándose en los pares que sobresalen, se construyó una serie de cinco componentes del banco hibridando oligonucleótidos complementarios en tubos separados:

5 **señal 1:**

5' -TAATACGACTCACTATAACCACAAGTTTGTACAAAAAAGCAGGCTCTATTC-3' y
y 5'-TAGGAAGAATAGAGCCTGCTTTTTTTGTACAAACTTGTGGTATAGTGAGTCGTATTA-3';

10 **señal 2:**

5'-TTCCTATGCAGTGGACCACTTTGTACAAGAAAGCTGGGTTGCAGT-3' y
5'-GCAACTACTGCAACCCAGCTTTCTTGTACAAAGTGGTCCACTGCA-3';

15 **señal 3:**

5'-AGTTGCTTGACGCCACAAGTTTGTACAAAAAAGCAGGCTTTGACG-3' y
20 5'-CGACATCGTCAAAGCCTGCTTTTTTTGTACAAACTTGTGGCGTCAA-3';

señal 4:

25 5'-ATGTCGAAGGGCGGACCACTTTGTACAAGAAAGCTGGGTAAGGGC-3' y
5'-GACAGGGCCCTTACCCAGCTTTCTTGTACAAAGTGGTCCGCCCTT-3';

señal 5:

30 5'-CCTGTCATGTGGACCACTTTGTACAAGAAAGCTGGGTTTCTATAGTGTCACCT
AAATC-3' y 5'-GATTTAGGTGACACTATAGAAACCCAGCTTTCTTGTACAAAGTGGTCCAC
AT-3';

35 **T7: 5'-TAATACGACTCACTATACCA-3'**

cebador T7-Cy5: 5'-TAATACGACTCACTATA-3'

40 **cebador SP6: 3'-AAGATATCACAGTGGATTTAG-5'**

Los componentes del banco (de 4 pmoles cada uno) se mezclaron a continuación y se ligaron utilizando 100 U de T4 ADN ligasa (NEB) en 1x tampón de ligasa a 25°C durante 15 minutos. La ligasa se inactivó a continuación a 65°C durante 20 min.

Se utilizaron 5 µl de la reacción de ligadura (50 µl) como plantilla en una reacción de PCR (50 µl) que contenía 1x tampón Thermopol (NEB), dNTP 0,05 mM, cebador T7 0,4 µM, cebador SP6 0,4 µM y 0,04 U/µl de Vent polimerasa (NEB). La PCR se inició en caliente (95°C durante 3 minutos antes de la adición de polimerasa) y se cicló 30 veces; 95°C 30 s; 55°C 30 s; 76°C, 30 s, utilizando un termociclador PTC-200 (MJ Research). Se analizaron 10 µl de la PCR en un gel de agarosa al 1,5% como se representa en la Figura 4. La fotografía del gel mostraba solamente una banda intensa correspondiente a aproximadamente 240 bp como cabía esperar (243 bp). El producto de PCR restante se extrajo dos veces con cloroformo y se precipitó utilizando etanol al 71% y NaAc 0,1 M. Se disolvió el ADN en agua y se secuenció. La secuencia confirmó que se generaba la cadena señal esperada (1-0-1-0-0).

Ejemplo 5

60 *Construcción de una cadena de 5x5 fragmentos que codifica la secuencia binaria utilizando un ciclo con ligadura seguido de un ciclo PCT o de dos ciclos de ligadura*

Este experimento demuestra la utilización de pares de cebadores complementarios para unirse a cadenas de fragmentos como alternativa a la estrategia de ligadura demostrada en el ejemplo anterior.

65 En este experimento 5 fragmentos de cadenas con 5 posiciones (fragmentos o bits) se ligan cada una por separado en el ciclo de ligadura 1 como se demostró al principio (Ejemplo 6). Estas cadenas de 5 fragmentos se amplían a continuación con 5 pares de cebador diferentes (se utiliza el par 1 para ampliar la cadena 1, el par 2 se utiliza para

ES 2 324 513 T3

ampliar la cadena 2, etc.). El segundo cebador en el par cebador 1 es complementario con el primer cebador en el par 2 cebador, el segundo cebador en el par 2 cebador es complementario con el primer cebador en el par 3 cebador, y así sucesivamente.

5 Se extrae a continuación una pequeña alícuota de cada una de las 5 reacciones de PCR y se lleva a cabo nuevas reacciones de PCR con los cebadores que son específicos al final de la cadena 1 y 5 señal. El procedimiento se ilustra en la Figura 5.

Materiales

10 Se seleccionan oligonucleótidos que se unen a la cadena del fragmento y actúan también como cebadores. Por lo tanto, por ejemplo, para las cadenas adyacentes puede estar unido utilizando por ejemplo los pares de cebador siguientes:

15 fragmento terminal de la cadena 2 (*con cebador unido*):

TTCTATAGTGTCACCTAAATC
20 AAGATATCACAGTGGATTTAGCCTACCAGTACATCCAACGGCAACT

fragmento terminal de la cadena 3 (*con cebador unido*):

25 GTCATGTAGGTTGCCGTTGATCCATCCTAATACGACTCACTATAGCA
ATTATGCTGAGTGATATCGT

Las zonas del cebador ampliado anteriores son complementarias y pueden unirse de este modo.

30 Como alternativa a este procedimiento, pueden utilizarse dos ciclos de ligadura en los que las cadenas de 5 fragmentos (generadas por ligadura) se ligan. Por lo tanto, varios ciclos de construcción para construir cadenas largas con señal. Después de la ligadura inicial en el primer ciclo de ligadura las cadenas de 5 fragmentos se amplían a continuación con cebadores que contienen una secuencia *FokI*. Se seleccionan cebadores de manera apropiada de modo que la digestión con *FokI* construya entonces prolongaciones no palindrómicas al final de cada cadena de fragmentos en la
35 que la prolongación generada en la cadena de 1 fragmento puede ligarse con la primera prolongación generada en la cadena de 2 fragmentos, el segundo saliente generado en la cadena de 2 fragmentos puede ligarse con el primer saliente generado en la cadena de 3 fragmentos, y así sucesivamente. La cadena de 5 fragmentos puede por consiguiente ligarse de manera controlada para generar una cadena final con 25 fragmentos (bits).

40 Si se quieren construir cadenas de fragmentos con 100 ó 500 fragmentos se puede repetir este procedimiento 1 ó 2 veces más. La capacidad de polimerasa, sin embargo, será un factor limitativo con respecto a cuántos ciclos de ligadura es posible realizar. Resultarán por lo tanto necesario utilizar otras estrategias para construir cadenas aun más largas.

45

50

55

60

65

REIVINDICACIONES

5 1. Procedimiento para la síntesis de una molécula de ácido nucleico bicatenario que contiene información en forma de código alfanumérico para almacenar información en forma de un código alfanumérico que comprende por lo menos las etapas siguientes:

10 i) generar n fragmentos de ácido nucleico bicatenario, en el que por lo menos n-2 fragmentos presentan zonas monocatenarias en ambos terminales y 2 fragmentos presentan zonas monocatenarias en por lo menos un terminal, en el que (n-1) zonas monocatenarias son complementarias con otras (n-1) zonas monocatenarias, produciendo de este modo (n-1) pares complementarios;

15 ii) poner en contacto dichos n fragmentos de ácido nucleico bicatenario, simultánea o consecutivamente, para efectuar la unión de dichos pares complementarios de las zonas monocatenarias; y

20 iii) ligar opcionalmente dichos pares complementarios simultánea o consecutivamente para producir una molécula de ácido nucleico constituida por n fragmentos;

en el que dichos fragmentos comprenden zonas que representan una unidad de información correspondiente a uno o más elementos del código.

25 2. Procedimiento según la reivindicación 1, en el que cada zona que representa una unidad del código alfanumérico está constituida por entre 4 y 10 bases.

3. Procedimiento según la reivindicación 1 ó 2, en el que cada fragmento está constituido por entre 8 y 25 bases.

4. Procedimiento según cualquiera de las reivindicaciones anteriores, en el que el código alfanumérico es binario.

30 5. Procedimiento según cualquiera de las reivindicaciones anteriores, en el que por lo menos 10 fragmentos de ácido nucleico bicatenario se hibridan conjuntamente en la etapa (i), para producir una molécula de ácido nucleico bicatenario que comprende por lo menos 10 fragmentos.

35 6. Procedimiento según cualquiera de las reivindicaciones anteriores, en el que una pluralidad de moléculas de moléculas de ácido nucleico bicatenario que comprenden una serie de fragmentos de ácido nucleico bicatenario se sintetizan y se unen de manera conjunta.

40 7. Procedimiento para identificar los elementos de código contenidos en una molécula de ácido nucleico preparada según un procedimiento según cualquiera de las reivindicaciones 1 a 6, en el que una sonda, portadora de unos medios de señalización, específica para uno o más elementos de código, está unida a dicha molécula de ácido nucleico y es detectada una señal generada por dichos medios de señalización, pudiendo identificarse uno o más elementos de código.

45

50

55

60

65

	fragmentos de partida "0"	fragmentos de partida "1"
Posición 1	GGGG GGGGAAA CCCCCCCCC	AAAAAAAAAAA TTTTTTTTT
Posición 2	GGGG GGGGAAC TTCCCCCCCCC	AAAAAAAAAAC TTTTTTTTTTT
	⋮	
Posición 7	GGGG GGGGCCG GCGCCCCCCCC	AAAAAAACCG GCGTTTTTTTT
Posición 8	GGGG GGGG GGCCCCCCCCC	AAAAAAA GGCTTTTTTTT

FIG. 1

	Fragmento 0	Fragmento 1
Posición 1.1	GGGG GGGGAAA CCCCCCCCC	AAAAAAAAAAA TTTTTTTTT
Posición 1.2	AAAGGGG GGGGAAA CCCCCCCCC	AAAAAAAAAAAAA TTTTTTTTT
Posición 1.3	AACGGGG GGGGAAA CCCCCCCCC	AACAAAAAAAAAA TTTTTTTTT
	⋮	
Posición 8.1	GGGG GGGG GCCCCCCCCCTTT	AAAAAAA GCTTTTTTTTTTT
Posición 8.2	GGGG GGGG GCCCCCCCCCTTG	AAAAAAA GCTTTTTTTTTTG
Posición 8.3	GGGG GGGG GCCCCCCCCCTTC	AAAAAAA GCTTTTTTTTTTC
	⋮	

FIG. 2

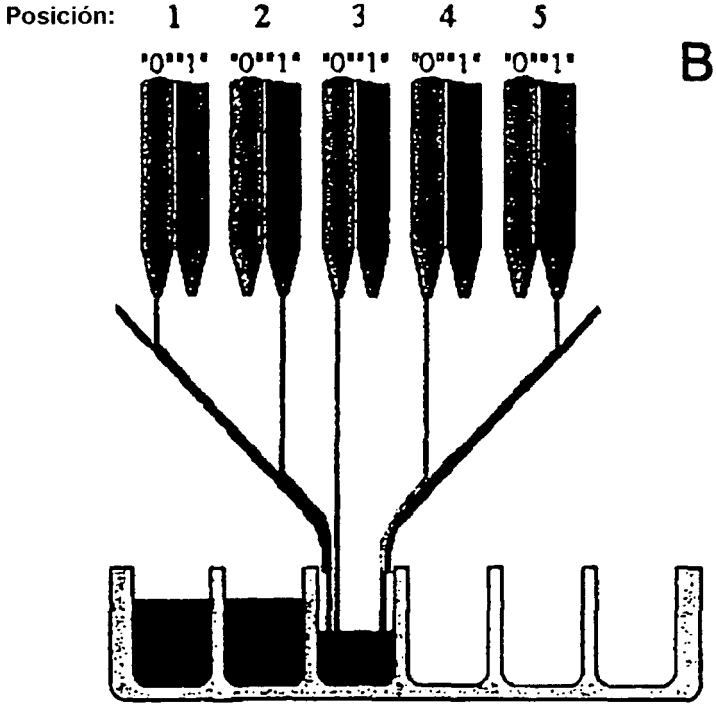
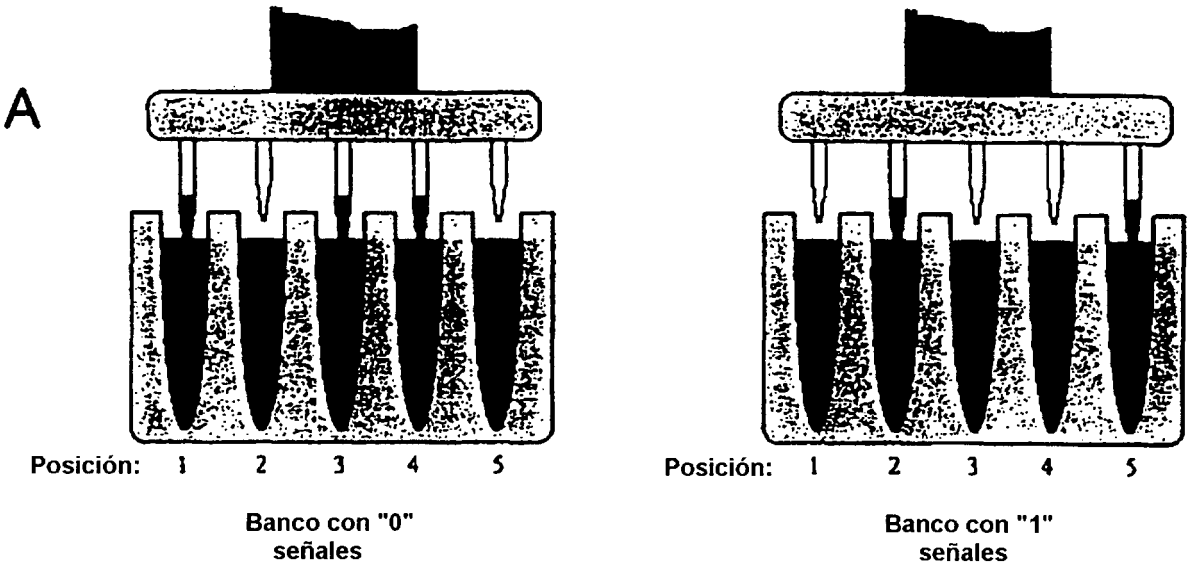


FIG. 3

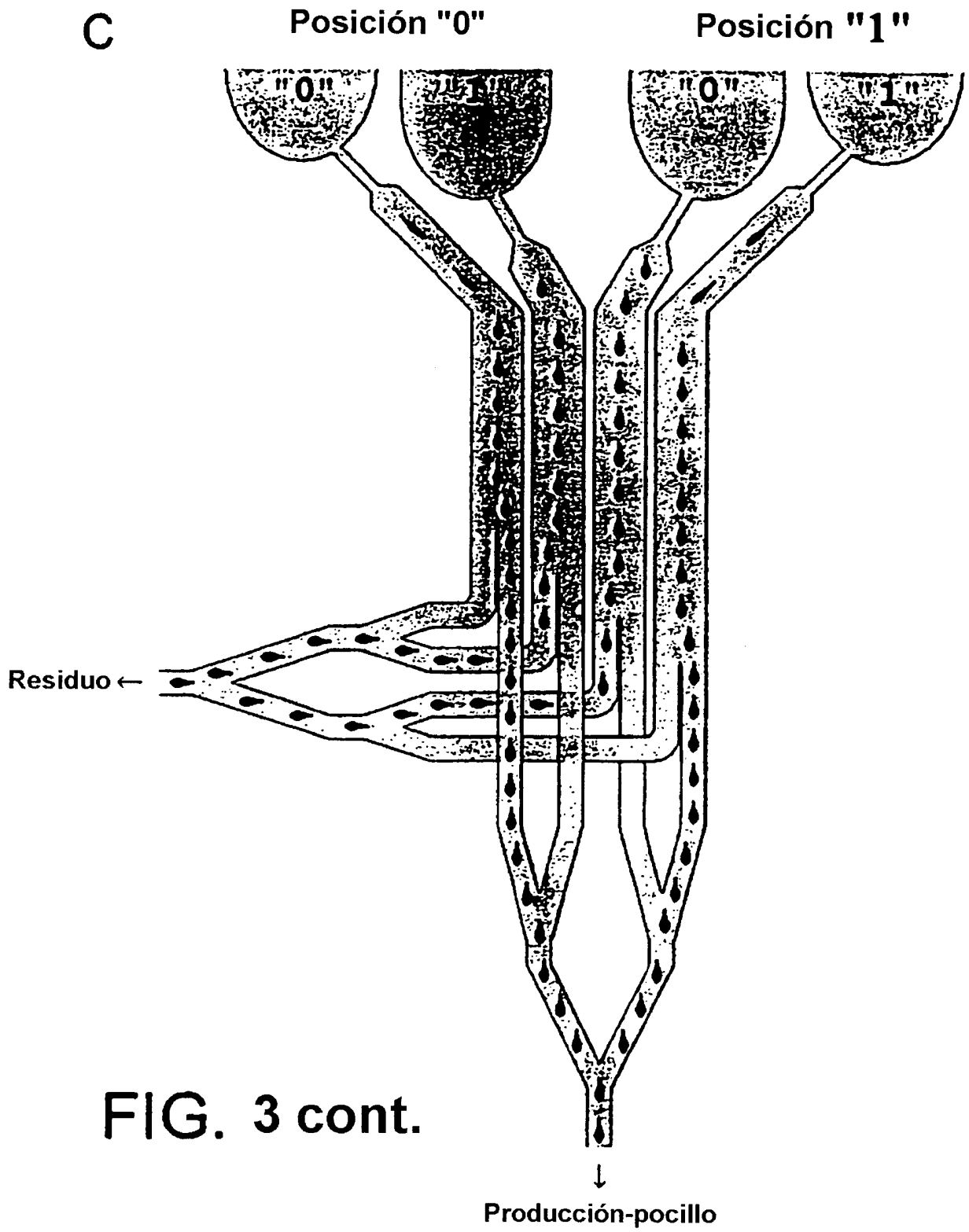


FIG. 3 cont.

escalón
1 kb

Cadena con señal 1-0-1-0-0
SP6/T7Cy SP6/T7

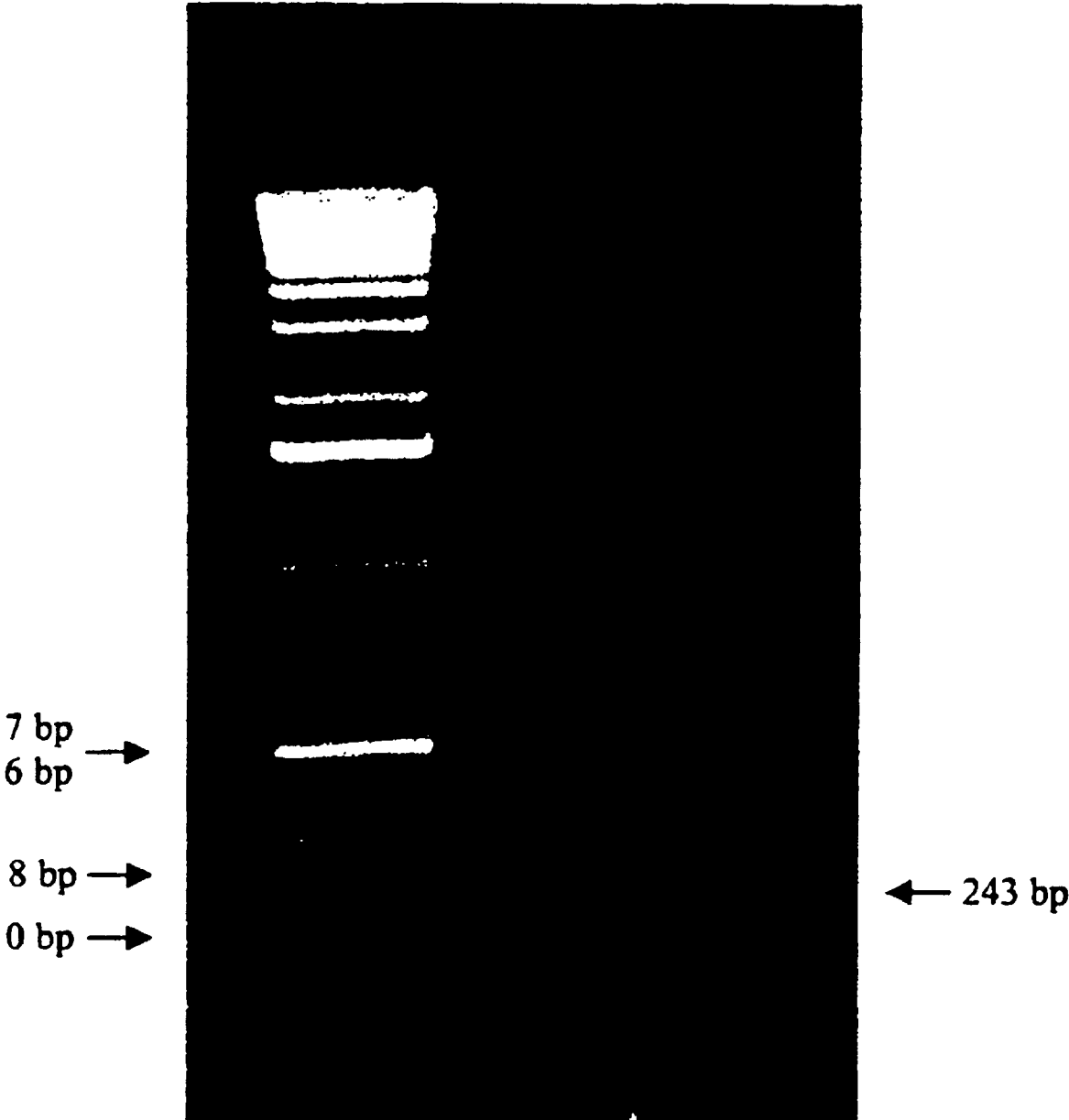


FIG. 4

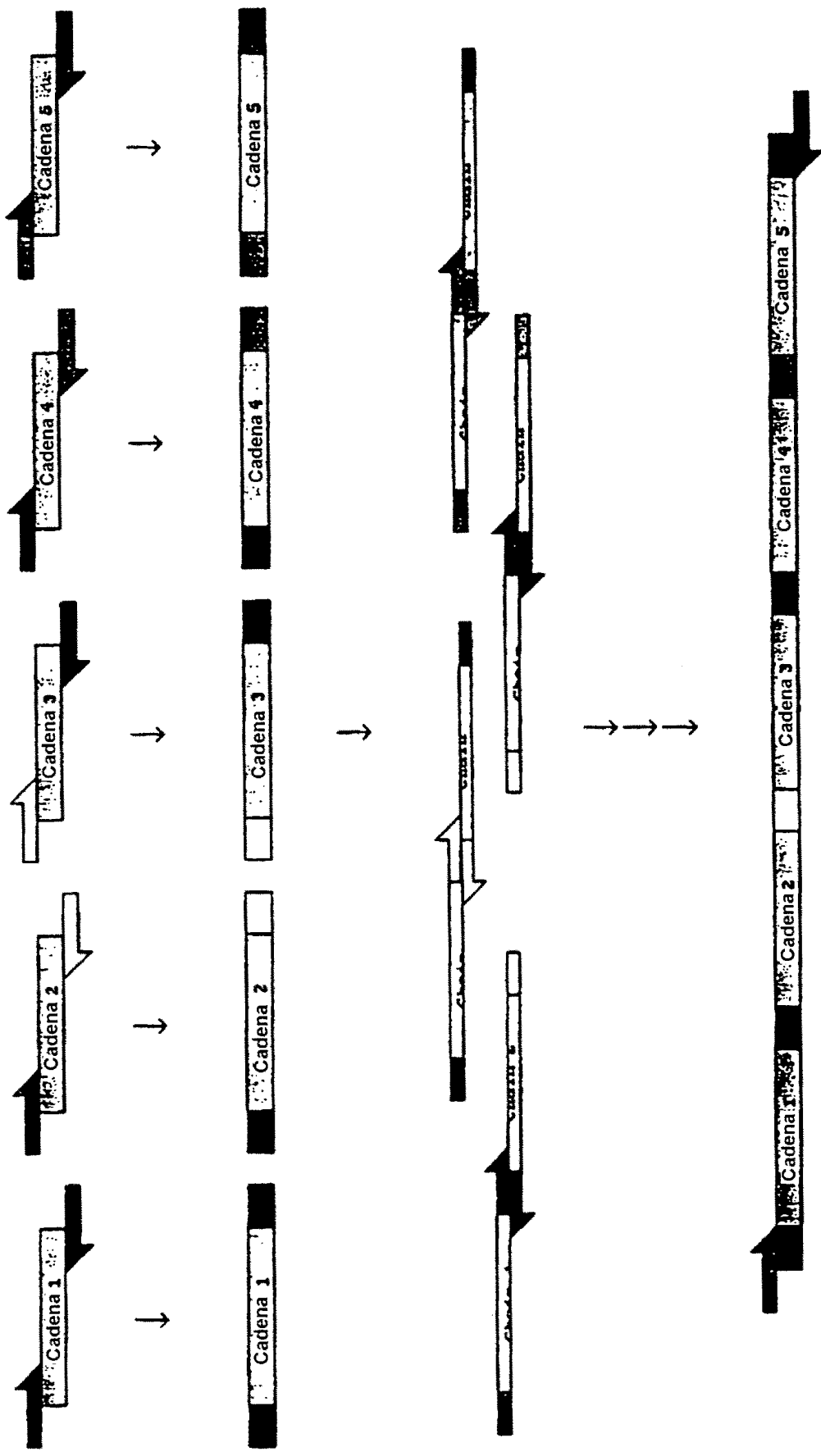


FIG. 5