

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 381 824**

51 Int. Cl.:
C12P 21/06 (2006.01)
C12P 19/34 (2006.01)
C12N 15/74 (2006.01)
C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **99932081 .5**
96 Fecha de presentación: **29.06.1999**
97 Número de publicación de la solicitud: **1092039**
97 Fecha de publicación de la solicitud: **18.04.2001**

54 Título: **Procedimientos para generar genotecas muy diversas**

30 Prioridad:
29.06.1998 US 90970 P

45 Fecha de publicación de la mención BOPI:
31.05.2012

45 Fecha de la publicación del folleto de la patente:
31.05.2012

73 Titular/es:
**BRISTOL-MYERS SQUIBB COMPANY
ROUTE 206 AND PROVINCE LINE ROAD
PRINCETON, NJ 08543-4000, US**

72 Inventor/es:
**WAGNER, Richard;
WRIGHT, Martin, C. y
KREIDER, Brent**

74 Agente/Representante:
Carpintero López, Mario

ES 2 381 824 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 para generar genotecas muy diversas

Antecedentes de la invención

En general, esta invención se refiere a procedimientos para generar y alterar genotecas recombinantes.

5 La capacidad de aislar una secuencia de ácidos nucleicos de aminoácidos deseada requiere la disponibilidad de genotecas recombinantes con un número y diversidad suficientes, de forma que una especie en particular esté representada en la genoteca y pueda ser identificada mediante una o más técnicas de cribado. Dichas genotecas facilitan el aislamiento de compuestos útiles, incluyendo terapéuticos, diagnósticos de investigación y reactivos agrícolas, así como sus secuencias codificantes.

10 Además, las genotecas deseables pueden diseñarse específicamente para que contengan grandes cifras de posibles variantes de un único compuesto. Este tipo de genotecas pueden usarse para cribar versiones mejoradas del compuesto, por ejemplo, de una variante de un compuesto con una eficacia terapéutica optimizada.

Para estas o cualquier otra aplicación, las metodologías generales para incrementar la diversidad de las genotecas son muy útiles y representan un foco importante en la industria de diseño de proteínas.

15 **Sumario de la invención**

En general, la presente invención presenta un procedimiento para generar una genoteca de ácidos nucleicos, implicando el procedimiento: (a) proporcionar una población de moldes de ácidos nucleicos monocatenarios, incluyendo cada uno de los moldes una secuencia codificante y una secuencia promotora unida operativamente; (b) 20 hibridar con la población de moldes de ácidos nucleicos monocatenarios una mezcla de fragmentos de ácidos nucleicos monocatenarios sustancialmente complementarios, siendo los fragmentos de una longitud más corta que el molde de ácido nucleico; (c) poner en contacto cada uno de los productos de hibridación de la etapa (b) tanto con una polimerasa de ADN que carezca de actividad de desplazamiento de hebra como con una ligasa de ADN en unas condiciones en las que los fragmentos actúen como cebadores para completar una segunda hebra de ácido nucleico que es sustancialmente complementaria del molde de ácido nucleico; y (d) poner en contacto los productos de la 25 etapa (c) con una polimerasa de ARN para generar una genoteca de ARN, siendo transcrita la genoteca a partir de la segunda hebra de ácido nucleico.

En las formas de realización preferidas, el procedimiento se usa para introducir una o más mutaciones en la genoteca; la mezcla de fragmentos de ácidos nucleicos monocatenarios sustancialmente complementarios se genera escindiendo una molécula de ácido nucleico bicatenario; la mezcla de fragmentos de ácidos nucleicos monocatenarios sustancialmente complementarios se genera mediante la síntesis de oligonucleótidos aleatorios; el 30 molde del ácido nucleico monocatenario se genera usando un fago M13 portador del ácido nucleico, mediante la digestión de una hebra de un molde de ácido nucleico bicatenario usando exonucleasa del gen VI o exonucleasa lambda, capturando una hebra simple de ácido nucleico biotinilada usando estreptavidina, o mediante una transcripción inversa del ARN; la mezcla de fragmentos de ácidos nucleicos monocatenarios sustancialmente 35 complementarios incluye al menos 100 especies diferentes de fragmentos de ácidos nucleicos; la etapa (b) se lleva a cabo usando entre 1 y aproximadamente 1.000 fragmentos por molde de ácido nucleico monocatenario; se usa una hebra simple del producto de la etapa (c) como molde de ácido nucleico y se repiten las etapas (b) y (c); las etapas (b) y (c) se repiten usando, en cada ronda, el producto de la etapa (c) como molde de ácido nucleico; el procedimiento implica adicionalmente proporcionar uno o más fragmentos de ácidos nucleicos monocatenarios que forman homodúplex con el molde de ácido nucleico monocatenario y llevar a cabo la etapa (b) en presencia de los fragmentos formadores del homodúplex; el promotor es un promotor T7; la polimerasa de ADN es la polimerasa de ADN T4; el procedimiento implica además amplificar el producto de la etapa (c) antes de dicha etapa de contacto (d); el procedimiento implica además la etapa de: (e) traducir la genoteca de ARN para generar una genoteca de 40 proteínas; el procedimiento implica además la etapa de: (e) unir al extremo 3' de la secuencia codificante de cada uno de sustancialmente todos los miembros de la genoteca de ARN a una molécula aceptora de aminoácidos; y el procedimiento implica además la etapa de (f) traducir la genoteca de ARN para generar una genoteca de fusión de ARN-proteínas.

En un segundo aspecto, la invención presenta un procedimiento para reducir la variación en la secuencia en una población de moléculas de ácidos nucleicos, implicando el procedimiento: (a) proporcionar una primera población de 50 moldes de ácidos nucleicos monocatenarios de secuencia variable, incluyendo cada uno de sustancialmente todos los moldes una secuencia codificante y una secuencia promotora unida operativamente; (b) hibridar los miembros de la primera población con una segunda población de fragmentos de ácidos nucleicos monocatenarios sustancialmente complementarios, siendo los fragmentos de una longitud más corta que el molde de ácido nucleico y teniendo los fragmentos una secuencia sustancialmente idéntica; (c) poner en contacto los productos de hibridación de la etapa (b) tanto con una polimerasa de ADN que carece de actividad de desplazamiento de hebra como con una ligasa de ADN en unas condiciones en las que los fragmentos actúen como cebadores para completar una segunda hebra de ácido nucleico que es sustancialmente complementaria al molde de ácido nucleico; 55 y (d) poner en contacto los productos de la etapa (c) con una polimerasa de ARN para generar una población de

moléculas de ARN, siendo transcrita la población de moléculas de ARN a partir de la segunda hebra de ácido nucleico y teniendo una variación reducida en la secuencia con respecto a la primera población de moldes de ácidos nucleicos monocatenarios.

5 En formas de realización preferidas, el procedimiento se usa para eliminar una o más mutaciones de la primera población de moldes de ácidos nucleicos monocatenarios; la etapa (b) implica la hibridación de la primera población de moldes de ácidos nucleicos monocatenarios con dos o más poblaciones diferentes de fragmentos de ácidos nucleicos monocatenarios sustancialmente complementarios; la segunda población de fragmentos de ácidos nucleicos monocatenarios sustancialmente complementarios se genera escindiendo una molécula de ácido nucleico bicatenario; la segunda población de fragmentos de ácidos nucleicos monocatenarios sustancialmente complementarios se genera mediante la síntesis de oligonucleótidos aleatorios; el molde de ácido nucleico monocatenario se genera usando un fago M13 portador del ácido nucleico, mediante la digestión de una hebra de un molde de ácido nucleico bicatenario usando exonucleasa del gen VI o exonucleasa lambda, capturando una hebra simple de ácido nucleico biotinilada usando estreptavidina, o mediante una transcripción inversa del ARN; la etapa (b) se lleva a cabo usando entre 1 y aproximadamente 1.000 fragmentos de ácidos nucleicos monocatenarios por molde de ácido nucleico monocatenario; se usa una hebra simple del producto de la etapa (c) como molde de ácido nucleico y se repiten las etapas (b) y (c); las etapas (b) y (c) se repiten usando, en cada ronda, el producto de la etapa (c) como molde de ácido nucleico; el promotor es un promotor T7; la polimerasa de ADN es la polimerasa de ADN T4; el procedimiento implica además amplificar el producto de la etapa (c) antes de dicha etapa de contacto (d); el procedimiento implica además la etapa de: (e) traducir la población de moléculas de ARN para generar una genoteca de proteínas; el procedimiento implica además la etapa de: (e) unir al extremo 3' de la secuencia codificante de cada uno de sustancialmente todos los miembros de la población de moléculas de ARN a una molécula aceptora de aminoácidos; y el procedimiento implica además la etapa de (f) traducir la población de moléculas de ARN para generar una genoteca de fusión de ARN-proteínas.

25 En un tercer aspecto, la invención presenta un procedimiento para generar una genoteca de ácidos nucleicos, implicando el procedimiento: (a) proporcionar una población de moldes de ácidos nucleicos monocatenarios, incluyendo cada molde una secuencia codificante; (b) proporcionar una población de moléculas de ácidos nucleicos monocatenarios de secuencia variable, siendo la población de moldes de ácidos nucleicos monocatenarios y la población de moléculas de ácidos nucleicos monocatenarios de secuencia variable sustancialmente complementarias; (c) hibridar la población de moldes de ácidos nucleicos monocatenarios con la población de moléculas de ácidos nucleicos monocatenarios de secuencia variable en unas condiciones suficientes para formar dúplex; y (d) poner en contacto los dúplex con una o más enzimas de escisión/repación en unas condiciones que permitan a las enzimas corregir los pares de bases mal apareados en los dúplex.

35 En formas de realización preferidas, el procedimiento implica además proporcionar una población de moldes monocatenarios derivados del producto de la etapa (d) y repetir las etapas (c) y (d); y las etapas (c) y (d) se repiten usando, en cada ronda, una población de moldes monocatenarios derivada del producto de la etapa (d).

40 En un cuarto aspecto, la invención presenta un procedimiento para generar una genoteca de ácidos nucleicos, implicando el procedimiento: (a) proporcionar una población de moldes de ácidos nucleicos monocatenarios, incluyendo cada molde una secuencia codificante; (b) hibridar con la población de moldes de ácidos nucleicos monocatenarios una mezcla de fragmentos de ácidos nucleicos monocatenarios sustancialmente complementarios, siendo los fragmentos de una longitud más corta que el molde de ácido nucleico; (c) poner en contacto cada uno de los productos de hibridación de la etapa (b) tanto con una polimerasa de ADN que carece de actividad de desplazamiento de hebra como con una ligasa de ADN en unas condiciones en las que los fragmentos actúen como cebadores para completar una segunda hebra de ácido nucleico que es sustancialmente complementaria al molde de ácido nucleico; y (d) poner en contacto los productos de la etapa (c) con una o más enzimas de escisión/repación en unas condiciones que permitan a las enzimas corregir los pares de bases mal apareados en los productos.

En formas de realización preferidas, el procedimiento implica además proporcionar una población de moldes monocatenarios derivados del producto de la etapa (d) y repetir las etapas (b) - (d); y las etapas (b) - (d) se repiten usando, en cada ronda, una población de moldes monocatenarios derivada del producto de la etapa (d).

50 En formas de realización preferidas del tercer y cuarto aspectos de la invención, el contacto con las enzimas de escisión/repación se lleva a cabo *in vivo* (por ejemplo, en una célula bacteriana); el contacto con las enzimas de escisión/repación se lleva a cabo *in vitro*; el molde de ácido nucleico monocatenario se genera usando un fago M13 portador del ácido nucleico mediante la digestión de un molde de ácido nucleico bicatenario usando exonucleasa del gen VI o exonucleasa lambda, capturando una hebra simple de ácido nucleico biotinilada usando estreptavidina, o mediante una transcripción inversa del ARN; la etapa (b) se lleva a cabo usando entre 1 y aproximadamente 1.000 moléculas de ácidos nucleicos monocatenarios de secuencia variable o fragmentos de ácidos nucleicos monocatenarios por molde de ácido nucleico monocatenario; el procedimiento implica además la etapa de: (e) amplificar el producto de la etapa (d); cada una de las secuencias codificantes está unida operativamente a una secuencia promotora; el procedimiento implica además la etapa de: (e) transcribir los productos de la etapa (d) para generar una genoteca de ARN; el procedimiento implica además la etapa de: (f) traducir la genoteca de ARN para generar una genoteca de proteínas; el procedimiento implica además la etapa de:

(f) unir al extremo 3' de la secuencia codificante de cada uno de sustancialmente todos los miembros de la genoteca de ARN una molécula aceptora de aminoácidos; y el procedimiento implica además la etapa de (g) traducir la genoteca de ARN para generar una genoteca de fusión de ARN-proteínas.

5 Según se usa en este documento, por "genoteca" se entienden al menos 10^8 , preferiblemente al menos 10^{10} , más preferiblemente al menos 10^{12} , y muy preferiblemente al menos 10^{14} moléculas con un componente ácido nucleico y/o un componente aminoácido.

Por una "mezcla" de fragmentos de ácidos nucleicos se entiende al menos 100, preferiblemente al menos 500, más preferiblemente al menos 1.000, y muy preferiblemente al menos 1.500 fragmentos de ácidos nucleicos.

10 Por una "secuencia promotora" se entiende cualquier secuencia de ácidos nucleicos que proporcione un sitio funcional de unión de la polimerasa de ARN y que es suficiente para permitir la transcripción de una secuencia codificante proximal.

15 Por "sustancialmente complementaria" se entiende que una hebra de ácidos nucleicos posee un número suficiente de nucleótidos que son capaces de formar pares de bases complementarios de Watson-Crick con una segunda hebra de ácido nucleico para producir una o más regiones bicatenarias entre los dos ácidos nucleicos. Se entenderá que cada nucleótido de una molécula de ácido nucleico no necesita formar un par de base apareado de Watson-Crick con un nucleótido de una hebra opuesta para ser sustancialmente complementario, y que en una "mezcla de fragmentos de ácidos nucleicos monocatenarios sustancialmente complementarios", una fracción significativa de los fragmentos contendrá uno o más nucleótidos que formen apareamientos incorrectos con el "molde de ácido nucleico monocatenario".

20 Por "actividad de desplazamiento de hebra" se entiende la capacidad de una polimerasa o de su helicasa asociada para disociar el apareamiento de bases entre dos hebras de ácidos nucleicos.

Por "mutación" se entiende cualquier cambio nucleotídico, e incluye alteraciones en la secuencia que dan como resultado diferencias genotípicas, así como cambios que son silenciosos.

25 Por "dúplex" se entiende una estructura formada entre dos hebras de ácidos nucleicos apareadas en las que existe una complementariedad de secuencias suficiente entre las hebras como para mantener un complejo de hibridación estable. Un dúplex puede ser un "homodúplex," en el que todos los nucleótidos de la primera hebra se aparean apropiadamente con todos los nucleótidos de la segunda hebra opuesta, o un heterodúplex. Por "heterodúplex" se entiende una estructura formada entre dos hebras apareadas de ácidos nucleicos en las que uno o más nucleótidos de la primera hebra no se aparean o no pueden aparearse apropiadamente con uno o más nucleótidos de la segunda hebra opuesta complementaria porque hay uno o más apareamientos incorrectos. Algunos ejemplos de diferentes tipos de heterodúplex incluyen aquellos que muestran un intercambio de uno o varios nucleótidos, y mutaciones por inserción o por delección.

30 Por "oligonucleótidos aleatorios" se entiende una mezcla de oligonucleótidos con una variación en la secuencia en una o más posiciones nucleotídicas. Los oligonucleótidos aleatorios pueden producirse usando metodologías sintéticas total o parcialmente aleatorias, o mediante la alteración intencional de un oligonucleótido de forma directa.

35 Por "molécula aceptora de aminoácidos" se entiende cualquier molécula capaz de ser añadida al C terminal de una cadena proteica en crecimiento o mediante la actividad catalítica de la función peptidil transferasa ribosómica. Normalmente, dichas moléculas contienen (i) una fracción de nucleótido o de tipo nucleótido (por ejemplo, adenosina o un análogo de adenosina (la dimetilación en la posición amino N-6 es aceptable)), (ii) un aminoácido o una fracción de tipo aminoácido (por ejemplo, cualquiera de los 20 aminoácidos D o L o cualquier análogo de aminoácido de los mismos (por ejemplo, O-metil tirosina o cualquiera de los análogos descritos por Ellman y col., Meth. Enzymol. 202: 301, 1991)), y (iii) un enlace entre ambos (por ejemplo, un enlace éster, amida o cetona en la posición 3' o, menos preferiblemente, en la posición 2'); preferiblemente, este enlace no altera significativamente el plegamiento del anillo de la conformación natural de ribonucleótidos. Los aceptores de aminoácidos también pueden poseer un nucleófilo, que puede ser, sin limitación, un grupo amino, un grupo hidroxilo o un grupo sulfhidrilo. Además, los aceptores de aminoácidos pueden estar formados por nucleótidos miméticos, aminoácidos miméticos o miméticos de una estructura combinada de nucleótidos y aminoácidos.

40 Por un grupo aceptor de aminoácidos unido al "extremo 3'" de una secuencia codificante se entiende que la molécula aceptora de aminoácidos está posicionada después del codón de finalización de esa secuencia codificante. Este término incluye, sin limitación, una molécula aceptora de aminoácidos que está posicionada precisamente en el extremo 3' de la secuencia codificante, así como aquella que esté separada del codón de finalización mediante la intervención de secuencias codificantes o no codificantes (por ejemplo, una secuencia correspondiente a un sitio de pausa). Este término también incluye constructos en los que secuencias codificantes y no codificantes siguen (esto es, son de 3' a) la molécula aceptora de aminoácidos. Además, este término engloba, sin limitación, a una molécula aceptora de aminoácidos que está unida covalentemente (bien directamente o bien indirectamente a través de la intervención de una secuencia de ácidos nucleicos) a la secuencia codificante, así como aquella que está unida a la secuencia codificante por algún medio no covalente, por ejemplo, mediante hibridación usando una segunda secuencia de ácidos nucleicos que se une a, o cerca de, el extremo 3' de la secuencia codificante y que por sí

misma está unida a una molécula aceptora de aminoácidos.

Por fusión de "ARN-proteína" se entiende cualquier molécula que incluya un ácido ribonucleico unido covalentemente a través de un enlace amida a una proteína. Este enlace covalente es resistente a la escisión por un ribosoma.

- 5 Por "proteína" se entiende cualquier par o más de aminoácidos naturales o modificados unidos por uno o más enlaces peptídicos. "Proteína", "péptido" y "polipéptido" se usan de forma intercambiable en este documento.

Por "una población de moldes monocatenarios de secuencia variable" se entiende que la especie de ácido nucleico de la población posee secuencias que difieren en una o más posiciones de nucleótidos.

- 10 Por "enzimas de escisión/repación" se entiende cualquier combinación de enzimas suficiente para sustituir un par de bases mal apareado o un bucle por un par de bases estándar (es decir, A:T o G:C).

Breve descripción de los dibujos

- 15 La FIGURA 1 es una representación esquemática de un procedimiento ejemplar de recombinación de fragmentos para generar genotecas de fusión de ARN-proteínas muy diversas. En este procedimiento los fragmentos derivan de una molécula de ADN bicatenario en la que se ha introducido una variación en la secuencia.

La FIGURA 2 es una representación esquemática de las etapas iniciales de un segundo procedimiento ejemplar de recombinación de fragmentos. En este procedimiento, los fragmentos son oligonucleótidos sintéticos en los que se ha introducido una variación en la secuencia.

Descripción detallada

- 20 La presente invención implica varios procedimientos nuevos y relacionados para la recombinación aleatoria de secuencias de ácidos nucleicos, facilitando la regeneración del ADN, el ARN y las genotecas de proteínas en las que se han introducido alteraciones genéticas. Según se describe con más detalle a continuación, en una forma de realización preferida, esta técnica se lleva a cabo *in vitro* y se usa para generar genotecas de proteínas tradicionales o genotecas de fusión de ARN-proteínas, cualquiera de las cuales puede usarse a continuación en combinación con cualquiera de los diversos procedimientos para la selección de las proteínas o péptidos deseados (por sus correspondientes secuencias codificantes) a partir de poblaciones de genotecas. Esta metodología general proporciona un medio para la introducción de mutaciones en genotecas de proteínas de una forma no sesgada, y también proporciona una técnica mediante la cual pueden limitarse mutaciones desfavorables a partir de una genoteca o conjunto seleccionado, o "retrocruzadas" en una población de moléculas durante las subsiguientes rondas de selección.

Recombinación de fragmentos

- 35 Según un procedimiento preferido de la invención, se genera una genoteca mediante la producción de fragmentos mutantes y la recombinación aleatoria de estos fragmentos con una secuencia no mutada (normalmente, natural). Un ejemplo de esta metodología general se muestra en la Figura 1. Según se indica en esta figura, en primer lugar se introducen aleatoriamente mutaciones en una secuencia de ADN bicatenario (denominada "dsDNA(init)"). Esto produce una población de secuencias de ADN bicatenarios mutantes, que, en la Figura 1, se denomina "dsDNA(mut)." Estas mutaciones pueden introducirse mediante cualquier técnica, incluyendo mutagénesis mediante PCR (que se basa en el deficiente mecanismo de corrección de errores de la polimerasa Taq), mutagénesis dirigida o mutagénesis dirigida al molde (por ejemplo, según se describe en Joyce e Inoue, Nucl. Acids Res. 17: 171, 1989).
- 40 El ADN de esta población que contiene mutaciones se fragmenta subsiguientemente usando cualquiera de los diversos procedimientos estándar. Por ejemplo, el ADN puede degradarse parcialmente usando una o más nucleasas (tal como ADNasa I, nucleasa microcócica, endonucleasas de restricción o nucleasa P1), o puede fragmentarse químicamente usando, por ejemplo, Fe-EDTA. Alternativamente pueden generarse fragmentos que contienen mutaciones mediante el consumo limitado de nucleótidos durante la polimerización (por ejemplo, durante la amplificación mediante PCR), o mediante un simple cizallamiento físico (por ejemplo, mediante la aplicación de ultrasonidos). Los tamaños preferibles de los fragmentos varían desde 25-1.000 pares de bases, y muy preferiblemente están en el intervalo de 50-150 pares de bases.

El documento WO97/20078 describe la fragmentación aleatoria de ADN bicatenario y el subsiguiente reensamblaje de dichos fragmentos en una reacción de tipo PCR para formar ADN bicatenario recombinante.

- 50 En la presente invención, los fragmentos de ADN se calientan y subsiguientemente se hibridan con un molde de ADN monocatenario completo que es idéntico al ADN inicial de la secuencia y que es la hebra no codificante (o menor) de ese ADN. Además, en esta mezcla de hibridación está incluido un segundo tipo de fragmento, denominado "fragmento de terminación" (Joyce e Inoue, Nucl. Acids Res. 17: 171, 1989). Este fragmento de terminación es complementario del extremo 3' del molde monocatenario y proporciona un cebador de la polimerización que se une al molde de una forma que es relativamente independiente del número o la naturaleza de los fragmentos hibridados aleatoriamente que contienen las mutaciones.

Los moldes monocatenarios pueden generarse mediante cualquier técnica estándar, por ejemplo, usando un fago M13 portador de la secuencia de ADN, mediante digestión de la hebra codificante de una molécula de dsDNA (init) usando la exonucleasa del gen VI (Nikiforov y col., PCR Methods Appl. 3: 285, 1994) o la exonucleasa lambda (Higuchi y Ochman, Nucl. Acids Res 17: 5865, 1989), mediante la captura de una hebra de ADN biotinilada usando estreptavidina inmovilizada (Joyce e Inoue, Nucl. Acids Res. 17: 171, 1989) o mediante la transcripción inversa del ARN. Para llevar a cabo la hibridación molde-fragmento, se mezclan los moldes con los fragmentos usando no menos de una molécula de fragmento por molécula de molde, y no más de aproximadamente 1.000 moléculas de fragmento por molécula de molde. Una baja proporción entre fragmentos y moldes produce hebras progenies que se parecen demasiado a los moldes, mientras que una alta proporción produce una progenie que se parece más a los fragmentos. Las condiciones de hibridación están determinadas por los procedimientos estándar, y se diseñan para permitir la formación de heterodúplex entre el molde y los fragmentos. Algunos ejemplos de técnicas de hibridación se describen en, por ejemplo, Stemmer. Patente de EE.UU. N° 5.605.793.

Una vez hibridados con el molde, los fragmentos se unen entre sí tratando con una polimerasa de ADN que carece de actividad de desplazamiento de hebra y con una ligasa de ADN. Las polimerasas de ADN útiles para este propósito incluyen, sin limitación, polimerasa de ADN T4 y DNA pol II reconstituida a partir de *E. coli* (véase, por ejemplo, Hughes y col., J. Biol. Chem. 266: 4568, 1991). Puede usarse cualquier ligasa de ADN (por ejemplo, ligasa de ADN T4). En esta etapa, los dúplex de ADN pueden tratarse en primer lugar con una polimerasa de ADN y después con la ligasa de ADN, o con ambas enzimas simultáneamente, y la etapa puede llevarse a cabo, por ejemplo, según se describe en Joyce e Inoue (Nucl. Acids Res. 17: 711, 1989). Según se muestra en la Figura 1, esta etapa genera una población de ADN bicatenarios (denominada "dsDNA (lib)"), cada miembro de la cual incluye una hebra que tiene introducidas normalmente una o más mutaciones. Debido a que tanto las mutaciones introducidas inicialmente como el número y la naturaleza de los fragmentos hibridados son aleatorios, diferentes dúplex de la población contienen diferentes secuencias mutantes.

Una alternativa a esta metodología general para generar una genoteca de ADN bicatenarios se muestra en la Figura 2. Mediante esta metodología alternativa se sintetizan fragmentos de nucleótidos monocatenarios que se corresponden con las porciones de la hebra codificante de una molécula de ADN bicatenario inicial. Estos fragmentos de oligonucleótidos varían preferiblemente entre 5-2.000 nucleótidos, y muy preferiblemente varían entre 20-100 nucleótidos de longitud y se generan, por ejemplo, usando cualquiera de las técnicas estándar de síntesis de ácidos nucleicos. Estos oligonucleótidos pueden sintetizarse con mutaciones completamente aleatorias o semialeatorias mediante cualquier técnica estándar. Preferiblemente, dichos oligonucleótidos incluyen hasta 3 apareamientos incorrectos introducidos por segmento de 20 nucleótidos, y están desprovistos de codones de detención en marco. Además, en ciertos casos, puede ser deseable o necesario aumentar el potencial de hibridación del oligonucleótido a través de la introducción de pares de bases no naturales incrementadores de la afinidad, tales como C-5 propina uridina o C-5 propina citidina. Estas técnicas se describen, por ejemplo, en Wagner y col., Science 260: 1510, 1993.

Estos fragmentos de oligonucleótidos que contienen mutaciones se hibridan a continuación con moldes monocatenarios que, como anteriormente, son hebras completas con secuencias idénticas a la hebra no codificante (o menor) del ADN inicial. Los fragmentos se unen entre sí usando polimerasa de ADN y ligasa de ADN, también según se describió anteriormente, para crear una genoteca de ADN bicatenarios (dsDNA(lib)). De nuevo, esta genoteca contiene una población de moléculas dúplex que contienen una matriz de diferentes hebras codificantes con mutaciones que difieren en número, posición e identidad.

Si se desea, pueden repetirse las etapas anteriores, para la metodología de fragmentos o de oligonucleótidos, para introducir cifras variables de mutaciones en una molécula de ADN. En particular, las hebras mutadas se transforman en los moldes monocatenarios iniciales, y los fragmentos o los oligonucleótidos mutantes se hibridan con esas hebras y se polimerizan y ligan.

Procedimientos de retrocruzamiento

Según se describió anteriormente de forma general, los procedimientos de la invención se usan para introducir mutaciones en una secuencia inicial de ADN. Además, estas técnicas pueden usarse para eliminar o reducir la frecuencia de mutaciones indeseables a partir de una genoteca de ADN. Según esta metodología, después de la fragmentación de los dsDNA(mut), pueden añadirse oligonucleótidos de secuencia natural o fragmentos específicos de ADN no mutado o natural (wtDNA) al molde monocatenario junto con los fragmentos o los oligonucleótidos de dsDNA(mut). Se separan las hebras de los fragmentos (si fuera necesario), se hibridan con el molde monocatenario completo y se unen entre sí usando polimerasa de ADN y ligasa de ADN, según se describió anteriormente. El uso de una elevada concentración de oligonucleótidos o fragmentos no mutados, con respecto al correspondiente fragmento mutante, permiten la generación de genotecas en las que se minimizan o eliminan las mutaciones indeseables.

Además, esta metodología puede usarse con genotecas existentes que contienen mutaciones para disminuir o eliminar de forma similar secuencias indeseables. Esta metodología implica una genoteca inicial que contiene secuencias mutantes y la hibridación, polimerización y ligación de fragmentos u oligonucleótidos de secuencia natural, según se describió de forma generalizada anteriormente.

ARN, proteína y genotecas de ARN-proteína

5 En una forma de realización preferida de la invención, las genotecas de ADN descritas anteriormente incluyen adicionalmente un sitio de unión de la polimerasa de ARN, por ejemplo, para la polimerasa T7 o SP6. Dichos sitios de unión se describen, por ejemplo, en Milligan y col., Proc. Natl. Acad. Sci. EE.UU. 87: 696, 1990. Este sitio está posicionado secuencia arriba de la secuencia codificante en una ubicación que permite la transcripción de la secuencia. Normalmente, dichos sitios se localizan a entre 5-2.000 pares de bases secuencia arriba de la secuencia codificante.

10 Las genotecas que contienen sitios de unión de la polimerasa de ARN pueden alterarse según se describió anteriormente. Tras la polimerización y la ligación, pueden transferirse directamente las dsDNA(lib), por ejemplo, usando un sistema de transcripción *in vitro*, para generar una genoteca de ARN. Alternativamente, las dsDNA(lib) pueden transcribirse y traducirse directamente, por ejemplo, usando sistemas de transcripción y traducción *in vitro*, para generar una genoteca de proteínas. Algunos ejemplos de sistemas de transcripción *in vitro* y de sistemas de traducción *in vitro* incluyen sistemas de transcripción T7, y sistemas de traducción de reticulocitos de conejo, germen de trigo, levadura y *E. coli*.

15 Si se desea, puede incrementarse el número de copias de cada ARN o proteína de la genoteca incluyendo una etapa de amplificación específica de hebra antes de la transcripción. Por ejemplo, puede llevarse a cabo una amplificación mediante PCR incorporando secuencias únicas de unión de cebador son incorporadas en la hebra mutante durante las etapas de polimerización y ligación. Estas secuencias pueden incorporarse como apareamientos incorrectos o como extensiones de la secuencia en uno o ambos extremos del ADN, permitiendo la
20 amplificación de la recién sintetizada hebra sin amplificar la hebra molde. Alternativamente, puede conseguirse una amplificación lineal mediante ciclos múltiples de hibridación y extensión de un único cebador oligonucleotídico que es complementario del extremo 3' de la recién sintetizada hebra. Las subsiguientes etapas de PCR y transcripción producen una mayoría de ARN correspondiente a las secuencias mutantes con únicamente una pequeña proporción de secuencias derivadas del molde.

25 En una metodología preferida, los anteriores procedimientos de introducción de mutaciones o de retrocruzamiento de mutaciones indeseables pueden usarse para producir genotecas muy diversas de ARN-proteína. Dichas genotecas pueden construirse ligando conectores que contienen una molécula aceptora de aminoácidos no hidrolizable, tal como puromicina, al extremo 3' de los ARN de una genoteca (por ejemplo, producidos según se describió anteriormente). Algunos ejemplos de técnicas para generar fusiones de ARN-proteínas se describen, por
30 ejemplo, en Szostak y col., U.S.S.N. 09/007,005; y en Roberts y col., Proc. Natl. Acad. Sci. EE.UU. 94: 12297, 1997. La subsiguiente traducción de estos ARN genera una genoteca de moléculas de fusión de ARN-proteínas que puede usarse subsiguientemente en experimentos de selección *in vitro*.

35 Además, si se desea, una vez elegidas las moléculas de ARN o de fusión de ARN-proteínas, pueden usarse como moldes en reacciones estándar de PCR para obtener la correspondiente secuencia codificante. Por lo tanto, este procedimiento proporciona un medio para llevar a cabo una recombinación de fragmentos, un retrocruzamiento molecular, la selección de proteínas y/o péptidos, y la selección de sus correspondientes secuencias codificantes, todo en un sistema *in vitro*.

Escisión/Reparación

40 Además de las metodologías de recombinación de fragmentos, también puede usarse una escisión/reparación para alterar las secuencias de las genotecas. Esta metodología puede usarse para generar genotecas de ADN, ARN y fusión de ARN-proteínas. Esta técnica se basa en el hecho de que las dsDNA(lib)s, producidas mediante cualquiera de los procedimientos descritos anteriormente, por su naturaleza, contienen un cierto número de pares de bases mal apareados. Para generar diversidad en las secuencias de las genotecas, estos apareamientos incorrectos son reparados *in vitro* por enzimas de escisión/reparación. Esto puede llevarse a cabo usando un sistema de reparación
45 por escisión (por ejemplo, según se describe en Jaiswal y col., Nucl. Acids Res. 26: 2184, 1998; o en Fortini y col., Biochemistry 37: 3575, 1998).

50 Alternativamente, la etapa de escisión/reparación puede llevarse a cabo transformando una dsDNA(lib) en una cepa bacteriana o de levadura y aprovechando los sistemas de reparación de la bacteria o la levadura *in vivo*. De nuevo, esta etapa puede llevarse a cabo transformando la genoteca en cualquier sistema de escisión/reparación estándar *in vivo*. Algunos ejemplos de sistemas se describen, sin limitación, en Campbell y col., Mutat. Res. 211: 181, 1989; Bishop y Kolodner, Mol. Cell Biol. 6: 3401, 1986; Fishel y col., J. Mol. Biol. 188: 147, 1986; y Westmoreland y col., Genetics 145: 29, 1997.

55 Debido a que los anteriores procesos de reparación son aleatorios, este procedimiento de escisión/reparación a veces da como resultado la introducción de mutaciones en una secuencia de una genoteca, y otras veces da como resultado el retrocruzamiento de alteraciones de la secuencia natural en la hebra codificante.

En una alternativa a las metodologías anteriores, también puede usarse directamente la escisión/reparación *in vitro* o *in vivo* para generar diversas genotecas usando como sustrato una mezcla de dsDNA(mut) (por ejemplo, producida según se describió anteriormente) y dsDNA(init) o wtDNA. En esta técnica, se separan y rehibridan las

hebras de la mezcla, y a continuación se incuban *in vitro* con las enzimas de escisión/repación o se transforman en bacterias para usar el sistema de escisión/repación bacteriano (por ejemplo, según se describió anteriormente). De esta forma pueden introducirse aleatoriamente mutaciones en una secuencia, y pueden retrocruzarse secuencias naturales en moléculas de dsDNA(mut).

REIVINDICACIONES

1. Un procedimiento para generar una genoteca de ácidos nucleicos, comprendiendo dicho procedimiento:

- (a) proporcionar una población de moldes de ácidos nucleicos monocatenarios, comprendiendo cada uno de dichos moldes una secuencia codificante y una secuencia promotora unida operativamente;
- (b) hibridar con dicha población de moldes de ácidos nucleicos monocatenarios una mezcla de fragmentos de ácidos nucleicos monocatenarios sustancialmente complementarios, siendo dichos fragmentos de una longitud más corta que dichos moldes de ácidos nucleicos;
- (c) poner en contacto cada uno de los productos de hibridación de la etapa (b) tanto con una polimerasa de ADN que carece de actividad de desplazamiento de hebra como con una ligasa de ADN en unas condiciones en las que dichos fragmentos hibridados aleatoriamente actúen como cebadores para completar unas segundas hebras de ácidos nucleicos que son sustancialmente complementarias a dichos moldes de ácidos nucleicos, produciendo así una población de hebras codificantes con mutaciones que codifican para hebras que difieren entre sí en número, posición e identidad de dichas mutaciones; y
- (d) poner en contacto los productos de la etapa (c) con una polimerasa de ARN para generar una genoteca de ARN, siendo transcrita dicha genoteca a partir de dichas segundas hebras de ácidos nucleicos.

2. El procedimiento de la reivindicación 1, en el que dicho procedimiento se usa para introducir una o más mutaciones en dicha genoteca.

3. Un procedimiento para reducir la variación en la secuencia en una población de moléculas de ácidos nucleicos, comprendiendo dicho procedimiento:

- (a) proporcionar una primera población de moldes de ácidos nucleicos monocatenarios de secuencia variable, comprendiendo cada uno de sustancialmente todos de dichos moldes una secuencia codificante y una secuencia promotora unida operativamente;
- (b) hibridar con los miembros de dicha primera población una segunda población de fragmentos de ácidos nucleicos monocatenarios, siendo dichos fragmentos de una longitud más corta que dichos moldes de ácidos nucleicos y siendo dichos fragmentos sustancialmente complementarios de dichos moldes de ácidos nucleicos, en los que una subpoblación de dicha segunda población comprende una secuencia sustancialmente idéntica e hibrida con una región de secuencia variable;
- (c) poner en contacto los productos de hibridación de la etapa (b) tanto con una polimerasa de ADN que carece de actividad de desplazamiento de hebra como con una ligasa de ADN en unas condiciones en las que dichos fragmentos actúen como cebadores para completar unas segundas hebras de ácidos nucleicos que son sustancialmente complementarias de dichos moldes de ácidos nucleicos; y
- (d) poner en contacto los productos de la etapa (c) con una polimerasa de ARN para generar una tercera población de moléculas de ARN, siendo dicha tercera población de moléculas de ARN transcrita a partir de dichas segundas hebras de ácidos nucleicos y teniendo una variación reducida en la secuencia con respecto a dicha primera población de moldes de ácidos nucleicos monocatenarios.

4. El procedimiento de la reivindicación 3, en el que dicho procedimiento se usa para eliminar una o más mutaciones de dicha primera población de moldes de ácidos nucleicos monocatenarios.

5. El procedimiento de la reivindicación 1 ó 3, en el que dicha mezcla de fragmentos de ácidos nucleicos monocatenarios sustancialmente complementarios se genera escindiendo moléculas de ácidos nucleicos bicatenarios o mediante la síntesis de oligonucleótidos aleatorios.

6. El procedimiento de la reivindicación 1 ó 3, en el que dicha mezcla de fragmentos de ácidos nucleicos monocatenarios sustancialmente complementarios comprende al menos aproximadamente 100 especies diferentes de fragmentos de ácidos nucleicos.

7. El procedimiento de la reivindicación 1 ó 3, que comprende además la generación de hebras complementarias de dichas segundas hebras de ácidos nucleicos, proporcionando dichas hebras complementarias como dicha población o dicha primera población de moldes de ácidos nucleicos monocatenarios, y repitiendo las etapas (b) y (c) durante una o más rondas, y repitiendo después la etapa (d).

8. El procedimiento de la reivindicación 1, en el que dicho procedimiento comprende además proporcionar uno o más fragmentos de ácidos nucleicos monocatenarios que forman un homodúplex con dichos moldes de ácidos nucleicos monocatenarios y llevar a cabo la etapa (b) en presencia de dichos fragmentos formadores de homodúplex.

9. El procedimiento de la reivindicación 1 ó 3, en el que dicho promotor es un promotor T7.

10. El procedimiento de la reivindicación 1 ó 3, en el que dicha polimerasa de ADN es polimerasa de ADN T4.

11. El procedimiento de la reivindicación 1 ó 3, en el que dicho procedimiento comprende además amplificar dicho producto de la etapa (c) antes de dicha etapa de contacto (d).

12. El procedimiento de la reivindicación 1 ó 3, que comprende además un procedimiento para generar una genoteca de proteínas, en donde dicho procedimiento comprende la etapa de:
- (e) traducir dicha genoteca de ARN para generar una genoteca de proteínas.
- 5 13. El procedimiento de la reivindicación 1 ó 3, que comprende además un procedimiento para generar una genoteca de fusión de ARN-proteínas, en donde dicho procedimiento comprende las etapas de:
- (e) unir una molécula aceptora de aminoácidos al extremo 3' de dicha secuencia codificante de cada uno de sustancialmente todos los miembros de dicha genoteca de ARN; y opcionalmente
- (f) traducir dicha genoteca de ARN para generar una genoteca de fusión de ARN-proteínas.
14. Un procedimiento para generar una genoteca de ácidos nucleicos, comprendiendo dicho procedimiento:
- 10 (a) proporcionar una primera población de moldes de ácidos nucleicos monocatenarios, comprendiendo cada uno de dichos moldes una secuencia codificante;
- (b) proporcionar una segunda población de moléculas de ácidos nucleicos monocatenarios de secuencia variable, siendo dicha primera población de moldes de ácidos nucleicos monocatenarios y dicha segunda población de moléculas de ácidos nucleicos monocatenarios de secuencia variable sustancialmente
- 15 complementarias;
- (c) hibridar dicha primera población de moldes de ácidos nucleicos monocatenarios con dicha segunda población de moléculas de ácidos nucleicos monocatenarios de secuencia variable en unas condiciones suficientes para formar dúplex; y
- 20 (d) poner en contacto dichos dúplex con una o más enzimas de escisión/repación en unas condiciones que permitan a dichas enzimas corregir los pares de bases mal apareados en dichos dúplex.
15. El procedimiento de la reivindicación 14, en donde dicho procedimiento comprende además, después de la etapa (d), proporcionar una nueva población de moldes de ácidos nucleicos monocatenarios que comprenden una secuencia codificante, derivando dichos moldes del producto de la etapa (d), sustituyendo dicha nueva población de moldes por dicha primera población de moldes, y repitiendo las etapas (c), y (d) durante una o más rondas.
- 25 16. Un procedimiento para generar una genoteca de ácidos nucleicos, comprendiendo dicho procedimiento:
- (a) proporcionar una población de moldes de ácidos nucleicos monocatenarios, comprendiendo cada uno de dichos moldes una secuencia codificante;
- (b) hibridar con dicha población de moldes de ácidos nucleicos monocatenarios una mezcla de fragmentos de ácidos nucleicos monocatenarios sustancialmente complementarios, siendo dichos fragmentos de una
- 30 longitud más corta que dichos moldes de ácidos nucleicos;
- (c) poner en contacto cada uno de los productos de hibridación de la etapa (b) tanto con una polimerasa de ADN que carece de actividad de desplazamiento de hebra como con una ligasa de ADN en unas condiciones en las que dichos fragmentos actúen como cebadores para completar una segunda hebra de ácidos nucleicos que es sustancialmente complementaria de dichos moldes de ácidos nucleicos; y
- 35 (d) poner en contacto los productos de la etapa (c) con una o más enzimas de escisión/repación en unas condiciones que permitan a dichas enzimas corregir los pares de bases mal apareados en dichos productos.
17. El procedimiento de la reivindicación 16, en donde dicho procedimiento comprende además las etapas de: proporcionar una nueva población de moldes de ácidos nucleicos monocatenarios que comprenden una secuencia codificante y que derivan del producto de la etapa (d), sustituir dicha nueva población de moldes por dicha población
- 40 de moldes de ácidos nucleicos monocatenarios y repetir las etapas (b)-(d) durante una o más rondas.
18. El procedimiento de la reivindicación 14 ó 16 en el que dicho contacto con dichas enzimas de escisión/repación se lleva a cabo *in vitro*.
19. El procedimiento de la reivindicación 14 ó 16, en el que dicho contacto con dichas enzimas de escisión/repación se lleva a cabo en una célula bacteriana.
- 45 20. El procedimiento de la reivindicación 1, 3, 14 ó 16 en el que dichos moldes de ácidos nucleicos monocatenarios se generan usando un fago M13 portador de dichos moldes de ácidos nucleicos, mediante digestión de una hebra de los moldes de ácidos nucleicos bicatenarios usando exonucleasa del gen VI o exonucleasa lambda, mediante la captura de hebras simples de ácidos nucleicos biotiniladas usando estreptavidina, o mediante una transcripción inversa del ARN.
- 50 21. El procedimiento de la reivindicación 1, 3, 14 ó 16, en el que la etapa (b) se lleva a cabo usando entre 1 y aproximadamente 1.000 moléculas de ácidos nucleicos monocatenarios de secuencia variable o fragmentos de ácidos nucleicos monocatenarios por cada uno de dichos moldes de ácidos nucleicos monocatenarios.
22. El procedimiento de la reivindicación 14 ó 16, en donde dicho procedimiento comprende además la etapa de:
- (e) amplificar dicho producto de la etapa (d).

23. El procedimiento de la reivindicación 14 ó 16, en el que cada una de dichas secuencias codificantes está unida operativamente a una secuencia promotora.
24. El procedimiento de la reivindicación 23, en el que dicho procedimiento comprende además la etapa de:
- (e) transcribir los productos de la etapa (d) para generar una genoteca de ARN.
- 5 25. El procedimiento de la reivindicación 24, que comprende además un procedimiento para generar una genoteca de proteínas, en donde dicho procedimiento comprende la etapa de:
- (f) traducir dicha genoteca de ARN para generar una genoteca de proteínas.
26. El procedimiento de la reivindicación 24, que comprende además un procedimiento para generar una genoteca de fusión de ARN-proteínas, en donde dicho procedimiento comprende las etapas de:
- 10 (f) unir una molécula aceptora de aminoácidos al extremo 3' de dicha secuencia codificante de cada uno de sustancialmente todos los miembros de dicha genoteca de ARN; y opcionalmente
(g) traducir dicha genoteca de ARN para generar una genoteca de fusión de ARN-proteínas.

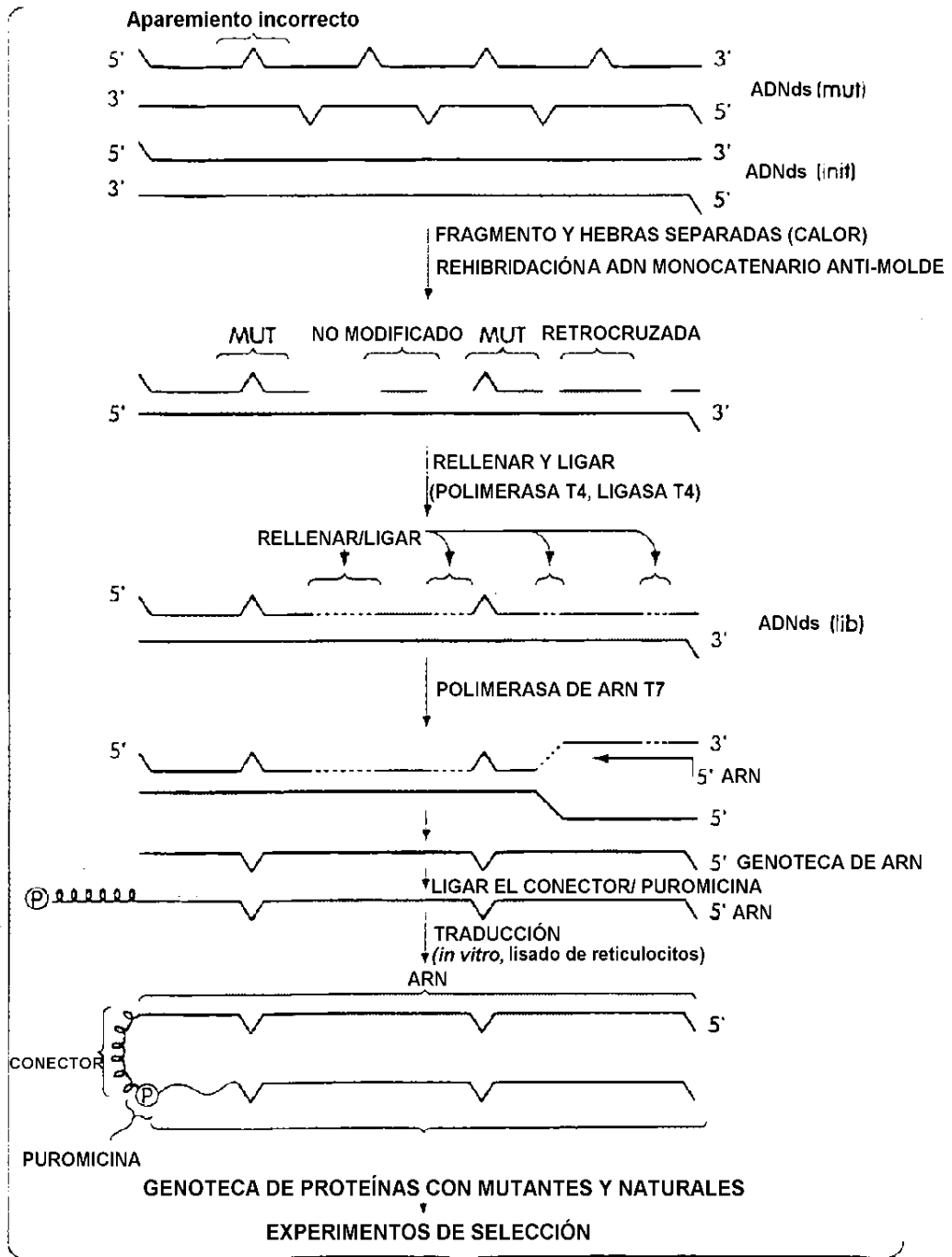


Fig.1

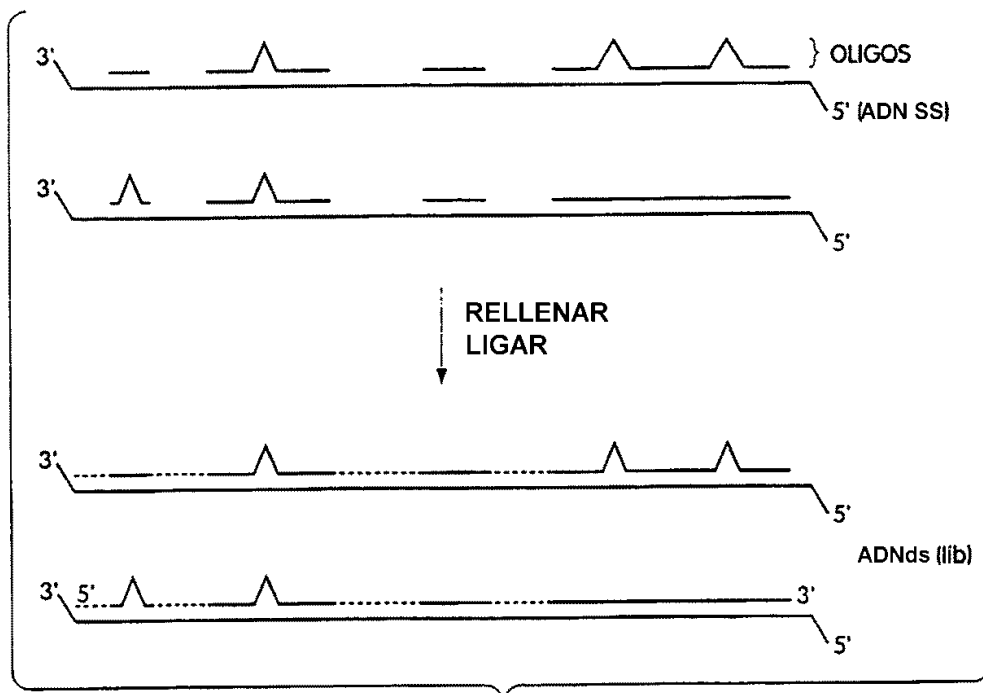


Fig. 2