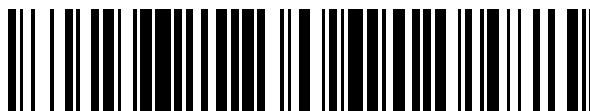


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 378 561**

51 Int. Cl.:
C12N 15/66 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **06765007 .7**
96 Fecha de presentación: **18.07.2006**
97 Número de publicación de la solicitud: **1907548**
97 Fecha de publicación de la solicitud: **09.04.2008**

54 Título: **Biblioteca de oligonucleótidos que codifican péptidos aleatorizados**

30 Prioridad:
22.07.2005 GB 0515131

45 Fecha de publicación de la mención BOPI:
13.04.2012

45 Fecha de la publicación del folleto de la patente:
13.04.2012

73 Titular/es:
**ASTON UNIVERSITY
ASTON TRIANGLE
BIRMINGHAM B4 7ET, GB**

72 Inventor/es:
**ASHRAF, Mohammed;
HUGHES, Marcus, Daniel y
HINE, Anna, Victoria**

74 Agente/Representante:
Sugrañes Moliné, Pedro

ES 2 378 561 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

Biblioteca de oligonucleótidos que codifican péptidos aleatorizados

5 La invención se refiere a la producción de bibliotecas de oligonucleótidos que codifican péptidos aleatorizados, a vectores y células huésped que contienen dichas bibliotecas y a kits para la producción de dichas bibliotecas.

10 Las técnicas combinatoriales para la producción de péptidos han estado disponibles desde la década de los años sesenta y han incluido la síntesis de péptidos en fase sólida y, adicionalmente, los procedimientos paralelos de síntesis en fase sólida desarrollados en los años ochenta. Estas técnicas se revisan en el artículo de Pinilla C., y otros (Nature Medicine (2003), Vol. 9, páginas 118-122).

15 La producción de bibliotecas de ADN que codifican diferentes péptidos es bien conocida *per se* en la materia. Las bibliotecas de genes aleatorizados tienen poco en común con la genómica convencional o con las bibliotecas de ADNc. Las bibliotecas convencionales consisten en clones que cubren colectivamente un genoma/transcriptoma completo y, generalmente, se seleccionan mediante hibridación de ácidos nucleicos. Por el contrario, las bibliotecas aleatorizadas generalmente contienen variaciones de un gen o de un fragmento de un gen que se seleccionan por su actividad novedosa. Los genes aleatorizados se expresan y seleccionan de forma convencional, por ejemplo, en fagos, bacterias o en técnicas de despliegue *in vitro*.

20 Convencionalmente, las técnicas de aleatorización de genes convencionales obligan a la clonación de un exceso de genes usando los codones NNN o NN^{G/T} donde N = A, C, G y T, y tienen que clonarse 64 o 32 codones, respectivamente, para asegurarse la representación de los 20 aminoácidos. En un principio, estos datos pueden parecer relativamente pequeños pero, si consideramos posiciones múltiples de aleatorización, existe una relación exponencial entre el número de genes clonados y el número de proteínas obtenidas. Hine y otros han descrito previamente un procedimiento alternativo para producir bibliotecas de ADN, la aleatorización "MAX", en la que se aleatorizan dos o más posiciones de modo que están representados los 20 aminoácidos, o un subgrupo de los mismos (publicación de PCT WO 00/15777). Esta metodología requiere un conjunto de oligonucleótidos para cada posición que se va a aleatorizar, que hibrida con un molde, convencionalmente (NNN), aleatorizado en las posiciones apropiadas. Esta aproximación permite que cada aminoácido esté codificado sólo una vez dentro de los conjuntos de oligonucleótidos, por tanto, el número de genes únicos generados es equivalente al número de proteínas codificadas, independientemente del número de posiciones de aleatorización. Aunque esta metodología representa una mejora significativa con respecto a las técnicas tradicionales, sigue habiendo un porcentaje relativamente alto de codones no MAX no deseados (~10%) en las posiciones aleatorizadas. Adicionalmente, debido a las limitaciones del proceso de hibridación sólo se producen pequeñas cantidades de las construcciones de ADN, las cuales son difíciles de manipular especialmente cuando codifican subconjuntos de aminoácidos.

35 Posteriormente, Hine y otros han hecho mejoras en esta metodología que eliminan prácticamente la presencia de secuencias no deseadas y mejoran el rendimiento de ADN (Hine, A.V., Hughes, M.D., Nagel, D.A., Ashraf, M. y Santos, A.F. (2002). MAX Codon Gene Libraries. Documento WO 03/106679; Hughes M.D., Nagel D.A., Santos A.F., Sutherland A.J. y Hine A.V. (2003). Removing the redundancy from randomised gene libraries. J. Mol. Biol. 331 (5), 973-979). Esta metodología emplea algunos oligonucleótidos adicionales que hibridan con la cadena molde pero que también proporcionan una extensión que no es complementaria a la cadena molde. Esto permite la amplificación selectiva de la cadena codificadora requerida aumentando de este modo el rendimiento y minimizando las secuencias no deseadas.

40 El problema asociado con los procedimientos previos de la técnica ha sido la producción de más de dos codones aleatorizados contiguos. Esto es importante para permitir extensiones de varios aminoácidos en la secuencia que se va a aleatorizar. Las variaciones de los procedimientos descritos en, por ejemplo, el documento WO 03/106679 sólo permiten aleatorizar dos codones contiguos ya que se requiere el uso de secuencias de oligonucleótidos flanqueantes que hibriden con las cadenas molde. Adicionalmente, estos procedimientos requieren la producción de una cadena molde aleatorizada que se producirá antes de usar los oligonucleótidos de selección. Potencialmente esto limita el número de codones que pueden aleatorizarse debido a la complejidad/masa del oligonucleótido molde.

55 Los inventores probaron diversas técnicas en las que se aleatorizaban tres o más oligonucleótidos consecutivos, lo que demostraba que era difícil obtener o, alternativamente producir, unos resultados satisfactorios. Entre estos se incluyen intentar ligar aleatoriamente tres o más trinucleótidos MAX en un oligonucleótido de cadena sencilla usando la ARN ligasa. Se comprobó que esta última técnica proporcionaba un resultado muy pobre y la posterior amplificación por PCR producía una mancha de la que los inventores no podían aislar muestras individuales. También se ha comprobado que es difícil la adición de codones MAX, que contienen 3 nucleótidos, ligados directamente al extremo 3' de un oligonucleótido.

60 La técnica identificada ahora y descrita por los inventores permite inesperadamente la producción de codones aleatorizados contiguos sin la necesidad de producir oligonucleótidos molde aleatorizados.

65

La invención proporciona un procedimiento para producir una biblioteca de oligonucleótidos como se define en la reivindicación 1.

Los oligonucleótidos aleatorizados difieren porque tienen diferentes codones aleatorizados.

Los oligonucleótidos utilizados son preferiblemente de ADN, sin embargo, pueden usarse otros nucleótidos de cadena doble, como ARN o análogos de ADN o ARN.

Preferiblemente, los oligonucleótidos iniciadores de cadena doble comprenden un extremo como al cual se ligan los oligonucleótidos de cadena doble aleatorizados.

El codón de aleatorización de la cadena codificadora y el codón complementario de la cadena no codificadora forman, preferiblemente, un extremo como en los oligonucleótidos de aleatorización y están preferiblemente ligados al extremo como del oligonucleótido iniciadores de cadena doble.

Al menos una porción del oligonucleótido iniciador al que se une el codón de aleatorización puede codificar una parte de un gen o de otra secuencia de nucleótidos que codifica una secuencia de aminoácidos predeterminada.

Preferiblemente, los oligonucleótidos aleatorizados y los oligonucleótidos iniciadores se ligan mediante una ADN ligasa. Las ADN ligasas son bien conocidas en la materia. Por ejemplo, *E. coli* y el fago T4 codifican una enzima, la ADN ligada, que sella mellas de cadena sencilla entre oligonucleótidos adyacentes en una cadena de ADN doble. Los requisitos de las diferentes enzimas son bien conocidos. Por ejemplo, la enzima de T4 necesita ATP mientras que la enzima de *E. coli* necesita NAD⁺. En cada caso, el cofactor se escinde y forma un complejo enzima-AMP. El complejo se une a uno de los lados de las cadenas de ADN que se quieren unir y establece un enlace covalente entre un 5'-fosfato de una cadena y un grupo 3'-OH de la cadena adyacente.

Por tanto, preferiblemente al menos uno de los extremos 5' de los oligonucleótidos que se van a ligar comprende un grupo fosfato que permite que los oligonucleótidos se ligen mediante una ADN ligasa.

Preferiblemente, ambos extremos 5' de los oligonucleótidos que se van a ligar comprenden un grupo fosfato. Preferiblemente la cadena adyacente a los grupos fosfato contendrán un grupo 3'-OH.

El sitio de escisión predeterminado para la endonucleasa está inmediatamente adyacente al codón de aleatorización.

Los oligonucleótidos ligados pueden amplificarse mediante, por ejemplo, PCR usando cebadores complementarios a, p. ej., secuencias de los oligonucleótidos iniciadores.

El producto de PCR producido puede purificarse y aislarse, por ejemplo, usando técnicas convencionales, como una electroforesis en geles de poliacrilamida (PAGE) y la escisión de la banda relevante de ADN antes del aislamiento del ADN y la digestión con la endonucleasa de restricción.

Se sigue la etapa (f) una o más veces para su uso en una biblioteca de oligonucleótidos que comprende diversos oligonucleótidos, teniendo cada oligonucleótido y biblioteca al menos dos codones aleatorizados contiguos. Más preferiblemente, el número de codones aleatorizados contenido en la biblioteca es de 1 o más, más preferiblemente 3, 4, 5, 6, 7, 8, 9 o 10 codones.

Preferiblemente el procedimiento proporciona ligar un oligonucleótido de finalización de cadena doble al oligonucleótido iniciador después del número requerido de codones aleatorizados que se ha añadido. El oligonucleótido de finalización puede comprender una secuencia de nucleótidos que comprende un sitio de reconocimiento de una endonucleasa de restricción que permite que la secuencia de nucleótidos aleatorizada se ayuste dentro de un gen de interés. Alternativamente, por ejemplo, el oligonucleótido de finalización puede codificar un fragmento no aleatorizado de un gen de interés.

La secuencia de nucleótidos unida al codón de aleatorización del oligonucleótido de aleatorización puede ser diferente en cada ronda de adición de codones de aleatorización. Es decir, cada conjunto de oligonucleótidos de aleatorización usado en cada ronda de las etapas (a) a (f), cuando se repite el procedimiento para añadir codones aleatorizados adicionales, puede contener una secuencia diferente unida al codón aleatorizado. Esto permite que los oligonucleótidos de cadena doble aleatorizados obtenidos después de cada ronda de adición aleatoria de codones se amplifiquen selectivamente mediante PCR con un cebador complementario a la secuencia diferente. Se espera que esto reduzca la necesidad de purificación por PAGE entre los ciclos de adición aleatoria de codones.

Preferiblemente, el sitio de reconocimiento de la endonucleasa de restricción y el sitio de escisión son:

5'-GAGTCNNNNN[^]-3' (SEQ ID. No. 1)

3'-CTCAGNNNNN^'-5' (SEQ ID. No. 2)

donde:

- 5 N = cualquier nucleótido
 ^ = el sitio de escisión de la endonucleasa de restricción.

10 Este sitio para la endonucleasa de restricción ha sido identificado como reconocido por dos endonucleasas de restricción: MlyI, que se obtiene de New England Biolabs, Inc. y Schl, disponible en Fermentas Life Sciences. Las dos enzimas son producidas por diferentes microorganismos. MlyI, por ejemplo, se describe en el documento US 6.395.531 y se encuentra en *Micrococcus lylae*. Ambas endonucleasas reconocen la secuencia de ADN de cadena doble 5'-GAGTC-3' y escinden 5 bases de ADN secuencia abajo generando extremos romos. Sin embargo, también puede usarse cualquier endonucleasa de restricción que se una a una secuencia de endonucleasa de restricción pero corte adyacente a la secuencia de reconocimiento, o varios nucleótidos secuencia arriba o abajo de la misma, permitiendo la separación del codón de aleatorización.

15 El código genético para los diferentes aminoácidos es bien conocido; por ejemplo, el código genético para los codones en el ARNm transcrito a partir de la cadena de codones del ADN es:

	U	C	A	G	
U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA* Stop UAG* Stop	UGU } Cys UGC } UGA* Stop UGG Trp	U C A G
C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
A	AUU } AUC } Ile AUA } AUG** Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
G	GUU } GUC } Val GUA } GUG** }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

- * Terminación de cadena o codones "sin sentido"
 ** También utilizado para especificar el iniciador fomial-Met-tRNA^{Met}. El triplete GUG de Val es, por tanto, "ambiguo" porque puede codificar tanto para valina como para metionina.

Los diferentes aminoácidos codificados por codones diferentes tienen propiedades diferentes. En algunas circunstancias puede ser deseable enfocar la biblioteca hacia grupos diferentes de aminoácidos con propiedades similares. Por tanto, preferiblemente se seleccionan los codones. Por ejemplo, pueden enfocarse hacia aminoácidos muy hidrófobos (como *Val, Ile, Leu, Met, Phe, Trp* o *Cys*), aminoácidos menos hidrófobos (*Ala, Tyr, His, Thr, Ser, Pro* y *Gly*), aminoácidos parcialmente hidrófobos (*Arg* y *Lys*), aminoácidos que contienen sulfuro (*Cys*), aminoácidos cargados positivamente (*Arg, His* y *Lys*), aminoácidos cargados negativamente (*Asp* y *Gly*), etc. Alternativamente, pueden seleccionarse para evitar un aminoácido en particular, como *Pro* o *Cys*.

Preferiblemente se usan al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 codones de aleatorización diferentes en cada ronda de adición de codones de aleatorización.

No todos los codones se utilizan de forma eficaz en cualquier organismo en particular. Por tanto, preferiblemente se seleccionan específicamente los codones que permiten que la biblioteca de ADN aleatorizada se utilice de forma eficaz en cualquier célula huésped para la expresión de la biblioteca de ADN y la producción de péptidos a partir de ésta. Los problemas asociados con la no utilización eficaz de las bibliotecas aleatorizadas se resumen en el artículo de Hughes M.D. y otros (J. Mol. Biol. (2003), Vol. 331, páginas 973-979).

Preferiblemente, se seleccionan los codones que favorecen la expresión de cada aminoácido en un organismo seleccionado. Por ejemplo, para *E. coli* los codones son preferiblemente GCG (A), TGC (C), GAT (D), GAA (E), TTT (F), GGC (G), CAT (H), ATT (I), AAA (K), CTG (L), ATG (M), AAC (N), CCG (P), CAG (Q), CGC (R), AGC (S), ACC (T), GTG (V), TGG (W) y TAT (Y). Las letras entre paréntesis indican el aminoácido codificado por el codón.

Preferiblemente, el extremo 3' de la cadena codificadora comprende un grupo bloqueante para prevenir el ligamiento no productivo de, por ejemplo, copias múltiples de los oligonucleótidos de aleatorización. El grupo bloqueante puede seleccionarse entre un aminoácido, un grupo fosfato, un resto glicerilo, un grupo tiol o, por ejemplo y resto polietilenglicol.

Adicional o alternativamente, la cadena no codificadora puede comprender, en su extremo 5', uno o más nucleótidos que se extienden más allá del extremo 3' de la cadena complementaria a la codificadora. Esto reduce de nuevo el número de ligamientos no productivos producidos.

Preferiblemente, los procedimientos de la invención comprenden adicionalmente las etapas de:

(g) Proporcionar un oligonucleótido iniciador de cadena doble aleatorizado producido mediante un procedimiento según cualquiera de las reivindicaciones precedentes;

(h) Proporcionar un oligonucleótido predefinido que comprende:

(i) una cadena codificadora, comprendiendo la cadena codificadora un codón predefinido que codifica un aminoácido predefinido; y

(ii) una cadena no codificadora sustancialmente complementaria,

donde el oligonucleótido predefinido comprende una secuencia de nucleótidos que codifica un sitio de reconocimiento de una endonucleasa de restricción capaz de ser reconocido por una endonucleasa de restricción, siendo la endonucleasa de restricción capaz de escindir el oligonucleótido predefinido secuencia arriba o abajo del reconocimiento de la endonucleasa en un sitio de escisión predeterminado para crear un corte de extremos romos;

(i) Ligar el oligonucleótido iniciador de cadena doble aleatorizado con el oligonucleótido predefinido para formar un oligonucleótido ligado;

(j) Amplificar el oligonucleótido ligado; y

(k) Digerir el oligonucleótido ligado con la endonucleasa de restricción para formar un oligonucleótido aleatorizado que comprende en un extremo el codón predefinido.

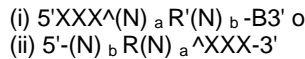
El codón predefinido está preferiblemente en un extremo del oligonucleótido predefinido.

Esto permite la inserción de un aminoácido conocido en una posición predeterminada del péptido producido mediante las bibliotecas de oligonucleótidos. La ventaja de esta técnica es que permite el efecto de cualquier aminoácido en particular en esta biblioteca predefinida que está siendo estudiada. Ejemplos de estas bibliotecas de selección posicional se describen en, por ejemplo, el artículo de Pinilla (Nature medicine (2003), Vol. 9, páginas 188-122). El procedimiento de la invención permite conseguir la producción de estas bibliotecas de selección posicional de forma relativamente rápida y relativamente fácil.

Se añaden uno o más codones aleatorios adicionales a un extremo del codón predefinido usando los procedimientos de la invención. El sitio de reconocimiento de la endonucleasa de restricción, la endonucleasa de restricción, el grupo bloqueante y/o los grupos de extensión opcionales al extremo 3' de la cadena no codificadora pueden ser como se define anteriormente.

5 El oligonucleótido aleatorizado que comprende el codón predefinido puede usarse a continuación como oligonucleótido iniciador para añadir uno o más codones aleatorios adicionales.

10 Preferiblemente, la cadena codificadora del oligonucleótido de aleatorización utilizada en el procedimiento de la invención comprende la secuencia:



15 donde:

XXX es el codón de aleatorización,

N es cualquier nucleótido,

a es un número entero de 0 a 10, preferiblemente 5,

20 b es un número entero de 0 a 40, preferiblemente de 1 a 20 o, preferiblemente, de 1 a 10.

^ es el sitio de escisión de la endonucleasa de restricción,

R' es el complemento inverso de la secuencia del sitio de reconocimiento de la endonucleasa de restricción,

R es la secuencia del sitio de reconocimiento de la endonucleasa de restricción,

25 B puede o no estar presente y puede seleccionarse entre -OH, -NH₂,

fosfato, un resto glicerilo, un tiol y un resto polietilénico.

El segundo oligonucleótido aleatorizado (ii) puede usarse ligado al extremo 5' del oligonucleótido iniciador.

30 Preferiblemente, los oligonucleótidos obtenidos por los procedimientos de la invención se insertan en un vector de expresión adecuado. Los vectores de expresión, por ejemplo, para expresar los péptidos codificados por los oligonucleótidos, son bien conocidos en la materia. Por ejemplo, son bien conocidas en la materia las bibliotecas de expresión de fagos en las que se insertan los oligonucleótidos dentro de la secuencia de nucleótidos que codifica las proteínas de la cubierta del fago, de modo que los péptidos codificados por estos se expresan en la superficie de las partículas del fago. Adicionalmente, los vectores de expresión en la superficie de bacterias, vectores de expresión de levaduras y otros vectores de expresión eucariotas son bien conocidos en la materia.

35 Preferiblemente, los oligonucleótidos de aleatorización usados en cada ciclo de adición de codones difieren en que tienen una secuencia diferente, (N)_b. Es decir, en un primer ciclo los oligonucleótidos de aleatorización tienen una primera secuencia (N)_b con codones de aleatorización diferentes. En un segundo ciclo posterior, los oligonucleótidos de aleatorización tienen una segunda secuencia (N)_b unida a codones de aleatorización. Esto permite que los oligonucleótidos aleatorizados obtenidos después de cada ciclo de adición de codones aleatorizados se amplifiquen con un cebador específico para la secuencia (N)_b' o (N)_b.

40 Preferiblemente, el vector de expresión se inserta en una célula huésped adecuada, como una célula bacteriana (p. ej., *E. coli*), de levadura, etc. La célula huésped expresa el péptido codificado por la biblioteca de ADN que, a continuación, se usa para su estudio adicional.

45 También se proporcionan los procedimientos para producir una biblioteca de péptidos aleatorizada que comprenden la expresión de la biblioteca de oligonucleótidos obtenidos mediante un procedimiento según la invención.

Se describen las bibliotecas de proteínas obtenidas por los procedimientos de las reivindicaciones.

50 También se describe una biblioteca de oligonucleótidos aleatorizada que comprende diversos oligonucleótidos, teniendo cada oligonucleótido 3 o más codones MAX aleatorizados contiguos que representan el uso óptimo de codones de un organismo predeterminado y en el que los codones MAX son diferentes entre los diferentes miembros de la biblioteca.

55 Preferiblemente, la biblioteca aleatorizada comprende 1 o 2, más preferiblemente 3, 4, 5, 6, 7, 8, 9 o 10 codones MAX. También se describen kits para producir un oligonucleótido mediante un procedimiento según las reivindicaciones.

Preferiblemente, el kit comprende diversos oligonucleótidos de aleatorización diferentes que comprenden:

60 (i) una cadena codificadora, comprendiendo la cadena codificadora un codón de aleatorización; y

(ii) una cadena no codificadora sustancialmente complementaria,

donde cada oligonucleótido de aleatorización de cadena doble comprende una secuencia de nucleótidos que codifica un sitio de reconocimiento de endonucleasa de restricción capaz de ser reconocido por una endonucleasa de restricción, siendo la endonucleasa de restricción capaz de escindir el oligonucleótido de aleatorización secuencia arriba o abajo del reconocimiento de la endonucleasa en un sitio de escisión predeterminado para crear un corte de extremos romos.

Preferiblemente, el sitio de reconocimiento es:

5'-GAGTCNNNNN[^]-3' (SEC ID N.º 1)
3'-CTCAGNNNNN[^]-5' (SEC ID N.º 2)

donde:

N = cualquier nucleótido
[^] = el sitio de escisión de la endonucleasa de restricción.

El kit puede comprender adicionalmente una enzima de restricción capaz de escindir el oligonucleótido de aleatorización y el sitio de escisión predeterminado. Preferiblemente, la endonucleasa de restricción se selecciona entre SchI y MlyI.

Preferiblemente, los codones de aleatorización consisten en codones MAX que representan el uso óptimo de codones de un organismo de interés predeterminado o una selección predeterminada de dichos codones MAX.

La cadena codificadora preferiblemente es el oligonucleótido de aleatorización de doble cadena que comprende un extremo 5' y un extremo 3', comprendiendo el extremo 3' de la cadena codificadora un grupo bloqueante.

Preferiblemente, el grupo bloqueante se selecciona entre un grupo amino, un grupo fosfato, un resto glicerol, un grupo tiol y un resto polietilenglicol.

Preferiblemente, la cadena no codificadora comprende un extremo 3' y un extremo 5', extendiéndose el extremo 5' de la cadena no codificadora uno o más nucleótidos más allá del extremo 3' de la cadena complementaria a la codificadora.

Preferiblemente, la cadena codificadora comprende la secuencia:

(i) 5'XXX[^](N)_aR'(N)_b-B3' o
(ii) 5'-(N)_bR(N)_a[^]XXX-3'

donde:

XXX es el codón de aleatorización,
N es cualquier nucleótido,
a es un número entero de 0 a 10, preferiblemente 5,
b es un número entero de 0 a 40, preferiblemente de 1 a 20 o, preferiblemente, de 1 a 10.
[^] es el sitio de escisión de la endonucleasa de restricción,
R' es el complemento inverso de la secuencia del sitio de reconocimiento de la endonucleasa de restricción,
R es la secuencia del sitio de reconocimiento de la endonucleasa de restricción,
B puede o no estar presente y, cuando está presente, puede seleccionarse entre -OH, -NH₂, fosfato, un resto glicerilo, un tiol y un resto polietilenglicol.

El kit puede comprender adicionalmente un oligonucleótido predefinido que comprende:

(i) una cadena codificadora, comprendiendo la cadena codificadora un codón predefinido que codifica un aminoácido predefinido; y

(ii) una cadena no codificadora sustancialmente complementaria,

donde el oligonucleótido predefinido comprende una secuencia de nucleótidos que codifica un sitio de reconocimiento de endonucleasa de restricción capaz de ser reconocido por una endonucleasa de restricción, siendo la endonucleasa de restricción capaz de escindir el oligonucleótido predefinido secuencia arriba o abajo del reconocimiento de la endonucleasa en un sitio de escisión predeterminado para crear un corte con extremos romos.

La invención también proporciona células huésped que comprenden una biblioteca de ADN obtenible mediante un procedimiento según las reivindicaciones.

5 Los kits pueden usarse para producir bibliotecas de péptidos para analizar el efecto de diferentes secuencias sobre, por ejemplo, la unión de ligandos o el efecto sobre una actividad biológica del péptido.

A continuación, la invención se describirá mediante un ejemplo sólo en referencia a las siguientes figuras:

10 En la **Fig. 1** se muestra esquemáticamente un procedimiento de producción de secuencias de ADN que contienen codones MAX en posiciones predeterminadas según la presente invención.

En la **Fig. 2** se muestra la distribución de codones MAX o codones no MAX en las posiciones predeterminadas dentro de las secuencias de ADN producidas mediante el procedimiento de la presente invención.

15 En la **Fig. 3** se muestra un diagrama esquemático que indica un procedimiento alternativo de la invención en el que se añaden codones MAX en el extremo opuesto al oligonucleótido iniciador como se muestra en la Figura 1.

EJEMPLO

20 En la figura 1 se muestra un esquema del procedimiento utilizado para generar la biblioteca de ADN aleatorizada que contiene codones MAX en las siete posiciones predeterminadas. MAX indica un codón que representa uno de los 20 codones favorecidos en *E. coli*. NH₂ representa un grupo amino presente en el extremo 3' de este oligonucleótido para minimizar los ligamientos no productivos. El recuadro gris representa el sitio para la endonucleasa de restricción *MlyI*

25

Las etapas principales implicadas en la producción de la biblioteca son:

1. Hibridación de los pares de oligonucleótidos complementarios
2. Ligamiento de los oligonucleótidos de inicio A y de aleatorización B
- 30 3. Amplificación por PCR del producto de ligamiento.
4. Digestión con *MlyI* y purificación del producto mediante PAGE del producto C
5. Repetición del proceso anterior para conseguir el número deseado de posiciones aleatorizadas contiguas.
6. Clonación de las construcciones de ADN en un vector apropiado

35

La hibridación, ligamiento y clonación se realizaron como se describe a continuación y las construcciones se transformaron en células de *E. coli* DH5α (genotipo: F' 80d1acZ(1acZYA-argF)U169 deoR recA1 endA1 hsdR17(rK-, mK+)phoA supE44 - thi-1 gyrA96 relA1/F' proAB+ lacIqZM15 Tn10(tetr)) químicamente competentes que fueron inducidas a captar ADN mediante choque térmico. Se aislaron los clones y se recuperó el ADN del plásmido. El ADN insertado se secuenció para establecer las secuencias presentes en las posiciones predeterminadas.

40

Puede usarse cualquier célula huésped adecuada y, de hecho, vectores de expresión alternativos, en lugar de esta cepa de *E. coli*.

MATERIALES Y PROCEDIMIENTOS

Síntesis del ADN iniciador

El par completamente complementario de oligonucleótidos de ADN iniciadores fue sintetizado por MWG Biotech.

50

Síntesis de oligonucleótidos de aleatorización

Los pares completamente complementarios de oligonucleótidos de aleatorización fueron sintetizados por MWG Biotech. Los oligonucleótidos de aleatorización se diseñaron para codificar un único codón MAX en el extremo 5' y un grupo amino en el extremo 3' de la cadena codificadora. La cadena no codificadora no estaba modificada con el grupo amino.

55

Fosforilación

Las reacciones de fosforilación en 5' de las cadenas codificadoras de los oligonucleótidos de aleatorización se realizaron en un volumen final de 50 µl, siempre que no se especificara otra cosa. Las reacciones consistían en tampón de ligasa 1 x (NEB - New England Biolabs), 10 unidades de T4 PNK, 300 pmoles de ADN y agua hasta un volumen final de 50 µl. La reacción se incubó a 37 °C durante 30 min y, a continuación, la reacción se detuvo elevando la temperatura a 65 °C durante 20 min.

65

Hibridación

Las reacciones de hibridación se establecieron usando cantidades iguales de dos oligonucleótidos. El volumen final de la reacción era de 50 µl. Esta reacción se calentó a 95 °C y se mantuvo a esta temperatura durante 2 min. Se dejó que la temperatura descendiera a 1°C/min hasta 4 °C y se dejó que las secuencias complementarias hibridasen.

Ligamientos

Se realizaron ligamientos de extremos romos entre los oligonucleótidos iniciadores y un conjunto de 20 oligonucleótidos de aleatorización. Los ligamientos se establecieron usando cantidades iguales de oligonucleótidos sentido e inversos (50 pmoles), tampón ligasa 1 x (NEB), 10 unidades de ligasa (NEB) y agua hasta un volumen final de 2011. La reacción se incubó a 26 °C durante la noche.

Reacción en cadena de la polimerasa (PCR).

Las reacciones de PCR contenían 1 unidad de polimerasa *Pfu* (Promega), 50 pmoles de cada uno de los cebadores de PCR, 1 µl de iniciador, dNTP 200 µM, tampón de *Pfu* 1 x y agua doble destilada hasta completar el volumen a 100 µl. El ADN se amplificó usando las condiciones siguientes: 94°C durante 30 segundos, 48°C durante 30 segundos, 72°C durante un minuto hasta un total de 35 ciclos. Las reacciones se completaron a 72 °C durante 7 minutos y las muestras se conservaron a 4°C.

Extracción de ADN con fenol cloroformo.

Se añadieron 0,1 volumen de acetato sódico 3M (pH 5,2) a la reacción del ADN que se estaba purificando, la cual se mezcló a continuación mediante agitación. Posteriormente, se añadió un volumen de fenol/cloroformo/alcohol isoamílico (25:24:1); la mezcla se agitó y centrifugó durante 2 min a 14.000 rpm. La capa acuosa que contenía el ADN se apartó con cuidado a un tubo de microfuga limpio. Se añadió un volumen de cloroformo y la mezcla resultante se agitó y centrifugó durante 2 min a 14.000 rpm. La fase acuosa se apartó a un tubo de microfuga limpio y se añadieron 2 volúmenes de etanol frío en hielo y la muestra se agitó. El tubo de microfuga se colocó a -20 °C durante la noche o a -70 °C durante 1 hora y, a continuación, se congeló antes de la centrifugación durante 20 minutos 14.000 rpm. Se eliminó el sobrenadante y el sedimento de ADN se lavó con 200 µl de etanol al 70%. Se eliminó el sobrenadante y se dejó secar al aire el sedimento antes de su resuspensión en agua destilada. Las muestras se conservaron a -20 °C hasta que fue necesario.

Digeridos de restricción del ADN.

Los digeridos de restricción se establecieron en un tampón 1 x apropiado (según las instrucciones del fabricante) con 0,1 µg/µl de ADN, 10 unidades de enzima de restricción y se ajustaron a 2011 con agua doble destilada. Los digeridos de restricción se incubaron a 37 °C o 55 °C durante dos horas según la recomendación del fabricante. A continuación, la temperatura se elevó a 65 °C durante veinte minutos para desnaturalizar la enzima. Los digeridos dobles se establecieron de la misma forma usando el tampón más apropiado para la combinación de enzimas.

Electroforesis en gel de poliacrilamida.

Los geles PAGE desnaturalizantes se prepararon disolviendo 21 g de urea, 12,5 ml de acrilamida al 40% (proporción 19:1 de acrilamida: bisacrilamida), 5 ml de TBE (10x) (Tris base 0,9 M, ácido bórico 0,9 M y EDTA 20 mM) y se completó hasta un volumen final de 50 ml con agua. Esta mezcla se agitó enérgicamente hasta que la urea se hubo disuelto, se añadieron 600 µl de persulfato amónico al 10% y, posteriormente, 80 µl de TEMED (N,N,N',N',tetraetilmetilendiamina). El equipo de electroforesis se programó según las instrucciones del fabricante y el gel se vertió y fijó. A continuación, las muestras se cargaron en el gel y se aplicó un voltaje de 20 V/cm durante aproximadamente 3 horas. Después se retiró el gel de las placas de cristal y se colocó en tampón TBE 1 x con bromuro de etidio a una concentración de 2 µg/ml. A continuación se colocó en un agitador a 150 rpm y se dejó durante cinco minutos. Después, el gel se retiró del tampón y se fotografió con luz UV. Cuando fue necesario un PAGE no desnaturalizante, el gel se hizo de la misma forma pero se omitió la urea.

Elución del ADN del PAGE.

Se cortó la banda de interés del gel y se colocó en una columna de elución de ADN/proteína. Se añadió tampón TAE 1 x (Tris base 100 mM, ácido acético 19 mM y EDTA 0,2 mM) a la columna y se ajustó la tapa para prevenir cualquier escape. A continuación, la columna se sumergió en una cubeta de electroforesis llena de tampón TAE 1x. Se aplicó un voltaje de 15 V/cm durante 20 min para que el ADN eluyera del gel. Después de 20 min se invirtió la corriente durante 30 s para liberar el ADN de la membrana de la columna de elución dentro del tampón y se retiró la columna de la gradilla. Se retiró con cuidado el tampón de la columna de elución y se precipitó el ADN.

Se añadieron 0,1 volúmenes de acetato sódico 3M (pH 5,2) y 2 µl de coprecipitante Pellet paint al tampón extraído. Se añadieron dos volúmenes de etanol al 100% y la muestra se agitó e incubó a temperatura ambiente durante 1 min. A continuación la muestra se centrifugó a 14.000 rpm durante cinco minutos, se retiró el sobrenadante y el sedimento se lavó con 200 µl de etanol al 70%. Se retiró el sobrenadante y el sedimento se secó antes de la resuspensión en una cantidad apropiada de agua doble destilada.

Preparación de *E. coli* (DH5α) químicamente competente

Se inoculó una única colonia de *E. coli* (DH5α) en 10 ml de SOB y se incubó durante la noche a 37 °C en un agitador a 250 rpm. Se transfirieron 8 ml del cultivo de la noche a 800 ml de LB. Este se incubó a 37 °C en un agitador a 250 rpm hasta que el cultivo estuvo a mitad de camino de la fase logarítmica (DO550 de ~0,45). A continuación, las células se enfriaron en hielo durante 30 minutos y después se sedimentaron mediante centrifugación a 4°C. Posteriormente se retiró el sobrenadante y las células se resuspendieron mediante pipeteo suave en 264 ml de RF1 (RbCl 100 mM, MnCl₂ 50 mM, acetato potásico 30mM, CaCl₂ 10 mM, glicerol al 15% ajustado a pH 5,8 con ácido acético 0,2 M). A continuación, las células resuspendidas se incubaron en hielo durante una hora. Las células se sedimentaron de nuevo y se retiró el sobrenadante. Las células se resuspendieron en 64 ml de RF2 (MOPS (ácido 4-morfolinpropanosulfónico) 10 mM), RbCl 10 mM, CaCl₂ 75 M, glicerol al 15%, ajustado a pH 6,8 con NaOH) y se incubaron en hielo durante 15 minutos. Se dispensaron en alícuotas de 200 µl en tubos de microfuga que, a continuación, se congelaron de forma instantánea en nitrógeno líquido a -70 °C hasta que fue necesario.

Transformación

La cepa de *E. coli* competente (DH5α) se congeló en hielo y se añadieron 100 µl a la mezcla de ligamiento (20 µl) removiendo ligeramente y se incubó en hielo durante 30 minutos. Se dio un choque térmico a las células a 37 °C durante 45 segundos y se devolvieron al hielo durante otros dos minutos. Se añadieron 100 µl de LB x 2 a las células y este se incubó a 37 °C en un agitador a 250 rpm durante 1 hora. Se añadieron 10 µl de IPTG y 50 µl de X-gal al 2% a las células antes de colocarlas en placas con medio selectivo.

Preparación del ADN plasmídico

El ADN plasmídico se preparó para secuenciación usando un kit Wizard miniprep de Promega según las instrucciones del fabricante.

Secuenciación del ADN

La secuenciación automática del ADN se realizó en el laboratorio de genómica de la Universidad de Birmingham en un secuenciador ABI 3700.

RESULTADOS

La Figura 2 muestra la distribución de los diferentes codones MAX obtenidos en las posiciones predeterminadas en los clones aislados. Se secuenciaron un total de 156 clones, que proporcionaron un total de 1.092 posiciones que codifican MAX. Todas las 20 secuencias codificadas estaban representadas en los clones analizados. La columna marcada como "no MAX" hace referencia a codones surgidos pero que no estaban especificados en la mezcla de aleatorización. La columna marcada como "N" hace referencia a aquellos codones en los que la secuencia no podría ser determinada debido a una falta de claridad en la secuenciación. La columna marcada como "Del" hace referencia a aquellos codones que no estaban presentes o contenían una deleción.

La técnica se ha utilizado en esta fase para producir al menos 7 codones aleatorizados contiguos.

En una realización alternativa puede añadirse el codón MAX en el extremo opuesto del oligonucleótido iniciador.

CONCLUSIÓN

La técnica proporciona un procedimiento de producción de oligonucleótidos aleatorizados que tienen codones aleatorizados.

PROCEDIMIENTO ALTERNATIVO

En la Figura 3 se muestra un procedimiento alternativo de adición de codones MAX. Los codones MAX se añaden en el extremo opuesto del oligonucleótido iniciador al que se muestra en la Figura 1. Las anotaciones son las mismas que para la Figura 1. Esto demuestra que esta técnica puede utilizarse para introducir codones MAX en cualquier extremo de una cadena codificadora.

La purificación por PAGE se indica entre paréntesis como en este y de hecho en el sistema mostrado en la Figura 1, la necesidad de PAGE puede reducirse usando oligonucleótidos aleatorizados diferentes con secuencias hacia arriba o hacia abajo del codón MAX diferentes en cada ronda de adición de codones MAX. Esto permite que los oligonucleótidos aleatorizados de cada doble obtenidos después de cada ronda de adición de codones se amplifiquen selectivamente mediante PCR con un cebador complementario a la secuencia diferente.

LISTADO DE SECUENCIAS

5 <110> Universidad de Aston

10 <120> Biblioteca de oligonucleótidos que codifican péptidos aleatorizados

<130> P713229PCT

15 <150> GB0515131.1
<151> 22/07/2005

<160> 2

20 <170> Patent In versión 3.3

<210> 1
<211> 10
<212> ADN

25 <213> Artificial

<220>
<223> Sitio de reconocimiento de endonucleasa de restricción

30 <220>
<221> misc_característica
<222> (6)..(10)
<223> n es a, c, g, t o u

35 <400> 1
gagtcnnnnn 10

<210> 2
<211> 10

40 <212> ADN
<213> Artificial

<220>
<223> Sitio de reconocimiento de endonucleasa de restricción

45 <220>
<221> misc_característica
<222> (1)..(5)
<223> n es a, c, g, t o u

50 <400> 2
nnnnngactc 10

REIVINDICACIONES

1. Procedimiento de producción de una biblioteca de oligonucleótidos que comprende diversos oligonucleótidos, teniendo cada oligonucleótido en la biblioteca al menos dos posiciones predeterminadas, un codón de aleatorización seleccionado entre un grupo definido de codones, codificando los codones dentro de dicho grupo definidos diferentes aminoácidos y comprendiendo dicho procedimiento las etapas de:
- (a) Proporcionar uno o más oligonucleótidos iniciadores de cadena doble, en el que los oligonucleótidos iniciadores tienen uno o más extremos romos;
- (b) Proporcionar diversos oligonucleótidos de aleatorización de cadena doble diferentes que comprenden:
- (i) una cadena codificadora, comprendiendo la cadena codificadora un codón de aleatorización; y
- (ii) una cadena no codificadora sustancialmente complementaria,
- donde cada oligonucleótido de aleatorización de cadena doble comprende una secuencia de nucleótidos que codifica para un sitio de reconocimiento de endonucleasa de restricción capaz de ser reconocido por una endonucleasa de restricción, la endonucleasa de restricción capaz de escindir el sitio de reconocimiento de endonucleasa secuencia arriba o secuencia abajo del oligonucleótido de aleatorización en un sitio de escisión predeterminado para crear un corte de extremos romos y en el que la endonucleasa escinde el oligonucleótido de aleatorización adyacente al codón de aleatorización proporcionado por el oligonucleótido de aleatorización.
- (c) Ligar cada oligonucleótido iniciador de cadena doble a un oligonucleótido de aleatorización de cadena doble para formar oligonucleótidos ligados;
- (d) Amplificar los oligonucleótidos ligados;
- (e) Digerir el oligonucleótido ligado con la endonucleasa de restricción para formar diversos oligonucleótidos aleatorizados de doble cadena, comprendiendo cada uno de ellos en un extremo un codón de aleatorización; y
- (f) Usar los oligonucleótidos iniciadores aleatorizados de cadena doble y repetir las etapas (a) a (e) del procedimiento para producir diversos oligonucleótidos aleatorizados de cadena doble comprendiendo cada uno un codón de aleatorización adicional; en el que la etapa (f) se sigue una vez, u opcionalmente más veces, para producir una biblioteca de oligonucleótidos que comprende diversos oligonucleótidos, teniendo cada oligonucleótido en la biblioteca al menos dos codones aleatorizados contiguos.
2. Procedimiento según la reivindicación 1, en el que el sitio de reconocimiento de la endonucleasa de restricción y el sitio de escisión son:
- 5'-GAGTCNNNNN[^]-3' (SEC ID N.º 1)
 3'-CTCAGNNNNN[^]-5' (SEC ID N.º 2)
- donde
- N = cualquier nucleótido
[^] = el sitio de escisión de la endonucleasa de restricción.
3. Procedimiento según la reivindicación 2 en el que la endonucleasa de restricción se seleccionan entre *SchI* y *MlyI*.
4. Procedimiento según cualquiera de las reivindicaciones precedentes que se centra en el grupo de codones aleatorizados.
5. Procedimiento según cualquier de las reivindicaciones precedentes en el que los oligonucleótidos aleatorizados utilizados en cada etapa (f) según se define en la reivindicación 1 comprende una secuencia de nucleótidos diferente unida al codón de aleatorización, comparado con los utilizados en los ciclos previos de adición de codones aleatorios (etapas (a) a (e)).
6. Procedimiento según cualquiera de las reivindicaciones precedentes que adicionalmente comprende ligar un oligonucleótido de finalización que tiene una secuencia predefinida al oligonucleótido aleatorizado después de que se haya añadido un número predeterminado de codones de aleatorización.
7. Procedimiento según cualquiera de las reivindicaciones precedentes, en el que los codones de aleatorización consisten en codones MAX que representan el uso óptimo de codones de un organismo de interés predeterminado o una selección predeterminada de dichos codones MAX.
8. Procedimiento según cualquiera de las reivindicaciones precedentes, comprendiendo la cadena codificadora del oligonucleótido de aleatorización de doble cadena un extremo 5' y un extremo 3', comprendiendo el extremo 3' de la cadena codificadora un grupo bloqueante.

9. Procedimiento según la reivindicación 8, en el que el grupo bloqueante se selecciona entre un grupo amino, un grupo fosfato, un resto glicerol, un grupo tiol y un resto polietilenglicol.
- 5 10. Procedimiento según cualquiera de las reivindicaciones precedentes, en el que la cadena no codificadora comprende un extremo 3' y un extremo 5', extendiéndose el extremo 5' de la cadena no codificadora uno o más nucleótidos más allá del extremo 3' de la cadena complementaria a la codificadora.
- 10 11. Procedimiento según cualquiera de las reivindicaciones precedentes, en el que se purifica el oligonucleótido iniciador de cadena doble aleatorizado obtenido en la etapa (e).
12. Procedimiento según cualquiera de las reivindicaciones precedentes que adicionalmente comprende las etapas de:
- 15 (g) Proporcionar un oligonucleótido iniciador de cadena doble aleatorizado producido mediante un procedimiento según cualquiera de las reivindicaciones precedentes;
- (h) Proporcionar un oligonucleótido predefinido que comprende:
- 20 (i) una cadena codificadora, comprendiendo la cadena codificadora un codón predefinido que codifica un aminoácido predefinido; y
- (ii) una cadena no codificadora sustancialmente complementaria,
- 25 donde el oligonucleótido predefinido comprende una secuencia de nucleótidos que codifica un sitio de reconocimiento de endonucleasa de restricción capaz de ser reconocido por una endonucleasa de restricción, siendo la endonucleasa de restricción capaz de escindir el oligonucleótido predefinido secuencia arriba o abajo del reconocimiento de la endonucleasa en un sitio de escisión predeterminado para crear un corte de extremos romos.
- (i) Ligar el oligonucleótido iniciador de cadena doble aleatorizado con el oligonucleótido predefinido para formar un oligonucleótido ligado;
- 30 (j) Amplificar el oligonucleótido ligado; y
- (k) Digerir el oligonucleótido ligado con la endonucleasa de restricción para formar un oligonucleótido aleatorizado que comprende en un extremo el codón predefinido.
- 35 13. Procedimiento según la reivindicación 12, que adicionalmente comprende las etapas de usar el oligonucleótido aleatorizado el codón predefinido como oligonucleótido iniciador y repetir las etapas (a) a (e) del procedimiento, y opcionalmente la etapa (f) como se define en la reivindicación 1, para añadir uno o más codones aleatorios adicionales a los oligonucleótidos.
- 40 14. Procedimiento según cualquiera de las reivindicaciones precedentes, que adicionalmente comprende la etapa de insertar un oligonucleótido obtenido de un procedimiento realizado como se define en cualquiera de las reivindicaciones precedentes en un vector de expresión.
- 45 15. Procedimiento según cualquiera de las reivindicaciones precedentes, en el que el vector de expresión se inserta en una célula huésped de expresión.
16. Procedimiento de producción de una biblioteca de péptidos aleatorizada que comprende la expresión de una biblioteca de oligonucleótidos obtenida por un procedimiento según cualquiera de las reivindicaciones precedentes.

Figura 1

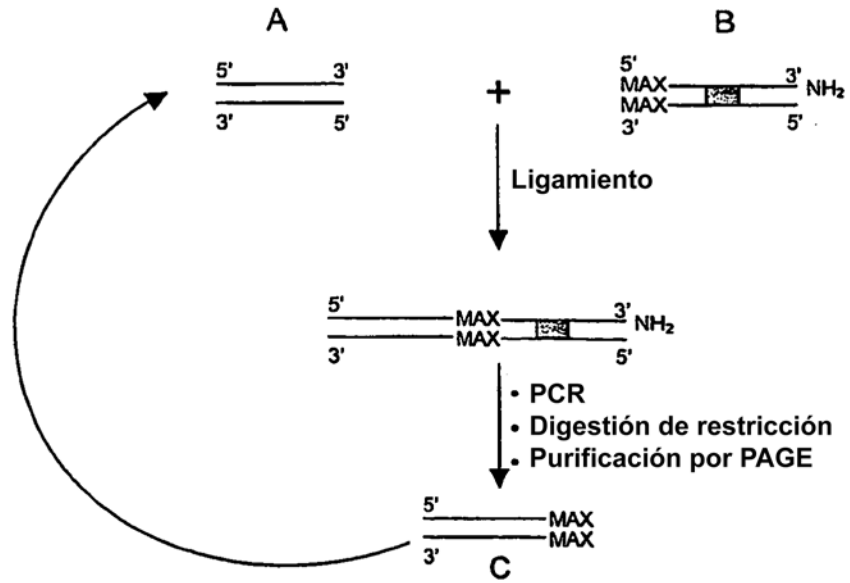


Figura 2

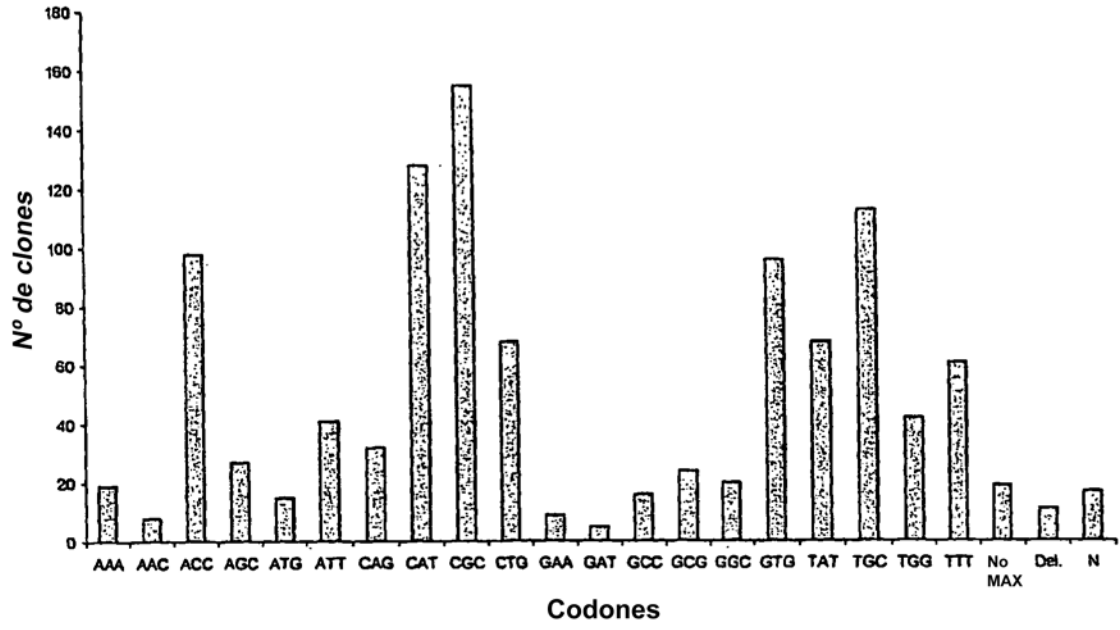


Figura 3

