



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 351 673**

51 Int. Cl.:
C12N 15/10 (2006.01)
C12Q 1/68 (2006.01)
C07K 14/56 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **01202350 .3**
96 Fecha de presentación : **17.12.1997**
97 Número de publicación de la solicitud: **1149905**
97 Fecha de publicación de la solicitud: **31.10.2001**

54 Título: **Métodos y composiciones para la ingeniería de polipéptidos.**

30 Prioridad: **18.12.1996 US 769062**

45 Fecha de publicación de la mención BOPI:
09.02.2011

45 Fecha de la publicación del folleto de la patente:
09.02.2011

73 Titular/es:
CODEXIS MAYFLOWER HOLDINGS, L.L.C.
515, Galveston Drive
Redwood City, California 94063, US

72 Inventor/es: **Patten, Phillip Andrew y**
Stemmer, Willem Peter Christiaan

74 Agente: **Ungría López, Javier**

ES 2 351 673 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

Antecedentes de la Invención

La recombinación repetitiva de secuencias supone realizar ciclos iterativos de recombinación y escrutinio o selección para "desarrollar" genes individuales, plásmidos completos o virus, agrupaciones de múltiples genes, o incluso genomas completos (Stemmer, Bio/Technology 13:549-553 (1995)). Tales técnicas no requieren el análisis y la computación exhaustivos requeridos por los métodos convencionales para la ingeniería de polipéptidos. La recombinación repetitiva de secuencias permite la recombinación de un gran número de mutaciones en un número mínimo de ciclos de selección, en contraste con los eventos de recombinación por pares, tradicionales.

De este modo, las técnicas de recombinación repetitiva de secuencias (RSR) proporcionan ventajas concretas ya que proporcionan la recombinación entre mutaciones en todas o cada una de éstas, proporcionado de así un modo muy rápido de explorar la manera en la cual diferentes combinaciones de mutaciones pueden afectar a un resultado deseado.

En algunos casos, no obstante, se encuentra disponible información estructural y/o funcional que, aunque no se requiere para la recombinación repetitiva de secuencias, proporciona oportunidades para la modificación de la técnica. En otros casos, la selección y/o el escrutinio de un gran número de recombinantes pueden ser costosos o llevar mucho tiempo. Un problema adicional puede ser la manipulación de moléculas grandes de ácido nucleico. La presente invención trata estas y otras cuestiones.

El documento WO 95/22625 describe métodos para desarrollar polinucleótidos o polipéptidos con una propiedad funcional deseada basándose en la fragmentación al azar de polinucleótidos molde, seguido de ciclos múltiples de desnaturalización, renaturalización e incubación en presencia de una polimerasa, y el escrutinio o selección para identificar un polinucleótido que tiene una propiedad funcional deseada.

Compendio de la Invención

La invención proporciona un método para desarrollar una proteína codificada por una molécula sustrato de ADN recombinando al menos una primera y una segunda moléculas sustrato de ADN donde las al menos primera y segunda moléculas sustrato difieren entre sí en al menos dos posiciones de nucleótidos y cada una comprende una pluralidad de segmentos definidos, comprendiendo el método:

1. (a) proporcionar un grupo de cebadores oligonucleotídicos puente, que comprende al menos un cebador para cada hebra de cada uno de una pluralidad de segmentos definidos, donde la secuencia del cebador es complementaria a al menos un empalme con otro segmento definido;
2. (b) amplificar los segmentos definidos de las al menos primera y segunda moléculas sustrato de ADN con los cebadores de la etapa (a) en una reacción en cadena de la polimerasa;
3. (c) ensamblar los productos de la etapa (b) para generar una genoteca diversa de moléculas sustrato de ADN recombinante;
4. (d) escrutar o seleccionar las proteínas codificadas por los productos de la etapa (c) en busca de una

propiedad deseada; y

5. (e) recuperar al menos una molécula sustrato de ADN recombinante de la etapa (d) que codifica una proteína desarrollada.

5

Otros rasgos de la invención se exponen en las reivindicaciones adjuntas.

10 Asimismo se describe en la presente memoria un método para desarrollar una proteína codificada por una molécula sustrato de ADN que comprende:

1. (a) digerir al menos una primera y una segunda moléculas sustrato de ADN, donde las al menos primera y segunda moléculas sustrato difieren entre sí en al menos un nucleótido, con una endonucleasa de restricción;
- 15 2. (b) ligar la mezcla para generar una genoteca de moléculas de ADN recombinante;
- 20 3. (c) escrutar o seleccionar los productos de la etapa (b) en busca de una propiedad deseada; y
4. (d) recuperar una molécula sustrato de ADN recombinante que codifica una proteína desarrollada.

25 Asimismo se describe en la presente memoria un método para desarrollar una proteína codificada por una molécula sustrato de ADN recombinando al menos una primera y una segunda moléculas sustrato de ADN, donde las al menos primera y segunda moléculas sustrato difieren entre sí en al menos un nucleótido y comprenden segmentos definidos, comprendiendo el método:

30

1. (a) proporcionar un grupo de cebadores oligonucleotídicos de PCR, que comprende al menos un cebador para cada segmento, donde la secuencia del
- 35

cebador es complementaria a al menos un empalme con otro segmento;

2. (b) amplificar los segmentos de las al menos primera y segunda moléculas sustrato de ADN con los
5 cebadores de la etapa (a) en una reacción en cadena de la polimerasa;

3. (c) ensamblar los productos de la etapa (b) para generar una genoteca de moléculas sustrato de ADN recombinantes;

10 4. (d) escrutar o seleccionar los productos de la etapa (c) en busca de una propiedad deseada; y

5. (e) recuperar una molécula sustrato de ADN recombinante de la etapa (d) que codifica una proteína desarrollada.

15

También se describe en la presente memoria un método de enriquecimiento de una población de fragmentos de ADN en secuencias mutantes que comprende:

20 1. (a) desnaturalizar y renaturalizar la población de fragmentos para generar una población de fragmentos de doble hebra híbridos en los que al menos un fragmento de doble hebra comprende al menos un emparejamiento erróneo de un par de bases;

25 2. (b) fragmentar los productos de la etapa (a) en fragmentos de aproximadamente 20-100 pb;

3. (c) purificar por afinidad los fragmentos que tienen un emparejamiento erróneo sobre una matriz de afinidad para generar una reserva de fragmentos de
30 ADN enriquecidos en secuencias mutantes; y

4. (d) ensamblar los productos de la etapa (c) para generar una genoteca de moléculas sustrato de ADN recombinante.

35

También se describe en la presente memoria un método

para desarrollar una proteína codificada por una molécula sustrato de ADN, recombinando al menos una primera y una segunda moléculas sustrato de ADN, donde las al menos primera y segunda moléculas sustrato comparten una región de homología de la secuencia de aproximadamente 10 a 100 pares de bases y comprenden segmentos definidos, comprendiendo el método:

1. (a) proporcionar regiones de homología en las al menos primera y segunda moléculas sustrato de ADN insertando una secuencia intrónica entre al menos dos segmentos definidos;
2. (b) fragmentar y recombinar las moléculas sustrato de ADN recombinante de la etapa (a), donde las regiones de homología están proporcionadas por los intrones;
3. (c) escrutar o seleccionar los productos de la etapa (b) en busca de una propiedad deseada; y
4. (d) recuperar una molécula sustrato de ADN recombinante a partir de los productos de la etapa (c) que codifican una proteína desarrollada.

También se describe en la presente memoria un método para desarrollar una proteína codificada por una molécula sustrato de ADN recombinando al menos una primera y una segunda moléculas sustrato de ADN, donde las al menos primera y segunda moléculas sustrato difieren entre sí en al menos un nucleótido y comprenden segmentos definidos, comprendiendo el método:

1. (a) proporcionar un grupo de cebadores de PCR oligonucleotídicos, donde para cada hebra de cada segmento se proporciona un par de cebadores, puenteando un miembro de cada par la unión en un extremo del segmento y puenteando el otro la unión

en el otro extremo del segmento, teniendo los extremos terminales de la molécula de ADN como miembro del par un cebador genérico, y donde se proporciona un grupo de cebadores para cada una de las al menos primera y segunda moléculas sustrato;

5

2. (b) amplificar los segmentos de las al menos primera y segunda moléculas sustrato de ADN con los cebadores de la etapa (a) en una reacción en cadena de la polimerasa;

10

3. (c) ensamblar los productos de la etapa (b) para generar una reserva de moléculas de ADN recombinantes;

4. (d) seleccionar o escrutar los productos de la etapa (c) en busca de una propiedad deseada; y

15

5. (e) recuperar una molécula sustrato de ADN recombinante de los productos de la etapa (d) que codifica una proteína desarrollada.

También se describe en la presente memoria un método para optimizar la expresión de una proteína desarrollando la proteína, donde la proteína está codificada por una molécula sustrato de ADN, que comprende:

20

1. (a) proporcionar un grupo de oligonucleótidos, donde cada oligonucleótido comprende al menos dos regiones complementarias a la molécula de ADN y al menos una región degenerada, codificando cada región degenerada una región de una secuencia de aminoácidos de la proteína;

25

2. (b) ensamblar el grupo de oligonucleótidos en una genoteca de genes completos;

30

3. (c) expresar los productos de la etapa (b) en una célula anfitriona;

4. (d) escrutar los productos de la etapa (c) en busca de una expresión mejorada de la proteína; y

35

5. (e) recuperar una molécula sustrato de ADN recombinante que codifica una proteína desarrollada de la etapa (d).

5 También se describe en la presente memoria un método para optimizar la expresión de una proteína codificada por una molécula sustrato de ADN desarrollando la proteína, donde la molécula sustrato de ADN comprende al menos un operador lac y una fusión de una secuencia de ADN que codifica la proteína con una secuencia de ADN que codifica un dímero de la porción de cabeza de lac, comprendiendo el método:

1. (a) transformar una célula anfitriona con una genoteca de moléculas sustrato de ADN mutagenizadas;
2. (b) inducir la expresión de la proteína codificada por la genoteca de la etapa (a);
3. (c) preparar un extracto del producto de la etapa (b);
4. (d) fraccionar la proteína insoluble de los complejos de proteína soluble y ADN; y
5. (e) recuperar una molécula sustrato de ADN que codifica una proteína desarrollada de la etapa (d).

25 También se describe en la presente memoria un método para desarrollar la expresión funcional de una proteína codificada por una molécula sustrato de ADN que comprende una fusión de una secuencia de ADN que codifica la proteína con una secuencia de ADN que codifica una proteína de fago filamentoso para generar una proteína de fusión, comprendiendo el método:

1. (a) proporcionar una célula anfitriona que produce partículas infecciosas que expresan una proteína de fusión codificada por una genoteca de moléculas

sustrato de ADN mutagenizadas;

2. (b) recuperar de la etapa (a) partículas infecciosas que presentan la proteína de fusión;
3. (c) purificar por afinidad las partículas que
5 presentan la proteína mutante utilizando un ligando para la proteína; y
4. (d) recuperar una molécula sustrato de ADN que codifica una proteína desarrollada de las partículas purificadas por afinidad de la etapa (c).

10

También se describe en la presente memoria un método para optimizar la expresión de una proteína codificada por una molécula sustrato de ADN que comprende una fusión de una secuencia de ADN que codifica la proteína con un
15 dímero de la porción de cabeza de lac, donde la molécula sustrato de ADN está presente en un primer vector plasmídico, comprendiendo el método:

1. (a) proporcionar una célula anfitriona transformada
20 con el primer vector y un segundo vector que comprende una genoteca de mutantes de al menos un gen de chaperonina, y al menos un operador lac;
2. (b) preparar un extracto del producto de la etapa (a);
- 25 3. (c) fraccionar la proteína insoluble de los complejos de proteína soluble y ADN; y
4. (d) recuperar el ADN que codifica un gen de chaperonina de la etapa (c).

30

También se describe en la presente memoria un método para optimizar la expresión de una proteína codificada por una molécula sustrato de ADN que comprende una fusión de una secuencia de ADN que codifica la proteína con un gen de un fago filamentoso, donde la fusión se lleva a
35 cabo en un fagémido que comprende una genoteca de

mutantes de genes de chaperonina, comprendiendo el método:

- 5 1. (a) proporcionar una célula anfitriona que produce partículas infecciosas que expresan una proteína de fusión codificada por una genoteca de moléculas sustrato de ADN mutagenizadas;
2. (b) recuperar de la etapa (a) las partículas infecciosas que presentan la proteína de fusión;
- 10 3. (c) purificar por afinidad las partículas que presentan la proteína utilizando un ligando para la proteína; y
4. (d) recuperar el ADN que codifica la chaperonina mutante de las partículas purificadas por afinidad
- 15 de la etapa (c).

También se describe en la presente memoria un método para optimizar la secreción de una proteína en un anfitrión desarrollando un gen que codifica una función

20 secretora, que comprende:

1. (a) proporcionar una agrupación de genes que codifican funciones secretoras;
- 25 2. (b) recombinar al menos una primera y una segunda secuencias en la agrupación de genes de la etapa (a) que codifican una función secretora, difiriendo las al menos primera y segunda secuencias entre sí en al menos un nucleótido, para generar, una genoteca de secuencias recombinantes;
- 30 3. (c) transformar un cultivo de células anfitrionas con los productos de la etapa (b), donde la célula anfitriona comprende una secuencia de ADN que codifica la proteína;
4. (d) someter el producto de la etapa (c) a
- 35 escrutinio o selección en busca de la secreción de

proteína; y

5. (e) recuperar el ADN que codifica un gen desarrollado que codifica una función secretora del producto de la etapa (d).

5

También se describe en la presente memoria un método para desarrollar una ADN polimerasa mejorada que comprende:

- 10 1. (a) proporcionar una genoteca de moléculas sustrato de ADN mutantes que codifican ADN polimerasa mutante;
2. (b) escrutar los extractos de células transfectadas con (a) y comparar la actividad con la ADN polimerasa de tipo salvaje;
- 15 3. (c) recuperar las moléculas sustrato de ADN mutantes de las células de la etapa (b) que expresan la ADN polimerasa mutante que tiene una actividad mejorada sobre la de la ADN polimerasa de tipo salvaje; y
- 20 4. (d) recuperar una molécula sustrato de ADN que codifica una polimerasa desarrollada de los productos de la etapa (c).

25 También se describe en la presente memoria un método para desarrollar una ADN polimerasa con una tasa de error mayor que la de la ADN polimerasa de tipo salvaje que comprende:

- 30 1. (a) proporcionar una genoteca de moléculas sustrato de ADN mutantes que codifican la ADN polimerasa mutante en una célula anfitriona que comprende un gen informador que tiene una mutación reversible, donde el gen informador es replicado por la ADN polimerasa mutante;
- 35 2. (b) escrutar los productos de la etapa (a) en busca

de revertientes del gen indicador;

3. (c) recuperar las moléculas sustrato de ADN mutantes de los revertientes; y
4. (d) recuperar una molécula sustrato de ADN que
5 codifica una polimerasa desarrollada de los productos de la etapa (c).

También se describe en la presente memoria un método para desarrollar una ADN polimerasa, que comprende:

10

1. (a) proporcionar una genoteca de moléculas sustrato de ADN mutantes que codifican una ADN polimerasa mutante, comprendiendo la genoteca un vector plasmídico;
- 15 2. (b) preparar preparaciones plasmídicas y extractos de células anfitrionas transfectadas con los productos de la etapa (a);
3. (c) amplificar cada preparación de plásmidos en una reacción de PCR utilizando la polimerasa mutante codificada por ese plásmido, estando presente la
20 polimerasa en el extracto de la célula anfitriona;
4. (d) recuperar los productos de la PCR de la etapa (c); y
5. (e) recuperar la molécula sustrato de ADN que
25 codifica una polimerasa desarrollada de los productos de la etapa (d).

También se describe en la presente memoria un método para desarrollar una p-nitrofenol fosfonatasa a partir de una fosfonatasa codificada por una molécula sustrato de ADN, que comprende:

30

1. (a) proporcionar una genoteca de mutantes de la molécula sustrato de ADN, comprendiendo la genoteca
35 un vector de expresión plasmídico;

2. (b) transfectar una célula anfitriona, donde el operón phn del anfitrión es suprimido;
3. (c) seleccionar el crecimiento de los transfectantes de la etapa (b) utilizando una p-nitrofenol fosfonatasa como sustrato;
4. (d) recuperar las moléculas sustrato de ADN de los transfectantes seleccionados de la etapa (c); y
5. (e) recuperar una molécula sustrato de ADN del apartado (d) que codifica una fosfonatasa desarrollada.

También se describe en la presente memoria un método para desarrollar una proteasa codificada por una molécula sustrato de ADN que comprende:

1. (a) proporcionar una genoteca de mutantes de la molécula sustrato de ADN, comprendiendo la genoteca un vector de expresión plasmídico, donde la molécula sustrato de ADN está conectada a un líder secretor;
2. (b) transfectar un anfitrión;
3. (c) seleccionar el crecimiento de los transfectantes de la etapa (b) en un medio de proteína complejo; y
4. (d) recuperar una molécula sustrato de ADN de la etapa (c) que codifica una proteasa desarrollada.

También se describe en la presente memoria un método para escrutar una genoteca de mutantes de proteasa presentados sobre un fago para obtener una proteasa mejorada, donde una molécula sustrato de ADN que codifica la proteasa está fusionada a un ADN que codifica una proteína de fago filamentoso para generar una proteína de fusión, que comprende:

1. (a) proporcionar células anfitrionas que expresan la proteína de fusión;

2. (b) recubrir las células anfitrionas con una red de proteína para atrapar el fago;
3. (c) lavar el producto de la etapa (b) para recuperar el fago liberado mediante digestión de la red de proteína;
4. (d) recuperar el ADN del producto de la etapa (c);
y
5. (e) recuperar un sustrato de ADN de la etapa (d) que codifica una proteasa mejorada.

También se describe en la presente memoria un método para escrutar una genoteca de mutantes de proteasa para obtener una proteasa mejorada, comprendiendo el método:

1. (a) proporcionar una genoteca de sustratos peptídicos, comprendiendo el sustrato peptídico un fluoróforo y un extintor de la fluorescencia;
2. (b) escrutar la genoteca de mutantes de proteasa en busca de la capacidad para escindir los sustratos peptídicos, donde se mide la fluorescencia; y
3. (c) recuperar el ADN que codifica al menos un mutante de proteasa de la etapa (b).

También se describe en la presente memoria un método para desarrollar un gen de interferón alfa que comprende:

1. (a) proporcionar una genoteca de genes de interferón alfa mutantes, comprendiendo la genoteca un vector de fago filamentoso;
2. (b) estimular células que comprenden un constructo informador que comprende un gen informador bajo el control de un promotor sensible al interferón, y donde el gen informador es GFP;
3. (c) separar las células que expresan GFP mediante FACS;

4. (d) recuperar el fago del producto de la etapa (c);
y
5. (e) recuperar un gen de interferón desarrollado del
producto de la etapa (d).

5

También se describe en la presente memoria un método para escrutar una genoteca de mutantes de un sustrato de ADN que codifica una proteína en busca de un sustrato de ADN desarrollado, que comprende:

10

1. (a) proporcionar una genoteca de mutantes,
comprendiendo la genoteca un vector de expresión;
2. (b) transfectar una célula anfitriona de mamífero
con la genoteca de la etapa (a), donde la proteína
mutante es expresada sobre la superficie de la
célula;
3. (c) escrutar o seleccionar los productos del
apartado (b) con un ligando en busca de la proteína;
4. (d) recuperar el ADN que codifica la proteína
mutante de los productos de la etapa (c); y
5. (e) recuperar un sustrato de ADN desarrollado de los
productos de la etapa (d).

15

20

25

También se describe en la presente memoria un método para desarrollar una molécula sustrato de ADN que codifica un interferón alfa, que comprende:

1. (a) proporcionar una genoteca de genes de interferón
alfa mutantes, comprendiendo la genoteca un vector
de expresión donde los genes de interferón alfa son
expresados bajo el control de un promotor inducible;
2. (b) transfectar células anfitrionas con la genoteca
de la etapa (a);
3. (c) poner en contacto el producto de la etapa (b)
con un virus;

30

35

4. (d) recuperar el ADN que codifica un interferón alfa mutante de las células anfitrionas supervivientes de la etapa (c); y
5. (e) recuperar un gen de interferón desarrollado a partir del producto de la etapa (d).

También se describe en la presente memoria un método para desarrollar la estabilidad en suero o la vida media en circulación de una proteína codificada por una molécula sustrato de ADN, comprendiendo la molécula sustrato de ADN una fusión de una secuencia de ADN que codifica la proteína con una secuencia de ADN que codifica una proteína de un fago filamentoso para generar una proteína de fusión, comprendiendo el método:

1. (a) proporcionar una célula anfitriona que expresa una genoteca de mutantes de la proteína de fusión;
2. (b) purificar por afinidad los mutantes con un ligando para la proteína, donde el ligando es una proteína de suero humana, una proteína específica de tejidos, o un receptor;
3. (c) recuperar el ADN que codifica una proteína mutante de los mutantes seleccionados por afinidad de la etapa (b); y
4. (d) recuperar un gen desarrollado que codifica la proteína del producto de la etapa (c).

También se describe en la presente memoria un método para desarrollar una proteína que tiene al menos dos subunidades, que comprende:

1. (a) proporcionar una genoteca de moléculas sustrato de ADN mutantes para cada subunidad;
2. (b) recombinar las genotecas en una genoteca de constructos de cadena sencilla de la proteína,

comprendiendo el constructo de cadena sencilla una molécula sustrato de ADN que codifica cada secuencia de la subunidad, estando conectada la secuencia de la subunidad por un conector a una secuencia de ácido nucleico que codifica el extremo amino de una subunidad a una secuencia de ácido nucleico que codifica el extremo carboxi de una segunda subunidad;

3. (c) escrutar o seleccionar los productos de la etapa (B);
4. (d) recuperar las moléculas sustrato de ADN del constructo de cadena sencilla recombinante de los productos de la etapa (c);
5. (e) someter los productos de la etapa (d) a mutagénesis; y
6. (f) recuperar una molécula sustrato de ADN del constructo de cadena sencilla desarrollado de la etapa (e).

También se describe en la presente memoria un método para desarrollar el acoplamiento de un receptor 7-transmembrana de mamífero a una ruta de transducción de la señal de levadura, que comprende:

1. (a) expresar una genoteca de mutantes de proteína alfa G de mamífero en una célula anfitriona, donde la célula anfitriona expresa el receptor 7-transmembrana de mamífero y un gen informador, estando expresado el gen del receptor bajo el control de un promotor sensible a feromona;
2. (b) escrutar o seleccionar los productos de la etapa (a) en busca de la expresión del gen informador en presencia de un ligando para el receptor 7-transmembrana; y
3. (c) recuperar el ADN que codifica un mutante de la

proteína alfa G desarrollada de los productos
escrutados o seleccionados de la etapa (b).

5 También se describe en la presente memoria un método
para recombinar al menos una primera y una segunda
moléculas sustrato de ADN, que comprende:

- 10 1. (a) transfectar una célula anfitriona con al menos
una primera y una segunda moléculas sustrato de ADN
donde las al menos primera y segunda moléculas
sustrato de ADN se recombinan en la célula
anfitriona;
2. (b) escrutar o seleccionar los productos de la etapa
(a) en busca de una propiedad deseada; y
- 15 3. (c) recuperar las moléculas sustrato de ADN
recombinantes de la etapa (b).

20 También se describe en la presente memoria un método
para desarrollar una secuencia sustrato de ADN que
codifica una proteína de interés, donde, el sustrato de
ADN comprende un vector, comprendiendo el vector ADN de
hebra sencilla, comprendiendo el método:

- 25 1. (a) proporcionar ADN vector de cadena sencilla y una
genoteca de mutantes de la secuencia sustrato de
ADN;
2. (b) recocer el ADN de hebra sencilla de la genoteca
de la etapa (a) con el ADN vector de hebra sencilla
de la etapa (a);
- 30 3. (c) transformar los productos de la etapa (b) en un
anfitrión;
4. (d) escrutar el producto de la etapa (c) en busca de
una propiedad deseada; y
- 35 5. (e) recuperar el ADN sustrato de ADN desarrollado de
los productos de la etapa (d).

Breve Descripción de los Dibujos

La Figura 1 representa el alineamiento de los cebadores de PCR oligo para el desarrollo de fosfatasa alcalina intestinal de ternera bovina.

La Figura 2 representa el alineamiento de secuencias de aminoácidos y de ácido nucleico de interferón alfa.

La Figura 3 representa el alineamiento de secuencias de aminoácidos de interferón alfa quimérico.

Descripción de las Realizaciones Específicas

La invención proporciona numerosas estrategias para desarrollar polipéptidos por medio de métodos de recombinación repetitivos. En algunas realizaciones, las estrategias de la invención se pueden clasificar generalmente como "barajado grosero" y "barajado fino". Como se describe con detalle más abajo, estas estrategias son especialmente aplicables en situaciones en las que se encuentra disponible cierta información estructural o funcional con respecto a los polipéptidos de interés, donde el ácido nucleico que se va a manipular es grande, cuando la selección o escrutinio de muchos recombinantes son engorrosos, etcétera. El "barajado grosero" implica generalmente el cambio o recombinación de segmentos de ácidos nucleicos, sin importar si se definen como dominios funcionales, exones, fragmentos de endonucleasas de restricción, o como otros segmentos definidos arbitrariamente. El "barajado fino" implica generalmente la introducción de variaciones de secuencia en un segmento, por ejemplo en los codones.

El barajado grosero y fino permite el análisis de la variación que existe en una secuencia de ácido nucleico, también denominada "búsqueda de espacios de secuencia".

Aunque ambas técnicas son meritorias, los resultados son cualitativamente diferentes. Por ejemplo, las búsquedas groseras a menudo están mejor adaptadas para optimizar agrupaciones de múltiples genes tales como operones de policétidos, mientras las búsquedas finas a menudo son óptimas para optimizar una propiedad tal como la expresión de la proteína utilizando genotecas de uso codónico.

Las estrategias generalmente implican el desarrollo de uno o varios genes o segmentos de los mismos para permitir la conservación de la función en una célula heteróloga o la mejora de la función en una célula homóloga o heteróloga. El desarrollo se realiza generalmente mediante un procedimiento denominado recombinación repetitiva de secuencias. La recombinación repetitiva de secuencias se puede lograr en muchos formatos diferentes y permutaciones de formatos, como se describe con mayor detalle más abajo. Estos formatos comparten algunos principios comunes. La recombinación repetitiva de secuencias implica ciclos sucesivos de recombinación para generar diversidad molecular, esto es, la creación de una familia de moléculas de ácido nucleico que muestran una identidad de secuencia sustancial entre sí pero que difieren en la presencia de mutaciones. Cada ciclo de recombinación está seguido de al menos un ciclo de escrutinio o selección en busca de moléculas que tienen una característica deseada. La molécula o las moléculas seleccionadas en una ronda forman los materiales de partida para generar diversidad en la siguiente ronda. En cualquier ciclo dado, la recombinación se puede producir *in vivo* o *in vitro*. Además, la diversidad resultante de la recombinación puede aumentar en cualquier ciclo aplicando métodos previos de mutagénesis (p. ej., PCR propensa a errores o

mutagénesis por inserción de casete, pase a través de cepas mutadoras bacterianas, tratamiento con mutágenos químicos, "carga" con una diversidad de secuencias de familias de genes homólogos) a cualquiera de los sustratos o a los productos de recombinación.

I. Formatos para la Recombinación Repetitiva de Secuencias

Algunos formatos y ejemplos para la recombinación repetitiva de secuencias, referida a veces como barajado de ADN, evolución, o cultivo molecular, han sido descritos por los autores de la presente invención y colaboradores en las solicitudes co-pendientes de la Solicitud de Patente de los Estados Unidos Núm. de Serie 08/198,431, presentada el 17 de Febrero de 1994, Núm. de Serie PCT/US95/02126, presentada el 17 de Febrero de 1995, Núm. de Serie 08/425,634, presentada el 18 de Abril de 1995, Núm. de Serie 08/537,874, presentada el 30 de Octubre de 1995, Núm. de Serie 08/564,955, presentada el 30 de Noviembre de 1995, Núm. de Serie 08/621,859, presentada el 25 de Marzo de 1996, Núm. de Serie 08/621,430, presentada el 25 de Marzo de 1996, Núm. de Serie PCT/US96/05480, presentada el 18 de Abril de 1996, Núm. de Serie 08/650,400, presentada el 20 de Mayo de 1996, Núm. de Serie 08/675.502, presentada el 3 de Julio de 1996, Núm. de Serie 08/721.824, presentada el 27 de Septiembre de 1996, y 08/722.660 presentada el 27 de Septiembre de 1996; Stemmer, Science 270:1510 (1995); Stemmer et al., Gene 164:49-53 (1995); Stemmer, Bio/Technology 13:549-553 (1995); Stemmer, Proc. Natl. Acad. Sci. U.S.A. 91:10747-10751 (1994); Stemmer, Nature 370:389-391 (1994); Cramer et al., Nature Medicine 2(1):1-3 (1996); Cramer et al., Nature Biotechnology 14:315-319 (1996).

En general, el término "gen" se utiliza en la presente memoria en sentido amplio para hacer referencia a cualquier segmento o secuencia de ADN asociado con una función biológica. Los genes se pueden obtener de una
5 variedad de fuentes, incluyendo la clonación a partir de una fuente de interés o la síntesis a partir de información de la secuencia conocida o pronosticada, y pueden incluir secuencias diseñadas para que tengan los parámetros deseados.

Se puede utilizar una amplia variedad de tipos celulares como receptor de los genes desarrollados. Las células de particular interés incluyen muchos tipos celulares bacterianos, tanto gram-negativos como gram-
15 positivos, tales como *Rhodococcus*, *Streptomyces*, *Actinomyces*, *Corynebacteria*, *Penicillium*, *Bacillus*, *Escherichia coli*, *Pseudomonas*, *Salmonella*, y *Erwinia*. Las células de interés también incluyen células eucarióticas, concretamente células de mamífero (p. ej., ratón, hámster, primate, humano), tanto líneas celulares como
20 cultivos primarios. Tales células incluyen células pluripotenciales, incluyendo células pluripotenciales embrionarias, cigotos, fibroblastos, linfocitos, ovario de hámster Chino (CHO), fibroblastos de ratón (NIH3T3),
25 células de riñón, hígado, músculo, y piel. Otras células eucarióticas de interés incluyen células vegetales, tales como maíz, arroz, trigo, algodón, soja, caña de azúcar, tabaco, y arabidopsis; peces, algas, hongos (*Penicillium*, *Fusarium*, *Aspergillus*, *Podospora*, *Neurospora*), insectos,
30 levaduras (*Picchia* y *Saccharomyces*).

La elección del anfitrión dependerá de numerosos factores, dependiendo del uso pretendido del anfitrión diseñado, incluyendo la patogenicidad, la gama de
35 sustrato, la dureza del entorno, la presencia de

intermedios clave, la facilidad de manipulación genética, y la probabilidad de transferencia promiscua de información genética a otros organismos. Un anfitrión preferido tiene la capacidad de replicar el ADN del vector, expresar las proteínas de interés, y transportar apropiadamente las proteínas de interés. Los anfitriones particularmente ventajosos son *E. coli*, *Lactobacilli*, *Streptomyces*, *Actinomyces*, hongos tales como *Saccharomyces cerevisiae* o *Pischia pastoris*, células de Schneider, células L, células COS, células CHO, y líneas de células B transformadas tales como SP2/0, J558, NS-1 y AG8-653.

El procedimiento de cultivo comienza con al menos dos sustratos que generalmente muestran una identidad de secuencia sustancial entre sí (esto es, una identidad de secuencia de al menos aproximadamente 50%, 70%, 80% o 90%), pero difieren entre sí en ciertas posiciones. La diferencia puede ser cualquier tipo de mutación, por ejemplo, sustituciones, inserciones y deleciones. A menudo, los diferentes segmentos difieren entre sí quizá en 5-20 posiciones. Para la recombinación para generar un aumento de la diversidad con respecto a las sustancias de partida, las sustancias de partida deben diferir entre sí en al menos dos posiciones de nucleótidos. Esto es, si solamente hay dos sustratos, debe haber al menos dos posiciones divergentes. Si hay tres sustratos, por ejemplo, un sustrato puede diferir del segundo en una única posición, y el segundo puede diferir del tercero en una única posición diferente. Los segmentos de ADN de partida pueden ser variantes naturales de cada uno de los otros, por ejemplo, variantes alélicas o de especie. Los segmentos también pueden ser de genes no alélicos que muestran cierto grado de relación estructural y normalmente funcional (p. ej., diferentes genes de una

superfamilia tal como la superfamilia de las inmunoglobulinas). Los segmentos de ADN de partida también pueden ser variantes inducidas de cada uno de los otros. Por ejemplo, se puede producir un segmento de ADN mediante replicación por PCR propensa a errores del otro, o mediante sustitución de un casete mutagénico. Los mutantes inducidos también pueden ser preparados propagando uno (o ambos) de los segmentos en una cepa mutagénica. En estas situaciones, en sentido estricto, el segundo segmento de ADN no es un único segmento si no una gran familia de segmentos relacionados. Los diferentes segmentos que forman las sustancias de partida tienen a menudo la misma longitud o esencialmente la misma longitud. Sin embargo, no es necesario que sea este caso. Por ejemplo; un segmento puede ser una subsecuencia de otro. Los segmentos pueden estar presentes como parte de moléculas más grandes, tales como vectores, o pueden estar en forma aislada.

Los segmentos de ADN de partida se recombinan mediante cualquiera de los formatos de recombinación repetitiva de secuencias proporcionados en la presente memoria para generar una genoteca diversa de segmentos de ADN recombinante. Semejante genoteca puede variar ampliamente en tamaño teniendo desde menos de 10 a más de 10^5 , 10^9 , o 10^{12} miembros. En general, los segmentos de partida y las genotecas recombinantes generadas incluyen secuencias codificantes completas y cualquiera de las secuencias reguladoras esenciales, tales como un promotor y una secuencia de poliadenilación, requeridas para la expresión. Sin embargo, si éste no es el caso, los segmentos de ADN recombinante de la genoteca pueden ser insertados en un vector común proporcionando las secuencias ausentes antes de realizar el escrutinio/selección.

Si el formato de recombinación repetitiva de secuencias empleado es un formato *in vivo*, la genoteca de segmentos de ADN recombinante ya existe en una célula, que es normalmente el tipo de célula en el que se desea la expresión de la enzima con una especificidad de sustrato alterada.

Si la recombinación repetitiva de secuencias se realiza *in vitro*, la genoteca recombinante se introduce preferiblemente en el tipo de célula deseado antes del escrutinio/selección. Los miembros de la genoteca recombinante pueden ser conectados a un episoma o virus antes de la introducción o pueden ser introducido directamente. En algunas realizaciones de la invención, la genoteca se amplifica en un primer anfitrión, y después se recupera de ese anfitrión y se introduce en un segundo anfitrión más susceptible de expresión, selección, o escrutinio, o cualquier otro parámetro deseable. La manera en la cual la genoteca es introducida en el tipo de célula depende de las características de absorción de ADN del tipo de célula, p. ej., de que tenga receptores virales, de que sea susceptible de conjugación, o de que sea naturalmente competente. Si el tipo de célula no es susceptible a la competencia natural o inducida por agentes químicos, pero es susceptible a la electroporación, se podría utilizar normalmente la electroporación. Si el tipo de célula tampoco es susceptible a la electroporación, se puede utilizar la biolística. La Pistola Génica PDS-1000 biolística (Biorad, Hercules, CA) utiliza una presión de helio para acelerar microportadores de oro o tungsteno recubiertos de ADN hacia las células diana. El procedimiento es aplicable a una amplia gama de tejidos, incluyendo plantas, bacterias, hongos, algas, tejidos animales intactos, células de cultivos de tejidos, y embriones

animales. Se puede emplear la liberación de pulsos electrónicos, que es esencialmente un formato de electroporación suave para tejidos vivos en animales y pacientes. Zhao, Advanced Drug Delivery Reviews 17:257-262 (1995). Se describen métodos novedosos para elaborar células competentes en la solicitud de Patente de los Estados Unidos co-pendiente Núm. de Serie 08/621.430, presentada el 25 de Marzo de 1996. Tras la introducción de la genoteca de genes de ADN recombinante, las células se propagan opcionalmente para permitir que se produzca la expresión de los genes.

A. Formatos In Vitro

Un formato para la recombinación repetitiva de secuencias utiliza una reserva de secuencias relacionadas. Las secuencias pueden ser ADN o ARN y pueden tener diferentes longitudes dependiendo del tamaño del gen o el fragmento de ADN que se va a recombinar o re-ensamblar. Preferiblemente las secuencias tienen de 50 pb a 100 kb.

La reserva de sustratos relacionados se puede fragmentar, normalmente al azar, en fragmentos de aproximadamente 5 pb a 5 kb o más. Preferiblemente el tamaño de los fragmentos al azar es de aproximadamente 10 pb a 1000 pb, más preferiblemente el tamaño de los fragmentos de ADN es de aproximadamente 20 pb a 500 pb. Los sustratos se pueden digerir mediante numerosos métodos diferentes, tal como digestión con ADNasa1 o ARNasa, corte al azar o digestión con enzimas de restricción. La concentración de fragmentos de ácido nucleico de una longitud concreta es a menudo menor del 0,1% o 1% en peso del ácido nucleico total. El número de fragmentos de ácido nucleico específicos diferentes de la

mezcla es normalmente de al menos aproximadamente 100, 500 o 1000.

La población mixta de fragmentos de ácido nucleico se desnaturaliza calentando de aproximadamente 80°C a 100°C, más preferiblemente de 90°C a 96°C, para formar fragmentos de ácido nucleico de hebra sencilla. Los fragmentos de ácido nucleico de hebra sencilla que tienen regiones de identidad de secuencia con otros fragmentos de ácido nucleico de hebra sencilla pueden ser recocidos después enfriando de 6°C a 75°C, y preferiblemente de 40°C a 65°C. La renaturalización se puede acelerar mediante la adición de polietilenglicol ("PEG") o sal. La concentración de sal es preferiblemente de 0 mM a 600 mM, más preferiblemente de 10 mM a 100 mM. La sal puede ser una sal tal como $(\text{NH}_4)_2\text{SO}_4$, KCl, o NaCl. La concentración de PEG es preferiblemente de 0% a 20%, más preferiblemente de 5% a 10%. Los fragmentos que se recuecen pueden ser de diferentes sustratos.

Los fragmentos de ácido nucleico recocidos se incuban en presencia de una polimerasa de ácido nucleico, tal como Taq o Klenow, Mg^{++} 1 mM - 20 mM, y dNTP (esto es dATP, dCTP, dGTP y dTTP). Si las regiones de identidad de secuencia son grandes, se puede utilizar Taq u otra polimerasa a elevada temperatura con una temperatura de recocido entre 45-65°C. Si las zonas de identidad son pequeñas, se pueden utilizar Klenow u otras polimerasas que son activas a una baja temperatura, con una temperatura de recocido entre 6-30°C. La polimerasa puede ser añadida a los fragmentos de ácido nucleico al azar antes del recocido, simultáneamente al recocido o después del recocido.

El ciclo de desnaturalización, renaturalización e

incubación de fragmentos de ácido nucleico al azar en presencia de polimerasa es referido a veces como "barajado" del ácido nucleico *in vitro*. Este ciclo se repite durante el número deseado de veces.

5 Preferiblemente el ciclo se repite de 2 a 100 veces, más preferiblemente la secuencia se repite de 10 a 40 veces. Los ácidos nucleicos resultantes son una familia de polinucleótidos de doble hebra de aproximadamente 50 pb a aproximadamente 100 kb, preferiblemente de 500 pb a 50

10 kb. La población representa variantes de los sustratos de partida que muestran una identidad de secuencia sustancial con ellos pero que también divergen en varias posiciones. La población tiene mucho más miembros que los sustratos de partida. La población de fragmentos

15 resultantes de la recombinación es amplificada primero preferiblemente mediante PCR, después clonada en un vector apropiado y la mezcla de ligación es utilizada para transformar células anfitrionas.

20 En una variación del barajado *in vitro*, se pueden generar subsecuencias de sustratos de recombinación amplificando las secuencias completas en condiciones que producen una fracción sustancial, típicamente al menos 20 por ciento o más, de productos de amplificación ampliados

25 incompletamente. Los productos de amplificación, incluyendo los productos de amplificación ampliados incompletamente son desnaturalizados y sometidos al menos a un ciclo adicional de re-recocido y amplificación. Esta variación, donde al menos un ciclo de re-recocido y

30 amplificación proporciona una fracción sustancial de productos ampliados incompletamente, se denomina "balbuceo". En la siguiente ronda de amplificación, los productos ampliados incompletamente se recuecen y ceban la ampliación de diferentes especies de moldes con

35 secuencia relacionada.

En una variación adicional, se puede llevar a cabo al menos un ciclo de amplificación utilizando una colección de fragmentos de ADN de hebra sencilla solapantes de secuencia relacionada, y diferentes longitudes. Cada fragmento puede hibridar y cebar la prolongación de la cadena de polinucleótidos de un segundo fragmento de la colección, formando de este modo polinucleótidos de secuencia recombinada. En una variación adicional, se pueden generar fragmentos de ADN de hebra sencilla de longitud variable a partir de un único cebador mediante ADN polimerasa Vent sobre un primer molde de ADN. Los fragmentos de ADN de hebra sencilla se utilizan como cebadores para un segundo molde de tipo Kunkel, que consiste en un ADN de hebra sencilla circular que contiene uracilo. Esto da como resultado múltiples sustituciones del primer molde en el segundo (véase Levichkin et al., Mol. Biology 29:572-577 (1995)).

Las secuencias de ácido nucleico pueden ser recombinadas mediante recombinación repetitiva de secuencias incluso si carecen de homología de secuencia. Se puede introducir homología utilizando oligonucleótidos sintéticos como cebadores de PCR. Además de las secuencias específicas para el segmento de ácido nucleico que está siendo amplificado, todos los cebadores utilizados para amplificar un segmento concreto se sintetizan para que contengan una secuencia adicional de 20-40 bases 5' con respecto al gen (secuencia A) y una secuencia de 20-40 bases diferente 3' con respecto al segmento (secuencia B). Se amplifica un segmento adyacente utilizando un cebador 5' que contiene la hebra complementaria de la secuencia B (secuencia B'), y un cebador 3' que contiene una secuencia de 20-40 bases diferente (C). De un modo similar, los cebadores para el siguiente segmento adyacente contienen secuencias C'

(complementarias a C) y D. De este modo, se introducen pequeñas regiones de homología, elaborando los segmentos en casetes de recombinación específicos del sitio. Después de la amplificación inicial de los segmentos individuales, se pueden mezclar los segmentos amplificados y someterlos a PCR sin cebadores.

Cuando los dominios de un polipéptido son barajados, puede no ser posible introducir secuencias colindantes adicionales en los dominios, debido a la restricción de mantener un marco de lectura abierto continuo. En lugar de eso, se sintetizan grupos de oligonucleótidos que son homólogos al extremo 3' del primer dominio codificado por uno de los genes a barajar, y los extremos 5' de los segundos dominios codificados por todos los demás genes que se van a barajar juntos. Esto se repite con todos los dominios, proporcionando de este modo secuencias que permiten la recombinación entre los dominios de proteína a la vez que mantienen su orden.

B. Formatos In Vivo

1. Recombinación Plásmido-Plásmido

Los sustratos iniciales para la recombinación son una colección de polinucleótidos que comprende formas variantes de un gen. Las formas variantes muestran normalmente una identidad de secuencia sustancial entre sí suficiente para permitir la recombinación homóloga entre sustratos. La diversidad entre los polinucleótidos puede ser natural (p. ej., variantes alélicas o de especie), inducida (p. ej., PCR propensa a errores o recombinación repetitiva de secuencias propensa a errores), o el resultado de una recombinación *in vitro*. La diversidad también puede ser resultado de volver a

sintetizar genes que codifican proteínas naturales con uso de codones alternativo. Debe haber al menos una diversidad suficiente entre sustratos para que la recombinación pueda generar productos más diversos que en las sustancias de partida. Debe haber al menos dos sustratos que difieran en al menos dos posiciones. Sin embargo, comúnmente se emplea una genoteca de sustratos de 10^3 - 10^8 miembros. El grado de diversidad depende de la longitud del sustrato que se está recombinando y del grado de cambio funcional que se vaya a desarrollar. Es típica una diversidad entre 0,1-25% de las posiciones. Los diversos sustratos se incorporan a los plásmidos. Los plásmidos son a menudo vectores de clonación convencionales, p. ej., plásmidos de múltiples copias bacterianos. No obstante, en algunos métodos que se van a describir más abajo, los plásmidos incluyen funciones de movilización (MOB). Los sustratos se pueden incorporar al mismo plásmido o a uno diferente. A menudo se utilizan al menos dos tipos diferentes de plásmidos que tienen diferentes tipos de marcadores seleccionables para permitir la selección en busca de células que contienen al menos dos tipos de vector. Asimismo, cuando se emplean diferentes tipos de plásmidos, los diferentes plásmidos pueden proceder de dos grupos de incompatibilidad distinta para permitir una coexistencia estable de dos plásmidos diferentes en la célula. No obstante, todavía pueden co-existir en la misma célula plásmidos del mismo grupo de incompatibilidad durante un tiempo suficiente para permitir que se produzca una recombinación homóloga.

Los plásmidos que contienen diversos sustratos son introducidos inicialmente en las células mediante cualquier método (p. ej., transformación química, competencia natural, electroporación, biolística, empaquetamiento en sistemas de fagos o virales). A

menudo, los plásmidos están presentes a una concentración saturante o próxima a ella (con respecto a la capacidad de transfección máxima) para incrementar la probabilidad de que entre más de un plásmido en la misma célula. Los plásmidos que contienen los diferentes sustratos pueden ser transfectados simultáneamente o en múltiples rondas. Por ejemplo, en el último enfoque las células pueden ser transfectadas con una primera alícuota de plásmido, los transfectantes seleccionados y propagados, y después infectados con una segunda alícuota de plásmido.

Habiendo introducido los plásmido en células, la recombinación entre sustratos para generar genes recombinantes se produce en las células que contienen múltiples plásmidos diferentes simplemente por propagación de las células. Sin embargo, las células que reciben solamente un plásmido son incapaces de participar en la recombinación y la contribución potencial de los sustratos de tales plásmidos al desarrollo no está totalmente explotada (aunque estos plásmidos pueden contribuir hasta cierto punto si se propagan en células mutadoras). La tasa de evolución se puede incrementar permitiendo que todos los sustratos participen en la recombinación. Esto se puede lograr sometiendo las células transfectadas a electroporación. Las condiciones para la electroporación son las mismas que las utilizadas convencionalmente para introducir ADN exógeno en células (p. ej., 1.000-2.500 voltios, 400 μ F y un espacio de 1-2 mM). En estas condiciones, se cambian los plásmidos entre las células permitiendo que todos los sustratos participen en la recombinación. Además los productos de la recombinación pueden experimentar rondas adicionales de recombinación entre sí o con el sustrato original. La tasa de evolución también puede ser incrementada mediante el uso de transferencias conjugativas. Para explotar la

transferencia conjugativa, se pueden clonar sustratos en plásmidos que tienen genes MOB, y también se proporcionan genes *tra* en posición *cis* o *trans* con respecto a los genes MOB. El efecto de la transferencia conjugativa es muy similar a la electroporación ya que permite que los plásmidos se muevan entre las células y permite que se produzca la recombinación entre cualquier sustrato y los productos de la recombinación previa, simplemente propagando el cultivo. La tasa de evolución también se puede incrementar fusionando las células para inducir el intercambio de plásmidos o cromosomas. La fusión puede ser inducida por agentes químicos, tales como PEG, o proteínas virales, tales como hemaglutinina de virus de influenza, gB y gD de HSV-1. La tasa de evolución también puede ser incrementada mediante el uso de células anfitrionas mutadoras (p. ej., Mut L, S, D, T, H en bacterias y líneas celulares humanas de Ataxia-telangiectasia).

El tiempo durante el cual las células se propagan y se permite que se produzca la recombinación, por supuesto, varía con el tipo de célula pero generalmente no es crítico, porque incluso un pequeño grado de recombinación puede incrementar sustancialmente la diversidad relativa con respecto a las sustancias de partida. Las células que portan plásmidos que contienen genes recombinados se someten a escrutinio o selección en busca de la función deseada. Por ejemplo, si el sustrato que está siendo desarrollado contiene un gen de resistencia a un fármaco, se podría seleccionar en busca de la resistencia al fármaco. Las células que sobreviven al escrutinio o selección pueden ser sometidas a una o más rondas de escrutinio/selección seguido de recombinación o pueden ser sometidas directamente a una ronda adicional de recombinación. Se pretende que

"escrutinio" según se utiliza en la presente memoria incluya "selección" como un tipo de escrutinio.

La siguiente ronda de recombinación se puede
5 realizar por medio de varios formatos diferentes
independientemente de la ronda previa. Por ejemplo, se
puede efectuar una ronda adicional de recombinación
simplemente reanudando la electroporación o la
transferencia celular mediada por conjugación de
10 plásmidos descrita antes. Alternativamente, se pueden
transfectar uno o varios sustratos de nueva aportación,
iguales o diferentes de los sustratos previos, a células
que sobreviven a la selección/escrutinio. Opcionalmente,
los nuevos sustratos están incluidos en vectores
15 plasmídicos que portan un marcador selectivo diferente
y/o de un grupo de incompatibilidad diferente que los
plásmidos originales. Como alternativa adicional, las
células que sobreviven a la selección/escrutinio se
pueden subdividir en dos subpoblaciones, y el ADN
20 plasmídico de una subpoblación se puede transfectar en la
otra, donde los sustratos de los plásmidos de las dos
subpoblaciones experimentan una ronda adicional de
recombinación. En cualquiera de las dos últimas opciones,
la tasa de evolución se puede incrementar empleando
25 extracción de ADN, electroporación, conjugación o células
mutadoras, como se ha descrito antes. En otra variación
adicional más, se puede extraer el ADN de las células que
sobreviven al escrutinio/selección y someterlo a
recombinación repetitiva de secuencias *in vitro*.

30 Después de la segunda ronda de recombinación, se
realiza una segunda ronda de escrutinio/selección,
preferiblemente en condiciones de mayor restricción. Si
se desea, se pueden realizar rondas adicionales de
35 recombinación y selección/escrutinio utilizando la misma

estrategia que para la segunda ronda. Con las sucesivas rondas de recombinación y selección/escrutinio, los sustratos recombinados supervivientes evolucionan hacia la adquisición de un fenotipo deseado. Típicamente, en este y otros métodos de recombinación repetitiva, el producto final de la recombinación que ha adquirido el fenotipo deseado difiere de los sustratos de partida en 0,1%-25% de las posiciones y ha evolucionado a órdenes de magnitud de velocidad por encima (p. ej., de al menos 10 veces, 100 veces, 1000 veces, o 10.000 veces) de la velocidad de evolución llevada por la mutación adquirida naturalmente de aproximadamente 1 mutación por 10^{-9} posiciones por generación (véase Anderson et al., Proc. Natl. Acad. Sci. U.S.A. 93:906-907 (1996)). El "producto final" puede ser transferido a otro anfitrión más deseable para la utilización del ADN "barajado". Esto resulta particularmente ventajoso en situaciones en las que el anfitrión más deseable es menos eficiente como anfitrión para los muchos ciclos de mutación/recombinación debido a la carencia de herramientas de la biología molecular o la genética disponibles para otros microorganismos tales como *E. coli*.

2. Recombinación Virus-Plásmido

La estrategia utilizada para la recombinación plásmido-plásmido también puede ser utilizada para la recombinación virus-plásmido; normalmente, la recombinación fago-plásmido. No obstante, algunas observaciones concretas para el uso de virus son apropiadas. Los sustratos iniciales para la recombinación son clonados en vectores tanto plasmídicos como virales. Normalmente no es crítico cuáles sustratos están insertados en el vector viral y cuáles en el plásmido,

aunque normalmente el vector viral debe contener uno o varios sustratos diferentes del plásmido. Como antes, el plásmido (y el virus) contiene típicamente un marcador selectivo. Los vectores plasmídicos y virales pueden ser introducidos ambos en las células mediante transfección como se ha descrito antes. No obstante, un procedimiento más eficaz consiste en transfectar las células con un plásmido, seleccionar los transfectantes e infectar los transfectantes con virus. Debido a que la eficacia de infección de muchos virus se aproxima al 100% de las células, la mayor parte de las células transfectadas e infectadas por esta ruta contienen tanto un plásmido como un virus que portan diferentes sustratos.

La recombinación homóloga se produce entre plásmidos y virus que generan tanto plásmidos recombinados como virus recombinados. Para algunos virus, tales como los fagos filamentosos, en los cuales existe ADN intracelular en forma tanto de doble hebra como de hebra sencilla, ambos pueden participar en la recombinación. Siempre que el virus no sea uno que destruya rápidamente las células, se puede aumentar la recombinación mediante el uso de electroporación o conjugación para transferir plásmidos entre células. También se puede aumentar la recombinación para algunos tipos de virus permitiendo que los virus de la progenie de una célula re-infecten otras células. Para algunos tipos de virus, las células infectadas con virus muestran resistencia a la superinfección. No obstante, semejante resistencia puede ser superada infectando a una multiplicidad elevada y/o utilizando cepas mutantes del virus en las cuales la resistencia a la superinfección está reducida.

El resultado de infectar células que contienen plásmidos con virus depende de la naturaleza del virus.

Algunos virus, tales como los fagos filamentosos, existen establemente con un plásmido en la célula y también extruden los fagos progenie de la célula. Otros virus, tales como lambda que tiene un genoma cosmídico, pueden existir establemente en una célula como plásmidos sin producir viriones progenie. Otros virus, tales como el fago T y lambda lítico, experimentan recombinación con el plásmido pero por último destruyen la célula anfitriona y acaban con el ADN plasmídico. Para los virus que infectan las células sin destruir el anfitrión, las células que contienen plásmidos y virus recombinantes pueden ser escrutadas/seleccionadas utilizando el mismo enfoque que para la recombinación plásmido-plásmido. Los virus progenie extrudidos por las células que sobreviven a la selección/escrutinio también pueden ser recogidos y utilizados en las rondas subsiguientes de recombinación. Para los virus que destruyen sus células anfitrionas, los genes recombinantes que resultan de la recombinación residen solamente en el virus progenie. Si el escrutinio o el análisis selectivo requiere la expresión de genes recombinantes en una célula, los genes recombinantes deben ser transferidos del virus progenie a otro vector, p. ej., un vector plasmídico, y retransfectados en células antes de realizar la selección/escrutinio.

Para los fagos filamentosos, los productos de la recombinación están presentes tanto en las células que sobreviven a la recombinación como en fagos extrudidos de estas células. La fuente dual de productos recombinantes proporciona algunas opciones adicionales con respecto a la recombinación plásmido-plásmido. Por ejemplo, se puede aislar ADN de partículas de fago para su uso en una ronda de recombinación *in vitro*. Alternativamente, el fago progenie puede ser utilizado para transfectar o infectar células que sobreviven a una ronda previa de

escrutinio/selección, o células de nueva aportación transfectadas con sustratos de nueva aportación para su recombinación.

5 **3. Recombinación Virus-Virus**

Los principios descritos para la recombinación plásmido-plásmido y plásmido-virus pueden ser aplicados a la recombinación virus-virus con unas pocas modificaciones. Los sustratos iniciales para la recombinación se clonan en un vector viral. Normalmente, se utiliza el mismo vector para todos los sustratos. Preferiblemente, el virus es uno que, naturalmente o como resultado de una mutación, no destruye las células. Tras la inserción, algunos genomas virales pueden ser empaquetados *in vitro* o utilizando una línea celular de empaquetamiento. Los virus empaquetados se utilizan para infectar células con una levada multiplicidad de manera que haya una elevada probabilidad de que una célula reciba múltiples virus que porten sustratos diferentes.

Después de la ronda de infección inicial, las siguientes etapas dependen de la naturaleza de la infección como se comenta en la sección anterior. Por ejemplo, si los virus tienen fagémidos (Sambrook et al., Molecular Cloning, CSH Press, 1987) genomas tales como cósmidos lambda o fagémidos M13, F1 o Fd, los fagémidos se comportan como plásmidos en la célula y experimentan recombinación simplemente mediante propagación en las células. La recombinación es particularmente eficaz entre las formas de hebra sencilla del ADN intracelular. La recombinación se puede aumentar mediante electroporación de las células.

35 Después de la selección/escrutinio, los cósmidos que

contienen genes recombinantes pueden ser recuperados de las células supervivientes, p. ej., mediante inducción por calor de una célula anfitriona lisogénica cos, o extracción de ADN mediante procedimientos convencionales,
5 seguido de re-empaquetamiento de ADN cosmídico *in vitro*.

Si los virus son fagos filamentosos, la recombinación del ADN en forma replicante se produce propagando el cultivo de células infectadas. La
10 selección/escrutinio identifica colonias de células que contienen vectores virales que tienen genes recombinantes con propiedades mejoradas, junto con partículas infecciosas (esto es, fagos o fagémidos empaquetados) extrudidas de tales células. Las siguientes opciones son
15 esencialmente las mismas que para la recombinación plásmido-virus.

4. Recombinación Cromosómica

Este formato se puede utilizar para desarrollar especialmente sustratos cromosómicos. El formato es particularmente preferido en situaciones en las que muchos genes cromosómicos contribuyen a un fenotipo o no se sabe la localización exacta del gen o de los genes
20 cromosómicos que van ser desarrollados. Los sustratos iniciales para la recombinación son clonados en un vector plasmídico. Si el gen o los genes cromosómicos que se van a desarrollar son conocidos, los sustratos constituyen una familia de secuencias que muestran un alto grado de
25 identidad de secuencia pero cierta divergencia del gen cromosómico. Si los genes cromosómicos que se van a desarrollar no han sido localizados, los sustratos iniciales constituyen normalmente una genoteca de segmentos de ADN de los cuales solamente un pequeño
30 número muestra identidad de secuencia con el gen o los
35

genes que se van a desarrollar. La divergencia entre el sustrato contenido en el plásmido y el gen o los genes cromosómicos puede ser inducida por mutagénesis o mediante la obtención de sustratos contenidos en plásmidos de una especie diferente a la de las células que portan el cromosoma.

Los sustratos contenidos en los plásmidos para la recombinación son transfectados en células que tienen uno o varios genes cromosómicos que se van a desarrollar. La evolución se puede producir simplemente propagando el cultivo, y se puede acelerar transfiriendo los plásmidos entre células mediante conjugación o electroporación. La evolución se puede acelerar más mediante el uso de células anfitrionas mutadoras o sembrando un cultivo de células anfitrionas no mutadoras que están desarrolladas con células anfitrionas mutadoras e induciendo la transferencia intercelular de plásmidos mediante electroporación o conjugación. Preferiblemente, las células anfitrionas mutadoras utilizadas contienen un marcador seleccionable negativo para facilitar el aislamiento de un cultivo puro de las células no mutadoras que están siendo desarrolladas. La selección/escrutinio identifica células que contienen cromosomas y/o plásmidos que han evolucionado hacia la adquisición de una función deseada.

Después las rondas de recombinación y selección/escrutinio prosiguieron de una manera similar a la descrita para la recombinación plásmido-plásmido. Por ejemplo, la recombinación adicional se puede llevar a cabo mediante propagación de las células que sobreviven a la recombinación combinada con electroporación o transferencia conjugativa de plásmidos. Alternativamente, los plásmidos que contienen sustratos adicionales para la

recombinación se pueden introducir en las células supervivientes. Preferiblemente, tales plásmidos son de un grupo de incompatibilidad diferente y contienen un marcador selectivo diferente del de los plásmidos originales para permitir la selección de células que contienen al menos dos plásmidos diferentes. Como alternativa adicional, se puede aislar ADN plasmídico y/o cromosómico de una población de células supervivientes y transfectarlo a una segunda subpoblación. Se puede aislar ADN cromosómico en un vector plasmídico antes de la transfección.

5. Recombinación Virus-Cromosoma

Como en otros métodos descritos más arriba, el virus es normalmente uno que no destruya las células, y es a menudo un fago o un fagémido. El procedimiento es esencialmente el mismo que para la recombinación plásmido-cromosoma. Los sustratos para la recombinación se clonan en el vector. Después se pueden transfectar los vectores que incluyen los sustratos a células o se empaquetan *in vitro* e introducen en células mediante infección. Los genomas virales se recombinan con los cromosomas del anfitrión simplemente propagando un cultivo. La evolución se puede acelerar permitiendo la transferencia intercelular de los genomas virales mediante electroporación, o reinfección de las células por viriones progenie. El escrutinio/selección identifica las células que tienen cromosomas y/o genomas virales que han evolucionado hacia la adquisición de una función deseada.

Existen numerosas opciones para las siguientes rondas de recombinación. Por ejemplo, se pueden transfectar los genomas virales entre células que

sobreviven a la selección/recombinación mediante electroporación. Alternativamente, los virus extrudidos de las células que sobreviven a la selección/escrutinio pueden ser reunidos y utilizados para superinfectar las células a una multiplicidad elevada. Alternativamente, se pueden introducir sustratos de nueva aportación para la recombinación en las células, ya sea en vectores plasmídicos o virales.

II. Aplicación de la Recombinación Repetitiva de Secuencias al Desarrollo de Polipéptidos

Además de las técnicas descritas más arriba, se describen más abajo algunas modificaciones adicionalmente ventajosas de estas técnicas para el desarrollo de polipéptidos. Estos métodos son referidos como barajado "fino" y "grosero". Los métodos groseros permiten intercambiar pedazos de material genético entre ácidos nucleicos sustrato, limitando de ese modo la diversidad en los recombinantes resultantes a intercambios o sustituciones de dominios, fragmentos de restricción, bloques de mutaciones codificadas por oligos, u otros segmentos definidos arbitrariamente, en lugar de introducir una diversidad más al azar en el sustrato. En contraste con el barajado grosero, el barajado fino permite la generación de todas las posibles recombinaciones, o permutaciones, de un grupo dado de mutaciones muy íntimamente relacionadas, incluyendo múltiples permutaciones, en un único segmento, tal como un codón.

En algunas realizaciones, las técnicas de barajado grosero o fino no se realizan como búsquedas exhaustivas de todas las posibles mutaciones en una secuencia de ácido nucleico. En lugar de eso, estas técnicas se

utilizan para proporcionar un muestreo de las variaciones posibles en un gen basado en la información de la secuencia o estructural conocida. El tamaño de la muestra se determina típicamente mediante la naturaleza del procedimiento de escrutinio o selección. Por ejemplo, cuando se realiza un escrutinio en un formato de microtitulación de 96 pocillos, puede ser preferible limitar el tamaño de la genoteca recombinante a aproximadamente 100 de tales placas de microtitulación por conveniencia para el escrutinio.

Las técnicas descritas en la presente memoria son especialmente útiles en la recombinación de genes de familias de genes, donde la diversidad en la secuencia de nucleótidos es proporcionada en su totalidad o en parte por las diferencias naturales en la secuencia de nucleótidos de los genes de la familia.

Se pretende que una "familia de genes" según se utiliza en la presente memoria incluya genes con una función similar, tales como, pero no limitados a, interferones o interleuquinas; los genes que se cree que derivan por descendencia de un ancestro común; y genes que codifican proteínas que son estructuralmente homólogas, tales como las proteínas del haz de cuatro hélices.

De este modo, por ejemplo las secuencias de ADN o proteínas se pueden alinear mediante algoritmos informáticos, tales como los descritos en la monografía de bioinformática de Schomburg y Lessel (Schomburg y Lessel, Bioinformatics: From Nucleic Acids and Proteins to Cell Metabolism, 9 de Octubre de 1995, Braunschweig, Alemania). Estos algoritmos pueden determinar la probabilidad de que dos secuencias, o subdominios de

secuencias, estén relacionados entre sí por descendencia de un ancestro común. Las secuencias que se cree que se obtienen mediante descendencia de un ancestro común comprenden una "familia de genes homólogos", y se puede
5 utilizar el barajado de ADN para acelerar la evolución de estas familias de genes.

Además, muchas secuencias de proteínas distintas son consistentes con plegamientos de proteínas similares, y
10 se puede decir que tales familias de secuencias comprenden familias de genes "estructuralmente homólogos". La superfamilia de proteínas de haces de cuatro hélices es una de tales familias. Aunque es una familia muy grande de proteínas muy diversas
15 funcionalmente que oscilan entre las citoquinas y las enzimas para las proteínas de unión al ADN que tienen este plegamiento, es poco probable que estas proteínas deriven de un ancestro común. Es más probable que hayan "evolucionado convergentemente" para tener plegamientos
20 de proteína similares. Ahora existen algoritmos funcionales (Dahiyat et al., Science 278:82-87 (1997)) que permiten diseñar proteínas con los plegamientos de proteína deseados, y tales algoritmos han sido utilizados para diseñar, por ejemplo, motivos en dedo de cinc que no
25 están relacionados en su secuencia primaria con ninguna de las proteínas naturales conocidas.

A. Uso de los Sitios para las Enzimas de Restricción para Recombinar Mutaciones

30 En algunas situaciones es ventajoso utilizar sitios de enzimas de restricción en ácidos nucleicos para dirigir la recombinación de mutaciones en una secuencia de ácido nucleico de interés. Estas técnicas son
35 particularmente preferidas en el desarrollo de fragmentos

que no se pueden barajar fácilmente por los métodos existentes debido a la presencia de ADN repetidos u otros motivos de secuencia primarios problemáticos. También son preferidos para barajar fragmentos grandes (típicamente mayores de 10 kb), tales como agrupaciones génicas que no pueden ser barajadas fácilmente ni "amplificadas por PCR" debido a su tamaño. Aunque se ha referido que se han amplificado fragmentos de hasta 50 kb mediante PCR (Barnes, Proc. Natl. Acad. Sci. (U.S.A.) 91:2216-2220 (1994)), esto puede resultar problemático para fragmentos de más de 10 kb, y de este modo se prefieren métodos alternativos para el barajado entre 10 - 50 kb y más allá. Preferiblemente, las endonucleasas de restricción utilizadas son del tipo Clase II (Sambrook et al., Molecular Cloning, CSH Press, 1987) y de estas, preferiblemente las que generan salientes de extremos cohesivos no palindrómicos tales como AlwI, Sfi I o BstXI. Estas enzimas generan extremos no palindrómicos que permiten un re-ensamblaje ordenado eficaz con la ADN ligasa. Típicamente, los sitios para las enzimas de restricción (o endonucleasas) se identifican mediante técnicas de mapeo de enzimas de restricción convencionales (Sambrook et al., Molecular Cloning, CSH Press, 1987), mediante análisis de la información de la secuencia para ese gen, o mediante introducción de los sitios de restricción deseados en una secuencia de ácido nucleico mediante síntesis (esto es mediante incorporación de mutaciones silenciosas).

Las moléculas sustrato de ADN que se van a digerir pueden ser de ADN replicado *in vivo*, tal como una preparación de plásmido, o de fragmentos de ácido nucleico amplificados por PCR que albergan sitios de reconocimiento para enzimas de restricción de interés, preferiblemente cerca de los extremos del fragmento.

Típicamente, al menos dos variantes de un gen de interés, que tienen cada una dos o más mutaciones, son digeridas con al menos una enzima de restricción determinada para cortar en la secuencia de ácido nucleico de interés. Los fragmentos de restricción se unen después con ADN ligasa para generar genes completos que tienen regiones barajadas. El número de regiones barajadas dependerá del número de cortes en la secuencia de ácido nucleico de interés. Las moléculas barajadas pueden ser introducidas en las células como se ha descrito más arriba y escrutadas o seleccionadas en busca de una propiedad deseada. Después se puede aislar el ácido nucleico de las reservas (genotecas) o clones que tienen propiedades deseadas y someterlo al mismo procedimiento hasta que se obtenga un grado de mejora deseado.

En algunas realizaciones, se aísla al menos una molécula sustrato de ADN o uno de sus fragmentos y se somete a mutagénesis. En ciertas realizaciones, la reserva o genoteca de fragmentos de restricción religados se somete a mutagénesis antes de repetir el procedimiento de digestión-ligación. "Mutagénesis" según se utiliza en la presente memoria comprende mecanismos conocidos en la técnica tales como mutagénesis por PCR, mutagénesis dirigida a oligonucleótidos, mutagénesis dirigida al sitio, etc., y recombinación repetitiva de secuencias mediante cualquiera de las técnicas descritas en la presente memoria.

Un ejemplo del uso de este formato es en la manipulación de agrupaciones de policétidos. Las agrupaciones de policétidos (Khosla et al., TIBTECH 14, Septiembre 1996) tienen típicamente de 10 a 100 kb de longitud, especificando policétidos grandes múltiples que se ensamblan en complejos multienzimáticos muy grandes.

Debido a la naturaleza modular de estos complejos y a la naturaleza modular de la ruta biosintética, los ácidos nucleicos que codifican módulos de proteínas pueden ser intercambiados entre diferentes agrupaciones de policétidos para generar policétidos quiméricos novedosos y funcionales. La introducción de sitios para endonucleasas de restricción raros tales como Sfil (reconocimiento de ocho bases, salientes no palindrómicos) en sitios no esenciales entre polipéptidos o en intrones diseñados en los polipéptidos proporcionaría "asas" con las cuales manipular el intercambio de segmentos de ácido nucleico utilizando la técnica descrita más arriba.

B. PCR de ensamblaje

Una técnica adicional para recombinar repetitivamente mutaciones en una secuencia de ácido nucleico utiliza "PCR de re-ensamblaje". Este método puede ser utilizado para ensamblar múltiples segmentos que se han desarrollado por separado en un molde de ácido nucleico completo tal como un gen. Esta técnica se realiza cuando se conoce una reserva de mutantes ventajosos de un trabajo previo o ha sido identificada mediante escrutinio de mutantes que pueden haber sido creados mediante cualquier mecanismo de mutagénesis conocido en la técnica, tal como mutagénesis por PCR, mutagénesis por inserción de casete, mutagénesis de oligos dopados, mutagénesis química, o propagación del molde de ADN in vivo en cepas mutadoras. Los límites que definen los segmentos de una secuencia de ácido nucleico de interés se encuentran preferiblemente en regiones intergénicas, intrones, o zonas de un gen que probablemente no tiene mutaciones de interés. Preferiblemente, los cebadores oligonucleotídicos

(oligos) son sintetizados para la amplificación por PCR de segmentos de la secuencia de ácido nucleico de interés, de manera que las secuencias de los oligonucleótidos solapan las uniones de dos segmentos. La
5 región de solapamiento tiene típicamente de aproximadamente 10 a 100 nucleótidos de longitud. Cada uno de los segmentos se amplifica con un grupo de tales cebadores. Los productos de la PCR son re-ensamblados después de acuerdo con los protocolos de ensamblaje
10 utilizados en las Secciones IA-B de más arriba para ensamblar los genes fragmentados al azar. En resumen, en un protocolo de ensamblaje los productos de la PCR son purificados primero mediante PCR de los cebadores, por ejemplo, mediante electroforesis en gel o cromatografía
15 de exclusión por tamaños. Los productos purificados se mezclan entre sí y se someten a aproximadamente 1-10 ciclos de desnaturalización, re-recocido, y ampliación en presencia de polimerasa y desoxinucleosido-trifosfatos (dNTP) y sales tampón apropiadas en ausencia de cebadores
20 adicionales ("auto-cebado"). Se utilizan PCR posteriores con cebadores que flanquean el gen para amplificar el rendimiento de los genes totalmente re-ensamblados y barajados. Este método es necesariamente "grosero" y por lo tanto solamente recombina mutaciones por bloques, una
25 ventaja para algunas búsquedas como cuando se recombinan variantes alélicas de múltiples genes con un operón.

En algunas realizaciones, los genes re-ensamblados resultantes son sometidos a mutagénesis antes del repetir
30 el procedimiento.

En algunas realizaciones, se utilizan oligonucleótidos que incorporan uracilo en los cebadores para la amplificación por PCR. Típicamente se incorpora
35 uracilo en un sitio en el oligonucleótido. Los productos

se tratan con uracilo glicosilasa, generando de ese modo un saliente de hebra sencilla, y se re-ensamblan de una manera ordenada mediante un método tal como el descrito por Rashtchian (Current Biology, 6:30-36 (1995)).

5

En una realización adicional, los cebadores de PCR para la amplificación de segmentos de la secuencia de ácido nucleico de interés se utilizan para introducir una variación en el gen de interés como sigue. Las mutaciones en los sitios de interés en una secuencia de ácido nucleico se identifican mediante escrutinio o selección, secuenciando homólogos de la secuencia de ácido nucleico, etcétera. Después se sintetizan los cebadores oligonucleotídicos de la PCR que codifican la información de tipo salvaje o mutante en los sitios de interés. Estos cebadores se utilizan después en la mutagénesis mediante PCR para generar genotecas de genes completos que codifican permutaciones de información de tipo salvaje y mutante en las posiciones designadas. Esta técnica es típicamente ventajosa en los casos en los que el procedimiento de escrutinio o selección es costoso, engorroso, o poco práctico con respecto al coste de secuenciación de los genes de los mutantes de interés y la síntesis de oligonucleótidos mutagénicos.

25

Un ejemplo de este método es el desarrollo de una polimerasa Taq mejorada, como se describe con detalle más abajo. Las proteínas mutantes resultantes de la aplicación de este método se identifican y se analizan en una reacción de secuenciación para identificar los mutantes con propiedades de secuenciación mejoradas. Esto se realiza típicamente en un formato de alto rendimiento (véase, por ejemplo, Broach et al. Nature 384 (Sup): 14-16 (1996)) para producir, después del escrutinio, un pequeño número, p. ej., aproximadamente de

35

2 a 100, de recombinantes candidato para la evaluación adicional. Después los genes mutantes se pueden secuenciar para proporcionar información acerca de la localización de la mutación. Se pueden sintetizar los correspondientes cebadores oligonucleotídicos mutagénicos a partir de esta información, y utilizarlos en una reacción de re-ensamblaje como se ha descrito antes para generar eficazmente una genoteca con una media de muchas mutaciones por gen. Una o más rondas de este protocolo permiten la búsqueda eficaz de variantes mejoradas de la polimerasa Taq.

C. Enriquecimiento de Información de Secuencias Mutantes

En algunas realizaciones de la invención, las reacciones de recombinación, tales como las comentadas antes, son enriquecidas en secuencias mutantes de manera que el espectro de múltiples mutantes, esto es las posibles combinaciones de mutaciones, sea muestreado más eficazmente. La base lógica de esto es la siguiente. Supongamos que se obtiene un número, n , de clones mutantes con actividad mejorada, donde cada clon tiene un único punto de mutación en una posición diferente en la secuencia de ácido nucleico. Si esta población de clones mutantes con una media de una mutación de interés por secuencia de ácido nucleico se coloca después en una reacción de recombinación, la población resultante todavía tendrá una media de una mutación de interés por secuencia de ácido nucleico como se define mediante la distribución de Poisson, dejando el espectro de múltiples mutaciones relativamente despoblado.

La cantidad de escrutinio requerido para identificar recombinantes que tienen dos o más mutaciones se puede reducir espectacularmente por medio de la siguiente

técnica. Las secuencias de ácido nucleico de interés se obtienen de una reserva de clones mutantes y se preparan en forma de fragmentos, típicamente mediante digestión con endonucleasas de restricción, sonicación, o
5 amplificación por PCR. Los fragmentos se desnaturalizan, después se permite que se vuelvan a recocer, generando de ese modo híbridos con emparejamientos erróneos donde una hebra de un mutante ha hibridado con una hebra complementaria de un clon mutante o de tipo salvaje
10 diferente. Los productos re-recocidos se fragmentan después en fragmentos de aproximadamente 20 - 100 pb, por ejemplo, mediante el uso de ADNase I. Esta reacción de fragmentación tiene el efecto de segregar regiones del molde que contienen los emparejamientos erróneos
15 (información mutante) de aquellos que codifican la secuencia de tipo salvaje. Los híbridos con emparejamientos erróneos se pueden purificar después por afinidad utilizando aptámeros, colorantes, u otros agentes que se unen al ADN con emparejamientos erróneos.
20 Una realización preferida es el uso de la matriz de afinidad de la proteína mutS (Wagner et al., Nucleic Acids Res. 23(19):3944-3948 (1995); Su et al., Proc. Natl. Acad. Sci. (U.S.A.), 83:5057-5061(1986)) con una etapa preferida de amplificación del material purificado
25 por afinidad in vitro antes de una reacción de ensamblaje. Este material amplificado se pone después en una reacción de PCR de ensamblaje como se ha descrito antes. Opcionalmente, este material puede ser titulado frente a la reserva de mutante original (p. ej., de
30 aproximadamente 100% a 10% de la reserva enriquecida en mutS) para controlar el número medio de mutaciones por clon en la siguiente ronda de recombinación.

Otra aplicación de este método es en el ensamblaje
35 de constructos génicos que están enriquecidos en bases

polimórficas que existen como variantes alélicas naturales o seleccionadas o como diferencias entre genes homólogos de especies relacionadas. Por ejemplo, se pueden tener diversas variedades de una planta que se cree que tienen una variación hereditaria en un rasgo de interés (p. ej., resistencia a la sequía). A continuación es interesante construir una genoteca de estos genes variantes que contienen muchas mutaciones por gen. La selección Muts puede ser aplicada combinada con las técnicas de ensamblaje descritas en la presente memoria para generar una reserva de recombinantes tal que esté altamente enriquecida en información polimórfica ("mutante"). En algunas realizaciones, la reserva de genes recombinantes se proporciona en un anfitrión transgénico. Los recombinantes pueden ser desarrollados adicionalmente mediante amplificación por PCR del transgén de organismos transgénicos que se ha determinado que tienen un fenotipo mejorado y aplicando los formatos descritos en esta invención para desarrollarlos adicionalmente.

D. Recombinación conducida por intrones

En algunos casos, las moléculas sustrato para la recombinación tienen una homología uniformemente baja, regiones de homología distribuidas esporádicamente, o la región de homología es relativamente pequeña (por ejemplo, aproximadamente 10 - 100 pb), tales como ligandos de péptidos presentados en fagos. Estos factores pueden reducir la eficacia y aleatoriedad de la recombinación en RSR. En algunas realizaciones de la invención, este problema se trata mediante la introducción de intrones entre exones codificantes en secuencias que codifican homólogos de proteínas. En realizaciones adicionales de la invención, se pueden

utilizar intrones (Chong et al., J. Biol. Chem., 271:22159-22168 (1996)).

5 En este método, se define arbitrariamente una
secuencia de ácido nucleico, tal como un gen o una
familia de genes, para que tenga segmentos. Los segmentos
son preferiblemente exones. Los intrones son diseñados
entre los segmentos. Preferiblemente, el intrón insertado
entre el primer y el segundo segmentos es divergente al
10 menos aproximadamente en 10% del intrón insertado entre
el segundo y el tercer segmentos, el intrón insertado
entre el segundo y el tercer segmentos es divergente en
al menos aproximadamente 10% de los intrones insertados
entre cualquiera de los pares de segmentos previos, y así
15 sucesivamente en los segmentos n y $n+1$. Los intrones
entre cualquier grupo de exones dado serán de este modo
inicialmente idénticos entre todos los clones de la
genoteca, mientras los exones pueden tener una secuencia
arbitrariamente divergente. Los intrones proporcionan por
20 los tanto secuencias de ADN homólogas que permitirán la
aplicación de cualquiera de los métodos descritos para
RSR mientras los exones pueden ser arbitrariamente
pequeños o de secuencia divergente, y pueden evolucionar
para lograr un grado arbitrariamente grande de
25 divergencia de la secuencia sin una pérdida significativa
en la eficacia de recombinación. También se pueden
diseñar sitios de restricción en la secuencia de ácido
nucleico intrónico de interés con el fin de permitir un
re-ensamblaje directo de los fragmentos de restricción.
30 El ADN del exón de partida puede ser sintetizado de novo
a partir de la información de la secuencia, o puede estar
presente en cualquier preparación de ácido nucleico (p.
ej., genómico, ADNc, genotecas, etcétera). Por ejemplo,
se pueden diseñar de 1 a 10 intrones no homólogos para
35 dirigir la recombinación de las secuencias de ácido

nucleico de interés colocándolas entre exones. La secuencia de los intrones puede ser total o parcialmente obtenida de secuencias de intrones conocidas. Preferiblemente, los intrones son auto-ajustables. Los grupos ordenados de genotecas de intrones y exones se ensamblan en genes funcionales mediante métodos convencionales (Sambrook et al., Molecular Cloning, CSH Press (1987)).

Se puede aplicar cualquiera de los formatos para la recombinación in vitro o in vivo descritos en la presente memoria para el barajado de exones repetitivo. Un formato preferido consiste en el uso de sitios de restricción no palindrómicos tales como Sfi I situados en las secuencias intrónicas para promover el barajado. Las reservas de clones seleccionados se digieren con Sfi I y se vuelven a ligar. Los salientes no palindrómicos promueven el re-ensamblaje ordenado de los exones barajados. Estas genotecas de genes se pueden expresar y escrutar en busca de las propiedades deseadas, después se someten a rondas repetitivas adicionales de recombinación mediante este procedimiento. En algunas realizaciones, las genotecas se someten a mutagénesis antes de repetir el procedimiento.

Un ejemplo de cómo se utilizaría la introducción de un intrón en un formato de genoteca de mamífero ventajosamente es el siguiente. Un intrón que contiene un sitio lox (Sauer et al., Proc. Natl. Acad. Sci. (U.S.A.), 85:5166-5170 (1988)) es introducido arbitrariamente entre los aminoácidos 92 y 93 en cada sustrato parental de interferón alfa. Se elabora una genoteca de 10^4 genes quiméricos de interferón para cada uno de los dos exones (residuos 1-92 y residuos 93-167), se clona en un vector plasmídico replicante, y se introduce en células diana. El número 10^4 se selecciona arbitrariamente por

conveniencia en el escrutinio. Un vector ilustrativo para la expresión en células de mamífero contendría un origen de SV40, expresando las células anfitrionas el antígeno T grande de SV40, con el fin de permitir la expresión transitoria de los constructos de interferón. Las células se sensibilizan con un virus citopático tal como el virus de la estomatitis vesicular (VSV) en un análisis de protección con interferón (p. ej., Meister et al., J Gen. Virol. 67:1633-1643, (1986)). Las células que sobreviven debido a la expresión del interferón se recuperan, las dos genotecas de interferón se amplifican por PCR, y se vuelven a clonar en un vector que puede ser amplificado en *E. coli*. Los plásmidos amplificados se transfectan después a una multiplicidad elevada (p. ej. 10 microgramos de plásmido por 10^6 células) en un anfitrión que expresa cre que puede soportar la replicación de ese vector. La presencia de cre en las células anfitrionas promueve la recombinación eficaz en el sitio lox en el intrón del interferón, barajando de ese modo los grupos seleccionados de exones. Esta población de células se utiliza después en una segunda ronda de selección mediante sensibilización viral y el procedimiento se aplica repetidamente. En este formato, la recombinasa cre es expresada preferiblemente de manera transitoria en una molécula co-transfectada que no puede replicar en el anfitrión. De este modo, después de la segregación de recombinantes a partir del plásmido que expresa cre, no se producirá más recombinación y se puede realizar la selección en permutaciones de exones estables genéticamente. El método puede ser utilizado con más de un intrón, con secuencias potenciadoras de la recombinación distintas de cre/lox (p. ej., int/xis, etc.), y con otros sistemas vectores tales como, pero no limitados a, retrovirus, adenovirus o virus adenoasociados.

5. Recombinación Mediada por Oligonucleótidos Sintéticos

1. Oligo puente a través del espacio de la secuencia

5 En algunas realizaciones de la invención, se desea la búsqueda de una región de espacios de la secuencia definida por un grupo de sustratos, tales como miembros de una familia de genes, que tengan una homología de menos de aproximadamente 80%, más típicamente, menos de
10 aproximadamente 50%. Esta región, que puede ser una parte o todo el gen o un gen se define arbitrariamente en segmentos. Las fronteras del segmento se pueden elegir al azar, basándose en la correspondencia con exones naturales, basándose en consideraciones estructurales
15 (bucles, hélices alfa, subdominios, dominios completos, núcleo hidrófobo, superficie, simulaciones dinámicas), y basándose en correlaciones con datos de mapas genéticos.

 Típicamente, los segmentos se amplifican después
20 mediante PCR con una reserva de oligonucleótidos "puente" en cada unión. De este modo, si el grupo de cinco genes se rompe en tres segmentos A, B y C, y si hay cinco versiones de cada segmento (A1, A2, ... C4, C5), se elaboran veinticinco oligonucleótidos para cada hebra de
25 la unión A-B donde cada oligo puente tiene 20 bases de homología con uno de los segmentos A y uno de los B. En algunos casos, el número de oligonucleótidos requeridos se puede reducir seleccionando límites del segmento que sean idénticos en algunos o todos los miembros de la
30 familia de genes. Se sintetizan de un modo similar los oligonucleótidos para la unión B-C. La familia de dominios A se amplifica mediante PCR con un cebador A genérico externo y la reserva de oligonucleótidos de la unión A-B; los dominios B con los oligonucleótidos puente
35 A-B mas B-C, y los dominios C con los oligonucleótidos

puente B-C mas un cebador exterior genérico. Elaborados los genes completos, se confeccionan después mediante PCR de ensamblaje o mediante los métodos de dUTP/uracilo glicosilada descritos antes. Preferiblemente, los productos de esta etapa se someten a mutagénesis antes del procedimiento de repetir el procedimiento de selección y recombinación, hasta que se obtiene un nivel de mejora o la evolución de una propiedad deseada. Esto se determina típicamente utilizando un escrutinio o selección apropiado para la proteína y la propiedad de interés.

Una ilustración de este método se muestra más abajo para la recombinación de once genes de interferón alfa humano homólogos.

2. Mutagénesis Dirigida al Sitio (SDM) con Oligonucleótidos que Codifican Mutaciones Homólogas Seguida de Barajado

En algunas realizaciones de la invención, se añade la información de secuencia de una o más secuencias sustrato a una secuencia "parental" de interés, con una recombinación posterior entre las rondas de escrutinio o selección. Típicamente, esto se realiza con una mutagénesis dirigida al sitio realizada mediante mecanismos bien conocidos en la técnica (Sambrook et al., Molecular Cloning, CSH Press (1987)) con un sustrato como molde y oligonucleótidos que codifican mutaciones individuales o múltiples de otras secuencias sustrato, p. ej. genes homólogos. Después del escrutinio o selección en busca de un fenotipo de interés mejorado, el recombinante o los recombinantes seleccionados se pueden desarrollar adicionalmente utilizando las técnicas RSR descritas en la presente memoria. Después del escrutinio

o selección, se puede realizar de nuevo la mutagénesis dirigida al sitio con otra colección de oligonucleótidos que codifiquen mutaciones homólogas, y repetir el procedimiento anterior hasta obtener las propiedades deseadas.

Cuando la diferencia entre dos homólogos es una o más mutaciones puntuales individuales en un codón, se pueden utilizar oligonucleótidos degenerados que codifiquen las secuencias en ambos homólogos. Un oligo puede incluir muchos de tales codones degenerados y permitir todavía una búsqueda exhaustiva de todas las permutaciones a lo largo de ese bloque de secuencia. Un ejemplo de esto se proporciona más abajo para la evolución de los genes del interferón alfa.

Cuando el espacio de la secuencia homóloga es muy grande, puede ser ventajoso restringir la búsqueda a ciertas variantes. De este modo, por ejemplo, se pueden utilizar herramientas de modelado por ordenador (Lathrop et al., J. Mol. Biol., 255:641-665 (1996)) para modelar cada mutación homóloga sobre la proteína diana y descartar cualquier mutación que se prevé que interrumpa groseramente la estructura y la función.

F. Recombinación Dirigida por la Maquinaria del Anfitrión

En algunas realizaciones de la invención, las moléculas sustrato de ADN se introducen en células, donde la maquinaria celular dirige su recombinación. Por ejemplo, se construye una genoteca de mutantes y se escruta o selecciona en busca de mutantes con fenotipos mejorados mediante cualquiera de las técnicas descritas en la presente memoria. Las moléculas sustrato de ADN que codifican los mejores candidatos se recuperan mediante

cualquiera de las técnicas descritas en la presente memoria, después se fragmentan y se utilizan para transfectar un anfitrión mamífero y se escrutan o seleccionan en busca de una función mejorada. Las moléculas sustrato de ADN se recuperan del anfitrión mamífero, por ejemplo mediante PCR, y el procedimiento se repite hasta obtener un nivel deseado de mejora. En algunas realizaciones, los fragmentos se desnaturalizan y se recuecen antes de la transfección, se recubren con proteínas estimuladoras de la recombinación tales como recA, o se co-transfectan con un marcador seleccionable tal como Neo^R para permitir la selección positiva en busca de células que reciben las versiones recombinadas del gen de interés.

Por ejemplo, se prefiere este formato para la maduración de la afinidad in vivo de un anticuerpo mediante RSR. En resumen, se genera una genoteca de anticuerpos mutantes, como se describe en la presente memoria para la maduración de la afinidad de 48G7. Esta genoteca es purificada por FACS con ligando para enriquecer en anticuerpos con una afinidad máxima de 0,1 - 10%. Los genes de las regiones V se recuperan mediante PCR, se fragmentan, y se co-transfectan o se someten a electroporación con un vector en el cual se pueden recombinar los genes de la región V re-ensamblados. Las moléculas sustrato de ADN se recuperan de las células co-transfectadas, y el procedimiento se repite hasta obtener el nivel de mejora deseado. Otras realizaciones incluyen el re-ensamblaje de las regiones V antes de la electroporación de manera que un exón de una región V intacta se pueda recombinar en un casete de expresión de anticuerpo. Otras realizaciones incluyen el uso de este formato para otros genes eucarióticos o para la evolución de virus completos.

G. Ensamblaje Basado en Fagémidos

En algunas realizaciones de la invención, se clona un gen de interés en un vector que genera ADN de hebra sencilla, tal como un fagémido. El sustrato de ADN resultante se mutageniza mediante RSR con cualquier método conocido en la técnica, se transfecta en células anfitrionas, y se somete a escrutinio o selección en busca de la propiedad deseada o de un fenotipo mejorado. El ADN de los fagémidos seleccionados o escrutados se amplifica, por ejemplo, mediante PCR o preparación de plásmidos. Esta preparación de ADN contiene las diferentes secuencias mutantes que se desean para la permuta. Este ADN es fragmentado y desnaturalizado, y recocido con el molde del fagémido de ADN de doble hebra (ADNdh) (ADNdh que codifica el gen de tipo salvaje y las secuencias del vector). Una realización preferida es el uso de cepas dut(-) ung(-) tales como CJ236 (Sambrook et al., Molecular Cloning CSH Press (1987)) para la preparación de ADNdh.

Los espacios del molde recocido se llenan con ADN polimerasa y se ligan para formar círculos relajados cerrados. Puesto que múltiples fragmentos se pueden recocer con el fagémido, la hebra recién sintetizada consiste ahora en secuencias barajadas. Estos productos se transforman en una cepa mutS de *E. coli* que es dut+ ung+. El ADN del fagémido se recupera del anfitrión transfectado y se somete de nuevo a este protocolo hasta obtener el nivel deseado de mejora. El gen que codifica la proteína de interés en esta genoteca de ADN de fagémido recuperado puede ser mutagenizado mediante cualquier técnica, incluyendo RSR, antes de repetir el procedimiento.

III. Expresión Mejorada de Proteínas

Si bien se ha demostrado que la tecnología de ADN recombinante es un método muy general para obtener cantidades grandes, puras y homogéneas de casi todas las secuencias de ácido nucleico de interés, todavía no se ha logrado una generalidad similar para la producción de grandes cantidades de proteína pura, homogénea en forma recombinante. Una explicación probable es que la expresión, plegamiento, localización y estabilidad de la proteína, es intrínsecamente más compleja e impredecible que para el ADN. El rendimiento de proteína expresada es una función compleja de las velocidades de transcripción, las velocidades de traducción, interacciones con el ribosoma, interacción del polipéptido nascente con chaperoninas y otras proteínas en la célula, eficacia de oligomerización, interacción con componentes de secreción y otras rutas de tráfico de proteínas, sensibilidad a la proteasa, y capacidad intrínseca del estado plegado final. La optimización de tales procedimientos complejos está bien adaptada a la aplicación de RSR. Los siguientes métodos detallan estrategias para la aplicación de RSR a la optimización de la expresión de proteínas.

A. Evolución de Genes Mutantes con Expresión Mejorada Utilizando RSR en Genotecas de Uso Codónico

El efecto negativo de los codones raros de *E. coli* sobre la expresión de proteínas recombinantes en este anfitrión ha sido claramente demostrado (Rosenberg, et al., J. Bact. 175:716-722 (1993)). No obstante, las reglas generales para la elección del uso de patrones de uso de codones para optimizar la expresión de proteínas funcionales han sido difíciles de alcanzar. En algunas realizaciones de la invención, la expresión de proteínas

se optimiza cambiando los codones utilizados en el gen de interés, basándose en la degeneración del código genético. Típicamente, esto se logra sintetizando el gen utilizando oligonucleótidos degenerados. En algunas realizaciones los oligonucleótidos degenerados tienen la estructura general de aproximadamente 20 nucleótidos de identidad con una molécula sustrato de ADN que codifica una proteína de interés, seguido de una región de aproximadamente 20 nucleótidos degenerados que codifican una región de la proteína, seguida de otra región de aproximadamente 20 nucleótidos de identidad. En una realización preferida, la región de identidad utiliza codones preferidos para el anfitrión. En una realización adicional, los oligonucleótidos son idénticos al sustrato de ADN al menos un nucleótido 5' y uno 3', pero tienen una homología de secuencia de al menos 85% con la molécula sustrato de ADN, con la diferencia debida al uso de codones degenerados. En algunas realizaciones, se utiliza un grupo de tales oligonucleótidos degenerados en el que cada oligonucleótido se solapa con otro mediante la fórmula general $n - 10$, donde n es la longitud del oligonucleótido. Tales oligonucleótidos tienen típicamente aproximadamente 20 - 1000 nucleótidos de longitud. Los genes ensamblados se clonan después, se expresan, y se escrutan o seleccionan en busca de la expresión deseada. Los genes ensamblados pueden ser sometidos a métodos de recombinación repetitiva como se ha descrito antes hasta lograr la mejora deseada.

Por ejemplo, esta técnica se puede utilizar para desarrollar fosfatasa alcalina intestinal bovina (BIAP) para su expresión activa en E. coli. Esta enzima se utiliza comúnmente como gen informador en formatos de análisis tales como ELISA. El gen clonado no puede ser expresado en forma activa en un anfitrión procariótico

tal como *E. coli* con un buen rendimiento. El desarrollo de un sistema de expresión podría permitir acceder a una tecnología de expresión poco costosa para BIAP e, importantemente, para variantes diseñadas con una actividad o con propiedades de acoplamiento químico mejoradas (tal como el acoplamiento químico a anticuerpos). Se proporciona un ejemplo detallado en la sección de Ejemplos Experimentales.

B. Plegamiento Mejorado

En algunas realizaciones de la invención, las proteínas de interés cuando son expresadas en exceso o expresadas en anfitriones heterólogos forman cuerpos de inclusión, encontrándose la mayor parte de la proteína expresada en productos agregados insolubles. Se pueden utilizar técnicas de recombinación repetitiva de secuencias para optimizar el plegamiento de tales proteínas diana. Existen diversas formas de mejorar el plegamiento, incluyendo el desarrollo mutante de la proteína diana de interés y el desarrollo de proteínas chaperoninas.

1. Desarrollo de una Proteína Diana

a. Selección por Fraccionamiento de Cuerpos de Inclusión Utilizando la Proteína de Fusión del Dímero de la Porción de Cabeza de lac

El "dímero de la porción de cabeza" del represor lac es una proteína pequeña que contiene dos dominios de la porción de cabeza conectados por un conector peptídico corto que se une al operador lac con una afinidad suficiente para que las fusiones de polipéptidos a este dímero de la porción de cabeza permanezca unido al

plásmido que los codifica por medio de un procedimiento de purificación por afinidad (Gates et al., J. Mol. Biol. 255:373-386 (1995)). Esta propiedad puede ser explotada, como sigue, para desarrollar proteínas mutantes de interés con propiedades de plegamiento mejoradas. La proteína de interés puede ser de mamífero, de levadura, bacteriana, etc.

Se construye una proteína de fusión entre el dímero de la porción de cabeza y una secuencia de una proteína diana, por ejemplo, como describe Gates (cita anterior). Este constructo, que contiene al menos un operador lac, se mutageniza mediante tecnologías comunes en las técnicas tales como mutagénesis por PCR, mutagénesis química, mutagénesis dirigida a oligos (Sambrook et al., Molecular Cloning CSH Press (1987)). La genoteca resultante se transforma en una célula anfitriona, y se induce la expresión de la proteína de fusión, preferiblemente con arabinosa. Se genera un extracto o un producto lisado a partir del cultivo de la genoteca que expresa el constructo. La proteína insoluble se fracciona de los complejos de proteína soluble/ADN mediante centrifugación o cromatografía por afinidad, y el rendimiento de los complejos de proteína soluble/ADN se cuantifica mediante PCR cuantitativa (Sambrook et al., Molecular Cloning, CSH Press, 1987) del plásmido. Preferiblemente, se utiliza un reactivo que es específico para una proteína plegada apropiadamente, tal como un anticuerpo monoclonal o un ligando natural, para purificar complejos de proteína soluble/ADN. El ADN plasmídico de esta etapa se aísla, se somete a RSR y se expresa de nuevo. Estas etapas se repiten hasta que el rendimiento de complejos de proteína soluble/ADN ha alcanzado un nivel de mejora deseado. Después se escrutan clones individuales en busca de la retención de las

propiedades funcionales de la proteína de interés, tales como la actividad enzimática, etc.

Esta técnica generalmente es útil para desarrollar solubilidad y otras propiedades tales como el tráfico celular de proteínas expresadas heterológamente en una célula anfitriona de interés. Por ejemplo, se podría seleccionar en busca de un plegamiento eficaz y una localización nuclear de una proteína fusionada al dímero de la porción de cabeza del represor lac codificando la proteína en un plásmido que codifique un origen de replicación de SV40 y un operador lac, y expresando transitoriamente la proteína de fusión en un anfitrión mamífero que exprese el antígeno T. La purificación de complejos de proteína/ADN de extractos HIRT celulares (Seed y Aruffo, Proc. Natl. Acad. Sci. (U.S.A.), 84:3365-3369 (1987)) permitiría seleccionar proteínas con un plegamiento y una localización nuclear eficaces.

b. Expresión Funcional de Proteínas Utilizando la Presentación en Fagos

Un problema encontrado a menudo en los métodos de presentación en fagos tales como los descritos por O'Neil et al. (Current Biology, 5:443-449 (1995)) es la incapacidad de expresar funcionalmente una proteína de interés en fagos. Sin estar limitado por ninguna teoría, el plegamiento inapropiado de la proteína de interés puede ser responsable de este problema. Se puede utilizar la RSR para desarrollar una proteína de interés para su expresión funcional en fagos. Típicamente, se construye una proteína de fusión entre el gen III o el gen VIII y la proteína diana y después se mutageniza, por ejemplo mediante mutagénesis por PCR. La genoteca mutagenizada se expresa después en un formato de presentación en fagos,

se elabora una solución de partida de fagos, y estos fagos se seleccionan por afinidad en busca de los que portan proteínas de fusión presentadas funcionalmente utilizando una matriz de afinidad que contiene un ligando conocido para la proteína diana. El ADN del fago seleccionado funcionalmente se purifica, y los genes de interés presentados se barajan y se vuelven a clonar en el formato de presentación en fagos. Las etapas de selección, barajado y reclonación se repiten hasta que el rendimiento de fago con la proteína funcional presentada ha alcanzado los niveles deseados definidos, por ejemplo, por la fracción de fagos que son retenidos sobre la matriz de afinidad del ligando o la actividad biológica asociada con el fago presentado. Los clones individuales pueden ser escrutados después para identificar los mutantes candidato con propiedades de presentación mejorada, nivel de expresión deseado, y propiedades funcionales de interés (p. ej., capacidad para unirse a un ligando o receptor, actividad de linfoquinas, actividad enzimática, etc.).

En algunas realizaciones de la invención, se utiliza un escrutinio o selección funcional para identificar una proteína desarrollada no expresada en un fago. La proteína diana, que no puede ser expresada inicialmente de manera eficaz en un anfitrión de interés, se mutageniza y se utiliza un escrutinio o selección funcional para identificar las células que expresan la proteína funcional. Por ejemplo, la proteína de interés puede complementar una función en la célula anfitriona, escindir un sustrato colorimétrico, etc. La recombinación repetitiva de secuencias se utiliza después para desarrollar rápidamente una expresión funcional mejorada a partir de semejante reserva de mutantes mejorados.

Por ejemplo, la transcriptasa inversa de AMV tiene una importancia comercial concreta porque es activa a una temperatura superior (42°C) y es más robusta que muchas otras transcriptasas inversas. No obstante, es difícil expresarla en anfitriones procarióticos tales como *E. coli*, y por consiguiente es costosa porque tiene que ser purificada de células de pollo. De este modo es muy deseable una transcriptasa inversa de AMV desarrollada que pueda ser expresada eficazmente en *E. coli*.

En resumen, el gen de la transcriptasa inversa de AMV (Papas et al., J. Cellular Biochem 20:95-103 (1982)) es mutagenizado mediante cualquier método común en la técnica. La genoteca de genes mutantes es clonada en un plásmido colE1 (resistente a Amp) bajo el control del promotor lac en un anfitrión *E. coli* polA12 (Ts) recA718 (Sweasy et al. Proc. Natl. Acad. Sci. U.S.A. 90:4626-4630 (1993)). La genoteca es inducida con IPTG, y llevada a la temperatura no permisiva. Esto selecciona los genes de transcriptasa inversa expresados funcionalmente en las condiciones selectivas referidas para la selección de los mutantes de transcriptasa inversa de VIH activos referidos por Kim et al. (Proc. Natl. Acad. Sci. (U.S.A.), 92:684-688 (1995)). Los genes RTX de AMV seleccionados se recuperan mediante PCR utilizando los oligonucleótidos que flanquean el gen clonado. Los productos resultantes de la PCR se someten a RSR in vitro, se seleccionan como se ha descrito más arriba, y se repite el procedimiento hasta que el nivel de expresión funcional es aceptable. Después se escrutan los clones individuales en busca de polimerización de ADN dependiente de ARN y otras propiedades de interés (p. ej. vida media a temperatura ambiente, tasa de errores). Los clones candidatos se someten a mutagénesis, y después se someten a ensayo de nuevo para dar una RT de AMV que

puede ser expresada en *E. coli* a niveles elevados.

2. Chaperoninas Desarrolladas

5 En algunas realizaciones de la invención, la
expresión en exceso de una proteína puede conducir a la
acumulación de intermedios de plegamiento que tienen
tendencia a agregarse. Sin estar limitado por ninguna
teoría, se piensa que el papel de las chaperoninas es
10 estabilizar tales intermedios de plegamiento frente a la
agregación; de este modo, la expresión en exceso de una
proteína de interés puede conducir a vencer la capacidad
de las chaperoninas. Los genes de las chaperoninas pueden
ser desarrollados utilizando las técnicas de la
15 invención, ya sea solos o combinados con los genes que
codifican la proteína de interés, para superar este
problema.

 Los ejemplos de las proteínas de interés que son
20 especialmente adecuadas para este enfoque incluyen, pero
no están limitados a: citoquinas; proteínas de la
envoltura de la malaria; receptores de células T;
anticuerpos; enzimas industriales (p. ej., proteasas
detergentes y lipasas detergentes); proteínas virales
25 para su uso en vacunas; y proteínas de almacenamiento de
semillas de plantas.

 Las fuentes de genes de chaperonina incluyen, pero
no están limitados a, genes de chaperonina de *E. coli* que
30 codifican proteínas tales como la tiorredoxina, Cro
ES/Gro EL, PapD, ClpB, DsbA, DsbB, DnaJ, DnaK, y GrpE;
chaperoninas de mamífero tales como Hsp70, Hsp72, Hsp73,
Hsp40, Hsp60, Hsp10, Hdj1, TCP-1, Cpn60, BiP; y los
homólogos de estos genes de chaperonina en otras especies
35 tales como levadura (J.G. Wall y A. Pluckthun, Current

Biology, 6:507-516 (1995); Hartl, Nature, 381:571-580 (1996)). Adicionalmente, se pueden utilizar genotecas genómicas heterólogas o de ADNc como genotecas para seleccionar o escrutar en busca de chaperoninas novedosas.

En general, la evolución de las chaperoninas se logra mutagenizando primero genes de chaperonina, escrutando o seleccionando en busca de la expresión mejorada de la proteína diana de interés, sometiendo los genes de chaperonina mutada a RSR, y repitiendo la selección o escrutinio. Como en todas las técnicas de RSR, esto se repite hasta obtener la mejora de la expresión deseada de la proteína de interés. Más abajo se proporcionan dos enfoques ilustrativos.

a. Desarrollo de Chaperoninas en Posición Trans con respecto a la Proteína de Interés con un Escrutinio o Selección en busca de una Función Mejorada

En algunas realizaciones se desarrollan los genes de chaperonina independientemente de los genes de la proteína de interés. La mejora en la chaperonina desarrollada puede ser analizada, por ejemplo, mediante escrutinio en busca de la actividad de la propia proteína diana o en busca de la actividad de una proteína de fusión que comprende la proteína diana y una proteína seleccionable o escrutable (p. ej., GFP, fosfatasa alcalina o beta-galactosidasa).

b. Operón de Chaperonina en cis

En algunas realizaciones, los genes de chaperonina y los genes de la proteína diana están codificados en el mismo plásmido, pero no necesariamente se desarrollan

juntos. Por ejemplo, se puede fusionar un dímero de la porción de cabeza de lac con la proteína diana para permitir la selección de plásmidos que codifican una proteína soluble. Los genes de chaperonina son proporcionados en el mismo plásmido ("cis") y son barajados y desarrollados en lugar de la proteína diana. De un modo similar, los genes de chaperonina pueden ser clonados en un plásmido fagémido que codifica una fusión del gen III o el gen VIII con una proteína de interés. Las chaperoninas clonadas se mutagenizan y, como en la selección descrita antes, se aíslan los fagos que presentan funcionalmente la proteína de fusión sobre una matriz de afinidad. Los genes de chaperonina de estos fagos se barajan y se aplica repetitivamente el ciclo de selección, mutación y recombinación hasta que las proteínas de fusión son presentadas eficazmente de forma funcional.

3. Localización Intracelular Mejorada

Muchas proteínas expresadas en exceso de interés biotecnológico son secretadas al periplasma o al medio para proporcionar ventajas en la purificación y en los análisis de actividad. La optimización de la secreción de alto nivel es difícil porque el procedimiento está controlado por muchos genes y por lo tanto la optimización puede requerir múltiples mutaciones que afecten al nivel de expresión y a la estructura de varios de estos componentes. Se sabe que la secreción de proteínas en *E. coli*, por ejemplo, está influenciada por muchas proteínas incluyendo: una ATPasa secretora (SecA), un complejo translocasa (SecB, SecD, SecE, SecF, y SecY), chaperoninas (DnaK, DnaJ, GroES, GroEL), peptidasas señal (LepB, LspA, Ppp), catalizadores de plegamiento específico (DsbA) y otras proteínas con una función peor

definida (p. ej., Ffh, FtsY) (Sandkvist et al., Curr. Op. Biotechnol. 7:505-511 (1996)). La producción en exceso de copias de tipo salvaje o mutantes de estos genes para estas proteínas puede incrementar significativamente el rendimiento de proteína secretada madura. Por ejemplo, la expresión en exceso de secY o secY4 aumentó significativamente el rendimiento periplásmico de IL6 humana madura a partir de una fusión hIL6-pre-OmpA (Perez-Perez et al., Bio-Technology 12:178-180 (1994)). Análogamente, la expresión en exceso de DnaK/DnaJ en *E. coli* mejoró el rendimiento de factor estimulador de colonias de granulocitos humanos secretado (Perez-Perez et al., Biochem. Biophys. Res. Commun. 210:254-259 (1995)).

La RSR proporciona una ruta para el desarrollo de uno o más de los componentes de la ruta secretora nombrados más arriba. Se emplea la siguiente estrategia para optimizar la secreción de proteína en *E. coli*. También se describen variaciones de este método, adecuadas para la aplicación a *Bacillus subtilis*, *Pseudomonas*, *Saccaromyces cerevisiae*, *Pichia pastoris*, células de mamífero y otros anfitriones. El protocolo general es el siguiente.

Se obtienen uno o más de los genes nombrados antes mediante amplificación por PCR de ADN genómico de *E. coli* utilizando una secuencia colindante conocida, y se clonan en disposición ordenada en un plásmido o vector cosmídico. En general estos genes no se encuentran naturalmente en agrupaciones, y por lo tanto estos comprenderán agrupaciones artificiales de genes. Los genes pueden ser clonados bajo el control de su promotor natural o bajo el control de otro promotor tales como los promotores lac, tac, arabinosa, o trp. Típicamente, se

colocan sitios de restricción raros tales como Sfi I entre los genes para facilitar el re-ensamblaje ordenado de los genes barajados como se describe en los métodos de la invención.

5

La agrupación de genes se mutageniza y se introduce en una célula anfitriona en la cual el gen de interés puede ser expresado induciblemente. La expresión del gen diana que se va a secretar y de los genes clonados es inducida por métodos convencionales para el promotor de interés (p. ej., adición de IPTG 1 mM para el promotor lac). La eficacia de la secreción de proteínas por una genoteca de mutantes se mide, por ejemplo, mediante el método de la transferencia de colonias (Skerra et al., Anal. Biochem. 196:151-155 (1991)). Se escogen aquellas colonias que expresan los niveles más elevados de proteína secretada (0,1 - 10% superior; preferiblemente 1% superior). El ADN plasmídico se prepara a partir de estas colonias y se baraja de acuerdo con cualquiera de los métodos de la invención.

10

15

20

25

30

Preferiblemente, cada gen individual se amplifica a partir de la población y se somete a RSR. Los fragmentos se digieren con Sfi I (introducida entre cada gen con salientes no palindrómicos diseñados para promover el re-ensamblaje ordenado por ADN ligasa) y se ligan entre sí, preferiblemente a una dilución baja para promover la formación de círculos relajados cerrados covalentemente (<1 ng/microlitro). Cada una de las poblaciones de genes amplificados por PCR puede ser barajada antes del re-ensamblaje en la agrupación génica final. Los productos de ligación se vuelven a transformar en el anfitrión de interés y se repite el ciclo de selección y RSR.

35

Se pueden emplear estrategias análogas en otros

anfitriones tales como *Pseudomonas*, *Bacillus subtilis*, levadura y células de mamífero. Los homólogos de los genes de *E. coli* enumerados antes son dianas para la optimización, y en efecto muchos de estos homólogos han sido identificados en otras especies (Pugsley, Microb. Rev. 57:50-108 (1993)). Además de estos homólogos, otros componentes tales como los seis polipéptidos de la partícula de reconocimiento de la señal, proteína de membrana asociada al péptido de exportación (TRAM), BiP, las proteínas Ssa y otros homólogos hsp70, y prsA (*B. subtilis*) (Simonen y Pulva, Microb. Rev. 57:109-137 (1993)) son dianas para la optimización por RSR. En general, se utilizan los vectores episómicos replicantes tales como SV40-neo (Sambrook et al., Molecular Cloning, CSH Press (1987), Northrup et al., J. Biol. Chem. 268(4):2917-2923 (1993)) para células de mamífero o los plásmidos de 2 micras o ars para levadura (Strathern et al., The Molecular Biology of the Yeast *Saccharomyces*, CSH Press (1982)). Se prefieren los vectores integrantes tales como pJM 103, pJM 113 o pSGMU2 para *B. subtilis* (Perego, Cap. 42, págs. 615-624 en: *Bacillus subtilis* and Other Gram-Positive Bacteria, A. Sonenshein, J. Hoch, y R. Losick, eds., 1993).

Por ejemplo, se puede desarrollar una ADN polimerasa termoestable secretada eficazmente, permitiendo de ese modo el funcionamiento de los análisis de polimerización de ADN con poca o ninguna purificación de la ADN polimerasa expresada. Semejante procedimiento sería preferido para la expresión de genotecas de mutantes de cualquier proteína que se desee someter a ensayo en un análisis de elevado rendimiento, por ejemplo cualquiera de las proteínas farmacéuticas enumeradas en la Tabla I, o cualquier enzima industrial. Los constructos iniciales se elaboran fusionando un péptido señal tal como el de

STII o OmpA al extremo amino de la proteína que se vaya a secretar. Las agrupaciones génicas de genes clonados que se cree que actúan en la ruta secretora de interés se mutagenizan y se co-expresan con el constructo diana. Los clones individuales se escrutan en busca de la expresión del producto génico. Las agrupaciones de genes secretores de clones mejorados se recuperan y se vuelven a clonar y se introducen de nuevo en el anfitrión original. Preferiblemente, se someten primero a mutagénesis antes de repetir el procedimiento. Este ciclo se repite hasta lograr la mejora deseada en la expresión de la proteína secretada.

IV. Propiedades del Polipéptido Desarrollado

A. Análogo del Estado de Transición Desarrollado y Unión al Sustrato

Existen muchas enzimas de interés industrial que tienen una actividad esencialmente sub-óptima sobre el sustrato de interés. En muchos de estos casos, se requiere que la enzima obtenida de la naturaleza funcione en condiciones que son muy diferentes de las condiciones en las cuales se desarrolló o tenga actividad hacia un sustrato que es diferente del sustrato natural.

La aplicación de las tecnologías evolutivas a enzimas industriales está a menudo limitada significativamente por los tipos de selecciones que se pueden aplicar y el modesto número de mutantes que pueden sobrevivir a los escrutinios. La selección de enzimas o anticuerpos catalíticos, expresados en un formato de presentación, para la unión a análogos del estado de transición (McCafferty et al., Appl. Biochem. Biotechnol. 47:157-171 (1994)) o análogos de sustrato (Janda et al.,

Proc. Natl. Acad. Sci. (U.S.A.) 91:2532-2536, (1994))
representa una estrategia general para seleccionar
mutantes con una eficacia catalítica mejorada.

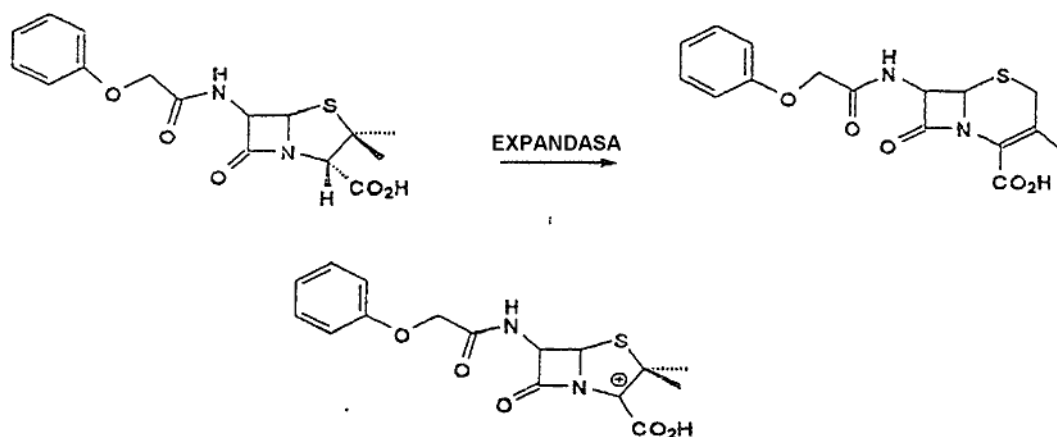
5 La presentación en fagos (O'Neil et al., Current
Biology 5:443-449 (1995) y los otros formatos de
presentación (Gates et al., J. Mol. Biol. 255:373-386
(1995); Mattheakis et al., Proc. Natl. Acad. Sci.
(U.S.A.) 91:9022-9026 (1994)) descritos en la presente
10 memoria representan metodologías generales para aplicar
las selecciones basadas en la afinidad a las proteínas de
interés. Por ejemplo, Matthews and Wells (Science
260:1113-1117 (1993)) han utilizado la presentación en
fagos de un sustrato de proteasa para seleccionar
15 sustratos mejorados. La presentación de enzimas activas
sobre la superficie de un fago, por otra parte, permite
la selección de proteínas mutantes con una unión al
análogo del estado de transición mejorada. Las mejoras en
la afinidad para los análogos del estado de transición se
20 corresponden con mejoras en la eficacia catalítica. Por
ejemplo, Patten et al., Science 271:1086-1091 (1996) han
demostrado que las mejoras en la afinidad de un
anticuerpo catalítico hacia su hapteno se corresponden
bien con las mejoras en la eficacia catalítica,
25 lográndose una mejora de 80 veces en k_{cat}/K_m para un
anticuerpo esterolítico.

 Por ejemplo, se puede desarrollar una enzima
utilizada en la biosíntesis de antibióticos para una
30 especificidad hacia un nuevo sustrato y una actividad en
las condiciones deseadas utilizando selecciones de
presentación en fagos. Actualmente se elaboran algunos
antibióticos mediante modificaciones químicas de
compuestos de partida producidos biológicamente. La
35 biosíntesis completa de las moléculas deseadas es

actualmente poco práctica debido a la carencia de una enzima con la actividad enzimática y la especificidad de sustrato requeridas (Skatrud, TIBTECH 10:324-329, Septiembre 1992). Por ejemplo, el ácido 7-amino-desacetoxicefalosporánico (7-ADCA) es un precursor de las cefalosporinas producidas semi-sintéticamente. El 7-ADCA se elabora mediante expansión química del anillo de penicilina G seguida de desacilación enzimática del grupo fenoxiacetal. El 7-ADCA se puede elaborar enzimáticamente a partir de desacetilcefalosporina C (DAOC V), que a su vez podría estar derivada de la penicilina V por expansión enzimática del anillo si se pudiera desarrollar una penicilina expandasa adecuadamente modificada (Cantwell et al., Curr. Genet. 17:213-221 (1990)). De este modo, el 7-ADCA podría ser producido en principio enzimáticamente a partir de penicilina V utilizando una penicilina N expandasa modificada, tales como formas mutantes del gen *cefE* de *S. clavuligerus* (Skatrud, TIBTECH 10:324-329, Septiembre 1992). No obstante, la penicilina V no es aceptada como sustrato por ninguna expandasa conocida con suficiente eficacia como para ser útil comercialmente. Como se esboza más abajo, se pueden utilizar las técnicas de RSR de la invención para desarrollar la penicilina expandasa codificada por *cefE* u otras expandasas de manera que utilicen penicilina V como sustrato.

La presentación en fagos u otras selecciones de formatos de presentación se aplican a este problema expresando genotecas de mutantes de penicilina expandasa de *cefE* en un formato de presentación, seleccionando la unión a sustratos o análogos de estados de transición, y aplicando la RSR para desarrollar rápidamente aglutinadores de alta afinidad. Los candidatos se escrutan adicionalmente para identificar mutantes con una

actividad enzimática mejorada sobre la penicilina V en las condiciones de reacción deseadas, tales como el pH, la temperatura, la concentración de disolvente, etc. La RSR se aplica para desarrollar adicionalmente mutantes con la actividad expandasa deseada. Numerosos análogos del estado de transición (TSA) son adecuados para esta reacción. La siguiente estructura es el TSA inicial que se utiliza para la selección de la genoteca de presentación de mutantes de cefE:



Se elaboran genotecas de las penicilina expandasas conocidas (Skatrud, TIBTECH 10:324-329(1992); Cantwell et al., Curr. Genet. 17:13-221 (1990)) como se describe en la presente memoria. La genoteca de presentación se somete a selección en busca de unión a penicilina V y/o un análogo del estado de transición dado antes para la conversión de penicilina V en DAOC V. Estas selecciones de unión se pueden realizar en condiciones de reacción no fisiológicas, tales como una temperatura elevada, para obtener mutantes que son activos en las nuevas condiciones. La RSR se aplica para desarrollar mutantes con una mejora de $2 - 10^5$ veces de la afinidad de unión para el ligando seleccionado. Cuando se ha obtenido el nivel de mejora de la unión deseado, se expresan los mutantes candidato en un formato de alto rendimiento y se

mide cuantitativamente la actividad específica para expandir la penicilina V a DAOC V. Los recombinantes con una actividad enzimática mejorada se mutagenizan y se repite el procedimiento para desarrollarlos más.

5

La retención de la unión de TSA por una enzima presentada (p. ej., presentación en fagos, dímero de la porción de cabeza de lac, presentación en polisomas, etc.) es una buena selección para la retención de la integridad global del sitio activo y por lo tanto puede ser explotada para seleccionar en busca de mutantes que conserven la actividad en condiciones de interés. Tales condiciones incluyen, pero no están limitadas a: óptimos de pH diferentes, óptimos de pH mas amplios, actividad en disolventes alterados tales como DMSO (Seto, et al., DNA Sequence 5:131-140 (1995)) o formamida (Chen et al., Proc. Natl. Acad. Sci. (U.S.A.) 90:5618-5622, (1993)) temperatura alterada, vida útil mejorada, especificidad de sustrato alterada o ampliada, o resistencia a proteasas. Más abajo se proporciona un ejemplo adicional, el desarrollo de p-nitrofenil esterasa, utilizando un formato de presentación en mamíferos.

10

15

20

B. Mejora de las ADN y ARN Polimerasas

25

30

35

Tienen una particular importancia comercial las polimerasas mejoradas para su uso en secuenciación de ácidos nucleicos y en reacciones en cadena de la polimerasa. Las siguientes propiedades son candidatos atractivos para la mejora de una polimerasa de secuenciación de ADN: (1) supresión de la terminación por inosina en un formato con cebadores marcados (H. Dierick et al., Nucleic Acids Res. 21:4427-4428 (1993)) (2) alturas de picos más normalizadas, especialmente con terminadores didesoxi marcados fluorescentemente (Parker

et al., BioTechniques 19:116-121 (1995)), (3) mejor
secuenciación de ADN con un elevado contenido en GC (>60%
GC), por ejemplo, tolerando >10% de DMSO (D. Seto et al.,
DNA Sequence 5:131-140 (1995); Scheidl et al.,
5 BioTechniques 19(5):691-694 (1995)), o (4) mejor
aceptación de los análogos de bases novedosos tales como
inosina, 7-desaza dGTP (Dierick et al., Nucleic Acids
Res. 21:4427-4428 (1993)) u otros análogos de bases
novedosos que mejoran las propiedades anteriores.

10 Se han descrito formatos de secuenciación novedosos
que utilizan espectroscopía de masas por ionización
asistida por una matriz por desorción con láser y
análisis por tiempo de vuelo (MALDI-TOF) para resolver
15 escalas didesoxi (Smith, Nature Biotechnology 14:1084-
1085 (1996)). Se observa en la revisión reciente de Smith
que la fragmentación del ADN es el rasgo singular que
limita el desarrollo de este método como alternativa
viable a la electroforesis en gel convencional para la
20 secuenciación del ADN. Los análogos de bases que
estabilizan el enlace N-glicosídico por medio de
modificaciones de las bases de purina a análogos 7-desaza
(Kirpekar et al., Rapid Comm. in Mass Spec. 9:525-531
(1995)) o del hidroxilo 2' (tal como 2'-H o 2'-F)
25 "mitigan enormemente la limitación del intervalo de masa"
de esta técnica (Smith, 1996). De este modo, las
polimerasas desarrolladas que pueden incorporar
eficazmente estos y otros análogos de bases que confieren
resistencia a la fragmentación en condiciones MALDI-TOF
30 son innovaciones valiosas.

Otras propiedades de interés de las polimerasas para
la mejora mediante RSR son la ADN polimerasa termoestable
de baja fidelidad para una mutagénesis más eficaz o como
35 una correlación útil para la aceptación de los análogos

de bases con los fines descritos más arriba; la polimerasa para PCR de alta fidelidad (Lundberg et al., Gene 108:1-6 (1991)); la transcriptasa inversa de mayor fidelidad para vehículos de terapia génica retroviral para reducir la mutación del constructo terapéutico y del retrovirus; la PCR mejorada para ADN rico en GC y PCR con bases modificadas (S. Turner y F. J. Jenkins, BioTechniques 19(1):48-52 (1995)).

De este modo, en algunas realizaciones de la invención, las genotecas de genes de polimerasa mutantes son escrutadas mediante escrutinio de alto rendimiento en busca de propiedades de secuenciación mejoradas. Los mejores candidatos se someten después a RSR. En resumen, se construyen genotecas de mutantes de polimerasas candidato tales como polimerasa Taq utilizando métodos convencionales tales como mutagénesis por PCR (Caldwell et al., PCR Meth. App. 2:28-33 (1992)) y/o mutagénesis por inserción de casete (Sambrook et al., Molecular Cloning, CSH Press (1987)). La incorporación de mutaciones en la ADN polimerasa Taq tal como el residuo del sitio activo de la polimerasa T7 que mejora la aceptación de los didesoxinucleótidos (Tabor y Richardson, J. Biol. Chem. 265:8322-8328 (1990)) y las mutaciones que inactivan la actividad exonucleasa 5' - 3' (R.S. Rano, BioTechniques 18:390-396 (1995)) son incorporadas a estas genotecas. La técnica de PCR de re-ensamblaje, por ejemplo, como se ha descrito antes es especialmente adecuada para este problema. De un modo similar, las genotecas de polimerasas quiméricas se elaboran cultivando las polimerasas termófilas existentes, secuenasa, y@"poll" de *E. coli* entre sí utilizando los métodos de oligonucleótidos puente descritos más arriba. Las genotecas se expresan en formatos en los que se utiliza la selección de colonias humanas o robóticas para

replicar las colonias individuales escogidas sobre placas de 96 pocillos donde se hacen crecer pequeños cultivos, y se induce la expresión de la polimerasa.

5 Se realiza una simple purificación a pequeña escala, de alto rendimiento para la polimerasa expresada en cada pocillo. Por ejemplo, se han descrito purificaciones de una única etapa simples de Taq etiquetada con His expresada en *E. coli* (Smimov et al., Russian J. Bioorganic Chem. 21(5):341-342 (1995)), y se podrían
10 adaptar fácilmente a un formato de expresión y purificación en 96 pocillos. Los datos se analizan para identificar los mutantes con propiedades de secuenciación mejoradas, de acuerdo con cualquiera de estos criterios:
15 escalas de mayor calidad en los moldes ricos en GC, especialmente más del 60% de GC, incluyendo puntos tales como menos productos de terminación artefactual y señales más fuertes que las dadas con la enzima de tipo salvaje; menos terminación de reacciones por inosina en reacciones
20 marcadas con cebador, p. ej., cebadores marcados con fluorescencia; menor variación en la incorporación de señales en las reacciones con desoxinucleótidos fluorescentes en cualquier posición dada; escalas de secuenciación más largas que las obtenidas con la enzima
25 de tipo salvaje, por ejemplo aproximadamente 20 a 100 nucleótidos; mejor aceptación de otros análogos de bases conocidos tales como 7-desaza-purinas; mejor aceptación de nuevos análogos de bases de genotecas de química combinatoria (Véase, por ejemplo, Hogan, Nature
30 384(Sup):17-1996).

Los mejores candidatos se someten después a mutagénesis, y después se seleccionan o escrutan en busca de las propiedades mejoradas descritas más arriba.

En otra realización, se realiza un escrutinio o selección como sigue. La replicación de un plásmido se puede colocar bajo el control obligado de una polimerasa expresada en *E. coli* u otro microorganismo. La eficacia de este sistema ha sido demostrada para realizar la replicación de plásmidos dependientes de polimerasa beta de mamíferos (Sweasy et al., Proc. Natl. Acad. Sci. (U.S.A.) 90:4626-4630, (1993)), polimerasa Taq (Suzuki et al., Proc. Natl. Acad. Sci. (U.S.A.) 93:9670-9675 (1996)), o transcriptasa inversa de VIH (Kim et al., Proc. Natl. Acad. Sci. (U.S.A.) 92:684-688 (1995)). El gen de polimerasa mutante se coloca en un plásmido que porta un origen colE1 y se expresa bajo el control de un promotor de arabinosa. La genoteca es enriquecida en polimerasas activas esencialmente como describen Suzuki et al., (cita anterior), siendo inducida la expresión de las polimerasas por la presencia de arabinosa en el cultivo.

Un escrutinio cuantitativo adicional utiliza la presencia de GFP (proteína fluorescente verde) sobre el mismo plásmido, cultivando en placa las réplicas sobre arabinosa a la temperatura no permisiva en ausencia de un antibiótico selectivo, y utilizando un fluorómetro para medir cuantitativamente la fluorescencia de cada cultivo. La actividad de GFP se corresponde con la estabilidad del plásmido y el número de copias que a su vez es dependiente de la expresión de la polimerasa nativa.

Una polimerasa con una tasa de error muy elevada sería una enzima de secuenciación superior, ya que tendría una señal más normalizada para la incorporación de análogos de bases tales como los didesoxi marcados fluorescentemente utilizados corrientemente puesto que tendrían una especificidad y una selectividad reducidas.

Las tasas de error de las polimerasas utilizadas corrientemente son del orden de 10^{-5} a 10^{-6} , órdenes de magnitud inferiores a los que se pueden detectar dado el poder de resolución de los sistemas de gel. Una tasa de error del 1%, y posiblemente tan alta como el 10%, podría no ser detectada por los sistemas en gel actuales, y de este modo hay un gran margen de oportunidad para aumentar la "falta de sistema" de la enzima. Una polimerasa de ciclación propensa a errores tendría otros usos tales como la hipermutagénesis de genes mediante PCR.

En algunas realizaciones, se utiliza el sistema descrito por Suzuki (Suzuki et al., Proc. Natl. Acad. Sci. (U.S.A.) 96:9670-9675 (1996)) para realizar la replicación de un plásmido informador dependiente de la polimerasa expresada. Este sistema sitúa la replicación de las primeras 200 - 300 bases cerca del origen ColE1 directamente bajo el control de la polimerasa expresada (Sweasy and Loeb, J. Bact. 177:2923-2925 (1995); Sweasy et al., Proc. Natl. Acad. Sci. (U.S.A.) 90:4626-4630 (1993)). Se coloca en esta región un gen informador escrutable o seleccionable que contiene codones de terminación, tal como uno que contiene LacZ alfa que contiene uno, dos o tres codones de terminación. Los constructos se hacen crecer sobre arabinosa a la temperatura no permisible, se deja que se recuperen, y se cultivan en placa sobre medio mínimo con lactosa selectiva que demanda la reversión de los codones de terminación en el casete informador. Las polimerasas mutantes se recuperan de los supervivientes mediante PCR. Los supervivientes se enriquecen en mutadores debido a que su fenotipo mutador aumenta la tasa de reversión de los codones de terminación en el fragmento informador lacZ alfa.

Los genes de polimerasa de los supervivientes se someten a RSR, después los mutantes de polimerasa se re-transforman en la cepa indicadora. Los mutadores pueden ser escrutados visualmente cultivando en placa sobre
5 placas de arabinosa/Xgal a la temperatura no permisiva. Las polimerasas mutadoras darán lugar a colonias con una elevada frecuencia de papilas azules debido a la reversión de los codones de terminación. Los formadores de papilas candidato pueden ser escrutados de nuevo
10 escogiendo una región no formadora de papilas de las colonias con una formación de papilas más densa (esto es, las "mejores" colonias) y volviendo a cultivar en placa sobre el medio indicador de arabinosa/Xgal para escrutar adicionalmente en busca de colonias con tasas aumentadas
15 de formación de papilas. Estas etapas se repiten hasta lograr una tasa de reversión deseada (p. ej., de 10^{-2} a 10^{-3} mutaciones por par de bases por replicación).

Las colonias que muestran una elevada frecuencia de
20 formación de papilas son candidatos para codificar una polimerasa propensa a errores. Estos candidatos se escrutan en busca de propiedades de secuenciación mejoradas esencialmente como en el escrutinio de alto rendimiento descrito más arriba. En resumen, las
25 proteínas Taq mutantes se expresan y purifican en un formato de 96 pocillos. Las proteínas purificadas se utilizan en reacciones de secuenciación y los datos de la secuencia se analizan para identificar mutantes que muestran las mejoras esbozadas en la presente memoria.
30 Los mutantes con propiedades mejoradas se someten a RSR y se vuelven a escrutar en busca de mejoras adicionales en su función.

En algunas realizaciones, se utiliza para la
35 construcción GFP que contiene codones de terminación en

lugar de lacZ alfa con codones de terminación. Las células con codones de terminación revertidos en GFP se seleccionan mediante un clasificador celular activado por fluorescencia (FACS). En general, la selección por FACS se realiza seleccionando las más brillantes aproximadamente 0,1 - 10%, preferiblemente el 0,1 a 1% superior, y se recogen de acuerdo con un protocolo similar al de Dangl et al., (Cytometry 2(6):395-401 (1982)). En otras realizaciones, el gen polA se flanquea con sitios lox u otras dianas de una recombinasa de sitio específico. La recombinasa es inducida, permitiendo de este modo suprimir induciblemente el gen polA (Mulbery et al., Nucleic Acid Res. 23:485-490 (1995)). Esto permitiría realizar selecciones de "tipo Loeb" a cualquier temperatura y en cualquier anfitrión. Por ejemplo, se podría ajustar dicha selección en un mesófilo o termófilo carente de recA colocando el homólogo de polA en un formato suprimible de manera inducible y de este modo aplicar la selección en busca de polimerasa activa en condiciones más generales.

En realizaciones adicionales, se prefiere este sistema general para la mutagénesis de genes in vivo dirigida. El gen diana se clona en la región próxima a un origen de replicación plasmídico que coloca su replicación obligatoriamente bajo el control de la polimerasa propensa a errores. El constructo se hace pasar a través de una cepa polA(ts) recA y se hace crecer a la temperatura no permisiva, mutagenizando de ese modo específicamente el gen diana a la vez que replica el resto del plásmido con una elevada fidelidad.

En otras realizaciones, la selección se basa en la capacidad de las ADN polimerasas mutantes para amplificar por PCR ADN en condiciones alteradas o utilizando

análogos de bases. Las polimerasas mutantes actúan sobre el molde que las codifica en una amplificación por PCR, replicando de ese modo esas polimerasas diferencialmente.

5 En resumen, se realiza una replicación en placa de las genotecas iniciales de mutantes. Las preparaciones de polimerasa se realizan en un formato de 96 pocillos. Las preparaciones de plásmido bruto se elaboran del mismo grupo. Cada preparación de plásmido es amplificada
10 mediante PCR utilizando la preparación de polimerasa derivada de ese plásmido en las condiciones para las cuales se desea optimizar la polimerasa (p. ej., con adición de DMSO o formamida, temperatura de
15 desnaturalización o extensión alterada, sales tampón alteradas, PCR con análogos de bases tales como a-tiol dNTP para su uso con la secuenciación por espectroscopía de masas, PCR de ADN rico en GC (>60% GC), PCR con análogos de bases novedosos tales como 7-desazapurinas, 2' fluoro dNTP, rNTP, PCR con inosina, etc.). Los genes
20 amplificados se reúnen, se clonan, y se someten a mutagénesis, y el procedimiento se repite hasta lograr una mejora.

25 C. Fosfonatasa Desarrollada

25 La fosfatasa alcalina es una enzima informadora ampliamente utilizada para los análisis ELISA, los análisis de fusión de proteínas, y en forma secretada como gen informador para células de mamífero. La
30 capacidad química de los sustratos de p-nitrofenil fosfato (pNPP) y la existencia de fosfatasas celulares que presentan reacción cruzada con pNPP son una importante limitación de la sensibilidad de los análisis utilizando este gen informador. Se puede desarrollar un
35 gen informador con una señal superior para las

propiedades de ruido basándose en la hidrólisis de p-nitrofenil fosfonatos, que son bastante más estables a la hidrólisis catalizada por álcalis que los correspondientes fosfatos. Adicionalmente, existen bastantes menos fosfonatasas de origen natural que fosfatasas alcalinas. De este modo una p-nitrofenil fosfonatasa es una sustitución atractiva para una fosfatasa alcalina porque el fondo debido a la hidrólisis química y enzimática es mucho menor. Esto permitirá realizar ELISA más sensibles para detectar concentraciones muy pequeñas de antígeno.

Chen et al. (J. Mol. Biol. 234:165-178 (1993)) han demostrado que una beta-lactamasa de *Staph. aureus* puede hidrolizar ésteres p-nitrofenil fosfonato con cinéticas de recambio único. El sitio activo Ser70 (el sitio activo nucleófilo para la hidrólisis de beta-lactama) forma un intermedio covalente con el sustrato. Esto es análogo a la primera etapa en la hidrólisis de las beta lactamas, y esta enzima puede ser desarrollada por medio de RSR para hidrolizar fosfonatos mediante un mecanismo análogo a la hidrólisis de beta lactamas. Metcalf y Wanner han descrito un fosfonato críptico utilizando el operón (phn) en *E. coli*, y han construido cepas que portan delecciones del operón phn (J. Bact. 175:3430-3442 (1993)). Esta publicación describe selecciones para el crecimiento de *E. coli* sobre medio mínimo sin fosfato donde el fósforo deriva de la hidrólisis de fosfonatos de alquilo por genes en el operón phn. De este modo, se podrían seleccionar p-nitrofenil fosfonatasas desarrolladas que sean activas utilizando selecciones bioquímicas en medios mínimos definidos. Específicamente, una fosfonatasa eficaz se desarrolla como sigue. Se construye una genoteca de mutantes de la beta lactamasa de *Staph. aureus* o de una de las enzimas phn de *E. coli*. La

genoteca se transforma en mutantes de *E. coli* donde el operón *phn* ha sido suprimido, y se seleccionan en busca del crecimiento sobre medio mínimo MOPS libre de fosfato que contiene p-nitrofenil fosfonato. Se aplica la RSR a los mutantes seleccionados para desarrollar adicionalmente la enzima para una mejor hidrólisis de los p-nitrofenil fosfonatos.

D. Proteasas Detergentes Desarrolladas

Se añaden proteasas y lipasas en grandes cantidades a detergentes para degradar enzimáticamente manchas de proteínas y lípidos en la ropa. La incorporación de estas enzimas a detergentes ha reducido significativamente la necesidad de tensioactivos en los detergentes con la consiguiente reducción en el coste de formulación de detergentes y la mejora en las propiedades de eliminación de manchas. Las proteasas con una actividad específica mejorada, un intervalo mejorado de especificidad de sustrato proteico, una vida útil mejorada, una estabilidad mejorada a temperatura elevada, y requerimientos reducidos de tensioactivos añadirían valor a estos productos.

Como ejemplo, se ha desarrollado la subtilisina como sigue. El gen de subtilisina clonado (von der Osten et al., J. Biotechnol. 28:55-68 (1993)) puede ser sometido a RSR utilizando selecciones de crecimiento en medios de proteínas complejos en virtud de que la subtilisina secretada degrada la mezcla de proteínas compleja. Más específicamente, se construyen genotecas de mutantes de subtilisina en un vector de expresión que dirige la proteína mutante a ser secretada por *Bacillus subtilis*. Los anfitriones de *Bacillus* transformados con las genotecas se hacen crecer en medio mínimo con una

formulación de proteína compleja como fuente de carbono y/o nitrógeno. Los genes de subtilisina son recuperados de cultivadores rápidos y sometidos a RSR, después son escrutados en busca de mejoras en la propiedad deseada.

5

E. Liberación de Fagos de una "Red de Proteína"

En algunas realizaciones, la selección de proteasas mejoradas se realiza como sigue. Se construye una
10 genoteca de genes de proteasa mutante sobre un fago de presentación y el fago se hace crecer en un formato de múltiples pocillos o sobre placas. Los fagos se revisten con una "red de proteína" que atrapa los fagos. La red puede consistir en una proteína o proteínas diseñadas con
15 disulfuros de superficie y después se entrecruza con una genoteca de conectores peptídicos. En una realización adicional se emplea una matriz auxiliar para atrapar adicionalmente el fago. Los fagos se incuban adicionalmente, después se lavan para recoger los fagos
20 liberados en los que la proteasa presentada era capaz de liberar el fago de la red de proteína. Los genes de proteasa se someten después a RSR para su desarrollo adicional. En una realización adicional se emplea una genoteca de proteasas codificadas por, pero no
25 presentadas en un fagémido donde la estreptavidina está fusionada a una píldora por medio de un conector peptídico. La genoteca de mutantes de proteasa se desarrolla para escindir el conector seleccionando los fagémidos sobre una columna de biotina entre rondas de
30 amplificación.

En una realización adicional, la proteasa no se proporciona necesariamente en un formato de presentación. Las células anfitrionas secretan la proteasa codificada
35 por, pero no presentada sobre, la superficie por un

fagémido, mientras está confinado en un pocillo, por ejemplo, en una placa de microtitulación. Se prefiere el formato de presentación en fagos donde un producto lisado de titulación elevada está encerrado en una matriz de red de proteína, y el fago expresa proteasas activas y de amplia especificidad que digieren la matriz para ser liberado en la siguiente ronda de amplificación, mutagénesis, y selección.

En una realización adicional, los fagos no están confinados en un pocillo, pero, en vez de eso, se utilizan filtros de unión a la proteína para elaborar una colonia de transferencia de placas y se escrutan en busca de actividad con sustratos cromogénicos o fluorogénicos. Las colonias o placas correspondientes a aplicaciones positivas sobre los filtros se recogen y se recuperan los genes de proteasa codificados, por ejemplo mediante PCR. Los genes de proteasa se someten después a RSR para su desarrollo adicional.

F. Escrutinios en busca de una Actividad Proteasa Mejorada

Los sustratos peptídicos que contienen fluoróforos anclados al extremo carboxi y radicales extintores de la fluorescencia en el extremo amino, tales como los descritos por Holskin, et al., (Anal. Biochem. 227:148-55 (1995)) (p. ej., (ácido 4,4'-dimetilaminofenazo)benzoyl-arg-gly-val-val-asn-ala-ser-ser-arg-leu-ala-5-(2'-aminoetil)-amino-naftaleno-1-sulfónico) se utilizan para escrutar mutantes de proteasa en busca de una especificidad ampliada o alterada. En resumen, se diseña una genoteca de sustratos peptídicos con un fluoróforo en el extremo amino y un potente extintor de la fluorescencia en el extremo carboxi, o viceversa. Los

sobrenadantes que contienen proteasas secretadas se incuban o bien por separado con diferentes miembros de la genoteca o bien con un cóctel complejo. Aquellas proteasas que son muy activas y tienen una amplia especificidad escindirán la mayoría de los péptidos, liberando de se modo el fluoróforo del extintor y dando una señal positiva en el fluorómetro. Esta técnica es susceptible de un formato de múltiples pocillos de alta densidad.

G. Mejora de proteínas farmacéuticas utilizando RSR

En la Tabla I se enumeran proteínas que tienen un particular interés comercial para la industria farmacéutica. Estas proteínas son todas candidatos al desarrollo mediante RSR para mejorar su función, tal como la actividad específica, la unión al ligando, la vida útil, la reducción de los efectos secundarios por medio de una especificidad intensificada, etc. Todas están bien adaptadas a la manipulación mediante las técnicas de la invención. Las realizaciones adicionales especialmente aplicables a esta lista se describen más abajo.

Primero, se aplican métodos de alto rendimiento para expresar y purificar genotecas de proteínas mutantes, similares a los métodos descritos antes para la polimerasa Taq, a las proteínas de la Tabla I. Estos mutantes se escrutan en busca de actividad en un análisis funcional. Por ejemplo, los mutantes de IL2 se escrutan en busca de resistencia a la degradación por las proteasas del plasma o tisulares; o en busca de la retención de actividad sobre el receptor de IL12 de baja afinidad pero con pérdida de actividad sobre el receptor de IL2 de alta afinidad. Los genes de mutantes con una actividad mejorada con respecto al tipo salvaje se

recuperan, y se someten a RSR para mejorar el fenotipo adicionalmente.

5 Preferiblemente, las genotecas se generan en un
formato de presentación de manera que la proteína plegada
madura se conecta físicamente a la información genética
que la codifica. Los ejemplos incluyen la presentación en
fagos utilizando fagos filamentosos (O'Neil et al.,
Current Biology 5:443-449 (1995)) o la presentación del
10 gen V en bacteriófago lambda (Dunn, J. Mol. Biol.
248:497-506 (1995)), péptidos sobre plásmidos (Gates et
al., J. Mol. Biol. 255:373-386 (1995)) donde el
polipéptido de interés se fusiona a un dímero de la
porción de cabeza de lac y el producto de traducción
15 naciente se une a un sitio del operador lac codificado
sobre el plásmido o producto de PCR, y la presentación en
polisomas (Mattheakis et al., Proc. Natl. Acad. Sci.
(U.S.A.) 91:9022-9026 (1994)) donde los ribosomas están
paralizados sobre moléculas de ARNm de manera que el
20 polipéptido naciente está expuesto para su interacción
con ligandos cognados sin interrumpir el complejo de
ribosoma paralizado/ARNm. Los complejos seleccionados se
someten a RT-PCR para recuperar los genes.

25 Cuando se presentan de este modo, a menudo se
realiza la unión por afinidad de los fagos recombinantes
utilizando un receptor para la proteína de interés. En
algunos casos es poco práctico obtener un receptor
purificado con retención de todas las características
30 biológicas deseadas (por ejemplo, receptores 7-
transmembrana (7-TM)). En tales casos, se podrían
utilizar células que expresan el receptor como sustrato
de la inmunopurificación. Por ejemplo, Barry et al. (Nat.
Med. 2:299-305 (1996)) han descrito una
35 inmunopurificación satisfactoria de genotecas de M13

frente a células completas para obtener fagos que se unen a células que expresan un receptor de interés. Este formato se podría aplicar generalmente a cualquiera de las proteínas enumeradas en la Tabla I.

5

En algunas realizaciones, se puede utilizar el siguiente método para la selección. Se puede utilizar una solución de partida de fagémidos que codifican mutantes de IFN alfa, por ejemplo, directamente a una dilución adecuada para estimular las células. El efecto biológico sobre las células puede ser leído mediante análisis convencionales (p. ej., proliferación o resistencia viral) o indirectamente a través de la activación de un gen informador tal como GFP (Crameri et al., Nat. Med. 14:315-.319 (1996)) bajo el control de un promotor sensible a IFN, tal como un promotor de clase I. En una realización, los fagémidos que quedan anclados después de la estimulación, la expresión y la purificación mediante FACS de las células sensibles, pueden ser purificados mediante FACS. Preferiblemente, se recogen las células más brillantes. Los fagémidos se recogen y su ADN se somete a RSR hasta lograr el nivel de mejora deseado.

10

15

20

25

30

35

De este modo, por ejemplo, se prepara IL-3 en uno de estos formatos de presentación y se somete a RSR para desarrollar un agonista con un nivel de actividad deseado. Se crea una genoteca de mutantes de IL3 sobre un vector de fago filamentoso y se selecciona por afinidad ("inmunopurificado") frente a receptor de IL3 purificado para obtener mutantes con una afinidad por el receptor mejorada o por IL-3 presentada con una potencia de fase mejorada. Los genes IL-3 mutantes se recuperan mediante PCR, se someten a RSR, y se vuelven a clonar en el vector de presentación. El ciclo se repite hasta lograr la afinidad o la actividad del agonista deseadas.

Muchas proteínas de interés son expresadas como dímeros o formas multiméricas de orden superior. En algunas realizaciones, se aplican preferentemente los formatos de presentación descritos antes a una versión de cadena sencilla de la proteína. Se puede utilizar la mutagénesis, tal como RSR, en estos formatos de presentación para desarrollar derivados de factores multiméricos de cadena sencilla mejorados que inicialmente tienen una actividad baja pero detectable. Esta estrategia se describe con mayor detalle más abajo.

H. Selecciones de Células Completas

En algunas realizaciones, la célula eucariótica es la unidad de selección biológica. Se puede utilizar el siguiente protocolo general para aplicar la RSR a la mejora de proteínas utilizando células eucarióticas como unidad de selección: (1) transfección o transducción de genotecas de mutantes en una célula anfitriona adecuada, (2) expresión de los productos génicos codificados transitoriamente o establemente, (3) selección funcional de células con un fenotipo mejorado (expresión de un receptor con afinidad mejorada por un ligando diana; resistencia viral, etc., (4) recuperación de los genes mutantes, por ejemplo, mediante PCR seguida de preparación de sobrenadantes HIRT con posterior transformación de *E. coli*, (5) RSR y (6) repetición de las etapas (1) - (5) hasta lograr el grado de mejora deseado.

Por ejemplo, un trabajo previo ha demostrado que se puede utilizar la presentación en la superficie de mamíferos para seleccionar funcionalmente las células que expresan los genes clonados, utilizando por ejemplo un anticuerpo para clonar el gen en busca de la proteína

expresada en la superficie (Revisado por Seed, Curr. Opin. Biotechnol. 6:567-573 (1995)). En resumen, las células son transfectadas transitoriamente con genotecas de genes clonados que residen en vectores episómicos replicantes. Se inmoviliza un anticuerpo dirigido contra la proteína de interés (cuyo gen se desea clonar) sobre una superficie sólida tal como una fuente de plástico, y las células transfectadas que expresan la proteína de interés se seleccionan por afinidad.

Por ejemplo, se puede mejorar la afinidad de un anticuerpo por un ligando utilizando la presentación en la superficie de mamíferos y la RSR. Los anticuerpos con una afinidad mayor por sus ligandos cognados son escrutados después en busca de la mejora de una o más de las siguientes propiedades: (1) propiedades terapéuticas mejoradas (incremento de la eliminación de células, neutralización de ligandos, activación de rutas de transducción de la señal por receptores de entrecruzamiento), (2) aplicaciones de imagen in vivo mejoradas (detección del anticuerpo mediante unión covalente/no covalente de un radionúclido o cualquier agente detectable fuera del organismo mediante medios no invasivos, tales como RMN), (3) aplicaciones analíticas mejoradas (detección de proteínas o moléculas pequeñas mediante ELISA), y (4) catalizadores mejorados (anticuerpos catalíticos). Los métodos descritos son generales y pueden ser ampliados a cualquier par receptor-ligando de interés. En la sección experimental se proporciona un ejemplo específico.

El uso del protocolo de una secuencia mutante-una célula transfectada es un rasgo de diseño preferido para los protocolos basados en RSR porque el punto consiste en utilizar la selección funcional para identificar mutantes

con fenotipos mejorados y, si la transfección no se realiza de un modo "clónico", el fenotipo funcional de cualquier célula dada es el resultado de la suma de múltiples secuencias transfectadas. La fusión de protoplastos es un método para lograr este fin, puesto que cada protoplasto contiene típicamente más de 50 copias de cada variante de plásmido individual. No obstante, es un procedimiento de eficacia relativamente baja (aproximadamente 10^3 - 10^4 transfectantes), y no funciona bien en algunas líneas celulares no adherentes tales como las células B. Los vectores retrovirales proporcionan una segunda alternativa, pero están limitados en el tamaño del inserto aceptable (<10 kb) y a veces es difícil lograr niveles de expresión elevados, constantes. La integración al azar da como resultado niveles de expresión variables, introduciendo de ese modo ruido y limitando la capacidad para distinguir entre las mejoras en la afinidad de la proteína mutante vs aumento de expresión. Una clase de estrategias relacionadas que se puede utilizar para lograr la transferencia de ADN "un gen-una célula" y niveles de expresión constantes para RSR consiste en utilizar un vector viral que contiene un sitio lox e introducir este en un anfitrión que expresa la recombinasa cre, preferiblemente de manera transitoria, y contenga uno o más sitios lox integrados en su genoma, limitando de ese modo la variabilidad de los sitios de integración (Rohlman et al. Nature Biotech. 14:1562-1565 (1996)).

Una estrategia alternativa consiste en transfectar con concentraciones limitantes de plásmido (esto es, aproximadamente una copia por célula) utilizando un vector que pueda replicar en las células diana, como en el caso de los plásmidos que portan orígenes de SV40 transfectados en células COS. Esta estrategia requiere

que la célula anfitriona o el vector suministren un factor de replicación tal como el antígeno T grande de SV40. Northrup et al. (J. Biol. Chem. 268:2917-2923 (1993)) describen una estrategia donde un transfectante estable que expresa el antígeno T grande de SV40 es transfectado después con vectores que portan orígenes de SV40. Este formato proporcionó una expresión transitoria consistentemente superior y una replicación plasmídica demostrable, analizada mediante la sensibilidad de la digestión por medio de Dpn I. La expresión transitoria (esto es, plásmidos no integrantes) es el formato preferido para las selecciones de presentación celular porque reduce el tiempo del ciclo e incrementa el número de mutantes que se pueden escrutar.

La expresión del antígeno T grande de SV40 u otros factores de replicación puede tener efectos deletéreos o pueden funcionar de modo ineficaz en algunas células. En tales casos, se aplica la RSR al propio factor de replicación para desarrollar mutantes con una actividad mejorada en el tipo de célula de interés. Un protocolo genérico para desarrollar semejante factor es el siguiente:

La célula diana es transfectada con GFP clonada en un vector que contiene el antígeno T grande de SV40, un origen de replicación de SV40, y un gen informador tal como GFP; un formato relacionado es la cotransfección con cantidades limitantes del vector de expresión del antígeno T grande de SV40 y un exceso de un informador tal como GFP clonada en un plásmido que contiene un origen de SV40. Típicamente después de 1-10 días de expresión transitoria, se purifican las células más brillantes mediante FACS. Los mutantes para el antígeno T grande de SV40 son recuperados mediante PCR, y sometidos a mutagénesis. El ciclo se repite hasta obtener el nivel

de mejora deseado.

I. Selección Autocrina

5 En algunas realizaciones, se seleccionan o se
escrutan proteínas mutantes basándose en su capacidad
para ejercer un efecto biológico de un modo autocrino
sobre la célula que expresa la proteína mutante. Por
ejemplo, se puede seleccionar una genoteca de genes de
10 interferón alfa en busca de la inducción de una actividad
antiviral más potente o más específica como sigue. Se
genera una genoteca de mutantes de interferón alfa en un
vector que permite la inducción de la expresión (esto es
bajo el control de un promotor de metalotioneína) y la
15 secreción eficaz en un formato de múltiples pocillos (96
pocillos por ejemplo) con uno o unos pocos clones
independientes por pocillo. En algunas realizaciones, el
promotor no es inducible, y puede ser constitutivo.

20 Se induce la expresión de los genes de interferón
clonados. Las células se sensibilizan con un virus
citotóxico contra el cual se desea desarrollar un
interferón optimizado (por ejemplo virus de la
estomatitis vesicular o VIH). Se recuperan las células
25 supervivientes. Los genes de interferón clonados se
recuperan mediante amplificación por PCR, se someten a
RSR, y se vuelven a clonar en el vector de transfección y
se retransfectan a las células anfitrionas. Estas etapas
se repiten hasta que se desarrolla el nivel de actividad
30 antiviral deseado.

 En algunas realizaciones, el virus de interés no es
fuertemente citotóxico. En este caso se clona en el virus
un gen condicionalmente letal, tal como la timidina
35 quinasa del virus del herpes simple humano y después de

la sensibilización con el virus y la recuperación, se aplican condiciones selectivas condicionalmente letales para eliminar las células que están infectadas con el virus. Un ejemplo de gen condicionalmente letal es TK de herpes, que se vuelve letal después de tratar las células que expresan este gen con el análogo de timidina aciclovir. En algunas realizaciones, la actividad antiproliferativa de los interferones clonados se selecciona tratando las células con agentes que eliminan las células en división (por ejemplo, agentes alquilantes de ADN).

En algunas realizaciones, se seleccionan citoquinas potentes expresando y secretando una genoteca de citoquinas en las células que tienen GFP u otro informador bajo el control de un promotor que es inducido por la citoquina, tal como el promotor de clase I del MHC que es inducido por variantes desarrolladas del interferón alfa. La ruta de transducción de la señal se configura de manera que la citoquina de tipo salvaje que se va a desarrollar proporciona una señal débil pero detectable.

J. Estabilidad en Suero y Vida Media en Circulación Mejoradas

En algunas realizaciones de la invención, se desarrollan proteínas mediante RSR para que tengan una vida media en la circulación o una estabilidad en suero mejoradas. Un método preferido para mejorar la vida media es desarrollar la afinidad de una proteína de interés a una proteína con una larga vida en suero, tal como un anticuerpo u otra proteína abundante en el suero. Los ejemplos de cómo la afinidad por un anticuerpo puede aumentar la vida media en suero incluyen la co-

administración de IL2 y anticuerpos anti-IL2 que incrementa la vida media en suero y la actividad anti-tumoral de la IL12 recombinante humana (Courtney et al., Immunopharmacology 28:223-232 (1994)).

5

Las ocho proteínas en suero humanas más abundantes son la albúmina del suero, las inmunoglobulinas, las lipoproteínas, la haptoglobina, el fibrinógeno, la transferrina, la alfa-1 antitripsina, y la alfa-2 macroglobulina (Doolittle, capítulo 6, The Plasma Proteins F. Putnam, ed.; Academic Press, 1984). Estas y otras proteínas abundantes del suero tales como la ceruloplasmina y la fibronectina son dianas primarias contra las que desarrollar sitios de unión en proteínas terapéuticas tal como en la Tabla I con el fin de prolongar la vida media. En el caso de los anticuerpos, la estrategia preferida consiste en desarrollar afinidad por las regiones constantes en lugar de las regiones variables con el fin de minimizar la variación individual en la concentración del epítipo diana relevante (el uso de la región V del anticuerpo entre diferentes individuos es significativamente variable).

Los sitios de unión de afinidad deseada se desarrollan aplicando la presentación en fagos, la presentación de péptidos sobre plásmidos o selecciones de presentación en polisomas para la proteína de interés. Como fuente de diversidad, se podría mutagenizar al azar un sitio de unión existente o una región definida de otro modo de la proteína diana, adjuntar una genoteca de péptidos al extremo N, el extremo C, o internamente como bucle no perjudicial funcionalmente, o utilizar "barajado de familia" de genes homólogos. El barajado de ADN resulta particularmente ventajoso para problemas en los que se desea optimizar simultáneamente dos o más

35

propiedades "no correlacionadas" tales como una afinidad mejorada por HSA a la vez que se conserva la actividad biológica.

5 En otras realizaciones de la invención, la vida media se mejora mediante derivatización con PEG, otros productos conjugados poliméricos o radicales químicos que prolongan la vida media. Estos son métodos establecidos para prolongar la vida media de las proteínas
10 terapéuticas (R. Duncan, Clin. Pharmacokinet 27:290-306 (1994); Smith et al., TIBTECH 11 397-403 (1993)) y pueden tener la ventaja añadida de reducir la inmunogenicidad (R. Duncan, Clin. Pharmacokinet 27:290-306 (1994)). Sin embargo, la derivatización también puede dar como
15 resultado una reducción de la afinidad de la proteína terapéutica por su receptor o ligando. La RSR se utiliza para descubrir sitios alternativos en la secuencia primaria que puedan ser sustituidos por lisina u otros residuos apropiados para la conjugación química o
20 enzimática con radicales químicos que prolonguen la vida media, y que den como resultado proteínas con una retención máxima de la actividad biológica.

25 Una estrategia preferida consiste en expresar una genoteca de mutantes de la proteína en un formato de presentación, derivatizar la genoteca con el agente de interés (esto es, PEG) utilizando una química que no inactive biológicamente el sistema de presentación, seleccionar basándose en la afinidad por el receptor
30 cognado, amplificar mediante PCR los genes que codifican los mutantes seleccionados, barajar, re-ensamblar, contar con el formato de presentación, y repetir hasta obtener un mutante con la actividad deseada post-modificación. Un formato alternativo consiste en expresar, purificar y
35 derivatizar los mutantes en un formato de alto

rendimiento, escrutar en busca de mutantes con la actividad optimizada, recuperar los genes correspondientes, someter los genes a RSR y repetir.

5 En realizaciones adicionales de la invención, los sitios de unión para las proteínas humanas diana que están localizadas en tejidos concretos de interés son desarrollados mediante RSR. Por ejemplo, se puede desarrollar un interferón para que contenga un sitio de
10 unión para una proteína de la superficie del hígado, tal como un receptor del factor de crecimiento de hepatocitos, de manera que el interferón se reparte selectivamente sobre las células del hígado y tiene una actividad antiviral específica superior sobre las células
15 del hígado. Semejante interferón desarrollado podría ser útil para el tratamiento de la hepatitis. Análogamente, se podría desarrollar la afinidad por epítomos abundantes sobre eritrocitos tales como los antígenos de la sangre ABO para localizar una proteína dada en la corriente
20 sanguínea.

 En realizaciones adicionales de la invención, la proteína de interés es desarrollada para que tenga una estabilidad a las proteasas incrementada. Por ejemplo, el
25 uso clínico de IL2 está limitado por los graves efectos secundarios que están relacionados con la necesidad de administrar dosis elevadas. Se requieren dosis elevadas debido a la corta vida media (3-5 min, Lotze et al., JAMA 256(22):3117-3124 (1986)) y a la consiguiente necesidad
30 de dosis elevadas para mantener un nivel terapéutico de IL2. Uno de los factores que contribuyen a las cortas vidas medias de las proteínas terapéuticas es la proteólisis por proteasas del suero. La catepsina D, una proteasa ácida renal principal, es responsable de la
35 degradación de IL2 en ratones Balb/c (Ohnishi et al.,

Cancer Res. 50:1107-1112 (1990)). Además, Ohnishi demostró que el tratamiento de ratones Balb/c con pepstatina, un potente inhibidor de esta proteasa, prolonga la vida media de la IL2 humana recombinante y aumenta la actividad asesina activada por linfoquinas en este modelo de ratón.

De este modo, el desarrollo de variantes de IL2 resistentes a proteasas o cualquiera de las proteínas enumeradas en la Tabla I que son resistentes a las proteasas del suero o riñón es una estrategia preferida para obtener variantes con vidas medias en suero prolongadas.

Un protocolo preferido es el siguiente. Se expresa una genoteca de la proteína mutagenizada de interés en un sistema de presentación con una etiqueta epitópica distante del gen (esto es en el extremo N de un constructo de presentación en fagos de manera que si es escindida por proteasas, la etiqueta epitópica se pierde). Las proteínas expresadas se tratan con proteasas definidas o con cócteles complejos tales como suero humano. Se realiza la selección por afinidad con un anticuerpo para la etiqueta distante del gen. Se realiza un segundo escrutinio o selección que demanda una función biológica (p. ej., unión al receptor cognado). Los fagos que conservan la etiqueta epitópica (y por lo tanto resistentes a la proteasa) se recuperan y se someten a RSR. El procedimiento se repite hasta que se logra el nivel de resistencia deseado.

En otras realizaciones, el procedimiento se realiza en un formato de escrutinio donde las proteínas mutantes son expresadas y purificadas en un formato de alto rendimiento y escrutadas en busca de resistencia a la

proteasa con retención de actividad biológica.

En realizaciones adicionales de la invención, se desarrolla la proteína de interés para que tenga un aumento de vida útil. Una genoteca de la secuencia de
5 ácido nucleico mutagenizada que codifica la proteína de interés es expresada en un formato de presentación o un formato de expresión de alto rendimiento, y expuesta durante diferentes intervalos de tiempo a condiciones para las cuales se desea desarrollar estabilidad (calor,
10 iones metálicos, pH no fisiológico, por ejemplo <6 o >8, liofilización, congelación-descongelación). Los genes son recuperados de los supervivientes funcionales, por ejemplo, mediante PCR. El ADN se somete a mutagénesis, tal como RSR, y el procedimiento se repite hasta lograr
15 el nivel de mejora deseado.

El caso del IFN presenta una oportunidad para desarrollar recombinantes con una vida media mejorada. Existen $> 10^{26}$ posibles recombinaciones de la diversidad
20 de aminoácidos de esta familia. Puesto que estos recombinantes se forman a partir de segmentos de genes de IFN de tipo salvaje, se crearán relativamente pocos, si se crean, epítomos de células T novedosos mediante el procedimiento. Es probable que las moléculas que son muy
25 activas se asemejen mucho a interferones naturales estructuralmente, y de este modo presenten pocos, si presentan, epítomos de células B novedosos. Esto crea una situación donde la capacidad para crear grandes genotecas de recombinantes se puede combinar con el poder de la
30 inmunopurificación de fagos para seleccionar recombinantes con afinidad por las proteínas del suero abundantes tales como la albúmina de suero humana. Se ha demostrado que las proteínas con afinidad por las proteínas del suero abundantes, de larga vida tienen
35 vidas medias en suero prolongadas. De este modo, se

podrían obtener recombinantes de IFN con vidas medias en suero prolongadas utilizando la inmunopurificación de fagos para seleccionar recombinantes que tengan afinidad por proteínas tales como HSA. Puesto que la unión a HSA o las mutaciones que crean afinidad por HSA pueden anular o reducir sustancialmente la actividad de IFN, se tendría que contrarrestar el escrutinio en busca de la retención de actividad de IFN potente. Aplicando la inmunopurificación de fagos, el escrutinio de actividad, y el barajado repetitivamente, se podrían obtener recombinantes con una elevada actividad y un nivel deseado de afinidad para las proteínas del suero diana. Las vidas medias de los IFN candidato pueden ser sometidas a ensayo en ratones transgénicos que expresan la proteína del suero humana como una neo-autoproteína.

Estos enfoques pueden ser generalizados para otras proteínas para las cuales existen múltiples formas alélicas o no alélicas humanas homólogas. El enfoque también puede ser generalizado adicionalmente para ser aplicado a proteínas sin homólogos humanos no alélicos, tales como IL2. El gen para IL2 sería barajado con genes IL2 de otros mamíferos, con una preferencia por los mamíferos íntimamente relacionados tales como los primates. Se espera que la recombinación de la "diversidad natural" definida por estos homólogos genere genotecas de muy alta calidad con muchas moléculas activas y superiores como se observó para las actividades de los interferones barajados en células humanas y de ratón.

K. Versiones Desarrolladas de Cadena Sencilla de Factores de Múltiples Subunidades

Como se ha comentado antes, en algunas realizaciones

de la invención, el sustrato para la evolución por RSR es preferiblemente una construcción de cadena sencilla. La posibilidad de realizar una mutagénesis asimétrica en constructos de proteínas homomultiméricas proporciona nuevas rutas importantes para una evolución adicional de tales constructos que no está abierta a las proteínas en sus estados homomultiméricos naturales. En particular, una mutación dada en un homomultímero dará como resultado que ese cambio esté presente en cada subunidad idéntica. En los constructos de cadena sencilla, no obstante, los dominios pueden mutar independientemente entre sí.

La conversión de proteínas de múltiples subunidades en constructos de cadena sencilla con propiedades nuevas y útiles ha sido demostrada para numerosas proteínas. Muy notablemente, se han conectado dominios variables de cadenas pesadas y ligeras de anticuerpos en Fv de cadena sencilla (Bird et al., Science 242:423-426 (1988)), y esta estrategia ha dado como resultado anticuerpos con una estabilidad térmica (Young et al., FEBS Lett 377:135-139 (1995)), o sensibilidad a la proteólisis (Solar et al., Prot. Eng. 8:717-723 (1995)) mejoradas. Se ha construido una versión de cadena sencilla funcional de IL5, un homodímero, que se ha demostrado que tiene afinidad por el receptor de IL5 similar a la de la proteína de tipo salvaje, y este constructo ha sido utilizado para realizar la mutagénesis asimétrica del dímero (Li et al., J. Biol. Chem. 271:1817-1820 (1996)). Se ha elaborado una versión de cadena sencilla del activador de plasminógeno de tipo uroquinasa, y se ha demostrado que el constructo de cadena sencilla es más resistente al inhibidor del activador de plasminógeno de tipo 1 que el homodímero nativo (Higazi et al., Blood 87:3545-3549 (1996)). Finalmente, se ha construido un híbrido de factor de crecimiento de tipo insulínico de

cadena sencilla I/insulina y se ha demostrado que tiene una afinidad mayor por los receptores de insulina/IGF-I quiméricos que la de cualquier ligando natural (Kristensen et al., Biochem. J. 305:981-986 (1995)).

5

10

15

20

25

30

35

En general, se construye un conector que une el extremo amino de una subunidad de una proteína de interés al extremo carboxilo de otra subunidad del complejo. Estas proteínas de fusión pueden consistir en versiones conectadas de homodímeros, homomultímeros, heterodímeros o heteromultímeros de orden superior. En el caso más simple, se añaden conectores polipeptídicos entre los extremos nativos que se van a unir. Se pueden elaborar dos variaciones significativas. Primero, se pueden construir diversas genotecas de variaciones de la secuencia de tipo salvaje en y alrededor de las uniones y en los conectores para facilitar la construcción de proteínas de fusión activas. En segundo lugar, Zhang et al., (Biochemistry 32:12311-12318 (1993)) han descrito permutaciones circulares de lisozima T4 en las cuales los extremos amino y carboxilo nativos se han unido y se han diseñado extremos amino y carboxilo novedosos en la proteína. Los métodos de permutación circular, las genotecas de conectores, y las genotecas de secuencias funcionales que flanquean los conectores permiten construir genotecas que tengan diversas estrategias de conexión topológica y en la secuencia primaria. Estas genotecas son expresadas y seleccionadas en busca de actividad. Se puede utilizar cualquiera de las estrategias mencionadas antes para el escrutinio o selección, siendo preferible la presentación en fagos en la mayoría de los casos. Los genes que codifican proteínas de fusión activas se recuperan, se mutagenizan, se vuelven a seleccionar, y se someten a protocolos de RSR convencionales para optimizar su función.

Preferiblemente, una población de constructos de cadena sencilla mutantes seleccionados es amplificado mediante PCR en dos reacciones de PCR separadas de manera que cada uno de los dos dominios es amplificado por separado. Los oligonucleótidos son obtenidos de los extremos 5' y 3' del gen y de ambas hebras del conector. Los dominios amplificados por separado son barajados en reacciones separadas, después las dos poblaciones son recombinadas utilizando el re-ensamblaje por PCR para generar constructos de cadena sencilla intactos para rondas de selección y evolución adicionales.

V. Propiedades Mejoradas de Proteínas Farmacéuticas

A. Especificidad Desarrollada para el Receptor o Tipo Celular de Interés

La mayoría de las proteínas enumeradas en la Tabla I son receptores o ligandos de interés farmacéutico. Muchos agonistas tales como las quimioquinas o las interleuquinas suscitan mas agonismo que un receptor. Los mutantes desarrollados con una especificidad mejorada pueden haber reducido los efectos secundarios debido a su pérdida de actividad sobre los receptores que están implicados en un perfil de efecto secundario concreto. Para la mayoría de estos ligandos/receptores, las formas mutantes con una afinidad mejorada tendrían propiedades farmacéuticas mejoradas. Por ejemplo, una forma antagónica de RANTES con una eficacia mejorada para CCR5 o CXCR4 o ambos debe ser un inhibidor mejorado de la infección por VIH en virtud del logro de una mayor ocupación del receptor para una dosis dada de fármaco. Utilizando las selecciones y escrutinios esbozados antes combinados con RSR, se pueden mejorar las afinidades y especificidades de cualquiera de las proteínas enumeradas

en la Tabla I. Por ejemplo, se podría utilizar el formato de presentación en mamíferos para desarrollar receptores de TNF con una afinidad mejorada por el TNF.

5 Otros ejemplos incluyen variantes de interferón alfa mejoradas que detienen la proliferación de las células tumorales pero no estimulan las células NK, variantes de IL2 que estimulan el complejo del receptor de IL2 de baja afinidad pero no el receptor de alta afinidad (o viceversa), superantígenos que estimulan solamente un subgrupo de las proteínas beta V reconocidas por la proteína de tipo salvaje (preferiblemente una beta V sencilla), formas antagónicas de quimioquinas que suscitan solamente el antagonismo de un receptor de interés, anticuerpos con una reactividad cruzada reducida, y factores quiméricos que activan específicamente un complejo receptor concreto. Como ejemplo de este último caso, se podrían elaborar quimeras entre IL2 e IL4, 7, 9, o 15 que también se unan a las cadenas alfa, beta y gamma del receptor de IL2 (Theze et al., Imm. Today 17:481-486 (1996)), y seleccionar en busca de quimeras que conservan la unión para el complejo del receptor de IL2 de afinidad intermedia en monocitos pero tienen una afinidad reducida por el complejo del receptor alfa, beta, gamma de IL2 de alta afinidad en células T activadas.

B. Agonistas Desarrollados con Potencia Incrementada

30 En algunas realizaciones de la invención, la estrategia preferida es la selección o escrutinio de mutantes con actividad agonística incrementada utilizando los formatos de células completas descritos antes, combinados con RSR. Por ejemplo, se expresa una genoteca de mutantes de IL3 en forma activa sobre fagos o

fagémidos como describen Gram et al. (J. Immun. Meth. 161:169-176 (1993)). Se preparan productos lisados clónicos resultantes de la infección con fagos purificados en placa en un formato de alto rendimiento tal como un formato de microtitulación de 96 pocillos. Se estimula una línea celular dependiente de IL3 que expresa un gen informador tal como GFP con las soluciones de partida de fago en un formato de microtitulación de 96 pocillos de alto rendimiento. Se recuperan los fagos que dan como resultado señales positivas a la mayor dilución de sobrenadantes de fago; alternativamente, se puede recuperar el ADN que codifica el IL3 mutante mediante PCR. En algunas realizaciones, se pueden estimular células individuales que expresan GFP bajo el control de un promotor sensible a IL3 con la genoteca de fagémidos, y se clasifican las células positivas mediante FACS. El ácido nucleico recuperado se somete después a PCR, y se repite el procedimiento hasta obtener el nivel de mejora deseado.

Tabla I

<u>POLIPÉPTIDOS CANDIDATO PARA SU DESARROLLO</u>	
<u>Nombre</u>	<u>Nombre</u>
Alfa-1 antitripsina	Proteínas Hedgehog (p. ej., Sonic, Indian, Desert)
Angiostatina	
Factor antihemolítico	Hemoglobina (para sustituto de la sangre; para radiosensibilización)
Apolipoproteína	
Apoproteína	Hirudina
Factor natriurético atrial	Albúmina de suero humano
Polipéptido natriurético atrial	Insulina

<u>POLIPÉPTIDOS CANDIDATO PARA SU DESARROLLO</u>	
<u>Nombre</u>	<u>Nombre</u>
Péptidos atriales	Interferón gamma
Toxinas de Bacillus thuringiensis (toxinas Bt)	Interleuquina 20 (gen 7 asociado a la diferenciación de melanoma)
Quimioquinas C (esto es, Linfotactina)	
Quimioquinas C-X-C (p. ej., T39765, NAP-2, ENA-78, Gro-a, Gro-b, Gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG)	Interleuquinas (1 a 18)
	Lactoferrina
	Leptina
	Factor Inhibidor de Leucemia (LIF)
Calcitonina	Luciferasa
Quimioquinas CC (p. ej., Proteína quimioatrayente de monocitos-1, Proteína quimioatrayente de monocitos-2, Proteína quimioatrayentes de monocitos-3, Proteína inflamatoria de monocito-1 alfa, Proteína inflamatoria de monocitos-1 beta, RANTES, 1309, R83915, R91733, HCC1, T58847, D31065, T64262)	Neurturina
	Factor inhibidor de neutrófilos (NIF)
	Oncostatina-M
	Proteína Osteogénica
	Hormona paratiroidea
	Proteína A
	Proteína G
	RANK (activador del receptor de NF- κ B)
Ligando CD40	Relaxina
Factor neurotrófico ciliar (CNTF)	Renina
Colágeno	Calcitonina de salmón

<u>POLIPÉPTIDOS CANDIDATO PARA SU DESARROLLO</u>	
<u>Nombre</u>	<u>Nombre</u>
Factor estimulador de colonias (CSF, G-CSF, GM-CSF, M-CSF)	Hormona de crecimiento de Salmón
	CD4 Soluble
Factor del complemento 5a	CD8 Soluble
Inhibidor del complemento	CD40 Soluble
Receptor del complemento 1	Ligando CD40 Soluble
Factor de crecimiento epidérmico (EGF)	CD80 Soluble (37-1)
Eritropoyetina	CD86 Soluble (87-2)
Factor IX	CD150 Soluble (SLAM)
Factor VII	CD152 Soluble (CTLA-4)
Factor VIII	Receptor de complemento soluble I
Factor X	I-CAM 1 Soluble
Fibrinógeno	Receptor IFN gamma soluble
Fibronectina	Receptores interleuquina soluble (IL-1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20)
Antagonista del receptor FLT-3	
Glucocerebrosidasa	
Gonadotropina	Receptor de leptina soluble
Hormona del crecimiento	RANK Soluble
	Receptor de TNF soluble
Somatomedina	Superóxido dismutasa
Somatostatina	Trombopoyetina

<u>POLIPÉPTIDOS CANDIDATO PARA SU DESARROLLO</u>	
<u>Nombre</u>	<u>Nombre</u>
Somatotropina	Timosina alfa 1
Factor de células pluripotenciales	Activador del plasminógeno tisular
Estreptoquinasa	Factor de crecimiento transformante beta
Superantígenos, esto es, enterotoxinas	Factor de necrosis tumoral beta (TNF beta)
Estafilocócicas (SEA, SEB, SEC1, SEC2, SEC3, SED, SEE),	Receptor del factor de necrosis tumoral (TNFR)
Toxina de síndrome de choque tóxico (TSST-1), Toxinas exfoliantes A y B,	Factor de necrosis tumoral alfa (TNF alfa)
Exotoxinas pirogénicas A, B, y C, y mitógeno de M. arthritidis	Uroquinasa
	Homólogos de IL10 Virales

C. Evolución de Componentes de la Transducción de la Señal o de las Rutas Transcripcionales Eucarióticas

5 Utilizando los escrutinios y selecciones enumerados antes, se puede utilizar la RSR de diferentes maneras para modificar la transducción de la señal o las rutas transcripcionales eucarióticas. Se puede desarrollar cualquier componente de una ruta de transducción de la

10 señal de interés, o de las regiones reguladoras y los activadores transcripcionales que interaccionan con esta región y con agentes químicos que inducen la transcripción. Esto genera sistemas reguladores en los cuales la transcripción es activada más potentemente por

15 el inductor natural o por análogos del inductor normal. Se prefiere esta tecnología para el desarrollo y

optimización de diferentes análisis de interés biotecnológico. Por ejemplo, docenas de receptores con 7 dominios transmembrana (7-TM) son dianas validadas para el descubrimiento de fármacos (véase, por ejemplo, Siderovski et al., *Curr Biol.*, 6(2):211-212 (1996); An et al., *FEBS Lett.*, 375(1-2):121-124 (1995); Raport et al., *Gene*, 163(2):295-299 (1995); Song et al., *Genomics*, 28(2):347-349 (1995); Strader et al. *FASES J.*, 9(9):745-754 (1995); Benka et al., *FEBS Lett.*, 363(1-2):49-52 (1995); Spiegel, J. *Clin Endocrinol. Metab.*, 81(7):2434-2442 (1996); Post et al., *FASEB J.*, 10(7):741-749 (1996); Reisine et al., *Ann NY Acad. Sci.*, 780:168-175 (1996); Spiegel, *Annu. Ref. Physiol.*, 58:143-170 (1996); Barak et al., *Biochemistry*, 34(47):15407-15414 (1995); y Shenker, *Baillieres Clin. Endocrinol. Metab.*, 9(3):427-451 (1995)). El desarrollo de análisis de alto rendimiento sensibles en busca de agonistas y antagonistas de estos receptores es esencial para explotar todo el potencial de la química combinatoria en el descubrimiento de tales ligandos. Adicionalmente, se pueden desarrollar biodetectores o biosensores para diferentes agentes químicos desarrollando 7-TM para que respondan agonísticamente a agentes químicos novedosos o proteínas de interés. En este caso, la selección sería en busca de constructos que sean activados por el nuevo agente químico o polipéptido que se va a detectar. El escrutinio se podría realizar simplemente con clasificación celular activada por fluorescencia o luz, puesto que la mejora deseada se acopla a la producción de luz.

Además de la detección de moléculas pequeñas tales como fármacos y contaminantes medioambientales, se pueden desarrollar biosensores que respondan a cualquier agente químico para el cual son receptores, o para el cual se pueden desarrollar receptores mediante recombinación

repetitiva de secuencias, tales como hormonas, factores de crecimiento, metales y fármacos. Los receptores pueden ser activadores intracelulares y directos de la transcripción, o pueden ser receptores unidos a la membrana que activan la transcripción de la señal indirectamente, por ejemplo por medio de una cascada de fosforilación. También pueden no actuar sobre la transcripción en absoluto, pero pueden producir una señal mediante alguna modificación post-transcripcional de un componente de la ruta de generación de la señal. Estos receptores también pueden ser generados fusionando dominios responsables de la unión de diferentes ligandos con diferentes dominios de señalización. De nuevo, se puede utilizar la recombinación repetitiva de secuencias para incrementar la amplitud de la señal generada para optimizar la expresión y el funcionamiento de receptores quiméricos, y para alterar la especificidad de los agentes químicos detectados por el receptor.

Por ejemplo, se pueden desarrollar proteínas G para acoplar eficazmente receptores 7-TM de mamífero a rutas de transducción de la señal de levadura. Actualmente existen 23 loci de proteína alfa G conocidos en mamíferos que se pueden agrupar mediante similitud de secuencia y funcional en cuatro grupos, Gs (Gna, Gna1), Gi (Gnai-2, Gnai-3, Gnai-1, Gnao, Gnat-1, Gnat-2, Gnaz), Gq (Gnaq, Gna-11, Gna-14, Gna-15) y G12 (Gna-12, Gna-13) (B. Numberg et al., J. Mol. Med., 73:123-132 (1995)). Poseen una actividad GTP-asa endógena que permite el acoplamiento funcional reversible entre receptores unidos a ligandos y efectores aguas abajo tales como enzimas y canales iónicos. Las proteínas alfa G forman complejos no covalentemente con las proteínas beta G y gamma G así como los receptores 7-TM cognados. La especificidad del receptor y la señal está controlada por la combinación

concreta de las subunidades alfa G, beta G (de la que existen cinco loci conocidos) y gamma G (siete loci conocidos). La activación del complejo heterotrimérico por el ligando unido al receptor da como resultado la disociación del complejo en monómeros alfa G y dímeros beta, gamma G que después transmiten señales por asociación con proteínas efectoras aguas abajo. Se cree que la subunidad alfa G es la unidad que contacta con 7-TM, y de este modo es el centro para el desarrollo de subunidades alfa G quiméricas o desarrolladas que pueden transmitir señales las 7-TM de mamífero a los genes aguas abajo de levadura.

Los bioanálisis basados en levadura para receptores de mamífero facilitarán enormemente el descubrimiento de ligandos novedosos. Kang et al. (Mol. Cell Biol, 10:2582-2590 (1990)) han descrito la complementación parcial de cepas de levadura que portan mutaciones en SCG1 (GPA1), un homólogo de las subunidades alfa de las proteínas G implicadas en la transducción de la señal en células de mamífero, mediante proteínas alfa G de mamífero e híbridos de levadura/mamífero. Estos híbridos tienen una función parcial, tal como la complementación del defecto de crecimiento en cepas scg1, pero no permiten el apareamiento y por consiguiente no complementan totalmente la función en la ruta de transducción de la señal de feromonas. Price et al. (Mol. Cell Biol. 15:6188-6195 (1995)) han expresado el subtipo 2 de receptor de somatostatina de rata (SSTR2) en levadura y han demostrado la transmisión de las señales de unión al ligando por medio de este receptor 7-TM a través de subunidades alfa G de mamífero y quiméricas de mamífero/levadura ("acoplamiento") a un gen informador HIS3, bajo el control del promotor sensible a feromonas FUS-1 que permite que células de otro tipo HIS3(-)

crezcan sobre medio mínimo que carece de histidina.

Tales cepas son útiles como cepas informadoras para receptores de mamífero, pero adolecen de limitaciones importantes como ilustra el estudio de Kang et al., donde
5 parece haber un bloqueo en la transmisión de señales desde los receptores de feromona de levadura a las proteínas G de mamífero. En general, para acoplar un receptor 7-TM de mamífero a rutas de transducción de la
10 señal de levadura se acopla el receptor de mamífero a las proteínas alfa G de levadura, mamífero, o quiméricas, y estas a su vez interaccionan productivamente con los componentes aguas abajo en la ruta para inducir la expresión de un promotor sensible a feromona tal como FUS-1. Semejante reconstitución funcional es referida
15 comúnmente como "acoplamiento".

Los métodos descritos en la presente memoria pueden ser utilizados para desarrollar el acoplamiento de receptores 7-TM de mamífero a rutas de transducción de la
20 señal de levadura. Un enfoque típico es el siguiente: (1) clonar un 7-TM de interés en una cepa de levadura con una ruta de respuesta a feromona modificada similar a la descrita por Price (p. ej., cepas carentes de FAR1, un regulador negativo de G, ciclinas, y carente de SST2 que
25 hace que las células sean hipersensibles a la presencia de la feromona), (2) construir genotecas de quimeras entre las proteínas alfa G de mamífero conocidas o que se piensa que interaccionan con GPA1 o proteínas alfa G de levadura homólogas, (3) colocar un gen informador seleccionable tal como HIS3 bajo el control del promotor
30 sensible a feromona FUS1 (Price et al., Mol. Cell Biol. 15:6188-6195 (1995)). Alternativamente, se puede colocar un gen escrutable tal como luciferasa bajo el control del promotor FUS1; (4) transformar la genoteca (2) en la cepa
35 (3) (HIS(-)), (5) escrutar o seleccionar en busca de la

expresión del informador en respuesta al ligando de interés, por ejemplo haciendo crecer la genoteca de transformantes sobre placas mínimas en presencia de ligando para demandar la expresión de HIS3, (6) recuperar
5 las células seleccionadas, y aplicar la RSR para desarrollar una expresión mejorada del informador bajo el control del promotor sensible a feromona FUS1.

Una segunda consideración importante en el
10 desarrollo de cepas con constructos informadores optimizados para las rutas de transducción de la señal de interés es optimizar la relación de señal/ruido (la razón de expresión del gen en condiciones de inducción vs no inducción). Muchas rutas 7-TM tienen pérdida de manera
15 que la inducción máxima de un gen informador típico es de 5 a 10 veces por encima del fondo. Este intervalo de señal/ruido puede ser insuficiente para detectar pequeños efectos en muchos análisis de elevado rendimiento. Por lo tanto, es interesante acoplar la ruta 7-TM a un segundo
20 sistema de amplificación no lineal que está sintonizado para que esté por debajo pero cerca del umbral de activación en el estado no inducido. Un ejemplo de un sistema de amplificación no lineal es la expresión de genes conducidos por el promotor P_L lambda. Las
25 interacciones cooperativas complejas entre el represor lambda unido a tres sitios adyacentes en el promotor cl dan como resultado una represión muy eficaz por encima de cierta concentración de represor. Por debajo de un umbral crítico se observa una inducción espectacular y hay una
30 ventana dentro de la cual un pequeño descenso en la concentración de represor conduce a un gran incremento en la expresión del gen (Ptashne, A Genetic Switch: Phage Lambda and Higher Organisms, Blackwell Scientific Publ. Cambridge, MA, 1992). Se observan efectos análogos para
35 algunos promotores eucarióticos tales como los regulados

por GAL4. La colocación de la expresión de un componente limitante de un factor de transcripción para semejante promotor (GAL4) bajo el control de un promotor sensible a 7-TM potenciado por GAL4 da como resultado pequeños niveles de inducción de la señal de la ruta 7-TM que es amplificada a un cambio mucho mayor en la expresión de un constructo informador también bajo el control de un promotor dependiente de GAL4.

Un ejemplo de semejante sistema acoplado consiste en colocar GAL4 bajo el control del promotor sensible a la feromona FUS-1 y tener el nivel de GL4 intracelular (por sí mismo un intensificador transcripcional) positivamente retroalimentado por sí mismo colocando un sitio de unión a GAL4 aguas arriba del promotor FUS-1. También se pone un gen informador bajo el control de un promotor activado por GAL4. Este sistema se diseña de manera que la expresión de GAL4 se auto-amplificará no linealmente y co-amplificará la expresión de un gen informador tal como luciferasa tras alcanzar un cierto umbral en la célula. Se puede utilizar la RSR con gran ventaja para desarrollar constructos informadores con las propiedades de señalización deseadas, como sigue: (1) Se elabora un único constructo plasmídico que contiene tanto el gen GAL4 regulado por la ruta de GAL4/feromona como el gen informador regulado por GAL4. (2) Este constructo es mutagenizado y transformado en la cepa de levadura diseñada apropiadamente que expresa 7-TM y una proteína de levadura/mamífero quimérica de interés. (3) Las células se estimulan con agonistas y se escrutan (o seleccionan) basándose en la actividad del gen informador. En un formato preferido, la luciferasa es el gen informador y la actividad es cuantificada antes y después de la estimulación con el agonista, permitiendo de ese modo una medida cuantitativa de la señal/ruido

para cada colonia. (4) Se recuperan las células con propiedades informadoras mejoradas, los constructos se barajan, y se aplica la RSR para desarrollar adicionalmente el plásmido para dar las características óptimas de señal/ruido.

Estos enfoques son generales e ilustran cómo cualquier componente de una ruta de transducción de la señal o factor de transcripción podría ser desarrollado utilizando la RSR y los escrutinios y selecciones descritos más arriba. Por ejemplo, estos métodos específicos podrían ser utilizados para desarrollar receptores 7-TM con especificidad por ligandos novedosos, especificidad de receptores nucleares por ligandos novedosos (por ejemplo para obtener la expresión de genes de interés inducibles por herbicidas u otras moléculas pequeñas en plantas transgénicas, de manera que un grupo dado de genes pueda ser inducido tras el tratamiento con un agente químico dado), especificidad de factores de transcripción para que sean sensibles a factores virales (induciendo de ese modo genes antivirales o letales en células que expresan este factor de transcripción (transgénicos o células tratadas con constructos de terapia génica), o especificidad de factores de transcripción por actividad en células cancerosas (por ejemplo células carentes de p53, permitiendo de este modo infectar con constructos de terapia génica que expresan genes letales condicionalmente de una manera específica del tumor).

Se ofrecen los siguientes ejemplos a modo de ilustración, en ningún modo de limitación.

EJEMPLOS EXPERIMENTALES

I. Desarrollo de BIAP

Una estrategia preferida para desarrollar BIAP es la siguiente. Se construye una genoteca de uso codónico a partir de oligonucleótidos de 60 unidades de manera que las 20 bases centrales de cada oligo especifican la proteína de tipo salvaje, pero codifican la secuencia de proteínas de tipo salvaje con codones degenerados. Preferiblemente, no se utilizan codones muy raros para el anfitrión procariótico de elección, tal como *E. coli*. Las 20 bases de cada extremo del oligo utilizan codones no degenerados, si no preferidos, en *E. coli*. Los oligonucleótidos se ensamblan en genes completos como se ha descrito más arriba. Los productos ensamblados se clonan en un vector de expresión mediante mecanismos bien conocidos en la técnica. En algunas realizaciones, la genoteca de uso de codones es expresada con una genoteca de secuencias líder secretoras, cada una de las cuales dirige la proteína BIAP codificada al periplasma de *E. coli*. Se utiliza una genoteca de secuencias líder para optimizar la combinación de secuencia líder y mutante. Los ejemplos de secuencias líder son revisados por Schatz et al. (Ann Rev. Genet. 24:215-248 (1990)). Los genes BIAP clonados son expresados bajo el control de un promotor inducible tal como el promotor de arabinosa. Las colonias inducidas por arabinosa son escrutadas mediante pulverización con un sustrato para BIAP, bromo-cloro-indolil fosfato (BCIP). Las colonias más azules se escogen visualmente y se someten a los procedimientos de RSR descritos en la presente memoria.

Los oligonucleótidos para la construcción de la genoteca de uso codónico se enumeran en la Tabla II. Las correspondientes localizaciones de estos promotores se proporcionan en la Figura 1.

Tabla II

1.	AACCCTCCAG TTCCGAACCC CATATGATGA TCACCCTGCG TAAACTGCCG
2.	AACCCTCCAG TTCCGAACCC CATATGAAAA AAACCGCT
3.	AACCCTCCAG TTCCGAACCC ATATACATAT GCGTGCTAAA
4.	AACCCTCCAG TTCCGAACCC CATATGAAAT ACCTGCTGCC GACC
5.	AACCCTCCAG TTCCGAACCC GATATACATA TGAAACAGTC
6.	TGGTGTTATG TC TGCTCAGG CDATGGCDGT DGAYTTYCAY CTGGTTCCGG TTGAAGAGGA
7.	GGCTGGTTTC GCTACCGTTG CDCARGCDGC DCCDAARGAY CTGGTTCCGG TTGAAGAGGA
8.	CACCCCGATC GCTATCTCTT CYTTYGCDTC YACYGGYTCY CTGGTTCCGG TTGAAGAGGA
9.	GCTGCTGGCT GCTCAGCCGG CDATGGCDAT GGAYATYGGY CTGGTTCCGG TTGAAGAGGA
10.	TGCCGCTGCT GTTCACCCCG GTDACYAARG CDGCD CARGT DCTGGTTCCG GTTGAAGAGG A
11.	CCCGGCTTTC TGGAACCGTC ARGCDGCDCA RGCDCTGGAC GTTGCTAAAA AACTGCAGCC
12.	ACGTTATCCT GTTCCTGGGT GAYGGYATGG GYGTDCCDAC CGTACCGCT ACCGTATCC

13.	AAACTGGGTC CGGAAACCCC DCTGGCDATG GAYCARTTYC CGTACGTTGC TCTGTCTAAA
14.	GGTTCCGGAC TCTGCTGGTA CYGCDACYGC DTAYCTGTGC GGTGTTAAAG GTAAC TACCG
15.	CTGCTCGTTA CAACCAGTGC AARACYACYC GYGGYAAAYGA AGTTACCTCT GTTATGAACC
16.	TCTGTTGGTG TTGTTACCAC YACYCGYGTG CARCAYGCDT CTCCGGCTGG TGCTTACGCT
17.	GTA CTCTGAC GCTGACCTGC CDGCDGAYGC DCARATGAAC GGTG GCCAGG ACATCGCTGC
18.	ACATCGACGT TATCCTGGGT GGYGGYCGYA ARTAYATGTT CCCGGTTGGT ACCCCGGACC
19.	TCTGTTAACG GTGTTCTGTA RCGYAARCAR AAYCTGGTDC AGGCTTGGCA GGCTAAACAC
20.	GAACCGTACC GCTCTGCTGC ARGCDGCDGA YGAYTCYTCT GTTACCCACC TGATGGGTCT
21.	AATACAACGT TCAGCAGGAC CAYACYAARG AYCCDACYCT GCAGGAAATG ACCGAAGTTG

22.	AACCCGCGTG GTTTCTACCT GTTYGTDGAR GGYGGYCGYA TCGACCACGG TCACCACGAC
23.	GACCGAAGCT GGATGTTTCG AYAAYGCDAT YGCDAARGCT AACGAACTGA CCTCTGAACT
24.	CCGCTGACCA CTCTCACGTT TTYTCYTTYG GYGGYTAYAC CCTGCGTGGT ACCTCTATCT
25.	GCTCTGGACT CTAÄATCTTA YACYTCYATY CTGTAYGGYA ACGGTCCGGG TTACGCTCTG
26.	CGTTAACGAC TCTACCTCTG ARGAYCCDTC YTAYCARCAG CAGGCTGCTG TTCCGCAGGC
27.	AAGACGTTGC TGTTTTGCT CGYGGYCCDC ARGCDCACT GGTTCACGGT GTTGAAGAAG
28.	ATGGCTTTTCG CTGGTTGCGT DGARCCDTAY ACYGAYTGYA ACCTGCCGGC TCCGACCACC
29.	TGCTCACCTG GCTGCTTMAC CDCCDCCDCT GGCDCTGCTG GCTGGTGCTA TGCTGCTCCT C
30.	TTCCGCCTCT AGAGAATTCT TARTACAGRG THGGHGCCAG GAGGAGCAGC ATAGCACCAG CC

31.	AAGCAGCCAG GTGAGCAGCG TCHGGRATRG ARGTHGCGGT GGTCCGAGC GGCAGGTT
32.	CGCAACCAGC GAAAGCCATG ATRTGHGCHA CRAARGTYTC TTCTTCAACA CCGTGAACCA
33.	GCGAAAACAG CAACGTCTTC RCCRCCRTGR GTYTCRGAHG CCTGCGGAAC AGCAGCCTGC
34.	AGAGGTAGAG TCGTTAACGT CHGGRCGRGA RCCRCCRCCC AGAGCGTAAC CCGGACCGTT
35.	AAGATTTAGA GTCCAGAGCT TTRGAHGGHG CCAGRCCRAA GATAGAGGTA CCACGCAGGG
36.	ACGTGAGAGT GGTCAGCGGT HACCAGRATC AGRGTRTCCA GTTCAGAGGT CAGTTCGTTA
37.	GAACATACCA GCTTCGGTCA GHGCCATRTA HGCYTTRTCG TCGTGGTGAC CGTGGTCGAT
38.	GGTAGAAACC ACGCGGGTTA CGRGAHACHA CRCGCAGHGC AACTTCGGTC ATTCCTGCA
39.	TCCTGCTGAA CGTTGTATTT CATRTCHGCH GGYTCRAACA GACCCATCAG GTGGGTAACA

40.	CAGCAGAGCG GTACGGTTCC AHACRTAYTG HGCRCCTGG TGTTTAGCCT GCCAAGCCTG
41.	TACGAACACC GTTAACAGAA GCRTCRTCHG GRTAYTCHGG GTCCGGGGTA CCAACCGGGA
42.	CCCAGGATAA CGTCGATGTC CATRTTRTTH ACCAGYTGHG CAGCGATGTC CTGGCAACCG
43.	CAGGTCAGCG TCAGAGTACC ARTTRCGRTT HACRGTRTGA GCGTAAGCAC CAGCCGGAGA
44.	TGG TAACAAC ACCAACAGAT TTRCCHGCTT TYTTHGCRCG GTTCATAACA GAGGTAACCT
45.	CACTGGTTGT AACGAGCAGC HGCRAHAACR CCRATRGTRC GGTAAGTACC TTTAACACCG
46.	ACCAGCAGAG TCCGGAACCT GRCGRTCHAC RTTGTARGTT TTAGACAGAG CAACGTACGG
47.	GGGTTTCCGG ACCCAGTTTA CCRTTCATYT GRCCYTTGAG GATACGGGTA GCGGTAACGG
48.	CCCAGGAACA GGATAACGTT YTTHGCHGCR GTYTGRATHG GCTGCAGTTT TTTAGCAACG

49.	ACGGTTCCAG AAAGCCGGGT CTCCTCTTC AACCGGAACC AG
50.	CCTGAGCAGA CATAACACCA GCHGCHACHG CHACHGCCAG CGGCAGTTTA CGCAGGGTGA
51.	ACCGGGGTGA ACAGCAGCGG CAGCAGHGCC AGHGCRATRG TRGACTGTTT CATATGTATA TC
52.	GCCGGCTGAG CAGCCAGCAG CAGCAGRCCH GCHGCHGCGG TCGGCAGCAG GTAGTTTCA
53.	AAGAGATAGC GA'CGGGGTG GTCAGHACRA TRCCCAGCAG TTAGCACGC ATATGTATAT
54.	CAACGGTAGC GAAACCAGCC AGHGCHACHG CRATHGCRAT AGCGGTTTTT TTCATATG
55	AGAATTCTCT AGAGGCGGAA ACTCTCCAAC TCCCAGGTT
56.	TGAGAGGTTG AGGGTCCAAT TGGGAGGTCA AGGCTTGGG
<p>Todos los oligonucleótidos son enumerados 5' a 3'. El código para las posiciones degeneradas es: R: A o G; Y: C o T; H: A o C o T; D: A o G o T.</p>	

II. Presentación en Superficie de Mamíferos

5 Durante una respuesta inmunitaria los anticuerpos experimentan naturalmente un procedimiento de maduración de la afinidad dando como resultado anticuerpos mutantes con afinidades mejoradas por sus antígenos cognados. Este procedimiento es dirigido por la hipermutación somática de genes de anticuerpos acoplados con la selección clónica (Berek y Milstein, Immun. Rev. 96:23-41 (1987)).

10

Patten et al. (Science 271:1086-1091 (1996)) han reconstruido el progreso de un anticuerpo catalítico desde la secuencia de la línea germinal, que se une a un hapteno de p-nitrofenilfosfonato con una afinidad 135 micromolar, a la secuencia madurada para la afinidad que ha adquirido nueve mutaciones somáticas y se une con una afinidad 10 nanomolar. La maduración de afinidad de este anticuerpo puede ser recapitulada y mejorada después de utilizar mutagénesis de casete de las CDR (o mutagénesis al azar por ejemplo con PCR), presentación en mamíferos, selección FACS en busca de unión mejorada, y RSR para desarrollar rápidamente afinidad mejorada recomblando mutaciones que codifican una unión mejorada.

Se construyen vectores lanzadera de expresión de anticuerpos genómicos similares a los descritos por Gascoigne et al. (Proc. Natl. Acad. Sci. (U.S.A.) 84:2936-2940 (1987)) de manera que las genotecas de exones de la región V mutante puedan ser fácilmente clonados en los vectores lanzadera. El constructo kappa es clonado en un plásmido que codifica resistencia a puromicina y la cadena pesada es clonada en un vector que codifica resistencia a neomicina. El ADNc derivado de las secuencias de la región variable que codifica las regiones V de las cadenas pesadas y ligeras maduras y de la línea germinal se reconfiguran mediante mutagénesis por PCR en exones genómicos flanqueados por sitios Sfi I con sitios Sfi I complementarios situados en localizaciones apropiadas en los vectores lanzadera genómicos. Los oligonucleótidos utilizados para crear los sitios Sfi I intrónicos que flanquean el exón VDJ son: 5' Sfi I: 5'-TTCCATTTCA TACATGGCCG AAGGGGCCGT GCCATGAGGA TTTT-3'; 3' Sfi I: 5'-TTCTAAATG CATGTTGGCC TCCTTGCCG GATTCTGAGC CTTCAGGACC A-3'. Se aplican protocolos de mutagénesis por PCR convencionales para producir

5 genotecas de mutantes donde los siguientes grupos de residuos (numerados de acuerdo con Kabat, Sequences of Proteins of Immunological Interest, U.S. Dept of Health and Human Services, 1991) son aleatorizados en codones NNK (GATC, GATC, GC) :

Cadena	CDR	Residuos mutados
V-L	1	30, 31 , 34
V-L	2	52, 53, 55
V-H	2	55, 56, 65
V-H	"4"	74, 76, 78

10 Se elaboran líneas transfectantes estables para cada uno de los dos constructos de las cadenas ligera y pesada (madura y de la línea germinal) utilizando el mieloma de células B AG8-653 (una donación de J. Kearney) como anfitrión utilizando protocolos de electroporación convencionales. Las genotecas de plásmidos mutantes que codifican las genotecas indicadas de mutantes V-L se transfectan en el transformante estable que expresa la línea germinal V-H; y los mutantes V-H se transfectan a la línea transfectante estable de la línea germinal V-L. En ambos casos, las genotecas se introducen mediante fusión de protoplastos (Sambrook et al., Molecular Cloning, CSH Press (1987)) para asegurar que la mayoría de las células transfectadas reciben una y solamente una secuencia plasmídica mutante (que no sería el caso para la electroporación donde la mayoría de las células transfectadas recibiría muchos plásmidos, expresando cada uno una secuencia mutante diferente).

20

25

El hapteno de p-nitrofenilfosfonato (JWJ-1) reconocido por este anticuerpo es sintetizado como

describen Patten et al. (Science 271:1086-1091 (1996)). JWJ-1 se acopla directamente a 5-(((2-aminoetil)tio)-acetil)fluoresceína (Molecular Probes, Inc.) mediante formación de un enlace amida utilizando una química de acoplamiento normalizada tal como EDAC (March, Advanced Organic Chemistry, Tercera edición, John Wiley and Sons, 1985) para dar una sonda JWJ-1-FITC monomérica. Se elabora un producto conjugado "dimérico" (dos moléculas de JWJ-1 acopladas a un marcador FACS) con el fin de obtener una sonda de mayor avidéz, haciendo de ese modo que las interacciones de baja afinidad (por ejemplo con el anticuerpo de la línea germinal) sean detectadas más fácilmente mediante FACS. Esto se genera mediante tinción con Rojo Texas conjugado con un anticuerpo anti-fluoresceína en presencia de dos equivalentes de JWJ-1-FITC. La estructura bivalente de IgG proporciona después un reactivo bivalente homogéneo. Se utiliza una columna giratoria para separar el exceso de moléculas de JWJ-1-FITC que no están unidas al reactivo anti-FITC. Se elabora un reactivo tetraivalente como sigue. Un equivalente de biotina se acopla con EDAC a dos equivalentes de etilendiamina, y después esto se acopla al carboxilato libre sobre JWJ-1. El producto JWJ-1 biotinilado se purifica mediante cromatografía de intercambio iónico y se caracteriza mediante espectrometría de masas. La avidina marcada con FITC se incuba con JWJ-1 biotinilado con el fin de generar una sonda tetraivalente.

Se realiza la selección FACS como sigue, de acuerdo con un protocolo similar al de Panka et al. (Proc. Natl. Acad. Sci. (U.S.A.) 85:3080-3084 (1988)). Después de la transfección de genotecas de genes de anticuerpo mutantes mediante el método de la fusión de protoplastos (con recuperación durante 36 - 72 horas), las células se

incubán sobre hielo con hapteno marcado fluorescentemente. La incubación se realiza sobre hielo para minimizar la pinocitosis del producto conjugado con FITC que puede contribuir a un fondo no específico.

5 Después las células son clasificadas en el FACS con o sin una etapa de lavado. El FACS sin etapa de lavado es preferible porque se espera que la disociación para el anticuerpo de la línea germinal antes de la maduración de afinidad sea muy rápida ($>0,1$ seg-1; Patten et al., Science 271:1086-1091 (1996)); una etapa de lavado añade

10 una variable de complicación. Se recogen 0,1 - 10% de las células más brillantes.

Se manipulan cuatro parámetros para optimizar la selección en busca de una unión incrementada: hapteno

15 monomérico vs dimérico vs tetramérico, concentración de hapteno utilizada en la reacción de tinción (la baja concentración selecciona Kd de elevada afinidad), tiempo entre lavado y FACS (el tiempo más largo selecciona disociaciones bajas), y selectividad en la distribución

20 (esto es tomar del 0,1% al 10% de más arriba, más preferiblemente el 0,1% de más arriba). Se utilizan los constructos que expresan la línea germinal, madura y ambas combinaciones de la línea germinal media como

25 controles para optimizar esta selectividad.

Se recuperan los plásmidos de las células seleccionadas por FACS mediante la transformación de un anfitrión *E. coli* con sobrenadantes Hirt.

30 Alternativamente, los exones del gen V mutante son amplificados mediante PCR a partir de las células seleccionadas por FACS. Los exones del gen V recuperados se someten a RSR, se reclonan en el correspondiente vector lanzadera genómico, y se aplica repetitivamente el

35 procedimiento hasta que aumenta la intensidad media de la

fluorescencia. Un control positivo relevante para la unión mejorada es la transfección con los exones 48G7 madurados por afinidad (Patten et al., op. cit.).

5 En un experimento adicional, se mezclan números iguales de la línea germinal y cada uno de los dos transfectantes de la línea germinal media. Se seleccionan las células más brillantes en las condiciones descritas antes. Los genes V son recuperados mediante PCR,
10 reclonados en vectores de expresión, y co-transfectados, dos plásmidos cualesquiera por *E. coli* seguido de fusión de protoplastos, o mediante electroporación masiva. La intensidad media de la fluorescencia de los transfectantes debe aumentar debido al enriquecimiento de
15 regiones V maduras con respecto a las de la línea germinal.

 Esta metodología se puede aplicar para desarrollar cualquier interacción receptor-ligando o compañero de
20 unión. Se pueden utilizar formatos de expresión naturales para expresar genotecas de mutantes de cualquier receptor para el cual se desea mejorar la afinidad por ligandos naturales o novedosos. Los ejemplos típicos serían la mejora de la afinidad de los receptores de las células T
25 por los ligandos de interés (esto es, complejos MHC/antígeno peptídico tumoral) o del receptor de TNF por TNF (se utilizan formas solubles de receptores de TNF terapéuticamente para neutralizar la actividad del TNF).

30 Este formato también se puede utilizar para seleccionar formas mutantes de ligandos que expresan el ligando en una forma unida a la membrana con un ancla de membrana diseñada mediante una estrategia análoga a la de Wettstein et al. (J. Exp. Med. 174:219-28 (1991)).
35 Después se realiza la selección FACS con el receptor

marcado fluorescentemente. En este formato se podrían desarrollar, por ejemplo, antagonistas de receptores mejorados a partir de los antagonistas de receptores de origen natural (antagonista del receptor de IL1, por ejemplo). También se podrían desarrollar en este formato formas mutantes de agonistas con afinidad mejorada por sus receptores cognados. Estos mutantes serían candidatos a agonistas mejorados o potentes antagonistas de receptores, análogos a formas mutantes antagónicas referidas de IL3.

III. Desarrollo de Interferón Alfa

Se encuentran disponibles 18 genes de interferón alfa (IFN- α) humano no alélicos conocidos, con estructuras primarias altamente relacionadas (idénticas en un 78-95%) y con una amplia gama de actividades biológicas. Se han descrito muchos interferones híbridos con actividades biológicas interesantes que difieren de las moléculas parentales (revisado por Horisberger y Di Marco, Pharm. Ther. 66:507-534 (1995)). Se ha construido sintéticamente un interferón alfa humano consenso, IFN-Con1, donde el residuo más común de los catorce IFN- α conocidos ha sido colocado en cada posición, y se compara favorablemente con los interferones de origen natural (Ozes et al., J. Interferon Res. 12:55-59 (1992)). Este IFN contiene 20 cambios de aminoácidos con respecto a IFN- α 2a, el IFN- α con el que está más íntimamente relacionado. IFN-Con1 tiene una actividad antiviral específica 10 veces superior a la de cualquier subtipo de IFN natural conocido. El IFN- α Con1 tiene actividades in vitro 10 a 20 veces superiores que las del IFN α -2a recombinante (el principal IFN utilizado clínicamente) en activación antiviral, antiproliferativa de células NK. De este modo, existe un interés considerable en producir

híbridos de interferón que combinen los rasgos más deseables de dos o más interferones. No obstante, dado el enorme número de híbridos potenciales y la carencia de una estructura cristalina del IFN- α o del receptor de IFN- α , se observa un impasse en el desarrollo de híbridos novedosos (Horisberger y Di Marco, Pharm. Ther. 66:507-534 (1995)).

Los efectos biológicos de los IFN- α son diversos, e incluyen propiedades tales como la inducción de un estado antiviral (inducción de factores que detienen la traducción y degradan el ARNm); inhibición del crecimiento celular; inducción del MHC de Clase I y Clase II; activación de monocitos y macrófagos; activación de células asesinas naturales; activación de células T citotóxicas; modulación de la síntesis de Ig en células B; y actividad pirogénica.

Los diferentes subtipos de IFN- α tienen un espectro único de actividades sobre diferentes células diana y perfiles de efectos secundarios únicos (Ortaldo et al., Proc. Natl. Acad. Sci. (U.S.A.) 81:4926-4929 (1984); Overall et al., J. Interferon Res. 12:281-288 (1992); Fish y Stebbing, Biochem. Biophys. Res. Comm. 112:537-546 (1983); Weck et al., J. Gen. Virol. 57:233-237 (1981)). Por ejemplo, el IFN α humano tiene efectos secundarios muy suaves pero poca actividad antiviral. El IFN α 8 humano tiene una actividad antiviral muy elevada, pero efectos secundarios relativamente graves. El IFN α 7 humano carece de actividad NK y bloquea la estimulación de NK por otros IFN α . El IFN- α J humano carece de capacidad para estimular las células NK, pero se puede unir al receptor de IFN- α en las células NK y bloquear la actividad estimuladora de IFN- α A (Langer et al., J. Interferon Res. 6:97-105 (1986)).

Las aplicaciones terapéuticas de los interferones están limitadas por diversos y graves perfiles de efectos secundarios que incluyen síntomas de tipo gripal, fatiga, trastornos neurológicos incluyendo alucinaciones, fiebre, elevación de las enzimas hepáticas, y leucopenia. La multiplicidad de efectos de los IFN- α ha estimulado la hipótesis de que hay más de un receptor o un receptor con múltiples componentes para la familia del IFN- α (R. Hu et al., J. Biol. Chem. 268:12591-12595 (1993)). De este modo, la existencia de una diversidad natural abundante en los IFN alfa humanos (y por tanto un gran espacio de secuencias de recombinantes) junto con la complejidad de los receptores y actividades de IFN- α crea una oportunidad para la construcción de híbridos superiores.

A. Complejidad del Espacio de Secuencias

La Figura 2 muestra las secuencias de proteínas de 11 IFN- α humanos. Se indican las diferencias de consenso. Las posiciones en las que un codón degenerado puede capturar toda la diversidad se indican con un asterisco. El examen de las secuencias alineadas revela que hay 57 posiciones con dos, 15 posiciones con tres, y 4 posiciones con cuatro posibles aminoácidos codificados por este grupo de genes de interferón alfa. De este modo, la diversidad potencial codificada por la permutación de toda esta diversidad natural es: $2^{57} \times 3^{15} \times 4^4 = 5,3 \times 10^{26}$. Entre estos híbridos, de los 76 polimorfismos diseminados a lo largo de un total de 175 sitios en los 11 genes de interferón, 171 de los 175 cambios pueden ser incorporados en genotecas homólogas utilizando codones degenerados individuales en las correspondientes posiciones. Por ejemplo, Arg, Trp y Gly pueden estar todos codificados por el codón degenerado (A,T,G)GG. Utilizando semejante estrategia, se pueden capturar $1,3 \times$

10²⁵ híbridos con un *único grupo* de oligonucleótidos degenerados. Como resulta evidente a partir de las Tablas III a VI, 27 oligonucleótidos son suficientes para barajar los once interferones alfa humanos. Virtualmente toda la diversidad natural es codificada de ese modo y completamente permutada debido a las degeneraciones de los nueve "bloques" de oligonucleótidos de la Tabla V.

B. Propiedades de una Búsqueda "Grosera" del Espacio de Secuencias Homólogas

La estructura modelada del IFN alfa (Kontsek, Acta Vir. 38:345-360 (1994)) ha sido dividida en nueve segmentos basados en una combinación de criterios de mantenimiento de elementos de la estructura secundaria como unidades únicas y colocación/selección de los límites de segmentos en regiones de elevada identidad. Por consiguiente, se puede capturar toda la familia con un *único grupo* de oligonucleótidos ligeramente degenerados. La Tabla III y la Figura 2 proporcionan localizaciones precisas de estos límites a nivel de proteína y ADN, respectivamente. Se debe hacer hincapié en que este esquema de segmentación concreto es arbitrario y que también se podrían perseguir otros esquemas de segmentación. La estrategia general no depende de la situación de los límites de recombinación en las regiones de alta identidad entre miembros de la familia ni en ningún algoritmo concreto para romper la estructura en segmentos.

Tabla III

Esquema de Segmentación para el Interferón Alfa			
Segmento	Amino-ácidos	Núm. Alelos	Núm. Permutaciones de todas las Variaciones de Secuencia
1	1-21	5	1024
2	22-51	10	$6,2 \times 10^4$
3	52-67	6	96
4	68-80	7	1024
5	81-92	7	192
6	93-115	10	$2,5 \times 10^5$
7	116-131	4	8
8	132-138	4	8
9	139-167	9	9216

Muchos de los IFN son idénticos a lo largo de algunos segmentos, y de este modo hay menos de once "alelos" diferentes de cada segmento. Así, una genoteca que constara de las permutaciones de los alelos del "segmento" tendría una complejidad potencial de $2,1 \times 10^7$ (5 segmentos Núm. 1 x 10 segmentos Núm. 2 x x 9 segmentos Núm.9). Esto es bastante más de los que podría ser examinado en la mayoría de los procedimientos de escrutinio descritos, y por lo tanto es un buen problema para utilizar la RSR para investigar el espacio de secuencia.

C. Estrategias Detalladas para Utilizar RSR para la Búsqueda del Espacio de Secuencias Homólogas de IFN-alfa

Los métodos descritos en la presente memoria para el barajado dirigido por oligos (esto es oligonucleótidos

puente) se emplean para construir genotecas de híbridos de interferón alfa, y los métodos generales descritos antes se emplean para escrutar o seleccionar estos mutantes en busca de una función mejorada. Como existen
5 numerosos formatos en los cuales escrutar o seleccionar en busca de actividad de interferón mejorada, muchos de los cuales dependen de las propiedades únicas de los interferones, más abajo se representan descripciones ilustrativas de los análisis basados en IFN.

D. Protocolo para una Búsqueda Grosera del Espacio de Secuencia del IFN Alfa Híbrido

En resumen, se construyen genotecas en las que se
15 permutan 11 formas homólogas de los nueve segmentos (obsérvese que en muchos casos dos homólogos son idénticos a lo largo de un segmento dado). Los nueve segmentos son amplificados por PCR de los once genes totales de IFN alfa con los dieciocho oligonucleótidos
20 enumerados en la Tabla IV, y se vuelven a ensamblar en los genes completos con recombinación dirigida por oligos. Se prepara un número arbitrario, p. ej., 1000, clones a partir de la genoteca en un formato de expresión/purificación de 96 pocillos. Se escrutan los
25 híbridos con las actividades antivirales más potentes. Se recupera el ácido nucleico mediante amplificación por PCR, y se somete a recombinación utilizando oligonucleótidos puente. Estas etapas se repiten hasta obtener candidatos con las propiedades deseadas.

E. Estrategias para Examinar el Espacio de $>10^{26}$ Híbridos Finos

En resumen, cada uno de los nueve segmentos es
35 sintetizado con un oligo degenerado por segmento. Las

degeneraciones se eligen para capturar toda la diversidad de IFN-alfa que puede ser capturada con un único codón degenerado sin añadir ninguna secuencia no natural. Se genera un segundo grupo de oligonucleótidos degenerados que codifican los nueve segmentos donde se captura toda la diversidad natural, pero se incluyen mutaciones no naturales adicionales en las posiciones en las que se necesita por las restricciones del código genético. En la mayoría de los casos se puede capturar toda la diversidad con un único codón degenerado; en algunos casos un codón degenerado captura toda la diversidad natural pero añadirá una mutación no natural; en unas pocas posiciones no es posible capturar la diversidad natural sin introducir un codón altamente degenerado que creará más de una mutación no natural. Es en estas posiciones en las que este segundo grupo de oligonucleótidos diferirán del primer grupo al ser más inclusivas. Cada uno de los nueve segmentos sintéticos es amplificado después mediante PCR con los 18 nucleótidos de PCR. Se generan genes completos que utilizan el método de recombinación dirigida por oligos, se transfectan a un anfitrión, y se analizan en busca de híbridos con las propiedades deseadas. Los mejores híbridos (p. ej., del 10%, 1% o 0,1% de más arriba; preferiblemente el 1% de más arriba) se someten a SRS y el procedimiento se repite hasta obtener un candidato con las propiedades deseadas.

F. Búsqueda Fina "No moderada"

Por otra parte, se podrían elaborar genotecas en las que cada segmento deriva de los oligonucleótidos sintéticos degenerados que codifican permutaciones al azar de la diversidad de homólogos. En este caso, la genoteca inicial buscará muy escasamente el espacio de $>10^{25}$ híbridos finos posibles que son posibles con esta

familia de genes. Se podría continuar cultivando los positivos juntos a partir de esta búsqueda. No obstante, habría numerosas diferencias entre los miembros independientes de tales genotecas, y por consiguiente el proceso de cultivo no sería muy moderado debido a que las reservas de los genes relativamente divergentes se recombinarían en cada etapa.

G. Búsqueda Fina "Moderada"

Un modo de elaborar este enfoque más "moderado" sería obtener un punto de partida candidato y buscar moderadamente a partir de allí. Este punto de partida podría ser uno cualquiera de los IFN alfa naturales (tales como IFN alfa-2a que es el que está siendo más ampliamente utilizado terapéuticamente), el interferón consenso IFN-Con1 caracterizado, o un éxito del IFN-alfa barajado antes. Dado un punto de partida, se elaborarían genotecas separadas en las que se cultivarían las genotecas de segmentos degeneradas una cada vez en la secuencia fundadora. Los éxitos mejorados de cada genoteca se cultivarían después juntos para construir moderadamente mutaciones en toda la molécula.

H. Análisis Celulares Funcionales

Los siguientes análisis, bien conocidos en la técnica, se utilizan para escrutar mutantes de IFN alfa: inhibición de la eliminación viral; error típico de 30-50%; inhibición de unidades formadoras de placas; error típico muy bajo (puede medir efectos pequeños); rendimiento viral reducido (útil para virus no formadores de placa, no letales); inhibición del crecimiento celular (análisis de absorción de timidina-H3; activación de células NK para eliminar células tumorales; supresión de

la formación de tumores por IFN humano administrado a ratones carentes de sistema inmunitario injertados con tumores humanos (tumores de piel por ejemplo).

5 La mayor parte de estos análisis son susceptibles de
escrutinio de elevado rendimiento. Las genotecas de
mutantes de IFN alfa recombinantes se expresan y
purifican en formatos de elevado rendimiento tales como
10 la expresión, lisis y purificación en un formato de 96
pocillos utilizando anticuerpos anti-IFN o una etiqueta
epitópica y una resina de afinidad. Las preparaciones de
IFN purificado se escrutan en un formato de elevado
rendimiento, se puntúan, y los mutantes que codifican las
15 actividades de interés más elevadas se someten a
mutagénesis adicional, tal como RSR, y el procedimiento
se repite hasta obtener un nivel de actividad deseado.

I. Presentación en Fagos

20 Se utilizan formatos de presentación en fagos
convencionales para presentar IFN biológicamente activos.
Las genotecas de genes de IFN quiméricos se expresan en
este formato y se seleccionan (positivamente o
negativamente) en busca de la unión (o unión reducida) a
25 una o más preparaciones de receptores de IFN purificados
o a uno o más tipos de células que expresan receptores de
IFN.

J. GFP o Luciferasa Bajo el Control de un Promotor 30 Dependiente de IFN-Alfa

Las proteínas expresadas por mutantes pueden ser
escrutadas en un formato de elevado rendimiento sobre una
línea celular informadora que expresa GFP o luciferasa
35 bajo el control de un promotor sensible a IFN alfa, tal

como un promotor de MHC de Clase I que conduce la expresión de GFP.

K. Estimulación de Células Diana con Partículas Infecciosas Intactas

La purificación de IFN activo limitará el rendimiento de los análisis descritos más arriba. La expresión de IFN alfa activo sobre el fago filamentoso M13 permitiría obtener preparaciones homogéneas de mutantes de IFN en un formato en el que fácilmente se podrían manipular miles o decenas de miles de mutantes. Gram et al. (J. Imm. Meth. 161:169-176 (1993)) han demostrado que la IL3 humana, una citoquina con un plegamiento de proteína con una topología similar a la del IFN alfa, puede ser expresada sobre la superficie de M13 y que el fago resultante puede presentar IL3 activa a líneas celulares dependientes de IL3. De un modo similar, Saggio et al. (Gene 152:35-39 (1995)) han demostrado que el factor neurotrófico ciliar humano, una citoquina del haz de cuatro hélices, es biológicamente activo cuando se expresa sobre fagos a concentraciones similares a las de la citoquina soluble. Análogamente, se pueden expresar genotecas de mutantes de IFN alfa sobre M13 y utilizar soluciones de partida de fago de título definido para presentar IFN biológicamente activo en los análisis de alto rendimiento y selecciones descritos en la presente memoria.

El siguiente cálculo apoya la viabilidad de aplicar esta tecnología al IFN alfa. Suponiendo (1) títulos de 1×10^{10} fagos/ml con cinco copias activas de interferón presentadas por fago, y (2) que el interferón presentado es equivalentemente activo al interferón recombinante soluble (bien podría ser más potente debido a la multi-

valencia), la cuestión es si se puede esperar razonablemente que sea biológicamente activo.

$$(1 \times 10^{10} \text{ fagos/ml}) \times (5 \text{ moléculas IFN/fago}) \times (1 \text{ mol}/6 \times 10^{23} \text{ moléculas}) \times (26.000 \text{ g/mol}) \times (10^9 \text{ ng/g}) = 2,2 \text{ ng/ml}$$

El intervalo de concentración utilizado en los análisis biológicos es: 1 ng/ml para la activación de NK, 0,1 - 10 ng/ml para la actividad antiproliferativa en células Eskol, y 0,1 - 1 ng/ml en células Daudi (Ozes et al., J. Interferon Res. 12:55-59 (1992)). Aunque algunos subtipos están glicosilados, el interferón alfa2 y el interferón consenso son expresados en forma recombinante activa en *E. coli*, de manera que al menos estos dos no requieren glicosilación para la actividad. De este modo, es probable que el IFN alfa expresado sobre fagos filamentosos sea biológicamente activo como solución de partida de fagos sin concentración adicional. Las genotecas de quimeras de IFN se expresan en formatos de presentación en fagos y se puntúan en los análisis descritos antes y más abajo para identificar mutantes con propiedades mejoradas para ser colocados en rondas de RSR adicionales.

Cuando un fago es suficiente para activar una célula debido al alto estado de valencia de la proteína presentada (cinco por fago en el formato del gen III; cientos por fago en el formato del gen VIII; decenas en el formato del gen lambda V), se puede utilizar una solución de partida de fagos directamente a una dilución adecuada para estimular células con un constructo informador GFP bajo el control de un promotor sensible a IFN. Suponiendo que el fago permanece anclado después de la estimulación, la expresión y la purificación FACS de las células sensibles, se podrían purificar después

directamente por FACS los híbridos con una actividad mejorada de genotecas muy grandes (hasta 10^7 y quizás mas fagos por ronda de FACS).

5 Un segundo modo en el que se utiliza FACS para aventajar en este formato es el siguiente. Las células pueden ser estimuladas en un formato de múltiples pocillos con una solución de partida de fago por pocillo y un constructo informador de tipo GFP. Todas las células
10 estimuladas son purificadas por FACS para recoger las células más brillantes, y los genes de IFN son recuperados y sometidos a RSR, con repetición del protocolo hasta obtener el nivel de mejora deseado. En este protocolo la estimulación se realiza con productos
15 lisados concentrados individuales y por lo tanto el requerimiento de que un único fago sea suficiente para estimular la célula se relaja. Además, se puede compartimentar para recoger las células más brillantes que, a su vez, deben tener el fago más potente anclado a
20 ellas.

L. Protocolo de Presentación en la Superficie Celular para Mutantes de IFN Alfa

25 A continuación se presenta un protocolo de muestra para la presentación en la superficie celular de mutantes de IFN alfa. Esta forma de presentación tiene al menos dos ventajas sobre la presentación en fagos. Primero, la proteína es presentada por una célula eucariótica y por
30 lo tanto puede ser expresada en una forma apropiadamente glicosilada que puede ser necesaria para algunos IFN alfa (y otros factores de crecimiento). En segundo lugar, es un formato de presentación de valencia muy elevada y se prefiere para detectar la actividad de mutantes muy
35 débilmente activos.

En resumen, se construye una genoteca de IFN mutantes donde se ha fusionado una señal polipeptídica para la adición de una cola de fosfoinositol al extremo carboxilo, dirigiendo de ese modo la proteína para la expresión en la superficie (Wettstein et al., J. Exp. Med. 174:219-28 (1991)). La genoteca se utiliza para transfectar células informadoras descritas más arriba (gen informador de luciferasa) en un formato de microtitulación. Se detectan los positivos con una cámara con un dispositivo de acoplamiento de carga (CCD). Se recuperan los ácidos nucleicos mediante HIRT y retransformación del anfitrión o mediante PCR, y se someten a RSR para el desarrollo adicional.

M. Protocolo de Presentación Autocrina para la Resistencia Viral

A continuación se presenta un protocolo de muestra para la presentación autocrina de mutantes de IFN alfa. En resumen, se genera una genoteca de mutantes de IFN en un vector, lo que permite la inducción de la expresión (esto es promotor de metalotioneína) y la secreción eficaz. La línea de células receptoras que portan un casete informador sensible a IFN (GFP o luciferasa) es inducida mediante transfección con los constructos de IFN mutantes. Los mutantes que estimulan el promotor sensible a IFN son detectados mediante FACS o cámara CCD.

Una variación de este formato consiste en sensibilizar transfectantes con virus y seleccionar los supervivientes. Se podrían realizar múltiples rondas de sensibilización viral y crecimiento en cada grupo de transfectantes antes de recuperar los genes. Múltiples rondas de eliminación y crecimiento permiten una amplificación exponencial de una pequeña ventaja y por lo

tanto proporcionan una ventaja en la detección de pequeñas mejoras en la eliminación viral.

Tabla IV

Oligonucleótidos necesarios para la recombinación por bloques: 18	
Oligonucleótidos para barajado de interferón alfa	
1.	5'-TGT[G/A]ATCTG[C/T]CT[C/G]AGACC
2.	5'-GGCACAAATG[G/A/C]G[A/C]AGAATCTCTC
3.	5'-AGAGATTCT[G/T]C[C/T/G]CATTTGTGCC
4.	5'-CAGTTCCAGAAG[A/G]CT[G/C][C/A]AGCCATC
5.	5'-GATGGCT[T/G][G/C]AG[T/C]CTTCTGGAACTG
6.	5'-CTTCAATCTCTTCA[G/C]CACA
7.	5'-TGTG[G/C]TGAAGAGATTGAAG
8.	5'-.GGA[T/A][G/C]AGA[C/G][C/G]CTCCTAGA
9.	5'-TCTAGGAG[G/C][G/C]TCT[G/C][T/A]TCC
10.	5'-GAACTT[T/G/A][T/A]CCAGCAA[A/C]TGAAT
11.	5'-ATTCA[T/G]TTGCTGG[A/T][A/T/C]AAGTTC
12.	5'-GGACT[T/C]CATCCTGGCTGTG
13.	5'-CACAGCCAGGATG[G/A]AGTCC
14.	5'-AAGAATCACTCTTTATCT
15.	5'-AGATAAAGAGTGATTCTT
16.	5'-TGGGAGGTTGTCAGAGCAG
17.	5'-CTGCTCTGACAACCTCCCA
18.	5'-TCA[A/T]TCCTT[C/A]CTC[T/C]TTAA

Los corchetes indican degeneración con una mezcla igual de bases especificadas en esas posiciones. El fin

de la degeneración es permitir que este grupo de cebadores cebe todos los miembros de la familia del IFNT con una eficacia similar. La elección de la recombinación dirigida por oligos es importante debido a que estos serán "sobreescritos" en cada ciclo de cultivo y por consiguiente no pueden desarrollarse simultáneamente con el resto de la secuencia a lo largo de muchos ciclos de selección.

10 **Tabla V**

Oligonucleótidos necesarios para la recombinación "fina" de diversidad natural sobre cada uno de los nueve bloques	
Bloque	Núm. Longitud de oligo requerida
1	76
2	95
3	65
4	56
5	51
6	93
7	50
8	62
9	80

Tabla VI

Aminoácidos que se pueden alcanzar mediante una única etapa de mutación en el codón de interés.	
Aminoácido de Tipo Salvaje	Aminoácidos alcanzables mediante una mutación
W	C, R, G, L
Y	F, S, C, H, N, D

Aminoácidos que se pueden alcanzar mediante una única etapa de mutación en el codón de interés.	
Aminoácido de Tipo Salvaje	Aminoácidos alcanzables mediante una mutación
F	L, I, V, S, Y, C
L	S, W, F, I, M, V, P
V	F, L, I, M, A, D, E, G
I	F, L, M, V, T, N, K, S, R
A	S, P, T, V, D, E, G
G	V, A, D, E, R, S, C, W
M	L, I, V, T, K, R
S	F, L, Y, C, W, P, T, A, R, G, N, T, I
T	S, P, A, I, M, N, K, S, R
P	S, T, A, L, H, Q, R
C	F, S, Y, R, G, W
N	Y, H, K, D, S, T, I
Q	Y, H, K, E, L, P, R
H	Y, Q, N, D, L, P, R
D	Y, H, N, E, V, A, G
E	Q, K, D, V, A, G
R	L, P, H, Q, C, W, S, G, K, T, I, M
K	Q, N, E, R, T, I, M

Basándose en esta Tabla, se han identificado las posiciones polimórficas en el IFN alfa en las que se puede capturar toda la diversidad por medio de un codón degenerado. Se sintetizan los oligonucleótidos de la longitud indicada en la Tabla V anterior con las

degeneraciones inferidas a partir de la Tabla VI.

N. Evolución del IFN- α Mejorado

5 1. Clonación

Se clonaron genes de IFN mediante amplificación por PCR a partir de ADN genómico con 12 grupos de cebadores degenerados mediante los métodos comentados generalmente más arriba. Los productos de la PCR fueron clonados en un vector de presentación en fagémidos convencional en forma de fusiones con el gen III del bacteriófago fd. Se secuenciaron treinta clones y se compararon con los genes de IFN alfa humano de las publicaciones. La mayor parte de la secuencia coincidía exactamente o casi exactamente (identidad de ADN >98%) con secuencias conocidas. Varios clones no correspondían bien con ningún IFN conocido (esto es, identidad de aproximadamente 93%) y son nuevos genes de IFN candidatos. Un gen era un claro recombinante que presumiblemente fue creado durante la PCR. Ocho de los diez clones fueron reunidos y barajados. Estas ocho secuencias contienen aproximadamente 66% de los cambios de aminoácidos conocidos en esta familia de genes.

25 2. Barajado

Los genes fueron barajados como sigue. Se prepararon reservas de fragmentos de 20-50 pb y 50-100 pb a partir de productos digeridos con ADNasa I parciales como se ha descrito más arriba. Adicionalmente, se prepararon fragmentos de 20-100 pb a partir de productos de PCR preparativa de ADN genómico humano con el mismo grupo de 12 cebadores. Estos fragmentos deben contener toda la diversidad de secuencias del locus de interferón alfa humano. Se ensamblaron las quimeras mediante PCR de

entrecruzamiento por medio de 20 ciclos (94°C x 60", 6°C x 60", 25°C x 120") seguido de dos rondas de dilución 1:10 en tampón para PCR y re-ensamblaje mediante 20 ciclos (94°C x 30", 40°C x 30", 72°C x (30+2n)") donde n = número de ciclos. Se rescataron genes completos mediante PCR con cebadores externos y el material se clonó en el vector de presentación en fagémidos mediante métodos convencionales. Se obtuvieron genotecas de complejidades de $2,5 \times 10^4$, $3,0 \times 10^5$ y 2×10^8 a partir de los fragmentos de PCR de 20-50 pb, 50-100 pb y genómicos, respectivamente. La secuenciación de quimeras al azar verificó que el barajado había funcionado eficazmente.

3. Validación la actividad biológica de los fagémidos

Se elaboraron preps a gran escala de partícula de fagémido mediante métodos convencionales, utilizando M13 VCS como fago coadyuvante. Se indujeron los genes de fusión IFN-gen III en la fase semi log mediante la adición de arabinosa al 0,02%. Las partículas de fagémido precipitadas con PEG formaron bandas con CsCl y se sometieron a diálisis. Las partículas de fagémidos presentaban IFN activo como evidenciaba la actividad biológica de las preps de fagémido que expresaban IFN-Con1, IFN2a, o los ocho IFN de tipo salvaje clonados en un análisis antiproliferación en células Daudi humanas (células humanas) (Tymms et al., Genet. Anal. Techn. Appl. 7:53-63 (1990)).

4. Escrutinio en busca de una actividad mejorada en el análisis con células Daudi

Se utilizaron dos estrategias de escrutinio para identificar los clones con una actividad mejorada:

análisis de actividad sobre clones elegidos al azar y análisis de actividad sobre reservas en bandas de CsCl seguido de identificación de los mejores clones de las reservas más activas.

5

Como ejemplo, de las ocho quimeras elegidas al azar, tres fueron más activas que Con1, una fue intermedia entre Con1 e IFN2a, y cuatro fueron negativas. La Figura 3 representa el alineamiento de las secuencias de aminoácidos de cuatro interferones quiméricos con IFN-Con1.

10

Un ejemplo de clones reunidos es el siguiente. Se combinaron 96 clones en ocho reservas diferentes de doce y se analizaron como reservas sobre células Daudi. Se elaboraron preps para CsCl a partir de los doce clones de la reserva más activa (P12.7, o Reserva "F"). Uno de estos clones, F4, era muy activo con una actividad aproximadamente 60x mayor que Con1 y aproximadamente 1000x mayor que IFN2a. Ninguno de los IFN parentales tuvo una actividad mayor que Con1. Por lo tanto esto representa un incremento de aproximadamente 60-veces con respecto al mejor clon parental. Este clon ha sido analizado en un análisis de protección de virus humanos (células WISH) (Jilbert et al., Microbial. Path. 1:159-168 (1986) y se ha encontrado que también es más activo que Con1 en este análisis, verificando de este modo una actividad del interferón auténtica en lugar de una toxicidad generalizada.

15

20

25

30

5. Desarrollo de actividad en células de ratón

Se amplificaron mediante PCR ocho genes de IFN de ratón de tipo salvaje mediante métodos convencionales y se clonaron en el vector fagémido. Uno de estos clones

35

fue muy activo en un análisis antiviral en ratón (células de ratón) (Beilharz et al., J. Interferon Res. 9:305-314 (1988) cuando se presentó en este vector. Los ocho clones de IFN parental humano e IFN2a fueron todos inactivos; y Con1 fue débilmente activo en el análisis antiviral en ratón. Una de las ocho quimeras humanas escrutadas al azar fue más activa que Con1. Una de las ocho reservas humanas de 12 clones (Reserva "G") fue activa en el análisis en ratón. La Reserva "G" produjo un clon muy activo, G8. Una de las dieciséis reservas de noventa y seis fue activa. Esta reserva de noventa y seis se rompió en ocho reservas de doce, y dos de estas reservas fueron muy activas.

6. Interpretación

Tomados juntos, estos datos demuestran que las técnicas de recombinación descritas en la presente memoria combinadas con los métodos de escrutinio descritos en la presente memoria pueden ser utilizadas para mejorar la actividad de los interferones ya potentes sobre las células humanas. Adicionalmente, se pueden utilizar los métodos para crear una actividad "relacionada" (actividad sobre células de ratón) que no existía previamente a un nivel detectable en la población de genes de partida. Los datos demuestran adicionalmente la aplicabilidad de la presente invención para crear poblaciones de genes recombinantes con distribuciones de actividades de tipo Gaussiano a partir de las cuales se pueden obtener fácilmente recombinantes superiores.

IV. Evolución de una Luciferasa Mejorada

La luciferasa de *Photinus pyralis* fue amplificada por PCR a partir de pGL2_basic (Promega Corporation,

Madison, WI). La luciferasa de *Luciola mingrellica* fue amplificada a partir de pJGR (Devine et al., Biochim. Biophys. Acta 1173:121-132 (1993)). Ambas fueron clonadas por su codón de inicio, codificado por NcoI, en pBAD24 (Guzman et al., J. Bacter. 177:4121-4130 (1995)). Para la digestión con ADNasa 1, se amplificaron mediante PCR los genes de luciferasa, incluyendo algunas regiones colindantes por medio de los cebadores BADup (TGCACGGCGTCACACTTTGCTA) y BADdown (TACTGCCGCCAGGCAAATTCT). Los productos de la PCR se mezclaron en cantidades equimolares y se digirieron parcialmente con ADNasa 1. Se purificaron en gel fragmentos de 70 a 280 pb. Se ensamblaron fragmentos de 5 µg en un volumen de 10 µl utilizando polimerasa Taq y los 15 ciclos siguientes en un Robocycler: 94°C, 30 segundos; 6°C, 60 segundos; 25°C, 180 segundos. La muestra se diluyó 1:6 y se sometió a un ciclo durante otros 20 ciclos utilizando una mezcla 1:1 de polimerasa Taq y Pwo en la máquina de ADN (94°C, 30 segundos; 40°C, 30 segundos, 72 °C, 30 segundos). La muestra se diluyó 1:4 y se sometió a otro ciclo durante 20 ciclos utilizando una mezcla 1:1 de polimerasa Taq y Pwo en la máquina de ADN (94°C, 30 segundos; 40°C, 30 segundos; 72°, 30 segundos). Para amplificar los fragmentos de ADN ensamblados, se diluyó la reacción de ensamblaje 1:10 a 1:100 y se incluyeron los cebadores núm. 773 (TAGCGGATCCTACCTGACGC) y núm. 297 (TGAAAATCTTCTCTCATCCG) utilizando en los siguientes 25 ciclos una mezcla 1:1 de polimerasa Taq y Pwo en la máquina de ADN (94°C, 30 segundos; 45°C, 30 segundos; 72°C, 110 segundos). Los productos de la PCR se digirieron con NcoI/HindIII y se ligaron en pCKX-GFP. pCKX-GFP es pBAD24, donde el casete de la unidad reguladora de Arabinosa ClaI, NcoI fue remplazada por una variante del sistema autoinductor *lux* de *Vibrio fischeri* de pJGR (Devine et al., Biochim. Biophys. Acta 1173:121-

132 (1993)). La ligación fue transformada en XL1-Blue. Las genotecas se cultivaron en placa sobre LB-Amp200 y se hicieron crecer ON a 37°C. Las colonias se seleccionaron en seis placas de 384 pocillos y se hicieron crecer durante la noche. Los cultivos se reticularon sobre nitrocelulosa y las colonias se hicieron crecer durante la noche (ON) a 30°C. La placa se incubó durante 45 min. a 60°C. Después el filtro de nitrocelulosa se colocó sobre un papel secante que contenía citrato de Na 100 mM pH 5 que contenía Triton X-100 al 0,2% y D-Luciferina 1 mM. Esto se colocó sobre película de plástico con la nitrocelulosa y las colonias mirando hacia abajo. Este ensamblaje se colocó sobre un BIOMAX MR en un casete de película de plástico durante 30 minutos. Después del revelado, la película se puntuó visualmente, se inocularon los clones más brillantes de las placas de 384 pocillos, y estos clones se hicieron crecer ON a 30°C en 75 µl de LB-Amp en un formato de 96 pocillos. La luciferasa se extrajo de estos cultivos como sigue. Se mezcló un volumen de cultivo de 20 µl con 20 µl de tampón de lisis I (Tris-Cl 100 mM pH 7,8, Triton X-100 al 5%, DTT 10 mM, EDTA 10 mM, 2 mg/ml de sulfato de Polimixina B). Después de sacudir, la mezcla de reacción se congeló durante 1 hora a -70°C y se descongeló después a la temperatura ambiente. Se añadieron 60 µl de tampón de lisis II (Tris-Cl 100 mM pH 7,8, ADNasaI de 0,25U/µl, 1,5 mg/ml de lisozima de huevo de gallina, MgSO₄ 40 mM) y la mezcla de lisis se incubó durante 30 min a la temperatura ambiente. Se incubaron alícuotas de los productos lisados durante 30 min a diferentes temperaturas entre 30°C y 42°C. Además, las alícuotas se dejaron a la temperatura ambiente durante varios días. La actividad luciferasa de 5 µl del producto lisado convencional y de los productos lisados tratados con calor se midió utilizando 50 µl de tampón de análisis completo (Tris-Cl 20 mM pH 7,8, MgSO₄

5 mM, ATP 0,5 mM, Coenzima A 0,5 mM, D-Luciferina 0,5 mM, DTT 5 mM) en un luminómetro Topcount. Varios clones mostraron un incremento en la actividad residual después del tratamiento con calor y cuando se dejaron a la temperatura ambiente durante varios días. Un clon mostró un incremento de la actividad luciferasa de 5 veces sobre el clon de tipo salvaje de *Luciola mingrelica* en extractos de *E. coli* cuando se trató durante 30 minutos a 39°C. Después de 4 días de incubación a la temperatura ambiente, el mismo clon mostró diez veces más actividad que la luciferasa de *L. Mingrelica* de tipo salvaje que había sido tratada idénticamente. Además, este clon mostró un incremento significativo (2 veces) en la actividad por encima del tipo salvaje cuando se hizo crecer a 37°C.

Estos resultados demuestran el desarrollo de una luciferasa con una estabilidad mejorada con respecto a sustratos moleculares de donantes parentales.

Aunque la invención anterior ha sido descrita con cierto detalle por medio de ilustración y ejemplos con el fin de aclarar su comprensión, será obvio que se pueden practicar ciertos cambios y modificaciones dentro del alcance de las reivindicaciones adjuntas.

LISTA DE SECUENCIAS

5 <110> MAXYGEN INC.

 <120> MÉTODOS Y CMPOSICIONES PARA DISEÑO DE POLIPÉPTIDOS

 <130> 0103.230EP

 <140>

10 <141>

 <150> 08/769,062

 <151> 1996-12-18

 <160> 98

 <170> PatentIn Ver. 2.0

15 <210> 1

 <211> 50

 <212> ADN

 <213> Secuencia Artificial

 <220>

20 <223> Descripción de la Secuencia Artificial:
 oligonucleótido degenerado utilizado para la genoteca del
 uso de codones

 <400> 1

25 aaccctccag ttccgaaccc catatgatga tcaccctgcg 50
 taaactgccg

<210> 2

<211> 38

<212> ADN

<213> Secuencia Artificial

5 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 2

10 aaccctccag ttccgaaccc catatgaaaa aaaccgct 38

<210> 3

<211> 40

<212> ADN

<213> Secuencia Artificial

15 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 3

20 aaccctccag ttccgaaccc atatacatat gcgtgctaaa 40

<210> 4

<211> 44

<212> ADN

<213> Secuencia Artificial

25 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 4

5 aaccctccag ttccgaaccc catatgaaat acctgctgcc gacc 44

<210> 5

<211> 40

<212> ADN

<213> Secuencia Artificial

10 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 5

15 aaccctccag ttccgaaccc gatatacata tgaaacagtc 40

<210> 6

<211> 60

<212> ADN

<213> Secuencia Artificial

20 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 6

25 tgggtgttatg tctgctcagg cdatggcdgt dgaytttycay 60
ctggttccgg ttgaagagga

<210> 7

<211> 60

<212> ADN

<213> Secuencia Artificial

<220>

5 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 7

10 ggctgggtttc gctaccggttgc cdcargcdgc dccdaargay 60
ctgggttccgg ttgaagagga

<210> 8

<211> 60

<212> ADN

<213> Secuencia Artificial

15 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 8

20 caccgccgatc gctatctctt cyttygcdtc yacyggytcy 60
ctgggttccgg ttgaagagga

<210> 9

<211> 60

<212> ADN

25 <213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 9

5 gctgctggct gctcagccgg cdatggcdat ggayatyggy 60
ctggttccgg ttgaagagga

<210> 10

<211> 61

<212> ADN

10 <213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

15 <400> 10

tgccgctgct gttcaccccg gtdacyaarg cdgdcargt dctgggtccg gttgaagagg 60
a 61

<210> 11

<211> 60

<212> ADN

20 <213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

25 <400> 11

cccggctttc tggaaccgtc argcdgdcda rgcdctggac 60
gttgctaaaa aactgcagcc

<210> 12

<211> 60

<212> ADN

<213> Secuencia Artificial

5 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 12

10 acgttatacct gttcctgggt gayggyatgg gygtgccdac 60
cgttaccgct acccgatatcc

<210> 13

<211> 60

<212> ADN

15 <213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

20 <400> 13

aaactggggtc cggaaacccc dctggcdatg gaycarttgc 60
cgtacgttgc tctgtctaaa

<210> 14

<211> 60

25 <212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

5 <400> 14

ggttccggac tctgctggta cygcdacygc dtayctgtgc 60
ggtgttaaag gtaactaccg

<210> 15

<211> 60

10 <212> ADN

<213> Secuencia Artificial

<220>

15 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 15

ctgctcgta caaccagtgc aaracyacyc gyggyaayga 60
agttacctct gttatgaacc

<210> 16

20 <211> 60

<212> ADN

<213> Secuencia Artificial

<220>

25 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 16

tctgttggtg ttgttaccac yacycgygtd carcaygcdt 60
ctccggctgg tgcttacgct

<210> 17

<211> 60

5 <212> ADN

<213> Secuencia Artificial

<220>

10 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 17

gtactctgac gctgacctgc cdgcdgaygc dcaratgaac 60
ggttgccagg acatcgctgc

<210> 18

15 <211> 60

<212> ADN

<213> Secuencia Artificial

<220>

20 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 18

acatcgacgt taccctgggt ggyggycgya artayatggt 60
cccggttggt accccggacc

25 <210> 19

<211> 60

<212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
 oligonucleótido degenerado utilizado para la genoteca del
 uso de codones

<400> 19

tctgttaacg gtgttcgtaa rcgyaarc ar aayctggtdc 60
 aggcttggca ggctaaacac

<210> 20

<211> 60

<212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
 oligonucleótido degenerado utilizado para la genoteca del
 uso de codones

<400> 20

gaaccgtacc gctctgctgc argcdgcdga ygaytcytct 60
 gttaccaccac tgatgggtct

<210> 21

<211> 60

<212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
 oligonucleótido degenerado utilizado para la genoteca del
 uso de codones

<400> 21

aatacaacgt tcagcaggac cayacyaarg ayccdacyct 60
gcaggaaatg accgaagttg

<210> 22

5 <211> 60

<212> ADN

<213> Secuencia Artificial

<220>

10 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 22

aacccgcgtg gtttctacct gtttygtdgar ggyggycgya 60
tcgaccacgg tcaccacgac

15 <210> 23

<211> 60

<212> ADN

<213> Secuencia Artificial

<220>

20 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 23

25 gaccgaagct ggtatgttcg ayaaygdat ygcdaargct 60
aacgaactga cctctgaact

<210> 24

<211> 60

<212> ADN

<213> Secuencia Artificial

<220>

5 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 24

ccgctgacca ctctcacggt ttytcttytg gyggytayac 60
cctgcgtggt acctctatct

10 <210> 25

<211> 60

<212> ADN

<213> Secuencia Artificial

<220>

15 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 25

20 gctctggact ctaaattctta yacytctyaty ctgtayggga 60
acgggtccggg ttacgctctg

<210> 26

<211> 60

<212> ADN

<213> Secuencia Artificial

25 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 26

5 cgттаacgac tctacctctg argayccdtc ytaycarcag 60
caggctgctg ttccgcaggc

<210> 27

<211> 60

<212> ADN

10 <213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

15 <400> 27

aagacgттgc tgttttcgct cgyggyccdc argcdcaуct 60
ggttcacggt gttgaagaag

<210> 28

<211> 60

20 <212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

25

<400> 28

atggcttttcg ctggттgcgt dgarccdtay acygaytgya 60
acctgccggc tccgaccacc

<210> 29

<211> 61

<212> ADN

<213> Secuencia Artificial

5 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 29

10 tgctcacctg gctgcttmac cdccdcddct ggcddtgctg gctgggtgcta tgctgctcct 60
c 61

<210> 30

<211> 62

<212> ADN

<213> Secuencia Artificial

15 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 30

20 ttccgcctct agagaattct tartacagrg thgghgccag gaggagcagc atagcaccag 60
cc 62

<210> 31

<211> 58

<212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

5 <400> 31

aagcagccag gtgagcagcg tchggratrg argthgcggt 58
ggtcggagcc ggcagggt

<210> 32

<211> 60

10 <212> ADN

<213> Secuencia Artificial

<220>

15 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 32

cgcaaccagc gaaagccatg atrtghgcha craargtytc 60
ttcttcaaca ccgtgaacca

<210> 33

20 <211> 60

<212> ADN

<213> Secuencia Artificial

<220>

25 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 33

gcgaaaacag caacgtcttc rccrcrctgr gtytcrghg 60
cctgcggaac agcagcctgc

<210> 34

<211> 60

5 <212> ADN

<213> Secuencia Artificial

<220>

10 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 34

agaggtagag tcgttaacgt chggrcgrga rccrccrccc 60
agagcgtaac ccgacccgtt

<210> 35

15 <211> 60

<212> ADN

<213> Secuencia Artificial

<220>

20 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 35

aagatttaga gtccagagct ttrgahgghg ccagrcraa 60
gatagaggta ccacgcaggg

25 <210> 36

<211> 60

<212> ADN

<213> Secuencia Artificial

<220>

5 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 36

acgtgagagt ggtcagcggg haccagratc agrgtrtcca 60
gttcagaggt cagttcgtta

<210> 37

10 <211> 60

<212> ADN

<213> Secuencia Artificial

<220>

15 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 37

gaacatacca gcttcgggtca ghgcatrta hgcyttrtcg 60
tcgtggtgac cgtggtcgat

20 <210> 38

<211> 60

<212> ADN

<213> Secuencia Artificial

<220>

25 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 38

ggtagaaacc acgcgggtta cgrgahacha crcgcaghgc 60
aacttcggtc atttcctgca

<210> 39

5 <211> 60

<212> ADN

<213> Secuencia Artificial

<220>

10 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 39

tcctgctgaa cgttgtatatt catrtchgch ggytcraaca 60
gacccatcag gtgggtaaca

15 <210> 40

<211> 60

<212> ADN

<213> Secuencia Artificial

<220>

20 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 40

25 cagcagagcg gtacggttcc ahacrtaytg hgcrccytgg 60
tgtttagcct gccaaagcctg

<210> 41

<211> 60

<212> ADN

<213> Secuencia Artificial

<220>

5 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 41

tacgaacacc gttaacagaa gcrtcrtchg grtaytchgg 60
gtccggggta ccaaccggga

10 <210> 42

<211> 60

<212> ADN

<213> Secuencia Artificial

<220>

15 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 42

20 cccaggataa cgtcgatgtc catrttrtth accagytghg 60
cagcgatgtc ctggcaaccg

<210> 43

<211> 60

<212> ADN

<213> Secuencia Artificial

25 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 43

5 caggtcagcg tcagagtacc arttrcgrrt hacrgtrtga 60
gcgtaagcac cagccggaga

<210> 44

<211> 60

<212> ADN

10 <213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

15 <400> 44

tggttaacaac accaacagat ttrcchgcyt tytthgcrcg 60
gttcataaca gaggtaactt

<210> 45

<211> 60

20 <212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

25

<400> 45

cactggttgt aacgagcagc hgcrghacr ccratrgtrc 60
ggtagttacc tttaacaccg

<210> 46

<211> 60

<212> ADN

<213> Secuencia Artificial

5 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 46

10 accagcagag tccggaacct ggcgrtchac rttrtargtt 60
ttagacagag caacgtacgg

<210> 47

<211> 60

<212> ADN

15 <213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

20 <400> 47

gggtttccgg acccagttta ccrttcatyt grccyttcag 60
gatacgggta gcggtaacgg

<210> 48

<211> 60

25 <212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

5 <400> 48

cccaggaaca ggataacggt ytthgchgcr gtytgrathg 60
gctgcagttt ttagcaacg

<210> 49

<211> 42

10 <212> ADN

<213> Secuencia Artificial

<220>

15 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 49

acggttccag aaagccgggt cttcctcttc aaccggaacc ag 42

<210> 50

<211> 60

20 <212> ADN

<213> Secuencia Artificial

<220>

25 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 50

cctgagcaga cataacacca gchgchachg chachgccag 60
cggcagttta cgcagggtga

<210> 51

<211> 62

5 <212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

10

<400> 51

accggggtga acagcagcgg cagcaghgcc aghgcratrg trgactgttt catatgtata 60
tc 62

<210> 52

<211> 59

15 <212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

20

<400> 52

gccggctgag cagccagcag cagcagrcch gchgchgcgg 59
tcggcagcag gtagtttca

<210> 53

25 <211> 60

<212> ADN

<213> Secuencia Artificial

<220>

5 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 53

aagagatagc gatcgggggtg gtcaghacra trcccagcag 60
ttagcacgc atatgtatat

<210> 54

10 <211> 58

<212> ADN

<213> Secuencia Artificial

<220>

15 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 54

caacggtagc gaaaccagcc aghgchachg crathgcrat 58
agcggttttt ttcatatg

20 <210> 55

<211> 39

<212> ADN

<213> Secuencia Artificial

<220>

25 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 55

agaattctct agaggcggaa actctccaac tcccagggtt 39

<210> 56

<211> 39

5 <212> ADN

<213> Secuencia Artificial

<220>

10 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para la genoteca del
uso de codones

<400> 56

tgagaggttg aggggtccaat tgggaggtca aggcttggg 39

<210> 57

<211> 18

15 <212> ADN

<213> Secuencia Artificial

<220>

20 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

<400> 57

tgtratctgy ctsagacc 18

<210> 58

<211> 23

25 <212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

<400> 58

ggcacaaatg vgmagaatct ctc 23

<210> 59

<211> 22

<212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

<400> 59

agagattctk cbcatttggtg cc 22

<210> 60

<211> 24

<212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

<400> 60

cagttccaga agrctsmagc catc 24

<210> 61

<211> 24

<212> ADN

5 <213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

10 <400> 61

gatggctksa gycttctgga actg 24

<210> 62

<211> 19

<212> ADN

15 <213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

20 <400> 62

cttcaatctc ttcascaca 19

<210> 63

<211> 19

<212> ADN

25 <213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

5 <400> 63

tgtgstgaag agattgaag 19

<210> 64

<211> 18

<212> ADN

10 <213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

15 <400> 64

ggawsagass ctcctaga 18

<210> 65

<211> 18

<212> ADN

20 <213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

25 <400> 65

tctaggagss tctswtcc 18

<210> 66

<211> 21

<212> ADN

<213> Secuencia Artificial

5 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

<400> 66

10 gaacttdwcc agcaamtgaa t 21

<210> 67

<211> 21

<212> ADN

<213> Secuencia Artificial

15 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

<400> 67

20 attcakttgcc tggwhaagtt c 21

<210> 68

<211> 19

<212> ADN

<213> Secuencia Artificial

25 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

<400> 68

5 ggactycatc ctggctgtg 19

<210> 69

<211> 19

<212> ADN

<213> Secuencia Artificial

10 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

<400> 69

15 cacagccagg atgragtcc 19

<210> 70

<211> 18

<212> ADN

<213> Secuencia Artificial

20 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

<400> 70

25 aagaatcact ctttatct 18

<210> 71

<211> 18

<212> ADN

<213> Secuencia Artificial

<220>

5 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

<400> 71

agataaagag tgattctt 18

10 <210> 72

<211> 19

<212> ADN

<213> Secuencia Artificial

<220>

15 <223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

<400> 72

tgggagggttg tcagagcag 19

20 <210> 73

<211> 19

<212> ADN

<213> Secuencia Artificial

<220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

<400> 73

5 ctgctctgac aacctccca 19

<210> 74

<211> 18

<212> ADN

<213> Secuencia Artificial

10 <220>

<223> Descripción de la Secuencia Artificial:
oligonucleótido degenerado utilizado para barajado de
interferón alfa

<400> 74

15 tcawtccttm ctcyttaa 18

<210> 75

<211> 166

<212> PRT

<213> interferón alfa consenso

20 <400> 75

- 186 -

Cys Asp Leu Pro Gln Thr His Ser Leu Gly Asn Arg Arg Ala Leu Ile
 1 5 10 15
 Leu Leu Ala Gln Met Gly Arg Ile Ser Pro Phe Ser Cys Leu Lys Asp
 20 25 30
 Arg His Asp Phe Gly Phe Pro Gln Glu Glu Phe Asp Gly Asn Gln Phe
 35 40 45
 Gln Lys Ala Gln Ala Ile Ser Val Leu His Glu Met Ile Gln Gln Thr
 50 55 60
 Phe Asn Leu Phe Ser Thr Lys Asp Ser Ser Ala Ala Trp Glu Gln Ser
 65 70 75 80
 Leu Leu Glu Lys Phe Ser Thr Glu Leu Tyr Gln Gln Leu Asn Asp Leu
 85 90 95
 Glu Ala Cys Val Ile Gln Glu Val Gly Val Glu Glu Thr Pro Leu Met
 100 105 110
 Asn Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr
 115 120 125
 Leu Tyr Leu Thr Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val
 130 135 140
 Arg Ala Glu Ile Met Arg Ser Leu Ser Phe Ser Thr Asn Leu Gln Lys
 145 150 155 160
 Arg Leu Arg Arg Lys Asp
 165

<210> 76

<211> 166

<212> PRT

5 <213> interferón alfa humano

<400> 76

- 187 -

Cys Asp Leu Pro Gln Thr His Ser Leu Gly Asn Arg Arg Ala Leu Ile
 1 5 10 15
 Leu Leu Ala Gln Met Gly Arg Ile Ser Pro Phe Ser Cys Leu Lys Asp
 20 25 30
 Arg His Asp Phe Gly Leu Pro Gln Glu Glu Phe Asp Gly Asn Gln Phe
 35 40 45
 Gln Lys Thr Gln Ala Ile Pro Val Leu His Glu Met Ile Gln Gln Thr
 50 55 60
 Phe Asn Leu Phe Ser Thr Glu Asp Ser Ser Ala Ala Trp Glu Gln Ser
 65 70 75 80
 Leu Leu Glu Lys Phe Ser Thr Glu Leu Tyr Gln Gln Leu Asn Asn Leu
 85 90 95
 Glu Ala Cys Val Ile Gln Glu Val Gly Met Glu Glu Thr Pro Leu Met
 100 105 110
 Asn Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr
 115 120 125
 Leu Tyr Leu Thr Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val
 130 135 140
 Arg Ala Glu Ile Met Arg Ser Leu Ser Phe Ser Thr Asn Leu Gln Lys
 145 150 155 160
 Arg Leu Arg Arg Lys Asp
 165

<210> 77

<211> 166

<212> PRT

5 <213> interferón alfa humano

<400> 77

Cys Asp Leu Pro Gln Thr His Ser Leu Gly Asn Arg Arg Ala Leu Ile
 1 5 10 15
 Leu Leu Ala Gln Met Gly Arg Ile Ser Pro Phe Ser Cys Leu Lys Asp
 20 25 30
 Arg Pro Asp Phe Gly Leu Pro Gln Glu Glu Phe Asp Gly Asn Gln Phe
 35 40 45
 Gln Lys Thr Gln Ala Ile Ser Val Leu His Glu Met Ile Gln Gln Thr
 50 55 60

- 188 -

Phe Asn Leu Phe Ser Thr Glu Asp Ser Ser Ala Ala Trp Glu Gln Ser
 65 70 75 80
 Leu Leu Glu Lys Phe Ser Thr Glu Leu Tyr Gln Gln Leu Asn Asn Leu
 85 90 95
 Glu Ala Cys Val Ile Gln Glu Val Gly Met Glu Glu Thr Pro Leu Met
 100 105 110
 Asn Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr
 115 120 125
 Leu Tyr Leu Thr Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val
 130 135 140
 Arg Ala Glu Ile Met Arg Ser Leu Ser Phe Ser Thr Asn Leu Gln Lys
 145 150 155 160
 Ile Leu Arg Arg Lys Asp
 165

<210> 78

<211> 166

<212> PRT

5 <213> interferón alfa humano

<400> 78

- 189 -

Cys Asn Leu Ser Gln Thr His Ser Leu Asn Asn Arg Arg Thr Leu Met
 1 5 10 15
 Leu Leu Ala Gln Met Arg Arg Ile Ser Pro Phe Ser Cys Leu Lys Asp
 20 25 30
 Arg His Asp Phe Glu Phe Pro Gln Glu Glu Phe Asp Gly Asn Gln Phe
 35 40 45
 Gln Lys Ala Gln Ala Ile Ser Val Leu His Glu Met Met Gln Gln Thr
 50 55 60
 Phe Asn Leu Phe Ser Thr Lys Asn Ser Ser Ala Ala Trp Asp Glu Thr
 65 70 75 80
 Leu Leu Glu Lys Phe Tyr Ile Glu Leu Phe Gln Gln Met Asn Asp Leu
 85 90 95
 Glu Ala Cys Val Ile Gln Glu Val Gly Val Glu Glu Thr Pro Leu Met
 100 105 110
 Asn Glu Asp Ser Ile Leu Ala Val Lys Lys Tyr Phe Gln Arg Ile Thr
 115 120 125
 Leu Tyr Leu Met Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val
 130 135 140
 Arg Ala Glu Ile Met Arg Ser Leu Ser Phe Ser Thr Asn Leu Gln Lys
 145 150 155 160

Arg Leu Arg Arg Lys Asp
 165

<210> 79

5 <211> 166

<212> PRT

<213> interferón alfa humano

<400> 79

- 190 -

Cys Asp Leu Pro Gln Thr His Ser Leu Gly Asn Arg Arg Ala Leu Ile
 1 5 10 15
 Leu Leu Ala Gln Met Gly Arg Ile Ser His Phe Ser Cys Leu Lys Asp
 20 25 30
 Arg His Asp Phe Gly Phe Pro Glu Glu Glu Phe Asp Gly His Gln Phe
 35 40 45
 Gln Lys Thr Gln Ala Ile Ser Val Leu His Glu Met Ile Gln Gln Thr
 50 55 60
 Phe Asn Leu Phe Ser Thr Glu Asp Ser Ser Ala Ala Trp Glu Gln Ser
 65 70 75 80
 Leu Leu Glu Lys Phe Ser Thr Glu Leu Tyr Gln Gln Leu Asn Asp Leu
 85 90 95
 Glu Ala Cys Val Ile Gln Glu Val Gly Val Glu Glu Thr Pro Leu Met
 100 105 110
 Asn Val Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr
 115 120 125
 Leu Tyr Leu Thr Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val
 130 135 140
 Arg Ala Glu Ile Met Arg Ser Leu Ser Phe Ser Thr Asn Leu Gln Lys
 145 150 155 160
 Arg Leu Arg Arg Lys Asp
 165

<210> 80

<211> 166

<212> PRT

5 <213> interferón alfa humano

<400> 80

Cys Asp Leu Pro Gln Thr His Ser Leu Gly His Arg Arg Thr Met Met
 1 5 10 15
 Leu Leu Ala Gln Met Arg Arg Ile Ser Leu Phe Ser Cys Leu Lys Asp
 20 25 30

- 191 -

```

Arg His Asp Phe Arg Phe Pro Gln Glu Glu Phe Asp Gly Asn Gln Phe
      35          40          45
Gln Lys Ala Glu Ala Ile Ser Val Leu His Glu Val Ile Gln Gln Thr
      50          55          60
Phe Asn Leu Phe Ser Thr Lys Asp Ser Ser Val Ala Trp Asp Glu Arg
      65          70          75          80
Leu Leu Asp Lys Leu Tyr Thr Glu Leu Tyr Gln Gln Leu Asn Asp Leu
      85          90          95
Glu Ala Cys Val Met Gln Glu Val Trp Val Gly Gly Thr Pro Leu Met
      100          105          110
Asn Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr
      115          120          125
Leu Tyr Leu Thr Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val
      130          135          140
Arg Ala Glu Ile Met Arg Ser Phe Ser Ser Ser Arg Asn Leu Gln Glu
      145          150          155          160
Arg Leu Arg Arg Lys Glu
      165

```

<210> 81

<211> 166

<212> PRT

5 <213> interferón alfa humano

<400> 81

- 192 -

Cys Asp Leu Pro Gln Thr His Ser Leu Arg Asn Arg Arg Ala Leu Ile
 1 5 10 15
 Leu Leu Ala Gln Met Gly Arg Ile Ser Pro Phe Ser Cys Leu Lys Asp
 20 25 30
 Arg His Glu Phe Arg Phe Pro Glu Glu Glu Phe Asp Gly His Gln Phe
 35 40 45
 Gln Lys Thr Gln Ala Ile Ser Val Leu His Glu Met Ile Gln Gln Thr
 50 55 60
 Phe Asn Leu Phe Ser Thr Glu Asp Ser Ser Ala Ala Trp Glu Gln Ser
 65 70 75 80
 Leu Leu Glu Lys Phe Ser Thr Glu Leu Tyr Gln Gln Leu Asn Asp Leu
 85 90 95
 Glu Ala Cys Val Ile Gln Glu Val Gly Val Glu Glu Thr Pro Leu Met
 100 105 110
 Asn Glu Asp Phe Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr
 115 120 125

 Leu Tyr Leu Met Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val
 130 135 140
 Arg Ala Glu Ile Met Arg Ser Phe Ser Phe Ser Thr Asn Leu Lys Lys
 145 150 155 160
 Gly Leu Arg Arg Lys Asp
 165

<210> 82

5 <211> 166

<212> PRT

<213> interferón alfa humano

<400> 82

- 193 -

Cys Asp Leu Pro Gln Thr His Ser Leu Gly Asn Arg Arg Ala Leu Ile
 1 5 10 15
 Leu Leu Ala Gln Met Arg Arg Ile Ser Pro Phe Ser Cys Leu Lys Asp
 20 25 30
 Arg His Asp Phe Glu Phe Pro Gln Glu Glu Phe Asp Asp Lys Gln Phe
 35 40 45
 Gln Lys Ala Gln Ala Ile Ser Val Leu His Glu Met Ile Gln Gln Thr
 50 55 60
 Phe Asn Leu Phe Ser Thr Lys Asp Ser Ser Ala Ala Leu Asp Glu Thr
 65 70 75 80
 Leu Leu Asp Glu Phe Tyr Ile Glu Leu Asp Gln Gln Leu Asn Asp Leu
 85 90 95
 Glu Ser Cys Val Met Gln Glu Val Gly Val Ile Glu Ser Pro Leu Met
 100 105 110
 Tyr Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr
 115 120 125
 Leu Tyr Leu Thr Glu Lys Lys Tyr Ser Ser Cys Ala Trp Glu Val Val
 130 135 140
 Arg Ala Glu Ile Met Arg Ser Phe Ser Leu Ser Ile Asn Leu Gln Lys
 145 150 155 160
 Arg Leu Lys Ser Lys Glu
 165

<210> 83

<211> 166

<212> PRT

5 <213> interferón alfa humano

<400> 83

- 194 -

```

Cys Asp Leu Pro Glu Thr His Ser Leu Asp Asn Arg Arg Thr Leu Met
 1           5           10           15

Leu Leu Ala Gln Met Ser Arg Ile Ser Pro Ser Ser Cys Leu Met Asp
          20           25           30

Arg His Asp Phe Gly Phe Pro Gln Glu Glu Phe Asp Gly Asn Gln Phe
          35           40           45

Gln Lys Ala Pro Ala Ile Ser Val Leu His Glu Leu Ile Gln Gln Ile
 50           55           60

Phe Asn Leu Phe Thr Thr Lys Asp Ser Ser Ala Ala Trp Asp Glu Asp
 65           70           75           80

Leu Leu Asp Lys Phe Cys Thr Glu Leu Tyr Gln Gln Leu Asn Asp Leu
          85           90           95

Glu Ala Cys Val Met Gln Glu Glu Arg Val Gly Glu Thr Pro Leu Met
          100          105          110

Asn Ala Asp Ser Ile Leu Ala Val Lys Lys Tyr Phe Arg Arg Ile Thr
          115          120          125

Leu Tyr Leu Thr Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val
          130          135          140

Arg Ala Glu Ile Met Arg Ser Leu Ser Leu Ser Thr Asn Leu Gln Glu
          145          150          155          160

Arg Leu Arg Arg Lys Glu
          165

```

<210> 84

<211> 166

<212> PRT

5 <213> interferón alfa humano

<400> 84

- 195 -

Cys Asp Leu Pro Gln Thr His Ser Leu Gly Asn Arg Arg Ala Leu Ile
 1 5 10 15
 Leu Leu Ala Gln Met Gly Arg Ile Ser Pro Phe Ser Cys Leu Lys Asp
 20 25 30
 Arg His Asp Phe Gly Phe Pro Gln Glu Glu Phe Asp Gly Asn Gln Phe
 35 40 45
 Gln Lys Ala Gln Ala Ile Ser Val Leu His Glu Met Ile Gln Gln Thr
 50 55 60
 Phe Asn Leu Phe Ser Thr Lys Asp Ser Ser Ala Ile Trp Glu Gln Ser
 65 70 75 80
 Leu Leu Glu Lys Phe Ser Thr Glu Leu Asn Gln Gln Leu Asn Asp Met
 85 90 95

Glu Ala Cys Val Ile Gln Glu Val Gly Val Glu Glu Thr Pro Leu Met
 100 105 110
 Asn Val Asp Ser Ile Leu Ala Val Lys Lys Tyr Phe Gln Arg Ile Thr
 115 120 125
 Leu Tyr Leu Thr Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val
 130 135 140
 Arg Ala Glu Ile Met Arg Ser Phe Ser Leu Ser Lys Ile Phe Gln Glu
 145 150 155 160
 Arg Leu Arg Arg Lys Ser
 165

<210> 85

5 <211> 166

<212> PRT

<213> interferón alfa humano

<400> 85

- 196 -

Cys Asp Leu Pro Gln Thr His Ser Leu Gly Asn Arg Arg Ala Leu Ile
 1 5 10 15
 Leu Leu Ala Gln Met Gly Arg Ile Ser Pro Phe Ser Cys Leu Lys Asp
 20 25 30
 Arg Pro Asp Phe Gly Leu Pro Gln Glu Glu Phe Asp Gly Asn Gln Phe
 35 40 45
 Gln Lys Thr Gln Ala Ile Ser Val Leu His Glu Met Ile Gln Gln Thr
 50 55 60
 Phe Asn Leu Phe Ser Thr Glu Asp Ser Ser Ala Ala Trp Glu Gln Ser
 65 70 75 80
 Leu Leu Glu Lys Phe Ser Thr Glu Leu Tyr Gln Gln Leu Asn Asn Leu
 85 90 95
 Glu Ala Cys Val Ile Gln Glu Val Gly Met Glu Glu Thr Pro Leu Met
 100 105 110
 Asn Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr
 115 120 125
 Leu Tyr Leu Thr Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val
 130 135 140
 Arg Ala Glu Ile Met Arg Ser Leu Ser Phe Ser Thr Asn Leu Gln Lys
 145 150 155 160
 Ile Leu Arg Arg Lys Asp
 165

<210> 86

<211> 166

<212> PRT

5 <213> interferón alfa humano

<400> 86

- 197 -

Cys Asp Leu Pro Gln Thr His Ser Leu Gly Asn Arg Arg Ala Leu Ile
 1 5 10 15
 Leu Leu Ala Gln Met Gly Arg Ile Ser His Phe Ser Cys Leu Lys Asp
 20 25 30
 Arg Tyr Asp Phe Gly Phe Pro Gln Glu Val Phe Asp Gly Asn Gln Phe
 35 40 45
 Gln Lys Ala Gln Ala Ile Ser Ala Phe His Glu Met Ile Gln Gln Thr
 50 55 60
 Phe Asn Leu Phe Ser Thr Lys Asp Ser Ser Ala Ala Trp Asp Glu Thr
 65 70 75 80
 Leu Leu Asp Lys Phe Tyr Ile Glu Leu Phe Gln Gln Leu Asn Asp Leu
 85 90 95
 Glu Ala Cys Val Thr Gln Glu Val Gly Val Glu Glu Ile Ala Leu Met
 100 105 110
 Asn Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr
 115 120 125
 Leu Tyr Leu Met Gly Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val
 130 135 140
 Arg Ala Glu Ile Met Arg Ser Phe Ser Phe Ser Thr Asn Leu Gln Lys
 145 150 155 160
 Gly Leu Arg Arg Lys Asp
 165

<210> 87

<211> 501

<212> ADN

5 <213> interferón alfa consenso

<400> 87

tgtgatctgc ctcagaccca cagcctgggt aataggaggg ccttgataact cctggcacia 60
 atgggaagaa tctctccttt ctctgcctg aaggacagac atgactttgg atttccccag 120
 gaggagtgtg atggcaacca gttccagaag gctcaagcca tctctgtcct ccatgagatg 180
 atccagcaga ctttcaatct cttcagcaca aaggactcat ctgctgcttg ggatgagagc 240
 ctctagaaaa aattttccac tgaactttac cagcaactga atgacctgga agcctgtgtg 300
 atacaggagg ttgggggtgga agagactccc ctgatgaatg aggactccat cctggctgtg 360
 aggaaatact tccaaagaat cactctttat ctgacagaga agaaatacag cccttgtgcc 420
 tgggaggttg tcagagcaga aatcatgaga tccttctctt tttcaacaaa cttgcaaaaa 480
 agattaagga ggaaggattg a 501

<210> 88

<211> 501

10 <212> ADN

<213> interferón alfa humano

<400> 88

```
tgtgatctgc ctcagaccca cagcctgggt aataggaggg ccttgatact cctggcacia 60
atgggaagaa tctctccttt ctctgcctg aaggacagac atgacttttg acttccccag 120
gaggagtttg atggcaacca gttccagaag actcaagcca tccctgtcct ccatgagatg 180
atccagcaga ccttcaatct cttcagcaca gaggactcat ctgctgcttg ggaacagagc 240
ctcctagaaa aattttccac tgaactttac cagcaactga ataacctgga agcatgtgtg 300
atagaggagg ttgggatgga agagactccc ctgatgaatg aggactccat cctggctgtg 360
aggaaatact tccaaagaat cactctttat ctaacagaga agaaatacag cccttgtgcc 420
tgggaggttg tcagagcaga aatcatgaga tccctctctt tttcaacaaa cttgcaaaaa 480
agattaagga ggaaggattg a 501
```

<210> 89

5

<211> 501

<212> ADN

<213> interferón alfa humano

<400> 89

```
tgtgatctgc ctcagaccca cagcctgggt aataggaggg ccttgatact cctggcacia 60
atgggaagaa tctctccttt ctctgcctg aaggacagac ctgacttttg acttccccag 120
gaggagtttg atggcaacca gttccagaag actcaagcca tctctgtcct ccatgagatg 180
atccagcaga ccttcaatct cttcagcaca gaggactcat ctgctgcttg ggaacagagc 240
ctcctagaaa aattttccac tgaactttac cagcaactga ataacctgga agcatgtgtg 300
atacaggagg ttgggatgga agagactccc ctgatgaatg aggactccat cctggctgtg 360
aggaaatact tccaaagaat cactctttat ctaacagaga agaaatacag cccttgtgcc 420
tgggaggttg tcagagcaga aatcatgaga tctctctctt tttcaacaaa cttgcaaaaa 480
atattaagga ggaaggattg a 501
```

10

<210> 90

<211> 501

<212> ADN

<213> interferón alfa humano

<400> 90

```
tgtaatctgt ctcaaaccca cagcctgaat aacaggagga ctttgatgct catggcacia 60
atgaggagaa tctctccttt ctctgcctg aaggacagac atgactttga atttccccag 120
gaggaaattg atggcaacca gttccagaaa gctcaagcca tctctgtcct ccatgagatg 180
atgcagcaga ccttcaatct cttcagcaca aagaactcat ctgctgcttg ggatgagacc 240
ctcctagaaa aattctacat tgaacttttc cagcaaatga atgacctgga agcctgtgtg 300
atacaggagg ttgggggtgga agagactccc ctgatgaatg aggactccat cctggctgtg 360
aagaaatact tccaaagaat cactctttat ctgatggaga agaaatacag cccttgtgcc 420
tgggaggttg tcagagcaga aatcatgaga tccctctctt tttcaacaaa cttgcaaaaa 480
agattaagga ggaaggattg a 501
```

15

<210> 91

<211> 501

<212> ADN

<213> interferón alfa humano

5 <400> 91

```
tgtgatctgc ctcagaccca cagcctgggt aataggaggg ccttgatact cctggcacaa 60
atgggaagaa tctctccttt ctcatgcctg aaggacagac atgatttcgg attccccgag 120
gaggagtgtg atggccacca gttccagaag actcaagcca tctctgtcct ccatgagatg 180
atccagcaga cttcaatct cttcagcaca gaggactcat ctgctgcttg ggaacagagc 240
```

```
ctcctagaaa aattttccac tgaactttac cagcaactga atgacctgga agcatgtgtg 300
atacaggagg ttgggggtgga agagactccc ctgatgaatg tggactccat cctggctgtg 360
aggaaatact tccaaagaat cactctttat ctaacagaga agaaatacag cccttgtgcc 420
tgggaggttg tcagagcaga aatcatgaga tccctctcgt tttcaacaaa cttgcaaaaa 480
agattaagga ggaaggattg a 501
```

<210> 92

10 <211> 501

<212> ADN

<213> interferón alfa humano

<400> 92

```
tgtgatctgc ctcagaccca cagcctgggt cacaggagga ccatgatgct cctggcacaa 60
atgaggagaa tctctctttt ctctgtctg aaggacagac atgacttcag atttccccag 120
gaggagtgtg atggcaacca gttccagaag gctgaagcca tctctgtcct ccatgaggtg 180
attcagcaga ctttcaatct cttcagcaca aaggactcat ctggtgcttg ggatgagagg 240
cttctagaca aactctatac tgaactttac cagcagctga atgacctgga agcctgtgtg 300
atgcaggagg tgtgggtggg agggactccc ctgatgaatg aggactccat cctggctgtg 360
agaaaatact tccaaagaat cactctctac ctgacagaga aaaagtacag cccttgtgcc 420
tgggaggttg tcagagcaga aatcatgaga tccttctctt catcaagaaa cttgcaagaa 480
aggttaagga ggaaggaata a 501
```

15 <210> 93

<211> 501

<212> ADN

- 200 -

<213> interferón alfa humano

<400> 93

```

tgtgatctgc ctcagaccca cagcctgcgt aataggaggg ccttgatact cctggcacia 60
atgggaagaa tctctccttt ctctctgcttg aaggacagac atgaattcag attcccagag 120
gaggagtgtg atggccacca gttccagaag actcaagcca tctctgtcct ccatgagatg 180
atccagcaga ctttcaatct cttcagcaca gaggactcat ctgctgcttg ggaacagagc 240
ctcctagaaa aattttccac tgaactttac cagcaactga atgacctgga agcatgtgtg 300
atacaggagg ttgggggtgga agagactccc ctgatgaatg aggactccat cctggctgtg 360
aggaaatact tccaaagaat cactctttat ctaatggaga agaaatacag cccttggtgcc 420
tgggaggttg tcagagcaga aatcatgaga tccttctctt tttcaacaaa cttgaaaaaa 480
gattaagga ggaaggattg a 501

```

<210> 94

5 <211> 501

<212> ADN

<213> interferón alfa humano

<400> 94

```

tgtgatctgc ctcagactca cagcctgggt aacaggaggg ccttgatact cctggcacia 60
atgcgaagaa tctctccttt ctctctgctg aaggacagac atgactttga attccccag 120
gaggagtgtg atgataaaca gttccagaag gctcaagcca tctctgtcct ccatgagatg 180
atccagcaga ctttcaacct cttcagcaca aaggactcat ctgctgcttt ggatgagacc 240
cttctagatg aattctacat cgaacttgac cagcagctga atgacctgga gtctgtgtg 300
atgcaggaag tgggggtgat agagtctccc ctgatgaatg aggacttcat cctggctgtg 360
aggaaatact tccaaagaat cactctatat ctgacagaga agaaatacag ctcttggtgcc 420
tgggaggttg tcagagcaga aatcatgaga tccttctctt tatcaatcaa cttgcaaaaa 480
agattgaaga gtaaggaatg a 501

```

10 <210> 95

<211> 501

<212> ADN

<213> interferón alfa humano

<400> 95

```

tgtgatctcc ctgagaccca cagcctggat aacaggagga ccttgatgct cctggcacia 60
atgagcagaa tctctccttc ctctgtctg atggacagac atgacttttg atttccccag 120
gaggagtgtg atggcaacca gttccagaag gctccagcca tctctgtcct ccatgagctg 180
atccagcaga tcttcaacct cttctccaca aaagattcat ctgctgcttg ggatgaggac 240
ctcctagaca aattctgcac cgaactctac cagcagctga atgacttgga agcctgtgtg 300
atgcaggagg agagggtggg agaaactccc ctgatgtacg cggactccat cctggctgtg 360
aagaaatact tccaaagaat cactctctat ctgacagaga agaaatacag cccttggtgcc 420
tgggaggttg tcagagcaga aatcatgaga tccttctctt tatcaacaaa cttgcaagaa 480
agattaagga ggaaggaata a 501

```

15

- 201 -

<210> 96

<211> 501

<212> ADN

<213> interferón alfa humano

5

<400> 96

```

tgtgatctgc ctcagaccca cagcctgggt aataggaggg ccttgatact cctggcacaa 60
atgggaagaa tctctccttt ctctgcctg aaggacagac atgacttttg attcccccaa 120
gaggagtttg atggcaacca gttccagaag gctcaagcca tctctgtcct ccatgagatg 180
atccagcaga cttcaatct cttcagcaca aaggactcat ctgctacttg ggaacagagc 240
ctcctagaaa aattttccac tgaacttaac cagcagctga atgacatgga agcctgcgtg 300
atacaggagg ttggggtgga agagactccc ctgatgaatg tggactctat cctggctgtg 360
aagaaatact tccaaagaat cactctttat ctgacagaga agaaatacag cccttgtgct 420
tgggaggttg tcagagcaga aatcatgaga tccttctctt tatcaaaaat ttttcaagaa 480
agattaagga ggaaggaatg a 501

```

<210> 97

<211> 501

<212> ADN

10

<213> interferón alfa humano

<400> 97

```

tgtgatctgc ctcagaccca cagcctgggt aataggaggg ccttgatact cctggcacaa 60
atgggaagaa tctctccttt ctctgcctg aaggacagac ctgacttttg acttccccag 120
gaggagtttg atggcaacca gttccagaag actcaagcca tctctgtcct ccatgagatg 180
atccagcaga cttcaatct cttcagcaca gaggactcat ctgctgcttg ggaacagagc 240
ctcctagaaa aattttccac tgaactttac cagcaactga ataacctgga agcatgtgtg 300
atacaggagg ttgggatgga agagactccc ctgatgaatg aggactccat cttggctgtg 360
aggaaatact tccaaagaat cactctttat ctaacagaga agaaatacag cccttgtgcc 420
tgggaggttg tcagagcaga aatcatgaga tctctctctt tttcaacaaa cttgcaaaaa 480
agattaagga ggaaggattg a 501

```

<210> 98

<211> 501

15

<212> ADN

<213> interferón alfa humano

<400> 98

```
tgtgatctgc ctcagactca cagcctgggt aataggaggg ccttgatact cctggcacia 60
atgggaagaa tctctcatth ctcctgcctg aaggacagat atgatttcgg attccccag 120
gaggtgtttg atggcaacca gttccagaag gctcaagcca tctctgcctt ccatgagatg 180
atccagcaga ctttcaatct cttcagcaca aaggattcat ctgctgcttg ggatgagacc 240
ctcctagaca aattctacat tgaacttttc cagcaactga atgacctaga agcctgtgtg 300
acacaggagg ttgggggtgga agagattgcc ctgatgaatg aggactccat cctggctgtg 360
```

```
aggaaatact ttcaaagaat cactctttat ctgatggaga agaaatacag cccttgtgcc 420
tgggaggttg tcagagcaga aatcatgaga tccttctctt tttcaacaaa cttgcaaaaa 480
ggattaagaa ggaaggattg a 501
```

REIVINDICACIONES

1. Un método para desarrollar una proteína codificada por una molécula sustrato de ADN recomblando al menos una primera y una segunda moléculas sustrato de ADN, donde las al menos primera y segunda moléculas sustrato difieren entre sí en al menos dos posiciones de nucleótidos y cada una comprende una pluralidad de segmentos definidos, comprendiendo el método:

(a) proporcionar un grupo de cebadores oligonucleotídicos puente, que comprende al menos un cebador para cada hebra de cada uno de la pluralidad de segmentos definidos, donde la secuencia del cebador es complementaria a al menos una unión con otro segmento definido;

(b) amplificar los segmentos definidos de las al menos primera y segunda moléculas sustrato de ADN con los cebadores de la etapa (a) en una reacción en cadena de la polimerasa;

(c) ensamblar los productos de la etapa (b) para generar una genoteca diversa de moléculas sustrato de ADN recombinantes;

(d) escrutar o seleccionar las proteínas codificadas por los productos de la etapa (c) en busca de una propiedad deseada; y

(e) recuperar al menos una molécula sustrato de ADN recombinante de la etapa (d) que codifica una proteína desarrollada.

2. El método de la reivindicación 1, donde las al menos primera y segunda moléculas sustrato de ADN se someten a mutagénesis antes de la etapa (a).

3. El método de la reivindicación 1, donde las al menos primera y segunda moléculas sustrato de ADN comprenden alelos de un gen.

5 4. El método de la reivindicación 1, donde las al menos primera y segunda moléculas sustrato de ADN comprenden una genoteca de mutantes.

10 5. El método de la reivindicación 1, donde los segmentos se definen por los sitios de las regiones intergénicas.

15 6. El método de la reivindicación 1, donde los segmentos se definen por los sitios de los intrones.

7. El método de la reivindicación 1, donde los cebadores comprenden una sustitución de uracilo en uno o más de los residuos de timidina.

20 8. El método de la reivindicación 7, donde los productos de la etapa (b) se tratan con uracilo glicosilasa.

25 9. El método de la reivindicación 1, donde se repiten las etapas (a) - (e).

30 10. El método de la reivindicación 1, donde las al menos primera y segunda moléculas sustrato de ADN comprenden cada una una agrupación de genes.

11. El método de la reivindicación 1, donde las al menos primera y segunda moléculas sustrato de ADN codifican cada una toda o parte de una ADN polimerasa.

12. El método de la reivindicación 1, donde al menos un cebador difiere de las al menos primera y segunda moléculas sustrato de ADN en al menos un nucleótido.

5 13. El método de la reivindicación 12, donde el cebador comprende una secuencia de nucleótidos de un mutante o polimorfismo de las al menos primera y segunda moléculas sustrato de ADN.

10 14. El método de la reivindicación 13, donde el cebador es degenerado y codifica las secuencias de nucleótidos de más de un mutante o polimorfismo de las al menos primera o segunda moléculas sustrato de ADN.

15 15. El método de la reivindicación 14, donde las al menos primera y segunda moléculas sustrato de ADN codifican cada una toda o una parte de una proteína seleccionada de la Tabla I o interferón alfa.

20 16. El método de la reivindicación 2, donde la mutagénesis comprende la recombinación repetitiva de secuencias.

25 17. El método de la reivindicación 1, donde los productos de la etapa (e) se someten a mutagénesis para producir moléculas sustrato de ADN recombinantes.

30 18. El método de la reivindicación 17, donde la mutagénesis comprende la recombinación repetitiva de secuencias.

19. El método de la reivindicación 17, donde las moléculas sustrato de ADN recombinantes mutagenizadas se utilizan en la etapa (b).

20. El método de la reivindicación 1, donde la molécula sustrato de ADN recombinante de la etapa (e) se utiliza como molécula sustrato de ADN en la etapa (b).

5 21. El método de la reivindicación 1, donde la molécula sustrato de ADN recombinante de la etapa (e) comprende una genoteca de moléculas sustrato de ADN recombinantes.

10 22. El método de la reivindicación 1, donde la genoteca comprende más de 10^5 miembros.

15

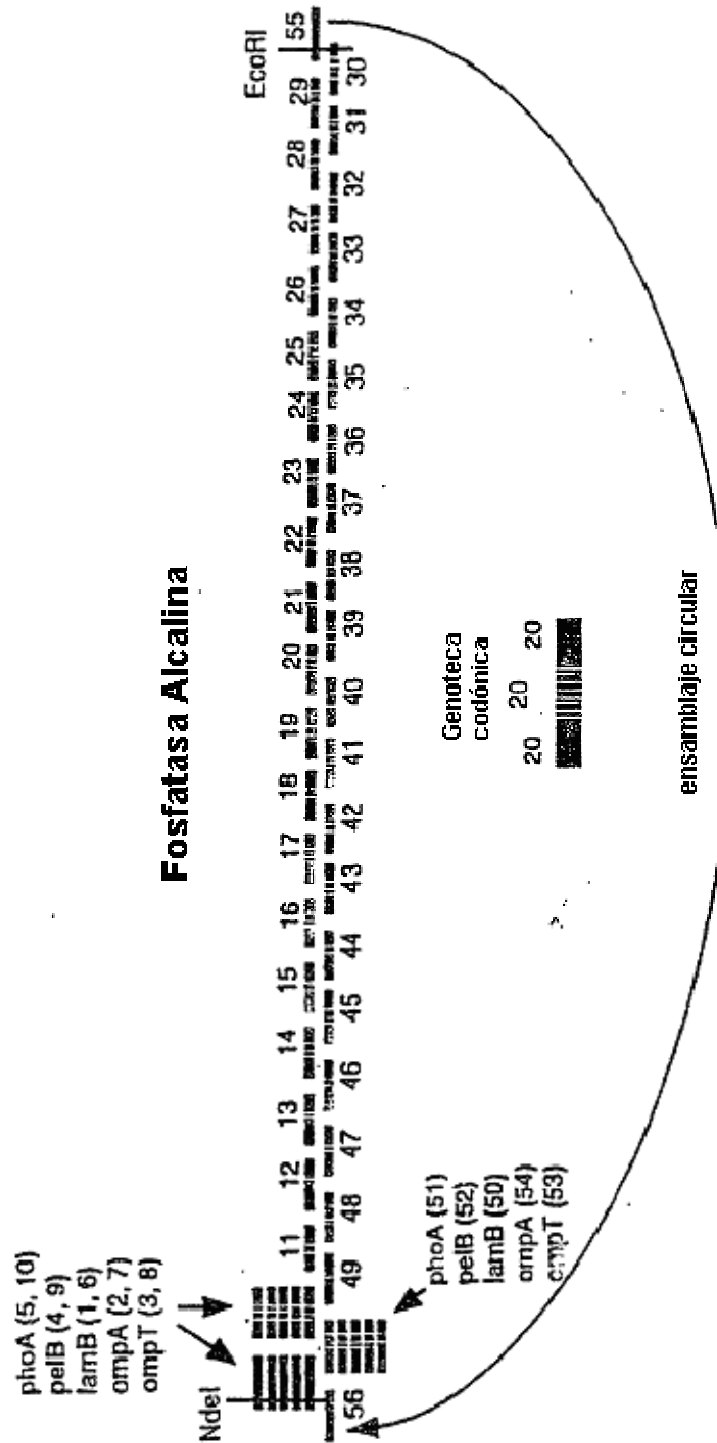


Figura 1

Figuras de Interferón

Secuencias de proteínas de interferones alfa a barajar

1. Consenso	C	D	L	P	Q	T	H	S	L	G	N	R	R	A	L	I	L	L	A	Q	M	G	R	I	S	P	F	S	C	L
	*		*	*							*				*	*	*	*		*		*			*	*	*			
2. alfa I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3. alfa C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4. alfa H	-	N	-	S	-	-	-	-	-	N	-	-	-	T	-	M	-	-	-	-	-	-	R	-	-	-	-	-	-	-
5. alfa 4B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	-	-	-	-
6. alfa 6	-	-	-	-	-	-	-	-	-	H	-	-	T	M	M	-	-	-	-	-	-	-	R	-	-	L	-	-	-	-
7. alfa 7	-	-	-	-	-	-	-	-	-	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8. alfa 8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	R	-	-	-	-	-	-	-
9. alfa D	-	-	-	-	E	-	-	-	-	D	-	-	T	-	M	-	-	-	-	-	-	-	S	-	-	-	S	-	-	-
10. alfa F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11. alfa I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12. alfa WA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	-	-	-	-

1. Consenso	K	D	R	H	D	F	G	F	P	Q	E	E	F	D	G	N	Q	F	Q	K	A	Q	A	I	S	V	L	H	E	M
	*		*				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
2. alfa I	-	-	-	-	-	-	-	L	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	P	-	-	-	-	-	
3. alfa C	-	-	-	P	-	-	-	L	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	
4. alfa H	-	-	-	-	-	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
5. alfa 4B	-	-	-	-	-	-	-	-	E	-	-	-	-	-	H	-	-	-	-	-	-	T	-	-	-	-	-	-	-	
6. alfa 6	-	-	-	-	-	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	E	-	-	-	-	-	V	-	
7. alfa 7	-	-	-	E	-	R	-	E	-	-	-	-	-	-	H	-	-	-	-	-	-	T	-	-	-	-	-	-	-	
8. alfa 8	-	-	-	-	-	E	-	-	-	-	-	-	-	-	D	K	-	-	-	-	-	-	-	-	-	-	-	-	-	
9. alfa D	M	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	-	-	-	-	-	L	-	
10. alfa F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
11. alfa I	-	-	-	P	-	-	-	L	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	
12. alfa WA	-	-	-	Y	-	-	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	F	-	-	

1. Consenso	I	Q	Q	T	F	N	L	F	S	T	K	D	S	S	A	A	W	E	Q	S	L	L	E	K	F	S	T	E	L	Y
	*		*						*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
2. alfa I	-	-	-	-	-	-	-	-	-	-	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3. alfa C	-	-	-	-	-	-	-	-	-	-	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4. alfa H	M	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	I	-	F	-
5. alfa 4B	-	-	-	-	-	-	-	-	-	-	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6. alfa 6	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7. alfa 7	-	-	-	-	-	-	-	-	-	-	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8. alfa 8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9. alfa D	-	-	-	I	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10. alfa F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11. alfa I	-	-	-	-	-	-	-	-	-	-	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12. alfa WA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

1. Consenso	Q	Q	L	N	D	L	E	A	C	V	I	Q	E	V	G	V	E	E	T	P	L	M	N	E	D	S	I	L	A	V
	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
2. alfa I	-	-	-	-	N	-	-	-	-	-	-	Q	-	-	-	-	M	-	-	-	-	-	-	-	-	-	-	-	-	-
3. alfa C	-	-	-	-	N	-	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-	-	-	-	-	-	-	-	-	-
4. alfa H	-	-	M	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5. alfa 4B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-
6. alfa 6	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-	W	-	G	G	-	-	-	-	-	-	-	-	-	-	-
7. alfa 7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	F	-	-
8. alfa 8	-	-	-	-	-	S	-	-	-	M	-	-	-	-	-	-	-	I	-	S	-	-	-	-	-	-	Y	-	-	-
9. alfa D	-	-	-	-	-	-	-	-	-	M	-	-	-	-	-	E	R	-	G	-	-	-	-	-	-	-	-	A	-	-
10. alfa F	-	-	-	-	M	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-
11. alfa I	-	-	-	-	N	-	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-	-	-	-	-	-	-	-	-	-
12. alfa WA	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	I	A	-	-	-	-	-	-	-	-	-

Figura 2 Página 1 de 7

1. Consenso	S	L	S	F	S	T	N	L	Q	K	R	L	R	K	D
	*		*		*	*	*	*	*	*		*	*		
2. alfa I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3. alfa C	-	-	-	-	-	-	-	-	-	I	-	-	-	-	-
4. alfa H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5. alfa 4B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6. alfa 6	-	F	-	S	-	R	-	-	-	E	-	-	-	-	E
7. alfa 7	-	F	-	-	-	-	-	-	K	-	G	-	-	-	-
8. alfa 8	-	F	-	L	-	I	-	-	-	-	-	K	S	-	E
9. alfa D	-	-	-	L	-	-	-	-	-	E	-	-	-	-	E
10. alfa F	-	F	-	L	-	K	I	F	-	E	-	-	-	-	E
11. alfa I	-	-	-	-	-	-	-	-	-	I	-	-	-	-	-
12. alfa WA	-	F	-	-	-	-	-	-	-	G	-	-	-	-	-

Secuencias de ADN de interferones alfa a barajar

1. Consenso	A A G G A C A G A C A T G A C T T T G G A T T T C C C C A G																				←	
	100										110					120						
2. alfa I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-
3. alfa C	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	C	-	-	-	-	-
4. alfa H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-
5. alfa 4B	-	-	-	-	-	-	-	-	-	-	T	-	-	C	-	-	-	C	-	-	G	-
6. alfa 6	-	-	-	-	-	-	-	-	-	-	-	-	-	C	A	-	-	-	-	-	-	-
7. alfa 7	-	-	-	-	-	-	-	-	-	-	A	-	-	C	A	-	-	-	C	-	A	G
8. alfa 8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-
9. alfa D	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10. alfa F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-
11. alfa I	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	C	-	-	-	-
12. alfa WA	-	-	-	-	-	-	-	T	-	-	-	T	-	C	-	-	-	C	-	-	-	-

1. Consenso	G A G G A G T T T G A T G G C A A C C A G T T C C A G A A G																													
	130										140										150									
2. alfa I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3. alfa C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4. alfa H	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A
5. alfa 4B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6. alfa 6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7. alfa 7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8. alfa 8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	T	-	A	-	-	-	-	-	-	-	-	-	-	-	-
9. alfa D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10. alfa F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11. alfa I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12. alfa WA	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

1. Consenso	G C T C A A G C C A T C T C T G T C C T C C A T G A G A T G																													
	160										170										180									
2. alfa I	A	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3. alfa C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4. alfa H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5. alfa 4B	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6. alfa 6	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-
7. alfa 7	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8. alfa 8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9. alfa D	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-
10. alfa F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11. alfa I	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12. alfa WA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-

1. Consenso	A T C C A G C A G A C C T T C A A T C T C T T C A G C A C A																													
	190										200										210									
2. alfa I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3. alfa C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4. alfa H	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5. alfa 4B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6. alfa 6	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7. alfa 7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8. alfa 8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-
9. alfa D	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	T	C
10. alfa F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11. alfa I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12. alfa WA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

1. Consenso	A A G G A C T C A T C T G C T G C T T G G G A T G A G A G C																													
	220										230										240									
2. alfa I	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3. alfa C	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4. alfa H	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5. alfa 4B	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6. alfa 6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	G
7. alfa 7	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8. alfa 8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9. alfa D	-	-	A	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G
10. alfa F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11. alfa I	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12. alfa WA	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C

[illegible]

Figura 2 Página 5 de 7

Figura 2 Página 6 de 7

1. Consenso	A	G	A	T	T	A	A	G	G	A	G	G	A	A	G	G	A	T	T	G	A
2. alfa I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3. alfa C	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4. alfa H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5. alfa 4B	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	A	-
6. alfa 6	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7. alfa 7	-	-	-	-	G	-	A	-	-	T	-	-	-	-	-	-	-	A	-	-	-
8. alfa 8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	A	-
9. alfa D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-
10. alfa F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11. alfa I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12. alfa WA	G	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figura 3